

UNIVERSITY OF LATVIA
FACULTY OF HUMANITIES
DEPARTMENT OF ENGLISH STUDIES

**USE OF CORPORA IN TEACHING ENGLISH
ACADEMIC WRITING**

**KORPUSU IZMANTOŠANA ANĢĻU AKADĒMISKĀS
RAKSTIŠANAS MĀCĪŠANĀ**

MASTER THESIS

Author: **Ieva Kraukle**
Matriculation Card No. **ik10176**
Adviser: doc. Zigrīda Vinčela

RIGA 2016

ANOTĀCIJA

Uz tekstu korpusu lietojuma balstītas valodu studijas ir atzītas par efektīvu metodi dažādu valodas aspektu mācīšanā, sākot ar korpusu analīzi mācību materiālu sagatavei līdz aktīvai korpusu izpētei mācību procesā. Viens no galvenajiem terciārā līmeņa akadēmiskās rakstīšanas apguves uzdevumiem ir leksiskās kohēzijas aktualizācija, tādēļ pētījuma darba mērķis ir veidot un veikt korpusu balstītas darbības, kas veicina studentu un mācībspēku izpratni par leksiskās kohēzijas nozīmi rindkopu rakstīšanā, kā arī veidot studentu korpusu pēc šo darbību pārbaudes. Gadījuma izpētes metodes ietvaros pielietotā kvalitatīvā, kvantitatīvā un korpusa analīze ļāvusi atzīt, ka korpusu balstītas darbības sekmē leksiskās kohēzijas kā svarīgas akadēmiskās rakstīšanas sastāvdaļas, īpaši rindkopu rakstīšanas, apzināšanos.

Atslēgvārdi: akadēmiskā rakstīšana, korpusu balstīta mācīšana, rindkopa, leksiskā kohēzija, British National Corpus, Wmatrix, studentu korpus.

ABSTRACT

Corpus-based language studies recently have been recognized as an efficient tool for a variety of linguistic aspects to be taught, ranging from corpora consultation for material development to approaches demanding active classroom corpus examination. One of the major objectives of tertiary level academic writing instruction being actualization of lexical cohesion, the goal of the current thesis is to introduce learner-centred and teacher-centred corpus-related activities for improving the aspect of lexical cohesion in students' writing, and design a learner corpus of paragraphs after piloting these activities. The performance of a case study within which qualitative and quantitative approaches have alternated with corpus analysis has permitted acknowledgment of corpus application as beneficial for promoting awareness of lexical cohesion as a vital component of academic writing, namely paragraph writing.

Key words: academic writing, corpus-based teaching, paragraph, lexical cohesion, British National Corpus, Wmatrix, learner corpus.

CONTENTS

Introduction	1
1 Application Of Corpora For Language Studies	5
1.1. Corpus linguistics and language studies	5
1.2. Methodologies of corpus implementation in language studies	7
2 Application Of Corpora For Teaching Writing Skill.....	13
2.1. Place of corpora in academic writing instruction	13
2.1.1 Background studies of corpus-based academic writing instruction.....	16
2.2. Stages and characteristics of academic writing	17
2.3. Structure and characteristics of paragraph writing.....	20
2.4. Lexical cohesion	22
3 Corpus-Based Teaching And Study Of Students' Written Texts.....	26
3.1. Research methods and data collection instruments	27
3.2. The research subjects and their text corpus	29
3.3. Research procedure and results	31
Conclusions	63
Theses	66
References	67
Appendix 1 Lesson plan 1: Recognition of grammar and lexis among registers	72
Appendix 2 Theory on grammar and lexis among registers.....	73
Appendix 3 Recognition tasks on grammar and lexis among registers.....	77
Appendix 4 Lesson plan 2: Comparison/contrast paragraph.....	79
Appendix 5 Tasks for students	82
Appendix 6 PowerPoint presentation on comparison/contrast paragraph.....	85
Appendix 7 Manual for operation in the British National Corpus	91
Appendix 8 UCREL Semantic Analysis System (USAS).....	94
Appendix 9 Distribution of semantic domains in students' paragraphs based on Wmatrix analysis	102
Appendix 10 Learner Corpus	104

INTRODUCTION

With the change of time, new teaching and learning approaches have become an inevitable part of educational system. Language teaching is not an exception. The teaching and learning approaches considered effective in the past might only partly meet the expectations of the contemporary educational practice. The need for innovations demands the linguistic instructors search for novel approaches to language teaching. Starting from the end of the previous century, one of such approaches is the implementation of the informational and methodological resources the computer provides. It has proved to be an effective tool facilitating various aspects of language teaching and learning. Computer implementation, as claimed by McEnery et al. (2006: 6), has brought changes in the field of language teaching and learning methodology.

One of the most appreciated methods of computer involvement in linguistic studies is related to text corpora. In recent decades, as uttered in theories by Aijmer and Altenberg (1991), Sinclair (1991), McEnery et al. (2006), the practice of corpus involvement in teaching and learning has become a topical issue as it reflects the system and variation of a language. Consequently, corpus linguistics has turned into a language acquisition theory (McEnery et al., 2006: 7). Tognini-Bonelli (2001) and McEnery et al. (2006) conclude that corpus linguistics is not framed within a narrow sense of static body of texts but is rather a tool for supporting the theoretical assumptions about the language, as well as for encouraging the understanding of the language through active examination of the evidence.

A range of scholars (Tribble, 2002; McEnery et al., 2006; Yoon, 2008; Römmer, 2008; Quinn, 2015) have described the practical investigations of the application of corpora in academic writing instruction. Their studies have ranged from the examination of students' habits of corpus assistance (Yoon, 2008) till the analysis of the concordance lines for gradually improving their written performances from the point of view of grammar and lexis (Tribble, 2002; Quinn, 2015). As a rule, the results have indicated a positive impact on the linguistic outcome. However, one of the main issues concerning the academic writing is the unconformity in the goal of the text and its linguistic manifestation. To overcome the gap, the information from text corpora can be manipulated with according to the needs. For example, the analysis of concordance lines is said to be beneficial to improve the text from the perspective of collocation, colligation, semantic prosody and semantic preference (McEnery et al., 2006: 82-85). These are the four aspects mostly responsible for the text cohesion in its grammatical and lexical manifestation, capable of transforming the linguistic outcome based

on intuition into the academic knowledge. As a rule, the results of the experiments appeared to be beneficial for both language teachers and learners. As claimed by Thurston and Candlin (1998), the application of concordancing programmes opens new opportunities for language learners to notice the linguistic regularities. For example, lexical cohesion is strongly concerned with the aspect of regularities.

Though, the studies reflecting about the role of corpus - related academic writing instruction is not a rare phenomenon among the linguistic community (Quinn, 2015; Park and Lu, 2015), so far, little attention appears to have been paid to the aspect of lexical cohesion within the scope of paragraph writing. Also, at an institution of higher education in Latvia, Corpus Linguistics happens to be delivered at the Master level from the point of view of its theoretical instruction. In addition, even if corpora are consulted at the preparatory stage of teaching materials for various language disciplines, active corpus usage is not a permanent part of teaching and learning of academic writing. Thus, the research paper intends to actualize the use of corpus-based activities for lexical cohesion instruction in academic writing context at a bachelor level. The focus is to apply corpora for the instruction and study of lexical cohesion within the boundaries of paragraph writing among the tertiary level English language students.

The **goal** of the present thesis is to introduce learner-centred (corpus-based) and teacher-centred (research of learner corpus) corpus-related activity design facilitating the composition of lexically cohesive paragraphs and compile a learner corpus.

The author has proposed the following **hypothesis**: the implementation of learner-centred and teacher-centred corpus-based activities in paragraph writing at tertiary level facilitates the students to develop lexically cohesive paragraphs as an important aspect of academic writing.

The following **enabling objectives** are established:

- analyse the secondary sources dealing with the corpus application in linguistic inquiry, academic writing and lexical cohesion in particular;
- analyse the sources reflecting about the previous investigations of incorporating corpus in academic writing instruction;
- collect the students' first written assignments composed in the framework of the *English Written and Spoken Communication* course to have an impression about the level of text cohesion;

- provide theoretical background of grammatical and lexical characteristics of formal and informal register as a basis for lexical cohesion followed by corpus-based recognition task;
- instruct the students about the comparison/contrast paragraph functions and structure followed by them composing a comparison or contrast paragraph;
- introduce the students with the corpus and its application for the acquisition of academic writing, in particular, types of lexical cohesion;
- provide the feedback on comparison or contrast paragraph with the focus on lexical cohesion and encourage the application of corpus (concordancing and synonymy) to improve the paragraph;
- perform analysis of the students' first written assignments in **Wmatrix** corpus analysis and comparison software for key word determination, followed by analysis of their collocates in **AntConc** concordancer;
- perform semantic analysis of the students' comparison and contrast paragraphs on **Wmatrix** corpus analysis and comparison software;
- create an unannotated **learner corpus**;
- Draw relevant conclusions and suggest future perspectives.

The theoretical research is based on the corpus-related language teaching and learning theories by Aijmer and Altenberg (1991), McEnery et al. (2006), Yoon (2008), O`Keeffe and McCarthy (2010), Quinn (2015), and other researchers, who have described the characteristics, advantages, and procedures – with the focus of passive and active (concordancing) corpora - of incorporating corpora as a language acquisition tool. Also, the features of academic writing in terms of their grammatical, lexical and compositional features (Biber et al., 1999; Tribble, 2002; Friedrich, 2008; Etherington, 2008; Cheng, 2012; Day, 2013) have been analysed, positioning academic paragraph writing as the focal point (Fiderer, 2002; Howe and Willoughby, 2005). Finally, the theoretical background deals with the aspect of lexical cohesion (Halliday and Hasan, 1976; Flowerdew and Mahlberg, 2009) and corpora as a tool for its exemplification (McEnery et al.: 2006).

Case study and **corpus analysis** are the **research methods** of the present thesis. Case study implies the pilot teaching conducted among the first year English Philology students, Group1 and Group2. Corpus analysis is manifested in five directions: (1) **British National Corpus (BNC)** has been consulted passively for the compilation of the teaching materials (2) The mentioned corpus has been applied actively in the framework of the analysis of

concordance lines and synonymy (3) **Wmatrix** corpus analysis programme has been applied for key word determination in students' first written assignments (4) **AntConc** concordancer has been used for the study of their collocates (5) The students' paragraphs have been analysed with the help of **Wmatrix** corpus analysis and comparison tool, on the basis of which the learner corpus has been designed, as well.

Data collection tool is qualitative in its nature. The students' written texts are examined in terms of their linguistic features. Qualitative approach enables the examination of the students' progress (if any) in shaping a written text's lexical cohesion. Yet, quantitative perspective has been occasionally applied regarding the calculation of the effectiveness of a range of activities piloted in the scope of the present study, where the correlation between corpus-based activities and number of students having managed these activities is the basis for further analysis.

The **subjects** of the thesis are the tertiary level first year English Philology students at the institution of higher education in Latvia, groups A1 and B1. They were exposed to the corpus-related academic paragraph writing within the framework of the *English Spoken and Written Communication 1* course. The two groups have undergone 3 piloting seminars each, ranging from the introduction of features of informal and formal register and ending up with the composition of paragraph and its improvement with the application of concordancing technique.

The first chapter entitled 'Application of Corpora for Language Studies' outlines the relation between Corpus Linguistics and text corpora and their role in language studies ranging from the historical emergence to methodological advantages. The second chapter 'Application of Corpora for Teaching Writing Skill' focuses on the characteristics of academic writing, paragraph writing and corpus-related methodologies for their attainment. It introduces lexical cohesion as an integral part of academic writing proficiency. The third chapter 'Corpus-Based Teaching and Study of Students' Written Texts' introduces the methodology and the results of the present study, followed by their discussion and conclusions.

1. APPLICATION OF CORPORA FOR LANGUAGE STUDIES

The following chapter outlines the theories regarding the history of the emergence of corpus linguistics - putting greater emphasis on text corpora - as a means for linguistic studies at the educational institutions. The chapter consists of two sub-chapters. The first entitled 'Corpus linguistics and language studies' reflects on the emergence of corpora as an educational tool, briefly touching upon its advantages. The second sub-chapter 'Methodologies of corpus implementation in language studies' is devoted to the description of theoretical and empirical models for corpus-related language teaching and learning, stressing the advantages of each pattern in respect to the teacher-student involvement and language proficiency development.

1.1. Corpus linguistics and language studies

Nowadays, when daily operations are highly dependent on the use of technologies, education is one of the spheres eager to use that potential. It is the interest of language teachers to present the subject matter and the corresponding materials in a way engaging and comprehensive for the students, where novelty, high level of enlightenment and appeal are of the primary concern. Also, it appears that nowadays students of any discipline, including linguistics, are keen on consulting technologies as an educational tool. In this respect, Sanders (2015) has made a note to a curious phenomenon: it is not only the content the students are eager to perceive, but it is also the fact of manipulating with technologies itself that is appealing for the learners (ibid.; Online 4).

Expanding the idea about technology use in education, Aijmer and Altenberg (1991: 1) and other scholars have stressed the importance of corpus linguistics and corpora development and the perspectives they provide for language teaching and learning. Their field of interest is concerned with the study of how language studies could be made more effective (ibid.). By definition proposed by McEnery and Hardie, 'Corpus Linguistics is the study of language data on a large scale – the computer-aided analysis of very extensive collections of transcribed utterances or written texts' (2012: xiii). While corpus linguistics is a term covering the whole branch of computer-assisted language research and education, text corpora is of the primary interest in the scope of language teaching and learning. It is the unit manipulated with most in various stages of language instruction. According to Cheng, corpus is 'a collection of texts that has been compiled for a particular reason.[...] A corpus, then, is a collection of texts based on a set of design criteria, one of which is that the corpus aims to be representative' (2012: 3). Consequently, McEnery et al. (2006: 7) conclude that corpus

linguistics covers the whole set of methods applicable for the corpora implementation in language studies.

Drawing parallels between corpus and language education is a relatively new practice, though inconsistency emerges among scholars interested in the subject field. According to Flowerdew (2010: 444), corpus as an educational tool for language instruction emerged in the middle of 1960s. ‘Data-driven language learning (DLL)’ (Baker, 2009: 206; Geist and Hahn, 2012: 125) was introduced by Tim Johns in 1991 (Geist and Hahn, 2012: 125). Following Talai and Fotovatnia, data-driven language learning is ‘a student-centered technique in which samples of authentic language are taken from linguistic corpora and be presented to language learners for improving language proficiency [...]’ (2012: 1526).

Continuing with the positive traits of DLL, it is possible to distinguish between a range of advantages supporting the corpus-related language teaching and learning process. As Talai and Fotovatnia (2012) claim, the language students are encouraged to observe the linguistic regularities and draw relevant rules and conclusions. The DLL acknowledges the student of a language as a central figure responsible for knowledge attainment. Consequently, the main points positioning data-driven language learning as a successful language acquisition methodology is its ability to promote students’ autonomy and critical thinking. Thus, they become language researchers at early stages of language learning at whatever level the particular teaching and learning methodology is practiced (ibid.). However, Johns (in Talai and Fotovatnia, 2012: 1526) and Geist and Hahn (2012:124) are in favour of the utilization of data-driven learning among advanced language learners due to their relatively high level of intelligence and motivation. Still, they conclude that age and language proficiency are not the only factors responsible for the effectiveness of data-driven language learning. The linguistic background of the students and the environment in which the technique is applied can have an impact on the learning outcome. For second or foreign language learners, the DLL is believed to indicate a higher risk of effectiveness fails as for the native language users (ibid.).

Sylviane Granger et al. (2002: 4) supported the implementation of corpus analysis as a powerful learning approach as one of its aspects – the frequency with which a linguistic feature is observed in authentic texts – asserts the validity of the language peculiarity a student has observed in a definite amount of sample texts. Talai and Fotovatnia (2012) conclude that the language teachers and instructors appreciating the advantages of the data-driven language learning claim this method to foster the acquisition of grammatical and lexical features of a language.

As Aijmer and Altenberg (1991: 2) and McCarten (2010: 423) have indicated, one of the major benefits of corpora use for language instruction is the authenticity and naturalness of the texts, providing an opportunity of a thorough analysis of ‘English lexis, syntax, discourse and prosody’ (Aijmer and Altenberg, 1991: 2) while being engaged in linguistic inquiry process. Still, the researchers have added that the aspect of naturally occurring discourses should not be regarded as a tool guaranteeing a complete language understanding and high-level proficiency. The tertiary level language students are focused on the attainment of native-like language proficiency. However, it has to be remembered that originally the target audience of the texts found in corpora is not a foreign language learner. Consequently, as summarized by Cheng (2010: 319), a student is incapable of participating in a discourse. Instead, a language learner becomes a ‘discourse observer’ (ibid.: 319). The scholar continues that the corpus material is analysed critically and the perception of language regularities is the primary goal of a learner. The ability to apply authentic language patterns in the learning process encourages the accumulation of the necessary knowledge. However, it has to be taken into account that computer being a technology is not capable of providing language learners with the linguistic product that completely corresponds to the notion of knowledge. In technical terms, Aijmer and Altenberg (1991: 14) summarize that corpora is a tool for sorting and arranging the textual evidence, but it is only in the power of a human being to analyse the data, see linguistic models and generate corresponding knowledge.

In the framework of learner corpus, the concept of text authenticity is argued by Sinclair (2004: 127). The scholar has classified texts into four categories: (1) fully natural (2) product of teaching process (3) controlled tasks (4) scripted (ibid.: 127).

To conclude with the concept of data-driven language learning, Talai and Fotovatnia (2012) have attributed the following key terms describing it: (1) ‘linguistic corpus (2) concordancing (3) induction (4) transmission from product to process (5) computer programs and linguistic corpus application for learning activities’ (2012: 1527). These characteristics introduce the methodologies described in the following sub-chapter.

1.2. Methodologies of corpus implementation in language studies

According to Aijmer and Altenberg (2009), corpus and language teaching are interrelated in three distinct aspects.

Aijmer (2009: 9, 47) and Frankenberg-Garcia (2012: 37) start with the use of language corpora in classroom. The general purpose is to introduce the students with the corpora as

such, providing instructions of its application. Such an approach is considered as an active corpus use (Chambers, 2010: 345) or ‘direct use of corpora’ (Aijmer, 2009: 45). Yet it appears to be a relatively seldom applied from the perspective of material developers (teachers), teachers (as active instructors), and pupils (Aijmer, 2009: 9, 47; Frankenberg-Garcia, 2012: 37). Geist and Hahn (2012) affirm a lack of information documenting the results of active corpora use in a classroom, supporting such a deficiency by a claim that ‘[...] the teacher/researcher in confronted with the challenge of reducing the academic complexity of corpora in order to enable learners to work with corpora effectively’ (ibid.: 124).

Yet, there are several attempts in favour of active application of corpora in a classroom, resulting in a positive linguistic-pedagogical outcome. An experiment by Johns (1991, discussed in Frankenberg-Garcia, 2012: 39) is a success as the students are claimed to have “investigated” the difference between synonyms *convince* and *persuade* (ibid.). Also, Krieger (2003, discussed in Frankenberg-Garcia, 2012: 39) has touched upon the direct corpora use for studying various grammatical aspects. Tribble (2002: 131-149) has studied the potential of corpora application in academic writing instruction, emphasizing the contextual and linguistic facets the students are capable of extracting from the text corpus and are expected to apply in their English for Academic Purposes (EAP) study process.

From Baker’s (2009) perspective, active application of corpora promotes language acquisition most successfully, as the students are subjected to cognitive language self-experience, i.e. the regularities of the language are discovered through the actual analysis of the texts. As claimed by McEnery et al. (2006: 99), the so called ‘three Ps’ (2006: 99) formula, respectively ‘presentation – practice - production’ (ibid.) is not applicable for corpus-based language teaching and learning due to the lack of the possibility to explore the language. The researcher happens to be in favour of the so called ‘Three Is’ (ibid.) theory, standing for ‘illustration – interaction - induction’ (McEnery et al., 2006: 99). The term “illustration” underlines the analysis of texts that are authentic and not modified. Next, “interaction” implies the discussion of the findings with the teacher and other students. The scholar concludes with “induction”, which stands for the learner-centred, bottom-up inductive knowledge processing. Johns (1991: 3) has described the same model using different terms. “Illustration” has been substituted by the observation of the concordance lines; “interaction” has been changed to classification of the language features noticed; generalization has been used as a synonym for “induction”.

Aijmer (2009: 12) and McCarten (2010) argue that, in the context of language classes, the corpus analysis mostly limits itself within the boundaries of the investigations of

concordance lines. Though being one of the most widely accepted and applied technique (i.e. concordancing), there are different approaches to corpus as a language teaching and learning aid capable of facilitating students' proficiency in various language areas and skills. Consequently, Cheng's (2012: 3) definition of corpus mentioned above supports corpora as a language teaching and learning method applicable for the search of specific linguistic features necessary for the compilation of a particular task. For instance, the analysis of the concordance lines provides the information about the grammatical and sense patterns of lexical units under investigation, and alike (McCarten, 2010: 423). As claimed by Cheng (2010: 326), the study of concordance lines fosters the student's awareness of different language investigation techniques, as well as improves the effectiveness and motivation towards language learning and linguistic outcome.

Cheng (2010: 326) continues by offering two approaches towards concordance line analysis. The first consists of four steps. The first step is to determine the research question. Secondly, it is the choice of the most appropriate research approach. Thirdly, a language learner is exposed to a great amount of samples from which relevant information for the research question is to be selected and critically analysed (ibid.). The scholar distinguishes it as the most vital step from all as this stage demands a sufficient ability to notice the regularities of the linguistic system. At the very first attempts of corpus analysis the teacher plays a crucial role in helping to recognize the language models, as well as in providing technical instructions of corpus analysis. Finally, the last step involves the formation of relevant conclusions, as Cheng (2010: 326) mentions. Regarding concordancing, Biber and Finegan (discussed in Aijmer and Altenberg, 1991: 209) have praised the manipulation with the concordance lines for the grammatical analysis as the language examiners are exposed to sentences, not the whole text, thus assuring a higher level of concentration on the issue under interest.

The second concordance line analysis technique, also exemplified by Cheng (2010: 319-332), is structured on the basis of bottom-up and top-down information processing models, each being realized through different applications to language study. The top-down approach has been applied in order to pay students' attention to the actual language use of the native speakers. The examination of corpus data supports the knowledge of the language use habits, thus making the students aware of its models and structures (ibid.). The bottom-up approach is based on the reciprocal language study technique –, the students are exposed to the learner corpus (Cheng, 2010: 322). The linguistic performance is developed by gradually improving students' own performances. The focus of the bottom-up approach is the students'

ability to recognize their own errors and accomplish the necessary activities for generating knowledge appropriate for error elimination. A study carried out by Chamber and O'Sullivan (discussed in Cheng, 2010: 326) illustrates the procedure of self-editing. Non-native English speakers were asked to compose a written text in English. Then, they were provided with the instructions for concordance line analysis in a corpus consisting of similar texts. After that, the students were invited to improve their written performances on the basis of the information and knowledge they have accumulated during corpus analysis. The essence of the bottom-up model has been realised through the students' ability to recognize and correct errors in terms of grammar, lexis, and spelling (ibid.).

Apart from concordance line analysis, language corpora available online are capable of extracting different information on the basis of students' needs and curriculum. For instance, the annotated British National Corpus (BNC) hosted by the Brigham Young University, one of the major corpora applied, provides frequency statistics, distribution peculiarities, and linguistic habits of different parts of speech, as well as other information (Online 1). So, the manner of working with text collections can be adjusted to the situational context.

As summarized by Aijmer and Altenberg (1991: 3), the basic advantages of the application of corpora in language classes, i.e., active or direct approach of corpora are: (1) it is multi-dimensional as it promotes the language description (2) it helps to formulate the models of language structure (3) it is practically usable. Though corpus linguists mention the application of data-driven language learning tools as beneficial for language studies, the lack of knowledge and motivation from the teachers and pupils still positions the active use of corpora as a relatively unfamiliar language teaching and learning technique (Cheng, 2010).

Another aspect of how text corpora and language studies are interrelated, referring to Sinclair (2004: 126) and Aijmer (2009: 9) is the passive or indirect use of corpora (also referred to as COBUILD or Birmingham approach), when a vast compilation of texts is used as a reference for material development. An advantage of passive corpora is the ability to modify and adjust the language material to the needs and proficiency level of the students. For instance, the materials can be modified so as they focus on one particular language issue the students are expected to study. Frankenberg-Garcia (2012: 40-41) offers the possible means of indirect use of language corpora. Claiming such an approach being more teacher-centred, the scholar reflects on a case study when Portuguese students of tourism have been exposed to the English language learning specific to the said field. After getting familiar with the problem areas of the students, the further teaching materials were based on the corpus data, in particular – the extraction of separate concordance lines have been implemented into the

handouts. Another point mentioned by the researcher is the different stages – ‘before, during, after’ (ibid.: 43) – when the corpus-based reference is suitable. Consequently, the indirect application of corpora is applicable in various stages of language studies when such an approach appears to be suitable and supportive. Frankenberg-Garcia gives a special prominence to the “after” stage, when a teacher is encouraged to consult corpora to provide a sufficient feedback on student’s performance.

Another way of involving corpora within language studies is the development of a learner corpus, which Sinclair defines as ‘systematic computerized collections of texts produced by learners’ (2004: 127). Following Sinclair, it is developed in order to study the pupils’ language proficiency development and the problem areas. For illustration, a quantitative and qualitative corpus-based study by Mattheoudakis and Hatzitheodorou (2012) aimed at examining the collocational patterning of specific lexical units in language learners’ written texts resulted in boosting the awareness of the students’ language use habits, which, if not corresponding to the set language norms, were a subject for improvement. Similar comparative analysis of learner corpora (Götz, 2012; Xu et al., 2012; Rogatcheva, 2012; Twardo, 2012) have been performed with the purpose of examining the level at which the language students’ knowledge and ability of practical application of different aspects of language correspond to the authentic use of language and its rules. Consequently, both the researches emphasize that a learner corpus is capable of storing and pointing out language command development of individuals and groups, as well as functions as a basis for a better understanding of the language issues students and teacher are to pay more attention to.

The three approaches towards corpora in language studies described above show that they are capable of functioning as the basis for pedagogical process. Not necessarily rejecting the “corpora-course book” relationship in a classroom, various combinations of corpora use can function also a successful methodology for solving language issues. Sinclair (2004: 126) has systematized the mutual dependence of various approaches towards corpora and their application for pedagogical purposes:

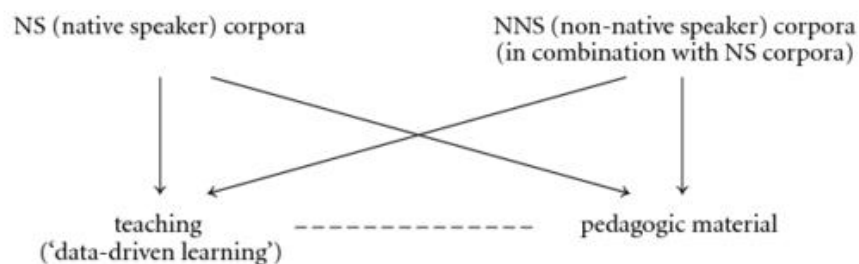


Figure 1.1 Relationship between corpora and language studies (Adapted from Sinclair, J. McH. (2004) *How to use Corpora in Language Teaching*, pp 126).

The Figure 1.1 above illustrates the division between the direct and indirect corpus application in language teaching. As Sinclair (2004: 126) suggests, both native and non-native speaker corpora can be incorporated within a language classroom, still with a difference in the purpose of doing so. As the scholar has revealed, the native-speaker corpora are applied: directly as a research material for students (foreign language learners) to analyse; indirectly for a teacher to develop study materials. Similarly, the non-native corpora are approached in alike manner, yet they are based on the non-native language students' composed texts. So, a language learner is simultaneously capable of encountering authentic language and his/her own linguistic outcome. Such a combination allows analysis and comparison of one's own language issues (if any) and the set use of different language aspects as capable to be observed in native speakers' corpora. Sinclair (2004: 127) concludes that it is the responsibility of both the teachers and learner to employ the corpora in a way beneficial for language improvement. The third possible application of corpora – not mentioned in the diagram by Sinclair – is concerned with the understanding of the language acquisition habits of the students (ibid.: 127). However, it has not been accredited by language instructors to a sufficient extent.

The present chapter has provided an insight into the development of corpus linguistics from an independent language discipline to the language teaching and learning fostering tool. The methodologies of corpus application have been introduced, ranging from the learner corpora and passive or indirect use of corpora for the materials' development till the active/direct analysis of concordance lines, revealing the authentic language peculiarities so necessary for the language learners to master. The following chapter and its sub-chapters will be focused on the academic writing instruction.

2. APPLICATION OF CORPORA FOR TEACHING WRITING SKILL

Patricia Friedrich has explained the acquisition of writing skill in a very simple, but precise phrase: ‘Writing is hard work. Teaching writing is even harder work’ (2008: 1). The sub-chapter ‘Place of corpora in academic writing instruction’ describes corpus-related academic writing teaching and learning, followed by the sub-chapter ‘Background studies of corpus-based academic writing instruction’ illustrating the research on corpus-based academic writing studies. The next sub-chapter ‘Stages and characteristics of academic writing’ is occupied with the compositional, grammatical and lexical features of academic writing, paying particular attention to the concept of lexical cohesion, which is described in the sub-chapter under the same title ‘Lexical cohesion’. The sub-chapter ‘Stages and characteristics of paragraph writing’ concentrates on the structure and language of the said type of the text.

2.1. Place of corpora in academic writing instruction

Several scholars (McEnery et al., 2006: 97; Yoon, 2008: 10), whose field of research has concerned corpora use in writing instruction, have acknowledged that the relatively rare cases of the application of corpora for language teaching are typical of the tertiary level studies (yet it cannot be claimed as an ultimate truth). Thus, the courses implementing corpora are found on universities’ study programmes, particularly in **writing courses**. As explained by Etherington (2008: 29), the main purpose of undergraduate and graduate university writing courses is to guide the students towards a writing proficiency recognized among the academicians.

Even though Biber (2010: 245) claim that the use of corpora for writing skill development in particular increases rapidly, the benefits of corpora remain only partly recognized due to an overall unfamiliarity with applying corpora for studying writing. McEnery et al. (2006: 97) and Yoon (2008: 32) add that it has not been strictly decided whether corpus linguistics should be taught as a distinct subject, or it should be a supplementary element for the acquisition of another language area or skill, for instance, academic writing. In this respect, McEnery et al. (2006: 97) and Flowerdew (2010: 444) have made a distinction between two concepts: ‘Teaching about’ and ‘Teaching to exploit’ (ibid.).

The first concept describes corpus linguistics taught as a distinct subject. However, its generality does not concentrate on the active application of corpora for language studies, which, in contrast, is the purpose of ‘teaching to exploit’ concept, as argued by McEnery et al.

(2006: 97). In ‘teaching to exploit’ (ibid.) the students are familiarized with the techniques of corpus analysis, thus language study is said to become learner-centred (ibid.). It turns out that students are encouraged to adjust their corpus analysis and investigation habits to the linguistic aspects relevant for the current teaching and learning context.

To introduce the basic notion about the possible corpora application in academic writing instruction, it is considered relevant to briefly comment on the assumptions about the focal points to be taught at the academic writing classes. Sian Etherington (2008: 27-36) has provided a well-balanced model of the major aspects to be taken into account when starting to teach academic writing for the first year university students. Firstly, the scholar mentions a necessity to stay focused on teaching English for General Academic Purposes (EAP) (ibid.: 27). The researcher means ‘a set of skills, text types, and language forms for Academic English which can be presented to any student’ (2008: 27). Etherington uses the terminology introduced by Meriel Bloor and Thomas Bloor in 1986 and refers to the phenomenon as ‘common core hypothesis’ (2008: 27). The core consists of the general academic writing skills, types of tasks and language models, ‘academic style, neutrality of tone, use of evidence and argument, [...]’ (ibid.: 28).

Secondly, Etherington (2008: 27-29) stresses an urgent necessity to introduce the above described “core” at the very beginning of the academic studies. Only after reaching a sufficient proficiency in the most fundamental characteristics of academic writing, it is possible to guide a student towards a more complex and subject-specific writing practice.

Thirdly, the acquisition of the peculiarities of the academic writing can affect the relationship between the community, like student-student or student-professor pairs (Etherington, 2008: 30). Thus, when the fundamentals are not learnt and used sufficiently, there is a risk of misunderstandings and decrease in performance. For instance, in educational settings it could be explained as an insufficient mark for a particular writing task. Next, the result of a student’s written performance appears to be directly dependent on the requirements set by the lecturer (ibid.).

The final and most appealing point in the scope of the present study is Etherington’s acknowledgment that research-based teaching is crucial for success, including academic writing studies (2008: 34). The scholar approves the corpora analysis as a means of study process and distinguishes between two approaches to linguistic study also emphasized previously in the present paper: the passive when the teachers do the corpus analysis in order to develop the necessary study aids and materials; the active when the language students themselves are the examiners of the linguistic material (ibid.: 35). In favour of the use of

corpora as such, the scholar mentions the lack of knowledge the first year university students have regarding the written task requirements. Even if they understand the purpose of the written task, they are not aware of the linguistic means by which to demonstrate that purpose. Thus, Etherington (2008: 30, 37) concludes that an active use of corpora data can promote the study of the language patterns and linguistic means applicable for the particular written task. The direct use of corpora is performed at three levels:

1. Examination of the context of the material under consideration.
2. Analysis of the structural and organizational patterns of the text.
3. Focusing on the lexical and grammatical characteristics of the text in order to take notice of the register-specific and task purpose-specific textual features (ibid.: 40).

These points mentioned above stand in hierarchy among each another regarding the level of difficulty students have in acquiring them. For instance, it might be easier to bear in mind that academic writing demands relative stylistic neutrality, while the lexical and grammatical features are causing difficulties to be acquired and used in writing. Consequently, Etherington (2008: 44) claims the corpora analysis to be performed on the grounds of: (1) vocabulary choice (2) collocations (3) sentence and text structure from the perspective of lexical cohesion.

The researcher suggests concordancing as an effective research approach for reflecting the last mentioned - lexical and grammatical patterns - typical of different text types (in terms of the goal and genre of a text). To practice the use of corpora in writing instruction, Etherington (2008) offers the following approach: students should be provided with the concordance lines from which the textual features are observed, analysed, and the knowledge is generated (ibid.). Flower continues that the basic feature characterizing corpora as highly beneficial for writing instruction is its ability to demonstrate the so called 'phraseological patterning' (2010: 444), which covers the study of collocations, synonymic preferences, prosodies and colligations. Flower states that 'Corpora are ideally suited to helping learners master such patterning in writing classes, as these phraseological features tend not to be easily accessible in either dictionaries or grammars' (2010: 444).

Apart from praising the active use of corpora, i.e., concordancing, Römer (2012: 71) has performed a large-scale, multi-dimensional study based on the use of learner corpora for teaching academic writing. The scholar stresses the potential of learner corpus in English academic writing instruction context, after collecting 829 written texts of various disciplines composed by undergraduate and graduate students of university in America. From the language instructors' perspective, the access to such a considerable amount of authentic

written texts is useful for compiling teaching materials, as the basic problem areas can be observed from the data. Also, the language students are eager to solve any issues by accessing the samples of academic texts and affirming them as appropriate models or samples for their own writing proficiency improvement. Without doubt, corpus-based academic writing instruction is appropriate for raising students' and teachers' awareness of the academic writing. The following sub-chapter exemplifies several successful corpus-based studies on academic writing instruction.

2.1.1. Background studies of corpus-based academic writing instruction

Tribble (2002: 131-149) has performed a study on how to incorporate corpus in academic writing teaching and learning, giving prominence to the necessity of actualizing both the contextual and linguistic aspects of a particular genre of academic writing. His research has been focused on the language students going through a set of context-based and language use-based questions that can be answered by approaching the corpus evidence. Once the context of the text is set, it is the grammar and lexis to be analysed with the help of corpus (ibid.: 136). Tribble continues by applying direct use of corpora – concordancing - and determining the frequency of the main keywords of the text (differentiating among various word classes), as well as their grammatical and lexical features. Simultaneously, it demonstrates the lexical items as belonging to definite semantic fields and forming collocations (ibid.).

An investigation carried out by Yoon (2008: 32) affirms the concerns regarding the poor distribution of corpora implementation in linguistic studies. He has discovered that the researchers interested in the field of corpora and writing proficiency tend to focus on 'genre-based text analysis and materials development, relatively few studies have examined students' writing experiences in association with corpus use' (Yoon, 2008: 32). Yoon reports on the investigations that only reflect on the students' attitude towards the implementation of corpora in writing classes and the influence an inductive corpora analysis has on the linguistic outcome, while there is a lack of actual study on text evidence reporting on the effect of active corpora application (ibid.). His research stands in agreement with the suggestion made by Friedrich (2008: 1) that writing instruction has to be systematic, concentrating on the whole process of text composition.

The perspective defending Friedrich's view has been implied in the research conducted by Cynthia Quinn (2015). A corpus has been implied as a reference source for writing improvement. The researcher has conducted her study relying on the procedure following.

Initially, the students majoring in English have been introduced with the corpus from theoretical and practical perspectives (ibid.: 167). Then, the students have been exposed to the practice of concordancing, simultaneously correcting their essays written before encountering corpus (ibid.). The students' essays were approached from the content and language errors (ibid.: 168). To emphasize the information drawn from the corpus, both inferior and corrected versions of the text had to be indicated (ibid.). However, the whole process of essay writing and editing consisted of eight stages: (1) writing an essay (2) providing feedback (3) improving the essay by consulting the corpus (4) indicating the items corrected (5) having a discussion with the teacher and reflecting on the findings, habits of corpus use, information found (6) improving the essay by consulting the corpus for the second time (7) submitting the final essay (8) grading the essay (ibid.: 168). Quinn (2015) has been capable of reporting on positive impact on the overall quality of the texts – the application of corpus, as well as the teacher's feedback have led to a final text sufficient in form, lexical and grammatical cohesion (ibid.).

Consequently, there is evidence for the positive effect the examination of naturally occurring texts has on the written text composition. The results of the studies function as evidence that, though, the initial stages of corpus analysis are hardly possible without additional feedback, corpus analysis fosters the linguistic dependence and critical-analytical skills of language students.

2.2. Stages and characteristics of academic writing

According to Jordan (1997), in comparison to other language skills, the attainment of a sufficient proficiency in academic writing is one of the most complex processes the undergraduate and even graduate students are engaged in. This intricacy manifests itself in two completely different perspectives, still interdependent, as revealed by Day (2013). Writing skill acquisition is a gradual activity which, in academic environment, is practiced by the application of a number of mutually related text-formation stages (ibid.). In other words, the writing is approached as a technical process. Another perspective is that students are encouraged to familiarize with the linguistic characteristics of the academic writing, i.e., the grammar and lexis appropriate for the particular text formality level (ibid.).

As Day (2013: 13-15) has noted, composition of academic text implies several stages. The first stage is planning. At this point, an individual is exposed to a vast preparatory work before actually writing a text: deciding upon the main points to be discussed, researching the

secondary sources relevant for the particular issue, taking notes on the main findings and ideas useful for one's own performance, as well as considering the whole process of writing, for instance, the time necessary for each part to be written (ibid.). The second stage is the composition of the first draft, when the thoughts are arranged in paragraphs. The third step is labelled as 'reviewing' (ibid.: 13). The draft can be assessed by both the students and the professor. The text is being evaluated, analysed, and considered from the point of view of the content relevance, formatting, use of language. However, three distinct sub-stages are observed in this stage: (1) 'developmental editing' (Day, 2013: 19) – rearranging relatively huge amounts of the text in order to arrange them in logical sequence. Most frequently, developmental editing is to take place after the completion of the draft. It assures the coherence of the text; (2) 'copy-editing' (ibid.: 20) – the stage includes the analysis of grammar, lexis, and punctuation, references, and quotations; (3) 'proofreading' –the final look at the text (ibid.). As continued by the scholar, this process of academic writing rarely is linear and static. In most cases, it is 'iterative' (ibid.). The writing process offered by Day (2013) has emphasized the academic writing process as cyclic activity, which is also the focal point in terms of academic writing and corpus implementation within the scope of the present research paper.

Biber et al. (1999) have introduced the basic differences of various communication genres, including academic writing. The author has enumerated the basic features characterizing academic writing, starting with an acknowledgement that it is not interactive and an immediate response is not expected. Next, the primary communicative goal of academic writing is to provide: information, explanation, argument (ibid.: 16). As a result, one of the features establishing the acquisition of academic writing as a sophisticated discipline to master is that the target audience of the texts consists of the specialists in the field. Reminding of the 'global dialect domain' (Biber et al., 1999: 16) of the academic writing, the language applied in academic writing has to follow the standard of its usage. The international target population of the academic texts demands definite language norms to be used in order to enhance the comprehension and prevent misunderstandings. The idea of language norms mentioned previously concerns itself with a set grammatical, stylistic and, most relevant to the present paper, lexical features.

Academic writing taught at tertiary level educational institutions aims at improving students' capacity to form a written text corresponding to formal register from the point of view of grammar and lexis. In this respect, the academic requirements are challenging for undergraduate students just having attained secondary education. Prominent figures in the

field of academic writing are Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan (1999), whose study of grammatical peculiarities of spoken and written English has attained recognition among linguistic society. The added value of their devotion is the analysis of corpus as the basis for the study. The following text will enumerate the most relevant grammatical-lexical characteristics of academic writing. Even though the focus of the present paper is lexical cohesion, Halliday and Hasan (1976) have stated grammar and lexis being in very close relationship.

Biber et al (1999: 421) and Biber (2006: 15) have claimed the major criterion for the academic writing to be neutrality, which is attained by the omission of first person personal pronouns. Instead, the passive voice is applied. In this respect, the corpus evidence has shown the so called 'short passive' (Biber et al., 1999: 477), i.e., the agent of the action is omitted (ibid.). The authors have noted that, in a highly informative text, about sixty per cent of the content words are nouns (Biber, 2006: 15). Consequently, nominalization is a typical feature, where the greatest part of the nouns denotes abstractions. Next, the noun phrases happen to be followed by several post-modifiers rich in complexity and length (Biber et al.: 1999: 640). Conditional, as well as concessional, clauses are valued as highly beneficial for academic environment as they support the argumentation established by the author of the text (ibid.: 825). Proceeding with word classes, the adjectives applied are mostly noun modifiers. In addition, the degree of comparison is more common than the superlative degree. To maintain the passive tone of the discourse, the adjectives often are preceded by the word *be*, as in 'be sure, important, difficult, likely, necessary, possible' (ibid.: 439). Concerning the verbs, their proportion is greatly smaller than that of a noun (ibid.: 396). Depending on the context, the academic writing favours past simple, past participle, or present perfect (ibid.: 396, 480). The present tenses are applied when the purpose is to demonstrate the relevance of the argument at the particular period of time (ibid.: 458). Phrasal verbs are not common in academic writing; instead, prepositional verbs, like 'be used in, depend on, be based on, be associated with' are more common (ibid.: 416). Finally, there are four suffixes that the scholars have investigated to be widely operated with in order to derive new lexical units: *-ity*, *-ise*, *-ify*, *-en* (ibid.: 402), while the leader in the list of prefixes is *re-* (ibid.: 403). Again, to refer to Biber et al., the vocabulary is technical and specific to the subject field represented (1999: 54, 409). A great variety of words are from Latin origin (ibid.: 325).

To summarize, the list of grammatical-lexical peculiarities could be expanded, but the major purpose of the text above was to demonstrate the very basic features differentiating academic writing from other texts. However, the following sub-chapter will be devoted to a

more detailed description of lexis typical of academic writing and the cohesive effect capable of being achieved by applying it.

2.3. Structure and characteristics of paragraph writing

As defined by Howe and Willoughby (2005: 72) a paragraph is a combination of interdependent sentences expressing a common idea. Paragraph, in academic writing context, can be perceived as both an independent piece of writing, as well as an element of a larger academic text, such as an essay (ibid.). However, the basic features characterizing a paragraph do not change because of the amount of the assignment the students are encouraged to fulfil. Thus the present sub-chapter aims at introducing the major functional, compositional, and linguistic features of an academic paragraph.

Fiderer (2002: 16) has classified paragraphs depending on their function. Thus, a distinction between ‘descriptive, narrative, expository, classification, contrast, comparison, process’ paragraphs can be made (ibid.). Howe and Willoughby (2005: 72) have added that an average paragraph consists of five up to ten sentences, which constitutes about seventy five till one hundred and fifty.

The authors have claimed the general principle of a paragraph structure reminding stable regardless of the paragraph’s function. The structural patterning explained by Howe and Willoughby (2005: 73) assures the topic sentence introducing the issue to be dealt with in the text. It has to be very precise in wording not to include the subjects not discussed later on in the text. The topic sentence demands a precision that is able to attract the attention of the reader (ibid.). Apart from revealing the major concern of the paragraph, a topic sentence may also introduce the possible disagreements between the issues discussed (ibid.). The scholars state that a paragraph positioned between different parts of the text has to convey a topic sentence signalling the thought-relation with the comments reflected in the preceding paragraph. Then, it is followed by the body sentences, or the supportive sentences (Fiderer, 2002: 12). Osmond (2015) continues by describing the body part of a paragraph, where a writer explains in greater detail the points relevant to the issue under consideration. Here, the grammatical and lexical features of academic writing discussed previously are of a great importance to enhance the appropriate level of formality and cohesion of a text. Also, in relation to lexis and lexical cohesion, Osmond (2015) has introduced the aspect of the so called signposting. It refers to the linking devices applicable for the smooth transition from one point to another. A variety of the linkers used depends on the function on the paragraph.

Following Howe and Willoughby (2005: 74), the sentences implied in the body can be of two types: (1) explanatory sentences (2) sentences illustrating the problem. The explanatory sentences are the cohesive force within the text, i.e., they link the aspects implied in the paragraph or among several paragraphs (ibid.). As claimed by the scholar, ‘they give theoretical structure to the evaluation of the abstract concept contained within the topic statement’ (ibid.: 74). Mostly, it is achieved by precise, valid examples, critical judgements, statistical evidence, paraphrasing (ibid.). At the end, the concluding paragraph is added, restating the major conclusions or aspects discussed in the body part (Fiderer, 2002: 12, Howe and Willoughby, 2005: 72). The paragraph structure proposed by Fiderer (2002), Howe and Willoughby (2005) and Osmond (2015) is believed to assure the coherence of a paragraph. The summary sentence is of a great importance as it determines whether the content of the whole paragraph would be remembered by the reader (Howe and Willoughby, 2005: 75). Also, it leads the reader to the following paragraph. Finally, it is the duty of the concluding sentence to encourage critical analysis of the topic being discussed (ibid.). The summary sentence can be of four types: (1) a reference to the topic sentence of the paragraph (2) concluding the main points stated in the body part of the paragraph (3) restate the ideas of the paragraph by different linguistic expression appealing to the reader (4) the sentence related the content of the paragraph to the wider context and co-text of the particular piece of text (ibid.).

Howe and Willoughby’s (2005) analysis of paragraph structure has implied also the aspect of linking the ideas between the sentences. Depending on the function and logics of the ideas expressed in separate sentences, it is possible to divide between the following ‘sentence connectors (ibid.: 76): (1) connectors of continuation (2) connectors of transition (ibid.). As the titles suggest, the first type is concerned with the continuation of an ideas among several sentences. In order to reach this goal, repetitions, pronouns, as well as demonstrative adjectives are used (ibid.). On the contrary, transitional connectors are the lexical units indicating the opposite relationship between the subjects expressed (ibid.). They are of nine types: ‘consequence, comparison, contrast, example, restatement, addition, concession, sequence, and summary’ (ibid.: 76 - 77).

Thus, paragraph writing is a demanding process as many structural and linguistic features are of the topmost significance to compose an academic paragraph meeting the task requirements.

2.4. Lexical cohesion

Halliday and Hasan in their work ‘Cohesion in English’ (1976) have defined cohesion as follows: ‘Cohesion is a semantic relation between an element in the text and some other element that is crucial for it’ (1976: 8). Sanders and Maat (2006) have recognized lexical cohesion as one of the linguistic forces insuring the overall cohesion of a text. In their point of view, lexical cohesion follows the formula ‘two elements share a lexical field’ (2006: 591), especially emphasizing collocation. It is not only Halliday and Hasan (1976), but also Flowerdew and Mahlberg (2009: 1), who define lexical cohesion as the mutual relationship between the lexical units in the text, and how these units ensure the overall coherence of the text, especially emphasizing the need for a shared reference. Flowerdew (2010: 444) expands the idea proposed above by claiming that lexical cohesion is achieved when the aspects of collocation, colligation, synonymy and prosody all are taken into account.

Halliday and Hasan (1976) are the two most acknowledged authorities in the field of lexical cohesion. They have differentiated between a range of linguistic forces capable of creating a ‘cohesive effect’ (1976: 274), giving a special prominence to the categories of reiteration and collocation (ibid.: 277 - 284). The following picture is a summative depiction of the lexical cohesion scheme as proposed by Halliday and Hasan (1976: 288):

Type of lexical cohesion:	Referential relation:
1. Reiteration	
(a) same word (repetition)	(i) same referent
(b) synonym (or near-synonym)	(ii) inclusive
(c) superordinate	(iii) exclusive
(d) general word	(iv) unrelated
2. Collocation	

Figure 2.1 Types of lexical cohesion and their referential nature (adapted from Halliday and Hasan, 1976: 228).

Figure 2.1 illustrates the major types of lexical cohesion and the relationship between the lexical units belonging to these types. The following text is a description of each of the lexical cohesion type.

To start, the first type of lexical cohesion is reiteration. It is equal to repetition of the same word or the application of a synonym or near-synonym yet both words standing for the same referent (ibid.: 278; Flowerdew and Mahlberg, 2009: 1). Halliday and Hasan (1976) see reiteration to be of four types: ‘(1) repetition of the same lexical unit (2) synonymy or near-synonymy (3) superordinate (‘any item that dominates the earlier one in the lexical taxonomy’

(Halliday and Hasan, 1976: 280)) and (4) general word' (Halliday and Hasan, 1976: 277-278). They have agreed that reiteration is purely concerned with the lexical aspect where the common referent is a vital precondition for lexical cohesion to be present in the text (ibid.), in contrast to the vague distinction between the emergence of grammar or lexis in the case of a general word category described below.

One of the sub-types of reiteration – general word or general noun class – as claimed by Halliday and Hasan, is '[...] a small set of nouns having generalized reference within the major noun classes [...]' (1976: 274). The nouns, according to scholars, are classified according to their referents, which can be divided into several smaller groups, such as ' (1) human, (2) non-human animate (3) inanimate concrete count (4) inanimate concrete mass (5) inanimate abstract (6) action (7) place (8) fact' (Halliday and Hasan, 1976: 274). This classification allows a language user controlling the amount of vocabulary applied for the transmission of a thought, not getting either too vague or restricted in lexical choice because general noun functions as a superordinate element. A good command of semantic knowledge facilitates the ability to choose the most appropriate vocabulary to reveal the topic intended.

The scholars (1976: 275) continue by revealing the very fragile borderline between grammatical and lexical cohesion within the boundaries of the category of general noun as the lexical units sharing referential relationship imply the usage of an article *the* in order to stand in anaphoric status in a text. For example, the mentioning of *a girl* followed by *the girl* in the part of text to continue illustrates the referential relationship the nouns and the parts of the text stand into. Halliday and Hasan conclude that '[...] the vocabulary, or lexis, is simply the open-ended and most 'delicate' aspect of the grammar of a language' (1976: 281). Then, even minor grammatical changes are capable of establishing a text as lexically cohesive or not.

Simultaneously, they offer another factor perplexing the idea of general noun class and lexical cohesion. Halliday and Hasan (1976: 282) conclude that two lexical units might share a cohesive force even without sharing one reference point. Again, there is no denying that grammar has the power of uniting the lexical items of interest, as the use of different grammatical forms of one and the same word, for example, *girl* and *girls* establishes the 'cohesive relation' (ibid.: 283). Thus, an individual labelled as *a girl* belongs to the general group of *girls* (here, the general noun class would be "human"). Sanders and Maat (2006: 591) have finalized that the category of general noun appears to depend greatly on the language user's semantic knowledge, as well as on the attentive control of the 'lexicosemantic relationship' (ibid.: 591).

The second major type of lexical cohesion is collocation. Similar to Halliday and Hasan (1976), Fellbaum sees collocation as ‘the regular and statistically discernible co-occurrence of words that can be expressed quantitatively’ (2007: 8). As the definition suggests, the frequency statistics of the word combinations or patterns provide information about the language structure and typical usage, which is an advantage when deciding in favour of the vocabulary usage, especially when in the framework of corpus-based language analysis and text composition. Again, Halliday and Hasan (1976) are sure that collocation is the most complicated aspect of lexical cohesion to be attained. A closer inspection of the semantic fields of the lexical items intended to be applied are of vital importance. The idea proposed by researchers leads towards a conclusion that, even if two or more words appear as synonyms, their specific contextual usage and meaning sense or colouring determines the preference of one word or another depending on the context, which, to draw parallels with the focal point of the present research paper, i.e., application of corpora in academic writing instruction, can be found out by the analysis of corpus data. Halliday and Hasan (1976: 285) and Brezina et al. (2015) have touched upon the concept of collocation also in a wider scale, stating that every pair or set of lexical units can be perceived as collocations when the written text is contextually focused and the vocabulary is in associative relationship (Halliday and Hasan, 1976: 285). Even the word classes the lexis belongs to are not of the primary importance if this association exists. Then, the words *garden* and *dig* also would be of collocational nature (ibid.). The collocational tie of words, which is the very basis for a sufficient use of a language, appears to be a sphere of language proficiency able to be mastered only by the exposure to sufficient textual evidence. Especially demanding it is for the second or foreign language learners, as their perception is greatly affected by the linguistic peculiarities of the native language.

Turning to the collocation, lexical cohesion and text corpora, the collocation is considered a seemingly complex linguistic sphere for language users to develop. McEnery et al. (2006) have defined collocation within the framework of corpus, claiming that ‘the term *collocation* refers to the characteristic co-occurrence patterns of words, i.e., which words typically co-occur in corpus data’ (2006: 56). As mentioned in chapter 1 of the present paper, concordancing is the most widely applied research tool for collocation study (ibid.). The analysis of a concordance line provides information about the grammatical and lexical items—the preceding and following lexical units are demonstrated. Kennedy’s (discussed in Aijmer and Altenberg, 1991: 98-100) corpus-based study of the synonyms *between* and *through* proved that the analysis of the adjoining lexical units (i.e., collocations) demonstrate the

morphological features, as well. The scholar's study concerns the agreement of words in terms of number and grammatical tense of verbs (ibid.: 100). His research emphasizes another point somehow forgotten when doing the analysis of concordance lines – the aspect of collocation concerns not only the preceding, but also the following lexical unit (ibid.). Apart from that, a difficulty in determining the collocation might appear because of the gap between the words formulated as a collocation – one or several words can be positioned between the two lexical units under investigation (ibid.: 101).

Even though explicitly set out as grammatical feature, colligation portrays some traits of lexical cohesion as well. It was not mentioned in Halliday and Hasan's (1976) scheme of lexical cohesion proposed above, yet it appears to have a relationship with lexical cohesion as well. Colligation refers to 'the grammatical patterns a word appears in and the grammatical functions it serves, including the grammatical categories it realizes' (Hoey et al., 2007: 8; Halliday and Hasan, 1976). In other words, colligation is the typical grammatical form that a particular lexical unit is used in. A closer inspection of the theory about colligation, as proposed by Halliday and Hasan (1976), leads towards drawing parallels between colligation and collocation as the co-occurrence typical of two words (i.e. collocation) also implies a specific grammatical category to be used within that pair of words.

The discussion of lexical cohesion by Halliday and Hasan (1976) is throughout linked to semantic fields. Bordwell claims that 'a semantic field is a set of relations of meaning between conceptual or linguistic units' (1989: 106).

The subchapter introduced the categories and sub-categories of reiteration and collocation, as well as demonstrated the aspect of colligation as the major forces having the power of improving text's lexical cohesion. The researchers referred to have emphasized the importance of semantics in composition of a text where the vocabulary is interconnected.

The chapter was an overview of the application of corpora in academic writing teaching and learning. It revealed the place of corpora in writing instruction, as well as reflected the characteristics of academic writing from the point of view of its function, form and linguistics. In this respect, academic paragraph as a vital component of the formal writing has been discussed. A particular attention has been paid to the lexical cohesion as the driving force for the attainment of appropriate proficiency in the subject field.

3. CORPUS-BASED TEACHING AND STUDY OF STUDENTS' WRITTEN TEXTS

The goal of the empirical part is multilateral as the study of the role of corpus-related activities focused on lexical cohesion in English academic writing is both learner-centred and teacher-centred. On the behalf of students, the corpus-related activities implying direct use of corpus are aimed at developing the receptive and productive skills relevant for learning distinct types of lexical cohesion (discussed in chapter 2.4). On the behalf of a teacher, it is the indirect use of corpora for teaching material development and text analysis the results of which can function as a basis for a better understanding of the students' language use, as well as exert influence upon the planning and development of further teaching materials and process.

Halliday and Hasan (1976) (see sub-chapter 2.4) have outlined a range of linguistic features contributing to text cohesion, i.e., reiteration and collocations. The scholars describe the cohesive force possible to be attained by applying synonymy, collocations, the class of general noun and semantic fields. Because the vision of the present thesis greatly corresponds to Halliday and Hasan's (1976) enumeration of the linguistic features responsible for text's lexical cohesion, several types of lexical cohesion, namely reiteration, collocations and semantic fields, has been used as the framework for the empirical investigation.

The research has been carried out in various stages as a sequential approach to the students' written assignments was considered as one of the priorities of the work. Table 3.1 below indicates the stages of the empirical study and the types of lexical cohesion examined in each of them (the stages of the research are described in detail further):

Table 3.1 Stages of empirical research

No.	Research process stage	Type of lexical cohesion				Corpora use approach
		Repetition	Collocations	Synonyms	Semantic fields	
1.	Analysis of students' pre-texts	+ (reiteration to determine key words/lexical words – N, V, Adj)	+ (lexical words (N, V, Adj))	-	-	Teacher-centred direct use of the Wmatrix and AntConc
2.	Development of corpus-	+ (Reception of	-	-	-	Teacher-centred

	based teaching materials	lexical words (N, V, Adj) in academic paragraphs (based on theoretical material, see Appendix 2))				indirect use of corpus – the BNC
3.	Direct use of corpora for improving paragraph drafts	-	+	+		Student-centred direct use of corpus – the BNC
4.	Analysis of the final version of paragraph	-	+	+	+	Teacher-centred direct use of Wmatrix
5.	Development of learner corpus					Learner corpus

Table 3.1 points at the types of lexical cohesion that have been the focus in each of the research stages, as well as provides a general overview of the corpus use approaches.

The following chapter provides a detailed outline of the empirical research methods and data collection tools applied in the present study. It covers the specifications of research methods and the process of their application, finalizing with the obtained results.

3.1. Research methods and data collection instruments

Case study has been applied as the research method in the scope of the present thesis. Griffée (2012: 96) reports the case study to be a method for social investigations. It aims at providing qualitative data on the research topic of interest. The case study often is confused with the ethnographic research, as both are aimed at investigating a particular feature in a particular context (ibid.). However, Griffée (2012) concludes that the case study covers a narrower research area, which, in the framework of the study, is the lexical cohesion in paragraph writing. Nunan (1992) has mentioned the possibility to implement the quantitative perspective in the case study in order to enhance data analysis, which is considered as an advantage in the context of the present study as the number of students participating is a vital component for the validity and reliability of the present research. In reference to the types of lexical cohesion examined in each stage of the present research (see Table 3.1), number of students' texts and linguistic features in them plays an important role in making relevant conclusions. Griffée (2012: 99) concludes the description of case study method with the acknowledgement that it

is aiming at detailed and deep analysis of the research data, rejecting superficiality and vagueness.

Within the scope of the present study, two of the research stages have been managed by the indirect and direct analysis of corpus data from the **British National Corpus (BNC)**. The BNC is a body of more than a hundred million – word spoken and written texts covering multiple disciplines. The BNC is a monolingual, synchronic, general corpus (Online 1). It comprises authentic English British texts (ibid.). The majority of the corpus is compiled by written texts representing language variety from the perspective of genre, formality, dialects, and other. Within the framework of the present research, it has been applied in two different perspectives. Firstly, the BNC has been consulted online for the preparation of the teaching materials **piloted** among the research participants (the more detailed specifications will be discussed further in the paper in relation to teaching materials). Secondly, the third stage of the empirical research, i.e., students self-editing their comparison or contrast paragraphs, was performed by direct use of the BNC online. After receiving a feedback on the comparison-contrast paragraph draft, the students were instructed about the procedure of the BNC analysis (discussed in detail further in the present work). The final stage of the present study implied the creation of a learner corpus, i.e., student text corpus.

The research object being lexical cohesion, the author of the present paper applied the **WMatrix** corpus analysis and comparison tool. Wmatrix allows the analysis of: (1) concordance lines (2) frequency statistics (3) main semantic categories represented in the texts under examination (Online 2). Wmatrix was used in order to study: (1) key words in students' pre-texts (2) the semantic fields and their distribution in students' written comparison and contrast paragraph final versions. The concordance lines are capable of demonstrating the students' language use habits, e.g., application of collocations. Thus, concordancing is a feature indicating the extent to which the written texts are built on collocations. Next, frequency statistics is vital for the distribution of both lexical words and semantic fields. Also, data supported by quantitative evidence can stimulate language teachers and investigators to think over the further language teaching and learning process as different aspects might demonstrate a need for more scrupulous attention. Finally, the information on the major semantic fields found in the students' written texts is believed to be a rich material for analysis reflecting the manner of how a writer categorizes the thought in mind into a linguistic form that is lexically coherent content and function-wise. Similar semantic analysis by Wmatrix has been carried out by Balossi (2014), whose study of parts of speech, their distribution, and use of semantic domains in Virginia Woolf's novel 'The Waves'

demonstrates the possibility to perform a great research in literary studies by applying Wmatrix software. Klebanov et al. (2008) has approached Wmatrix for determining the success of political speeches, resulting in the findings that the opposite parties have earned the voter's loyalty each boosting their speeches by different semantic domains – medical care versus manufacturing and defence field. The list of the investigations based on Wmatrix analysis could be continued with more examples, such as the one by Ooi et al. (2007), who have approached one of the latest trends – blogging – from the perspective of genre-based language use. Finally, **AntConc** concordancer was used to examine collocations of the key words found in students' pre-text.

3.2. The research subjects and their text corpus

The research subjects were two groups of the 1st year English Philology Bachelor study program students at the institution of higher education in Latvia. For convenience, the groups would be referred to as Group1 and Group2.

Before the pilot lessons, the author of the present study was not familiar with the students, as well as did not have any impression about their English language proficiency, except for the assumption that Group1 could demonstrate a better language command than Group2. The personal experience of being a student, as well as the division of students into groups practiced in the concrete educational institution, i.e., division according to the results of the 12th year English language state exam, allowed making such an assumption. Not knowing the language proficiency level of the students might have been considered as a disadvantage as it could have affected the process of compiling and piloting the teaching materials, but it was made clear for the researcher beforehand to prepare materials and activities of low complexity as corpus-based language teaching and learning has not been a typical practice experienced by the research participants. Consequently, it was considered as a logical suggestion being in accordance with the students' educational background and experience at tertiary level studies. The researcher was introduced with the major characteristics of each group before entering the classroom.

Taking into account student group formation on the basis of the year 12 examination results, the teaching process in Group1 was believed to be brisker. Though, at the end of the research it was possible to conclude that the students from Group1 really have indicated a greater tendency towards active participation during seminars, the linguistic performance during the seminars did not differ greatly between the Group1 and Group2, though, the latter

being more introvert and reserved. Group1 indicated an overall interest in active cooperation with the teacher, as they asked questions, as well as commented and answered the points emphasized by the teacher. They were able to maintain alert and focused throughout the seminars. In contrast, the members from Group2 were more focused on individual and uninterrupted learning process. However, such a tendency of the group members did not have an explicitly negative impact on the students' results. It appears to be just a matter of personal preferences and methods of studying. Throughout the period of time ascribed for the research, the students became more open towards interacting with the instructor.

In the scope of the present research, the students were asked to submit 4 tasks: (1) the very first written texts, i.e., pre-texts, composed in the study course (2) responses to the corpus-based task aimed at recognizing lexical features of academic writing; (3) a draft of a comparison or contrast paragraph submitted for the instructor to analyse and provide feedback (later labelled as paragraph 1); (4) the final version of the comparison or contrast paragraph improved after the direct use of corpus (later labelled as paragraph 2).

The analysis of all the tasks was focused on lexical cohesion, still slightly differentiating in the lexical cohesive devices tested in each of the task (see Table 3.1). Thus, the piloting of the corpus-based task was aimed at investigating the students' perception skills, whereas all the rest of the tasks indicated participants' ability to put into effect various lexical cohesive devices, namely synonymy, collocations, and semantic fields.

Table 3.2 below demonstrates the number of papers submitted in each of the four tasks described above:

Table 3.2 Number of papers submitted by students

Task	Group1	Group2
Pre-texts	19	19
Task based on corpus-derived materials	18	18
Draft – paragraph 1 (receiving feedback)	16	15
Paragraph 2 (with corpus-based changes)	12	12

Table 3.2 depicts the number of responses achieved for each of the tasks. As observed, a drop in the number of responses emerged with every task. It ought to be explained by the observed absence during the seminars, misunderstanding of the requirements, or a drop into the number of students. To prevent a possibility of further ambiguities regarding the difference between the numbers of paragraph 1 and paragraph 2, it has to be mentioned that 9

students (both groups together) also had received the teacher's feedback on their written performance, but the corrected version was either not received at all (in 4 cases in both groups) or the improvements have not been made with the help of the BNC corpus (5 cases in both groups). Also, 2 students managed to compose and submit the paragraph during the last lecture. So, these paragraphs were excluded from the analysis.

Consequently, a learner corpus was created. In total, the learner corpus designed within the scope of the present study implies pre-texts, comparison or contrast paragraph drafts (paragraph1) and final versions of the said paragraphs (paragraph2). The texts are codified following the formula "STxx_p" for pre-texts (xx denotes the number ascribed for a paper, for instance, "ST01_p", while the letter *p* ascribes pre-text), whereas "STxx_d" stands for comparison or contrast paragraph drafts and "ST_xx" for their final versions. The codes for final versions implies also letter(s) E, CT, CA denoting to the topic of the paragraph (E = Education in Country X and Country Y; CT = Cinema and Theatre; CA = Childhood and Adulthood).

3.3. Research procedure and results

Research procedure comprised 5 stages (see Table 3.1). Within the first stage, the author of the thesis analysed pre-texts the particular research subjects had composed in the framework of the course in academic writing. The students were required to write a 250-word essay about travelling.

The overall purpose of the analysis of those texts was to gain an initial impression of the students' texts, especially the cohesive devices used in those texts. Within the framework of this study, the texts were researched from the angle of reiteration and collocations (Halliday and Hasan, 1976). The research was considered relevant in order to analyse the use of lexical cohesive devices among the students just entering their studies. The following is an average sample text from the collection of students' pre-texts. The instances marked in bold are the places where the word combinations, i.e., collocations, have attracted the researcher's attention:

Travelling

Nowadays travelling is a very popular way to relax, but not everyone notices that it has a lot of purposes.

As everything, travelling has its advantages and disadvantages. In my opinion, the main advantage is that travelling gives you a great opportunity to learn something new. You can learn a lot about the country you have been travelled to. You can meet a lot of wise and friendly people, who maybe will play a **big role** in your life. Also it has an educational purpose. You have a chance to learn about history and culture of this country also it is a way to improve your knowledge. About disadvantages. I think that travelling doesn't have any disadvantages at all. Person can always avoid unpleasant situations if he or she **picks the right travel agency** and there will not be any misunderstandings. You just have to do everything right.

For me **travelling is the best invention** the human had ever done. I do like to visit another countries of our big planet. It is always exciting to go to a place, where everything is not like in your home country. For many people travelling is just a way to relax, but for me it is a wonderful way to escape from everyday's routine.

Regarding the use of lexical cohesive devices in pre-texts, the general observation concerned frequency of particular lexical words (key words) and the collocational relationship they share with the adjoining lexical units. The study of the pre-texts started with examination of the most frequently applied key words from three major word classes: nouns, verbs and adjectives. As the objective of the first research stage was to gain only a general idea about the lexical cohesion of the pre-texts, 3 most frequently applied key words from each word class (nouns, verbs, adjectives) were chosen for further examination. Word frequency was analysed by **Wmatrix** tool. At this stage of the research, the information on part of speech and frequency was relevant. The analysis revealed the most frequently used nouns, verbs and adjectives to be the following:

- Nouns: people (123); traveling (78); world (51);
- Verbs: traveling (50); travel (49); see (33);
- Adjectives: new (55); different (46); great (23).

The next step of analysis concerned the study of collocations of the key words by using **AntConc** concordancer. The whole set of pre-texts was uploaded in the programme and the previously studied key words were written one by one in the search bar. The collocations were examined within the concordance lines. Figure 3.1 is a sample of **AntConc** workarea and data presentation.

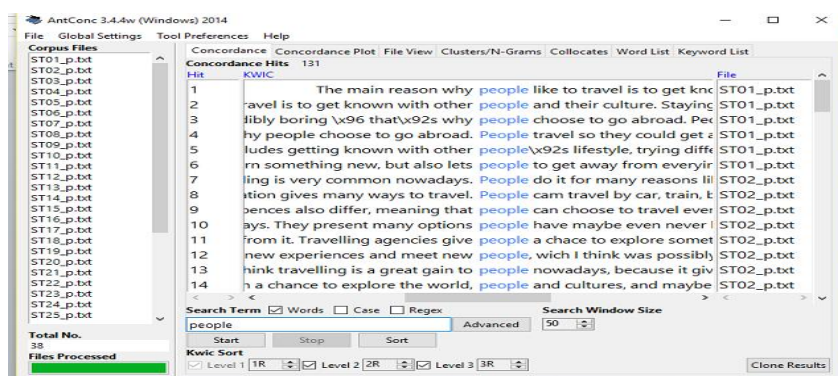


Figure 3.1 Concordance line presentation in AntConc.

The collocation research exemplified in Figure 3.1 was performed to all the key words under examination. The results indicated a tendency of the majority of the key words to collocate with a modal, preposition, auxiliary word, conjunction or pronoun preceding or following it. In cases, where the collocation was mainly built on the basis of key word and function word combination, no mistakes were observed. However, in each word class that the key words belonged to gaps in combinations with content words were observed. Some of the examples are provided below:

To explore few people, peoples' horizons, it lets people to [...], important place in people, luxurious traveling, traveling increases, culture traveling, traveling location, wide world, [...] try to find new meeting [...], new knowledges, different history.

Consequently, the collocational analysis of the key words in pre-texts revealed that the students were only partly aware of the way words collocate as a means for the production of a lexically cohesive text. Either the collocation was not successful at all (e.g., *to explore few people; to find new meeting*) or it appeared to be partly appropriate (e.g., *luxurious traveling*). The main intention of the students obviously had been to reflect their thought, the idea they had in mind about the particular topic. The semantic and collocational harmony of the text occasionally appears to be of the secondary importance. Here, several possible reasons might be responsible for this: (1) students are not familiar with the concept of collocation as such yet, as this is the very beginning of their tertiary level language studies (2) students' language proficiency is only at the beginning to be developed till advanced level, thus, they have not acquired the aspect of synonymy adequately, i.e., the amplitude of the word stock is on its way to be broadened. The acquisition of the style of the language, including the aspect of lexical cohesion and collocations, develops only with practice. The mutual relationship between synonymy and collocation has not been developed properly. The analysis demonstrated a general preference to the most general word from the set of synonyms (e.g., *it lets people to [...]*). For the reader, evidently, the message is still understandable, but, when

analysed in respect to collocations, some parts of the text feels too pretentious or “technical”. The influence for students’ native language could be noticed. Thus, the example “it lets people” resembles a typical Latvian phrase “tas ļauj cilvēkam”.

The second stage of the research (see Table 3.1) was dedicated to the piloting of corpus-based teaching materials (indirect use of corpus). The materials for the tasks were compiled by the researcher after consulting the British National Corpus (BNC) and adjusting the materials extracted from the corpus to the learners’ needs and study context. The activities presented to the students were aimed at developing perception skill, in particular – lexical features of academic writing.

It is important to explain the relationship between the goal of the particular activity and the study framework within which it was delivered.

The teaching and learning process of academic writing – comparison and contrast paragraph writing, in particular – was based on both theoretical guidance and practical activities. Even though, the course specifications and plan at the time of piloting corpus-based teaching materials envisaged the introduction of register differences in writing (The beginning of the research took place at the time when the students were supposed to start acquiring the topic ‘Formal and informal writing. Register. Paragraph structure. Topic sentence’ (taken from the calendar plan), it was also the lexical cohesion in paragraph writing that was incorporated within that topic. The goal of the task was to develop students’ perception skills, i.e., focus on linguistic features of academic writing. Ability to encounter authentic, yet slightly modified texts was believed to foster the recognition of the general lexical framework applied in academic writing, and the knowledge generated could function as a reference source for the students when engaged in further academic writing tasks. The procedure consisted of introducing the main grammatical and lexical features of informal and formal writing, followed by practical activities. To exclude any deviations from the course plan, the researcher’s teaching materials were adapted to the existing materials introduced by the course developers and teaching staff (see Appendix 2). After getting familiar with the theoretical background supposed to be taught to the learners, the author of the present research consulted the BNC to search for paragraphs, exerting indirect use of corpus. The texts implied in the task were chosen from the section “Academic” and slightly adapted in terms of length and topic sentence precision. In addition, the texts were searched under the topic “cinema” (writing in the word in the query box) so as to sound convenient for the students as this theme is included in the spoken part of the study course.

Both groups underwent 3 pilot seminars. The first pilot seminar was devoted to the theoretical introduction of the main grammatical and lexical features of informal and formal register, followed by a corpus-based practical activity believed to foster the perception of lexical features of academic writing as a vital component for lexical cohesion (see Appendix 3). The texts for the task were taken and slightly adapted from the British National Corpus (BNC), the primary sources being Slocum (2000), Hope and Martin (1998), and Teachout (2009). The BNC is an interface of the Brigham Young University offering less up-to date linguistic material, but exercising British English, which is the variety of English adhered at the particular institution of higher education. The variety of English functioned as the main reason for the choice of the text. Approximately 45 minutes of the seminar were devoted to the research procedure, during which the investigator briefly introduced the aspect of formal and informal register. Then, the students were exposed to a practical activity. The task could be classified as a recognition task. Firstly, the students were offered to read all the short paragraphs under the topic “Cinema”, the formality level of paragraphs differentiating. Then they had to fill in the provided table (see Appendix 3) by indicating the lexical and grammatical features observed in each of the paragraphs. These features have been mentioned in the handouts also provided and allowed to be consulted (see Appendix 2).

The following table indicates the level of formality of each paragraph included in the activity, distribution of lexical words (nouns, verbs, adjectives) in each paragraph, the presented lexical features on the basis of the theoretical material developed by the teaching staff of the course. Such a table is designed to demonstrate what approximately the paragraphs were alike and what lexical features were expected to be noticed by the students. The table below does not concern grammatical aspects as they were not the primary concern of the present research. Yet, the corpus-based activity is believed to have been a suitable material for acquiring grammatical features peculiar for different registers, as well.

Table 3.3 Distribution of lexical features among paragraphs

Number of paragraph	Number of words in paragraph				Register	Lexical features
	N	V	Adj	Other		
1.	34	17	20	53	Formal	<ul style="list-style-type: none"> - ‘High lexical density - High nominal style - Greater variety of lexis - Complex vocabulary, abstract terms - Limited use of ‘run on’ expressions - Limited use of personal

						pronouns' (Appendix 2).
2.	37	31	12	76	Informal	<ul style="list-style-type: none"> - 'Low lexical density - Less abstract lexis - Extensive use of personal pronouns' (ibid.).
3.	57	19	26	85	Formal	<ul style="list-style-type: none"> - 'High lexical density - High nominal style - Variety of lexis - Complex vocabulary - Limited use of 'run on' expressions - Limited use of personal pronouns' (ibid.).

A curious feature has been observed with the second paragraph representing informal register. Even though informal texts are claimed to be of low nominal style, indicating to the dominance of the verbs, it is not exactly the case in the particular text. The amount of nouns is even higher than verbs – 37 against 31. However, the first impression of the text, especially after reading the paragraph 1 which is written in formal style, might lead to a conclusion that verbs appear to dominate in the text. Only a scrupulous examination of the passage can reveal the actual situation.

After reading the paragraphs, the students had to fill in the table, mentioning the texts' register and the grammatical and lexical features supporting the choice made (see Appendix 3). To remind, the purpose of the activity was to develop students' perception skills of lexical features typical of academic texts as a feature shaping lexical cohesion of a text. The lexical characteristics were of the primary interest and importance from the perspective of the present research paper. In various cases, students had indicated grammatical characteristics in the space ascribed form lexical features, and vice versa. As a result, the researcher had to be particularly attentive during the data collection process.

In total, 36 response papers (18 from each group) of the task were collected. Even though, the quality of the task was considered prior to the quantity of the responses from each group, the final results depicted in the Figure 3.1, Figure 3.2, Figure 3.3, when observed also from the perspective of quantity, indicated several diametrically opposite traits regarding students' perception skills. The students' answers were counted manually.

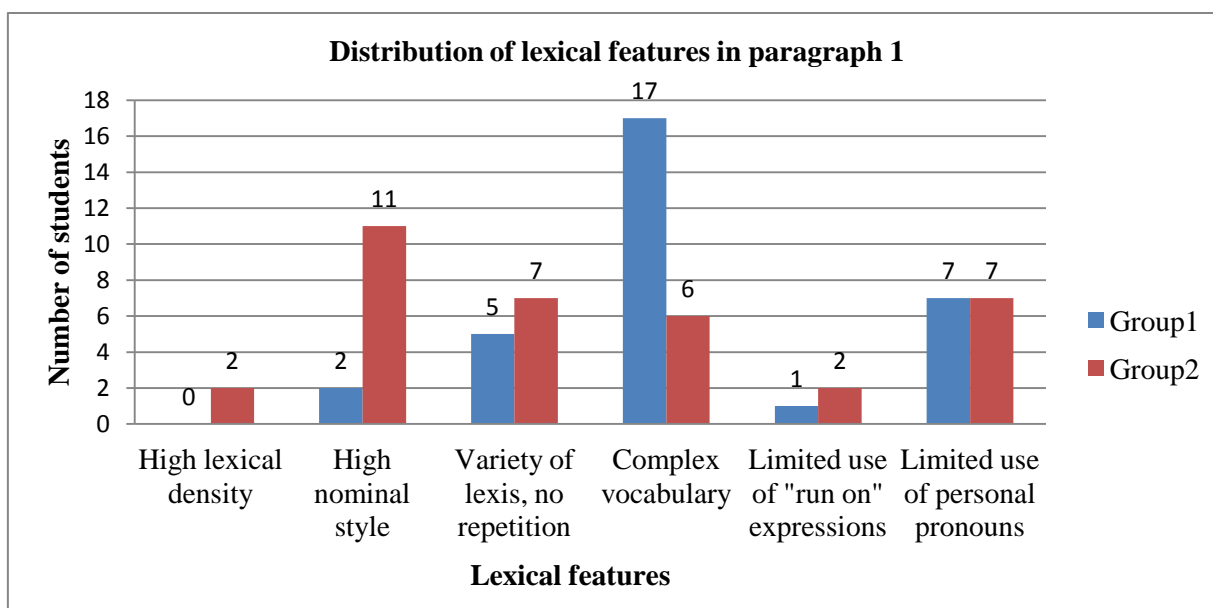


Figure 3.2 Recognition of lexical features in paragraph 1 among Group1 and Group2

The Figure 3.2 demonstrates the students' capability of differentiating between a range of lexical features in authentic English language text of formal register taken from the corpus. The results indicate a surprising unconformity among the representatives of both groups of students. Group1 - 17 students out of 18 - dominates in the instances of recognizing the text to possess complex and abstract vocabulary, while Group2 is in charge of all the rest of features mentioned except for the complex vocabulary and limited use of personal pronouns, which an equal number of students from both groups – 7 students from each – have noticed. As described above, the paragraph is considered to be of high lexical density and high nominal style. The nouns, verbs and adjectives found in the text are explicitly content words. As such, the text is saturated with information. However, a very small proportion of the students have mentioned high lexical density as one of the leading lexical features. In Group2, only 2 students have marked it, while in none of the works in Group1 such a lexical feature has been observed.

Still, a closer inspection of the results leads to seeing compatibilities among the range of answers devoted to: (1) high lexical density (2) high nominal style (3) variety of lexis (4) complex vocabulary. Group 1 students have identified the complex vocabulary used in the paragraph, but not lexical density. Similarly only 2 students from Group 2 has noticed high lexical density of the paragraphs, whereas the majority of the students had noticed high nominal style, variety of lexis and complex vocabulary in the paragraphs that they had analysed. The possible cause for the particular results happen to be noticed – the students in both groups actually were capable of perceiving the formal paragraph as containing a great

deal of content words, i.e., high lexical density as a linguistic peculiarity processed in the text, but they have attributed this characteristic under the labels of:

- Group1 - complex vocabulary;
- Group2 – high nominal style, variety of lexis, complex vocabulary.

In the case of Group1, explicitly, the content and the “load of information” of the vocabulary are as if understood, but the terminology offered to classify the phenomenon might have caused uncertainties among the students. Thus, they have chosen to name the great amount of the content words as “complex vocabulary”. A similar occurrence cannot be excluded among the distribution of the results from Group2, where the factor of high lexical density might have been dispersed under the headings of high nominal style, variety of lexis, and complex vocabulary.

As regards the limited use of run-on expressions, 1 student from the Group1 and 2 students from Group2 have noticed the particular lexical feature, which is a good sign as the paragraph really does not imply any such expressions. However, the small amount of students having mentioned this feature reveals that it is easier for the students to make notice of the present rather than absent lexical features. From this point of view, the use of authentic texts taken from corpus encourages the discernment of the linguistic congruencies.

Finally, 7 students from each group have mentioned the limited use of personal pronouns in the text under study. Even though 7 students out of 18 in each group do not reach even a half of the participants, this might be one of the easiest linguistic characteristics to spot when drawing parallels with the students’ own habits of using personal pronouns in their written compositions (pre-texts) mentioned below.

To sum up, the analysis of the responses devoted to the first – formal register – paragraph have revealed two major trends. Firstly, even though the background knowledge of the students from both groups was believed not to demonstrate highly opposite indications in the results, the actual situation has proved partly diametrical traits in their ability to perceive linguistic features in a text. Secondly, despite these differences, the application of corpus-based material has justified the expectations – the students have been encouraged to notice the major lexical features of a formal text, i.e., high lexical density (here, distributed among such aspects as high nominal style, complex vocabulary, and varied lexis) and limited use of personal pronouns.

Figure 3.3 depicts students’ analysis of linguistic features of paragraph 3, which also was written in formal register. Though, in comparison to the first paragraph, the latter is slightly less formal, the general lexical features are the same. The most prominent are:

(1) high lexical density (2) high nominal style.

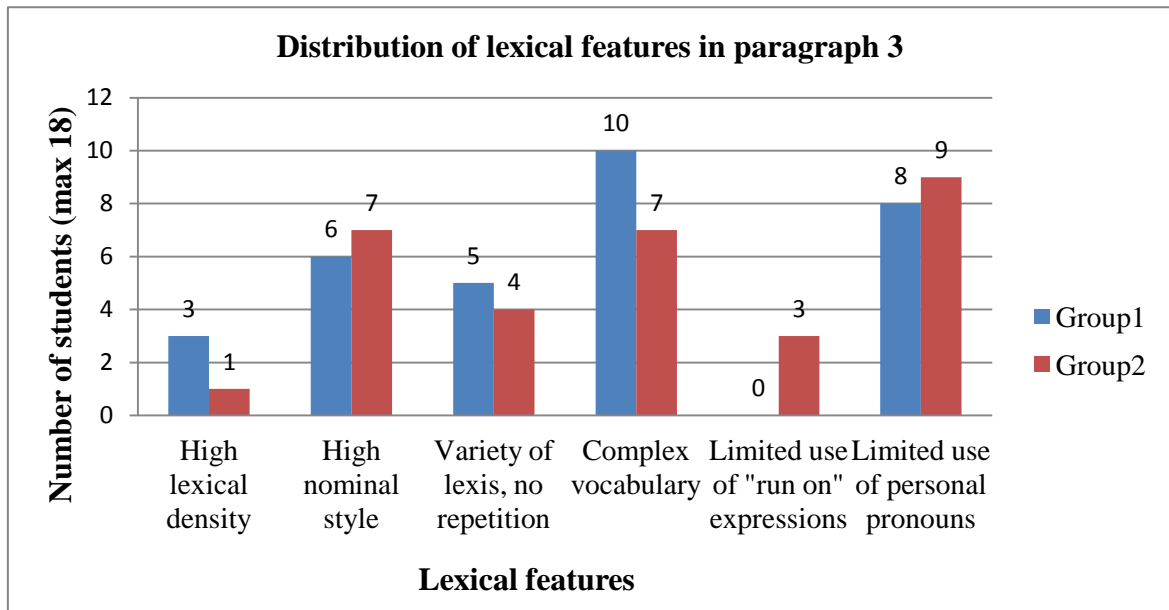


Figure 3.3 Recognition of lexical features in paragraph 3 among Group1 and Group2

Even though, the data in figure 3.3 show the general tendency of students' linguistic analysis of the paragraph similar to that performed for the previous paragraph (see Figure 3.1), i.e., complex vocabulary and limited use of personal pronouns reaching the highest recognition statistics, a remarkably greater range of students has noticed also such features as high lexical style and high nominal style. To obtain a more detailed analysis, the results from Group1 and Group2 were viewed separately.

To start with the results from Group1, 3 students have noticed the presence of high lexical density being one of the features of the text. Similarly, in comparison to paragraph 1, recognition of high nominal style has increased, reaching 6 students. Consequently, it is believed that success in noticing these lexical features might also promote the comprehension of lexical cohesion as a significant component of academic writing. In addition, the increase in recognizing high lexical density and nominal style appears to compensate the decline in the number of students having named the concept of complex vocabulary. Consequently, the initial impression is fallacious – the drop in mentioning complex vocabulary actually signals progress in the students' ability to differentiate between more scrupulous lexical features, such as high lexical density and high nominal style. Further, the limited use of personal pronouns and “run-on” expressions differ from the results found in relation to paragraph 1 only within the boundaries of +/- 1 response.

It could be concluded that students from Group1 have performed a more detailed analysis of lexical features in paragraph 3, which, hopefully would promote the consideration

of lexical cohesion in formal academic texts. Even though, the reiteration in the particular case does not concern itself with the repetition of one and the same lexical item, it appears that students are capable of noticing reiteration of different lexical items yet sharing a common feature, e.g., high nominal style or lexical density.

Turning to the linguistic analysis of paragraph 3 performed by Group2, an opposite tendency in comparison to Group1 has emerged. The results show an inferior recognition level of particular lexical features when compared to these features in paragraph 1 (though, both paragraphs are, for instance, of high lexical density). To illustrate the case, only 1 student has managed to mark high lexical density; 7 – high nominal style; 4 – greater variety of lexis. The rest of the lexical features offered in the list, i.e., complex vocabulary, limited use of personal pronouns and “run-on” expressions have increased for only 1 or 2 responses in comparison to the analysis performed for paragraph 1.

To conclude with the analysis of the lexical characteristics of formal register, a range of important conclusions can be drawn. There are two so called corner-stones that the corpus-based materials have brought up. Firstly, it is the overall complexity of the vocabulary used in formal, academic texts. Due to the students’ prior written proficiency and experience, the text analysis is believed to have actualized the lexical features characterizing an academic writing, building the very basis for the exploration of lexical cohesion. Secondly, the implementation of two paragraphs labelled as formal has turned out to be a successful strategy as the number of cases when the students have been able to notice a particular feature differ in number between the both paragraphs (here, it is important to remind that both paragraphs actually were similar in terms of lexical characteristics). Next, the paragraphs under consideration implied a variety of words of high lexical density, including nouns, verbs and adjectives. As the focus of the particular research stage was to examine the reiteration of lexical words, it has to be emphasized again that here the central criteria was not the repetition of one and the same word (though it cannot be excluded that paragraphs comprised also this kind of repetition), rather recognition of words sharing one and the same lexical feature of formal, academic written texts.

Turning away from the formal register and the lexical characteristics typical of it, the second paragraph in this activity presented informal register. Figure 3.4 informs about the lexical features the students have noticed. Nevertheless, the text for the analysis also has been found in the BNC under the section “academic”.

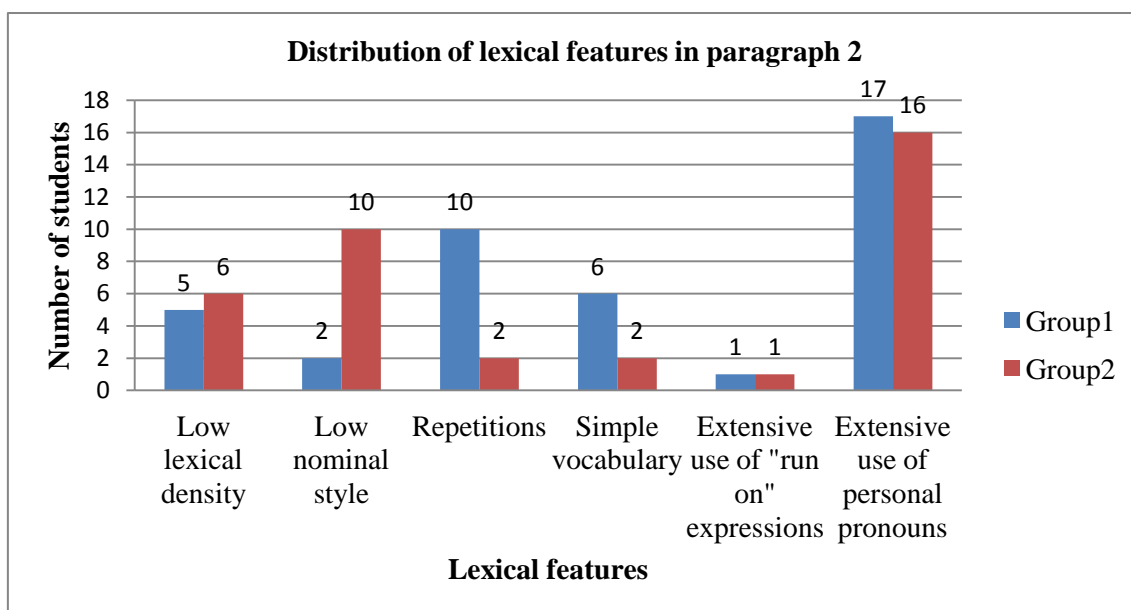


Figure 3.4 Recognition of lexical features in paragraph 2 among Group1 and Group2

The analysis of the lexical features in a less formal written text is considered to have caused fewer complexities as the students had a chance to compare this paragraph with the formal ones and their personal experience in writing. Still, the diversities in lexical features noticed among the students from Group1 and Group2 have reminded. The positive aspect, in comparison to the analysis of formal register paragraphs, is triplication in the number of students having paid attention to lexical density. However, it is still only one third of the students from each group (18 students in a group) participating in the activity that have managed to notice this linguistic feature. Next, the following two lexical features – nominal style and range of repetitions – indicate the opposite distribution of responses among the both groups. In Group2, 10 students have found the text to contain a considerable number of verbs, while in Group1 only 2 students have noticed this linguistic feature. In contrast, the opposite situation concerns the instances of word repetition. Thus, 10 students from Group1 have considered the variety of lexis as low, while in Group2 the same linguistic feature has been noticed by only 2 students.

As a result, Figure 3.3 re-assures that corpus-based teaching materials are capable of advancing the level of recognition of various lexical features and lexical cohesive force. The representatives from Group2 appear to pay more attention to word classes, while their colleagues from Group1 concentrate on the features that are more lexico-stylistic related, i.e., range of repetitions, formality level of the vocabulary applied.

Further, conformity exists in terms of the range of “run-on” expressions and use of personal pronouns. The author of the present paper has liberated herself to assume that the high number of instances when the extensive use of personal pronouns has been mentioned is derived from the students’ personal writing habits. The analysis of the students’ pre-texts affirmed an extensive use of personal pronouns.

To summarize, the piloting of the corpus-based teaching materials, i.e., the analysis of the students’ responses, has actualized the following lexical aspects of academic writing:

1. The corpus-based teaching materials have promoted the students’ perception and recognition skills. A detailed inspection of the paragraphs has allowed students making comparisons of the distribution of various lexical aspects assigned to different formality levels.
2. The analysis of the written texts of formal register has urged a necessity to rely on the students’ personal writing experience at secondary school from two perspectives: (1) the absence of a range of lexical features in students’ own texts is believed to foster the recognition of these features in the formal corpus texts offered for the analysis. Here, the emphasis is put on the aspects of lexical density, nominalization and complex vocabulary in particular (2) on the contrary, the presence of a set of lexical characteristics in students’ own informal writing practice also is considered to be responsible for the ability to affirm their presence or absence in the text(s).
3. Even though a detailed division between the two groups undergoing the pilot teaching was not considered as one of the priorities of the present study, the final results of the second research stage implying paragraph analysis have proved to be a rich source to be taken into account when planning the further English language teaching process in general. This discovery has emerged from the diversity in the number of students from each group recognizing different lexical features in paragraphs. The ability to examine their priorities in text analysis (with priorities meaning the lexical features observed) might have provided useful background information to be taken into consideration when developing further teaching materials and planning lectures.
4. Despite the final results of the second research stage regarding recognition of lexical features among different register as the basis for lexical cohesion in academic writing reached the expectations of the author, the results in several individual research stages might have been believed to be better. However, this phenomenon has a logical explanation. The students participating in the investigation undertook their very first steps at tertiary level education. So far, the students had mostly worked with the text books and grammar books. Then, the texts

and materials in books have been adapted to the particular topic of the chapter. In contrast, not forgetting about the necessity to link the corpus-based materials with the course description and calendar plan, the corpus-based materials developed by the author of the present study have been taken from the collection of authentic texts. i.e., corpus. Then, the low level of particular lexical feature recognition is explainable, as the students were not used to working with such corpus materials yet. But it is worth remembering that the results of the second research stage have to be perceived holistically, not paying too much attention to recognition of some particular lexical features implied in the texts of the task. If so, the students' have indicated good results in recognising different lexical features typically used in formal and informal academic texts.

The third stage of the present study was focused on the students' composition process of comparison or contrast paragraphs. The following themes were suggested:

- Theme 1: Childhood and Adulthood
- Theme 2: Theatre and Cinema
- Theme 3: Education in country X and country Y

The topics were suggested by the author of the present research. Three themes were offered to provide the students with the possibility to choose the theme that they might be more interested in. Taking into account that there were ERASMUS exchange students among research participants, these students were encouraged to tailor themes, especially Theme 3, according to the context of their own country or institution they study. Apart from the 3 students being in favour of the said topic, the rest chose between *Childhood and Adulthood* and *Theatre and Cinema*. The topics were presented in the second seminar when the students were exposed to theoretical guidance about the goal, structural and lexical peculiarities of the said type of paragraphs (see Appendices 4-6). The composition of a paragraph was set as a home task were instructed to submit paragraph drafts electronically before the next seminar.

In total, 31 comparison or contrast paragraph drafts were received before the next seminar for being granted with a feedback from the instructor till the next seminar. The feedback was focused on the lexical cohesive devices analysed in the third research stage - synonyms and collocations (see Table 3.1).

Before turning to the specific features of the drafts received, it has to be mentioned that a time-span took place between the previous and this stage of the research. After the piloting of corpus-based materials for the instruction of formal and informal register (research stage 2) the course description and calendar plan suggested a range of different aspects of academic writing to be studied. In accordance with the academic regulations and the arrangements made

with the professor delivering the particular study course, the section devoted to the comparison and contrast paragraph writing was selected as the most appropriate period for the realization of the third step of the present research. Consequently, advancement in the students' general writing proficiency was to be taken into account by the instructor of the research during the analysis of the draft of the comparison or contrast paragraph (paragraph 1).

The third stage of the study comprised four sub-stages, and each of the sub-stages was devoted to a particular writing activity aimed at the development of a comparison or contrast paragraph. So, the first sub-stage included the analysis of comparison and contrast paragraph examples as a part of theoretical guidance. The author of the present study introduced basic theoretical information regarding paragraph writing followed by practical exercises on structural and lexical features of comparison and contrast paragraphs (see Appendix 5 and 6). The activities of this sub-stage were done individually, followed by a group discussion.

In the second sub-stage, the students were asked to compose their own comparison or contrast paragraph. In the very first writing steps the instructor's assistance was ensured to guide the students. The researcher observed the students and their progress in writing. Due to time restraints and the number of students in a group it was not possible to have an individual approach for each and every student, but the researcher focused her efforts to assist as many students as possible. At the very beginning of starting to compose the paragraphs, the students were urged to be specific about the goal/function of the paragraph – either comparing or contrasting, as well about the points to be discussed. The precision at the introductory part of the paragraph is vital not only for helping the students in maintaining the text focused, but also for the researcher performing further analysis. The students were given time to finish their comparison or contrast paragraph drafts till the next seminar taking place in a week's time. They had to submit them to the researcher electronically.

During the said period of time the researcher analysed the submitted paragraph drafts (i.e., the third sub-stage). The study of the drafts of the comparison or contrast paragraph of Group1 and Group2 revealed the following:

- Out of 31 paragraph drafts received, 23 were aimed at contrasting, whereas 8 – at comparing the two subjects implied in the topic of the paragraph.
- In 28 drafts the topic sentence or the introductory part of the paragraph included the points intended for comparison or contrasting. Such concretization of the points focused the researcher's attention towards the semantic fields of the vocabulary and their appropriateness for the attainment

of the aspect of lexical cohesion and function of a text (discussed in detail in the fourth stage of the empirical research).

The analysis of the paragraph drafts proved no necessity to differentiate between whether a paragraph has been written by a representative from Group1 or Group2 as in both groups the texts were observed to be similar. The paragraph draft analysis ended by preparing feedback of the drafts for the students.

As the focus of the present study is lexical cohesion, the grammatical peculiarities of the drafts were not the primary concern. Yet any serious mistakes have been indicated to be considered. The framework suggested by Halliday and Hasan (1976) was taken into mind when providing comments for the improvements of the drafts. The majority of the feedback's suggestions reflected about the three concepts: (1) use of pronouns (2) collocations (3) repetition, i.e., synonyms. Use of pronouns, though, not indicated as a point for improvement in very many paragraphs, related to the instances where a lack of anaphoric reference or logically understandable referent could not be observed. Even though, not implied in the schematic representation of the types of lexical cohesion proposed by Halliday and Hasan (1976) (see Figure 2.1), it provoked considerations for further studies to be performed in the subject field. To continue, collocations and synonymy were mostly implied in the suggestions for contextually appropriate word choices. Collocations and synonyms are two types of lexical cohesion that are strongly related as the formation of an appropriate collocation to a great extent depends on the choice of an appropriate word from a string of synonyms. As a result, the peculiarity of the topmost interest was word choice from the perspective of synonymy (collocations) and repetition avoidance.

The following Example 1 is one of the best contrast paragraph drafts submitted for the feedback:

Example 1:

Education in Belgium and Latvia

In aspect of education such as preliminary study approaches in preschool institutions and schools, Belgium *strongly differs* from Latvia.

First of all, preschool institutions in Belgium have a very unusual method of preparing children for school; they are forced to understand a lot of tasks which are *closely related* to practical things in life. *Learning process* occurs mostly outdoors parallel with physical activities. At the age of 3 they are already self-sufficient, and possess the basics of different natural sciences. On the contrary, in Latvian preschool institutions children are mostly taught basics of mathematics and some grammar. It gives a good entry-level knowledge for school; however, the process is not as complex and creative as in Belgium. Secondly, schools in Belgium have a system of dividing students into the fields of their skills; they can choose juristic, engineering, medical, etc. directions of studying, so they are better prepared for the

university. In Latvian schools, students are more ‘independent’, they are not offered any special studying directions; instead, a good general knowledge is gained before starting the bachelor’s degree.

In spite of that these countries are in the European Union, both are very antithetical in terms of preliminary study approaches in preschool institutions and schools.

The instances in *italics* are the ones that are considered a good example for collocations, e.g., *closely related*. These words typically form collocates in the English language. While the phrases underlined, for example, possess the basics, are the ones suggested to the author of the paragraph to be re-considered. The underlined cases show that a successful collocation is not formed. In addition, the student has been suggested to consider the use of a more appropriate word, i.e., synonym for *possess*. A similar procedure would be offered for the phrase starting bachelor’s degree, and other underlined cases.

The evidence from Example 1 highlights a range of positive aspects, as well. Firstly, contrasting has been specified as the aim of the text, as well as the point – preliminary study approaches – is defined clearly. Secondly, the student has not indicated a necessity to provide feedback on the use of personal pronouns.

The Example 1 illustrates the overall tendency to form the contrastive function of the text by the application of: (1) verbs expressing opposition (*differs*) (2) linking devices expressing opposition (*on the contrary; however*) (3) phrases expressing opposition with the help of comparative degree of adjectives (*is not as complex as*) (4) adverbs indicating inequality between the concepts contrasted, as well as adverbs as intensifiers (*better; more; strongly*) (5) adjectives expressing opposition (*antithetical*); conjunctions (no contrasting conjunctions observed in the sample paragraph, yet present in other students’ texts). Consequently, Example 1 above could have been assessed successful in terms of lexical cohesion of the text, yet indicating a need for several instances to be reconsidered. It belongs to the most qualitative paragraphs.

Despite the considerably sufficient list of lexical units marking a contrast, their distribution points at an opposite phenomenon – the analysis of the whole set of students’ paragraphs asserted the dominance of the use of linking devices and conjunctions as a lexical technique to achieve the goal of the text, i.e., comparing or contrasting. Then, the structural pattern of the majority of the drafts often resembled a description of the subjects (e.g., theatre and cinema) within the boundaries of the points specified in the introductory part or the topic sentence of the paragraph (e.g., production and audience), while the function of the text –

comparison or contrasting between the concepts – was indicated by the insertion of a linking device or conjunction.

Example 2 is lower in quality in comparison with the text of Example 1:

Example 2:

Theatre and cinema

Theatre and cinema are two similar and at the same time dissimilar forms of art. Theatre includes real people performing in front of crowd where movies are pre-filmed and no actual pressure of outside is felt. Each mistake is at the essence, because movie artist can go over same thing again where in theatre actors can't.

On the first hand, theatres and cinemas both complete field of art, expressing emotions and features where real people are thrown into complex characters and their life. Actors are forced to perform at their best and most professional levels, but standing in front of crowd is more difficult than making a movie.

On the second hand, controlling facial expressions and emotions are required while filming too, because it will be viewed by more people compared with theatre so every expression is at its best. Delivering a quality performance is essential and preparation is needed as well as in the theatres.

In conclusion, both of these art forms are important, but they cannot be compared, because they both work in specific fields and require specific knowledge and experience.

Example 2 demonstrates a lower level of lexical cohesion. It is noticeable by the various instances (underlined) where a feedback has to have been provided in terms of word choice, i.e., use of vocabulary, synonyms in order to form appropriate collocations. The author of the paragraph is likely to have received the following suggestions to be considered for improvement of the text:

- “dissimilar” – a synonym would be preferred to prevent one-stem word repetition in single, short sentence;
- “crowd” – in the context of theatre, the people attending the performance are referred to by another term.
- “pressure of” – wrong preposition used.
- and others.

Even more, the function of the paragraph appears to be confusing: The linking devices – *on the first hand, on the second hand*, as well as conjunctions – *but, as well as* – appear as insertions informing about the diversities between the aspects commented on in the paragraph. Only the concluding sentence claims explicitly that the function of the paragraph was contrasting.

Consequently, the study of the first drafts of the comparison or contrast paragraph left the researcher with a positive impression about the students' writing skills. Remembering the pre-texts, the drafts had grown in terms of the information amount and sophistication. Though

not of the primary concern, the grammatical patterns were noticed to have developed in complexity. Still, the spelling and punctuation mistakes could not be excluded completely, the latter being a point logically explained by the impact of the native language, mostly Latvian or Russian. The draft analysis could be concluded with a general acknowledgement that the students have indicated gaps in forming appropriate collocations. Then the goal of the fourth sub-stage was to promote corpus-based analysis and correction of these gaps.

The fourth sub-stage consisted of the students themselves upgrading the paragraphs by the active application of the BNC for the examination of: concordance lines; synonyms for the lexical units of interest. The fourth sub-stage began with the students receiving the researcher's feedback. After familiarizing with it and receiving brief comments, the students were asked to continue their work at the paragraph, especially focusing on the word choice. Before that, the researcher distributed the manual for operating within the British National Corpus webpage (Online 1) and helped/instructed the students on how to log in the system. None of the students claimed to be familiar with the BNC. The second step was to register on the British National Corpus page.

After registration, a sample query, as demonstrated in Figure 3.5, was exemplified and performed together with the students in order to familiarize with the design and information extraction manner.

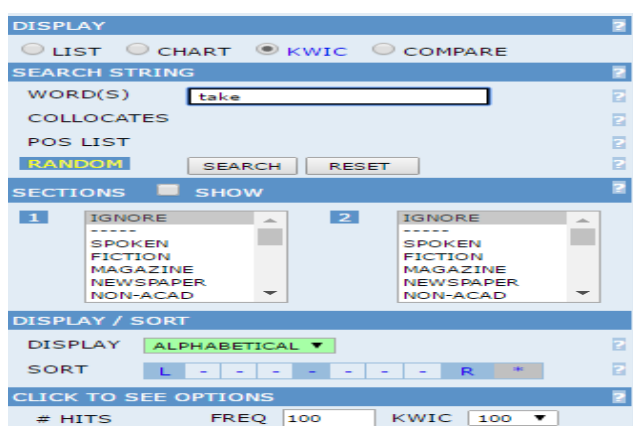


Figure 3.5 Sample query 'take' entered into the search string

As indicated in Figure 3.5, the students were asked to a word 'take' into the search bar. Due to the time restrictions and the necessity to remain precise and focused in terms of the linguistic information, i.e. collocations, intended to be emphasized, the students were asked to choose 'KWIC' (Key Word in Context) option. The contextualization stressed with the use of different colors fosters the perception of the language use of interest, paying particular attention to the parts of speech distribution. The analysis of concordance lines offers the information about typical collocations.

CLICK FOR MORE CONTEXT				[?]
1	H7K	W_non_ac_nat_science	A B C	bottles and beer cans . The old-fashioned bottle , shaped to take a cork to figure 7.3a) was replaced about 1900 by the
2	CEL	W_newsp_other_commerce	A B C	They should be rewarded when things are going well but should take a financial hit when the chips are down . ' # Pit
3	ASV	W_newsp_brdsht_nat_social	A B C	at their portrayal as lazy gypsies . But those who can take a loke have ranged from children to adults -something that the
4	CEP	W_newsp_other_sports	A B C	gun . ' Lennox is like an old-time champion who can take a man out with one punch . ' Duva was reacting to
5	K5M	W_newsp_other_report	A B C	Denny Macgee # ALFRED McTEAR , the first Scot preparing to take a tobacco company to court blaming it for his lung cancer and
6	CR7	W_pop_lore	A B C	from 5% in 1992 to zero in 1997 . Textile tariffs take a year longer to reach zero . The agreements also allow the
7	B7G	W_non_ac_nat_science	A B C	elephant handlers told the Indonesians that a 50 km drive would take a year to The Sri Lankans move their herds in stages ,
8	J6Y	W_ac_polit_law_edu	A B C	upholding party autonomy and would not rewrite the contract to take account of subsequent concerns . # 6.9.4 Public policy For a
9	G2F	W_pop_lore	A B C	You will be notified if a longer delay is expected . Take advantage of low phone-in services : Access/visa cardholders
10	AD7	W_misc	A B C	must have felt horribly unbalanced he made no attempt to take advantage of this unwanted freedom and run away . Winter
11	BNG	W_non_ac_polit_law_edu	A B C	great deal of adult help at first , young children can take an active role in caring for equipment and tidying it away .
12	K8T	W_fict_prose	A B C	have a double fascination for Charlie , as he started to take an interest in the advertisements displayed on almost every
13	H8B	W_fict_prose	A B C	in the parish represents the bishop to the people . Priests take an bath of allegiance to the bishop at their induction into their
14	J75	W_ac_polit_law_edu	A B C	awards of compensation , relates the basic steps we will take and suggests a few of the things you can do to help
15	ABS	W_pop_lore	A B C	lady just shook her head and said : ' ! You must take another ticket to Then maybe I will call you . Now I
16	CR8	W_pop_lore	A B C	of Germany 's new policy has been winning Polish agreement to take back expelled asylum-seekers ; a quarter of Germany 's influx
17	KDM	S_conv	A B C	.(SP:PSORS) Yeah . (SP:PSO50) Alright love (pause) take care wo in you ? (SP:PSOPN) It 's clean Margaret !
18	KD1	S_conv	A B C	what 's wrong (SP:KD1PSUNK) (unclear) (SP:PS03E) take care (SP:PS03A) he might have a bottle when he get him home
19	CEE	W_biography	A B C	later to supervise journeys to and from school , or to take charge of home when both parents are out . Of course some
20	AYJ	W_commerce	A B C	rates are usually rather immature and may not be able to take criticism so whatever approach you adopt you might well

Figure 3.6 'KWIC' results for the sample query 'take'

After guiding the students towards the information possible to be extracted from the 'KWIC' option, the next step was to introduce the procedure for searching the synonyms. To provide a sample search, the students were asked to tick 'LIST' option. An adjective 'beautiful' was written in square brackets ([]) and preceded by an equality symbol (=) in the query bar. The corpus offered a list of the synonyms and the frequency information on their usage (see Figure 3.7).

DISPLAY [?]

LIST CHART KWIC COMPARE

SEARCH STRING [?]

WORD(S) [?]

COLLOCATES [?]

POS LIST [?]

SECTIONS SHOW [?]

1 IGNORE

SPOKEN

FICTION

MAGAZINE

NEWSPAPER

NON-ACAD

2 IGNORE

SPOKEN

FICTION

MAGAZINE

NEWSPAPER

NON-ACAD

Figure 3.7 Entrance of a sample query for synonymy investigation

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]				COMPARE	[?]	SIDE BY SIDE
	<input type="checkbox"/>	CONTEXT	FREQ			
1	<input type="checkbox"/>	BEAUTIFUL [S]	8307			
2	<input type="checkbox"/>	LOVELY [S]	6000			
3	<input type="checkbox"/>	ATTRACTIVE [S]	5000			
4	<input type="checkbox"/>	WONDERFUL [S]	4626			
5	<input type="checkbox"/>	STRIKING [S]	2540			
6	<input type="checkbox"/>	SUPERB [S]	2009			
7	<input type="checkbox"/>	MAGNIFICENT [S]	1948			
8	<input type="checkbox"/>	HANDSOME [S]	1570			
9	<input type="checkbox"/>	CHARMING [S]	1324			
10	<input type="checkbox"/>	DELIGHTFUL [S]	1054			
11	<input type="checkbox"/>	STUNNING [S]	914			
12	<input type="checkbox"/>	PICTURESQUE [S]	602			

Figure 3.8 Results for synonyms for word "beautiful"

As observed in Figure 3.8, the list of synonyms allows a deeper analysis of the lexical unit of interest. Depending on the students' interest and the linguistic necessity, clicking on a

word provided on the list encourages the examination of the concordance lines implying the lexical unit. Though, the understanding of the sense differences among the synonyms demands a definite language proficiency level in order to be able to form a proper collocation, the list of synonyms structures a hunch of lexical variety which, consequently, diminishes word repetitions and fosters lexical cohesion of a text.

Consequently, the students were introduced with the major aspects of interest in the particular stage of the research, i.e., keyword in context (KWIC) and synonymy. The rest of the seminar was devoted to students improving their comparison or contrast paragraph drafts by examining corpus from the perspective of the suggestions provided by the researcher. The whole process was observed by the researcher. In general, the corpus-based activity went without considerable disturbances, except for the computers occasionally being slow or freezing. The students were also instructed to highlight the places improved with the direct use of corpus in their paragraphs.

24 students submitted the paragraphs that were improved with the help of corpus evidence. Language issues approached in corpus happened to differ. The possible reasons are believed to be the different skills in the work with corpus engine, as well as individual working pace and time restraints. However, the majority of students had managed to mark the places in the paragraph improved after studying corpus material, thus aiding the selection of the works relevant for further analysis. The minimal number of corrections demanded to enlist the paragraph as suitable for the examination was 1.

Table 3.4 depicts the distribution of the linguistic features searched in the corpus and adapted to the students' own written texts:

Table 3.4 Distribution of linguistic features and number of paragraphs containing these features

Linguistic Feature	Number of paragraphs
Synonym(s)	14
Preposition(s)	3
Synonym(s) and preposition(s)	7
	Total: 24

The most frequently searched and used linguistic feature was synonym, as out of 24 instances of corrections 21 (14+7) dealt with the changes in word choice (see Table 3.4). Next, even though not of the primary concern, 10 (3+7) students had managed to operate with the corpus engine make changes in the use of prepositions. So, the aspect of grammatical cohesion also appears to be touched upon by the students during their paragraph improvement process.

Keeping with the word choices, namely synonymy, the following are the excerpts from students' drafts and the final versions (improved after the direct use of the BNC corpus) of the paragraphs:

Table 3.5 Samples of BNC-based synonymy improvements

Excerpt from the draft indicated to be improved	Excerpt from comparison/contrast paragraph final draft improved after direct use of the BNC
Obligatory dress code	Strict dress code
Every scholar can afford	Every student can afford
Make-up is used more	Make-up is worn more
You can do a lot of stuff	You can do a lot of activities
Trust and depend on	Trust and rely on
You can never touch or interact with them	[...] or communicate with them
Strongest sides	Best qualities
Manner of presentation	Manner of expression
In front of crowd	In front of spectators
Similar and dissimilar forms of art	Similar and diverse forms of art
A big part of film	A considerable part of film
It has become separate	It has become autonomous
There were many takes	There were many attempts

Examples included in Table 3.5 demonstrate how a preference to a more suitable word/synonym is capable of shaping contextually appropriate collocations, thus resulting in improvements in lexical cohesion of a text. The substituted lexical items (e.g., *dissimilar*, *crowd*, *depend*) represent a variety of word classes not being restricted or especially emphasizing a particular word class or part of speech. One of the most prominent and important aspects in the boundaries of the present research study is the students' preference of the vocabulary that tends to be of a higher formality level than that used in the draft, for instance, *crowd* → *spectators*; *separate* → *autonomous*; *dissimilar* → *diverse*, and other examples. Thus, the groups have indicated that the direct use of corpus information has functioned not only as a source for theoretical basis for enriching one's vocabulary knowledge but simultaneously it has offered an insight in the variety of formality levels of the language.

Apart from synonyms, 10 drafts contained the preposition corrections. These improvements have been made after studying of the concordance lines of 'KWIC' engine display. The table below covers several examples:

Table 3.6 Samples of BNC-based preposition improvements

Example from a draft	Example after corpus-based improvements
Differs a lot by price	Differs a lot in price
Pros and cons to being	Pros and cons of being
To get on trouble	To get into trouble
People in both age groups	People from both age groups
Actors in the screen	Actors on the screen

The students having made changes in the use of prepositions are to be praised for their approach to linguistic inquiry. On the one hand, occasionally the issue with prepositions is even more complicated to be solved by the analysis of the concordance lines, as the key word entered into the query box might also present the samples misleading in a particular context. To exemplify, the search for the correct preposition in a phrase ‘*people in both age groups*’ might position a student into an inconvenient situation as not being sure about the key word to be written in the search bar. Whether the input of several key words might be necessary for corpus to display a sample equal to the one desired as the correct version, or the appropriate preposition might partly depend on the ability to notice compatibilities in grammatically and/or lexically similar samples. Whatever the option adapted by the students, they have indicated a rather advanced approach to data analysis as the introductory stage of the corpus usage did not cover guidance of such a sophistication.

The fourth stage of the empirical research was dedicated to the researcher’s analysis of the semantic fields of the students’ comparison and contrast paragraphs – final drafts - by the application of **Wmatrix** corpus analysis tool (Online 2). The purpose of the stage was to study the semantic fields represented in the students’ written comparison or contrast paragraphs. The final drafts of the comparison and contrast paragraphs (have undergone the improvements by the direct use of the BNC) were collected and combined in a **learner corpus** (discussed in page 31).

To perform the analysis of the semantic categories represented in the students’ final paragraphs, each text had to be converted into plain text (*txt*) format for the tool to recognize it as a valid source for processing. Also, the final versions of the paragraphs had to be re-studied from the point of view of formatting, ensuring the absence of all the unnecessary symbols, comments, or any other features not accredited by the corpus analysis tool. Before uploading the texts into the Wmatrix, they were sorted according to the topics, i.e., *Childhood and Adulthood; Cinema and Theatre, Education in country X and country Y*, resulting in the formation of 3 separate folders/learner corpora created in the system after uploading all the plain text files for the examination. Such a division was a logical step to be undertaken as the

differences into topics would not permit the researcher to comment on the results as if applicable to the set of texts as a whole.

One of the wide amplitude of the procedures Wmatrix system performs is an automatic semantic tagging process of a corpus text. The tool offers 21 semantic categories each consisting of a range of semantic sub-categories, 231 in total. The said corpus analysis and comparison engine applies the UCREL Semantic Analysis System (USAS) (Online 3). The following is a list of the semantic categories recognized by Wmatrix:

- ‘General and abstract terms
- The body and the individual
- Arts and crafts
- Emotional actions, states and processes
- Food and farming
- Government and the public domain
- Architecture, buildings, houses and the home
- Money and commerce
- Entertainment, sports and games
- Life and living things
- Names and grammatical words’ (Online 3)
- Movement, location, travel and transport
- Numbers and Measurement
- Substances, materials, objects and equipment
- Education
- Linguistic actions, states and processes
- Time
- The world and our environment
- Psychological actions, states and processes
- Science and technology

Once a text is uploaded in the system, Wmatrix performs an automatic part-of-speech (POS), frequency, and semantic tagging analysis. However, the primary focus of the particular thesis was semantic domains and, secondly, their frequency. Then, the results are presented in as in Figure 3.9:

Word	Semtag	Frequency	Relative Frequency	
using	A1.5.1	1	0.65	Concordance
a_lot	A13.3	1	0.65	Concordance
quite	A13.5	1	0.65	Concordance
rather	A13.5	1	0.65	Concordance
at_least	A13.7	1	0.65	Concordance
especially	A14	1	0.65	Concordance
depends	A2.2	1	0.65	Concordance
are	A3+	1	0.65	Concordance
is	A3+	3	1.95	Concordance
specific	A4.2+	1	0.65	Concordance
advantages	A5.1+	1	0.65	Concordance
disadvantages	A5.1-	1	0.65	Concordance
different	A6.1-	1	0.65	Concordance
differs	A6.1-	1	0.65	Concordance
differ	A6.1-	2	1.30	Concordance
usually	A6.2+	1	0.65	Concordance
variety	A6.3+	1	0.65	Concordance
can	A7+	2	1.30	Concordance
according	A9-	1	0.65	Concordance
dress	B5	1	0.65	Concordance
evening_dresses	B5	1	0.65	Concordance

Summary information:		
Number of types shown: 104		
Total frequency of types shown: 154 (100.00%)		
Total frequency overall: 154		
Number of items shown with a given frequency:		
Frequency	Types	Tokens
1	78 (75.00%)	78 (50.65%)
2	16 (15.38%)	32 (20.78%)
3	4 (3.85%)	12 (7.79%)
4	2 (1.92%)	8 (5.19%)
5	1 (0.96%)	5 (3.25%)
6	2 (1.92%)	12 (7.79%)
7	1 (0.96%)	7 (4.55%)
8	(0.00%)	(0.00%)
9	(0.00%)	(0.00%)
10	(0.00%)	(0.00%)
> 10	0 (0.00%)	0 (0.00%)

Figure 3.9 Sample output of Wmatrix analysis of semantic domains

The picture in Figure 3.9 is used only to illustrate the manner of data representation. The data comprises the lexical units of the text, each followed by the semantic category it belongs to according to the USAS tagging system, as well as the frequency of the word's appearance of the text. Even though, as mentioned previously, the aspect of frequency is of the secondary importance, yet in several instances it was worth paying attention to, when, for instance, a lexical unit has been repeated in one paragraph several times.

Consequently, the type of analysis depicted above was performed to all the final draft comparison or contrast paragraphs, in total 24. Next, the data output for all the texts was researched, marking the semantic domains represented and their distribution. While a complete table revealing a full and detailed analysis of the general semantic fields and their sub-categories can be found in Appendix 8, a table in Appendix 9 shows the results on the type of the paragraphs (i.e., according to the topic of the paragraphs: *Childhood and Adulthood; Cinema and Theatre; Education in Country X and Country Y*) and the usage of the semantic fields in a hierarchical order based on the frequency counts (the numbers in the brackets denote the instances when the semantic domain has been observed, NOT the number of cases when a lexical unit tagged to belong to the particular semantic domain has been noticed in a text being analysed). The results generated by Wmatrix corpus analysis tool and summarized by the researcher are to be described from the perspectives of texts' content and function, as the semantic tagsets proposed by the engine permits it. Yet, in order to support the following discussion with a textual example, Example 3 is a sample final draft of a paragraph:

Example 3:

There are many pros and cons of being a child and being an adult, like power, the lack of power and responsibilities. First, when one is a child, they do not have any real responsibilities. They only have to play, learn about the surroundings and try not to get into trouble, but they do not have that much control over their time table – the parents usually decide for them. Adults, in contrast, have many responsibilities, including the children – they need to take care of them, give them a place to live and make sure, that the children actually avoid trouble, but adults are the ones who hold the power. Another major difference between a child and an adult is that children do not know that they know little, but adults actually know they have a lot to earn. In conclusion, neither – the childhood nor the adulthood – is better than the other. All adults wish they could be teenagers again, but all children want to grow up as soon as possible.

The underlined words in Example 3 denote different periods of life and some of the characteristic features typical of those periods. From the perspective of content of the paragraphs, the statistics for the set of paragraphs (including Example 3) under the topic “Childhood and Adulthood” indicate an absolute dominance of the lexical terms that express general actions or states, as well as abstractions. It is not surprising as the category appears to cover a wide range of various sub-categories pronouncing universal states and processes, for instance, existence, changing of concepts, degrees boosting or diminishing different qualities, and others. As the title of the semantic domain suggests, it is necessarily to use words of this category, especially remembering the topic of the paragraph (*Childhood and Adulthood*) and the purpose of the paragraph (comparison or contrast). The said periods of life each have its own specifications and characteristics, which, when performing a more scrupulous examination, are explicitly expressed by the use of 3 semantic sub-categories: general actions, making; existing; getting a possession. Recalling the contextual nature of the students’ paragraphs, the application of these categories makes sense. For instance, one of the perspectives a range of students have acquired to illuminate the similarities or differences between children and adults were to write about duties and possessions. The Wmatrix engine, then, has affirmed that the texts imply words falling into the semantic category of possessions (see Appendix 8).

The second most widely applied set of words among the same paragraphs belongs to social phenomena. The sphere of social factors is dominated by the lexical units falling into the groups of people, kinship, obligations and power relation between the members of different social groups. Drawing parallels with the semantic domains for the expression of general states and abstractions, the students have given a priority to clearly indicating a borderline of being a child and adult with all the consequences following. Example 3 and the underlined words above highlight it (e.g., *make sure*, *learn*, *responsibilities*, and others).

Next, a rather surprising observation was the semantic group of numbers and measurement taking the 3rd place in the list, as the concept of the 4th place – time – would seem to be a more logical choice due to the fact that two different time spans of a human life are to be commented on. However, the measurement stands higher in the hierarchy mostly due to one reason – the students have made use of such indicators of quantity as “much” and “many”. The application of this type of adverbs indicate the students’ urge to stay rigid in describing the concepts offered in the topic of the paragraph – the lexical choices are believed to strengthen and support the reliability of the content of the text, presumably elevating students’ competence in the subject area. Consequently, the points described so far lead towards a finding that, yet not paying attention to the words pointing towards the function of the texts, a typical feature of the learners is to apply the semantic fields clearly categorizing the two subjects of the paragraph as individual entities.

The semantic category being the 5th most frequent – psychological actions, states, and processes – affirms the presence of lexical units mirroring the mental conditions typical of the members of the said age groups. Then, the results show that students have approached the topic not only from the physiological and explicitly rational point of view, but also have implied vocabulary for psychological processes.

Keeping with the content-wise study, not all of the paragraphs under the topic “Childhood and Adulthood” were lexically cohesive. At the very first stages of the composition of the text, the students were particularly instructed to maintain the precision in the topic sentence of the paragraph, stating the function, as well as the specific aspects to be compared or contrasted in the text. It was believed to ensure students’ concentration. Even though, every aspect can be approached from various perspectives, there exist some limitations in thematic and, respectively, linguistic choices. In a range of paragraphs, these limitations did not appear to be considered, or were partly thought over. Referring back to one of the paragraphs written by a student from Group1, the topic sentence is as follows:

Although, childhood and adulthood are age-wise pretty close compared to the length of the rest of the life, there are great differences in, for example, responsibilities and independence.

From this sentence, the following points are to be written about are: (1) responsibilities (2) independence. Following the USAS semantic tagset, the domains of responsibilities (renamed as *obligation and necessity* in the tagset, see point S6 in Appendix 8), as well as independence falls into the category of “Social actions, states, and processes” (ibid.). The paragraph would be expected to be rich in lexical units of these semantic meanings. According to the analysis of the particular paragraph performed by Wmatrix, only

2 words – *need* and *responsibilities* - are recognized as belonging to the semantic field. In addition, the term *responsibilities* happens to be repeated twice in the paragraph, from which one instance being used in the topic sentence would not count exactly as a statistical unit worth analysing. The table 3.8 below portrays the situation in terms of lexical cohesion (content-wise) in the 8 paragraphs written under the topic “Childhood and adulthood”. It demonstrates the points the students have included in the topic sentences as the focus of the text, supported by the information on the dominant semantic fields on the basis of USAS tagset and the paragraphs. It is important to recall and remember the fact that the semantic field labelled as “General and abstract terms” is still dominant in all of the paragraphs analysed within the scope of the present research. Thus, even not included in the following table, still it stands in front of all the rest implied.

Table 3.8 Conformity between topic and semantic domains in “Childhood and Adulthood” texts

Nr.	Points indicated in the topic sentence	Classification of the points according to the USAS tagset	Dominant semantic fields in the paragraph
1.	Relationships, ownership, colour of preference	1. Social actions, states, processes; 2. Substances, materials, objects, equipment	1. Time; 2. Psychological actions, states, processes 3. Social actions, states, processes
2.	Power, lack of power, responsibilities	1. Social actions, states, processes	1. Social actions, states, processes; 2. Psychological actions, states, processes
3.	Responsibilities, independence	1. Social actions, states, processes	1. Numbers and measurement/ Social actions, states, processes
4.	Enjoyment, fears	Emotional actions, states, processes	1. Emotional actions, states, processes; 2. Time
5.	Responsibility, work, self-determination	1. Social actions, states, processes; 2. Money, commerce	1. Social actions, states, processes 2. Numbers, measurement
6.	Responsibility, acting	Social actions, states, processes	1. Social actions, states, processes; 2. Psychological actions, states, processes
7.	Responsibilities, duties, rights	Social actions, states, processes	1. Numbers, measurement; 2. Social actions, states, processes; 3. Time
8.	Age, way of thinking, actions	1. Time; 2. Psychological actions, states, processes; 3. Social actions, states, processes	1. Time; 2. Social actions, states, processes; Psychological actions, states, processes

Table 3.8 affirms that the distribution of the lexical items belonging to the semantic field of social actions, states and processes is also the one most widely applied in the texts. Yet, some paragraphs, like paragraph 1 and 3, indicate that the student has drifted away from the focus of the paragraph, as it happens to be governed by vocabulary from inappropriate semantic fields.

A more scrupulous attention appeared to be worth paying to the distribution of the sub-categories of the USAS tagsets from “Social actions, states, and processes”. Despite it really being the semantic domain which lexical items support the points expressed in the topic sentences, often the dominance of the field is achieved by the use of words belonging to other sub-categories. Here, the cause for the phenomenon is a thematic-wide spectrum of semantic fields forming a greater group. For example, in paragraph 6, out of 9 words ensuring the semantic field of social affairs to be the most prominent one (after the domain of “General and abstract terms”), only 3 lexical items actually have the semantic meaning shared with the issues mentioned as the focus of the paragraph in the topic sentence. The rest of the instances cover different aspects of social states that, no doubt, have an influence on the text as such, but they do not promote keeping the text focused. Consequently, one of the major findings the Wmatrix corpus analysis has revealed is gaps in maintaining lexical cohesion of the text based on the central points claimed to be described.

Similar analysis as the one just described could have been performed for all the 21 semantic groups offered by USAS tagset. Nevertheless, it was the intention of the researcher to pay particular attention to the most frequently used ones in order to get a general idea about the conformity between the subjects of the comparison or contrast paragraph and the contextually appropriate semantic fields applicable for their characterization. In addition, the semantic domains standing in the end of the list (see Table 3.6) were not considered a fruitful material for a deeper analysis as their low distribution does not portray the common trends in the students’ lexical choices. Rather, they affirm the overall richness of the semantic domains capable of being differentiated as such (for more details, see Appendix 8).

The second major group of paragraphs (13) were written under the topic “Cinema and Theatre”, for which an identical analysis was performed. In comparison to the texts described above, the subjects, i.e., cinema and theatre, have appeared to impose more visible restrictions in the basic semantic fields to be applied for the composition. Interestingly, almost all of the students have decided to write about the same aspects of both forms of art, only with slight deviations. Thereof, the table to follow is arranged in slightly different format, where all the

13 paragraphs are not listed separately. Instead, the facets the students have pointed out in the topic sentence are followed by the most widespread semantic fields.

Table 3.9 Conformity between topic and semantic domains in “Cinema and Theatre” texts

Points indicated in the topic sentences	Classification of the points according to the USAS tagset	Dominant semantic fields in the paragraphs
<ul style="list-style-type: none"> • Production • Performance • Audience • Price • Etiquette • Genre • Emotional state • Entertainment • Origins 	<ul style="list-style-type: none"> • Linguistic actions, states, processes • Arts and crafts • Emotional actions, states, processes • Social actions, states, processes • The body and the individual • Money and commerce • Psychological actions, states, processes • Entertainment, sports, games 	<ul style="list-style-type: none"> • Linguistic actions, states, processes • Entertainment, sports, games • Psychological actions, states, processes • Time • Social actions, states, processes • Movement, location, transport • Emotional actions, states, processes

According to the obtained data (see Table 3.9), a partial congruence emerges between the USAS semantic tagset and the distribution of semantic fields in students’ texts, where the categories of “Time” and “Movement, location, travel, and transport” happen to be applied quite often. However, this habit again could be explained by the notable range of sub-categories for each group.

When comparing the two thematic sets of paragraphs (*Childhood and Adulthood – Cinema and Theatre*), each marks different areas for consideration in terms of lexical cohesion. When Wmatrix analysis for *Childhood and Adulthood* texts directed towards a problem in the distribution in semantic sub-categories building the general vocabulary framework of a text, the compositions for *Cinema and Theatre* amazed the researcher by the vagueness in lexical choices exactly in the topic sentence. Some of the paragraphs undergoing the study demanded a scrupulous reading of the texts, as the naming of the aspects implied in the topic sentence initially appeared to be too abstract to be classified under any of the semantic fields. Of course, any lexical item belongs to one or another lexical domain, but the context of the final draft occasionally alerted the investigator to double-check the data gathered from Wmatrix analysis. For instance, the topic sentence below is a typical sample of the issue:

‘Cinema and theatre differ in specific features, such as: atmosphere, public, and price’.

So, as Wmatrix is an automatic program based on artificial intelligence principle, the word *atmosphere* falls into the semantic category entitled “The world and our environment”. But the continuation of the paragraph is as follows:

‘The atmosphere in theatre is rather formal. Women are wearing luxurious dresses and men prefer suits. People try to communicate politely, without any slang words’.

So, it turns out that the term *atmosphere* found in the topic sentence actually refers to the *etiquette* in theatre. Thus, it was the duty of the researcher to mark these imperfections and find a more appropriate semantic field for the words.

Similarly, *public* would be better paraphrased as *audience*, even though both variants would remind under one and the same semantic category.

Finally, 3 students had decided to write about the educational systems in two countries of their choice. Without doubt, the USAS semantic fields to govern the texts would be (1) Education (2) Social actions, states, processes (3) Psychological actions, states, processes (as the category implies sub-domains of learning and knowledge) (4) Names and Grammatical words (as the category implies sub-domains of Geographical names and Proper names). These paragraphs, maybe because of the small number of students having chosen the particular topic, did not indicate any peculiar features to be subjected for a very detailed commenting. The distribution of the semantic domains offered by Wmatrix was in complete accordance with the ones applied in the paragraphs (see Table 3.6).

To summarize with lexical cohesion from content-based perspective, the data gathered after analysing comparison or contrast paragraph final drafts by Wmatrix corpus analysis and comparison tool have brought forward the following conclusions. The paragraphs appear lexically coherent in spite of the simplicity and occasional abstractions of the vocabulary. Nevertheless, a more scrupulous investigation of the semantic sub-categories provided in USAS tagset would be suggested as a future research for a very significant reason: the vast amplitude of semantic groups building a bigger semantic set might be misleading when the central peculiarity of the language teacher or investigator is the interaction between the topic sentence and the supporting text (especially in relatively short texts, such as paragraphs). The paragraphs studied in detail could have an effect upon the learners in regards to their academic writing process, but the particular procedure and findings are believed to be more teacher-centred as the absorption of the information demands definite competencies, such as relative proficiency on working with the Wmatrix engine.

In the period of the present study the concept of lexical cohesion was approached also for the perspective of a text’s function. Often the maintenance of the functionally stable

academic text is a struggle for the students. The academic environment is a formal environment, where the capability of analytical thinking and precision is one of the major components both in content and function of a text.

To perform the function-based final draft examination, a thematic division of the paragraphs was not necessary, as the function of comparing or contrasting was made clear at the very first steps of the piloting lessons. The findings are based on the Wmatrix engine output.

Out of 24 paragraphs, the function of 3 was comparison, while the very majority – 21 texts were aimed at contrasting. Following the semantic tagging system accredited by Wmatrix, the lexis for the said language functions belong the category “General and abstract terms”, sub-fields from A6 to A6.3- (see Appendix 8). The distribution of the semantic fields in the paragraphs is shown in Table 3.10:

Table 3.10 Conformity between function of text and semantic distribution

Language function (number of paragraphs)	Number of lexical units falling into the corresponding semantic field
Comparison (3)	32
Contrast (21)	60

So far, the results depicted in Table 3.10 would be comfortable for affirmation that the paragraphs were lexically coherent also from the functional matrix. But a more attentive examination of the statistics revealed the inevitable cause-effect relationship between the function of the text (here, it would be compared with the cause) and the lexical choices to reach that function (element of effect). To remind, the proportion between the comparison and contrast paragraphs is 3:21, while the correlation between distributions of the lexis is 32:60. Consequently, regardless of the comparison paragraphs being 7 times less frequent than contrast texts, they were richer in vocabulary holding the semantic force of comparison. The Wmatrix analysis affirmed the fact that for a text to be lexically cohesive, the function of that text might be one of the leading factors defining the success or failure of the writer. On the other hand, it can also be a matter of student’s language proficiency level. Yet, it should not be acknowledged as a universal truth as every sociolinguistic situation is unique.

Keeping with the function of a text, this was the stage of the research where the repetition of one lexical unit present in the paragraph became a crucial aspect of analysis. Despite the word repetitions not being important so far in the run of Wmatrix corpus analysis, the frequency of separate lexical items set the ground for the investigation of lexical cohesion in relation to the text’s function. The frequency observation was performed to gain a general

view on the variety of the vocabulary applied to express the function of the text. Here, it is extremely vital not to mix or relate the results in Table 3.10 with the ones to come – the following results are based on the fact that they do not count the repetition of one lexical unit, but summarize different words regardless of their frequency in a particular paragraph.

So, this frequency analysis indicated that the word repetitions vary among the paragraphs. There is a group of students capable of diversifying lexical choices, while still the majority of the learners prefer repetition of a small set of words. Nevertheless, the function of the paragraph still remains correct. Here is a couple of the textual samples:

Student1: Function: contrast. Number of semantically appropriate lexical items: 2 (*differences, different*). Without the fact that both are one-stem words, *different* is repeated twice in the text.

Student2: Function: contrast: Number of semantically appropriate lexical items: 7 (*parallel, antithetical, contrary, different, differs, divided, instead*). To add, none of the words is repeated.

To go deeper, there are two more points to briefly touch upon. Firstly, the analysis has affirmed that the students undergoing the research are capable of staying lexically cohesive to fulfil the function of the text. At this stage, a crucial role happens to be played by a grammatical concept – in almost every paragraph out of 24 examined, the majority of the words labelled as indicating comparison or contrast (according to USAS tagset) is a part of a grammatical linker or conjunction, such as *in contrast, instead, also, however, but, one the one hand, on the other hand*, and others. As mentioned previously in the scope of the present study, the paragraphs often resemble a description of the concepts. It is not a surprise referring to Halliday and Hasan (1976) who have acknowledged the very fragile borderline between grammar and lexis. Also, as the learners were the first year students, it would not be appropriate to expect texts highly diverse in lexis.

Secondly, the comparative or contrastive nature is strengthened by the application of synonyms or antonyms, or words of semantic domains declaring oppositions. Especially noticeable it is on the texts aimed at contrasting. So, in one of the contrasting paragraphs a student has enumerated differences between childhood and adulthood, using such pairs of words as: *joy – working; impulses – training; no limitations – acceptable*, and others. However, it would be a point worth investigation in further investigations to come.

To finalize the research on lexical cohesion in academic writing instruction described in the present thesis, a learner corpus has been compiled as a reference source for the present study, as well as for scholars interested in the topic or willing to continue the research in the

particular field. A part of the learner corpus (pre-texts and comparison or contrast paragraph final versions) is available at Wmatrix corpus analysis and comparison tool, approachable by registration. The texts are available in plain text (*txt*) format. In addition, other information except for the semantic field distribution is available, for instance, part of speech tagging, their frequency, and other aspects of interest. The learner corpus is attached in the Appendix 10 of the present thesis as well. The data output files can be saved on personal storage devices printed out for various purposes, if necessary. Also, the data analysis can be adjusted to different proficiency and complexity levels.

To conclude, the present empirical research has affirmed the usefulness and effectiveness of corpus application in academic writing instruction, lexical aspects in particular. It aimed at supporting corpus-based activities as beneficial for both language learners and teachers. The indirect use of corpus was applied for teaching material development, whereas direct use of the BNC was practiced by the students for their own text editing. In both cases the corpus-based instruction of lexical cohesion can be affirmed as having left a positive effect on the students' ability to compose lexically cohesive texts. The AntConc concordancer and Wmatrix corpus analysis and comparison tool revealed significant information about the extent to which the research subjects were capable of composing texts appropriate in collocational force and contextually appropriate semantic fields.

Though, being only a small step into the field of corpus-based language teaching and learning, the findings of the present research are believed to be a successful starting material for further investigations in the field of academic writing and lexical cohesion.

CONCLUSIONS

The goal of the present thesis was to introduce learner-centred (corpus-based) and teacher-centred (research of learner corpus) corpus-related activity design facilitating the composition of lexically cohesive paragraphs and compile a learner corpus.

Despite the application of corpora being a relatively new, yet successful trend in linguistic studies, the analysis of secondary sources has indicated lack of corpus-based studies focused exactly on the aspect of lexical cohesion in language teaching. Nevertheless, the various aspects of how a corpus can be implemented within language teaching and learning bring out the importance of direct, indirect and learner corpus implementation for language proficiency development. The study of authentic texts is a means of promoting the language learners to approach the study material more independently and generate knowledge inductively. The language regularities and their frequency characterizes text corpus as a useful tool for studying various linguistic features.

Emphasizing academic writing instruction, corpus-based studies, especially concordancing and study of authentic texts, are capable of extracting characteristics typical of academic writing and the various genres it comprises. To link corpus-based language research and study of academic writing, both the theory and empirical study revealed that corpus analysis can be successfully implemented as a writing activity step facilitating students self-editing their written texts.

The empirical research indicated that the development and practice of corpus-based activities has revealed the importance of lexical cohesion in academic writing. The learner-centred materials and activities were successfully implemented within the study course devoted to academic writing, resulting in positive effect on the students' creation of lexically cohesive paragraphs, ranging from the reception to production skills.

In respect to teacher-centred corpus analysis, the empirical research reflecting about the distribution of semantic fields in students' comparison and contrast paragraphs has brought out a conclusion worth consideration for further studies: despite the paragraphs being generally lexically cohesive, often the focal points of the paragraph are not supported by the lexical items belonging exactly to the contextually most appropriate semantic fields. Such gaps in the use of words belonging to the most appropriate semantic fields as a means for achieving a lexically cohesive text are related also with synonymy and, respectively, collocations. Even if two words are considered to be synonyms, they cannot be used interchangeably without considering the situations and concepts that each of the synonymous word typically describes.

The present research has brought forward the linguistic means by which the students demonstrate the function of their composed texts. The preference has been given to the linking devices expressing comparison or contrast, or a very narrow set of words classified as fulfilling the said functions. The variety of lexically cohesive devices for comparing and contrasting is still a field for development.

Next, the present research has only very briefly mentioned the role of antonyms and semantically opposite lexical words applicable for fulfilling the function of comparing or contrasting. Thus, a corpus-based analysis of these also appears to be a promising research area for further investigations.

The analysis of theories and the results from the empirical research shows that corpus-based activities can create a positive attitude towards applying corpus for investigating various language disciplines, for instance, grammar.

The theoretical analysis and empirical research showed that the subject area investigated is an appreciative field for further research. Firstly, a more extended introductory phase of the corpora in actual learning settings could lead towards language learners of high linguistic awareness and inductive approach towards studies. Longer period of time ascribed for students getting familiar with the corpora could widen their capability of operating with different corpora, as the operations and types of data possible to be extracted among various online corpora also happen to differ. Thus, the technological competence could even be adjusted to the linguistic problem eager to be solved.

The congruence between the quantitative and qualitative data assures about the advantageous nature of corpus-based teaching materials and activities resulting in texts lexically more precise.

THESES

1. The present research is an investigation aimed at promoting corpus-based teaching and learning of lexical cohesion.
2. The aspect of lexical cohesion in academic writing is vital both content-wise and function-wise.
3. Lexical cohesion is one of the major aspects of concern in academic writing, as its attainment appears to cause difficulties for students in terms of contextually suitable words (synonyms) and semantic fields.
4. Corpora and the type of information possible to be extracted from them for educational purposes are applicable directly and indirectly, depending on the students' needs and language instructors' preferences.
5. The use of corpus for teaching and learning one specific language aspect has the potential of being recognized as a useful tool also for the acquisition of different language area(s).
6. Corpus-based language study promotes independent, inductive learning.
7. A proper time span of instructions is required to become successful corpus-based language students and researchers.
8. One and the same corpus-based activities can result in various aspects of language to be noticed among different groups of learners.
9. Corpus-based analysis of synonymy and concordance lines has facilitated students' ability to edit their own written texts in order to improve their collocational force.
10. Computer-based analysis of students' comparison and contrast paragraphs from the perspective of semantic fields is a successful mean of determining the traits of lexical choices and their overall influence on lexical cohesion of a text.
11. Learner corpus is applicable as a starting point for a scrupulous and large scale research.
12. Further corpus-based investigations on lexical cohesion could concern a deeper analysis of the use of personal pronouns and their referential relationship in students written texts.

REFERENCES

1. Aijmer, K. (ed.), (2009) *Corpora and Language Teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
2. Aijmer, K. and Altenberg, B. (eds.), (1991) *English Corpus Linguistics*. London: Longman.
3. Baker, P. (ed.), (2009) *Contemporary Corpus Linguistics*. London: Continuum International Publishing Group.
4. Balossi, G. (2014) *A Corpus Linguistics Approach to Literary Language and Characterization*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
5. Biber, D. (2006) *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
6. Biber, D. (2010) What can corpus tell us about registers and genres? In A. O’Keeffe and M. McCarthy (eds.) *Routledge Handbook of Corpus Linguistics* (pp. 241-255). Abingdon and New York: Routledge.
7. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Essex: Longman.
8. Brezina, V., McEnery, T., Wattam, S. (2015) Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20 (2): 139-173.
9. Bordwell, D. (1989) *Making Meaning*. Cambridge and London: Harvard University Press.
10. Chambers, A. (2010) What is data-driven learning? In A. O’Keeffe and M. McCarthy (eds.) *Routledge Handbook of Corpus Linguistics* (pp. 345-359). Abingdon and New York: Routledge.
11. Cheng, W. (2010) What can a corpus tell us about language teaching? In A. O’Keeffe and M. McCarthy (eds.) *Routledge Handbook of Corpus Linguistics* (pp. 319-332). Abingdon and New York: Routledge.
12. Cheng, W. (2012) *Exploring Corpus Linguistics: Language in Action*. Oxon: Routledge.
13. Day, T. (2013) *Success in Academic Writing*. Hampshire and New York: Palgrave and Macmillan.

14. Etherington, S. (2008) Academic writing and the disciplines. In P. Friedrich (ed.) *Teaching Academic Writing* (pp. 27-58). London: Continuum.
15. Fellbaum, Ch. (2007) *Idioms and Collocations*. London: Continuum.
16. Fiderer, A. (2002) *Paragraph Power*. New York: Scholastic Professional Books.
17. Flowerdew, J. and Mahlberg, M. (eds.), (2009) *Lexical Cohesion and Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
18. Flowerdew, L. (2010) Using corpora for writing instruction. In A. O’Keeffe and M. McCarthy (eds.) *Routledge Handbook of Corpus Linguistics* (pp. 444-458). Abingdon and New York: Routledge.
19. Frankenberg-Garcia, A. (2012) Integrating corpora with everyday language teaching. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 36-54). Brno: Masaryk University Press.
20. Friedrich, P. (ed.), (2008) *Teaching Academic Writing*. London: Continuum.
21. Geist, M. and Hahn, A. (2012) Using a corpus for written production: A classroom study. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 123-136). Brno: Masaryk University Press.
22. Götz, S. (2012) Temporal fluency variables in native and non-native English speech: Corpus findings and pedagogical implications. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 229-244). Brno: Masaryk University Press.
23. Granger, S., Hung, J., Petch, S., Petch-Tyson, S. (eds.), (2002) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
24. Griffee, D. T. (2012) *An Introduction to Second Language Research Methods: Design and Data*. Berkeley: TESL-EJ Publications.
25. Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London and New York: Routledge.
26. Hoey, M., Mahlberg, M., Stubbs, M., Teubert, W. (2007) *Text, Discourse and Corpora. Theory and Analysis*. London and New York: Continuum.
27. Howe, B. M. and Willoughby, H. A. (2005) *Introduction to Academic English Writing*. Korea: Ewha Womans University Press.
28. Johns, T. (1991) Should you be persuaded: Two samples of data-driven learning materials. T. Johns and P. King (eds.) *Classroom Concordancing: ELR Journal*, (4): 1-16.

29. Jordan, R. R. (1997) *English for Academic Purposes: A Guide and Resource Book for Teachers*. Cambridge: Cambridge University Press.
30. Klebanov, B. B., Diermeier, D., Beigman, E. (2008) Automatic detection of semantic fields for political science research. *Journal of Information Technology and Politics*, 5 (1): 95-120.
31. Mattheoudakis, M. and Hatzitheodorou, A. (2012) The lexical patterning of light verbs in GRICLE and native corpora: A comparative corpus-based study. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 213-229). Brno: Masaryk University Press.
32. McEnery, T. and Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
33. McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies*. London and New York: Routledge.
34. Nunan, D. (1992) *Research methods in language learning*. Cambridge: Cambridge University Press.
35. McCarten, J. (2010) Corpus-informed course book design. In O’Keeffe, A. and McCarthy, M. (eds.) *Routledge Handbook of Corpus Linguistics* (pp. 413-428). Abingdon and New York: Routledge.
36. Ooi, V B. Y., Tan, P. K. W., Chiang, A. K. L. (2007) Analyzing personal weblogs in Singapore English: the Wmatrix approach. *Towards Multimedia in Corpus Studies*, 2nd ed. Available from http://www.helsinki.fi/varieng/series/volumes/02/ooi_et_al/
37. Osmond, A. (2015) *Academic Writing and Grammar for Students*, 2nd ed. London: SAGE Publications, Ltd.
38. Park, K. and Lu, X. (2015) Automatic analysis of thematic structure in written English. *International Journal of Corpus Linguistics*, 20 (1): 81-101.
39. Quinn, C. (2015) Training L2 Writers to Reference Corpora as a Self-Correction Tool. *ELT Journal*, 69 (2): 165-177.
40. Rogatcheva, S. (2012) Measuring learner (mis)use: Tense and aspect errors in the Bulgarian and German components of ICLE. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 258-273). Brno: Masaryk University Press.
41. Römer, O. (2012) Corpora and teaching academic writing: Exploring the pedagogical potential of MICUSP. In J. Thomas and A. Boulton (eds.) *Input, Process and Product*.

- Developments in Teaching and Language Corpora* (pp. 70-82). Brno: Masaryk University Press.
42. Sanders, J. (2015) *5 Reasons Students Want Technology in the Classroom*. Available from <https://www.gaggle.net/speaks/5-reasons-students-want-technology-in-the-classroom/> [Accessed on 9 May 2016].
 43. Sanders, T. and Maat, H. P. (2006) *Cohesion and Coherence: Linguistic Approaches*. Elsevier, Ltd.: The Netherlands.
 44. Sinclair, J. McH. (1991) *Corpus Concordance Collocation*. Oxford: Oxford University Press.
 45. Sinclair, J. McH. (2004) *How to Use Corpora in Language Teaching*. Amsterdam and Philadelphia: John Benjamins B.V.
 46. Talai, T. and Fotovatnia, Z. (2012) Data-driven Learning: A student-centered technique for language learning. *Theory and Practice in Language Studies*, 2 (7): 1526-1531.
 47. Thurstun, J. and Candlin, C. (1998) Concordancing and the teaching of the vocabulary of academic English. *English for Special Purposes*, (17): 267-280.
 48. Tognini - Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Company.
 49. Tribble, Ch. (2002) Corpora and corpus analysis: new window on academic writing. In J. Flowerdew (ed) *Academic Discourse* (pp. 131-149). London and New York: Routledge.
 50. Twardo, S. (2012) Selected errors in the use of verbs by adult learners of English at B1, B2, and C1 levels. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 273-282). Brno: Masaryk University Press.
 51. Xu, J., Morgan, M., McKenny, J. (2012) Chinese learners' use of formulaic sequences in spoken interaction. In J. Thomas and A. Boulton (eds.) *Input, Process and Product. Developments in Teaching and Language Corpora* (pp. 244-258). Brno: Masaryk University Press.
 52. Yoon, H. (2008) More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning and Technology*, 12 (2): 31-48.

Sources for teaching materials

1. Boardman, C. A. and Frydenberg, J. (2008) *Writing to Communicate. Paragraphs and Essays*, 3rd edition. New York: Pearson.
2. Hope, B. and Martin, P. (1998) A century of hope. In *Saturday Evening Post*, 270 (2): 32-70.
3. Zemach, D. E. and Rumisek, L. A. (2006) *Academic Writing: From Paragraph to Essay*. London: Macmillan.
4. Oshima, A. and Hogue, A. (2008) *Writing Academic English*, 4th ed. New York: Pearson.
5. Slocum, J. D. (2000) Film violence and the Institutionalization of the cinema. *Social Research*, 67 (3): 649-681.
6. Teachout, T. (2009) The trouble with Alfred Hitchcock. *Commentary*, 127 (2): 43-46.
7. Lahav, D. (n.d.) *Business English. Paragraph structure. Comparing and Contrasting*. Ruppin Academic Center. Available from <http://www.slideshare.net/debbielahav/paragraph-structure-compare-contrast> [Accessed on 20 November 2015].
8. <http://english120.pbworks.com/w/page/19006833/contrast%20and%20comparison%20paragraph> [Accessed on 20 November 2015].
9. <http://lrs.ed.uiuc.edu/students/fwalters/compcnt.html> [Accessed on 20 November 2015].
10. http://www.slideshare.net/abdullahkhalifah/example-of-compare-contrast-paragraphs1?next_slideshow=1 [Accessed on 20 November 2015].

Internet sources

- 1) *British National Corpus*. Available from <http://corpus.byu.edu/bnc/> [Accessed on 12 November 2015].
- 2) *Wmatrix corpus analysis and comparison tool*. Available from <http://ucrel.lancs.ac.uk/wmatrix/> [Accessed on 20 March 2016].
- 3) UCREL Semantic Analysis System (USAS). Available from <http://ucrel.lancs.ac.uk/usas/> [Accessed on 20 March 2016].
- 4) *Why Integrate Technology into the Curriculum?: The Reasons are Many*. Available from <http://www.edutopia.org/technology-integration-introduction> [Accessed on 9 May 2016].

Appendix 1

Lesson plan 1: Recognition of grammar and lexis among registers

Date: November 5th and 10th, 2015

Location: University of Latvia, Faculty of Humanities

Topic of the paragraphs used for classroom work (according to the spoken part of the course): **Cinema**

Lesson Plan Components	Topic of the lecture: <i>Formal v. informal writing. Register. Paragraph structure. Topic sentence.</i> * *Register, topic sentence, paragraph structure – taught by the professor.			
Level of students	1 st year English Philology BSP students, groups 1A and 1B			
Timetable fit	Students have acquired punctuation, sentence structure, and differences in American and British English.			
Main aim(s)	1. Introduce the grammatical and lexical characteristics of formal and informal writing. 2. Practice analytical skill/critical text analysis – recognition of grammatical and lexical characteristics in authentic texts (incorporates the reading skill).			
Objectives	1. Encourage the students in discussing theory (eliciting answers). 2. Promote decision-making skill by analysing texts.			
Assumptions	1. Students might feel unmotivated to study; 2. The materials(texts) used for the activity might not cover the topic interesting to the students; 3. The materials (texts) used for the activity might appear to be too complicated for their comprehension; 4. Unknown vocabulary might appear.			
Anticipated problem	The time ascribed for the explanation of theory and practical exercise might deviate from the expected.			
Possible solution	Teacher's assistance has to be provided.			
Timing (40 min)	<i>Procedure</i>	<i>Stage Aims</i>	<i>Aids and Materials</i>	<i>Interactional Pattern</i>
2 min.: Lead-in activity	Greetings. Brief activation of schemata. Introduction of the topic.	Establishing relationships with the students.	No specific materials. Teacher's speech dominates.	Teacher – students.
15 min.	Introduction of the	1. Provide theory	Handout on	Teacher –

	theory: <i>grammatical and lexical characteristics of formal and informal writing</i> (see Appendix 1). Clarifying the unknown vocabulary, concepts.	necessary for the accomplishment of the practical task. 2. Encourage the students to speak (students might learn from one another).	the theory discussed (found in Moodle).	students (when clarifying concepts, individual students provide answers).
23 min.	Practical activity: reading 3 paragraphs, defining the register; providing textual evidence.	Develop reading skill and critical text analysis skill. Application of vocabulary of the theoretical concepts.	Written texts (see Appendix 2) and the table (see Appendix 3) available on Moodle.	1. Initially: Teacher – students (exemplifying the task); 2. Individual work.

Appendix 2 Theory on grammar and lexis among registers

Grammar

Informal writing	Formal writing
In very informal writings there might be a tendency to ellipsis (for instance, 'don't mind' instead of 'I don't mind')	Full phrases and clauses.
The abbreviation of words (contractions) are extensively used (he's , they're)	Full forms are used (he is, they are)
Simple and short clauses , with little elaboration (papildināšana)	Standard grammar in terms of word order and sentence construction
Incidence (bieži izmantots) of co-ordinated	Longer and more elaborate complex clauses

clauses ; the use of 'and' tend to be used as loose continuation marker.	with embedded phrases and clauses, particularly in the form of densely informative noun phrases; explicit and varied marking of clause relations, for instance, use of subordinating conjunctions between clauses (although, whereas, if, as) or sentence adverbs (however, in addition)
Personal constructions , e.g. You may find it difficult to incorporate an additional workload into heavy agenda.	Impersonal constructions , e.g. It is often difficult to incorporate an additional workload into heavy agenda.
Informal negative forms : not any, not much, not many, e.g. The analysis didn't yield any new results. This problem doesn't have many viable solutions.	Formal negative forms : no, little, few, e.g. The analysis yielded no new results. This problem has few viable solutions.
Use of direct questions What can be done to lower costs?	Limited use of direct questions We now need to consider what can be done to lower costs. We now need to consider how costs may be lowered.
Adverbs often occur as clauses at the beginning or end of sentences , e.g. <u>Then</u> the solution can be discarded. The blood is withdrawn <u>slowly</u> .	Adverbs are normally placed in midposition , e.g. The solution can <u>then</u> be discarded. The blood is <u>slowly</u> withdrawn.
	-ing forms used, not so much relative clauses
	Use of conditionals (if... then)
Active voice	Passive voice

Vocabulary/Lexis

Informal writing	Formal writing
Low <u>lexical density</u> , which means that more grammatical or function words (determiners – the that, conjunctions, prepositions, modals, pronouns, auxiliary v, quantifiers some, both) are used than lexical or content words	High <u>lexical density</u> , which means that more lexical or content words are used than grammatical or function words
<u>Low nominal style</u> , which means greater use of verbs than nouns to construct meanings	<u>High nominal style</u> , which means greater use of nouns than verbs
Repetitions of vocabulary	Greater variety in choice of vocabulary with lower levels of repetition
Tendency for less abstract vocabulary ; tendency for more generalised and simpler vocabulary, and semantically 'empty' fillers , such as 'you know', 'well'; use of terms that depend on the context of production for their understanding.	Complex vocabulary and the use of more abstract terms with a higher incidence of words of Greek or Latin origin;
Quite extensive use of such 'run on'	Limited use of such 'run on' expressions as

expressions/abbreviations as 'and so forth, etc.'	'and so forth, etc.'
These semiconductors can be used in robots, CD players, etc.	These semiconductors can be used in robots, CD players, and other electronic devices.
Extensive use of personal pronouns: 'you and I'	Considerably limited use of personal pronouns 'you and I' except in letters and textbooks
	Compound-nouns

Discourse

Discourse is a continuous stretch of language larger than a sentence, and it is perceived to be meaningful, unified, and purposive [Cook, 1990: 158; Crystal 1992:72; McEnery, Hardie 2012)

Informal writing	Formal writing
<u>Discourse</u> is produced by more than one participant	<u>Discourse</u> is presented as product of a single participant
There might be some incidence of markers of interpersonal dynamics (tag-questions, interruptions, unfinished clauses)	Few markers of interpersonal or personally-oriented <u>discourse</u>
Explicit indication of text organisation , such as paragraphing new lines.	Explicit indication of text organisation , such as paragraphing or the term 'see below'
<u>Punctuation</u> marks not only help to organise the text and chunk the information but also are used to express emotions. Therefore there might be a very high incidence of <u>question and exclamation marks, ellipsis, dash and brackets.</u>	<u>Punctuation</u> marks help to organise the text and chunk the information, however, there is a <u>lower incidence of such punctuation marks that help show the writers' emotions:</u> question, exclamation marks, <u>ellipsis</u> , dash, brackets.
Emotions in informal writing are used to compensate the tone of the voice and gestures to communicate feelings. Thus, for example, some email users apply special symbols to communicate their state of mind: :-)= happy. Similarly it is possible to place words in brackets followed by exclamation mark. For example, 'I will need at least three more days to finish my report (sigh), but then I will visit my best friend in Paris smile!). These strategies are fine to use in emails for friends.	Emotions are expressed with the help of a carefully chosen vocabulary and <u>punctuation</u>

Mechanics

Informal writing	Formal writing
Capitalisations are used to emphasise words or even sentences.	It is important carefully follow the guidelines regarding capitalisation .
In informal messages it is possible to abbreviate any words as far as the message	It is important carefully follow the guidelines regarding abbreviations . In case it <u>is more</u>

<p>sender is sure that the recipient will understand the message. Sometimes even the abbreviations used in faxes are used also in informal email messages.</p>	<p>convenient to use the abbreviation of the title of an organisation or project, it is essential <u>first of all give the full title and its abbreviation in brackets. Further in the text it is possible to operate with the abbreviation.</u></p>
<p>It is <u>very advisable</u> to choose carefully between BrE or AmE spelling depending on the purpose, readership and communicative context. Once the choice is made attempt to remain consistent throughout the whole writing.</p>	<p>It is <u>essential</u> to choose carefully between BrE or AmE spelling depending on the purpose, readership and communicative context. Once the choice is made attempt to remain consistent throughout the whole writing.</p>

Appendix 3

Recognition tasks on grammar and lexis among registers

Read the paragraphs. Fill in the table below indicating the level of formality (register) of each paragraph. Support your choice by indicating the lexical and grammatical features found in the texts.

Paragraph 1:

The relationship between cinema and the social norms it presents is complicated. [...] Cinema regulated itself in order to avoid governmental intervention, maintained prevailing social norms and critiqued regnant values, and profited from movements both to maintain and to contest existing social structures. In other words, little consistency emerges from an overview of the history of violent film images in terms of their seeming conformity with or deviation from existing norms. There likewise appears no deterministic relationship between instances in which mainstream cinema operates as an agent of social control and those in which it serves as an institution of change. Yet the correspondences persist and urge closer consideration of specific historical episodes when violent cinema functions as a cultural practice through which the role of cinema is negotiated.

(Slocum, J. D. (2000) Film Violence and the Institutionalization of the Cinema. In *Social Research*, 67 (3): pp. 649-681)

Paragraph 2:

I think of my participation in cinema as very successful. When I hit Los Angeles to play the Hill Street Theater, a Hollywood agent, Bill Periberg, called me and asked, "How'd you like to make a movie test?" I was large about it. I said that, as it happened, we had a few days before playing San Diego, so I thought I could work a test in. Bill didn't seem too impressed by my willingness to give the cinema a chance. "Get out to the Pathe lot next Thursday," he said. At Pathe, when the camera began to turn, I did my stuff. Then Louise Troxell came on and worked with me. The stagehands and crew laughed, and I went away convinced that all the movie industry needed to put it on a firm basis was me. I went on down to San Diego with Louise, with my little pointed head filled with such happy thoughts.

(Hope, B. and Pete, M. (1998) A Century of Hope. In *Saturday Evening Post*, (270):2, pp. 32-70)

Paragraph 3:

The middle of the 20th century marked a turning point into the field of cinema. The first step in his (Hitchcock's) elevation to the pantheon came with the publication in 1967 of

Hitchcock/Truffaut, a book in which the French director Francois Truffaut, a central figure in the "new wave" of postwar European cinema, interviewed his older colleague at length about the making of Hitchcock's 50-odd feature films. Not coincidentally, Truffaut's book appeared at the moment when the long-accepted distinction between "commercial" and "art" films was breaking down. Throughout the first half-century of filmmaking, it had been generally assumed that Europe was the center of artistically serious cinema. As early as the 50's, though, a new generation of critics was starting to pay closer attention to the work of such Hollywood directors as Ford, Hawks, and Hitchcock, and before long their films were being taken as seriously as the rock critics of the day took the music of Bob Dylan and the Beatles. In the same way that the pop culture of the 60's attained critical respectability, so did the maker of Psycho and Catch.

(Teachout, T. (2009) The Trouble With Alfred Hitchcock. In *Commentary*, (127):2, pp. 43-46)

Register: The words, style, and grammar used by speakers and writers in a particular situation or in a particular type of writing (e.g., formal/informal) (*Longman Dictionary of Contemporary English*)

	Register	Lexical evidence	Grammatical evidence
Paragraph 1			
Paragraph 2			
Paragraph 3			

Appendix 4

Lesson plan 2: Comparison/contrast paragraph

NB! The course is delivered in the form of **seminars** (focusing on students engaged in writing process)

Date: December 8th, 2015

Lesson Plan Components	TOPIC OF THE LESSON: STRUCTURE OF A COMPARISON/CONTRAST PARAGRAPH. LINKING OF THE IDEAS IN COMPARISON/CONTRAST PARAGRAPH – LEXICAL COHESION
Level of students	1 st year English Philology BSP students, Group A1
Timetable fit	Previous lecture: classification paragraph structure and topic sentence, very brief introduction of comparison/contrast paragraph concept and structure.
Main aim(s)	<ol style="list-style-type: none"> 1. Continue the introduction of the notion of comparison/contrast paragraph 2. Revise the structural patterns of comparison/contrast paragraph 3. Introduce linking devices and the concept of lexical cohesion of comparison/contrast paragraph 4. Consolidate the theoretical knowledge through practical exercises 5. Engage the students in paragraph writing process
Objectives	<ol style="list-style-type: none"> 1. Inclusion of the following skills: <ul style="list-style-type: none"> - Listening (presentation, theory) - Reading (practical exercises aimed at critical text analysis – recognition of textual evidence (skimming and scanning)) - Writing (composition of a comparison/contrast paragraph)
Assumptions	<ol style="list-style-type: none"> 1. Students might encounter difficulties with applying the appropriate terminology.
Anticipated problem	The time ascribed to each activity might deviate from the plan.
Possible solution	If some problems have been encountered, teacher's assistance has to be provided.

<i>Timing</i> (90 minutes)	<i>Procedure</i>	<i>Stage Aims</i>	<i>Aids and Materials</i>	<i>Interaction Pattern</i>
5 minutes Lead-in activities	Students are reminded of the topic of the <u>previous</u> lesson (point-by-point v. block structure) A question-answer session	<ol style="list-style-type: none"> 1. Consolidate previous knowledge 2. Activate students' schemata regarding paragraph structure 3. Encourage speaking 	No specific materials For consolidation: students might make reference to their notes, Power Point Presentation (see Appendix 2)	Teacher-students
10 minutes	Exercise 4: arrangement of sentences/ideas to form a logical (cohesive/coherent) paragraph	Consolidate the knowledge of paragraph structure	The task found in Students Post (See Appendix 1)	Individual work
5 minutes	Theory: coherence and cohesion: lexical devices	Introduce the main linking words/phrases in comparison/contrast paragraph	The task found in Students Post (See Appendix 1)	Individual work
5 minutes	Exercise 5: Read the paragraph and mark the	<ol style="list-style-type: none"> 1. Recognize linking devices 2. Promote text analysis skills 	The task found in Students Post (See Appendix 1)	<ol style="list-style-type: none"> 1. Individual work 2. After

	transition devices			finishing – peer work (checking)
10 minutes	Theory: Lexical cohesion – cohesive force	1. Promote listening skills 2. Introduce the aspect of lexical cohesion	Power Point presentation	Teacher - students
10 minutes	Exercise 6: Reading and recognition exercise: fill in the table	1. Practice lexical cohesion force (synonyms, etc).	The task found in Students Post (See Appendix 1)	Peer work
43 minutes	Exercise 7: Writing comparison/contrast paragraph	Consolidate the theoretical information from the lecture by practical exercise	Computer - New Word document (to be included in the students' portfolio)	Individual work Teacher's guidance /advice provided
2 minutes	Rounding up: questions, next lecture	Clarify the future issues	-	Teacher - students

Appendix 5
Tasks for students

COMPARISON/CONTRAST PARAGRAPH
Structure. Linking devices. Lexical Cohesion.

Exercise 1. Read the paragraph. Identify the topic sentence (underline) and the structural pattern of the paragraph.

My hometown and my college town have several things in common. First, both are small rural communities. For example, my hometown, Gridlock, has a population of only about 10,000 people. Similarly, my college town, Subnormal, consists of about 11,000 local residents. This population swells to 15,000 people when the college students are attending classes. A second way in which these two towns are similar is that they are both located in rural areas. Gridlock is surrounded by many acres of farmland which is devoted mainly to growing corn and soybeans. In the same way, Subnormal lies in the center of farmland which is used to raise hogs and cattle. Thirdly, both of these towns are similar in that both contain college campuses. Gridlock, for example, is home to Neutron College, which is famous for its Agricultural Economics program as well as for its annual Corn-Watching Festival. Likewise, the town of Subnormal boasts the beautiful campus of Quark College, which is well known for its Agricultural Engineering department and also for its yearly Hog-Calling Contest.

Structural pattern of the paragraph: _____
(modified from Lahav, n.d.).

Exercise 2. Now, when you have identified the structural pattern of the paragraph, please, fill in the table by providing examples from the text.

<u>TOPIC SENTENCE</u>	
POINT #1	
DETAIL (a)	
DETAIL (b)	
POINT #2	
DETAILS (a)	
DETAILS (b)	
POINT #3	
DETAILS (a)	
DETAILS (b)	

(modified from Online 2)

Exercise 3. Read the paragraph. Identify the topic sentence and the structural model applied for the text organization.

Even though Arizona and Rhode Island are both states of the U.S., they are strikingly different in physical size, population, natural environment. Arizona is large, having an area of 114,000 square miles. Next, it has about four million people living in it. Arizona is a very dry state, consisting of large desert areas that do not receive much rainfall every year. In addition, Arizona is a landlocked state and thus has no seashore. In contrast, Rhode Island is only about a tenth the size of Arizona, having an area of only 1,214 square miles. Rhode Island's population is less than one million. Regarding the environment, Rhode Island is located in a temperate zone and receives an average of 44 inches of rain per year. Rhode Island lies on the Atlantic Ocean and does have a significant coastline. Although representing the same region, both countries are so diverse.

Structural pattern of the paragraph: _____

(modified from Online 2)

Exercise 4. In the left side of the table you see sentences of a paragraph arranged in the wrong order. Put these sentences into the correct order (in the right side of the table) to form a comparison/contrast paragraph.

The Weather in Chicago and Miami

<ol style="list-style-type: none">1. In the end, my cousin and I learned that each of our hometowns has unique weather.2. Our discussion centred on two differences between the weather in our hometowns.3. Miami, in contrast, has only two seasons: a very mild winter and a very long summer.4. My cousin and I recently had a discussion about whether his hometown, Chicago, or my hometown, Miami, has better weather.5. Another major difference in the weather between our two cities is that Chicago's worst weather occurs in the winter.6. First, Chicago has all four seasons, but Miami does not. Chicago enjoys summer, fall, winter, and spring weather.7. On average, the high temperature reaches only around 32 degrees, and the low each night goes down to about 20degrees.8. In the summer, the temperature reaches 95 degrees in the daytime and drops only to 75 or so at night.9. The problem in Miami is not the cold but rather the heat.	
---	--

(modified from Online 3)

Exercise 5. Look back at the paragraph in the **right** side of the table **Exercise 4** and find the transitional devices. Mark them in different colour.

Exercise 6. Look at the paragraph from Exercise 1. Identify the lexical cohesion force (reiteration, synonym, near synonym, superordinate, collocation). Complete the table below. Some examples are already provided.

My hometown and my college town have several things in common. First, both are small rural communities. For example, my hometown, Gridlock, has a population of only about 10,000 people. Similarly, my college town, Subnormal, consists of about 11,000 local residents. This population swells to 15,000 people when the college students are attending classes. A second way in which these two towns are similar is that they are both located in rural areas. Gridlock is surrounded by many acres of farmland which is devoted mainly to growing corn and soybeans. In the same way, Subnormal lies in the center of farmland which is used to raise hogs and cattle. Thirdly, both of these towns are similar in that both contain

college campuses. Gridlock, for example, is home to Neutron College, which is famous for its Agricultural Economics program as well as for its annual Corn-Watching Festival. Likewise, the town of Subnormal boasts the beautiful campus of Quark College, which is well known for its Agricultural Engineering department and also for its yearly Hog-Calling Contest.

General noun	Synonyms	Near-synonyms	Collocations
Human: community Place: Hometown Non-human animate: hogs Inanimate concrete: college Inanimate abstract: Festival	community–population	people – students	Things in common Rural communities Consists of

(modified from Online 1)

Exercise 7. Choose one of the topics provided below and write a comparison/contrast paragraph (150 – 200 words). Follow the **point-by-point** structure.

Topics:

1. Childhood and Adulthood
2. Theatre and Cinema
3. Education in country X and country Y

Send the paragraph **before the next lecture to the e-mail: kraukle.ieva@inbox.lv**

Appendix 6

PowerPoint presentation on comparison/contrast paragraph

Slide 1

Slide 2

COMPARISON-CONTRAST PARAGRAPH

Structure
Linking devices
Lexical Cohesion

Ieva Kraukle
MA2, Methodology

THIS LECTURE:

- What is comparison-contrast paragraph?
- Structure of a paragraph
- Linking devices
- **Lexical Cohesion**
- Practical exercises
- Home task

Coherence and Cohesion

Slide 3

3

COMPARISON PARAGRAPH

- The comparison paragraph compares two subjects and discusses how they are **alike** and lists a few examples.

(Boardman and Frydenberg, 2008: 40)

- A paragraph which discusses the **similarities** between two subjects.

(Online 1; Online 2)

Slide 4

4

CONTRAST PARAGRAPH

- **Contrast** points out the **differences** between (at least two) persons, places, things, ideas, or situations.

(Zemach and Rumisek, 2006: 40)

Slide 5

5

STRUCTURE OF PARAGRAPH

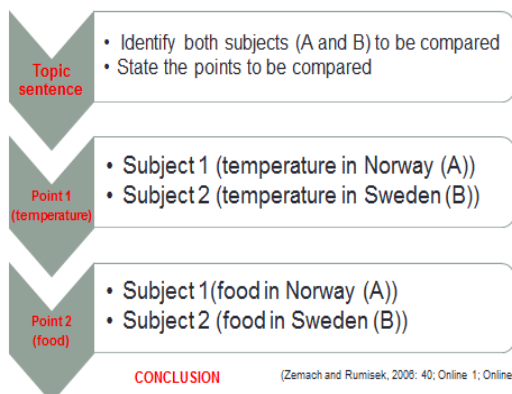
1. Point-by-point paragraph
2. Block paragraph

(Online 1)

Slide 6

6

POINT-BY-POINT PARAGRAPH



(Zemach and Rumisek, 2006: 40; Online 1; Online 2)

Slide 7

Slide 8

7

BLOCK PARAGRAPH

Topic sentence

- Identify both subjects (A and B) to be compared
- State the points to be compared (temperature and food)

A
Norway

- Temperature and food in Norway

B
Sweden

- Temperature and food in Sweden

CONCLUSION ibid.

8

POINT-BY-POINT PARAGRAPH

Topic sentence

- Identify subjects (A and B) to be contrasted
- State the points to be contrasted

Point 1
(temperature)

- Subject 1 (temperature in Norway (A))
- Subject 2 (temperature in Sweden (B))

Point 2
(food)

- Subject 1 (food in Norway (A))
- Subject 2 (food in Sweden (B))

CONCLUSION (Zemach and Rumisek, 2006: 40; Online 1; Online 2)

Slide 9

Slide 10

9

BLOCK PARAGRAPH

Topic sentence

- Identify both subjects (A and B) to be contrasted
- State the points to be contrasted (temperature and food)

A
Norway

- Temperature and food in Norway

B
Sweden

- Temperature and food in Sweden

CONCLUSION ibid.

10

Exercise 1

Read the paragraph and:

- identify the topic sentence
- identify the structural model applied for the text organization

My hometown and my college town have several things in common. First, both are small rural communities. For example, my hometown, Gridlock, has a population of only about 10,000 people. Similarly, my college town, Subnormal, consists of about 11,000 local residents. This population swells to 15,000 people when the college students are attending classes. A second way in which these two towns are similar is that they are both located in rural areas. Gridlock is surrounded by many acres of farmland which is devoted mainly to growing corn and soybeans. In the same way, Subnormal lies in the center of farmland which is used to raise hogs and cattle. Thirdly, both of these towns are similar in that both contain college campuses. Gridlock, for example, is home to Neutron College, which is famous for its Agricultural Economics program as well as for its annual Corn-Watching Festival. Likewise, the town of Subnormal boasts the beautiful campus of Quark College, which is well known for its Agricultural Engineering department and also for its yearly Hog-Calling Contest.

(modified from Labat, s.4)

Slide 11

Slide 12

11

Exercise 2

TOPIC SENTENCE	
POINT #1	
DETAIL #1 (a)	
DETAIL #1 (b)	
POINT #2	
DETAIL #2 (a)	
DETAIL #2 (b)	
POINT #3	
DETAIL #3 (a)	
DETAIL #3 (b)	

12

Exercise 3

Read the paragraph and:

- identify the topic sentence
- identify the structural model applied for the text organization
- support your choice by textual evidence

Even though Arizona and Rhode Island are both states of the U.S., they are strikingly different in physical size, population, natural environment. Arizona is large, having an area of 114,000 square miles. Next, it has about four million people living in it. Arizona is a very dry state, consisting of large desert areas that do not receive much rainfall every year. In addition, Arizona is a landlocked state and thus has no seashore. In contrast, Rhode Island is only about a tenth the size of Arizona, having an area of only 1,214 square miles. Rhode Island's population is less than one million. Regarding the environment, Rhode Island is located in a temperate zone and receives an average of 44 inches of rain per year. Rhode Island lies on the Atlantic Ocean and does have a significant coastline. Although representing the same region, both countries are so diverse.

Slide 13

Slide 14

Exercise 4

- Arrange the sentences to form a paragraph (the task provided in the handouts).

(mine, modified from Online 3).

Slide 15

COHERENCE, COHESION

« [...] The sentences must hold together; that is, the movement from one sentence to the next must be **logical** and **smooth**. »

(Oshima and Hogue, 2008: 21; Boardman and Frydenberg, 2008: 18)

Methods:

- use consistent **pronouns**;
- use **transition signals** to link ideas;
- arrange ideas in logical order.

(Oshima and Hogue, 2008: 22)

Slide 16

LINKING DEVICES - COMPARISON

Transitional expressions used in comparison:

*in the same way
and, also, in addition
as well as
both, neither
each of
just as...so
similarly
like
too
the same*

(Boardman and Frydenberg, 2008: 42; Online 3)

Slide 17

Exercise 5

Look back at the paragraph on the **right** side of the table in **Exercise 4** and find the transitional devices. Mark them in different colour.

LINKING DEVICES - CONTRAST

Short Conjunctions	Subordinating Conjunctions
However, In contrast, By contrast, ..., but ..., yet	On the other hand, even though + [sentence] although + [sentence] whereas + [sentence] unlike + [sentence] while + [sentence] nevertheless,

(Zemach and Rumisek, 2006: 43; Online 1; Online 2)

Slide 18

LEXICAL COHESION

«The cohesive effect achieved by the selection of vocabulary»
(Halliday and Hasan, 1976: 274)

Lexical force:

1. **General noun**
- a small set of nouns having generalized reference within the major noun classes (human, non-human, place).
2. **Reiteration (repetition)**
- synonyms, near-synonyms, superordinate.
3. **Collocation**
- lexical items that regularly re-occur.

(Halliday and Hasan, 1976: 274 - 292)

Slide 19

General noun

- Can you tell me where to stay in Geneva? I have never been to the place.

(ibid.: 275)

Slide 20

Reiteration (repetition)

- *reiteration:*
There was a large mushroom... She stretched to see above the edge of the mushroom.
- *Synonym (near-synonym):*
- I took leave, and turned to the ascent of the peak. The climb is perfectly easy.
- *Sir Bedivere*..clutched the sword. The great brand made lightnings in the splendour of the moon.
- *superordinate:*
- Henry bought a new Jaguar. He practically lives in the car. (ibid.: 277 - 278)

Slide 21

Collocation

- Words that frequently occur together.
 - *the whole day*
 - *all day long*
 - *to make dinner*
 - *richly decorated*
 - *fully aware*
 - *barking dog*
 - *regular meal*
 - *consists of*
 - *etc.*

Slide 22

Exercise 6

General noun	Synonyms	Near-synonyms	Collocations
Human: community, population, people, residents, students Place: hometown, college town, towns, center, farmland, campuses, department Non-human animate: hogs, cattle Inanimate concrete: college, department, program, farmland, Neutron College, Quark College Inanimate – abstract: festival, contest	community – population people – residents famous – well-known	people – students rural area – farmland grow- raise annual – yearly festival - contest	things in common rural communities population of consists of local residents attending classes located in rural areas surrounded by acres of farmland colleague campus famous for in the center of well-known for

(modified from Outline 1)

Slide 23

Exercise 7 / Home Task

Choose one of the topics provided below and write a **comparison or contrast** paragraph (150 – 200 words).

Send it **before the next lecture to the e-mail:**

kraukle.leva@inbox.lv

1. Childhood and Adulthood
2. Cinema and Theatre
3. Education in Country X and Country Y

Slide 24

Next lecture

- Corpus-based research
- Self-editing (comparison/contrast paragraphs)

REFERENCES

1. Boardman, C. A. and Frydenberg, J. (2008) *Writing to Communicate: Paragraphs and Essays*, 3rd edition. New York: Pearson.
2. Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English*. London and New York: Routledge.
3. Zeman, D. E. and Rumisek, L. A. (2006) *Academic Writing: From Paragraph to Essay*. London: Macmillan.
4. Oshima, A. and Hogue, A. (2008) *Writing Academic English*, 4th edition. New York: Pearson.
5. Lahav, D. (n.d.) *Business English: Paragraph structure: Comparing and Contrasting*. Ruppel Academic Center. Available from <http://www.slideshare.net/DebbieLahav/paragraph-structure-compare-contrast> [Accessed on 20 November 2015]

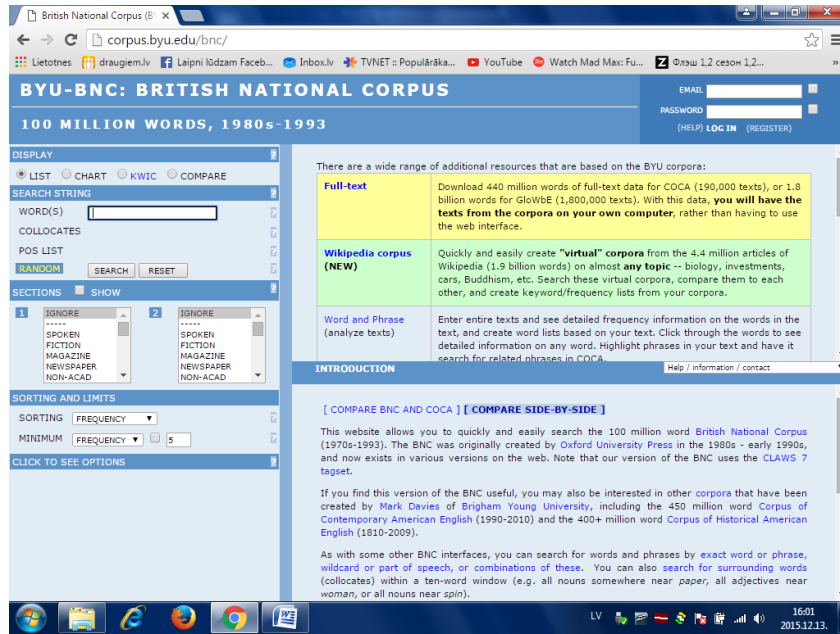
Internet sources

1. <http://enqllsh120.pbworks.com/w/page/19006833/contrast%20and%20comparison%20paragraph> [Accessed on 20 November 2015]
2. <http://irs.ed.uiuc.edu/student/walters/compoort.html> [Accessed on 20 November 2015]
3. http://www.slideshare.net/abdullahkhalifa/sample-of-compare-contrast-paragraphs17ne.it_slideshow-1 [Accessed on 20 November 2015]

Appendix 7

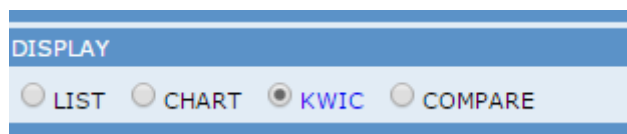
Manual for operation in the British National Corpus

1. Go to: <http://corpus.byu.edu/bnc/>
2. Press 'ENTER'
3. The following should be seen:

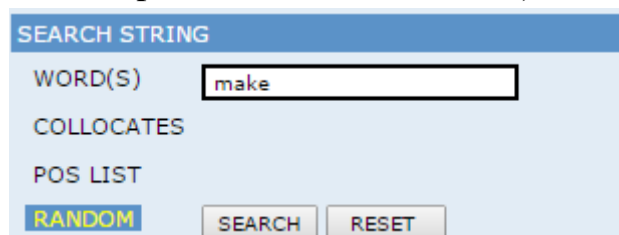


a.

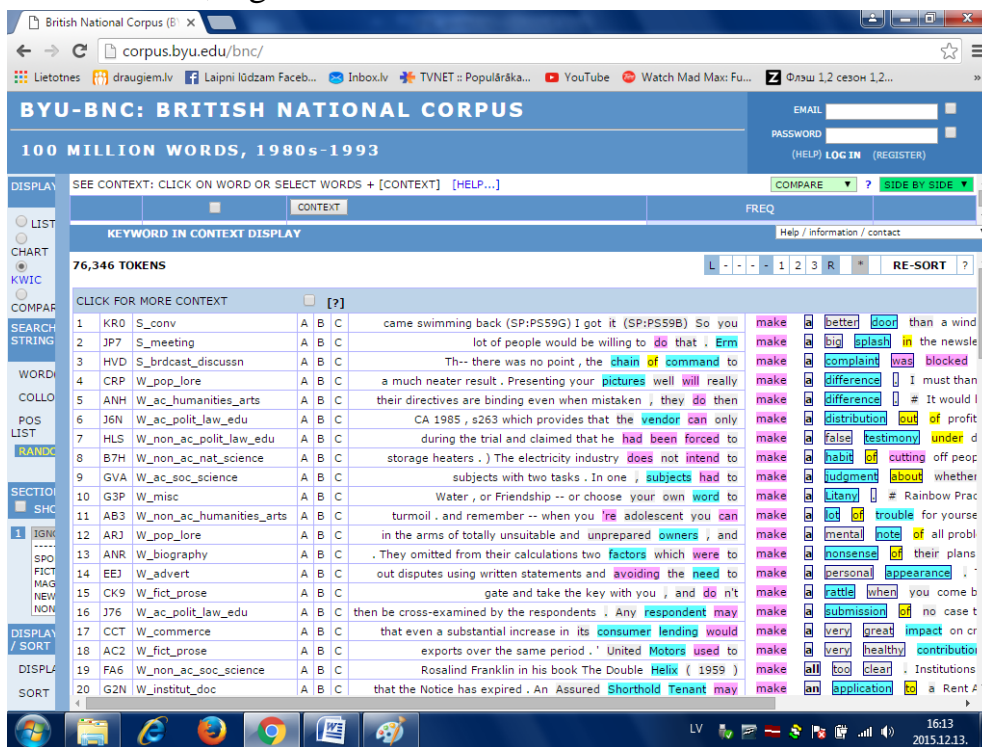
4. Press "Register" on the right upper corner of the screen and fill in the registration form.
5. Log in into the BNC webpage by using your username and password.
6. On the left side of the screen, under the section "DISPLAY" tick "KWIC":



7. Turn back to your comparison/contrast paragraph and make notice of the mistakes/suggestions.
8. Under the section "SEARCH STRING" type in a name of your interest (as in the example below, the word *make*). Press "SEARCH":



9. On the right side of the screen, the examples of the usage of the word have been indicated, e.g.:



10. The different colors indicate different word classes and parts of speech (nouns, verbs, adjectives, adverbs, prepositions, articles, etc). The middle column – pink – indicates the word you have searched for (e.g. *make*). The preceding and following text in each line provides information about the typical usage of the word: collocations, expressions, etc.

11. Turn back to your comparison/contrast paragraph and make appropriate changes in the text.

GUIDELINES for searching for SYNONYMS:

12. In the search bar enter the name you are interested in, positioning it into the square brackets, and preceded by an equality mark, for example, [=beautiful]. Press “ENTER”. You will see the following screen:

	CONTEXT	FREQ
1	BEAUTIFUL [S]	8307
2	LOVELY [S]	6000
3	ATTRACTIVE [S]	5000
4	WONDERFUL [S]	4626
5	STRIKING [S]	2540
6	SUPERB [S]	2009
7	MAGNIFICENT [S]	1948
8	HANDSOME [S]	1570
9	CHARMING [S]	1324
10	DELIGHTFUL [S]	1054
11	STUNNING [S]	914
12	PICTURESQUE [S]	602

13. The information possible to be found concerns the most typical synonyms of the word *beautiful* followed by their frequency in hierarchical order. If

necessary, press on the word in the list interesting for you and observe the results provided.

14. Turn back to your comparison/contrast paragraph and make the necessary changes.

IF ANY QUESTIONS, DO NOT HESITATE TO ASK FOR HELP!

Appendix 8

UCREL Semantic Analysis System (USAS)

A1 General And Abstract Terms
A1.1.1 General actions / making
A1.1.1- Inaction
A1.1.2 Damaging and destroying
A1.1.2- Fixing and mending
A1.2 Suitability
A1.2+ Suitable
A1.2- Unsuitable
A1.3 Caution
A1.3+ Cautious
A1.3- No caution
A1.4 Chance, luck
A1.4+ Lucky
A1.4- Unlucky
A1.5 Use
A1.5.1 Using
A1.5.1+ Used
A1.5.1- Unused
A1.5.2 Usefulness
A1.5.2+ Useful
A1.5.2- Useless
A1.6 Concrete/Abstract
A1.7+ Constraint
A1.7- No constraint
A1.8+ Inclusion
A1.8- Exclusion
A1.9 Avoiding
A1.9- Unavoidable
A2 Affect
A2.1 Modify, change
A2.1+ Change
A2.1- No change
A2.2 Cause&Effect/Connection
A2.2+ Cause/Effect/Connected
A2.2- Unconnected
A3 Being
A3+ Existing
A3- Non-existing
A4 Classification
A4.1 Generally kinds, groups, examples
A4.1- Unclassified
A4.2 Particular/general; detail
A4.2+ Detailed
A4.2- General
A5 Evaluation
A5.1 Evaluation: Good/bad
A5.1+ Evaluation: Good
A5.1- Evaluation: Bad
A5.2 Evaluation: True/false
A5.2+ Evaluation: True
A5.2- Evaluation: False
A5.3 Evaluation: Accuracy
A5.3+ Evaluation: Accurate
A5.3- Evaluation: Inaccurate
A5.4 Evaluation: Authenticity
A5.4+ Evaluation: Authentic

A5.4- Evaluation: Unauthentic
 A6 Comparing
 A6.1 Comparing: Similar/different
 A6.1+ Comparing: Similar
 A6.1- Comparing: Different
 A6.2 Comparing: Usual/unusual
 A6.2+ Comparing: Usual
 A6.2- Comparing: Unusual
 A6.3 Comparing: Variety
 A6.3+ Comparing: Varied
 A6.3- Comparing: Unvaried
 A7 Probability
 A7+ Likely
 A7- Unlikely
 A8 Seem
 A9 Getting and giving; possession
 A9+ Getting and possession
 A9- Giving
 A10 Open/closed; Hiding/Hidden; Finding; Showing
 A10+ Open; Finding; Showing
 A10- Closed; Hiding/Hidden
 A11 Importance
 A11.1 Importance
 A11.1+ Important
 A11.1- Unimportant
 A11.2 Noticeability
 A11.2+ Noticeable
 A11.2- Unnoticeable
 A12 Easy/difficult
 A12+ Easy
 A12- Difficult
 A13 Degree
 A13.1 Degree: Non-specific
 A13.2 Degree: Maximizers
 A13.3 Degree: Boosters
 A13.4 Degree: Approximators
 A13.5 Degree: Compromisers
 A13.6 Degree: Diminishers
 A13.7 Degree: Minimizers
 A14 Exclusivizers/particularizers
 A15 Safety/Danger
 A15+ Safe
 A15- Danger
 B1 Anatomy and physiology
 B2 Health and disease
 B2+ Healthy
 B2- Disease
 B3 Medicines and medical treatment
 B3- Without medical treatment
 B4 Cleaning and personal care
 B4+ Clean
 B4- Dirty
 B5 Clothes and personal belongings
 B5- Without clothes
 C1 Arts and crafts
 E1 Emotional Actions, States And Processes General
 E1+ Emotional
 E1- Unemotional
 E2 Liking
 E2+ Like
 E2- Dislike

E3 Calm/Violent/Angry
 E3+ Calm
 E3- Violent/Angry
 E4 Happiness and Contentment
 E4.1 Happy/sad
 E4.1+ Happy
 E4.1- Sad
 E4.2 Contentment
 E4.2+ Content
 E4.2- Discontent
 E5 Bravery and Fear
 E5+ Bravery
 E5- Fear/shock
 E6 Worry and confidence
 E6+ Confident
 E6- Worry
 F1 Food
 F1+ Abundance of food
 F1- Lack of food
 F2 Drinks and alcohol
 F2+ Excessive drinking
 F2- Not drinking
 F3 Smoking and non-medical drugs
 F3+ Smoking and drugs abuse
 F3- Non-smoking / no use of drugs
 F4 Farming & Horticulture
 F4- Uncultivated
 G1 Government and Politics
 G1.1 Government
 G1.1- Non-governmental
 G1.2 Politics
 G1.2- Non-political
 G2 Crime, law and order
 G2.1 Law and order
 G2.1+ Lawful
 G2.1- Crime
 G2.2 General ethics
 G2.2+ Ethical
 G2.2- Unethical
 G3 Warfare, defence and the army; weapons
 G3- Anti-war
 H1 Architecture, houses and buildings
 H2 Parts of buildings
 H3 Areas around or near houses
 H4 Residence
 H4- Non-resident
 H5 Furniture and household fittings
 H5- Unfurnished
 I1 Money generally
 I1.1 Money and pay
 I1.1+ Money: Affluence
 I1.1- Money: Lack
 I1.2 Money: Debts
 I1.2+ Spending and money loss
 I1.2- Debt-free
 I1.3 Money: Cost and price
 I1.3+ Expensive
 I1.3- Cheap
 I2 Business
 I2.1 Business: Generally
 I2.1- Non-commercial

I2.2 Business: Selling
 I3 Work and employment
 I3.1 Work and employment: Generally
 I3.1- Unemployed
 I3.2 Work and employment: Professionalism
 I3.2+ Professional
 I3.2- Unprofessional
 I4 Industry
 I4- No industry
 K1 Entertainment generally
 K2 Music and related activities
 K3 Recorded sound
 K4 Drama, the theatre and show business
 K5 Sports and games generally
 K5.1 Sports
 K5.2 Games
 K6 Children's games and toys
 L1 Life and living things
 L1+ Alive
 L1- Dead
 L2 Living creatures: animals, birds, etc.
 L2- No living creatures
 L3 Plants
 L3- No plants
 M1 Moving, coming and going
 M2 Putting, pulling, pushing, transporting
 M3 Vehicles and transport on land
 M4 Sailing, swimming, etc.
 M4- Non-swimming
 M5 Flying and aircraft
 M6 Location and direction
 M7 Places
 M8 Stationary
 N1 Numbers
 N2 Mathematics
 N3 Measurement
 N3.1 Measurement: General
 N3.2 Measurement: Size
 N3.2+ Size: Big
 N3.2- Size: Small
 N3.3 Measurement: Distance
 N3.3+ Distance: Far
 N3.3- Distance: Near
 N3.4 Measurement: Volume
 N3.4+ Volume: Inflated
 N3.4- Volume: Compressed
 N3.5 Measurement: Weight
 N3.5+ Weight: Heavy
 N3.5- Weight: Light
 N3.6 Measurement: Area
 N3.6+ Spacious
 N3.7 Measurement: Length & height
 N3.7+ Long, tall and wide
 N3.7- Short and narrow
 N3.8 Measurement: Speed
 N3.8+ Speed: Fast
 N3.8- Speed: Slow
 N4 Linear order
 N4- Nonlinear
 N5 Quantities
 N5+ Quantities: many/much

N5- Quantities: little
 N5.1 Entirety; maximum
 N5.1+ Entire; maximum
 N5.1- Part
 N5.2 Exceeding
 N5.2+ Exceed; waste
 N6 Frequency
 N6+ Frequent
 N6- Infrequent
 O1 Substances and materials generally
 O1.1 Substances and materials: Solid
 O1.2 Substances and materials: Liquid
 O1.2- Dry
 O1.3 Substances and materials: Gas
 O1.3- Gasless
 O2 Objects generally
 O3 Electricity and electrical equipment
 O4 Physical attributes
 O4.1 General appearance and physical properties
 O4.2 Judgement of appearance
 O4.2+ Judgement of appearance: Beautiful
 O4.2- Judgement of appearance: Ugly
 O4.3 Colour and colour patterns
 O4.4 Shape
 O4.5 Texture
 O4.6 Temperature
 O4.6+ Temperature: Hot / on fire
 O4.6- Temperature: Cold
 P1 Education in general
 P1- Not educated
 Q1 Linguistic Actions, States And Processes; Communication
 Q1.1 Linguistic Actions, States And Processes; Communication
 Q1.2 Paper documents and writing
 Q1.2- Unwritten
 Q1.3 Telecommunications
 Q2 Speech
 Q2.1 Speech: Communicative
 Q2.1+ Speech: Talkative
 Q2.1- Speech: Not communicating
 Q2.2 Speech acts
 Q2.2- Speech acts: Not speaking
 Q3 Language, speech and grammar
 Q3- Non-verbal
 Q4 The Media
 Q4.1 The Media: Books
 Q4.2 The Media: Newspapers etc.
 Q4.3 The Media: TV, Radio and Cinema
 S1 Social Actions, States And Processes
 S1.1 Social Actions, States And Processes
 S1.1.1 Social Actions, States And Processes
 S1.1.2 Reciprocity
 S1.1.2+ Reciprocal
 S1.1.2- Unilateral
 S1.1.3 Participation
 S1.1.3+ Participating
 S1.1.3- Non-participating
 S1.1.4 Deserve
 S1.1.4+ Deserving
 S1.1.4- Undeserving
 S1.2 Personality traits
 S1.2.1 Approachability and Friendliness

- S1.2.1+ Informal/Friendly
- S1.2.1- Formal/Unfriendly
- S1.2.2 Avarice
- S1.2.2+ Greedy
- S1.2.2- Generous
- S1.2.3 Egoism
- S1.2.3+ Selfish
- S1.2.3- Unselfish
- S1.2.4 Politeness
- S1.2.4+ Polite
- S1.2.4- Impolite
- S1.2.5 Toughness; strong/weak
- S1.2.5+ Tough/strong
- S1.2.5- Weak
- S1.2.6 Common sense
- S1.2.6+ Sensible
- S1.2.6- Foolish
- S2 People
- S2- No people
- S2.1 People: Female
- S2.1- Not feminine
- S2.2 People: Male
- S3 Relationship
- S3.1 Personal relationship: General
- S3.1- No personal relationship
- S3.2 Relationship: Intimacy and sex
- S3.2+ Relationship: Sexual
- S3.2- Relationship: Asexual
- S4 Kin
- S4- No kin
- S5 Groups and affiliation
- S5+ Belonging to a group
- S5- Not part of a group
- S6 Obligation and necessity
- S6+ Strong obligation or necessity
- S6- No obligation or necessity
- S7 Power relationship
- S7.1 Power, organizing
- S7.1+ In power
- S7.1- No power
- S7.2 Respect
- S7.2+ Respected
- S7.2- No respect
- S7.3 Competition
- S7.3+ Competitive
- S7.3- No competition
- S7.4 Permission
- S7.4+ Allowed
- S7.4- Not allowed
- S8 Helping/hindering
- S8+ Helping
- S8- Hindering
- S9 Religion and the supernatural
- S9- Non-religious
- T1 Time
- T1.1 Time: General
- T1.1.1 Time: Past
- T1.1.2 Time: Present; simultaneous
- T1.1.2- Time: Asynchronous
- T1.1.3 Time: Future
- T1.2 Time: Momentary

T1.3 Time: Period
T1.3+ Time period: long
T1.3- Time period: short
T2 Time: Beginning and ending
T2+ Time: Beginning
T2- Time: Ending
T3 Time: Old, new and young; age
T3+ Time: Old; grown-up
T3- Time: New and young
T4 Time: Early/late
T4+ Time: Early
T4- Time: Late
W1 The universe
W2 Light
W2- Darkness
W3 Geographical terms
W4 Weather
W5 Green issues
X1 Psychological Actions, States And Processes
X2 Mental actions and processes
X2.1 Thought, belief
X2.1- Without thinking
X2.2 Knowledge
X2.2+ Knowledgeable
X2.2- No knowledge
X2.3 Learn
X2.3+ Learning
X2.4 Investigate, examine, test, search
X2.4+ Double-check
X2.4- Not examined
X2.5 Understand
X2.5+ Understanding
X2.5- Not understanding
X2.6 Expect
X2.6+ Expected
X2.6- Unexpected
X3 Sensory
X3.1 Sensory: Taste
X3.1+ Tasty
X3.1- Not tasty
X3.2 Sensory: Sound
X3.2+ Sound: Loud
X3.2- Sound: Quiet
X3.3 Sensory: Touch
X3.4 Sensory: Sight
X3.4+ Seen
X3.4- Unseen
X3.5 Sensory: Smell
X3.5- No smell
X4 Mental object
X4.1 Mental object: Conceptual object
X4.1- Themeless
X4.2 Mental object: Means, method
X5 Attention
X5.1 Attention
X5.1+ Attentive
X5.1- Inattentive
X5.2 Interest/boredom/excited/energetic
X5.2+ Interested/excited/energetic
X5.2- Uninterested/bored/unenergetic
X6 Deciding

X6+ Decided
X6- Undecided
X7 Wanting; planning; choosing
X7+ Wanted
X7- Unwanted
X8 Trying
X8+ Trying hard
X8- Not trying
X9 Ability
X9.1 Ability and intelligence
X9.1+ Able/intelligent
X9.1- Inability/unintelligence
X9.2 Success and failure
X9.2+ Success
X9.2- Failure
Y1 Science and technology in general
Y1- Anti-scientific
Y2 Information technology and computing
Y2- Low-tech
Z0 Unmatched proper noun
Z1 Personal names
Z2 Geographical names
Z3 Other proper names
Z4 Discourse Bin
Z5 Grammatical bin
Z6 Negative
Z7 If
Z7- Unconditional
Z8 Pronouns
Z9 Trash can
Z99 Unmatched

Appendix 9
Distribution of semantic domains in students' paragraphs based on
Wmatrix analysis

Type of paragraph →	Childhood and Adulthood	Cinema and Theatre	Education in Country X and Country Y
Nr.	Semantic domain		
1.	General, abstract terms (164)	General, abstract terms (327)	General, abstract terms (73)
2.	Social actions, states, processes (76)	Numbers, measurement (125)	Numbers, measurement (36)
3.	Numbers, measurement (65)	Linguistic actions, states, processes (90)	Psychological actions, states, processes (23)
4.	Time (59)	Entertainment, sports, games (73)	Education (16)
5.	Psychological actions, states, processes (42)	Psychological actions, states, processes (72)	Social actions, states, processes (15)
6.	Movement, location, travel, transport (31)	Time (60)	Names and grammatical words (14)
7.	Emotional actions, states, processes (18)	Social actions, states, processes (55)	Movement, location, travel, transport (10)
8.	Linguistic actions, states, processes (14)	Movement, location, transport (42)	Time (10)
9.	Money, commerce (13)	Substances, materials, objects, equipment (28)	Linguistic actions, states, processes (10)
10.	Life, living things (13)	Emotional actions, states, processes (21)	Money, commerce (3)
11.	Substances, materials, objects, equipment (12)	Money, commerce (17)	The body, the individual (2)
12.	Entertainment, sports, games (6)	The body, the individual (16)	Emotional actions, states, processes (2)
13.	Architecture, buildings, houses, the home (5)	Arts, crafts (13)	The world, our environment (2)
14.	Food and farming (4)	Foods and farming (10)	Science, technology (2)
15.	Govt., the public domain (3)	Life, living things (9)	Govt., the public domain (1)
16.	Education (2)	Science and technology (9)	Life, living things (1)
17.	The body, the individual (1)	Govt., the public domain (6)	Substances, materials, objects, equipment (1)
18.	Science, technology (1)	Architecture, buildings, houses, the home (5)	Arts, crafts (0)
19.	The world, the environment (1)	The world, the environment (5)	Food, farming (0)
20.	Arts, crafts (0)	Education (2)	Architecture, buildings, houses, the home (0)
21.	Names and grammatical words*	Names and grammatical words*	Entertainment, sports, games (0)

*Due to a great proportion of grammatical words (e.g., conjunctions, articles, pronouns, and other aspects of grammar) and their inevitable presence in any English text, they were not counted specifically, except for the paragraphs under the topic "Education in Country X and Country Y", where 2 sub-categories were relevant for the analysis: 'Geographical names', 'Other proper names'.

Appendix 10
Learner Corpus