

LATVIJAS UNIVERSITĀTE

MAGISTRA DARBS

RĪGA 2014

UNIVERSITY OF LATVIA
FACULTY OF HUMANITIES
DEPARTMENT OF ENGLISH STUDIES

**APPLYING CORPORA TO TEACHING SUMMARY
WRITING**

**KORPUSU LIETOŠANA
KOPSAVILKUMU RAKSTĪTMĀCĪŠANĀ**

MASTER THESIS

Author: **Ivans Sturovs**

Matriculation Card No. is07074

Adviser: as. prof. Zigrīda Vinčela

RIGA 2014

DECLARATION OF ACADEMIC INTEGRITY

I hereby declare that this study is my own and does not contain any unacknowledged material from any source.

Date:

Signed:

ACKNOWLEDGEMENTS

I would like to express my gratitude to the teaching staff of the University of Latvia, Department of English Studies, for the support in the thesis writing process. I especially wish to thank doc. Zigrīda Vinčela for assistance with planning the research and conducting the pilot lessons and asoc. prof. Monta Farneste for the initial stages of research planning and instructions on academic writing.

ANOTĀCIJA

Teksta korpusi jau vairākus gadus tiek izmantoti vairākiem tādiem mērķiem, kā vārdnīcu kompilēšana, pētniecība, tulkošana un literāras studijas. Pēdējā laikā gadījās veiksmīga korpusa balstītas metodes rašana mācīšanā, it īpaši akadēmiskajā līmenī, kur audzēkņiem ir spēja veikt pareizu korpusu datu izmantošanu daudzos izaicinājumos. Pētījuma darba mērķis ir veidot korpusa balstītas darbības, lai attīstītu terciārā līmeņa studentu akadēmiskās rakstīšanas prasmes, īpaši kopsavilkumu rakstīt mācīšanas prasmi, un izpētīt šo darbību efektivitāti, veicot izmēģinājuma pētījumu, kurā izstrādāti materiāli tiek pārbaudīti. Šis pētījums parāda korpusa metožu integrācijas iespējas augstākās izglītības kontekstā, īpaši attiecībā uz kopsavilkumu rakstīt mācīšanu, izvērtot veidus, kādos studenti izpētītu korpusu ar instruktora palīdzību un cik lielā mērā šāda iedarbība ietekmē akadēmiskās rakstīšanas prasmes un zināšanas par tās īpašajām leksikas funkcijām.

Atslēgvārdi: korpusi un mācīšana, concordances, korpusa lingvistika, akadēmiskā rakstīšana, angļu valoda akadēmiskajiem mērķiem, kopsavilkumi, kopsavilkumu rakstīšana, COCA korpus.

ABSTRACT

Corpora have long been utilised for purposes as diverse as compiling dictionaries, research, translation and literary studies. Recent times have witnessed a successful emergence of corpus-based methods in teaching, especially at the academic level where learners are fully capable of making proper use of corpus data for a multitude of challenges. The goal of the research paper is to create a set of corpus-based activities for the purpose of developing tertiary level students' academic writing skills, namely summarising skills, and investigate the effectiveness of these activities by carrying out a pilot study where the developed materials are tested. This research shows the opportunities the integration of corpus methods poses before an instructor in a tertiary education context, specifically as regards teaching summary writing, highlighting the ways in which the students explore the corpus under a guided instruction and the extent to which such an exposure affects their academic writing skills and awareness of specific lexical features thereof.

Key words: corpora and teaching, concordances, corpus linguistics, academic writing, EAP, summary writing, summarising, COCA Corpus.

Contents

Introduction.....	1
1 Corpora and English Language Teaching.....	4
1.1 Corpus Linguistics and Text Corpora	4
1.2 Corpus Linguistics and Language Teaching.....	5
1.3 Corpora in the Classroom: Advantages and Disadvantages.....	7
2 Corpus-based Teaching of Summary Writing	11
2.1 Features of Academic Writing.....	11
2.2 Features of Summary Writing.....	12
2.3 Corpora in Academic Writing Teaching.....	14
3 Corpus-based Research of Student-Composed Summaries.....	17
3.1 Methods and Research Tools.....	17
3.2 Research Subjects and the Pre-Questionnaire	17
3.3 Corpus-based Teaching Materials.....	26
3.4 Research Procedure and Results.....	28
Conclusions.....	65
Theses.....	67
References.....	68
Appendix 1 Questionnaire 1	71
Appendix 2 Handout: Sample Summary.....	73
Appendix 3 Lesson Plan 1	75
Appendix 4 Lesson Plan 2	78
Appendix 5 Post-Questionnaire.....	80
Appendix 6 Slides on Summarising Theory.....	82
Appendix 7 Feedback for Students on Pre-Summary Writing.....	87

Introduction

The usage of corpora in teaching has been characterised as revolutionary at one point (Sinclair, 2004 in O’Keeffe and McCarthy, 2010) and has since been popularised by many instructors and writers (Tribble, Aston). The aims of corpus-based learning vary greatly, from theoretical instruction on what corpora exist to practical application thereof in various areas of research activity to development of teaching materials (Renouf, 1997 in O’Keeffe and McCarthy, 2010). The current thesis deals with the implementation of corpus-based tasks to facilitate the teaching of summary writing to tertiary level students.

Even though the application of corpora in academic writing has been covered extensively in journals and research papers, the same cannot be said regarding its incorporation to summary writing specifically, which is the role the present research paper attempts to assume.

Conrad S., one of the leading contemporary specialists in the field of corpus linguistics, offers such advice to those seeking to enrich the progress of this scientific area and help guide it into a new direction, among other pieces of advice: ‘study students’ use of concordancing and their ability to apply the techniques to their own new searches’ (Conrad, 2013: Online 7), emphasising the importance of students’ individual ability to apply the wealth of corpus-based data for their own benefit and its value in research. This idea has been central to the pilot lessons conducted within the framework of the current thesis paper, as students were given an example of how the data the corpus offers could be utilised for the improvement of their academic writing skills; for instance, in error correction, assisting with the writing process and solving specific linguistic issues.

The approach behind the many of the corpus-related actions in the current research is **corpus-based**, as opposed to being corpus-driven. The difference between the two approaches is mainly related to the claims of the users and not the actual usage of corpora (McEnery et al., 2006:9). The claims or, more precisely, the absence of such claims include the respect paid to the existing theory in the researched area, which is denied in the case of corpus-driven linguists (McEnery et al., 2006:10), the usage of a small learner corpus for the sake of analysis as opposed to a large one lacking the desired niche and the lack of any claim that would insist upon the corpus findings being final and capable of rejecting or replacing the existing theoretical data within the scope of a single research (McEnery et al., 2006:9-11).

The **goal** of the thesis is to create a set of corpus-based activities for the purpose of developing tertiary level students' academic writing skills, namely summary writing skills, and investigate the effectiveness of these activities by carrying out a pilot study where the developed materials are put into practice.

The **hypothesis** of the paper is that the integration of corpus-based activities into developing English Academic writing skills, namely summary writing skills, is effective as it augments students' ability to distinguish between registers, use correct grammar and select the appropriate collocates, creating extra opportunities that facilitate the writing process **and accentuating the peculiar linguistic features typical of the summary genre.**

The **enabling objectives** are as follows:

- Analyse theoretical material and previous research in areas relevant to the paper to pinpoint possible directions for the research paper.
- Conduct a questionnaire to discover learners' backgrounds, issues with academic writing and expectations from the summary writing practice.
- Test students' summary writing skills prior to both theoretical instruction on summary writing and being acquainted with corpus analysis methods by conducting a pre-summary activity.
- Carry out lessons where students are familiarised with corpora and corpus applications in developing writing skills.
- Test the students throughout the available period of time to assess the impact of the piloting and specifically the usage of corpus materials.
- Create a learner corpus of students' summaries written before the piloting and after it for the purpose of analysing the linguistic features within.
- Analyse the written records, lessons that took place and the opinions students expressed as a single unity to determine the effectiveness of the methods applied.
- Evaluate the future of such techniques and possible improvements for a more efficient and meaningful implementation of corpus methods.

The research methods are **case study** and **corpus analysis**. It is a case study as it addresses a specific group of people – first year students of English philology BA, reporting both factual information and additionally giving importance to the subjective views and experiences of the students involved. The corpus analysis methods are applied to the scrutiny

of the learner corpora to discover regularities and linguistic features typical to students' summary writing.

The **data collection methods** are subjected to **triangulation**; namely, the interaction between quantitative and qualitative research in its three methods of gathering data:

- Pre-questionnaire/Post-questionnaire are administered prior to the piloting of corpus methods and the theoretical instructions and following the lessons. The first questionnaire attempts to investigate the students' prior experience, pinpoint their perceived difficulties in academic writing and checks if they have earlier familiarity with the corpus. The two questionnaires integrate both qualitative and quantitative research methods.
- Pre-summary/Post-summary writing coupled with interim tasks that help fortify the students' command of the corpus and their ability to use and rely on the theoretical data. Besides providing a wealthy amount of material for the learner corpus, the summary writing process contributes greatly to the qualitative research methods in the empirical part.
- Learner corpora analysis – examines the quantitative aspects of said features' usage and allows to efficiently locate any given grammatical feature in every student-produced text it has occurred.

The **subjects** of the research are tertiary level year 1 English philology students at the University of Latvia. Five groups (1A, 1B, 1C, 1D and 1FIN) are taught to incorporate evidence by means of summarising while making optimal use of the **British National**

Corpus as far as time restraints permit. The proficiency levels, attitudes, preferences and experience vary significantly depending on the group and particular students.

1. Corpora and English Language Teaching

The current chapter represents the theoretical materials that served as the background for the present research paper, laying out the foundation for the methods chosen and theories that proved relevant for conducting the research. It is divided into three subchapters – the first one titled ‘Corpus Linguistics and Text Corpora’ outlines a brief history of the field known as Corpus Linguistics and provides general information regarding text corpora. The next subchapter is titled ‘Corpus Linguistics and Language Teaching’ and features an explanation of possible applications of corpora known in English Language Teaching (ELT) as well as theories and views on the subject from prominent linguists. The final subchapter ‘Corpora in the Classroom: Advantages and Disadvantages’ is preoccupied with the perceived benefits and drawbacks of the corpus integration into teaching and attempts to reconcile these diametrically opposing views by pinpointing a way how some of the supposed difficulties of corpus application could be dealt with effectively.

1.1. Corpus Linguistics and Text Corpora

A corpus can be defined as ‘a large and principled collection of natural texts.’ (Biber et al., 1998: 12)

Corpus linguistics as a separate branch of linguistics is a fairly new discipline, having emerged in the 1960s. It had not become popular in use until decades later, obscured by the approach of Chomskyan and cognitive linguistics representatives and subjected to harsh criticism from their end.

Teubert (2004: 51) states the following about the earliest experiment in this area of science: ‘In the English-speaking world, the first large-scale project to collect language data for empirical grammatical research was Randolph Quirk’s Survey of English Usage which later led to what became the standard English grammar for many decades: A Comprehensive Grammar of the English language. The data was only computerized in the mid-1980s in Quirk and Greenbaum’s subsequent project now known as the International Corpus of English (ICE)’.

Gradually, corpora have been found use for in virtually all areas of linguistics and even outside of this science. Their use in compiling dictionaries has become increasingly

widespread. The first ever general language dictionary based exclusively on a corpus was published in 1987: it was the Collins COBUILD English Language Dictionary.

According to Teubert (2004: 37), corpus linguistics sees language as a social phenomenon.

' Meaning is, like language, a social phenomenon, he states. Teubert argues that freedom is a key aspect of discourse and therefore also corpus linguistic in a democratic society: 'It is something that can be discussed by the members of a discourse community. There is no secret formula, neither in natural language nor in a formal calculus, that contains the meaning of a word or phrase. There is no right or wrong. What I call a weapon of mass destruction differs probably a lot from what President George W. Bush calls a weapon of mass destruction. What I call a baguette is not the same as what many supermarkets sell as a baguette. What I call love may not be what my partner calls love. Different people paraphrase words or phrases in different ways. They do not have to agree. In a democracy, everyone's opinion is as good as anyone else's.'

He contrasts corpus linguistics to cognitive linguistics, explaining that the former deals with meaning whilst the latter is preoccupied with understanding (Teubert, 2004:38), and that the discipline is more effective than other disciplines in linguistics: 'corpus linguistics can tell us more about the meaning than either Chomskyan linguistics or standard linguistics.' (Teubert, 2004:39)

Teubert (2004: 134) is convinced of corpus linguistics potentially being an eminently useful source to all those interested in discovering the peculiarities of others' speech habits:

'Corpus linguistics puts us into a position where we can inform ourselves what use others have made of language.' He believes that such knowledge enables the corpus users to make their contribution to the discourse of their respective environment (Teubert, 2004: *ibid.*).

1.2. Corpus Linguistics and Language Teaching

McEnery et al. (2008: 97) observe that an increasing interest in corpus linguistics has been a growing trend since the early 1990s, reflected by the amount of published literature, found both in scientific journal articles and separate volumes, focusing on subjects such as corpus-based language description, corpus analysis in the context of teaching as well as learner corpora.

Leech in Wichmann et al., 1997b, as cited in McEnery et al. (2008: *ibid.*) points out the apparent convergence between language corpora and teaching. This connection, he claims, can be marked by three main foci:

- ‘Direct use of corpora in teaching
- Indirect use of corpora in teaching
- Further teaching-oriented corpus development’ (Leech, 1997b in McEnery et al.: *ibid.*)

The first focus is further divided into three specific direct kinds of usage:

- ‘Teaching about
- Teaching to exploit
- Exploiting to teach’

Indirect uses of corpora include such usage of corpora in which the user is not in direct access of the corpus, for instance in lexical or grammatical studies, exploiting dictionaries whose choice and sequential arrangement of meanings are derived from corpus data, as well as materials development and language testing. Teaching-oriented corpora, on the other hand, offer considerable help in languages for specific purposes in addition to research on first and second language (McEnery et al., 2008: *ibid.*). This division is also shared by Fligelstone (1993) cited in O’Keeffe and McCarthy (2010:319).

The focus that was most topical for the current research was undoubtedly the direct use of corpora, both by the teacher and the students. Prior to the piloting, the majority of the research subjects had most likely had unconscious indirect contact with corpora by making use of the wealthy amount of materials available that were built using one of the existing corpus engines, such as dictionaries. Enabling the learners to make direct contact with a corpus was one of the enabling objectives in attaining the desired goal of the research. Furthermore, Johns (1991:30) in his DDL or ‘data-driven learning’, cited in O’Keeffe and McCarthy (2010:320), expressed the viewpoint that corpora permit learners to combine the roles of active language learners and language researchers, an idea that is of utmost topicality to the present research paper:

‘What distinguishes the DDL approach is the attempt to cut out the middleman as much as possible and give direct access to the data so that the learner can take part in building his or her own profiles of meanings and uses. The assumption that underlines this approach is that effective language learning is itself a form of linguistic research, and that the concordance printout offers a unique resource for the stimulation of inductive learning strategies – in

particular, the strategies of perceiving similarities and differences and of hypothesis formation and testing.’

Though the students were not entirely without a mediator during the present piloting of corpus-based tasks and assignments, as instructions were required to permit the young users of corpora to take their first steps and understand its mechanics, the idea of the instructor playing the role of a mediator was taken up for the purposes of the pilot lessons held in five groups to grant the learners the autonomy that was expected to enable their swift progress.

Cheng in O’Keeffe and McCarthy (2010:320) discusses the interrelation between implicit and explicit knowledge in language learning in light of their relevance to DDL studies. Implicit knowledge, he argues, is knowledge that lacks awareness, whereas instruction plays an auxiliary role in making that knowledge explicit. ‘Language items, patterns and rules’ therefore speed up the learning process (ibid.).

Having entered the University of Latvia, all of the students that were observed in the present research had a certain implicit knowledge of the English language that they were in the process of making explicit for themselves and others, and some of the activities offered for them that were facilitated by corpus-based materials were created with such intent.

The inevitable conclusion is that among the existing foci in corpus application to studies it is the direct usage of corpora that is of utmost importance to the present research paper, while the theories on learner autonomy in their active role as language researchers whose activities are merely mediated by the instructor and the progression from implicit to explicit command of the English language have served to be fundamental for the research conducted.

1.3. Corpora in the Classroom: Advantages and Disadvantages

With an increase in interest in corpus-based studies in the recent years, more research has been carried out to investigate the application corpora could potentially have specifically in the field of teaching English in an academic context. Among such investigations, it is necessary to point out those by Gavioli and Aston (2001), Thurstun and Candlin (1998) and Conrad (1999), each evaluating the situation from the perspective that they found to be an immediate concern in their respective area of expertise and working conditions.

Gavioli and Aston (2001) polemicise against two common sentiments shared by corpus sceptics, Carter and Gook (1998): firstly, the doubt that corpus analysis is

representative of what they refer to as ‘real English’; secondly, the topicality of linguistic data extracted from a corpus for the purposes of learning English.

Mishan (2004:220) points out that the original form of the texts, both spoken and written, is obscured due to their incorporation into another medium that is foreign to the initial one:

‘What was a blaring 72-point font newspaper headline appears in the same size and typeface as a medicine instructions leaflet; a memo that was originally typed, and advertisement copy that was designed to appear in an arty, colourful font, are now indistinguishable. And more importantly, words spoken from the mouth of a feisty London hairdresser are identical in appearance with words from a dull legal document.’

Acknowledging that an average adult speaker’s knowledge of the language (Gavioli and Aston, 2001) still exceeds the entirety of the contents of British National Corpus (at least as of 2001), Gavioli and Aston go on to argue that such generalisations concerning English language use as ones made by Carter (1998: 43), for instance that the world ‘real’ can only have positive connotations, can be tested using corpus analysis methods and, in the particular case of Carter’s assertion, clearly prove the generalisation to be a hasty one. Additionally, Gavioli and Aston stress the convenience of corpus use in the process of choosing which aspects of vocabulary to dwell on and which to emphasise as being important to learn. However, they warn (ibid.) against relying on corpora when deciding what to teach, instead using them as helpful auxiliary material when making an informed decision.

The present research took the aforementioned piece of advice very seriously, opting against conducting theoretical instruction based on the data found within a corpus and instead offering the students to investigate the contents of the corpus pertaining to their concern of the moment, i.e. writing summaries and summarising, and make appropriate conclusions based on their individual discoveries, all in a setting of autonomy that would be merely mediated by the instructor. That way the pitfall of overt reliance on corpora erroneously as a complete source of all necessary knowledge of information was avoided, in contrast to what is sometimes seen as the shortcomings of the corpus-driven approach (McEnery et al., 2006:8).

Gavioli and Aston (2001) treat the identified problem of corpus data’s representation of ‘real English’ with equal criticality. Confessing that, in the event that a teacher needs to bring a model of conversation at the hairdresser’s, the teacher is unlikely to use concordance lines as a typical sample of such dialogue, they insist that corpora sources can nevertheless be consulted for comparison sake and as a general source of ideas.

Authenticity ought to be distinguished from the notion of genuineness, the contrast originally and eminently described by Widdowson (1978:80) in Mishan (2004:221): ‘Genuineness is a characteristic of the passage itself and is an absolute quality. Authenticity is a characteristic of the relationship between the passage and the reader and it has to do with appropriate response.’ Addressing the supposed insufficiency of corpus use based on Widdowson’s dichotomy between *genuineness* and *authenticity*, Gavioli and Aston (2001) claim that only the former category can be applied to assessing the worth of corpora, while the authenticity of corpus application is not challenged by the fact that corpus learners are not present in the discourses corpora contain as participants, stating that the role of an observer, one capable of reviewing the available data analytically. The importance of subjective, analytical interpretation of corpus data is stressed in their work, and is seen as one of the benefits of corpus usage by students and teachers alike.

Thurstun and Candlin (1998) summarise the application available for concordancing programs:

- ✓ new and exciting directions for developing teaching materials
- ✓ enabling students themselves to make direct discoveries about language
- ✓ an aid to course design

Based on the frequency counts found in the *Micoconcord Corpus of Academic Texts*, Thurstun and Candlin (ibid.) produced a corpus-based list of vocabulary that suits specific rhetoric functions identified, as seen below. They found these ‘semi-technical’ pieces of vocabulary to be of higher importance than highly technical jargon, based on professors’ concern that students produce logically coherent material, rather than simply combine the terms they have to learn in a forced manner. The rhetoric functions and most common lexical units discovered were as follows:

- 1) stating the topic of your writing (factor, issue, concept)
- 2) referring to the research literature (evidence, research, source)
- 3) reporting the research of others (according to, suggest, claim)
- 4) expressing opinions tentatively (may, possible, unlikely, probably)
- 5) explaining procedures undertaken in a study (identification, analysis, criteria)
- 6) drawing conclusions (conclude, summarize, it is clear, thus)’ (ibid.)

According to Thurstun and Candlin (ibid.), students’ awareness of the ways corpora can be implemented into the study process is ‘accompanied by guided opportunities for research, practice and improvisation.’ Similarly to Gavioli and Aston, they emphasise the role corpus analysis tools play in challenging conventional generalisations about language; for

example, the prohibition of utilising the plural *researches* in academic context when corpus data points to the moderate frequency of such usage.

Likewise, Conrad (1999) subjects the restricted number of textbooks on corpora for teachers to criticism, encouraging teachers to not limit themselves to small-scale analyses, such as lexical and lexico-grammatical ones. She proposes that teacher independently focus on complex grammatical and syntactical analyses, citing ‘frequency, syntactic forms of the feature, typical lexical items realizing the feature, and differences in use across language varieties’ (Conrad, 1999) among productive applications for corpus analysis for a teacher. She also points out that students and teachers alike should share their insights into language that they gain from corpus linguistics.

To conclude this subchapter, the presence of both strong empowering elements and unfortunate shortcomings are evident in the application of corpus-based materials and methods to the teaching of English. It is important to remember the aforementioned disadvantages as a potential risk undermining the successful implementation of said materials, and follow some of the pieces of advice offered by defenders of corpus implementation in teaching, such as authenticating materials not originally intended for the newfound purposes, while also discovering own ways of overcoming various obstacles depending on the context and the resources available.

The current chapter of the MA thesis addressed some of the relevant theory existing in the field of corpus linguistics that would be of relevance of the work and guide it in such a direction that the students’ progress and behaviour could be interpreted in light of the theoretical sources examined, allowing for an appropriate evaluation of the research methods and tools chosen and a framework in which the students’ actions could help provide additional insight into the aforementioned issues of corpus linguistics and perhaps also help to resolve them. Unlike the present chapter, the next one will continue the theoretical evaluation of available information of relevance to the present research with an added focus unto EAP (English for Academic Purposes), namely that of academic writing.

2. Corpus-based Teaching of Summary Writing

Summarising, alongside paraphrasing and direct quoting, is a key skill both within and outside the academic framework. As potential readers tend to lack the desired time to thoroughly scrutinise a particular text in its entirety, a summary providing a brief, concise insight into the subject matter can be often seen as an appealing alternative. It is directly intertwined with academic reading via note-taking or note-making (Jordan, 1989: 170).

The Practice of teaching the skill of summary writing has evolved progressively throughout the years; for instance, Johns (1988b) in Jordan (1989: 170) argues that rigid guidelines such as reinforcing the idea of a summary being precisely a third of the original's length may be detrimental to the coherence and quality of the final product. Björk and Räisänen (1997: 127) also destroy another harmful stereotype that every paragraph should have representation in the summary. This is unnecessary unless the paragraph in question contains important information. In the case of summarising a problem/solution text, Jordan (1989: *ibid.*) recommends sorting the ideas according to the categories of situation, problem, responses/solutions and evaluation (Jordan, 1989: *ibid.*). Additionally, Björk and Räisänen (1997: 122), as cited below, share Jordan's view about the importance of main supporting points, yet emphasise that the controlling idea should be indicated. These ideas are grouped into boxes that serve as a framework for the summary to come. The same principle can arguably be applied to summarising other text types, such as causal analysis and argumentation essays (Björk and Räisänen, 1997).

McCormack and Slaght (2009: 24) have designed the so-called *N.O.W. approach*, the abbreviation of which stands for 'Note', 'Organize', 'Write'. Students are advised to begin the writing process by making notes on the key points of the text they are summarising, organising them in a logical order and finally writing a summary based on these notes.

2.1. Features of Academic Writing

As summary writing is part of the discipline of academic writing, many of the features of the latter are therefore applicable to the former. The Using English for Academic Purposes (UEFAP, Online 5), developed and maintained by Gillett A., offers access to an elaborate and highly valuable list of said features that was adopted for students during the pilot lessons, contributing greatly to the number of linguistic features to be learnt (see

Appendix 6). UEFAP (ibid.) divides the features into complexity, formality, precision, accuracy, objectivity, explicitness and responsibility.

Complexity is by far the largest of the categories, including the usage of noun based phrases, subordinate and complement (that-, to- and of+ing-) clauses, sequences of prepositional phrases, participles, passive voice, lexical density, lexical complexity, nominalisation and attributive adjectives.

Formality, on the other hand, entails avoiding the usage of informal lexis, contracted forms and phrasal verbs. Students doing their best to avoid sounding informal are advised to not ask questions, including rhetorical ones, and abstain from ordering items into lists, opting for summary markers instead.

Precision bears the implication of referring to things, people and places directly instead of using vague language, while accuracy is understood as using words in their correct sense, differentiating between shades of meaning properly.

Objectivity advises the students against the usage of personal pronouns, most notably the first person *I, me, my* (opinion), and instead exploit impersonal constructions like *it is evident that...*

Explicitness of academic writing demands that the writer makes his or her point clear by means of discourse markers and signalling; for instance, *however* and *on the other hand*.

Finally, responsibility urges writers to give credit where credit is due, citing the authors correctly and avoiding plagiarism this way.

Generally speaking, academic writing features extrapolate soundly unto the field of summary writing as a genre. Perhaps the sole point of contradiction is the advice to avoid the usage of personal pronouns while using summary genre-specific constructions such as ‘the author suggests’, which would lead one to resort to personalised phrases such as ‘she also adds’ so as to avoid repetition. The less personal ‘it is stated’ summary marker remains a more formal alternative.

2.2. Features of Summary Writing

In addition to the universal guidelines of choosing appropriate linguistic means in order to maintain the fitting academic style of writing, certain features are typical to the genre of summary as such and can be differentiated from academic writing as such.

Björk and Räisänen (1997: 122-123) identify the following features of the summary as a genre:

1. “A summary should include full and clear information about the original text: name of author; title; and, if printed, the title of the printed source: newspaper, journal, book, etc.
2. A summary should indicate the *overall purpose*, i.e. the controlling idea, of the original text.
3. A summary should also include the *main supporting points* of the controlling idea.
4. In a summary you should use *your own vocabulary* as much as possible. You should quote only such words and phrases that cannot be expressed satisfactorily in your own vocabulary. A summary that consists of clauses and sentences patched together from the original is a poor summary: it does not show that you have understood the original text. Remember the old insight: “you have not understood what you cannot formulate in your own words.”
5. A summary should be as objective as possible and *not* comment on the original.
6. A summary should be coherent: i.e. you should indicate how the various points included are related to each other and to the overall controlling idea. It is more important that the summary is coherent than that it follows the chronology of the original.
7. A summary should include summary markers to remind the reader that the ideas are not yours: “The author argues, maintains, suggests...”
8. A summary should, in short, give the reader, who has not read the original text, an adequate understanding of the nature and development of the ideas in that text. “

McCormack and Slaght (2009: 24) distinguish summarising as a means of incorporating evidence in one’s academic writing as being distinct from paraphrasing and direct quotations. Unlike a paraphrase, a summary focuses exclusively on the main idea(-s), omitting minor details, and can be seen as a compressed version of the original text, in contrast to a paraphrase which rarely ever, they argue (2009: *ibid.*), turns out shorter as every detail needs to be rendered. They maintain the view that summarising is preferred to direct quotations as the former approach indicates that the writer has a clear understanding of the text incorporated within their own work. The same view on summarising is shared by Oshima and Hogue (1998: 94): “a summary is similar to a paraphrase except that a summary is shorter.” On the other hand, Jordan (1989: 170) holds the view that summarising is an umbrella term that also includes paraphrasing, claiming that expressing somebody’s views in one’s own manner is only one aspect of summarising.

Summarising is a difficult skill to develop, Jordan argues (1989: *ibid.*); however, it is highly useful as even an academic text that is as original as possible is obliged to meet the formal requirements of referring to previous work in the field by means of referencing and quoting, including paraphrasing and summarising (Campbell, 1990 in Jordan, 1989: 171). According to Campbell (1990: *ibid.*), the unfortunate reality is that students usually opt to copy another piece of work without trying to avoid plagiarism; therefore, summarising is a skill that requires much work to be developed and polished.

As it can be deduced from the present subchapter, there are recurring ideas present that lead one to acknowledge that summary writing is just a narrow branch of academic writing – the conformity to the main idea, the effort to avoid plagiarism by citing the source and using appropriate means for it and the use of formal language all point to the homogeneousness between summarising and academic writing. Perhaps the one reason to single out summarising among other genres found in academic writing is the presence of highly specific summary markers that make it a distinct feature.

2.3. Corpora in Academic Writing Teaching

There has been a recent increase and development of application of corpus materials and corpora as such in the academic context. Researchers in this specific area generally report that corpus-based studies in the academic field provide promising alternatives to the materials teachers are accustomed to, and, whilst there are various issues with their practical implementation, appear to have a bright future. In particular instances, language instructors could use their familiarity with corpora as an effective tool in reaction to specific issues encountered in the classroom.

Fagan (2005: 3) implements a corpus-based approach to mediating the learning of peculiar résumé writing vocabulary to advanced ESL university students in response to the perceived lack of authenticity (or, as she expresses it, the absence of clarity in testifying such authenticity) in the selection of exemplary list of relevant vocabulary for students to use. She stresses (Fagan, 2005: *ibid.*) the importance of showing the students samples of authentic usage of appropriate terms as opposed to limiting the instruction to a simple list of words that student would struggle to utilise in their writing.

Fagan's research is highly topical for providing an example of effective corpus integration in an academic context in response to occurring problems; however, it can be argued that no list can exhaust the virtually limitless number of expressions and combinations

students can produce independently in the process of improving their writing. It could be hypothesized that the teaching of general corpus skills would enable the students to make use of the corpus for various purposes at their own convenience when such need arises. It is presumable that the level of independence the students exhibit in the natural process of language acquisition could affect the effectiveness of the implementation of corpus-based studies where the teacher adapts the role of a mediator, requiring additional direction and interference for less independent students.

Yoon (2008) introduces L2 students of different educational backgrounds to the COBUILD Corpus to assist them in their independent writing process, asking students to email their corpus searches to him and combined them for the mutual benefit of the entire group. Students made use of the corpus to solve their sentence-level writing problems and in addition to that carried out error correction of their groupmates with the help of the corpus. Special attention was directed to the attitudes of the participants and the transformation of their views on language.

Yoon (2008:45) emphasises the importance of language chunks available via corpora in enriching students' repertoire of English expressions, adding that independence and self-monitoring benefit greatly from corpus-based activities.

In a long study such as Yoon's, the idea to survey students' search queries is highly effective for monitoring the students' use of the corpus and providing feedback on the possible ways how a corpus can aid the learners with solving specific linguistic issues. Though the effectiveness of this method declines with a lack of continuous guided corpus-based studies; supervising students' search entries could still be of great help for improving future focus in corpus-based studies and adjustment of own instruction to highlight such uses of corpus that are elusive to inexperienced corpus users.

Kaur And Hegelheimer (2005: 298) point out that with Coxhead's Compleat Concordancer introduced to the them, the students are inclined to consult the dictionary instead so as to remain in the comfort zone and also given the existing time constraints. They nevertheless deduce that out of two groups, one familiarised with the concordancer and the other not, the former demonstrate a higher number of correct word use in their writing assignments, thus the concordancer helps them in their study of word knowledge.

Kaur et al. (2005) provide the following guidelines for improving the situation:

- Guidance and motivation
- More time

- More detailed instruction

It can be concluded from this chapter that the application of corpus-based methods has produced the desired visible results in different contexts; however, this application does not avoid the inevitable issues of which different researchers warn. Those include, namely, the difference in students' language proficiency levels, insufficient time for efficient implementation of corpora, lack of motivation and inadequate instruction in light of the students' habitual predilection for the familiar dictionaries over the newfangled corpus technology. Therefore, a more detailed and also accessible instruction about the practical application of corpora and an emphasis on the nature of corpus application in contrast to that of the dictionary could help the students feel more comfortable about using corpus data for self-monitoring during the writing tasks.

3. Corpus-based Research of Student-Composed Summaries

3.1. Methods and Research Tools

The most suitable research method for the current research is the **case study**, as it investigates the behaviour of a particular social group of people in a specific context, the subjects of the research – first year students of English philology. According to Yin (2003: 1), it is one of the most challenging research methods to be applied in social studies.

Yin distinguishes case studies from other research methods on the basis of the former method's tendency to aim to respond to the questions 'why' and 'how' (ibid.), leading to research that, in its nature, is either explanatory, exploratory or descriptive. Since the goal of the present research paper is to investigate the application and success of corpus-based materials and tasks in the classroom when teaching summarising skills and academic writing skills, it attempts to explore the progress made by students and measure how successful such application turned out to be, describing the research process, thus fitting the description.

3.2. Research Subjects and the Pre-Questionnaire

The research subjects –first year students of English Philology BA at the University of Latvia - were largely unknown to the conductor of the research prior to the beginning of the piloting lessons, aside from the information about the groups offered by their professor. It was known that there are five groups total – 1A, 1B, 1C, 1D and 1FIN (a term that further on will refer to the English-Finnish group specialising in the English and Finnish language and culture studies), and that the participants within the same group, much like the researcher remembers himself several years ago, were divided in accordance with their performance in the state exams in secondary school, thus creating an expectation that the students from groups 1A and 1B would be most competent. Additionally, Online 1 and Online 2 show that prior to summarising texts, the students had learnt how to write paragraphs and essays. However, it was not until the pilot lessons unfolded and the pre-questionnaire was administered that the research subjects were made acquaintance with.

As the researcher / instructor was warned by the students' professor beforehand, the groups' willingness to socialise was generally diametrically opposite to their performance. Groups 1A and 1B were able to perform the best while not responding much to the instructor.

Meanwhile, their peers in the less competent lower groups were more outgoing than their peers, and it was harder to focus their attention on the instruction and prevent them from spending time chatting among themselves and using social networks in-between or even in the midst of the assignments. They were equally talkative with the instructor as well, being very eager to ask questions and answer the ones posed by the researcher, approach the instructor after the class to clarify the less transparent aspects of the instruction and even ask for advice with their own research to come.

The analysis of the Pre-Questionnaire, first of the two questionnaires administered to the five groups, sought to discover the students' background, their perceived difficulties when doing academic writing and prior familiarity with corpora.

Knowing that the exact number of students in each group cannot be adequately measured from first-hand experience, with little clarity concerning which students are temporarily inactive for individual reasons or entirely gone from the course without notifying the university authorities, no such attempt was made; however, upon analysing questionnaires and the written pieces of work produced by students during the course, statistics were gathered to compare the number of students in each group that filled in the questionnaires and the number thereof that submitted the summary assignments before the set deadline. Both in the tables below in the current subchapter and in other sections of the present paper, 1D/FIN represents the students of groups 1D and 1FIN (the so-called English-Finnish group) as a unity, as certain assignments made it more difficult to differentiate between these two groups.

Table 3.1. Questionnaire (Pre- and Post-) Turnover.

Group	Pre-Questionnaire	Post-Questionnaire
1A	15	10
1B	11	6
1C	12	16
1D/FIN	17	13
Total	55	45

Table 3.2. Summary (Pre-, Revised and Final) Turnover.

Group	Pre-Summary	Revised Summary	Final Summary
1A	14	15	16
1B	8	8	8
1C	9	9	9
1D/FIN	14	14	15
Total	45	46	48

Tables 3.2.1 and 3.2.2 depict the number of students in each group that filled in the questionnaires and submitted three of the summary activities on time (the first one submitted during the first lesson, the latter two – entered into the e-studies system online at home).

As it is evident from the tables, the attendance dropped substantially for groups 1A, 1B and 1D/FIN, since attendance is directly connected with the capacity to fill in the questionnaire, whereas completing the assignments is slightly more demanding. In particular, the attendance of group 1B was reduced nearly in half, which could be interpreted either as a random occurrence or a decreasing interest in the subject matter. Group 1C, on the other hand, increased from 12 to 16 students if questionnaires 1 and 2 are contrasted. The overall number of students who were present to fill in the provided questionnaire dropped from 55 to 45.

The number of students who submitted the summaries in a timely manner, on the other hand, remained mostly static, though it ought to be noted that 9 students in group 1C, for instance, were not necessarily the same students, as there were instances when specific students failed to submit either the pre-summary or the revised summary, possibly misinterpreting the given assignment.

The order of questions posed in the questionnaire attempted to first determine the national and academic background of the students to find out whether a substantial part of the students came from different cultural backgrounds and if anybody has had prior tertiary education to the English philology Bachelor’s programme, followed by three specific questions on academic writing in which students were to evaluate themselves according to offered criteria and pinpoint what they perceived to be their weaknesses, finally rounding up with a question that posed the issue of corpus familiarity, as one of the goals of the research was to convince the students to continue using the corpus upon initial familiarisation with it.

Table 3.3. Questionnaire 1: What is your native language?

Group	Latvian	Russian	Other
1A	14	1	0
1B	5	4	2
1C	6	6	0
1D/FIN	12	5	0
Total	37	16	2

Table 3.4. Questionnaire 1: What is your current education?

Group	Secondary	Percentage	Other	Percentage
1A	15	100%	0	0%
1B	10	90.9%	1	9.1%
1C	11	91.7%	1	8.3%
1D/FIN	16	94.1%	1	5.9%
Total	52	94.5%	3	5.5%

Table 3.2.3 depicts the native language of all students which in all instances but two was either Latvian (37 students) or Russian (16 students). Only two students joined the course from abroad and both of them were from the same group, that being 1B. The two students came to study to Latvia from Italy and Czech Republic, which was reflected in their choice of language for the question. The results of this particular question of the questionnaire testify that research subjects come from generally the same academic background, an overwhelming majority of them being compatriots, and thus form a largely homogeneous group.

Table 3.2.4, on the other hand, is preoccupied with the students' educational background and shows another overwhelming majority in this research, this time showing that nearly all the students had only finished secondary education prior to the course. Group 1A consisted entirely of students who had graduated from school, whereas groups 1B, 1C and 1D/FIN (not distinguished among one another during the questionnaire piloting process) all featured exactly one student with another higher education, compromising 9.1%, 8.3% and 5.9% respectively for a total of 5.5% students with other higher education, in contrast to 94.5% students with secondary education, which is over 17 times the number.

Table 3.5. Questionnaire 1: How would you evaluate the difficulty of academic writing

Group	Highly	%	Mod.	%	Tolerable	%	-	%
1A	1	6.7	12	80	2	13.3	0	0
1B	2	18.2*	9	81.8*	1	9.1	0	0
1C	0	0	10	83.3	2	16.7	0	0
1D/FIN	5	26.3	9	47.4	5	26.3	0	0
Total	8	14.5	40	72.7	10	18.2	0	0

(respectively – highly challenging, moderately challenging, tolerable, not a problem at all)

* - a student in group 1B ticked two answers in spite of the instructions to tick just one.

Table 3.2.4 illustrates the results of the first question present in the questionnaire that deals with students' attitudes towards academic writing. In this closed-ended question students were asked to evaluate the difficulty of academic writing, offered 'highly challenging', 'moderately challenging', 'tolerable' and 'not a problem at all' as the options.

As was the case with most of the questions to come, the students opted against voting for extreme answers, which in this question was expressed by 72.7% students choosing the ‘moderately difficult’ option to evaluate the difficulty of academic writing for themselves. The group with the largest discrepancy by far was 1D/FIN with only 47.4% of the students selecting this option, in contrast to the other three groups where the same option was represented by just above 80% of the respondents. The rest of the votes in group 1D/FIN were evenly divided between ‘highly challenging’ and ‘tolerable’, with 5 students ticking each box, or 26.3% of all students in that group. In spite of being one of the two poorest performing groups, the other being 1C, this group was the most optimistic with the largest percentage of students claiming that academic writing was merely tolerable for them.

On the other hand, the opposite extremity was also true, as group 1D/FIN also had the largest percentage voting for ‘highly challenging’ among the four groups. In the other weaker group, 1C, there was not a single student to have selected ‘highly challenging’ out of 12 respondents (a number that increased for the second questionnaire). It could perhaps be argued that the better-performing students were more aware of the possible issues that arise in the process of academic writing or set higher standards for the final product before themselves. Alternatively, the disproportionately high confidence of groups 1C and 1D/FIN could be attributed to their generally good confidence, such as witnessed in verbal communication with the instructor, while the stronger groups could be seen as more reserved, patient and possibly less secure. It is worth noting that not a single student in either group selected the option that academic writing is not problematic at all, so a certain level of homogeneity can be observed nevertheless.

Table 3.6. Questionnaire 1: What aspects of academic writing would you most like to improve on?

Aspect	1A	%	1B	%	1C	%	1D	%	Total	%
Coherence	2	13.3	1	9.1	3	25	5	29.4	11	20
Vocabulary	6	40	3	27.3	0	0	6	35.3	15	27.2
Register	5	33.3	2	18.2	3	25	4	23.5	14	25.5
Referencing	2	13.3	0	0	1	8.3	1	5.9	4	7.3
Punctuation	1	6.7	3	27.3	0	0	1	5.9	5	9.1
Essays	0	0	0	0	0	0	3	17.6	3	5.5
None	1	6.7	3	27.3	2	16.7	0	0	6	11
Other	3	20	3	27.3	5	41.7	4	23.5	15	27.2

The fourth question (Table 3.2.6) posed aimed to discover the exact areas of academic writing students have most trouble with. 27.2% of the students or 15 people out of 55 cited

vocabulary as causing them the most trouble, which was followed by register and formality issues in topicality (selected by just one respondent less). The third most popular option was coherence, selected by 11 respondents or exactly one fifth of the students. This was followed by punctuation (9.1%), referencing to avoid plagiarism (7.3%) and essay writing, which was chosen by 3 students in the group 1D/FIN. 6 students out of 55 failed to offer an answer, which could be partially explained by the intentional choice on the researcher's behalf to pose this question before question 5, which did mention some of the potential problems. The goal of this sequencing of questions was to enable the students to mention any specific problems they may have had with academic writing, whether or not it was covered in the following question, and have a fuller reflection of students' needs than the researcher could have deduced individually. For instance, 'keeping the writing impersonal', as suggested in the following question, was one of the least popular answers alongside others in the 'Other' group. Likewise, the most popular option of 'vocabulary' was not mentioned at all in the following question, allowing the students to freely express their concerns with regards to academic writing without being affected by the researcher's prompts.

The 'Other' category featured a variety of different answers, having received 15 students' responses – the same number as the most popular response 'vocabulary'. Among the answers united in this category, group 1A offered issues with sounding impersonal as well, grammar and coming up with ideas during the pre-writing stage. Group 2B also mentioned the pre-writing stage, possibly influenced by the essay writing lessons that preceded the piloting of corpus-based summary writing, also mentioning trouble with capitalisation and summarising as such, perhaps to fit with the stated goal of the studies. Group 1C provided such unique answers as thinking analytically, choosing appropriate stylistic means, as well as difficulty related to grammar, coming up with ideas and summarising in general. Group 1D/FIN, finally, singled out stylistic concerns, staying impersonal in one's writing and one student mentioned spelling as a relevant concern. Though some of the answers repeated at least once, they were all united in the 'Other' group for their rarity. In general, the students were not affected greatly by the following question on the list.

In Tables 3.7.1 – 3.7.4 seen below, the students were requested to evaluate themselves on a scale from 1 to 5, 1 being poor and 5 – highly proficient, regarding the following: maintaining the formal register, keeping the writing impersonal, using abundant and appropriate discourse markers and avoiding errors leading to plagiarism while quoting another source. The tables depict the number of students that picked each number, also expressed in

percentage form, as well as the average score computed by calculating the average of all scores within a group as well as for all research subjects regardless of groups.

Table 3.7.1. Questionnaire 1: Maintaining the formal register

Evaluation	1A	%	1B	%	1C	%	1D	%	Total	%
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	8.3	0	0	1	1.8
3	7	46.7	7	63.7	5	41.7	7	41.2	26	47.3
4	7	46.7	4	36.4	6	50	9	52.9	26	47.3
5	1	6.7	0	0	0	0	1	5.9	2	3.6
Average	3.6		3.36		3.42		3.65		3.51	

Table 3.7.2. Questionnaire 1: Keeping the writing impersonal

Evaluation	1A	%	1B	%	1C	%	1D	%	Total	%
1	0	0	0	0	0	0	0	0	0	0
2	1	6.7	0	0	0	0	2	11.8	3	5.5
3	3	20	2	18.2	1	8.3	5	29.4	11	20
4	9	60	8	72.7	9	75	9	52.9	35	63.6
5	2	13.3	1	9.1	2	16.7	1	5.9	6	10.9
Average	3.8		3.91		4.08		3.53		3.83	

Table 3.7.3. Questionnaire 1: Using abundant and appropriate discourse markers

Evaluation	1A	%	1B	%	1C	%	1D	%	Total	%
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	2	11.8	2	3.6
3	11	73.3	7	63.7	8	66.7	10	58.8	36	65.4
4	4	26.7	3	27.3	3	25	3	17.6	13	23.6
5	0	0	1	9.1	1	8.3	2	11.8	4	7.3
Average	3.27		3.45		3.41		3.29		3.35	

Table 3.7.4. Questionnaire 1: Avoiding errors leading to plagiarism while quoting another source

Evaluation	1A	%	1B	%	1C	%	1D	%	Total	%
1	0	0	0	0	0	0	0	0	0	0
2	2	13.3	0	0	1	8.3	0	0	3	5.5
3	3	20	7	63.7	4	33.3	8	47.1	22	40
4	9	60	2	18.2	5	41.7	7	41.2	23	41.8
5	1	6.7	2	18.2	2	16.7	2	11.8	7	12.7
Average	3.6		3.55		3.67		3.65		3.62	

In contrast with the preceding question, this particular part of the pre-questionnaire aimed to see how the students were ready for summarising prior to their theoretical

instructions. All the four criteria refer to particular issues or skills related to summary writing and also academic writing as such, allowing the research to compare students' expectations shown in their self-evaluation to the results of the corpus study of the summaries they had produced.

It is visible from Tables 3.7.1 – 3.7.4 that students were sober in their personal evaluation according to the offered criteria, with only one reaching an average rating of above 4 – this was group 1C regarding their ability to keep the writing impersonal. Once again, it is observable that the research subjects abstained from opting answers expressing an extremity – this can be seen in the general absence of 1, 2 and 5, with 3s and 4s making the majority of scores chosen by the students.

For all four (five if separating group 1D from the first year English-Finnish group 1FIN) groups, keeping the writing impersonal appeared to be the easiest task, with an average score of 3.83, though group 1D/FIN ranked lower compared to the other groups with as many as two students giving themselves a 2, which exceeded the number of 2s in the three other groups put together. 1D/FIN were nowhere as critical concerning their ability to maintain the formal register however, giving themselves an average of 3.65 – the highest among the groups polled. The average for this problem was only 3.51, group 1B somehow rating themselves the lowest. The second highest average score among all groups was that for avoiding errors leading to plagiarism. The average for it was 3.62 and all groups rated themselves approximately the same (group 1A – 3.6, 1B – 3.55, 1C – 3.67, 1D/FIN – 3.65). The results reveal that one of the biggest summary writing-related problems students face was the usage of discourse markers in abundance and appropriately in the context given (see Table 3.7.3), with scores above 3 being very sparse for all respondents. This was the only criterion of evaluation where 3s exceed 4s and quite significantly so – 65.4% against 23.6%.

With 3s and 4s being the highest scores, it is interesting to note the way they correlate – the number of 3s and 4s is even or almost even for maintaining a formality level and avoiding plagiarism, while the other two criteria divided the students greatly. 63.6% of the respondents gave themselves a score of 4 for their ability to keep the writing impersonal, compared to 20% who opted for a 3 instead. The exact opposite happened concerning the abundant and appropriate usage of discourse markers, as already outlined above.

Table 3.8. Questionnaire 1: Are you familiar with language corpora?

Group	Yes	%	No	%
1A	1	6.7	14	93.3
1B	4	36.4	7	63.6
1C	1	8.3	11	91.7
1D	0	0	17	0
Total	6	10.9	49	89.1

The final question (Table 3.8) asked in the pre-questionnaire inquired the students regarding their prior acquaintance with corpora, without specifying whether they simply heard some description during their lectures or elsewhere or have had exploited any of the existing corpora for various purposes, allowing those who provide an affirmative response to explain where their acquaintance springs from in an open-ended space following the question.

The results of this question were unambiguous – the vast majority of students responded to the question negatively, with one student with some acquaintance with corpora in groups 1A and 1C, none in group 1D but 4 students out of 11 polled in group 1B, which was among the best-performing, or 36.4% of all students in that group. This is contrasted with only 10.9% students total among groups who filled in the pre-questionnaire. The explanation is rather simple: though some of the students may have brought up that the mention of corpora was part of a lecture they were present at, the lack of direct experience with a corpus might have prevented some of their peers from remembering the same part of the lecture, as the corpus described was difficult to connect with the corpus brought up in the pre-questionnaire which, it should be mentioned, was not introduced in any way prior to the theoretical instruction and practice with the BNC engine of the COCA Corpus.

It is important to remember the 89.1% of students who could not recollect any familiarity with corpora, as the piloting changed this trend significantly, largely affecting students' attitude towards and expectation from corpora further in their studies and active work.

As it is evident, most of the background information relevant to the case study gathered regarding the students tracks to the pre-questionnaire whose statistical and qualitative insight is referred to and addressed later in the work, checked by students' own work during the piloting of the corpus-based materials.

3.3. Corpus-based Teaching Materials

Both the teaching materials and the research process have been produced with a direct assistance of corpus-based methodology. Firstly, the samples for the approximately 30-minute long theoretical instruction on summarising and features of general summary writing were all found using the Corpus of Contemporary American English (COCA Corpus). Second, during the first of the two pilot lessons (for each of the four groups), the students were taught to log into their COCA Corpus accounts that they had just created and conduct a few simple searches, which were tied to what the researcher viewed as useful applications of the corpus to the solution of immediate issues of linguistic usage. As British English is the standard variety of English for the English philology students at the University of Latvia, the BYU-BNC engine of the COCA Corpus was utilised so that the students would not be misled by the difference of the American English which is observable in the main COCA Corpus, even though there is no doubt that most of the students are familiar with the variety in one way or another. Finally, a learner corpus was produced out of the three summaries produced by the students during the piloting process, with an aim to investigate the success of the integration of corpus-based materials in the teaching of summary writing.

The Corpus of Contemporary American English is the most sizeable corpus of those that are available freely to all (Online 8), and claims itself to be the only corpus of American English that is large in addition to boasting a satisfactory level of balance (*ibid.*). The sheer size of the COCA Corpus comprises over 450 million words recorded in genres as diverse as ‘spoken, fiction, popular magazines, newspapers, and academic texts’ (*ibid.*), all sorted in divisions that permit its users to pinpoint the context or register most commonly associated with a desired unit of language or linguistic problem the autonomous researcher attempts to resolve. According to the information offered to users in Online 8, in addition to searching by genre and over time, the users of the COCA Corpus are able to enter any combination of words or phrases, lemmas, part of speech, wildcards as a query, or enter one separately.

The closest equivalent to the COCA Corpus among the freely available corpora online is the British National Corpus (BNC); however, it loses to the former in size considerably, namely around 4 times (Online 9). Since the BNC search engine has been made part of the website of the Corpus of Contemporary American English, the cited page on COCA Corpus’s site provides a comparison between these two major corpora. What they share is a high level of balance and are very sizeable; on the other hand, the COCA Corpus is a clear victor in terms of topicality of the texts presented, with all of them belonging to the past two or so

decades, whilst the BNC has not been updated in 20 years, resulting in inferior topicality of texts found within and a glaring absence of such words found commonly in contemporary usage as ‘download’, ‘globalisation’ or ‘terrorist’ (ibid.). The BNC compensates for this shortcoming by covering less formal, mundane conversation as represented by a greater variety of spoken sub-genres, though it is still highly recommended to make use of both (ibid.).

Though the comparative obsolescence and a niche favouring a plethora of spoken sub-genres typical of the smaller British National Corpus are reasons against choosing the BNC engine for the students of English philology who aspire to write in a manner resembling the standard observed in Great Britain, it is precisely the limitation to one variety of English, appropriate in the given academic context, that served as the reason for choosing the BNC engine over the larger COCA Corpus offering a fruitful focus on the American English variety.

The first usage of the COCA Corpus in the work saw presence in the creation of materials for the theoretical instruction on summary writing, where features of academic writing were adapted from the online resource UEFAP (University Centre for Computer Corpus Research on Language, Online 5); however, all the examples were taken from the COCA Corpus by conducting queries that would result in the expected kind of sentences, an activity usual for the research as an avid user of corpora since studying at Bachelor’s level.

The samples from the COCA Corpus were used namely to vividly explain the features of summary writing and related academic writing aspects, for example:

- *A poll commissioned by someone a while ago showed that much of the population declared themselves to be broadly satisfied with living in the Albion, more compared to some other European countries.*
- *A poll commissioned by the EEC in 1977 showed that 82 per cent of the population declared themselves to be broadly satisfied with living in Britain, compared with 68 per cent in France and 59 per cent in Italy.*

The second sentence originates from the COCA Corpus, whereas the first one was modified by the instructor / researcher to teach the concept of precision, drawing clear contrast between writing lacking in precision and excelling in such. Numerous examples were used for other features of academic writing cited in the theoretical chapter 2.1 of the present research, and the presentation ended with a couple of slides with information on the COCA Corpus, which immediately lead to the corpus-based activities that students would engage in for several assignments to come.

3.4. Research Procedure and Results

Firstly, upon filling in the pre-questionnaire, the students were asked to write a summary prior to being subjected to any theoretical instruction, which certain students found a dubious decision and questioned the instructor aloud. The goal of preceding the presentation with a pre-writing part was related with both pedagogical and research objectives – first of all, having to produce a text in a certain genre that is not entirely familiar to the students arouses natural curiosity expressed in wanting to learn the precise rules and conventions of such kind of writing. Those students who struggled with the pre-writing part in particular would be expected to pay extra attention to the presentation and memorise the features displayed on the slides with additional attentiveness. What this decision contributed to the research was the possibility of comparing each group's performance prior to and following the presentation and the connected activities.

The lesson plan for the first contact with the groups can be found in Appendix 3 to the present paper, showing the aims and stages of the premiere lesson in full. The groups were allotted 35 minutes of time to complete this assignment, summarising an online text giving valuable advice on staying competitive on the job market, containing a useful peace of advice particularly for this age group, for whom finding a job is a topical issue. The students generally worked independently and were asked to stop writing when the allotted time came to its end.

In general, the students' pre-summaries left a very lasting, positive impression. The research subjects made effort to avoid plagiarism by utilising synonymy pairs, for instance using the word 'annoying' instead of 'irritating' or 'persistent' instead of 'resilient and determined'. They were often conscious of register, using expressions such as 'The second part indicates', 'Another suggestion is' or 'it is very advisable', and made frequent use of passive voice constructions along the lines of 'a few ideas to be thought about'. They were often lexically rich, one example being 'it would extinguish the employer's suspicion'.

There were certain problems with the pre-summaries too, however. First of all, not all students managed to ensure a word count of approximately 200 words as was asked of them. Some of them also used inappropriate formatting such as numbering, and many of them used an unnecessary abundance of paragraphs.

Another issue was the inclusion of unnecessary details, which goes against the rules of summary writing which requires only the main supporting points to be included. Furthermore,

students were not as impersonal as it had been expected of them, often resorting to first and second person pronouns such as ‘I’ and ‘you’.

In conclusion you have to remember that always have to ask questions because it shows that you are really interested in the job.

As seen from the quote above, springing from one of the students’ pre-summaries, there was also an issue of stated conclusions not the content matching such conclusions. For instance, the cited sentence states it is the final or main idea of the article that asking questions is important, although it is not either of these but instead merely one of the pieces of advice offered.

The pre-summaries written at this stage of the piloting process often resembled the following:

„Scoring a job: Interview preparation”

The main idea of the text is - if someone wants to get a job he or she should think of number of things on how to prepare for the interview and what to say when being interviewed.

The very first thing to do is to explore the company where someone is to work in. This makes him or her much more competent and unique among other competitors. Learning about the company and it’s structure also allows person to find it’s weak places and propose a solution for the problem.

Another good thing to do before the interview is to read carefully the job specification. It gives a person an opportunity to get ready for specific questions.

It is also good to mention your special skills and previous experience. It is no matter is it bad or good one. The main point of it is to show your abilities and that you are capable of being

If there is someone at home it is also a good idea to make a fake interview at home with that person. It gives a possibility to hear of what you are actually saying and again, to get ready for unexpected questions

The last but not least to do is to ask for additional questions if there is such an option. This shows that you are interested in the job and that

The showcased summary belongs to a student from 1C, one of the poorer-performing groups that took part in the research. As it can be observed in the summary, understanding the text and paraphrasing the information did not pose immense difficulty even to the weaker students, though such areas as grammatical accuracy, choice of appropriate register, word order and repetition were some of the recurring issues in the groups that did not quite reach

the same command of writing formally as the stronger 1A and 1B groups. In contrast, a typical pre-summary from group 1A resembled the following:

Scoring a job: Interview preparation

The Independent's student guide to snagging that first job

Before the interview, you should get to know the company better – how it functions, what makes it unique, what competitors it has. You should also be able to tell why you've chosen that company, not any other, what it could improve on, what solutions you could offer. Your dedication could show them that you are more valuable and innovative than other candidates. If you read the job/person specification thoroughly, it should show you the required skills and from that you can derive what kind of questions might be asked in the interview. Afterwards, you should prepare scenarios, in which you demonstrate the following skills: communication skills; leadership skills; initiative; problem-solving skills; organisational skills, commitment. Don't forget to include a "bad" story, an occasion in which you failed; it would extinguish the employer's suspicion that you've somehow too perfect. Don't forget to act it all out in mock interviews. It's great for obtaining feedback. At the end of the interview, use the opportunity to ask questions – ask them everything you're curious about. If at the end you're not chosen, do not get discouraged. You ought to be determined to get a job in these times. Think about your strengths, weaknesses, what makes you special, your experience, to fully prepare for the interview.

Though preceding the theoretical instruction in time, the cited summary from group 1A exhibited some of the traits that were later pointed out in the presentation featuring aspects of summary and academic writing, such as being one single paragraph and including the full title of the original text. This summary is also arguably better-written than the previous one, featuring greater grammatical precision; however, the writing is also noticeably informal in some of its elements, which may have not been obvious to its author: the abundance of the second person singular pronoun 'you' is a major sign of informality that is inappropriate to summary writing as a genre of academic writing. The level of writing exhibiting generally correct word order, having good language and syntax was a common trait shared among the students in groups 1A and 1B.

Scoring a job: Interview preparation

The Independent's student guide to snagging that first job

Presummary

This text is about how to prepare for a job interview. Before going to a job interview it is important to do a research on company that you are applying. Even if you don't have experience you have to find a way how to make yourself stand out. Think of some kind of possibly innovative solutions you could offer the company. It is important to be prepared for any questions your potential employer might ask. Also understanding what they expect from you will be useful as well. That way you will know what to offer.

Not being afraid to show that you are suitable for the job is important as well. Telling what kind of skill you have might help in that case. It is important to keep in mind that everybody makes mistakes, so you should not be afraid to show you have „bad” stories as well.

Doing a dry run might help to see what is working and what not.

At the end of the interview it is important to ask questions, because that way you will look more intelligent and interested in vacancy. If you don't get the job, don't feel discouraged, because you never know what was the reason for employer to have picked someone else.

To cite one last pre-summary, this time from the English-Finnish group, the poorer-performing groups were at times such as this better aware of the need to introduce other people's ideas with phrases such as “the main idea is” or “the text is about”, showing some other aforementioned inappropriate aspects in summarising such as the usage of the second person singular pronoun ‘you’ and its derivatives, and featuring general errors such as misspelling certain words (*potentian* for potential, *afraid* instead of *afraid*), using contracted forms like “don't” instead of full “do not”, omission or inappropriate usage of articles.

In general, the pre-summary stage revealed that the students in all groups are not lost without instruction, capable of producing work expected of their level, though some of the research subjects manage to accomplish this at a faster pace and with better quality than others, which is naturally expected. This stage also prepared the students to listen in for the correct features of summary writing and academic writing in the theoretical presentation that followed the pre-summary writing process immediately.

The presentation showcased the slides with brief theory on summary writing (see the theoretical subchapter 2.2), the entirety of which (the slides) is available in Appendix 6 of the

current MA thesis, and the students were also given handouts with a sample summary with peculiar features of summarising underlined, as seen in Appendix 2.

Following the theoretical instruction, the groups were asked to register on the COCA Corpus page (Online 8) where they repeated sample queries offered to them by the researcher.

Figure 3.1. *Registering at the COCA Corpus webpage*

Please fill out the brief form below. Within one or two minutes, you will receive an email. Simply click on the link in that email, and you will be able to continue using the corpus. (Note: If you have already registered, please [log in](#) to see your profile)

Name	<input type="text"/> <input type="text"/> (e.g. Mary Smith)
Email address	<input type="text"/>
Password	<input type="password"/> (you will use this to log on to the corpora)
Category	<ul style="list-style-type: none"> <input type="radio"/> University professor (or post-doc): languages/linguistics 3 RESEARCHER <input type="radio"/> Graduate student: languages or linguistics <input type="radio"/> University professor (or post-doc): not languages/linguistics 2 SEMI-RESEARCHER <input type="radio"/> Teacher: not university; not graduate student <input type="radio"/> Creator of language-related blog, or professional translator <input type="radio"/> Graduate student: not languages or linguistics 1 NOT RESEARCHER <input type="radio"/> Student (undergraduate) <input checked="" type="radio"/> Other
<input type="button" value="SUBMIT"/> <input type="button" value="RESET"/> PROBLEMS ??	

As depicted in Figure 3.1, students in every group register for their COCA Corpus accounts (which also provides them with access to the BNC engine of the COCA Corpus). This process is remarkable simple as students are given the link to the corpus’s webpage, affected only slightly by some of the computers in the classroom freezing or performing inadequately. Some time is lost as a result of the corpus not giving its users immediate access to the search function upon registering, requiring them to browse the recent corpus-based research hosted on the website. In spite of the fact that simply following the link and then backtracking to the corpus circumvents this necessity, which would be too time-consuming under already stiff time restraints limiting the introduction to the corpus, a certain amount of precious was lost without any chance of recovery.

Figure 3.2. Entering a query 'ubiquitous' into the corpus



Once logged into the corpus, the students were offered to enter a query into the search bar – perhaps the simplest of the search functions available. The researcher thought it would be fitting to offer to search for a highly obscure word and asked the group to choose one (which the researcher required assistance with spelling correctly at times), while the one group that hesitated to mention a word were simply offered the above search of ‘ubiquitous’ (Figure 3.2). Table 3.3 pinpoints the concordance lines whose importance in seeing the context was emphasised to the students. The young people were given some time to read the examples of real language in use, trying to understand the obscure (or for some, easily forgotten) word.

Figure 3.3. The concordance lines resulting from the entered query

KEYWORD IN CONTEXT DISPLAY										Help / Information / contact
SECTION: NO LIMITS										PAGE: << < 1 / 26 > >> SAMPLE: 100 200 500 1000
CLICK FOR MORE CONTEXT										[?] SAVE LIST CHOOSE LIST ----- CREATE NEW LIST [?]
1	2012	FIC	Bk:AnAmericanSpy	A	B	C	platform, bumping into hunched men lighting cigarettes, he gazed up at the freshly ubiquitous spiderweb ceiling of steel and glass. How much h			
2	2012	FIC	Bk:AnAmericanSpy	A	B	C	, the idea of " the other " becomes faceless and broad, its tentacles ubiquitous enough to hide in every crevice. # " So what do you want from			
3	2012	FIC	Bk:VariationsNovel	A	B	C	a girl so thin the white extravagance of skin: so clean and helpless and ubiquitous . So much flesh in the world, all forbidden to him. How could			
4	2012	FIC	Bk:VariationsNovel	A	B	C	a lack of realism by those who expected him to be if not omniscient certainly ubiquitous , tending to each one individually all at once. " Be here,			
5	2012	FIC	Bk:ReeducationCherry	A	B	C	request to see the city her parents once knew, before the Microsoft billboards and ubiquitous Kodak photo shops. In the Cholon district, Granda			
6	2012	FIC	Analog	A	B	C	starling life history, and by extension, about how one might begin eradicating this ubiquitous bird. Within a few days she, too, was driving the de			
7	2012	FIC	NewEnglandRev	A	B	C	absolutely every soap opera. Infidelity, along with alcoholism and drug abuse, is ubiquitous , the most prevalent open secret of most present-da			
8	2012	FIC	AmerTheatre	A	B	C	We bear the automated voice at the end of the line - it is the ubiquitous female voice for most banks, doctors offices, etc. # Timothy - sharply			
9	2012	MAG	Astronomy	A	B	C	928757069 Small, dark, and ubiquitous , Bok globules offer astronomers an inside look at how low-mass stars form. #			
10	2012	MAG	CountryLiving	A	B	C	squares on walls and wearing them as headbands or scarves. Such textiles were as ubiquitous then as campaign bumper stickers are now. You			
11	2012	MAG	Esquire	A	B	C	the Embody chair to accommodate the length of your thighs. # 2. The ubiquitous Aeron chair supports the natural lean of your pelvis. Plus, it's			
12	2012	MAG	HarpersBazaar	A	B	C	Meadham Kirchhoff are turning heads. # Celebrations of London's vibrant sartorial scene are ubiquitous this season, from exhibitions in the new			
13	2012	MAG	Newsweek	A	B	C	how can he solve it? " # The biggest reason negative ads are so ubiquitous in politics, but much less common in commercial advertising, is this			
14	2012	MAG	PCWorld	A	B	C	being developed with touch in mind, and Windows 8 is sure to become a ubiquitous operating system, meaning that software developers who d			
15	2012	MAG	PCWorld	A	B	C	# FIRST WE HAD laptops and PDAs. Then the iPhone led the way to ubiquitous smartphones, and the iPad ushered in the era of tablets. These c			

As the choice to display a list of concordance lines frequently proves to be insufficient for the purpose of efficiently revealing the meaning of the word and traits of its issues, the students were asked to tick ‘KWIC’ (Keyword in context) instead of ‘list’. The consistently

coloured syntactic units and parts-of-speech (see Figure 3.4) enable a far faster investigation into a linguistic issue of choice.

Figure 3.4. *Keyword-in-context results for the query ‘ubiquitous’*



Switching back to ‘list’ listing, the next step was introducing wildcard searches (Figure 3.5). The noun ‘student’ was preceded by an asterisk (*) in the query entered into the corpus search bar, causing the corpus to produce a list of the most common collocations with the second unit being the noun ‘student’, ordered by frequency count. As it is evident in Figure 3.6, many of the examples were disposable for the purposes of linguistic analysis or even real-world knowledge, something Aston G. (Online 12) describes as ‘encyclopaedic and oracular use’. A query to circumvent the issue that arose was required.

Figure 3.5. *Searching for collocations with wildcard queries (wildcards represented by an asterisk)*

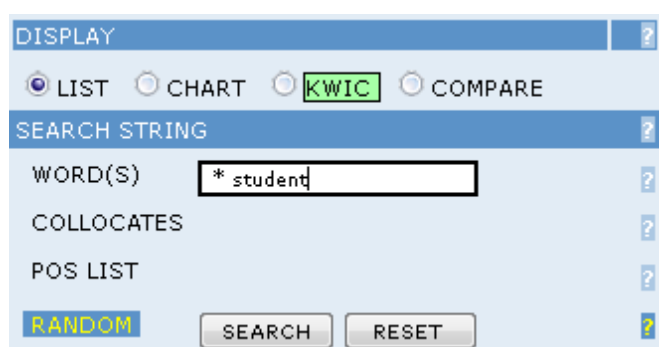


Figure 3.6. *The frequency counts of the query's results*

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

COMPARE ? SIDE BY SIDE

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	THE STUDENT	12714	
2	<input type="checkbox"/>	A STUDENT	10312	
3	<input type="checkbox"/>	OF STUDENT	3448	
4	<input type="checkbox"/>	AND STUDENT	2201	
5	<input type="checkbox"/>	GRADUATE STUDENT	2000	
6	<input type="checkbox"/>	EACH STUDENT	1997	
7	<input type="checkbox"/>	COLLEGE STUDENT	1863	
8	<input type="checkbox"/>	, STUDENT	1735	
9	<input type="checkbox"/>	. STUDENT	1259	

At this stage, the wildcard marked by an asterisk was replaced by the letter J in-between square brackets, denoting the algorithm for an adjective. The students were informed that some of the other popular algorithms they would find themselves continuously entering would be [n] for noun and [v] for verb. Additionally, they were given a hyperlink to Online 5, providing access to the entire CLAWS7 tagset in accordance which the COCA Corpus search function is built, should they attempt to make significantly more complex searches than the ones offered to them.

Figure 3.7. *Searching for collocates with adjectives [j] preceding the noun 'student'*

DISPLAY ?

LIST CHART KWIC COMPARE

SEARCH STRING ?

WORD(S) ?

COLLOCATES ?

POS LIST ?

RANDOM ?

SEARCH RESET

Figure 3.8 shows the effectiveness of a search limited by part-of-speech algorithms, revealing such common adjective + noun collocation with the word 'student' as the second unit as 'medical student', 'fellow student', 'serious student' and 'brilliant student'.

Figure 3.8. *The most frequent adjective + noun collocations featuring the noun 'student'*

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	MEDICAL STUDENT	64	
2	<input type="checkbox"/>	FELLOW STUDENT	39	
3	<input type="checkbox"/>	MATURE STUDENT	35	
4	<input type="checkbox"/>	NEW STUDENT	27	
5	<input type="checkbox"/>	YOUNG STUDENT	22	
6	<input type="checkbox"/>	INDIVIDUAL STUDENT	21	
7	<input type="checkbox"/>	OVERSEAS STUDENT	21	
8	<input type="checkbox"/>	OLD STUDENT	19	
9	<input type="checkbox"/>	PART-TIME STUDENT	14	
10	<input type="checkbox"/>	FIRST-YEAR STUDENT	13	
11	<input type="checkbox"/>	SERIOUS STUDENT	13	
12	<input type="checkbox"/>	BRILLIANT STUDENT	11	

Lemma-oriented searches (Figure 3.9) were a mandatory algorithm to be introduced to the students, as the idea to look for both singular and plural of the same noun is productive for the sake of obtaining as many concordance lines showcasing the desired linguistic feature. The query is placed in square brackets, forming the basic algorithm that surveys all lemmas of one word within the corpus.

Figure 3.9. *Searching for all lemmas of the verb 'be'*

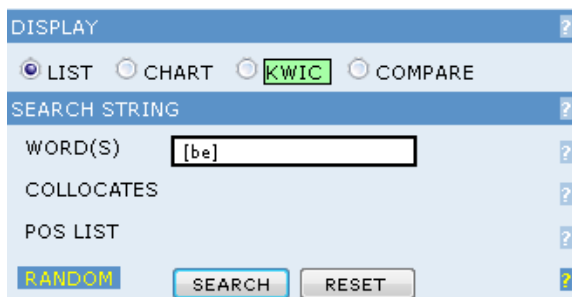


Figure 3.10 depicts the results for the above query, listing all lemmas of the verb 'be' ordered by frequency count.

Figure 3.10. *The most frequent lemmas of the verb 'be' in the corpus*

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	IS	973497	
2	<input type="checkbox"/>	WAS	872618	
3	<input type="checkbox"/>	BE	643901	
4	<input type="checkbox"/>	ARE	459496	
5	<input type="checkbox"/>	'S	342303	
6	<input type="checkbox"/>	BEEN	256779	
7	<input type="checkbox"/>	BEING	82135	
8	<input type="checkbox"/>	'RE	79931	
9	<input type="checkbox"/>	'M	61994	
10	<input type="checkbox"/>	AM	24203	

Figure 3.11. Attempting to search for all synonyms of the word 'brave'

DISPLAY ?

LIST CHART KWIC COMPARE

SEARCH STRING ?

WORD(S) ?

COLLOCATES ?

POS LIST ?

?

The researcher thought it highly topical to show how to locate synonymous units of language using the corpus, as students writing summaries could hesitate to find a close equivalent to a word while attempting to avoid repetition and sometimes plagiarism that is associated with such inappropriate repetition of the language from the original source. The entered word is placed into square brackets not unlike lemma searches; however, it is now preceded by an equality sign within the brackets, which differentiates the synonymy algorithm from the lemma algorithm. The results of the query (Figure 3.11) are depicted in Figure 3.12 below.

Figure 3.12. The most frequent results of the search

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	FACE [S]	32434	
2	<input type="checkbox"/>	BEAR [S]	5537	
3	<input type="checkbox"/>	SUFFER [S]	3374	
4	<input type="checkbox"/>	BRAVE [S]	1704	
5	<input type="checkbox"/>	BOLD [S]	1241	

Figure 3.13. *Revised search to include only*

SEARCH STRING ?

WORD(S) [=brave].[*j] ?

COLLOCATES ?

POS LIST ?

RANDOM ?

SEARCH RESET

With the query ‘brave’, the results were not quite what the user of the corpus may have had in mind, as ‘face’, ‘suffer’ and ‘bear’ are all synonymous to a different kind of the word ‘brave’ intended; namely, the adjective ‘brave’. The students were shown that the corpus does indeed make it feasible to limit the entered queries by parts of speech; however, the algorithm is considerably more complex. Since the search is limited to solely adjectives, the previous search is complemented by the algorithm for (any) adjective, expressed as the letter J in-between square brackets, preceded by an asterisk inside the brackets (following the code with the asterisk seems to produce the same result, but students need not have complicated the corpus engine for themselves any further)

Figure 3.14. *The modified results of adjective synonyms of ‘brave’ only*

	<input type="checkbox"/>	CONTEXT	FREQ	
1	<input type="checkbox"/>	BRAVE [S]	1317	
2	<input type="checkbox"/>	BOLD [S]	1176	
3	<input type="checkbox"/>	HEROIC [S]	578	
4	<input type="checkbox"/>	DARING [S]	452	
5	<input type="checkbox"/>	COURAGEOUS [S]	358	
6	<input type="checkbox"/>	VALIANT [S]	158	
7	<input type="checkbox"/>	FEARLESS [S]	154	
8	<input type="checkbox"/>	PLUCKY [S]	55	
		TOTAL	4248	

The introduction to the COCA Corpus engine then moved on to a more complicated, context-dependent (as opposed to query-dependent) utility of the corpus; specifically, the solution of specific linguistic issues. The students were offered to investigate whether the pronoun ‘neither’ is to be followed by a singular or plural verb. The query was expressed as ‘neither is/are’, and would list frequency counts for either variant (see Figures 3.15-3.16). According to the BNC database, the singular ‘is’ has a greater occurrence compared to the plural ‘are’ within the corpus, as if leading to the appropriate conclusion.

Figure 3.15. Checking the usage of the verb ‘be’ after the pronoun neither

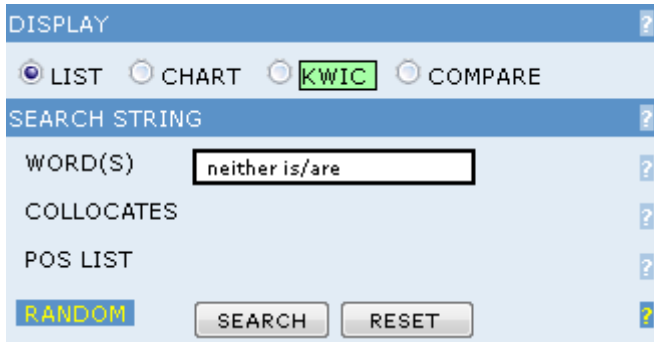


Figure 3.16. The frequency counts of ‘neither is’ and ‘neither are’

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]

	<input type="checkbox"/>	CONTEXT	FREQ
1	<input type="checkbox"/>	NEITHER IS	188
2	<input type="checkbox"/>	NEITHER ARE	70
		TOTAL	258

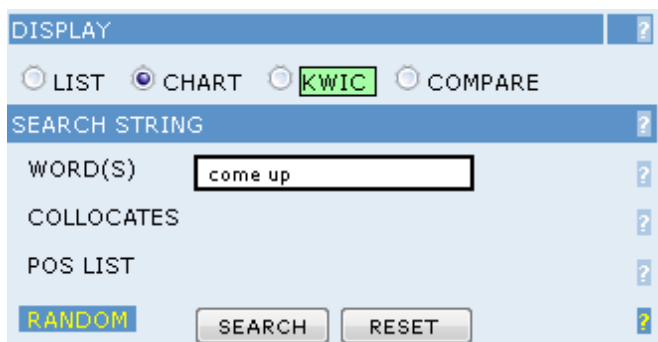
However, as students were taught to mind the context when interpreting corpus data, they were asked to click the results of ‘neither is’, the more popular variant, and see the specifics of that particular phrase’s occurrence. Instead of pointing to any conclusions, it was suggested that there is no guarantee that the usage of ‘neither’ is the same between the offered variants, permitting the students to deduce a conclusion autonomously, and keeping the researcher’s role to that of a mediator.

Figure 3.17. The concordance lines for ‘neither is’

too close to call, which is not good news for Mitt Romney. And **neither is** the latest nationwide poll which shows Rick Santorum with a two-point lead Hate Romney? " And hate is not a motivating season to follow somebody. **Neither is** anger. But passion is. And Santorum shows some passion, espe this case, Anderson, investigating it, they're not releasing the name. **Neither is** the district attorney in this case. So it's hard for us to real, and it is widespread, and it is not going away. But **neither is** a much broader category of industries known to use enslavement in order to mak spit of land that doesn't even reach all the way across. # But **neither is** it the same lake. Brisson's never seen fog on Lake Garner, 's black gums onto the damp earth. I am in no hurry, and **neither is** she. The sprint seems to have cost the dog most of her remaining # " He's not nuts about nuts, " said another. # " **Neither is** my mother, " said a small voice from the back. # All year, a very bad year for housing. Pension contributions aren't taxed; **neither is** employer-provided health insurance, a policy that redistributed \$17 a prerequisite for God to respond positively to our prayers. For that matter, **neither is** being perfect King David was a murderer, adulterer, and liar. ` limited partnerships had 5.6 percent yields, according to the Alerian MLP index. But **neither is** risk-free. Another downturn and collapse in the rental do occur. It doesn't include swearing, though some swearing is slang. **Neither is** it restricted to the spoken language to the extent that it once was. had been exposed to potentially contaminated factor VIII also died of an unrelated illness (**neither is** included in the table). # Table 2. Incubation pe marketing strategy of Focus Features, whose advertising campaign he echoes rather uncritically. **Neither is** the romance between Ennis and Jack a : fact that can be discerned by the methods of the natural sciences alone, but **neither is** it simply a social one discernable without those methods. It i well through responsible management. Its management would not be simple or foolproof, but **neither is** the system that we have in place now. If it itself. " # The American West is no longer what it was, and **neither is** the American Jew. Portland is the most progressive big city in the country a job-crusher. You know, call it what you will. CHRISTIANE-AMANPOU# (Voiceover) And **neither is** Senator Lee. SENATOR-MIKE-LEE-# The shooter i But this software engineer isn't buying it. And back at his table, **neither is** his friend. CUSTOMER-1MALE2-# You can go ahead and do what you wan As you can imagine, neither idea is very comforting to Mr. Heward. JAY-SCHADLER-1-AB# (Voiceover) **Neither is** this. The cops now want him to go

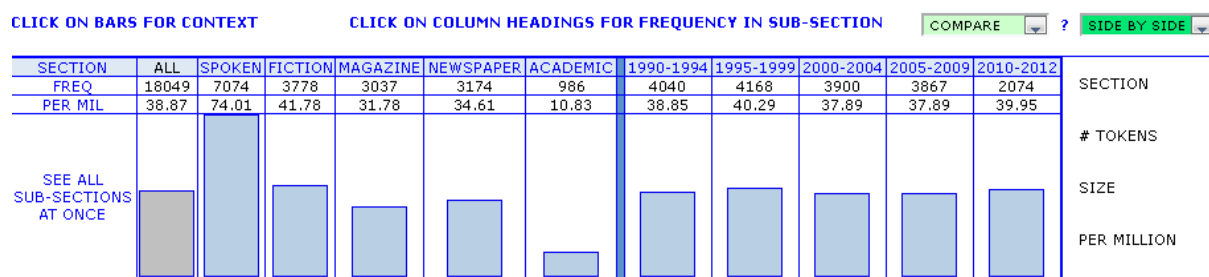
Finally, the students were taught to distinguish between ‘chart’, ‘list’ and ‘KWIC’ functions of the search engine, the so-far unfamiliar ‘chart’ function offering genre differentiation of the entered query, algorithm and/or unit of language. An attempt to enter a phrasal verb (Figure 3.18) was made, with the ‘chart’ listing enables with a tick.

Figure 3.18. Searching for the frequency of the phrasal verb ‘come up’



The outcome left little ambiguity concerning the disposition of the phrasal verb ‘come up’ to occur primarily in spoken ‘texts’ of the BNC, occurring rarely in magazine context and barely appearing at all in the academic genre. Such knowledge empowers students to check their own writing for formality appropriate to the chosen medium of writing, and allowing them to distinguish between contexts in which they would opt for or against the usage of specific units of language.

Figure 3.19. ‘Come up’ frequency in all five genres in the COCA Corpus



The same introduction was presented to all four groups, and the time restrictions were at times in the way of giving the students sufficient wait-time to follow the researcher’s directions and queries entered. As the students were prompted to register on the website of the COCA Corpus in order to avoid being blocked from the corpus for exceeding the maximum number of queries allotted to unregistered usage (around 10-15 at the time of the research), it was rather surprising to hear students’ accounts of being blocked from the corpus in the midst

of the lesson regardless, which seemed highly improbable considering such a limit for registered users is as high as 100 for users who are ‘non-researchers’ (Online 13). As students were expected to be reminded of certain algorithms, search queries and tags, a link to the corpus explanation thereof and the CLAWS7 Tagset was available for them in their e-course.

With no time left at the end of the introduction to the corpus, the students were given their home assignment – to revise their pre-summary or write it anew should they find it necessary (the so-called revised summary) and find 10 corpus examples of the summary/academic writing features outlined in the presentation, the presentation available to them in the e-course as a downloadable .ppt document.

The revised summaries were submitted next week, allowing for a comparison between what students thought of summarising prior to the piloting and the theoretical instruction and what they thought of it afterwards, as reflected in their improved writing. Being the second piece of summary writing produced by the students during the piloting out of three, followed by the final summary of an entirely new text (Online 11), the comparative corpus-based analysis of the student summaries is postponed until the final summaries are received, but so far it is feasible to see how some of the students improved over time. Below is a pre-summary from a student in group 1A, followed by the same student’s revised summary of the same exact text:

Pre-Summary

You, the student on the job hunt, should get on the same wavelength with the company you’re hopeful to be a part of. Try to understand the ins and outs of the company, understand how it works, what are the company’s goals and how they are achieved. Comprehend why would you want to work here and what benefits can the company get from you. Read about the vacancy and apprehend what they will ask from you. Think of examples to illustrate your strong suites. If you have someone who can help you out, try role-playing a mock interview, try to understand from the vacancy advert what questions might they ask.

When the actual interview is coming to a close you’ll have the opportunity to ask questions, actually asking the employer some questions shows that you are truly interested in the position, and you have a chance to better understand the processes going on inside the company.

You have to be prepared to get rejected even when your interview has went smoothly and you and the employer have clicked, a rejection doesn’t denote your ability or inability to do a job,

often candidates don't get picked for trivial reasons. Don't let it get you down and stay perseverant during these tough times.

Revised Summary

The author opens the article by pointing the reader to a web page with tips about CV writing. Then she carries on and starts advising the reader on how to ace an interview. Firstly, she points out that the applicant should do research on the company they are hoping to be hired by. Secondly, it is said that the would-be employee ought to understand what the employer might ask from them by carefully reading the job specification, that would help the applicant comprehend what kind of questions they might get asked at the interview. Further on the author advises to prepare real life examples that would nicely demonstrate that the applicant possesses the required skill set. After that the author suggests trying a made up interview with a friend, this might help in getting prepared for the actual interview. Further on she encourages the prospective candidate to ask questions during the interview or at the very least prepare a list of questions they want to clarify. The author elaborates that this shows the employer that the applicant is serious and is truly interested in the vacancy. The author then concludes that the applicant must be ready to get rejected, she states that often there isn't a fault with the candidate and that they must stay perseverant.

By Sophie Warnes

<http://www.independent.co.uk/student/career-planning/getting-job/scoring-a-job-interview-preparation-8231208.html>

The difference, evidently, is dramatic and it can be hard to believe that the summary was written by the same person. Both summaries were excellent each in their own way, though the former (the pre-summary) was rather informal and had some other features that were not characteristic of or even against the summary writing conventions outlined in this work. The student (and most students in general) paid great attention to the theoretical instruction and has used summary markers in abundance in the revised summary: 'the author opens the article by pointing...', 'it is said', 'the author suggests' and 'the author elaborates' are some of the summary markers utilised in the latter piece of writing. The usage of the pronoun 'you' is entirely eliminated, replaced by impersonal constructions ('it is said') and

correct usage of pronouns, though not imperfect, for example the tenses in sentences such as the following do not correlate: ‘...that would help **the applicant** comprehend what kind of questions **they** might get asked at the interview’. The plural ‘applicants’ would have been preferable, yet the progress and the appropriateness to the genre of summary writing is transparent, at least in the writing of students from the stronger groups (1A and 1B). The tendency to reference the material quotes at least outside the text is also a typical feature of the better-written revised summaries at this stage.

In contrast, the theoretical instruction alone was insufficient to change the writing level of the poorer-performing groups, such as 1C, considerably. Below are the pre-summary and the revised summary belonging to a student in group 1C:

Pre-Summary

To find a good job, first of all, you have to know some information about the company, in which person wants to work. It is very important to know the job or person specification, because it will help you to answer some questions during the interview. In your CV you have to write some good points of you, for example: communication and leadership skills, examples of scenarios where you have proved that you have a certain skill or ability. It will be good if you mention these things again. Do a mock interview with someone else, give them potential questions, it will show how confident you are. In the end you will have a chance to ask questions. Use it, because it shows that you are interested in the job. Try to start with simple questions, and explain why you asking about them. Sometimes it happens that person does not get a job. But you do not have to despair about it, because it can be many reasons why you do not get a job. It is not necessarily that you are not good enough. You have to try to get a job several times.

Revised Summary

To find a good job for yourself is not easy. Person has to know a few points that will help to get a job. For instance, man has to know all information about company: what the company does, how it achieves that, who their competitors are. Person has to ask a question about whether he likes the company, whether he wants to work there. Next point is to answer some questions, which include the knowledge about specification of the job and about the skills of new employee. You have to bring a list of examples with skills from work or from university. Next point is to write a good CV. It should include some skills, such as communication skills; leadership skills; initiative; problem-solving skills; organizational skills. You have to show

not only good sides, but also bad sides. You have to do a mock interview with someone else, because it will prepare you for real interview. In the end of your interview you can ask some questions, it will show that you are interested in future work. Ask only easy questions and the questions that you are interested in. Do not be afraid if you do not get a job, you must try more times.

The student's summary has clearly been revised, whole parts of it re-written seemingly entirely. Some changes caused by the theoretical instruction are observable very transparently; for instance, there is evident effort to move on from one idea to another, though this is not achieved very effectively – by means of repetition of 'next point'. When the pre-summary immediately launched into the listing of the points covered in the source text, the revised version has a semblance of introduction: 'To find a good job for yourself is not easy'. There are some attempts to leave a few impersonal constructions, such as 'Person has to know a few points' or 'Person has to ask a question', yet not very consistently as there are still numerous instances of the usage of the second person pronoun 'you' ('You have to do a mock interview with someone else') and even the imperative ('Do not be afraid if you do not get a job'). There is still a significant amount of work to do, as complete improvement cannot happen overnight, but a certain level of improvement is observable nevertheless.

Besides re-writing their pre-summaries, students were expected to bring along their home assignment of 10 corpora examples of summary writing or academic writing to class, although the submission thereof had to wait until over a week later, allowing students to correct their individual examples after watching what sort of queries other students have prepared. The success of this activity varied depending on the group, with the better groups having highly imaginative and often technically impressive examples, while the weaker groups often failed to comprehend the task.

Some of the progress of the weaker groups was hindered by the students' choice to find such examples that corresponded with grammatical features they did not understand or just happened to misinterpret, as seen below:

That clauses

*Or even the second or the third, but **that** doesn't mean a lack of abilities or skills.*

Query – that [cs]*

Attribute adjectives

*It used to be old fashioned, **the kind** that can be seen written by professors or in written journal articles, but now it has developed due to a democracy and internationalization.*

*Query - * kind*

The student that produced these queries had the right idea concerning the assignment – the corpus is to be searched for examples of specific features of summary/academic writing; all that remained to do was think of appropriate queries. For example, the first of the examples quoted is not a that-clause, not to mention that the query cited actually does not bring the outcome the student copied, instead listing instances where ‘that’ is followed by subordinating conjunctions, which differs from the sentence which was most probably generated through other means. In the second example taken from the same student, even though the word ‘kind’ does happen to be an adjective in most contexts, it is a noun in the example and the student could have forgotten to read the sentence to check if this is an appropriate example to feature.

In many instances in groups 1C and 1D/FIN, students showed their failure to follow the task given, conducting corpus searches of actual names of features, as opposed to thinking of the possible algorithm that would cause the desired feature to appear in the concordance lines. Admittedly, the assignment was disproportionately difficult for said groups, as it required imagination and what was called explicit grammatical knowledge in the theoretical part of the present research. Among instances of such experiments are the following:

Paraphrasing – In general, if you can't restate someone else's ideas in your own words, it is likely that you do not yet understand them sufficiently to be able to make appropriate use of them. The writer of this extract may have felt that she or he was paraphrasing rather than plagiarising, and might point to the reference made in the extract to Leech and Short as evidence that she or he was not being dishonest. Nevertheless, the rewritten version conforms so closely to the original, with just isolated words changed, that it can not reasonably be called a paraphrase.

** clause - Make and have are at the top of the binding hierarchy and take the " highly non-verbal " bare infinitive; order is lower on the scale and takes the to infinitive; next comes insist, which is followed by a subordinate clause containing a finite subjunctive form, and then think, construed with a subordinate clause in the indicative.*

In these cases, the student simply used the corpus to find the words ‘paraphrasing’ and ‘subordinate clause’ within the corpus. While this has very minimal utility in getting acquainted with how searching for a query operates, it achieves nothing in terms of teaching

the students how to summarise, and it is the combination of the two that were the aim of the corpus-based activities of the present research.

Regarding failure to interpret the theory, there was also a difficulty in differentiating the participle from the verb in the past tense for the weaker groups, such as:

*that he dreamed of having a Madras curry. Two weeks ago Roy and Fiona **went** for a curry. They ate early, before the restaurant got too full*

Meanwhile, some of the students' peers from the same weaker groups and the vast majority of the tougher groups like 1A and 1B showed a surprising level of autonomy and volition so as to skim through the links offered to them that would let them add to the verbal instruction with additional information of their own. There were many solid, acceptable examples such as:

Passive voice

[is made]

*It seems to be clear that not enough use **is made** of solicitors for the solution of legal problems.*

Subordinate clause

[having typed]

*Mrs Padmore, **having typed the interpretation of the remark**, has gone back and altered its semantic status to fit the interpretation.*

The students effectively listed features not only of academic writing, but also the specifics of summarising as such:

Query: *author [v]*

Result: *Summary markers*

Example: *been kept alive.and well in influential papers such as Die Zeit and elsewhere.*

*The **author recalls** a visit to the German Bundestag where the British were roundly berated for being*

A high number of the students in groups 1A and 1B were able to successfully pinpoint the corpora instances of the participle, in contrast with the less successful students in the lower groups (albeit with the word 'query' misspelt):

Participles

To see God's love expressed to us in the prayers and giving of people far away is a great encouragement and blessing to us

Query: *love *ed*

A wide variety of features were covered, even ones seemingly difficult to pinpoint:

Precision

It's now known that more than 100 people are HIV positive following treatment here.

Query: 100 people

Some of the students opted to organise their examples in table, making their submission very cleanly organised and readable, though in the weaker groups this organisational effort had the misfortune to not correspond with an equal level of accuracy with the designed task. Nevertheless, they found ways how to deal with the task in ways they found fitting:

fear of [vvg]	Fear of making mistakes.	Living magically: a new vision of reality. Edwards, Gill. London: Judy Piatkus (Pubs) Ltd, 1991, pp. 45-160. 2798 s-units.
suggests [vvg] / 55 results	The manual suggests using Card No. 1 to start with.	Machine Knitting Monthly. Maidenhead: Machine Knitting Monthly Ltd, 1992, pp. ?? 1485 s-units.

Among the examples collected in this assignment, there were ones highlighting great creativity, imagination and complex technicality some of the gifted students from groups 1A and 1B boast. These examples make it evident that students approached the assignment creatively and analytically, composing complex strings of algorithms to result in outcomes appropriate to the task:

Of+ing-clauses

*The normal **method of securing** is to drive a long nail through the end of the sleeper at an angle into the one beneath.*

*Query: “[*nnl*] of [v?g*] is”*

Sequences of prepositional phrases

*If he says the second it means that there is a better chance **of winning** 1000 **by drawing** a red marble from the urn.*

Query: “[i] [v?g*] * [i*] [v?g*]”*

Here is another creative take on the appointed task, from a student who took it upon himself to render sentences found in the academic genre section of the corpus:

'When Wittgenstein says that the linguistic expression is a substitute for the natural expression [...]'

Query: [np] [[=claim]].[v*] (Here and further search results are restricted to 'academic' section)*

Feature: summary markers, the reporting manner.

During the lesson, the instructor carried out another presentation (see Appendix 7), praising students for their pre-summaries (this preceded the checking of the revised summaries) so that they remain motivated and then listed some of the common errors (concerning lack of suitability to fit the genre of summary writing, as opposed to unrelated, though undoubtedly important nevertheless, grammatical rules and norms of syntax) that were recurring in the pre-summaries checked. The instructor also used the corpus engine once again to highlight some of the possible ways how the COCA corpus could be helpful in assisting during the writing process, complementing the previous instruction on corpora. Spellchecking using the corpus and investigating the appropriateness of the linguistic means chosen to the necessary register were some of the functions emphasised. The researcher did stress that the corpus is only one of the ways how students could assist their writing, dictionaries being another option that is sometimes more efficient.

During the second lesson, much of the time following the presentation was invested into having students co-operate in groups, in which they shared some of the corpora examples of summary writing and academic writing among each other with the final aim to present a couple of the ones they agreed to be the best before the class.

The examples differed greatly depending on the group. Some of the examples highlighted above as highly impressive technically were once again exposed in the presentations of group 1A and 1B; for instance:

Attributive adjectives

MICHAEL JACKSON is' killing himself' by hiding away and refusing to face **the real world**.

Query: "the [j*] [nn*]"

Meanwhile, some of the examples the weaker groups presented were more straightforward, simplistic or sometimes even unsatisfactory for the task given:

QUERY: WRITING [V*]

Source - Curricula for diversity in education. Swann, W; Booth, T; Masterton, M; Potts, P. London: Routledge & Kegan Paul plc, 1992, pp. 9-113. 2297 s-units.

Example - There are so many writers who say that writing is a process of discovery, or investigation, rather than a record of it

The presentations generally confirm what the written submissions showcase and show interesting patterns of student co-operation – some groups clearly focused on one student’s impressive work, while others simply had each person choose one effective example from their submission. Both were undoubtedly productive with their choices.

With whatever little time left after the presentation, since much time was spent with miscellaneous infrastructure issues such as how to get all the students grouped together, getting them to finish in spite of very high dedication to delivering a presentation ‘more perfect’ than expected and getting them to stand up and speak before their peers, the students were given their last task for the second lesson, which was to use the corpus to correct some of the errors committed during the pre-summary writing process (see Appendix 7).

The amount of time left for this assignment differed depending on the group, as different groups took different amounts of time and effort of coercion on researcher’s behalf to present in a relatively timely manner, with group 1B having no time left for the assignment and group 1C having too little time to plunge them into an activity that would be rather complex for their level, even though some of the errors were too simple for the stronger groups to correct without any help of corpus assistance whatsoever.

The students put some effort into correcting the errors, often correctly but sometimes failing to locate the problem:

*The article „Scoring a job: Interview preparation” **advices** young students on how to get their first job.*

*Query: the [n] advi*es*

Conclusion: advices is the incorrect form of the noun advice , the verb advise if necessary.

*...they could have simply **cut down** applicants to the ones who live closer because of their multiplicity.*

Query: display (chart)

Conclusion: Cut down is not used in academic context, instead the word “decrease” could be used.

*If the applicant **won’t get** the job, author recommends staying calm and trying to send his CV to other companies and eventually the job will be his.*

Query: will not [get]

Conclusion: Contracted forms cannot be used in academic writing, instead will not should be used.

The first example above shows a correct exploitation of the corpus to prove the difference between the verb 'advise' and the noun 'advice'. However, in the second quoted error correction, the student emphasised the inappropriateness of the phrasal verb to academic context instead of investigating the reality of the collocation given. The same can be seen in the third error correction attempt, where attention is paid to the contracted form instead of the wrong tense. Another student in the same group pinpointed the issues more easily with these sentences, albeit seemingly without any usage of corpora, as no queries were cited for these particular errors:

*...they could have simply **cut down** applicants to the ones who live closer because of their multiplicity.*

The phrase could be left if the sentence is improved with the noun 'number', i.e. 'simply cut down the number of applicants'.

The weaker group 1D/FIN, on the other hand, struggled much more, often making more errors than the original sentence:

*The article „Scoring a job: Interview preparation” **advices** young students on how to get their first job.*

*Query: advice**

Conclusion: word advices could be found only 6 times, word advice more than 10000 times

Correct: advice

Not all of them found the task too difficult however, though additional time could have definitely enabled the students to focus on working with the corpus more thoroughly, perhaps asking the instructor for advice. Nevertheless, even this group had some good error corrections, as much as the time permitted:

*The article „Scoring a job: Interview preparation” **advices** young students on how to get their first job. Query: advice.[v*]; Conclusion: advice is not used as a verb. Correct: The article ... gives advice*

After the class, the students found the second text they would be summarising, which was their final summary. The pre-summary, the revised summary and the final summary were all converted separately in .txt documents and distributed into folders according to the summary typing and student group, so that a corpus analysis of the information therein would be feasible to conduct.

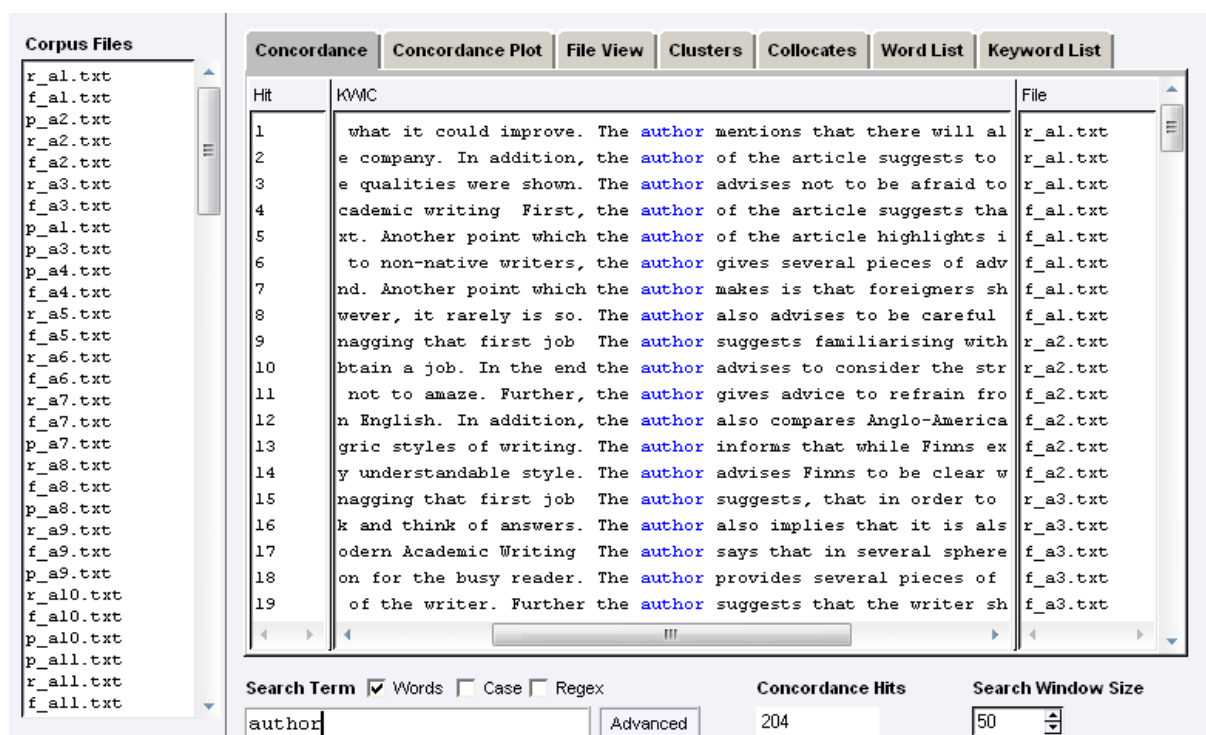
Table 3.9 below reflects the manner in which the texts within the corpus were arranged. The folder with the files was divided into categories of ‘tagged and ‘untagged, the former coded using the CLAWS7 tagset and the latter preserved in their original shape without suffering any change. The AntConc Concordancer tool (Online 6) was downloaded online and utilised to produce the learner corpus part of the research.

Table 3.9. *Location of texts in the student corpus*

		Tagged	Untagged
Summary	Group	Location Within the Corpus	
Pre-Summary	1A	p_a1_t.txt	p_a1.txt
	1B	p_b1_t.txt	p_b1.txt
	1C	p_c1_t.txt	p_c1.txt
	1D	p_d1_t.txt	p_d1.txt
	1FIN	p_f1_t.txt	p_f1.txt
Revised Summary	1A	r_a1_t.txt	r_a1.txt
	1B	r_b1_t.txt	r_b1.txt
	1C	r_c1_t.txt	r_c1.txt
	1D	r_d1_t.txt	r_d1.txt
	1FIN	r_f1_t.txt	r_f1.txt
Final Summary	1A	f_a1_t.txt	f_a1.txt
	1B	f_b1_t.txt	f_b1.txt
	1C	f_c1_t.txt	f_c1.txt
	1D	f_d1_t.txt	f_d1.txt
	1FIN	f_f1_t.txt	f_f1.txt

The files from each respective folder were uploaded to the AntConc program in order to locate and calculate the frequency of features related to summary writing and academic writing that were possible to measure. The simplest one by far was the word ‘author’, which was sufficient to produce concordance lines with nothing except summary markers (see Figure 3.20). The rest were the product of an analysis of well over a hundred of tags available in the CLAWS7 Tagset that were deemed suitable for determining whether students’ summaries matched the required standards or not.

Figure 3.20. Concordance line of 'author' in AntConc, untagged



The following tags were selected from the tagset (Online 5):

Precision

Lack of Precision

DA1 singular after-determiner (e.g. little, much) **(lack of) precision**

DA2 plural after-determiner (e.g. few, several, many) **(lack of) precision**

DD determiner (capable of pronominal function) (e.g. any, some) **(lack of) precision**

Presence of Precision

MC (MC1) cardinal number, neutral or singular (two, three..) **precision**

NNT (NNT1/NNT2) temporal noun, singular or plural (e.g. day, week, year) **precision**

NP (NP1/NP2) proper noun, any number (e.g. IBM, Andes) **precision**

Impersonality

Presence of impersonality

PPH1 3rd person sing. neuter personal pronoun (it) **impersonality**

PN1 indefinite pronoun, singular (e.g. anyone, everything, nobody, one) **impersonality**

Lack of impersonality

APPGE possessive pronoun, pre-nominal (e.g. my, your, our)

PPX1/2 singular or plural reflexive personal pronoun (e.g. yourself, itself)

Impersonality

PPY 2nd person personal pronoun (you)

Formality

XX not, n't **formal register (not vs nt)**

Finally, the corpus was sought for the most common discourse markers used by the students, as a result of which the following list was compiled:

Coherence

However

Nevertheless

Namely

In addition

Also

Furthermore

Additionally

Moreover

Lastly

Finally

Hence

Therefore

Thus

Precision Lack of precision

Table 3.10. *singular after-determiner in the student corpus*

DA1	singular after-determiner (e.g. little, much)				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	2	0	1	2	5
Revised	1	1	1	3	6
Final	1	0	1	0	2

Table 3.11. *plural after-determiner in the student corpus*

DA2	plural after-determiner (e.g. few, several, many)				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	5	3	5	9	22
Revised	5	7	7	12	31
Final	7	3	4	9	23

Table 3.12. *determiner (capable of pronominal function) in the student corpus*

DD	determiner (capable of pronominal function)				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	2	2	5	1	10
Revised	0	2	3	1	6
Final	2	0	1	3	6

The tags above represent a lack of precision, an attribute that was expected to decline following the theoretical instruction and introduction of corpus-based activities. The chosen tags were chosen in opposition to more precise units of language: for example, ‘little’, ‘much’ and ‘several’ over exact numbers, the latter contributing to precision to a greater degree. Two aspects of change in this direction are visible from Tables 3.10-3.12. First of all, the expected tendency did not occur, with the numbers increasing as much as they decline, often with an inclination towards the former. On the other hand, the expected tendency is indeed observable with determiners capable of pronominal function such as ‘any’ or ‘some’, as there seems to have occurred a conscious decline of such words from the pre-summary to the final summary. No change is observed between the revised summary and the final summary; however, it should be remembered that the two texts were different and the content of the texts cannot have escaped affecting the features found in students’ summaries.

Presence of precision

Table 3.13. *cardinal number, neutral or singular in the student corpus*

MC (MC1)	cardinal number, neutral or singular				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	11	6	2	2	21
Revised	10	3	5	10	28
Final	18	10	12	16	56

Table 3.14. *temporal noun, singular or plural in the student corpus*

NNT (NNT1/NNT2)	temporal noun, singular or plural				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	10	2	5	4	21
Revised	4	3	5	6	18
Final	11	4	10	9	34

Table 3.15. *proper noun, any number in the student corpus*

NP (NP1/NP2)	proper noun, any number				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	7	3	6	16	32
Revised	21	8	6	31	66
Final	47	14	13	46	120

Tables 3.13-3.15, on the other hand, refer to the tags associated with the presence of precision as a positive quality of academic writing. Though more than three tags above were found as indicating precision, only these ones produced any results upon the respective researches, allowing the information to be interpreted. The expected outcome was for the presence of said features to increase when progressing from one summary to a later one. In general, the tendency proved true in all instances but one (usage of temporal nouns, either singular or plural, between the pre-summary and the revised summary), which is expressed in the most pronounced manner in the instance between the usage of cardinal numbers between the revised summary and the final summary, and later in the case of proper noun usage between any two summaries, as the usage doubles with each new stage.

Impersonality Presence of impersonality

Table 3.16. *3rd person singular neuter personal pronoun in the student corpus*

PPH1	3rd person singular neuter personal pronoun				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	78	33	24	47	182
Revised	68	28	35	45	176
Final	57	15	41	45	158

Table 3.17. *indefinite pronoun, singular in the student corpus*

PN1	indefinite pronoun, singular				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	33	17	21	21	92
Revised	25	11	20	29	85
Final	8	5	6	9	28

A similar approach was taken for impersonal tags. Usage of third person and indefinite pronouns, as opposed to first and second person ones, is associated with a better command over the theoretical material provided. This was perhaps the least successful step of the corpus

analysis, as the results were opposite from the expectations, as seen in Tables 3.16-3.17. The lack of indefinite pronouns in final summaries is especially noticeable.

Two conclusions can be drawn from this failure to have the corpus tools provide insight into the features indicating impersonality – first of all, not every tag is capable of expressing the state of things as they are in terms of compatibility with the summary and academic writing featured mentioned in this research paper, and second, the data points to the fact that students focus on some features thereof at the expense of others, as it is highly unlikely that all features would be proportionately represented in the students’ summaries.

Lack of impersonality

Table 3.18. *possessive pronoun, pre-nominal in the student corpus*

APPGE	possessive pronoun, pre-nominal				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	42	17	36	30	125
Revised	10	14	8	21	53
Final	14	12	34	34	94

Table 3.19. *singular or plural reflexive personal pronoun in the student corpus*

PPX1/2	singular or plural reflexive personal pronoun				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	11	1	4	8	24
Revised	3	1	3	8	15
Final	2	1	0	3	6

Table 3.20. *2nd person personal pronoun in the student corpus*

PPY	2nd person personal pronoun				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	121	37	63	116	337
Revised	1	0	9	52	62
Final	1	0	2	9	12

The tags in Tables 3.18-3.20 were intended for the opposite – a lack of impersonality, an attribute which was expected to decline. The intended outcome matched the expectations perfectly, as there was a significant drop in the features denoting a lack of impersonality, therefore proving that the students’ summaries had grown more impersonal. Table 3.18, depicting the occurrence of possessive pronouns, slightly contradicts the main conclusion showing an increase in possessive pronoun occurrences from the revised summary to the final summary; however, it should be noted that the system include rather impersonal pronouns such as ‘their’ in the mix, meaning this is the least accurate of the tags chosen.

On the other hand, second person personal pronouns are connected to the category of impersonality in the most direct manner and Table 3.20 exposes an enormous drop from 337 to 62 second person pronouns between the pre-summary and the revised summary, culminating in the slump observable in the final summary, at mere 12 occurrences of second person pronouns. Similar results can be witnessed in Table 3.19, albeit with less data to interpret.

Formality

Table 3.21. *Full negative forms in the student corpus*

XX - not	not, n't				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	34	25	18	33	110
Revised	31	17	21	31	100
Final	42	12	32	27	113

Table 3.22. *Contracted negative forms in the student corpus*

XX -nt	not, n't				
Summary	1A	1B	1C	1D/FIN	Total
Pre-	15	0	10	8	33
Revised	0	1	2	3	6
Final	0	0	0	0	0

The category of formality was represented rather poorly among the tags found in the CLAWS7 tagset, the tag [XX] being arguably the only viable tag for the stated purposes. The use of full negatives such as ‘do not’ and ‘cannot’ are expected to reflect an awareness of formality, while the contracted forms such as ‘can’t’ and ‘don’t’ are expected to represent the opposite.

It is evident from Table 3.21 that no significant change in the frequency of full negative forms had occurred between the subsequent summaries, to say nothing of the opposite trend seen with group 1B, where the number of such linguistic units halved between the pre-summary and the final summary. On the other hand, the contracted forms disappeared entirely following the theoretical instruction in the stronger groups and the weaker groups followed after the feedback on their pre-summaries, taking slightly longer to correct their errors (some other possible variables would include students being absent during a certain part of the piloting and contributing work while oblivious of the knowledge and experience shared by their peers).

Coherence Usage of Discourse Markers

Table 3.23. *Discourse Markers in the student corpus*

Pre-Summary					
Discourse Marker	1A	1B	1C	1D/FIN	Total
However	2	1	0	3	6
Nevertheless	0	0	0	0	0
Namely	1	0	0	0	1
In addition	1	1	0	0	2
Also	20	10	7	13	50
Furthermore	2	0	0	1	3
Additionally	1	0	0	0	1
Moreover	0	1	0	1	2
Lastly	1	0	0	0	1
Finally	2	2	1	1	6
Hence	0	0	0	0	0
Therefore	2	2	0	1	5
Thus	1	0	1	0	2
Total	33	17	9	20	79
STUDENTS	14	8	9	14	45
Markers per 1 student	2.36	2.13	1	1.43	1.76

Table 3.24. *Discourse Markers in the student corpus*

Revised Summary					
Discourse Marker	1A	1B	1C	1D/FIN	Total
However	5	2	0	4	11
Nevertheless	0	0	0	1	1
Namely	0	0	0	0	0
In addition	2	1	0	0	3
Also	17	8	8	12	45
Furthermore	5	1	0	1	7
Additionally	2	0	0	0	2
Moreover	1	1	0	1	3
Lastly	2	0	0	0	2
Finally	4	4	1	0	9
Hence	1	0	0	0	1
Therefore	1	3	0	2	6
Thus	2	1	2	0	5
Total	42	21	11	21	95
STUDENTS	15	8	9	14	46
Markers per 1 student	2.8	2.63	1.22	1.5	2.07

Table 3.25. Discourse Markers in the student corpus

Final Summary					
Discourse Marker	1A	1B	1C	1D/FIN	Total
However	7	1	1	3	12
Nevertheless	0	0	0	0	0
Namely	1	0	0	0	1
In addition	2	2	0	0	4
Also	14	3	9	12	38
Furthermore	5	2	1	1	9
Additionally	0	0	0	1	1
Moreover	2	3	0	0	5
Lastly	4	0	1	0	5
Finally	4	0	1	2	7
Hence	0	0	0	0	0
Therefore	6	0	0	3	9
Thus	2	0	1	2	5
Total	47	11	14	24	96
STUDENTS	16	8	9	15	48
Markers per 1 student	2.94	1.38	1.56	1.6	2

The most time-consuming process of the corpus analysis was pinpointing the most widespread discourse markers used by the students and then locating their frequency at all stages of the summary writing process – the pre-summary, the revised summary and the final summary.

Tables 3.23-3.25 reveal the general sparsity of discourse markers occurring in student summaries. ‘Also’ is by far the most widespread discourse marker that can be found in summaries, as one would expect; however, the rest of the specifics vary significantly according to group. The outcome of the corpus analysis proves that groups 1A and 1B generally use discourse markers nearly twice as often as their peers in groups 1C and 1D/FIN, yet group 1B shows anomalous behaviour yet again by suffering a decrease from 2.63 discourse markers per one’s student summary at the revised summary stage to 1.38 in their final summaries. The best explanation for this exception to an expected trend is once again the aforementioned impossibility to focus on all features at once, as there is a necessity to focus on something and the variety of features of summarising and academic writing could overwhelm students, pushing them to use a limited number of ones they are already familiar with or were ready to memorise more easily than others.

Summary Markers

Table 3.26. *Summary markers in the student corpus*

Summary	Author				Total
	1A	1B	1C	1D/FIN	
Pre-	1	0	4	10	15
per student	0.07	0	0.44	0.71	0.33
Revised	33	9	6	26	74
per student	2.2	1.1	0.67	1.86	1.61
Final	41	14	20	40	115
per student	2.56	1.75	2.22	2.67	2.40

The final Table 3.26 displays the occurrence of summary markers featuring the word ‘author’, such as ‘author suggests’, ‘author claims’, ‘according to the author’ and so forth. This was one of the most explicit parts of the corpus research, depicting the extent to which the occurrence of summary markers skyrocketed following the theoretical instruction and once again after the presentation with feedback on student errors. From 15 occurrences of ‘author’ in the pre-summary, the frequency of summary marker usage increased to nearly five times as much – 74 in revised summaries (or from 0.33 occurrences of ‘author’ per student’s summary to 1.61). Group 1B once again showed anomalous behaviour by not meeting the expectations and using summary markers more sparsely than their peers in other groups.

3.5. Post-Questionnaire

After the piloting, the students received copies of the Post-Questionnaire (Appendix 5) whose aim was to receive feedback on the piloting process and obtain insight about the possible improvements in further prospect implementations of corpora in the field of academic writing and other areas of utility to English philology students.

Table 3.27. *What auxiliary tools did you use during the writing process for the two piloting lessons involving corpus-integrated methods in summary writing?*

Group	Only Corpus	Only Dictionary	Multiple / Depends
1A	4	2	4
%	40	20	40
1B	1	0	5
%	16.7	0	83.3
1C	1	1	14
%	6.3	6.3	87.5
1D/FIN	7	2	4
%	53.8	15.4	30.8
Total	13	5	27
%	28.9	11.1	60

The first question addressed to the students at the Post-Questionnaire stage dealt with the facilities available to the students. The goal was to learn whether students preferred to use corpora over dictionaries during the corpus-integrated activities, or whether they used multiple facilities depending on the problem at hand.

The vast majority of students – 60% - acknowledged that they had used multiple facilities, 87.5% of multi-facility users in group 1C. Among students who used corpora solely, the biggest corpus enthusiasts were in group 1A (40% reported having used only corpora) and group 1D/FIN (53.8%). In total, 28.9% students in all groups stated that they limited themselves to using corpora only in the corpus-based activities, while 11.1% preferred to use dictionaries.

It is evident that a large number of students have the capacity to distinguish between the facilities available to them and choose the one that is the most appropriate and efficient for a particular task that requires solving.

Table 3.28. *What do you believe would improve the integration of corpus materials in the summary teaching lessons?*

Group	More time	Simpler interface	Detailed instruction
1A	3	6	5
%	30	60	50
1B	1	3	5
%	16.7	50	83.3
1C	4	5	10
%	25	31.3	62.5
1D/FIN	4	6	7
%	30.8	46.2	53.8
Total	12	20	27
%	26.7	44.4	60

Following this inquiry, the students were asked to reflect upon the lessons that had just passed and offer their opinion concerning the possibility of improving the corpus integration in the teaching of summary writing. 26.7% of the respondents admitted the lack of time (which was significantly more explicit for the researcher), while 44.4% found the corpus interface to be too complicated to be understood and used efficiently. Finally, 60% of the students in all groups, and all but one student in group 1B, would enjoy learning about corpora in a more detailed manner, which the current research did not permit to happen due to time restraints.

Table 3.29. *Do you expect yourself to continue using the corpus?*

Group	Yes	No
1A	8	2
%	80	20
1B	5	1
%	83.3	16.7
1C	12	4
%	75	25
1D/FIN	7	6
%	53.8	46.2
Total	32	13
%	71.1	28.9

The students were then asked if they foresaw themselves continuing to use corpora in the nearest future, with a total of 71.1% students responding to this question affirmatively and 28.9% - negatively. The level of enthusiasm for corpora was lower in the weaker groups, which also found the assignments hard to comprehend, as the previous pages show, and the corpus interface too complex to master fully, for example, 6 students out of 13 stated that they would not use corpora, which is nearly half of the respondents at 46.2%. In contrast, only one student responded negatively in group 1B; however, the small size of the respondents answering the Post-Questionnaire (nearly a half of the Pre-Questionnaire's polled group) casts doubt on the value of this piece of information.

The results of this particular question are highly important, as they indicate a change from being unaware about corpora to showing volition to consult them whenever they may be of use.

Table 3.30. *Which fields would you say the corpus is most helpful in?*

Group	Writing	Ling. issues	Research	Translation	Other
1A	3	3	8	1	2
%	30	30	80	10	20
1B	6	2	2	0	0
%	100	33.3	33.3	0	0
1C	7	9	5	3	1
%	43.8	56.3	31.3	18.8	6.3
1D/FIN	9	9	3	2	0
%	69.2	69.2	23.1	15.4	0
Total	25	23	18	6	3
%	55.6	51.1	40	13.3	6.7

Table 3.30 pinpoints the students' opinion concerning the fields they could apply corpora to the greatest effect. Assisting writing and solving linguistic issues were in the lead with 55.6% and 51.1% respectively, followed closely by corpus-based research (40%). Only 13.3% of the respondents found corpora useful in translation. 6.7% of the respondents offered other answers, including one who acknowledged not knowing how to best apply corpora. The largest discrepancy between the groups was group 1A's agreement on corpus's suitability for research purposes. It is not entirely unlikely that some of the group's student will be conducting their own corpus-based research somewhere in the nearest future as of the writing of this research paper. Group 1B, on the contrary, favoured the usage of corpora in assisting writing (100% answered this way), while group 1D/FIN could not decide between writing and the solution of linguistic issues.

Finally, the students were left with an open-ended questions concerning their comments. As it was observed during the pre-questionnaire and the initial contact, the weaker groups showed the most initiative and left the most feedback. Group 1B provided no comments whatsoever, whereas there was only one comment in group 1A which read: "I really would like to see a more user-friendly interface".

Group 1D/FIN left two comments critical of the corpus they were shown, and statistics above do point to their dissatisfaction with it relative to other groups:

"I believe that dictionaries are more useful than corpus."

"More detailed instruction on how to use Corpus would help the students, because this system is very hard to understand."

Group 1C was responsible for the largest amount of feedback, both critical and positive:

"I don't have any comments!"

“These lessons were very informative and changed our view on the writing process.”

“It is quite complicate to use the corpus. There are more comfortable alternatives.”

“Its hard to understand corpus if haven’t experience with that.”

“Thank you for your lectures and for showing us the corpus.”

“I think there could be more explanations what is and what for is corpus seemed for.”

“I don’t understand how to use corpus.”

“Thank you for showing corpus! Really useful thing! ☺”

With students either grateful for the introduction of corpora or seeking a more detailed approach, it is clear that there may be high potential for further implementation of corpus-based activities.

To conclude the current empirical chapter, there is still sufficient space for improvements in the area of corpus application to academic writing teaching and summary writing, namely with additional allocation of time, possibly distributed evenly during the lessons of the study year, longer, more detailed instructions and more precise corpus analysis tools with convenient accessibility. The present case study managed to show possible ways to teach and use the corpus in the solution of specific tasks, namely that of writing a summary and solving specific linguistic issues. It can be estimated that many of the research subjects will continue to use corpora and improve their command thereof, discovering new methods of their application besides ones discussed in the present research.

Conclusions

The goal of the present research paper was to create a set of corpus-based activities for the purpose of developing tertiary level students' academic writing skills, namely summary writing skills, and investigate the effectiveness of these activities by carrying out a pilot study where the developed materials are put into practice.

Having completed the piloting and the analysis of the relevant texts and information, the research paper can safely claim the accomplishment of the practical aims it had posed before itself. The intended corpus-based materials meant for teaching summary writing were integrated into the curriculum not without problems, but achieved visible results and caused changes, including quantitatively measurable ones, in students' capacity to produce quality texts in the genre of academic summary writing as well as their ability and volition to use corpora in different areas of activity, from research and solution of specific linguistic issues to aiding creative and academic writing and translation.

In certain ways, the outcome of the research is a triumph for the application of corpora materials to teaching tertiary level students. The piloting process had succeeded in turning what had previously been an unfamiliar obscurity to the students to a desirable tool that is helpful in a number of contexts and preferable to existing facilities commonly consulted by them prior to the current research.

Besides bringing tertiary level students closer to corpus activities and promoting corpus usage, the current research has managed to facilitate the teaching of summary writing by its selected means of corpus-based materials and raise students' awareness of miscellaneous linguistic issues by empowering them to research them with a number of tools and facilities available, thus making their linguistic knowledge more explicit than before.

Numerous implementations and developments have the potential of making further research in the area feasible. First of all, any progress in freely accessible concordancers, including less restrained and more flexible search, would benefit corpus research of learner corpora to come. Second, allotting additional time for corpus activities, preferably evenly distributed in the curriculum to provide a sufficient amount of time and space for students to absorb the new methodology and tools offered to them and take slow but steady steps in implementing them in the solution of current problems topical to them. Finally, research would only benefit from an increased number of research subjects, activities and collected texts.

The research is small and unnoticeable, yet a victory in the application of corpora in teaching on a local level and could be recommended to anybody interested in corpora and teaching utilising corpus-based materials.

Theses

1. The present study is corpus-based and as such it is pious towards the available theory in relevant areas and does not consider itself to be capable of replacing theory by means of corpus research alone.
2. Corpora can potentially be integrated either directly or indirectly into teaching, depending on the choice of methods and materials.
3. By using corpora autonomously with slight mediation from a more experienced corpus user, students can successfully combine the roles of corpus user and corpus researcher.
4. The problems of misplaced context in corpus-based studies can be compensated for by authenticating the corpus texts to the context relevant to its users.
5. Writing a proper summary requires knowledge of specific features of academic writing as well as of features special to summarising such as complexity and impersonality.
6. A certain level of autonomy is expected of students to become skilled users of corpora assuming a moderate level of mediation.
7. The research subjects tend to abstain from any kind of extremities in their judgment, including sober evaluation of their current skills and talents.
8. Various groups can appear anomalous in their behaviour, failing to produce expected outcomes or perform in a predictable manner.
9. A broader scope of accessible corpus tools and their improved performance has the capacity of facilitating corpus analysis of learner corpora, which is crucial for further research in the area.
10. All groups share the trait of capturing the fundamentals of theory explained to them, varying in selectivity concerning details that are not on the surface.
11. The majority of the students express volition to continue working with the corpus following the piloting of the corpus-based activities, indicating a bright future for the application of corpora to teaching on local and global levels.
12. The students who express less satisfaction with corpora than their peers, primarily those in weaker groups, believe more time with corpus activities and detail of instruction would improve the piloting.

References

1. Biber, D., Conrad S., and Reppen R. (1998) *Corpus linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
2. Björk, L. and Räisänen, C. (1997) *Academic Writing: A University Writing Course*. Sweden: Studentlitteratur.
3. Fagan D. (2005) *Using Corpus Linguistics to Teach ESL*. San Diego: San Diego State University.
4. Fligelstone, S. (1993) Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal*, 17: 97–110.
5. Johns, T. (1991) From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal*, 4: 27–45.
6. Jordan R. R. (1989) *English for Academic Purposes: A guide and resource book for teachers*. Cambridge: Cambridge Language Teaching Library.
7. Kaur J. And Hegelheimer V. (2005) ESL Students' use of concordance in the transfer of academic word knowledge: an exploratory study. *Computer Assisted Language Learning*, 18 (4), 287-310. Routledge: Taylor & Francis Group.
8. Leech, G. (1997b) Teaching and language corpora: a convergence in Wichmann A., Fligelstone S., McEnery A. and Knowles G. (eds.) *Teaching and Language Corpora*, pp. 1-23. London: Longman.
9. McCormack J. and Slaght J. (2009) *Extended Writing & Research Skills*. Reading: Garnet Publishing Ltd.
10. McEnery T., Xiao R. and Tono Y. (2006) *Corpus-Based Language Studies: an advanced resource book*. London and New York: Routledge.
11. Mishan, F. (2004) Authenticating corpora for language learning: a problem and its resolution. *ELT Journal*, 58 (3): 219-227.
12. O'Keeffe A. and McCarthy M. (2010) *The Routledge Handbook of Corpus Linguistics*. London and New York: Routledge.
13. Oshima A. and Hogue A. (1998) *Writing Academic English: Third Edition*. New York: Addison Wesley Longman.
14. Renouf, A. (1997) *Teaching Corpus Linguistics to Teachers of English*, in Wichmann A., Fligelstone S., McEnery T. and Knowles G. (eds) *Teaching and Language Corpora*. London: Longman, pp. 255–66.

15. Sinclair J. (2004) *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
16. Teubert W. and Cermakova A. (2004) *Corpus Linguistics: A Short Introduction*. London: Athenaeum Press, Gateshead, Tyne & Wear.
17. Widdowson, H. G. (1978) *Teaching Language as Communication*. Oxford: Oxford University Press.
18. Yin R.K. (2003) Case study research: design and methods. *Applied Social Research Methods Series*, 5(3): 1-53. Available from <http://vivauniversity.files.wordpress.com/2012/10/1-yin-case-study-research-3rd.pdf> [Accessed 15 December 2013]
19. Yoon H. (2008) More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language Learning & Technology*, 12 (2): 31-48.
- Online 1. luis.lv. *English Spoken and Written Communication I*. Available from https://luis.lu.lv/pls/pub/kursi.kurss_dati?p_kods=2VAL1584&l=2&p_par= [Accessed 2 December 2012]
- Online 2. luis.lv. *English Spoken and Written Communication II*. Available from https://luis.lu.lv/pls/pub/kursi.kurss_dati?p_kods=2VAL1585&l=2&p_par= [Accessed 2 December 2012]
- Online 3. UEFAP. *Features of academic writing*. Available from <http://www.uefap.com/writing/feature/featfram.htm> [Accessed 2 December 2012]
- Online 4. Brigham Young University. *BYU-BNC: British National Corpus*. Available from <http://corpus.byu.edu/bnc/> [Accessed 16 December 2012]
- Online 5. University Centre for Computer Corpus Research on Language. *UCREL CLAWS7 Tagset*. Available from <http://ucrel.lancs.ac.uk/claws7tags.html> [Accessed 16 December 2012]
- Online 6. Anthony L. *Laurence Anthony's Software: AntConc*. Available from <http://www.antlab.sci.waseda.ac.jp/software.html> [Accessed 8 November 2013]
- Online 7. Conrad S. *Ideas for M.A. Thesis Topics from Susan Conrad (5/6/13 version)*. Available from <http://web.pdx.edu/~conrads/Thesis%20Ideas.pdf> [Accessed 8 November 2012]
- Online 8. Brigham Young University. *Corpus of Contemporary American English*. Available from <http://corpus.byu.edu/> [Accessed 2 December 2012].

- Online 9. Brigham Young University. Compare: The Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). Available from <http://corpus.byu.edu/coca/compare-bnc.asp> [Accessed 10 December 2013].
- Online 10. Warnes S. *Scoring a job: Interview preparation*. Available from <http://www.independent.co.uk/student/career-planning/getting-job/scoring-a-job-interview-preparation-8231208.html> [Accessed 2 January 2013].
- Online 11. Norris C. B. (2013) *Academic Writing English*. Available from <http://www.helsinki.fi/kksc/language.services/AcadWrit.pdf> [Accessed 2 January 2013].
- Online 12. Aston G. (2000) Learning English with the British National Corpus. In Battaner M.P. and López C. (eds.) *VI jornada de corpus lingüistics*. Barcelona: Institut universitari de lingüística aplicada, Universitat Pompeu Fabra. 15-40. Available from <http://www.sslmit.unibo.it/~guy/barc.htm> [Accessed 4 November 2013]
- Online 13. Corpora, size, queries = better resources, more insight. Available from http://corpus.byu.edu/userCategories_e.asp?s=y [Accessed 10 December 2013]

Appendix 1

Questionnaire 1

Please take your time to answer the following questions concerning your experience in education, writing and, if applicable, corpus linguistics.

1. What is your native language?

2. What is your current education?

- secondary
- other higher education

3. How would you evaluate the difficulty of academic writing?

- highly challenging
- moderately challenging
- tolerable
- not a problem at all

4. What aspects of academic writing would you most like to improve on?

5. On a scale from 1 to 5, how good would you say you are at the following (circle the relevant number):

maintaining the formal register

keeping the writing impersonal

using abundant and appropriate discourse markers

avoiding errors leading to plagiarism while quoting another source

6. Are you familiar with language corpora?

- yes
- no

If positive, tell briefly about your experience:

Appendix 2

Handout

Sample Summary

In “Euthanasia – Murder or Mercy” the author takes a position in favour of legalising euthanasia in order to reduce suffering and secure civil rights for those terminally ill and physically incapable of committing suicide. Weeks and months of unnecessary suffering can be avoided. Furthermore, she argues, to legalise euthanasia is after all only to grant terminally ill people the same rights as those have who take their own life. While acknowledging the potential misuse of legislation permitting euthanasia, by for instance greedy relatives or profit-seeking hospital administrators, the author maintains that satisfactory legal steps can be taken to protect the interests of the patient: since we have satisfactory laws regulating the beginning of life, i.e. artificial insemination, the legal system should be able to safeguard the end of life as well.

Summary	Characteristic Features
In “Euthanasia – Murder or Mercy” the author takes a position in favour of legalising euthanasia in order to reduce suffering and secure civil rights for those terminally ill and physically incapable of committing suicide. Weeks and months of unnecessary suffering can be avoided. <u>Furthermore</u> , she argues, to legalise euthanasia is after all only to grant terminally ill people the same rights as those have who take their own life. <u>While acknowledging</u> the potential misuse of legislation permitting euthanasia, by for instance greedy relatives or profit-seeking hospital administrators, the author maintains that satisfactory legal steps can be taken to protect the interests of the patient: <u>since</u> we have satisfactory laws regulating the beginning of life, i.e. artificial insemination,	<p><u>Source information</u>: only the title; the author is anonymous and the text unpublished.</p> <p><u>Overall purpose of the original</u>: “the author takes a position in favour of euthanasia”, i.e. a clear stand on an issue is indicated.</p> <p><u>Supporting arguments</u>: human suffering and civil rights issues.</p> <p><u>Objectivity</u>: no opinion about the original text is given.</p> <p><u>Vocabulary</u>: no clauses or sentences are quoted.</p> <p><u>Coherence</u>: the underlined words help to indicate relationships in the text.</p> <p><u>Summary markers</u>: “the author takes... she argues... while acknowledging... the author</p>

the legal system should be able to safeguard the end of life <u>as well</u> .	maintains.”
--	-------------

From: Björk, L. and Räisänen, C. (1997) *Academic Writing: A University Writing Course*.
Sweden: Studentlitteratur, p. 140

Appendix 3

Lesson Plan 1

Themes: summary writing, corpus linguistics

Level: tertiary level 1st year students of English philology, upper-intermediate and higher levels depending on group and student (groups 1A, 1B, 1C, 1D, 1FIN).

Length of the class: 90 min.

Aim of the lesson: to instruct the students on summary writing and introduce them to language corpora.

Objectives of the lesson:

1. Learn students' background and past experience, issues connected with academic writing students face and, if present, experience with linguistic corpora.
2. Observe how summarising organically unfolds prior to theoretical instruction.
3. Instruct the students on features of summary writing in connection with academic writing.
4. Accentuate students' attention on specific linguistic features of both summary writing and academic writing.
5. Enable the students to use the different queries in the Corpus of Contemporary American English (COCA Corpus) while logged into their COCA Corpus accounts.

Activities enabling objectives: listening, speaking, reading, writing, individual work, pair work, corpus activities.

Preparation:

- Computers for every student
- Video projector and a laptop with a .ppt presentation
- Files stored in estudijas, including the powerpoint presentation of the theory on summarising, a link to the text to be summarised and a link to the British National Corpus (BNC) engine within the Corpus of Contemporary American English
- A folder in 'studpasts' (students' post) containing a copy of the text to be summarised, in the event there are problems with computer speed, divided into group folders, where students submit their pre-summaries
- Handouts :
 - Questionnaire 1 – 1 copy per student.
 - A sample summary with features of a summary underlined - 1 copy per student

Assumptions: It is presumed that the students have the basic IT skills necessary to access the appropriate folders on the screen, log into their estudijas course and know the formal requirements of writing an academic text (name, font, text size et al.). They are also expected to have the appropriate grammar knowledge of 2nd term first year tertiary level students, such as distinguishing between active and passive voice, the tenses and similar knowledge. The students are acquainted with different kinds of paragraphs, which helps them with types of summary, as summary types match those of the texts they are summarising.

Procedure:

Task/Theme	Time
<p>1) Introduction/Questionnaire 1</p> <p>The teacher (T) introduces himself and the goals of his piloting. Students (S) receive handouts with questionnaire 1.</p>	10 min
<p>2) Pre-summary</p> <p>Ss are given the text they are to summarise and are instructed on the formal requirements of the file formatting and the location of the files in students' post. They read the text and summarise it in their preferred way without receiving any instruction. T supervises the process and observes the progression of the writing process, answering questions when necessary.</p>	35 min
<p>3) Theoretical instruction</p> <p>Ss listen to the theory on summarising, its definition and place in incorporating evidence in academic context, its features, as well as general expectations from academic writing. T asks Ss to read out examples (selected from T's corpus searches) from the slides so as to keep their attention to the subject matter. T moves on from summarising to corpus linguistics, giving brief information about corpora in general and the BNC in particular.</p>	25 min
<p>4) Corpus activity (The British National Corpus)</p> <p>Ss are asked to work in pairs, removing the danger of computer slowdown and integrating communication skills. They register on the COCA website and log into the corpus. T types queries into the corpus, explaining their significance and application, especially in connection to aiding the writing process; Ss repeat the queries and follow with their own.</p>	20 min

Home work: Students are asked to revise their pre-summary, either updating the first draft with their new knowledge of summarising theory and corpus usage. They are also to find 10 corpus examples of different features of summarising/academic writing until the next lecture.

Next class: Continuation of summary writing instruction.

Appendix 4

Lesson Plan 2

Themes: summary writing, corpus linguistics, error correction

Level: tertiary level 1st year students of English philology, upper-intermediate and higher levels depending on group and student (groups 1A, 1B, 1C, 1D, 1FIN).

Length of the class: 90 min.

Aim of the lesson: to reinforce the theoretical knowledge on summary writing by means of guided corpus and peer activities.

Objectives of the lesson:

1. Exemplify key features of summary writing by citing examples from students' own pre-summaries.
2. Reinforce the theoretical knowledge and practical corpus skills from the previous class.
3. Enable the students to improve their collaboration and presentation skills with the help of the group activity.
4. Gather a list of students' corpus search histories for analysis purposes.

Activities enabling objectives: listening, speaking, reading, writing, individual work, pair work, group work, corpus activities, presenting.

Preparation:

- Computers for every student
- Video projector and a laptop with a .ppt presentation
- Files stored in estudijas, including the powerpoint presentation of the theory on summarising, a link to the text to be summarised and a link to the British National Corpus (BNC) engine within the Corpus of Contemporary American English
- A folder in 'studpasts' (students' post) containing a copy of the text to be summarised, in the event there are problems with computer speed, divided into group folders, where students submit their pre-summaries
- A .ppt file featuring recurring errors in students' pre-summaries for the purpose of the corpus activity.

Assumptions: Same assumptions as before lesson 1 (see previous appendix) apply. Students are assumed to have been present during the previous class and benefited from both the theoretical and the practical instructions. The availability of all necessary materials in the e-course ensures absent students or those who need to revise the material can easily access the information in the comfort of their homes. Students absent from the previous class are able to

reach their peers owing to the pre-summary analysis and group work where teammates enter queries to produce corpus examples.

Procedure:

Task/Theme	Time
<p>1) Questions</p> <p>Ss ask questions concerning their two homework assignments and possibly generally about the theoretical and practical material introduced during the previous lesson. The presence of two instructors facilitates the process when several students in different parts of the room demand feedback or instruction.</p>	10 min
<p>2) Analysis of Pre-Summaries</p> <p>T analyses the pre-summaries written by the five groups, showcasing Ss' attention on some recurring errors and problems. The examples of errors appear on the slides complemented by appropriate rules or traditions of academic writing/summarising. T shows those aspects of summarising Ss rendered correctly prior to theoretical instruction.</p>	15 min
<p>3) Presentation of Corpus Examples</p> <p>Ss are divided into groups of 3-4, depending on the number of students in the class, and asked to share their corpus findings amongst each other, choosing 3 most interesting or creative examples of summarising/academic writing they found at home. The groups then present their findings to the whole class.</p>	35 min
<p>4) Error Correction</p> <p>Ss are given a file in students' post containing examples of typical errors in pre-summary writing carried out during the previous class and asked to find as many corpus examples of correct usage as possible.</p>	25 min
<p>5) Query Collection</p> <p>Ss enter their COCA search history to Word documents and save them anonymously with gibberish titles.</p>	5 min

Home work: Students write the final summary utilising all of the theoretical and practical knowledge attained during the two lectures.

Next class: Students continue studying Written Communication II without the researcher-teacher; they fill in Questionnaire 2 (post-questionnaire).

Appendix 5.

Post-Questionnaire

Please take your time to answer the following questions concerning the two pilot lessons in which corpus materials were integrated into the study of summary writing. Tick one option only, unless specified. The questionnaire is anonymous.

1. What auxiliary tools did you use during the writing process for the two piloting lessons involving corpus-integrated methods in summary writing? Tick one variant.

- corpus
- dictionary
- other (specify) :

Multiple / depends on the task/problem at hand

2. What do you believe would improve the integration of corpus materials in the summary teaching lessons? You may tick multiple boxes.

- more time with the corpus assignments
- simpler corpus interface
- clearer/more detailed instruction
- other (specify):

3. Do you expect yourself to continue using the corpus?

- Yes
- No

4. Which fields would you say the corpus is most helpful in? You may tick multiple boxes.

- more creative/academic writing
- simpler resolving linguistic issues
- research
- translation

other (specify):

4. Any additional comments that you may have:

Appendix 6.

Slides on Summarising Theory.

Summary Writing

Definition

- A short clear description that gives the main facts or ideas about something
- Incorporating evidence
in academic writing**
- Direct quotation
 - Paraphrasing
 - Summarising

Features of a Summary: what to do

- Full and clear information about the source text:
- Author's name
- Title
- Title of the printed source (newspaper, journal, book)
- (Björk and Räisänen, 1997: 122-123)

Features of a Summary: what to do

- Indicates the overall purpose/controlling idea
- *Main* supporting points
- Use personal vocabulary as much as possible
- As objective as possible

Features of a Summary: what to do

- Coherent (points linked to one another and the controlling idea)
- Summary markers:
- *The author suggests/argues/maintains...*

Features of a Summary: what not to do

- Patching together clauses and sentences from the original (=plagiarism)
- Not commenting on the original
- Chronology over coherence
- Using tenses other than present

In summary

- “give the reader, who has not read the original text, an adequate understanding of the nature and development of the ideas in that text”

- (ibid.)

Academic writing

- Features of academic writing tend to apply to summary writing

Complexity

- **Noun based phrases**
- Using nouns over verbs
- *Such evaluation (**and what it implies**) is already vaguely present in all the circumstances of ordinary life.*
- *-> Such evaluation (**and its implication**) is already vaguely present in all the circumstances of ordinary life.*

Complexity

- **Subordinate clauses**
- *I arrived late so I was not able to piece together the particulars of the argument, but enjoyed the climax just the same.*
- *-> **Having arrived late**, I was not able to piece together the particulars of the argument, but enjoyed the climax just the same.*

Complexity

- **Complement clauses**
- **That-clauses**
- *It was stated **that** solicitors would not find it possible to prepare the written briefs, and that it would be unfair to expect them to do so.*

Complexity

- **Complement clauses**
- **To-clauses**
- *This is not to suggest (as some have) that the attempt **to** oust Mrs Thatcher was initiated by the media.*

Complexity

- **Complement clauses**
- **Of+ing-clauses**
- *The danger **of suppressing** enterprise is an issue of which the American courts are also particularly conscious.*

Complexity

- **Sequences of prepositional phrases**
- *Unfortunately the report is so lacking **in detail** as to be virtually useless **in identifying** individual objects, and there are no accompanying photographs.*

Complexity

- **Participles**
- -ed (or appropriate irregular form)
- *The sections that follow (3.1-3.9) provide ordinary-language characterisations for reading experiences **located** at each of the idealised nodes on the network.*
- -ing

- Carrell (1983,1987), who has been much occupied in **determining** the effects of schematic knowledge on reading comprehension, refers to such knowledge constructs as 'content schemata'.

Complexity

- **Passive voice**
- **AV:** *God, or the Deity, existed because **somebody** had had to work out the mathematics in the first place.*
- **PV:** *If the all-provident mother of the hunter-gatherer societies had been lost with the coming of agriculture and weaning so that a divine substitute had had **to be found** in heaven in the shape of the mother-goddess.*

Complexity

- **Lexical density**
- written English has a higher lexical density than spoken English (Halliday, 1996, p. 347)

Complexity

- **Lexical density**
- *Obviously the government is frightened of union reaction to its move to impose proper behaviour on unions.*
- > (more lexically dense)
- *Obviously the government is frightened **how the unions will react** if it tries to make them behave properly.*

Complexity

- **Lexical complexity**
- Usage of prefixes and suffixes.
- *Restructure, **undertake**, **interpolate***
- *Quantify, **compromise**, **countless***

Complexity

- **Nominalisation**
- Judgment > judge
- Development > develop

Complexity

- **Attributive adjectives**
- The large number > the number is large (predicative)

Formality

- **No** informal lexis. E.g. *stuff, lots of, kinda*.
- **No** contracted forms. E.g. *can't -> cannot, didn't -> did not*.
- **No** phrasal verbs. *Put off -> postpone, go through -> experience*.
- **No** questions (including rhetorical ones)
- **No** numbering, sub-heading or bullet points.

Precision

- *A poll commissioned by someone a while ago showed that much of the population declared themselves to be broadly satisfied with living in the Albion, more compared to some other European countries.*

Precision

- *A poll commissioned by the EEC in 1977 showed that 82 per cent of the population declared themselves to be broadly satisfied with living in Britain, compared with 68 per cent in France and 59 per cent in Italy.*

Accuracy

- Misusing ‘irony’ in literary context
- Distinguishing shades of meaning between synonyms
- *Panorama vs landscape* etc.

Objectivity

- Avoid *I, me, my opinion* and so on
- Avoid personal pronouns like *you, one*
- *One can see in Europe that something fundamental had happened*
- *-> It is evident...*

Explicitness

- Signalling/transition signals/discourse markers
- Make it clear what you are stating
- *For instance* – signifies you are giving an example
- *However* – moves on to the next point

Responsibility

- Avoiding plagiarism
- Cite the author appropriately
- Give credit where it is due

Types of Summary

- Causal analysis
- Problem solving
- Argumentation
- Mixed
- Etc.

Corpus Linguistics

- **corpus**, plural **corpora** A collection of linguistic data, either compiled as written texts or as a transcription of recorded speech. The main purpose of a corpus is to verify a hypothesis about language - for example, to determine how the usage of a particular sound, word, or syntactic construction varies. **Corpus linguistics** deals with the principles and practice of using corpora in language study. A **computer corpus** is a large body of machine-readable texts.
- (cf. Crystal, David. 1992. *An Encyclopedic Dictionary of Language and Languages*. Oxford, 85)

British National Corpus

- The only large freely available corpus of BrE

- 100 million words
- Work: 1991-1994

References

- 1. Björk, L. and Räisänen, C. (1997) *Academic Writing: A University Writing Course*. Sweden: Studentlitteratur.
- Online 1. UEFAP. *Features of academic writing*. Available from <http://www.uefap.com/writing/feature/featfram.htm> [Accessed 2 December 2012].

Links

- <http://corpus.byu.edu/coca/compare-bnc.asp>
- <http://www.vivquarry.com/wkshts/prefixex.html>
- http://linguaspectrum.com/do_the_quiz.php?id=65

Appendix 7.

Feedback for Students on Pre-Summary Writing.

Pre-Summary Writing Analysis

Strengths

Synonyms

- irritating -> annoying
- resilient and determined > persistent

Conscious of register

- *The second part indicates*
- *Another suggestion is*
- *it is very advisable*

Passive Voice

- *a few ideas to be thought about*

Lexis

- *it would extinguish the employer's suspicion*

Problems Size

- ~200 words
- Formatting: no numbering
- One single paragraph

Unnecessary details

- *Even though it might seem annoying, at the end of the interview – ask questions.*
- *Almost at the end of the text, an encouragement is offered*
- *The general idea of the first section is that*
- Only **include** main supporting points!

Be Impersonal

- *First you would like to search out the job's structure and the industry more broadly.*
- *I suggest not to lose hope and keep trying.*

Matching person, number et al.

- *The article „Scoring a job: Interview preparation” advises young students on how to get his first job.*
- *One must get to know your surroundings, your requirements, the company, their rivals, their goals, it will show your determination.*

Gender

- *An applicant during an interview should not forget to ask questions when he's given the chance.*

Conclusions

- *In conclusion you have to remember that always have to ask questions because it shows that you are really interested in the job.*

Paraphrase!

- *First of all, you need to get into the vibe, or at the very least, get good at pretending to be really interested in what the company does.*
- -> plagiarism (borrows from the original without changing lexis/grammar)

Improving Register

- *Don't forget to act it all out (no contracted forms)*
- *you and the employer have clicked*
- *Think of some kind of possibly innovative solutions you could offer the company.*
- *This means that to try again and again in many other places until you get the job.*
- Avoid repetition

Signalling

- *As to questions, author claims that it is important to ask the interviewer about your potential job – this will show that you were really thinking about this job.*

Break into Sentences

- *When the actual interview is coming to a close you'll have the opportunity to ask questions, actually asking the employer some questions shows that you are truly interested in the position, and you have a chance to better understand the processes going on inside the company.*
- OR use the semicolon ;

Word Order

- *understand how it works, what are the company's goals and how they are achieved*

Tense

- Use present simple
- *Author is giving 5 steps*
- *Even if you don't have experience you have to find a way how to make yourself stand out.*

Modal Verbs

- *a person needs to have a small training session*

How A Corpus Can Help

- *Refusal is not a **reflexion** of your ability to do a job.*
- Refle*ion

How A Corpus Can Help

- *Understanding what kind of skills or attributes company wants their new employee to have might help you to imagine what questions will be asked **on the interview***
- * the interview

How A Corpus Can Help

- *Sophie Warnes notes that before an interview a student should get as much information as possible about the job he's applying to, **what the criteria's are** for the applicants and find information about the industry itself.*
- [Criteria]

How A Corpus Can Help

- *Doing a **dry run** might help to see what is working and what not.*
- Sections: academic

- Dry run

How A Corpus Can Help

- *they could have simply cut down **appliers** to the ones who live closer because of their multiplicity*
- Applier
- [=candidate]

How A Corpus Can Help

- *one must not frown, but must be **determinate***
- determinate

How A Corpus Can Help

- *Sometimes the main **reasons concerning** the rejection might not be even **worth worrying***
- [Reason] concerning
- Worth worrying

How A Corpus Can Help

- *It is also good to mention your special skills and previous experience.*
- [=good]