

Lexical Repetition in Academic Discourse:
A Computer-aided Study of the Text-organizing Role of Repetition

PhD Dissertation

Doktori (PhD) disszertáció

PhD Programme in Language Pedagogy

Doctoral School of Education

Eötvös Loránd University,

Budapest

Adorján Mária

Supervisor: Károly Krisztina, PhD, habil.

2015

Committee:

Chair: Medgyes Péter, DSc.

Internal Opponent: Pohl Uwe, PhD.

External Opponent: Prószéky Gábor, DSc.

Secretary: Tartsayné Németh Nóra, PhD.

Members: Pődör Dóra, PhD

Jasmina Sasdowska, PhD

Tankó Gyula, PhD

Acknowledgements

I would like to express my sincere gratitude to all the members of the Doctoral School of Language Pedagogy at Eötvös Loránd University for the most enjoyable three years of my life as a student, and for the continuous guidance they gave me during my PhD studies and related research.

I am immensely grateful for the assistance of my supervisor, Dr. Krisztina Károly, who patiently guided me with her insightful comments and encouragement throughout the whole process.

Last but not the least, I would like to thank my husband and children for their unwavering support.

ABSTRACT

Due to the various functions and diverse attitudes to lexical repetition in discourse, it is an aspect of cohesion which creates difficulty for raters when assessing L2 academic written discourse. Current computer-aided lexical cohesion analysis frameworks built for large-scale assessment fail to take into account where repetitions occur in text and what role their patterns play in organizing discourse. This study intends to fill this gap, by applying a sequential mixed method design, drawing on Hoey's (1991) theory-based analytical tool devised for the study of the text-organizing role of lexical repetition, and its refined version, Károly's (2002) lexical repetition model, which was found to be capable of predicting teachers' perceptions of argumentative essay quality with regard to its content and structure. It first aims to test the applicability of the previous models to assessing the role of lexical repetition in the organization of other academic genres, then propose a more complex, computer aided analytical instrument that may be used to directly assess discourse cohesion through the study of lexical repetition.

In order to test the applicability of Károly's model on other academic genres, two small corpora of thirty-five academic summaries and eight compare/contrast essays were collected from English major BA students at Eötvös Loránd University. The lexical repetition patterns within the corpora were analyzed manually in the case of the summaries, and partially with a concordance program in the case of the compare/contrast essays. The findings revealed that in both genres lexical repetition patterns differed in high and low-rated texts.

Given that in its present form the model cannot be used on large-scale corpora, in the third stage of the research, a computer-aided model was designed for large-scale lexical repetition analysis. First, by employing the theoretical, empirical and methodological results gained from the corpora, several new analytical steps were proposed and built into a modular format. Next, in order to better align the new computer-aided analysis to its manual version, parallel processes were identified between the new analytical model and an existing socio-cognitive framework. The newly proposed model may help teachers to assess discourse cohesion, or can be used as a self-study aid by visualizing the lexical net created by semantic relations among sentences in text.

Table of Contents

1. Introduction	1
1.1 <i>Background to the study</i>	1
1.2 <i>Aims of the present research</i>	5
1.3 <i>An overview of the dissertation</i>	6
 2 Theoretical Framework	 8
2.0 <i>Overview</i>	8
2.1 <i>Coherence and cohesion</i>	8
2.1.1 <i>Definitions of coherence and cohesion</i>	8
2.1.2 <i>Types of cohesion</i>	11
2.2 <i>Lexical cohesion and lexical repetition</i>	12
2.2.1 <i>Categories of lexical cohesion</i>	13
2.2.2 <i>Lexical chains or a lexical net?</i>	17
2.3 <i>Hoey's (1991) Repetition Model</i>	21
2.3.1 <i>The theoretical background of the model</i>	21
2.3.2 <i>Hoey's (1991) taxonomy of lexical repetition</i>	23
2.3.3 <i>Links and bonds creating a lexical net</i>	25
2.3.4 <i>The steps of the analysis</i>	28
2.3.5 <i>Applications of Hoey's (1991) model</i>	29
2.3.6 <i>Inconsistencies within Hoey's (1991) model</i>	34
2.3.7 <i>The link triangle and the mediator missing</i>	36
2.3.8 <i>The questions of anaphora resolution</i>	38
2.3.9 <i>Prescriptiveness or descriptiveness of the model</i>	39

2.4	<i>Károly's (2002) Repetition Model</i>	40
2.4.1	Károly's (2002) taxonomy of lexical repetition	40
2.4.2	Károly's (2002) method of analysis	41
2.4.3	Károly's empirical investigation	42
2.4.4	A corpus-based investigation using Károly's (2002) taxonomy	44
2.5	<i>Summary</i>	48
3	Methodological background: the academic writing context	49
3.0	<i>Overview</i>	49
3.1	<i>The nature of academic discourse</i>	50
3.1.1	General features of English academic discourse	50
3.1.2	The types of writing tasks required at university	53
3.1.3	Disciplinary differences in academic discourse	55
3.1.4	Implications for language pedagogy	56
3.1.5	Independent vs. integrative writing tasks	57
3.2	<i>Task variables influencing academic discourse quality</i>	59
3.2.1	The classification of variables in academic writing	59
3.2.2	Contextual variables of integrated academic discourse quality	62
3.2.3	Cognitive variables of integrated academic discourse quality	66
3.2.4	Summary writing as a complex task	70
3.2.5	Writing a compare/contrast essay	79
3.3	<i>Assessing academic discourse</i>	85
3.3.1	'Traditional' and recent academic essay assessment practices	85
3.3.2	Validity in L2 academic writing assessment	88
3.3.3	Generalizability of judgement on academic discourse quality	89

3.3.4	Reliability of perceived discourse quality	90
3.3.5	Text quality requirements by course teachers	95
3.3.6	Explicit instruction on coherence, cohesion and lexical repetition in higher education	96
3.3.7	Automated assessment of text quality	98
3.3.8	Controversial views on the automated assessment of essay quality	100
3.4	<i>Summary</i>	103
4	Aims and Research Questions	105
5	Research design and procedures of analysis	109
5.1	<i>A sequential mixed design</i>	109
5.2	<i>Stage 1: Analysis of academic summaries</i>	109
5.2.1	The summary writing task	109
5.2.2	Corpus size and representativity	111
5.2.3	Context validity evaluation of the summary writing task	112
5.2.4	Features of the input text	114
5.2.5	Quality assessment of the corpus	117
5.2.6	Methods of data analysis in Stage 1	125
5.3	<i>Stage 2: Analysis of compare/contrast essays</i>	131
5.3.1	The compare/contrast essay writing task	131
5.3.2	Quality assessment of the corpus	133
5.3.3	Methods of data analysis in Stage 2	134
6	Results of the lexical repetition analysis of academic summaries	141
6.1	<i>General features of the summaries</i>	141
6.2	<i>Results related to repetition type</i>	142

6.3	<i>Results related to the combination of links and bonds</i>	146
6.4	<i>Methodological outcomes</i>	151
6.5	<i>Summary</i>	151
7	Results of the lexical repetition analysis of compare/contrast essays	156
7.1	<i>General features of the compare/contrast essays</i>	156
7.2	<i>Results related to repetition type</i>	157
7.3	<i>Results related to the combination of links and bonds</i>	159
7.4	<i>Features not detected</i>	161
7.5	<i>Methodological outcomes with automation in mind</i>	162
7.6	<i>Summary</i>	163
8	The design of a new LRA model for large-scale analysis	165
8.1	<i>The newly proposed LRA model: the three modules of the analysis</i>	165
8.2	<i>Phase 1: Preparation of the corpus</i>	166
8.2.1	Plagiarism check.....	166
8.2.2	L2 special corpora treatment / Error treatment.....	167
8.2.3	POS tagging.....	169
8.2.4	POS tagging for lower level L2 texts.....	171
8.2.5	Using WordNet with the existing taxonomy.....	172
8.2.6	Using WordNet with errors in a learner corpus.....	173
8.3	<i>Phase 2: Finding links</i>	173
8.3.1	Theoretical considerations: altering the taxonomy.....	173
8.3.2	Introducing the concept of ‘key term’ into the coding process.....	175
8.3.3	Lexical unit identification in the case of multiword units.....	176

8.4	<i>Special use of the model for academic summary writing</i>	177
8.5	<i>Visual representation of links and bonds</i>	178
8.6	<i>Connecting the new LRA model to a cognitive framework</i>	180
8.7	<i>The scope and limitations of the new LRA model</i>	182
9	Conclusions	183
9.1	<i>Summary of main results</i>	183
9.2	<i>Pedagogical implications</i>	184
9.3	<i>Limitations</i>	185
9.4	<i>Terminology issues</i>	186
9.5	<i>Suggestions for further research</i>	186
	References	190
	APPENDICES	212

List of Tables

Table 1 Types of cohesive devices in Halliday and Hasan (1976) with the researcher's examples	12
Table 2 The changes of lexical cohesion taxonomies based on Halliday and Hasan's (1976) and Hasan's (1984) models	14
Table 3. Types of repetitions based on Hoey's (1991) taxonomy.....	24
Table 4 Types of lexical relations in Károly's taxonomy with examples (examples based on Károly, 2002, p. 104, and these two corpora)	40
Table 5. Variables with the strongest predictive power in Károly's (2002) lexical repetition analysis research	43
Table 6. The difference in explicitness caused by phrasal vs. clausal modification (Biber & Gray, 2010)	52
Table 7. Features of task setting and features of input text within contextual variables (based on Weir's (2005) socio-cognitive framework).....	62
Table 8. Cognitive variables involved in integrative academic discourse based on Chan (2013) and Chan, Wu, & Weir (2014).	66
Table 9. The differences between writing a summary 'essay' vs summarizing.....	71
Table 10. Whole text summary task and guided summary task questions for idea selection based on Tankó, 2012, p. 119.....	72
Table 11 Mental processes during summarization as suggested by four models.....	73
Table 12. The constructs, their divisions, and the cognitive processes involved in the summary task of the research (Stage 1)	74
Table 13. Assessing written academic discourse	87
Table 14. Subscales and descriptions of the IELTS Writing Module (based on Cambridge IELTS webinar 26. February, 2014)	92
Table 15. University teacher's requirement (based on Moore & Morton, 1999) analyzed from a language pedagogy angle.....	96
Table 16. The various uses of Coh-Metrix in analyzing L2 writing.....	102
Table 17. Summary of the intended approach of data collection and analysis	106
Table 18. Basic features of the summary task.....	110
Table 19. The features of task setting in Stage 1.	112
Table 20. Features of the input text for Stage 1 summary task.....	114
Table 21. The constructs, their divisions, and the cognitive processes involved in the summary task of the research (Stage 1)	121
Table 22. Overview of assessments, methods and their aims in Stage 1	124
Table 23. Summary of the categories of repetition (based on Károly, 2002)	125
Table 24. The number of bonds pointing backward and forward within Text 3.	129
Table 25. Compare/contrast essay evaluation: the organization and discourse control rubrics..	134
Table 26. Károly's (2002) taxonomy for LRA with examples from the compare/contrast essay corpus.....	135
Table 27. The summary construct.....	141
Table 28. The frequency of types of repetition in high- and low-rated summaries	142
Table 29. The difference between the mean frequencies of SUR and DUR in high-and low-rated summaries	143

Table 30. Comparison of results related to lexical repetition types between Károly's (2002) argumentative essay corpus and Stage 1 summary corpus	146
Table 31. Frequency of links and bonds and density of bonds	147
Table 32. Possible lexical repetition patterns in one or multiple paragraph summaries.....	154
Table 33. The frequency of types of repetition in high- and low-rated essays. Abbreviations: SUR: Same unit repetition. DUR: Different unit repetition	157
Table 34. Comparison of results related to lexical repetition types between Károly's (2002) argumentative essay corpus and Stage 2 compare/contrast essay corpus	158
Table 36. Two interpretations of a learner sentence and their analyses (based on Dickinson & Ragheb, 2013)	172
Table 37. Proper names as possible topics in four disciplines	176
Table 35. The contextualization of the new LRA model in this study: the parallel processes between the model and the cognitive processes during reading (Khalifa & Weir, 2009), with explanations	181

List of Figures

Figure 1 Illustration of a semantic network for business trips (based on Grimm, Hitzler, & Abecker, 2005, p. 39) Nouns represent the concepts (in rectangles), the arrows specify the relationships between the concepts.	15
Figure 2. Visual representation of a gene ontology within the field of biology (based on the online training material of the European Bioinformatics Institute).....	16
Figure 3. Topical Structure Analysis indicating semantic links between sentences (Lautamatti, 1987, p. 102).....	18
Figure 4. A net and a chain of lexical repetition in two studies (Hoey, 1991, p. 81; and Barzilay & Elhadad, 1999, p. 116).....	21
Figure 5. The link triangle (Hoey, 1991, p. 65).....	37
Figure 6. The General—Particular relationship in text (Hoey, 1995, p.135).....	38
Figure 7. Synonyms offered for the word <i>bank</i> in WordNet.....	46
Figure 8. Three paragraphs of a sample compare/contrast essay indicating some of the lexical repetition links (adjectives/adverbs). Text: <i>Oxford Advanced Learners' Dictionary</i> 8 th ed. (OUP, 2008).....	47
Figure 9. A priori (before the test) components of Weir's (2005) validation framework.....	61
Figure 10. Weir's (2005) whole socio-cognitive framework.....	62
Figure 11. The Cognitive Process Model of the Composing Process (Flower & Hayes, 1981, p. 370).....	68
Figure 12. The notion of comparison (Mitchell, 1996).....	80
Figure 13. Suggested outline for the point-by-point pattern (based on the online academic writing guidelines of Humber University).....	83
Figure 14. The block-by-block pattern.....	84
Figure 15. The hook-and-eye technique connecting major thoughts on discourse level (in Cresswell, (2007, p. 59).....	97
Figure 16. A detail of the repetition matrix of Text 3, itemized and classified.....	127
Figure 17. The matrix showing the number of links between each sentence of Text 3.....	128
Figure 18. The span of bonds in Text 3.....	130
Figure 19. The strength of connection between bonded sentences in Text 3.....	130
Figure 20. A detail of the headword frequency list (Text 4). Headwords are listed according to occurrence. (N.= number, % = the percentage of the occurrence in the text).....	138
Figure 21. Another detail of the headword frequency list (Text 3). The numbers represent the sentences in which the words occur. (No.1 = the title).....	139
Figure 22. The span of bonds in Text 1. The two figures (e.g., 2-3) show the two sentences connected by bonds.....	160
Figure 23. The three big modules of the new LRA model.....	165
Figure 24. The steps of the new LRA model.....	166
Figure 25. Teich and Fankhauser's lexical repetition analysis links in text view (2005).....	179
Figure 26. Visual representation of topic sentence and conclusion sentence identified as central by LRA (text based on Oshima & Hogue, 2006).....	179

1. Introduction

1.1 Background to the study

Cohesion and coherence in text have become widely researched areas within the field of discourse analysis, and a great deal of attention has been given to the subjects of lexical cohesion and lexical repetition due to their significant discourse function (e.g., Halliday, 1985; Halliday & Hasan, 1976; Hoey, 1991; Reynolds, 1995, 2001; Tyler, 1994, 1995). Discourse is “a unit of language larger than a sentence and which is firmly rooted in a specific context” (Halliday & Hasan, 1990, p. 41). Lexical cohesion was defined by Hoey (1991) as “the dominant mode of creating texture” because it is “the only type of cohesion that regularly forms multiple relationships” in text (p.10). He called these relationships lexical repetition, using repetition in a broader sense, referring not only to reiterations but also various other forms of semantic relatedness, such as synonyms, antonyms, meronyms, as well as other paraphrases.

Based on Halliday and Hasan’s (1976) empirical investigation of cohesive ties in various text types, Hoey concluded that lexical cohesion accounted for at least forty percent of the total cohesion devices (1991). In a more recent corpus linguistic study Teich and Fankhauser (2004) claimed that nearly fifty percent of cohesive ties consist of lexical cohesion devices (p. 327), thus making lexical cohesion the most pronounced contributor to semantic coherence.

The study of lexical cohesion is relevant to language pedagogy because what and how to repeat in English written text causes disagreement among native and non-native language users alike. Most teachers would agree with Connor (1984), for instance, who found that in students’ texts repeated words were both a sign of limited vocabulary and of poor text structuring. The problem is more complex however, because lexical choice depends not only on language proficiency level, but on various other factors as well. According to Reynolds (2001), for example,

lexical repetition used by writers changes in relation to (1) writing topic, (2) cultural background, and (3) development of writing ability; the third being the most determining factor. Myers (1991) also found that scientific articles generally require more reiterations than popular articles because exact concepts in this field cannot be replaced by synonyms. Therefore what should and what should not be repeated in academic writing is context dependent, and this complexity calls for more research into lexical cohesion in general, and into texts produced by language learners on various topics and genres in particular.

Lexical cohesion is studied both in text linguistics (discourse analysis) and corpus linguistics, which two terms cover related but not the same kind of approaches to the study of text. In discourse analysis, first the various cohesive devices are categorized according to semantic relatedness criteria and a theoretical framework is built, which is later tested on a small number of texts. Lexical repetition patterns are analyzed quantitatively and manually (e.g., the researcher counts how many times certain categories are represented in the text) as well as qualitatively (e.g., conclusions are drawn observing the types, location and lexical environment of repeated words). The main problem with this type of analysis is that only a small number of texts can be observed; therefore, the data gained do not permit generalizations.

The other approach to lexical cohesion analysis is offered by corpus linguistics, which allows for automated analysis of large linguistic data. A disadvantage of this method is that individual differences within texts in a corpus cannot be observed. Reviewing best practice in text-based research, Graesser, McNamara, and Louwse (2011) maintain that the recent shift in discourse analysis is characterized by moving from “theoretical generalizations based on empirical evidence observing a small corpus to large-scale corpus-based studies” (p. 37), and the results have

changed “from deep, detailed, structured representations of a small sample of texts to comparatively shallow, approximate, statistical representations of large text corpora” (p. 37).

Several manual and computer-aided methods exist to analyze lexical features in text. Of particular interest are frameworks capable of not only identifying and classifying linguistic elements but also providing information on their patterns and roles in structuring text. Hoey’s (1991) theory-based analytical tool designed for the study of lexical repetition is the first one of the frameworks devised to offer a manual analytical method for studying the text-structuring role of lexical repetition. This framework explores the semantic network (**links, bonds and the lexical net**) of text and distinguishes between central and marginal sentences by finding lexical repetition patterns. With this method it is possible to summarize certain types of discourse.

Hoey’s (1991) comprehensive analytical model was later revised by Károly (2002) who made significant changes in the categories. Károly also extended the model by introducing several new analytical steps to reveal the organizing function of lexical repetition in texts. Hers was the first application of Hoey’s model in a Hungarian higher education setting. Károly’s (2002) research results showed that her theory-driven ‘objective’ analytical tool not only offered a descriptive function, but with her analytical procedures the tool was capable of predicting the ‘intuitive’ assessment of teachers judging argumentative essay quality with regard to its content and structure.

Given that in holistic scoring teachers assign more weight to content and organization than to any other components (Freedman, 1979), and given that these two components comprise the concepts of cohesion and coherence, responsible for textuality (Halliday & Hasan, 1976), it is of little surprise that lexical repetition analysis (LRA) can detect the difference between valued and poor writing. The results of Károly’s (2002) analysis proved that the texts, which had previously

been judged by experienced university instructors, differed significantly in both repetition types and patterns. Post-tests conducted with another group of teachers confirmed these findings indicating that the analytical measures devised can reliably predict how teachers perceive essay quality, and the results may be generalized for a wider sample.

Hoey (1991) found predictable lexical repetition patterns in news articles, whereas Károly (2002) studied the academic argumentative essay genre in this respect. Due to the fact that summary and compare/contrast essay are the two most commonly used genres across the disciplines at universities (Bridgeman & Carlson, 1983; Moore & Morton, 1999), these integrative (reading into writing) tasks also deserve such thorough investigation. Therefore, research needs to be extended to the predictive power of Károly's model in genres most likely faced by EFL (English as a Foreign Language) students across universities in Hungary. At the moment no study exists applying Károly's (2002) lexical repetition analysis (LRA) model for the genres of the academic summary and compare/contrast essay. Neither does a tool exist with a similar theoretical basis that can be applied to large-scale corpora using the same method for other academic genres.

Large-scale essay assessment applications (such as E-rater¹, Intelligent Essay Assessor²) have been in use for decades to test essay writing skills in the EFL context. However, these applications were developed by major testing agencies and are not available to the public. These essay scoring programs measure cohesion and coherence quantitatively, using statistical methods and natural language processing (NLP) techniques. Their methods focus on cohesion and coherence on the local level, mainly by comparing adjacent sentences semantically, on the assumption that words in adjacent sentences form semantic chains which can be identified for topic

¹ <https://www.ets.org/erater/about>

² http://www.lt-world.org/kb/ipr-and-products/products/obj_80512

progression. This can be called the **lexical chain** principle. However, these chains are linear in nature, and indicate cohesion on the local level, whereas discourse also shows global cohesion. If text is considered to create (or to be created by) **lexical nets**, it is necessary to observe semantic links between all the sentences in the text, even if they are located far from each other, in other words: if the lexical chain principle is switched for the lexical net principle. Károly's (2002) lexical repetition analysis tool, which was based on Hoey's (1991) Repetition Model, can "measure" discourse cohesion, yet in its present form it is not fit for large-scale application.

1.2 Aims of the present research

With the above in mind, the purpose of the study is

- (1) to extend the use of Hoey's (1991) and Károly's (2002) lexical repetition model to the academic summary and the compare/contrast genres by analyzing two Hungarian EFL university student corpora;
- (2) to test whether Károly's (2002) analytical tool can predict teachers' judgement regarding discourse quality in the case of these two genres, too, and
- (3) to alter this analytical tool to enable large-scale analysis of EFL student corpora
- (4) in order to be able to design the steps and modules necessary for a computer-assisted lexical repetition analysis.

The main questions guiding this research are therefore the following:

- (1) Is Károly's (2002) theory-based lexical repetition model, a revised version of Hoey's (1991) Repetition Model applicable to the study of summaries and compare/contrast essays written by Hungarian EFL university students?
- (2) What modifications are needed in Károly's (2002) theory-based lexical repetition model to be applicable to large-scale EFL learner corpora?

The study uses a mixed methods design including both qualitative and quantitative methods, as suggested by Creswell (2007); and a sequential mixed design paradigm, as described by Tashakkori and Teddlie (2003). The rationale behind using qualitative and quantitative methods is that, according to Tyler (1995) and Károly (2002), quantitative analysis alone cannot inform research about the real role of lexical repetition in organizing discourse. In the first stage of this study, the model is applied to the summary genre. The second stage utilizes results gained from the first stage and continues to test the model on compare/contrast essays. In this second stage, a concordance analyzer³ is introduced at the initial step of the analysis. In the third stage, the theoretical, empirical and methodological results of the previous stages form the basis of the design of the new, semi-automated analytical tool. Therefore, results gained from each stage inform the next stage of research according to the sequential mixed design paradigm.

This study is multidisciplinary in nature, aiming to contribute to the fields of (a) applied linguistics, more closely to discourse analysis and corpus linguistics; (b) language pedagogy, especially to the teaching and evaluating EFL academic writing; and (c) computer science, to enhance educational software development.

1.3 An overview of the dissertation

This dissertation is organized into nine parts: Chapter 2 surveys the literature on theoretical aspects of coherence, cohesion and lexical repetition. Hoey's (1991) and Károly's (2002) repetition models are compared, followed by a description of the major computer-aided applications of these models.

³ a program which displays every instance of a specified word with its immediate preceding and following context

Chapter 3 focuses on the context of academic discourse: from task setting to assessment. First the chapter enumerates the typical written academic genres across disciplines, followed by the description of the task variables influencing discourse quality. The requirements of the summary and compare/contrast essay genres, the two genres investigated in this study, are described next. The final part of the chapter deals with academic discourse assessment, with particular emphasis on teachers' perceptions of coherence, cohesion and lexical repetition in students' texts. Basic findings of research into automated summarization and automated essay scoring are also introduced.

Chapter 4 enumerates and explains the research questions. Chapter 5 describes the research design used in this study, introducing the detailed steps of analysis in both Stages 1 and 2 of the research, where Stage 1 aims to test Károly's (2002) repetition model on academic summaries and Stage 2 extends the model to compare/contrast essays. Chapters 6 and 7 give account of the results of Stages 1 and 2, respectively. Chapter 8 presents the new computer-aided model designed for the lexical repetition analysis of larger corpora. Chapter 9 summarizes the main findings of the study, and considers the implications, limitations and the possible areas of future research.

2 Theoretical Framework

2.0 Overview

The aim of this chapter is to provide a theoretical background to the study of the text organizing role of lexical repetition in order to be able to propose a new computer-aided analytical model later on. It offers a brief introduction to the theories behind the two basic concepts of textuality: coherence and cohesion, with special emphasis on lexical cohesion and lexical repetition, giving definitions of how these key terms are used in this paper. After the presentation of Hoey's (1991) and Károly's (2002) lexical repetition models, as well as Károly's (2002) empirical investigation, which is the starting point of this research project, some examples follow of how these models have been applied on large corpora. The strengths and weaknesses of previous analyses are also highlighted so as to fulfil the theoretical, empirical, and methodological aims of the current investigation.

2.1 Coherence and cohesion

2.1.1 Definitions of coherence and cohesion

The complex nature of coherence and cohesion offers grounds for a wide spectrum of interpretations. Károly (2002), in her concise review of the most influential schools of English written text analysis, distinguishes three groups among the various descriptive models of coherence (1) those focusing on the surface elements of text and their combinations, (2) those defining coherence as an interaction between textual elements and the mind, and (3) those claiming that coherence is created in people's minds entirely.

In the first group belong Halliday and Hasan (1976), and later Halliday (1985), who see coherence as created by surface textual elements (e.g., identity and similarity chains).

'Interactionalists', who comprise the largest of the three groups, such as van Dijk and Kintsch (1983), define coherence as a cognitive interaction between the reader and the textual elements. According to de Beaugrande and Dressler (1981), coherence refers to "how the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant" (pp. 3-4). Hoey also contends that coherence is a "facet of the reader's evaluation of a text" (1991, p. 12). More recently, coherence was defined by Crossley and McNamara as "the understanding that the reader derives from the text" (2010, p. 984), its main factors being prior knowledge, textual features, and reading skills (McNamara, Kintsch, Songer, & Kintsch, 1996). The relatively few theoreticians in the third group, such as Sanders, Spooren and Noordman (1993), approach the notion of coherence, as not being a property of discourse, rather a mental representation of it in people's brains (p. 94).

A more consistent approach can be observed towards textual cohesion, because most researchers agree that cohesion is a property of the text. Halliday and Hasan (1976) maintain that cohesion is "the relation of meaning that exists within the text, and that define it as a text" (p.4). According to their interpretation, cohesion occurs "where the interpretation of some element in the text is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it."(ibid.). Thus, cohesion largely (but not exclusively) contributes to coherence.

De Beaugrande and Dressler (1981) see cohesion as one of their six criteria of textuality: cohesion, coherence, intentionality, acceptability, situationality and intertextuality. They claim that cohesion is indispensable for a text to be a text. Enkvist (1990) and Hoey (1991) also see cohesion as a property of text and therefore being objectively observable. Hoey's definition is perhaps a little blurred because he concentrates on the textual roles of sentences: "Cohesion may be crudely

defined as the way certain words or grammatical features of a sentence can connect that sentence to its predecessors (and successors) in a text” (1991, p. 3). The nature of cohesion might be better captured instead focusing on the text, as “[c]ohesion refers to the presence or absence of explicit clues in the text that allow the reader to make connections between the ideas in the text” (Crossley & McNamara, 2010, p. 984). Such explicit clues can be “overlapping words and concepts between sentences” or connectives such as *therefore* or *consequently* (ibid.).

Widdowson (1978) takes a different approach towards cohesion. He argues that cohesion is neither necessary, nor sufficient for coherence. His famous example for this is the following conversation (p. 29):

A: That’s the telephone.

B: I’m in the bath.

A: O.K.

Even though this short exchange is an example of spoken discourse where coherence can be detected across turn boundaries (what A says /what B says /what A says), it still demonstrates that coherence can exist without the explicit markers of cohesion. As Widdowson puts it: coherence is by nature interactive, while cohesion is within discourse, which is by nature static (as cited in Fulcher, 1989, p. 148).

Given that there is disagreement in the literature about coherence and cohesion, for the purposes of this research project, the following two similar sets of definitions will be used for cohesion and coherence:

- (1) Coherence is “the understanding that the reader derives from the text” (Crossley & McNamara, 2010, p. 984).

- (2) “Cohesion refers to the presence or absence of explicit clues in the text that allow the reader to make connections between the ideas in the text” by (Crossley & McNamara, 2010, p. 984).

and;

- (1) “[C]oherence is the quality that makes a text conform to a consistent world picture and is therefore summarizable and interpretable.” (Enkvist, 1990, p. 14)

- (2) “Cohesion is the term for overt links on the textual surface [.]” (Enkvist, 1990, p. 14)

From these two coherence definitions, the phrases “reader derives from the text”, “quality”, “consistent world picture”, and “interpretable” are the key terms. From the two cohesion definitions “presence or absence”, “overt links”, “explicit clues”, “connections between ideas”, and “textual surface” are the most important terms for this study, because the presence or absence of overt links and explicit clues will be observed first quantitatively, and next qualitatively, in order to interpret readers’ quality judgements derived from the text.

2.1.2 Types of cohesion

Halliday and Hasan (1976) in their analytic model identify the semantic and lexicogrammatical elements which are responsible for creating texture in English. The five categories are **reference**, **substitution**, **ellipsis**, **conjunction**, and **lexical cohesion**, shown in Table 1 with examples. The first four are mainly grammatical categories, and as such, they are fairly straightforward. The category of **lexical cohesion** seems more problematic, with its two subclasses: **reiteration** and **collocation**. (This category will be analyzed in Section 2.2.1). The cohesive relation between any two of these lexical elements is called a **cohesive tie**. These ties form **cohesive chains**, and the interactions among chains further cause global “cohesive harmony” (Hasan, 1984) in text.

Type of cohesive relation	Example
Reference	<i>Rob</i> has been driving. <i>He</i> is tired.
Substitution	Do you <i> speak </i> Irish? I <i> do </i> a little.
Ellipsis	Are you <i> sick </i> ? Yes, I am. (<i> sick </i> is missing)
Conjunction	This register is academic. <i> Therefore </i> , formal vocabulary is necessary.
Lexical cohesion	This <i> girl </i> is pretty. That <i> girl </i> is ugly, though. This girl is <i> pretty </i> . That girl is <i> ugly </i> , though.

Table 1. Types of cohesive devices in Halliday and Hasan (1976) with the researcher's examples

In her later work, Hasan (1984) changed the categories of lexical cohesion considerably, indicating that semantics is an area where items are particularly difficult to classify.

2.2 Lexical cohesion and lexical repetition

Lexical cohesion, lexical organization and their roles in establishing coherence have been the focus of several influential studies (e.g., Halliday & Hasan, 1976, 1985; Hoey, 1991; Reynolds, 1995, 2001; Sinclair, 1998; Tyler, 1992, 1994, 1995). Hoey focused his research on cohesion instead of coherence, on the assumption that markers of cohesion appear in the text as observable features. In his view, the study of coherence is outside the scope of textual analysis because only explicitly manifest data can be analyzed (1991). He maintains that lexical repetition items are suitable for 'objective' analysis as they appear on the surface of text as categorizable and countable items.

Lexical cohesion was defined by Hoey (1991) as "the dominant mode of creating texture", because it is "the only type of cohesion that regularly forms multiple relationships" in text (p.10), making it unique among cohesive devices. His empirical investigation indicated that lexical cohesion accounted for more than forty percent of the total cohesion devices in the various texts he studied (1991). In a more recent corpus linguistic study it was claimed that nearly fifty percent

of a text's cohesive ties consist of lexical cohesion devices (Teich & Fankhauser, 2004), thus making it the most pronounced contributor to semantic coherence.

A further argument for the relevance of lexical repetition studies is offered by Stubbs (2004), in the *Handbook of Applied Linguistics*. In his chapter on corpus linguistics, he makes the following observation when describing the importance of word frequency lists: "A few, mainly grammatical, words are very frequent, but most words are very rare, and in an individual text or smallish corpus, around half the words typically occur only once each" (p.116). If we reverse this logic, this statement also implies that half of the words in the text **do** occur at least twice in any individual text or smallish corpus. Even if we do not count further types of lexical repetition, such as repeating by synonyms or antonyms, etc., according to this observation, the number of repeated words in any text seems impressive and certainly warrants relevance for further studies, (regardless of the fact that these items occurring in multiple occasions are most probably function words, e.g., *be, and, I*).

2.2.1 Categories of lexical cohesion

Lexical cohesion plays an important role in creating texture (Halliday & Hasan, 1976): it is the central device that makes the text hang together, defining the "aboutness" of text (ibid, Chapter 6). It is interesting to observe the shifts of categories within lexical cohesion in two different models: Halliday and Hasan's 1976 model and Hasan's 1984 revision. Table 2 offers a comparison of these two models.

It is noteworthy that exact repetition of a word comes first in both models although in Halliday and Hasan's (1976) model it is called **reiteration**. The reason for this might be that repeating the same semantic item is the most obvious and easily recognizable way to create semantic relatedness. Reiteration however, in the first model also comprises synonymy, and

superordinates, as well as the vague category ‘general’ item, thus widening the concept of reiteration beyond the traditional grammatical sense. Later Hasan (1984) changed the categories, separating repetition of the same base form from any other, orthographically different, semantic items.

Halliday and Hasan (1976)	Hasan (1984)
1. Reiteration	1. Repetition
1.a same item repetition	2. Synonymy
1.b synonymy / hyponymy	3. Antonymy
1.c superordinate	4. Hyponymy
1.d ‘general’ item	5. Meronymy
2. Collocation	6. Instantial relations
(including:antonymy, meronymy)	6.a equivalence
	6.b naming
	6.c semblance

Table 2. The changes of lexical cohesion taxonomies based on Halliday and Hasan’s (1976) and Hasan’s (1984) models

Another interesting category is that of **collocation**, which subsumes antonymy, meronymy, and, as Hoey put it, a “ragbag of lexical relations” (1991, p. 7). Word pairs classified as collocation are, for instance, *laugh – joke*, *try – succeed*, or *ill – doctor* (Halliday & Hasan, 1976, pp. 285-286). The term collocation in connection with co-occurrence was thoroughly discussed in studies in the sixties (see e.g., Halliday, 1966; or for a review: Langedoen, 2009). In more recent publications collocations are referred to as words frequently appearing together within a sentence, including phrases such as *bright idea*, or *talk freely*, and are described as “essential building blocks of natural-sounding English”⁴. In the case of the above examples, however, it is not a requirement for the word pairs to occur intrasententially: they can appear in different sentences.

Two observations need to be made here concerning Halliday and Hasan’s (1976) lexical cohesion classification. The first is that, even though several discourse analysts (e.g. Hoey, 1991;

⁴ examples and description from the online version of Oxford Collocations Dictionary, <http://oxforddictionary.so8848.com>

Martin, 1992; Tanskanen, 2006) criticized this taxonomy as offering categories which were partly unjustified and partly too vague (typically meaning the category of collocation), time has justified Halliday and Hasan's decision to include non-systematic sense relations in their taxonomy, such as the above examples. The pairs *laugh – joke*, *try – succeed*, or *ill – doctor* are all semantically related; and all of them, although they cannot be easily categorized, are perceived as cohesive ties. These types of relations are dealt with in psycholinguistics, and can now be analyzed using computer-aided corpus linguistic techniques, after a long wait of almost twenty years in the history of automated textual analysis.

One way of analyzing text for such non-systematic sense relations as *ill – doctor* is applying knowledge representation (KR) frameworks commonly built for AI (Artificial Intelligence) purposes (Davis, Shrobe, & Szolovits, 1993). Semantic networks and maps are built to represent world knowledge (i.e., everything humans know but a computer does not, for example: When we feel *ill*, we go to the *doctor*). Figure 1 illustrates the semantic network for *business trip*.

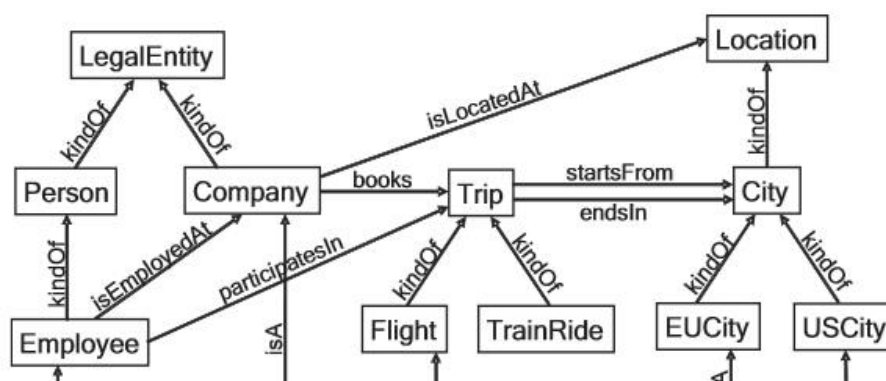


Figure 1. Illustration of a semantic network for business trips (based on Grimm, Hitzler, & Abecker, 2005, p. 39) Nouns represent the concepts (in rectangles), the arrows specify the relationships between the concepts.

These networks not only contain concepts (nouns, adjectives, etc.), but also the hierarchy relations between them (*X is a part of Y*) are defined and taught to the program. Rules, such as *IF X, THEN Y*; or logic, such as *X is an employee, therefore X must be a person*, are also provided to

improve the knowledge base of the application. This way ontologies (hierarchies) are built for each domain (each field of knowledge) to be able to serve the Semantic Web (see e.g., Grimm, Hitzler, & Abecker, 2005 for more on this topic). Figure 2 shows part of a visual ontology for the Biology field. As can be seen, based on the examples provided in Figures 1 and 2, world knowledge constitutes a large part of perceived coherence and cohesion in texts.

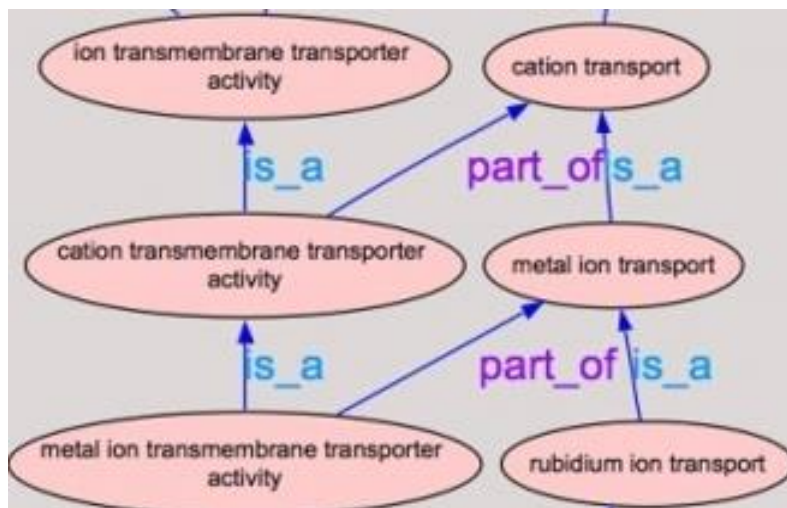


Figure 2. Visual representation of a gene ontology within the field of biology (based on the online training material of the European Bioinformatics Institute⁵)

Another new way of analyzing text for non-systematic sense relations is applying Word Association Norms (WAN) to a corpus. WANs are an alternative lexical knowledge source to analyze linguistic computational creativity, in order, for example, to explore lexical associations common in poetic texts (Netzer, Gabay, Goldberg, & Elhadad, 2009). These norms are a collection of cue words and subsequent sets of free associations collected from people as respondents to the cue words. WANs are used with statistical measures to analyze, for example, the semantic distance of associated terms in poetry. These new applications are but a few made possible by the advances in technology that can prove that Halliday and Hasan were right to include collocation-type lexical

⁵ <https://www.ebi.ac.uk/>

relations in their taxonomy because these word pairs greatly contribute to lexical cohesion, however there are no grounds for naming them **collocations**.

The second observation regarding the above taxonomy is that although Hoey was one of the linguists who criticized the ‘ragbag’ nature of **collocation**, he himself proposed a rather similarly ragbag category, that of **paraphrase** which includes synonymy, antonymy, superordinates, and the even more obscure subgroup: **link triangle**. Hoey’s (1991) categories will be explored further in Sections 2.3.2 and 2.3.6 of this chapter.

2.2.2 Lexical chains or a lexical net?

According to Hasan (1984), not every word is equally important in a text with regard to its cohesive power. Tokens (i.e., actual words) of a text may or may not form semantic relationships with other words, called **cohesive ties**. If they are not parts of chains, they are called **peripheral tokens**, whereas tokens which are parts of chains are **relevant tokens**, which are central to the text. Centrality is a recurring but ever changing concept in discourse analysis. Mann and Thompson (1988) in their Rhetorical Structure Theory, differentiate between **nuclei** (the units that are most central to the writer’s purposes) and **satellites** (less central supporting or expanding units). They call the produced patterns **schemas**. The hierarchy of important and less important ideas are also described by van Dijk and Kintsch (1983), which they refer to as a system of **macro-** and **microstructures**. Hoey (1991) is similarly concerned with centrality, distinguishing between **central** and **marginal sentences**. A major importance of chains as cohesive ties is that central tokens within chains connect several discourse levels: words connect sentences, sentences connect paragraphs, and the list can be continued to chapter or whole book-length level. The longer the chain, the longer the writer “stays on topic”.

Other influential models devised for analyzing coherence also attempt to recognize chains, be they lexical or phrasal, even if this fact is not mentioned explicitly in the name of the model. Topical Structure Analysis (TSA), for instance, by Lautamatti (1987) examines semantic relationships between sentence topics and overall discourse topics: it looks at the internal topical structure of paragraphs as reflected by the repetition of key words and phrases (see Figure 3). Thus, the aim of the model is to provide insights into the organizational patterns of discourse by observing chains in the text.

ST2		No. of sub-topics
1	babies	1
2	baby	1
3	baby	1
4	he	1
5	animals**	2
6	they	2
7	man's survival	3
8	babies	1
9	this capacity	4
10	babies (S)	1
11	learning to speak	4
No. of different sub-topics: 4		

Figure 3. Topical Structure Analysis indicating semantic links between sentences (Lautamatti, 1987, p. 102)

By focusing on lexical chains in discourse, several conclusions can be drawn regarding coherence requirements. For example, comparative studies between languages show that English paragraphs tend to have a higher use of internal coherence than Spanish paragraphs (Simpson, 2000), making lexical chains in English texts a relevant field for study.

Chains not only connect discourse structures; they also divide parts of the text. Identifying where chains begin and where chains end is used for both text summarization and text segmentation in corpus linguistics. Barzilay and Elhadad (1999) created summaries by extracting strong lexical chains from news articles: chain strength was scored according to length (“the number of occurrences of members of the chain”, p. 116)” and the homogeneity index (“the number of distinct occurrences divided by the length”, p. 116). As far as segmentation is concerned, close correspondence between the starting and ending points of lexical chains, and paragraph boundaries (structural unit boundaries) was found, for example, by Morris and Hirst (1991), and Berber Sardinha (2000).

Hasan (1984) describes two types of chains. **Identity chains** are text-bound and held together by the semantic bond of co-referentiality. In other words, identity chains are made up of words which have the same referent (*John – he – the boy*). **Similarity chains** are not-text bound, and are based on co-classification or coextension. To give an example, if in a text something was mentioned as *easy*, and later in the text, something else is also mentioned as *easy*, a similarity chain will be formed between these two elements, i.e., between the two mentions of *easy*. The concept of similarity chain is close to the psycholinguistic concept of word associations and is also a basic tenet of intertextuality.

Hoey (1991) maintains that the presence of chains does not guarantee coherence: it is the interaction of chains that matters in this respect. Therefore, Hasan’s contribution to clarifying the relationship of coherence and cohesion, according to Hoey, is that Hasan abandoned the classificatory approach, and introduced an integrated approach. In other words, the combination of ties within chains is a more important idea in Hasan’s model than observing and classifying the ties without their context.

Besides considering texts as holders of lexical chains, they can also be viewed as containers of a **lexical net** (or network). The lexical net concept, however, is more connected to research on semantic networks in literature, rather than to research in discourse analysis. Such studies, for example, in psycholinguistics describe the mental lexicon (the arrangement of words in one's mind), or more recently, studies utilize Princeton WordNet⁶, an online semantic network database. According to Hoey (1991), who is a major advocate for the lexical net concept, the main difference between chains and a net is that the constituents of chains (i.e., the ties) have directions, pointing either backward or forward, whereas a net is a two-dimensional map of words disregarding directionality. In order to prove scientifically whether the chain or net representation of text is more accurate, more research is necessary concerning the two types of lexical repetition patterning. It is possible, for example, that there are generic differences and, for certain genres or registers, chain patterns would be more suitable, while other genres would call for a net-like organization.

Lexical chains and nets are comparable in certain aspects if they are represented visually. Figure 4 shows a net from Hoey (1991) and a chain as illustrated in Barzilay and Elhadad (1999). At first sight, they look very similar. The second illustration (the chain) is a little unusual in this form because it looks more like a net. However, the spread-out form only serves as an aid to highlight the **interaction** between the items. The division below *information*, for example, indicates that this word is repeated further as *area* and *datum*.

⁶ <https://wordnet.princeton.edu>

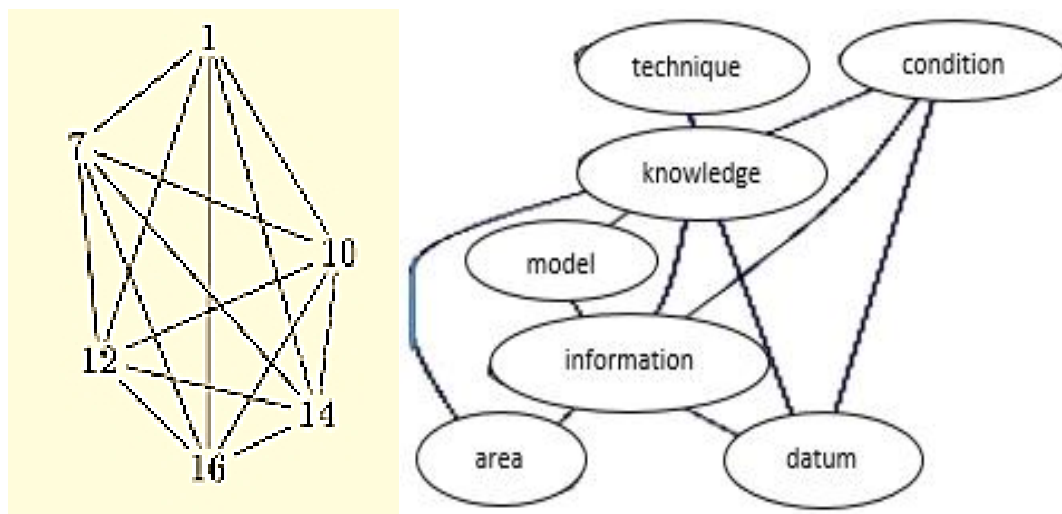


Figure 4. A net and a chain of lexical repetition in two studies (Hoey, 1991, p. 81; and Barzilay & Elhadad, 1999, p. 116)

The main difference between these two representations of semantic relations is **what** they connect: in the first illustration, provided by Hoey, **sentences** are shown which are connected by three, or more than three, lexical repetitions. The numbers indicate which sentences in the text bond, i.e., are significantly connected semantically. The second illustration, on the other hand, shows instead the actual **words** that link sentences. Hoey's net therefore, visualizes a higher layer of discourse.

2.3 Hoey's (1991) Repetition Model

2.3.1 The theoretical background of the model

Hoey (1991) was the first to provide a comprehensive analytical model which reveals the organizing function of lexical repetition in texts. His great contribution to knowledge in discourse analysis and corpus linguistics was that he recognized the information content of the lexical net which was created by lexical repetition.

In his view, the role of grammatical cohesion is less significant than that of lexical cohesion; therefore, he focuses only on words with lexical meaning. In creating his model, he

draws mostly on the theories of Hasan (1984), Winter (1977, 1979) and Phillips (see Hoey, 1991, pp. 14-25). The connection between Hasan's and Hoey's models has already been highlighted in the previous sections (Sections 2.1 and 2.2). The contributions of Winter (1977, 1979) and Phillips are briefly described below.

Hoey (1991) adopted the broad interpretation of repetition from Winter (1977, 1979), even though they disagreed on the function of repetition. According to Winter, the function of repetition is to focus attention on the word which has been replaced. Thus, in the *John – he* word pair the focus is on John. Although Hoey did not refute this explicitly, the main problem with this for him must have been the directionality, more precisely the anaphoric direction of repetition assumed by Winter. On the other hand, Winter's other assumption that replacement (exact repetition or repetition by another item) also needs to be observed at a clause level, was favoured by Hoey. According to Winter, if we repeat an item, the position of new versus old information will also change in the clauses involved. With this assumption, Winter integrates two levels: lexical analysis and clausal analysis.

Hoey translated Winter's main conclusions for his own research in the following way:

1. "If cohesion is to be interpreted correctly, it must be interpreted in the context of the sentences where it occurs.
2. We are more likely to arrive at a satisfactory account of how cohesion works if we concentrate on the way repetition clusters in pairs of sentences.
3. It is the common repeating function of much cohesion that is important, not the classificatory differences between types of cohesion.
4. There is informational value to repetition, in that it provides a framework for interpreting what is changed.

5. Relations between sentences established by repetition need not be adjacent and may be multiple.” (Hoey, 1991, p. 20)

Similarly to Hoey’s critique of Hasan’s (1984) work, he again stresses the importance of interactionality over classification (in conclusion No. 3, above). He maintains that repetition defines a context in which sentences interact not only with neighbouring sentences but on a wider distance as well (in No. 4 and 5). Perhaps this is why Hoey proposed the term organization instead of structure when he described texts in his study (1991).

Besides Winter (1977, 1979) and Hasan (1984), Phillips (1985) influenced Hoey methodologically. Phillips analyzed long texts by computer, which made Hoey broaden his enquiries to book-length texts, and observe long-distance lexical relations. A noteworthy result of Phillips, as cited by Hoey (1991, p. 24), is that academic texts contain many more long-distance clusters of repetition than other types of text, and these consistent mentions have an important organizational function. The fact that Phillips used automated means of research drove Hoey to work on his methodology with automation in mind.

2.3.2 Hoey’s (1991) taxonomy of lexical repetition

As mentioned in the previous section, Hoey himself did not regard the classification of links of primary importance compared to other aspects of his work, namely the role of lexical repetition patterns in organizing discourse. The key concepts of his taxonomy can be grouped into Lexical and Non-lexical repetition. Lexical repetition is categorized, as shown in Table 3, in the following way:

Type of repetition	Example		
I. Lexical repetition	simple	<i>bear-bears</i>	
	complex	<i>drug-drugging</i>	
II. Paraphrase	simple	<i>produce-cause</i>	
	complex	antonymy	<i>hot-cold</i>
		link triangle	<i>writer-author-writing</i>
		the “mediator” missing	<i>writer-(author)-writing</i>
		other	superordinates (<i>biologists-scientists</i>) co-reference (<i>Augustus- the Emperor</i>)
III Non-lexical repetition	substitution links	e.g. <i>personal pronouns, demonstrative pronouns, modifiers, Halliday and Hasan’s substitutes, etc.</i>	

Table 3. Types of repetitions based on Hoey’s (1991) taxonomy

Simple lexical repetition occurs “when a lexical item that has already occurred in a text is repeated with no greater alternation than is entirely explicable in terms of a closed grammatical paradigm” (p. 55). It includes exact repetitions, and repetitions of the same word with inflectional changes.

Complex lexical repetition occurs “either when two lexical items share a lexical morpheme, but are not formally identical, or when they are formally identical, but have different grammatical functions” (p. 55). Hoey’s example for this is *drug* as a noun and *drugging* as in *making sleepy* (-ing form of a verb).

Simple paraphrase occurs “whenever a lexical item may substitute another in context without loss or gain in specificity and with no discernible change in meaning” (p. 62). Such are synonyms, e.g. *hot—cold*.

Complex paraphrase occurs when “two lexical items are definable such that one of the items includes the other, although they share no lexical morpheme”. This category is broken down into three subcategories: 1. **antonymy**, 2/a **link triangle**, 2/b **the “mediator” missing**, 3. Other types of complex paraphrase: **superordinates** and **co-reference** (p. 64). The categories link

triangle and the mediator missing are very difficult to interpret, therefore they will be discussed within Section 2.3.7 among the problematic features of Hoey's taxonomy.

Non-lexical repetitions are substitution links, and as such, grammatical categories: personal pronouns, demonstrative pronouns and modifiers.

Words of a text form links with other words according to these main categories. The links need to be coded according to repetition type, counted and their positions recorded (more details in Section 2.3.4). According to Hoey's description, only content words (words with lexical meaning, e.g., nouns, verbs, adjectives, etc.) can be part of a link. Grammatical items and other semi-grammatical categories, such as connectives, although they play a role in cohesion, are not analyzed within his framework of lexical cohesion. However, substitutions, such as pronouns need to be replaced by the original item, thus resolving the anaphora, i.e. the backward reference created between the pronoun and the missing noun. This theoretical, as well as methodological, problem is analyzed in Section 2.3.8.

2.3.3 Links and bonds creating a lexical net

Hoey (1991) claimed that "lexical items form links when they enter into semantic relationships" (p. 91). These links, however, are only realized between two sentences, not inside a sentence. Therefore, if two words are repeated within one sentence, these will not be analyzed. The reason for this, according to Hoey, is that intra-sentential repetitions do not play a role in structuring discourse, even if they have an important function, e.g., emphasis (*They laughed and laughed and laughed uncontrollably.*). Hoey differentiated his concept of **link** from Hasan's cohesive *tie* in two aspects. Firstly, his categories were greatly different from those of Hasan's. Secondly, he emphasized that links have no directionality.

Hoey's important claim is that certain sentences play a more central role in organizing discourse than others. Sentences sharing three or more links are significant for the organization of discourse because they form **bonds**, a higher level connection. Marginal sentences, with fewer than three links, do not contribute essentially to the topic, therefore if omitted, do not disrupt the flow of the topic (Hoey, 1991).

Bonded sentences lead to nets, which ultimately organize text, in a manner similar to Hasan's (1984) identity and similarity chains. Hoey found that bonded sentences are central to text, as they are the core bearers of information (resembling the concept of macropropositions by van Dijk and Kintsch, 1983, see also Section 3.2.4.2). Hoey's main claim that links created via lexical repetition may form bonds which subsequently create significant sentences, was later reaffirmed by Reynolds (1995) and Teich and Fankhauser (2004, 2005).

Hoey also defines the minimum level of linkage necessary to create a bond, i.e., identifies certain constraints to making sentences significant. He sets three links as the threshold, and turns to Sinclair's (1988) work on word sense disambiguation to support this claim. Word sense disambiguation is a necessary step to identify the right meaning of a word in a sentence. The English language contains many polysemous and homonymous⁷ words, therefore differentiation of meaning is a relevant problem for discourse analysis. Sinclair, who uses corpus linguistic techniques, finds word sense disambiguation one of the most problematic issues. He recommends to look at the collocational pattern of words for sense disambiguation because different senses of words more than likely also have different collocational profiles.

⁷ "Homonymous words exhibit idiosyncratic variation, with essentially unrelated senses, e.g. bank as FINANCIAL INSTITUTION versus as NATURAL OBJECT. In polysemy [...] sense variation is systematic, i.e., appears for whole sets of words. E.g., lamb, chicken, and salmon have ANIMAL and FOOD senses." (more about how to distinguish them by automated means, in Utt & Padó, p. 265) <http://www.aclweb.org/anthology/W11-0128>

Let us relate this to the problem of threshold calculation for bonding. Supposing we have to distinguish between two senses of the word *bank* (sense 1: financial institution; sense 2: part of a river), by observing the preceding and following words, we will find significant lexical differences regarding the two senses. If, in a text which is about money matters, we find a sentence about a river, the collocational profile of this sentence will be so different from the other sentences that the sentence will “stick out”. The reason therefore, why three is the number of minimum links to form a bonded sentence is that if the sentence is central to the topic, it will be linked in at least three places: minimum once as a key word, plus in two other places as collocations to this key word, which collocations probably reappear in the text, comprising further links with other sentences. Even though the above might shed light to Hoey’s decision regarding the threshold, it still does not explain why he chose three, instead of four or five as the limit for bonding.

Hoey (1991) used a news article and a short section of a non-narrative book for his analysis. Based on the results gained from this small data pool, he showed three methods of how abridgements (summaries) could be created of a longer text: (1) by deleting marginal sentences; (2) by collecting central sentences; or (3) by selecting topic opening and topic closing sentences. He admitted that these modes would summarize different aspects of the original text with shifts in meaning. Nevertheless, he emphasized that the patterns of lexical repetition he examined are characteristic of non-narrative prose, as opposed to narration which has a more linear structure.

Although Hoey presented these models as possible means to create summaries, he did not give guidance on how to evaluate the quality of these summaries. He argued, however, that the lexical repetition patterns revealed by his analytical tool can indicate differences in text quality. He put forward two claims, a weak and a strong one:

“The weak claim: each bond marks a pair of sentences that is semantically related in a manner not entirely accounted for in terms of its shared lexis.” (p.125)

“The strong claim: because of the semantic relation referred to in the weak claim, each bond forms an intelligible pair in its context.” (p.126)

What follows from this, firstly, is that bonded sentences hold important information content; and secondly, bonded sentences have a special discourse function in text. Hoey did not claim that the observed lexical repetition patterns are present in every text type, in fact he excluded narrative texts from his analysis maintaining that they are structurally different from the news article he experimented with.

2.3.4 The steps of the analysis

Hoey’s analytical steps consist of three phases: (1) identifying the lexical items which enter into relationship with other items, (2) classifying these relationships according to the taxonomy (i.e. finding the links), and (3) identifying which sentence pairs have three or more than three links (i.e., finding the bonds). The detailed steps of the analysis also appear in Appendix A, illustrated with diagrams.

1. Coding the text according to the taxonomy. Finding links between every sentence pair, including the title, which also counts as one sentence. (1—2, 1—3, 1—n, etc., and in the same way 2—3, 2—4, 2—n).
2. Writing the links into a connectivity matrix where each cell represents a sentence, as Hoey put it: “to trace a sentence’s connections with other sentences in the text” (1991, p. 85). All links should be written into the cells.
3. The information in the matrix should be written into another matrix in a number format.

4. Cells containing three, or more than three links should be highlighted because these are the bonded sentences. In the following only these sentences will be examined.
5. The locations of bonded sentences need to be found in the text, and they should be highlighted.
6. If the purpose of the analysis is to create a summary, either the bonded sentences should be collected, or the marginal sentences should be deleted (same procedure).
The third procedure is to collect all the topic opening and topic closing sentences.
The bonded sentences will give the basis of the summary.

2.3.5 Applications of Hoey's (1991) model

As the above method is laborious, the first question is whether it can be applied to a large corpus or not. In his book on lexical repetition patterns, Hoey analyzed a 5-sentence long article in detail, as well as the first forty sentences of the first chapter of a non-narrative book. He concluded that theoretically it is possible to create summaries of texts of "unlimited length" applying his repetition model, but he did not give instructions on how to do so in practice. Furthermore, the process of *comparing* lengthy texts with their summaries was not examined, either.

2.3.5.1 The model as the basis for automated summarization

To the knowledge of the researcher, the first text-processing computer application based on Hoey's model (Tele-Pattan) was created by Benbrahim and Ahmad in 1994 (de Oliveira, Ahmad, & Gillam, 1996). It represented a computer implementation of two of Hoey's four lexical repetition categories: Simple Repetition and Complex Repetition. The program created five summaries of the same stock exchange news, which were then evaluated by four traders and five university students. The outcome was that 60% of the raters felt that essential information was missing, and

that participants evaluated the summaries differently. As the texts were not available in the research paper, the experiment cannot be replicated. However, it can be argued that the text processing program was limited in use because (1) it incorporated only two of Hoey's categories, and perhaps as a consequence (2) the resulting summaries were rated differently, even though the type of text (stock exchange news) did not allow for a wide lexical or syntactic variety. The result is all the more surprising because, for a non-expert, it would seem relatively easy to summarize such a functionally 'predictable' genre.

As the size of available and searchable corpora increased significantly, British Telecom financed a project, lead by Hoey and Collier, to design a "software suite" for the abridgement of electronic texts (Collier, 1994), by automatically selecting central sentences, i.e., sentences containing the main ideas in text. The program was able to create a matrix of links in seconds, but again, only for the two basic repetition categories: simple and complex repetition. According to Collier (1994), thesaural links were added manually to analyze antonyms, but this step resulted in only a minor improvement in the program. His research plan lists several semantic and structural difficulties in automating central concordance line selection and he concludes that further research is necessary into these areas.

Two programs evolved from the original version: a document similarity tracer (Shares), and an automatic document summarization/abridgement system (Seagull). A demo version of both can be accessed at the Birmingham City University Research and Development Unit for English Studies website⁸. Several other attempts have been made to use computer discourse analysis programs based on Hoey's taxonomy (de Olivera, et al., 1996, Monacelli, 2004), however, no research we know of has utilized the whole of Hoey's framework without alterations.

⁸ <http://rdues.bcu.ac.uk/summariser.shtml>

As Collier (1994) described above, the automated identification of repetition links was attempted using a concordance selector. Due to the extremely laborious nature of data collection, many studies utilize a concordance program (e.g., AntConc⁹, or Concordance 3.3¹⁰) to search discourse data in the area of investigating lexical repetition patterns. As data is textual, a frequency analysis software is helpful in counting how many times certain words appear in the text. It is also possible, using a concordancer, to count how many times certain pairs are repeated. The software is able to show in which sentences the repetitions occur. It cannot evaluate qualitative data, however, without a human observer to process information.

Since its first implementation, automated summarization has been widely used for a variety of purposes (e.g., Barzilay & Elhadad, 1999). These summarization applications, however, use algorithms different from the one Hoey provided. Mani (2001) and Spärk Jones (1999) gave detailed descriptions of the latest developments in this field.

2.3.5.2 The model as the basis for a discourse segmentation tool

Hoey's (1991) claims regarding central sentences, particularly topic initial and topic closing sentences instigated research into text segmentation (i.e., dividing a text into meaningful units). While he aimed to synthesize text by collecting topic opening and topic closing sentences which were revealed by bonds, other researchers wanted to achieve the exact opposite: segmenting discourse by identifying paragraph boundaries.

Hearst (1994) used a computer program to locate segment boundaries by lexical repetition counts, a different method from Hoey's. She compared equally long text parts and tried to find similarities. If the two text parts were very different lexically, they did not constitute a single

⁹ <http://www.laurenceanthony.net/software/antconc>

¹⁰ <http://www.concordancesoftware.co.uk/>

coherent paragraph. This method is called text tiling by Hearst, who computed simple repetition only and used a technical text which is more likely to contain repetition of terminology.

Berber Sardinha (2000) on the other hand, who criticized Hearst for her practice of comparing equally long discourse parts, looked for an alternative method and attempted segmentation using Hoey's analytical framework. He soon found, however, that the lexical net organization pattern is an obstacle to segmentation because the sentences forming the net were all parts of a large group connecting, rather than segmenting the whole of the text. Therefore, Berber Sardinha diverged from Hoey's framework and looked for link sets, more resembling cohesive chains. He calculated the link set medians, which provided meaningful information about where link sets begin and end.

The prevalent problem with these early computational applications was that they only searched for simple repetition because computerized thesauri were not available. This trend changed with the implementation of Princeton WordNet¹¹, an online thesaurus. Its synsets (systematic semantic relations) are now the basis for most lexical analysis on semantic relatedness in corpora.

2.3.5.3 Large-scale lexical repetition analysis with WordNet

Teich and Fankhauser (2004, 2005) connected Hoey's framework with WordNet¹² as a thesaurus in order to observe differences between registers in the Brown Corpus¹³ concerning lexical repetition patterns. As they noted, the flexibility of Hoey's categories facilitated the creation of a **multi-layer** corpus, the building of which was the focus of attention around the turn of the

¹¹ <https://wordnet.princeton.edu/>

¹² <https://wordnet.princeton.edu/>

¹³ The Standard Corpus of Present-Day Edited American English; the manual of the corpus is available from <http://clu.uni.no/icame/brown/bcm.html>

millennium. This kind of corpus is annotated at multiple linguistic levels, such as e.g., the syllable, word, phrase, clause and text levels. Teich and Fankhauser's results shed light on interesting aspects of lexical repetition. It was found for example, that register-specific vocabulary forms stronger cohesive ties than general vocabulary (2005), as the words typical of the specific register form longer chains within text than general vocabulary does. Their qualitative analysis also revealed that the texts in the **learned** and in the **government** registers have longer-stretching lexical chains than texts from **press** or **fiction**. This might be the result of academic texts in English being more linear (c.f. Lautamatti, 1987), or showing higher lexical density (the number of lexical words divided by the number of total words), or due to nominalization (Halliday and Martin, 1993)

2.3.5.4 The model as the basis for validating multiple choice test prompts

A unique implementation of Hoey's lexical repetition analysis was carried out by Jones (2009) who investigated question—answer pairs in an EFL context. He analyzed reading passages and related question prompts in the Pearson Language Test Reading Comprehension Section. He investigated the semantic similarity between the wording of the questions and the wording of the related answers by looking for semantic links between the sentences drawing on Hoey's categories. As it was a manual analysis, he was able to use all the categories in Hoey's (1991) taxonomy. His original assumption was that an answer is more difficult for EFL learners if the semantic links between the prompt and the answer are semantically less related, i.e., they can be found further down in Hoey's taxonomy table. Thus, for instance, a word in the question which is repeated exactly (simple repetition) in the answer, makes the answer fairly easy. If the question and answer pair contains derived repetitions, or superordinate terms, finding the right answer is linguistically more demanding for the student. Jones, in this pilot study was able to lay down the theoretical

foundations for further possible studies to scientifically measure the semantic distance of multiple choice item pairs in the Pearson Reading Comprehension test prompts.

2.3.6 Inconsistencies within Hoey's (1991) model

Several inconsistencies exist within Hoey's (1991) model. Károly (2002) pointed out that it contains three weaknesses: (1) theoretical problems with the taxonomy, such as several obscure category labels, and the unclear definition of the basic unit of analysis, (2) weaknesses of the method of analysis, such as not examining intra-sentential repetition, or the missing theoretical foundation for choosing the number of bonds to be seen as significant connections, (3) research methodological problems, such as making strong claims based on a single text type.

The inconsistencies of the model derive from two areas: Hoey's inductive approach in his data collection and analysis, and his claim that the classification of links is of lesser importance than the patterns gained by the interactivity of links. He analyses data gained from a short text and, based on these results, he draws conclusions implying that the same can apply to longer texts. This inductivity can be observed, for example, when Hoey gives guidance on how to categorize certain problematic words.

It needs to be mentioned that approaches which start by analyzing existing data without a previous hypothesis are the standard procedure for corpus linguistics, therefore it is not a unique characteristic of Hoey's research. The problem is rather that he uses his data as illustration, and makes decisions on a case-to-case basis, which makes the model difficult to apply.

The first problem is that his categories are difficult to identify, and the second is that the categorization is unjustified. He did not use traditional grammatical categories and subcategories, perhaps because he wanted to create a new multi-layered analytical tool, and found the existing

categories too restrictive in this sense. It might be suggested that these invented categories are rather confusing than helpful during the coding process.

A further problem is that the categorization is not justified. The category **Simple Repetition** seems to be the most straightforward, nevertheless even this is problematic, especially when observed with automation in mind or from a theoretical point of view. Regarding the latter issue, Hoey contemplates that when we repeat the same word in the same form, is it still the same word, or has its meaning changed by the fact that it was repeated.

Another interesting problem, which Hoey recognizes is that words have different functions within a text, even though he does not explicitly state this: he calls them accidental repetition in his book *On the surface of text* (1995, p. 108). He suggests observing the functions of the word *reason* as an example.

*No faculty of the mind is more worthy of development than the **reason**. It alone renders the other faculties worth having. The **reason** for this is simple.*

According to the author's explanation, this word appears with two different meanings in the text, therefore the two examples within this paragraph cannot be considered a repetition link, since the second mention of *reason* has a referential function in this context similar to metadiscoursal¹⁴ phrases in text, such as *in the following, let us see*. Hoey does not offer a list of such problematic words, or a rule on the basis of which these umbrella terms, or special discourse function words should be excluded from the analysis. Such nouns were later compiled by Hyland (1998) which he collected by corpus analytical means from eighty research articles from eight academic disciplines. (See in Appendix B.)

¹⁴ Metadiscourse is a common concept in academic discourse research, the narrow definition being text about the text (Swales & Feak, 2010).

During the analysis of a large corpus, the frequency of the errors caused by misidentified links as the above might be scarce compared to the frequency of the correctly identified links. However, if we analyze a corpus of short essays by computer, the lack of unified instruction could be a large problem. Given the fact that academic discourse is notorious for using metadiscoursal nouns (issue, problem, reason), the error rate of the analysis can be higher.

A serious, but different kind of difficulty arises when we attempt to analyze Paraphrases. **Complex paraphrase** by definition can be a one word unit, a phrase, a clause, a sentence or even “a stretch of text” (Hoey, 1995, p.112)! It is difficult to prepare coders for such intricacies regarding the unit of analysis. Even though the complex nature of any text cannot be denied, it is still questionable whether all these features are necessary to be analyzed within the same single framework in order to yield data on textual cohesiveness.

The above mentioned problems with the taxonomy are but a few of the inconsistencies regarding the framework and the over-flexible nature of the units of analysis. Due to these unresolved issues, identification of units and annotation of text is prone to low inter-coder agreement, thus hindering reliability. It seems, Hoey’s ground-breaking idea to collate several discourse layers still needs to be further refined, and more experiments are needed for its implementation.

2.3.7 The link triangle and the mediator missing

Hoey introduces two categories which are unique in the reviewed literature: these are the link triangle (shown in Figure 5) and the mediator missing categories,. According to Hoey’s (1991) explanation, if two words of a text are connected in a certain link and form a pair, this will cause a putative link between the two items otherwise previously not connected. An alternative version

of this concept is the mediator missing, which exists when one of the elements does not appear directly in the text, instead it is referred to by a pronoun.

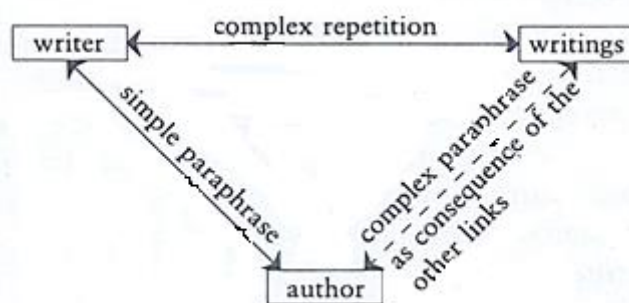


Figure 5. The link triangle (Hoey, 1991, p. 65)

Hoey rightly maintains that semantic triangles exist, however, the problem is that these word associations can take the form of any other shapes as well, for example, a square (by the fourth repeated item), an octagon (by the eighth mention), etc. Therefore, even if Hoey reveals another important feature of discourse by introducing the triangle concept, there is no theoretical basis for calling such formations triangle and insisting on placing them in his link taxonomy.

Besides theoretical considerations, the concept of link triangle is also problematic from a methodological point of view. Hoey's category inadvertently confuses his own taxonomy by looking for connections between more than two elements at the same time. While data on frequencies and locations of inter-sentential relationships between lexical units can be observed and analyzed relatively easily, triangle-type relationships would be more difficult to detect and record. Triangle frequencies could also prove to be impossible to interpret alongside the other types of data. For instance, supposing we find three links between the first and the 20th sentence, according to Hoey, we can claim that these two sentences are bonded. In other words, there is a strong semantic and structural relationship between them with a distance of over 20 sentences. However, it is not described what procedure should be followed if there is another word in sentence

21, which may be a candidate for a link triangle: where should this be recorded in the connectivity matrix and how this would influence the overall number and length of bonds within the text. This is a threat for both manual and computerized analysis, even considering the latest advancements in technology. As far as the mediator missing category is concerned, such links should be manually coded (and possibly inserted), on a case-by-case basis. This would seriously slow down the analysis, if not make it impossible.

Hoey further elaborates on the triangle concept in his other influential book, “*On the surface of discourse*” (1983), where he observes triangles one level higher: at the discourse level, and he describes their discourse functions such as the problem – solution or the general – particular patterns. See Figure 6 for the general – particular relationship.

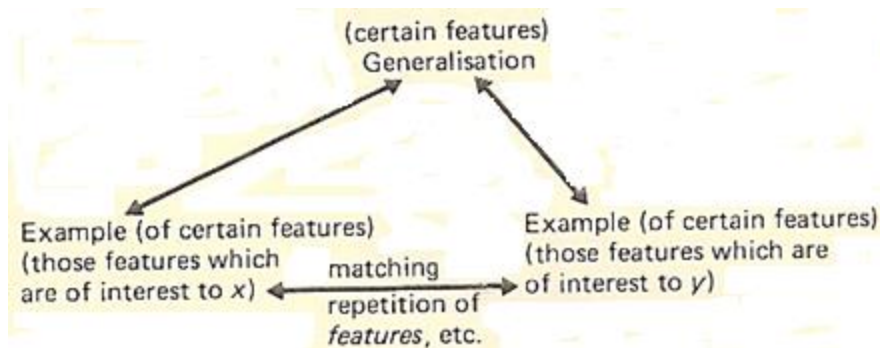


Figure 6. The General—Particular relationship in text (Hoey, 1995, p.135).

2.3.8 The questions of anaphora resolution

Another problematic area is whether to resolve pronominal anaphoric reference; one of the major contributors to cohesion, affecting at least every second sentence in any English text. According to Hoey’s (1991) methodology, if a word is repeated by a substitution (i.e., replaced by a pronoun in the following sentence), the original noun (who or what the pronoun refers to) should be inserted in the place of the pronoun in order to recreate the original lexical link. Thus, in the

sentence pair *John wants to go to the theatre. He is a great fan of contemporary drama, he should* be replaced by *John* for the purpose of the analysis.

If we want to create summaries using Hoey's model, reestablishing links is a logical and necessary step to improve cohesion. However, this treatment cannot be applied if we want to connect Hoey's model with text quality research. If we followed Hoey's advice and replaced the pronouns with their preceding referents in their original form, the number of simple repetition would increase considerably, distorting perceptions of discourse quality. (Not to mention dubious cases, when it is difficult to decide who or what was meant by the author as referent.)

2.3.9 Prescriptiveness or descriptiveness of the model

Tyler (1995) criticized Hoey on the grounds that quantitative investigation of lexical repetition alone cannot capture the difference between well and badly formed texts: qualitative analysis is necessary to explore how repetition is used. Connor (1984) went further by suggesting that it is possible for a text to be lacking in lexical cohesive links, still be better organized than another text containing more lexical repetition links, but at the same time not having a well-formed argument. Therefore, if we accept the assumption that the quantity of lexical repetition is not a major dimension in discourse quality, the next obvious question is: if we examine the cohesion **patterns** formed by lexical repetition links in texts, will we be able to judge discourse quality or not?

Tyler's (1992, 1994) empirical study indicated that repetition in itself was not sufficient to cause cohesion because the perceived quality difference of native and non-native speakers' language production is influenced by **what** and **how** is repeated. This issue is not addressed in Hoey's studies. Nevertheless, Tyler did not contradict Hoey's main claim regarding the function of bonds as text-building devices. Reynolds (1995, 2001) also found differences between the usage

of lexical repetition among native and non-native speakers. He applied Hoey’s coding system and methodology on students’ expository essays. His findings revealed that EFL writers did not use the lexical repetition devices optimally: bonded sentences were not used to form the bases of developing the argument structure. Reynolds’ conclusion was that “the content of what is being repeated is as important as the quantity” (1995, p. 185). Thus, it can be concluded that Hoey’s model has great potential for studying lexical repetition analysis, particularly if it is completed with a content-based approach.

2.4 Károly’s (2002) Repetition Model

Károly (2002) applied Hoey’s model to explore the text-organizing role of lexical repetition. She revised Hoey’s (1991) taxonomy putting it into a wider perspective, using the model in a Hungarian EFL academic context.

2.4.1 Károly’s (2002) taxonomy of lexical repetition

Categories of lexical relations		Examples
I. Same unit repetition		
1. repetition	simple	<i>writers – writers</i>
	derived	<i>writers – write</i>
II. Different unit repetition		
2. synonymy	simple	<i>to exercise – (after) working out</i>
	derived	<i>built – construction</i>
3. opposites	simple	<i>small – major</i>
	derived	<i>hatred – like</i>
4. hyponymy		<i>languages – English</i>
5. meronymy		<i>hands – fingers</i>
III. Text-bound relations		
6. instantial relations		<i>manager – O’Leary</i>

Table 4. Types of lexical relations in Károly’s taxonomy with examples (examples based on Károly, 2002, p. 104, and these two corpora)

Károly introduced the term *lexical unit* as the basic unit of her analysis. This is a unit “whose meaning cannot be compositionally derived from the meaning of its constituent elements”

(Károly, 2002, p. 97), i.e., the individual words placed one after the other mean something different than each word means standing alone. A lexical unit can be a one-word unit, an idiom or a phrasal compound (words expressing a unique concept, e.g., *non-native English speaking teachers, non-NEST-s*). She also proposed a new taxonomy of the lexical repetition types, as indicated in Table 4.

Table 4 shows that Károly (2002) uses more traditional grammatical terms than Hoey (1991), and her units of analysis are linguistic constituents which can be more easily identified than those of Hoey's. The instantial relations category introduces a semantic category which is temporarily bound by context, and resembles Hasan's (1994) instantial lexical cohesion category, which was originally broken down to equivalence, naming and semblance. Károly also argues for the differentiation between inflection and derivation within the category same unit repetition because inflectional differences are only syntactical variants, therefore represent closer semantic connections than derivation which changes the meaning of the word, irrespective of whether it happens with or without word class change. Hoey's original idea that a unit is as small as a word but can be stretched as far as a passage if these two passages are paraphrases of each other (e.g., 1995, p. 110) is lost by Károly's more rigorous categorization. As a consequence, the semantic flexibility Hoey's analysis offered is sacrificed. On the other hand, the clarity of the categories and the traditional grammatical terminology enhances the reliability of the coding, inasmuch as coders do not have to make many ad-hoc decisions.

2.4.2 Károly's (2002) method of analysis

Károly also found several weaknesses in Hoey's methodology. Such were not examining intra-sentential repetition, or the missing theoretical foundation for choosing the number of bonds to be seen as significant connections. As far as Hoey's research methodology is concerned,

Károly's criticism was that Hoey made strong claims about the role of lexical patterns in discourse based on a single text type.

Károly (2002) not only revised the categories but also introduced a number of new analytical steps related to the combination of links and bonds to extend the research capacity of the analytical tool. Her method of analysis focused on new aspects of bonds, such as their position, length, and strength between sentences with special discourse function (SDF), such as the title, the thesis statement, the topic sentences, and the concluding sentences.

For instance, the **length of bonds** category indicates how far apart bonded sentences are located from each other, and the distinction between **adjacent bonds** and **non-adjacent bonds** indicates which sentences form mutual relationships. The **strength of bonds** was calculated to reveal how many links connect sentences in the given text. Károly's new quantitative analytical measures are shown in Appendix C.

2.4.3 Károly's empirical investigation

Károly (2002) investigated the organization of ten high-rated and ten low-rated argumentative EFL essays written by English BA majors at a Hungarian university. Her main hypothesis was that her revised tool is able to differentiate between high-rated and low-rated essays, based on the role lexical repetition plays in structuring texts.

Károly used a number of variables, which she later reduced to five. These were: the frequency of derived repetition, the relative use of bonds at paragraph boundary, the density of bonds, the frequency of adjacent bonds, and the amount of bonds between the title and the topic sentences of the essay. They proved to be capable of predicting raters' quality judgements of essays with 95% certainty. The variables with the most predictive power are shown in Table 5.

Károly's (2002) research results showed that her theory-driven "objective" analytical tool not only offered a descriptive function, but with her analytical measures, it was capable of predicting the "intuitive" assessment of teachers evaluating the essays with regard to the content and structure of EFL academic argumentative essays. Her main hypothesis was that her revised tool is able to differentiate between high-rated and low-rated essays.

Variables with most predictive power by Károly (2002)	Functions of the variables
frequency of derived repetition	quality of repetitions
density of bonds	quantity of repetitions
frequency of adjacent bonds	characteristic length of bonds
title's bonds with topic sentences	special discourse function sentence behavior
relative use of bonds at paragraph boundary	potential topic opening and closing sentences

Table 5. Variables with the strongest predictive power in Károly's (2002) lexical repetition analysis research

The results of Károly's analysis proved that the texts, which had previously been rated high or low by experienced university instructors, differed significantly in both **repetition amount and types**. Her results indicated that high-rated essays contained significantly more repetition links, including more same unit repetition, and within this, derived repetition; as well as more simple opposites and instantial relations.

An interesting finding was that the analytical tool could *not* discriminate between high-rated and low-rated essays based on the **combination of bonds**. The four aspects observed here were the quantity of bonds, the amount of adjacent and non-adjacent bonds, the length of bonds and the strength of bonds. Therefore as a next step, a content-based approach was used to investigate the sentences with special discourse function (SDF). Such sentences were for instance, the title, the thesis statement and the topic sentences.

The novel outcome of the analysis was that the **amount of bonds connecting SDF sentences** was significantly higher in high-rated essays: particularly the title and the topic

sentences and the title and the rest of the sentences. Károly's results revealed thus far hidden dimensions of lexical repetition, such as the first result, which means that even high-rated essays contained many repetition links, although it is common teachers' practice to advise against using repetitions in texts. Post-tests conducted with another group of teachers confirmed these findings, thus indicating that the analytical measures devised are reliable and the results may be generalized for a wider sample.

Another interesting aspect of the teachers' perceptions of essay quality was also uncovered by Károly's (2002) analysis: one essay which was predicted to be low-rated by the model due to the lack of appropriate bonding, was still scored high by the teachers. A content-based analysis revealed that this particular essay utilized a number of rhetorical devices, such as illustrative examples, parallelisms, and rhetorical questions for supporting the argument. This, as well as irony, such as the following example, indicates that the model cannot capture certain features perceived as significant in overall discourse quality.

What a shame that there are such inhuman beings living around us as journalists, you think when reading through the passage.

2.4.4 A corpus-based investigation using Károly's (2002) taxonomy

To my knowledge, no research has been carried out by manual coding using Károly's (2002) model. A recent computer-aided empirical investigation based on Károly's (2002) taxonomy aimed to compare shifts in lexical cohesion patterns between translated and authentic Hungarian texts (Seidl-Péché, 2011). Seidl-Péché found that authentic Hungarian and translated Hungarian texts differ in lexical cohesion patterns. Her quantitative analysis was facilitated by language technology modules provided by Orosz and Laki (Laki, 2011; Novák, Orosz & Indig, 2011), whose linguistic parser (analyzer) program helped to automate the analysis. The Hungarian

WordNet Program¹⁵ (Prószéky & Miháltz, 2008) was used to explore semantic links between sentences.

Although Seidl-Péché's (2011) study was the first to utilize Károly's lexical repetition analysis model for a multilingual corpus-based investigation, it cannot be considered as a model for our research for several reasons. Firstly, parts of her methodological decisions were determined by her special research focus, namely studying the quality of translation. Secondly, the scope of Seidl-Péché's research limited her investigation to nouns. (Interestingly, in the results section, however, the screenshots revealed that the software also analyzed pronouns, e.g., *azt* [*that* in object form], *arra* [*onto that*], p. 135. It is possible that anaphoric references (i.e., repetitions by pronoun) were also included in the sum of repetitions. If not, it would be important to know for future research such as this one, how they were discarded.)

Thirdly, Seidl-Péché (2011) did not provide enough details on how the application analyzed the texts exactly. She did not explain, for example, how lexical sense disambiguation occurred precisely. An English example would be this: How did the application decide which the synonym sets for *bank* were? As lexical repetition analysis (LRA) sets out to identify semantic relations, and polysemy and homonymy are frequent in English, the key methodological question is whether the software offered this word for the researcher to manually choose the right meaning in the given context, or the application selected the right meaning on its own, entirely automatically. In the latter case, it is of key importance to examine how the program decided which meaning was relevant.

¹⁵ <http://www.inf.u-szeged.hu/rgai/HuWN>

To explore this feature in the English version of WordNet, on which the HunWordNet was based, I experimented with the word *bank* to find out which meaning is considered first: the most frequent, the most likely¹⁶ or whether some other factors are considered? The result was that bank as sloping land was offered before bank as financial institution (as shown in Figure 7, which is a proof that the WordNet application was trained on a general corpus (Brown Corpus) as a database to calculate frequencies, and not on specialized texts, such as texts from the business domain. A more detailed description would have been helpful so that further research can replicate the treatment of how synonyms or antonyms were coded using the HunWordNet in Seidl-Péché's (2011) research.

Noun

- S: (n) **bank** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
- S: (n) **depository financial institution, bank, banking concern, banking company** (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
- S: (n) **bank** (a long ridge or pile) *"a huge bank of earth"*
- S: (n) **bank** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- S: (n) **bank** (a supply or stock held in reserve for future use (especially in emergencies))
- S: (n) **bank** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- S: (n) **bank, cant, camber** (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- S: (n) **savings bank, coin bank, money box, bank** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- S: (n) **bank, bank building** (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*
- S: (n) **bank** (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning)) *"the plane went into a steep bank"*

Verb

- S: (v) **bank** (tip laterally) *"the pilot had to bank the aircraft"*
- S: (v) **bank** (enclose with a bank) *"bank roads"*
- S: (v) **bank** (do business with a bank or keep an account at a bank) *"Where do you bank in this town?"*

Figure 7. Synonyms offered for the word *bank* in WordNet

¹⁶ I looked up several words in the WordNet dictionary related to the meaning of *bank* – as institution to find out whether the software ‘remembers’ the previous requests when I asked it to define *bank*. It did not remember. (This was only an unorthodox trial-error test to explore this feature, it is not based on the literature.)

The above detailed two methodological questions were already addressed in Sections 2.3.3 and 2.3.8 when Hoey's (1991) anaphora resolution and link threshold decisions were discussed as theoretical decisions. Now both reappeared as research application issues. Nevertheless, Seidl-Péché's (2011) research indicates that it is feasible to automate Károly's (2002) framework.

Another issue to consider is that Seidl-Péché limited the scope of her research to nouns due to the fact that the HunWordNet contains noun synsets. It seems appropriate to do more research into texts which contain more adjectives or verbs than usual, to explore their text structuring significance. One genre where adjectives and adverbs are also frequently compared with their opposites is the compare/contrast essay. Figure 8 shows part of such an essay, presented here as an example of how much lexical cohesion would be lost without the analysis of adjectives and adverbs. (Note: In reality, many more adjectival/adverbial repetition links start from these two paragraphs than indicated in Figure 8. They are not visible now because only link pairs between the second and third paragraphs are illustrated.)

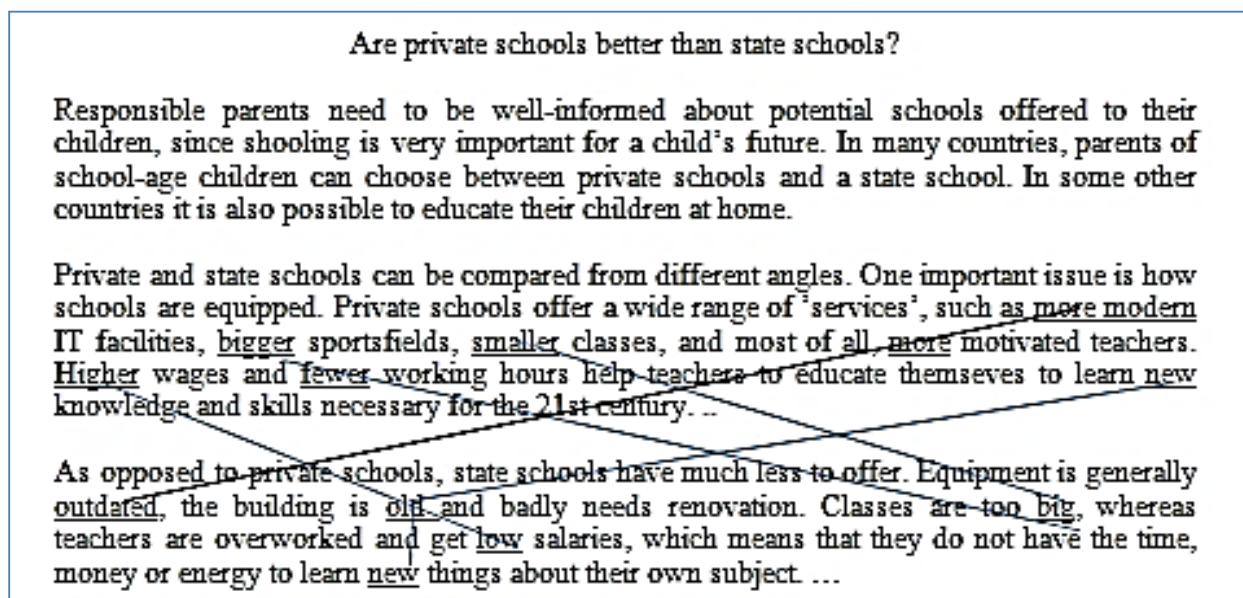


Figure 8. Three paragraphs of a sample compare/contrast essay indicating some of the lexical repetition links (adjectives/adverbs). Text: *Oxford Advanced Learners' Dictionary* 8th ed. (OUP, 2008)

2.5 Summary

The aim of this chapter was to provide a theoretical background to the study of cohesion and coherence in discourse. Whereas there seems to be more of a consensus as far as cohesion, an overt textual feature is concerned, coherence, which involves more hidden, cognitive factors, is looked at in various contradictory ways. Within the concept of cohesion, Hoey's (1991) comprehensive lexical repetition analysis model was discussed, revealing that the patterns lexical repetition creates play a key role in organizing discourse, thus their investigation can be used for various purposes (for instance, for text summarization). Károly's (2002) extended and revised version of Hoey's original model was capable of adding a discourse quality judgement dimension to the framework. The current investigation follows these lines and extends the use of the model to two other academic genres and further designs a method which will enable the model to be applied to larger corpora, too.

3 Methodological background: the academic writing context

3.0 Overview

The academic context from where the corpora of this study originate plays a significant role in this research for several reasons. The first reason is that this research project draws on theories and practices from discourse analysis and corpus linguistics, both of which areas deal with “real”, authentic texts created within an existing social context. In this study, both the summary and the compare/contrast essay corpora consist of real assignments at a university: the former task is written in a disciplinary course (English for Tourism), and the latter is an Academic Writing course assignment. This has an undeniable effect on the texts as products.

The second reason why it is important to observe the academic context is that the students’ writings in the corpora were all first evaluated by their course tutors. The tutors’ judgement is consequential for this study, because their quality judgement on the texts is collated with the predictive power of the lexical repetition analysis tool tested here on two genres. Therefore, it is important to find out how clear tutors’ perceptions are about the three major concepts (coherence, cohesion and lexical repetition) discussed in the previous chapter. It is also interesting to find out how clear these concepts are in a relatively new field: automated essay assessment. Thus, this chapter is the ‘melting pot’ of viewpoints regarding these three concepts examining discourse analysis theory, everyday assessment practicalities and language technology.

This interdisciplinarity requires us to draw on the relevant literature to describe what **quality of writing** really means in a university context, with particular focus on the research into EFL (English as a foreign language) student writing. In order to better understand the notion of quality in academic discourse, three major areas need to be discussed: what kind of writing tasks students are assigned; which are the variables (contextual and cognitive factors) that influence

their final products; and how teachers make their quality judgements about these products. These two latter areas, as they have instigated a large body of research in their own right, are only examined from a narrower, EFL pedagogical angle.

First, academic writing tasks will be described which have been found to be the most prominent in university contexts generally. This is followed by enumerating the task variables of academic writing, following through from task setting to assessment. Next, the two major integrative tasks in the focus of this research are introduced: summary writing and compare/contrast essay writing. The last part of the section also deals with the theoretical and empirical aspects of evaluating writing assignments. Given that student texts are assessed both ways, manual and automated assessment practices will also be discussed, with particular focus on how teachers and computers ‘judge’ particular aspects of cohesion and coherence.

3.1 The nature of academic discourse

3.1.1 General features of English academic discourse

In this study the research focus is on written discourse, more precisely on written academic discourse¹⁷. Discourse as a product appears in the form of particular genres. Genre is defined by Swales as

“a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre. This rationale shapes the

¹⁷ Discourse is “a unit of language larger than a sentence and which is firmly rooted in a specific context” (Halliday & Hasan, 1990, p. 41). For the purposes of this study it will be regarded as a similar concept as text, which, according to Halliday and Hasan, comprises “any passage, spoken or written, of whatever length, that does form a unified whole” (1976, p. 1).

schematic structure of the discourse and influences and constrains choice of content and style” (Swales, 1990, p. 58).

Typical written academic genres are, for instance, the journal article or the research proposal, whereas spoken genres are the lecture or the student presentation. The communicative events (i.e. the kinds of written and oral products) of the discourse communities which Swales describes above have evolved simultaneously with the emergence of the scientific communities, thus the same members comprise the scientific as well as the discourse communities. Therefore typical genres,¹⁸ text types, and structural conventions might differ across disciplines.

Besides typical genres and text types, another discourse feature of academic writing is the academic register, as opposed to, say, legal or media registers. The notion of register is defined by Biber et al. (1998) as a “cover term for varieties defined by their situational characteristics” considering the “purpose, topic, setting, interactiveness, mode, etc.” of the situation (p. 135). The academic English register, particularly written discourse, is objective, complex, and formal. It is also more structured both in its spoken and written form than general English

Objectivity in academic discourse is achieved by using an impersonal style, for example, by avoiding first person singular pronouns, as well as using referencing when introducing other authors’ ideas. Higher lexical density, i.e., a greater use of content words (verbs and nouns) than structure words (words with grammatical function), is also pervasive in this register. Particular grammatical forms, such as passive structures and nominalization make academic texts more static and more abstract. Nominalization is characterized by Francis (1990) as “a synoptic interpretation of reality: it freezes the processes and makes them static so that they can be talked about and evaluated” (p.54).

¹⁸ Biber (1988) differentiates between genre and text type. The former is categorized by external criteria, the latter are texts similar in linguistic pattern, and can take the form of various genres.

Biber and Gray (2010) discuss nominalization as a strategy to make a text more condensed. They argue that academic discourse is wrongly characterised as elaborated; on the contrary, it is compressed and implicit due to nominalization and passivisation. They illustrate the gradual shift towards reduced explicitness with the following three phrases (p.11):

(a) someone manages hazardous waste

(b) hazardous waste is managed

(c) hazardous waste management.

In example *(b)* the agent is omitted, in example *(c)* “it is not even explicit that an activity is occurring” (p. 11). It is also obvious that while the first two examples can stand on their own as complete sentences, example *(c)*, which is the most likely in academic discourse, will need some further syntactic elements to build up a whole sentence, which, as a consequence, will result in more information content in the particular sentence. The relevance of this observation for our research is that due to this compression, in academic discourse a high number of sentence elements will enter into links with other sentences in the text.

Besides nominalization, the lack of elaboration in academic writing is caused by the relative lack of clausal modifiers. Instead, the above mentioned phrasal modification is employed in cases where relative clauses would provide extra information in order to facilitate comprehension. The following Table 6 shows how condensed phrasal modification can be, with some examples from the same Biber and Grey (2010) corpus-based study (p. 9).

Phrasal modification in a research article	Revised expression made explicit by clausal modifiers
the participant perspective	the perspective that considers the participant’s point of view
a systems, theoretical orientation	an orientation which is theoretical and which focuses on the analysis of systems

Table 6. The difference in explicitness caused by phrasal vs. clausal modification (Biber & Gray, 2010)

Apart from difficulties caused by conventions and syntax, language learners also face problems created by academic vocabulary. Coxhead (2000) set out to collect the word families which appear with high frequency in English-language academic texts. The Academic Word List (AWL) contains 570 word families commonly found in academic journals, textbooks, lab manuals, and course notes, thus comprising the “general academic English”. It does not include terminology characteristic of specific disciplines. Academic words cover ten percent of an average academic text (Reid, 2015), yet their abstractness is an obstacle for comprehension.

3.1.2 The types of writing tasks required at university

Several studies have assessed the typical writing tasks required from students at universities (Bridgeman & Carlson, 1983; Hale et al., 1996; Horowitz, 1986; Huang, 2010; Leki & Carson, 2012). Some of these studies identify genres common across universities. According to data collected by Horowitz (1986) from 29 mainly undergraduate courses across 17 departments in a US university context, the most common genres students have to write are **summary** of/reaction to a reading, **annotated bibliography**, **report** on a specified participatory experience, **case study**, and **research project**. He also describes connection of theory and data, and synthesis of multiple sources as tasks also frequently required.

A decade later, a survey conducted across various discipline areas from Accounting to Visual Arts in Australian universities (Moore & Morton, 1999) resulted in almost similar findings. The following genres were found the most prevalent: **essay** (most common with 60% of all tasks), **review**, **literature review**, **experimental report**, **case study report**, **research report**, **research proposal**, and **summary**.

Other studies do not use the term genre when they refer to students’ tasks, such as description, summarization, explanation, etc. The terminology is varied in this respect, the terms:

tasks, rhetorical functions, abilities or skills are all used when referred to these smaller assignments. In a large-scale survey (Bridgeman & Carlson, 1983) covering 190 university departments in Canada and the USA, for example, when teachers were asked which tasks they perceived as the most typical, both at undergraduate and postgraduate levels, the two most common text types mentioned were **description and interpretation** of non-verbal input and **comparison and contrast** plus taking a position. In the Moore and Morton (1999) survey these were **evaluation** (67%), **description** (49%), **summarization** (35%), **comparison** (35%), and **explanation** 28(%)

Rosenfeld, Courtney, and Fowles (2004) identified the specific tasks required both at undergraduate and graduate levels. Their large-scale survey included more than 700 faculty members from thirty institutions in the United States. The most important subtasks the students were required to perform across disciplines are presented in a combined table in Appendix D.

These findings suggest that requirements gradually grow from undergraduate to doctoral levels. Johns and Swales (2002) also found that there is an upward progression in terms of assignment length, complexity of resources utilized, and sophistication expected from students. The table shows that across the three levels, the three most important requirements are the same: proper crediting of sources, writing coherently and writing in standard written English.

As the focus of this dissertation is coherence and cohesion, it is important to note here that the requirements regarding coherence and cohesion are considered very important for all three levels. Another observation is that summarization and comparison and contrast, the two types of discourse this dissertation observes, appear at Master's and Doctoral levels as core requirements. Even though these studies were based on Anglo-Saxon higher education practice, the tasks mentioned above can be considered typical also in the Hungarian academic context.

3.1.3 Disciplinary differences in academic discourse

In order to be able to understand teachers' requirements regarding students' written tasks, specific, disciplinary differences also need to be examined. In line with this, more recent studies set out to map the different writing tasks across faculties. For instance, in the social sciences, humanities and arts domains the **review**, **proposal**, **case study** and **summary** genres were found to be typical, whereas only **library research papers** and **project reports** were commonly assigned across all the disciplines (Cooper & Bikowski, 2007). The same genres also differ in many aspects depending on the subject domain: a report or case study, for instance, written for the biology department might differ considerably from the same genre submitted to the sociology department. This difference, according to Carter (2007) lies in the fact that various fields of studies have their own "ways of doing and ways of knowing", which, as a consequence, formed their "ways of writing" practices (p. 393). New disciplines form their own standards. As Haswell (2013) observes "[e]ach new field of study seems to function out of a different epistemology requiring a different set of writing skills – unfamiliar composing processes, novel genres and tasks, shifting standards and expectations" (p. 416).

Carter (2007) argues that four main types of writing can be distinguished across academic fields: **problem-solving assignments**, **research from sources**, **empirical inquiries**, and **performance assignments**. Problem-solving tasks are typical in the business and engineering domains, where students need to write business plans, project reports or project proposals. Research from sources is dominant in the humanities, a typical task being literary criticism. Empirical inquiry relies on data collected by the students themselves and characteristic in science disciplines, an example of which could be the lab report. In performance assignments the artistic value of the text is as important as the content. Performance text types are creative writing tasks

or poetry, typical in arts classes (examples from Olson, 2013). Carter (2007) assumes that identifying these four types of “metagenres”, i.e., higher categories than individual genres, is more informative than analyzing texts according to the conventions of the genre alone. This “overarching allegiance” in disciplinary communication was also observed by Widdowson (1979), who argued that “scientific exposition is structured according to certain patterns of rhetorical organization which [...] imposes conformity on members of the scientific community no matter what language they happen to use” (p. 61).

3.1.4 Implications for language pedagogy

Several large-scale surveys mentioned in Section 3.1.1 had an L2 (second language) test development focus. For instance, the findings of both the Bridgeman and Carlson (1983) and the (Moore, Morton, & Price, 2011; Moore & Morton, 1999) surveys were later incorporated into internationally recognized language proficiency exams: the TOEFL Computer-Aided Written Exam and the IELTS Academic Module Writing paper, respectively. Additional written genres were introduced with more diverse rhetorical functions, thus connecting examination and university requirements and as a consequence, improving exam task authenticity.

The majority of surveys collected data on academic writing task types (see also Grabe, 2003; Johns, 1980, 1981) in order to align university EAP (English for Academic Purposes) curricula to the requirements of disciplinary courses¹⁹. Some of their pedagogical implications in the teaching of academic English suggested that EAP courses should prepare students for their content courses by providing explicit instruction, i.e., raising genre-awareness, explaining the conventions relevant in their field, modelling rhetorical functions by thinking aloud and simulating real-life test situations. Horowitz (1986) suggests that EAP teachers should be aware of

¹⁹ Disciplinary courses and content courses are synonymous terms.

departmental requirements across the university in order to understand more clearly such contextual features as audience, topic, task definition, or grading. An EAP teacher's task is to raise the awareness of students to the general and specific academic writing conventions, thus such courses should begin with the learner and the situation, as opposed to general English classes where language acquisition is in the focus (Hamp-Lyons, 2001; Jordan, 1997).

Being familiar with the required formal and content criteria of genres is essential for students to be able to complete their assignments. Some of these requirements are easy to learn, such as collecting, organizing and presenting data in a controlled manner (Horowitz, 1986), or building arguments on evidence. For instance, students learn very early in their first semester that using sources is an indispensable part of their essays, and information about authors need to be referenced according to the required documentation style (APA, MLA, Harvard, etc.). However, learning how to write efficiently for academic purposes is a gradual and slow process, because numerous aspects and requirements are implicit "rules of the 'game'" (Casanave, 2002), thus "pose a 'clueless' challenge" to academic writers (p. 19).

3.1.5 Independent vs. integrative writing tasks

As it is clear from the previous lists of genres, most writing tasks at university presuppose reading source texts before writing. Moore and Morton (1999), for instance, found that almost all tasks required the use of primary or secondary sources (94%), which were either provided by the teacher or had to be collected by the students. These types of tasks are called integrative (reading-to-write, reading-into-writing, hybrid) tasks. Independent writing tasks, on the other hand, are stand-alone writing assignments that do not require previous source texts.

There has been an ongoing debate on which task type measures better academic writing competence: independent or integrative. The main argument against independent tasks was that

the completion of such assignments draws on examinees' background knowledge on the topic, which may differ considerably, thus raising questions about test fairness. The topic has been found to influence students' performance (Weigle, 2004). Similarly, various topic prompts have also lead to significant differences in students' scores (Brown, Hilgers, & Marsella, 1991). This is a major concern regarding the cognitive validity (Weir, 2005) or interactional authenticity (Bachman & Palmer, 1996) of such tests.

Another concern regarding the validity of independent writing tasks is that they do not reflect the extent to which test takers are able to incorporate information from source texts. As the majority of assignments at university involve "discourse synthesis" (Spivey, 1991, see more of this in Section 3.2.4.2), independent writing tasks question what Bachman and Palmer (1996) call situational authenticity of the test.

Advocates of integrated writing tasks for testing purposes showed ample evidence for the positive impact of reading-to-write assignments. Besides authenticity, such were positive washback effect (Grabe, 2003; Leki & Carson, 1994, 1997; Belcher & Hirvela, 2001; Campbell, 1990; Cumming et al., 2005 (Weigle, *Assessing writing*, 2002); Tierney & Shanahan, 1991; Weigle, 2004), or good pedagogical value because such assignments were found to support literary development (Armbruster, Anderson, & Ostertag, 1987).

For the above considerations, international testing agencies (IELTS Academic Writing module) in increasing numbers use integrative tasks in their written exams where examinees need to interpret visual data or read sources as prompts (Gebril & Plakans 2009; Weigle 2002, 2004). The IELTS Academic Writing module, for instance, consists of two tasks, one essay and one integrated. This examination board not only introduced a new integrated task type but also renewed

the corresponding subscales for better assessment of the Cambridge Proficiency Writing Test (Cambridge English TV, webinar, 2013)²⁰.

Even though proponents of integrated academic writing tasks are growing in numbers, there are also voices of caution for careful implementation (Cumming et al., 2005; Plakans, 2008). Due to the complex nature of the writing process, numerous variables need to be controlled for during testing in order to keep the validity of the test, i.e., not mixing the testing of reading comprehension and writing ability.

3.2 Task variables influencing academic discourse quality

So far we have seen what types of writing tasks EFL students are likely to face at university. Next, it will be discussed which variables of academic discourse have a major influence on perceived quality of writing.

3.2.1 The classification of variables in academic writing

Academic discourse quality is an elusive concept with aspects difficult to pinpoint. Several descriptive models have been proposed to address the complex nature of writing with varying category labels and classifications. Such labels are, for instance, **dimensions** (Hale et al., 1996), **constraints** (Kirkland & Saunders, 1991), **parameters** (Weir, 2005), or the general term from social sciences: **variables**. Models considered best by theoreticians are the ones that attempt to capture both the writing process and the written product. Hale et al. (1996) provide a scheme for classifying variables of university assignments in the following way: locus (in-class or out-of-class), prescribed length of product, genre, cognitive demand, rhetorical task (narration, description, exposition, argument), and pattern of exposition (classification, comparison, contrast, cause-effect, etc.).

²⁰ https://www.youtube.com/watch?v=N_CeEM47tYU

Kirkland and Saunders (1991) make a clearer differentiation between “inside” and “outside” forces, which they call internal and external constraints. In their framework the six internal constraints are: (1) L2 proficiency, (2) content schemata, (3) affect, (4) formal schemata, (5) cognitive skills, and (6) metacognitive skills. Content schemata refers to the reader’s/ writer’s prior knowledge of the content area, whereas formal schemata is concerned with the knowledge of organization of a text, on local and global levels, including rhetorical organization. Affect addresses the role of cultural factors and religion in writing. External constraints are (1) purpose and audience of the assignment, (2) features of the assignment, (3) discourse community conventions, (4) nature of the input text, (5) time constraints, and (6) the working environment.

Kirkland and Saunders (1991) in their model focus on the **test taker** (in their research context that is the summary writer) when they describe these variables as constraints (i.e., potential restrictions or obstacles for the writer). Another model, Weir’s (2005) socio-cognitive framework, classifies these variables from the **test maker**’s point of view. It is the clearer distinction of components within context validity (setting and demands), and cognitive validity (processing and knowledge) which makes Weir’s model more suitable for a study setting out to build a summary and a compare/contrast essay corpora.

Weir’s (2005) socio-cognitive framework has been used as a validation tool by the University of Cambridge ESOL Examinations and by other international research bodies. The cognitive dimension of the model investigates the test taker’s mental constructs, while the social dimension addresses writing as a social and linguistic phenomenon.

The model comprises five components, all of which describe important elements test developers need to pay attention to in order to ensure test validity (Figure 9). The first two components need to be observed before the test event, i.e., they are a priori components. **Context**

validity contains external variables, which derive from the task setting context (discoursal, social and cultural variables), and their demands on the test taker. The component **cognitive validity** (also referred to with the somewhat more obscure term theory-based validity) represents the individual differences in processing and knowledge, similarly to Kirkland and Saunders' (1991) internal constraints.

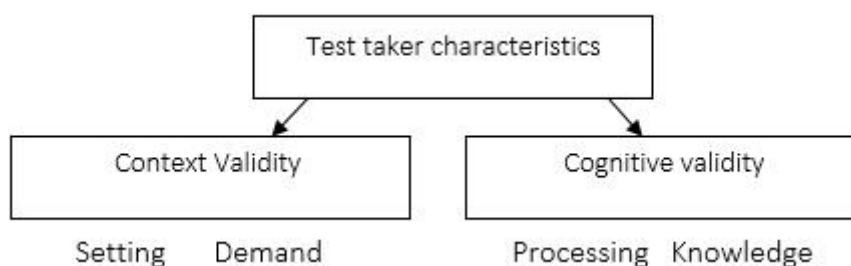


Figure 9. A priori (before the test) components of Weir's (2005) validation framework

The further types of validity components which need to be considered after the test event (a posteriori) are: scoring validity and external validity; which latter is further broken down to criterion-related validity and consequential validity. The a posteriori validity components refer to how the task performance should be evaluated and how this evaluation reflects students' later performance (Figure 10).

In this study, the first two components, **context validity** and, to a lesser extent, **cognitive validity** are considered because they account for the contextual and cognitive variables influencing integrated academic discourse quality: contextual variables control for the social embeddedness of academic texts as a product, whereas cognitive variables refer to the mental operations necessary for academic writing as a process. It is the test creator's responsibility to ensure context validity by examining all the contextual variables during the piloting phase. Information on the cognitive validity of the test can be gained by verbal reports from the examinees and observation of the test

results. In the next section the components of context validity and cognitive validity will be described.

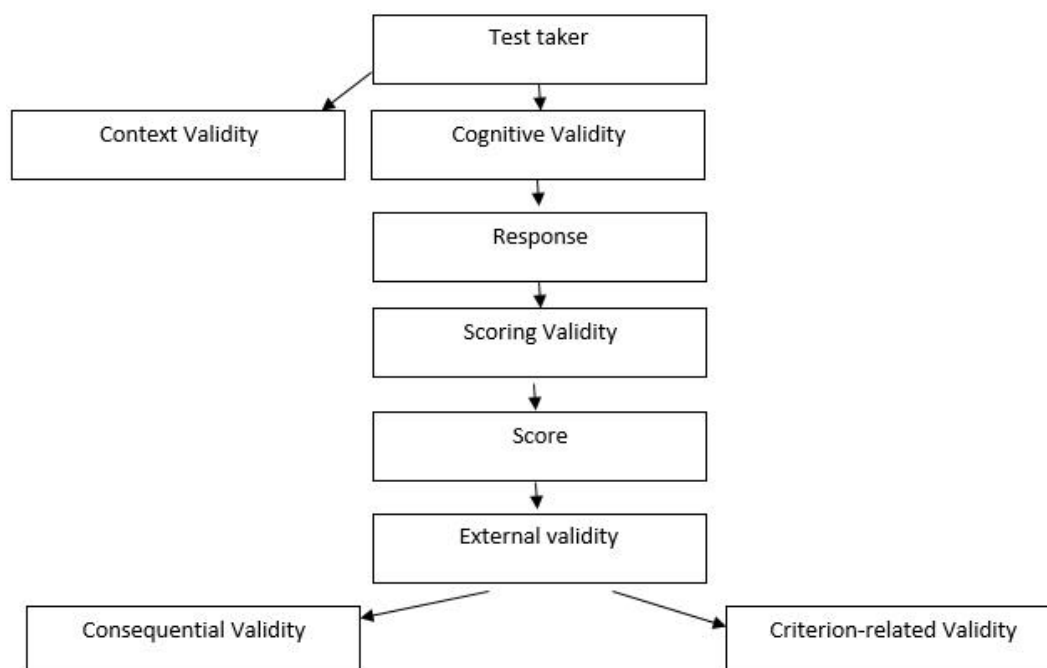


Figure 10. Weir's (2005) whole socio-cognitive framework

3.2.2 Contextual variables of integrated academic discourse quality

Features of task setting	Features of input text
Clarity of purpose	Input format
Intended audience	Verbal/ Non-verbal input genre
Knowledge of marking criteria	Discourse mode
Topic domain	Concreteness/Abstractness of ideas
Genre	Explicitness of textual organization
Cognitive demands	Cultural specificity
Language functions to perform	Linguistic complexity:
	lexical complexity
	syntactic complexity
	degree of cohesion

Table 7. Features of task setting and features of input text within contextual variables (based on Weir's (2005) socio-cognitive framework)

Table 7 summarizes the contextual variables influencing academic discourse quality. Two groups of variables can be distinguished: **features of task setting** and **features of the input text**. The latter is of particular importance when we intend to test students' summary writing abilities.

3.2.2.1 Features of task setting

In optimal cases, the task is set in a clear social framework where students are aware of their own role, the expected audience and the purpose for writing. For instance, students might be asked to write a reader-based or a writer-based summary (Hill, 1991). A reader-based summary can be an alternative to a multiple-choice test on a certain topic learnt in a course, with the teacher as the intended audience. A writer-based summary can be written by, and for, the writer himself, for example, for revision or as a reminder (Hidi & Anderson, 1986). Academic writing is required to be reader-based. It means that the writer should form his ideas in a way that the intended audience can understand the message, using verbal and visual cues that guide the reader. Such cues can be, for example, text structuring, connecting phrases, or the visual arrangement of the text.

The first two variables, **clarity of purpose** and the **intended audience** refer to what Bachman and Palmer call **interactional authenticity** (1996). Another important feature of task setting which might influence discourse quality is **knowledge of marking criteria**. Shaw and Weir (2007) argued that test takers should be fully aware of which criteria are to be used in the marking as it affects students' monitoring and revising processes.

Topic and **domain** are two hierarchical concepts. Within the four domains (personal and daily life, social, academic, and professional) numerous topics can be set as writing tasks. The familiarity of topic affects the writer positively. Bonzo (2008) found that texts written on self-selected topics exhibited significantly higher fluency than those written on teacher-assigned topics. Familiarity with the domain is equally important for writing quality. Tests devised for the highest

levels of language proficiency typically involve topics from the academic or professional domains, which pose greater cognitive demand on the writer due, among many things, to their vocabulary and formal conventions.

The concept of **genre**, and text type and the types of genres required at university were addressed previously in Sections 1.1 and 3.1.1. The category **language functions to perform** refers more to the text types than to the genres. For instance, summarization and synthetization are two language functions to perform, both of which can be parts of an extended argumentative essay. More language functions are, for instance, discussion, expressing opinion, or justification.

The last concept to discuss within the Features of the task setting category is **cognitive demand**. It refers to the complexity of the task, and is placed among the task setting variables because even the easiest task can be made difficult by hardening some other variables. For example, despite having been taught how to write a well-structured compare/contrast essay, students' writing will be of low quality if they cannot understand the two literary texts they are to compare. Similarly, writing an argumentative essay based on a familiar topic might be cognitively very demanding if it is written as an in-class test with a short time limit. These suggest that the above described features interact with each other and all need to be carefully examined while designing a writing assignment.

3.2.2.2 Features of the input text

Integrative writing tasks use one or more source texts as input, the features of which greatly influence the quality of the output. Previous research by Yu (2009) on L2 summarization tasks proved that the features of the input text have a more pronounced effect on students' summarizing skills than their L2 proficiency level. She described macro-organization, frequency of unfamiliar words, topic familiarity, and length of the source text, as the main contributors to summarizability.

This is in line with Kirkland and Saunders' (1991) results, who identified familiarity, complexity and length as key features in the source text to observe before a writing task is set.

Within Weir's (2005) socio-cognitive framework **input format** refers to single or multiple sources, which can be verbal or non-verbal. These latter are for example visual prompts such as a diagram that students need to describe. Writing quality also depends on the **mode of discourse** e.g., narrative, descriptive, expository, or argumentative. The latter two types of analytical writing are more characteristic of academic discourse. The required depth of analysis is easier to achieve if the textual organization is explicit. A text can be more easily followed if it is "signposted" by connecting phrases, such as *moreover* or *on the contrary*.

In the first section of this chapter (Section 3.1.1), where the general features of academic discourse was discussed, linguistic complexity was described as the main contributor for difficulties for language learners. Both lexical complexity and syntactic complexity manifest in a higher degree of cohesion in academic English texts than usual. Evaluating input texts, therefore, is a necessary step when designing a writing task.

Evaluating text difficulty can be expert-based, reader-based, or text-based (Schriver, 1997). An expert-based method is, for example, peer-reviewing an article submitted for publication. Actual readers of the texts can also be asked to comment on the clarity of the writing. Text-based methods, as a third option, utilize computer programs. Various text readability formulae or vocabulary profiler applications exist to evaluate language level and complexity. These three methods can be used separately or in combination. In this research the text-based method was used to analyze the features of input text because it can be considered the most objective of the three.

3.2.3 Cognitive variables of integrated academic discourse quality

Weir's (2005) socio-cognitive framework focuses also on internal features when describing validity measures. He distinguishes five cognitive variables conceptualizing the writer's mental operations. It is interesting to note how these concepts "resemble" the previous list of categories. This similarity is not by chance because contextual variables are directly related to cognitive variables, see Table 8 for the latter.

Cognitive phases	Cognitive processes
Conceptualization	Task representation
	Macro-planning
Meaning and discourse construction	High-level reading processes
	Connecting and generating
Organizing	Organizing
Low-level monitoring and revising	Low-level editing
High-level monitoring and revising	High-level editing

Table 8. Cognitive variables involved in integrative academic discourse based on Chan (2013) and Chan, Wu, & Weir (2014).

For instance, as Table 8 shows, the cognitive process **task representation** is a mental operation occurring during identifying the **purpose** and the **intended audience**, the first two variables from the previous list. This is so because the writer uses his previous knowledge and schemata when starting the writing process. Schemata, as Brown and Yule maintain, "lead us to expect or predict aspects in our interpretation of discourse" (1983, p. 248), thus help focusing on what the task requires.

For the purposes of this study, from Weir's (2005) cognitive variable categories **meaning and discourse construction** and **organizing** are key terms because they directly affect writers' decisions on lexical patterning. Therefore, in the following we will focus on these two variables only. Given that there are several other influential models which describe these two variables from various perspectives, they are introduced first in the next section.

3.2.3.1 *Cognitive processes in writing*

Extensive research has been carried out to map the mental operations involved in writing academic discourse, and several models have been proposed to describe the processes, the influencing factors, and the individual differences (e.g., Flower & Hayes, 1981, later modified by Hayes, 1996), motivated by the assumption that text is not only a product of academic inquiry but also is a process. In the traditional view, writing consists of five sequential stages: pre-writing, drafting, revising, editing, and publishing. The term "stages" might imply that one stage precedes the next. Research has proven that processes throughout writing rarely occur in fixed steps, rather, writing is modelled as consisting of multiple, sometimes overlapping recursive/recurring processes. The most influential models were proposed by Flower and Hayes (1981) and Hayes (1996), and Scardamalia and Bereiter (1987).

Composing a text involves three main cognitive processes according to Flower and Hayes (1981): **planning**, **translating** and **revising/reviewing**. Planning means generating ideas for writing, organizing them as well as goal setting; translating (or text generation) stands for turning ideas into written text; while revising/reviewing comprises evaluating the written text and making any necessary improvements. Monitoring is also mentioned as essential for evaluating the progress of each process.

The model also identifies two further interactional components: task environment and writer's long-term memory, referring to external and internal variables, respectively. The latter component (writer's long-term memory) includes a variety of information stored in the brain, for example knowledge of conventions about the particular genre. (Translating in this context refers to converting concepts into a linguistic form, rather than translating between languages.) The components are illustrated with Figure 11.

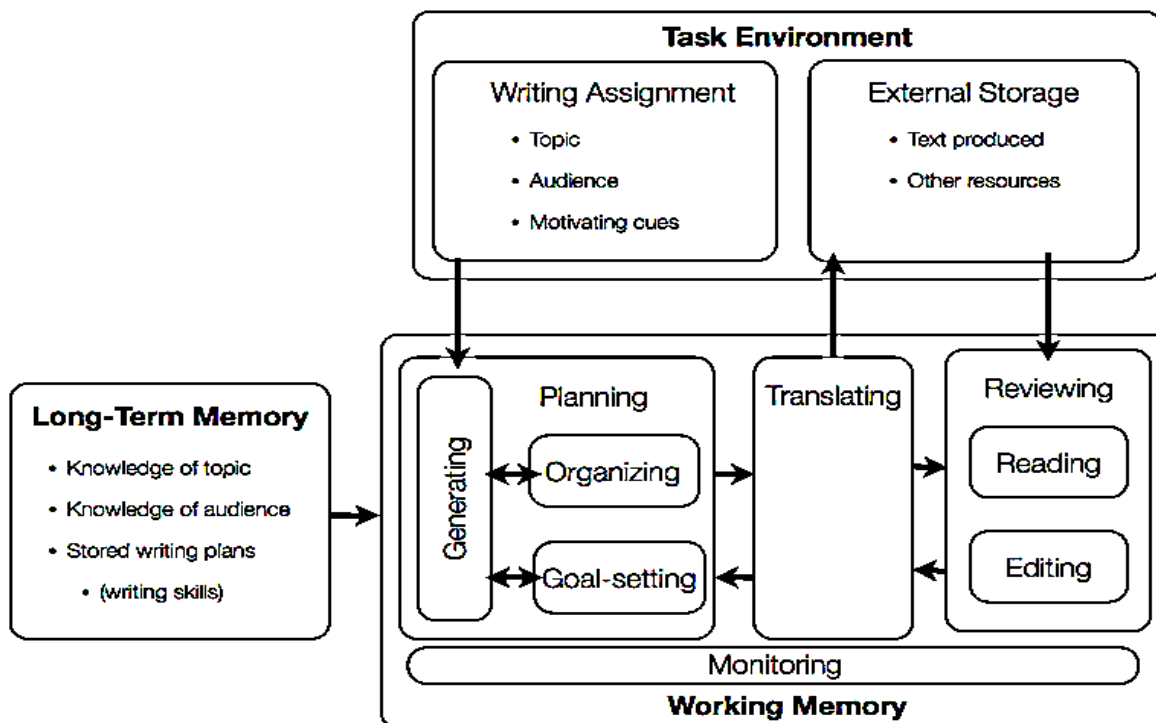


Figure 11. The Cognitive Process Model of the Composing Process (Flower & Hayes, 1981, p. 370)

Even though the Flower and Hayes (1981) model did not perceive the writing processes to be linear, the labelling might have implied that three distinct stages occur. Therefore, in his revised model, Hayes (1996) changed the original labels to **reflection**, **text production** and **text interpretation**, respectively, indicating perhaps more clearly the recurring nature of the writing process. Two important differences are that planning, which term could have been perceived as the initial stage of writing in the first cognitive model, was subsumed by reflection, and revision became part of text interpretation. (A more detailed description of these models as well as enumeration of various other cognitive models of writing is found in Galbraith, 2009).

The other influential model describing two approaches to writing is that of Scardamalia and Bereiter's (1987), who distinguished between **knowledge telling** and **knowledge transforming** as two modes of meaning construction. In knowledge telling, writers use source

texts to generate ideas; they do not reconstruct these ideas by connecting them to previous knowledge. In knowledge transforming, writers synthesize the ideas they read and also use their background knowledge from their memory, thus creating text in a different order of importance which reflects the writer's goals instead of the structure of the input texts.

It is common procedure to use think-aloud protocols with participants to identify the processes involved in writing (e.g., Flower & Hayes, 1981). Even though a lot can be learned about mental operations with this method, it is still controversial, as writers need to reflect on mental processes that they might not be aware of. Besides the traditional think-aloud protocols and questionnaires, more objective methods have been used recently for the study of the complex processes during writing. These clearly show the overlapping nature of cognitive processes.

In several recent studies, for instance Wengelin et al. (2009), set out to study the reading and writing processes that occur during typing by utilizing eye-tracking and keyboard-logging mechanisms. They found that both text composing and reading one's own text takes place simultaneously, doubling the cognitive workload. This resulted for instance in more backward reading during text composing than it was measured during reading static texts. They argued that the reason behind this could be a number of revisions and checks for cohesion in one's own text.

3.2.3.2 Implications for language pedagogy

Two conclusions can be drawn from the literature on the cognitive processes used in integrated writing tasks. The first conclusion for language pedagogy is that EAP teachers need to know that, according to research, academic writing is not a linear process or a sequence of linear processes. In light of this, it is interesting to find that writing teachers usually give out guidelines for students, where the "steps" or "stages" of how a writing task should be carried out are numbered. It can only be hoped that they also mention the recurring nature of these steps.

The second conclusion is that we have to distinguish between the sequentiality in the text creation process and the sequentiality in written prose as a product. The former is recurring (or overlapping) in nature, whereas the latter should follow a strict linear order, at least in the English academic writing tradition, which is known for requiring linear topic progression²¹ (Lautamatti, 1987).

To what extent such a connection exists between processing patterns and discourse organization seems an interesting field for further investigation. One work addressing this question is that of Lenski and Johns (1997), who watched middle-school participants doing a research project in their mixed interview/ observation study. Their aim was to find out how students use readings in their writing assignments, and to what extent the organizational pattern of their final text follows their researching pattern. The students used three different approaches for researching: sequential, spiral, or recursive. Sequential and spiral approaches to using the source texts resulted in texts with a summary-type format. The student (only one of the six participants) who used a recursive researching pattern wrote an integrated paper. These are interesting pedagogical implications because they indicate that researching patterns might influence discourse organization, which observation should be exploited in academic writing classes.

3.2.4 Summary writing as a complex task

3.2.4.1 Definitions and purposes of summaries

A number of definitions are available for summaries, all of which agree that a summary contains a gist of a source text or source texts by presenting its/their main points. It should be comprehensive, brief and it recapitulates previously stated facts. Definitions also differ, depending

²¹ When a topic is progressing linearly, new information is transformed into known information as sentences follow each other in order to achieve coherence in the text.

on the purpose of the summary and subsequently, the various purposes determine the required features of the text in terms of structural and linguistic elaboration.

In an educational context, two main uses of summary writing have to be distinguished: the summary essay and summarizing a short passage from a source text within one's writing. In the first case, students' task is to write a summary based on a reading or readings, i.e., to write down the gist of that text without including their own opinion on the subject. In the second case, students incorporate other writers' work into their own writing either for reviewing the literature, or using the summarized passage for supporting their own argument. When a source text is used as support, direct quoting or paraphrasing is an alternative to summarizing, as shown in Table 9.

Writing a summary 'essay'	Summarizing
involves the skill: summarizing	involves the skill: summarizing
structured similarly to essay: introduction, body paragraph(s) with topic sentence(s), conclusion	part of writer's own text only a short passage of source text
only gist of the source text, not containing writer's opinion	source is used for literature review or as support for writer's argument
	alternative for direct quoting or paraphrasing

Table 9. The differences between writing a summary 'essay' vs summarizing

Two main types of summaries are produced in academic contexts: complete source text summaries and guided summaries (Tankó, 2012). In the case of the complete source text summaries, the students are asked to collect all the main ideas of the given text, whereas in the cases of guided summaries only the ideas need to be found that relate to a certain topic or topics.

Table 10 shows the differences between the questions asked during the process.

Idea selection questions of whole text summary	Idea selection questions of guided summary
What are the main ideas of the text?	What are those main ideas in this text that are related to topic X/to aspect N of topic X?
What are the supporting details of the main ideas?	What are the supporting details of the main ideas related to topic X/to aspect N of topic X?

Table 10. Whole text summary task and guided summary task questions for idea selection based on Tankó, 2012, p. 119.

3.2.4.2 The cognitive demand behind writing a summary

During the completion of an integrated writing task more processes are activated than during writing an essay using prompts, as the roles of reader and writer alternate. Hayes (1996) distinguishes three purposes of reading when writing: (a) reading to define the writing task, (b) reading source texts to obtain writing content, and (c) reading and evaluating text produced so far.

Van Dijk and Kintsch (1983) when explaining discourse coherence both at the local and global levels, identified the important role the structure of the source text plays in summary writing. According to their model, each text has macrostructures and microstructures which can be found on different levels. Macrostructures represent the global meaning of text, and local microstructures need to be interpreted within this global framework which provides cohesion.

The first use of macrostructures is to organize complex microstructures. The second use is reduction. Microinformation can be turned into macrostructures by applying the derivative rules called macrorules. These are (1) **deletion** of irrelevant information, or its reverse, **selection**, keeping important information; (2) **generalization**, creation of a global proposition; and (3) **construction**, creation of a new proposition. These rules have become also known as summarization rules because they comprise the cognitive processes summarizers employ.

Spivey conducted several studies to investigate the processes unique to reading more than one source texts for writing (e.g., Spivey, 1991; Spivey & King, 1989). Her term “discourse synthesis” is defined as “a process in which readers read multiple texts on a topic and synthesize them” (Spivey & King, 1989, p. 11). Three processes of discourse synthesis occur while reading multiple sources for the purpose of writing: selecting relevant content, organizing content according to writing goals, connecting content from source texts and generating links between ideas. Better and less abled readers, as well as different age groups use these skills differently because the two main factors influencing composing from sources found to be comprehension and sensitivity to text structure. Expert readers, Spivey found, therefore develop more complex writing plans. Table 11 summarizes the two above models, as well as shows two other classifications regarding summarizing processes.

van Dijk & Kintsch (1983)	Spivey & King (1989)	Johnson (1983)	Hidi & Anderson (1986)
deletion of irrelevant information	selecting relevant content	comprehending individual propositions	comprehension
selection, keeping important information	organizing content	establishing connections between them	evaluation
generalization, creation of a global proposition	connecting content	identifying the structure of the text remembering the content	condensation
construction, creation of a new proposition	generating links between ideas	selecting the information to be included in the summary	frequent transformation of ideas
		formulating a concise and coherent verbal representation	

Table 11. Mental processes during summarization as suggested by four models

Interestingly, none the above four well-known and often-cited models mention “classifying information gained from the source text” as one of the major components of summarization. Spivey and King (1989) identify **organizing-**, and **connecting content**, as well as **generating links** between ideas, however, these can refer to the actual formation of the text, rather than the mental process of categorization of the ideas from the input. Similarly, Johnson also refers to the process of **establishing connections** between individual propositions, yet again, it is not specifically spelt out that it includes creating mental categories, or creating textual connective cues, or perhaps both.

For proof of the importance of classifying ideas, the construct of the summary writing in this research can be brought here as an example. The task in Stage 1 was to summarize a text with regard to the history and the business model of a company. The latter had to be further broken down into its strengths and weaknesses. Therefore, for the instruction “*Summarize Ryanair’s history and business model, highlighting its strengths and weaknesses*” the students also had to generate hierarchies, as well as evaluate content. Furthermore, evaluation had to be done on a higher level: not only evaluating information of the source text on the **relevant—irrelevant information** scale, but also on the **negative result—positive result** dimension. The cognitive processes students had to employ is illustrated in the table of the summary construct in Table 12. As they read a structurally non-explicit source, the information had to be gained from different parts of the text.

Main constructs of the summary task	Cognitive process involved	Divisions of main constructs	Cognitive process involved
History	<i>selecting relevant information</i>		
Business model	<i>selecting relevant information</i>	strengths	<i>evaluating information as positive or negative</i>
		weaknesses	

Table 12. The constructs, their divisions, and the cognitive processes involved in the summary task of the research (Stage 1)

The above review of the various models indicates that summarization is a complex skill, which is difficult even in one's own language. Eileen Kintsch (1990) found, for example, that even more advanced students are incompetent in using macrorules in expository texts. Several other studies examined the relationship between age and proficiency in summarization, for instance comparing high-school and college students. Results suggested that college students' performance was better in finding the main ideas. They also outperformed the younger students in condensing more ideas into the same number of words (Hill, 1991; Brown, Day, & Jones, 1983; Brown, Smiley, & Lawton, 1978). Others, such as Winograd (1984), however, maintain that the main reason behind students' summarizing difficulties is the lack of strategic skills, particularly reading skills.

A recent large-scale questionnaire study by Li (2014) revealed several interesting aspects of summarization about EFL students' reading and writing strategies during the task. It was found that the most frequent reading strategies were **skipping unknown words** and **re-reading for clarification**, while the least frequently employed ones were **checking inference** (checking whether the students understood what the source text really meant) and **re-reading the instruction**.

The most frequent writing strategy was **planning content from source**, whereas the least frequently employed were **reorganizing source pattern** and **editing for punctuation**. These findings have serious implications for both language pedagogy and test making. Particularly the frequency of the writing strategies mentioned needs attention. Students very often used the strategy planning content from source. This possibly means that they kept turning back to the source text for ideas to be included into their summaries. However, it might also imply that the students copied directly large amounts of the source text.

The finding that the students' least employed strategy was reorganizing source pattern might indicate that they extensively relied on the structure of the source text. In other words, they

did not attempt to synthesize ideas from different paragraphs, rather they linearly read the source text and treated the content of each paragraph as separate entities. This result calls for more emphasis on teaching summarization skills in EAP classes, and it should be taken into consideration while planning summarization tasks.

3.2.4.3 Computer generated summaries: lessons to be learned about quality

Learning about computerized discourse creation can inform pedagogy about the processes performed in writing. Particularly the unresolved issues are of interest because they highlight areas which might involve unique or unconscious human abilities. Within the field of computer science, where the rapid growth of textual data available has stimulated extensive research, summarization is considered a natural language processing²² (NLP) task, its main uses being information extraction and question answering.

The main purpose of computer generated summaries is to yield factual information from a written text in various genres, e.g., news articles, or scientific articles; however it is also possible to summarize spoken language, e.g., news broadcasts, meetings or recorded lectures. To a lesser extent it is used for summarizing opinion, which is much more challenging as speech is less structured than written text, both at syntactic and discourse levels.

Computerized summary writing can result in two types of summaries: extractive and abstractive (or generative) summaries (Spärck Jones, 1999). Extractive summaries create long sentences based on the key information of the source text. Abstractive summaries break down these lengthy sentences into smaller ones by eliminating not central information such as a relative clause while observing syntactic information such as sentence structure, and at the same time observing higher levels of discourse to recognize such elements as the theme/rheme structure of text.

²² NLP tasks involve enabling computers to derive meaning from human (i.e., natural) language input.

There is a distinction between computer-generated and human-generated summaries in purpose and structure, as well as linguistic elaboration. As a consequence, the definition of a computer-generated summary does not contain **construction** (cf. van Dijk & Kintsch, 1983, summarization rules, Section 3.2.4.2), which is a major qualifying element for academic summaries. A definition of a summary in the language technology domain is “a reductive transformation of source text to summary text through content condensation by selection and/or generalization what is important in the source.” (Spärck Jones, 1999, cited in 2007, p. 6) This definition suggests that nothing is required from the summarizer program “where something can go wrong”, such as coming to a wrong conclusion or inventing an idea not present in the text. It can also tell us that **construction** is the most demanding cognitive skill employed by human writers.

Summaries which are similar to extractive summaries are not acceptable as reader-based summaries at universities: firstly, because they contain word-for-word information from the source text, which is considered plagiarism, secondly, because academic summaries need to contain generalization and invention (van Dijk & Kintsch, 1983), and thirdly, because they are not textually well-formed. In other words, they do not contain the cohesive and stylistic elements which are required in academic discourse.

As a consequence for this study, the academic summaries to be analyzed will be high-rated if they have the following features: (a) containing the key information from the source text avoiding word-for-word copying; (b) using generalization and construction; (c) being textually well-formed.

3.2.4.4 Computer generated summaries in language pedagogy: a virgin territory?

The previous detour into automated summarization also shed light on a number of problems surrounding summarization in an academic context. The first, surprisingly obvious, however, seemingly never tackled question is how it is possible that test creators in an academic context do not resort to using automated summarization techniques for selecting important information in the input text while giving out a summary task. Instead, several teachers are asked to manually select important information from the source text and then reach a consensus on what should be included in the summaries as main constructs (see for example the Methods sections in recent studies, e.g., Keck, 2014; Yu, 2009).

This practice has two consequences for assessment. Firstly, task creators' decisions about what is considered important in the source text will be based on consensus but left without justification. It will not be triangulated objectively against computerized measures. Given that the original texts are not examined with any of the above described summarizing techniques either in university content courses or in EAP writing courses, the score given to the Content rubric in the assessment will remain subject to the teachers' judgement regarding the important information in the source text. In order to justify the quality judgement of summaries, it would be optimal for both the source text and the created summary to be analyzed with the same discourse analytical methods. At the moment no such practice is described in the assessment literature.

The second consideration for using some form of computer-assisted analysis on the source text might be that the Language (within this: Style and Vocabulary) rubric would also benefit from objectively analyzing the original text. Comparison of lexical complexity, for instance, might indicate that the source text uses more complex elements which the particular student's summary

does not reflect in vocabulary, therefore his lower score in the Language rubric could be justified by providing factual information, such as figures or percentages. The Organization rubric could benefit in the same way as regards rhetorical features. Thus, it can be concluded that the lexical repetition analysis (LRA) tool this study is aiming to design addresses a clear gap in research in these areas.

3.2.5 Writing a compare/contrast essay

3.2.5.1 *The cognitive demand behind comparing and contrasting*

Writing a compare/contrast text is a complex cognitive process, in which various higher order thinking skills are utilized, mainly analyzing and evaluating during the idea generating phase, and synthesizing and creating²³ in the text structure planning and writing up phase. Similarly to academic summaries, compare/contrast text types are widely used as a means of assessment, being capable of measuring not only subject knowledge but also reading comprehension. They can substitute for true/false or multiple choice questions.

The content of such a text needs to be guided by careful structuring, the logical arrangement of ideas being as important as criteria for comparing. Describing the major characteristics of the comparison process, Mitchell observes that it is “a double process of equating and differentiating things, seeing them as similar or the same, on the one hand, and as distinct or different on the other” (Mitchell, 1996, p. 321). He concludes that comparison is “never just finding similarities, identities, or equivalences, nor is it just finding differences. It is the dialectic between similarity and difference, the process of finding differences between things that appear to be similar, and of

²³ The higher order thinking skills mentioned here are part of the new version of Bloom’s (1956) taxonomy of the cognitive domain, revised in 2001. The main difference is that the highest two categories have been reversed and renamed: **creating** is the highest, **evaluating** is now second (Anderson & Krathwohl, 2001).

finding similarities between things that appear to be different.” This nature of comparison is highlighted in Figure 12.

These two signs are similar in x aspect. O – O	These two signs are different in x aspect. O – O
In what way are they different?	In what way are they similar?

Figure 12. The notion of comparison (Mitchell, 1996)

The cognitively demanding nature of the compare/contrast writing task can be eased if the teacher provides some guidance while setting the task. Usually the two topics that need to be compared and/or contrasted are given, providing only a broad framework. Further guidance is given when the aspects used as criteria for comparison are also offered. Besides helping students meet the content requirements, another aid can be the visual essay planner. This makes the task easier because students can start planning their essays by grouping ideas, thus having a clear focus when further developing and organizing their writing.

3.2.5.2 *The compare/contrast essay genre*

Similarly to the act of summarizing, the act of comparing and contrasting can refer to creating a certain text type and a genre. The latter is called the compare/contrast essay. It is interesting to note that some course materials²⁴ on academic writing discuss comparing and contrasting and its two structural patterns within their chapter on argumentative writing (e.g. Oshima & Hogue, 2006), regarding it as a text type, while other sources discuss comparing and contrasting as a separate genre of its own right. Oshima and Hogue’s treatment also indicates that these two genres (argumentative and compare/contrast essays) are similar in structure, the difference lies in the argumentation/description employed, respectively. In the following, the

²⁴ also the Harvard Online Writing Lab guide (Walk, 1998) <http://writingcenter.fas.harvard.edu/>

compare/contrast essay genre will be described. (The meaning of the term genre was already explored in Section 3.1.1, where the nature of academic discourse was discussed.)

3.2.5.3 The content of compare/contrast essays

Instructional materials address four special areas of concern for compare/contrast essay writers: the title, the frame of reference, the grounds for comparison, and the thesis. Guidebooks emphasize that students should read the given essay title carefully, paying particular attention to the instruction verb (compare, contrast, or compare and contrast), and should not do both if specifically instructed to do only one. (Gillett, Hammond, & Martala, 2009)

Handbooks point out that the frame of reference, i.e., the selection of the two concepts to be compared should be meaningfully comparable. The more similar the two items are, the more focused the essay will be. Typical items to compare in an academic context are two texts, two theories, two historical figures, or two scientific processes, as suggested by the Harvard Online Writing Lab guide (Walk, 1998)²⁵.

Course books also emphasize the careful selection of criteria for comparison. A common mistake of inexperienced academic writers is that they do not apply the same criteria for both concepts they try to describe. Some guides, mostly online university sources, prescribe quite rigid “rules” in this respect; for instance, they suggest selecting no more than three criteria for comparison, each for a separate paragraph in the body (Zemach & Rumisek, 2003a, 2003b).

Handbooks also warn that the task is not a mere description but emphasis needs to be placed on showing the ways these two items are different or alike. The introduction should explain the reason for the comparison or contrast (e.g., which is more desirable of the two options). The thesis

²⁵ <http://writingcenter.fas.harvard.edu/pages/resources>

statement should clearly state (1) the two items to be compared, (2) the criteria for the comparison /contrast (i.e., the main points), and (3) the writer's judgement (Oshima & Hogue, 2006).

An exception to the above described, rather didactic, "instruction manual" approach, mainly created for novice writers, is the Harvard Online Writing Lab (Walk, 1998) where a more sophisticated structure is suggested for writing extended compare/contrast essays (2,500 or more words).

3.2.5.4 The structure of compare/contrast essays

Few research papers have addressed the compare/contrast essay from a structural point of view; even though the compare/contrast rhetorical pattern is universal in academic discourse (Newby & Penz, 2009). Explicit instruction in this structure may have several positive effects on students (Hidi & Anderson, 1986). Although it is scarcely in the focus of scientific research, the compare/contrast textual pattern appears in most teachers' guides and self-study handbooks on academic writing.

The compulsory formal elements of the genre are (an) introductory paragraph(s), followed by body paragraphs, and (a) conclusion paragraph(s). The plural forms are meant for the case of extended essays, with cca. 2500 word length. Academic writing course books suggest two main types of organization for the compare/contrast essay to present information in an orderly manner: the point-by-point and block-by-block patterns. (e.g., Oshima and Hogue, 2006).

The point-by-point (slice or alternating) pattern, as the term suggests, compares the items one point at a time, usually employing three or more criteria as bases for the comparison. In each paragraph, the topic sentence, according to the guides, should focus on the criterion compared rather than the item which is compared, or contrasted.

Figure 13 outlines the point-by-point method used for comparing colleges and high schools using three criteria. It also gives suggestions for the content of the thesis statement in the introduction covering three areas: the items compared, the reason for comparison and the main points. Similarly, the conclusion paragraph is suggested as a brief summary or restatement of the main points, as well as an indication of preference for one of the items compared (Humber University, Department of English, The Writing Center²⁶).

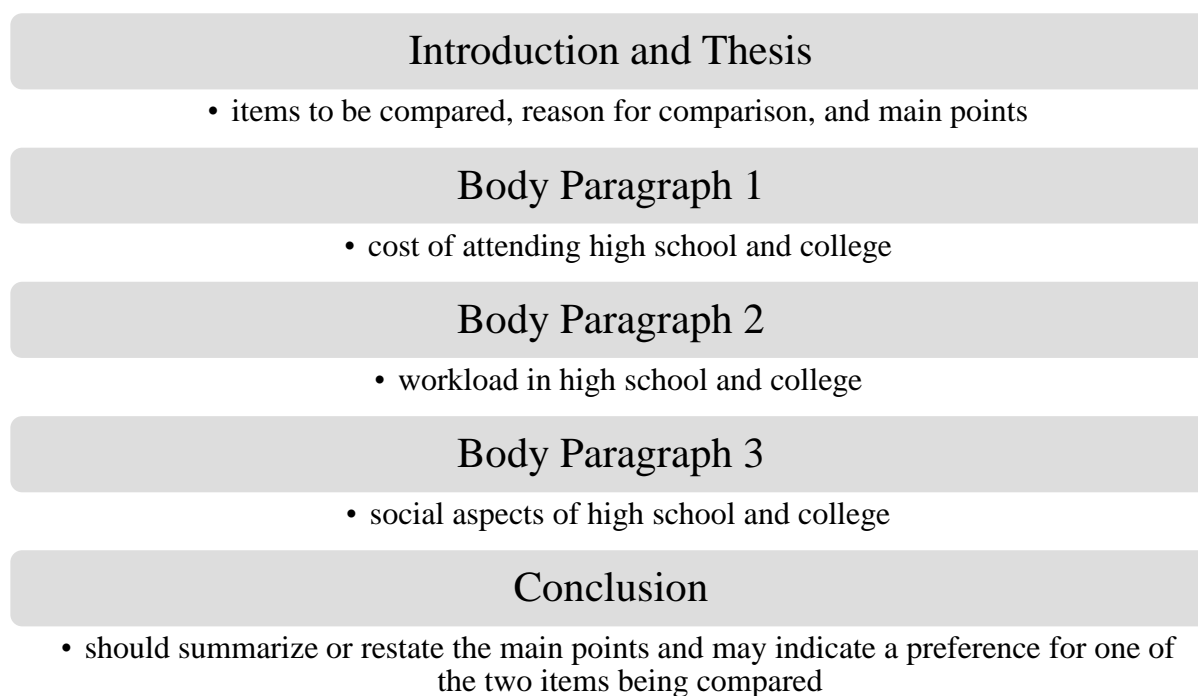


Figure 13. Suggested outline for the point-by-point pattern (based on the online academic writing guidelines of Humber University)

The second pattern appears in various terms: block, block-by-block, text-by-text, subject-by-subject, chunks, clustered pattern (Figure 14). When this method is applied, there will be only two body paragraphs in the case of a short essay, whereas for longer essays, which use more than three criteria for comparison, an even number of body paragraphs will be more desirable, as an

²⁶ <https://www.humber.ca/liberalarts/las-writing-centre/Student%20Resources>
<https://www.humber.ca/liberalarts/sites/default/files/COMPARE%20AND%20CONTRAST%20ESSAY.pdf>

equal amount of description should be devoted to both topics. Advice on the presentation of the comparison emphasizes that the main points should be grammatically parallel so as to make sentences natural and clearer in style, and they must apply equally to both items.

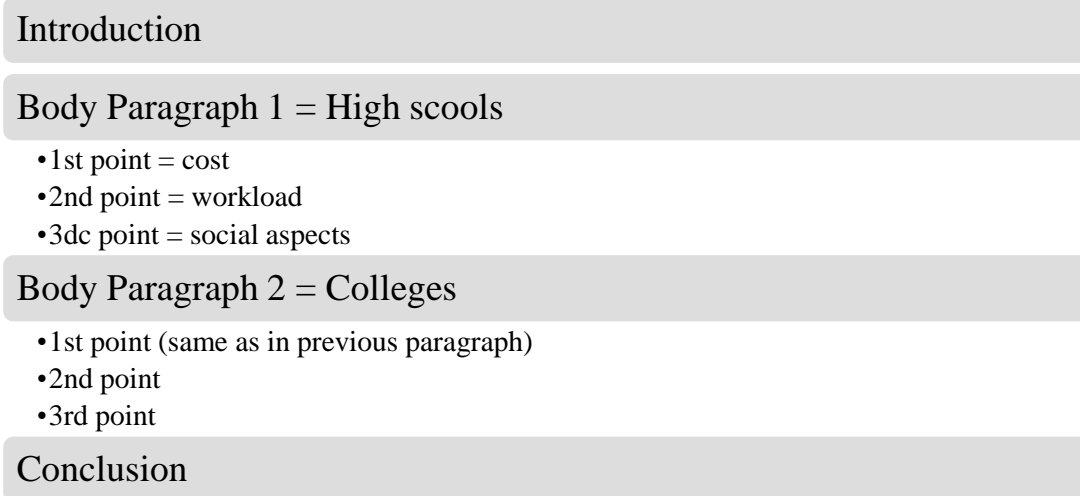


Figure 14. The block-by-block pattern

Teachers' experience suggests that these two, seemingly easy structures are difficult to keep during essay writing, particularly when students need to read sources to write their compositions. Explicit instruction in both summarization skills and text structuring are necessary for an optimal compare/contrast writing (Hammann & Stevens, 2003).

So far we have seen what kind of academic writing tasks are required from students and we dwelt on certain contextual and cognitive aspects of academic discourse, particularly on the various aspects of summary and compare/contrast essay writing. The next section within this chapter will examine how teachers decide whether these texts are of good or bad quality.

3.3 Assessing²⁷ academic discourse

One of the main aims of this dissertation is to propose a computer-aided model which can analyze the organization of student writing by studying lexical repetition patterns, and based on these results, can evaluate text quality. In order to perform this task, it is important to see how teachers assess students' texts in real life. The literature on assessment theory and practices is extensive but some of its aspects, particularly theoretical concepts such as validity, reliability and generalizability of assessment, are relevant for the purposes of this study. Validity has partly been dealt with in the previous sections: the contextual and cognitive variables of academic writing have been explored.

Another key issue is reliability, in other words: to what extent teachers assess students' writing consistently. Within this field, the teachers' judgements regarding coherence, cohesion and lexical repetition will be explored, i.e., how perceptions of these features of text influence the overall judgement of text quality. Finally, "traditional" human and automated evaluation methods will be compared. By "traditional" assessment I refer to the manual evaluation of students' assignments and tests at universities, mostly carried out by the class teacher; as opposed to automated or semi-automated assessment made, entirely or partly, with the help of a computer.

3.3.1 'Traditional' and recent academic essay assessment practices

Assessing written academic discourse requires great effort and time, still most teachers, as a rule, use 'traditional' paper-and-pen methods when they correct assignments at universities. The two most widely employed assessment techniques are either holistic or analytic. The third, less frequently used, assessment method is called primary trait evaluation (Crusan, 2010).

²⁷ For the purposes of this study, we treat the terms **assessment**, **evaluation**, and **scoring** as synonyms. In the same fashion, **assessor**, **rater**, and **evaluator** all refer to the person performing the quality judgement of texts.

During holistic scoring, raters form an overall opinion of the compositions. There are rubrics to guide the rater to judge content, organization, language and grammar; however, the rater is asked to give one overall grade for each text, usually employing six rather than five-point scales. The reason for this is that it is easier to place a writing into the “better than average” or into the “worse than average” dimensions when the rater can choose from six scores. Appendix E shows such a holistic scale used in the TOEFL exam.

Holistic scoring has been criticized for two main reasons. The first reason is, according to Cumming, et al., 2001, p. 3) that “the exact nature of the constructs they assess remain uncertain”. They are too general and the raters cannot interpret them well. The second reason is that it is not clear how raters weight each construct, thus it is possible that every rater interprets the same construct differently.

Analytic scoring includes several writing components, such as content, organization, cohesion, register, and accuracy (Weigle, 2002). Each of these components are scored separately, and some of the features can be scored higher than the others. The third type of scoring (primary trait scoring) is a variation of analytic scoring however, in this case only one feature of writing is analyzed, e.g. vocabulary use, or grammar. Two examples of such analytic scoring rubrics are shown in Appendix F.

As a rule, the features selected for assessment should be based on the specific context of the assignment (Crusan, 2010). For instance, there should be a different scoring guide prepared for a persuasive essay or a paragraph writing exercise. Analytic scoring was criticized by White (1984) maintaining that writing subskills are not separable, and also by Myford and Wolfe (2003), who concluded that raters cannot separate the scales while rating.

Due to these problems, there is a call for quicker, more objective, text-based methods in assessment. A growing tendency can be seen to justify one's decisions by utilizing textual analysis programs, such as Coh-Metrix (McNamara, Louwerse, Cai, & Graesser, 2005) to gain detailed information on discourse features. An example for one such textual analysis can be seen in Appendix G. It describes the input text used for this study, in Stage 1. Such analytical tools can help teachers to compare various features of texts, and are also useful when giving written feedback on the writers' performance.

Table 13 shows the continuum from the traditional towards increasingly computerized techniques of assessing written academic discourse employed today, from fully manual assessment to fully automated assessment.

	Manual assessment	Semi-automated assessment	Fully automated assessment
method	holistic vs. analytical (scales/rubrics)	using statistical programs (e.g., counting word frequencies)	by applying statistical means vs. LSA (see Appendix G)
type of data analysis	mainly qualitative	mainly quantitative	mainly quantitative
feedback	written comments or numerical (grades) on the general quality of the essay, or separate scores for each rubric	a calculated list of features, only certain features are/can be checked, e.g., key words, lexical bundles (chunks, n-grams)	a calculated list of features converted into written comments, grades (calibrated), or in a table format (%)
reliability measures	induction sessions for assessors before assessment; two or more experienced assessors whose grades are averaged; using validated scales	simple statistics; settings should be accurate before calculations are made	statistical analytical methods; large number of variables; machine learning; two experienced human assessors' judgement, or one human and machine
treatment of lexical repetition	repetition warning, unclear "rules"	lexical diversity index	lexical diversity index, ratio of exact repetition vs. synonym only general advice on usage

Table 13. Assessing written academic discourse

3.3.2 Validity in L2 academic writing assessment

In theory, validity in testing means the extent to which the test measures what it is supposed to measure (Denzin & Lincoln, 1994). In practice, validity is a complex and context-dependent issue. A test is said to be valid if it "reflects the psychological reality of behaviour in the area being tested" (Hamp-Lyons, 1990, p 71). In an EAP context, effort should be made to create "authentic tasks" for testing. An authentic task is similar to the tasks the language user tends to encounter in the target situation (Bachman & Palmer, 1996), i.e., in their content courses. The cognitive and contextual variables of writing need to be considered not only for task setting but also for assessment.

In Moore and Morton's (1999) study where they compared the IELTS Academic Writing Task 2 to Australian university disciplinary course assignments as regards validity, they found that in both contexts, essays were the predominant task. However, a closer inspection revealed that there are major differences in rhetorical function. While in both contexts, evaluation was the most common rhetorical function, with comparison also present, in IELTS testing the second most common was exhortation (arguing for the notion of necessity, should-ness, inciting) which was rarely present at university disciplinary courses. Such an exhortatory sample task they collected from IELTS tests was for example the following:

Television nowadays features many programs of a violent nature. For this reason, it is necessary for parents to impose strict controls on their children's viewing habits. (Moore & Morton, 1999, p. 100)

The rhetorical functions in IELTS Task 2 items were less varied, with some university tasks such as summarization and description, for example, totally missing. In their explanation why exhortation was the predominant rhetorical mode in testing, they posited that the background

knowledge required for this task is less than for other university tasks that need search of the topic, especially if the objects of enquiry were general knowledge topics such as studying abroad, health care, or computers in education. Nevertheless, these results initiated a change towards test tasks which have more predictive validity in university settings.

3.3.3 Generalizability of judgement on academic discourse quality

According to Messick, generalizability in educational measurement is “an aspect of construct validity” which can be interpreted as reliability or transfer. Generalizability as reliability thus, is “the consistency of performance across the tasks, occasions, and raters of a particular assessment”, whilst generalizability as transfer is “the range of tasks that performance on the assessed tasks is predictive of” (Messick, 1989, p. 250).

The above two notions and their descriptions refer to the test construct. Generalizability as transfer thus, refers to the task the test takers are required to perform. However, the range of tasks the **raters** perform is also an informative aspect of assessment because it can indicate how exactly they decide on writing quality. A large-scale think-aloud study into this area (Cumming, Kantor, & Powers, 2001) asked professional essay assessors of ESL compositions to record what they are thinking about, deciding and doing while holistically rating sixty written essays of varying language proficiency levels. The descriptive framework of raters’ decision-making behaviors can be found in Appendix H. Most raters, all native speakers of English in the study, read the composition at hand quickly and made evaluative reflections on it as a whole, summarizing the key features which affected their impressions most. They paid a balanced attention to content, argumentation and language.

There are general rules about academic writing that apply unanimously, but the quality of writing is also genre and discipline specific to a great extent. Research motivated by the Writing

Across the Curriculum /Writing in the Disciplines (WAC/WID) movement described significant differences in the teaching of writing in diverse fields in higher education. Such differences exist, for example, in the required amount of “technical expertise demonstrated, the relative importance attached to content and style, in the importance ascribed to truthfulness, confidentiality, register, format, evidence, and authorial point of view” (Haswell, 2013, p. 416; also Anson, Schwiebert, & Williamson, 1993). These are just a few of the factors influencing teachers’ scores when assessing the quality of student writing. As a consequence, it would be sensible for a tool to measure general academic English conventions disregarding discipline specificity, and concentrating on EAP course requirements when defining quality.

3.3.4 Reliability of perceived discourse quality

As a criticism of present evaluation practices, Douglas (2000) described evaluation criteria as tending not to be well specified in task rubrics. As a consequence, teachers/raters do not interpret the descriptors unanimously, thus inter-rater and intra-rater reliability is a continuous issue to be concerned with in assessment. Within the large area of reliability the following issues will be examined: (1) the relationship between the content and organization subscales; (2) coherence, cohesion and perceptions of text quality; and (3) unclear perceptions of cohesion and coherence in assessment.

3.3.4.1 The relationship between the content and organization subscales

Even though content and organization are defined as two distinct characteristics of writing according to most analytical scales, several researchers found evidence for their interconnected nature. Freedman (1979) found that evaluators give more weight to **content** and **organization** than to other rubrics when given a 4-point holistic rating scale to judge essay quality. In order to prove this, she conducted an experiment in which she manipulated some argumentative essays to make

them better or worse in one of the four characteristics (**content, organization, sentence structure and mechanics**). The rewriting guide contained strict ‘rules’ as how to change the essays. To make organization weaker for instance, the rule was: “Delete inter- and intraparagraph transitions; vary the lexical items chosen for the key words and avoid using transition words and phrases appropriately (p. 330).” Freedman found that essays rewritten to be strong in content were rated significantly higher than the ones rewritten to be weak in content. The second most important contributor for perceived quality was the strength of organization. In other words, the findings might be interpreted that these two features have a ‘halo effect’ on the whole quality of essays. The other interesting finding of the same research is that Content and Organization are not only weighted more but are also interconnected. This resembles Spivey and King’s findings (1989) who experienced a connection of comprehending content and sensitivity to text structure as two major indicators of good discourse synthesizer readers.

3.3.4.2 Coherence, cohesion and perceptions of text quality

Whether evaluating essays holistically or scoring them against detailed performance descriptors (rubrics), besides content, **coherence** and **cohesion** will be considered as two of the most important features of perceived ‘goodness’ of text. While trying to model expert examiners’ judgement of essay quality, Crossley and McNamara (2010) for instance, found that human judgement of text coherence was the most predictive variable as regards holistic essay scores, explaining no less than 60 percent in their variance. Yet, practice reveals that essay raters’ perceptions of coherence and cohesion are often subjective and more intuitional than based on clear principles, even though they are provided with detailed guides for scoring.

The following example is taken from the training material of IELTS (International English Language Testing Services), an internationally recognized examination, the scores of which are

widely accepted by universities for their entry requirements. The raters are provided with detailed scoring rubrics in order to help them assess test takers' performance in four areas: (1) task achievement, (2) coherence and cohesion, (3) lexical resource, (4) grammatical range and accuracy.

Subscale	Descriptions
Task Achievement (for Task 1) / Task Response (for Task 2)	Candidates are required to formulate and develop a position in relation to the given prompt. Ideas should be supported by evidence. There is a 250 word length criterion.
Coherence and Cohesion	It is concerned with the overall clarity and fluency of the message: how the response organizes and links information, ideas and language. Coherence refers to the linking of ideas through logical sequencing. Cohesion refers to the varied and appropriate use of cohesive devices (e.g., logical connectors, pronouns, and conjunctions) to assist in clarifying the conceptual and referential relationships between and within sentences.
Lexical Resource	It refers to the range of vocabulary used, also the accuracy and appropriacy in context.
Grammatical Range and Accuracy	It refers to grammatical resource as manifested in the candidate's writing at sentence level.

Table 14. Subscales and descriptions of the IELTS Writing Module (based on Cambridge IELTS webinar 26. February, 2014)

It is interesting to observe how the two concepts: coherence and cohesion are described in this guide for practical purposes. According to the description, cohesion manifests on the surface level in countable explicit elements, and the role of cohesion is to help achieve coherence. Cohesion is believed to work between and within sentences in the form of clarifying ideas: it has a referencing function, as well as a conceptual function. Coherence, on the other hand, is achieved by the logical sequencing of ideas. The developers of the scales supposedly think about the progression of ideas in the text created by semantic linkages. Coherence is achieved if the interaction between the reader's understanding of text and the text is clear.

The subscale Lexical Resource is also worth attention from the angle of cohesion, more specifically from the angle of lexical cohesion. Lexical Resource refers to the range of vocabulary used, also the accuracy and appropriacy in context. Neither in this subscale, nor in the previous Coherence and Cohesion rubrics appears the concept of lexical cohesion as a device for either (1) a text organizing device or (2) a semantic means to contribute to the logical flow of ideas. However, ‘range’ can also mean a variety of semantic devices, such as repetition, synonym, superordinates, etc., implying that if the writer is resourceful, he will use a range of vocabulary.

3.3.4.3 *Unclear perceptions of cohesion and coherence in assessment*

In order to understand how essay raters form their opinions on coherence and cohesion, as an illustration, some comments of experienced examiners are shown here taken from the same IELTS training material. In each comment my assumption, whether the particular section of the comment refers to coherence, cohesion, or any other of the four criteria, will be in brackets.

(1) There is a good variety of cohesive devices (cohesion) and the message can be followed quite easily (coherence).

This is a positive comment referring to both cohesion and coherence. It is not entirely clear however, whether the text could be followed easily **because of** the appropriate use of cohesive devices, or the two features strengthened each other without a causal relation.

(2) The answer flows quite smoothly (coherence) although connectives are overused or inappropriate (cohesion) and some of the points do not link up well (coherence? cohesion?).

This comment sheds no more light on the coherence/cohesion dichotomy. In teachers’ jargon “*flows quite smoothly*” is traditionally synonymous with good topic progression. What is interesting about this comment is that the examiner positively rated the overall coherence of the text in spite of acknowledging the problems with the connective devices. It is unclear where this

perceived smoothness came from, especially after reading the obscure “some of the points do not link up well” part of the comment. “*Linking up*” might refer to either cohesion or coherence, or both: problems with linking devices causing a cohesion problem, and/or illogical presentation of ideas, perhaps contradictions. In either way, the smooth flow of the text most probably originated from its semantic features, more specifically from its lexical cohesion. As a separate rubric is assigned for Lexical Resources in the IELTS writing test, lexical cohesion was probably not considered as a cohesive device by the examiner.

(3) *The answer can be followed* (coherence) *although quite repetitive* (coherence? cohesion? lexical resource?) *and cohesive devices are overused* (cohesion).

This comment states that the examiner was able to understand the writing (*answer can be followed*). However, the assessed text showed signs of repetitiveness, which might be interpreted in several ways. Repetitions might have appeared on the lexical or phrasal level, such as the same word or the same phrase repeated (rather a Lexical Resource problem according to the IELTS rubric), but it could also mean that there were problems on a higher, discoursal level. For instance, the topical progression might have been too slow because stretches of text were repeated, and perhaps too many cohesive devices were used, or perhaps only a few were used but in a repetitive manner. The fact that the rater commented on repetition in connection with text flow clearly indicates the important role lexical repetition plays in textual organization.

(4) *There are quite a lot of ideas* (task achievement/response, i.e., content) *and while some of these are supported better than others* (task achievement), *there is an overall coherence to the answer* (coherence). *Organizational devices are evident* (cohesion) *although some areas of the answer become (sic) unclear* (task achievement? coherence? lexical resource? grammar resource?) *and would benefit from more accurate use of connectives* (cohesion).

This comment is an example of how even experienced examiners use the given rubrics as a starting point, but while interpreting their scores, justify their judgements from a holistic viewpoint. This is not only shown by the fact that seemingly contradictory content and coherence judgements are put into one sentence, but in the case of a negative comment, it seems impossible to select one aspect which determines “*unclearness*”. Lack of clarity might result from content problems, lexical errors, sentence structure problems as well as problems with cohesive devices; or all of these. Therefore, negative comments that involve judgement of cohesion and/or coherence are more difficult to interpret, showing the complex nature of texts and the complexity of text quality assessment.

Rather surprisingly, similarly to the table describing the four subscales (Table 14, previously), none of these comments mention lexical cohesion, even when they refer to connective devices (such as *moreover*, *despite*). This might mean that teachers are unaware of what role lexical cohesion plays in the text; they only notice errors of cohesion when the writer violates some rules related to it.

3.3.5 Text quality requirements by course teachers

Given that university teachers’ quality judgement of students’ texts depends on various factors, it is important to find out how conscious they are of coherence, cohesion and lexical cohesion. In Moore and Morton’s (1999) study an Economics teacher (who can be regarded a “layperson” concerning English language teaching) gave an account of how (s)he perceives writing quality:

“I’m asking them to write a coherent piece of work, not a set of dot points or scattered ideas. It is essential that they construct the arguments that they present with examples and

relevant evidence. The tasks that I set usually ask them to compare and contrast, do you agree or disagree, to what extent is this statement relevant, or I have a quote, do you agree.”

These requirements are can be interpreted in the following way for an EAP teacher, see

Table 15.

...university teacher’s requirement (based on Moore & Morton, 1999)	Researcher’s comment: “translated” task for EAP teacher, what to teach to students
<i>...I’m asking them to write a coherent piece of work...</i>	coherence in text, flow of ideas, staying on-topic, using surface cohesion features (coherence/cohesion)
<i>...not a set of dot points or scattered ideas...</i>	organizing ideas, structuring text
<i>...it is essential that they construct the arguments...</i>	argumentative writing, thesis statement, topic sentences
<i>...that they present with examples and relevant evidence...</i>	paragraph structure: topic sentence, supporting evidence, citing sources (quoting/paraphrasing/summarizing, referencing styles)
<i>...usually ask them to compare and contrast...</i>	compare/contrast essay type
<i>...do you agree or disagree... I have a quote, do you agree</i>	argumentative essay type, understanding author’s point of view, identifying conflicting sources
<i>...to what extent is this statement relevant...</i>	descriptive, argumentative essay type, presenting both sides of argument

Table 15. University teacher’s requirement (based on Moore & Morton, 1999) analyzed from a language pedagogy angle

Coherence is mentioned explicitly and cohesion is implied in the requirements, whereas lexical cohesion and repetition are not referred to. It is another proof that they remain unnoticed until the rules are broken.

3.3.6 Explicit instruction on coherence, cohesion and lexical repetition in higher education

The concepts of coherence, cohesion and lexical repetition are often mentioned in academic course books as aspects of writing which need caution and attention. As an illustration, Cresswell’s

(2007) hook-and-eye technique is mentioned here, which teaches writers how to self-check textual cohesion in their drafts. The technique, which Creswell credits to Wilkinson (1991), connects words and phrases referring to the same idea by drawing hooks and eyes across sentences. The words can be followed further in a chain-like manner to highlight progression of topic on the discourse level, in overarching paragraphs. The following illustration is an example from Creswell's book (2007, p.59) where he offers strategies for writing up research.

They sit in the back of the room not because they want to but because it was the place designated to them. Invisible barriers that exist in most classrooms divide the room and separate the students. At the front of the room are the "good" students who wait with their hands poised ready to fly into the air at a moment's notice. Slouched down like giant insects caught in educational traps, the athletes and their following occupy the center of the room. Those less sure of themselves and their position within the room sit in the back and around the edge of the student body.

The students seated in the outer circle make up a population whom for a variety of reasons are not succeeding in the American public education system. They have always been part of the student population. In the past they have been called disadvantaged, low achieving, retards, impoverished, laggards and a variety of other titles (Cubar, 1989; Presselsen, 1988). Today they are called students-at-risk. Their faces are changing and in urban settings their numbers are growing (Hodgkinson, 1985).

In the past eight years there has been an unprecedented amount of research on the need for excellence in education and the at-risk student. In 1983 the government released a document entitled *A Nation At-Risk* that identified problems within the American education system and called for major reform. Much of the early reform focused on more vigorous courses of study and higher standards of student achievement (Barber, 1987). In the midst of attention to excellence, it became apparent the needs of the marginal student were not being met. The question of what it would take to guarantee that all students have a fair chance at a quality education was receiving little attention (Hamilton, 1987; Toch, 1984). As the push for excellence in education increased, the needs of the at-risk student became more apparent.

Figure 15. The hook-and-eye technique connecting major thoughts on discourse level (in Creswell, (2007, p. 59)

Creswell (2007) describes this text as highly coherent because these connections can easily be made. He suggests that it is not necessary for this writer to add transitional phrases here to

establish a clearer logical sequence. Even though he does not mention lexical repetition as such, the chain of reiterations, synonyms, superordinate terms used with pronouns clearly show the distinctive discourse organizational role of repetition.

Another of Creswell's (2007) suggestions is the use of consistent terms throughout the text. He also adds that it is advisable to refrain from using synonyms for these terms. It is important to note that this is a more useful recommendation than the rather simplistic advice for students in various other guides to avoid repetition in general. He also mentions staging of new information and foreshadowing ideas as building coherence and readability of texts.

3.3.7 Automated assessment of text quality

The increasing number of test takers and the demand for fast and reliable feedback has given rise to automated essay quality assessment, particularly in US higher education. Due to the fact that text quality is influenced by many factors, these assessment systems examine multiple features that correspond to various properties of texts. The first fully automated essay quality assessment applications examined texts as mainly quantifiable entities. They computed basic features such as text length, sentence length, word length, frequencies of items; the next step was examining ratios of items, and sentence-level features. Qualitative analysis of these features followed, particularly the classification of words based on psycholinguistics and corpus studies: classifying words according to age of acquisition (K1/K2/AWL) facilitated by the appearance of word lists helped assessment because these new findings were incorporated into the increasingly more sophisticated assessment systems.

Complex automated scoring systems (e.g., E-rater²⁸, Intelligent Essay Assessor²⁹) have been designed to assist in the evaluation of essays, particularly for standardizing large-scale exams. Such programs are used to complement a rater's judgement, in other words instead of two raters, a human and a program evaluates the essays. Research suggests that the inter-rater reliability for these programs is rather high (even though many of such reliability assessments are carried out by the developers themselves). A rarer case is when human judgement is entirely left out of the process.

At the moment, automated scoring systems are unable to evaluate a wide variety of texts due to the number of task variables in writing. Instead, they can judge only one or two specific types of essays. For instance, E-rater can be applied to texts which discuss an issue or requires students to analyze an argument. In order to be able to identify main features of such essays, the program first has to be trained, i.e., 'fed' by essays previously scored by expert raters. This way the program can learn which features were scored high and which were scored low. This process has to be repeated for each essay prompt because different essay titles require a different range of vocabulary. Program developers have already identified more than a hundred important variables in essays which influence text quality. In practice, regression analysis is used to select those with the most predictive power.

E-rater v2 focuses on five key areas in writing: (1) grammar, usage, mechanics, and style; (2) organization and development; (3) topical analysis; (4) word complexity; and (5) essay length (Ben-Simon & Bennett, 2007; Quinlan, Higgins, & Wolff, 2009). See Appendix I for more details of assessed features and microfeatures. The program is designed to give weight to features in essays resembling the judgement of human raters, and it quantitatively analyzes these areas. For instance, it counts the rhetorical features important to show logical order in text, (e.g., *firstly*,

²⁸ <https://www.ets.org/erater/about>

²⁹ http://www.lt-world.org/kb/ipr-and-products/products/obj_80512

additionally), as well as phrases connected to expressing opinions or presenting an argument (*certainly, possibly, particularly*). (For a comprehensive introduction of current approaches in Automated Essay Scoring (AES) applications, see Dikli, 2006)

3.3.8 Controversial views on the automated assessment of essay quality

Teachers are also increasingly aware of the “comprehensive introduction of technology” (Chapelle & Douglas, 2006, p. 1) to language assessment. Besides reports published by developers introducing their constantly improving applications and rather sparingly describing their methods, there are several studies carried out on human assessors’ beliefs and experiences with these systems. There is a clear discrepancy between the two concerning the usefulness of automated essay scoring software.

Teachers have mixed feelings about automated assessment. This is to some extent due to the fact that teachers have difficulties understanding the jargon used in published materials on the subject. Those who do not embrace the idea of computer-aided scoring, form anxious, or even angry opinions, such as the petition against machine scoring of essays entitled “Professionals Against Machine Scoring Of Student Essays In High-Stakes Assessment”³⁰, which can be found in Appendix J. The concise and heavily research-based petition, even though it explains the limitations of automated essay correction, also depicts several misunderstandings and non-understandings regarding the purpose and mechanisms of automated essay scoring.

Firstly, teachers wrongly identify the purpose of machine scoring: the programs are not designed to leave teachers out of the scoring process entirely, because their quality judgement serves as a model for the computer. Thus, teachers are necessary at the preparation stage. What computers can recognize depends on the initial essays the computer is trained on. Secondly, to our

³⁰ http://humanreaders.org/petition/research_findings.htm

present knowledge, it is unrealistic to require the application to recognize a nonsense paper submitted to the machine in order to deceive it, as described in the petition as a key argument against machine scoring.

Furthermore, as regards the mechanisms of automated essay scoring, the petition does not distinguish between machine model and machine algorithm. The next misunderstanding is that reductivity is not as simple and straightforward as the teachers manifesto suggests: AES models do not look for features one by one and separately give scores for each, but they find sequences of features (in a linear regression) and if they have found all the sequences necessary, they assign a score (based on human raters' score).³¹

The above example clearly shows that language teachers need a solid understanding of all aspects of computer-aided assessment because they prepare students for the exams and help learners to develop self-assessment strategies (Chapelle & Douglas, 2006). The main purpose of automated essay scoring is not to eliminate human involvement from the assessment, rather it aims to objectivize the process by improving reliability. The versatility of such tools is illustrated by describing some of uses of Coh-Metrix in language teaching.

Coh-Metrix provides textual analysis on over 200 variables of writing such as cohesion, language, and readability. Its modules are capable of analyzing the lexis, the parts of speech, the syntax, as well as above sentence level features of the text. Standard text readability formulas are used in order to scale texts on difficulty by relying on e.g., word length or sentence length. Coh-Metrix can also measure cohesion, world knowledge, and discourse characteristics. In order to observe cohesion, it offers as many as twelve indices for referential cohesion; eight LSA³²

³¹ <http://www.vikparuchuri.com/blog/on-the-automated-scoring-of-essays/>

³² Latent Semantic Analysis, introduction to: <http://www.vikparuchuri.com/blog/on-the-automated-scoring-of-essays/>

components, such as similarity between sentences, and paragraphs; and nine indices for connective phrases. Apart from the example of a Coh-Metrix analysis shown in Appendix G, the various uses of the program are summarized in Table 16.

The various uses of Coh-Metrix in analyzing L2 writing	Focus	Main findings
Crossley, Salsbury, & McNamara (2009)	L2 lexical growth, use of hypernyms, a 1-year longitudinal study	lexical diversity and hypernymic values correlate significantly
Crossley & McNamara (2009)	lexical differences in L1 vs L2 writing	L1 writers: greater lexical associations, more coherent text
Crossley & McNamara (2010)	cohesion, coherence, teachers 'evaluation of writing proficiency	"expert raters evaluate coherence based on the absence of cohesive cues in the essays rather than their presence"
Crossley, Salsbury, McNamara, & Jarvis (2011)	predicting lexical proficiency in L2 texts	"lexical diversity, word hypernymy values and content word frequency explain 44% of the variance of the human evaluations of lexical proficiency"
Brandon, Crossley, McNamara (2012)	paraphrasing by expert writers using condensing	syntactically more complex, less familiar words used
Crossley, McNamara (2012)	cohesion and linguistic sophistication in L2 levels	lexical diversity, word frequency, word meaningfulness, aspect repetition and word familiarity differ
Guo, Crossley, McNamara (2013)	predicting human judgement of writing quality: integrated vs independent task	significantly can predict scores in both: but 2 sets of predictive features
Kyle & Crossley (2014)	lexical sophistication tool	language proficiency can and should be automatically assessed
Crossley & McNamara (2014)	syntactic complexity L2 writing	L2 syntactic complexity grows, but EAP needs different syntactic sets

Table 16. The various uses of Coh-Metrix in analyzing L2 writing

3.4 Summary

The main aim of this section was to introduce the academic context in which typical L2 student writing takes place by exploring the tasks, and their contextual and cognitive variables which influence the quality of written work. Within the varieties of genres, the most typical two (summaries and compare/contrast essays) were analyzed as these are the focuses of this research.

Next, current assessment practices were highlighted, particularly how teachers perceive coherence and cohesion. We have seen that the teachers were struggling in formulating their opinions when they referred to these two concepts, and their treatment of coherence and cohesion was not systematic. Given that lexical cohesion, or lexical repetition were not mentioned, we may conclude that teachers are not aware entirely of how these two concepts work in writing. Thus, coherence, cohesion and lexical repetition may be defined clearly in theory but they are not in their right place in assessment practice. Teachers using manual assessment methods, whether holistic or analytic, often find it problematic to interpret these features and feedback regarding them is rather vaguely worded, confusing students. Teachers need an “objective” tool for assessing such elements.

Finally, computer-aided assessment techniques were described in order to find out what features of text they base their judgement on, and whether they measure cohesion and coherence. It was found that both are present in automated analysis. The quantitative measures used may indicate the differences between high and low-rated essays however, several questions remain unanswered. While cohesion is viewed by these applications in quantitative terms, corpus linguistic techniques do not give answers to questions on discourse organization: cohesion and word repetition is not researched qualitatively. The applications focus on how many of such features a text contains. Discourse theory informed applications would be necessary to find out what is/should be repeated and where in the text. Without this, quality judgements of texts are only

partially justified. At the moment of writing no application is described as being capable of indicating the location of lexical repetition within text. Literature on the subject reveals that no model utilizes either Hoey's (1991) or Károly's (2002) lexical repetition analysis frameworks as a whole.

4 Aims and Research Questions

Three types of research questions guide this study: theoretical, empirical and methodological ones. The main theoretical question guiding this study is whether other academic genres form detectable lexical repetition patterns, or this feature is exclusive to argumentative essays. In order to answer this question, Károly's (2002) LRA (lexical repetition analysis) framework is applied to two academic genres: a summary and a compare/contrast essay corpora.

The empirical questions that follow from the above are: can clear lexical repetition patterns be detected, and can these patterns be connected to essay quality? In other words, the quantitative and qualitative differences between lexical repetition patterns (i.e., their frequencies, types, or locations) do or do not coincide with teachers' writing quality judgements?

The findings in this respect the researcher hopes to gain are limited, because the generalizability of these results would require a large data size. However, due to the laborious nature of the analysis, this cannot be achieved reliably without partial automation. Therefore, the next stage focuses on the design of a computer-aided tool whose modular format will enable future developers to further improve the various features of the tool independently. See Table 17 for a detailed outline.

Research Questions	Methods of data collection	Methods of data analysis
<p>Stage 1 RQ1: How can Károly's (2002) theory-based lexical repetition analysis model be applied to the study of summaries written by EFL university students?</p> <p>.....</p> <p>Stage 2 RQ2: How can Károly's (2002) theory-based lexical repetition analysis model be applied to the study of compare/contrast academic essays written by EFL university students?</p> <p>RQ3: How can <i>Concordance/Version 3.3</i> software facilitate the analysis of the text-organizing function of lexical repetition patterns?</p>	<p>Stage 1 initial corpus collection rated in 4 modes holistic rating analytical rating1 analytical rating2 automated textual analysis (<i>Coh Metrix</i>) final corpus collection coding steps of Károly's LRA</p> <p>Stage 2 corpus collection coding steps of Károly's LRA preparation of corpus for computerized research highlighting errors of mechanics and spelling in learners' text highlighting spelling mistakes and other linguistic features incompatible with applications</p>	<p>Stages 1 and 2 descriptive analyses of the expert judgement's results qualitative and quantitative discourse analysis using Károly's (2002) lexical repetition analysis framework, Appendix C</p> <p>Stage 2 only frequency/concordance analysis using <i>Concordance/Version 3.3</i> software</p>
<p>Stage 3 RQ4: How can Károly's (2002) theory-based lexical repetition model be applied to large corpora?</p> <p>SubRQ 1: How should the learner corpora be prepared prior to the computer-based analysis?</p> <p>SubRQ 2: What computer-aided applications enable the analysis of the text-organizing role of lexical repetition in large corpora?</p> <p>SubRQ 3: What theoretical/ methodological changes are necessary in the existing framework and analytical steps for large-scale application?</p>	<p>Stage 3 highlighting linguistic features incompatible with applications</p> <p>.....</p> <p>experimenting with alternative applications using results from Stage 2</p>	<p>Stage 3 descriptive analyses of errors of mechanics and spelling in learners' text descriptive analyses of inconsistencies reading previous research on alternative applications comparing the existing taxonomy, and analytical steps with applications' taxonomies and analytical steps</p>

Table 17. Summary of the intended approach of data collection and analysis

In the first phase, Károly's (2002) framework is applied to the texts entirely manually, as it was originally developed, whereas in the second phase the initial steps of the analysis are assisted by a concordance program. These two phases employ a discourse analytic approach which is described by Graesser, McNamara, and Louwse (2011) as working with "theoretical generalizations based on empirical evidence observing a small corpus" (p. 37).

The operationalized research questions are:

Stage 1: RQ1: How can Károly's (2002) theory-based repetition analysis model be applied to the study of summaries written by Hungarian EFL university students?

Stage 2: RQ2: How can Károly's (2002) theory-based lexical repetition analysis model be applied to the study of compare/contrast academic essays written by EFL university students?

RQ3: How can *Concordance/ Version 3.3* software facilitate the analysis of the text-organizing function of lexical repetition patterns?

In the third phase, the design of a computer-aided tool is attempted, in order to facilitate a large-scale lexical repetition analysis. The research questions to guide the investigation are:

Stage 3: RQ4: How can Károly's (2002) theory-based lexical repetition model be applied to large EFL learner corpora?

SubRQ1: How should the learner corpora be prepared prior to the computer-based analysis?

SubRQ2: What computer-aided applications enable the analysis of the text-organizing of lexical repetition in large corpora?

SubRQ3: What theoretical/methodological changes are necessary in the existing framework and analytical steps for large-scale application?

The third stage is theoretical and methodological, based on the findings of the first two stages. The three subquestions aim at exploring two aspects of computerized analysis, namely that the EFL learners' texts should be suitable for machine reading, and the applications to be used should be suitable for analyzing the role which lexical repetition plays in discourse.

This dissertation aims to create the theoretical basis for the design of a lexical repetition analysis (LRA) tool. It does not aim to physically build the actual software modules. Once such modules are built, the results of both Stages 1 and 2 need to be reconfirmed on a large corpus in future research, and the methods and modules can then be fine-tuned based on this future application.

5 Research design and procedures of analysis

5.1 A sequential mixed design

This study follows a mixed-approach design, consisting of several methods of inquiry, as recommended by Creswell (2007) among others. The approach of data collection and analysis is based on theoretical and research methodological assumptions drawn from the literature reviewed, which implies that written discourse is a complex phenomenon therefore, textual analysis should involve quantitative and qualitative elements (Tyler, 1995; Károly 2002). Due to the exploratory nature of such studies, the aim is not hypothesis testing, it is rather hypothesis creation.

Furthermore, according to the sequential mixed design paradigm (as described by Tashakkori and Teddlie, 2003) data collected and analyzed from one phase will be used to inform the other phase. Sequential mixed designs, as Chen (2006) points out, are especially desirable in theory driven evaluations when contextual information is collected to help the interpretation of findings in the previous phases.

5.2 Stage 1: Analysis of academic summaries

5.2.1 The summary writing task

The initial corpus consisted of 35 summaries collected from a group of second year English major BA students at Eötvös Loránd University in 2011. The summaries were collected in an electronic form. Spelling and punctuation mistakes were corrected only for automatic lexical analysis purposes. It has to be noted that the summaries contained hardly any mistakes of spelling and mechanics because they were supposed to be submitted following a spelling check required by the course teacher.

In the task the students were instructed to summarize Ryanair's history and business model, highlighting its strengths and weaknesses. The constructs in this summary task were (1) company history, (2a) strengths of the business model, (2b) weaknesses of the business model. According to summary writing rules, the students had to select the relevant information from the source text and restructure this information under the three labels.

The original summary task was first piloted in 2011 (Adorján, 2011, 2013) and revised the next year, motivated by the constructive criticisms of Gyula Tankó, for which I am grateful. The main changes affected the wording of the instruction which, as a consequence, affected the construct. In the original task, more guiding prompts were given to the students, and also certain key business terms were suggested for inclusion. This was partly due to the fact that an authentic summarization task, which originated from an ESP Tourism course, was used as the basis for this research.

This guidance was criticized to be superfluous and rather confusing for the students because the prompts mixed up the construct, thus interfered with the reading and macroproposition extraction processes. Thus, the instruction which asked students to use certain words and expressions, which was a valid prompt in the ESP class, was deleted. The new guide simplified the students' task, making the construct clearer.

Task instruction	Input	Output
<i>“Summarize Ryanair’s history and business model, highlighting its strengths and weaknesses.”</i>	single text, cca 7405 words Wikipedia entry (electronic genre), http://www.wikipedia.org/Ryanair	max 500 words

Table 18. Basic features of the summary task

5.2.2 Corpus size and representativity

The theoretical, empirical and methodological results of this study are based on data collected from a small corpus. For corpus, we use Atkins, Clear and Ostler's (1992) definition which states that "a corpus is a body of text assembled according to explicit design criteria for a specific purpose" (1992, p. 5). The corpus collected for this research contained 3500 words. This data size, due to the complex nature of manual analysis, was similar to that of international and Hungarian empirical studies in this specialized field of discourse analysis. Flowerdew draws attention to the fact that analyzing small corpora is useful because it can inform pedagogy. The teacher can take the role of a "mediating specialist informant" (2005, p. 329), by using the data gained from the corpus in the classroom.

The issue of representativity of the texts collected is a key consideration for this study. The corpus is built to match the research questions. The summary corpus was selected as being representative of 'real-life' academic written discourse, in other words, they reflect the disciplinary environment. In contrast with timed essay writing often used for testing purposes, the summaries in question were written as home assignments without time constraints. Given that academic writing is a slow process which comprises a considerable amount of reading, planning, editing, and redrafting, it is our strong belief that such a corpus is better able to reflect students' summarization skills, particularly regarding organization and lexical repetition patterns. Advocates of the alternative assessment movement similarly suggest that instead of in-class time-bound exams, writing tests should be altered to resemble more the actual processes involved in academic writing (Alderson & Banerjee, 2001).

5.2.3 Context validity evaluation of the summary writing task

Features of task setting	Feature evaluation
Clarity of purpose	clear
Intended audience	clear
Knowledge of marking criteria	clear
Topic domain	professional (business/finance)
Genre	academic summary
Cognitive demands	summarizing/organizing given ideas
Language functions to perform	summarizing, classifying

Table 19. The features of task setting in Stage 1.

5.2.3.1 *Clarity of purpose and intended audience*

The summary writing task was an authentic “real-life” task collected from a Tourism English ESP (English for Specific Purposes) course offered at the university to English major students. Therefore, the summarizing assignment was a real course task. The Tourism English ESP course put an emphasis on employability skills, providing situative exercises, simulations and tasks that will be performed in working situations later. This was reflected in the wording of the summarization task. The instruction placed the task into a clear social framework where students were aware of their own role (employee in the tourism sector), the expected audience (their boss) and the purpose for writing (summarizing a “best practice” in order to improve their company).

5.2.3.2 *Knowledge of marking criteria*

The students were aware of how they would be scored by their course teacher when the original task was set. At the time of writing they were not aware that their texts would be used for research purposes therefore this did not affect their performance. (Their permission was asked later for the use of their texts anonymously.) The students were reminded to use the general rules of academic writing. They were also told to submit their summaries electronically.

5.2.3.3 *Topic / domain and genre*

The topic / domain was business/financial, more explicitly, touristic. The students were familiar with the topic and the summary genre. Students were familiar with the conventions of the academic summary genre because they attended an Academic Writing course at the university the previous year³³.

5.2.3.4 *Cognitive demands and language functions*

Three variables needed control in this category: (a) cognitive demands created by the input text (e.g., understanding vocabulary in the text); (b) cognitive demands due to the variety of rhetorical functions students had to perform during the summarization task (e.g., selecting main ideas, deleting unnecessary information); and (c) cognitive differences among students.

The language level of the input text is of major concern when reading-to-write tasks are used for testing language proficiency, e.g., measuring reading comprehension ability. The language level of this input text was deliberately chosen to be lower than the language level of the students, therefore understanding it was not cognitively demanding for the students. (The analysis of the linguistic features of the source text is detailed in the next section.) The only feature of the input that might be regarded as demanding was its length, therefore it seemed suitable to assign it as home work instead of an in-class test.

The students had to apply three types of reading: skimming, scanning and close reading, a main subtask was also monitoring for standard business English. The students did not have to read the whole text with close reading, only looking for the information prompted by the instruction.

³³ It is interesting to note that teachers claim students “forget to apply” these conventions in their content courses once their EAP courses are over. This is partly confirmed by the results of this study, although the rhetorical functions necessary to perform a summarization task were known and practised by students in their previous Academic Writing course.

They were familiar with van Dijk and Kintsch's (1983) summarization "rules" (see Section 3.2.4 for details).

Cognitive differences among students was the third factor which had to be minimized so as not to distort the cognitive validity construct. Therefore, for this research on summaries a task with a prompt was chosen. As we did not want to measure the students' summarizing abilities or language comprehension, it was of no concern to find out whether the students were able to select the main ideas in the text or not. Instead, the research focus was on observing textual organization of their texts. In other words, we wanted to observe the patterns of lexical repetition visible in the writers' texts: where in the summaries main ideas are placed and how different lexical patterning creates quality differences in cohesion. Therefore, a prompt was felt necessary to facilitate finding the information selected to be important from the text.

5.2.4 Features of the input text

Features of input text	Features evaluation
Input format	single text, 7405 words
Verbal/ Non-verbal input genre	Wikipedia entry (electronic genre), http://www.wikipedia.org/Ryanair
Discourse mode	mostly descriptive and in parts analytical, narrativity 24% (Coh-Metrix)
Concreteness/Abstractness of ideas	Word concreteness 54% (Coh-Metrix)
Explicitness of textual organization	less explicit than usual, not academic paragraph format
Cultural specificity	international topic
Linguistic complexity:	
Lexical complexity	Syntactic simplicity 61%, rather simple text
Syntactic complexity	Referential cohesion 12%, deep cohesion
Degree of cohesion	84%, connective words present, added support (Coh-Metrix)

Table 20. Features of the input text for Stage 1 summary task

The input text was an Internet source, a Wikipedia entry (electronic genre) describing the history of a company. The students were familiar with the structure of the format and genre of such

entries. The length of the text³⁴ was 7405 words, however, it was perceived shorter by the two teachers who took part in the analysis. This might be due to the fact that the text was web-based and appeared in a continuous flow. Keck (2014) found that L2 students who are novice academic writers are more prone to direct copying from the input text than their more experienced peers. This finding reinforced the choice for a lengthy source text.

The features of the input text were assessed partly by observation, partly by utilizing the Coh-Metrix (McNamara, Louwerson, Cai, & Graesser, 2005) textual analysis program, Text Easibility Assessor measures. It analyzed five features of the input text: Narrativity, Syntactic Simplicity, Word Concreteness, Referential Cohesion, and Deep Cohesion. The results in percentile scores can be seen in Table 20.

The text was mostly descriptive and in parts analytical. A narrative text was to be avoided (Hoey, 1991) because such texts do not exhibit clear lexical repetition patterns similar to the one observed by Hoey. Other features, such as linguistic complexity were measured by Coh-Metrix, and the results are shown in Appendix G. The interpretation of the index as regards abstractness of lexis is as follows: "Words that score low on the concreteness scale include *protocol* (264) and *difference* (270) compared to *box* (597) and *ball* (615)." (McNamara, Louwerson, Cai, & Graesser, 2005)

When selecting the input text concerning the explicitness of textual organization, the main methodological decision was to use a text with no overt organizational pattern. My main motivation behind this was twofold. Most studies on summary writing use abridged/ simplified texts as input (see e.g., Yu, 2008, 2009). These texts are usually from course books or news popular

³⁴ as accessed from <http://www.wikipedia.org/ryanair>, on 2 February, 2011

science articles, and have a rather clear, didactic organization. This feature of the source texts make it possible for the students to “copy” their organizational pattern.

Segev-Miller (2004) distinguishes between **summary** and **synthesis** writing, depending on the number of source texts. In the case of one source text, students can apply the same structure as the input, (in fact they might feel bound to use the same structure, as an implied requirement of the teacher), thus resorting to selecting the main points of the source text by **knowledge telling** instead of **knowledge transforming** (for the discussion of this see Section 3.2.3.). However, when multiple texts are used for input, a “superproposition” (an integrating idea) needs to be created from sources, each with their own macropropositions. As far as discourse organization is concerned, the writers have to invent a structure depending on their understanding of the superproposition. Therefore, Internet Wikipedia entries seemed a good choice.

The main reason for choosing this text from internet Wikipedia was that the text complied with all the requirements for this research, particularly as its text organization was not clearly visible. Other texts used for summarization in academic settings were simplified versions of original texts resulting in a rather “didactic” structure with too clear paragraph organization, signaled by more than average linking phrases. It was important to find a text devoid of these clear textual marks in an attempt to examine students’ text organizing practices. The second consideration for choosing a text with a non-explicit structure was that according to Li’s (2014) findings, the least frequently employed summary writing strategy was **reorganizing source pattern**, which I wanted to avoid.

5.2.5 Quality assessment of the corpus

5.2.5.1 *Holistic evaluation of summary quality*

The summary task was a ‘real-life’ home assignment for a Tourism English ESP (English for Special Purposes) course. The course teacher corrected the thirty-five summaries according to the original aim of the course and gave a grade from 1-5. Content, organization, language used, and grammar were considered the main features observed. Content was evaluated by the presence or lack of required information, as well as the presence or lack of irrelevant information. For the purposes of this study, the grades were multiplied by four to distinguish better between the low-rated and high-rated groups. The mean of the scores was 16.19 for the higher quality summaries, and 11.5 for the lower quality summaries out of a score of 20.

5.2.5.2 *Content similarity evaluation*

The aim of the next stage of assessment was to identify direct copying from the source text within students’ summaries. Two methods of identification are provided in the literature for the assessment of academic summaries in this respect. The first one is a content similarity evaluation based on Keck’s (2006) model. Keck classified students’ paraphrasing attempts into four groups (Near Copy, Minimal Revision, Moderate Revision, and Substantial Revision) for her L2 summary corpus. However, this type of deep manual analysis was not possible in the case of this summary corpus because the input text was considerably longer than in Keck’s study. The second method is Friend’s (2001) Paraphrase and Integration identification by classifying T-units taken from the source text using a coding scheme. (The difference between paraphrase and integration lies in the length of source text used: when paraphrasing, writers rephrase text within one paragraph; while integrating, writers use a longer stretch of text than a paragraph.)

Both Keck's (2006) and Friend's (2001) methods assume that the input text is short, thus it can be easily compared to the output text. In the case of this task, similarly to other source-based 'real-life' tasks, the input text did not make it possible to use either method. Another important difference was that the domain of the input text was business/financial, whereas in both mentioned studies the input texts were general popular scientific articles. No previous studies were found describing the acceptable length of original text within sentences in summaries concerning texts from the business domain, where certain key phrases need to be repeated instead of synonymized.

Therefore, as a third option, an experiment was carried out with an online text similarity tool³⁵. This tool can compare two texts by looking for exact subsequent word-for-word matches, and gives a percentage score to indicate text similarity. In order to be able to interpret the percentage score of the tool, the following testing method was used. A random section was selected from the input text (Wikipedia/ Ryanair/Controversial advertising). This section of 667 words can be considered representative of the source text because it is located in the middle section, after the introduction and before the conclusion part.

The selected section was copied into the "First string" slot of the test similarity tool, while four randomly selected complex sentences of the same section (with 73 words) were copied into the "Second string" slot, simulating the direct copying of four sentences from the original document. The resulting similarity index was 18.16 percent as a strings match score. The interpretation of this score is that if a writer copies out approximately 10 percent of the original text (73 words in four sentence-strings in this case), it will result in a cca.18 percent similarity score.

³⁵ www.tools4noobs.com/online_tools/string_similarity

Because the four sentences chosen were exact copies, another test had to be carried out concerning the connection between one short original section and its paraphrased version in the summary. The question was now whether the similarity score would be higher or lower in the case of paraphrasing. The two sections compared were the following:

Original section: *Ryanair operates around 300 Boeing 737-800 aircraft. The airline has been characterized by its rapid expansion, a result of the deregulation of the aviation industry in Europe in 1997 and the success of its low-cost business model. Ryanair's route network serves 28 countries in Europe and also Morocco.*

Paraphrased version: *Ryanair has more than 300 airplanes which fly 28 countries in Europe and Morocco. It is expanding rapidly due to the deregulation of the aviation industry and the success of the low-cost model.*

The strings match this time was 54.44 percent. This surprisingly high number of exact strings indicate that the online tool used here cannot discriminate between acceptable and unacceptable paraphrases, only capable of identifying what percentage of the original text was used. Given that university instructors interpret a score of more than cca.15 percent text similarity score in Turnitin³⁶ (a widely used plagiarism detecting application) unacceptable, my decision for this study was to allow less than 15 percent similarity score as detected by this online tool when the *whole* summary is compared to the *whole* source text. All the summaries of the initial corpus were checked for text similarity. None of them scored above 15 percent, the mean similarity score being 14 percent. Thus, based on direct copying none of the summaries were excluded from the corpus.

³⁶ <http://turnitin.com/>

5.2.5.3 *Main idea collection*

Besides the holistic evaluation of the summaries, for the purposes of this study, analytical evaluation of the content and the organization were also necessary. To address the former issue, the next procedure in the corpus quality assessment was to collect the ideas the writers had to include in the summaries according to the prompt. As it was mentioned in Section 3.2.4.4, no previous research into academic summary writing describes using computer-aided summarization as an option for this daunting task for test creators. In Garner's (1985) text summarization study, first "idea units" were identified in the source text. Next, the level of importance within these "idea units" were ranked by five raters using 3 scores: 3 = important idea, to be included in the summary; 2 = of medium importance, could or could not be included; 1 = not important, omit if summarizing text. According to the average values given by raters, 2.5-3 were very important, 1.5-2.4 were rather important and ideas 1.4 and below were considered not essential. This ranking formed the objective basis of student text judgement.

This method was modified and simplified for the purposes of this study informed by Yu's (2008) main idea collection method in the following way: two native speakers were given the task description and the source text. They were asked to read the texts and select the main ideas from the source text based on the summary construct. Their decisions were discussed by the researcher and compared to the source text. When agreement was reached about what the main ideas were, these were evaluated with a similar scheme as in Garner (1985), with two exceptions. The first difference was that the "ideas to omit" category was not used due to the length of the input. The second difference was necessary due to the classification of the main ideas into negative and positive results, as Table 21 indicates.

Main constructs of the summary task	Divisions of main constructs	Sample from the input text
History		<i>O'Leary became head of the company in ...</i>
Business model	strengths (+)	<i>more competitive prices than other companies</i>
	weaknesses (-)	<i>arrogant communication; extra charge for obese passengers</i>

Table 21. The constructs, their divisions, and the cognitive processes involved in the summary task of the research (Stage 1)

This ranking became the basis of content judgement for the students' summaries. Idea units categorized as (2), had to be included, e.g., *O'Leary became head of the company*. Idea units (1) if present, were not considered erroneous, rather additional features, possibly included as supporting ideas, e.g. *extra charge for obese passengers*. If any other points were found in the student summaries, which had not been previously identified as (2) or (1), they were considered irrelevant information thus an error, and were given a (-1). If an idea was incorrectly classified as a strength or weakness, it was given a (-1). The maximum possible score was 26. All the summaries were scored by the researcher with this method.

All summaries above score 13 were considered high-rated. None of the summaries misclassified a weakness for a strength, or vice versa. All the summaries the class teacher judged as high-rated were also found to be good quality with a mean score of 20.8 out of 26. The texts scored low-rated by the holistic evaluation reached 15.4 as a mean value. This means that these summaries also contained many of the main ideas, however they did not include all of them; or included unimportant details which resulted in taking off some scores. In most cases, these summaries also reached as high scores as the holistically high-rated texts. The reason for this might

be that the main points in the input were easy to identify. However, there were other features in these texts which lowered their overall scores.

5.2.5.4 Judgement of discourse control

Another type of analytical rating addressed the summaries as pieces of academic discourse. The evaluation was carried out using Hamp-Lyons's (1991) six-band scale from the Michigan Writing Assessment Scoring Guide (see Appendix K). This scoring scheme was already used for summaries by Yu (2009). A positive aspect of this scoring guide is that it separates the organization and discourse control subscales. While the organization rubric contains judgements about clarity of point of view; unity; paragraphing; present thesis/purpose statement; conclusion and effectiveness of topic sentences; the discourse control rubric draws attention to features such as logical coherence; the use of cohesive devices (key words, pronouns, references, transitions, etc.); and the clarity and coherence of the presentation of ideas.

Two experienced teachers were asked to evaluate the summaries, however, they were asked to observe only the discourse control subscale, and grade according to those descriptors in the scoring guide. These two teachers did not read the source text thus, their judgement of quality was based on the texts as self-contained pieces of discourse. This was made possible by the fact that in this scoring guide the organization rubric contains features clearly connected to content, e.g. the clarity of point of view, thesis/purpose statement. These, as well as topic sentence effectiveness can be better judged if, at the same time, the content is judged, whereas the discourse control rubric can gain meaningful data about the text formation quality without having to interpret content-related information. For the purpose of this study, vocabulary, sentence structure, and mechanics were not examined either, because they are variables referring to the "traditional" accuracy and

appropriacy scales and, according to Károly (2002), they are not relevant for investigating the organizing function of lexical repetition in text (p. 127).

The mean discourse control score of the summaries which were considered high-rated holistically was 3.33, whereas in the low-rated group it was 2.78. The teachers' judgement of discourse control rubric coincided with the holistic rating of the class teacher. Most summaries reached 3 for a score with only two summaries getting 5 from the two teachers.

5.2.5.5 Cohesion analysis using Coh-Metrix

Motivated by the successful employment of Coh-Metrix for the easibility assessment of the input text (see 5.2.4, previously), as the last step in the evaluation of the summaries, the tool was used again as a means of a more objective, computerized analysis. The students' summaries were submitted one-by-one into the analyzer, and out of the 106 indices the following were selected as referring to cohesion: Referential cohesion (10 indices), Latent Semantic Analysis (8 indices), Lexical diversity (4 indices), and Connectives (9 indices). A sample text and its analytical results are shown in Appendix L.

Unfortunately, the data gained from this analysis did not prove meaningful in distinguishing the quality of texts. A closer observation revealed that this might be caused by the fact that even though the program analyzed many textual features, the resulting indices showed mostly mean values: they did not give information about what words were repeated and where these repetitions occurred. For example, in the sample text, the index: Noun overlap, adjacent sentences, binary, mean was 0.474 and the Noun overlap, all sentences, binary, mean was 0.283 in a high-rated summary. In another, low-rated summary these were 0.541, and 0.4, respectively. Given that the application presented these results in a table format, without parameters offered, it was impossible to interpret these figures. The previous assessment of the input text, on the other

hand, offered useful information about how easy the input was for students by showing for instance, the readability index Flesch-Kincaid Grade level, because the parameters for the various levels had already been identified.

5.2.5.6 *The final summary corpus*

The thirty-five essays were first divided into high-rated and low-rated according to the teacher's original holistic assessment. This initial assessment was compared to the main idea collection and discourse control judgement analytical assessments. Given that most summaries received a high score in the main idea collection assessment, and the discourse control judgement coincided with the holistic scoring, five-five summaries were randomly selected from each group for the lexical repetition analysis, as the final corpus for Stage 1. The reliability measures concerning the summary assessment in Stage 1 are presented in Table 22.

Type of assessment	Method of assessment	Aim
class teacher evaluated the summaries	holistic, one grade multiplied by 4 to gain a 20point scale.	evaluation of content, organization, language and grammar/mechanics
content similarity evaluation with text similarity program	(a) testing a text similarity online application and (b) checking summaries for direct copying	text originality, how similar text is to the source text, eliminating too similar texts from the corpus to leave only texts in the corpus suitable for lexical repetition analysis
two other teachers who did not read the original text: judgement of discourse control	analytical, checking overt cohesive devices	assessing textual coherence and cohesion, observing texts as self-contained pieces of discourse
automated textual analysis	Coh-Metrix cohesion	using an objective tool to assess cohesion

Table 22. Overview of assessments, methods and their aims in Stage 1

5.2.6 Methods of data analysis in Stage 1

In the following the text preparation, coding, link and bond establishing steps are described. The procedures of the data analysis closely follow those used by Károly (2002), and described in detail in Section 2.4.2. The analytical steps and the list of variables are presented in Appendix C.

5.2.6.1 Preparation of the texts

Given that in Stage 1 the summaries were manually analyzed, the texts did not need much preparation. In each summary the sentences were numbered and written in a new line (to be cut up for printing or to be analyzed more easily on the screen). Paragraph beginnings and endings were annotated (by a simple sign, not using XML).

5.2.6.2 Coding

The analytical framework applied in the study was based on Károly's (2002) revised version of Hoey's (1991) lexical repetition model. Károly's taxonomy of the lexical repetition types is shown in Table 23. First the two coders (the researcher and another coder) analyzed a randomly chosen sample text from the corpus with the help of Table 23, and discussed their coding decisions.

Categories of lexical relations		Examples
I. Same unit repetition		
1. repetition	simple	<i>company – companies</i>
	derived	<i>decisions – decide</i>
II. Different unit repetition		
2. synonymy	simple	<i>company – firm</i>
	derived	<i>lead – managing</i>
3. opposites	simple	<i>small – major</i>
	derived	<i>successful – failure</i>
4. hyponymy		<i>marketing – advertisement</i>
5. meronymy		<i>(hands – fingers)</i>
III. Text-bound relations		
6. instantial relations		<i>manager – O'Leary</i>

Table 23. Summary of the categories of repetition (based on Károly, 2002)

Note: no meronyms were found in the summary corpus

Even with this taxonomy, which can be better operationalized than that of Hoey's (1991), some disagreements occurred about the analytical decisions. The first of these concerned the unit of analysis itself: in a few cases the coders had problems identifying multi-word items as one or two lexical units. One such problematic case was the combination *low-cost airline*. The main argument for considering it as one lexical unit was that the meaning of this expression is already fossilized, it is a special kind of airline. Also, another word cannot come between the adjective and the noun. However, seeing that the source text compared this type of airline and other high-charging airlines, still it did not contain any antonymous fossilized expressions, only *other airlines*, the decision was that in this particular case low-cost airline will be considered as two lexical units.

Categorizing lexical connections containing the company name (Ryanair – company or Ryanair – airline) also required consideration. According to the taxonomy, these lexical connections should have been categorized as instantial relations. However, it became clear during the discussion of the analysis that these lexical variations were used as synonyms throughout the summaries. The reason for this was that the subject was a specific agent (Ryanair) and the available synonyms to be used instead of the proper name were limited (airline, company, firm, carrier). Therefore it seemed more appropriate to classify the *company name – synonym* repetition links as simple synonyms. The argument behind this was that instantial relations by the original definition (Halliday & Hasan, 1976) were used in texts to serve as occasional text-based connections, not a central idea of the text. Here, on the other hand, the topic of the text was the company name.

The classification of other lexical units, such as non-derivational differences of the same unit were counted as simple repetition, whereas zero derivation was counted as derived repetition. The differences between identifying derivational and non-derivational cases were resolved but slowed down the coding process.

According to Károly (2002), lexical relations loosen from same unit repetition being the strongest, to instantial relations being the most distant semantically (see also Halliday & Hasan, 1976). Therefore following Károly's method of analysis, if a lexical unit could fall into two different categories, the closer relation was recognized.

After the above described disagreements were resolved and analytical decisions were recorded, the two coders continued the classification procedure individually for each summary. Finally, the researcher compared the two sets of coded texts. The further steps of the data analysis in this phase were as follows: a coding matrix was created for both coders, in which cells represented possible links between sentences. The title also counted as a sentence (was marked as zero). With the help of this matrix the number of cells was calculated. The links were each written into the appropriate cell of the matrix, also indicating their classification from the taxonomy with capitalized abbreviation in Figure 16 such as SR, SS, etc. It indicates part of a repetition matrix for Text 3, as an example, with the repetition links itemized and classified among the title (0) and the first three sentences.

		0	
S1	Ryanair- ~ SR		S1
S2		airline- ~ SR, company- carrier SS, low-cost - ~ SR	S2
S3	Ryanair- ~ SR	Ryanair - ~ SR, runs -operates SS	airline- Ryanair SS

Figure 16. A detail of the repetition matrix of Text 3, itemized and classified (Abbreviations: SR: simple repetition, SS: simple synonym, 0: the title, S1, S2: sentences)

At this point the two coders exchanged matrices and again compared their analytical results which were now in visual format. Due to the initial standardization of the coding, at this point no major theoretical problems were detected. Besides minor differences, the coding decisions showed

much similarity. The reason for this could be that the summaries did not differ in their lexis considerably: they were based on the same source texts and the vocabulary comprised mainly from the business/tourism domain. (This was noted by the researcher as a caution to be avoided and allow for a wider variety of lexis for the second stage of the research.) However, there were several occasions when both coders missed links between sentence pairs. These were all recorded on each final matrix. From this point onwards, the researcher continued the analysis alone. In order to gain numerical data, another matrix was drawn, indicating the number of links counted in each cell. Cells with three or more links were highlighted. These sentences are called bonded, and they are central to the text. Figure 17 shows the same text as in Figure 16, however here the whole text is illustrated.

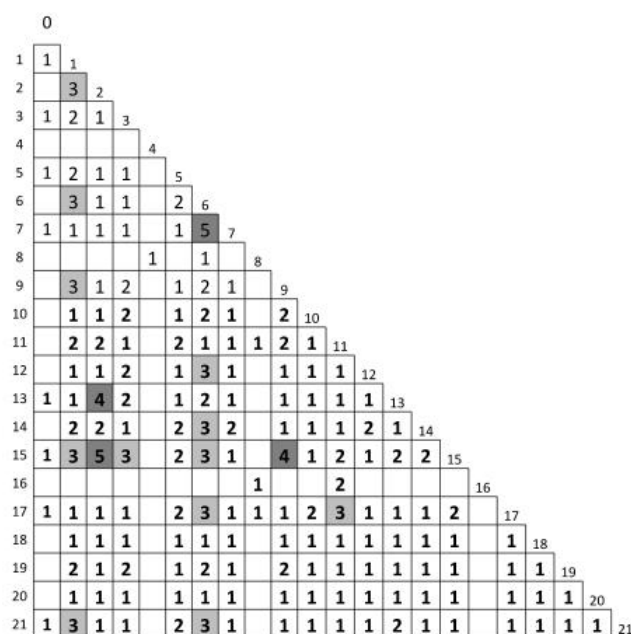


Figure 17. The matrix showing the number of links between each sentence of Text 3.

A chart was also created next to indicate the position of and directions of bonds. Table 24 shows the number of bonds pointing backward and forward within Text 3. In brackets there are the number of links which create the bonds, in other words, the strength of bonds. It can be seen, for

instance, that Sentence 1, which is indicated in 0;5, has no bonds with the title, however, it has five bonds with other sentences which are detailed in the last column: Sentences 2, 6, 9, 15, and 21. All these bonds contain three links.

Sentence	Number of bonds pointing backward and forward	The bonded sentences (No. of links in brackets)
0 (title)	- ; 0	
1	0 ; 5	1-2 (3)
2	1 ; 2	1-6 (3)
3	0 ; 1	1-9 (3)
4	0 ; 0	1-15 (3)
5	0 ; 0	1-21 (3)
6	1 ; 6	2-13 (4)
7	1 ; 0	2-15 (3)
8	0 ; 0	3-15 (3)
9	1 ; 1	6-7 (5)
10	0 ; 0	6-12 (3)
11	0 ; 1	6-14 (3)
12	1 ; 0	6-15 (3)
13	1 ; 0	6-17 (3)
14	1 ; 0	6-21 (3)
15	5 ; 0	9-15 (4)
16	0 ; 0	11-17 (3)
17	2 ; 0	
18	0 ; 0	
19	0 ; 0	
20	0 ; 0	
21	2 ; -	

Table 24. The number of bonds pointing backward and forward within Text 3.

The span of bonds and the cumulative bond span as illustrated in Figure 18. It indicates the bond span in Text 3. The first two numbers (e.g., 1-2) show the two sentences connected by bonds, and the number in brackets indicates the number of bonds the sentences are connected by. The shortest span involves two adjacent sentences (e.g., sentences 1-2), and the longest span is between the first and the last sentences of the summary (sentences 1 and 21).

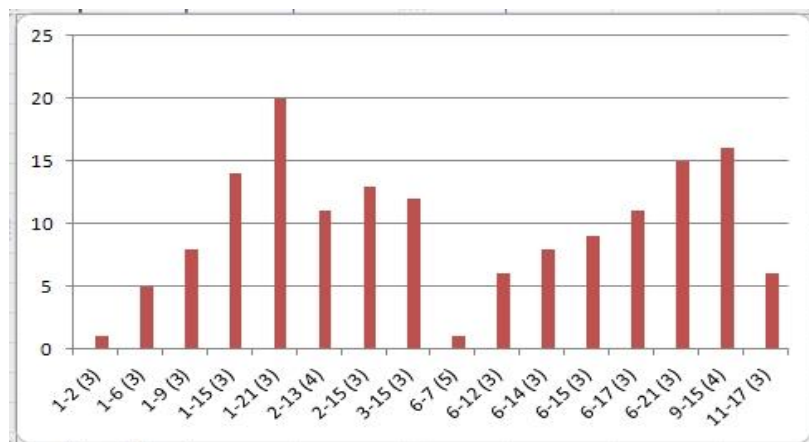


Figure 18. The span of bonds in Text 3

The strength of the connection was determined and shown in Figure 19 for Text 3. It illustrates how many links connect the sentence pairs. It shows that in Text 3 the average connection between bonded sentences consists of three links. The strongest connections are between sentences 6 and 7, with five links.

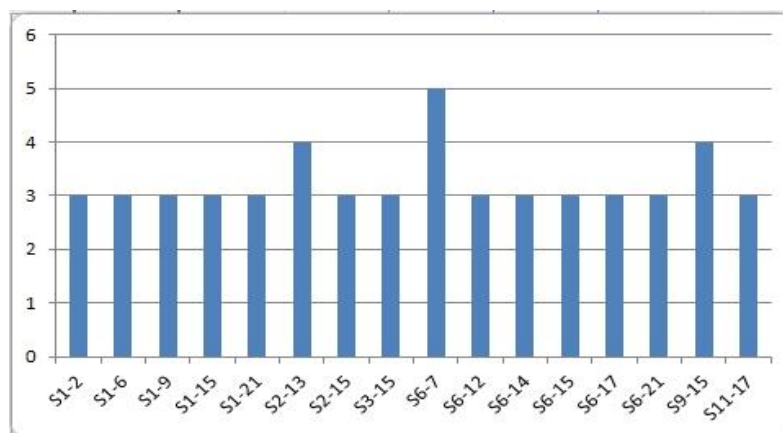


Figure 19. The strength of connection between bonded sentences in Text 3

After the above described procedures had been repeated with each summary, the qualitative measures (Appendix C) were calculated: the frequencies and ratios of links and bonds. Finally, according to Measures related to the combination of links and bonds (Appendix C), the frequency of links and bonds, as well as the density of bonds were calculated. For calculating the frequency

of links, the number of links were divided by the number of sentences, and the same was repeated for calculating the frequency of bonds. The number of bonds were divided by the number of cells to count the density of bonds. All these steps were repeated for each summary, and the results are presented in Chapter 6.

5.3 Stage 2: Analysis of compare/contrast essays

The present research project followed a mixed sequential design, according to which the results of the first stage inform the second stage. Thus, the results regarding the summary corpus had to be taken into account when the methods for Stage 2 were designed. Similarly, the methodological decisions which arose in Stage 1 were also considered in Stage 2. Given that most of the analytical decisions had already been made in the first stage, and the procedures of the analysis were repeated in the second stage, the description of the methods in this section is considerably shorter.

The aims of this stage were: (a) to test Károly's (2002) LRA tool on the compare/contrast essay genre; (b) to test an existing computer-aided method to ease the steps of the analysis; and (c) to draw conclusions (theoretical, empirical, and methodological) to inform the design of a computer-aided analytical framework.

5.3.1 The compare/contrast essay writing task

Selecting a corpus for Stage 2 of this study was more challenging than it was for Stage 1. A key consideration was to obtain texts from the same pool of participants as in Stage 1 because they were a homogeneous group of students with the same language background and language proficiency. Given that the usage of lexical cohesion devices is greatly determined by language level (Reynolds, 2001), it was important to collect texts from learners who had already passed the Proficiency Test in English administered by the university (CEFR level C1).

Getting access to compare/contrast essays produced by such students proved difficult. Even though during Academic Writing courses at Eötvös Loránd University argumentative essays and summaries are often set as assignments, not many EAP teachers require a compare/contrast essay task, simply because time does not permit it. The second important issue was to find suitable topics for the texts. In Stage 1, the selected input text turned out to be problematic because it was based around a specific agent (*Ryanair*, and to a lesser extent, *O'Leary*), which resulted in a very high ratio of instantial relations in summaries due to the high number of repeated proper nouns. To avoid this problem, the texts for the compare/contrast essays were required to be about a general topic.

The compare/contrast essay corpus to which I gained access were assignments during an EAP course. They consisted of eight texts from a small group of students, whose task was to write a compare/contrast academic essay on an applied linguistics topic of approximately 600 words. The source texts (books and journal articles) were selected by the students. The course teacher randomly chose 4 high-rated and 4 low-rated essays after she holistically scored them.

The content of the compare/contrast academic essay corpus covered the following applied linguistics topics: native language acquisition, language gender, native and non-native language teachers, language differences, dyslexic language learners. Each essay had a multi-word title and consisted of 4-7 paragraphs, containing a separate introduction and conclusion paragraph. The corpus contained of 4,971 words. This data size, due to the complex nature of manual analysis, was similar to that of international and Hungarian empirical studies in this specialized field of discourse analysis.

5.3.2 Quality assessment of the corpus

There were several differences between the corpus collection in Stage 1 and Stage 2. One of the main differences was that the input texts were provided by the students. The second difference was that the course teacher did not provide the scores for the essays, apart from marking them as high or low-rated. She also checked them for plagiarism individually, therefore the essays did not contain an excessive amount of text from the sources. Therefore, the quality assessment of the corpus was restricted to the analytic scoring.

Similarly to the procedures in Stage 1, two experienced teachers were asked to evaluate the essays using Hamp-Lyons's (1991) Michigan Writing Assessment Scoring Guide (see Appendix K). However, this time not only discourse control, but also organization was evaluated. The reason for this was that the compare/contrast essay is a genre where, besides content, the positioning of key ideas also highly influence teachers' judgement of quality.

Given that this scoring guide separates the organization and discourse control subscales, the following features were assessed separately: clarity of point of view; unity; paragraphing; present thesis/purpose statement; conclusion and effectiveness of topic sentences within the organization rubric, while logical coherence; the use of cohesive devices (key words, pronouns, references, transitions, etc.); and the clarity and coherence of the presentation of ideas within the discourse control rubric. Accuracy and appropriacy related features, similarly to Stage 1, were not examined.

The mean scores of the compare contrast essays are provided in Table 25. The aggregated scores of the two rubrics again coincided with the holistic rating of the class teacher, which might be further proof of Freedman's (1979) findings that, besides content, organization is weighted more than other features while evaluating essays; as well as Crossley and McNamara's (2010)

results, namely that coherence was the most predictive variable as regards holistic essay scores, explaining 60 percent in their variance.

Organization	Discourse control	TOTAL
3.38	3.03	6.41
2.58	2.51	5.09

Table 25. Compare/contrast essay evaluation: the organization and discourse control rubrics

5.3.3 Methods of data analysis in Stage 2

5.3.3.1 Coding

Two coders (the researcher and the same coder as in Stage 1) coded the essays using the same framework for identification of the types of links. Although the essays were all from the applied linguistics domain, the texts were more varied in their lexis, therefore the standardization part of the coding process (where the two coders worked together on a randomly chosen essay) took longer and involved more disagreement. Particularly the identification of multi-word lexical units (whether they form a single lexical unit or can be coded as separate entities) caused problems. The reason behind this might be that applied linguistics is a relatively new field and its terminology is still changing. Terms which are already widely used, such as *non-native speaker (NNS)*, were identified as one entity easily.

The coders agreed that if the multi-word expression can be found in the Longman Handbook of Applied Linguistics as a separate entry, it will be regarded as a single lexical unit of analysis, otherwise the words will be considered separately for potential links. For convenience, Károly's (2002) taxonomy is repeated here, with examples from the current corpus, in Table 26.

REPETITION		Examples	
I. Same unit repetition	repetition	simple repetition	teacher—teachers
		derived repetition	speak—speech
II. Different unit repetition	synonymy	simple synonymy	acquisition-- learning
		derived synonymy	acquisition—to learn
	opposites	simple opposites	learner—teacher
		derived opposites	learner—teach
	hyponymy		course book— materials
	meronymy		brains—head
instantial relations		ventriloquist— Medgyes	

Table 26. Károly's (2002) taxonomy for LRA with examples from the compare/contrast essay corpus

The researcher anticipated disagreement regarding different unit repetitions (DUR), especially during identifying possible synonyms, hypernyms and meronyms, due to the applied linguistics content, therefore several online thesauri were used. WordNet did not prove useful, perhaps because its semantic synsets are based on common English words from the Brown Corpus, whereas the ambiguous cases in the essays were mostly academic/applied linguistics expressions. One such case was, for instance, whether acquisition and learning can be considered synonyms or not. In these cases the coders referred to the Handbook of Applied Linguistics again. Due to these difficulties, the two coders analyzed all the eight essays together to find and categorize the links. Still more empirical research is necessary into how links can be identified in a (more) exact way regarding specialist terminology.

5.3.3.2 *Preparing the corpus for computerized analysis*

The manual analysis of summaries was replaced by a partially computerized analysis in the case of the compare/contrast essays. These essays were written using Microsoft Word in the format required by the teacher. In order to utilize the Concordance/ Version 3.3³⁷ software application,

³⁷ <http://www.concordancesoftware.co.uk/>

however, alterations were necessary in the text format and structure. First, each sentence was broken into a new line. This was necessary for the program to handle the data sentence by sentence. The title was also treated as one sentence because it was also searched for concordances. Next, misspelt words had to be corrected, because the built-in headword recognition dictionary would not have recognized them.

Multi-word phrases were united by placing a hyphen between them, otherwise the program would have counted each word separately. For instance, one essay compared Native and non-native English speaking teachers, later referring to them as *NESTs* and *non-NESTs*. As the original intention of the writer was to use the abbreviation (NEST) as a simple repetition for *native English speaking teacher*, the uploading of the four-word expression had to reflect reference to one lexical unit (i.e., *native-English-speaking-teacher*). This was especially important because the basis of the analysis was the ‘lexical unit’, described in Section 2.3 in detail.

The essays were saved as text files (.txt) and loaded into the concordance program. Each text was annotated to retain its basic structural features such as title and paragraphs, to facilitate further research of sentences at paragraph boundaries. This annotation also contained a description of paragraphs for further analysis where possible, naming them introductory paragraph, paragraph(s) describing similarities, paragraph(s) describing differences, and summary paragraph. This step was non-compulsory, as the printed versions could also have been used to determine paragraph features.

5.3.3.3 *Setting the program for the analysis*

First, the basic measures were calculated automatically by the concordance program: number of words (types), number of words (tokens), type/token ratio³⁸, number of sentences, out

³⁸ In the sentence *Do it again and again* the number of tokens is 5, the number of types is 4 because *again* is repeated. Tokens are the actual words, types are the kinds of words in the text.

of which only the number of sentences and number of tokens data were used in this study. Next, a full search was carried out on concordances in order to establish simple repetitions. Although the program contained a built-in lemmatization³⁹ tool, it distinguished between lemmas as types and tokens, and made no distinction between inflections and derivations, therefore it could not be used to distinguish between simple or derived repetitions. Due to the fact that the program did not offer a sentence parser, thorough examination was necessary to distinguish between cases such as *uses* – plural noun vs. *uses* – present tense verb, third person singular. This step had to be taken so as not to distort the simple /derived repetition ratio in the analysis.

Some concordances had to be eliminated from the headword lists:

1. Concordances of non-content words, as they were not within the scope of the lexical repetition analysis. They were ignored by the program when they were loaded into the ‘stop list’, an inside dictionary which can be extended by the user.
2. Nouns which were part of a discourse organizer phrase, whenever they served the purpose of conjunctions in the paragraphs were also eliminated, such as *hand* in the expression of *on the one hand / on the other hand*. In the same fashion, when the noun *summary* was part of the introductory phrase *in summary* in the conclusion paragraph, it was eliminated from the wordlist however, when the concordance *summary /summarize* appeared as content word-pairs in one of the essays, it was treated as relevant to the analysis and kept as part of the list.
3. According to Hoey’s (1991) and Károly’s (2002) work, concordances within sentences (repetition inside sentences) do not contribute to the organizing function of lexical cohesion of texts, therefore, these concordances were ignored.

³⁹ A lemma is the dictionary entry of a word. Run, ran, running, runs are inflected or derivated forms of *run* as lemma.

4. In the cases of non-integral citation, the name of the author and the date was deleted, but the name was kept when an integral citation was used. The reason for this was that in non-integral citations the author and the date were indicated as additional information (in brackets), and not as an integral part of the sentence, whereas the integrally cited author could have been part of an instancial repetition link.
5. Types of concordances added to the headword lists are the following:
- inflected words, such as singular and plural forms of the same noun (regular and irregular forms: *situation, situations, man, men*),
 - possessive cases in singular and plural (*child, child's*),
 - verbs conjugated (third person singular and plural, simple present and simple past tense forms, regular and irregular).

5.3.3.4 Listing concordances

The measures related to repetition type (Appendix C) were calculated with the help of the concordance application. First, simple repetition frequency lists were drawn up. Figure 20 indicates the wordlist counted by the program, organized according to frequency. In this text (Text 4) the words *gender, identity* and *mother* appeared most.

Headword	N.	%
GENDER	15	4,032
IDENTITY	15	4,032
MOTHER	15	4,032
FORMATION	9	2,419
ROLE	7	1,882
WOMEN	7	1,882
MALES	6	1,613
MEN	6	1,613
PERIOD	6	1,613
DIFFERENCES	5	1,344
BETWEEN	4	1,075
CASE	4	1,075
FATHER	4	1,075
FEMALES	4	1,075
OEDIPAL	4	1,075

Figure 20. A detail of the headword frequency list (Text 4). Headwords are listed according to occurrence. (N.= number, % = the percentage of the occurrence in the text)

Next, a chart was prepared manually to show the sentential position of the words. Columns B-Z indicate the sentences in which the given headword occurs. Number 1 represents the title. With the help of the table, conclusions can be drawn for all types of repetition links. This way intra-sentential links, such as the simple synonyms of *men* and *male* (Rows 4 and 6) can be eliminated from the link-count, as they both occur in sentence 23.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	language/s	1		3						9						15	16	17			20	21		23	24	25
2	gender/s																		18	19			22			
3	women, women's		2			5		7	8			11		13		15								23		25
4	men, men's		2				6	7	8						14	15								23		25
5	female	1		3						9											20			23		24
6	male	1		3						9											20			23		24
7	example		2																18					23		
8	devices											11	12	13	14						19					
9	hedging											11			14											
10	Japanese																16	17	18	19						
11	speaker																			19			22			
12	talking					5					10															
13	speech		2		4																			23		
14	talk						6	7	8		10															
15	difference/s	1			4											15	16	17			20	21		23		25
16	different		2	3																18					24	
17	forms																				19		21	22		
18	lexical															15		17			20					
19	morphological															15		17			20					
20	use/d											11	12	13	14					18						
21	aspects		2	3																						
22	cars						6	7																		
23	certain																					21	22			
24	choice														15										24	
25	computers						6	7																		
26	topic/s				4				7							15									24	
27	modern																16	17								
28	person									10																
29	phonetical																				20			23		
30	relationships					5		7																		
31	distinctions															15					20		22			
32	verb																					21	22			
33	explains									11												21				
34	states					5				10																
35	prefer/red				4	5																				
36	preference/s		2						7																	
37	rule/s									8														22		

Figure 21. Another detail of the headword frequency list (Text 3). The numbers represent the sentences in which the words occur. (No.1 = the title)

This step was the last where the concordance could be used for the lexical repetition analysis. For the rest of the analysis, the steps described in Stage 1 were repeated: for the other

lexical repetition types, a “classic” coding matrix was created, in which cells represented possible inter-sentential links. The types of the links were determined, and put into the matrix cells, where all links were itemized, and their types of repetition were identified. The data indicated by the frequency table (Figure 21) and the matrix representing all types of repetitions were checked to obtain the final number of repetition links for each essay. Finally, the frequency of all repetition types were calculated, and the results were presented in a table, establishing the combination of links and bonds. These steps are illustrated in Section 5.2.6.

6 Results of the lexical repetition analysis of academic summaries

This chapter discusses the results referring to Stage 1, the lexical repetition analysis on the academic summary corpus. After a brief review of the corpus of summaries, the results will be presented in three sections:

1. Empirical results related to repetition type
2. Empirical results related to the combination of links and bonds
3. Methodological results

6.1 General features of the summaries

In the summary task the students were instructed to summarize Ryanair's history and business model, highlighting its strengths and weaknesses. The students first needed to find all the information on the company history and on the business model. Next, they had to reorganize the information under the strengths and weaknesses categories, as seen in the construct model, in Table 27.

Main constructs	Divisions of main constructs
History	
Business model	strengths
	weaknesses

Table 27. The summary construct

Most students were able to find the main ideas with regards the main constructs and the second construct's two divisions. The differences in the qualities of the summaries were more due to the structuring of the information under these headings and the subheadings.

As far as rhetorical patterns are concerned, an academic summary should contain a title, an introduction, a thesis sentence, topic sentences and a conclusion. The title was not given, therefore

most students gave the company name (*Ryanair*) as a title. One summary was not divided into shorter sections, the others consisted of two or three paragraphs.

6.2 Results related to repetition type

After Basic measures and Measures related to repetition type (Appendix C) were performed for each text, the following results were found (Table 28).

code	Frequency of types of repetition										
	Same unit repetition			Different unit repetition							
	Simple	Derived	SUR	Synonymy		Opposites		Hyponymy	Meronymy	Instantial relations	DUR
				Simple	Derived	Simple	Derived				
L1	1.6	0	1.6	1.33	0.06	0.06	0	0	0	0	1.45
L2	2.97	0.11	3.08	2.42	0.08	0	0	0	0	0.08	2.58
L3	1.6	0	1.6	1.46	0.06	0.06	0	0	0	0	1.58
L4	2.86	0.17	3.03	1.96	0.1	0	0.07	0	0	0	2.13
L5	4.12	0.12	4.24	2.47	0	0	0	0	0	0	2.47
H1	2.93	0.46	3.39	1.8	0.2	0.066	0.13	0	0	0.066	2.26
H2	5.14	0.57	5.71	1.86	0.07	0	0	0.14	0	0.21	2.28
H3	2.73	0.33	3.06	2.066	0.2	0.066	0.066	0	0	0.066	2.46
H4	8.97	0.48	9.45	6.22	0	0.03	0.06	0	0	0	6.31
H5	5.33	0.33	5.66	4.47	0.28	0.14	0	0	0	0.09	4.98

Table 28. The frequency of types of repetition in high- and low-rated summaries
Abbreviations: SUR: Same unit repetition. DUR: Different unit repetition

According to the table, high-rated summaries contain more repetition links in general. A tendency can be observed towards more frequent use of simple repetition, simple synonymy, and derived opposites in higher-rated summaries. This group also contains a higher number of derived same unit repetition (this is highlighted with grey in Table 28). The mean frequency of derived same unit repetition is **0.08** in low-rated summaries, while it is as high as **0.434** in high-rated summaries. There is also a difference between the two groups in the usage of instancial relations (other than the *Ryanair* -- *synonym* links, which were excluded from this relation during the coding process) with higher-rated summaries containing more of this repetition type. At this point,

unfortunately, statistical significance⁴⁰ of these differences cannot be calculated due to the small size of the sample.

If we analyze the results by using Károly's broader categories (same unit repetition vs. different unit repetition), the following conclusions can be drawn: the results indicate that there is a clear difference between the two groups concerning same unit repetition (SUR), because the average frequency in low-rated summaries is **2.71**, whereas in high-rated summaries it is more than double (**5.45**). The average frequency of different unit repetition (DUR) in low-rated texts is **2.042**, and in high-rated summaries it is **3.658** as indicated in Table 29.

code	SUR	DUR
L1	1,6	1,45
L2	3,08	2,58
L3	1,6	1,58
L4	3,03	2,13
L5	4,24	2,47
mean frequency	2,71	2,042
H1	3,39	2,26
H2	5,71	2,28
H3	3,06	2,46
H4	9,45	6,31
H5	5,66	4,98
mean frequency	5,45	3,658

Table 29. The difference between the mean frequencies of SUR and DUR in high-and low-rated summaries

Abbreviations: SUR: Same unit repetition, DUR: Different unit repetition

These results are in line with what Károly (2002) found in the case of argumentative essays, where results indicated that high-rated essays used more same unit repetition, particularly derived repetition, simple opposites and instantial relations. A tendency was also observed to a more

⁴⁰ Two groups each with a sample size of 10 and means: 30.5 and 31.5 respectively, $t=2.15$, $p > 0.05$. This difference is not statistically significant. But with two groups of $n=100$ items each, and the same means: 30.5 and 31.5, $t=2.15$, $p < 0.05$, the difference is significant.

frequent use of simple repetition and synonymy, particularly derived synonymy in high-rated essays in Károly's data.

A surprising result in the summary corpus was the general lack of meronymy and hyponymy. (See rows in Table 28, previously), particularly the lack of the latter was an interesting phenomenon in the corpus. It could be assumed that the generalization rule (van Dijk & Kintsch, 1983) presupposes using a more general term for summarizing lists. Baba (2009), for instance, found that the ability to write definitions is the most important trait in summary writing. A closer look on the texts revealed that the writers of high-rated summaries resorted to a different structure when explaining content. They did use generalization instead of listing, however, instead of using superordinate terms they used adjective + general noun structure (*economizing processes, financial measures*) in which the adjective classified the noun. These nouns were not in hyponym—hyperonym relation with the nouns which were members of the list in the source text, rather these nouns can be described as a broader term that could not be classified within the taxonomy, for example: *Such financial measures, introduced by O'Leary proved to be successful in the long run for the company.* These nouns seem similar, although not the same as the list Hyland (1998) compiled and which was already discussed in Section 2.3.6, in connection with Hoey's (1991) analytical decisions. (See the list in Appendix B.)

Another reason why Baba's (2009) results and the results of this study differ in this respect might be that definitions generally comprise the higher term and the lower term within one sentence (e.g., *the chair is a kind of furniture*), whereas lexical repetition analysis (LRA) explores links only above sentence level. Thus, hierarchical semantic relations within a sentence remain unnoticed.

This interesting phenomenon, namely the missing superordination, could also result from the features of the source text. More specifically, the original text did not contain hierarchical concepts that could lend themselves to creating an equally hierarchical organization in the summaries. Therefore, more experiment is necessary with different source texts regarding topic domain and hierarchical organization of concepts contained in the input texts.

After observing the number of links, we can conclude that their quantitative differences alone could not account for the teachers' different perceptions of the two groups of summaries. In other words, although some differences could be observed in the number of links, it cannot be stated that high-rated summaries were perceived as better as a consequence of containing more repetition links. Similarly, it cannot be stated either that low-rated summaries were worse because they contained fewer lexical repetition links. This second observation is perhaps more surprising if we consider the purpose of a summary, which is to condense a longer text. We would hypothesise that a shorter version of a longer text should contain less repetition. However, we need to remember that **repetition** is used in a broad sense in this study, containing synonyms, antonyms, etc. Thus the act of staying on topic in itself implies that the text will contain numerous repetition links because the words are selected from the same semantic field. (At this point the researcher tentatively reconsiders "who to believe": de Beaugrande and Dressler (1981), who claimed that cohesion is a **precondition** of textuality, or Halliday and Hasan (1976) who state that cohesion is a **contributor** to coherence.)

To observe the link types qualitatively, we can conclude that even though the high-rated and low-rated summaries differed in the type of repetition links used, these differences were not as significant that they could have been the cause of perceived text quality differences.

Károly's (2002) results	Stage 1 results
higher-rated essays contain more repetition links, this difference is significant	confirmed
high-rated essays contain more same unit repetition	significance cannot be confirmed
high-rated essays contain more derived repetition	confirmed

Table 30. Comparison of results related to lexical repetition types between Károly's (2002) argumentative essay corpus and Stage 1 summary corpus

6.3 Results related to the combination of links and bonds

The next to observe is the patterning of the links. The location of bonds in the summaries show varied patterning. In nine of the ten summaries, no bonds can be found between the title and the summary. The reason for this was that these texts contained the company name as title (one-word title), therefore maximum one link was available to tie with any other sentences. The one summary title which consisted of more words (*Ryanair, Europe's largest low-cost carrier*) had only one bond with the first sentence. Therefore, the analytical steps concerning the bonds between the title and the other sentences in the summaries could not gain informative results. This finding is not surprising if we approach the summary genre from a task authenticity angle. When a summary task is given to students, usually there are no formal requirements for the title unlike in the case of an essay. In the latter case the title should refer to the essay content and one-word titles cannot express the complexity of the topic.

The frequency of links and bonds, as well as the density of bonds were calculated, as indicated in Table 31. For calculating the frequency of links, the number of links were divided by the number of sentences, and the same was repeated for calculating the frequency of bonds. The number of bonds were divided by the number of cells to count the density of bonds. As Table 31 shows, in the case of the summary corpus several differences can be observed both for bond frequency and density. In all three categories (frequency of links, frequency of bonds, and density of bonds) high-rated summaries contained higher values.

Code	Frequency of links	Frequency of bonds	Density of bonds
L1	3,36	0,05	0,008
L2	5,55	0,055	0,002
L3	3,2	0,06	0,008
L4	5,24	0,034	0,002
L5	6,88	0,05	0,006
mean	4,846	0,05	0,0052
H1	5,46	0,266	0,033
H2	8	0,71	0,09
H3	5,53	0,266	0,033
H4	12,7	0,9	0,056
H5	10,66	0,76	0,69
mean	8,47	0,58	0,18

Table 31. Frequency of links and bonds and density of bonds

In high-rated texts, the frequency of links and bonds as well as the density of bonds were higher. (It is not surprising if we consider that there is a connection: if a text contains more links, these links will be bonded with the given sentences and as a consequence, the text will be more densely bonded.) A reason for this could be that good summaries contained sentences which were more ‘compacted’; in other words, the writers arranged the information from the source text in a logical order, with introductory sentences.

A qualitative investigation into the location of the bonded sentences shows that in high-rated summaries they are more likely to appear in the first three-quarters of the texts. This may be caused by the type of information the students had to collect, as the task they were given consisted of organizing two types of information. The first topic they had to summarize (history of the company) involved facts and figures, which were rather listed than summarized in low-rated summaries. Concerning lexical repetition patterns, this resulted in fewer cohesive devices because adjacent sentences did not rely on each other. Good summaries, on the other hand, contained such information with introductory phrases or a general topic sentence followed by listing. To illustrate

the different structuring of important information, two content-wise almost identical examples are given:

(1) *He introduced quick turn-around times, 'no frills' and no business class, as well as collecting ancillary revenue.* (from a low-rated summary)

(2) *His measures have become models for the low-fared airlines, which included quick turn-around times, no business class, only a single model of aircraft in operation and the use of regional airports instead of international ones.* (from a high-rated summary)

As it can be observed, both sentences contain almost the same information. The first sentence is a so-called marginal sentence as it has no links or bonds with any other sentences. However, if it were not present, essential information would be lost. The underlined units in the second sentence contain links to other sentences, and it is also bonded to two others.

The second sentence is also better because the ideas it contains were collected from different paragraphs in the source text. If we consider extractive vs. abstractive summaries in the automated summarizing literature, this sentence resembles the abstractive type of summary as it contains a classification/definition type of head. This sentence classified the information collected from the source text, and built a hierarchy in the topic. Thus, it shows knowledge of the topic, and is considered more coherent. This structure is more highly valued in academic texts as it is more focused.

The second sentence contains a noun that was already highlighted while discussing the lack of hypernyms. These nouns are not only similar to Halliday and Hasan's **general items** (a subcategory within reiterations), but are also addressed by Hoey (1991, p. 105), who found the sentences containing such nouns marginal in his analysis. He called these sentences signaling sentences because their function is to relate earlier sentences to later ones. He claims that these can

be considered sentence-length conjuncts. However, in this corpus they were found to form several links with other significant sentences, perhaps because the texts in question are summaries, therefore short and concise.

So far we may conclude that bonded SDF (special discourse function) sentences contain the most important information in good summaries. Therefore Károly's (2002) analytical tool devised for argumentative essays can be assumed to distinguish between high-rated and low-rated summaries as well. It was seen in Table 31 that high-rated summaries had a higher frequency of bonds and it was illustrated with the two examples that in good summaries it was bonded sentences which contained the important information. However, there are some doubts about this, especially when we perform a content-based analysis on the summaries.

For instance, there are cases when bonding between sentences is superfluous. The following two sentences (sentence 2 and 13) are taken from the same summary.

(S2) The airline was founded in 1985 in Ireland by Christopher Ryan, and it has become today's largest low-cost carrier in Europe.

(S13) By 2003, Ryanair was among the largest carriers in Europe.

The sentences have four links in common and both are bonded with other sentences in the text. The problem is that S13 contains hardly any new information compared to S2, and could have been incorporated into another sentence. According to the framework however, S13 is also part of the net of bonds, frequencies, cumulative bond span and density of bonds, adding to their value. It might be possible that another type of content-based element should be incorporated into the present framework to investigate the flow of information content and filter unnecessary repetitions between sentences. A further argument for this might be that by providing superfluous information, the principle of providing an optimal flow of information is violated. Another framework has

already been devised for analyzing this issue: the Topical Structure Analysis (TSA) by Lautamatti (1987). Several findings indicate that there is a difference between high-rated and low-rated EFL student texts e.g., Schneider and Connor (1990) regarding their topical progression.

One low-rated summary was of particular interest regarding topic progression/flow of information. The teacher's comment was that it uses too many equivalences intra-sententionally (inside the sentence). This redundancy was caused because the same referent (*Ryanair*) was circumlocuted in various ways (*company, airline, carrier, firm* and wrongly **aircraft*) within sentences. This type of repetition is not analyzed within the present framework because only inter-sentential repetition links are examined, it is however noticeable for raters. It seems, lacking repetition links, as well as too many repetition links are both detected by teachers. Therefore, not 'normal' but a diversion from 'normal' is what is noticed.

Another doubt about the importance of central sentences is based on Tyler's (1995) critique of Hoey (1991). Tyler claimed that several marginal sentences were essential in the good summary of a given text. This investigation showed that this might be true when information in the source text is listed and appears only once, but have considerable content value and without which the information of the original would be partially lost, therefore these sentences are not to be considered merely holding additional information. According to her findings (1995), "the articulation of relevant concepts is the property that makes a sentence peripheral or central to the text, not the number of lexical repetitions or links it contains" (p. 150). If there had been a strict word limit for the summaries, the one containing sentence (1) might have been rated higher than (2).

The last two measures the **mean of the cumulative bond span** and the **strength of bonds** values were higher in high-rated summaries than in low-rated summaries. However, as one of the

low-rated texts contained no bonds, and two summaries contained one bond, the mean values indicate limited information concerning the quality of texts. The mean value of the strength of bonds was 3.4 for high-rated summaries, which indicates that three links were the average to connect bonded sentences in this corpus. This figure is less than two (1.8) in low-rated summaries, and both the frequency of links and the frequency of bonds are lower. Károly (2002) concluded that quantitative investigation of bonding (in general) did not distinguish between the argumentative essays, however, here a tendency can be observed towards more bonding in high-rated texts.

6.4 Methodological outcomes

It was an interesting outcome of Stage 1 of this research that the coding process was more difficult than expected. During the coding procedure special effort was made to improve inter-rater reliability, however, both the link finding procedure and the lexical unit identification took a long time. A possible reason for this lies in the discourse domain of the texts, which consequently influenced the vocabulary use. The coding procedures and the analytical decisions were detailed in the Research design section, Section 5.2.6, and possible resolution regarding some of the difficulties is presented in Chapter 8.

6.5 Summary

Stage 1 resulted in novel findings in theoretical, empirical and research methodological aspects. The main theoretical finding of the research is that Károly's analytical tool was capable of detecting structural differences between the lexical organizations of high and low-rated summaries. It was also capable of predicting teachers' judgement regarding cohesion in the texts analyzed. Given that rater judgement on cohesion is connected to discourse quality, it can be said that the tool can differentiate between structurally well-formed and structurally ill-formed

summaries, thus can predict rater judgement in this respect. As a consequence, this is the first discourse analytical study which can demonstrate that high-rated and low-rated summaries show different lexical repetition patterns. Taken together, these findings shed light on the underlying interactions of lexical repetition elements which organize discourse within the summary genre.

An interesting outcome of the analysis was that in high-rated summaries the main ideas to be collected were organized into sentences with special discourse functions, particularly at the first section of the summaries where the thesis (or in the case of paragraphs the topic sentence) should be placed. It can be assumed that high-rated summary writers resorted to their knowledge transforming skills, as opposed to the less consciously structured knowledge telling practice (Scardamalia & Bereiter, 1987). Writers of high-rated summaries used the generalization and construction strategies (van Dijk and Kintsch, 1983), even though how exactly the general nouns are used in these sentences needs further research. Some measures of the original model still need further investigation on a larger pool of summaries.

Two conclusions can be drawn from the LRA in Stage 1. The first one is that Károly's (2002) model, which was originally devised for EFL argumentative essays, can be used for the LRA for summaries. This is a 'somewhat surprising' result given that argumentative essays seem to have a more visible organizational pattern within the introduction – body paragraphs – conclusion division: they need to include a clear thesis statement, as well as topic sentences and a concluding sentence. Summaries are at a first glance less structured. They depend on the source text in their content and also in their organization. Yet, this study revealed that summaries are also structured along these 'rules'.

The second conclusion that can be drawn is that Károly's tool is capable of predicting the pedagogical evaluation of the summaries. Those summaries were rated higher by teachers and the

analytical tool which contained the important information in 'topic sentences', at the first third of the summaries. The original article started describing company history based around a time-line, with dates connected to various new facts appearing in almost every sentence. Later the text changed into listing several measures the company introduced to improve their situation. Students who did not approach the text linearly, but had an overall view of it, realized that the presented list of information could be organized into groups, evaluating it as weakness or strength. This resulted in a more condensed information content, fewer numbers of sentences, fewer simple repetitions and simple synonyms within inter-sentential relations.

The analytical tool in its present form may predict subjective perceptions of the quality of the type of summary observed. No conclusions can be drawn on its reliability and validity in cases when the original document to be abridged uses a different narrative form. Limitations arise if we consider measures calculating bonds between structural elements not present or different from those in the argumentative essay: the title, the thesis statements, and in some cases the central sentences.

Future studies are necessary into how the preset length limit of summaries influence the quantity and quality of lexical repetition. It is likely that a shorter limit in length necessitates different structural organization. However, it is not clear how this shorter text length affects lexical repetition patterns. Higher lexical density (i.e., more content words per sentence), necessitated by condensing more information from the source text into the summary, might or might not appear in the lexical repetition patterning.

Let us suppose that the student has a one-paragraph summary to write, and he resorts to the technique of condensing every paragraph from the source text into a sentence. These sentences may consist of 15 words or fewer. Finally he polishes his final draft by adding other cohesive

devises such as transition phrases to make the summary coherent. In this case, it is difficult to anticipate how many lexical repetitions will appear in the text. Probably the density of repetition links will depend on the density of macropropositions the source text contains: if a lot of different ideas are present in the original text, the summary might have fewer repetition links. If however, the original text describes one main idea from various angles, the key words of the source text might be repeated more often in the summary. Table 32 illustrates the possible links in one-paragraph and multiple-paragraph summaries.

Summary with one paragraph	Lexical repetition patterns	Summary with multiple paragraphs	Lexical repetition patterns
title	<i>possible links (bonds) with topic sentence and concluding sentence(s)</i>	title	<i>possible links (bonds) with thesis sentence, topic sentence and concluding sentence(s)</i>
topic sentence	<i>possible link (bonds) with title and concluding sentence(s)</i>	thesis sentence	<i>possible link (bonds) with title, topic sentences and concluding sentence(s)</i>
supporting sentences	<i>links are possible, bonds are less possible</i>	topic sentence paragraph 1	<i>possible link (bonds) with title and concluding sentence(s)</i>
concluding sentence(s)	<i>see above</i>	supporting sentences paragraph 1 concluding sentence(s)	<i>links are possible, bonds are less possible see above</i>

Table 32. Possible lexical repetition patterns in one or multiple paragraph summaries

Methodological results based on the coding process and the analysis revealed that two aspects of the analysis are prone to errors: identifying the appropriate units of analysis (lexical units) and finding all the links within a text, therefore both issues need further research: whether a

general definition of the unit of analysis is sufficient, or topic/domain based definition would be necessary to ensure better inter-rater reliability of the coding. A further methodological issue related to this is to suggest ways how the units of analysis can be defined in each case: whether to use a thesaurus, specialized dictionaries or other handbooks related to the topic domain in question in ambiguous cases. The last issue to mention is the consistent treatment of text-bound lexical relations, such as proper names. For a large corpus, consistent methodology is necessary to classify e.g., instantial relations.

7 Results of the lexical repetition analysis of compare/contrast essays

This chapter discusses the results referring to Stage 2, the lexical repetition analysis on the compare/contrast essay corpus. After a brief review of the general features of the compare/contrast essays, the results will be presented in three sections:

1. Empirical results related to repetition type
2. Empirical results related to the combination of links and bonds
3. Methodological results

First, results related to lexical repetition types are discussed, followed by conclusions drawn on characteristics of bonds between the sentences. The predicting capability of the analytical tool is discussed afterwards with observations on the use of the Concordance 3.3 program.

7.1 General features of the compare/contrast essays

Seven of the compare/contrast essays in the corpus were organized in the block pattern. This means that after the introduction all the features of one topic were discussed, and following this, all the features of the other topic were mentioned. One essay used the point-by-point organization pattern. This text discussed three periods of identity formation in its three body paragraphs. Both comparing and contrasting were present in all the essays; in other words, the writers collected similarities, as well as differences. The fact that only one essay contained the point-by-point organization was a disappointing feature of the corpus for the purposes of this study because showing the differences between the two patterns as regards quality differences could have given an extra dimension to the lexical repetition analysis. Nevertheless, both the holistic evaluation of the course teacher, and the analytic assessment of the two teachers who checked the

essays for organization and discourse control, revealed clear quality differences between the two groups of essays, which is a good starting point for testing Károly's analytical tool on this genre.

7.2 Results related to repetition type

As Table 33 indicates, all of the essays contain a number of various repetition types. In line with other previous research results (Károly, 2002; Hoey 1991), the most frequent is simple repetition, whereas derived repetition is also frequent. This was also observed by Teich and Fankhauser (2005) whose computer-assisted analysis results showed a 50% frequency of repetition (simple and derived) among lexical cohesion devices in academic texts, and who interpreted these results as a characteristic of the learned register, mainly because of the need to avoid ambiguity, by using consistent terms.

code	Frequency of types of repetition										
	Same unit repetition			Different unit repetition							
	Simple	Derived	SUR	Synonymy		Opposites		Hyponymy	Meronymy	Instantial relations	DUR
				Simple	Derived	Simple	Derived				
H1	0.35	0.67	1.02	0.03	0.02	0.06	0.007	0.02	0	0	0.137
H2	0.49	0.19	0.68	0.06	0.01	0.05	0.05	0.12	0	0.002	0.292
H3	0.62	0.07	0.69	0.02	0.01	0.01	0.01	0.02	0	0	0.07
H4	0.5	0.03	0.53	0.04	0.1	0.06	0.07	0.06	0	0	0.33
L1	0.38	0.02	0.4	0.009	0.008	0.02	0.006	0.02	0	0	0.063
L2	0.47	0.08	0.55	0.039	0.2	0.066	0.05	0.02	0.001	0.066	0.442
L3	0.48	0.01	0.49	0.042	0.07	0.03	0.008	0.033	0	0	0.183
L4	0.45	0.05	0.5	0.02	0.2	0.066	0.066	0.004	0	0	0.356

Table 33. The frequency of types of repetition in high- and low-rated essays. Abbreviations: SUR: Same unit repetition. DUR: Different unit repetition

Similarly, a clear tendency can be observed towards the more frequent use of simple synonyms and simple opposites as opposed to derived ones. In some cases the ratio of repeated

words is surprisingly high, especially if we consider the fact that only content words were calculated in the study, grammatical words were entirely ignored. Hyland (2006) explains this phenomenon with the observation that a high proportion of content words in relation to grammar words is characteristic of the academic register, thus adding to the high lexical density of such texts.

Károly's (2002) results	Stage 2 results
higher-rated essays contain more repetition links, this difference was significant	confirmed significance cannot be confirmed
high-rated essays contain more same unit repetition	confirmed
high-rated essays contain more derived repetition	confirmed

Table 34. Comparison of results related to lexical repetition types between Károly's (2002) argumentative essay corpus and Stage 2 compare/contrast essay corpus

Another feature of all the essays is the relatively high number of hyponyms and hyperonyms used, which could also be an academic content-related feature. The essays compare abstract academic topics such as gender identity formation or native language acquisition models, therefore often employ clarifications, or give definitions of terms. This is a new phenomenon compared to Stage 1, where the lack of such links were noted. It is a proof that the topic has a definite influence on writers' decisions whether they want to resort to explanations and hierarchical organizations of ideas or not.

Initial assumption would suggest that hyperonyms appear in the introductory paragraph, where the topic is explained and in the final paragraph, where conclusions are drawn and ideas summarized. No clear evidence of this is visible, however, rather a scattered positioning of hyponyms and hyperonyms can be observed. A closer investigation reveals that this type of repetition appears in initial and closing sentences within paragraphs.

It was also found that high-rated essays contain slightly more repetitions than low-rated ones. Károly's (2002) previous study revealed that quantitative measures alone cannot predict

textual quality, therefore a content-based analysis is also necessary in this respect to draw conclusions.

7.3 Results related to the combination of links and bonds

Both high-rated and low-rated essays contain more marginal sentences than bonded sentences, but high-rated essays have a higher frequency of bonds. The ratio of marginal/bonded sentences also differs between the two groups (the average ratio is **0.3** and **0.215** in high and low-rated essays, respectively). The main bond-related differences fall into two areas: the relative high use of bonds at paragraph boundaries (especially in the introductory and in the concluding paragraphs), and in the span of bonds between sentences. High-rated essays connect paragraph-initial and paragraph-final sentences more frequently to each other, providing a more structured framework to of the topic explained within the paragraph. Similarly, lexical links of the main sentence in the introductory paragraph reappear in high ranked essays in one of the concluding sentences, typically in the first one. Figure 22 indicates the location of bonded sentence-pairs in the best-rated essay. The two numbers (e.g., 2-3) show the two sentences connected by bonds. It also reveals the span of bonds between sentences, which means how far the connected sentences are located from each other in the text. The shortest span involves two adjacent sentences (e.g., sentences 2-3), and the longest span is between the first and the penultimate sentence of the summary (sentences 2 and 25). The connection between the title and the essay body is also represented in Figure 22.

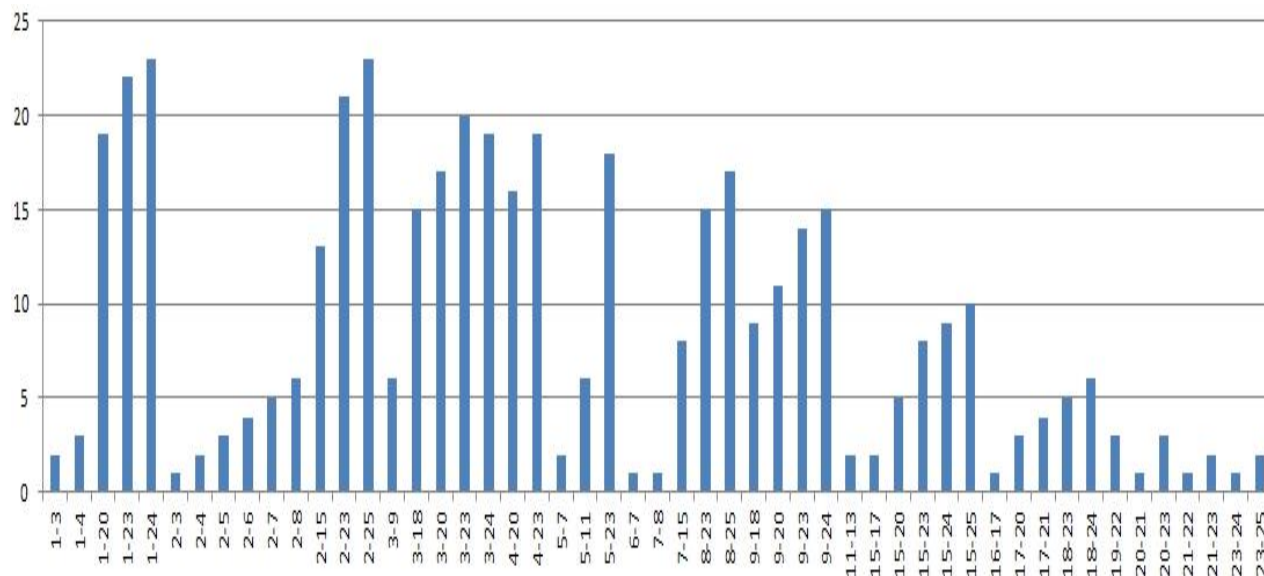


Figure 22. The span of bonds in Text 1. The two figures (e.g., 2-3) show the two sentences connected by bonds.

Seven essays contain bonds between the title and the rest of the text, however, high ranked essays contain more bonds (average = 4.3), the highest number of bonds being 7; while one low ranked essay contains only two links but no bonds between the title and the essay. Figure 22 indicates that Text 1 has five bonds with the title, two with introductory sentences (sentences 3 and 4, and three with sentences towards the end of the essay, one of which is the summary statement. No. 23). Sentence No. 2 and 23 are the key sentences, the former having 9 bonds looking forward, and the latter containing 11 bonds referring backward in the text, providing a structural framework for the essay. Therefore it can be argued that there is a relationship between the position of bonded sentences and the span of bonds, and they together influence text quality. In the case of low-ranked essays, the tendency described above is not so clearly represented.

As was the case with the examined summary corpus, in high quality essays the length of bonds stretched out to the very end of the text. This was also observed by Teich and Fankhauser (2005), who studied texts of various registers (learned, religious, press, fiction, etc.) and found that in academic articles the length of lexical chains is the longest. The reason for this might be that the

learned register “exhibits a strong topic continuity” (Teich & Fankhauser, 2005, p. 140), in other words, such texts are not supposed to have topic shifts or diversions. Although lexical chains are formed by semantically related words stretching out across sentences, thus not entirely the same as Hoey’s **length of bonds** creating lexical nets, the similarities in their textuality forming functions are obvious.

7.4 Features not detected

As it was mentioned in the description of the essays (Section 7.1), the corpus contained mostly essays with the block pattern, therefore whether the analytical tool can distinguish between the two patterns of compare/contrast essays reliably, cannot be confirmed. The clustering of bonds showed some tendencies to indicate where the paragraphs began and ended in the texts. Bonded sentences appeared in ‘paragraph-like’ arrangements in the essays. More essays with the point-by-point pattern are necessary to reveal how generalizable this observation is.

One of the main principles of lexical repetition pattern research is that the scope of investigation is restricted to inter-sentential lexical repetition, and is not concerned with intra-sentential repetition. Further examination is necessary into sentence density, as the length of the sentences determine the number of words per sentence, also determining the amount of information content spread out inter-sententially, or condensed into one sentence. Sometimes sentence boundaries set by essay writers of this corpus seemed relatively subjective or unestablished. It was especially the case when no conjunctions were used between sentences. A ‘minor’ difference in punctuation, however, means a great difference in repetition frequencies, if Hoey’s (1991) and Károly’s (2002) instructions are followed. As an illustration for sentence length as a determining factor, one of the complex sentences is shown here from Text 7, in its original form, and with the same content spread into two separate sentences. In the second case, several

links between the two sentences, which were not part of the original analysis, are revealed. (They are underlined.)

(S2) For instance, in Europe the general way adults turn to children is the child-centered communication model, while in most non-European communities the situation-centered pattern is the common one.

(S2a) For instance, in Europe the general way adults turn to children is the child-centered communication model.

(S2b) However, in most non-European communities the situation-centered pattern is the common one.

7.5 Methodological outcomes with automation in mind

The third research question referred to how the Concordance program can facilitate researching the text-organizing function of lexical repetition patterns. It certainly helped organizing the vast amount of data collected from the corpus, especially in the first phase of the analysis when the number of words, sentences and paragraphs were counted. The program, if used in further research, might also add to the depth of the analysis with a new element, namely establishing the sentence density or, as the program defines it: **the density of words per sentence**, which will need further research. The Sorting Lemmatized Headwords function also contributed considerably to the lessening of occasionally overseen or missed concordances.

Unfortunately, the functionality of the program was found to be limited for the current research in the phase of counting concordances within the texts, due to three reasons. One reason was that the specific categories of the present analytical tool did not match exactly those in the software. Especially problematic was the type/token count instead of the inflection/derivation distinction expected by the present analytical tool. Secondly, the table of links, which was drawn

on the basis of the computerized headword frequency count (Figure 21), although served as a good means of visual representation for the simple repetition links, in fact doubled the workload, since it could not represent any other repetition types, which had to be itemized and classified in a cell matrix, as well as drawn up in another matrix in a number format. (The latter two matrices were used by Hoey (1991) and Károly (2002), and also in Adorján (2013) as a compulsory element of the analysis.) Thirdly, the program was not suitable for counting frequencies, or establishing bonds between sentences.

Further research is needed into the application of a suitable computer program which is flexible enough to handle textual data according to the categories defined by Károly's (2002) taxonomy, as well as capable of recognizing more types of links and their connections to ease the workload lexical repetition pattern analysis requires, so as it could be extended to larger corpora.

Conclusions based on results regarding the usage of computerized data analysis indicate that Concordance, Version 3.3 is unable to capture certain important features of the framework, for example, except for reiterations, it is unable to locate other lexical repetition categories. Furthermore, the application is unable to store data or collate data into charts or matrices as it is required by LRA. Specific categories of the present analytical framework require the development of a software application with matching categories to enable research using large corpora.

7.6 Summary

The aim of the research in this stage was to investigate whether Károly's (2002) lexical repetition analysis framework is capable of predicting the quality of compare/contrast essays as regards their lexical repetition patterns, thus providing an "objective" tool for pedagogical evaluation. The analysis in Stage 2 therefore, contributes to the further testing of the analytical tool, using the genre of comparison-and contrast essay. Due to the small number of

compare/contrast essays, it can only be claimed that the LRA tool revealed certain tendencies in high-rated essays towards more repetition links, and a kind of bonding which resembled paragraph boundary patterns, more clearly in high-rated essays than in the case of low-rated essays. The LRA tool is certainly worth further testing on a larger pool of data.

Further content-based investigations might also be necessary into three fields. The first area of interest was mentioned in Section 7.4, whether intra-sentential repetitions should be incorporated into the analysis in some way. The second question for further study could be the categorization of repeated vocabulary in the corpus. This issue was considered in Section 6.2 in connection with general nouns. There are other aspects of noun usage in the corpus, however, which might warrant noun categorization as an additional element in the analysis. According to Nation (2001), vocabulary falls into three main groups: high frequency words, academic vocabulary (comprising 8-10 percent of the words in academic texts), and technical vocabulary, which differs by subject area (covering up to 5 percent of texts). It would be interesting to investigate the ratio of the various types of repetition links according to this classification to see how much of the repetitions are meta-discoursal, how much are content-related and, within this area, technical vocabulary used in the special field.

The third issue to consider is connected to Károly's (2002) taxonomy of lexical repetition types. In both Stages 1 and 2, non-derivational differences of the same unit were counted as simple repetition, zero derivation was counted as derived repetition, and derivational differences were counted as derived repetition. Whether this classification can be kept via automation, is an important question. (For instance, identifying **zero derivation** will need sentence parsing). It is also important, whether the difference is significant or not between simple and derived repetition.

8 The design of a new LRA model for large-scale analysis

The results of Stages 1 and 2 informed the last stage of this project in which a new analytical model for the study of the text organizing role of repetition is proposed. This is the first lexical repetition analysis (LRA) model in educational context designed to address and cope with large data size.

8.1 The newly proposed LRA model: the three modules of the analysis

The new analytical phases considered necessary for a computer-aided lexical repetition analysis (LRA) were identified during Stages 1 and 2 of this research. These phases follow a strict sequence. The first phase of the new LRA model, **Preparation of the corpus**, became necessary due to the change from manual to computer-aided analysis. The second and third phases, **Establishing links** and **Establishing bonds**, are based on Hoey's (1991) original LRA framework which was further developed by Károly (2002). These are also modified however, only to a lesser extent. The phases are considered modules because they can be independently developed and comprise different actions. They need to be linked to be part of a LRA computer program. Figure 23 shows the three modules of the new LRA model.

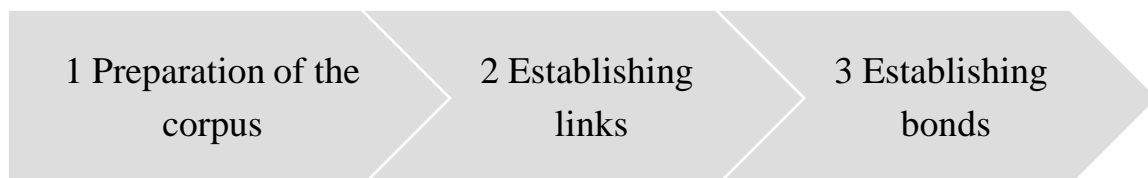


Figure 23. The three big modules of the new LRA model

An overview of the proposed model for computer-assisted LRA is shown in Figure 24. This is followed by a discussion considering particularly the first two modules, which involve

several new steps, as well as new theoretical decisions. Some of these steps can utilize existing computer applications, however, they all need further testing on larger data.

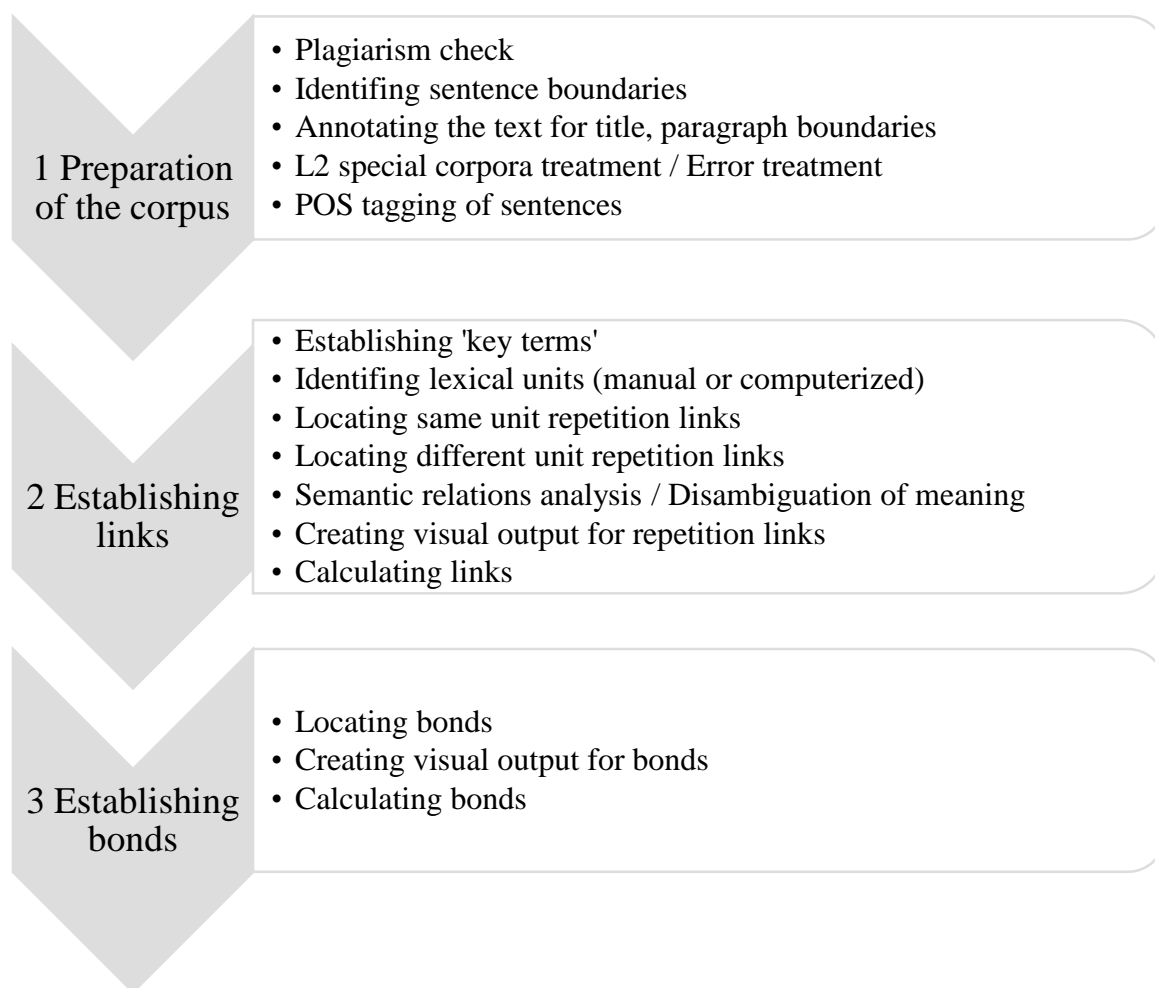


Figure 24. The steps of the new LRA model

8.2 Phase 1: Preparation of the corpus

8.2.1 Plagiarism check

As the first step in the preparation of the corpus, it is essential to establish the originality of each text. The first reason for this is to ensure that the academic integrity policy of the university is maintained. Secondly, due to the fact that the analysis draws on lexical features of the text, it is essential to make sure the text was not copied from another source. It could be seen in the Research

design section of this project (5.2.5.2), where several methods were tried out and some failed, that it is a rather difficult procedure, especially with regard to summaries. In the case of the integrated compare/contrast essays, where students read source texts to reflect on, the suggested program specializing on text similarity recognition is Turnitin⁴¹, or a similarly commercially available program.

As far as summaries are concerned, three methods were experimented with during the assessment phase of the initial corpus within content similarity evaluation (Section 5.2.5). The two manual methods described in literature could not be implemented due to the length of the source text. In the case of automated textual analysis, text originality parameters need to be set: how many percentages of text similarity, and/or how many word strings (adjacent words in a sentence in a row) are still acceptable. This is a special consideration in the case of summaries. The suggested method for testing existing text similarity tools is described in Section 5.2.5.2, but in this respect, further testing of available programs is necessary, particularly analyzing how the program rates the original text – copied version vs. original text – paraphrased version dimensions.

8.2.2 L2 special corpora treatment / Error treatment

While teachers correct essays, they look for lexical, grammatical, structural and mechanical errors. Before a learner corpus is handed over to a program for textual analysis, however, decisions need to be made about error corrections: whether to change anything in the text or not; and if errors are corrected: what and how to correct. In corpus linguistics, learner errors are analyzed using two alternative approaches: Contrastive Interlanguage Analysis or Computer-aided Error Analysis (Granger, 2002). The first approach analyzes the differences between e.g., native and non-native corpora, concentrating on norms and deviations from norms.

⁴¹ <http://turnitin.com/>

The second approach uses a special computer program to analyze errors systematically, which presupposes devising a taxonomy of errors before tagging all the linguistic violations in the texts. Granger and Wynne's (1999) study claims that using a common text retrieval program, e.g., WordSmith Tools⁴², might prove to be rather useless in learner texts even for the most typical searches, such as type/token ratio counts, due to the high rate of non-standard forms present in learner corpora (Granger & Wynne, 1999).

The above mentioned corpus linguistic approaches can be used when the aim is to draw conclusions based on the learner mistakes in the text. The aim of this research was, however, to find semantic links in the learner texts, therefore, several errors had to be eliminated. To this end, the following treatment is suggested for computer-aided LRA:

- Errors of mechanics regarding the sentence should be corrected: each sentence initial word needs to be capitalized, each sentence should end in a period.
- Spelling errors need to be corrected otherwise links will be missed.
- Multiword units which the writer clearly meant as a one-word unit need to be hyphenated or otherwise signaled to the program to interpret as a single word (e.g. *non-native-speaker-English-teacher* = *non-NEST*).
- Semantic errors should not be altered (e.g. errors of style, register)
- Lexico-grammatical mistakes need to be considered on a one-by-one basis.
- Errors in syntax do not need treatment.
- Errors on the discourse level should not be altered.

⁴² <http://www.lexically.net/wordsmith/>

8.2.3 POS tagging

POS tagging is an important part of sense disambiguation. As it was described in Section 2.3.3, the English language contains a great number of polysemous and homonymous nouns. In order to find the appropriate link pairs, the right meaning of the word needs to be selected. A POS tagger, which marks the syntactical function of the words may assist in the selection process. For instance, the word *dance* in the sentence *this was our last dance* stands as a noun, in *we dance all night* stands as a verb, and in *I would like to dance* stands as an infinitive. Thus, the coded link pairs could range from simple same unit repetition (exact mention) to derived same unit repetition (e.g., when the word class changes).

So as to be able to suggest a POS tagger available for the analysis, I made a brief experiment with three online tagger applications⁴³. I tested which program is able to recognize some ‘problematic cases’ in derivations. The following short sentences were entered into the taggers:

This was a good read. John's painting is hung in the hall. John's careful painting of the wall made me jealous. John carefully painting the wall made me jealous. Painting can be dangerous. You can bank on it.

Results for 1st POS tagger, with an asterisk for the error

1. **This/DT was/VBD a/DT good/JJ read/VBN./ (VBN= verb, past participle)*
2. *John/NNP 's/POS painting/NN is/VBZ hung/VBN in/IN the/DT hall/NN ./.*
3. **John/NNP 's/POS careful/JJ painting/NN of/IN the/DT wall/NN made/VBD me/PRP jealous/JJ ./.* (NN = noun, singular or mass)
4. *John/NNP carefully/RB painting/VBG the/DT wall/NN made/VBD me/PRP jealous/JJ ./.*
5. ***Painting/VBG** can/NN be/VB dangerous/JJ ./.* (VBG=verb, gerund or participle!)
6. *You/PRP can/MD bank/VB on/IN it/PRP ./.*

⁴³ <http://nlpdotnet.com/services/Tagger.aspx/> <http://textanalysisonline.com/nltk-pos-tagging>
<http://ucrel.lancs.ac.uk/cgi-bin/claws71.pl>

Results for 2nd POS tagger

1. *This*|DT was|VBD a|DT good|JJ read|NN .|.
2. *John*|NNP 's|POS painting|NN is|VBZ hung|VBN in|IN the|DT hall|NN .|.
3. **John*|NNP 's|POS careful|JJ painting|NN of|IN the|DT wall|NN made|VBD me|PRP jealous|JJ .|. (NN = noun, singular or mass)
4. *John*|NNP carefully|RB painting|VBG the|DT wall|NN made|VBD me|PRP jealous|JJ .|.
5. *Painting|NNP can|MD be|VB dangerous|JJ .|. (NNP= proper noun)
6. *You*|PRP can|MD bank|VB on|IN it|PRP .|.

Results for 3rd POS tagger

1. ----- PUN
2. 1. *This*_DT0 was_VBD a_AT0 good_AJ0 read_NN1 ._SENT ----- PUN
3. 2. *John*_NP0 's_POS painting_NN1 is_VBZ hung_VVN in_PRP the_AT0 hall_NN1 ._SENT ----- PUN
4. 3. *John*_NP0 's_POS careful_AJ0 painting_NN1 of_PRF the_AT0 wall_NN1 made_VVD me_PNP jealous_AJ0 ._SENT ----- PUN
5. 4. *John*_NP0 carefully_AV0 painting_VVG the_AT0 wall_NN1 made_VVD me_PNP jealous_AJ0 ._SENT ----- PUN
7. 5. *Painting*_NN1 can_VM0 be_VBI dangerous_AJ0 ._SENT ----- PUN
8. 6. *You*_PNP can_VM0 bank_VVI on_PRP it_SENT ._PUN

The first two applications misidentified sentence 3 in which *painting* referred to an ongoing action. The first POS tagger did not recognize that *read* in that context is a noun premodified by a determiner and an adjective. This tagger uses the Penn Treebank tagset⁴⁴, whose accuracy is about 97.1 percent on ordinary texts. According to the tagset description, VBG means verb, gerund *or* participle, making no distinction between the three. As a consequence for our research, this tagger cannot distinguish between derived and inflected verbs therefore, cannot distinguish between same unit repetition and derived unit repetition. The second POS tagger misinterpreted *painting* in a sentence initial position, identifying it as a proper noun. The third tagger (CLAWS) was developed by UCREL (University Centre for Computer Corpus Research on Language, Lancaster). Based on the results, this proved to be the most reliable of the three, thus this is suggested for the analysis.

⁴⁴ https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

8.2.4 POS tagging for lower level L2 texts

The summary and the compare/contrast essay corpora were written by near-proficiency level language learners. However, texts written by lower level students might contain errors which can be an obstacle for the POS taggers we analyzed above. Dickinson and Ragheb (2013) give a detailed description of their recommended POS annotation practice for a learner corpus with the annotation tool *Brat*⁴⁵ devised for learner corpus.

Dickinson and Ragheb's assumption is that the linguistic violations in any learner text are characteristic of the learner's interlanguage (i.e. the stage of language development the language learner is at the moment of writing), therefore, annotation should focus on the context and not on the mistake. Their goal is to mark syntactic and morpho-syntactic information with as little error encoding as possible. Their advice for annotators is:

“Try to assume as little as possible about the intended meaning of the learner. ... Specifically, do the following: (1) fit the sentence into the context, if possible; (2) if not possible, at least assume that the sentence is syntactically well-formed (possibly ignoring semantics) and (3) if that fails, at least assume the word is the word it appears to be (i.e., do not substitute another word in for the word which is present).” (p.3).

In other words, do not try to guess what the learner wanted to mean because this is prone to mistakes. Given that in some cases several interpretations of the same sentence are possible, a suggested annotation practice is illustrated in Table 36.

⁴⁵ <http://brat.nlplab.org/>

Interpretation	Analysis of interpretation
<i>I think that my life is mine. <u>In Korea, however Parents's think is very important.</u></i>	
not suggested: In Korea, however, what parents think is very important.	parents (common noun, plural, nominal) think (verb, present simple, singular) <i>Parents 's</i> misformed possessive <i>what</i> missing word
suggested: In Korea, how parents think is very important	parents (common noun, plural. nominal) think (verb, present simple, singular) 's minor orthographical issue <i>however</i> lexical issue; used instead of <i>how</i> the sentence is well-formed: <i>Parents 's</i> = subject of <i>think</i> <i>however</i> = adjunct of <i>think</i>

Table 35. Two interpretations of a learner sentence and their analyses (based on Dickinson & Ragheb, 2013)

For our purposes this suggested treatment of linguistic violations seems a viable option, because if we put extra words into the text on assumption of what the writer intended, we might put a link into the text where it was originally not intended. Further detailed descriptions of error coding in learner corpora can be found for instance, in a Cambridge Learner Corpus error analysis study by Nicholls (2003).

8.2.5 Using WordNet with the existing taxonomy

For establishing the types of links other than same unit repetition, WordNet is suggested as an online option. Still, the types of links in the existing LRA model need further study. Károly (2002) already revised Hoey's (1991) taxonomy in order to make it less ambiguous and to better serve the coding process. However, the question arises, whether the taxonomy should be aligned with the WordNet thesaural categories for the sake of making the analysis more viable by computerized means. This change to Károly's categorization, nevertheless, will raise difficulties in determining the 'right' ratio for the different link categories because the WordNet thesaurus

defines more types of semantic relations (Fellbaum, 1998) than Károly's original model does. Whether employing more categories would help or hinder the LRA, will have to be resolved by further research.

8.2.6 Using WordNet with errors in a learner corpus

How WordNet can be used with a learner corpus needs further testing. As it is described in the Frequently Asked Questions section of WordNet, the program looks for the base forms of words "stripping common English endings until it finds a word form present in WordNet". It also "assumes its input is a valid inflected form. So, it will take "childes" to "child", even though "childes" is not a word"⁴⁶. What is important for this research is that although pre-treatment of the corpus is necessary for learner errors, some spelling mistakes such as overgeneralizing certain existing grammar rules (e.g., the *-s* vs. *-es* plural suffix) may not cause problems for WordNet in the identification of words.

8.3 Phase 2: Finding links

8.3.1 Theoretical considerations: altering the taxonomy

A key decision is whether the taxonomy should be changed or should remain intact. This needs to be decided in light of the results in Stages 1 and 2 in connection with the ratio of lexical repetition types. Even though the sample size did not make it possible to find significant differences, both Károly's (2002) results based on argumentative essays, and the results of this research (in both stages) indicate that there is a strong positive correlation between the frequency of derived repetition and cohesion, and as a consequence, between the frequency of derived

⁴⁶ <https://wordnet.princeton.edu/faq>

repetition and coherence, thus having a major influence on discourse quality. An example for this could be the following two adjacent sentences:

(S5) *O'Reily **decided** to change flight timetables in order to increase turnaround times.*

(S6) *This **decision** resulted in more income for the company.*

These two sentences are connected by a derived repetition link which is very common in academic writing. More typical example in adjacent sentences could be e.g., *X claimed that ... -- This claim* As mentioned above, this structure as a cohesive device is typical between two **adjacent** sentences in academic discourse. In Károly's (2002) model, there is a clear distinction between simple and derived repetition, the basis of which is the distinction between inflected and derived word forms. These appear as two distinct categories in the taxonomy. She claims that high-rated essays contain more derived repetitions, therefore the revised taxonomy to be developed for large-scale analysis should keep the original two categories, unless large-scale research does not suggest otherwise. The concordancing program used in Stage 2 however, could not make a distinction between the two repetition types.

If no alternative concordancer is found which can distinguish between inflection and derivation, two possible solutions are suggested: (1) either sacrifice this distinguishing function of the model, or (2) trace for derived repetition links between adjacent sentences with a sentence parser in order to observe whether they contain such elements. Following this, the ratio of such links should be collated with the higher category same unit repetition links (derived repetition links in adjacent sentences divided by ratio of same unit repetition links). Large-scale investigations then will be able to inform us whether the observation of this discourse feature indicates statistically significant differences between high and low-rated texts or not.

8.3.2 Introducing the concept of ‘key term’ into the coding process

In Stage 1, one of the first issues to be solved was how to code the proper noun *Ryanair*. As already described in the Research design chapter (5.2.6), the task was to summarize Ryanair’s history and business model. Therefore, it contained several mentions of the proper noun, as well as synonyms to describe the company (*company, firm, airline carrier, etc.*). Seeing this, the coders contemplated that the frequency of the category instantial relations will grow ‘out of proportion’ compared to the other types of links in the summaries. The rather ad-hoc decision of the coders was to disregard *Ryanair – company* link pairs as instantial relations, which would have been the procedure if Károly’s taxonomy was to be strictly followed. Rather, they chose to consider them simple synonyms ‘for these texts only’. The analysis of the findings revealed that the decision to regard such links **synonym** was justified because the lexical repetition patterning of the texts showed a similar ratio of the link types as in Károly’s previous analysis, and it was in line with the intentions of the summary writers.

At the time of coding, this solution was seen by the coders as a diversion from Károly’s original model, and a negative feature emerging due to the characteristics of the particular corpus, even raising doubts about corpus representativity. However, as research evolved, it became clear by examining summary tasks across various disciplines that it is highly likely that students are asked to summarize a source in which the main topic is specific and is referred to with a proper name (i.e., the name of a person, a company, an action, a theory, a model): in other words, a specific instead of a general concept. Therefore, the specific nature of the topic was not as unique as the coders had thought. The following are examples from different disciplines of the possible topics that students need to summarize, all involving a proper name, and the possible synonyms writers would use as substitution.

Domain	Topic	Possible links
History	Queen Elizabeth I.	queen, emperor
Biology	Gibson Assembly	approach, method
Informatics, Business	ASUS VivoBook S200E	notebook, computer

Table 36. Proper names as possible topics in four disciplines

Further investigation of Károly's original analysis revealed that the task prompt writers had to reflect on in their argumentative essays focused on the ethics of journalism. The students had to develop one aspect of the text, which move implied making generalizations and the abstraction of ideas. The realization, that the input topic highly influences the repetition types, motivated further research and resulted in Stage 2 in selecting a domain offering more abstract topics, namely applied linguistics.

Given that topic specificity needs a unified treatment, the following analytical decision is suggested: if the topic is referred to as a proper noun, this noun should be considered a **key term**, and lexical repetition links this term enters into should be considered simple repetition (in cases of word-for-word repeating) or simple synonym (in cases of *proper noun—equation* mentioning). Only one proper noun lexical unit should be treated as key term in each text. The 'keyness' of a noun can be easily recognized by examining the title, which in academic discourse is informative, and as such, defines the topic and its controlling aspect(s), e.g. *the role of women in **Romeo and Juliet***. This example shows a multiword key term, which will be treated as one lexical unit.

8.3.3 Lexical unit identification in the case of multiword units

Proper identification of the units of analysis is key to gaining valid and reliable results in lexical repetition analysis. We distinguish between one-word lexical units and multi-word lexical units, the most problematic for this research being noun compounds. Most noun compounds are dictionary entries. According to Barzilay and Elhadad (1999), around 50,000 noun compounds are

included in WordNet such as *sea level*. They also mention collocations found in WordNet such as *digital computer*. However, there are two problems. The first is that, as Barzilay and Elhadad observe, the English language produces a great number of noun compounds and each new domain brings along more of their specific new noun compounds not present in WordNet. The second problem, already mentioned earlier, is that WordNet was trained on the Brown Corpus (a general corpus of English) not on special corpora for academic English or even more specific subcorpora based on professional registers. Specialist dictionaries and handbooks are necessary, just as it was done 'intuitively' in the coding phase.

Prószéky and Földes (2005) describe their comprehension assistant software as capable of identifying multiword expressions. According to their explanation, this was achieved by using several big capacity dictionaries and restructuring their entries by splitting up original entries, thus separating single and multiword expressions; the latter then could be entered as new headwords. This way the program became capable of detecting multiword expressions, which might seem a solution to the problem of identifying context-specific multiword units.

8.4 Special use of the model for academic summary writing

The new model can have two distinct uses in the academic summary writing process. Firstly, it can be used during the input text selection phase, when the main ideas need to be extracted from the source text, according to Hoey's (1991) original idea. By applying the steps (Establishing links and Establishing bonds), the central sentences can be collected and the resulting abridged version of the input text can substitute for the manual collection of the main ideas, which so far have been generated by teachers.

Alternatively, the main points of the text can be collected by teachers in the traditional way, and their decisions can be 'objectively' tested by applying the tool on the text. It has to be noted,

however, that this kind of main idea collection gives only valid results if the students need to write a whole-text summary. If the task is guided summary writing, the content points students need to select will probably differ from the main ideas of the input text, and in this case the central sentences in the input text will not coincide with the information required.

The second use of the model is similar to the one described in Stage 1 of this research, when it was used to distinguish the quality of summaries, observing the quantity and types of links and bonds within texts. However, using the computer-aided model it will also be possible to compare the organizations of both the input and output texts with the same analytical tool, which might reveal so far hidden similarities or differences in their lexical patterning.

8.5 Visual representation of links and bonds

Identified links and bonds can be illustrated in various ways. During the manual analysis (Section 5.2.6) a connectivity matrix was used, and its cells were filled with the link types found. Later this was converted into another matrix to represent the number of links, which was informative of where the link pairs appeared in the text. In Teich and Fankhauser's (2005) study the matrix format was substituted by indexing each sentence with link location numbering. Figure 25 shows part of their annotated data in text view. Even though this might be of use for a researcher, a language teacher or learner could not benefit from it fully in order to practise academic writing. Another straightforward option, the table format of the results (e.g., Table 33 in Section 7.2) would also be too complicated for educational purposes.

[1,0,6] It is obvious enough that linguists(1,1,1) in_general have been less successful in coping_with tone_systems(2,1,1) than with consonants or vowels. [2,0,0] No single explanation is adequate_to account_for this. [3,0,1] Improvement(3,1,1), however, is urgent, and at_least three things will be needed.

[4,0,31] The first is a wide-ranging sample of successful tonal(4,1,1) analyses(5,1,1). [5,0,2] Even beginning students(6,1,1) in linguistics are made familiar with an appreciable variety of consonant_systems(7,1,1), both in their general outlines and in many specific details. [6,2,2] An advanced student(6,2,2) has read a considerable number(8,1,1) of descriptions of consonantal_systems(7,2,2), including some of the more unusual types(9,1,1). [7,2,2] By contrast, even experienced linguists(1,2,2) commonly know no_more of the range of possibilities in tone_systems(2,2,2) than the over-simple distinction between register(2,3,2) and contour_languages(2,4,2). [8,1,0] This limited familiarity with the possible phenomena has severely hampered work with tone(4,2,2). [9,3,7] Tone(4,3,3) analysis(5,2,2) will continue to be difficult and unsatisfactory until a more representative selection of systems is familiar to every practicing field(10,1,1) linguist(1,3,3). [10,1,2] Papers(11,1,1) like these four(12,1,1), if widely read, will contribute importantly to improvement(3,2,2) of our analytic work.

Figure 25. Teich and Fankhauser's lexical repetition analysis links in text view (2005)

One solution seems to be visualizing the lexical net, similarly as in Hoey's (1991) study to show the pattern of bonded sentences, as in Figure 4, Section 2.2.2. An even more sophisticated option could be representing the actual text with either various lexical repetition link types highlighted, or central sentences highlighted. This way the building blocks of the text, such as the paragraphs, their boundaries, and the sentences with special discourse function could also be represented (Figure 26) within their discourse functions, and the teachers could comment immediately, for instance, on ill-formed, unconnected sentences by referring to their missing centrality which was identified by the model.

Synonyms

Synonyms, words that have the same basic meaning, do not always have the same emotional meaning. For example, the words stingy and frugal both mean "careful with money." However, calling someone stingy is an insult, but calling someone frugal is a compliment. Similarly, a person wants to be slender but not skinny, aggressive but not pushy. Therefore, you should be careful in choosing words because many so-called synonyms are not really synonymous at all.

Figure 26. Visual representation of topic sentence and conclusion sentence identified as central by LRA (text based on Oshima & Hogue, 2006)

8.6 Connecting the new LRA model to a cognitive framework

As part of L2 test validation, it is common practice to ensure that test takers' mental processes during an exam are similar to the mental processes they employ in target situations. In the same manner, in automated essay assessment research, when developing a preliminary analytical framework, validating automated essay scoring methods, or evaluating the construct coverage of a scoring engine, human scorers' decision making is observed to find out how they assess various essay features so as to better align the automated assessment tool to human rating (Cumming, Kantor, & Powers, 2001; Quinlan, Higgins, & Wolff, 2009; see also Section 3.3.3. for more details). Therefore, I also felt necessary to contextualize the new LRA model by connecting its steps to an existing cognitive framework.

In order to contextualize our new model, we again turn to Weir's (2005) socio-cognitive framework (described in Section 3.2.1) more precisely, we use the framework devised by Khalifa and Weir (2009) to describe the mental processes during reading⁴⁷. A parallel can be drawn between the processes described in their framework and the steps of our computer-aided lexical repetition model, Table 35. It shows the cognitive processes during reading in the first column. In the second column these stages are operationalized for a reading test situation (Bax, 2013), and are explained further in column three. In the last column the steps of our new, computer-aided lexical repetition model are linked to each cognitive process in the Khalifa and Weir model. Interestingly, the reading processes in column one appear in ascending order of difficulty, word recognition being the least cognitively demanding for human readers, whereas difficulty is

⁴⁷ We use the framework now devised for *reading* and not *writing*, as we previously did, because the modules need to analyze finished texts; we are not interested in the writing process.

reversed in the case of computer-aided analysis: identifying and disambiguating semantic relations being the most difficult, and establishing links and bonds being a routine operation for a computer.

Processes of reading (Khalifa & Weir, 2009)	Operationalized processes (adapted by Bax, 2013)	Explanation of processes	Steps of the new, computer-aided LRA model
word recognition	word matching, word-level	reader identifies same word in question and text	identifying lexical units establishing same unit repetition links
lexical access	synonym, antonym and other related word matching, word-level	identifying word meaning and word class	establishing different unit repetition links establishing key terms
syntactic parsing	grammar/syntax parsing, clause/sentence-level	reader disambiguates word meaning and identifies answer	POS tagging of sentences disambiguation of meaning
establishing propositional meaning	establishing propositional meaning, sentence-level	reader establishes meaning of a sentence	~ establishing links
inferencing	inferencing, sentence/paragraph / text level	beyond literal meaning to infer further significance	~ establishing bonds
building a mental model	building a mental model, text-level	using several features of text	~ establishing bonds
creating a text level representation	understanding text function, text-level	using genre knowledge to identify text structure and purpose	~ establishing bonds
creating an intertextual representation	not in test situation, between texts	comparing texts	not relevant, but present in Hoey's (1991) original model: establishing links/bonds between texts

Table 37. The contextualization of the new LRA model in this study: the parallel processes between the model and the cognitive processes during reading (Khalifa & Weir, 2009), with explanations

8.7 The scope and limitations of the new LRA model

It is important to emphasize that this lexical repetition analysis model is designed as an aid intending to gain results in connection with cohesion, a text-internal concept. It is out of the scope of this model to address coherence, a concept which is text-external. In other words, the lexical repetition links within the text, whose patterns the model is attempting to capture, are overt, 'countable' cohesive links. On the other hand, coherence is described in this study as the interaction between the reader and the text, and is therefore seen as a subjective concept, and as such, the model does not intend to interpret it.

For the same reason, the design is computer-aided rather than fully automated, observing lexical links as data, and disregarding other aspects of discourse quality, such as syntactical, grammatical and stylistic variables, or their interactions. However, given that human readers' overall quality judgements on texts are influenced by cohesion, the model aims at positively correlating with readers' overall quality judgements in this respect.

9 Conclusions

9.1 Summary of main results

In comparison with the abundance of research on the discourse-cohesive function of repetition, there is still a dearth of empirical research which utilizes a theory-driven ‘objective’ analytical tool devised for studying the text-organizing function of lexical repetition and, at the same time, predicting ‘subjective’ rater judgement of discourse quality. Such is Károly’s (2002) analytical tool, which is a revised and extended version of Hoey’s (1991) Repetition Model used in a Hungarian EFL academic context.

Two of the principle aims of this study were to extend the use of Károly’s (2002) lexical repetition model to (1) the academic summary and (2) the compare/contrast genres in order to test whether the analytical tool can predict teachers’ judgement regarding cohesion/discourse quality. This study revealed that Károly’s model is suitable for the analysis of the text-organizing role of lexical repetition in both genres (Sections 6 and 7). It revealed that the structures of high-rated summaries and compare/contrast essays were different from low-rated ones: in both genres the main ideas were organized into sentences with special discourse functions, such as the theses and topic sentences.

The third main aim of this study was to alter the analytical tool to make it apt for large-scale analysis of EFL student corpora in order to be able to design the steps and modules necessary for computer-assisted LRA. This last stage of the study was motivated by and based on the results gained in the previous stages. Its basic assumption was that large-scale, computerized application of the tool is necessary but at the moment not resolved. The research methodology in this part was both theoretical and experimental. Its main aim was to reveal whether it is necessary to alter any features of the analytical tool for the large-scale analysis of EFL student corpora. First the question

of preparing the students' text prior to the computer-based analysis was addressed. Next, the stages (modules) of the computer-aided analysis were identified, and these were broken down into steps necessary for the analysis. This was followed by matching the appropriate applications for the steps wherever such were available.

The investigation revealed that several applications can be used to help computer-aided lexical repetition analysis. It can be concluded that the coding process can be automatized if (1) the EFL learners' texts are stripped of errors of mechanics (for sentence recognition, Section 8.2.2) and spelling (in order not to miss links, Section 8.2.2), (2) the unit of analysis is appropriately identified (in the case of multi-word units, Section 8.3.3), (3) word sense disambiguation is resolved (so as to identify the right meaning of words in the sentences, Sections 8.2.2, 8.2.5).

9.2 Pedagogical implications

This tool, if ever it is fully complete, will never take the place of a teacher. It was not designed to judge the overall quality of essays because for this, I believe, some human intervention is necessary. The tool designed in this project, however, might help teachers and students in several ways. Firstly, it can aid teachers to get an objective picture of a student's text with regard to cohesion, to be more precise, with regard to how lexical repetition patterns cluster and thus organize discourse. We have seen in Section 3.3.4.3, where feedback comments on students' texts were analyzed, that even experienced essay raters were struggling in formulating their opinions when referring to coherence and cohesion, and their treatment of these two concepts was not systematic. Furthermore, because the concepts of lexical cohesion and lexical repetition were not mentioned at all in the teachers' comments, we may conclude that teachers are not entirely aware of how these two concepts work in writing: they only notice errors of cohesion when the writer violates some rules related to it.

Therefore, the designed new lexical repetition analysis model might provide information on missing or misplaced cohesive links, which might further help the teacher in her decision on scoring discourse cohesion. In the same fashion, the tool might be used as a demonstration of discourse patterning in students' texts, visualizing the lexical net created by links and bonds. It can thus be used as a self-study aid.

A teacher might incorporate the feedback of the tool into her explicit teaching of various EFL genres. For instance, a five-step scaffolding instruction is suggested for teaching academic arguments (which can also be generalized to other types of writing), with the stages (1) building the context, (2) modeling and deconstructing texts, (3) constructing texts jointly, (4) constructing texts independently and (5) linking related texts (Bacha, 2010). The tool might be used in all of these five stages.

9.3 Limitations

Two main factors limited this study: the small size of the sample and the variables of the task. Due to the fact that the original model was devised for manual analysis, only a limited number of student writing could be analyzed. Therefore, although several interesting results were revealed by the model, statistical significance could not be calculated due to the small sample size, only certain tendencies could be observed. Such finding was, for instance, that perhaps contrary to assumption, high-rated summaries contained not only a higher number of simple synonymy and derived opposites, but also more simple repetition links.

As far as task variables are concerned, the analytical tool in its present form may predict subjective perceptions of the quality of the type of summary observed, however, no conclusions could be drawn on its reliability and validity in cases when the original document to be abridged uses a different narrative form. Similarly, the influence of the length limit of summaries were not

examined either. In the case of the compare/contrast essay corpus, the uneven ratio of the two essay patterns (block and point-by-point patterns) made it impossible to draw reliable conclusions on whether the model can distinguish between the two patterns with regard to the clustering of bonds according to patterning. These limitations motivated the study to investigate how to apply the model on larger corpora.

9.4 Terminology issues

The interdisciplinary nature of the work required to read the literature and use approaches derived from three distinct areas: **discourse analysis** for the theoretical background, **language pedagogy** for describing assessment practice, and **language technology** for automated text analysis; each bringing along an entirely different set of technical vocabulary. To make matters worse, some common terms refer to different concepts in these areas. An example for the clash between linguistic terms and teachers' jargon was the genre – text type distinction: the latter is widely used in books for teachers, however, it is seen as too vague for a theoretically based paper. Terminology in automation studies which means *evaluation* were *human scoring*, *rater judgement*, *assessing writing*, and *judging essay quality*; whereas in a pedagogical context they are called *correcting essays*, *giving feedback*, *summative or formative assessment of writing*, and *grading assignments*. Some terms needed to be redefined and some had to be discarded. I hope, this did not cause distress for the readers.

9.5 Suggestions for further research

This multidisciplinary study aimed to contribute with new results to the fields of (1) applied linguistics, more closely discourse analysis and corpus linguistics; (2) language pedagogy, especially for the teaching and evaluating EFL academic writing; and (3) computer science, for

educational software development. The following section suggests further theoretical and methodological investigations within these three fields regarding lexical repetition research.

The first issue that would deserve further investigations is the treatment of collocations (analyzed in Section 2.2.1). These are non-systemic semantic relations which are excluded from the analysis. Morris and Hirst (2004) report a study where readers had to identify words in a general news article, which, in their view, were semantically related. With 63% agreement the result showed that the readers identified word **groups** and not word **pairs** when they had to give labels for the relations. This finding might suggest that perhaps cohesion is perceived in a formation which is different from links. Hoey (1991) already attempted to describe the nature of these formations with the **link triangle/the mediator missing** concept (Section 2.3.7), but did not further elaborate on this idea.

Another interesting finding in Morris and Hirst (2004) is that most identified word **pairs** were collocations. These relations, which represent world knowledge, are perceived strongly by the reader as semantically related. Even though it is a fact that collocations mostly appear within the same sentence, it would be interesting to study to what extent such (rather intra-sentential) semantic relationships influence discourse cohesion, and how these could be incorporated into the proposed model.

The next theoretical issue to consider is the language level of the writers and its consequences for the written product with regard to the present lexical repetition model. Among the many possible difficulties that might arise from learner errors, only the issue of faulty sentence creation is mentioned now, as a problem area. Several language learners violate the two basic formal rules of sentence building, namely that sentences should start by a capital letter and end in a period, question mark, or exclamation mark. This type of error is prominent particularly below

levels IELTS 5.5 / CEFR B1, and has to be manually corrected, although the sentence boundaries are not always certain.

Apart from violating the formal rules such as punctuation, students also make mistakes in writing by creating sentence fragments, i.e., they leave out the finite verb from the sentence. This appears to be a recurring problem, particularly in the case of compound sentences, when students mistakenly identify the participle for a verb. When such basic rules of sentence building are violated, the analytical tool might not be able to gain valid results, given that inter-sentential links provide the basis of lexical repetition patterns.

A further reason for not gaining valid results when observing inter-sentential links might be that sentence creation is only partially based on fixed compulsory elements: there is also room for writer creativity. The same information content can be packaged into one sentence or divided between two sentences. For example, a compound sentence with *and* can be rewritten as two separate sentences with the same meaning spread out in two sentences using a connective, such as *moreover*, *furthermore*, etc., (see Section 7.4). This is a key issue if we want to analyze text with a tool based on inter-sentential relations. The same problem seemed to arise in the case of a research study utilizing Latent Semantic Analysis (Landauer, Laham, & Foltz, 2003), which analytical tool assesses semantic relatedness between adjacent sentences using vector-based similarity.

The last area for further research is connected to language technology. During this research a number of existing programs were analyzed, typically those which have already been reviewed in scientific journals, or the ones that offer publicly available manuals. Such were, for instance, Concordancer 3.3; Coh-Metrix; several POS taggers; or WordNet. It is possible, however, that other commercially available programs exist, perhaps in modular format, which might be suitable

for certain steps of this analytical process. If not, such a modular program can be built aligned with the newly designed LRA model. Especially if it graphically represents links and bonds, such program can be sold as a self-contained product or an online writing tool to assist academic writing teachers and their EFL students.

References

- Adorján, M. (2011). *Predicting rater appeal of summaries based on Károly's (2002) study of the text organizing role of lexical repetition*. Unpublished doctoral seminar paper. Eötvös Loránd University, Budapest, Hungary.
- Adorján, M. (2013). Explorations in lexical repetition analysis: The outcomes of manual vs. computer automated research methods. *Working Papers in Language Pedagogy*, 7, 1–28.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment. *Language Teaching*, 34, 213-236.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Addison Wesley Longman.
- Anson, C. M., Schwiebert, J., & Williamson, M. M. (1993). *Writing across the curriculum: An annotated bibliography*. Westport, CT: Greenwood Press.
- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly*, 22(3), 331-346.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing*, 7, 1-16.
- Attali, Y. (2004). Exploring the feedback and revision features of Criterion. *Paper presented at the National Council of Measurement in Education, April 12-16*. San Diego, CA.
- Retrieved from
http://www.ets.org/Media/Research/pdf/erater_NCME_2004_Attali_B.pdf

- Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing, 18*, 191-208.
- Bacha, N. (2010). Teaching the academic argument in a university EFL environment. *Journal of English for Academic Purposes, 9*, 229-241.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
- Barzilay, R., & Elhadad, M. (1999). Using lexical chains for text summarization. In I. Mani, & M. T. Maybury (Eds.), *Advances in automatic text summarization* (pp. 111-121). Cambridge, MA: MIT Press.
- Bax, S. (2013). Readers' cognitive processes during IELTS reading tests: Evidence from eye tracking. *ELT Research Papers, 13*(6).
- Bazerman, C., Little, J., Chavkin, T., Fourquette, D., & Garufis, J. (2006). *Reference guide to writing across the curriculum. Reference guides to rhetoric and composition*. West Lafayette, Indiana: Parlor Press and The WAC Clearinghouse. Retrieved January 13, 2015, from http://wac.colostate.edu/books/bazerman_wac
- Bean, T. W., & Steenwyk, F. L. (1984). The effect of three forms of summarization instruction on sixth graders' summary writing and comprehension. *Journal of Reading Behavior, 16*(4), 297-306.
- Belcher, D., & Hirvela, A. (2001). *Linking literacies: Perspectives on L2 reading-writing connections*. Ann Arbor, MI: The University of Michigan Press.
- Ben-Simon, A., & Bennett, R. E. (2007). Toward more substantively meaningful automated essay scoring. *The Journal of Technology, Learning, 6*(1), 1-47. Retrieved from

<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Toward+More+Substantively+Meaningful+Automated+Essay+Scoring#0>

- Berber Sardinha, T. (2000). *Investigating discourse organisation in corpora*. Retrieved from DIRECT Papers 43: <http://www2.lael.pucsp.br/direct/DirectPapers43.pdf>
- Biber, D. (1988). *Variations across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20.
doi:10.1016/j.jeap.2010.01.001
- Bonzo, J. (2008). To assign a topic or not: Observing fluency and complexity in intermediate foreign language writing. *Foreign Language Annals*, 41(4), 722-735.
- Bridgeman, B., & Carlson, S. B. (1983). *Survey of academic writing tasks required of graduate and undergraduate foreign students*. ETS Research Report. Educational Testing Service.
- Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education*, 25(1), 27-40.
- Brock, M. N. (1995). Computerized text analysis: Roots and research. *Computer Assisted Language Learning*, 8(2-3), 227-258.
- Brown, A. L., Day, J. D., & Jones, R. S. (1983). The development of plans for summarizing texts. *Child Development*, 54, 968-979.
- Brown, A. L., Smiley, S. S., & Lawton, S. (1978). The effects of experience on the selection of suitable retrieval cues for studying texts. *Child Development*, 49, 829-835.

- Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
- Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. *Written Communications*, 8, 533-556.
- Carter, M. (2007). Ways of knowing, doing, and writing in the disciplines. *College Composition and Communication*, 58(3), 385-418.
- Casanave, C. P. (2002). *Writing games: Multicultural case studies of academic literacy practices in higher education*. Mahwah, NJ: Lawrence Erlbaum.
- Chan, S. H. (2013). *Establishing the validity of reading-into-writing test tasks for the UK academic context*. PhD Dissertation: University of Bedfordshire.
- Chan, S. H., Wu, R. Y., & Weir, C. J. (2014). Examining the context and cognitive validity of the GEPT Advanced Writing Task 1: A comparison with real-life academic writing tasks. *The LTTC - CRELLA Collaboration Project RG-03*. Retrieved from <https://www.ltcc.ntu.edu.tw/lttc-gept-grants>
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge: Cambridge University Press.
- Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, 13(1), 75-83.
- Collier, A. (1994). *A system for automating concordance line selection*. Retrieved from Birmingham City University, Research and Development Unit for English Studies: http://rdues.bcu.ac.uk/publ/AJC_94_02.pdf
- Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics: International Journal of Human Communication*, 17, 301-3016.

- Connor, U., & Kaplan, R. (1987). *Writing across languages: Analysis of L2 text*. (U. Connor, & R. Kaplan, Eds.) Reading, MA: Addison-Wesley.
- Cooper, A., & Bikowski, D. (2007). Writing at the graduate level: What tasks do professors actually require? *Journal of English for Academic Purposes*, 6(3), 206-221.
- Coxhead, A. (2000). The New Academic Word List. *TESOL Quarterly*, 34(2), 213–238.
Retrieved from http://203.72.145.166/TESOL/TQD_2008/VOL_34_2.pdf
- Creswell, J. (2007). *Qualitative enquiry and research design*. London: Sage Publications.
- Crossley, S. A., & McNamara, D. S. (2009). Computational assessment of lexical differences in L1 and L2 writing. *Journal of Second Language Writing*, 18, 119–135.
doi:10.1016/j.jslw.2009.02.002
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. *32nd Annual Conference of the Cognitive Science Society*, 984–989.
- Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66–79. doi:10.1016/j.jslw.2014.09.006
- Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring l2 lexical growth using hypernymic relationships. *Language Learning*, 59(June), 307–334. doi:10.1111/j.1467-9922.2009.00508.x
- Crossley, S., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, 28(4), 561–580.
doi:10.1177/0265532210378031
- Crusan, D. (2010). *Assessment in the second language writing classroom*. Ann Arbor, MI: University of Michigan Press.

- Cumming , A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5-43.
- Cumming, A., Kantor, R., & Powers, D. (2001). Scoring TOEFL essays and TOEFL 2000 prototype writing tasks: An investigation into raters' decision making and development of a preliminary analytic framework. *TOEFL Monograph Series*, 1-95.
- Cumming, A., Rebuffot, J., & Ledwell, M. (1989). Reading and summarizing challenging texts in first and second languages. *Reading and Writing: An Interdisciplinary Journal*, 2, 201-219.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, Avon: Multilingual Matters.
- Cunningham, J. W. (1982). Generating interactions between schemata and text. In J. A. Niles, & L. A. Harris (Eds.), *New Inquiries in Reading Research and Instruction* (pp. 42-47). Rochester, NY: National Reading Conference.
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a Knowledge Representation? *AI Magazine*, 14(1), 17-33.
- de Beaugrande, R., & Dressler, W. U. (1981). *Introduction to text linguistics*. London: Longman.
- de Oliveira, P., Ahmad, K., & Gillam, L. (1996). A Financial News summarisation system based on lexical cohesion. Retrieved from http://antigo.univille.br/arquivos/4694_Paper3.pdf
- Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.
- Dickinson, M., & Ragheb, M. (2013). *Annotation for learner English. Guidelines, v.0.1*. Indiana University, Bloomington.

- Dikli, S. (2006). An overview of automated scoring of essays. *Journal Of Technology Learning And Assessment*, 5(1). Retrieved from <http://www.jtla.org>
- Douglas, D. (2000). *Assesing language for specific purposes*. Cambridge: Cambridge Univesity Press.
- Enkvist, N. E. (1990). Seven problems in the study of coherence and interpretability. In U. Connor, & A. M. Johns (Eds.), *Coherence in writing: Research and pedagogical perspectives* (pp. 9-28). Washington, DC: TESOL.
- Enright, M. K., & Quinlan, T. (2010). Complementing human judgment of essays written by English language learners with e-rater scoring. *Language Testing*, 27(3), 317-334.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge: Cambridge University Press.
- Flower, L., & Hayes, J. R. (1981). The Cognitive Process Theory of writing. *College Composition and Communication*, 32(4), 365-387.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Francis, G. (1990). Theme in the Daily Press. *Occasional Papers in Systemic Linguistics*, 4, 51-87.
- Freedman, S. W. (1979). How characteristics of student essays influence teachers' evaluations. *Journal of Educational Psychology*, 71(3), 328-338.
- Friend, R. (2001). Effects of strategy instruction of summary writing of college students. *Contemporary Educational Psychology*, 26(1), 3-24.

- Fulcher, G. (1989). Cohesion and coherence in theory and reading research. *Journal of Research in Reading*, 12(2), 146-163.
- Galbraith, D. (2009). Cognitive models of writing. *German as a Foreign Language*, 2-3, 7-22.
- Gardner, R. (1985). Text summarization deficiencies among older students: Awareness or production ability? *American Educational Research Journal*, 22(4), 549-560.
- Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7, 47-84.
- Gillett, A., Hammond, A., & Martala, M. (2009). *Successful academic writing*. London: Pearson Longman.
- Grabe, W. (2003). Reading and writing relations: Second language perspectives on research and practice. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 242–262). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139524810
- Graesser, A., McNamara, D., & Louwse, M. (2011). Methods of automated text analysis. In M. L. Kamil, D. Pearson, E. Moje, & P. Afflerbach (Eds.), *Handbook of reading research* (pp. 34-54). New York: Routledge.
- Granger, S. (2002). A bird's eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.
- Granger, S., & Wynne, M. (1999). Optimising measures of lexical variation in EFL learner corpora. In J. Kirk (Ed.), *Corpora galore*. Amsterdam and Atlanta: Rodopi.

- Grimm, S., Hitzler, P., & Abecker, A. (2005). *Knowledge representation and ontologies: Logic, ontologies and Semantic Web*. Retrieved from Data Semantics Laboratory Wright State University, USA: <http://daselab.cs.wright.edu/resources/publications/kr-onto-07.pdf>
- Guba, E. (1990). *The alternative paradigm dialog. The paradigm dialogue*. Newbury Park, CA: Sage.
- Hale, G., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1995). A study of writing tasks assigned in academic degree programs. *ETS Research Reports Series*(2), 1-61. doi:10.1002/j.2333-8504.1995.tb01678.x
- Hale, G. A., Taylor, C., Bridgeman, B., Carson, J., Kroll, B., & Kantor, R. (1996). *A study of writing tasks assigned in academic degree programs*. ETS Research Report. Educational Testing Service.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell (Ed.), *In memory of J. R. Firth* (pp. 148-162). Longman.
- Halliday, M., A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A.K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: Palmer.
- Hammann, L. A., & Stevens, R. J. (2003). Instructional approaches to improving students' writing of compare-contrast essays : An experimental study. *Journal of Literacy Research*. doi:10.1207/s15548430jlr3502
- Hamp-Lyons, L. (2001). English for academic purposes. In R. Carter, & D. Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp. 126-130). Cambridge: Cambridge University Press.

- Hasan R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension* (pp. 181-219). Delaware: International Reading Association.
- Haswell, R. H. (2013). Teaching of writing in higher education. In C. Bazerman (Ed.), *Handbook of research on writing: History, society, school, individual, text*. New York, N.Y.: Routledge.
- Haswell, R., Donnelly, W., Hester, V., O'Neill, P., & Schendel, E. (2012). An annotated bibliography of writing assessment: Machine scoring and evaluation of essay-length writing. *The Journal of Writing Assessment*, 5(1). Retrieved June 3, 2013, from <http://www.journalofwritingassessment.org/article.php?article=58>
- Hearst, M. (1994). Multi-paragraph segmentation of expository text. In *ACL '94 Proceedings of the 32nd annual meeting on Association for Computational Linguistics* (pp. 39–50). Retrieved from <http://dl.acm.org/citation.cfm?id=981734>
- Hayes, J. (1996). A new framework for understanding cognition and affect in writing. In M. Levy, & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum.
- Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. *Review of Educational Research*, 56(4), 473-493.
- Hill, M. (1991). Writing summaries promotes thinking and learning across the curriculum- but why are they so difficult to write? *Journal of Reading*, 34(7), 536-539.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hoey, M. (1983). *On the surface of discourse*. London: George Allen and Unwin.

- Horowitz, D. M. (1986). What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20(3), 445–462. doi:10.2307/3586294
- Huang, L.-S. (2010). Seeing eye to eye? The academic writing needs of graduate and undergraduate students from students' and instructors' perspectives. *Language Teaching Research*, 14(4), 517–539. doi:10.1177/1362168810375372
- Huot, B. (2002). *(Re)articulating writing assessment for teaching and learning*. Logan, Utah, Utah State University Press.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30(4), 437-455.
- Hyland, K. (2006). *English for academic purposes. An advanced resource book*. London: Routledge.
- Johns, A. M. (1980). English and the undergraduate non-native speaker: Some directions for research. *ESL in Higher Education*, 2(2), 25-26.
- Johns, A. M. (1981). Necessary English: A faculty survey. *TESOL Quarterly*, 15(1), 51-57.
- Johns, A. M. (1985). Summary protocols of "underprepared" and "adept" university students: replications and distortions of the original. *Language Learning*, 35(4), 495-512.
- Johns, A. M., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. *Applied Linguistics*, 11(3), 253-271. doi:10.1093/applin/11.3.253
- Johns, A. M., & Swales, J. M. (2002). Literacy and disciplinary practices: opening and closing perspectives. *Journal of English for Academic Purposes*, 1(1), 13-28.
- Johnson, N. (1983). What do you do if you can't tell the whole story? The development of summarization skills. In K. Nelson (Ed.), *Children's language* (pp. 315-383). Hillsdale, NJ: Erlbaum.

- Jones, G. (2009). Lexical cohesion as a predictor of item difficulty. *EALTA Conference 2009*. Retrieved from <http://www.ealta.eu.org/conference/2009/docs/friday/Jones.pdf>
- Jordan, R. R. (1997). *English for academic purposes: a guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Károly, K. (2002). *Lexical repetition in text*. Frankfurt am Main: Peter Lang.
- Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing, 25*, 4–22. doi:10.1016/j.jslw.2014.05.005
- Khalifa, H., & Weir, C. J. (2009). *Examining reading: Research and practice in assessing second language reading, Studies in Language Testing 29*. Cambridge: Cambridge University Press.
- Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*, 161-179.
- Kirkland, M. R., & Saunders, M. A. (1991). Maximizing student performance in summary writing: managing cognitive load. *TESOL Quarterly, 25*(1), 105-121.
- Kormos, J., & Csölle, A. (2000). *A brief guide to academic writing*. Budapest: Műszaki Kiadó.
- Kyle, K., & Crossley, S. A. (2014). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly, 0*(0), 1–30. doi:10.1002/tesq.194
- Laki, L. J. (2011). Statisztikai gépi fordítási módszereken alapuló egynyelvű szövegelemző rendszer és szótövesítő. [Monolingual text analyzer and lemmatizer based on statistical computational methods.]. *Paper presented at the 8th Hungarian Computational Linguistics Conference*. Szeged.

- Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. D. Shermis, & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87-112). Mahwah, NJ: Erlbaum Associates.
- Langendoen, D. T. (2009). Reviewed works: In Memory of J.R. Firth by J. R. Firth ; C. E. Bazell ; J. C. Catford ; M. A. K. Halliday ; R. H. Robins. *Foundations of Language*, 5(3), 391-408. Retrieved from <http://www.jstor.org/stable/25000391>
- Lautamatti, L. (1987). Observations on the development of the topic in simplified discourse. In U. Connor, & R. B. Kaplan (Eds.), *Writing across languages: Analysis of L2 text* (pp. 87-114? 92-126?). Reading, MA: Addison-Wesley.
- Leki, I., & Carson, J. (1994). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly*, 29, 235-260.
- Leki, I., & Carson, J. (1997). "Completely different worlds" EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly*, 31(1), 39-69.
- Leki, I., & Carson, J. G. (2012). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101.
- Lenski, S. D., & Johns, J. L. (1997). Patterns of reading-to-write. *Reading Research and Instruction*, 37(1), 15-38.
- Lenski, S. D., & Johns, J. L. (1997). Patterns of reading-to-write. *Reading Research and Instruction*, 37(1), 15-38.
- Li, J. (2014). The role of reading and writing in summarization as an integrated task. *Language Testing in Asia*, 4(3), 1-19. Retrieved from <http://www.languagetestingasia.com/content/4/1/3>

- Mani, I. (2001). *Automatic summarisation*. Amsterdam: John Benjamins.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281.
- Martin, J. R. (1992). *English text: System and Structure*. Amsterdam: John Benjamins.
- McKeown, K., Hirschberg, J., Galley, M., & Maskey, S. (2005). From text to speech summarization. *ICASSP, Special Session on Human Language Technology Applications and Challenges for Speech Processing*. Retrieved December 8, 2010, from http://www.cs.columbia.edu/nlp/papers/2005/mckeown_al_05a.pdf
- McNamara, D. S., Louwse, M. M., Cai, Z., & Graesser, A. (2005, January 1). Coh-Metrix version 1.4. Retrieved December 8, 2012, from <http://cohmetrix.memphis.edu>
- McNamara, D., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1-43.
- McNeil, J., & Donant, L. (1982). Summarization strategy for improving reading comprehension. In J. A. Niles, & L. A. Harris (Eds.), *New Inquiries in Reading Research and Instruction* (pp. 215-219). Rochester, NY: National Reading Conference.
- Messick, S. (1989). Validity. In L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York: Macmillan.
- Mitchell, W. (1996). Why comparisons are odious? *World Literature Today*, 70(2), 321-324.
- Monacelli, C. (2004). A tightrope with a net. *Meta: Translators' Journal*, 49, 891-909.
- Moore, T., & Morton, J. (1999). Authenticity in the IELTS Academic Module Writing Test: A comparative study of Task 2 items and university assignments. *IELTS Research Reports*, 2.

- Moore, T., Morton, J., & Price, S. (2011). Construct validity in the IELTS Academic Reading Test : A comparison of reading requirements in IELTS test items and in university study Authors. *IELTS Research Reports*, 11(1998), 1–86.
- Morris, J., & Hirst, G. (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17(1), 21-48.
- Myers, G. (1991). Lexical cohesion and specialized knowledge in science and popular science texts. *Discourse Processes*, 14, 1-26.
- Myford, C. M., & Wolfe, E. W. (2003). Monitoring rater performance over time: A framework for detecting differential accuracy and differential scale category use. *Journal of Educational Measurement*, 46(4), 371-389. doi:DOI: 10.1111/j.1745-3984.2009.00088.x
- Nakamura, Y. (2004). A comparison of holistic and analytic scoring methods in the assessment of writing. *The Interface Between Interlanguage, Pragmatics and Assessment: Proceedings of the 3rd Annual JALT Pan-SIG Conference*. Tokyo. Retrieved May 2, 2012, from <http://jalt.org/pansig/2004/HTML/Nakamura.htm>
- Nation, I. S. (2001). *Learning vocabulary in another language*. New York: Cambridge University Press.
- Netzer, Y., Gabay, D., Goldberg, Y., & Elhadad, M. (2009). Gaiku: Generating haiku with word association norms. *Proceedings of the NAACL HLT Workshop on Computational Approaches to Linguistic Creativity* (pp. 32-39). Boulder, Colorado: Association for Computational Linguistics.
- Newby, D., & Penz, H. (Eds.). (2009). *Languages for social cohesion: Language education in a multilingual and multicultural Europe*. Strasbourg: Council of Europe Publishing.

- Nicholls, D. (2003). The Cambridge Learner Corpus - error coding and analysis for lexicography and ELT. *Lancaster University Computer Corpus Research on Language*, pp. 572-581.
Retrieved from <http://ucrel.lancs.ac.uk/publications/cl2003/papers/nicholls.pdf>
- Novák, A., Orosz, G., & Indig, B. (2011). Javában taggelünk. [We are a-tagging.] . *Paper presented at the 8th Hungarian Computational Linguistics Conference*. Szeged.
- Olson, B. (2013, February). *Academic writing across the disciplines*. Retrieved December 11, 2013, from Literary Design Collaborative: <https://ldc.org>
- Oshima, A., & Hogue, A. (2006). *Writing academic English* (4th ed.). White Plains, NY: Pearson Longman.
- Paltridge, B. (1996). Genre, text type and the language learning classroom. *50*(3), 237-243.
- Pearson, C. R. (1981). Advanced academic skills in the low-level ESL class. *TESOL Quarterly*, *15*(4), 413-423.
- Phillips, M. (1985). *Aspects of text structure: An investigation of the lexical organization of text*. Amsterdam: North-Holland.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111-129.
- Professionals Against Machine Scoring Of Student Essays*. (2013). Retrieved from Human Readers: http://humanreaders.org/petition/research_findings.htm
- Prószéky, G., & Földes, A. (2005). Between understanding and translating: A context-sensitive comprehension tool. *Archives of Control Sciences*, *15*(4), 625-632.
- Prószéky, G., & Miháltz, M. (2008). Magyar WordNet: az első magyar lexikális szemantikai adatbázis. [HunWordNet: the first Hungarian lexical semantic database]. *Magyar Terminológia*, *1*(1), 43-57.

- Purves, A. (Ed.). (1988). *Writing across languages and cultures*. Newbury Park, CA: Sage.
- Quinlan, T., Higgins, D., & Wolff, S. (2009). *Evaluating the construct-coverage of the e-rater scoring engine*. ETS Research Report. Educational Testing Service.
- Reid, J. (2015). *Assessing English proficiency for university study*. Basingstoke: Palgrave MacMillan.
- Reynolds, D. W. (1995). Repetition in nonnative speaker writing: More than quantity. *Studies in Second Language Acquisition*, 17(2), 185-209.
- Reynolds, D. W. (2001). Language in the balance: Lexical repetition as a function of topic, cultural background, and writing development. *Language Learning*, 51(3), 437-476.
- Rinehart, S. D., Stahl, S. A., & Erickson, L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 21(4), 422-438.
- Rosenfeld, M., Courtney, R., & Fowles, M. (2004). Identifying the writing tasks important for academic success at the undergraduate and graduate levels. ETS Research Report. Educational Testing Service.
- Sanders, T., Spooren, W., & Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2), 93-133.
- Scardamalia, M., & Bereiter, C. (1987). Knowledge telling and knowledge transforming in written composition. In S. Rosenberg (Ed.), *Advances in Applied Psycholinguistics: Reading, writing and language learning* (Vol. 2, pp. 142-175). Cambridge: Cambridge University Press.
- Schiver, K. A. (1989). Evaluating text quality. *IEEE Transactions on Professional Communication*, 32(4), 238-255.

- Schneider , M., & Connor, U. (1990). Analyzing topical structure in ESL essays: Not all topics are equal. *Studies in Second Language Acquisition*, 12(4), 411-427.
- Schriver, K. A. (1997). What document designers can learn from readers. *Dynamics in Document Design*, 443-495.
- Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature*, 4(1), 5–33. doi:10.1023/B:ESLL.0000033847.00732.af
- Seidlhofer, B. (1995). *Approaches to summarization: Discourse analysis and language education*. Tübingen: Narr.
- Seidl-Péché, O. (2011). *Fordított szövegek számítógépes összevetése [Computerized comparison of translated texts]*. Eötvös Loránd University, Budapest.
- Shaw, S., & Weir, C. J. (2007). *Examining writing: Research and practice in assessing second language writing*. *Studies in Language Testing* 26. Cambridge: Cambridge University Press.
- Simpson, J. M. (2000). Topical structure analysis of academic paragraphs in English and Spanish. *Journal of Second Language Writing*, 9(3), 293-309.
- Sinclair, J. (1998). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Spärck Jones, K. (2007). Automatic summarising: A review and discussion of the state of the art. *Mergent*, 4, 4-67.
- Spärk Jones , K. (1999). Automatic summarising: Factors and directions. In I. Mani, & M. T. Maybury (Eds.), *Advances in automatic text summarisation* (pp. 1-14). Cambridge, MA: MIT Press.

- Spivey, N. N. (1991). The shaping of meaning: options in writing the comparison. *Research in the Teaching of English*, 25, 390-418.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7-26.
- Stern, A. (1976). 'When is a paragraph?'. *College Composition and Communication*, 27(3), 253-257.
- Stubbs, M. (2004). Language corpora. In A. Davies, & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 106-132). Malden, MA: Blackwell Publishing.
- Swales, J. M. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J., & Feak, C. (1994). *Academic writing for graduate students*. Ann Arbor, MI: University of Michigan Press.
- Swales, J., & Feak, C. (2010). From text to task: Putting research on abstracts to work. In M. Ruiz-Garrido, J. Palmer-Silveira, & I. Fortanet-Gomez (Eds.), *English for professional and academic purposes* (pp. 167-180). Amsterdam-New York, NY: Rodopi.
- Tankó, G. (2012). *Professional writing: The academic context*. Budapest: Eötvös University Press.
- Tanskanen, S. K. (2006). *Collaborating towards coherence*. Amsterdam: John Benjamins Publishing Company.
- Tashakkori, A., & Teddlie, C. (2003). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 671-701). Thousand Oaks, CA: Sage.

- Teich, E., & Fankhauser, P. (2004). WordNet for lexical cohesion analysis, (Section 3), 326–331.
- Teich, E., & Fankhauser, P. (2005). Exploring lexical patterns in text : Lexical cohesion analysis with WordNet. *Heterogeneity in Focus: Creating and Using Linguistic Databases - Interdisciplinary Studies on Information Structure 02, 02*, 129–145.
- Tierney, R. J., & Shanahan, T. (1991). Research on the reading-writing relationship: Interactions, transactions, and outcomes. In Barr R, M. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 246-280). New York, Longman.
- Tyler, A. (1992). Discourse structure and specification of relationships. A crosslinguistic analysis. *Text, 12*, 1-18.
- Tyler, A. (1994). The role of repetition in perceptions of discourse coherence. *Journal of Pragmatics, 21*(6).
- Tyler, A. (1995). Co-constructing miscommunication: The role of participant frame and schema in cross-cultural miscommunication. *Studies in Second Language Acquisition, 17*, 129-152.
- Utt, J., & Padó, S. (2011). Ontology-based distinction between polysemy and homonymy. *IWCS '11 Proceedings of the Ninth International Conference on Computational Semantics*, (pp. 265-274). Retrieved from www.aclweb.org/anthology/W11-0128
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Walk, K. (1998). How to write a comparative analysis. Retrieved from <http://writingcenter.fas.harvard.edu/pages/how-write-comparative-analysis>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9(9), 27-55.
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Wengelin, A., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., & Johansson, R. (2009). Combined eye-tracking and keystroke-logging methods for studying cognitive processes in text production. *Behavior Research Methods*, 41(2), 337-351.
- White, E. M. (1984). Holisticism. *College Composition and Communication*, 35, 400-409.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.
- Wilkinson, A. M. (1991). *The scientist's handbook for writing papers and dissertations*. Edgewood Cliffs, NJ: Prentice Hall.
- Winograd, P. (1984). Strategic difficulties in summarizing texts. *Reading Research Quarterly*, 19, 404-425.
- Winter, E. (1977). A clause relational approach to English texts: a study of some predictive lexical items in written discourse. *Instructional Science*, 6(1), 1-92.
- Winter, E. (1979). Replacement as a fundamental function of the sentence in context. *Forum Linguisticum*, 4(2), 95-133.
- Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? *Language Testing*, 25(4), 521-551.
- Yu, G. (2009). The shifting sands in the effects of source text summarizability on summary writing. *Assessing writing*, 14(2), 116-137.

Zemach, D., & Rumisek, L. (2003a). *College writing. From paragraph to essay*. Oxfo:
Macmillan Education.

Zemach, D., & Rumisek, L. (2003b). *College writing. From paragraph to essay. Teacher's book*.
Oxford: Macmillan Education.

Computer software

Anthony, L. (2014). TagAnt (Version 1.1.2) [Computer Software]. Tokyo, Japan: Waseda
University. Available from <http://www.laurenceanthony.net/>

Watt, R. J. C. (1999-2009). Concordance (Version 3.3 July, 2003) [Computer software].
Retrieved in November, 2011, from <http://www.concordancesoftware.co.uk/concordance-software-download.htm>

Miller, G. (1995). WordNet: A Lexical Database for English. [Computer software]. Retrieved in
November, 2011 from <http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=8106>

APPENDICES

Appendix A

Hoey's (1991) original passage with three links provided and the matrix for repetition links

1. Coding the text according to the taxonomy. Finding links between every sentence pair, including the title, which also counts as one sentence. (1—2, 1—3, 1—n, etc., and in the same way 2—3, 2—4, 2—n). In the example below, sentences 1—2, 1—4, 2—4 are linked by various forms of the word ‘drug’.)

1 A **drug** known to produce violent reactions in humans has been used for sedating grizzly bears *Ursus arctos* in Montana, USA, according to a report in The New York Times. 2 After one bear, known to be a peaceable animal, killed and ate a camper in an unprovoked attack, scientists discovered it had been **tranquillized** 11 times with phencyclidine, or ‘angel dust’, which causes hallucinations and sometimes gives the user an irrational feeling of destructive power. 4 To avoid potentially dangerous clashes between them and humans, scientists are trying to rehabilitate the animals by **drugging** them and releasing them in uninhabited areas.

2. Writing the links into a connectivity matrix where each cell represents a sentence, as Hoey put it: “to trace a sentence’s connections with other sentences in the text” (1991, p. 85). All links should be written into the cells, see Section 5.2.6.2, p.127.
3. The information in the matrix should be written into another matrix in a number format, see Section 5.2.6.2, p.128.
4. Cells containing three, or more than three links should be highlighted because these are the bonded sentences. In the following only these sentences will be examined.
5. The locations of bonded sentences need to be found in the text, and they should be highlighted.
6. If the purpose of the analysis is to create a summary, either the bonded sentences should be collected, or the marginal sentences should be deleted (same procedure). The third procedure is to collect all the topic opening and topic closing sentences. The bonded sentences will give the basis of the summary.

Appendix B

The most frequent metadiscoursal nouns (preceded by *this*) in the Hyland (1998)

corpus (based on Swales, 2005, p.10)

Discipline	Most frequent metadiscoursal nouns				
Dentistry	study	finding	result	patient	process
Medicine	study	group	difference	procedure	technique
Biology	result	observation	study	difference	finding
Electrical Engineering	approach	algorithm	method	paper	technique
Mechanical Engineering	paper	method	approach	type	figure
Applied Linguistics	study	result	experiment	difference	finding
Marketing	study	paper	cluster	approach	research
Philosophy	account	article	argument	conclusion	claim
Sociology	article	model	paper	process	group
Physics	effect	approach	behavior	contribution	figure

Appendix C

Quantitative measures in Károly's (2002) the LRA analysis (based on Károly, 2002, p. 144), also used in Stages 1 and 2 in this study

1. Basic measures:	number of sentences number of paragraphs number of links number of cells
2. Measures related to repetition type:	frequency of simple repetition derived repetition same unit repetition simple synonymy derived synonymy synonymy (simple and derived) simple opposites derived opposites hyponymy meronymy instantial relations different unit repetition ratio of same unit repetition to different unit repetition ratio of simple repetition to derived repetition
3. Measures related to the combination of links and bonds:	frequency of links frequency of bonds density of bonds frequency of adjacent bonds frequency of non-adjacent bonds cumulative bond span frequency of central sentences frequency of marginal sentences relative use of bonds at paragraph boundary strength of connection (1-8 links) bonds between: title & essay title & thesis statement title & topic sentences thesis statement & topic sentences thesis statement & essay

Appendix D
Tasks in academic writing across three levels (in order of perceived importance, with mentions of coherence in bold, summarize and compare/contrast in italics)

Upper-Level Undergraduate Tasks	Master's-Level Tasks	Doctoral-Level Tasks
Credit sources appropriately (e.g., use attribution, footnotes, or endnotes)		
Organize ideas and information coherently		
Use grammar and syntax that follow the rules of standard written English, avoiding errors that distract the reader or disrupt meaning		
Avoid errors in mechanics (e.g., spelling and punctuation)		Abstract or <i>summarize essential information</i> (e.g., from speeches, observations, or texts)
Revise and edit text to improve its clarity, coherence , and correctness	Abstract or <i>summarize essential information</i> (e.g., from speeches, observations, or texts)	Analyze and synthesize information from multiple sources (includes <i>comparison and contrast</i>)
Write precisely and concisely, avoiding vague or empty phrases	Analyze and synthesize information from multiple sources (includes <i>comparison and contrast</i>)	Examine the reasoning in a given argument and discuss its logical strengths and weaknesses (e.g., the legitimacy of claims, the soundness of assumptions, the sufficiency of support, or the distinction between correlation and causation)
	Integrate quoted and referenced material appropriately into the students' own text	Interpret data within a relevant framework by applying the findings to new situations, asking insightful questions, identifying the need for further information, or drawing conclusions
	Develop a well-focused, well-supported discussion, using relevant reasons and examples	Present data and other information in a clear and logical manner, offering explanations that make the material understandable to a particular audience (includes tables and charts as well as text)
	Write clearly, with smooth transitions from one thought to the next	Develop a well-focused, well-supported discussion, using relevant reasons and examples
	Write precisely and concisely, avoiding vague and empty phrases	Choose words effectively
	Revise and edit text to improve its clarity, coherence , and correctedness [sic].	Write fluently, avoiding plodding or convoluted language
	Work independently to plan and compose text	Avoid errors in mechanics (e.g., spelling and punctuation)
		Work independently to plan and compose text

Appendix E

TOEFL holistic scoring guide (for placement purposes)

<http://www.etstechnologies.com/html/TOEFLscoringguide.htm>

6-- Demonstrates clear competence in writing on both the rhetorical and syntactic levels, though it may have occasional errors.

A paper in this category:

- * effectively addresses the writing task
- * is well organized and well developed
- * uses clearly appropriate details to support a thesis or illustrate ideas
- * displays consistent facility in the use of language
- * demonstrates syntactic variety and appropriate word choice

5-- Demonstrates competence in writing on both the rhetorical and syntactic levels, though it will probably have occasional errors.

A paper in this category:

- * may address some parts of the task more effectively than others
- * is generally well organized and developed
- * uses details to support a thesis or illustrate an idea
- * displays facility in the use of language
- * demonstrates some syntactic variety and range of vocabulary

4-- Demonstrates minimal competence in writing on both the rhetorical and syntactic levels.

A paper in this category may reveal one or more of the following weaknesses:

- * addresses the writing topic adequately but may slight parts of the task
- * is adequately organized and developed
- * uses some details to support a thesis or illustrate an idea
- * demonstrates adequate but possibly inconsistent facility with syntax and usage
- * may contain some errors that occasionally obscure meaning

3-- Demonstrates some developing competence in writing, but it remains flawed on either the rhetorical or syntactic level, or both.

A paper in this category may reveal one or more of the following weaknesses:

- * inadequate organization or development
- * inappropriate or insufficient details to support or illustrate generalizations
- * a noticeably inappropriate choice of words or word forms
- * an accumulation of errors in sentence structure and/or usage

2-- Suggests incompetence in writing.

A paper in this category is seriously flawed by one or more of the following weaknesses:

- * serious disorganization or underdevelopment
- * little or no detail, or irrelevant specifics
- * serious and frequent errors in sentence structure or usage
- * serious problems with focus

1-- Demonstrates incompetence in writing.

A paper in this category:

- * may be incoherent
- * may be undeveloped
- * may contain severe and persistent writing errors

Appendix F

A sample persuasive essay writing and a paragraph writing analytic rubric⁴⁸

	Criteria				Pts.
	4	3	2	1	
Position Statement	Position is clearly stated and consistently maintained. Clear references to the issue(s) are stated.	Position is clearly stated and consistently maintained. References to the issue(s) at hand are missing .	Position is stated, but is not maintained consistently throughout work.	Statement of position cannot be determined.	—
Supporting Information	Evidence clearly supports the position; evidence is sufficient .	Evidence clearly supports the position; but there is not enough evidence .	Argument is supported by limited evidence .	Evidence is unrelated to argument.	—
Organization	Structure of work is clearly developed.	Structure developed reasonably well, but lacks clarity .	Some attempt to structure the argument has been made, but the structure is poorly developed .	There is a total lack of structure.	—
Tone Of Letter	Tone is consistent and enhances persuasiveness.	Tone enhances persuasiveness, but there are inconsistencies .	Tones does not contribute to persuasiveness.	Tone is inappropriate to purpose.	—
Sentence Structure	Sentence structure is correct .	Sentence structure is generally correct. Some awkward sentences do appear.	Work contains structural weaknesses and grammatical errors .	Work pays little attention to proper sentence structure.	—
Punctuation & Capitalization	Punctuation and capitalization are correct .	There is one error in punctuation and/or capitalization.	There are two or three errors in punctuation and/or capitalization.	There are four or more errors in punctuation and/or capitalization.	—
				Total ---->	—

⁴⁸ retrieved from http://www.teach-nology.com/web_tools/rubrics/

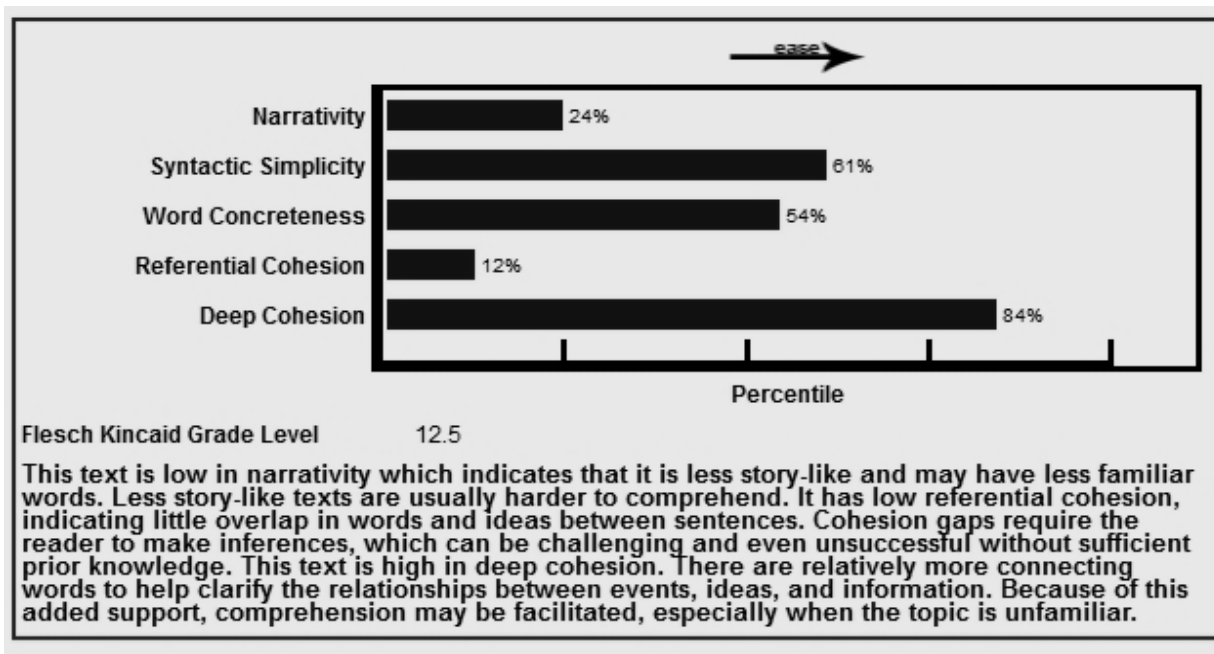
	Criteria				Points
	4	3	2	1	
Main/Topic Idea Sentence	Main/Topic idea sentence is clear, correctly placed, and is restated in the closing sentence.	Main/Topic idea sentence is either unclear or incorrectly placed, and is restated in the closing sentence.	Main/Topic idea sentence is unclear and incorrectly placed, and is restated in the closing sentence.	Main/Topic idea sentence is unclear and incorrectly placed, and is not restated in the closing sentence.	—
Supporting Detail Sentence(s)	Paragraph(s) have three or more supporting detail sentences that relate back to the main idea.	Paragraph(s) have two supporting detail sentences that relate back to the main idea.	Paragraph(s) have one supporting detail sentence that relate back to the main idea.	Paragraph(s) have no supporting detail sentences that relate back to the main idea.	—
Elaborating Detail Sentence(s)	Each supporting detail sentence has three or more elaborating detail sentences.	Each supporting detail sentence has at least two elaborating detail sentences.	Each supporting detail sentence has one elaborating detail sentence.	Each supporting detail sentence has no elaborating detail sentence.	—
Legibility	Legible handwriting, typing, or printing.	Marginally legible handwriting, typing, or printing.	Writing is not legible in places.	Writing is not legible.	—
Mechanics and Grammar	Paragraph has no errors in punctuation, capitalization, and spelling.	Paragraph has one or two punctuation, capitalization, and spelling errors.	Paragraph has three to five punctuation, capitalization, and spelling errors.	Paragraph has six or more punctuation, capitalization, and spelling errors.	—
				Total ---->	—

Appendix G

Percentile scores of the input text features in Stage 1, summary writing task by Coh-

Metrix, all explanations from the Coh-Metrix manual

(McNamara, Louwerse, Cai, & Graesser, 2005)



Referential cohesion refers to overlap in content words between local sentences, or *co-reference*. Coh-Metrix measures for referential cohesion vary along two dimensions. First, the indices vary from local to more global. Local cohesion is measured by assessing the overlap between consecutive, adjacent sentences, whereas global cohesion is assessed by measuring the overlap between all of the sentences in a paragraph or text. Latent Semantic Analysis (LSA; Landauer et al., 2007) provides measures of semantic overlap between sentences or between paragraphs. Coh-Metrix 3.0 provides eight LSA indices. Each of these measures varies from 0 (low cohesion) to 1 (high cohesion).

Example:

Text 1: *The field was full of lush, green grass. The horses grazed peacefully. The young children played with kites. The women occasionally looked up, but only occasionally. A warm summer breeze blew and everyone, for once, was almost happy.*

Text 2: *The field was full of lush, green grass. An elephant is a large animal. No-one appreciates being lied to. What are we going to have for dinner tonight?*

In the example texts printed above, Text 1 records much higher LSA scores than Text 2. The words in Text 1 tend to be thematically related to a pleasant day in an idyllic park scene: *green, grass, children, playing, summer, breeze, kites, and happy*. In contrast, the sentences in Text 2 tend to be unrelated.

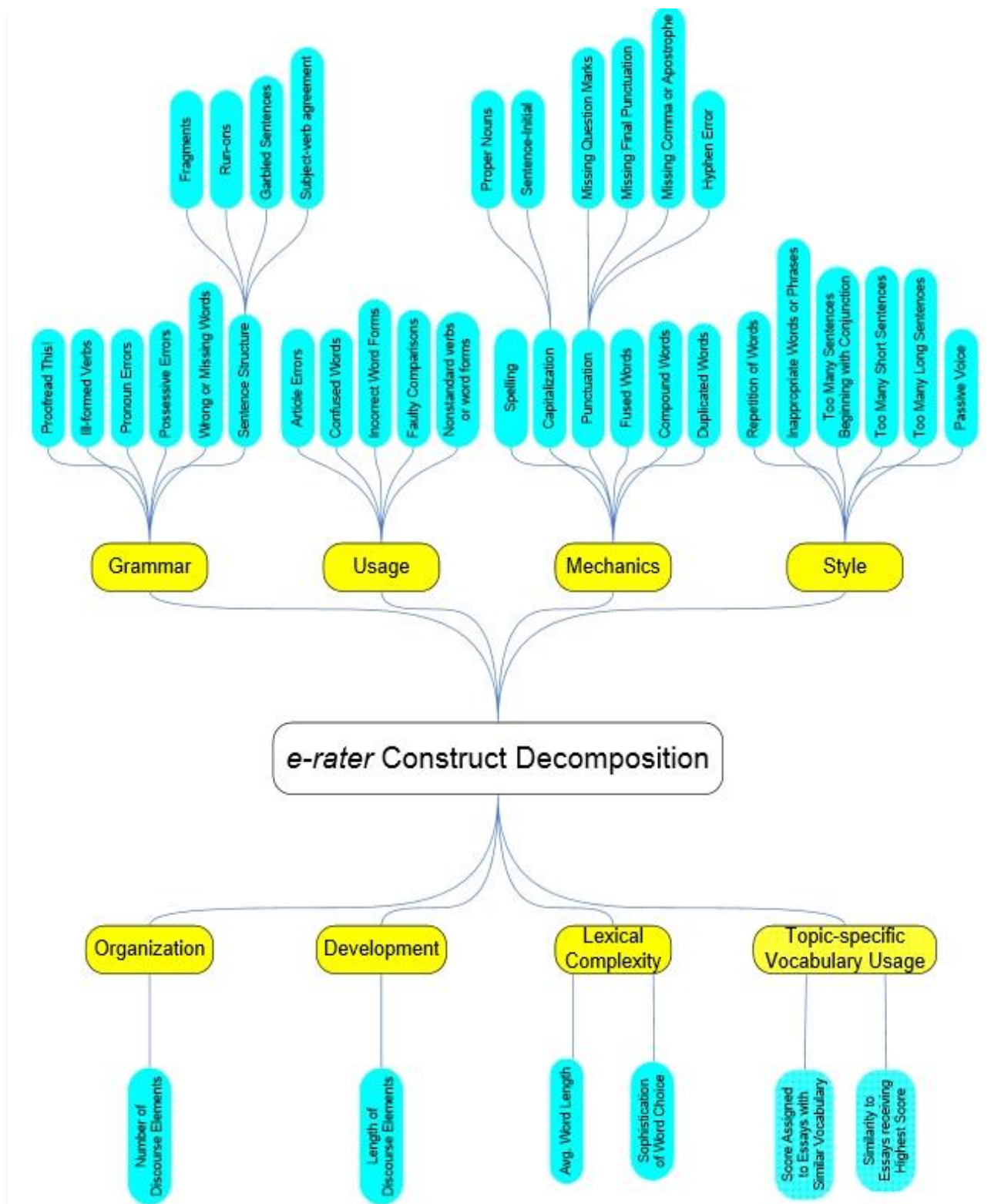
Appendix H.

The descriptive framework of raters' decision-making behaviors (Cumming, Kantor, & Powers, 2001).

Self-monitoring focus	Rhetorical and ideational focus	Language focus
Interpretation strategies		
<ul style="list-style-type: none"> • read or interpret prompt and/or task input • read or reread composition • envision personal situation of writer 	<ul style="list-style-type: none"> • discern rhetorical structure • summarize ideas or propositions • scan whole composition or observe layout 	<ul style="list-style-type: none"> • classify errors into types • interpret or edit ambiguous or unclear phrases
Judgement strategies		
<ul style="list-style-type: none"> • decide on macro-strategy for reading and rating; compare with other compositions; or summarize, distinguish, or tally judgements collectively • consider own personal response or biases • define or revise own criteria • articulate general impression • articulate or revise scoring decision 	<ul style="list-style-type: none"> • assess reasoning, logic, or topic development • assess task completion or relevance • assess coherence and identify redundancies • assess interest, originality, or creativity • assess text organization, style, register, discourse functions, or genre • consider use and understanding of source material • rate ideas or rhetoric 	<ul style="list-style-type: none"> • assess quantity of total written production • assess comprehensibility and fluency • consider frequency and gravity of errors • consider lexis • consider syntax or morphology • consider spelling or punctuation • rate language overall

Appendix I.

E-rater (automated essay assessor) features and microfeatures (Quinlan et al., 2009).



**Professionals Against Machine Scoring Of Student Essays In High-Stakes Assessment:
Petition of teachers against machine scoring of essays⁴⁹**

RESEARCH FINDINGS SHOW THAT

no one—students, parents, teachers, employers, administrators, legislators—can rely on machine scoring of essays:

1. computer algorithms cannot recognize the most important qualities of good writing, such as truthfulness, tone, complex organization, logical thinking, or ideas new and germane to the topic (Byrne, Tang, Truduc, & Tang, 2010)
2. to measure important writing skills, machines use algorithms that are so reductive as to be absurd: *sophistication of vocabulary* is reduced to the average length or relative infrequency of words, or *development of ideas* is reduced to average sentences per paragraph (Perelman, 2012b; Quinlan, Higgins, & Wolff, 2009)
3. machines over-emphasize grammatical and stylistic errors (Cheville, 2004) yet miss or misidentify such errors at intolerable rates (Herrington & Moran, 2012)
4. machines cannot score writing tasks long and complex enough to represent levels of writing proficiency or performance acceptable in school, college, or the workplace (Bennett, 2006; Condon, 2013; McCurry, 2010; Perelman, 2012a)
5. machines require artificial essays finished within very short time frames (20-45 minutes) on topics of which student writers have no prior knowledge (Bridgeman, Trapani, & Yigal, 2012; Cindy, 2007; Jones, 2006; Perelman, 2012b; Streeter, Psotka, Laham, & MacCuish, 2002; Wang, & Brown, 2008; Wohlpart, Lindsey, & Rademacher, 2008)
6. in these short trivial essays, mere length becomes a major determinant of score by both human and machine graders (Chodorow & Burstein, 2004; Perelman, 2012b)
7. machines are not able to approximate human scores for essays that do fit real-world writing conditions; instead, machines fail badly in rating essays written in these situations (Bridgeman, Trapani, & Yigal, 2012; Cindy, 2007; Condon, 2013; Elliot, Deess, Rudniy, & Joshi, 2012; Jones, 2006; Perelman, 2012b; Powers, Burstein, Chodorow, Fowles, & Kukich, 2002; Streeter, Psotka, Laham, & MacCuish, 2002; Wang & Brown, 2008; Wohlpart, Lindsey, & Rademacher, 2008)
8. high correlations between human scores and machine scores reported by testing firms are achieved, in part, when the testing firms train the humans to read like the machine, for instance, by directing the humans to disregard the truth or accuracy of assertions (Perelman, 2012b), and by requiring both machines and humans to use scoring scales of extreme simplicity
9. machine scoring shows a bias against second-language writers (Chen & Cheng, 2008) and minority writers such as Hispanics and African Americans (Elliot, Deess, Rudniy, & Joshi., 2012]
10. for all these reasons, machine scores predict future academic success abysmally (Mattern & Packman, 2009; Matzen & Hoyt, 2004; Ramineni & Williamson, 2013)

⁴⁹ http://humanreaders.org/petition/research_findings.htm

AND THAT machine scoring does not measure, and therefore does not promote, authentic acts of writing:

1. students are subjected to a high-stakes response to their writing by a device that, in fact, cannot read, as even testing firms admit (Elliott, 2011)
2. in machine-scored testing, often students falsely assume that their writing samples will be read by humans with a human's insightful understanding (Herrington & Moran, 2006)
3. conversely, students who knowingly write for a machine are placed in a bind since they cannot know what qualities of writing the machine will react to positively or negatively, the specific algorithms being closely guarded secrets of the testing firms (Frank, 1992; Rubin & O'Looney, 1990)—a bind made worse when their essay will be rated by both a human and a machine
4. students who know that they are writing only for a machine may be tempted to turn their writing into a game, trying to fool the machine into producing a higher score, which is easily done (McGee, 2006; Powers, Burstein, Chodorow, Fowles, & Kukich, 2001; see item 6, above) by rote-learned phrases placed into their writing, this is difficult to catch even by human raters, on the other hand: is this really something teachers should look at negatively? in fact this is actively encouraged in the L2 context also by textbooks providing „samples of language”
5. teachers are coerced into teaching the writing traits that they know the machine will count—surface traits such as essay length, sentence length, trivial grammatical mistakes, mechanics, and topic-related vocabulary—and into *not* teaching the major traits of successful writing—elements such as accuracy, reasoning, organization, critical and creative thinking, and engagement with current knowledge (Council, 2012; Deane, 2013; Herrington & Moran, 2001; National, 2010)
6. machines also cannot measure authentic audience awareness, a skill essential at all stages of the composing process and correlative with writing competence of students both in the schools (Wolmann-Bonilla, 2000) and in college (Rafoth, 1985)
7. as a result, the machine grading of high-stakes writing assessments seriously degrades instruction in writing (Perelman, 2012a), since teachers have strong incentives to train students in the writing of long verbose prose, the memorization of lists of lengthy and rarely used words, the fabrication rather than the researching of supporting information, in short, to dumb down student writing.

Appendix K ACADEMIC WRITING Scoring Rubric (weighted)⁵⁰

adapted by Christine Bauer-Ramazani, 2006, from MELAB (L. Hamp-Lyons, 1992)

Task Completion/ Format/Layout (x1)	
5	The assignment includes all required components of the task and follows formatting guidelines for layout (headings), spacing, alignment, indentations, etc. Format and layout make the assignment exceptionally attractive .
4	Almost all required components are included; formatting guidelines for layout (headings), spacing, alignment, indentations, etc. are almost always followed. 1-2 problems in format and layout, but readability and attractiveness are not affected.
3	Some required components are included; formatting guidelines for layout (headings), spacing, and alignment are sometimes followed. 3-4 problems in format and layout , but the assignment is easy to read.
2	Required components are mostly absent; formatting is repeatedly inconsistent in layout (headings), spacing, and alignment, reducing readability and attractiveness.
1	Off-topic; formatting guidelines for layout (headings), spacing, and alignment are not followed, making the assignment unattractive or hard to read .
Score	
Topic Development (x3)	
5	full and rich development of content (<i>purpose, relevance, focus, explanations, support</i>); shows knowledge of subject matter through effective use of topic-related concepts to support argument; sophistication in fluency of expression
4	clear and complete development of content; high level of fluency in expression (clarity); evidence of logical progression of ideas; mostly effective use of topic-related concepts to show knowledge of subject matter to support argument
3	development of content adequate , but lacks clearly stated positions/argument or supporting information; some use of topic-related concepts to show knowledge of subject matter and to support argument; fluency of expression may be halting or awkward
2	development of content restricted ; may be incomplete or unclear; little evidence of argument; little evidence of knowledge of subject matter (use of topic-related concepts) to support argument; lack of fluency in expression
1	simplicistic statement of content; often copied from sources or lists of information; no argument; no use of topic-related concepts to support argument
Score	
Organization (x2)	
5	organization fully appropriate and effective for topic (<i>point of view, unity, paragraphing</i>); very strong thesis/purpose statement & conclusion; paragraphs have effective topic sentences
4	organization controlled ; generally appropriate to topic; paragraphing, sequencing, purpose statement/thesis & conclusion, topic sentences evident and appropriate
3	organization controlled but limited ; some problems with paragraphing or sequencing, limited purpose statement/thesis or conclusion; limited use of topic sentences
2	little or poorly controlled organization; incorrect paragraphing or sequencing; purpose statement/thesis, conclusion, topic sentences may be missing or incomplete
1	minimal attempt at paragraphing, often unsuccessful; strings of sentences; no purpose statement/thesis, conclusion, or topic sentences

⁵⁰ http://academics.smcvt.edu/cbauer-ramazani/AEP/EN104/Rubrics/acad_writ_rubric-weighted.htm

Score	
Vocabulary (x2)	
5	broad and fluent range of vocabulary; elaboration and detail achieved through full use of <i>topic-related terms</i> ; concise and appropriate use of vocabulary (formal English); correct use of <i>word forms and word choice</i> .
4	flexibility in range; appropriate use of <i>topic-related terms</i> and other vocabulary in a variety of situations; mostly correct use of <i>word forms</i> and <i>word choice</i> ; occasional wordiness or colloquialisms (informal English)
3	adequate range; no precise use of subtle meanings displayed; <i>topic-related terms</i> only used occasionally; vocabulary sometimes used inappropriately; 3-4 instances of wordiness or colloquialisms (informal English); often incorrect use of <i>word forms</i> .
2	narrow range; many word form errors; <i>topic-related terms</i> and other vocabulary often used inappropriately; only basic and elementary meanings are conveyed; 5 or more instances of wordiness or colloquialisms (informal English)
1	simple vocabulary, often inappropriately used; no control of <i>word choice and word forms</i> ; no attempt to use <i>topic-related terms</i> ; many instances of wordiness or colloquialisms (informal English)
Score	
Discourse Control (x1)	
5	full control (<i>logical coherence</i>) and excellent use of <i>cohesive devices</i> (key words, pronouns, references, transitions, etc.); presentation of ideas extremely clear and coherent
4	good control of <i>cohesive devices</i> (key words, pronouns, references, transitions, etc.) used successfully in a range of situations; coherence apparent
3	generally adequately connected; presentation of ideas generally clear and coherent ; <i>cohesive devices</i> (key words, pronouns, references, transitions, etc.) could be used more often, more effectively, or more accurately
2	connections awkward ; <i>cohesive devices</i> (key words, pronouns, references, transitions, etc.) may be missing or are used inaccurately; lack of logical sequencing of ideas
1	connections (<i>cohesive devices</i>) not present or unsuccessful ; presentation of ideas unclear and confusing
Score	
Sentence structure (x2)	
5	full range of sentence patterns (simple, compound, complex), effectively used; error-free sentence-level grammar
4	mastery of sentence patterns demonstrated; may have occasional grammatical errors on the sentence level (e.g. possessives, word deletions; 1-2 run-on sentences or fragments)
3	sentence patterns most often successfully used; several grammatical errors on the sentence level (e.g. word deletions, possessives, prep., tense); 3-4 fragments, or run-on sentences)
2	many sentence structure problems; 5-7 fragments or run-on sentences; grammatical errors distract from meaning
1	replete with errors in sentence structure and sentence grammar; more than 7 fragments or run-on sentences
Score	

Mechanics (x1)

5	correct use of in-text citations & references; spelling, capitalization, and punctuation error-free
4	spelling, capitalization, punctuation, and citation/reference errors few and not distracting
3	occasional errors in spelling, capitalization, punctuation, and citation/references; sometimes distracting
2	spelling, capitalization, punctuation, and citation errors are frequent and distracting
1	spelling, capitalization, punctuation, and citation/reference errors throughout
Score	

NAME: _____

DRAFT #: _____

SCORE: _____

ASSIGNMENT: _____

**Calculation of score (in %): Total
points / 60**

GRADE: _____

Appendix L

A sample summary and its Coh-Metrix analysis for cohesion

Ryanair is an Irish low-cost airline which was established in 1985. The founders were Christopher Ryan, Liam Lonergan (who is the owner of Irish Travel Agent Club) and Tony Ryan (Irish businessman, after whom the company was named). The company operates more than 250 Boeing 737-800 aircraft. Ryanair organises about 1,100 routes across Europe and even in Morocco from 44 bases.

It started as a small airline but the company is expanding rapidly due to the deregulation of the aviation industry in Europe and the low-cost business model. At the beginning the airline ran at a loss, so restructuring was necessary. Michael O’Leary was appointed to make the airline profitable. He had several good and controversial decisions. The strength of the company are due to the following. He realised that the key to low fares was to implement quick turn-around times for aircraft, that the planes should have no business class and frills, and that the company should operate only one model of aircraft. Another money-saving idea was that the flights were scheduled into regional airports, which offered lower landing and handling charges than international airports. As a result of these innovations, by 1995 Ryanair celebrated its 10th birthday by carrying 2.25 million passengers.

However, not everyone is happy with Ryanair’s relationship with customers. Ryanair's training policies, security procedures and aircraft hygiene were also criticised. The passengers complained about the ticketing policies, and strict measures, such as extra fees for obese customers. They also wanted to introduce standing “seats” to carry more passengers at a time! Ryanair is also said to blackmail airports to lower their taxes or else they withdraw their flights from these destinations. This happened to Ferihegy Airport recently.

All in all, we can conclude that Ryanair is one of the world’s best airlines and due to its success by 2011 air 2010 traffic grew by 10% from 65 million to over 72 million passengers.

Referential Cohesion			
28	CRFNO1	CRFBNlum	0.474 Noun overlap, adjacent sentences, binary, mean
29	CRFAO1	CRFBAlum	0.579 Argument overlap, adjacent sentences, binary, mean
30	CRFSO1	CRFBSlum	0.579 Stem overlap, adjacent sentences, binary, mean
31	CRFNOa	CRFBNaum	0.283 Noun overlap, all sentences, binary, mean
32	CRFAOa	CRFBAaum	0.317 Argument overlap, all sentences, binary, mean
33	CRFSOa	CRFBSaum	0.31 Stem overlap, all sentences, binary, mean
34	CRFCWO1	CRFPClum	0.07 Content word overlap, adjacent sentences, proportional, mean
35	CRFCWO1d	n/a	0.062 Content word overlap, adjacent sentences, proportional, standard deviation
36	CRFCWOa	CRFPCaum	0.042 Content word overlap, all sentences, proportional, mean
37	CRFCWOad	n/a	0.069 Content word overlap, all sentences, proportional, standard deviation
LSA			
38	LSASS1	LSAassa	0.157 LSA overlap, adjacent sentences, mean
39	LSASS1d	LSAassd	0.126 LSA overlap, adjacent sentences, standard deviation
40	LSASSp	LSAapssa	0.161 LSA overlap, all sentences in paragraph, mean
41	LSASSpd	LSAapssd	0.123 LSA overlap, all sentences in paragraph, standard deviation
42	LSAPP1	LSAppa	0.452 LSA overlap, adjacent paragraphs, mean
43	LSAPP1d	LSAppd	0.159 LSA overlap, adjacent paragraphs, standard deviation
44	LSAGN	LSAGN	0.275 LSA given/new, sentences, mean
45	LSAGNd	n/a	0.113 LSA given/new, sentences, standard deviation
Lexical Diversity			
46	LDTTRc	TYPTOKc	0.776 Lexical diversity, type-token ratio, content word lemmas
47	LDTTRa	n/a	0.599 Lexical diversity, type-token ratio, all words
48	LDMTLD	LEXDIVTD	103.518 Lexical diversity, MTLTLD, all words
49	LDVOCD	LEXDIVVD	119.956 Lexical diversity, VOCD, all words
Connectives			
50	CNCAll	CONi	99.071 All connectives incidence
51	CNCCaus	CONCAUSi	40.248 Causal connectives incidence
52	CNCLogic	CONLOGi	43.344 Logical connectives incidence
53	CNCADC	CONADVCONi	12.384 Adversative and contrastive connectives incidence
54	CNCTemp	CONTEMPi	9.288 Temporal connectives incidence
55	CNCTempx	CONTEMPEXi	6.192 Expanded temporal connectives incidence
56	CNCAdd	CONADDi	52.632 Additive connectives incidence
57	CNCPos	n/a	0 Positive connectives incidence
58	CNCNeg	n/a	0 Negative connectives incidence

¹Declaration form
for disclosure of a doctoral thesis

I. The data of the doctoral thesis:

Name of the author: Adorján Mária

MTMT-identifier: 10030659

Title and subtitle of the doctoral thesis: Lexical Repetition in Academic Discourse: A Computer-aided Study of the Text Organizing Role of Repetition

DOI-identifier²: 10.15476/ELTE.2015.180

Name of the doctoral school: Doctoral School of Education

Name of the doctoral programme: Programme in Language Pedagogy

Name and scientific degree of the supervisor: Károly Krisztina, PhD, habil

Workplace of the supervisor: Eötvös Loránd University, Centre for Teacher Training

II. Declarations

1. As the author of the doctoral thesis,³

a) I agree to public disclosure of my doctoral thesis after obtaining a doctoral degree in the storage of ELTE Digital Institutional Repository. I authorize Madar Veronika, the administrator of the Student Affairs and Registrar's Department to upload the thesis and the abstract to ELTE Digital Institutional Repository, and I authorize the administrator to fill all the declarations that are required in this procedure.

b) I request to defer public disclosure to the University Library and the ELTE Digital Institutional Repository until the date of announcement of the patent or protection. For details, see the attached application form;⁴

c) I request in case the doctoral thesis contains qualified data pertaining to national security, to disclose the doctoral thesis publicly to the University Library and the ELTE Digital Institutional Repository ensuing the lapse of the period of the qualification process;⁵

d) I request to defer public disclosure to the University Library and the ELTE Digital Institutional Repository, in case there is a publishing contract concluded during the doctoral procedure or up until the award of the degree. However, the bibliographical data of the work shall be accessible to the public. If the publication of the doctoral thesis will not be carried out within a year from the award of the degree subject to the publishing contract, I agree to the public disclosure of the doctoral thesis and abstract to the University Library and the ELTE Digital Institutional Repository.⁶

2. As the author of the doctoral thesis, I declare that

a) the doctoral thesis and abstract uploaded to the ELTE Digital Institutional Repository are entirely the result of my own intellectual work and as far as I know, I did not infringe anyone's intellectual property rights.;

b) the printed version of the doctoral thesis and the abstract are identical with the doctoral thesis files (texts and diagrams) submitted on electronic device.

3. As the author of the doctoral thesis, I agree to the inspection of the thesis and the abstract by uploading them to a plagiarism checker software.

Budapest, 2015. November 1.

.....

Signature of thesis author

¹ Endorsed by Senate Rule CXXXIX/2014. (VI. 30.) on the amendment of the Doctoral Regulations of ELTE. Effective date: 01 July 2014.

² Filled by the administrator of the faculty offices.

³ The relevant part shall be underlined.

⁴ Submitting the doctoral thesis to the Disciplinary Doctoral Council, the patent or protection application form and the request for deferment of public disclosure shall also be attached.

⁵ Submitting the doctoral thesis, the notarial deed pertaining to the qualified data shall also be attached.

⁶ Submitting the doctoral thesis, the publishing contract shall also be attached.