

UNIVERSITY OF LATVIA
FACULTY OF HUMANITIES
DEPARTMENT OF ENGLISH STUDIES

**SYSTEMIC FUNCTIONAL ANALYSIS OF THE YEAR
9 ENGLISH LANGUAGE EXAMINATION CORPUS**

**DEVĪTĀS KLASES ANĢĻU VALODAS EKSĀMENA
KORPUSA SISTĒMISKI FUNKCIONĀLĀ ANALĪZE**

MASTER THESIS

Author: **Līga Ločmele**

Matriculation Card No. 1109294

Adviser: assoc. prof. V. Kalnbērziņa

RIGA 2015

ACKNOWLEDGEMENTS

I express my deepest gratitude to my advisor, Dr. Phil. V. Kalnbērziņa, whose passion for foreign language testing and expert guidance inspired me throughout the process of writing this thesis.

I would also like to thank the schools participating in the research, and the National Centre for Education for the assistance with collecting the writing scripts necessary for the study.

I am also immensely grateful to my family and friends, including my fellow students, for their love, support, and invaluable comments, which helped me to improve my work.

ANOTĀCIJA

Šis pētījums izstrādāts, sakarā ar kohēzijas zemāka līmeņa rakstu darbos, kā arī tekstu korpusu izpētes trūkumu Latvijā, izveidojot un analizējot rakstveida 9. klases angļu valodas eksāmena tekstu korpusu. Mērķis izpētīt gramatiskās kohēzijas lietojumu skolēnu rakstu darbos angļu valodā tika sasniegts, apvienojot uz tekstu korpusu balstītu statistisko analīzi un sistēmiski funkcionālu diskursa analīzi, izmantojot programmu UAM CorpusTool. Rezultāti liecina, ka skolēni ar zemāku angļu valodas līmeni spēj uzrakstīt saistītu tekstu, ļaujoties uz vietniekvārdiem, saikļiem *un, vai, bet* un attieksmes vietniekvārdiem. Nākamajā līmenī saikļu daudzveidība pieaug, vietniekvārdu blīvums tekstā samazinās, un pareizrakstības kļūdas izzūd. Pētījums apstiprināja, ka veiksmīgs gramatiskās kohēzijas lietojums ir atkarīgs no rakstīšanas prasmes, taču teksta resursu dažādība pakāpeniski pieaug attiecībā pret angļu valodas līmeni.

Atslēgvārdi: gramatiskā kohēzija, sistēmiski-funkcionālais skatījums, eksāmenu tekstu korpus, UAM CorpusTool, statistika, diskursa analīze.

ABSTRACT

The present thesis addresses the gap of research on cohesion in lower level writing, as well as corpus studies in Latvia, by combining and analysing a written Year 9 English language examination corpus. The aim of exploring the use of grammatical cohesion in learners' writing was achieved with a combination of corpus-based statistics and systemic functional discourse analysis, using the UAM CorpusTool. The results show that learners with lower level English proficiency are able to create cohesive texts by relying on pronouns, conjunctions *and*, *or*, *but*, and relative items. At the next level, the variety of conjunctions increases, the density of pronouns decreases, and spelling mistakes disappear. The study confirmed that successful use of grammatical cohesion is dependent on one's writing skill, while the variety of textual resources gradually increases alongside one's English language proficiency.

Key words: grammatical cohesion, systemic-functional approach, examination corpus, UAM CorpusTool, statistics, discourse analysis.

CONTENTS

INTRODUCTION	1
1. GRAMMATICAL COHESION IN EFL WRITING	4
1.1. Reference.....	5
1.2. Substitution and Ellipsis.....	8
1.3. Conjunctions.....	9
1.4. Cohesion and EFL.....	12
1.5. Summary of Chapter 1	14
2. CORPUS ANALYSIS OF WRITTEN LEARNER LANGUAGE.....	16
2.1. Learner Language.....	16
2.2. Written Language.....	19
2.3. Assessing Written Text	21
2.4. Learner Corpora	23
2.5. Summary of Chapter 2	26
3. GRAMMATICAL COHESION IN THE YEAR 9 ENGLISH LANGUAGE EXAMINATION CORPUS	27
3.1. Description of the Research Methodology.....	27
3.2. Research Context.....	29
3.2.1. Research Population and Sample	29
3.2.2. Data Collection	30
3.2.3. Research and Data Analysis Instruments	31
3.2.3.1. Statistical Analysis for Language Tests.....	31
3.2.3.2. UAM CorpusTool 2.8.14.....	34
3.2.3.3. Systemic Functional Approach to Discourse Analysis.....	37
3.2.4. Research Procedure and Data Analysis	38
3.3. Presentation and Interpretation of the Findings	38
3.3.1. Research Question 1	38
3.3.2. Research Question 2	46

3.3.3. Research Question 3	49
3.4. Summary of Chapter 3	56
CONCLUSIONS	57
THESES	59
REFERENCES	60
Appendix 1: Descriptive Statistics of Grammatical Cohesive Devices	64
Appendix 2: Comparative Statistics in Low and Medium Performance Groups	65
Appendix 3: Comparative Statistics in Medium and High Performance Groups.....	67
Appendix 4: Comparative Statistics in Low and High Performance Groups.....	69
Appendix 5: Statistical Representation of Unsuccessful Cohesive Ties	71

INTRODUCTION

The role of learner corpora research in foreign language teaching and testing has substantially increased over the past years. A rapidly growing number of language teachers, testers, and researchers are developing unique software and methodologies for analysing learner language via learner-corpora. To name but a few, English Profile (English Profile, n.d.), project Merlin (Online 1), the TREACLE project (O'Donnell, Murcia, Garcia, Molina, Rollinson, MacDonald, Stuart, Boquera, 2009) among others, are invested in a bottom-up approach to finding linguistic material that would profile language proficiency levels.

Some of them, such as project Merlin, are linked to the descriptors that would define the levels of the Common Europe Framework of Reference (CEFR) which have so far 'been produced by language specialists largely using their insight as expert users and teachers of the language' (Online 1). It can be therefore observed that interlanguage is widely recognized as a body of developing language, with its own functions and competences at various levels. As such, it should be analysed and described in its natural form, and, regrettably, in Latvia, this approach has presently not gained the attention it deserves.

The model of language ability, as summarized by Bachman and Palmer (1996), goes far beyond the scope of this paper. The focus is therefore put on the discourse competence – the ability to formulate logical, organized and coherent utterances (Council of Europe, 2011: 123). To be more specific, the main research interest is in cohesion, which manifests itself in the skill of linking various parts of a text into a unified whole. This has been an attractive aspect of research, ever since the seminal work on cohesion by Halliday and Hassan (1976) was published.

Nevertheless, it appears that the majority of studies are concerned with cohesion in either samples of native language or learner language with high language proficiency, for example, cohesion in academic writing by Pearson and Pennock Speck (2005), Leo (2012), Zhao (2014), and many others. As a teacher of low to intermediate English proficiency level learners, the author of this paper has grown to appreciate the interlanguage at lower proficiency levels and is highly interested in a deeper analysis of how the meaning-making mechanism of a language learner operates at a cohesive level. The notion of treating language as a meaning-making body and analysing authentic language coincides with the theories propounded by systemic functional linguists, thus this study follows systemic functional analysis approach.

To summarize the above, the present paper attempts to fill **the gap** of corpus-based studies of learner language in Latvia and lack of research on cohesion in lower level writing

by compiling an examination corpus comprising learner texts produced in the writing test of Year 9 English examination, and applying a systemic functional analysis to it. The examination corpus evidently represents low to intermediate English proficiency, roughly corresponding to the CEFR proficiency levels A1 to B1. However, since the national Year 9 English examination in Latvia is not presently linked to the CEFR levels, the CEFR framework could not be applied to the study in a reliable manner. Thus a distinction between low to intermediate language proficiency levels is made instead, to account for the different performances collected in the corpus. The present study could be useful for English language teachers, testers, and curriculum designers in Latvia as a source of data derived from authentic language on cohesion and, consequently, learner language development in Year 9 English writing.

This Master thesis **aims** at investigating how learners with low to intermediate proficiency of English use grammatical cohesive devices to achieve cohesion in writing.

The following **research questions** are put forward:

1. Can cohesive markers expressing reference, substitution, and conjunctive cohesion be detected in the Year 9 English language examination corpus and, if so, which types are represented and to what extent?
2. Can the use of these cohesive markers distinguish between levels of English language proficiency?
3. What are the parameters that render the use of cohesive devices unsuccessful in the Year 9 English language examination corpus, and how could the cohesive links be repaired?

Six **enabling objectives** are formulated in order to reach the goal and answer the research questions:

1. to do a summative analysis of Halliday and Hassan's model of grammatical cohesive devices, relevant research on cohesion in learner's writing, interlanguage, and corpora in language testing;
2. to build an examination corpus by collecting and digitalizing writing scripts, from Year 9 English language examination in 2014,
3. to create relevant systemic networks based on Halliday and Hassan's theoretical model of grammatical cohesive devices and annotate the examination corpus accordingly;
4. to perform a statistical analysis of the categorical data yielded by the annotation procedures in order to identify the types of cohesive textual devices used in learners' writing and to extract language samples from the corpus representing

different types of cohesive markers used in the Year 9 English language examination corpus;

5. to perform a discourse analysis of cohesive chains in a subsample to make more detailed observations on the use of cohesive ties in lower English proficiency level writing;
6. to present and discuss the results, as well as to draw relevant conclusions.

The following **research methods** are used to achieve the goal of this paper:

1. Theoretical method: a review of the literature on cohesive devices by Halliday and Hassan (1976), McCarthy (1991), Halliday and Matthiessen (2004), Eggins (2004), etc.; Corder (1974), Selinker (1974), Saville-Troike (2006) on learner language and error analysis, Cushing Weigle (2002) on assessing writing, and Sinclair (1996), Granger (2002), Barker (2010) on learner corpus.
2. Empirical methods: mixed-method approach of corpus-based statistical and discourse analysis, using the UAM Corpus Tool software for annotation and corpus statistics.

The present Master thesis comprises an introduction, three chapters, conclusions, theses, references, and relevant appendices. The first chapter provides a framework for a systemic functional analysis by summarizing existing literature on grammatical cohesive devices. The second chapter discusses issues pertaining to analysis of written learner language, including interlanguage, assessing writing, and corpus-based analysis of learner language. Finally, the third chapter takes a corpus-based approach to the learner language by analysing textual cohesive devices in 60 scripts collected from the Year 9 English language writing test in 2014.

1. GRAMMATICAL COHESION IN EFL WRITING

The present chapter explores M.A.K. Halliday and R. Hassan's (1976) proposed model of grammatical cohesion, which enables a detailed analysis of cohesion and text organisation patterns in written discourse. Even almost forty years after its publishing, 'Cohesion in English' (1976) still remains the most comprehensive treatment of cohesion (Brown and Yule, 1983: 190, Mahlberg, 2009: 104), and as such, it is widely used in EFL (English as a Foreign Language) research (Pearson and Pennock Speck, 2005; Mahlberg, 2009).

Halliday and Hassan (1976: vii) define cohesion as semantic 'relations between two or more elements in a text that are independent of the structure' either within a sentence or above sentence level. The notion of cohesion is considered to be inextricably linked with Halliday's systemic functional linguistics (SFL) theory – an approach to language that focuses on its meaning-making potential in order to answer such functional questions as 'how do people use language?' (Martin, 2001: 35), by looking at authentic examples of language use (Eggins, 2004: 4).

Within the framework of systemic functional analysis, intuition or expert knowledge is not considered to be a reliable source of information, thus only examples of authentic language in natural contexts validate the analysis (ibid.: 4). The same principle holds true for corpus-linguists (Sinclair, 1996; Granger, 2002), making the merger of the two fields even more convenient.

Following Eggins, SFL is universally accepted 'as a very useful descriptive framework for viewing language as a strategic, meaning-making resource' (2004: 2). This understanding of language as a semiotic system enables an analysis that treats any sample of language as a meaning-making mechanism, and allows asking questions about the learner language without going into a complex error analysis. Thus, studying cohesion from such a perspective can provide answers about (learner) language use in the process of creating meaning via texts. A significant argument to consider at this point is that not every piece of writing is considered to be a text. Halliday and Hassan (1976) see text as a semantic, unified whole of stretches of language in use, thus distinguishing connectivity as the pivotal property in setting text apart from unrelated sentences in the meaning-making process, in other words, text has the property of texture – the quality of 'being a text' (1976: 2).

Connectivity is not only to be understood as lexicogrammatical links among the various parts of a text or co-text, but semantic ties with the context it represents outside of the text. However, texture is manifested mostly within the text.

The English language has a certain set of textual resources for establishing texture, by reaching past the boundaries of a single clause, and these are linguistic and grammatical elements which, according to a much recent publication by Halliday and Matthiessen (2004: 538), are placed in 'either the grammatical zone or the lexical zone of the lexicogrammatical continuum'. For this reason, the overall model of cohesion is divided into two branches, grammatical and lexical cohesion respectively. Mahlberg suggests that 'corpus linguistic research has been accumulating evidence that the boundaries between lexical and grammatical categories are less clear-cut than traditional approaches seem to suggest' (2009: 108). This is an interesting argument, which suggests that a textual resource could belong to both types of cohesion, possibly creating a stronger semantic tie.

Similarly as most research done in the field, the present paper has chosen a narrow aspect of the model of cohesive devices, that is, grammatical cohesive devices. Grammatical cohesion is realized through three distinct types of grammatical systems, which are: reference, substitution and ellipsis, and conjunction (Halliday and Hassan, 1976). According to Halliday and Matthiessen, these semantic links are grammatical in the sense that:

[t]he point of origin of each of these systems falls within one or more particular grammatical unit; and terms within these systems are realized either by grammatical items that have some particular place within the structure of that unit or (in the case of ellipsis) by the absence of elements of grammatical structure. For example, the systemic environment of conjunction is that of the clause. (Halliday and Matthiessen, 2004: 538)

To summarize the above, cohesion is a semantic link between two or more lexicogrammatical elements in a text, establishing texture. Grammatical systems that maintain a semantic link are analysed in this paper, and these textual resources are further discussed in the following subchapters.

1.1. Reference

Reference is defined as a semantic link between two textual elements or a textual element and an item outside of the text (Halliday and Hassan, 1976). Such link possesses referential meaning – information required for a successful interpretation of the textual elements, which can be found in another text element (endophoric reference) or situational context (exophoric reference) (ibid.: 33). These semantic links create texture due to the same meaning being present in the text repeatedly. Although exophoric reference does not operate on a textual level, it is sometimes not considered to be strictly cohesive, however McCarthy (1991: 35) insists on including exophoric reference in a discussion on cohesion, as it forms an integral part of the meaning-making process for the reader/listener.

Considering reference in general, Halliday claims that '[w]ithout reference, you cannot construe your experience as meaning' (2010: 11'49"). This general statement refers to isolation of experiences and lack of semantic links thereof, which rob them of meaningfulness and create ambiguity. Similarly, grammatical resources devoid of referential meaning cannot be considered as successfully used textual elements.

On the linguistic level, reference is discussed vis-à-vis text participants. When people, places, or things participating in the text are introduced in a text, they have to be linked with other mentions of the same participants to ensure continuity. Eggins (2004: 33) asserts that the main function of reference is introducing and tracking these participants throughout the text by linking them back to previous (anaphoric reference) or upcoming (cataphoric reference) mentions.

When such cohesive tie is not properly created, the text becomes ambiguous or less comprehensible, since the identity of the item cannot be retrieved and the referential meaning – established. Halliday and Hassan (1976: 33) emphasize that in different types of discourse, including written language, reference is primarily textual, not situational. To summarize this argument, Brown and Yule (1983: 205) point out that

the concept which interests the discourse analyst is not that of correct (true) reference, but *successful reference*. Successful reference depends on the hearer's identifying, for the purposes of understanding the current linguistic message, the speaker's intended referent, on the basis of the referring expression used.

Thus, in the context of learner language, particular attention should be paid to the *success* of each reference.

Halliday and Hassan (1976: 37) name three textual categories that establish grammatical reference in the English language: personals (category of person), demonstratives (location) and comparatives (identity or similarity), and identifies the textual resources that fall under each category as

Table 1.1 illustrates.

The category of personals is realized through the linguistic resources of pronouns, providing an existential reference, and determiners – giving a possessive reference (ibid.: 38). In addition, there seems to be no reason why other types of pronouns, such as reflexive pronouns, should be excluded from this taxonomy.

Demonstrative reference is expressed through determiners *this, these, that, those*, signalling proximity, adverbs *here, now, there, then*, and the definite article *the* (ibid.). There are two issues concerning grammatical items belonging to the category of demonstrative reference deserving particular attention, and those are the construction *there + be* used as a subject, and the definite article.

Table 1.1 Textual Categories of Reference (Halliday and Hassan, 1976: 38-9)

Personals		Demonstratives		Comparatives	
Pronouns	I, me, you, we, us, he, him, she, her, they, them, it, one	Deter- miners	This, these That, those	Adjectives	Same, identical, equal, similar, additional, other, different, else, Better, more + other comparative adjectives and quantifiers
Deter- miners	mine, my, yours, your, ours, our, his, hers, her, theirs, their, its, one's	Adverbs	Here, now There, then	Adverbs	Identically, similarly, likewise, so, such, differently, otherwise so, more, less, equally
		Deter- miner	the		

First of all, a problematic case of referent *there* is the construction *there + be*, since some grammar reference materials argue that it is void of referential meaning, for example, an extract from an online grammar resource maintained by the British Council (British Council, n.d.):

We often use *there + to be* and *It...* as a subject but they do not refer to any object. *There is / are* is used to introduce a topic, or say that something exists. *It ...* is often used for the weather, time and distance.

On the other hand, in some instances it could be argued that *there* cataphorically provides a locational reference. A grammar reference and practice book for elementary learners (Murphy, 1998: 80) gives the following examples of the construction in question:

- There's a big tree in the garden.
- There's a good film on the TV tonight.
- We can't go skiing. There isn't any snow.

The first two examples include a mention of a participant – place, thus a locational reference between *there* and *the garden* in the first sentence and *there* and *the TV* could be traced. However, in the third example, the location is not introduced and the referential meaning is presumed to be exophoric. Following this argument, the case of *there + be* used as the so-called *dummy subject* will be treated as cohesive in cases where the referent of place is present.

Second of all, the definite article has numerous uses to refer to text participants that are, for example, unique, generalizations of countable nouns in singular, or with adjectives in the superlative degree (Hewings, 2005: 90-4). However, the particular instance of use when the definite article provides textual reference is described in 'Advanced Grammar in Use' as: '[t]he is used when we assume the listener or reader will understand who or what we are

referring to, or when other words in the noun phrase make the reference specific’ (ibid.: 94). In other words, not all instances of the definite article are cohesive.

The third category of grammatical cohesive devices is that of comparative reference. Comparatives of identity or similarity can be expressed through adjectives and adverbs, referring to either general (identity, similarity, difference) or particular (enumerative, epithet) comparison (Halliday and Hassan, 1976: 76). Eggins (2004: 35) notes that ‘[w]ith comparative reference, the identity of the presumed item is retrieved not because it has already been mentioned or will be mentioned in the text, but because an item with which it is being compared has been mentioned’. Clearly, there are much more textual resources for establishing comparative reference than for personal and demonstrative reference.

Admittedly, it is quite complicated for English language learners to use comparatives correctly, especially all the forms of adjectives and adverbs in the comparative degree. Murphy’s most popular grammar reference book for intermediate students (1994: 104) name five rules and four exceptions, a list of variations in spelling, and at least nine constructions, proving the argument that it is a challenging grammar point.

To summarize, reference is a semantic link between inter-textual or inter- and extra-textual elements, realized through grammatical resources of personals, demonstratives, and comparatives. The second category in the model under discussion is substitution and ellipsis.

1.2. Substitution and Ellipsis

Halliday and Hassan define substitution as a type of cohesive tie that links together linguistic words and phrases on the grammatical level ‘as the replacement of one item by another’ (1976: 88). Substitution is often discussed together with ellipsis, for the reason that they, essentially, are two forms of the same cohesive tie. Ellipsis is understood to be ‘the omission of an item’, thus – a substitution by zero (ibid.).

The authors note that in the English language, since one linguistic item can substitute either a noun, a verb, or a clause, there are three different types of substitution, namely, nominal, verbal, and clausal (ibid.:90). Table 1.2 summarizes the lexical items that can function as substitutes for each of the three types:

Table 1.2 Substitution (Halliday and Hassan, 1976: 90-141)

Nominal	Verbal	Clausal
one, ones same, the same	do (does, did, doing, done)	so, not

These lexical items can have many different functions, as explicitly demonstrated by Halliday and Hassan (1976: 90-141) – for example, *one* can function as a substitute, a personal pronoun (function – generalized reference), cardinal numeral, determiner (indefinite article), or a general noun. Thus, the same lexical item can either provide a cohesive tie or not, and if it does, it can be either reference, if it functions as a personal pronoun, or substitution.

Although in cases of substitution and ellipsis, words, phrases, or even clauses are substituted or omitted, it is important to note that the meaning is implied and encoded within the lack of words and understood by all parties involved nonetheless (*ibid.*). Thus, the message is communicated without using actual linguistic resources. Accordingly, an instance where words are omitted leaving the meaning ambiguous or sentence structurally impaired, could not be viewed as substitution/ellipsis.

The reason that makes ellipsis possible and successful is the very nature of language: ‘ellipsis as a notion is probably a universal feature of languages, but the grammatical options which realise it in discourse vary markedly’ (McCarthy, 1991: 43). In typical situations, language functions as a text, and the situation around the communicator functions as the co-text (*ibid.*), thus there is a great deal of information to be received and interpreted, even above the textual level. This means that the message often does not have to be expressed verbally and it can be implied in the context, for example, through exophoric ellipsis. Yet, it should be noted that in the written discourse there is evidently much less context present than in spoken discourse, thus ellipsis should be less frequent.

In the present thesis, both grammatical links of replacement and omission are referred to under the term *substitution*. The final type of grammatical cohesive devices in Halliday and Hassan’s (1976) model is conjunctive cohesion.

1.3. Conjunctions

Egins (2004: 47) affirms that ‘[c]onjunctive cohesion adds to the texture of text, helping to create that semiotic unity that characterizes unproblematic text.’ Conjunctions are lexical items, operating as structural devices that are used for creating an anaphoric or cataphoric semantic link between two or more linguistic elements, such as sentences, clauses, or phrases (Halliday and Hassan, 1976: 227). Conjunctions can consist of one lexical unit (e.g. *and*, *but*, *if*) or several (e.g. *as soon as*, *in order to*) (Wekker and Haegeman, 1996: 61). The most typical environment of conjunctions is that of a clause, thus clausal relationship will be further discussed to understand this type of cohesion.

In SFL, the relationships between clauses are of two types – structurally inter-dependant or tactic, and logico-semantic (combination of logical and semantic) (Halliday, 1994: 218). In standard grammar textbooks, clauses and conjunctions accordingly are subdivided into coordinators (conjunctions that join clauses with the same ‘status’) and subordinators (conjunctions that link clauses with subordinate relations) (Miller, 2002: 66). Going back to the systemic functional approach, Halliday (1994: 218) prefers a distinction between hypotaxis (two elements entered in dominant-subordinate relations) and parataxis (two elements of equivalent status, where no element modifies another). Conjunctions can therefore link elements both in hypotaxic and parataxic relationship.

The logico-semantic distinction is made between expansion and projection, where expansion refers to a secondary clause *expanding* the idea of a primary clauses, while projection comprises clauses where a secondary clause *projects* the primary clause through quoted or reported speech (locution) and quoted or reported thought (idea) (ibid.: 219). Expansion is often achieved with the help of conjunctions, which is the reason it is directly linked with the notion of cohesion.

Originally, Halliday and Hassan (1976: 238) adopted the following classification of conjunction, including four types: additive, adversative, causal, and temporal. More recently, Halliday has further developed the model in Halliday and Matthiessen (2004), focusing more on the semantic relations expressed through conjunctions. Halliday and Matthiessen (2004: 538-49) distinguish between three types of conjunctive cohesion – elaboration, extension and enhancement – which are seen as a part of the logico-semantic system of the English clause. It appears that systemic functional analysts now prefer this model (Eggins, 2004), thus it will be used in the present research.

The first type of conjunctive cohesion, elaboration, is ‘a relationship of restatement or clarification, by which one sentence is (presented as) a re-saying or representation of a previous sentence’ (Eggins, 2004: 47). One clause enriches the existing meaning of another, not by adding new information, but a descriptive attribute, or by restating, clarifying, exemplifying, or commenting (Halliday and Matthiessen, 2004: 396-99). The relationship of elaboration can be introduced by a relative item (ibid.: 396). As examples of relative items could be mentioned the pronouns *which, that, who, whom, whose* and relative words *when, whereby, where, why* (Hewings, 2005: 106-8).

The second type is extension, which is seen as ‘a relationship of either addition (once sentence adds to the meaning made of another) or variation (one sentence changes the meaning of another, by contrast or by qualification)’ (Eggins, 2004: 47). According to

Halliday and Matthiessen, (2004: 406), the most popular additive conjunctive items are *and*, *but*, *or*.

Thirdly, enhancement ‘refers to ways by which one sentence can develop on the meanings of another, in terms of dimensions, such as time, comparison, cause, condition or concession’ (Eggins, 2004: 48). Under this category fall the majority of conjunctions from the original model of cohesion. Table 1.3 provides a summary of the functions, some examples, and the symbols systemic functional linguistics use for each type:

Table 1.3 **Conjunction: examples** (Halliday, 1994; Halliday and Matthiessen, 2004)

Type	Elaboration	Extension	Enhancement
Symbol	=	+	x
Function	restates, comments, exemplifies	new information, exception, alternative	circumstantial features of time, place, cause/reason, condition, result
Examples	<i>for example, like</i> relative items	<i>and, or, but</i>	<i>because, so, yet, then</i>

To summarize, the final type of grammatical cohesion is expressed through the linguistic resources of conjunctions, operating mainly in the environment of a clause, creating parataxical or hypotaxical structures. In SFL they are subdivided into conjunctions creating the relationships of elaboration, extension, or enhancement.

Concerning the whole model of grammatical cohesion, remarkably, quite a number of these textual resources or grammatical cohesion markers can be found in the New General Service List – an updated corpus-based list of the core vocabulary of English (see Figure 1.1).

1. the	11. he	21. this	31. go	41. can
2. be	12. for	22. as	32. so	42. get
3. and	13. they	23. she	33. all	43. think
4. of	14. not	24. at	34. about	44. like
5. to	15. that	25. but	35. if	45. more
6. a	16. we	26. from	36. one	46. who
7. in	17. I	27. by	37. would	47. when
8. have	18. on	28. will	38. know	48. what
9. you	19. with	29. or	39. there	49. time
10. it	20. do	30. say	40. which	50. see

Figure 1.1 **New General Service List: Top 50** (Browne, Culligan, and Phillips, 2013)

The most frequently used lexical item in English is the definite article. Additionally, pronouns *you, it, he, they, we, I, she* can be found on the list, alongside conjunctions *and, but, or, if* and determiners *that* (also relative item), *this, there*, relative items *which, that* (also determiner), *who, when, what*.

The discussion on cohesion is continued in the following subchapter by approaching cohesion from the EFL perspective, including foreign language testing, and summarizing existing research done on cohesion in learners' writing.

1.4. Cohesion and EFL

As previously mentioned, in order to successfully create texture, all cohesive ties should be 'tied up'. The function of cohesive ties may be labelled as – to reduce uncertainty and help the reader decode and process the text with less effort (James, 1998: 159). Yet, the interpretation of the meaning of cohesive ties depends not only on the writer's skilful use of cohesive ties, but also on the readers pre-existing knowledge on the subject, as well as reading skills (McCarthy, 1991: 27).

For this reason, cohesion should not be neglected in the EFL classroom, since English language learners may not be explicitly aware of English-specific patterns governing texture. As Mahlberg (2009: 104) leads us to believe, awareness of cohesive ties is an essential part of reading and writing skills, especially if one wishes to achieve native-like English proficiency. Furthermore, current teaching practices do not seem to be aware of this problem or simply neglect this issue due to time-restrictions. The present paper author's experience emphasizes this point, since there have been several complaints from her students that a text was misunderstood, because the other end of a cohesive tie could not be traced. As Mahlberg words this problem: '[i]t is clear that issues that we encounter in pedagogic approaches to cohesion are not merely a consequence of the requirements of the classroom, they also reflect general linguistic beliefs' (ibid.: 105). Hence, the discourse competence is generally not attributed the same linguistic significance as grammatical and lexical competences.

Prior to undertaking an analysis of cohesion as a means of successful text creation, two additional points have to be considered. Firstly, intersentence cohesion, that is, cohesive ties above the sentence level, should be of particular interest in describing a text, since it is the variable aspect that signals differences in texts (Halliday and Hassan, 1976: 9). Secondly, in reference to the density of cohesive ties in a text, Halliday and Hassan maintain that a text can be described and analysed in regard to the number and types of ties present in it (ibid.: 4), and this argument will be one of the building blocks of the analysis in this paper.

Hence, a question arises to what extent lack of cohesive markers could be considered as a signal for a less developed discourse competence. James (1998: 142) propounds that textual errors are created by misuse or omission of cohesive ties. On the other hand, the author also argues that using cohesive devices is not always required in order to create or decode the

meaning of a text, and at the times when the reader can easily make inferences and connect the elements of a text, cohesive devices can be undesirable (ibid.: 159). Thus, one must be careful in analysing cohesive markers or lack thereof, and interpretations should be carefully made on the basis of the actual need for a cohesive tie.

Shifting the discussion above the sentence level, from a linguistic perspective, cohesion is seen as a semantic aspect of discourse competence, a relationship between the lexical elements and sentences that create texture. The term *discourse* in SFL is used to refer to the semantic level of a spoken or written text – ‘relations of meaning across text’, above lexical and grammatical level (Eggs, 2004: 24). In discourse analysis, cohesion is often discussed together with coherence, which is seen as extra-textual continuity of the text in social and cultural dimensions (ibid.). Since coherence is an extra-textual element, it is not discussed in this paper for practical considerations.

From a testing perspective, cohesion alongside with coherence is widely acknowledged in writing tests. For instance, in the analytical scale used for marking writing tasks of one of the major international tests for English language learners, the IELTS exams, coherence and cohesion is one of the four major categories (Exam English Ltd, 2014). Likewise, the CEFR proficiency level descriptors include cohesion under discourse competence. Yet, the CEFR cannot provide a reliable account of research done on the development of discourse competence, judging by the use of hedge *likely*:

In learning a foreign language, a learner is likely to start with short turns, usually of single sentence length. At higher levels of proficiency, the development of discourse competence, the components of which are indicated in the section, becomes of increasing importance. (Council of Europe, 2011: 123)

It also implies that in lower levels discourse competence is not of importance, an implication with which it is difficult to agree. Perhaps, regarding cohesion as not significant for lower level learners is the reason that in later stages it causes difficulties for writing tasks and tests. Table 1.4 presents the CEFR descriptors for cohesion and coherence.

These descriptors, translated in Halliday and Hassan’s terms, imply that the lowest level learners do not use above-sentence cohesion. Moreover, creation of texture is not possible, since they can link ‘words or groups of words’, not clauses. The next level enables the use of simple conjunctions, while the level above that allows the learners to establish such texture of a text as manifested in ‘connected, linear sequence’. Finally, upon reaching the level above that, learners are able to form ‘coherent discourse’. Evidently, only the higher levels of English language proficiency enable the learners to ‘create coherent and cohesive text[s]’. Thus, it is to be expected that in lower level writing basic conjunctions will be used to link groups of words, without achieving texture that characterizes an effortlessly written text.

Table 1.4 Cohesion and Coherence: CEFR descriptors (Council of Europe, 2011: 125)

C2	Can create coherent and cohesive text making full and appropriate use of a variety of organisational patterns and a wide range of cohesive devices.
C1	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
B2	Can use a variety of linking words efficiently to mark clearly the relationships between ideas.
	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some ‘jumpiness’ in a long contribution.
B1	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
A2	Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.
	Can link groups of words with simple connectors like ‘and’, ‘but’ and ‘because’.
A1	Can link words or groups of words with very basic linear connectors like ‘and’ or ‘then’.

Turning to the research done in the field of cohesion and EFL writing, the articles are mainly concerned with high English language proficiency and academic writing. Nevertheless, the following conclusions are relevant for the present paper. McCarthy (1991: 40-2) notes that exophoric and cataphoric reference can create problems for language learners, since exophoric reference may contain information not shared by the learner’s cultural background, and cataphoric reference may not seem straightforward for an inexperienced English language user. As a possible solution, McCarthy suggests implicit teaching of cataphoric reference (*ibid.*: 42). Cabrejas (2002: 81) has discovered that elegant texture is very dependent on vocabulary range, and there are no differences in the cohesive density between skilled and unskilled writers (Quoted in Pearson and Pennock Speck, 2005: 275-6). Similarly, Pearson and Pennock Speck have found that the main processing problems are not caused by the definite article and determiners, and reference is mostly pronominal (*ibid.*: 275). These findings point to possible problems encountered by Year 9 English language learners in Latvia that could manifest themselves in the examination corpus.

1.5. Summary of Chapter 1

Chapter 1 has provided a detailed account of Halliday and Hassan’s (1976) model of grammatical cohesive devices – the building block for the present research – including a more recent development on conjunctive cohesion by Halliday and Matthiessen (2004). Grammatical cohesive devices are the grammatical resources used for creating successful textual connectivity, falling under the categories of reference (personal, demonstrative, and comparative), substitution/ellipsis (clausal, nominal, verbal), and conjunction (elaboration, extension, enhancement). Having summarized the research on cohesion and respective CEFR

descriptors, it is expected that writing at lower English proficiency level is characterized by use of pronominal reference and basic conjunctions, whereas effortless texture is not achieved. The following section summarizes theory on written learner language, its assessment, and the main issues in corpus linguistics, which enables a more informed corpus-based study of learner language, and analysing a task of a writing test.

2. CORPUS ANALYSIS OF WRITTEN LEARNER LANGUAGE

Chapter 2 is devoted to issues pertaining to corpora-based studies of written learner language. The theoretical discussion begins with an exploration of the notion of learner language, including a brief note on the treatment of errors. Secondly, development of written learner language is described, combining Halliday's SFL approach with the CEFR's more practical perspective on written language proficiency. Next, assessment of written learner language is briefly dealt with to address those basic concepts in testing, which are directly related to the practical analysis of the examination corpus. Finally, the second chapter concludes with a theoretical outline of the major points on the use of learner corpora vis-à-vis learner language analysis and assessment.

2.1. Learner Language

To facilitate the construct validity of the present study, a distinction between language analysis and learner language or interlanguage analysis is drawn. The present subchapter therefore defines learner language, describes factors influencing the process of target language (TL) development, as well as takes a systemic functional approach to meaning-making in the context of interlanguage.

The term *interlanguage* was coined by Selinker (1974: 35) to mean 'a separate linguistic system based on the observable output which results from learner's attempted production of a TL norm'. It is placed on the L2 acquisition continuum in between one's innate ability to acquire a second language and knowledge of the first language (L1) on the one end, and multilingual competence on the other end, as outlined by Saviile-Troike (2006: 16). In other words, learner language is language in the process of evolving, a term reserved for attempted learning of a second or foreign language, henceforth – a target language – in contrast with native language acquisition.

According to Saviile-Troike, interlanguage is characterized by systematic, dynamic, and variable acquisition of new language phenomenon, reduced form (grammar) and function (communicative needs) (ibid.: 41). As a challenge for learner language analysis, Gass and Selinker note that although patterns can be traced, interlanguage development is primarily individual, as each speaker creates their own language system (1994: 49). This argument seems debatable, since it questions the concept of unified language proficiency levels and standardization in general. Perhaps it is more likely that although interlanguage development

follows the same steps and can be classified according to a standardized set of proficiency levels, individual differences can be traced.

In 1970's, when the notion and importance of interlanguage was brought to light, it was considered that the learner language system is influenced by such factors as language transfer, intralingual interference, sociolinguistic situation, modality, age, stability of interlanguage and a universal hierarchy of difficulty (Richards and Sampson, 1974). A much more recent summary on learner language by Saville-Troike argues that the necessary condition for language learners is language input, while facilitating factors include feedback, aptitude, motivation, and instruction (2006: 20). The basic processes of interlanguage development comprise positive or negative transfer from L1, which is successful or unsuccessful cross-linguistic influence, for example, on the phonetic level a negative transfer is a foreign-sounding accent (ibid.: 19). Another type of transfer is intralingual transference, which occurs when a learner makes correct or incorrect observations about the target language structures (Richards and Sampson, 1974: 6). In the context of cohesion, these arguments seem to suggest that transfer is an important factor to be considered.

Since cohesion is probably a language universal, although expressed through different textual means, it could be hypothesized that English language learners at lower proficiency levels transfer grammatical cohesive devices from the native language. As, generally, principles of achieving grammatical cohesion are not explicitly taught at lower levels, their presence require a different explanation, other than instruction. If the learners compose the message in the Latvian language, which is probably the case, and, as a result, translate these sentences in English before writing them down, they automatically try to observe innate cohesive patterns of the Latvian language.

It could be assumed that cohesive use of pronouns, determiners, comparatives, relative items, and basic conjunctions is acquired through a lexical transfer from L1. Textual resources that cannot be transferred with such ease from the Latvian language include the definite article, since the category of definiteness is expressed through adjectival inflection (Skujiņa, 2005), and construction *there + be*, which in Latvian is achieved through word order. Furthermore, the comparative degree of adjectives follow different principles in the English language, as mentioned in Chapter 1.

In addition to the previously discussed characteristics of interlanguage, certain creativity on the part of language learner as opposed to merely mimicking the input material is also expected (Saville-Troike, 2006: 18). Creativity here is used to refer to attempts of combining language forms and functions to arrive at new meanings. It can be assumed that, in general, some learners are more fearless of mistakes than others, and they attempt language structures

which they have not fully mastered yet. Since the boundaries of terms *mistakes* and *errors* seem to be fuzzy at this stage, an explanation of terminology is the next point of discussion.

In reference to errors made by English language learners, two schools of thought exist. The first belief is that errors point to inadequate teaching methodologies, while the second follows the approach of errors viewed as a normal part of the imperfect world we live in (Corder, 1974: 20). The present paper supports the second approach and maintains that errors are an integral part of creative learning process and therefore should be treated as an authentic source of learner language development.

Corder in his seminal work on error terminology, distinguished between slips of tongue or unsystematic chance *mistakes*, and systematic *errors* that serve as an evidence of transitional competence, with the remark that it may not be easy to distinguish between the two (ibid.: 25). According to Corder (ibid.), errors provide data on the progress of learner goal achievement for the teacher, research evidence on language acquisition, and a device for learning (the previously discussed aspect of creativity). Flowerdew (2009:86) asserts that error analysis allows establishing the level of learner's language development. The most appropriate summary of this point is a statement by Saville-Troike, in which she claims that although cognitive processes taking place during second language (L2) acquisition cannot be directly studied and analysed, it is possible to draw inferences by looking at the language learners produce at different stages of learner language development (2006: 19), including different types of errors.

The SFL approach to language acquisition analysis is a functional one, recognizing seven functions of a language that every child sets to master (Halliday, 1973). Saville-Troike (2006: 54) maintains that the SFL model of language acquisition has proven beneficial in the context of learner language analysis, showing that TL acquisition adds multilingual meaning potential and new linguistic forms.

On interpretations of meanings, SFL linguistics claim that a meaningful text consists of a meaningful use of phonological and lexico-grammatical codes (Eggins, 2004: 27). Coming back to the previous discussion on texture and cohesion, Eggins (ibid.: 28) asserts that "sense' is what we're always looking for in language. If text is a 'unified whole', it is a whole unified in terms of meanings, not in terms of form.' This emphasizes the importance of semantic links, as opposed to correct use of form as significant factors in the meaning-making process, which, if created successfully, could render the text meaningful despite certain grammatical or lexical errors. On the other hand, even though the text may follow conventional grammar and use appropriate vocabulary, it can be difficult for most readers to make sense of this text. Eggins suggests, it can be due to the text 'not hanging together' (ibid.:

29). In other words, lack of cohesion or coherence can render text links irretrievable, and, as a consequence – interlanguage becomes meaningless.

2.2. Written Language

The present study analyses learner language in the written mode, thus such matters as written language in connection to spoken language, and differing writing skills need to be addressed to enable a more informed practical discussion on written learner language.

To start the discussion on written language; perhaps it is worth quoting Kramiņa (2000: 109) who states that:

[w]riting makes a special contribution to the way people think. When we write we compose meanings. We put together facts and ideas and make something new. We create an intricate web of meaning in which sentences have special relationships to each other.

Hence, it can be assumed that writing has an important role not only in cognitive, but also in social meaning-making processes. The examples of diaries as an interaction with emotions, written online communication for social transaction, or research papers for knowledge accumulation show that writing has a very substantial role across a range of communicative purposes.

From a slightly more technical perspective, the CEFR define writing or written production as a set of activities as the result of which a written text is produced for the audience of one or more readers (Council of Europe, 2011: 61). Cushing Weigle points out that ‘written language is [...] a distinct mode of communication, involving among other things very different sociocultural norms and cognitive processes’ (2002: 19).

Writing is frequently contrasted with speaking, and they are often seen as the opposite poles of a binary productive communication continuum. Although speaking used to be held in a higher value, nowadays neither of the modes of communication are regarded as a superior to the other, yet the differences among the two across several dimensions, such as, textual features, contexts of use, mental processes, etc., are recognized (Cushing Weigle, 2002: 15). According to Lewis (1986: 154), expert knowledge of written language is impossible without first mastering spoken forms, and a spoken fluency requires a mastery of written forms. This claim shows that both modes of communication are inextricably linked.

On the subject of socio-cultural aspect of writing, Cushing Weigle notes that

[t]o some extent, the ability to write indicates the ability to function as a literate member of a particular segment of society or discourse community, or to use language to demonstrate one’s membership in that community. (2002: 22)

That is the reason why writing is an important skill to be learned at schools, since it is also an instrument of expressing one's knowledge, and on the basis of congruent applications of textual norms, the text is to be assessed and the writer evaluated by the respective community.

However, several authors emphasize that neither all native speakers, nor all non-native speakers possess the same level of writing skill ability (Lewis, 1986; Cushing Weigle, 2002; Pearson and Pennoch Speck (2005)). This leads to several important conclusions in connection with the empirical research of the present thesis.

Firstly, Pearson and Pennoch Speck remark that non-native errors do not necessarily derive from negative transfer or lack of learner language development (2002: 276). A possibility that the creator of the text lacks the necessary writing skills should be considered.

Secondly, Lewis (1986: 154) notes that 'writing well is a minority skill', both in the native language and the target language. From this it is to be assumed that only a minority of language learners are expected to be excellent writers.

Thirdly, as to writing in the TL, it is believed that writing skill and expertise can be transferred from the L1 to the TL, taking into regard also the language proficiency level (Cushing Weigle, 2002: 35). Cushing Weigle mentions some characteristics of expert writers in contrast with novice writers, namely, that expert writers spend more time planning and organizing the text, considering the target audience, as well as analysing global (organisational, etc.) errors, not only local (textual) errors (2002: 22-3). Yet, these are features of writing process, not production, and are not particularly helpful in text analysis. To summarize all these points, a writing test will invariably test not only language ability and the more specific construct it is supposed to measure, but also, the writing skill as such.

Inevitably, the process of reading and meaning-construction in the text produced by a non-native language user is also inextricably linked with writing skills. Since English is considered to be a writer-responsible language (Cushing-Weigle, 2002: 21), cohesive devices are particularly important means of lessening the difficulty of text processing for the reader. Especially in regard to lower levels of English, there is a considerable difficulty in understanding the exact meaning of the speaker who has a limited set of linguistic resources. Writing mastery is a skill that requires a combination of time and effort; especially so in a foreign language, and as such, some form of cohesion should be present even in low level writing for the reader to be able to decode the communicative message, which could be obstructed by incorrect structures or inappropriate vocabulary.

There are a lot of matters to be discussed if one wishes to properly investigate written language. The point that was necessary to be made in this paper is that writing skill is an

important factor of text construction, and, more importantly, that it differs among learners regardless of their language proficiency.

2.3. Assessing Written Text

A separate section of this paper is devoted to assessment of written text, to deal with questions that pertain to description of an examination corpus, such as rating scales and proficiency levels. In the fields of language teaching and testing, writing is traditionally regarded as one of the four skills, using the visual channel and the productive mode of communication (Cushing Weigle, 2002: 14). Bachman and Palmer (1996: 76) consider writing and speaking not as separate skills, rather as different manifestations of the language ability, and for testing purposes, as different task-types. In order to assess one's written language, the first step is to construct an appropriate test that allows learners to channel their language ability, and writing skill through the written script.

As suggested by Hughes (2003: 83), these are the three underlying issues that any writing test should address:

1. Writing tasks should represent tasks that students are expected to perform.
2. Writing tasks should result in representative samples of learner language.
3. These samples are to be scored in a valid and reliable manner.

Firstly, there are certain text types that students are expected to be able to produce at certain levels, and others, which they are not. Secondly, the writing script should represent only the writing ability, not general intelligence or creativity skills. On the other hand, it is virtually impossible to separate writing ability from writing skills, in the sense that some people are simply better at writing than others, and thus are prone to get better results in a written test, as previously discussed. Coulthard (1994: 1) urges the test developers to acknowledge that some test-takers are better writers than others, and 'to try to account not simply for difference and for how existing texts mean, but also for quality and for why one textualization might mean more or better than another.'

Finally, for the purpose of reliable and valid scoring of the samples, Hughes offers the following options: setting tasks which can be reliably scored, setting as many tasks as possible, restricting candidates, giving no choice of the tasks, ensuring long enough samples, and creating appropriate scales for scoring (2003: 94).

For the reason that the examination corpus dealt with in this paper is collected from national English language examination in Latvia, which is considered to be a high-stakes test,

it is assumed that the test has elicited representative samples, which reflect students writing ability.

The following point of discussion is the rating scales that are used to score written tests. A grading scale illustrates achievement in a test, and such scales are typically used in subjectively marker papers (including writing) by an expert rater (Council of Europe, 2011: 41). Classically, there are two types of scales used to assess writing samples: holistic and analytic. Holistic scoring means that a single score is assigned to a sample based on the test rater's general impression of the writing, with the advantage of yielding quick results and allowing several raters to evaluate the sample (Hughes, 2003: 94-5). Fulcher and Davidson (2007: 98) argue that these scales are not intuitive, as they might at first appear, but based, not only on theoretical models, but also language users' experience.

Analytic scoring, on the other hand, involves assigning a score for each of the aspects that are tested, for example, task achievement, vocabulary range, orthography, etc., resulting in a slightly more reliable manner of scoring, but being more time-consuming (Hughes, 2003: 100). Naturally, the more raters assess the script, the more reliable the final score. In practice, where possible, at least two raters are employed to mark the sample, as in the national English language examination for Year 12. However, Hughes refers to the number four, which research accepts as the most reliable number of raters (2003: 95). Practically, such high number of raters is somewhat far-fetched.

Another important argument put forward by Hughes is that the main job of a rating scale is to simply serve its purpose, since not every scale will be equally valid and reliable anyway (2003: 95). Thus, if a rating scale provides reliable data on the construct the test attempts to measure, and, for example, discriminates between the test-takers in an achievement test; it is to be considered successful.

The results of tests are often used to place the learner on an arbitrary proficiency level scale for standardisation purposes. The CEFR is most notable for its proficiency scales, which define levels of proficiency on a vertical continuum (Council of Europe, 2011: 40). Kramiņa (2000: 29) words the problem of learner language level consistency in the following way:

However, in foreign language teaching practice the notion of a certain level of attainment is rather indistinct. There is a lack of co-ordination between the levels of language proficiency and the teaching aids offered. There is a discrepancy also in terminology used when referring to the language teaching/learning process, teaching aids meant for a definite category of language learners and the language performance rating scales.

The problem of proficiency levels is also encountered in the present paper, since the examination corpus uses a system of grading that is presently not compatible with the widely recognized CEFR proficiency level system.

Having dealt with the most important issues of learner language, written language and its assessment, the next section introduces relevant theory on learner corpora as a tool for analysing written learner language.

2.4. Learner Corpora

The present subchapter discusses the notion of learner corpora in the context of this paper by defining corpora and learner corpora, outlining the main features and applications of a learner corpus, and discussing the use of corpus in the context of SFL.

Granger (2002: 3) defines corpora as ‘electronic collections of naturally occurring texts’. Although the collection and analysis of authentic bodies of texts have not always been digital, last decades have brought about significant changes in practicality and use of technology, allowing computerized language corpora to be built and analysed with more ease and speed. Granger emphasizes the fact that the digital approach to large amounts of texts has a huge potential for analysing language (ibid.). It can be derived that corpus linguistics is a field that is certainly benefitting from recent advances of the modern technologies, and this potential can still be explored.

Despite the term *corpus* being used to refer to different kinds of collections of texts, for a body of language to be truly qualified as a corpus and to be analysed as such, it has to comply with several criteria. Sinclair (1996: Online) sets four criteria that have to be met: quantity, quality, simplicity, and documentation. In case a corpus does not correspond to any of those criteria, the deviation should be specified, and then it is called a special corpus (ibid.).

The principle of quantity refers to the amount of words a corpus contains, and typically the number is rather large, since one of the main benefits of a corpus is its size. However, in reality, the quantity of a corpus represents how difficult it is to collect samples of the language required (ibid.).

The second aspect is quality, which corresponds to the idea of authentic language use. Sinclair argues that ‘all the material is gathered from the genuine communications of people going about their normal business’ (ibid.). It means that the language collected cannot be affected by the fact that it is being collected, nor can it be produced with the intention of building a corpus.

As the third criterion, simplicity or plainness of a text is discussed. According to Sinclair, ‘[t]his means that the user can expect an unbroken string of ASCII characters, with

any mark-up clearly identified, and separable from the text' (ibid.) In other words, a corpus texts have to be organized in clear, simple, and standardised manner.

The process of standardisation is also linked to the fourth criterion, documentation, which means that there has to be separate information about the language collected for a corpus (ibid.). In that way, a researcher can easily access data about the language sample, for example, where the text is taken from, who has produced it, etc., which might be relevant for the research, teaching or testing purpose at hand.

Consequently, a learner corpus is a specific type of corpus, since the values for these criteria differ. Granger provides a definition for learner corpora, which is based on the Sinclair's previously discussed definition of a corpus, that is:

Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance. (Granger, 2002: 3)

Granger also emphasizes the type of language collected for a learner corpus, and that is, foreign or second learner language.

Thus, the value of authenticity for language used has to be considered from a different angle, since the main value of native language is that it does not contain any mistakes, and can serve as a model or provide examples for language learners. However, learner language or interlanguage is often characterized by non-standard use of language, or attempted learning, as previously mentioned. Consequently, non-native speaker data is not treated as authentic for a native speaker corpus and is not included in such, but it is authentic for the purposes of a learner corpus. Interestingly, however, Granger claims that learner language is rarely 'natural' in that it has an undercurrent of artificiality to it (2002: 8). For example, activities set during controlled practice of language, scaffolding, prompts given in the task, by the teacher or other student – they all help the student to construct language, interfering in natural production.

As a consequence, it has to be remembered that learner corpora include samples of controlled or, perhaps, elicited language, at least to some (immeasurable) degree (ibid.). Granger concludes that learner language is authentic as far as it is 'the product of a natural classroom activity'; and therefore 'learner corpora of essay writing can be considered to be authentic written data' (2002: 8). Similarly, conversations resulting from natural classroom activity can be considered as authentic spoken data and can be used to create a learner corpus. However, language that results from controlled activities, for example, translations, is not typically included in learner corpora (Nesselhauf, 2004: 128). As refers to authentic language, it is to be assumed that written language produced in test settings is to be treated as authentic, since, for example, the Cambridge English Corpus contains examination texts (Cambridge

University Press, n.d.). Similarly, also written scripts produced in Year 9 English language examination in Latvia are authentic examples of learner language.

Regarding the aspect of quantity, it is important to mention that the amount of a learner corpus represents the size of language that can be collected from learners. The most straightforward option is collecting students' essays written as a part of their examination, similarly as done in the present research; however the problem that arises is accessibility to that kind of information. For example, Cambridge examination providers are known to keep their examination samples a very strict secret, not granting access to every researcher interested in learner language analysis.

On the other hand, essays written by students during national Year 12 examination in Latvia can, in theory, be accessed through the Ministry of Education, and Year 9 - locally through schools. Yet, every year these papers are destroyed during the summer period. A point already mentioned is that, until recently, it was rather time-consuming to collect and analyse all such data in large amounts, which is why, at present, use of learner corpora is a field still in the process of developing, thus it can be expected that in the near future samples for learner corpus will be more easily accessible.

Another issue worth discussing is the aim of a learner corpus. Barker (2010: 638) asserts that the purpose of learner corpora is usually a specific one, which limits their use. For example, essays collected from Year 12 examination do not represent academic learner English and cannot be used for analysing anything else but learner language at that level. Nesselhauf (2004: 127), on the other hand, suggests that learner corpora 'should not be intended merely for use in one particular study (or a limited number of studies) but for more general use'. So that, taking into regard the learner level and non-native language that is represented in a learner corpus, it can be used for several research purposes and several studies, although limited and carefully narrowed in scope.

The aim of the corpus also influences the selection of language data. The criteria are usually external, such as, the level, mother tongue, the school, etc., so that the collection of language samples is representative, and there are not too many variables making the process of analysis too difficult (*ibid.*).

Specifically in testing, there are many different uses of learner corpora, not only informing syllabi, analysing typical mistakes for speakers of certain languages or ages, but also developing a deeper understanding of interlanguage at a particular stage. The aspect that has gained most attention recently is using data from learner corpora to illustrate the nature of language proficiency levels, in particular, identifying the linguistic markers that are characteristic for each level (Barker, 2010: 636-7). Baker notes that this approach is crucial

for the reason that it helps test developers and markers in assessing test takers at a particular proficiency level, as compared with the whole proficiency scale (ibid.: 636).

Alderson (2007: 407) has conducted a study that proves the point of the importance of corpus-based descriptors by comparing the expert raters' judgements with corpora data, and finding surprising inconsistencies (Quoted in Barker, 2010: 637). It suggests that corpus-based studies should be conducted, as they could contribute greatly to the existing descriptors of proficiency levels.

Flowerdew (2009: 87) points out that a useful application of corpus in the context of learner language analysis is the convenient extraction and investigation of collective and individual errors made by learners at different stages of language development. Additionally, tagging enables the researcher to collect large amount of examples illustrating the same type of an error or a case of item overuse and analyse them in the given textual or social (proficiency-level related) context (ibid.).

Inevitably, corpus studies go hand in hand with systemic functional linguistics, since both fields advocate the use of authentic language. Halliday and Matthiessen (2004: 34) assert that 'corpus is fundamental to the enterprise of theorizing language.' The reason it is so central to language analysis is that the language produces and the language thought to be produced differ considerably, especially in experimental conditions (ibid.)

2.5. Summary of Chapter 2

The second chapter has covered a rather wide range of issues in relation to learner language, written language ability and assessment, as well as corpus-based studies. To summarize, learner language or interlanguage is a distinct linguistic system produced by a language learner during the process of learner language development. It is characterized by several factors, the most important of which in the context of creation of textual cohesion was deemed to be positive or negative cross-linguistic influence, or intralingual transference, and errors. Errors are systematic, unsuccessful attempts at creating structures, which the learner has not yet fully mastered, and can be easily analysed with the help of corpus tool. A learner corpus is an electronic collection of authentic samples of learner language, collected with a specific goal in mind and documented in a standardized way. Similarly as corpus linguistics, also systemic functional linguistics rely on authentic language as the only trustworthy source of language data. These approaches are combined in the following chapter.

3. GRAMMATICAL COHESION IN THE YEAR 9 ENGLISH LANGUAGE EXAMINATION CORPUS

The final chapter introduces the methodology of the study, summarizes and presents the results of the empirical inquiry, which was carried out with the aim of exploring how learners with comparatively low proficiency of English use grammatical cohesive devices to achieve cohesion in writing. A systemic functional approach was taken to the Year 9 writing scripts from 2014 English language national examination in Latvia, compiling a corpus, annotating it in accordance with Halliday and Hassan's (1976) model of grammatical cohesive devices, and using the methods of corpus statistics and discourse analysis to answer the following research questions:

1. Can cohesive markers expressing reference, substitution, and conjunctive cohesion be detected in the Year 9 English language examination corpus and, if so, which types are represented and to what extent?
2. Can the use of these cohesive markers distinguish between levels of English language proficiency?
3. What are the parameters that render the use of cohesive devices unsuccessful in the Year 9 English language examination corpus, and how could the cohesive links be repaired?

The first section is a discussion of the research methodology used in this study, which is followed by a presentation and interpretation of the findings.

3.1. Description of the Research Methodology

A mixed-method approach was adopted, including quantitative and qualitative approaches in order to facilitate methodological triangulation. Triangulation is understood to be the use of multiple methods in data collection level or the analysis level to maximize the validity and reliability of the results, and it is presently considered to be the most advantageous approach to carrying out research (Cohen, Manion, Morrison, 2007: 141, Dörnyei, 2007: 24).

After collecting the data and building the examination corpus, the corpus containing learner language was annotated using a systemic functional model. It was followed by a corpus-based statistical analysis of the categorical data yielded by the annotation of the examination corpus. Simply put, 'statistics is a set of logical and mathematical procedures for analysing quantitative data' (Bachman, 2004: 3) and it was found useful in discriminating the

test scores of the scripts and describing the percentage of cohesive items, their density in the corpus, as well as calculating significant differences in the use of grammatical cohesive devices among different groups of test-takers.

Next, a further corpus-based discourse analysis of linguistic elements carrying a grammatical cohesive link in the corpus, as well as analysis of cohesive chains in a subsample was performed to gather in-depth information on the textual resources which were used to achieve grammatical cohesion, and creation of unsuccessful cohesive ties. The method of discourse analysis was selected, since it 'is concerned with the study of the relationship between language and the contexts in which it is used' (McCarthy, 1991: 5). Martin (2001: 35) describes the aim of the method of discourse analysis as: 'to build a model that places texts in their social contexts and looks comprehensively at the resources which both integrate and situate them.' An inquiry in the link between the nature of the language, in particular, cohesive resources, and the social context of language use, being the restricted learner language in test settings, is the focus of interest in the present paper. Moreover, McCarthy (1991: 38) acknowledges the role of discourse analysis in the field of foreign language learning and teaching, as it can yield invaluable insights in grammatical problems.

The downsizing of the sample in the third step of this multiple level analysis can be justified with the fact that qualitative analysis is usually performed on a smaller amount of data than quantitative analysis due to practical purposes, such as, workload and time constraints. Furthermore, Dörnyei reports that such practice is relevant in cases when a larger group is analysed using quantitative methods, and then a more intensive qualitative examination of this sample is necessary, thus a subsample representing the whole sample is selected for the analysis (2007: 273). This is the exact case in this study, where it was important to establish which cohesive devices are used in the whole examination corpus. Yet, it would be clearly impractical to qualitatively approach the creation of cohesive ties in each of the 120 texts, thus a subsample was analysed to provide answers to the third research question.

To summarize the point on the research methods in the present Master thesis using the codes suggested by Johnson and Christensen (2004: 418-9), the approach could be described as *QUAN*→*QUAL*, where *QUAN* stands for quantitative research, *QUAL* for qualitative research, and the arrows indicate the order of the phases of data collection and analysis. In addition, all capital letters demonstrate that the parts are of equal weighting.

3.2. Research Context

The next section of the paper provides insights into the research context of the present thesis, opening with a description of the research population and data collection of the sample, followed by a detailed account of the instruments used in the process of research and data analysis. Finally, the steps taken during the research procedure and data analysis are accounted for to ensure external validity of the results.

3.2.1. Research Population and Sample

The research population in the present study is Year 9 learners of English as a foreign language in Latvia. Despite various learning backgrounds and levels of achievement, the group is believed to represent low to intermediate proficiency of English. In order to assess each learner's achievement, at the end of the Year 9 the learners take part in a national examination in English, consisting of five tests: language use, listening, reading, speaking, and writing (VISC, n.d.). The exams are marked locally by the respective English teachers and the results of each test are expressed in scores, which are rendered into a mark, ranging from 1 to 10. Only the writing test was analysed in the present thesis.

Concerning the selection of the sample for the analysis, the issues of convenience and random selection are of significance, since the Year 9 examination scripts are stored locally in schools, yet it is not mandatory to preserve the scripts after the results have been released. This factor inconvenienced the collection of the scripts. Based on the availability concerns of the data, it was decided that 60 papers, comprising 2 tasks each, would be a sizable sample to outline the cohesive devices used by learners with low to intermediate English proficiency level. Moreover, following Dörnyei, the aim of sampling in a qualitative study is to 'find individuals who can provide rich and varied insights into the phenomenon under investigation so as to maximize what we can learn' (2007: 126). This corresponds to the goal of this research, which is to explore how cohesion is achieved in low to intermediate level writing.

The sample collected for this study consists of 60 examination scripts, comprising 2 writing tasks each. To ensure the representativeness of the sample, examination scripts were collected in clusters. Dörnyei (2007: 98) explains cluster sampling as a method of selection by choosing several larger units of the population, for example, schools, and analysing students within these units. The units were chosen to represent different schools in Latvia, including 2 primary schools, 2 secondary schools and 2 grammar schools (gymnasiums), which correspond to the Latvian system of *pamatskola*, *vidusskola*, *ģimnāzija*. Additionally,

geographical distribution of both rural and town schools was considered and is represented in the sample.

To address the issues of research ethics, the names of the schools and learners participating in the research have been deleted. The only information preserved is about the type of the school, the gender of the learner, and the points awarded for each task.

3.2.2. Data Collection

The collected data consists of both, qualitative and quantitative data. Firstly, language or qualitative data from the Year 9 examination writing test in 2014 was digitalized and turned into an examination corpus. Table 3.1 summarizes the specifications for each of the two tasks in the writing test:

Table 3.1 Test Specifications for the Writing Test in 2014

	Task 1	Task 2
Max Points	15	25
Required Length	40 – 60 words	100 – 120 words
Estimated Time	15 min	25 min
Text Type	paragraph for a website	personal letter
Function	description	giving advice, stating preferences, comparing alternatives
Domain	personal (friendship) public (website)	personal
Requirements	<i>Write a paragraph for an English website describing what makes a good friend.</i>	<i>Write a reply to Sam:</i> <ul style="list-style-type: none"> • <i>give advice on what he/she should do;</i> • <i>write about which school club you would like to attend, and why.</i>
Input Text	-	<i>Hi!</i> <i>I want to join a school club but don't know which club to choose – music, drama or sports. Unfortunately, I don't have time for all of them. What should I do?</i> <i>Yours,</i> <i>Sam</i>
Marking	locally, 1 rater, using an analytical marking scale	
Constructs + points awarded	<ul style="list-style-type: none"> • task achievement (4) • vocabulary (4) • grammar (4) • organisation (3) 	<ul style="list-style-type: none"> • task achievement (5) • vocabulary (5) • grammar (5) • organisation (5) • spelling (5)

Due to the differing text types, function and the number of points awarded for each task, they were entered into the corpus as separate scripts. Thus, all the 120 scripts were digitalized, and pasted into a separate plain text document. The name of each .txt fail contains information about the test-taker: the task, points received, gender and the type of school, for example: T1_07P_F_G_33. Table 3.2 provides the key to the file names:

Table 3.2 Key to File Names

	Variables	Category
1	T1 – Task 1 T2 – Task 2	Task type
2	Any number from 00P to 25P	Points awarded, ranging from 00 to 25
3	F – Female M – Male	Gender
4	S – Secondary school, town R – Primary school, rural G – Gymnasium, town	School type
5	Any number from 01 to 60	Number of the examination paper in chronological order of digitalization, ranging from 01 to 60

Subsequently, these files were entered into the corpus and annotated according to the model of grammatical cohesive devices, discussed in the first theoretical chapter of this paper, to arrive at quantitative categorical data which was submitted for a statistical analysis. Finally, the language data was analysed with the help of a corpus-based discourse analysis.

3.2.3. Research and Data Analysis Instruments

For the research and data analysis, a combination of instruments from the fields of corpus-linguistics, statistics, discourse analysis, and systemic functional linguistics were used. The following subchapters present the corpus tool, statistical measurements used in the study, and the concept of cohesive chains.

3.2.3.1. Statistical Analysis for Language Tests

A rather unfortunate limitation of the analysis of Year 9 examination scripts in Latvia as opposed to Year 12 examination scripts is the fact that the writing scripts of Year 9 examination are marked locally. Furthermore, an analytical marking scale is used, and two scores ranging from 0-15 and 0-25 points for each task respectively is awarded to each script, as described previously. It diminishes the inter-rater validity of the test scores, since the raters

are not calibrated and the tests are marked by only one rater, and does not discriminate proficiency levels in the examination corpus.

Nevertheless, a scope of low to intermediate proficiency levels is evidently represented in the corpus, since points attributed to scripts stretch from minimum to maximum values, as histograms in Figure 3.1 and Figure 3.2 illustrate.

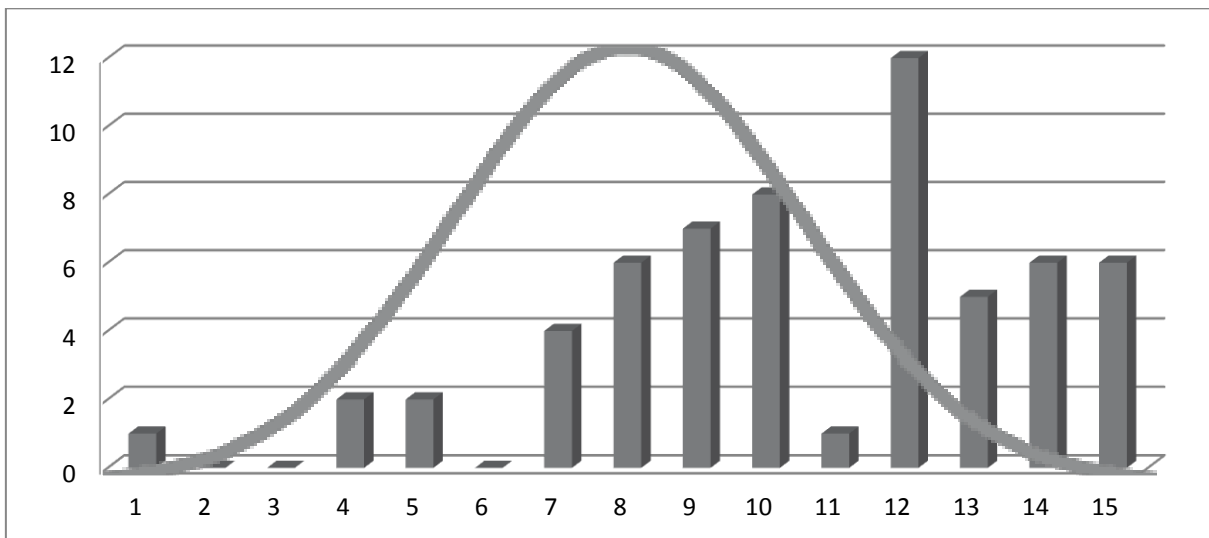


Figure 3.1 Sample Point Distribution for Task 1

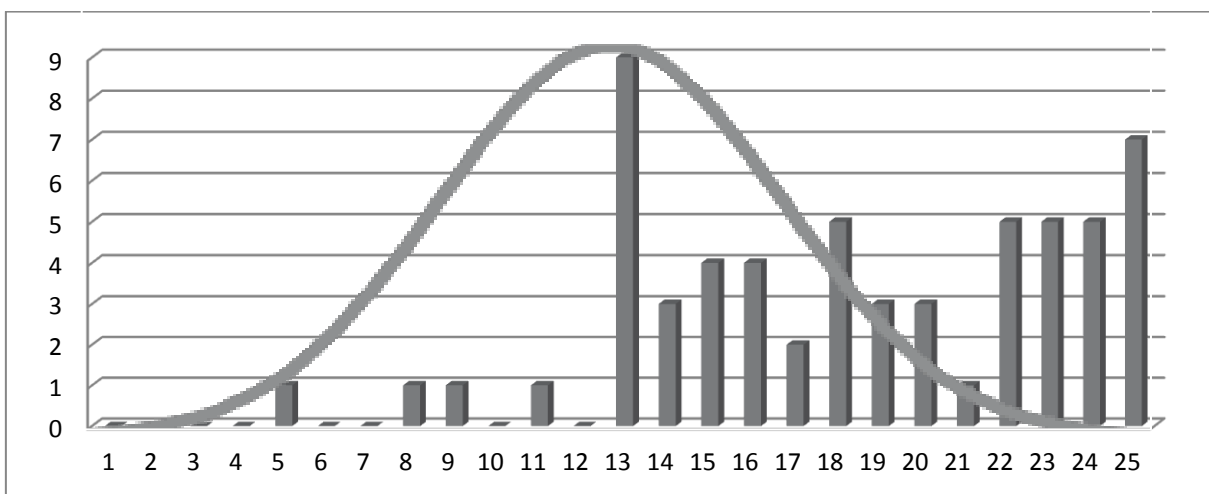


Figure 3.2 Sample Point Distribution for Task 2

It is even possible that levels slightly above intermediate level are represented, since the *bell curve* of a normal distribution cannot be observed. Instead, the first histogram appears to be somewhat negatively skewed, possibly indicating that the marking scale targets lower language proficiency than the test-takers are ready to show. Remarkably, the histogram of the second task differs considerably from the histogram illustrating point distribution for the first task.

Although the current trend in testing is to express scores in the appropriate CEFR Common Reference Levels, it was deemed too unreliable and thus impractical for the scope of

this paper to attempt a reassessment of the sample according to the proficiency levels in the CEFR. For this reason, a different set of categories needed to be established to account for the possible differences in use of grammatical cohesive devices among the writing scripts.

As a result, a simple norm-referencing solution of using standard deviations to compare the scripts with other scripts based on their scores was arrived at. Three levels of performance were identified – low, medium, and high. It is important to note at this point that in these terms *high* does not reflect high language proficiency, rather a high performance in the given task. A set of formulae in the Microsoft Excel was used to calculate means and standard deviations for each task in the sample, and the total score. The results were compared with the means and standard deviations of the whole population (VISC, n.d.). The data is presented in Table 3.3.

Table 3.3 Comparison of Means and Standard Deviations

	Mean (points)	St. dev. (points)	Mean (%)	St. dev. (%)
T1: Sample	10,51	3,22	.70	.21
T2: Sample	18,3	4,99	.73	.20
Total: Sample	28,81	7,88	.72	.20
Total: Population	unknown	unknown	.53	.23

It appears that the means and standard deviations in the sample differ slightly from those of the population. It can be therefore assumed that the findings of this study should be generalized with care, since the sample does not entirely represent the whole population. Nonetheless, as it is neither the goal nor the obligation of this paper to discriminate between the performance groups of test-takers in a way that would correspond to the population division of Year 9 students across the whole country, this will be only considered as a slight drawback.

Next, three groups were discriminated, starting with high and low scores, which were calculated by adding and subtracting one standard deviation from each mean score. The scores from mean plus one standard deviation up to maximum points were assigned high performance, while scores lower than mean minus one standard deviation distinguished low performance group in each tasks. Accordingly, the scores in between exposed the medium performance group. The results are summarized in Table 3.4.

Table 3.4 Distribution of Performance Groups

	T1	T2
Low performance	0-7	0-13
Medium performance	8-12	14-23
High performance	14-15	24-25

As a consequence of these calculations, the corpus was subdivided in three performance groups, which enabled comparative analysis of the grammatical cohesive devices used by each group.

3.2.3.2. UAM CorpusTool 2.8.14

The types of cohesive devices used in the Year 9 examination corpus were derived from a corpus analysis of the learner language. For a systemic functional analysis of cohesive devices the UAM CorpusTool, version 2.8.14 (O'Donnell, 2012a) was selected. UAM CorpusTool is an annotation tool, which enables both, automatic and manual analyses on multiple annotation layers, allows using a predesigned annotation scheme or creating a new one, offers several search options, and, finally, performs descriptive and comparative statistics on the annotated layers (O'Donnell, 2012b).

The main reason for choosing this tool was that the UAM CorpusTool is suited for systemic functional research, since it uses hierarchical system for the annotation – organization of systems into a network, borrowed from SFL (O'Donnell, 2012b :16). Each system network can consist of numerous systems (choices/types) and features (alternatives within the choice/subtypes), which can further contain new systems and features, thus very complex schemes can be created. The networks can be designed prior to annotation, and new systems and features can be added at any point. In addition, glosses or comments can be added to each feature to define its meaning so that the researcher can consult them in the process of annotation. The annotation process means assigning a feature from the network to a whole document or a segment by tagging and coding it.

For the purposes of this investigation, three annotation layers for each script were created: Grammatical Cohesion, Performance, and Task, as Figure 3.3 illustrates.

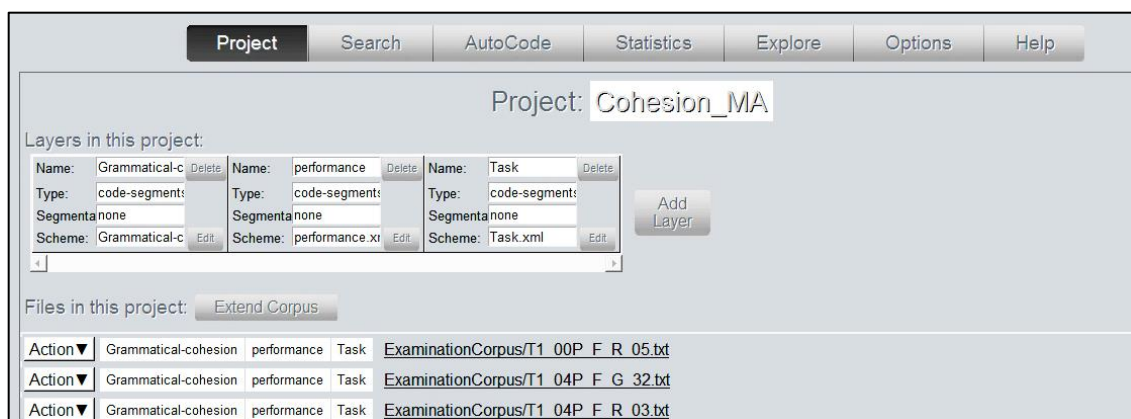


Figure 3.3 Main Windows of the UAM CorpusTool 2.8.14, Examination Corpus Open

Accordingly, three annotation schemes were designed. The first scheme was devoted to grammatical cohesion. The model was based on the previous theoretical discussion on grammatical cohesive devices in Chapter 1. As it is customary in a quantitative study, the majority of the categories were summarized prior to data analysis, not derived from it. The reason for this was that we believe the model of cohesion to be exhaustive and reliable, and a prior model was a prerequisite for the corpus analysis, since annotation required an annotation scheme. Nonetheless, another set of categories were added during the analysis to account for the interpretation problems specific to learner language. Figure 3.4 contains the final version of the systemic network used as the annotation scheme:

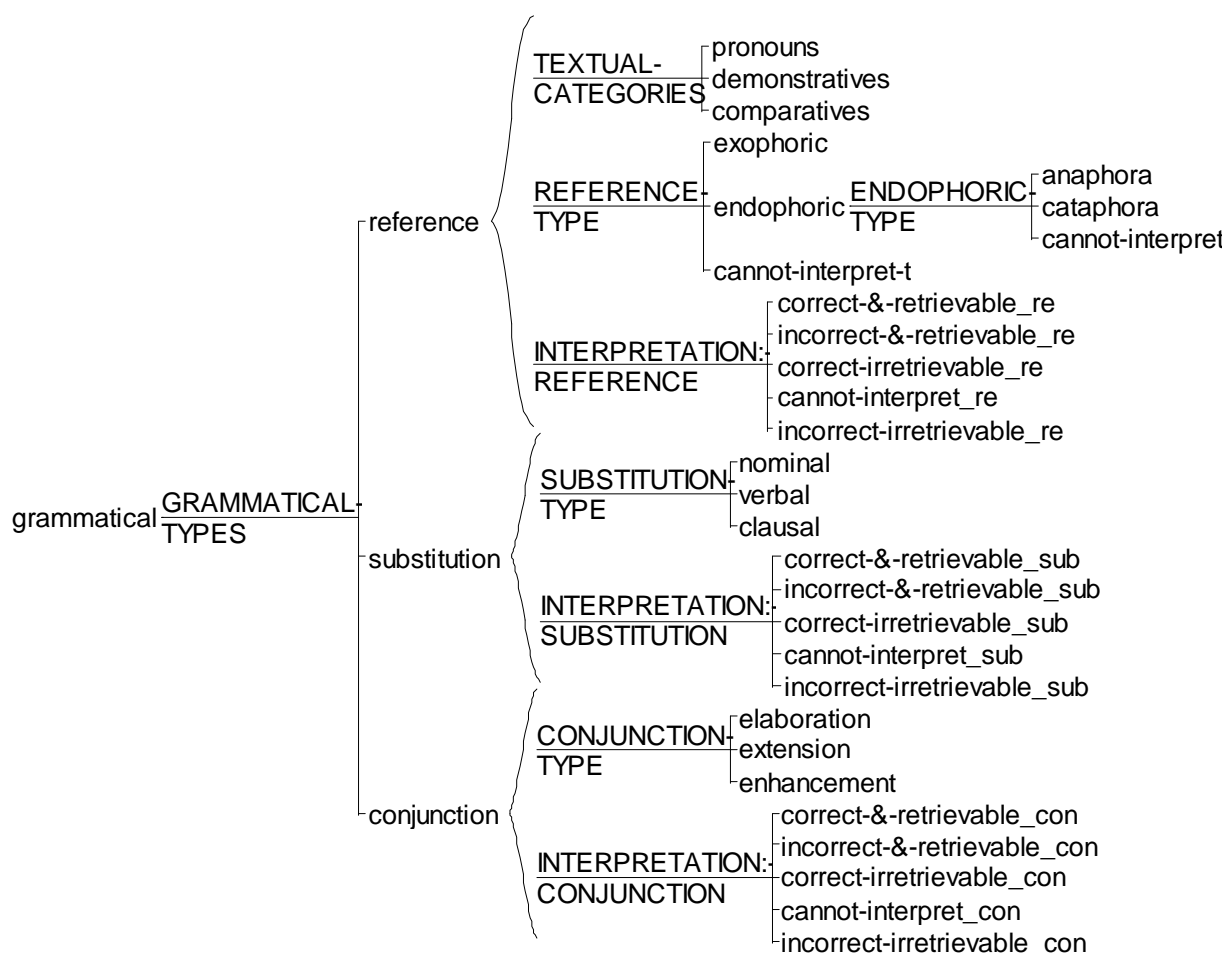


Figure 3.4 Annotation Scheme: Grammatical Cohesion

The system network is titled ‘grammatical cohesion markers’ and it comprises three systems (reference, conjunctions, substitution/ellipsis), and each system has sub features: type and interpretation, which are both assigned to the same segment. Labels used for interpretation are explained in detail in Section 3.3.3. Following this scheme, each file in the corpus was analysed and annotated.

Two additional annotation schemes were created for the programme to be able to distinguish between performance levels (Figure 3.5) and the two tasks (Figure 3.6). Although

the performance level and task types were assigned to the whole script, not particular segments, in order to enable the statistics section, it was found that the option of segment annotation must be selected while creating the layer and then the whole script must be highlighted and the feature of task and performance type assigned to it.

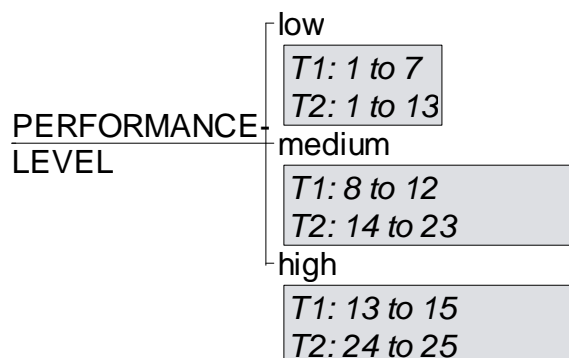


Figure 3.5 Annotation Scheme: Performance Level



Figure 3.6 Annotation Scheme: Task Type

Apart from the annotation feature, the UAM CorpusTool 2.8.14 also offers a statistics section, enabling descriptive and comparative analysis using general text statistics or feature coding options. General text statistics section displays information about the text length and complexity in the selected unit, thus such measurement as density of grammatical cohesive devices in the corpus could be derived. Feature coding shows the distribution of annotated features in the whole corpus or in particular segments, which were performance group and task type. For example, grammatical cohesive devices in low performance group, in Task 1 could be selected and displayed separately from Task 2 or any other performance group.

In comparative analysis section, the programme automatically calculates statistical significance of the findings. Statistical significance, expressed through a probability coefficient (p), shows the possibility of the results to be replicated in the whole population. (Dörnyei, 2007: 210). The UAM CorpusTool 2.8.14 calculates statistical significance with the help of two measurements – T-Statistics and Chi-Squared statistics. It is indicated after each entry and expressed through the following symbols, as shown in Table 3.5.

Table 3.5 Symbols: Statistical Significance in UAM CorpusTool 2.8.14 (O'Donnell, 2012b: 35)

None	Not significantly different.
+	Significant at the 90% level (10% chance of error).
++	Significant at the 95% level (5% chance of error).
+++	Significant at the 98% level (2% chance of error).

Additionally, the search function of the UAM CorpusTool 2.8.14 enables display of lexical items that carry an assigned feature in different categories. The results can be presented by file, as a summary, as a lemmatised summary, or in the form of a concordance, as showed in Figure 3.7:

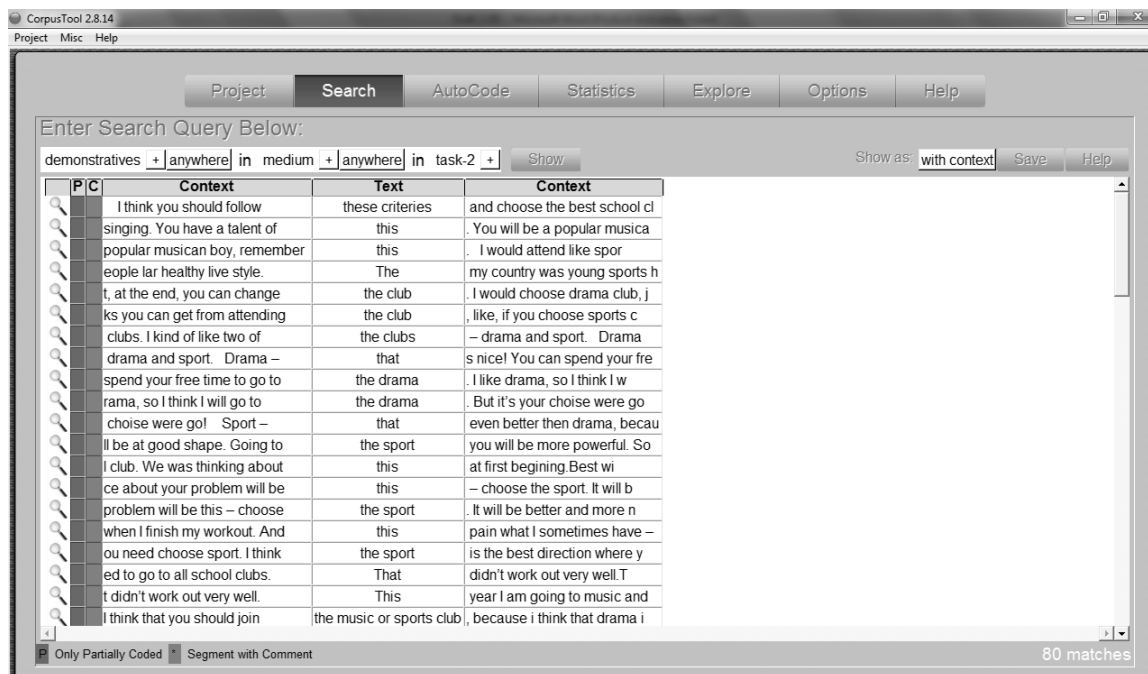


Figure 3.7 Search Feature of the UAM CorpusTool 2.8.14, Examination Corpus Open

The search feature easily extracted language data for each type of grammatical cohesive devices in the whole corpus, as well as different performance groups.

3.2.3.3. Systemic Functional Approach to Discourse Analysis

The instruments for analysing written learner discourse from a systemic functional analysis perspective are borrowed from Eggins (2004) as described in *Introduction to Systemic Functional Linguistics*. The author at length discusses qualitative analysis of referential and conjunctive cohesion in a written text.

Cohesive chains is a term used by Eggins to explain and exemplify a way of applying a systemic functional analysis to surface cohesion in a written text as a ‘convenient way to capture the reference patterns in a text is simply to trace through mentions of text participants’ (2004: 37). The analysis starts with an identification of the major or all the text participants (people, places, objects) and linking them with all of the mentions in the text (ibid.). Dominance of some referents over others could lead to interesting assumptions about text

overall cohesion. In addition, it can be seen whether all mentions of the referents are properly linked together.

As to analysing conjunctive cohesion, the approach of listing clauses as related by logico-semantic relations of elaboration, extension, and enhancement, and marking them using the SFL symbols discussed in Chapter 1 is suggested (Eggins, 2004: 48). Such analysis enables the researcher to see patterns of conjunctive cohesion in the text and trace links among clauses and sentences.

3.2.4. Research Procedure and Data Analysis

Having stated the above, the research procedure consisted of the following steps:

1. Collecting the sample, digitalizing 60 scripts, and entering them into the UAM CorpusTool 2.8.14.
2. Making decisions and carrying out procedures regarding discrimination between the performance groups.
3. Creating the annotation schemes and annotating the corpus.
4. Performing descriptive and comparative statistics on the categorical data.
5. Using the search function of the corpus tool to gather language data.
6. Selecting a subsample for qualitative analysis to be analysed with the method of discourse analysis.
7. Presenting and interpreting the findings, which are described in the following section of the paper.

3.3. Presentation and Interpretation of the Findings

The following subchapter addresses the three research questions by presenting and interpreting the findings of the study.

3.3.1. Research Question 1

The first research question essentially addresses the Halliday and Hassan's (1976) model of grammatical cohesive devices, which was described in the first theoretical chapter, to see if it could be implemented in learner language analysis. It is expected that by answering this question, insights will be gained into how frequently learners employ the three types of

cohesive markers and what categories of textual resources are used under each of the types. The answer to this question will serve as a building block for the next two questions which explore differences among the scripts and unsuccessful use of grammatical cohesive devices accordingly.

The first research question set was:

- Can cohesive markers expressing reference, substitution, and conjunctive cohesion be detected in the Year 9 English language examination corpus and, if so, which types are represented, to what extent, and how?

In order to answer this threefold question, the statistics section of the UAM Corpus Tool 2.8.14 was used, to display the coded features and percentage of use in a form of a systemic network (see Appendix 1). Overall, 3152 instances (3344 words) of grammatical cohesive markers were discovered and tagged in the Year 9 examination corpus comprising 11,301 words in total. All three types of grammatical cohesive devices could be detected in the corpus, if not represented equally, as demonstrated by Figure 3.8.

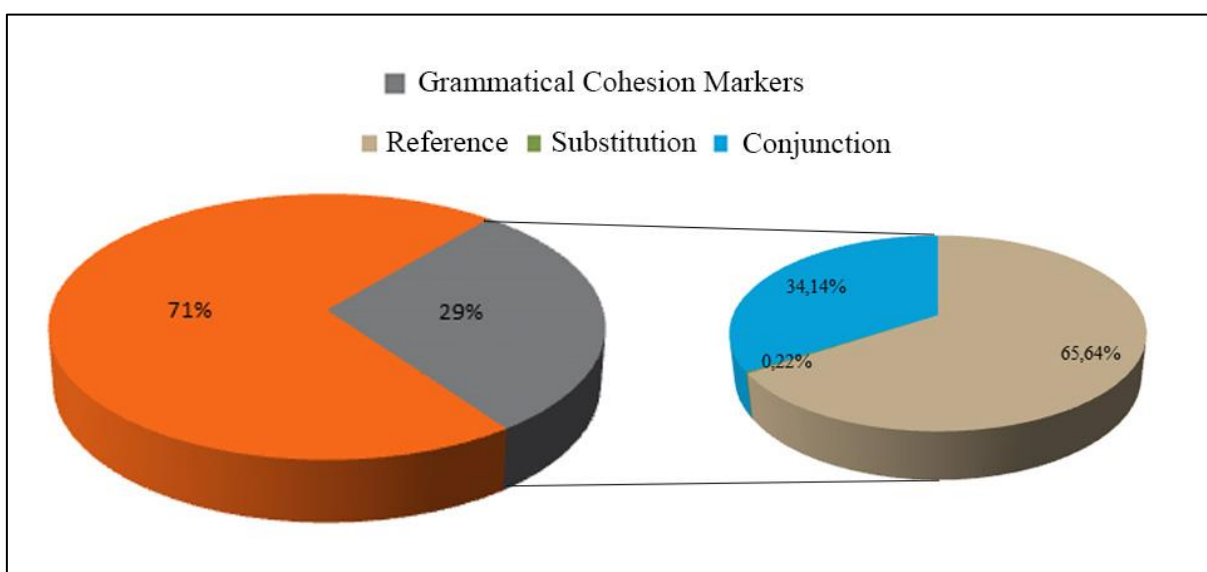


Figure 3.8 Representation of Grammatical Cohesive Devices in the Corpus

As it can be derived from the pie charts, nearly 30% of the words in the whole corpus contain a grammatical cohesive link. The prevailing category is reference, accounting for 65.64% of all the cohesive markers, followed by conjunction, a proportion of 34.14%. There were only 7 instances of substitution, leaving it at a mere 0.22% mark. These results are not surprising, as it was expected that reference and conjunction would be frequently encountered in the corpus. However, to find even 7 cases of substitution in 120 scripts was unexpected, as it was presumed that substitution would appear at later stages of English language proficiency.

Figure 3.9 shows a further subdivision of these three types into subcategories, preserving the colour codes of the previous chart, namely, different shades of blue stand for conjunction, and brown - for reference.

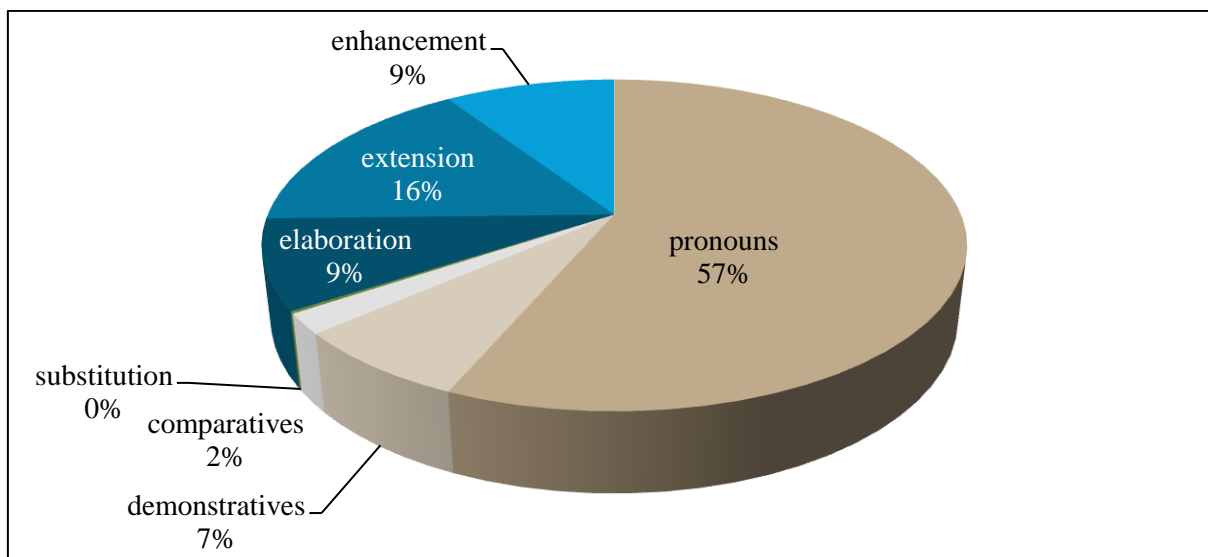


Figure 3.9 Representation of Grammatical Cohesive Devices: Subcategories

From all the instances of reference, pronominal reference was the prevailing type of reference used in the texts, reaching 86.08% mark, and demonstrative and comparative reference was present only in 10.87% and 3.04% of all the cases respectively. This finding does not confirm with the New General Service list, where the definite article occupied the first place, and *that* and *this* was used more frequently as many pronouns. This also implies the possibility that establishing comparative reference is difficult for learners at that level, since one of the text types required comparing three alternatives, thus much higher coverage of comparative reference would be expected. Demonstrative reference requires special attention in the context of Latvian learners, as this category includes the definite article, which is a category not directly present in the Latvian language, thus being notorious for causing difficulties. This issue will be discussed in the section dealing with unsuccessful reference.

As to the types of substitution, only 7 cases of nominal substitution were traced, and only half of these instances were deemed to be correctly used and the cohesive link to be retrievable. Looking for substitution proved to be rather challenging, since the quality of the texts was considerably low, meaning that in many cases the message of a sentence was hard to interpret. Thus, in some cases it seemed at first that a sentence contained an example of substitution, only to appear that the verb or noun was simply omitted, seemingly not intentionally.

Statistically, elaboration (25.74%) and enhancement (27.60%) – two types of conjunctions – appear to be quite similarly represented, with the third type – extension – covering a double of that amount (46.65%). These numbers point to a predominance of additive conjunctions, which was to be expected, however it was initially believed that additive conjunctions would take up a larger portion of all the conjunctions.

As refers to the use of conjunctions, this type appears to hold the most cases of correct uses and retrievable links (88.85%), if compared to reference (62.88%) and substitution (57.41%). It could be partly attributed to the large amount of exophoric pronominal references with irretraceable referents, and excluding this feature from the statistics, the numbers could be evened, without suggesting that conjunctions are used more successfully than pronouns, determiners, and comparative devices to create grammatical cohesion.

To enhance the understanding of how these grammatical cohesive links are achieved, it is important to look at the words learners actually use, which could be easily extracted via the search function, displaying a summary for each coded feature. Unfortunately, the numbers of occurrences of each of these items are not reliable, since the programme differentiates between capital letters and non-capitalized letters, thus counting, for example, *I* and *i* as two different items.

(1a) Personal Reference. The most frequently used textual category is second person personal pronoun *you*, which is followed by first person pronoun *I*, including the not properly capitalized version *i*. Next, possessive determiners *your* and *my* are common in the corpus. This finding corresponds, both, to the text type representing personal domain, and the idea that learners at lower level discuss ideas in the personal domain and shows that they rarely distance themselves from the text. Other pronouns used include *it*, *me*, *he*, *we*, *they*, *him*, *them*, *his*, *she* and a couple of instances of *her*, *our*, *us*.

Additionally, attempts to spell the following pronouns are: *hes*, *youre*, *ou*, *Id*, *youes*, *hi*, *A*, *may*, etc. Furthermore, there are some instances of the politically correct *he/she* and *him or her* to refer to a hypothetical friend, whose gender is textually unknown, but in most such cases third person singular pronoun *he* is used.

(1b) Demonstrative Reference. Determiners commonly used include *that*, *this*, and *there*, and much less common are *these* (alternatives: *dees*) and *those*. Misspellings of these determiners that could be traced are: *thet*, *thiss*, *thear*, *ther*, *tha*, however it appears that these are individual spelling problems.

The use of the definite article in the examination corpus is quite rare, with less than 70 instances of the definite article as a device for establishing a cohesive link in the whole corpus. The majority of article and noun combinations are variations of the

following: *the sports club, the drama club, the music club, and the good friend*. It could be debatable whether this use is actually cohesive, since there isn't any single mention of a particular club or a friend, all of the texts speculate about possibilities of joining a club, be it sports, drama or music in the second task, and good friends in the first task. Nevertheless, these instances of the definite article appear in the corpus statistics, since in their respective texts there seemed to be a cohesive link of a kind.

- (1c) Comparative Reference. Due to the nature of the second task, it was expected that it would require the use of comparative reference – comparing the advantages and disadvantages of the three alternatives of joining the three school clubs and giving arguments in favour of one of them. Additionally, it could be also present in the first task, which required a description of a good friend, and a frequent technique appeared to be describing the opposite. Yet, using comparative reference seems not to be a popular strategy.

The most frequently used textual categories are *too, like, better* and *more*. There are 3 attempts at using the structure '*as ... as*' and one simile – *flew like a rocket*. There are no comparative forms of adverbs and only some instances of adjectives in the comparative degree with ending -er: *better, faster, easier*, and structure more + adjective: *more interesting, more needly, more simple*. As refers to comparative use of *than* – only one instance was traced. Such rare occurrence of comparative forms suggests that providing comparative reference is of difficulty to lower level learners. As a result, the first steps are using words *too, like, more, more* + an adjective, and the comparative form of adjective *good*, namely, *better*.

- (1d) Exophoric and endophoric reference. An important aspect that should be addressed and its particular solution justified in the process of analysis, is the fact that almost one third of all the references was interpreted as exophoric, thus, according to Halliday and Hassan (1976) not relating to textual cohesion per se. The use of the first and second person singular and plural pronouns *I, me, we, us, you* and determiners *my, mine, your, yours, our* accounts for the clear majority of exophoric references in the text, and it required serious deliberation whether to mark them as grammatical cohesive devices at all, and, if yes, which individual instances would be endophoric, and which – exophoric.

Two major approaches to this particular issue were identified: analysing the whole examination paper as a whole, including the test rubrics, etc., or extracting the written scripts and studying them as individual stretches of discourse. If the co-text should be taken into account, that is, the whole examination sheet, which requires the test-taker to write their name and surname on the top of the page, such approach automatically

would render all the first person pronoun instances endophoric → anaphoric. On the other hand, it appears that the co-text (name, surname, test rubrics) views itself as an arbitrary entity, since it asks the test-takers to compose texts for different contexts altogether – a paragraph for a webpage and a personal letter to a friend, in the latter complying with all the norms and standards of an informal letter.

Although a test is not an arbitrary environment per se, the intention of the writing task is not only to produce samples of the English language, but also, even more importantly, texts that correspond to the conventions of the two given genres outside of the test environment. From the point of discourse analysis, a more interesting approach is analysing text construction as separate entities from the test environment. Hence, having considered both options, the analysis, similarly as the test itself, treats the co-text of the test as arbitrary, and analyses the scripts written by the test-takers not as test scripts, but as a paragraph for a webpage and a personal letter accordingly. From such perspective, first person pronouns *I, my, me, we, us* are endophoric only in the cases where the name of the author is present in the text. Such is the case with most personal letters, thus in these cases first person pronouns were deemed to be endophoric → cataphoric.

Nonetheless, the use of second person pronoun *you* and, accordingly, possessive determiner *your*, is not so clear-cut, since it could refer to one of the following:

- (1) the rater of the test and/or the English teacher;
- (2) an abstract reader;
 - (2a) an abstract reader of the webpage that describes good friends (for Task 1);
 - (2b) an abstract reader of the personal letter to a friend (for Task 2);
- (3) an abstract person in general (impersonal use of *you* = one);
- (4) a recipient of the personal letter (for Task 2):
 - (5a) Sam, the sender of the first letter the test-takers were asked to reply to;
 - (5b) any other friend, existing or imaginary, the test takers chose to address despite the task requirements.

Table 3.6. exemplifies these interpretation problems:

Table 3.6 Examples of Exophoric Unknown Referent – Second Person Pronoun ‘You’

Possible Referent	Example
(2a), (3) and (1) unlikely, but possible	<i>If you want to get atleast one good friend you should start think more about of good things and do good stuff. (T1_10P_M_G_25)</i>
(2a), (3)	<i>I'm sure most of you can agree. (T1_15P_M_G_40)</i>
(2a), (3) and (1) unlikely, but possible	<i>I you want to be good friend you have to be comunicable because you is under stend him. (T1_10P_M_R_09)</i>
(2a), (3)	<i>Hello, dear readers! [...] Good friend always help you in all situation, you can tell him all your secrets. (T1_12P_F_G_26)</i>
(2a), (5a), (5b) and (1) unlikely, but possible	all the letters in Task 2 with no introductory remarks or mention of the friends body in the text
(3) or (5a)	<i>Dear Sam, [...] Drama is more for people that are good in front of auditory. There you learn plays etc. (T2_25P_M_G_29)</i>

In point of fact, the question of interpreting exophoric pronominal reference leads to the issue of an ambiguous addressee of the text and/or use of impersonal pronouns in learner texts. Unless indicated by the author of the text, the two are virtually indistinguishable. Oftentimes, the poor quality of the texts in question was an attributing factor to interpretation difficulties. Additionally, the English proficiency level of the majority of the authors of the texts appeared to be too unreliable for them to be in full control of the polysemantic personal pronoun *you*, as it would be in a text composed by a native-speaker or a learner with a higher English language proficiency. Also, the use of these reference items would fall into the category of lexical repetition, thus creating a cohesive link between the pronouns, without actually pointing to the referent, and demonstrating a close connection between the lexicogrammatical systems of the English language in terms of cohesive devices.

- (2a) Nominal Substitution. Following the model by Halliday and Hassan (1976), 7 instances of substitution in 6 different texts were traced (Table 3.7):

Table 3.7 Concordance: Substitution with Context

Context	Text	Context
o!l think you should choose	the one	you want the most. If you don
which you want the most, try	the one	, which you think is more inte
So I think you will take	one	that you really like. I give
help you to choose the better	one	.[name]
d friend, of course if you do	the same	to him/her.
I club but can't decide which	one	to join? If you can't join al
I think you should choose	the one	witch will give you the most

As this corpus extract demonstrates, all cases fall into the category of nominal substitution, using *the one/one* instead of *the club*. There is one instance of nominal substitution using *the same*, which appears not to be cohesive in this sentence. However,

taking in regard the whole paragraph where the device is present, it appears that *the same* refers to all the previously described features that a good friend should do for his/her friend, thus it is (con)textually cohesive, linking the sentences together.

- (3a) Conjunctions for Elaboration. The following relative items are used to create relations of elaboration: *that* (also *that, the, tha* and *then*), *what, who (hu), when, which (witch, wich), how (hou), where (wher)*, even one correctly used instance of *whom*. The many cases of misspellings demonstrate that relative items, which create cohesive ties of elaboration, are easy to notice despite orthographical deviations.

Additionally, a cohesive link of exemplification between clauses or sentences is created with the help of items: *for example* (4) and the more informal *like* (2). They are not frequent in the corpus, showing that the function of exemplification is possibly acquired at later stages of learner language development.

- (3b) Conjunctions for Extension. The function of conjunctions classified under the category of extension is to create a relationship of addition or contrast. Around 90% of all such conjunctive items present in the examination corpus under analysis are: *and (an, un), or, but*, which confirms the claim that these three are the most frequent additive conjunctions. In addition, there are 19 instances of additive conjunction *also*, 4 cases of *furthermore* and two attempts at *one the other hand, from one hand/from other hand*, which could be classified as negative intralingual transference.

- (3c) Conjunctions for Enhancement. Conjunctive items maintaining temporal relations between clauses or phrases traced in the corpus are: more frequently – *then*, and individual uses of *after, yet, since, firstly, at first, second, third, in the end*. Causal relationships are expressed via conjunction *because (becouse, becace, becase, beause, beacuse)* – with over 130 matches, and at least 12 cases of *so* and 4 cases of *that's why*. Conjunction *if* was also used to develop the idea of condition in approximately 80 sentences, similarly item *then* was used.

To summarize the above, grammatical cohesion markers expressing reference, conjunctive cohesion, as well as some instances of substitution can be detected in the Year 9 examination corpus. All subcategories of these three types are present, except for clausal and verbal substitution. Additionally, there are relatively few cases of demonstratives and comparatives. The most frequently used cohesive markers for creating grammatical cohesion are pronouns (first and second person), additive conjunctions *and, or, but*, causal conjunction *because*, conditional conjunction *if*, and relative items *that, what, who, when, which, how, where*.

This is not to say that all of these cohesive links were retrievable or the appropriate marker was selected, neither do these results imply that all the scripts were equally well written. Thus, the following section addresses the question of proficiency levels and cohesion.

3.3.2. Research Question 2

The second research question asks the question of grammatical cohesive devices in connection with language proficiency levels. As a consequence of different proficiency levels represented in the examination corpus under analysis, the use of grammatical cohesive devices cannot be expected to be equally distributed among the different scripts. It leads to the second research question:

- Can the use of these cohesive markers distinguish between levels of English language proficiency?

The answer to this question is twofold, consisting of quantitative and qualitative differences.

- (1) Quantitative Differences. To compare the overall statistics of the use of grammatical cohesive devices among the three groups, corpus statistics section was used. Table 3.8 reflects quantitative differences in general statistics among the three performance groups pertaining to the number of scripts, length, and grammatical cohesion density across the three categories.

Table 3.8 Corpus Data Statistics: General Statistics

	Low			Medium			High		
	Total	T1	T2	Total	T1	T2	Total	T1	T2
- Nr. of scripts:	22	9	13	69	34	35	29	17	12
LENGTH:									
- Words in scripts:	1930	547	1379	6715	2607	3968	2656	1086	1560
- Av. Word Length:	4.05	4.44	3.87	3.97	4.04	9.39	4.04	4.13	3.97
- Av. Script Length:	87.73	60.78	106	97.32	76.68	116.7	91.59	63.88	130
GR. COHESION DENSITY (% of text):									
- Reference density:	.21	.18	.20	.19	.15	.22	.18	.14	.10
- Substitution density:	-	-	-	.0011	-	.002	.0011	.0011	-
- Conjunction density:	.08	0.07	.089	.09	.09	.10	.10	.09	.05

- (1a) Number of Scripts. Evidently, the normal distribution is somewhat present across the three major categories, since there are 22 scripts classified as low performance, while the majority consisting of 69 scripts in total are of medium performance, and 29 are comparatively excellent scripts. Looking at the division between the two tasks, it appears, however, that these numbers are not represented in the same way, thus the two tasks in some scripts composed by the same test-takers would be placed in different

categories. The numbers show that it is easier to get a higher score in Task 1 than Task 2, which is attributed to the fact that Task 1 targets lower language ability, and, as a consequence, Task 2 discriminates more, by eliminating students of lower language proficiency.

- (1b) Length. It can be argued that the length of a script can be an indicator of the test-taker's overall language proficiency. In the light of this argument, it seems reasonable that the lowest performance group would write the shortest texts and the length of the texts would slightly increase for the medium and high performance groups. The results derived from corpus statistics partly confirm this assumption, showing that the shortest texts, comprising, in average, 87 words per script, were written by the low performance group. Curiously, it appears that the medium performance group composed the longest texts, + 10 words per each script, followed by a decrease in the text length for the high performance group – 5.73. This could be attributed to the fact that the test-takers with higher language ability are able to deliver their message more clearly without circumlocution, thus do not need to use so many words to communicate their ideas. However, if the length of the texts is compared between the two tasks, it can be seen that the high performance group composed the longest letters (Task 2).

The average word length measurement indicates that there are many short words in the script, which could be function words, including pronominal reference, articles, prepositions, etc.

- (1c) Grammatical Cohesion Density. The third measurement is the density of grammatical cohesion in the corpus, which is a percentage of the words tagged as carrying a grammatical cohesive link in the whole text (divided by the number of words in the text). It can be observed that the density of conjunctions increases in accordance with the performance levels, while the density of reference decreases. Naturally, it is assumed that vocabulary range increases in higher performance groups, and thus there would be less saturation of function words in the text, including markers establishing grammatical cohesion. As a part of vocabulary expansion is also acquisition of more conjunctive items, which would explain increase in the use of conjunctions, along with the most popular additive items used at lower level.
- (1d) Significance of Differences. Apart from general corpus statistics, statistics displaying feature coding shows significant differences among the three groups. Appendix 2, Appendix 3, and Appendix 4 contain the settings used to arrive at the results, all the three tables of comparison between any two data sets, namely low and medium, medium and high, low and high, as well as Chi Squared statistics, where relevant. From

this data highlighting the significance in differences, it can be derived that the use of personal reference, endophoric reference, anaphoric reference, correct and retrievable referents and conjunctive items increases in medium and high performance groups. Moreover, use of demonstratives, cataphoric reference, cohesive links that cannot be interpreted, and incorrectly used or selected markers as cohesive devices decreases. There are no significant differences in the use of conjunction types and substitution.

(2) Qualitative Differences. The use of textual resources for the creation of grammatical cohesive ties across the three performance groups can be categorized and summarised in the following manner:

- Personal Reference. There are no misspellings in the high and medium performance groups, in contrast with the low performance group. The use of pronouns slightly increases from the low performance group to the high performance group.
- Demonstrative Reference. No misspellings in the high performance group (except *this*) were found, however some misspellings in the medium performance group and several in the low performance group were indicated. The plural of demonstrative *these* appears only at the high (4 instances) and medium (1 instance) level performance groups and *those* (2) + (1). 14 cases of the definite article *the* in the low performance group were traced, while there were 22 cases in medium and 23 cases in the high performance group. Thus, keeping in mind the distribution of the scripts, there is no significant increase in the use of articles between the low and medium performance groups. However, the high performance group appear to use comparatively more articles, since this group is represented by a much lower number of scripts.
- Comparative Reference. There was only 1 instance of comparative reference in the low performance group, 9 in medium, and 5 in high performance groups. Thus, virtually, comparative references appear only in the medium and high performance groups.
- Substitution. Use of substitution as a grammatical cohesive device, however rare, appears only in medium and high level performance groups.
- Elaboration. No misspellings of conjunctive items expressing the relationship of elaboration were traced in the high performance group. *Like* and *for example* appear only in medium performance group, while textual items *especially* and *whom* were present only in the high level performance group. Nonetheless, the

rare instances of these items could be due to individual learner language systems.

- Extension. The low performance group used only conjunctive items *and*, *but*, *or*. The medium group introduces the conjunction *also* and individual attempts at: *furthermore*, *from one hand*, *from other hand* could be observed. As regards the high performance group, the use of *also* and *furthermore* increased, if ever so slightly, and individual attempts at *on the one hand*, *besides* are observable.
- Enhancement. Conjunctive cohesion expressing enhancement in the low performance group is characterized by the use of conjunction *because* and frequent misspellings, as well as items *than/then*, some instances of *if* (4), and individual uses of *since*, *after*, *so*. The medium performance group employs conjunctions: *so*, *at first/firstly*, and individual cases of *yet*, *that's why*, *since*, *however*, *until* show possible attempts at reaching a higher performance level. Finally, in the high performance group the variety conjunctions increases to *so*, *first of all*, *instead*, *secondly*, *that's why*, *after*, *however*, *third*, *second*, *firstly*, but it should be pointed about, that all these instances are still very individual.

To summarize the answer to the second research question, use of the linguistic elements providing grammatical cohesion can distinguish between levels of English language proficiency on the quantitative scale by use of pronominal reference, endophoric reference, anaphoric reference, correct and retrievable referents and conjunctive items, and decreased use of demonstratives, cataphoric reference, cohesive links that cannot be interpreted, and incorrectly used or selected cohesive markers. Qualitatively, spelling mistakes decrease with higher proficiency level learners and range of cohesive items increases by adding the plural form of demonstratives, conjunctions carrying out the functions of exemplification and numbering.

3.3.3. Research Question 3

The third part of the study is concerned with unsuccessful use of grammatical cohesive devices, therefore the final research question put forward was:

- What are the parameters that render the use of cohesive devices unsuccessful in the Year 9 English language examination corpus and how could the cohesive link be repaired?

In order to answer this question, the two tasks will be discussed separately as dependant variables, since they represent two different text types, and the three groups will be

distinguished to account for differences among proficiency levels. The success of the use of grammatical devices was judged according to two parameters – correct use of the textual device (e.g. spelling, choice), and traceability of the cohesive link. Thus, the options for annotation were:

- 1) correct & retrievable – correct use of the textual device, and the cohesive link is retrievable;
- 2) incorrect & retrievable – incorrect use of the textual device, yet the cohesive link appears to be retrievable;
- 3) correct & irretrievable – textual device seemingly correctly used, but the cohesive link is ambiguous or irretrievable;
- 4) incorrect & irretrievable – the text is clear enough to signal that the choice of the textual device is incorrect and the cohesive link is irretrievable;
- 5) cannot interpret – a textual device typically carrying grammatical cohesive link appears to be used, but the text is too unclear.

The first type (correct and retrievable) represents successful use of cohesive devices and cohesive link, while the two later categories account for texts that are too badly written. Since these categories were derived from the learner texts, it can be concluded that there could be two potential types of unsuccessful cohesive ties: correct use of a textual device that is not properly linked to another element, and an incorrect use of a textual element, yet the text makes the link retrievable. If the aim of the cohesive device is to provide a link of the text element to another text element, only the first type is cohesively unsuccessful, while the second type pertains to the field of grammatical or lexical errors. Nevertheless, both types are further discussed.

The presentation and interpretation of the findings in this subchapter is divided into three parts. The first part of this section discusses general quantitative differences among the sample, the second section analyses general qualitative differences, while the third section chooses a subsample and performs a closer analysis on cohesive chains and conjunctive relations in the texts.

- (1) General Quantitative Differences. Corpus-based statistical analysis on the interpretation of grammatical cohesive devices in low to intermediate level writing reveals many instances of the problem cases, namely, incorrect and retrievable, correct and irretrievable references and use of conjunctive items. The full table presenting the statistics on interpretation in the examination corpus can be seen in Appendix 5.
- (1a) Reference. There are a substantially lower percentage of correct and retrievable reference items in T1 than in T2 across all three groups. A wide range is represented,

from 25% in the low level performance group in T1 up to 85% in the high level performance group in T2. The percentage of incorrect and retrievable references steadily decreases from low to high level groups, with almost no instances in the high level performance group in T2. Correct and irretrievable references are more typical for T1 than for T2 in all performance groups, especially for the high performance group. References that cannot be interpreted are characteristic only to the low level performance group, especially in T1.

- (1b) Conjunctions. The majority of cohesive ties created by cohesions are correct and retrievable, ranging from 74% in the lower performance level group to 96% in the higher performance group. There are no incorrect and retrievable conjunctive links in the low performance group in T1, while there are 16% incorrectly used conjunctions in T2, and 6% in the medium group both tasks, with the number fluctuating between Tasks 1 and 2 in high performance group. Correct and irretrievable conjunctive items are not frequently represented in the corpus, with only 4 cases in the low performance group and 13 cases in the medium performance group.

These numbers allow the following conclusion to be made: in general, unsuccessful use of cohesive ties is more characteristic for the lower performance group. Successful use of reference appears to be much more problematic for all learners in comparison with conjunctions. As regards conjunctions, the problem seems to be lexical or grammatical, not cohesive.

- (2) General Qualitative Differences. To understand the problem cases in more detail, a search was performed for all the categories separately.
- (2a) Personal Reference. Under the category of correct and irretrievable personal reference fall the items classified as exophoric, since in such case the item is not textually linked to any presupposition. The cohesive tie could be repaired by introducing an agent, e.g. Hello, dear readers! (*T1_12P_F_G_26*), Hi, Sam/Laura/Kate (conventions for a personal letter). As incorrect and retrievable personal reference, object pronouns and possessives were tagged, including the careless non-capitalization of personal pronoun I, as the following examples demonstrate:
- follow they (*T1_04P_F_R_03*)
 - she talking for they boyfriend (*T1_04P_F_G_32*)
 - he or her (*T1_07P_F_S_56*)
 - apostrophe in *your, hes*
 - capitalization of *i*

It indicates that these structures could be simply acquired at later stages of learner language development.

- (2b) Demonstratives. Correctly used demonstratives with irretraceable cohesive links were mostly uses of *the* and *this ,that*, while incorrectly used where the plurals of the demonstratives:

- This is my thinks. (T2_13P_F_R_02), which supports the argument that the plural form is not acquired yet.

Although the determiners were problematic due to exophoric reference, the definite article is an issue for all Latvian English language learners at low to intermediate proficiency levels, as the results of the present study confirm. There were also instances of negative transfer from the Latvian language, falling under the category of incorrect use of the device with a retrievable cohesive link, for example: *The good friend is that who live outside, no't inside, to computer.* (T1_05P_F_R_01) (LV: labs draugs ir tāds, kurš ...) In addition, examples of sentences that could not be interpreted where observed, such as: *The club very much, but I like that a club.* (T2_11P_F_S_54)

- (2c) Comparative Reference. The discussion in the sections of the first two research questions asserted that creating comparative reference is of difficulty for low to intermediate level learners, and that attempts, which are only made at medium to high performance level groups are often faulty. The present section supports this finding, and the examples have already been provided in the previous sections.
- (2d) Substitution. Overall, it appears that intentional and successful use of substitution, and also ellipsis, is acquired at later stages of learner language development.
- (2e) Elaboration. Incorrect and retrievable cases include:
- confusion between *which* and *who*: [...] *drama or sport who you like more.* (T2_13P_F_S_45)
 - negative transfer: *You have to choose your own what you like better.* (T2_13P_M_R_06)
- (2f) Extension. The only problematic case with conjunctions used to achieve the logico-semantic relationship of extension is with the additive item *and*, when both ends of the cohesive tie that are structurally linked do not appear to be semantically linked: *You need choose where your heart want and I believe you can choose.* (T2_13P_F_S_13)
- (2g) Enhancement. Correct and irretrievable cases of conjunctive items under the category of enhancement include:

- *because*, since what follows does not seem to be either cause, nor consequence, eliminating the explanation of a possible confusion with the conjunctive *so*: *I like twitter, because – I have a twitter conts and twitter was found new friend.* (T1_04P_F_R_03)
- Conditional constructions often do not follow standard patterns. *If you have some things you can discuss with him.* (T1_10P_M_R_09)

Incorrect and retrievable conjunctive cohesive links are created due to the following reasons:

- Spelling of *because* and *that's why* is a difficulty.
- Conditional constructions, which, possibly, should have been tagged as correct and retrievable uses of conjunctions, since the success of the construction is not judged, only the logico-semantic ties in the clause. A case in point: *If you will join, your life will become more interesting and active.* (T1_22P_F_S_52)
- Wrong choice of a conjunction: *In my opinion sport makes you a friends, furthermore a team sports.* (T1_10P_M_G_17)

In summary, cohesively unsuccessful devices include personal pronoun *you*, which should not be exophoric in a personal letter in accordance with informal letter writing conventions, and demonstratives, including the definite article with no traceable cohesive links. The list of textually unsuccessful, yet cohesively successful devices is much longer, and it includes: object pronouns and possessives, capitalization of personal pronoun *I*, plural of *this* and *that*, negative transfer of conjunctive items, comparatives: form, *which* vs. *who*, conditional constructions with *if*, spelling of *because*.

(3) Discourse Analysis of Cohesive Devices in Context. Finally, in order to gather information on overall success of cohesion in a text, discourse analysis of grammatical cohesion was performed in the context of three low level learner texts, representing satisfactory text cohesion, unsatisfactory text cohesion, and no observable cohesion.

(3a) The first text, a script written for the completion of Task 1, was composed by a female learner from a gymnasium, and it scored seven points, placing it into the low performance group:

As it can be seen, there are three text participants:

- the author (I think);
- exophoric *you* = the reader, perhaps (helps you – you always can tell - your secrets – tells you – believe you – trust you – you are sad – walk with you – talk with you);

- a good friend (good friend - he or she – tell him – he tells – he needs to believe – he wants).

Example 1: (T1_07P_F_G_33)

I think god friend makes that he or she always helps you on difficult situeishent. You always can tell him your seecrets and he tels you. He heed to belive you and trust you. When you are seed he wants to go walk whit you and talk whit you abot that.

There are also three cases of extension (+). It could be argued whether ellipsis is present in the second sentence; however, it seems that does not fall into the category of grammatical cohesion, as ‘he tells you his’ would.

It appears that the text is linked through repeated mentions of a good friend and exophoric *you*. Overall, textual cohesion in Example 1 is satisfactory – the density of textual devices is quite high, they are linked properly (except *you*, which is not linked to any participant outside of the text, but all the mentions are internally linked), yet there is not much variety in the use of devices. Despite low performance on the writing test, the text is cohesive, which indicates that cohesion is dependent on writing skill, rather than one’s language proficiency level.

- (3b) The second text is a script written as the Task 1 by a female learner from a rural primary school, and it scored only four points. This script is interesting in the sense that it does not follow the test rubric and instead of describing what makes a good friend for a website, it describes a website that makes a good friend. This text is more creative in comparison with the first text, since the learner attempts to form constructions which she clearly has not mastered, at the expense of interpretable meaning, as well as loosing points on the test.

Example 2: (T1_04P_F_R_03)

I going tell about for twitter. I think, twitter was very good website for meating good friend. I like twitter, because – I have a twitter conts and twitter was found new friend. Every year twitter celebrate birthday in December. I was twitter for 14 years old, since I was a twitter, I don’t work a homework, and I was very angry, because Twitter don’t let my work school homework. This website was found friends pictures, tweets, and follow they. Twitter is my best friend.

There are only two text participants:

- the author (I going tell – I think – I like – I have – I was – I was – I don’t work – I was – my work – my best friend)

- twitter (about for twitter - twitter as – like twitter - twitter conts – twitter was found – twitter celebrate – twitter for 14 years old – I was a twitter – twitter don't let – this website – tweets – they – twitter).

There is no variety in the use of the cohesive devices to link these participants, and, what is more, there is a considerable difficulty in interpreting the meaning of the text. Probably the only semantic links are within these two text participants and conjunctive items and (+), because (x), and since (x).

All in all, cohesion in Example 2 is below satisfactory, and this appears to be largely influenced by the learner's poor English.

(3c) The third example was selected for analysis, since the script did not score any points in Task 1.

Example 3: (T1_00P_F_R_05)

English website is very big. Website who written is littl. I like twitter.com. Twitter frend written is lovly, hangry and emocional. I English website written english. I like internet. Internet is very good and very popular. Internet is teacher life. Internet application is popular. English website it chat written english. Website it written a planet and new frend.

Upon a closer inspection of this script, it seems that there is no texture at all. There are several unconnected text participants: English website, I, twitter, friend, internet, teacher, chat, and planet, with almost no semantic links between them. There are three uses of additive *and* (+), however one of them is cohesively unsuccessful. Since the meaning of the paragraph cannot be retrieved, and there appears to be no textual cohesion, unfortunately, the script can hardly be labelled as a text.

These three examples of successful, unsuccessful overall text cohesion and no texture lead us to believe that in low level writing grammatical cohesion is achieved through personal reference of a small number of text participants and basic, retrievable conjunctive items. Too many textual participants, which are not properly linked, create confusion for the reader. Moreover, successful use of cohesion depends more on one's writing skill than one's language proficiency level.

To answer the final research question, the reasons that render the use of cohesive devices cohesively unsuccessful is not introducing the text participants properly or introducing too many text participants, as well as not linking demonstratives with their referents. The cohesive link could be repaired by mentioning the referents that are at the other end of the cohesive link of demonstrative reference. Unsuccessful textual use of cohesive

markers does not disrupt the cohesive link, but can be improved by acquiring the proper form and use of object pronouns and possessives, plurals of demonstratives, comparatives, conjunctions, and fixing spelling mistakes.

3.4. Summary of Chapter 3

To summarize, the present section has provided in-depth insights into grammatical cohesion existing in low to intermediate written learner language and shown that even test-takers with low English language ability can create meaningful and cohesive texts. It was affirmed that learners at low proficiency levels rely on pronominal reference and basic conjunctions, while learners with a comparatively higher proficiency level use a larger variety of conjunctive items. In addition, the definite article is a problematic issue for Year 9 learners in Latvia. The implications of the main theoretical and empirical findings are summarized in the Conclusion section.

CONCLUSIONS

The present research was carried out with the aim of investigating the use of grammatical cohesive devices in low to intermediate level writing. Three research questions were put forward, addressing the use of reference, substitution and conjunctions; proficiency levels, and unsuccessful cohesion respectively. In order to reach the research goal and answer the questions, literature on grammatical cohesion, written learner language, and learner corpora was reviewed, and a systemic functional approach to corpus-based statistical and discourse analysis was taken.

The theoretical discussion summarized the framework for the analysis of grammatical cohesive devices and concluded that the use of reference, substitution and conjunctions enable semantic links contributing to overall text cohesion. The density and successfulness of the cohesive ties are the factors that set different texts apart. However, written learner language at lower English proficiency levels is not expected to contain effortless text cohesion. Learner language is a separate language system, and an analysis of systematically made errors enables conclusions about its development. Learner language that is created as a product of natural activity is considered to be authentic language data, and held in high regard by corpus linguists and systemic functional linguists alike.

It was affirmed in the empirical chapter that low to intermediate level learners in Latvia successfully use pronouns (first and second person), additive conjunctions *and, or, but*, as well as causal conjunction *because*. In addition, conditional conjunction *if*, and relative items *that, what, who, when, which, how, where* are used to create cohesive ties between clauses and sentences. Moreover, the textual device does not have to be properly spelled or even selected, for example, due to negative transfer, for the cohesive link to be retrievable. Thus, a cohesive marker can be textually unsuccessful, but cohesively successful. This finding implies that the proficiency level of the English language is not directly linked with successful use of cohesive ties.

In order to repair unsuccessful cohesive links, all the text participants should be introduced prior to demonstratives *that* and *this*, and the definite article. Perhaps, the Latvian learners should be given additional exercises to understand the underlying cohesive connections between *this, that, the*, and their respective referents, to become aware of the whole mechanism of pointing and linking. It appears that it would be worth approaching these three items from the cohesive perspective, perhaps, with visual demonstrations of the text links, thus increasing the learners' awareness of cohesion. In addition, second person pronoun *you* should not be left exophoric in a personal letter.

The present study also brought to light some common errors made by Year 9 learners in Latvia, such as, incorrect use or lack of object pronouns and possessives, plural of *this* and *that*, negative transfer of conjunctive items, *which* vs. *who*, conditional constructions with *if*; and spelling, which gradually improved. In addition, Year 9 learners are not comfortable with the function of comparing, and do not tend to use substitution and ellipsis.

The study also shows the development of learner language. For example, the variety of conjunctions increases gradually throughout the proficiency levels, as learners attempt to use other conjunctive items apart from the three basic ones, and acquire the functions of exemplification and numbering. Also, spelling issues and negative transfer pertaining to grammatical cohesive devices gradually disappear with a proficiency level increase.

Perhaps the most interesting conclusion disproves the notion that a text composed by a learner with low English proficiency level is not cohesive. As the examples demonstrated, such learners use grammatical cohesive devices to link not only words, but also clauses and sentences, and create cohesively successful texts. As a consequence, it can be concluded that successful cohesion in a written learner text depends more on the writing skill, than one's language proficiency level.

The results of the present thesis could be useful for teachers, testers, as well as curriculum and textbook developers focusing on low to intermediate English level learners, since it summarizes the major problems of cohesive writing and provides examples of language development. The main strength of this Master thesis lies in its research methodology, a combination of corpus-based studies with language testing, which is a novelty in the context of the majority of the learners in Latvia – the so far neglected lower level learners. An unfortunate limitation of the study was the fact that the test scores were not expressed in the CEFR proficiency levels, thus the research could not be linked to the CEFR. Thus, a further study could be conducted (1) to link the performance levels in writing tasks with the CEFR proficiency levels, (2) to include the Year 12 examination corpus in the scope, (3) to compare Latvian examination corpus with corpora of other language learners, (4) to increase the scope of cohesive devices by analysing lexical cohesion.

THESES

1. Systemic functional linguistics is a school of thought, concerned with answering the questions relating to the meaning-making potential of authentic bodies of language.
2. A text has the property of texture, which is expressed through semantic links, maintained with the help of various grammatical and lexical cohesive devices.
3. Grammatical cohesion is a semantic link between two or more textual elements as expressed via grammatical resources of reference, substitution and ellipsis, and conjunctions.
4. Density and proper linking of cohesive ties are the two factors that point to differences in overall success of text cohesion.
5. Learners do not possess the same level of writing skill ability, and it can be transferred from the native language into the target language.
6. Learner corpora collection and annotation is a useful tool that enables an in-depth statistical and discourse analysis of authentic learner language systems in the process of development.
7. The most frequently used textual resources for creating grammatical cohesion in low to intermediate level writing are pronouns (first and second person), additive conjunctions *and, or, but*, causal conjunction *because*, conditional conjunction *if*, and relative items *that, what, who, when, which, how, where*.
8. Spelling and appropriate selection of the textual resources are not the decisive factors that enable text cohesion.
9. For a successful overall cohesion of the text in low level learner writing, a text should not contain many participants, and all the text participants should be grammatically linked.
10. Learners at lower proficiency levels are able to connect not only words and phrases, but also clauses and sentences, thus creating meaningful and cohesive texts.
11. Successful cohesion in a written learner text is more dependent on the learner's writing skill, than on the language proficiency level.

REFERENCES

1. Alderson, J. C. (2007) Judging the Frequency of English Words. *Applied Linguistics* 28(3): 383–409.
2. Bachman, L. F. and Palmer, A. S. (1996) *Language Testing in Practice*. Oxford: Oxford University Press.
3. Bachman, L. F. (2004) *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
4. Barker, F. (2010) How Can Corpora Be Used In Language Testing? In A. O’Keeffe and M. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics* (pp. 633-645). Abingdon: Routledge.
5. British Council, (n.d.) *There is/There are and It*. Available at <http://learnenglishteens.britishcouncil.org/grammar-vocabulary/grammar-videos/there-are-and-it#sthash.ILwV1WFV.dpuf> [Accessed 16 May, 2015].
6. Brown, G. and Yule, G. (1983) *Discourse Analysis*. Cambridge: Cambridge University Press.
7. Browne, Ch., Culligan, B., and Phillips, J. (2013) *New General Service List (1.01)* Available at <http://www.newgeneralservicelist.org/> [Accessed 16 May 2015].
8. Cabrejas Peñuelas, A. (2002). Cohesion, Coherence and Writing Quality: A Comparison. In F. Fernández Fernández (ed.) 71-83.
9. Cambridge University Press. (n.d.) *Cambridge English Corpus*. Available at <http://www.cambridge.org/lv/cambridgeenglish/about-cambridge-english/cambridge-english-corpus> [Accessed 23 April 2015].
10. [CEFR] *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (2011) Council of Europe. Cambridge: Cambridge University Press.
11. Cohen, L., Manion, L. and Morrison (2007) *Research Methods in Education*, 6th edit. London and New York: Routledge.
12. Coulthard, M. (1994) On Analysing and Evaluating Written Text. In M. Coulthard (ed.) *Advances in Written Text Analysis*. London and New York: Routledge.
13. Corder, S. P. (1974) The Significance of Learners’ Errors. In J. C. Richards (ed.) *Error Analysis: Perspectives on Second Language Acquisition* (pp. 19-28). Abingdon: Routledge.
14. Cushing Weigle, S. (2002) *Assessing Writing: Cambridge Language Assessment*. Cambridge: Cambridge University Press.

15. Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.
16. Eggins, S. (2004) *An Introduction to Systemic Functional Linguistics*. Second Edition. London: Continuum International Publishing Group.
17. *English Profile* (n.d.) Available from: <http://www.englishprofile.org> [Accessed 17 January 2015].
18. Exam English Ltd. (2014) *IELTS Writing Mark Schemes*. Available from: http://www.examenglish.com/IELTS/IELTS_Writing_MarkSchemes.html [Accessed 17 May 2015].
19. Flowerdew, J. (2009) Use of Signalling Nouns in Learner Corpus. In J. Flowerdew and M. Mahlberg (eds.) *Lexical Cohesion and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing.
20. Fulcher, G. and Davidson, F. (2007) *Language Testing and Assessment: An Advanced Resource Book*. Abingdon and New York: Routledge.
21. Gass, S. and Selinker, L. (1994) *Second Language Acquisition: An introductory course*. Lawrence Erlbaum Associates Publishers.
22. Granger, S. (2002) A Bird's Eye View of Learner Corpus Research. In S. Granger, J. Hung, and S. Petch-Tyson (eds.) *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (pp. 3-33). Amsterdam & Philadelphia: Benjamins.
23. Halliday, M. A. K. (1973) The Functional Basis of Language. In J. Webster (2003) *On Language and Linguistics. M. A. K. Halliday*. London: Continuum.
24. Halliday, M.A.K. (1994). *An Introduction to Functional Grammar*. 2nd ed. London: Arnold.
25. Halliday, M. A. K (2010) *Language Evolving: Some Systemic Functional Reflections on the History of Meaning*. Webcast: plenary session at the 37th International Systemic Functional Congress. University of British Columbia. Available at <https://www.youtube.com/watch?v=nC-blhaIUCk> [Accessed March 28 2015]
26. Halliday, M. A. K. and Hassan, R. (1976) *Cohesion in English*. London: Longman Group Ltd.
27. Halliday, M. A. K. and Matthiessen, C. M. I. M. (2004) *An Introduction to Functional Grammar*. Third Edition. London: Edward Arnold.
28. Hewings, M. (2005) *Advanced Grammar in Use: A self-study reference and practice book for advanced learners of English*. 2nd ed. Cambridge: Cambridge University Press.

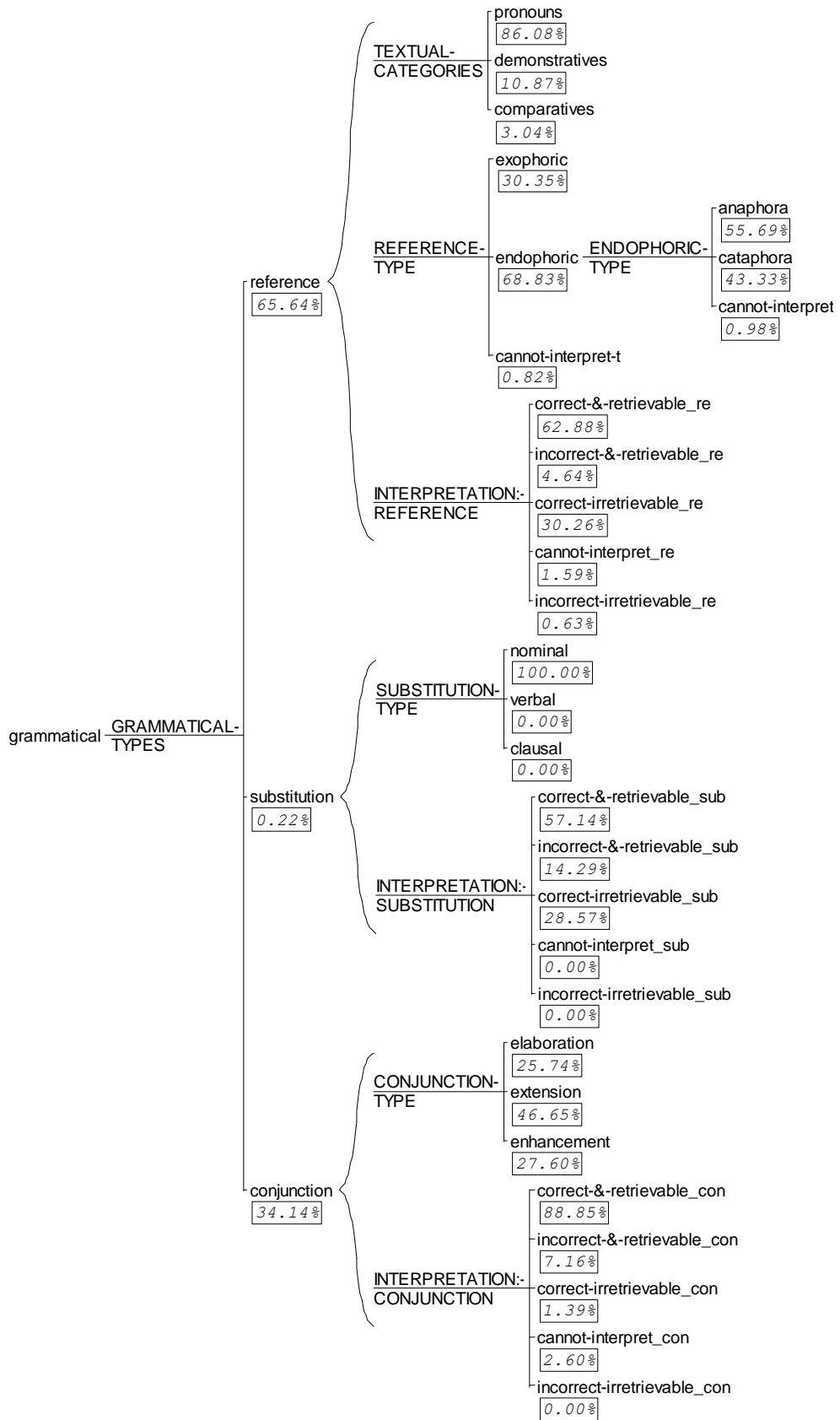
29. James, C. (1998) *Errors in Language Learning and Use: Exploring Error Analysis*. Harlow: Longman Limited.
30. Johnson, B. and Christensen, L. B (2004) *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. 2nd ed. Boston: Allyn and Bacon.
31. Hughes, A. (2003) *Testing for Language Teachers*. Second Edition. Cambridge: Cambridge University Press.
32. Kramiņa, I. (2000) *Linguo-didactic Theories Underlying Multi-Purpose Language Acquisition*. Rīga: The University of Latvia.
33. Leo, K. (2012) Investigating Cohesion and Coherence Discourse Strategies of Chinese Students with Varied Lengths of Residence in Canada. *TESL Canada Journal*, 29 (6): 157-179.
34. Lewis, M. (1986) *The English Verb: An Exploration of Structure and Meaning*. Hove: Language Teaching Publications.
35. Mahlberg, N. (2009) Lexical Cohesion. Corpus Linguistic Theory and Its Application in English Language Teaching. In J. Flowerdew and M. Mahlberg (eds.) *Lexical Cohesion and Corpus Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing.
36. Martin, J. R. (2001) Cohesion and Texture. In D. Schiffrin, D. Tannen, H. E. Hamilton (eds.) *The Handbook of Discourse Analysis*. Malden and Oxford: Blackwell Publishers.
37. Miller, J. (2002) *An Introduction to English Syntax*. Edinburgh: Edinburgh University Press Ltd.
38. McCarthy, M. (1991) *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.
39. Murphy, R. (1994) *English Grammar in Use: A self-study reference and practice book for intermediate students. Second Edition*. Cambridge: Cambridge University Press.
40. Murphy, R. (1998) *Essential Grammar in Use: A self-study reference and practice book for elementary students of English. Second Edition*. Cambridge: Cambridge University Press.
41. Nesselhauf, N. (2004) Learner Corpora and Their Potential For Language Teaching. In J. McH. Sinclair. *How to Use Corpora in Language Teaching* (pp. 125-146). Amsterdam & Philadelphia: Benjamins.
42. O'Donnell, M. (2012a) *UAM CorpusTool* Available at <http://www.wagsoft.com/CorpusTool/download.html> [Accessed 28 March 2015].

43. O'Donnell, M. (2012b) *UAM CorpusTool Version 2.8 User Manual*. Available at <http://www.wagsoft.com/CorpusTool/documentation.html> [Accessed 9 April 2015].
44. O'Donnell, M., Murcia, S, Garcia, R., Molina, C., Rollinson, P., MacDonald, P., Stuart, K., and Boquera, M. (2009) *Exploring the Proficiency of English Learners: The TREACLE Project*. Proceedings of the Fifth Corpus Linguistics, Liverpool. Available at <http://www.wagsoft.com/CorpusTool/publications.html> [Accessed 9 April 2015].
45. Pearson, R. and Pennoch Speck B. (2005) Coherence in English Essays Written by Non-Native Students of Sociology. *Quaderns de Filologia. Estudis Lingüístics*. 10. 261-278.
46. Richards, J.C. and Sampson, G. P. (1974) The Study of Learner English. In Richards, J. C. (ed.) *Error Analysis: Perspectives on Second Language Acquisition* (pp. 3-18). Abingdon: Routledge.
47. Saville-Troike M. (2006) *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.
48. Selinker, L. (1974) Interlanguage. In Richards, J. C. (ed.) *Error Analysis: Perspectives on Second Language Acquisition* (pp. 31-54). Abingdon: Routledge.
49. Sinclair, J. McH. (1996) EAGLES. *Preliminary Recommendations on Corpus Typology*. Available at <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html> [Accessed 2 November 2014].
50. Skujiņa, V. (2005) *Gramatikas kategoriju loma terminu sistēmas veidošanā*. Available at <http://termini.lza.lv/article.php?id=13> [Accessed 22 April 2015].
51. VISC, (n.d.) *Valsts pārbaudes darbu programmas 2014./2015. m.g.* Available at http://visc.gov.lv/vispizglitiba/eksameni/dokumenti/programmas/9/svesval9_2015.pdf [Accessed 15 May 2015].
52. Wekker, H. and Haegeman, L. (1996) [1985] *A Modern Course in English Syntax*. London and New York: Routledge.
53. Zhao, C. (2014) Lexical Cohesion of Sino-British College Students' EAP Writing. *Theory and Practice in Language Studies*. 4 (10): 2123-2128.

Internet Source

1. [Online 1] Available from: <http://www.merlin-platform.eu/> [Accessed 8 June 2014].

Appendix 1: Descriptive Statistics of Grammatical Cohesive Devices



Appendix 2: Comparative Statistics in Low and Medium Performance Groups

Project:	Cohesion_MA					
Unit:	Grammatical-cohesion:grammatical in Task:task					
Counting:	local					
Set1:	performance:low					
Set2:	performance:medium					
	low		medium			
Feature	N	Percent	N	Percent	Chisqu	Signif.
GRAMMATICAL-TYPES	N=555		N=1891			
reference	382	68.83%	1239	65.52%	2.10	
substitution	0	0.00%	5	0.26%	1.47	
conjunction	173	31.17%	647	34.21%	1.78	
TEXTUAL-CATEGORIES	N=382		N=1239			
pronouns	312	81.68%	1077	86.92%	6.56	+++
demonstratives	63	16.49%	118	9.52%	14.29	+++
comparatives	7	1.83%	44	3.55%	2.83	+
REFERENCE-TYPE	N=382		N=1239			
exophoric	91	23.82%	408	32.93%	11.37	+++
endophoric	287	75.13%	818	66.02%	11.17	+++
cannot-interpret-t	4	1.05%	13	1.05%	0.00	
ENDOPHORIC-TYPE	N=287		N=818			
anaphora	164	57.14%	417	50.98%	3.24	+
cataphora	122	42.51%	389	47.56%	2.18	
cannot-interpret	1	0.35%	12	1.47%	2.29	
INTERPRETATION:-REFERENCE	N=382		N=1239			
correct-&-retrievable_re	239	62.57%	753	60.77%	0.39	
incorrect-&-retrievable_re	25	6.54%	61	4.92%	1.53	
correct-irretrievable_re	98	25.65%	400	32.28%	6.03	+++
cannot-interpret_re	18	4.71%	15	1.21%	17.95	+++
incorrect-irretrievable_re	2	0.52%	10	0.81%	0.32	
SUBSTITUTION-TYPE	N=0		N=5			
nominal	0	0.00%	5	100.00%	0.00	
verbal	0	0.00%	0	0.00%	0.00	
clausal	0	0.00%	0	0.00%	0.00	
INTERPRETATION:-SUBSTITUTION	N=0		N=5			
correct-&-retrievable_sub	0	0.00%	2	40.00%	0.00	
incorrect-&-retrievable_sub	0	0.00%	1	20.00%	0.00	
correct-irretrievable_sub	0	0.00%	2	40.00%	0.00	

cannot-interpret_sub	0	0.00%	0	0.00%	0.00			
incorrect-irretrievable_sub	0	0.00%	0	0.00%	0.00			
CONJUNCTION-TYPE	N=173		N=647					
elaboration	45	26.01%	163	25.19%	0.05			
extension	84	48.55%	311	48.07%	0.01			
enhancement	44	25.43%	173	26.74%	0.12			
INTERPRETATION:-CONJUNCTION	N=173		N=647					
correct-&-retrievable_con	132	76.30%	580	89.64%	21.25	+++		
incorrect-&-retrievable_con	20	11.56%	45	6.96%	3.97	++		
correct-irretrievable_con	4	2.31%	11	1.70%	0.28			
cannot-interpret_con	17	9.83%	11	1.70%	27.33	+++		
incorrect-irretrievable_con	0	0.00%	0	0.00%	0.00			

Appendix 3: Comparative Statistics in Medium and High Performance Groups

Project:	Cohesion_MA					
Unit:	Grammatical-cohesion:grammatical in Task:task					
Counting:	local					
Set1:	performance:medium					
Set2:	performance:high					
	medium		high			
Feature	N	Percent	N	Percent	Chisqu	Signif
GRAMMATICAL-TYPES	N=1891		N=697			
reference	1239	65.52%	442	63.41%	0.99	
substitution	5	0.26%	2	0.29%	0.01	
conjunction	647	34.21%	253	36.30%	0.97	
TEXTUAL-CATEGORIES	N=1239		N=442			
pronouns	1077	86.92%	386	87.33%	0.05	
demonstratives	118	9.52%	44	9.95%	0.07	
comparatives	44	3.55%	12	2.71%	0.71	
REFERENCE-TYPE	N=1239		N=442			
exophoric	408	32.93%	125	28.28%	3.25	+
endophoric	818	66.02%	317	71.72%	4.82	++
cannot-interpret-t	13	1.05%	0	0.00%	4.67	++
ENDOPHORIC-TYPE	N=818		N=317			
anaphora	417	50.98%	212	66.88%	23.37	+++
cataphora	389	47.56%	104	32.81%	20.22	+++
cannot-interpret	12	1.47%	1	0.32%	2.68	
INTERPRETATION:-REFERENCE	N=1239		N=442			
correct-&-retrievable_re	753	60.77%	307	69.46%	10.54	+++
incorrect-&-retrievable_re	61	4.92%	10	2.26%	5.70	+++
correct-irretrievable_re	400	32.28%	124	28.05%	2.72	+
cannot-interpret_re	15	1.21%	0	0.00%	5.40	++
incorrect-irretrievable_re	10	0.81%	1	0.23%	1.69	
SUBSTITUTION-TYPE	N=5		N=2			
nominal	5	100.00%	2	100.00%	0.00	
verbal	0	0.00%	0	0.00%	0.00	
clausal	0	0.00%	0	0.00%	0.00	
INTERPRETATION:-SUBSTITUTION	N=5		N=2			
correct-&-retrievable_sub	2	40.00%	2	100.00%	2.10	

incorrect-&-retrievable_sub	1	20.00%	0	0.00%	0.47			
correct-irretrievable_sub	2	40.00%	0	0.00%	1.12			
cannot-interpret_sub	0	0.00%	0	0.00%	0.00			
incorrect-irretrievable_sub	0	0.00%	0	0.00%	0.00			
CONJUNCTION-TYPE	N=647		N=253					
elaboration	163	25.19%	67	26.48%	0.16			
extension	311	48.07%	109	43.08%	1.82			
enhancement	173	26.74%	77	30.43%	1.24			
INTERPRETATION:-CONJUNCTION	N=647		N=253					
correct-&-retrievable_con	580	89.64%	241	95.26%	7.15	+++		
incorrect-&-retrievable_con	45	6.96%	12	4.74%	1.50			
correct-irretrievable_con	11	1.70%	0	0.00%	4.35	++		
cannot-interpret_con	11	1.70%	0	0.00%	4.35	++		
incorrect-irretrievable_con	0	0.00%	0	0.00%	0.00			

Appendix 4: Comparative Statistics in Low and High Performance Groups

Project:	Cohesion_MA					
Unit:	Grammatical-cohesion:grammatical in Task:task					
Counting:	local					
Set1:	performance:low					
Set2:	performance:high					
	low		high			
Feature	N	Percent	N	Percent	Chisqu	Signif.
GRAMMATICAL-TYPES	N=555		N=697			
reference	382	68.83%	442	63.41%	4.03	++
substitution	0	0.00%	2	0.29%	1.60	
conjunction	173	31.17%	253	36.30%	3.62	+
TEXTUAL-CATEGORIES	N=382		N=442			
pronouns	312	81.68%	386	87.33%	5.06	++
demonstratives	63	16.49%	44	9.95%	7.75	+++
comparatives	7	1.83%	12	2.71%	0.71	
REFERENCE-TYPE	N=382		N=442			
exophoric	91	23.82%	125	28.28%	2.11	
endophoric	287	75.13%	317	71.72%	1.22	
cannot-interpret-t	4	1.05%	0	0.00%	4.65	++
ENDOPHORIC-TYPE	N=287		N=317			
anaphora	164	57.14%	212	66.88%	6.07	+++
cataphora	122	42.51%	104	32.81%	6.05	+++
cannot-interpret	1	0.35%	1	0.32%	0.00	
INTERPRETATION:-REFERENCE	N=382		N=442			
correct-&-retrievable_re	239	62.57%	307	69.46%	4.35	++
incorrect-&-retrievable_re	25	6.54%	10	2.26%	9.24	+++
correct-irretrievable_re	98	25.65%	124	28.05%	0.60	
cannot-interpret_re	18	4.71%	0	0.00%	21.29	+++
incorrect-irretrievable_re	2	0.52%	1	0.23%	0.50	
SUBSTITUTION-TYPE	N=0		N=2			
nominal	0	0.00%	2	100.00%	0.00	
verbal	0	0.00%	0	0.00%	0.00	
clausal	0	0.00%	0	0.00%	0.00	
INTERPRETATION:-SUBSTITUTION	N=0		N=2			
correct-&-retrievable_sub	0	0.00%	2	100.00%	0.00	
incorrect-&-retrievable_sub	0	0.00%	0	0.00%	0.00	
correct-irretrievable_sub	0	0.00%	0	0.00%	0.00	
cannot-interpret_sub	0	0.00%	0	0.00%	0.00	

incorrect-irretrievable_sub	0	0.00%	0	0.00%	0.00		
CONJUNCTION-TYPE	N=173		N=253				
elaboration	45	26.01%	67	26.48%	0.01		
extension	84	48.55%	109	43.08%	1.24		
enhancement	44	25.43%	77	30.43%	1.26		
INTERPRETATION:-CONJUNCTION	N=173		N=253				
correct-&-retrievable_con	132	76.30%	241	95.26%	33.89	+++	
incorrect-&-retrievable_con	20	11.56%	12	4.74%	6.87	+++	
correct-irretrievable_con	4	2.31%	0	0.00%	5.91	+++	
cannot-interpret_con	17	9.83%	0	0.00%	25.89	+++	
incorrect-irretrievable_con	0	0.00%	0	0.00%	0.00		

Appendix 5: Statistical Representation of Unsuccessful Cohesive Ties

	low in Task:task-1		low in Task:task-2		medium in Task:task-1		medium in Task:task-2		high in Task:task-1		high in Task:task-2	
FEATURE	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent	N	Percent
INTERPRETATION:-REFERENCE	N=90		N=264		N=396		N=817		N=157		N=149	
correct-&-retrievable_re	23	25.56%	190	71.97%	133	33.59%	609	74.54%	57	36.31%	128	85.91%
incorrect-&-retrievable_re	10	11.11%	14	5.30%	18	4.55%	39	4.77%	7	4.46%	1	0.67%
correct-irretrievable_re	46	51.11%	51	19.32%	232	58.59%	157	19.22%	92	58.60%	20	13.42%
cannot-interpret_re	9	10.00%	9	3.41%	3	0.76%	12	1.47%	0	0.00%	0	0.00%
incorrect-irretrievable_re	2	2.22%	0	0.00%	10	2.53%	0	0.00%	1	0.64%	0	0.00%
INTERPRETATION:-CONJUNCTION	N=42		N=123		N=238		N=392		N=94		N=81	
correct-&-retrievable_con	32	76.19%	92	74.80%	211	88.66%	355	90.56%	89	94.68%	78	96.30%
incorrect-&-retrievable_con	0	0.00%	20	16.26%	15	6.30%	27	6.89%	5	5.32%	3	3.70%
correct-irretrievable_con	3	7.14%	1	0.81%	4	1.68%	7	1.79%	0	0.00%	0	0.00%
cannot-interpret_con	7	16.67%	10	8.13%	8	3.36%	3	0.77%	0	0.00%	0	0.00%
incorrect-irretrievable_con	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%

Dokumentārā lapa

Maģistra darbs „Systemic Functional Analysis of The Year 9 English Language Examination Corpus” (Devītās klases angļu valodas eksāmena korpusa sistēmiski funkcionālā analīze) izstrādāts LU Humanitāro zinātņu fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Līga Ločmele

Rekomendēju darbu aizstāvēšanai

Vadītāja: asociētā profesore Dr. Phil. Vita Kalnbērziņa.....

Recenzents:

Studiju metodiķe: Sintija Zankovska

Darbs iesniegts Anglistikas nodaļā 28. 05. 2015.

Darbu pieņēma:

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

2015. gada..... jūnijā, prot. Nr., vērtējums

Komisijas sekretāre: