



**UNIVERSITY  
OF TURKU**

# **The Differences in Use of Connectors between British L1 Users of English and Swedish L2 Users of English**

Patrik Matinlassi

Master's Thesis

Degree Programme in Language Learning and Teaching, Department of English

School of Languages and Translation Studies

Faculty of Humanities

University of Turku

May 2024

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

**Master's Thesis**

**Degree Programme in Language Learning and Teaching, Department of English**

**Patrik Matinlassi**

**The Differences in Use of Connectors between British L1 Users of English and Swedish L2 Users of English**

**Number of pages:** 40 pages, 10 pages in appendices

This is an MA thesis that studies how L1 English users from the UK and L2 English users from Sweden use connectors in their texts to create cohesion, as well as if there are major differences in the types of connectors used. It will use contrastive rhetoric as a framework to ground the study in. The texts come from two corpora, the British Academic Written English corpus and the Uppsala Student English corpus. Both corpora collected texts from university level students to make the texts as comparable as possible. Due to the British corpus containing a higher number of texts, the numbers were normalized by multiplying the Uppsala text count by 5.33. The texts were then put into the WordSmith software where it was possible to search for the key terms from all the texts.

The results of the study did show differences between the two groups, but it is difficult to draw very strict conclusions where one group would have performed better, due to cohesion being created in many different ways and different subjects requiring different levels of explicit cohesion. In general however, the British corpus featured more varied vocabulary and their Uppsala peers did not use a single connector that was not used at least once by the British group. However, the Swedish group then used certain words more commonly than the British group. Overall in terms of semantic categories, the Swedish group used sequential connectors the most, but the British group dominated with additive, causal and adversative connectors.

**Key words:** Cohesion, connector, contrastive rhetoric, corpus study, CEFR

<b>Table of contents</b>	
<b>1. Introduction</b>	<b>4</b>
<b>2. Theoretical background</b>	<b>6</b>
<b>2.1. Contrastive rhetoric</b>	<b>6</b>
<b>2.2. Cohesion</b>	<b>9</b>
<b>2.3. Connectors</b>	<b>13</b>
<b>2.4. CEFR and Language Competence</b>	<b>18</b>
<b>3. Methodology</b>	<b>23</b>
<b>3.1. Corpus analysis</b>	<b>24</b>
<b>4. Results</b>	<b>27</b>
<b>4.1. General differences in connector use</b>	<b>29</b>
<b>4.2. Semantic categories</b>	<b>31</b>
<b>5. Discussion</b>	<b>34</b>
<b>6. Conclusion</b>	<b>37</b>
<b>References:</b>	<b>39</b>
<b>Appendices</b>	<b>41</b>
Table 1. Types of connectors. ....	14
Table 2. Conjunctions in Moini & Kherkhah.....	17
Table 3. Frequency of conjunctions in subcategories. ....	18
Table 4. Competence in cohesion according to the CEFR. ....	21
Table 5. List of all connector words used in the corpus study. The underlined words have been added to the list provided by Hasan & Halliday (1996, 243).....	24
Table 6. Most commonly used connectors for Uppsala corpus.....	27
Table 7. Most commonly used connectors for British corpus.....	28
Table 8. Connectors only used in British texts.....	30
Image 1. Results for the word <i>despite</i> using the Concord tool in WordSmith. ....	26

## 1. Introduction

Academic skills are very important for writing essays and other texts at higher education levels. Not only do students need to have a lot of theoretical knowledge on subjects they study, but they need to be able to write in a succinct manner with cohesive sentence structures in order to let anyone reading their texts get a good sense of what they want to communicate. Good vocabulary and grammar may not be enough for effective communication if correct sentences do not produce easily comprehensible messages. Education systems of course aim to tackle this, but different environments may have various effects on how students learn these mechanisms that create cohesion in their texts. Additionally, academic proficiency can vary a lot from ordinary communication, as explained by Jim Cummins and his theory (Cummins 2000, 58) on *basic interpersonal communicative skills* (BICS) and *cognitive academic language proficiency* (CALP). He makes a distinction between these types of proficiency as studies have shown that students who have fluent skills in academic writing yet show a lower level of proficiency when it comes to verbal language performance (Cummins 2000, 58). In terms of the current study, CALP is naturally the area of language in the forefront, with interpersonal communication not truly factoring in directly. According to Cummins (1980, 177) CALP includes those areas of language proficiency that relate to literacy skills in both L1 and L2 language use. However, he argues that this proficiency is also separate from the BICS aspects like oral fluency and sociolinguistic competence in one's L1. This connection in CALP between L1 and L2 will also connect to the ideas of contrastive rhetoric I examine in my thesis, as this view then suggests that the development of language faculties used in L1 can have a great influence on L2 academic language production too. It is also noted that this connection may be affected by other factors and is not always that clear (Cummins 1980, 180). It is of course possible that the results for interpersonal communication could differ from the results found in this study, but academic competence is naturally the focus here as I will be analysing text found from academic essays. With the Common European Framework of Reference (CEFR) the different levels of language competence also become more comparable. I will examine the CEFR later in the thesis in its own separate section, as it is a metric that can be used to compare language users.

In my MA thesis, I will study differences in academic essays between L1 users of English in the UK and L2 users of English in Sweden. The Swedish group serves as a group that would be expected to perform well in English even though it is not a native language with the UK group being a comparison point as a native language user group. I will examine them

through the lens of *contrastive rhetoric* and cohesion present in the texts. Halliday and Hasan (8, 1976) describe cohesion as semantic relations of different parts in a text. It goes beyond sentence structure and makes parts of text connect for easier understanding. To focus on a more particular aspect, I will study connectors in particular as a way to achieve cohesion. These are words or phrases that link sentences together and can be categorised into subsets that I will describe later in the theory section. These connectors can for example include *in addition*, *thus* and *firstly*. Contrastive rhetoric has a long history as a field of study, but it also has more room to develop. I will explore some of the unexplored space in my thesis, utilising the framework for cohesion and connectors in particular as a tool for cohesion. The primary material in my thesis consists of two corpora: the Uppsala Student English (USE) corpus and the British Academic Written English (BAWE). Both of these corpora include samples of English texts used by university students but for students with different language backgrounds, either second language or native language. In particular, my research questions are the following: Firstly, what kinds of differences can be found between the L1 users of English from the UK and L2 users of English from Sweden when it comes to using connectors to create cohesion in text? This will be measured by relative word count of these connectors in all the texts. Secondly, are there any particular differences among the semantic categories both within one group and between the groups? I will try to examine if there are trends among the groups that go beyond individual variance, potential differences not fully explained by the language background. I expect the groups to be more similar within themselves rather than getting results that show nothing shared among members of one group in the beginning, before conducting the research itself. I will first go over the theoretical background of my thesis which will further contain subsections for contrastive rhetoric, cohesion and the Common European Framework of Reference. This will then be followed by methodology. The methodology section will contain a thorough description of the corpus analysis. Once that is done, I will go over the results in the discussion section and explore what kinds of conclusions can be drawn from them. Finally, the conclusion will summarise everything.

## **2. Theoretical background**

This section will detail the theoretical background I use for my thesis. I will first explain the field of contrastive rhetoric in terms of how it is useful for my study. This will then be followed by theory on cohesion, which will play an important part in my analysis here. The cohesion section is further divided into sub-sections discussing connectors in particular as well as the Common European Framework of Reference (CEFR) to further focus on what I am examining in my thesis. The CEFR is important for pointing towards reasons why people may want to take interest in cohesion and related phenomena when it comes the second language teaching in general, not just in regard to my study. The section is then concluded in a short summary of what I have discussed.

### **2.1. Contrastive rhetoric**

Firstly, it is important to define what contrastive rhetoric means as a field of study. Ulla Connor speaks of the field as follows: “Contrastive rhetoric is an area of research in second language acquisition that identifies problems in composition encountered by second language writers and, by referring to the rhetorical strategies of the first language, attempts to explain them” (Connor 1996, 5). Contrastive rhetoric was first introduced by Robert Kaplan in 1966. Connor noted that the way we use words differs between different languages even beyond just lexical and grammatical dimensions, which she noted from the studies of Kaplan (1966, 2). He goes on to note that just because a student can write an essay in their own native language, they would be able to do so in a second language (1966, 4). While it is true that they will be able to write in some sense of the word, it is also notable that they might not be able to write an essay at a level of competence that native speakers would, due to cultural differences and the lack of knowledge around conventions that the native users of a language learn implicitly. Kaplan takes note that English speakers are used to certain types of paragraph structures, inductive and deductive, which an English text should be expected to be written in typically (Kaplan 1966, 5). However, Arabic opts for different structures, a series of parallel constructions. There’s no reason why these languages could not use any of these paragraph structures if you only look at vocabulary and grammar, yet writing in a non-standard style would certainly feel strange to a native speaker. Therefore, there are aspects other than traditional language rules in play, often learned more implicitly where the native speaker might not be able to fully explain these things, yet they hold the knowledge and can use it smoothly in their own writing. In terms of more experienced users of a second language, it is

possible that they may retain some of this implicit knowledge as well, however I will see if there is anything to showcase that in the results of my study.

Naturally, as the field was only in its early stages during that time, it has gained new perspectives and ideas since then to better explain new findings and to get more reliable results. Connor (2002) describes contrastive rhetoric as follows: “Contrastive rhetoric examines differences and similarities in ESL and EFL writing across languages and cultures as well as across such different contexts as education and commerce”. She continues to explain that Kaplan’s original idea was based on the notion that different language cultures emphasise different writing norms and these differences manifest in second language writing as transfer. As my area of interest is the difference between L1 Swedish and L1 English users of English in their texts, with both the UK and Sweden being located in the northern parts of Europe, they are not necessarily geographically very far from each other and even culturally there are many differences with the areas. Still, I wonder if the Nordic countries have certain cultural differences to the British culture, which may manifest in the writings between the two groups. These differences could relate to aspects like the density of connectors in text. Geographically they are not too distant but of course they do not share land borders so direct exchange does not happen easily. Additionally, for people who can communicate decently in English these geographic lines may not matter too much with online communication.

Contrastive rhetoric builds upon other areas of study, taking influence from other theories to form a starting point upon which to form new ideas. Connor (1996, 10) proposes a list of theories that influence contrastive rhetoric. Firstly, theory of rhetoric is one relevant theory in terms of this subject, naturally. What it offers here is inspection into ways in which writers try to affect readers with their writings. Theory of applied linguistics is useful for establishing ideas related to language acquisition, as well as a point of view focused on applied problems. This is very useful in the sense of examining these very concrete ways differences can manifest between L1 and L2 users of English when it comes to the contrastive rhetoric approach. Linguistic relativity, along with the Sapir-Whorf hypothesis is another influence on contrastive rhetoric, with the focus being that one’s own language has an influence on thought processes and therefore different first languages can make a difference in the way ideas are processed between people. This idea has also received criticism from other researchers on the topic. Ying (2000, 260) asserts that the Sapir-Whorf hypothesis does not work with contrastive rhetoric, as Kaplan’s idea on the importance of culture conflicts with the deterministic ideas of the Sapir-Whorf hypothesis. In addition to these, the theory of text linguistics is utilised, and it gives a framework that can be used to analyse aspects of the texts

such as cohesion, which is what my own study will also have a focus on. Connor also lists theory of discourse types and genres as relevant here, however the texts that I personally examine should be similar to each other and thus there is little need to compare text types under this framework. Still, identifying typical features of the academic texts should be useful. Another theory listed is theory of literacy, which has the purpose of explaining how different writing cultures may put emphasis on different writing styles, which can then naturally affect products of target language. Theory of translation is finally suggested as a relevant field, as it shares a lot with contrastive rhetoric. Translation theory is concerned with ideas like how audiences receive translated texts, and these ideas can be adopted by contrastive rhetoric research as well (Connor 1996, 122).

One thing to keep in mind when discussing typical writing conventions and styles in this manner is, of course, that people from a single culture do not all write in the same manner. Concepts such as “Swedish academic culture” are far too large to lead to a uniform style of writing and thus, there are naturally differences within the group. Also notably, different genres may call for different conventions as well. It is very important not to generalise these ideas to every writer of a given language or culture. It may be hard to show a clear relationship with the L1 of the speaker and their written production being a certain way. This comes back to the idea of genres having varying writing norms. Still, there are certain aspects of language that seem to be typical for one group. Narita, Sato and Sugiura (2004) studied Japanese speakers of English as a foreign language and found out that while Japanese people used some of the logical connectors in similar ways to native speakers, “the Japanese EFL learners significantly overuse the connectors in sentence-initial position” (Narita et al. 2004, 1173). Faruk (2006, 1358) says that similar overuse of certain logical connectors happens in other languages, potentially due to differences in French and English in how the languages tend to use connectors compared to other cohesive devices. These cohesive devices refer to tools that are used to achieve cohesion in a text, such as connectors. Connectors can be words such as *briefly* or longer word clusters like *at this moment*. I will go into further detail on connectors when discussing cohesion.

Yet another criticism given to contrastive rhetoric is ethnocentrism (Connor 1996, 162), an idea which manifests in the belief that Anglo-American writing has features of being logical and direct, compared to other cultures that are then implied to be illogical to different extents. This however shows a bias from Anglo-Americans mostly, as representatives of other cultures may not agree to this standard of logic. Another similar idea that Connor then presents is the fact that Finnish writers tend to use less transition in texts, which is attributed

to the idea that Finnish writers respect their readers in interpreting the correct meanings even with less explicit transitions (Connor 1996, 163). Similarly to non-Anglo-Americans finding their lack of logic potentially hurtful, these Anglo-American writers may find it offensive that they as writers apparently do not respect their readers. Such concepts are important to take note of and a way to approach this is to avoid harsh stereotypes among one writing culture.

## **2.2. Cohesion**

An aspect of language I will focus on is cohesion. Tanskanen (2006, 7) says “cohesion refers to the grammatical and lexical elements on the surface of a text which can form connections between parts of the text”. A related concept is brought up with coherence, which “resides not in the text, but is rather the outcome of a dialogue between the text and its listener or reader” (Tanskanen 2006, 7). These two concepts are not in opposition, but they are somewhat distinct, albeit closely related concepts I decided to focus on cohesion because I find it interesting and previous research has noted that additional studies on cohesion and connectors are needed (Crossley and McNamara 2012, 131). There are conflicting results in existing research and additional studies can help learn more to settle the conflicting ideas found. Additionally, Crossley and McNamara (2012, 115) note that there’s a gap in research regarding “underlying meaning and intentions of the discourse” in L2 language studies. With the help of contrastive rhetoric, I will be examining cohesion in the texts produced by the two groups. For this purpose, the meaning of cohesion must of course be defined. To be more precise, the cohesion I am interested in can be categorised as conjunction. According to Halliday and Hasan (1976, 6), conjunction can be placed between the categories of lexical and grammatical cohesion, as it has elements of grammar, but it also resembles lexical methods of cohesion.

For further clarity, Flowerdew and Mahlberg (2009, 1) give the following definition for lexical cohesion: “Lexical cohesion is about meaning in text. It concerns the way in which lexical items relate to each other and to other cohesive devices so that textual continuity is created” (2009, 1). In other words, it is about how the words work together in a sentence and goes beyond the word level. It is separate from grammatical correctness and vocabulary proficiency. This can be illustrated in a few examples: Flowerdew and Mahlberg (2009, 1) borrow from Hasan & Halliday (1976), material I also use in this thesis. Hasan & Halliday describe that there are two primary ways to create cohesion in text: reiteration and collocation. The former refers to a lexical item being used multiple times within one text either in the

same form or with a synonymous term. Most simply, you can look at a construction like *John has a ball. The ball is red.* Repeating the word *ball* creates a sense that the sentences are related and speak of a single main topic. The latter option, collocation, is achieved by utilising words that are semantically related and thus achieving a sense of a coherent whole. You can for example say *There are many kinds of life in a forest. The environment creates protection for fungi, smaller plants and animals.* There is no direct repetition, but the words are related to a single topic, making it easier to understand conceptually.

Cohesion can be expressed with sentence structure as well as words that are used mainly to create higher levels of cohesion. In addition to lexical cohesion, there is grammatical cohesion. Grammatical cohesion is expressed via grammar. It can be divided into categories of 1) reference, 2) substitution and 3) ellipsis (Halliday and Hasan 1976, 5). Reference is when a word like *there* or *that* is referring to another known term, either by earlier linguistic context or social cues for example (Hasan & Halliday 1976, 30). They are specific words with low inherent meaning, relying on other context to carry on meaning. Substitution is replacing words with other similar words in the same category (Hasan & Halliday 1976, 88), for example in a series of sentences such as *I like these stones. The pink ones are especially pretty.* Here *ones* is substituting for *stones* as a word that can take on a similar structure and is clearly referring back to the earlier sentence. It can also happen with verbs, substituting a verb with *does* in a later sentence for example, as well as even full clauses. Finally, ellipsis is a type of substitution, where you substitute something with nothing (Hasan & Halliday 1976, 142). One can consider a pair of sentences like *Joan had three apples. Diana had two.* Here the latter sentence omits the word *apple* but it is nevertheless assumed by the context surrounding it. Hasan & Halliday (1976, 88) note that the division into the former subcategories is not entirely strict and the categories do overlap to some extent.

Cohesion is listed in the Common European Framework of Reference (CEFR) as one of the aspects that form pragmatic competence, which in turn is a part of communicative language competence (Council of Europe 2001, 13). Considering these are widely recognised criteria used for measuring language competence, it is important to investigate the potential ways lexical cohesion can be expressed differently between different language groups.

Crossley and McNamara (2012) studied L2 writing from Hong Kong high school level students, with a sample size of 1200 essays, 200 for each possible passing grade given to the students. Their interest was in cohesion and sophistication within English L2 texts. The essays were of approximately 500 words in length, written within 1 hour and 15 minutes. From all of

the essays, only 514 were selected for their study, due to having similar word counts so they can be compared in a relevant manner. They also note that essay length is often correlated with scores so the validity of the study would be at risk if the text length varied too greatly. For the variables they chose three different aspects to monitor: syntactic complexity, word frequency and finally hypernymy and polysemy indices (Crossley & McNamara 2012, 121). The first one is measured by how many words are used before the main verb in a sentence, mean number of high-level constituents per word in a text, and finally an index that checks consistency of syntactic structures. This way of viewing syntactic complexity from multiple different angles is used to gain a wide understanding of the texts. Word frequency is then another measurement, checking how often words used in the essays are generally used in the English language, utilising a large corpus. The researchers note that more skilled writers are better at utilising words that are less frequently used. Then lastly the hypernymy and polysemy indices measure the use of words with several potential meanings and potential hierarchical structures a word is included in. In terms of my own thesis, I will mostly be examining the word frequency over the other aspects, as I see that working most effectively with connectors. As they are their own separate words, searching for the number of times these words appear works well, compared to other structures that may require more qualitative analysis because searchable strings of words are absent.

Surprisingly, the study mentioned in the previous paragraphs found out that higher language proficiency did not always predict better use of cohesive devices, which are different methods that can be used to achieve cohesion in a text, for example connectors or repetition of key terms. Texts written by more proficient writers displayed higher lexical diversity, which is understandable due to their more developed proficiency, however this can decrease cohesion when key words are not repeated to create connections between sentences and paragraphs even (Crossley & McNamara 2012, 130). Furthermore, it is reported that essays judged to have higher proficiency of language often provided the reader with less context and fewer connectors for cohesion. Crossley and McNamara suggest that the less cohesive texts produced by those who write at a more proficient level are a result of assumptions on the writer's part that make them assume their readers have high language proficiency and appropriate knowledge on the topic as well. Such a reader would need less help from cohesive devices as a familiar topic generally needs less help from text cohesion and naturally lower proficiency writers and readers then use cohesion to connect ideas to each other when their knowledge might not be sufficient to understand these ideas if it were not for

use of cohesive devices. The writers also note this contradicts previous research on the field (Connor 1990, Jin 2001), so it is not wise to conclude too much from this alone.

Jin (2001) studied Chinese users of English. With Chinese being much more distant to English both linguistically and culturally, the languages are quite different not only on a grammatical and syntactical level, but Jin (2001, 2) explains that the structuring of texts in Chinese writing is quite different. He also notes (2001, 3) that in interviews, the Chinese students noted that their Chinese writing may use explicit connective devices less, rather relying on more implicit, logical means to create cohesion. His study was asking if there is any variability in overall cohesiveness of Chinese ESL graduate students' writing across genres and proficiency levels, if they exhibit central tendencies or common features in terms of coding textual cohesion, and finally if there's a relationship between Chinese ESL student's linguistic ability and writing ability (2001, 4). These questions were examined with quantitative research, examining various genres and proficiency levels among the Chinese ESL students, with Halliday and Hasan's framework (1976) used. By the end, only six students were chosen, for a total of 18 texts they wrote during a period of one academic year (2001, 8). As alluded to in the previous paragraph, the study did find differences between different proficiency levels, with the advanced group showing a higher use of personal pronouns to help with cohesion, as well as a significantly higher use of temporal conjunctions as well as a milder difference when it comes to additive and adversative conjunctions (2001, 20). The advanced writers also used more synonyms than their counterparts, a more sophisticated method of creating lexical cohesion than word repetition (2001, 22). Both groups overall still used a great amount of lexical repetition, which was attributed to that being typical in Chinese writing (2001, 27). Jin notes that it is commonly argued that proficiency in written and spoken language are not linked directly but points out that the results of this study indicate that there is a more direct link, however it is not an absolute relationship there either. Still, there was added cohesiveness in the writings of those more proficient in English (2001, 28). The group of six students did demonstrate these trends quite well between the advanced and intermediate groups, but the sample size of six people is naturally quite small, even if that allowed a deeper analysis that would get much more complicated with a larger group.

### 2.3. Connectors

In order to measure differences in text cohesion between Swedish and English backgrounds for users of English, there has to be a framework that can be used for operationalising the phenomenon. I have decided to use connectors. These connectors are a way to form cohesion in the text and they are particularly important for understanding texts that contain new information for the reader (Goldman & Murray 1992, 504). Connectors help explain relationships between clauses and sentences and as such, mastery of them is useful for effective writing. They are a type of cohesive device used in texts to make it flow better. There are other types of cohesive devices too, like conjunctions and linking adverbials. Trebits explains the following: “Conjunctions are important cohesive devices and the mastery of their use has a huge impact on students’ textual competence” (Trebits 2009, 200). Some words can serve the role of a conjunction or a connector, with conjunctions linking clauses and connectors typically linking sentences together (Yale Center for Teaching and Learning, 1). Thus, it is important to study the use of connectors by students of English. It is then good to define connectors so that it is clear what the focus here is. According to Goldman and Murray (1992, 504): “Connectors are a type of text device the primary function of which is to link adjacent text propositions in such a way as to maintain local coherence”. Connectors can be words such as *however*, *therefore*, *because* and *in addition*. They establish ways in which ideas are related to one another and they are used between sentences. The Yale Center for Teaching and Learning (3) notes differences for when connectors and conjunctions should be used. Connectors work when combining sentences that both provide new information, where as conjunctions are used when information in the first part is assumed information. *If it’s sunny tomorrow, I will go out shopping* has the *if* part suggest there is reason to believe it could be sunny. Comparatively a sentence pair like *It could be sunny tomorrow. If so, I will go shopping* has two separate sentences linked by a connector, which goes between the two sentences directly as opposed to the conjunction. Connectors also allow a more precise degree of separation between the ideas, when conjunctions are more used to connect ideas on an equal level. However, as both of these types of cohesive devices achieve similar purpose and will get noted by the software when examining the corpora, I will not discriminate between conjunctions and connectors in this way, adding any kind of use like this together.

Goldman and Murray note: “Research and teacher reports indicate that mastering appropriate use of connectors is extremely difficult” (1992, 505). In their study, they had participants choose between different types of connectors, judging by which one seemed the

most appropriate to put inside a sentence where the original connector word had been removed. The participants had different levels of proficiency in English. The four types of connectors used in this study were additive (additional information without a strong relationship between the ideas; in addition, moreover), causal (expresses causality; as a result, therefore), adversative (expresses contrast; however, in contrast) and sequential (conveying logical relationships and indicating temporal sequence or order of things; first, second, finally). These are listed in Table 1 below for added clarity.

Table 1. Types of connectors.

Connector	Explanation	Examples
Additive	Add information, weak relation between items	<i>In addition, moreover</i>
Causal	Express causality between items	<i>As a result, therefore</i>
Adversative	Contrast between items	<i>However, in contrast</i>
Sequential	Temporal order between items	<i>Finally, in summary, subsequently</i>

The results seem to indicate that those more proficient in English performed better in terms of choosing the correct connectors, however the different levels of proficiency did not seem to change which type of connectors the participants had most issues with. So, while the proficient, native users of English had more correct answers, the same types of connectors still had the most issues. Goldman and Murray speculate that additive and causal connectives were correctly identified and added back into the text due to being more common in text and therefore the participants had enough exposure to them prior to the test so they could place them correctly (1992, 512).

Trebits (2009) also studies cohesion and connectors in her study. Her focus was use of conjunctions in English texts for the European Union and her study is also a corpus study. She first used the British National Corpus (BNC) to find which conjunctions appeared most frequently in British texts, to gain a point of comparison for the studied data. In this study, both conjunctions and sentence or linking adverbials are treated as one category and both are counted (Trebits 2009, 202). This corpus includes over 89 million words in it. The mainly studied corpus is the Corpus of EU English and the total number of words in her material adds up to a little more than 200 000 words. The large difference in the size of the corpora means

that the numbers had to be adjusted to something comparable. Trebits chose to count words per million. For the analysis, Trebits used WordSmith Tools software, as well as The Lexical Frequency and Range Programs (2009, 205). The counting of these words is however not entirely straightforward, and one has to take steps to make sure the numbers are not skewed. There are multiple issues that need to be accounted for. Firstly, some of the conjunctions that consist of multiple words include words that appear in other conjunctions too. To make sure this does not cause problems, the instances of other conjunctions must be subtracted from the total number of the word count. Trebits notes (2009, 205) that the number of “for example” was subtracted from the count of “for” as an example of this. Secondly, there are words that are used in multiple contexts and a word like “since” can be used with several meanings despite it having the same form in these instances and this needs to be considered during the count. The results show that certain connectors have similar levels of use in both corpora (*and* was the most common in both), however many of them were also used in much different quantities (*nor* appears many more times in the BNC). When examining the subordinating conjunctions, it was noted that further variation appears between the corpora. Certain causal conjunctions that are rare in the British corpus were surprisingly frequent in the EU texts, for example “with a view to” and “so as to”. Adversative conjunctions were on the other hand more common in the BNC, including words like *although*, *though*, *if* and *unless*. There were even many conjunctions that frequently appear in the BNC but did not appear even once in the EU texts (*in spite of*, *as though*, *by comparison*, *next* and *in conclusion*). It is then concluded that these results indicate that study material should focus on these rarely or never occurring words when teaching students who have an interest in English as a language in the European Union (Trebits 2009, 207).

Faruk Syed Md Golam (2016) discusses cohesion and the use of cohesion in texts produced by Bangladeshi Non-Native Speakers (BNNS). He notes that English teachers in Bangladesh complain how their students do not have enough cohesion in their texts. The study then aims to find out which methods the students use the most for creating cohesion and if they indeed underuse or overuse said methods. For the sake of comparison, the study also checked the use of cohesive methods in Swedish, Japanese, French and Chinese non-native speakers of English. The material for these was gathered from previous studies, which drew from International Corpus of Learner English. All of the aforementioned sub-corpora, including the Bangladeshi essays were further compared to native essays in the Louvain Corpus of Native Essay Writing (LOCNESS). Faruk explains (2016, 1357) that connectors are optional tools in the English language that can help cohesion but if used inappropriately

can cause confusion. The hypothesis of the study is that non-native users of English tend to overuse connectors, especially in sentence-initial positions. It is also noted that this hypothesis is not supported by all findings. The actual research question is if advanced users of English in Bangladesh use connectors in a similar manner to the other groups listed earlier. The results of this study showed that the Bangladeshi group showed signs of using connectors which were similar to commonly used English connectors in Bangladesh (2016, 1361). This shows that there is some credence to the idea that the native language affects second language even if it might not be a conscious process. The study also shows that the Swedish group involved showed high usage of connectors in a few categories (2016, 1391), like the connector *for example*, where they not only get used a lot but more than native speakers of English. The resultive connector *therefore* is somewhat overused in Swedish compared to native speakers, whereas *so* was overused in some languages but comparing English and Swedish produced very similar results. The results discussed here are a little limited, but overall there were some similarities between Swedish and native users of English here in this study, which could possibly be noted in my own study too.

Moini and Kheirkhah compare the use of cohesive devices in novels written for adults and novels written for children. For their study, they chose *Danny, the Champion of the World* by Roald Dahl as the children's book, *Coming Up for Air* by George Orwell for the adult novel. Their study is not limited to only connectors and includes other categories, such as *simple repetition*, which is just repeating a word to create a sense of cohesion within a text. I pay minimal attention to this as it is not related to my study here. In terms of conjunctions, they utilise a model, which divides conjunctions into elaboration, extension and enhancement. Elaboration includes subcategories of either restating or clarifying previous ideas. Extension includes the categories of additive and adversative conjunctions which I also use in my study. It also includes variative conjunctions which is used to present alternative ideas. Enhancement includes the categories of temporal, causal, comparative, conditional, concessive, and matter, the first two being categories in the present study as well. The enhancement category overall includes items that "are used to show how the elements in text relate to each other in terms of cause and effect, time and space or what has been discussed elsewhere in the text" (Moini & Kheirkhah 2016, 16). For example, temporal conjunctions here can include phrases like *at the moment* or *in the early morning*. Causal conjunctions then include phrases like *therefore* and *hence* that express causality. Comparative conjunctions compare and contrast phrases and sentences with words like *likewise* or *alternatively*. Conditional conjunctions express conditionality, for example with *unless* or *if*. Concessive conjunctions express somehow

surprising ideas with words like *although* or *despite*. Matter conjunctions are not features in this study. The results are presented as the frequency of these words in a table (Moini & Kheirkhah 2016, 18), which I will present below in Table 2.

Table 2. Conjunctions in Moini & Kherkhah

	F (child novel)	NF (child novel)	Percentage (child novel)	F (adult)	NF (adult)	Percentage (adult)
Elaboration	23	2	6%	75	5	12%
Extension	98	10	25%	235	16	37%
Enhancement	265	27	69%	314	22	51%
Total	386	39	100%	624	43	100%

F = Frequency

NF = Normalised frequency

The table here describes the distribution of conjunctions in the texts. While the frequencies of these words differ between the two books, the categories are in the same order of frequency overall with elaboration being the rarest and enhancement being the most popular category. The columns showing the frequency and normalised frequency of conjunctions count how many total uses of these conjunctions appear in the texts. The percentage then shows how much the percentage is compared to the other conjunctions. Because the texts are of differing lengths, the normalised numbers are to make them more comparable between each other. Table 3 below will describe how these can be further divided into their subcategories. I will only add the categories relevant to my study to keep it simple, with temporal conjunctions matching with sequential connectors in the present study. Some of the framework used in this study is different from the one I have chosen from Hasan & Halliday, which was more comprehensive. This will lead to the percentage totals not adding up to 100 percent overall, but otherwise the numbers are still correct.

Table 3. Frequency of conjunctions in subcategories.

	F (child)	NF	Percentage	F (adult)	NF	Percentage
Additive	50	5	13%	145	10	23%
Adversative	41	4	11%	67	5	11%
Temporal	185	19	48%	213	15	34%
Causal	42	4	11%	36	2.5	6%

Almost half of all the conjunctions in the children's novel are temporal, indicating that the writing is more focused on describing a clear order of events. The adult book has slightly less focus on this but it is still the largest single category so there seem to be certain trends common between both books even if there is some variance. The authors of the study note (2016, 20) that the differences between the novels can most likely be explained by the fact that the children's story is more dynamic and includes more dialogue compared to the novel for adults, but even so the differences do not change the overall direction of the results. The study goes into more detail in the subcategories and their differences, but I will keep things more simple as I am more interested in the larger categories in the present study. One weakness of this study on the two novels is that the sample size is only concerned with two books and no matter how good the quality of those books is, it is not a very representative sample. Still, both of them still seem to point to a similar direction. When it comes to academic English, on average it should match more to the adult novels, potentially even higher absolute use of these words because the language of academic contexts can be very formal and focused on following these conventions, where novels can be more free form or varied.

#### **2.4. CEFR and Language Competence**

In this sub-section, I will have a look at what the CEFR is in order to further support cohesion being an important aspect of language assessment and why this research therefore is important for the field. It is a framework designed to make language teaching learning and assessment comparable between different European countries. It sets standards and expectations of language users within Europe and thus it stands to reason to examine what is written about writing there. In my thesis I will study English users from Sweden and the UK, so they would both be European countries where this framework is used to measure language competence. As noted, cohesion is part of the category discourse competence. This is part of Canale and

Swain's theory on communicative competence (Canale & Swain 1980). They bring up how these words have been used typically. The weaker definition of competence, according to Canale and Swain, has referred to "knowledge of grammar and of other aspects of language" (1980, 3). It is contrasted with performance, which is putting this competence to use in a practical situation. In terms of studying people, it can of course be difficult to distinguish between theoretical knowledge and the ability to apply that knowledge and they are also quite directly linked together. They also bring up the stronger definition of competence, which is a linguistic system internalised by an ideal native speaker, and in reference to this, performance includes psychological aspects of being able to produce speech based on this system, as well as memory limitations and other issues that may affect one's ability to use language according to one's skill level. However, with all of this in mind, Canale and Swain propose that communicative competence includes both grammatical competence and socio-linguistic competence (Canale & Swain 1980, 6). It is still also distinguished from communicative performance, which is how this knowledge is practically used once mastered. They further maintain that when doing second language teaching should incorporate both grammar and communication even when concerned with communication first. With further development of their theory, communicative competence is divided into three categories: strategic competence, psychophysiological competence and lastly language competence. The last one is further divided into four different competencies, which are grammatical, discourse, illocutionary and socio-linguistic competence (Canale 1987, 68). Out of these, discourse competence is the one that matters for cohesion and for the framework of my thesis.

The CEFR (Council of Europe 2001, 123) refers to discourse competence as "the ability of a user/learner to arrange sentences in sequence so as to produce coherent stretches of language". This is the level of language I am interested in with my data. There are other aspects to discourse competence beyond cohesion. I will briefly discuss these and how they achieve higher discourse competence to contrast them with cohesion, which is what this study is about more directly. They are measured by the ability to order sentences in terms of the following:

1. Topic/focus: When formulating a sentence, one starts with a more general topic to familiarise the topic with the receiver, then going into a more specific idea and making the message easier to understand that way.
2. Given/new information: This is type of sentence structure where one starts a sentence with given information that the receiver of the message will already be aware of, to

make it easier to parse. Following up with new information will be easier to connect to the known information.

3. Natural sequencing: Temporal sequencing is one way the language user can sequence a message so that it feels natural to parse. When telling a story, it is easier to follow if things are explained in the order they happened, compared to any random order.
4. Cause and effect: Similarly to the previous point, it is easier to follow an explanation where the cause is established first, then the effect feels more natural.
5. Coherence and cohesion: Connecting sentences in a message with cohesive devices can help make them understand. Connectors, conjunctions and other cohesive devices can be used so that sentences and clauses form a cohesive whole that is easier to interpret as the receiver.
6. Logical ordering: Overall it is important to take note of what aspects of the message one is trying to communicate are established first.
7. Style and register: When communicating with other language users, it is good to note when to use more natural registers and when more specified words or slang may be appropriate.
8. Rhetorical effectiveness: Often it is not only the pure content of words that is important, but that they are expressed in a way that will appeal to the recipient.
9. The co-operative principle: People make certain assumptions when communicating to make sure all parties go as smoothly as possible. This is achieved by making your contributions have good quality, i.e. your contributions are true. Also, you should have proper quantity in your contribution: state everything you need to but not more than that. Relevance is another important aspect, try to talk in a way that contributes to the topic at hand. Finally, one should take note of manner, so that the message is laid out clearly, briefly and without unnecessary obscurity.

The CEFR notes then (2001, 123) that not meeting these criteria should only happen if trying to achieve some special purpose by doing so, not because one lacks the ability to follow them.

I will not delve into these further as they are irrelevant for the purposes of my study here.

They could certainly work as topics for further research and have their own merits, however connectors seem more interesting to me, because methods of cohesion feel more concrete and present in text on a clearer level than the other aspects I listed. These additional aspects would also distract from the main focus of the thesis. However, I found that they are good to bring

up here, as they will also contribute to the clarity of any message that is laid out, however for the scope of this thesis they will deviate too much from my aims.

The CEFR has a table describing how language users of varying proficiency use the language in terms of proficiency, (Common European Framework of Reference, 2001, 125).

Table 4. Competence in cohesion according to the CEFR.

	Coherence and cohesion
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
B2	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.
A2	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'.

As shown by Table 4 above, assessments for lower levels of proficiency expect simple methods of cohesion with easy connectors already and when higher levels of proficiency are in question, there is a lot more to be expected than just connectors of course, but additionally they include more complex connectors. When this is contrasted with the study conducted by Crossley and McNamara (2012) however, it should be noted that the ability to use these

cohesive methods does not mean that the language user will employ them in their texts, if they find them unnecessary. Notably those using English at the university level should in general be at the C2 level but of course there are various factors going into measuring this and if the students are equally good at all types of English (written, spoken etc.) or if they need to use English for their goals even if they still have issues with certain aspects of the language. Even the lower levels of proficiency here include limited use of these cohesive devices as shown above, with B2 level of proficiency already requiring a variety of linking words and certain levels of cohesion even if the proficiency is below C1 and C2 levels.

All in all, the framework of contrastive rhetoric is useful as I have presented earlier in this section. Cohesion has mixed results in how it is expressed by writers, with a potential divide between ability to create cohesion in text and the decision not to use it to keep language shorter and more efficient with a possible cost of making parts of your message less easy to understand for people less familiar with a given topic. The knowledge of connectors and cohesive devices is also important as evident by the CEFR criteria of language assessment. In the following section of the thesis, I will look more into the practical applications of these theoretical ideas presented in this section. I believe it will tie in the connections of the contrastive rhetoric and the cohesion elements achieved by connectors.

### 3. Methodology

In this section I will go over the corpora I chose for my thesis as well as the methods I will use to analyse them in ways relevant to my thesis. As stated before, the research questions here are the following: How do the two different groups of English speakers use connectors to create cohesion in the text compared to each other? Secondly, how do the used connectors get used in regard to their semantic categories? The data in the corpora is anonymised, but given code names so all texts produced by any one person can be identified to have been produced by the same person. For the Swedish speakers, I will be using the Uppsala Student English corpus. This corpus includes samples of L1 Swedish users of English at an appropriate level for my thesis, i.e. university students. They would be assumed to have a CEFR level of C2, very proficient users of English. One potential issue here is that students with lower proficiency could in fact have more native-like methods for cohesion that have been replaced by styles more akin to the British users of English. And as noted above, some students may be in a way get into using English in their studies to achieve goals they have even if their competence level is not as high as it ideally should be at the university level. I cannot say for sure this is the case but it is something to keep in mind. It has 1,221,265 words in total. As for the L1 corpus, I am using the British Academic Written English (BAWE) corpus. It contains texts by different levels of students from Great Britain and it has 6,506,995 words in total. The texts were collected from three universities: Oxford Brookes, Reading and Warwick. This corpus does mention that English is not the first language for everyone involved, but the texts are collected from British universities, so I believe they are close enough to that with room for a small margin of error. The same is true for the Uppsala students but in smaller numbers. Both corpora were chosen on the basis of containing usage of English by students, so that they are as similar as possible to each other in terms of everything but the country or university of origin. The corpora will be studied with quantitative methods, counting instances of each applicable word but with relative word count to take into account the fact that the corpora are of very different sizes. This way it is possible to find how many times each word appears comparatively. These results can then be shown as percentages. First I will describe the texts in both corpora, who the writers are and how many texts there were in both overall. The results will then be normalised so they can be compared to each other more clearly. By dividing the number of words in the British corpus by the number of words in the Uppsala corpus, you get 5.33, which is used to multiply each instance of the Uppsala word count. After that, I will describe the method of analysis in the following sub-section.

### 3.1. Corpus analysis

For the corpus analysis, I have gathered all the text samples found in my two corpora. I have chosen appropriate connector words, divided into four categories as described in the previous section, categorised by Halliday and Hasan (1976, 242-243). These sections are additive, causal, adversative and sequential connectors. As noted before in the theory section, these will include both traditional connectors and conjunctions, furthermore just listed as connectors.

Table 5 contains the comprehensive list of connectors used in this study.

Table 5. List of all connector words used in the corpus study. The underlined words have been added to the list provided by Hasan & Halliday (1996, 243).

Connector type	
Additive (17)	And also, or else, furthermore, in addition, besides, alternatively, incidentally, by the way, I mean, in other words, for instance, for example, thus, likewise, similarly, in the same way, by contrast
Causal (23)	hence, therefore, consequently, for this reason, on account of this, as a result, in consequence, for this purpose, with this in mind, because, it follows, on this basis, to this end, in that case, in such an event, that being so, under the circumstances, otherwise, in this respect, in this regard, with reference to this, in other respects, aside from this.
Adversative (14)	though, but, however, nonetheless, despite, in fact, as a matter of fact, on the other hand, at the same time, on the contrary, in any case, in either case, at any rate, <u>in contrast</u>
Sequential (33)	after that, just then, previously, before that, finally, at last, in the end, at once, thereupon, soon, after a time, next time, on another occasion, next day, an hour later, meanwhile, until then, at this moment, firstly, secondly, thirdly, in conclusion, up to now, hitherto, at this point, from now on, henceforward, to sum up, to summarize, in short, briefly, to resume, <u>subsequently</u> .

However, while inspecting the words listed in Halliday and Hasan, I discovered that the list they had included was missing words that I think fit these categories and that are represented in the corpora, so added such words to the list for a more comprehensive inspection. For example, *in contrast* and *subsequently* were not in the original text. Additionally, any

connectors that come up with 0 hits in both corpora will be dismissed and not listed in the end result. Finally, separating instances of *and* where it is used as a connector and where it is used for lists is not a feasible task so it will be left out of the list. The same is true for *that is, or, nor, so, for, then, yet, only, actually, instead, at least, rather, next* and *here*. I will proceed to list all the connectors per category, later on adding them all together to examine potential differences among the categories as well as all connectors together. Overall this list contains 87 different connectors. These are listed in the appendices. In order to analyse the texts, I used the WordSmith Tools software (Scott 2015). I ran every text through the Concord tool in WordSmith, which provides a list of every instance a specific word is used in alongside the context it appears in. This way it is possible to ascertain that the words are categorised correctly.

The corpora both contain texts from university level students to make them as comparable as possible, reducing factors other than the writers' native language to a reasonable extent. The Uppsala Student English Corpus (USE) contains 1,489 essays, collected from 440 Swedish university students, mostly from students doing their first term of studies. It contains 1,221,265 words in total among all these texts. They have around 700-800 words each typically and cover various different subjects. The essays are marked to show which were written by which student, with their actual identities hidden. Other information that could help identify the writer was also erased from the material. The students were also informed of the collection of these texts beforehand.

The British Academic Written English was collected from the universities of Oxford Brookes, Reading and Warwick. The texts are from 4 different groups: art and humanities, life sciences, physical sciences and social sciences. Overall the corpus contains a total of 2761 assignments, with a total of 5,506,995 words in all. It contains information on the year each student who wrote the text is and from which discipline they are while hiding their true identities.

The actual process of gathering the data is that I collected all texts from one corpus together in the Concord tool in WordSmith. Once I did this, I could enter a search term I was interested in and the program went through all of the instances that word was used in while showing the surrounding context. As shown in image 1 below, the program shows the results in their context, making it possible to see if there is anything notable in the words why they might not be suitable for the analysis. No such issues occurred, but it was useful to have that option available when checking the results. The program shows other information too, such as the files they are from but I will not need to take note of that for my analysis.

Image 1. Results for the word *despite* using the Concord tool in WordSmith.

N	Concordance	Set	Tag	ent.	Sent.	Para.	lead.	head.	ect.	Sect.	File	Date	%	
					Pos.	Pos.	Pos.	Pos.		Pos.				
1 079	nurses and midwives in modern day <b>practice</b> , <b>despite</b> the vast clinical difference between				6	84	0	539	n/a	n/a	0	0	3034d.txt 116/12/12 00.00	7,3%
1 080	where they become involved in <b>practice</b> <b>despite</b> the prohibitions (whether in IHL or				34	981	0	370	n/a	n/a	0	0	0423c.txt 116/12/12 00.00	64,3%
1 081	of traditional religious culture and <b>practice</b> . ↔ <b>Despite</b> this lack of consensus, religious				3	1	0	581	n/a	n/a	0	0	0144c.txt 116/12/12 00.00	4,2%
1 082	a straightforward ordering of <b>preferences</b> . ↔ <b>Despite</b> this theoretical backing, however,				27	1	0	282	n/a	n/a	0	0	0138b.txt 116/12/12 00.00	25,2%
1 083	box of 100 atoms was therefore still <b>present</b> , <b>despite</b> the disappearance of the peak in the				80	78	0	338	n/a	n/a	0	0	6137b.txt 116/12/12 00.00	64,9%
1 084	P that her support mechanisms are still <b>present</b> <b>despite</b> finishing her treatment. ↔One to				68	72	0	639	n/a	n/a	0	0	0260b.txt 116/12/12 00.00	92,5%
1 085	unconfirmed (15). 3.2 Histological <b>Presentation</b> </heading> <b>Despite</b> the proposed		ig>		109	177	0	619	n/a	n/a	0	0	0047a.txt 116/12/12 00.00	24,7%
1 086	direct effect on wages, through wage <b>pressure</b> <b>despite</b> excess supply in the labour market,				17	187	0	054	n/a	n/a	0	0	0077g.txt 116/12/12 00.00	45,1%
1 087	against the scarcity of food and of its <b>price</b> . ↔ <b>Despite</b> them acting in a way that was				133	1	0	955	n/a	n/a	0	0	0144a.txt 116/12/12 00.00	86,7%
1 088	of inherent influence and <b>privilege</b> . ↔ <b>Despite</b> equality advances in recent years, it				116	1	0	684	n/a	n/a	0	0	0005a.txt 116/12/12 00.00	85,6%
1 089	and the complicated recruitment <b>procedures</b> . ↔ <b>Despite</b> the relatively lax application				50	1	0	580	n/a	n/a	0	0	0350a.txt 116/12/12 00.00	29,9%
1 090	or end of decision-making <b>processes</b> . ↔ <b>Despite</b> discretionary powers; it merely				39	1	0	392	n/a	n/a	0	0	0209e.txt 116/12/12 00.00	31,1%
1 091	the same electron carriers. Organelle <b>Processes</b> </heading> <b>Despite</b> having different		ig>		96	124	0	003	n/a	n/a	0	0	0045a.txt 116/12/12 00.00	67,7%
1 092	of the explanation of atomic <b>processes</b> , <b>despite</b> the fact that we now understand the				66	149	0	727	n/a	n/a	0	0	6129e.txt 116/12/12 00.00	92,6%
1 093	in very poor prices paid for livestock <b>produce</b> . ↔ <b>Despite</b> efforts to improve productivity				50	1	0	579	n/a	n/a	0	0	6015d.txt 116/12/12 00.00	94,0%
1 094	and fair allocation of its joint <b>products</b> . ↔ <b>Despite</b> critics being quick to point out that				45	1	0	531	n/a	n/a	0	0	0244k.txt 116/12/12 00.00	26,1%
1 095	foods with ready-made food <b>products</b> . ↔ <b>Despite</b> industrialized societies reaching a				27	1	0	865	n/a	n/a	0	0	6012a.txt 116/12/12 00.00	33,7%
1 096	confidence to the auditing <b>profession</b> , <b>despite</b> irritating the firms themselves. ↔				107	394	0	337	n/a	n/a	0	0	0253g.txt 116/12/12 00.00	92,0%
1 097	towards creating a secular legal <b>profession</b> . ↔ <b>Despite</b> the religious influences this at least				44	1	0	978	n/a	n/a	0	0	0411b.txt 116/12/12 00.00	58,3%
1 098	on, for example the health care <b>profession</b> . ↔ <b>Despite</b> this, it can be argued that we are				81	1	0	660	n/a	n/a	0	0	6012a.txt 116/12/12 00.00	94,7%
1 099	will eventually result in a fall in the rate of <b>profit</b> <b>despite</b> an increase in the rate of surplus. ↔				96	106	0	559	n/a	n/a	0	0	0135f.txt 116/12/12 00.00	58,9%
1 100	the direct involvement forms less <b>profoundly</b> . ↔ <b>Despite</b> this fact, the ICE regulations				60	1	0	647	n/a	n/a	0	0	0219d.txt 116/12/12 00.00	68,6%
1 101	was part of the government <b>program</b> . ↔ <b>Despite</b> the "construction of more than				61	1	0	117	n/a	n/a	0	0	0261c.txt 116/12/12 00.00	71,7%
1 102	Kent County Council and Gap Activity <b>Projects</b> . </note> <b>Despite</b> the criticisms which can be		te>		24	218	0	858	n/a	n/a	0	0	3042e.txt 116/12/12 00.00	26,8%
1 103	had control over women's moveable <b>property</b> . ↔ <b>Despite</b> then the position of women				4	1	0	563	n/a	n/a	0	0	0144a.txt 116/12/12 00.00	3,1%
1 104	↔Saved's very title is optimistic, even <b>prophetic</b> : <b>despite</b> its potential for irony, Bond chooses				3	50	0	729	n/a	n/a	0	0	0126b.txt 116/12/12 00.00	4,9%
1 105	to illustrate consistency in Leveller <b>proposals</b> , <b>despite</b> the apparent advocacy of a				79	81	0	206	n/a	n/a	0	0	0040c.txt 116/12/12 00.00	20,1%
1 106	economy, most tragically into <b>prostitution</b> . ↔ <b>Despite</b> the fact that 'sex for hire (is) not a				96	1	0	200	n/a	n/a	0	0	0417d.txt 116/12/12 00.00	80,1%
1 107	for example crime, and <b>prostitution</b> , <b>despite</b> the fact that much of this was not				38	86	0	659	n/a	n/a	0	0	0001a.txt 116/12/12 00.00	79,3%
1 108	immediately in the reader for his <b>protagonist</b> , <b>despite</b> the fact that, until the bottom of the				12	175	0	265	n/a	n/a	0	0	0229c.txt 116/12/12 00.00	7,0%
1 109	pressure groups and organised <b>protests</b> . ↔ <b>Despite</b> this, it is true that many of they				97	1	0	293	n/a	n/a	0	0	0275c.txt 116/12/12 00.00	77,1%
1 110	of their theory. ↔What they do <b>provide</b> , <b>despite</b> the problem of their errant				64	22	0	335	n/a	n/a	0	0	0255c.txt 116/12/12 00.00	36,8%
1 111	are a number of mobility genes on <b>pXO1</b> . ↔ <b>Despite</b> genomic variability in the <i>B. cereus</i>				60	1	0	072	n/a	n/a	0	0	0141g.txt 116/12/12 00.00	48,0%
1 112	destiny as opposed to a personal <b>quest</b> . ↔ <b>Despite</b> Aeneas' seeming desire to stay with				48	1	0	739	n/a	n/a	0	0	6192b.txt 116/12/12 00.00	65,4%

I could then check how many times the word appeared and wrote down that number. The same process was then repeated for the other corpus. In order to analyse the numbers properly, the Uppsala numbers were multiplied by 5.33 so the numbers were comparable and the different sizes of the corpora do not make a big difference. This has its own issues, where some smaller numbers might be amplified disproportionately in a way that would not reflect a larger sample size in the corpus, but that is a minor issue to be expected in order to gain the ability to compare the results generally. Otherwise the results achieved here should work well and could be repeated in other studies for different corpora for further analysis.

## 4. Results

Because of the practical unfeasibility of counting some of the more common connectors which have several meanings, I had to dismiss many words from this study which will potentially skew data between the categories a little, especially because a word such as *and* is featured a large number of times in English texts. However, I argue that this is not a problem for my thesis because the relative word count between the listed words in each corpus is what I am more interested in.

Going through all the instances of connectors picked up by WordSmith, there was a total of 121,741 instances throughout all of the texts. A thorough division of all these numbers can be found in appendix 1 where they are divided for each connector and both corpora. In Table 6 and 7 below I show the most used connectors in both corpora.

Table 6. Most commonly used connectors for Uppsala corpus

	Uppsala	British
Because	12163	6867
But	7073	15147
However	5159	12230
Therefore	3907	7815
Soon	1679	430
Finally	1610	1193
On the other hand	1567	840
For instance	1492	777
Furthermore	1269	1315
In fact	1189	1200

Table 7. Most commonly used connectors for British corpus

	British	Uppsala
But	15147	7073
However	12230	5159
Therefore	7815	3907
Because	6867	12163
Thus	4483	1433
Hence	2094	213
Though	2035	963
Despite	1618	645
Furthermore	1315	1269
As a result	1278	522

The results show that neither group overwhelmingly used all types of connectors more than the other but there is more nuance to the results. I had expected the native speakers to have a much greater use of these words but upon examining the results, this is not the case. A total of 41 words entered into the program showed that they appeared more often in the Uppsala corpus. The normalisation process might skew the results slightly: while I consider the normalisation to be necessary for more even data, the fact that the British texts contained five times more words but there were only around twice as many texts means that the texts are structured slightly differently. Still, while the results are not a clear-cut case of one group using all words more, there are still many things to note. Certain words do have quite significant differences here. Perhaps the clearest difference is in *hence*, which is the sixth most used connector from the British text with 2094 times of use, and while it is used in the Uppsala corpus more than just a handful of times, 213 is still significantly less. However, the Uppsala students do use other causal connectors plenty, such as *therefore* and *because*. In contrast to that, the top 10 most common connectors in the British texts did not include a single sequential connector. As shown in the appendix, these are used to a decent extent too, with *finally* showing up 1193 times in the British texts too, which while not a small number is still relatively rarer than the more commonly used ones. According to these results, the British texts did not highly favour sequential structures but instead focused more on causality, but it should not be taken too strictly as the sequential words are still not neglected either.

These results are not entirely clear cut – cohesive methods besides connectors can be used for a similar semantic or structural meaning and these will not be noticeable by the methodology used here. Another potential issue is the specificity of the search terms, especially with multi-word connectors where words can be slightly altered for practically identical meaning. However, I believe these do not present a problem for the thesis as these issues would mostly alter the absolute numbers but not necessarily the relation between the numbers, so the question of the relative use of connectors between the groups can still be answered with these incomplete data.

#### **4.1. General differences in connector use**

In this section I will go over the first research question: how do the two groups differ in general in their connector use? Firstly, the vocabulary of the British group was larger in terms of the chosen words. There were 19 words that only appeared in the British corpus and they never appeared in the Uppsala corpus. However, the words that were never used there had very few appearances in the British counterpart in all cases, so they are just words that both groups did not favour to a great extent. These seem to be more of an anomaly, however there were no instances of a word only appearing in the Uppsala corpus texts. Connectors that only appeared in the British texts are the following: *incidentally, by contrast, on account of this, on this basis, to this end, in such an event, in this regard, with reference to this, in other respects, in either case, at any rate, just then, thereupon, on another occasion, until then, henceforward, to resume*. None of these were used more than 100 times in the British corpus and most of them were close to 0 even then. Table 8 below shows the number of each of these words used in the British corpus.

Table 8. Connectors only used in British texts.

Connector	Use in BAWE
Incidentally	20
By contrast	61
On account of this	1
On this basis	13
To this end	12
In such an event	1
That being so	1
Under the circumstances	3
In this regard	37
With reference to this	1
In other respects	2
In either case	4
At any rate	9
Just then	1
Thereupon	76
On another occasion	1
Until then	7
Henceforward	1
To resume	7

Secondly, there still are many words that are clearly favoured by one group more. As noted above, 41 words were used more commonly by the Uppsala group whereas 47 words were used more often by their British counterparts. This difference is not that large of course, but it is notable that both groups did have quite a large number of words more commonly used by them. By only examining this data it is not possible to explain what causes these differences of course. There could be more of an emphasis for certain sentence structures for one area, especially for the Uppsala corpus since those texts are collected from one university only, while the British texts were collected from three separate universities. This means that conventions taught in one area might differ from others even more than the varied background of students already might cause. For more concrete examples, the British group used *in*

*addition* 1,221 times overall while the Uppsala students have a total of 538 instances of it after the differing word counts are taken into account. Another word with a great representation for the British texts was *thus*, which appeared 4,483 times for the British data, but only 1,433 times for the Swedish. It was popularly used in both cases but much more frequently in one.

Other words that were much more popular for British use were *alternatively, similarly, hence, therefore, as a result, though, but, however, despite, in contrast, previously, meanwhile, firstly, subsequently*. On the other hand, *by the way* and *I mean* appear 155 and 469 times, respectively, in the Uppsala corpus and only 14 and 27 times in the British corpus, which is quite a large gap overall. Additionally, *for instance* appears 1,492 times in the Uppsala corpus and only 777 times in the British one. There are other words that are relatively more represented in the Uppsala corpus: *besides, on the other hand, because, in that case, as a matter of fact, at the same time, on the contrary, after that, at the same time, soon, in conclusion, to sum up*. Others not mentioned in this section did have some variation, but the variation was relatively small. For a thorough examination of all the numbers, Appendix 1 contains every word and how many times they appeared when adjusted for normalisation of the two corpora.

## 4.2. Semantic categories

In this sub-section I will then go over the second research question: how are these connectors used in terms of their semantic categories? As listed earlier in the theoretical background, the categories I will use in my thesis are additive, causal, adversative and sequential. I will go over them one by one, listing potential differences that the two groups had between each other. None of the semantic categories showed an overwhelming preference among one group: both corpora showed plenty of use in all of them although there were still certain words among them that proved particularly popular.

When it comes to additive connectors here, both groups had words they used more than the other. Notably, *thus* was used 1,433 times among the Uppsala students but 4,483 times by the British group. Words like this might be a little less active in the vocabulary of second language users of English then. *Thus* is the most commonly occurring additive connector overall, but relatively speaking there are some larger differences too. The British students used *similarly* 713 times while it only occurs 53 times in the Uppsala students' texts. On the other hand, *I mean* is used 469 times in the Uppsala texts but only 27 by the British

students, so there are words that are more popular among them too. *In addition* is used 538 in the Uppsala group, 1,221 times by the British group. However there is an inverse relation with *besides*, used 576 times by the Uppsala students and 276 times by the British students. The overwhelming numbers show more absolute use of these by the British students comparatively but there's variation among particular words.

Causal connectors then show more uniform results with a slight exception. Most of the words in this category show more common use by the British student group, with a few more only showing very slightly higher numbers by the Swedish students. These are words like *hence, therefore, consequently, as a result, it follows, in this respect*. To be exact, 19 out of the 23 items in causal connectors were more common in the British texts. However, one single connector, *because*, is used 12,163 times by the Uppsala group and a little more than half of that by the British group, 6,867 times. While that is still one of the more common words used by the British students as well, it is a large difference in terms of absolute numbers. In terms of just categories, this one difference equalises the groups a lot in terms of casual connector use. Finally there are a few connectors that while rarely used by the British students, do not show up in the texts by their Swedish peers once. These are *to this end, on this basis, on account of this, with reference to this, in other aspects*. Notably these are all used less than a hundred times by the British texts too and are overall not common there either.

Adversative connectors are overall more represented by the British users' texts but there is no single clear trend with them either. Here there is another large difference, with *but* appearing 7,073 times in the Uppsala texts and a twice as high number of 15,147 in the British texts. A similar relative difference can be found in the 5,159 times of *however*, contrasted with the 12,230 times it was used in the British texts, as well as *though* with a respective 963 and 2,035 times. *Despite* is also used 645 by the Swedish group, 1,618 by their peers. But it is not all in the favour of the British either. *In fact* is used approximately 1200 times by both groups, so quite commonly but similarly by both groups. *On the other hand* appears 1,567 times in Uppsala texts and 840 times in the British texts. *On the contrary* 352 times and 114 times respectively, and *at the same time* 1,157 and 507 times. Finally there's a couple of rare ones with *in either case*, used 4 times by the British and *at any rate*, 9 times. Both of these are not used by the Swedish students once.

The final group of sequential connectors on the other hand seems to overall trend with higher numbers among the Uppsala students. Again it is not just that simple and the British students do have some connectors that they use more often. 21 out of 33 words from the list

of sequential connectors here are used more commonly by the Uppsala students. Overall it is not a large difference. Words used clearly more often by the Swedish are *after that* (240 to 65), *at the same time* (1,157 to 507), *finally* (1,610 to 1,193), *at last* (80 to 39), *soon* (1,679 to 430), *in conclusion* (437 to 127) and *to sum up* (378 to 63). Still, while fewer in number, the British do have their own words they prefer with *previously* (899 to 165), *meanwhile* (143 to 64), *firstly* (864 to 448) and *subsequently* (349 to 37). Finally there are again words that used once by the Swedish students, but a few times by the British in *thereupon* (76), *just then* (1), *on another occasion* (1), *until then* (7), *henceforward* (1) and *to resume* (7).

## 5. Discussion

With these results then, it is time to discuss what kinds of conclusions can be drawn. In the previous section I have gone over the most important differences in data for a general look of the numbers. As evident by the result, there were no overwhelming differences inside one single semantic group of connectors, however certain words were certainly overrepresented in texts of one group. The reasons for this can only be guessed as there is no real data to tell why these differences exist.

By the nature of the British group consisting of native speakers of English, it is perhaps not surprising that there are no instances of a connector used only by the Uppsala group. Still, there is no word that is very popular among the British users of English while not at all used by the Swedish group. It also seems that the Swedish group, while having enough proficiency to use connectors to create cohesion in their texts, may rely on some more common ones a little more. This is exemplified by *because* in the group of causal connectors, where it is used 12,163 times to be very common, when their peers use this word only around half of that many times. On the other hand, *however* is also used a comparative number of times by the British writers 12,230 times and an even higher number of use in *but*, 15,147 in total. So it is not fair to say that the Swedish students are the only ones heavily relying on some relatively simple one-word connectors. It is also notable that the semantic meaning of a few of these connectors is almost identical, with only slight variation in how they are expressed on a syntactic level. The Swedish student for example never used *in this regard* but did use *in this respect* and these serve very similar purposes in a sentence, so the quality of writing is mostly not affected, if not for potential repetition with fewer alternatives used.

To reflect on Crossley and McNamara (2012), it is also good to bring up again their results where texts that otherwise utilised complex language and high levels of proficiency in general, cohesion was not necessarily reflecting this high proficiency. Some of the more competent students suffered in expressing cohesion, and their texts may have suffered from using complex language, at least when it comes to creating cohesion. Repetition and other methods of cohesion may occur more in a simpler text too. This is not necessarily a flaw on the complex text however, since more advanced language might be appropriate when discussing more complex topics and these texts may just assume more from the reader, thus taking short cuts with cohesion when discussing things that require little explanation from those already familiar with the subject matter. It is also important to mention that just because some more varied connectors were used by the British group in total, it is not indicative of them being more cohesive overall. In regard to this, focusing only on connectors for the scope of this study may ignore other methods of cohesion so these results should be noted just for connector use, not to imply that those who use connectors more are achieving higher levels of cohesion.

When it comes to cohesion and goals of achieving communicative competence as per the Common European Framework of Reference, both of these group seem to be using connectors to achieve these goals. Of course, the second language group also consists of university level students,

with long-time exposure to English as part of their curriculum and in many cases, during their free time as people in the modern world are often exposed to English in many contexts, even if the particular expose is difficult to measure across texts. The C1 level of language competence in the CEFR, as listed in Table 4 of this paper, lists the use of connectors as a method to create cohesion as an important metric for someone to have achieved that level. Furthermore, the C2 level in the same table indicates more varied mastery of cohesive devices, which does fall outside of the scope of this thesis partially but again highlights the importance of connectors among other things. Without explicitly measuring how well these connectors are used, it is clear that both groups have varied use of them and while the particular connectors may vary overall, both groups show high levels of competence and the second language group has acquired many connectors they use in texts.

When examining what Trebits (2009) said, it is possible that these students also get influenced by teachers and the surrounding social context to explicitly prefer certain connectors. Her study noted that there were several words that were very common in the European Union and its official contexts, while being not in much use if at all for native, British contexts, at least in the corpus used in this study. In response she suggested teaching words like this to students or other people interested in working for the European Union explicitly to make them familiar with associated language choices. Such ideas can affect my results, if students are instructed to explicitly prefer certain words in an academic context, or even just implicitly by seeing certain words used more often. If such an influence is lacking for the other group, the pressure to use certain words might not be present. As such, it is also to be noted that this academic language use can differ greatly from other language use in the same areas.

Compared to Jin 2001, I believe the differences are not quite as large. Certainly differences exist between Swedish and British culture and language too, but they are closely connected languages that are not quite as distant as languages spoken in China for example. I think this is evident in the fact that many of the more popular connectors were common in both the L1 and L2 users of English, albeit differences existed here too. As pointed out in this study, cohesion can also be created with other means like word repetition and synonyms and such differences remain unexamined in my thesis.

To summarize, the differences certainly exist between these two groups but it is difficult to make out very clear general trends in terms of one group of connectors being dismissed or underutilised by either of the groups. This can be due to a variety of factors. Swedish and English are linguistically not too far from each other, and geographically both the UK and Sweden are both in Europe and not too far away from each other. This is not to say there are no differences between them but they are also similar in many ways, and perhaps many conventions of English writing would already be learnt by Swedish speakers when it comes to their native language, making it easier to apply to a new language that operates in similar ways. When comparing this to Faruk (2016), the cultural differences and norms might not vary as much from the beginning, so these results might rather indicate that Swedish natives are able to easily adopt the norms used in British English, where

the differences are not too overarching even when they exist. Naturally there were other limitations too. Not all of the students have the same level of proficiency, although with corpus studies like this, outliers should not look too prominent with the other texts surrounding them. Still, the texts are not from the same fields of study either and are from a limited number of universities too. In the case of the British corpus they are collected from three different universities however to make it slightly more representative of a larger group, but that remains limited too. And as noted before in the discussion section, some of these connector words get lost when searching a larger corpus, and as such the full picture is not quite shown.

## 6. Conclusion

The results of this thesis did not present a clear one-way conclusion where one group uses connectors more or less in general, however there are certain trends where one group had a clear preference for certain words. The reasons for this cannot be determined for certain however. My research questions I set out to find answers from the start where to see what kinds of differences would I find in connector usage between the two different language groups, and secondly the connector usage differences between the different semantic groups. So, to answer the research questions, there is certainly a lot of variation in what kinds of words are used. Whether through informal or formal language learning there are certain trends to be found although in many cases connectors that are popular among one group are still somewhat popular with the other even if less so. Many connectors are naturally used to a similar extent between both texts. The number of potential connectors included was fairly large, so it is expected that some words are common or uncommon in both groups even if they have differences on the large scale. This was of course shown in words like *but* and *because*, which were very high in both language groups at the 4 most highly used words for both languages. Certain words had such great differences between the compared groups that I think it shows a clear enough difference between the groups to exist. A great example of this is *hence*, which was used over 10 times more in the British texts while it was not in high use for the Uppsala corpus texts. However both groups also use connectors in flexible ways that it would not be right to say one group uses them in a better fashion. As pointed up above, *hence* was a very popular word with 2094 uses for the British group but it was fairly unused by the Swedish students at 213 uses. So while they did use it more than many other words it was not too prominent.

Many of the connectors also have semantic overlap where one can convey similar relationships even when using varied methods. When it comes to the semantic groups and the second research question, there were some notable differences with the British group shining in the diversity of connectors used in most groups, but even then the Swedish group always had certain connectors they were using more than their peers. When looking at sequential connectors, the Swedish group also lead in more connectors, with the British students also having relatively low use in sequential connectors and none of them were in their ten most used connectors. It is also notable however that the British students used many sequential connectors that their peers completely neglected to use. *Thereupon* was not used once in the Uppsala corpus texts for example, but a total of 76 times for the British texts. This is not a very high number either, but it is a theme that a few of these words are used a small number of times for the British texts and none at all for the Uppsala text. The reverse was not true for any connector: any connector used by the Uppsala students was used at least once by their peers, even if the Uppsala students did have higher numbers for many words as well.

There are certain limitations in this study, especially when it comes to scale. As a one-person corpus study, there are physical limitations on how deeply I can analyse any particular text, with just

one person going through the texts. While I can reasonably look at some, overall I will need to rely on aggregate data and general tendencies. This means that small things may escape my notice even if something noteworthy appears that I have not accounted for before-hand. Certain multi-word connectors might even be missed, particularly those with similar variants already listed in this study, yet with the lists I have gathered from previous material on similar topics, I believe most such instances have been accounted for. Still, I believe this is only a minor setback and does not detract from my results, only that there is more to look into potentially. Another thing relating to corpus size here is that the corpora had quite a significant difference in the number of texts and words. While this can be alleviated by normalising the number of words as I have done, it is not equivalent to having the authentic unmodified numbers. I believe this to not affect too much, but when it comes to smaller numbers, it might skew them a bit when the Uppsala group numbers get multiplied to an extent where the corrected for results make these rare words overrepresented a little bit. Larger numbers are less of an issue here, as the nature of being largely represented itself makes the multiplication more likely to produce accurate results instead of amplifying anomalies. Finally, cohesion is caused by many things with connectors only being a subset of that, so any results about the use of connectors might be influenced by any single group using other methods of cohesion in its place, so the results only relate to the use of connectors, overall cohesion cannot be properly evaluated by these results.

In terms of future research, there could certainly be a smaller sample size of texts too, but a much more thorough look through them, to refer back to the limitations. The current dataset is useful for going through a larger number of texts, but it also becomes much more complicated to focus on the particular texts within the restrictions of this paper, where the large-scale conclusions are still reliable, but some smaller details might not receive the care they need. Other aspects of discourse competence or more specifically cohesion could be taken into consideration too, in order to gain a more whole idea of how these ideas are communicated in text when comparing these groups. Naturally it would also be possible to repeat the methods in this paper with different corpora and different language groups to see if they show similar or different results and what those potential differences could be caused by.

## References:

### Primary sources:

USE = Uppsala Student English corpus, compiled by Margareta Westergren Axelsson and Ylva Berglund, the Department of English, Uppsala University, 1999-2001.

BAWE = The British Academic Written English corpus, compiled by Alois Heuboeck, Jasper Holmes and Hilary Nesi. Universities of Warwick, Reading and Oxford Brookes. June 2010.

### Secondary sources:

Canale, Michael. 1987. "The Measurement of Communicative Competence." *Annual Review of Applied Linguistics*. 867–884. Cambridge University Press

Canale, M. and Merrill Swain. 1980. "Theoretical bases of communicative approaches to second language teaching and testing." *Applied Linguistics 1*. 1-47. Oxford University Press.

Connor, Ulla. 1996. "Contrastive Rhetoric: Cross-Cultural Aspects of Second Language Writing." *Cambridge Applied Linguistics*. Cambridge: Cambridge University Press.

Connor, Ulla. 2002. New Directions in Contrastive Rhetoric. *TESOL quarterly* 36, no. 4: 493–510.

Connor, Ulla, Robert B. Kaplan, and Nils Erik Enkvist. 1986. "Writing Across Languages: Analysis of L2 Text. Reading (MA): Addison-Wesley.

Connor, Ulla. 1990. "Linguistic/rhetorical measures for international persuasive student writing." *Research in the Teaching of English* 24, 67-87.

Crossley, Scott A. and Danielle S. McNamara. 2012. "Predicting Second Language Writing Proficiency: The Roles of Cohesion and Linguistic Sophistication." *Journal of Research in Reading* 35, no. 2: 115–135.

Cummins, Jim. 1980. "The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue." *TESOL quarterly*. Volume 14, No 2. 175–187.

Cummins, Jim. 2000. "Language, Power and Pedagogy Bilingual Children in the Crossfire." *Multilingual Matters*. Clevedon.

Faruk, Syed Md Golam. 2016. "Bangladeshi EFL Learners' English Connectors: Overused or Underused?" *Theory and Practice in Language Studies* 6, no. 7: 1357–1363.

Flowerdew, John, and Michaela Mahlberg. 2009. "Lexical Cohesion and Corpus Linguistics." Amsterdam: John Benjamins Publishing Company.

Godo, Agnes M. 2008. "Cross-Cultural Aspects of Academic Writing: A Study of Hungarian and North American College Students L1 Argumentative Essays." *International journal of English studies* 8, no. 2: 65–111.

- Goldman, Susan R, and John D Murray. 1992. "Knowledge of Connectors as Cohesion Devices in Text: A Comparative Study of Native-English and English-as-a-Second-Language Speakers." *Journal of educational psychology* 84, no. 4: 504–519.
- Halliday, M. A. K., and Ruqaiya. Hasan. 1976. "Cohesion in English". London: Longman,
- Jin, Wenjun, 2001. "A quantitative study of cohesion in Chinese graduate students' writing: Variations across genres and proficiency levels."
- Kaplan, R. B. 1966. "Cultural Thought Patterns in Intercultural Education." *Language Learning* 16: 1-20.
- Narita, Masumi, Chieko Sato, and Masatoshi Sugiura. 2004. "Connector Usage in the English Essay Writing of Japanese EFL Learners." *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. European Language Resources Association. 1171-1174. Lisbon, Portugal
- Mohammad Raouf Moini, and Fateme Kheirkhah. 2016. "Use of Cohesive Devices in Children and Regular Literature: Conjunctions and Lexical Cohesion." *International Journal of Comparative Literature and Translation Studies* 4, no. 4: 12–23.
- Scott, Mike. 2015. *WordSmith Tools Manual*. Oxford: Oxford University Press.  
<https://lexically.net/downloads/version6/wordsmith6.pdf>
- Tanskanen, Sanna-Kaisa. 2006. "Collaborating Towards Coherence: Lexical Cohesion in English Discourse." John Benjamins Publishing Company
- Council of Europe. 2001. "The Common European Framework of Reference for Languages: Learning, teaching, assessment." Council of Europe. <https://rm.coe.int/1680459f97>
- Trebits, Anna. 2009. "Conjunctive Cohesion in English Language EU Documents – A Corpus-Based Analysis and Its Implications." *English for Specific Purposes (New York, N.Y.)* 28, no. 3: 199–210.
- Tsemach, Ehud, and Anat Zohar. 2021. "The Intersection of Gender and Culture in Argumentative Writing." *International journal of science education* 43, no. 6: 969–990.
- Viana, Vander. Sonia. Zyngier, and Geoff. Barnbrook. 2011. "Perspectives on Corpus Linguistics". Amsterdam. John Benjamins Publishing Company
- Yale Center for Teaching and Learning. *Linking Words (Conjunctions and Connectors)*. Accessed March 25<sup>th</sup> 2024. [https://poorvucenter.yale.edu/sites/default/files/files/GWC\\_LinkingWords-1.pdf](https://poorvucenter.yale.edu/sites/default/files/files/GWC_LinkingWords-1.pdf)
- Ying, H.G. 2000. "The Origin of Contrastive Rhetoric Revisited." *International Journal of Applied Linguistics* 10, no. 2: 259–268.

## Appendices

Appendix 1. Number of all words. Uppsala count multiplied by 5.33 for normalisation.

### **Additive:**

and also: Uppsala 1082 British 901  
or else: Uppsala 59 British 30  
furthermore: Uppsala: 1269 British 1315  
in addition: Uppsala 538 British 1221  
besides: Uppsala 576 British 276  
alternatively: Uppsala 5 British 134  
incidentally: Uppsala 0 British 20  
by the way: Uppsala 155 British 14  
I mean: Uppsala 469 British 27  
in other words: Uppsala 495 British 337  
for instance: Uppsala 1492 British 777  
thus: Uppsala 1433 British 4483  
likewise: Uppsala 75 British 161  
similarly: Uppsala 53 British 713  
in the same way: Uppsala 314 British 213  
by contrast: Uppsala 0 British 61

### **Causal:**

hence: Uppsala 213 British 2094  
therefore: Uppsala 3907 British 7815  
consequently: Uppsala 672 British 807  
for this reason: Uppsala 101 British 178  
on account of this: Uppsala 0 British 1  
as a result: Uppsala 522 British 1278  
in consequence: Uppsala 11 British 19  
for this purpose: Uppsala 43 British 47  
with this in mind: Uppsala 59 British 35  
because: Uppsala 12163 British 6867  
it follows: Uppsala 5 British 72  
on this basis: Uppsala 0 British 13  
to this end: Uppsala 0 British 12

in that case: Uppsala 96 British 18  
in such an event: Uppsala 0 British 1  
that being so: Uppsala 0 British 1  
under the circumstances: Uppsala 0 British 3  
otherwise: Uppsala 778 British 514  
in this respect: Uppsala 37 British 97  
in this regard: Uppsala 0 British 37  
with reference to this: Uppsala 0 British 1  
in other respects: Uppsala 0 British 2  
aside from this: Uppsala 5 British 3

**Adversative:**

though: Uppsala 963 British 2035  
but: Uppsala 7073 British 15147  
however: Uppsala 5159 British 12230  
nonetheless: Uppsala 42 British 265  
despite: Uppsala 645 British 1618  
in fact: Uppsala 1189 British 1200  
as a matter of fact: Uppsala 240 British 13  
on the other hand: Uppsala 1567 British 840  
at the same time: Uppsala 1157 British 507  
on the contrary: Uppsala 352 British 114  
in any case: Uppsala 64 British 44  
in either case: Uppsala 0 British 4  
at any rate: Uppsala 0 British 9  
in contrast: Uppsala 245 British 496

**Sequential:**

after that: Uppsala 240 British 65  
just then: Uppsala 0 British 1  
previously: Uppsala 165 British 899  
before that: Uppsala 90 British 7  
finally: Uppsala 1610 British 1193  
at last: Uppsala 80 British 39  
at once: Uppsala 117 British 76  
thereupon: Uppsala 0 British 76  
soon: Uppsala 1679 British 430

after a time: Uppsala 5 British 1  
next time: Uppsala 91 British 16  
on another occasion: Uppsala 0 British 1  
next day: Uppsala 21 British 11  
an hour later: Uppsala 5 British 2  
meanwhile: Uppsala 64 British 143  
until then: Uppsala 0 British 7  
at this moment: Uppsala 43 British 15  
firstly: Uppsala 448 British 864  
secondly: Uppsala 634 British 548  
thirdly: Uppsala 181 British 123  
in conclusion: Uppsala 437 British 127  
up to now: Uppsala 16 British 11  
hitherto: Uppsala 5 British 20  
at this point: Uppsala 250 British 180  
from now on: Uppsala 37 British 8  
henceforward: Uppsala 0 British 1  
to sum up: Uppsala 378 British 63  
to summarize/summarise: Uppsala 75 British 58  
in short: Uppsala 165 British 109  
briefly: Uppsala 112 British 196  
to resume: Uppsala 0 British 7  
subsequently: Uppsala 37 British: 349

## Appendix 2: Finnish Summary.

### 1. Aloitus

Kielen osaaminen on monimutkainen kokonaisuus ja siihen kuuluu monia osa-alueita. Erityisesti kouluissa on helppo jakaa kielen oppiminen sanaston ja kieliopin hallitsemiseen ja on totta, että ne ovat näkyvämpiä osia kielen osaamisesta. On kuitenkin myös olennaista osata luoda tekstejä, joita on helppo ja sujuva lukea. Tätä tekstin sujuvuutta ja kykyä saada viesti perille onnistuneesti kutsutaan koheesioksi. Tämä tutkimus keskittyy siihen, miten koheesiota luodaan konnektorien avulla. Erityisesti keskityn siihen, millaisia eroja löytyy englantia äidinkielenä puhuvien välillä ja englantia toisena kielenä puhuvien opiskelijoiden välillä. Tätä varten olen tutkinut Uppsala Student English (USE) ja British Academic Written English (BAWE) korpuksia. Käytän niihin kerättyjä tekstejä, jotta pystyn vertailemaan näitä kahta ryhmää keskenään. Tutkimuskysymykseni ovat seuraavat: 1. Miten tutkittavien kielitaustat vaikuttavat konnektorien käyttöön kokonaisvaltaisesti? 2. Mitä eroja konnektorien käytössä semanttisten kategorioiden sisäisesti.

### 2. Teoria

Käytän tätä tutkimusta varten useita teorioita, jotka reflektioivat tutkimukseen- Tärkeimpänä yksittäisenä teoriana käytän kontrastiivisen retoriikan teoriaa. Sen lisäksi käyn läpi yleistä teoriaa liittyen koheesioon, konnektoreihin ja viimeiseksi käyn läpi Common European Framework of Referencen mukaan kielen osaamista ja miten se liittyy erityisesti koheesioon.

#### 2.1. Kontrastiivinen retoriikka

Ulla Connor (1996, 5) kuvailee kontrastiivista retoriikkaa niin, että se havaitsee ongelmia toisen kielen kirjoittamisessa ja pyrkii selittämään ongelmat äidinkielen retorisen strategian avulla. Ensimmäisen kerran konseptin esitti Robert Kaplan vuonna 1966. Hän esitti (1966, 4), että toisella kielellä kirjoittamiseen ei kokonaisuudessaan riitä vain, että osaa kieltä, mutta osa kirjoittamisesta vaatii kulttuuritietoa, joka usein poikkeaa kielten välillä ja on myös implisiittistä, ei erityisesti opetettua. Kaplan nostaa esimerkiksi Arabian kielen, jossa suositaan rinnakkaisia rakenteita ilmaisemaan viesti, verrattuna Englantiin, joka sen sijaan käyttää induktiivisia ja deduktiivisia rakenteita, jotka ovat lineaarisempia. Kuitenkaan kielitaito sanaston ja kieliopin kannalta ei kummassakaan kielessä johda suoraan siihen tulokseen, että käytössä oleva rakenne on erityisesti parempi. Tämän opettaminen on myös vaikeampaa kuin esimerkiksi sanaston, sillä se on abstraktimpi

konsepti verrattuna sanastoon tai moneen kieliopin käsitteeseen. Omaan tutkimukseeni nähden on siis relevanttia tutkia, millaisia eroja voi löytyä ruotsia äidinkielenä puhuvien ja englantia äidinkielenä puhuvien välillä. Verrattuna esimerkkinä käytettyyn arabiaan, on kuitenkin mahdollista löytää enemmän samankaltaisuuksia näiden kielten väliltä, sillä ne ovat läheisiä sukukieliä. Ei myöskään ole välttämättä mahdollista sanoa, soveltaako yksittäinen kirjoittaja äidinkielen kanssa opittuja käytänteitä, vai onko hän oppinut kohdekielen mukaiset käytänteet niin, että pystyy soveltamaan niitä kun kirjoittaa toisella kielellään.

Kontrastiivinen retoriikka rakentuu monien muiden teorioiden esittämien ajatusten pohjalta. Connor (1996, 10) kuvailee näitä teorioita. Ensimmäisenä on luonnollisesti retoriikka itsessään, miten tekstin kirjoittaja yrittää vaikuttaa lukijaan tekstissä ja saada viestinsä perille. Toisena esimerkkinä Connor ottaa esille lingvistisen relativismin ja Sapir-Whorfin hypoteesin. Kuten kontrastiivisessa retoriikassa, siihen kuuluu ajatus siitä, että ihmisen äidinkieli vaikuttaa suoraan ajatteluun, joten äidinkielen konventiot voivat vaikuttaa myös siihen, miten käyttää vierasta kieltä. Tätä vertausta on kuitenkin myös kritisoitu, sillä Sapir-Whorfin hypoteesi ilmaisee paljon deterministisemmän kuvan kielestä kuin kontrastiivinen retoriikka (Ying 2000, 260).

## 2.2. Koheesio

Yksi tärkein konsepti tutkimustani koskien on koheesio. Sen voi määritellä kieliopillisten ja leksikaalisten elementtien luomiin yhteyksiin tekstissä. Crossley ja McNamara (2012, 131) toteavat, että koheesio ja konnektorien välinen yhteys on vielä tutkimusaihe, johon pitää keskittyä tulevaisuudessa lisää. Koheesiota voi luoda monin eri tavoin, mutta omaa tutkimustani varten keskityn nimenomaan konnektorien luomaan koheesioon. Se on osittain kielippiin, osittain leksikaalinen tapa luoda koheesiota. Flowerdew ja Mahlberg (2009) määrittelevät leksikaalisen koheesio sanojen luomana koheesiona, mutta se menee syvemmälle kuin yksittäiset sanat ja keskittyy kokonaiskuvaan. Leksikaalista koheesiota voi luoda toistamalla sanoja ja fraaseja tai luomalla yhteyksiä saman aihepiirin sanoilla (Hasan & Halliday 1976). Kieliopillinen koheesio puolestaan luodaan joko lauserakenteella tai sanoilla, jotka itsessään luovat koheesiota. Sanoja voi korvata sanoilla, jotka itsessään viittaavat aiemmin mainittuun sanaan, esimerkiksi pronomineilla. Voi myös käyttää yleistä kontekstia viittaamaan aiempiin sanoihin, tai jopa jättää sanoja sanomatta, jos se on kontekstista selvä vaikka sen pudottaa lauseesta. Koheesio mainitaan myös CEFR:ssä yhtenä kategoriana, joka mittaa kielen käyttäjän kielitaitoa. Kuitenkin, vaikka koheesio luominen on osa kielitaitoa, tai ainakin kyky tehdä niin, Crossley ja McNamara (2012) totesivat myös, että korkea kielitaito voi johtaa myös monimutkaisempiin rakenteisiin, jotka eivät ole aina niin koheesiivisia, eli suhde ei ole täysin lineaarinen.

### 2.3. Konnektorit

Kuten aiemmin mainitsin, koheesiota voi luoda monin keinoin. Kuitenkin tämän tutkimuksen fokuksena on konnektorit. Ne ovat sanoja ja fraaseja, jotka yhdistävät lauseita toisiinsa. Ne selkeyttävät millaisia suhteita lauseiden välillä on ja siten luovat koheesiota ja auttavat ymmärtämään. Sanat kuten *however*, *because* ja *in addition* ovat konnektoreita. Läheinen konsepti konnektorien kanssa ovat konjunktiot, mutta koska ne ovat leksikaalisesti niin samankaltaisia, en diskriminoin niiden välillä. Niiden suurin ero on se, yhdistävätkö ne lauseita vai virkkeitä ja ne voivat monissa tapauksissa olla muuten hyvin samankaltaisia.

### 2.4. Eurooppalainen viitekehys

Eurooppalainen viitekehys, tai Common European Framework of Reference (CEFR) mainitsee koheesion tärkeänä kielen osaamisen osa-alueena. CEFR määrittelee yleisen pohjan kielen osaamisen arvioinnille Euroopassa, jotta eri maiden välillä olisi yhteisesti määritelty standardi vertailukelpoiseen kielen osaamiseen. Tämä tutkimus keskittyy brittiläisiin ja ruotsalaisiin englannin puhujiin, joten se soveltuu erinomaisesti omiin käyttötarkoituksiini. CEFR puhuu diskurssikompetenssista, joka on osa Canalen ja Swainin teoriaa kommunikaatiokompetenssista (Canale & Swain 1980). Se määrittelee sen kykynä järjestää lauseita järjestykseen, joka muodostaa koherenttia kieltä (Council of Europe 2001, 123). Koheesion lisäksi diskurssikompetenssin muodostaa muutama muu kielenkäytön piirre. Korkeatasoinen diskurssikompetenssi pystyy ilmaisemaan asioita aihe/fokus -rakenteella tai vanha/uusi informaatio -rakenteella, muodostaa luonnollisia sekvenssejä, kuvailee syy-seuraussuhteita oikein, on koherenttia ja koheesiivista, jäsentää lauseet loogisesti, käyttää sopivaa kielityyliä, on retorisesti vaikuttavaa ja pyrkii kommunikatiiviseen yhteistyöhön ja ilmaisee sanoman selkeästi. Teoriaosion päätteeksi haluan vielä korostaa, että koheesio on olennainen osa hyvää kommunikaatiota, mutta se on hyvin monimutkainen kokonaisuus, jota on vaikea arvioida numeerisesti. Konnektorit ovat yksi selkeämmistä metodeista ilmaista koheesiota, mutta niitäkin arvioimalla jää puutteellinen kuva kokonaisuudesta. Nämä ovat kuitenkin tärkeitä aihealueita kielen osaamisessa, kuten voi todeta siitä, miten CEFR korostaa niiden osaa hyvässä kielitaidossa. Kontrastiivinen retoriikka teoriana ottaa taas kantaa siihen, kuinka konnektorien käyttö voi poiketa kielten välillä, sillä eri kielissä ihmisillä on eri odotuksia sopivaan kielenkäyttöön.

## 3. Metodologia

Tässä osuudessa käyn läpi metodologian, jota käytän tutkimuksessani. Käyn läpi tekstejä kahdesta eri korpuksesta, ruotsin tausta yhdessä ja englannin tausta toisessa. Sekä Uppsala Student English korpus ja British Academic Written English korpus sisältävät tekstejä yliopisto-opiskelijoilta, jotta ne ovat sen suhteen vertailukelpoisia. Tekstit korpuksissa on anonymisoitu, mutta merkitty niin, että taustatiedot kirjoittajista löytyvät. USE-korpus sisältää 1221265 sanaa kokonaisuudessaan, jotka käyn läpi konnektoreita etsien. Tekstit on kerätty Uppsalan yliopistosta. BAWE-korpus sen sijaan sisältää 6506995 sanaa. Sen tekstit on kerätty Oxford Brookesin, Readingin ja Warwickin yliopistoista. Tekstit BAWE-korpuksesta muodostavat paljon suuremman kokonaismäärän, yli viisinkertaisen verrattuna USE-korpukseen. Tämän takia olen normalisoinut tulokset niin, että Uppsalan korpuksen tulokset kerrotaan 5,33:lla. Tämä prosessi tarkoittaa, että varsinkin pienemmät numerot voivat vääristyä hieman ja yksikin esiintyminen sanalla näyttää suhteessa suuremmalta, mutta kun puhutaan suuremmista numeroista tämä suhteellinen vääristyminen ei enää vaikuta samalla tavalla. Sanojen laskeminen itsessään tapahtuu WordSmith ohjelmalla. Muutamia konnektoreja jätin myös pois, sillä ne esiintyvät liian paljon eri konteksteissa, joten niiden tutkiminen voisi vääristää dataa, kuten *that is* tai *next*.

### 3.1. Korpusanalyysi

Korpusanalyysia varten keräsin konnektoreita Hallidayn ja Hasanin listan mukaan (1976, 243). Kategoriat joita käytän tässä ovat additiivinen, kausaalinen, adversatiivinen ja sekventaalinen, riippuen konnektorin käyttötarkoituksesta. Kun selasin korpuksia, löysin myös lisää konnektoreita, jotka kuuluvat näihin kategorioihin ja lisäsin ne listaan. Listassa oli myös sanoja, jotka poistin sillä ne eivät esiintyneet kertaakaan kummankaan korpuksen teksteissä. Sanat, jotka esiintyivät vain yhdessä korpuksessa, pysyivät listassa. Listaani tuli kaiken kaikkiaan 87 konnektoria, jotka listaan liitteissä. Uppsalan korpus (USE) sisältää 1489 esseetä, jotka kerättiin Uppsalan yliopisto-opiskelijoilta ja esseet ovat tyypillisesti 700-800:n sanan pituisia. Yhteensä korpuksessa on 1221265 sanaa. Brittiläinen korpus (BAWE) sisältää 2761 tekstiä yhteensä, kerätty Warwickin, Readingin ja Oxford-Brookesin opiskelijoilta. Se sisältää yhteensä 6506995 sanaa. Molemmissa korpuksissa on käytetty useita tieteenaloja. Laitoin kaikki tekstit WordSmith-sovellukseen ja sen Concord-työkalua käyttäen löysin jokaisen konnektorin ja kuinka monta kertaa se esiintyy teksteissä. Sovellus näyttää myös sanan esiintymiskontekstin, jotta pystyy tarkistamaan, että sitä käytetään oikeassa kontekstissa.

## 4. Tulokset

Kokonaisuudessaan WordSmith löysi 121741 konnektoria näistä teksteistä listani mukaan.

### 4.1. Konnektorien käyttö kielten välillä

Tässä osuudessa käyn läpi eroja USE-korpuksen ja BAWE-korpuksen välillä ja miten ryhmät erosivat toisistaan kokonaisuudessaan. Ensimmäinen ero on se, että jälkimmäinen ryhmä käytti suurempaa sanavarastoa, ja 19 sanaa esiintyivät ainakin kerran BAWE-korpuksessa, mutta eivät kertaakaan USE-korpuksessa. Kuitenkin, kaikki näistä sanoista olivat suhteellisen epäsuosittuja myös BAWE-korpuksessa, mikään näistä sanoista ei esiinny edes 100 kertaa BAWE-korpuksessa. Kyseessä ei ole siis erityisen suuri ero.

#### 4.2. Konnektorien käyttö semanttisten kategorioiden mukaan

Tässä osiossa taas käsittelen konnektoreita niiden semanttisten kategorioiden mukaan. Kategoriat joihin olen jakanut ne ovat additiiviset, kausaaliset, adversatiiviset ja sekventaaliset konnektorit. Kokonaisuudessa voi sanoa, ettei kumpikaan korpus sisältänyt erityisen suurta eroa siinä, mistä kategoriasta konnektoreita suosittiin.

Additiivisten konnektorien osalta *thus* oli suosituin, ja se esiintyi BAWE-korpuksessa 4483 kertaa ja USE-korpuksessa 1433 kertaa. *Similarly* esiintyi 713 kertaa brittiläisissä teksteissä, mutta vain 53 kertaa Uppsalan teksteissä, mikä on suhteellisen suuri ero. Kuitenkin *besides* on konnektori joka esiintyy BAWE-korpuksessa 276 kertaa, mutta USE-korpuksessa 576 kertaa, joten molemmilla ryhmillä on eri konnektoreita suosiossa.

Kausaalikonnektorien osalta BAWE-korpus käyttää useammin suurinta osaa, vain neljä konnektoria ovat suosittumia USE-korpuksessa 23:sta. On kuitenkin huomioitava, että *because* tästä kategoriasta esiintyy 12163 kertaa USE-korpuksessa ja vain 6867 kertaa BAWE-korpuksessa. Suurin osa tämän kategorian tuotannosta Uppsalan opiskelijoilla muodostuu tämän yhden sanan erittäin yleisestä käytöstä.

Adversatiivisten konnektorien tapauksessa ne esiintyvät useammin BAWE-korpuksessa myös. *But* on hyvin suosittu molemmissa ja esiintyy BAWE-korpuksessa 15147 kertaa, mutta 7073 kertaa USE-korpuksessa, joten se on molemmissa hyvin yleinen. Samoin myös *however* esiintyy näissä korpuksissa 12230 kertaa ja 5159 kertaa, taas brittiläiskorpuksen eduksi.

Viimeiseksi sekventaaliset konnektorit ovat poikkeavia siinä, että niiden suhteen Uppsalan opiskelijat käyttävät niitä enemmän, varsinkin suhteellisesti. 21 33:sta konnektorista tässä kategoriassa ovat suosittumia Uppsalan teksteissä, mutta kuitenkin monet ovat runsaassa käytössä siis myös toisen ryhmän teksteissä.

#### 5. Diskussio

Tuloksia katsoessa ei voi todeta systemaattisia eroja suurella skaalalla. Kuitenkin molemmissa ryhmissä on selkeästi tiettyjä tendenssejä sen suhteen, mitkä sanat ovat erityisen suosittuja. Nämä erot eivät kategorisesti kuitenkaan sano mitään tarpeeksi tarkkaa, että voisi yleistää. Yksi olennainen huomio on, että Uppsalan teksteissä ei esiintynyt yhtään konnektoria, joka ei myös esiinny BAWE-

teksteissä. Tämä on ymmärrettävää, sillä vertailussa on äidinkielen puhujia ja toisena kielenä englantia puhuvia. Tästä huolimatta, sanat jotka ovat suosittuja BAWE-korpuksessa ovat kaikki kohtuullisessa käytössä myös USE-korpuksessa, vaihtelevasti.

Palaten Crossleyn ja McNamaran tutkimukseen (2012), he mainitsivat, ettei hyvä kielitaito aina automaattisesti tuota parempaa koheesiota. Siihen nähden kuitenkin sanoisin, että molemmat ryhmät käyttivät kohtuullisesti eri konnektoreita. Kuitenkin myös todettava, etten huomionnut muita tapoja luoda koheesiota tekstiin, joten laajamittaisempia johtopäätöksiä on vaikeampi tehdä. CEFR:ään verraten taas, sanoisin, että molemmat ryhmät osaavat konnektoreilla luoda koheesiota tekstiin monipuolisesti, joka viittaa korkeaan kielitasoon. Trebitsin (2009) tutkimus toteaa, että Euroopan Unionin töissä käytettävä englanti sisältää paljon sanoja, jotka ovat harvinaisia tavallisissa brittitexteissä taas, mikä on mielenkiintoista sen suhteen, että äidinkielenä englantia puhuvat eivät ole tottuneet käyttämään vahvasti sanoja, jotka kuuluvat tällaiseen virallisempaan kontekstiin. Hän mainitsee, että opiskelijoita voi johdattaa tiettyjen sanojen käyttämiseen ja tällainen ajattelu voi siten mielestäni pistää ajattelemaan, että opiskeluympäristö voi hyvin ohjata käyttämään tietynlaisia sanoja, ja akateeminen kieli voi sisältää suuria eroja tavallisempaan kielenkäyttöön.

Tiivistääkseni, sanoisin että ryhmien välillä on tiettyjä merkittäviä eroja, mutta niistä on vaikea tehdä isoja johtopäätöksiä suuremmalla skaalalla. Erot ovat monesti sanojen välillä ja silloinkin kun kategorisempia eroja voi havaita, ne eivät ole niin selkeitä, että voi olla varma selkeästä yhdenmukaisesta erosta. Verrattuna kieliin, joissa on suurempia eroja lähtökohtaisesti, voi olla, että ruotsin ja englannin taustat eivät eroa erityisen vahvasti toisistaan, toisin kuin esimerkiksi arabian ja englannin. Tutkimuksessani on tiettyjä rajoituksia, jotka voivat antaa syytä lisätutkimukseen, kuten keskittyminen muuhunkin kuin vain konnektoreihin, tai tarkempaan tutkimukseen pienestä määrästä tekstejä, jolloin kokonaiskuva on pienempi mutta keskittyneempi.

## 6. Johtopäätökset

En voi sanoa, että tutkimuksestani tulisi yksi selkeä vastaus siihen, käyttikö toinen ryhmä paremmin konnektoreita. Molemmissa löytyi eroja, mutta monet niistä eroista voivat olla subjektiivisia sen suhteen, että toinen ryhmä käyttää muita sanoja ja metodeja koheesion luomiseen. Vastatakseni tutkimuskysyksiin, eroja ryhmien välillä löytyi niin, että brittiläisten teksteissä esiintyi laajemmin eri konnektoreita, mutta silti myös ruotsalaisissa teksteissä oli monia konnektoreita, jotka olivat suosittuja ja myös suuremmissa määrin kuin vastaavissa brittiläisteksteissä. Kategorioiden mukaan, BAWE-korpus käytti useammin kaikkia muita kategorioita paitsi sekventaalisia, joten Uppsalan tekstit sisälsivät enemmän sanoja määrittelemään lauseiden suhdetta aikaan. Sekään ei ollut yksipuolista ja mikään kategoria ei ollut erityisen poikkeava kummankaan ryhmän osalta. Monet sanat, jotka ovat harvinaisia toisessa ryhmässä voi myös korvata synonyymeillä tai melkein vastaavilla sanoilla, sillä kielenkäyttö on monimutkaista. Tulevat tutkimukset voivat käyttää samoja tai tämän pohjalta

tehostettuja metodeita tai eri kieliryhmiä, jotta voi selvittää vahvemmin konnektorien suhdetta koheesioon kontrastiivisen retoriikan pohjalta.

