

**UNIVERSITY OF TARTU
TRANSLATION AND INTERPRETING CENTRE**

**RHETORIC OF MASTER'S THESES OF ESTONIAN AND
ANGLOPHONE ECONOMICS STUDENTS**

MA thesis

KAUR MÄNNAMAA

SUPERVISOR: ASSOC. PROF. RAILI MARLING

TARTU

2014

ABSTRACT

This thesis aims to give an overview of the differences in the academic English used by Estonian Master's students and Master's students with an Anglophone academic background. As the field as such is extremely broad, this thesis will focus on comparing the use of various linguistic instruments used for creating cohesion and coherence, as well as hedging. The comparison involves corpora from two universities (University of Tartu and MIT) and relies on computer based text analysis.

The thesis consists of five main parts. The introductory part gives a short overview of the aims of the author. Chapter 1 contains the discussion of the key concepts of the thesis and the difficulties involved in defining them. Chapter 2 offers an overview of the field of comparative rhetoric and also introduces similar studies that have been carried out before. Chapter 3 introduces the methods used in conducting the analysis for this thesis and contains discussion on the results reached. The conclusion is a short summary of the aims and results of the study carried out for this thesis.

TABLE OF CONTENTS

ABSTRACT	2
LIST OF ABBREVIATIONS	4
INTRODUCTION	5
1 CHAPTER I: KEY CONCEPTS OF THE THESIS	8
1.1 Hedging.....	8
1.1.1 Defining a Hedge.....	9
1.2 Cohesion and Coherence	12
2 CHAPTER II: LITERATURE REVIEW	15
2.1 Culture and Discourse.....	16
2.2 Comparative Rhetoric	18
2.3 English Academic Writing.....	21
2.4 Academic Writing of Non-native English Speakers.....	25
2.5 Academic Writing in Estonian.....	33
3 CHAPTER III: EMPIRICAL ANALYSIS	37
3.1 Data.....	37
3.2 Method.....	38
3.3 Results.....	42
3.3.1 Lexical Bundles	42
3.3.2 Discourse Markers	48
3.3.3 Hedges	51
CONCLUSION	55
REFERENCES	58
RESÜMEE	64

LIST OF ABBREVIATIONS

DM – discourse marker

EAP – English for academic purposes

EFL – English as a foreign language

FBA – frequency based approach

LB – lexical bundle

MIT – Massachusetts Institute of Technology

PTW – per thousand words

UT – University of Tartu

INTRODUCTION

It is hardly a surprise that an EFL speaker (someone for whom English is a foreign language) makes mistakes when writing in English. The mistakes need not always be language mistakes per se; there may be oddities regarding style, collocations that are rarely, if ever, used, unexpected argumentation patterns etc. This in itself is not that surprising but it implies an interesting question. What is it that makes us realise a text has been written in the native language of the author? Is it the lexis used, the structure or something else?

As I attempted to answer these questions, it became clear that I had to narrow down my area of research. There are many styles and traditions of academic writing but considering the length of this thesis and the limited resources it would not be possible to study all of them. Hence, I concentrated on a particular field: hedges and cohesion/coherence in MSc / MA theses in the field of economics. For the sake of cultural comparison, the corpus to be studied was to include theses from at least two academic cultures, one of which had to be English-speaking and the other Estonian due to my research focus. After considering possible alternatives I chose theses from the digital repositories of the University of Tartu (UT) and Massachusetts Institute of Technology (MIT). In order to ensure that the differences would mostly be those of academic cultures and not of academic disciplines I chose theses from the same discipline, economics. A clear advantage of analysing such material is the fact that the corpus is authentic (i.e. it has not been created with the present study in mind) and thus gives a realistic idea of the writing conventions applied in two academic cultures.

Among the many considerations to be noted when studying the papers was the question of the largely unknown background of the authors (e.g. their first language or prior contact with

academic English). We can, however, make the assumption that if an author has reached the stage of writing a Master's thesis, they are probably subject to the norms and linguistic traditions characteristic to the university they are studying at. In other words, this thesis compares academic traditions rather than foreign or native speakers of English.

Research in the field of academic discourse is anything but new and papers involving Estonian from such perspective have been written as well (e.g. Rummel 2005, 2010, Helstein 2004). Helstein wrote a Master's thesis in which he compared theological academic articles of Estonian, English and German authors. Similar studies have also been carried out in other countries (e.g. Kreutz and Harres 1997) and to some extent their results can be used for comparison in the present thesis as well. Rummel's theses are also of considerable importance as she has studied the writings of students, in addition to the writings of academics, from the perspective of coherence and cohesion. However, none of the aforementioned papers discussing Estonian writings have concentrated specifically on student writing or used computer-based text analysis. This paper aims to fill this gap to some extent, as the corpus used consists of the theses of graduate students and will be subject to computer analysis.

The thesis does not aim to show that there is a difference in how people write English as that is somewhat obvious and has been done before (Dahl 2004, Kaplan 1966, Povolná 2012 etc.). Instead, it outlines the main characteristics of the English used by Estonian English speakers comparing it to the English used by people from a university where it is the working language. More specifically, I aim to see if the authors in the UT corpus could be clearly distinguished from those in the MIT corpus in terms of hedging and cohesion/coherence.

Should I succeed in describing the differences, the results might later be used to create analysis software that could assess the English used in a text using a set of predefined criteria.

Even though this thesis alone could not offer sufficient data, it might be seen as the beginning of a bigger project. Not only because of the results that could be used in writing the program but also because of the methods used in gathering the data. If the program was web based and public, it would be easy to expand the corpus by including works by volunteers after they have been briefly reviewed. Although such software could offer advice rather than definite rules for text creation it might still prove useful, particularly if fine-tuned for discovering the possible patterns characteristic to Estonian writers.

More importantly, the results of this study might also prove useful for anyone interested in carrying out similar research as the thesis covers one possible approach to text analysis and offers additional material on how authors influenced by the academic traditions dominating in Estonia use academic English and adapt to the Anglophone tradition.

There are five main parts in the thesis:

The introduction offers an overview of the intentions of the author and the initial choices made when planning the thesis. Chapter 1 introduces the key concepts used in the thesis, offering an overview of the notions of hedging, cohesion and coherence. Chapter 2 introduces research that has been carried out before in the field of comparative rhetoric and discusses some more relevant works more thoroughly. Chapter 3 contains details on the empirical analysis carried out for this study and discusses the findings and their possible implications. The conclusion contains discussion on the thesis and the results reached.

1 CHAPTER I: KEY CONCEPTS OF THE THESIS

Texts can be approached from many angles and discussing all of them in one paper would be impossible within the limits set for an MA theses at the University of Tartu, which is why the field of study was narrowed down to cohesion, coherence and hedging. Such a choice stems from my belief that neither hedging nor cohesion or coherence are concepts that could be conclusively defined, which makes them hard to learn without extensive practice. In addition, the way they are applied (if at all) differs greatly by academic cultures. As shall be demonstrated below, this belief is supported by other authors as well, which is why I expect these particular aspects to be heavily influenced by the academic background of an author and believe that the specific differences between the theses from different cultures can be rather informative in describing what makes some texts seem more appropriate in English than others.

1.1 Hedging

There are many ways to present a statement. We can say that X is true, we can say that we believe X to be true, and we can say that X might be true. Although the core of such claims is the status of X, there is a difference in the effect they have on the reader and in the position they place the author in. And there is much more to shaping the reader-writer relationship than different modals. To look at just one additional example: the hypothetical fragments one could use for introducing their work might be: 'A study was conducted in order to...' and 'I conducted a study in order to...' It is such tiny details that will eventually make a big difference in how the reader understands what we are saying. In the first fragment the author weakens the connection between themselves and their research, whereas the second fragment clearly states that it was the author who was behind the study.

1.1.1 Defining a Hedge

When the hedges were first introduced by George Lakoff in 1973, he described them as 'words whose meaning implicitly involves fuzziness – words whose job it is to make things fuzzier or less fuzzy' (Lakoff 1973: 471). Even though the central concept has remained largely the same, the possible classifications of hedges have grown rather complicated since then so that the term has by now, in treatments by various authors, come to denote things from commitment to truth value to devices used for encouraging debate (Kreutz and Harres 1997: 182-184). A possible way of classifying hedging devices 'at a lexico-grammatical and syntactic level' referred to by Alonso et al is that of Martín (2003). It distinguishes four main subtypes of hedges that can be described as follows (Alonso et al 2012: 51-53):

1. Devices used to indicate that a statement is not to be taken as definite and final but rather as something uncertain. Examples include indicators of epistemic modality, such as modal verbs related to possibility '(may/might/can)', as well as verbs related to cognition (e.g. 'seem to') or the probability of a statement being true (e.g. 'to assume') (Alonso et al 2012: 52). Additionally there are modal adverbs (e.g. 'possibly'), modal nouns (e.g. 'suggestion') and modal adjectives (e.g. 'probable') that can be used to a similar effect (Alonso et al 2012: 52). The same subtype also includes approximators (e.g. 'generally', 'frequently') (Alonso et al 2012: 52).
2. '[M]etalinguistic operators', such as 'in fact, generally speaking' etc. (Alonso et al 2012: 53), the aim of which is to attempt to shift the attention and hence possible criticism of the reader from the proposition expressed.
3. '[P]ersonal pronouns followed by verbs of cognition' used for making a proposition seem subjective to give the reader a chance to agree or disagree. The category also

contains phrases used to express the doubts of the author(s) (e.g. 'to our knowledge'), as well as expressions meant to make the reader believe that what is being said is considered important, judging by the way it is done (e.g. 'extremely interesting') (Alonso et al 2012: 53).

4. Constructions used to make propositions impersonal so that their connection to the author becomes less evident (e.g. 'the results suggest' (Alonso et al 2012: 53) that creates the impression as if the author had nothing to do with whatever is suggested).

Despite such classifications, it is not easy to say what exactly can be considered hedging. From a linguistic perspective one could say that hedging is the process of softening statements for whatever reason. However, different authors have offered other definitions that sometimes differ greatly. For example, some say (while also pointing out alternative views) that words used for amplifying statements (e.g. 'particularly important') also classify as hedges (Martín in Alonso et al 2012: 53). Such hedges are not included in the present study.

Irrespective of which subtype we study and which classification we draw on, there is a rather obvious but complicated factor that is to be considered when talking about hedges. It is the background of the author as the use of hedges varies greatly across cultures. For example, Chinese L2 (Alonso et al 2012: 49), as well as Finnish EFL writers, (Ventola 1997: 174) appear to use stronger, i.e. more authoritative, modals than English writers

Moreover, predicting how a hedge affects a reader is complicated, to say the least. In order to realise where the complexity stems from we can consider an example given by Lewin (2005: 167) who asked a group of twelve authors, nine of whom were native speakers of English, to mark what they considered 'downtoners' (a possible branch of hedging) in their

(published) articles and give reasons for using them. Somewhat surprisingly, it turned out that what the authors considered a downtoner was often quite different from what linguists or a possible reader (Lewin 2005: 170-172) recognised as one. For example, half the occurrences of modality were not considered as downtoning by the authors and the vast majority did not consider using personal attribution ('I think that X is true') as downtoning either (Lewin 2005: 167, 169). That is, although we can describe a text by the hedges in it, we should be careful in our assumptions about the actual intentions of an author. Namely, authors who were asked about the motivation behind their hedging implied that they were trying to remain true to the facts rather than to express uncertainty or politeness towards the reader (Lewin 2005: 169-170) as is often thought.

Studies have confirmed that the use of hedging devices varies considerably depending on whether the person writing the text uses English as a native language or not, at least as for modal verbs. This is not only to say that EFL speakers necessarily use fewer or more hedging devices than those representing the English-language tradition but that they seem to differ in their choice of the words / phrases to use for doing so, which also affects the strength of the hedge (Ventola 1997: 174). English has a rather diverse system of modal verbs, so that a writer can present their claims in different degrees. However, such 'secondary function' of modals can be rather ambiguous. Even though we can talk about a hierarchy of modals, it cannot be defined exactly as it 'varies according to situation' (Longman English Grammar 2007: 208). It seems that it is hard for non-native speakers of English to comprehend 'the pragmatic value and force of the English modality system' (Pérez-Llantada 2013: 252).

Hedges in general depend on the situation as well. According to Kreutz and Harres (1997), they can, for example, be seen as 'approximators' that change 'the propositional content' or as

'shields' that 'indicate degrees of speaker commitment' depending on the context they appear in. To prove their point, Kreutz and Harres (1997: 183) consider two different uses of the phrase 'sort of', noting that while it functions mainly as a shield in the sentence 'I was sort of disappointed with him', it functions as an approximator in the sentence 'It has sort of a green hue'. The examples are also worth considering because they give us an idea of the diversity of the possible uses for hedging; finding such phrases from an academic text is highly unlikely but finding a phrase the purpose of which is the same is very likely; the hedges used in EAP (English for academic purposes) are just different.

While hedges and hedging is too broad a topic to lend itself to manual analysis, the scope of this thesis is limited to lexical elements of hedging in that regard. The choice can also be justified by the fact that these elements seem to be the most frequently used means for hedging (Varttala 1999: 183). The aim is to examine the way hedges are used in the MIT and UT theses and to describe the peculiarities. Except for the above mentioned limitation, the criteria set for a hedge in this thesis are quite loose so that no word classes are preferred. After all, one can successfully use (modal)verbs, adverbs, adjectives etc. to hedge.

1.2 Cohesion and Coherence

Due to the complex nature of cohesion and coherence, this thesis aims to see whether and how authors differ in their usage of different linguistic means for making their text easier to comprehend.

A part of the writing process that may be considered as important as using hedges involves making what you write into a whole instead of offering many unrelated facts in a row. While attempting to do so seems self-evident, it is not that easy. Even though the terms 'cohesion' and 'coherence' appear to be similar, they are rather different. According to McNamara,

'cohesion' means 'cues that are directly observable in a text', whereas 'coherence' refers to 'connections that are formed in the mind of the reader or listener' (McNamara 2013: 590). Palmer (1999), who has also studied the field of coherence and cohesion, highlights the difference as well, also drawing on various other authors. He points out that if we look at what has been said by Widdowson, we can conclude, in spite of some differing views, that cohesion is not a necessary condition for coherence as we are sometimes perfectly capable of understanding texts that do not contain any cohesive links (Widdowson in Palmer 1999: 67). The opposite situation, however, leads to nonsense. Just because a text contains certain word patterns and even grammatically valid sentences does not mean it is coherent. There have even been successful attempts in developing text generators that could fool automated essay-grading software by compiling texts based on the keywords they are fed. A text generator that was given the keyword 'privacy' came up with the following sentences: 'Privateness has not been and undoubtedly never will be lauded, precarious, and decent. Humankind will always subjugate privateness.' (Kolowich 2014: para 2). Although the first sentence contains a clear link, it is absurd as a whole. Still, even though we can see the absurdity, it would be wrong to say this is due to some universal definitions of coherence and cohesion. In fact, cohesive devices need not always be explicit. It may well be that the opposite holds true, so that coherence is achieved without using explicit cohesive devices (Crossley et al 2014: 188), which makes the assessment of cohesion and coherence even more complicated. What is more, even achieving cohesion, whatever the means used to do so, need not necessarily ensure a text of higher quality (McNamara 2013: 580). We can describe some extremes and make a basic distinction between cohesion and coherence but the topic is at least as ambiguous as that of hedging and, as it turns out, quite context dependent as well. However, the nature of this thesis sets the focus to cohesion as this is something that can be studied by working with

corpora. The reason why the terms are often used together throughout the thesis is that even though I cannot offer a valid analysis on coherence, I still find it important to treat the notions together to emphasise that the two are often connected. Cohesive devices studied will include lexical bundles (LBs) and discourse markers (DMs).

Although we can describe and apply cohesive devices, it is hard to identify universal rules for doing so. A text that makes sense for one person may seem incoherent to another. This is why the unidimensional approach to cohesion has been criticised (McNamara 2013: 586-587). Despite the fact that we can conclude something about the cohesiveness of a sentence by the time needed to read it or by the number of lexical overlaps it contains, this is still but a part of a bigger picture when it comes to determining how well a text is understood (McNamara 2013: 580-581). Firstly, as McNamara notes, there is the fact that the potential readers have different skills of text processing. Second, which is even more important, every reader has a different intellectual background concerning the material they read (McNamara 2013: 582-583). Clearly, an author could try to make sure that everyone would understand their work by providing sufficient amounts of information but such an approach could never work; not just because there are limits to the resources of every writer but because the more informed readers may not reach the possible connections with the knowledge they already have if the text is too cohesive (McNamara 2013: 584).

This implies that a writer must not only know the discourse they are writing in but they must also be able to sense their target audience. This is actually much more than just perceiving whether the potential reader will be an expert or not; even an informed reader may need certain triggers that would help them to realise that they actually know what the author means (cf. McNamara 2013: 582). A possible example to give here is that of drawing

parallels; if a reader has knowledge about X, they may need the help of the author to realise that X is actually very similar to the Y they are reading about. All in all, as Rummel puts it, we can say that both the reader and the writer affect the nature of a written discourse (Rummel 2010: 24).

2 CHAPTER II: LITERATURE REVIEW

Texts range from a few sentences to lengthy novels, from easy reading to complex academic writings and hence the expectations we have about them vary. They are described by using a system of genres in which each genre has a set of distinct features. From the perspective of academic discourse, genres can be described as follows: '[they] are conceived of functionally as configurations of expected forms and meanings that a given community associates with particular situations and acknowledges as distinct from each other' (Chovanec 2012: 5-6).

The process of text creation is nearly always subject to some kind of rules dependent on the genre (e.g. novels, short-stories, poems etc. in the case of fiction). While the line between them is often somewhat vague and subjective, it is nevertheless there. In the case of academic texts things are no different. There are certain rules writers of such texts must adhere to if they aim for comprehensibility and credibility. To some extent the requirements set can be generalised but there are aspects that certainly differ by the culture, language or discipline an author is operating in. It can even be said that in the case of academic texts the influence the norms set by a particular discipline have on a text is sometimes no smaller than that of the author (cf. Dahl 2004). The latter is obviously more or less free in saying whatever they desire but there are various written and non-written rules as for how they are expected to do this. One

of the most obvious and widely known examples of such requirements is that of verification in the world of science but there are many more ('authoritative stance', 'stepwise logical argumentation', etc. (Snow and Uccelli 2009: 117)).

2.1 Culture and Discourse

The language an author uses is also important as this is what determines the main linguistic means and writing conventions one can, or is more likely to, use. Language, in turn, is related to a culture and certain expectations that stem from it. While we often readily acknowledge that there are differences, for example, between the USA and China in that regard (cf. Mao 2003), the issue is in fact more complicated as differences can also be found in languages and cultures that are related more closely or even within the same linguistic community but affected by class or age, for example. Such local variation may seem to be of little importance, especially if we consider the fact that during recent decades English has established itself more and more as the lingua franca and thus other languages are used less in the academic world. Yet, the traditions the authors follow may not have changed as fast as the language they use; a non-English author writing in English can often be identified as such. Not only because of the vocabulary or non-native constructions but also because of the general style and writing conventions they have acquired using their mother tongue.

In spite of the fact that different fields, different languages and different people often use different ways of writing, there are also linguistic devices and techniques that are used in almost every genre and that make up what we know as rhetoric. According to Lloyd Bitzer, a rhetorician who created the concept of rhetorical situation¹, genuine rhetoric can only be

1 '.../ the nature of those contexts in which speakers or writers create rhetorical discourse.' (Bitzer 1968: 1)

created in an actual situation, in which it can be used for changing reality (Bitzer 1968: 4). He adds, though, that it would be more precise to say that a particular discourse comes into being because a certain situation calls for it (Bitzer 1968: 2). It has to be said though, that the whole concept of rhetorical situation has been criticised as well. For instance, Richard E. Vatz believes that we should not over-estimate the power of a situation. According to Bitzer, it is the situation that calls for a certain response by the rhetor or has 'certain /.../ characteristics' but Vatz says that the way a situation is perceived depends on the perceiver; the choices on what to do, what to convey to the audience, lay with the one who acts in a situation, which in case of rhetoric means the rhetor. There are no meanings in a situation, neither can it call for certain actions (Vatz 1973: 155-157).

Regardless of whether we see the choices of an author as affected by the situation as such or believe them to stem from the goal the author has, it is clear that people usually follow certain rules in certain rhetoric settings. Vatz's idea of the greater responsibility of the author makes sense. It would be hard to imagine a situation or 'exigence' containing the appropriate or correct action to be taken in itself (Vatz 1973: 156). Even though an author writing for a certain audience is likely to be influenced by it, the choices they make depend not so much on the situation as on what they decide to do about it. However, it would be unfair to say that Bitzer denied the influence of the rhetor. He does admit that an author brings a set of constraints with them; it is just that he seems to under estimate the overall importance of the choices made by a rhetor.

When it comes to academic writing, choosing sides in this debate is not easy. I agree with Vatz, on the importance of choices made by the author, yet our choices, however free they may be, serve a certain purpose and are thus affected by the traditions involved. For example,

the aim of writing a thesis is, to put it very simply, to get a degree. If we want that to happen, we are inevitably bound to make choices keeping our goal in mind, which means that the choices follow certain norms. These norms, however, vary greatly across cultures.

2.2 Comparative Rhetoric

One of the first to look at such differences across cultures was Robert B. Kaplan who analysed the structure of paragraphs in about seven hundred works by EFL students in the 1960s in order to see how they differed from those written by native speakers of English (Kaplan 1966: 15). Kaplan starts his discussion by drawing our attention to the fact that the way people think and use their language can differ by culture and that there is a connection between the two. A simple example he uses involves a European and an Eskimo. While the former sees things happening as a result of their will, an Eskimo believes that things and phenomena happen **to** him (Kaplan 1966: 12).

Kaplan found rather remarkable differences in his study. He implies that knowing a language is not only about knowing the words and grammar. It is about acquiring a way of thinking (Kaplan 1966: 20). An example provided by him involves the way people speaking Arabic develop their paragraphs. Their style seems to involve many parallel constructions with a greatly varying level of complexity. For an English speaking reader, such a text might seem overtly 'archaic' and ambiguous (Kaplan 1966: 15). In the case of the French, Kaplan points out that we can see quite a bit of what an English language reader would consider digression. Furthermore, he notes that it may have to do with both, the speciality of the writer and the peculiarities of French. In the context of this thesis it is worth noting that Kaplan also hints at writers transferring their native style of writing to another language when they are using it (Kaplan 1966: 18).

While Kaplan's work has been criticised by many for presenting English as a universal norm (Matalene 1985), not making a clear enough distinction between 'thought patterns' and 'rhetorical patterns' (Severino 1993) etc. and his cultural generalisations are somewhat outdated, he can be considered a person whose initiative has by today led to comparative rhetoric as a separate field of study (Mao 2003: 402). Briefly put, it has been defined as '/.../ the cross-cultural study of rhetorical traditions' (Kennedy 2006), the aim of which is to compare different traditions and reveal their peculiarities that might not be noticed otherwise. The idea took root in the second half of the 20th century when some scholars started to express the concern that the study of rhetoric drew too much on the Western tradition, creating the false impression that rhetoric as such could only be Western (Kennedy 2006). Such attitude can be seen in a statement by John Morrison, according to which 'Japanese culture does not have the necessary institutional ingredients to nourish a rhetorical tradition' (Morrison (1972), cited in Mao 2003: 407). In addition to highlighting the differences, comparing obviously allows us to see the similarities, meaning that comparative rhetoric can be treated as an attempt towards defining the core of rhetoric as such regardless of the culture and society (Kennedy 2006).

Clearly, the cultural differences I expect to find conducting the analysis in my thesis are likely to be less drastic than those described, for instance, by Kaplan just because the cultural variations between Estonia and the Anglo-American world are probably somewhat smaller than the ones he dealt with. Still, this is not to say that Estonian and Anglophone cultures are alike; there are rather important possible differences between cultures when it comes to writing (cf. Kaplan 1966: 12) and thus they can be expected also between Estonian and Anglophone cultures. These differences will be introduced in the following sections. At that, it is worth to keep in mind that due to globalisation and advances in using computers for

linguistic analysis, comparative discourse analysis has by today gone through significant developments. Today's studies do not usually make broad generalisations like Kaplan did but offer insight into more specific topics be it the use of politeness strategies, LBs, hedging devices or something strictly characteristic to a particular corpus.

What makes considering the differences complicated is the status of English today. Since it is the lingua franca, the number of people using it is huge and speaking of standards has become complicated. English has its norms and grammar but the English academic discourse should not be associated with native speakers only; even though British and American English can still be considered the yardsticks against which the 'linguistic proficiency' of English language academic writing is measured, it is not always easy to say what to do about the culture related characteristics of a text written by a non-native author. In fact, we seem to be facing a double edged sword. On the one hand, as Chovanec points out, referring to Swales (1997), that the huge popularity of English may eventually impoverish (the academic discourse) of other languages (Chovanec: 2012: 6-7). Crystal adds that English may even replace many other languages (Crystal 1999: 8). On the other hand, the dominant position of English means that it is becoming harder and harder to say whether we should condemn the non-standard English used by non-native speakers (Chovanec 2012: 7).

The nature of a discourse is not rigid but rather flexible (Chovanec 2012: 9), especially in the humanities where there 'may be not be a single pattern or formulaic structure that could be used as a mould /.../ in the case of articles or other written genres' (Chovanec 2012: 7, cf. Cooley & Lewkowicz 1997: 121). The writing experience we have gained when writing texts in our native tongue may not be of much use to us when trying to address an international audience, as its expectations differ from those of the local audience (cf. Rummel 2010: 27).

What we know about writing for a local audience may even work against us, should we have our work translated instead of writing it in the language it is to be published in (Chovanec 2012: 14). However, Chovanec draws our attention to the fact that a native speaker may not always have an advantage here either. It can happen that a paper submitted by a native speaker gets sent back due to deficiencies in language use and the editor suggests it should be proofread by a native speaker (Chovanec 2012: 7-8). Such misunderstanding can hardly be due to cultural differences but might as well, according to Chovanec, be a question of 'personal academic style /.../' instead (Chovanec 2012: 13).

2.3 English Academic Writing

Irrespective of the field they study or work in, a native speaker may not have as many advantages compared to an EFL writer as is usually thought. They may have less trouble with grammar and sentence structures but, as noted by Duff, this in itself does not make them proficient in creating academic texts or delivering presentations (Duff 2007: 01.5-01.6).

Although schools and universities demand that students use what is called 'academic language' in their writings, relatively little empirical research has been done on academic writing when it comes to native speakers of English (Snow and Uccelli 2009: 113). The reason for that may be that the idea of academic writing is hard to describe due to the lack of clear-cut universal rules.

Despite the difference between what is written at secondary schools and what people are expected to write at a university, it is the secondary school that introduces us to academic writing. It is often there that we get the first experience of reader feedback and are told to mind our language by refraining from using too colloquial expressions and to make sure that our arguments follow a logical pattern (Uccelli et al 2013: 39-40). At some point we are asked

to change the descriptive style characteristic to children for a more argumentative one. Doing that means the focus of the text will shift from the feelings and thoughts of the author to more objective claims the writer is expected to verify.

This is not to say that authors are to remain entirely unnoticed. However, they must present themselves and what they say 'under particularly challenging conditions' (Snow and Uccelli 2009: 122). As for expressing one's opinion, an author is expected to use 'stepwise argumentation' (often achieved by using special linguistic constructions) and hedging. In short, an author is expected to sense the limits of what they can state and to word the statements accordingly (Uccelli et al 2013: 40).

Knowing how to speak or write academic English is important, whatever the field (Snow and Uccelli 2009: 114). Sentences in academic writing are expected to be dense, containing only the most important (Snow and Uccelli 2009: 118) so that the reader can get the information as easily as possible. As academic language is mostly used to speak about things that are considerably more complex than what we encounter in our everyday life, be it either due to lack of background knowledge, abstract concepts or something else, it is but natural that the language cannot be simple (Snow and Uccelli 2009: 122-123). The complexity should, however, never be associated with the complexity of comprehending the sentences but stem from other aspects. For example, Snow and Uccelli refer to Halliday (Halliday (1994) in Snow and Uccelli 2009: 120-121), who notes that in the case of academic language we are expected to see a broader picture as for the possible things words can denote as while children associate nouns with things and verbs with processes etc., academic language contains 'grammatical metaphors'; nouns to refer processes, verbs refer to relationships, etc.

Academic writing is also characterised by the fact that we have to assume as little as

possible about the reader. Furthermore, even if we know precisely who will be reading our text, the writing itself should be such that it relies little on what the reader knows (Snow and Uccelli 2009: 124). True, the requirement has its limits, set by the genre of the text that has an effect on the words an author uses (McNamara 2013: 589).

In addition to these and other characteristics of academic writing, there is still the question of text quality or, more precisely, what determines it. A possible example to consider involves essays by last year secondary school students (Uccelli et al 2013: 42). Uccelli et al found, for example that in the case of stance markers it was not the diversity but overall frequency that could be positively linked to the quality of writing. Similarly, it appears that the general complexity of sentence structures,² and the diversity of the words used seem to have a positive effect on the quality of the essay or the grade it receives (Crossley et al 2014: 189, McNamara 2013: 591). However, the authors admit that it would be wrong to draw definite conclusions from such findings due to the limited corpora and the fact that contradictory claims have been presented as well, for example about the connection between organisational markers (such as causal connectives) and writing quality (Uccelli et al 2013: 55). Rogers (2004: 141-142) has also pointed out that following recommendations often given for academic writing, such as placing topics at the beginning of a meaningful unit does not seem to be in positive correlation with writing quality, whereas delayed topics seem to show the opposite correlation. All this implies that there are no universal formulae for successful writing, rather only general guidelines.

A more thorough insight into the matter has been given by Crossley et al who conducted a computerised statistical analysis of various student papers written by students in the ninth and

2 The sentence as such must be easy to understand despite the complexity of the structures.

eleventh grades as well as by college freshmen. The authors focus on proving that considering grading and writing essays a linear process is somewhat narrow minded (2014: 185, 189, 202) and this is something worth keeping in mind in the case of theses as well. They do not argue with the belief that there seems to be a set of properties by which the success of an essay writer can be predicted. For instance, a writer who is well versed in the language they are writing in can focus more on the 'argumentative and rhetorical' structures (Crossley et al 2014: 187). Similarly, evidence suggest that successful writers write longer essays that are more sophisticated in syntax and rhetorical structures (Crossley et al 2014: 188-189, McNamara 2013: 591-592). However, they also point out that there are many ways for writing a good essay (Crossley et al 2014: 189).

Another fact relevant in the context of this thesis is the way they classify successful essays into categories as a result of their analysis. In short, all categories (four in total) exhibit considerable variation and only one of them is described as 'academic', containing 'little causal cohesion, /.../ few explicit cohesion devices, and /.../ few pronominal references'. This does not mean that essays from such a category are not coherent, just the means for achieving the coherence are different (Crossley et al 2014: 202-203). Even though one can always ask what exactly qualifies as a what the authors refer to as academic, such findings have significant implications. First, the authors imply that there are indeed properties and texts that can be described as academic. All the same it seems self-evident that an editor reviewing an article or a lecturer assessing a thesis does not count the cohesive devices in a text but judges the quality of what they read by whether they find it comprehensible.

2.4 Academic Writing of Non-native English Speakers

As pointed out by Rummel (2005: 79, 2010: 22), non-native users of English usually have many other things to worry about, such as basic grammar, when it comes to writing so that they often do not think about English written discourse in general. This does not mean that EFL writers are completely unaware of the aspects of text creation; after all, they exist in other languages as well, the question is what happens if we try to transfer them. The result of such transfer is often an odd mixture of languages; a variation of English with peculiarities in stance, modality, overall argumentation and many other aspects (Pérez-Llantada 2013: 251-252).

An example to look at involves DMs, which are a means for guiding the reader and can hence be attributed considerable relevance in adding cohesion to academic writing. In an ideal world one sentence follows another and the reader can instantly grasp what the author has in mind. Thus, we often feel the need to emphasise a connection or contrast just to make sure our line of reasoning makes sense for the reader. The way casual and contrastive DMs are used in the Master's theses of Czech students of English has been analysed by Povolná (2012). I believe her findings to be relevant as there seem to be considerable similarities between the English texts of Czech and Estonian writers. It has to be said, though, that the focus of Povolná's thesis is not on drawing parallels with English speaking writers as much as it is on describing how Czech students use DMs in English. According to Povolná, who refers to Biber et al (1999), DMs are particularly relevant in academic writing where both reasoning and supporting one's claims are of great importance (Povolná 2012: 133). The focus of Povolná's analysis is on causal and contrastive relations occurring at the 'clausal and higher levels of discourse' where the 'markers relate two separate messages, hence functioning as

DMs.' She omits the analysis of relations expressed at a 'lower level' where the markers are 'conjunctions within single messages'. Her example of a marker functioning as a DM (the kind she studies) is as follows: 'The spoken and written language of the Britons gradually altered its form both structurally and lexically. **As a consequence**, a great amount of English word stock is of Latin and French origin.' Her example of a conjunction is: 'This attitude refers to Cato's forfeit of God and truth and **in consequence** his damnation.' Both sentences are authentic, i.e. from the corpus she analyses. While the first example is about linking two different thoughts, 'in consequence' in the second sentence is not used to introduce new ideas (Povolná 2012: 134).

Povolná points out, citing Fraser, that although morphologically the markers belong to specific classes, being mainly either conjunctions, adverbs or prepositional phrases, they do not act as members of their class 'but instead are separate from the propositional content' (Fraser (1990) in Povolná 2012: 134) and have a procedural meaning instead of a conceptual one, so that removing a marker does not change the content of discourse segments but makes it harder for a reader to reach the interpretation intended by the author (Povolná 2012: 134). Syntactically Povolná distinguishes between discourse markers expressing paratactic (e.g. 'as a result', 'thus') and hypotactic (e.g. 'because', 'since') relations (Povolná 2012: 135, 137).

Summing up the results of her study she arrives at a conclusion similar to what Dontcheva-Navratilova concluded about LBs (Dontcheva-Navratilova 2012: 56). EFL students (novice writers) are rather limited in their use of causal and contrastive markers (Povolná 2012: 135), although she adds that there is a difference here between causal and contrastive markers. More precisely, the set of contrastive DMs novice writers of academic texts use is broader than their set of causal markers. A possible simple explanation offered by her is the fact that there are

just more contrastive markers than there are causal markers (Povolná 2012: 139). She also finds that in the case of causal discourse markers, explicit hypotactic relations ('as', 'because', 'since') are marked somewhat more clearly in academic discourse than explicit paratactic relations, which she found to occur rather infrequently across the corpus as only 8 of the 15 paratactic markers she searched for occurred with a frequency worth noting (Povolná 2012: 136-137).

In the case of contrastive DMs, Povolná finds, contrary to her initial belief, that expressing relations using hypotactic markers is 'considerably less frequent' than doing so using paratactic markers. It turns out that only fifteen out of the thirty eight contrastive markers studied appear in the whole corpus fifteen or more times and only six of the fifteen are hypotactic. Additionally, Povolná points out that the paratactic marker 'but' (566 occurrences) appears about as often as all the other paratactic markers together (586) and is 'almost six times more frequent than the most common hypotactic marker 'although' (101) /.../' (2012: 138).

Povolná's study also confirms that there is a clear difference in how students from different fields use DMs. The repertoire of those with a background in linguistics seems to be rather broad but they tend to overuse some discourse markers. The most drastic example involves a contrastive or causal marker in nearly every second sentence (Povolná 2012: 142-143). Despite field specific trends, she also notes that there can be considerable variation within a field, which she attributes to the individual knowledge of the students as well as the influence of their supervisor (Povolná 2012: 136).

Another linguistic instrument closely related to the coherence of a text is called a 'lexical bundle' (LB) (also known as 'fixed expressions', 'lexical phrases', etc.). In short, LBs are 'the most frequent recurrent sequences of three and more words in a register or genre' (Biber and

Barbieri (2007) cited in Dontcheva-Navratilova 2012: 38) that form an important part of a discourse and support the hypothesis first introduced by Sinclair, according to which the way speakers and writers of a language think is phrase-based rather than word-based (Dontcheva-Navratilova 2012: 38). In other words, as noted by Hyland (who relies on various other authors), what 'we know about a word' depends on our previous experience with it, which means that when we try to say something, the 'wordings we choose' are similar to those we have encountered before (2008: 6). This explains the importance of LBs in academic writing; it has been found by multiple authors that they add a touch of predictability to the text making it easier to comprehend and process (Hyland 2008: 5, Simpson-Vlach and Ellis 2010: 487). Thus, knowing which combination to use is an important part of knowing a language (Simpson-Vlach and Ellis 2010: 488-489).

LBs in the context of academic writing have been studied, among others, by Dontcheva-Navratilova who used a corpus of English Master's theses written by Czech students of English. The aim of her study is somewhat similar to that of this thesis as her main goal is to see if and how Czech students of English use the LBs characteristic to expert academic discourse (Dontcheva-Navratilova 2012: 37). Different studies have used different criteria of frequency³ when analysing texts for linguistic bundles (Dontcheva-Navratilova 2012: 39). It is worth noting that studies often set an additional criteria regarding the number of texts a bundle has to appear in to make sure that it can really be considered a generally used bundle and not an idiosyncratic characteristic of a particular author (Simpson-Vlach and Ellis 2010: 488-489).

Although the number of words considered a bundle may vary, Dontcheva-Navratilova (2012: 42-43) studies four word bundles and works with bundles preselected on the basis of

3 Dontcheva-Navratilova refers to frequencies from 10 to 40 per million words (Dontcheva-Navratilova 2012: 39).

earlier research.

Dontcheva-Navratilova says, referring to Biber and Barbieri (2007) that LBs can be clearly distinguished from other formulaic expressions as they are not idiomatic, 'lack perceptual salience' and do not usually represent complete structural units like most idioms do (Dontcheva-Navratilova 2012: 39). In other words, the meaning of an idiom usually cannot be devised from the idiom itself, whereas the meaning of a bundle is usually self evident (Wei and Lei 2011: 156), such as 'it is important to'.

Dontcheva-Navratilova divides bundles into functional types and examines how they are used by Czech Master's and doctoral students in their theses. She also mentions that classifying the bundles is everything but clear-cut as LBs are often multifunctional (Dontcheva-Navratilova 2012: 40). Although the classification and naming conventions differ, the general meaning behind the categories seems to be the same in case of different authors. The following are some examples of the divisions Dontcheva-Navratilova (2012: 40-41) uses:

1. 'Referential bundles', that 'express ideational meanings related to the representation of reality', such as 'at the end of the', 'a little bit of', etc.
2. 'Discourse organisers' (term by Biber & Barbieri (2007)) that 'convey textual meanings concerned with the organisation of a text and the development of argumentation' (e.g. 'on the other hand', 'as a result of', 'as shown in Figure', etc).
3. 'Attitudinal bundles' that 'convey interpersonal meanings', such as 'the fact that the', 'are more likely to', 'it should be noted that' etc.

Dontcheva-Navratilova concludes that Czech university students tend to keep to the devices they have been taught at school and do not usually use discourse devices specific to their field,

meaning that the 'soft' and 'hard' fields are not distinguished as clearly as in the case of native writers. That Czech students are more limited in using LBs and tend to write in a way that expects more from the reader than native English speakers do (Dontcheva-Navratilova 2012: 50) is not particularly striking if we consider the cultural differences in the respective traditions of academic writing but there are some more interesting aspects that Dontcheva-Navratilova mentions. For example, it turns out that there are clear differences in the distribution of the functional types of bundles in different fields (Dontcheva-Navratilova 2012: 54).

According to her paper, those with a background in linguistics tend to use bundles more like native speakers of English when it comes to frequency and context. Yet, she also points out that according to previous studies, the former are also likely to over-use some phrases (Dontcheva-Navratilova 2012: 45) and still fall behind the rates of 'expert academic discourse' when it comes to using discourse organising and attitudinal bundles (Dontcheva-Navratilova 2012: 56).

Such speciality-specific characteristics have also been, among others, studied by Dahl (2004), who compared 180 academic articles in total, written in Norwegian, French and English, trying to establish what influences us more while writing, our cultural background or speciality. The fields she looked at were economics, medicine and linguistics (2004: 1814). She concentrates mainly on two categories: the phrases that say something about the text (e.g. 'The analysis now turns to...') and those that are meant to inform the reader of what the writer is doing (e.g. 'We conclude that...') (Dahl 2004: 1812-1813).

Dahl concludes that the discipline a text is part of sometimes seems to play an important role in what the text is like. Of the three disciplines compared, medical texts clearly have a

distinct structure and style spanning across languages. As their structure is so fixed, there is also less need for metatext that would guide the reader. The necessary information can yet be quickly extracted by anyone familiar with the standards (Dahl: 2004: 1819).

Although linguistics and economics had common traits, they were not that extensive. Texts in economics, for example, tended to have a general description of the outline in the introductory section (Dahl 2004: 1820). Again, this does not mean that discipline is always the only decisive factor. Dahl finds that in the case of texts written in economics and linguistics, the influence of national and cultural traditions has a part to play (Dahl 2004: 1882).

In addition to the professional background and the native language of a writer, general writing experience is also of great importance. Wei and Lei conducted a study in which they analysed the way LBs were used by advanced Chinese EFL students in Chinese universities in writing their doctoral thesis. The results were compared to the use of LBs by professional writers. The conclusion was that the use of LBs by students, as well as intermediate or advanced non-native learners, seems to be considerably different from the use of LBs by professional writers (writing in six recognised journals of applied linguistics) (Wei and Lei 2011: 158). Wei and Lei found that while the corpus consisting of theses contained 154 bundles, the 'control corpus' consisting of the works of professional writers contained only 87 bundles⁴. What is more, the corpora differed in terms of the most frequent bundle. The most popular bundle used in the theses was 'the present study' but in the articles the most frequent bundles was 'on the other hand'. In short, learners used considerably more bundles than professional authors (Wei and Lei 2011: 159).

⁴ The criteria set for a bundle were different due to the different number of texts in the corpora (Wei and Lei 2011: 159).

Hyland (2008: 9), who used a mixed corpus of MA/MSc theses, PhD dissertations and articles from various disciplines, also found 'on the other hand' to be clearly the most frequently used bundle. Without knowing the exact ratio of different types of text in his corpus, it is hard to say, whether his finding might in a way contradict the claim by Wei and Lei, but it does give me reason to assume that 'on the other hand' will prove to be the most frequent bundle in the corpora used in this thesis. Even though the results by Dahl and Wei and Lei allow us to assume that the cultural background of an author, although important, may not be decisive when it comes to professional academic writing, there are two important things to keep in mind in that regard. First, the notion of 'cultural background' should be treated with caution, as it may not be as evident as it seems. There are studies that remind us that the way texts are written also depends on what is or has been going on in the society, maybe even to a greater extent than on the more fixed qualities / traditions of a culture (Shaw and Vassileva 2009: 301). Additionally, someone conducting a study usually has no means for determining the true background of the subjects of their study. Just because we know how someone is named and where they live does not mean we actually know the formative factors they have been subject to (family, beliefs, etc.) Second, we must not forget that both Dahl and Wei and Lei used academic articles and thus their material is not strictly parallel to the present thesis and its corpus of Master's theses. Hyland, however, says that MA theses, particularly those involving 'hard-sciences' seem to stand out by the large number of procedure-describing bundles in them, possibly because of the expectations associated with them, that is to demonstrate the ability of an author to use various 'research methods' (Hyland 2008: 15). Although economics is not a hard science, I still believe that the corpora studied in this thesis will also contain a considerable number of bundles related to research methodology.

2.5 Academic Writing in Estonian

For obvious historical reasons, the tradition of writing Estonian academic texts is rather short and hence it has not been extensively studied. There is some research, however, that allows us to make some general statements about academic writing in Estonian. In addition, we can draw parallels with other countries and cultures. Of those, German and Finnish (cf Rummel 2005, 2010, Helstein 2004) seem to be widely referred to but there are other cultures, such as Czech, that seem to make a valid basis for comparison in many respects.

The common problems Estonians face while writing in English have been studied most extensively by Rummel (Rummel 2005, 2010). While the central question of her work was different from that of this thesis, she conducted a survey in which she asked undergraduate students, academics and English language teachers to describe what they considered the most problematic fields (Rummel 2005: 58). The results can be considered rather telling. It turned out that even experienced academics often found sentence structure to be difficult to achieve. At that, it was generally believed that achieving cohesion and text organisation was not too hard and only about a third of the respondents considered a clear and simple style an important characteristic of a text (Rummel 2005: 69, 75). In fact, 75% of the academics found that 'text overall organisation' was not an issue (Rummel 2005: 72, 75). As such claims derive from a questionnaire in the case of which the respondents had to describe their own behaviour, this does not mean that there are no issues in the organisation of the respondents' writing, but only that they have not been trained to see textual organisation as a critical feature of academic writing and instead focus on sentence level problems.

Additionally, there is always the question of whether authors writing in their native tongue face fewer problems with style and cohesion or is the problem as sharp for them. Even though

it is hard to argue against the fact that an author writing in a foreign language must indeed deal with many additional issues they do not face in their native language, we cannot say for sure that the problems are language specific. The often poor quality of writing may rather be due to an important fact mentioned by Rummel herself. She notes that writing argumentative essays is not, contrary to the United States, for example, something generally done much in Estonian schools (Rummel 2005: 23). An Estonian student graduating from secondary school does have the experience of argumentative writing but it may indeed be often limited to the native language. What is more, it seems likely that the standards set at school are somewhat lower than those followed in professional academic writing. Yet, despite the difficulties in determining the source of such linguistic problems, Rummel (2005: 72, 75) notes that most of the academics seemed to believe that L1 writing was not that different from L2 writing, which made her to assume that transfer of L1 writing knowledge occurs. Similar trends have also been observed by other authors, such as Alonso et al (2012).

In comparison to Anglophone traditions, an Estonian author seems to be used to writing much more complex texts both in vocabulary and structure; we concentrate more on presenting facts than on presenting arguments or making sure the reader understands what we want to say (Rummel 2010: 60-63). Relying on various other authors and her own work, Rummel presents a list of factors characteristic to English texts written by Estonians. The most important of them, as far as this thesis is concerned, is the scarce use of links and other devices that are meant to help the reader, as well as lower audience awareness (Rummel 2010: 60-61). There are also non-linguistic factors that affect the coherence of a text. Cooley & Lewkowicz (1997: 20) draw our attention to the fact that according to supervisors, the most serious mistakes made by non-native English writers often involve lack of logic and argument development. It is true that such difficulties can be attributed to a different linguistic

background but there is also the chance that a writer has just lacked the necessary knowledge of their subject at the time of writing. The extent of one's knowledge on a particular field affects putting it all into words. Crossley et al (2014: 187), for example, suggest that having 'high domain knowledge' allows a writer to concentrate more on generating 'relevant ideas'.

Many of the claims made by Rummel confirm the observations made by Jana Chamonikolasová who analysed the English writings of Czech students. According to Chamonikolasová, Czech authors are more modest in their writings than their English counterparts. The difference can be described by the various lexical and syntactic means used. Czech writers may find the assertive phrases ('The analysis indicates that' etc.) used by Anglophone authors to be odd, as in their tradition it is not uncommon to find modesty even in titles (Chamonikolasová 2005: 9). We may assume that a similar kind of modesty may be present in English texts written by Estonians as Helstein also describes the phenomenon in his thesis, noting that compared to English and German texts, the ones written in Estonian contain about 30% fewer phrases meant for amplifying statements⁵ (Helstein 2004: 8, 33), such as 'it is clear that', to use an English example. Anglophone authors, on the other hand, might frown at an author for their habit of using the authorial plural instead of 1st person singular, which may seem wrong in English unless the text indeed has multiple authors (Chamonikolasová 2005: 8). The Estonian texts studied by Helstein seemed to be rather personal when compared to the others in that the authors did not use many passive constructions when actually referring to something that could be linked to specific persons (Helstein 2004: 21) (e.g. 'it has been said'). Such a finding contradicted what he had expected to find due to Estonian academic traditions (Helstein 2004: 32).

5 Helstein (2004: 33) notes that the results may reflect something characteristic to theology.

In the case of hedging, it all depends on what kind of hedging devices one looks at. Helstein observes that as ideational hedging devices are mostly related to objective uncertainty about the things said, it is of no surprise that he saw few differences between languages in that regard (Helstein 2004: 34). When it comes to interpersonal hedging devices, the main aim of which is to soften the effect of what is being said, not because the author thinks a particular claim could not be proved but because they want to express their respect to the possibly different opinion of a reader (Helstein 2004: 25). Helstein says that Estonian authors seem to be somewhere in the middle; they definitely do not hedge as much as English speakers but neither do they emphasise their claims. It is, however, worth noting that Estonian authors seem to use somewhat more of the ideational hedging devices, the aim of which is to give an author the chance to avoid criticism (Helstein 2004: 24, 34-35). Considering that the sample used by Helstein was rather small and specific, it would be wrong to assume that they represent more general truths. For example, Alonso et al suggest that German scholars writing in English hedge even more than native English speakers (Alonso et al 2012: 49). All in all, we might say that while native English speakers can often be cautious in how they say things they certainly do not tolerate vagueness in the broader sense, whereas the structure and argumentation of the English texts written by Estonians can be rather loose.

The observations made by Rummel and Chamonikolasová are also to be treated with caution. They present the findings of localised studies, rather than the complexities of real life. More than one author has said that things have already started, at least as for certain aspects, to change towards a more global tradition just because this is what the world expects (Chamonikolasová 2005: 6, Dontcheva-Navratilova 2012: 50).

3 CHAPTER III: EMPIRICAL ANALYSIS

In short, the aims of the analysis are:

- To see whether the theses written at the University of Tartu (UT) contain fewer hedging devices than those written at the MIT. The question derives from the supposition that Anglophone authors use hedges extensively.
- To describe the various techniques used for hedging and to see whether the MIT and UT theses differ in some respects.
- To see whether the UT theses differ from those of the MIT as for cohesion achieved using contrastive and causal discourse markers, as well as some LBs, and whether parallels could be drawn with similar studies on Czech student theses.

3.1 Data

As always with such analyses it is the corpus that sets the first limits. It can certainly be said that the corpus compiled for this thesis is relatively small but an attempt was made to create two clear, reliable and comparable corpora. Although there are many sources of Master's theses written in English, many of them also have various constraints that had to be taken into account.

The first criterion was that the theses had to be written at a university with a high enough reputation, to minimise the possibility of the thesis being either of poor disciplinary or linguistic quality. The other important criterion concerns the systems provided for searching the theses. There are many repositories on-line that contain theses but from which it is difficult to filter out a certain field or timespan of publishing. The third criterion that had to be considered was whether enough English papers had been written in the field in an Estonian

academic environment.

All these factors eventually made me choose theses from the Department of Economics at the University of Tartu and at the Massachusetts Institute of Technology. Only theses that have been written in the 21st century and were available in the public on-line databases of the two universities were included in the corpus. The number of texts in the databases was relatively small. With no other limitations applied, both corpora amounted to seven theses in total. Due to its size, the resulting corpus should enable a rather thorough investigation of the general trends. The findings cannot be considered conclusive in any way but they might reveal aspects worth further research. Even though the number of theses was equal, the corpora were very different in size as there were roughly 70 thousand word tokens in the MIT corpus and about 178 thousand in the UT corpus due to the different length requirements apparently set for the theses in the two universities.

3.2 Method

Manual analysis of each thesis would be a difficult process even in the case of a small corpus. As all the texts were Portable Document Format (pdf) files, which in the case of the MIT corpus appeared to be the result of scanning, the first step was to extract the text from all files so that the entire corpus would consist of UTF-8 encoding text files. The procedure for doing so is not completely error-free, particularly in the case of scanned materials that, in addition, contain non-standard alphanumeric symbols often found in various economics related formulae. However, random samples of the material converted were studied and there is every reason to believe that the number of conversion errors is so small that it does not have a remarkable impact on the results of the analysis.

To focus the analysis on textual elements containing academic rhetoric,

acknowledgements, abstracts, annexes, title pages, tables of contents, and lists of references were omitted from the texts. The remaining material was divided into two different corpora, one named UT (texts from the University of Tartu) and the other named MIT (texts from the Massachusetts Institute of Technology).

The corpora were loaded into the concordance program AntConc (Anthony 2014) to easily identify and describe in-text occurrences of different phrases and phenomena. Due to the context dependent nature of the material studied, occurrences were often described considering the context they appeared in. All searches conducted were case-insensitive.

AntConc was chosen as it is freeware and has all the basic capabilities for text analysis. The simplest function it has is that of a search engine that also offers statistics such as the frequency a word is used with. The program was particularly useful for me as it also has a simple means for frequency based extraction of bundles, using the criteria of total frequency, number of words to be considered a bundle and range (defining how many texts in the corpus have to contain the bundle). As the corpora were so different in terms of size, using the same frequency and range criteria for both would have made little sense, which is why the criteria were matched so that the number of results returned was relatively similar (9 phrases from the MIT and 11 from the UT corpus). More precisely, a bundle in the UT corpus was expected to appear in at least 4 texts and 15 times in total, whereas a bundle in the MIT corpus was expected to appear in 3 texts, 5 times in total. In the case of both corpora, a bundle was defined as consisting of 4 words, 'since their functional specification tends to be clearer than that of 3-word bundles, and they are more frequent and show less variation than 5- and more word bundles' and because 4-word bundles are often used in other similar studies (Dontcheva-Navratilova 2012: 42). Even despite the matching, it is clear that a mere comparison of

frequencies would be of little use due to the different number of word tokens. To balance the situation, the frequencies were additionally calculated per thousand words (PTW) in the case of bundles as well as in the case of hedges and DMs. Unless otherwise stated, the number will be presented in parentheses after the frequencies.

Unlike LBs, DMs cannot be searched for using simple automatic systems, and hence a choice had to be made as for which markers to look for. It was based on the lists of most popular markers given in the article by Povolná (2012: 137-139). As her corpora were those of Czech students, one might say that using her list to search for DMs in the MIT corpus may lead to biased results, as authors influenced by the Anglophone traditions could be using different DMs. While this may be so, the DMs listed by Povolná still offer valid grounds for describing some of the possible differences between the UT and MIT corpora, which is the aim of this thesis.

The list of hedges to look for was also compiled from those mentioned by other authors (Aull and Lancaster 2014: 176), with the main focus on the categories of modals, approximators and evidential verbs. Some additional entries were added. The grammatical categories of the hedges were not limited so that the data obtained would be more extensive. Overall, the procedure of analysis was similar to that applied to LBs. Statistical data on both corpora were compared and similarities or differences discussed in more detail.

Even though the frequency a LB occurs with has often been analysed, such an approach also has some deficiencies, which have been discussed, for example, by Liu (2012) who tried to establish whether there would be a remarkable difference in the most frequent multi word constructions⁶ depending on the method used. Simpson-Vlach and Ellis (2010: 490) quoted in

6 A term broader than 'lexical bundle' as it includes lexical bundles, idioms, collocations etc. (2012: 25).

Liu (2012: 27) argue that the frequency based approach (FBA) is 'methodologically straightforward but results in a long list of recurrent word sequences that collapse distinctions that intuition would deem relevant', which in turn means that we may also detect bundles that have little psycholinguistic salience, such as 'to do with the' or 'I think it was'. Vlach and Ellis additionally note that the frequent occurrence of some n-grams⁷ can be explained by the 'high frequency of their component words', whereas phrases whose components co-occur rarely may wrongly be considered irrelevant using the FBA (Simpson-Vlach and Ellis: 2010: 493).

The reason for this seems to be widely accepted even by the people using the FBA. The FBA leads to many bundles that are incomplete, as far as structure or semantics are concerned (Liu 2012: 27). In order to address the limitations of the FBA, Vlach and Ellis used an alternative method to investigate LBs. Their approach is based not on frequency but on importance of a bundle (referred to as formulas by Simpson-Vlach and Ellis). This is determined by looking at many factors but simply put they focus on association, i.e. the probability that two words do not occur together by chance (Simpson-Vlach and Ellis 2010: 487, 490).

Liu set out to compare the two methods, additionally seeking answers to other questions, such as 'which [multi word constructions] are most common and what main functions they play' (Liu 2012: 28). What makes his study special is the fact that the corpora he used (the academic sub-corpora of Corpus of Contemporary American English and British National Corpus were enormous with their roughly 98 million words compared to corpora of five million words or less (Liu 2012: 31) that had been used before. A list of about 2000 idioms

7 'A sequence of n letters or characters /.../' (OED Online).

and 2000 phrasal/prepositional verbs compiled using intuition and lists of previous studies was compared against these corpora and it was found that the two methods of identifying LBs indeed lead to different results (Liu 2012: 28-29, 31). More precisely, many of the items in the previous lists of popular LBs did not make it into the list compiled by Liu. However, as the material used in this thesis is so small and the author can review most of the examples, the possible difference should not be of concern. Yet, as identifying hedges and DMs (hundreds of occurrences at times) is much more demanding and also depends more on context (e.g. 'since'), data on them should be treated with caution because the numbers used in this thesis reflect total frequencies. Many examples were, however, reviewed and the observations give me reason to believe that the words and phrases discussed below were mostly used as DMs or hedges. In the case of low frequencies, all examples were reviewed and respective comments have been added to the results.

3.3 Results

The results of the analysis revealed some considerable differences in all categories compared. This was not surprising but there were occasions on which the trends in both corpora were clearly inverted, for example. The analysis also revealed some constructions used only by the authors in the UT corpus. Even though explaining the reasons for using such phrases is not always possible, there were some constructions that can be considered indicative of more general tendencies in the UT corpus.

3.3.1 Lexical Bundles

In terms of LBs, the lists of most frequent bundles were quite different by corpora. The following tables lists five most popular bundles in both corpora (the list has been compiled using the restrictions mentioned in the *Methods* chapter).

Table 1. The most frequent bundles in the MIT corpus

Rank	Bundle	Total frequency	Number of authors
1	On the other hand	16	4
2	The results of the	11	4
3	The results for the	9	4
4	Of the firm's	8	3
5	Are more likely to	6	3

Table 2. The most frequent bundles in the UT corpus

Rank	Bundle	Total frequency	Number of authors
1	On the other hand	43	6
2	It is important to	37	6
3	At the same time	30	6
4	To be able to	28	6
5	As well as the	21	5

In agreement with previous research (Hyland 2008: 7, Wei and Lei 2011: 159)⁸, it turned out that the most frequent LB in both corpora was 'on the other hand' and the frequencies it was used with did not differ much. With regard to other LBs, the differences were noteworthy.

8 Wei and Lei found 'on the other hand' to be among the most frequent bundles but not the most frequent (Wei and Lei 2011: 159).

Out of the 11 most frequent phrases in the UT corpus, only 3 matched with the 8 in the MIT corpus. More importantly, there were bundles in the UT corpus that were not used in the MIT corpus or were used very scarcely. A prominent example is the phrase 'it is important to'. It was the second most frequent and widely used phrase in the UT corpus, occurring 37 times (0.2), whereas it was not used at all in the MIT corpus. Since the difference was so remarkable, I decided to test the MIT corpus for possible derivatives of 'it is important to', such as 'it is very important to'. but the difference remained. The MIT corpus contained only two instances of 'important to' (0.03), one of which could be considered equal in meaning with the original bundle.

It is worth noting that there was another bundle in the UT corpus similar to 'it is important to'. The phrase 'it is necessary to' was used on 6 (0.03) occasions in the UT corpus, which is not often but it is very similar to the previously discussed 'it is important to', which is why using the phrases was studied in more detail. In the case of one author, there were 3 consecutive sentences quite at the beginning of the thesis, all containing either of the two phrases, so that the frequent occurrence may be due to last minute writing or poor editing. Again, it is hard not to notice that the phrases are often redundant. One might as well assume that if something has been written down it has been considered relevant enough to do so.

The third and fourth most popular phrases in the UT corpus seem to follow a similar pattern. 'At the same time' was used 30 times in 6 theses out of seven in the UT corpus while the MIT corpus contained only 2 such instances. This is not to say that it is wrong to use 'at the same time' for introducing contrast but the fact remains that the MIT corpus contained less of it. One possible explanation might be that the authors in the UT corpus tend to translate Estonian contrastive DMs, such as '*samas*'.

'To be able to', the fourth most popular bundle in the UT corpus also revealed a remarkable difference between the corpora. It appeared 28 times in the UT corpus, which is a good example of the redundancy disapproved in Anglophone academic writing. While the phrase is, again, grammatically correct and can be used, it might often be made shorter by using words like 'to'. That the MIT corpus contained no instances of the longer phrase can also be considered a clear indicator that it should be avoided.

Inspired by the differences in the most frequent bundles, I decided to conduct an additional analysis of a list of 4 word bundles that a remarkable part of the UT corpus contained and that were not present in the MIT corpus. The criteria for compiling the list were not set equal as the corpora varied so much in size and I wanted to get a relatively long list of bundles, in order for the comparison to be broad. Whereas the UT corpus had to contain a bundle in three texts and at least six times in total, it was considered sufficient if the same phrase was present at least once in two texts in the MIT corpus. To compare the resulting lists, I wrote a computer program that provided another list which contained information on whether a bundle in the UT corpus even appeared in the MIT corpus and if so, then which were the differences between the corpora in terms of the popularity of the bundle. The final list was rather extensive and as it was also based on frequency of use, I will hereby discuss just the first 10 entries. This allows me to make some more specific assumptions, regarding the peculiarities and justified use of the bundles in the UT corpus. Some of the phrases in the list are those already discussed above, so they will not be discussed in further detail. The most frequent bundle in the MIT corpus (ranking 0) has been omitted as it was a meaningless array of letters that was probably due to a conversion error. It is worth noting that the program did not distinguish bundles with exact same frequency so it is technically possible that, for example, a bundle that ranked 1st was used exactly as often that a bundle that ranked 2nd.

Table 3. Popular lexical bundles in the UT corpus (in comparison with the MIT corpus)

Bundle	MIT ranking	UT ranking	UT Ranking – MIT ranking
on the other hand	1	0	-1
it is important to	Not present	1	N/A
at the same time	Not present	2	N/A
to be able to	Not present	3	N/A
as well as the	14	4	-10
is one of the	174	5	-169
it was decided to	Not present	6	N/A
the size of the	26	7	-19
it is possible to	Not present	8	N/A
in the context of	Not present	9	N/A

A phrase that appeared among the top ten and was quite similar to 'to be able to', in terms of frequency was 'it is possible to' that occurred just once in the MIT corpus (0,01) but was used 19 times (0.10) in the UT corpus. Further study of the examples in the UT corpus revealed that the phrase could often be replaced with 'it can be'. A possible reason why this was not done in the UT corpus lies in the fact that such a replacement might require a change in word order,

which is something non-native English users may feel insecure about. For example, the sentence 'Still, to make practical use of the model, it is possible to analyse the current situation and opportunities for long-term development from an organisational perspective based on the model.' could be rewritten as '/.../, the current situation and opportunities can be analysed from an organisational perspective based on this model'. It seems likely that the frequent use of 'it is possible to' in the UT corpus is again an example of translating a phrase from Estonian, as there is the phrase 'on võimalik'.

Another phrase that was used broadly in the UT corpus and little in the MIT corpus was 'is one of the' that was used 21 times by 6 authors in the UT corpus, and 2 times in the MIT corpus. While it might again be explained with a translation from Estonian, the reasons behind the differing usage are probably somewhat more complicated. For some reason, the authors in the UT corpus seem to use the phrase in superlatives, in constructions such as 'is one of the best', 'is one of the most'. Studying the MIT corpora additionally for occurrences of 'one of the' revealed that it was not used at all in superlative constructions, if not to consider one of the two occurrences mentioned above.

In addition to such examples, the UT corpus also contained phrases that could be made shorter; 'considered to be' that occurred in five UT thesis, 24 times in total, can be replaced by 'considered'. It is noteworthy that 'considered' was used 12 times in total in the MIT corpus and 4-5 instances were such that the author could have used 'considered to be' had they wanted to do so). However, there were no instances of 'considered to be' in the MIT corpus, which may indicate that the authors in the MIT corpus disprefer such longer alternatives.

In conclusion, it may be said that there were few, if any, such bundles in the MIT corpus that could be used in a clearly redundant function. Examples include bundles like 'the results

for the', 'the results of the', 'are more likely to', 'the size of the', 'we assume that the', etc., none of which can be considered redundant. In fact, 5 of the 9 most popular bundles in the MIT corpus contain either 'of the' or 'for the', so that shortening the constructions seems rather difficult. The UT corpus, on the other hand, was somewhat different, which allows us to assume that the authors in the MIT corpus are more aware of the need to avoid redundant constructions.

3.3.2 Discourse Markers

There were DMs in the corpora that some authors seemed to use much more than others. 'Thus', which I consider synonymous with the word 'therefore', was such a word. The latter appeared 129 times in the UT corpus, but on only 13 occasions in the MIT corpus. At that, the PTW ratios of the phrase were 0.72 and 0.18 respectively. It has to be said though that 84 of the 129 occurrences stemmed from two texts, which clearly indicates that the finding is to be considered a personal preference rather than a widespread trend.

As 'therefore' had proved so popular in the UT corpus, it seemed reasonable to expect 'thus' to appear but a few times, particularly considering Povolná's (2012: 143) observation that authors – unless they have a particularly wide repertoire of markers – usually tend to use either of the two phrases. The frequencies seem to contradict this, as 'thus' was used in every text of the UT corpus and occurred 87 (0.49) times. The more telling PTW ratio of the MIT corpus, however, was 1.2 compared to 0.49 of the UT corpus. Also considering that most appearances of 'therefore' in the UT corpus stemmed from two authors, it becomes clear that 'thus' is indeed seldom used in the UT corpus when compared to MIT. Povolná refers to Biber et al (1999) in saying that 'thus' and 'therefore' are among the most common markers in academic texts by native speakers (Povolná 2012: 143). In other words, it seems that the UT corpus resembles

the MIT corpus in that regard, only the specific phrase used is different.

In terms of unequal distribution, there is a DM that was very unevenly distributed in both corpora. The phrase 'as a result' appeared 27 times in total in the UT corpus but 18 of the cases were in the same text. In the MIT corpus the phrase appeared in 4 theses, 6 times in total, with half the occurrences in one text. The word 'consequently' was another one that stood out due to its uneven use. In the UT corpus, one author was responsible for ~ 73% of the occurrences and in the MIT corpus the same percent was ~ 66% (somewhat smaller as the word was used by 3 authors instead of 2). 'Since' was no different as there were authors in both corpora who used it much more than others; while some had used roughly thirty or even fifty times, there were those that used it only a few times.

Despite the fact that there were various differences between the corpora, I was most interested in the extremes. The following tables contain data on DMs in the case of which corpora based differences are the biggest. The greater the PTW UT – PTW MIT value, the more the phrase is used in the UT corpus (compared to the MIT corpus).

Table 4. Five most frequent DMs in the MIT corpus (in comparison with the UT corpus)

	DM	Total (UT)	PTW (UT)	Total (MIT)	PTW (MIT)	PTW UT – PTW MIT
1	since	122	0.68	132	1.87	-1.19
2	thus	87	0.49	91	1.29	-0.80
3	so that	31	0.17	34	0.48	-0.31
4	though	11	0.06	8	0.11	-0.05
5	consequently	15	0.08	9	0.13	-0.05

Table 5. Five most frequent DMs in the UT corpus (in comparison with the MIT corpus)

	DM	Total (UT)	PTW (UT)	Total (MIT)	PTW (MIT)	PTW UT – PTW MIT
1	therefore	129	0.72	13	0.18	0.54
2	while	173	0.97	31	0.44	0.53
3	because	222	1.25	64	0.91	0.34
4	however	286	1.61	91	1.29	0.32
5	instead	73	0.41	7	0.10	0.31

'Therefore' occurred but 13 times (0.18) in the MIT corpus while it was present in all seven UT texts, appearing 129 (0.72) times in total. The same can be said about 'instead' that was, again, present in all seven UT texts, occurring 63 (0.35) times in total, whereas only 2 texts in the MIT corpus contained the phrase and it appeared 7 (0.03) times. It is worth noting that the use of 'instead of' revealed an inverted trend in the corpora. In the UT corpus, about 66% of the occurrences of 'instead' were those of 'instead of', whereas in the MIT corpus the same percentage was about 42. 'Yet' showed similar behaviour. While 5 authors used it in both corpora, using it as a DM differed by corpora: in the MIT corpus about a third of the occurrences qualified as DMs but in the UT corpus about 43% of the total appearances could be classified as DMs.

Further examination of the tables revealed another interesting pattern. The authors in the UT corpus stand out due to their extensive use of 'because' and 'however'. Also considering the fact that 4 out of 5 entries in the MIT table are causal in nature, gives us reason to assume that the authors in the MIT corpus generally use a much broader set of causal markers when

compared to the UT corpus. It would be wrong to make any assumptions about the background of the UT authors, the finding may agree with Povolná's observation that novice writers' repertoire of contrastive markers is broader than their set of causal markers (Povolná 2012: 139). However, this is not to say that the authors in the UT corpus did not use other causal markers than 'therefore' and 'because'. The two were just preferred to the other options at the cost of underusing the others when compared to the MIT corpus.

3.3.3 Hedges

The words 'can' and 'might' occurred 714 (4.0) and 86 (0.4) times in the UT corpus, while the MIT corpus contained only 115 (1.6) and 18 (0.25) of them respectively. The way 'might' was used at times in the UT corpus revealed an interesting trend. Some of the authors who had used the hedge had clearly tried to make it even stronger, which led to constructions such as 'at least in theory, one might' and 'in some ways, there might'.

The number of modals alone is not enough to decide on whether a text contains hedging or not. However, studying phrases like 'can be' and 'may be' may narrow things down to some extent. It is worth noting that the statistics for the phrases was clearly inverted at rather even distribution rates. The UT corpus contained more of 'can be' and the MIT corpus contained more of 'may be'. The number of occurrences was so high that it was hard to assess which were true hedges but the difference was so remarkable ('can be' 394 (2.21) and 38 (0.54), 'may be' 69 (0.98) and 59 (0.33)) that we may assume it to be considerable in the case of hedges containing the construction. As the phrases are interchangeable in many contexts, although not necessarily equal in meaning, it is hard to draw any major conclusions from the observation but a possible explanation is that the writers in the UT corpus do not consider the difference between modals as relevant, which is also suggested by Ventola (1997) in the case of Finns.

She also mentions the possibility that Finns just prefer stronger modals but the extent of variation in her study suggests that difficulties in distinguishing the phrases are a more likely explanation (Ventola 1997: 174). ESL speakers' difficulties in understanding the English modality system have also been suggested by other authors, such as Llantada (2013: 265).

In addition to modals there were other hedges that were used with varying frequency in the corpora. For example, while the MIT corpus contained just one instance of 'probably' and even that was part of a quote, the word appeared in half the theses in the UT corpus and occurred 9 times. 'Quite' was used with a similar frequency; it was present in 5 UT papers and was used 22 times in total, whereas 5 occurrences of the word in the MIT corpus were in the same text. 'Rather' (not including 'rather than') was used 40 times in total by four authors in the UT corpus and 17-18 instances could be classified as hedging. Considering that 31 of the 40 stemmed from one author, we can clearly say that the number is biased but even so it remains noteworthy that the MIT corpus contained the word (again, not including 'rather than') just 5 times and only one instance of the 5 classified as hedging.

'Sometimes' was not present at all in the MIT corpus, whereas it appeared in 6 out of 7 theses in the UT corpus, appearing 27 times in total. The difference with the MIT corpus remains noteworthy although 10 of the UT occurrences stemmed from one author. 'Appear' was another word that the MIT corpus did not contain as a hedge, even though it was used 5 times in total by 3 authors. In the UT corpus, on the other hand, the word occurred 14 times and was used by 6 authors, 4 of whom used the word for hedging (6 occurrences). 'Indication' and 'fairly' also did not occur in the MIT corpus, while the UT corpus contained 6 instances of both (used by 3 authors). The MIT corpus contained but a few instances of 'indicate' in contrast to the UT corpus in which 6 authors had used the phrase so that it occurred 51 times

in total, although there was again an author who used it considerably more than the rest. Another word with a similar usage pattern was 'often' that appeared 3 times in the MIT corpus and 86 times in the UT corpus.

The authors in the UT corpus used a considerably greater number of some approximators, such as 'almost' and 'around'. An exception worth mentioning concerns the word 'about'. Although all authors in the UT corpus used it, so that it appeared 165 times in total, there were only about 3 instances of hedging in total by three authors. In the MIT corpus, about 13 of the 47 occurrences could be classified as hedges.

An increase in the impact of a hedge can also be achieved by creating a distance between a statement and the author. For example, the UT corpus contained the word 'claim' about seven times more than the MIT corpus. 'Claim' was present in the latter as well but only half the occurrences could be considered hedges, while the rest often denoted receiving (e.g. 'claim benefits'). In the UT corpus, such instances formed only about ten percent of the total, while there were phrases like 'some claim', 'many claim'. Another way to hedge is to use a certain adjective or adverb and the authors in the UT corpus are certainly well aware of that. For example, the word 'usually' was used 62 times (0.35), whereas the word appeared only seven times (0.1) in the MIT corpus. 'Generally', on the other hand, which is quite similar in meaning, was used rather equally in the corpora.

There were also some hedges that were used more by the authors in the MIT corpus. For example, the PTW value for 'typical' was 0.1 in the MIT corpus and only 0.07 in the UT corpus. The PTW value for 'typically' was also higher in the MIT corpus but 7 of the 10 occurrences stemmed from one author. Other words that were more popular in the MIT corpus include 'possibly', 'suggest(s)', 'tend(s)', 'indicate(s)', 'likely', 'suggest', and 'about'. As for words

that were used scarcely in both corpora, the list was somewhat surprising for me. 'Presumably', 'probably', and 'perhaps' were all rare or used mainly by one author, although all of them seem like good choices for hedging (cf. Varttala 1999: 187).

In conclusion, it can be said that although both corpora contained interesting findings, the data gathered are not sufficient to say whether authors in one of the corpora hedge more. However, the study revealed two important findings. First, the authors in the UT corpus seem to use stronger modals than the authors in the MIT corpus and second, there seems to be a set of hedges that the MIT authors do not use. Although the existence of such hedges may indicate that the authors in the UT corpus hedge more, further analysis, preferably using corpora more equal in size, is needed before further conclusions can be made.

CONCLUSION

That a non-native speaker of a language differs from a native speaker in how they use the language, seems perfectly natural. However, describing the difference is not as simple as it seems. After all, a language involves different elements such as words, sentence structures etc. Defining the difference becomes even more intriguing if we were to consider the specific case of academic English that has become the global norm in the academia. In addition to the possible linguistic differences, the academia sets standards of its own, some more general, some rather field-specific. Authors all over the world are expected to adapt to the norms and in a way they usually do, the question is how.

The aim set in this thesis was to conduct a comparative analysis of Estonian and English academic writing habits, concentrating on hedging, cohesion and coherence. The introductory part of the thesis contains a brief insight into the intentions of the author. The first chapter is an overview of what has been done before in the field of comparative rhetoric. Chapter 2 gives an overview of comparative rhetoric and contain a discussion of some works published before that are relevant for this thesis. Chapter 3 contains the empirical part of the study, introducing the methods used, analysing their limitations and discussing the results reached during the study conducted for this thesis. The conclusion sums up the entire thesis and contains the key findings.

The study involved Master's theses written by students at the department of Economics at the University of Tartu and at the Massachusetts Institute of Technology. All papers included were processed using the text analysis tool AntConc (Anthony 2014). The specific phrases and words analysed were determined either by a broader set of predefined criteria, such as the

number of words in a phrase and its frequency of occurrence or by the results of previous research in the field.

The results of the analysis, although not conclusive, led to some interesting conclusions. First, it seems that the students in the two universities differ greatly as for the types of LBs used extensively in the theses. The fact that authors influenced by Anglophone traditions use bundles not present in the writings of authors writing in a non-Anglophone tradition is no surprise but the fact that some of the bundles used extensively in the UT corpus were not at all used in the MIT corpus was a bit unexpected. Further examination of the findings revealed that the most popular bundles in the UT corpus were often such that could either be expressed using a more compact alternative or omitted altogether. If we assume that the authors in the UT corpus are affected by Estonian, we can say that a considerable number of the lengthy constructions stem from attempts to translate an Estonian expression into English. The hypothesis is supported by the fact that the authors in the MIT corpus did not use some constructions at all. Yet, it must be noted that the most frequent bundle 'on the other hand', the popularity of which has also been confirmed by other studies (Hyland 2008, Wei and Lei 2011), was the same in both corpora.

The analysis of discourse markers revealed a clear difference between the UT and MIT corpora, as the range of frequently used causal markers was narrower in the UT corpus. At that, 'therefore', one of the two causal markers used extensively in the UT corpus, was used very little in the MIT corpus.

Drawing major conclusions in terms of hedging is somewhat more difficult as hedges are hard to classify and define conclusively. It could be assumed that the authors in the UT corpus hedged more but it has to be kept in mind that the authors in the MIT corpus did not use some

hedges at all, which is why the motives of the authors should be studied further in order to reach more conclusive results. Furthermore, in terms of modals, the study actually indicated that the authors in the UT corpus may hedge less, i.e. use stronger modals; the MIT corpus contained more of 'may be' than the UT corpus, whereas the latter contained more of 'can be' than the MIT corpus. Even though all such constructions need not be hedges, the finding may prove the hypothesis presented by other authors (Ventola 1997), which states that non-native English writers tend to use stronger modal constructions than native-writers.

All in all, it can be said that the thesis helped to outline some notable differences between the academic traditions dominating in the universities studied. Further research is needed to say whether the findings represent more general tendencies characteristic to the two academic traditions but the results reached could be used as a material of comparison in further research as the number of similar studies appears to be rather small.

REFERENCES

- Alexander, Louis G. 2007. *Longman English Grammar*. Harlow: Longman.
- Alonso Alonso, Rosa, Maria Alonso Alonso and Laura Torrado Mariñas. 2012. Hedging: an exploratory study of pragmatic transfer in nonnative English readers' rhetorical preferences. *Iberica*, 23, 47-64.
- Anthony, Laurence. (2014). AntConc (Version 3.4.1w) [Computer Software]. Tokyo, Japan: Waseda University. Available at <http://www.antlab.sci.waseda.ac.jp/>, accessed 6th July 2014.
- Aull, Laura L. and Zak Lancaster. 2014. Linguistic markers of stance in early and advanced academic writing: a corpus-based comparison. *Written Communication*, 31: 2, 151-183.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, Douglas and Federica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26: 3, 263-286.
- Bitzer, Lloyd F. 1968. The rhetorical situation. *Philosophy & Rhetoric*, 1: 1, 1-14.
- Chamonikolasová, Jana. 2005. Comparing the structures of academic texts written in English and Czech. Available at <http://www.phil.muni.cz/stylistika/studie/comparing.pdf>, accessed 6th July 2014.

- Chovanec, Jan. 2012. Written academic discourse in English: from local traditions to global outreach. *Brno Studies in English*, 38: 2, 5-16.
- Cooley, Linda and Jo Lewkowicz. 1997. Developing awareness of the rhetorical and linguistic conventions of writing a thesis in English: addressing the needs of EFL/ESL postgraduate students. In Anna Duszak (ed). *Culture and Styles of Academic Discourse*, 113-129. Berlin: Mouton de Gruyter.
- Crossley, Scott A., Rod Roscoe and Danielle S. McNamara. 2014. What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31: 2, 184-214.
- Crystal, David. 1999. World English: past, present, future. Available at <http://www.davidcrystal.com/?fileid=-4031>, accessed 6th July 2014.
- Dahl, Trine. 2004. Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics*, 36: 10, 1807-1825.
- Dontcheva-Navratilova, Olga. 2012. Lexical bundles in academic texts by non-native speakers. *Brno Studies in English*, 38: 2, 37-58.
- Duff, Patricia A. 2007. Problematizing academic discourse socialisation. In Helen Marriott, Tim Moore and Robyn Spence-Brown (eds). *Learning Discourses and the Discourses of Learning*, 1.1 - 1.18. Melbourne: Monash University ePress.
- Durkin, Kathy. 2008. The middle way: East Asian master's students' perceptions of critical argumentation in U.K. universities. *Journal of Studies in International Education*, 12: 1, 38-55.

- Helstein, Alar. 2004. *Keeleline viisakus ja žanrikonventsioonid saksa, inglise ja eesti teoloogilistes teadusartiklites*. [Linguistic Politeness and Genre Conventions in German, English and Estonian Theological Articles.] Available at http://valdre.com/andres/KIT/MA/MA_Alar.pdf, accessed 6th July 2014.
- Halliday, Michael A. K. 1994. A language development approach to education. In Norman Bird, et al. (Eds.). *Language and learning. Papers Presented at the Annual International Language in Education Conference*, Hong Kong, 1993.
- Hyland, Ken. 2008. As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27: 1, 4-21.
- Kaplan, Robert. 1966. Cultural thought patterns in inter-cultural education. *Language and Learning*, 16: 1-2, 11-25.
- Kennedy, George A. 2006. Comparative rhetoric. In Thomas O. Sloane (ed). *Encyclopedia of Rhetoric* (e-reference edition). New York: Oxford University Press.
- Kolowich, Steve. 2014. Writing instructor, skeptical of automated grading, pits machine vs. machine. Available at <http://chronicle.com/article/Writing-Instructor-Skeptical/146211/>, accessed 7th June 2014.
- Kreutz, Heinz and Annette Harres. 1997. Some observations on the distribution and function of hedging in German and English academic writing. In Anna Duszak (ed). *Culture and Styles of Academic Discourse*, 181-201. Berlin: Mouton de Gruyter.
- Lakoff, George. 1973. Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2: 4, 458-508.

- Liu, Dilin. 2012. The most frequently-used multi-word constructions in academic written English: a multi-corpus study. *English for Specific Purposes*, 31: 1, 25-35.
- Martín, Pedro. 2003. The pragmatic rhetorical strategy of hedging in academic writing. *Vigo International Journal of Applied Linguistics*, 0: 1, 57-72.
- McNamara, Danielle S. 2013. The epistemic stance between the author and reader: a driving force in the cohesion of text and writing. *Discourse Studies*, 15: 5, 579-595.
- Mao, LuMing. 2003. Reflective encounters: illustrating comparative rhetoric. *Style*, 37: 4, 401-425.
- Matalene, Carolyn. 1985. Contrastive rhetoric: an American writing teacher in China. *College English*, 47: 8, 789-808.
- Morrison, John L. 1972. The Absence of a Rhetorical Tradition in Japanese Culture. *Western Speech*, 36: 2, 89-102.
- OED Online. June 2014. Oxford University Press. Available at <http://www.oed.com/view/Entry/245099?redirectedFrom=n-gram&>, accessed 7th July 2014.
- Palmer, Juan C. 1999. Coherence and cohesion in the English language classroom: the use of lexical reiteration and pronominalisation. *RELC Journal*, 30: 2, 61-85.
- Pérez-Llantada, Carmen. 2013. 'Glocal' rhetorical practices in academic writing. *European Journal of English Studies*, 17: 3, 251-268.
- Povolná, Renata. 2012. Causal and contrastive discourse markers in novice academic writing. *Brno Studies in English*, 38: 2, 131-148.

- Rogers, Sandra H. 2004. Evaluating textual coherence: a case study of university business writing by Efl and native English-speaking students in New Zealand. *RELC Journal*, 35: 2, 135-147.
- Rummel, Kärt. 2005. How To Write Reader-Friendly Texts: Common Problems in the English Academic Writing of Estonian Writers. Available at <http://dspace.utlib.ee/dspace/bitstream/handle/10062/707/rummel.pdf?sequence=5>, accessed 6th July 2014.
- Rummel, Kärt. 2010. *Creating Coherent Texts in English as a foreign language: theory and practice*. Tartu: Tartu University Press.
- Severino, Carol. 1993. The 'doodles' in context: qualifying claims about contrastive rhetoric. *Writing Center Journal*, 14: 1, 44-61.
- Shaw, Philip and Irena Vassileva. 2009. Co-evolving academic rhetoric across culture; Britain, Bulgaria, Denmark, Germany in the 20th century. *Journal of Pragmatics*, 41: 2, 290-305.
- Simpson-Vlach, Rita and Nick C Ellis. 2010. An academic formulas list: new methods in phraseology research. *Applied Linguistics*, 31: 4, 487-512.
- Snow, Catherine E and Paola Uccelli. 2009. The challenge of academic language. In David R. Olson and Nancy Torrance (eds). *The Cambridge Handbook of Literacy*, 112-133. New York: Cambridge University Press.
- Swales, John. 1997. English as a tyrannosaurus Rex. *World Englishes*, 16: 3, 373-382.

- Uccelli, Paola, Christina L. Dobbs and Jessica Scott. 2013. Mastering Academic language: organization and stance in the persuasive writing of high school students. *Written Communication*, 30: 1, 36-62.
- Ventola, Eija. 1997. Modalization: probability – an exploration into its role in academic writing. In Anna Duszak (ed). *Culture and Styles of Academic Discourse*, 157-179. New York-Berlin: Mouton de Gruyter.
- Varttala, Teppo. 1999. Remarks on the communicative functions of hedging in popular scientific and specialist research articles on medicine. *English for Specific Purposes*, 18: 2, 177-200.
- Vatz, Richard E. 1973. The myth of the rhetorical situation. *Philosophy & Rhetoric*, 6: 3, 154-161.
- Wei, Yaoyu and Lei Lei. 2011. Lexical bundles in the academic writing of advanced Chinese EFL learners. *RELC journal*, 42: 2, 155-166.
- Widdowson, Henry G. 1978. *Teaching Language as Communication*. Oxford: Oxford University Press.

RESÜMEE

Tartu Ülikool

Tõlkekeskus

Kaur Männamaa

RHETORIC OF MASTER'S THESES OF ESTONIAN AND ANGLOPHONE ECONOMICS STUDENTS

EESTI JA INGLISE KEELT KÕNELEVATE MAJANDUSTEADUSKONDADE
ÜLIÕPILASTE MAGISTRITÖÖDES KASUTATAV RETOORIKA

Magistritöö

2014

Lehekülgede arv: 66

Annotatsioon:

Töö autori eesmärgiks oli uurida, mille poolest erineb inglise keelt kõneleva inimese keelekasutus selle inimese inglise keelest, kelle jaoks inglise keel ei ole emakeel. TA võrdles Tartu Ülikoolis (TÜ) ja Massachusettsi Tehnoloogiainstituudis (Massachusetts Institute of Technology) (MIT) (USA) 21. sajandil majanduse teemal kirjutatud magistritöid, keskendudes

neis leiduvate nn 'põiklusväljendite' (Helstein 2004: 7) (*hedges*) ja teksti sidususega seonduvate konstruktsioonide (*discourse markers, lexical bundles*) uurimisele. Tekstide analüüsimisel kasutab autor arvutianalüüsi, ent otsitavad tunnused ise pärinevad paljuski varasematest uurimistöödest.

Analüüsi tulemusena selgus, et mitmed väljendid, mida MIT kirjutistes ei esine üldse või esineb vähe, olid TÜ-st pärit töödes äärmiselt populaarsed. Ühe võimaliku põhjusena pakutakse välja TÜ autorite soov tõlkida eesti keeles sageli esinevad väljendid otse inglise keelde. Põiklusväljendeid leitakse TÜ-st pärinevatest töödest kohati isegi rohkem kui MIT korpusest, mida võib pidada pisut üllatuslikuks. Samas tekib taoline mulje paljuski tänu sellele, et MIT korpuses mõningaid väljendeid üldse ei kasutata, mistõttu teema vajab lähemat uurimist. Liiasi on TÜ korpuses ka aspekte, mis mõjuvad, võrreldes MIT korpusega küllaltki jõulisena. Samuti on TÜ töödes teatud tüüpi väljendite varieeruvus oluliselt madalam kui analüüsitud MIT töödel.

Ehkki siinsed tulemused on vaid esialgsed ning üldisemate ilmingute kaardistamiseks oleks tarvis täiendavat tööd, peaks kogutud materjal ja tehtud tähelepanekud pakkuma ometi mõningast abi inimesele, kes soovib tulevikus sarnast valdkonda uurida.

Märksõnad: akadeemiline kirjutamine, võrdlev retoorika, korpusanalüüs

Lihlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kaur Männamaa (sünnikuupäev: 06.02.1989)

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) enda loodud teose

„Rhetoric of Master’s Theses of Estonian and Anglophone Economics Students“

mille juhendaja on Raili Marling,

1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace’i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 08.07.2014