

Vilnius University
Faculty of Philology

Lukas Ciesiūnas

From Complex to Simple: A Comparison of Legal and Media Discourses

Thesis submitted in partial fulfilment of requirements for the degree of MA in Language for
Specific (Legal) Purposes

Supervisor: Assoc. Prof. Dr. Rita Juknevičienė

Vilnius

2021

Contents

Abstract	3
1. Introduction	4
Literature Review	5
2. Data and Methods	9
3. Results and Discussion.....	11
3.1 Word frequency analysis.....	11
3.1.1 Delfi AS v Estonia Case	11
3.1.2 Georgia v Russia Case.....	18
3.2 Lexical Bundle Analysis	20
3.2.1 Structural Analysis	20
3.2.2 Functional Analysis.....	23
3.2.3 Delfi AS v Estonia Case Qualitative Analysis.....	28
3.2.4 Georgia v Russia Case Qualitative Analysis	40
4. Conclusions	46
Summary in Lithuanian.....	48
References	50

Abstract

Legal language and terminology is more relevant to a layman than ever before. Because of that, it is getting increasingly important to understand legal discourse even without having any legal education. The focus of this paper is the comparison of legal discourse and media discourse to see if any features of simplification can be identified. Two case studies were performed, four corpora in total were compiled, two corpora for each study. One corpus represented legal language and consisted of the official case report from the European Court of Human Rights, while the other corpus represented media discourse and consisted of various online news articles from different internet portals that discussed specific legal cases. With the help of AntConc (Anthony 2020) a corpus analysis software, word frequency lists were generated and analyzed. Using the same software 4-word most frequent lexical bundles lists were also generated and investigated. Functional and structural analyses of the lexical bundles were carried out as well, they were based on a previous study done by Biber et al. (2004). It was found that certain legal terms, words and expressions were avoided in media discourse and replaced with less complex alternatives. It was also found that that certain details of the cases like specific legal acts, articles of conventions and directives were more often than not omitted in media discourse.

1. Introduction

Over time, the legal sphere has gotten closer and more relevant to our everyday lives. One might be signing legally binding contracts in a grocery store without actually knowing that they did that. Because of that, understanding information and implications related to the law is more important than ever. However, without certain knowledge or education, legal language is not that easy to understand and interpret for the common person. Since legal language includes words, phrases, legal terms that are unknown to a regular person, also there are words that are easy to identify however, they take on a different meaning in a legal context (Hyatt 2018). This description of legal language is supported by various other scholars as well. Legal texts have long and complex sentences, include technical vocabulary that is peculiar to the legal register (Trosberg 1997), archaic words, unusual sentence structure, nominalizations and passives, multiple negation and impersonal constructions and redundancy as well (Tiersma 1999). It is also said that legal language is considered to be formulaic and is composed of technical terms, common terms with uncommon meaning, archaic expressions, doublets, formal items and unusual prepositional phrases (Trosberg 1997). It is also worth noting that the Institute for the Languages of Finland (https://www.kotus.fi/en/on_language/plain_language) is in works of developing a plain legal language that would be comprehensible to the masses. That just proves that understanding legal language is not easy and that there is a necessity for a simplified version of it. Because of these kinds of nuisances, the legal language is often simplified in news articles related to the law, so that the majority of the people would still be able to process the information.

Even though there are some studies in regard to the difficulty of legal English, there are not many studies that focus on the simplification itself of legal English, how it occurs and what sort of features are present. That is why this study will compare the two types of discourse. The official case reports of the European Court of Human Rights will represent legal discourse, while news articles of the specific cases chosen will represent media discourse. One of the aims of the present study is to see if simplification can be identified in any way while comparing the two types of discourse. Another aim is to identify the differences and the similarities between the discourses. In the case that simplification is present and identifiable, the final aim of the study is to determine what sort of techniques are used to simplify the text in media discourse and to see what is added and what is lost during simplification, how professional communication turns into broad communication. The hypothesis is that simplification will definitely be present and identifiable in daily news articles.

Literature Review

Simplification

Without certain knowledge or education, legal language is not that easy to understand and interpret for an ordinary person. Since legal language includes words, phrases, legal terms that are unknown to a regular person, also some words are easy to identify however, they take on a different meaning in a legal context (Hyatt 2018). That is why simplification of legal texts seems to be necessary for them to be more accessible. Simplification is often practiced in language teaching, the purpose of which is to simplify the text in terms of content and form for the learners. That is done by bringing the contents of the text within the area of language that is already known by the learners (Bhatia 1983). However, simplifying legal texts is not an easy task, oftentimes it is difficult to replace legal terms with something else, because the synonyms for them do not exist or they do not carry the same meaning and part of the meaning will get lost or the text will become distorted and that is a serious issue. Of course, if a legal term is difficult to substitute, one can provide explanations in the simplified text, however as we can see in Bhatia's (1983) work those explanations can become quite long and elaborate:

- 1) *A tort is any wrongful act or omission that injures another's person, property, or reputation. The party whose person, property, or reputation is damaged has a remedy in the form of a right to sue and recover damages from the wrongdoer. A contract duty is voluntarily assumed, but a tort duty is one imposed by law.*
- 2) *A tort is a legal term which can be regarded as a violation of a kind of 'social contract' which we all enter into, in exchange for the amenities of a common social life in a civilized society. This 'social contract' establishes standards of conduct which we all must meet whether we like it or not. If we fail to meet these standards, we will be committing a tort and thus may be in peril of paying compensation for injuries that we might thus cause to others. The injuries should not be taken literally to mean physical injuries alone. Any person who suffers an injury to his person, property, or reputation has a right to claim compensation in terms of money from the one who happens to be the cause of it. This kind of social contract is not the same as an agreement or a promise. An agreement or a promise is entered into by two persons out of their free will, but a 'social contract' is enforced by law even against our wishes.*

As pointed out by Baker, simplification is connected with explicitation: “simplification involves making things easier for the reader (but not necessarily more explicit), but it does tend to involve also selecting an interpretation and blocking other interpretations, and in this sense, it raises the level of explicitness by resolving ambiguity” (1996:182).

According to Ritter (1981), the longer the category of a legal instrument, the more complicated the language of such instruments; what becomes paramount is the knowledge of the hidden context of case law and statute law. As a result, legal instruments that are usually encountered by ordinary people, such as wills, guarantees or conveyances of real property, are most difficult to understand without legal expertise, whereas custom-drafted agreements for complex commercial ventures are quite comprehensible even for a layman. In her research Ritter (1981) analyzed the Bank’s of Nova Scotia new mortgage form that was supposed to be simplified, so that it would be less confusing to the reader. It was found to be easier to read, but that does not mean that it was easier to understand. In fact, because it was easier to read, it can create an illusion to an ordinary person that they understand all of the implications of the contract when that might not be the case. From this research paper, we can see that when a legal text is not simplified properly, it can become even more confusing. Another study by Masson & Waldron (1994) investigated if a set of legal contracts that were redrafted in three stages would increase the comprehension of the text to a regular reader. In the first stage of redrafting archaic and redundant terms were replaced or removed, in the next stage the word and sentence structure was simplified and in the final stage legal terms were defined or replaced with simpler terms. Comprehension was measured through question and answer tasks and paraphrasing. It was found that through redrafting the legal text was more comprehensive however, it was still not comprehensive enough for the participants to not get confused.

Hartley’s (2000) research explores Redish's (1979) seven propositions concerning legal text. The aforementioned propositions are as follows: (i) many legal documents cannot be read and understood by lay persons; (ii) people without legal training have to read and understand legal documents; (iii) much legal writing is unintelligible, even to lawyers; (iv) tradition – not necessity – and a lack of understanding of the audience – are the major reasons why legal language is so obscure; (v) legal language can be made clear without losing its precision; (vi) it is not the technical vocabulary but the complex sentence structure that makes legal writing difficult to understand; and (vii) clarity is not the same as simplicity, brevity or ‘plain English’. Most of the propositions were found to be still valid, with two possible exceptions: the fifth and the sixth. It is fair to mention that the research paper was published 20 years ago

and during that time legal language could have definitely evolved and perhaps not as many propositions are as relevant as they were in the 2000s, but I do believe that many of them are still relevant. This could be a whole separate research paper if one replicated the study in the present day. Also, since Redish's propositions were published 41 years ago and they still seem to be relevant today, it would seem that legal English is not changing/evolving as rapidly as one would like. But there could be several reasons for the lack of change, tradition being one of them.

The case of simplification is not only present in the sphere of law, but in other important spheres as well, like medicine and science. In a research article by C. Zarcadoolas (2011), the simplification of medical texts was explored. They were trying to see if the information that is provided in a simplified manner would increase the subjects' health literacy, however, it was noticed that through the process of simplification some important information could sometimes be omitted. In that specific study, it was noticed that perhaps the process of simplification was not the best approach to make the medical text seem more comprehensible. The study deals with an interesting case: since simplification was not enough, it was found that "... a richer, more theory-based understanding of text structures and functions, along with other powerful constructs, including cultural appropriateness, relevancy and context, are needed to close the gaps between health messages, health messengers and patients/the public" (Zarcadoolas 2011:1). There were also some studies on simplification being used in second language acquisition and learning. (Atkinson et al., 2018; Leow, 2003; Petersen & Ostendorf, 2007;)

Jury Duty

As mentioned before, understanding legal texts is highly important these days, especially in the US, since the concept of jury duty exists there. Since the jurors are ordinary people, they need to understand instructions and legal context. Randall (2014) in his research paper explores if the jurors understand instructions better that are written in legal English or in "plain" English. It was found, to the surprise of no one, that the simplified versions of the text were more comprehensible to the jury. The results also showed that comprehension varied with linguistic complexity, significantly worse on instructions containing passive verbs and presupposed information, factors known to increase processing load (Randall 2014). However, in the study conducted by Goodman & Greene (1989), simplified instructions were not enough for the jurors, and when asked to paraphrase some legal concepts, less than half of the paraphrases were correct. Another study conducted by Robert and Veda Charrow (1979) also used a similar methodology and had the jurors paraphrasing their received simplified and

regular instructions. It was found that certain grammatical constructions and discourse structures rather than legal terminology caused the most confusion among the jurors. An important question should be raised regarding this piece of research. If the jury may not understand the instructions that are provided fully, due to legal English, can they make a fair decision in judgment without understanding the whole context (Charrow & Charrow 1979)? Perhaps, in this case, it would be better that all of the jury's instructions would be simplified and more comprehensible to a lay person.

Lexical Bundles

As mentioned before, legal discourse is considered to be formulaic (Trosberg 1997). The Google dictionary that is provided by Oxford Languages defines the word *formulaic* as “constituting or containing a set form of words” and lexical bundles are defined as the most often occurring lexical sequences or a set of words in a register (Biber et al. 1999). That is why analyzing lexical bundles is an excellent method of investigation of formulaic languages. Even the lexical bundles themselves used to be called ‘formulas’ among other names like, ‘routines’, ‘fixed expressions’, ‘prefabricated patterns’ (Biber et al. 2004). In Biber's et al. 2004 study the authors analyzed lexical bundles that were present in two different university registers, one being classroom teaching while the other register was textbooks. The lexical bundles were categorized based on their function and structure and an approach that was based on this research was applied to the current study as well. Here are some more studies that analyze lexical bundles in different registers. (Adel & Erman, 2021; Allen, 2010; Biber & Barbieri, 2007; Breeze, 2013; Byrd & Coxhead, 2010; Chen & Baker, 2010; Cortes, 2004; Cortes, 2006; Jablonkai, 2009; Jablonkai, 2010; Juknevičienė, 2009; Pan et al., 2016; Tomankova, 2016;)

As one may notice, some of the above studies are quite old, but it indicates a huge gap in research focusing on simplification and legal language. Therefore the present paper will compare two types of discourse to identify the similarities and differences among the discourses and to see if specific features of simplification can be identified. European Court of Human Rights case reports will represent legal discourse and news articles of specific cases will represent media discourse.

2. Data and Methods

For the purpose of this research, two case studies were carried out. Two legal cases were chosen. The *Delfi AS v Estonia* case was chosen for the first case study. The case was covered by the media quite extensively because the ruling was considered to be quite controversial. It was found that news websites like Delfi can be found accountable for defamatory comments that are left on the website by anonymous users, even though those comments are removed when requested. This ruling was unexpected since if websites can be considered liable for anonymous comments that are left by other people there, to protect themselves, the websites could potentially start censoring comments and our freedom of speech on the internet could be compromised. The next case for analysis that was chosen is *Georgia v Russia (II)*. The case revolves around the military conflict between the two countries when Russia invaded Georgia in 2008. Interestingly a judgment on the case was made only 13 years later, that Russia did commit multiple human rights violations during their attack. The case was widely covered by the media, since it is considered to be a landmark judgment and a landmark case, due to the circumstances of the case being so unusual.

The case reports of the two cases were taken from the official European Court of Human Rights website. The case reports contain the procedure of the case, the facts, relevant law and the judgment. These case reports represented legal discourse and each case report was turned into a legal discourse representative corpus. Then news articles that discussed these specific cases were collected from various internet news portals and blogs that cover legal cases, including *The Guardian*, *SLATE* and others. From these articles, two media discourse representative corpora were compiled.

A total of four corpora were compiled, two for each case study, one representing legal discourse and the other representing media discourse. The legal discourse corpus contained only the official case report from the European Court of Human Rights. The legal corpus of the *Delfi AS v Estonia* case contained 38,626 words, while the legal corpus of the *Georgia v Russia (II)* case contained 116,895 total words. As for the media corpora, the *Delfi AS v Estonia* media corpus was compiled from 8 news articles and the official press release of the European Court of Human Rights. The total number of words contained in the corpus was 13,300. The media corpus of the *Georgia v Russia (II)* case contained 8 news articles as well,

along with the official press release of the European Court of Human Rights. The total number of words contained in the corpus was 15,549.

The analysis of corpus data was performed in two stages. First, word frequency lists were generated using the corpus analysis software AntConc (Anthony 2020). The top 100 most often occurring words were collected from each corpus and analyzed. The purpose was to identify differences on the level of single words. The second stage involved analysis of multi-word units, namely, lexical bundles. Using the same software top 200 4-word lexical bundle lists were also generated for each corpus and analyzed. A structural analysis was performed, categorizing the lexical bundles into these categories: lexical bundles incorporating fragments of noun phrases (henceforth: NP-based); lexical bundles incorporating fragments of verb phrases (henceforth: VP-based); lexical bundles incorporating fragments of dependent clauses (henceforth: DepCl-based); lexical bundles incorporating fragments of prepositional phrases (henceforth: PP-based); lexical bundles incorporating fragments of adjective phrases (henceforth: AP-based); lexical bundles incorporating fragments of adverb phrases (henceforth: AdvP-based). This classification draws on Biber et al. (1999) and Biber et al. (2004). A functional analysis of the lexical bundles was carried out as well. The analysis was based on Biber's et al. (2004) study in which lexical bundles were categorized into the following categories, *referential expressions*, *discourse organizers* and *stance expressions*.

3. Results and Discussion

In this part of the study, all of the analyses that were mentioned in the methods and data section will be discussed for the two case studies. Specific, eye-catching cases are discussed and an interpretation of potential implications those cases may carry was endeavored.

3.1 Word frequency analysis

The top 100 words from each list are analyzed and compared to see what kind of words occur most often in legal discourse and media discourse respectively. Since the news articles and the official case report discuss the same case, by comparing the word frequency lists we can see if the terms and language used in media discourse are more simple and if specific legal terms are avoided and are replaced with more common words so that the text would be more understandable to the general public. Normalized frequency for each term is also provided since the corpora are different in size. Bearing in mind the size of the corpora, absolute frequencies were normalized per thousand words.

3.1.1 Delfi AS v Estonia Case

When looking at the word frequency lists, the top 5 most frequently occurring words in both lists are, *the*, *of*, *to*, *and*, *in*. However, these sorts of words do not carry any informational value for our study so they will be ignored. The word *comments* was occurring often in both corpora. It was the 11th most often occurring word in the legal corpus and the 9th most often occurring word in the media corpus. The normalized occurrence rate for the word *comment* in the legal corpus is 9,5 times per 1,000 words, while in the media corpus the normalized occurrence rate is 14,4 times per 1,000 words. Even though the word does seem to carry some informational value in this case, since the case is related to comments that are left on the news website Delfi, it is no surprise that the word is occurring quite often in both media discourse and legal discourse in this case.

The second most frequent word that stood out is *Delfi*. In the legal corpus word frequency list *Delfi* is only in the 29th place however, in the media corpus word frequency list *Delfi* is in the 8th position. In terms of normalized frequency, the word *Delfi* in the legal corpus occurs 4.5 times per 1,000 words, while in the media corpus it occurs 15.9 times per 1,000 words. There is a significant occurrence rate discrepancy between the two corpora. Since in the media corpus the word occurs more than 3 times more often per 1,000 words than in the legal corpus. However, in the legal corpus word frequency list, the word *applicant* appears in the 25th position and occurs 5 times per 1,000 words. That is the case because in the official case

report it is stated that *Delfi* will be referred to as *the applicant company* further on in the document. On the other hand, in the media corpus word frequency list, the word *applicant* occurs only 5 times in the whole list and it occurs only 0.4 times per 1,000 words. Knowing this, we can see that in the official case report *Delfi* was oftentimes replaced with the phrase *the applicant company* while in the news articles the word *applicant* occurred only 5 times and the phrase *the applicant company* was only seen twice in the whole corpus. It can be stated that most writers of the news article have avoided this legal term and just used *Delfi* instead, since using the word *applicant* might not be very reader-friendly and we could state that in this case proof of simplification has been found. At the same time, due to the phrase *the applicant company* the word *company* in the legal corpus occurs, 5.4 times per 1,000 words, while in the media corpus the same word occurs only 2.1 times per 1,000 words, which is significantly less. Also, it should be mentioned that in the media corpus there were 28 total instances of the word *company* and 18 (64.3%) of the instances referred to a ferry company and not *Delfi*. The ferry company sued *Delfi*, because of the defamatory comments that were left on the website in an article relating to that company. In the legal corpus, most of the cases of the word *company* were referring to *Delfi*. Since news articles surrounding the case do try to summarize it, it is no surprise that the ferry company is mentioned quite a bit. It could be stated that using *Delfi* instead of *the applicant company* is incentivized even more since the reader could get confused which company the phrase refers to since the ferry company can be considered an applicant company as well in the case where they sued *Delfi* because of the defamatory comment's that were left on their website.

The word *Judgment* is in the 24th position on the legal corpus word frequency list and occurred 5.2 times per 1,000 words. In the media corpus word frequency list, the word *judgment* is in the 68th place and occurred 2.1 times per 1,000 words. So, the word *Judgment* has occurred more than twice as often in the legal corpus than in the media corpus. Interestingly, the word *decision* that is a synonym for *judgment*, occurs only 5 times in the whole official case report and occurs 0.1 times per 1,000 words. However, *decision* in the media corpus occurs 2.2 times per 1,000 words and takes up the 64th place in the word frequency list. So again, we can see that journalists might be leaning toward simplifying legal terms and might choose to use the word *decision* instead of *judgment*. The noun *Judgment* does have several meanings; however, it is quite often used in legal language as seen by the word frequency list of the legal corpus. While, the word *decision* is much more neutral and is not really tied to court proceedings, even though it can still definitely occur in legal discourse, just quite a bit less often than in media discourse.

The word *information* is also an interesting case. It takes up the 20th position in the legal corpus word frequency list and it occurs 5.7 times per 1,000 words. In the media corpus word frequency list, the word *information* took up the 96th place and only occurred 1.6 times per 1,000 words. In the legal corpus, the word *information* is used most often while establishing what legal provisions and acts were followed, such as, *Information Society Services Act*, *The E-commerce Directive* and others. While some of the directives and acts are explained in some of the news articles, they are of course summarized greatly, because otherwise a layman would be bombarded with information and it could be quite easy to get lost between the different directives and acts and what sort of implications all of them carry. The *Information Society Services Act* alone was mentioned 19 times in the whole legal corpus and it appears 0.5 times per 1,000 words, so every 2,000 words we could see this act being mentioned. While in the media corpus, the act itself was mentioned only 3 times and it equals to 0.2 times per 1,000 words. The act was found once in the official press release of the European Court of Human Rights and it was also found once in two other news articles, while the rest of the articles have not mentioned this specific act. It is not surprising that the act itself does not appear more in news articles since they have to summarize the case, so they cannot include every single act and directive that was relevant to the proceedings. It is believed however that this sort of summarization and omission of specific acts could also be considered as signs of simplification of the text and legal language.

Since acts were mentioned, the word *act* took up the 64th place in the legal corpus word frequency list and occurred 2 times per 1,000 words. In the media corpus word frequency list, the word *act* took up 132nd place and occurred 1.2 times per 1,000 words. It can be stated that the word *act* occurs 40% fewer times in the media corpus than in the legal corpus. Since the word *act* can be a noun and a verb, it was noticed that in the legal corpus in 10.5 % of all occurrences of *act*, the word appeared as a verb, while in the 89.5 % of occurrences *act* appeared as a noun, usually referring to a specific legal act. In the media corpus, in 18.75 % of occurrences *act* acted as a verb, while in 81.25 % of occurrences the word *act* appeared as the noun. The noun *act* in the media corpus, usually also indicated some sort of legal act, however, there were two instances in one article, where the noun *act* was used to separate the text into act one and act two. Since the word *act* can have many meanings it can be used differently in different contexts, but it is definitely part of the legal lexicon since legal acts are a very important part of law in general. In this case, legal acts provide context for the Delfi AS v Estonia case, but as mentioned before, news articles cannot and should not include every single act that was mentioned in the case since that would be too much information for

the layman, that is why it is believed there is a discrepancy in how often the word *act* appears in the legal corpus and the media corpus.

The word *law* took up the 33rd place in the legal corpus word frequency list and occurred 4.1 times per 1,000 words in the corpus. In the media corpus word frequency list, the word *law* took up the 51st position and occurred 0.9 times in the corpus. It is quite surprising that there is such a large discrepancy here since the word *law* in a legal sense is hard to replace, there are no real synonyms to it, but it can also depend on the context. However, in the legal corpus, there is a large number of different types of law indicated there. For example:

3) ... *whereas libel and slander are the weak corresponding rights in Anglo-American law.*

(DvE-LC)

4) *The applicant company argued that there was no domestic law according to which...*

(DvE-LC)

5) *The Chamber was furthermore satisfied that the relevant provisions of the civil law – although they were quite general and lacked detail in comparison with, for example, the Information Society Services Act – along with the relevant case-law, made it clear that a media publisher was liable for any defamatory statements made in its publication.*

(DvE-LC)

6) *The issue was that there were two laws and the applicant company believed that the Directive was applicable as European Union law and as special law, whereas the Supreme Court took the view that the other law was applicable, because the service provider was a publisher.*

(DvE-LC)

7) *This issue, at the crossroads of civil law (persons, contracts, property), criminal law and human rights law and also civil and criminal procedure, has given rise to an extensive body of case-law at domestic and European level, based on various sources*

of national law (Civil Code, Criminal Code, Data Processing and Civil Liberties Act, Freedom of the Press Act) and international law...

(DvE-LC)

As seen through the examples, numerous types of laws are constantly mentioned and repeated throughout the official case report (legal corpus). In the case of the media corpus, here are some examples of the word *law* that occurs there:

8) *The comments complained of were clearly unlawful under Estonian law and included hate speech and incitements to violence.*

(DvE-MC)

9) *Thus, it did not address the issue under EU law and limited itself to the question of whether the Supreme Court's application of the domestic law to Delfi's situation had been foreseeable.*

(DvE-MC)

10) *That potential liability is at odds with the prevailing view under English law...*

(DvE-MC)

11) *... the Estonian Courts had failed properly to apply Estonian national law implementing Europe's safe harbour regime...*

(DvE-MC)

12) *This case demonstrates the difficulty for site operators in a multi-platform reality where the boundaries between UGC and content are too blurred to accurately apply previous case law.*

(DvE-MC)

In the media corpus, different types of law are mentioned, but they do seem to be mostly generalized (Estonian law, EU law, domestic law, Estonian national law). When compared to the types of law that are mentioned in the official case report, they do mention the more generalized types of law (Anglo-American law, domestic law, national law), however; in the legal corpus, specified types of law are found often as well (civil law, criminal law, human rights law, case law at domestic and European level). In the media corpus these specified types of law are not found, but since the article writer's job is to summarize the case and make it as reader-friendly as possible, it is no surprise that not every specified type of relevant law is mentioned there and the types of law are generalized to Estonian law or domestic /

national law. In this case, perhaps it could be stated that by generalizing the relevant law of the case it could be considered as a form of simplification as well.

A similar case happened with the word *article*. In the legal corpus word frequency list, the word takes up the 35th place and occurs 4 times per 1,000 words. In the media corpus word frequency list, the word takes up the 29th position and occurs 5.2 times per 1,000 words. So surprisingly the word *article* occurs more often in the media corpus. It should be mentioned that the word contains several meanings, two meanings were used in both the media corpus and the legal corpus. One of the meanings being that the word refers to news articles, for example, news articles that Delfi wrote. The other meaning of the word refers to articles of different directives and conventions. The second meaning was used significantly more often in both legal and media corpora. However, it would be fair to note, that the number of different articles of conventions and directives differs quite significantly. In the media corpus, there were 8 different articles mentioned, while in the legal corpus there were 20 different articles mentioned in the text. Similarly, as with laws and acts, it is not expected of the article writers to mention every single article that is relevant to the case because it would overcomplicate the text to a layman and it is believed that news articles should be concise and get to the point. That is why even though the word *article* occurs more often in the media corpus, that is the case because they are emphasizing the main relevant articles of the case.

Another interesting case that is worth analyzing is the verb *may*, it takes up the 63rd position in the legal corpus word frequency list. In terms of occurrence, it occurs 2 times per 1,000 words in the legal corpus. In the media corpus, it takes up the 143rd place in the word frequency list and occurs 1.1 time per 1,000 words. The verb *may* in the legal corpus occurs almost twice as often when comparing to the media corpus. The verb itself can definitely be considered a part of legal language since in legal texts it is very prominent and often occurring. For example, if we would take up the synonym *can*, in the legal corpus word frequency list it takes up the 160th position and occurs 0.8 times per 1,000 words in the legal corpus, which is significantly less when compared to the verb *may* that occurs 2 times per 1,000 words. In the media corpus word frequency list, *can* takes up the 100th place and occurs 1.5 times per 1,000 words, which is more often than the occurrence rate of 1.1 of the word *may* in the media corpus. Based on this statistic we can state, that *may* is definitely more prominent in legal language and is part of the legal formal register. While in the media discourse it would seem that the verb *can* is more preferable than *may*. Journalists are not expected to follow the legal register, but it is a significant difference when comparing the two discourses. Another verb that is considered to be quite prominent in legal discourse is the verb

shall. The verb itself, occurred 0.8 times per 1,000 words in the legal corpus, however, it did not appear at all in the media corpus, which is quite interesting. However, the verb *will* which would be considered the closest synonym to *shall* and is not as formal does occur in the media corpus 1 time per 1,000 words. Another often occurring verb in legal discourse is *must*. In the legal corpus, it occurred 1.2 times per 1,000 words, while in the media corpus the occurrence rate was 0.5 times per 1,000 words. It can be seen that even though the news articles and the case report discuss the same case, the common verbs in legal discourse *must*, *may* and *shall* are not as prevalent and as often occurring in media discourse, and less formal verbs such as *can* and *will* are more common in media discourse. That is quite understandable since the media discourse target audience is the layman and the articles should not be too formal, the more informal register could also be considered a sign of simplification when comparing the two discourses.

An interesting case was noticed regarding the word *offensive*. In the legal corpus, the word occurred only 9 times in total and took up the 546th position in the word frequency list. The occurrence rate per 1,000 words was 0.2 times. However, in the media corpus word frequency list, the word *offensive* took up the 84th place and occurred 1.7 times per 1,000 words. This discrepancy was unexpected. It was noted that the most often occurring collocation to the word *offensive* was *comments*. As a result of that, collocations of the word *comments* were checked in the legal corpus as well as the media corpus, to see if there are any synonyms present that could have been used instead of the adjective *offensive*. Plenty of synonyms were found. The phrase *offensive comments* occurred 0.1 times per 1,000 words in the legal corpus. In the media corpus, the same phrase occurred 1.1 times per 1,000 words. Then, the phrase *defamatory comments* occurred 0.4 times per 1,000 words in the legal corpus. In the media corpus, the phrase occurred 0.8 times per 1,000 words. Another phrase that was found was *derogatory comments* it only occurred once in the whole legal corpus and did not occur in the media corpus at all. Next, *impugned comments* was another phrase that was found, it occurred 0.1 times in the legal corpus and 0.2 times in the media corpus per 1,000 words. There was also the phrase *sensitive comments* it occurred once in the legal corpus and did not occur at all in the media corpus. Another phrase that was found was *racist comments* it occurred twice in the legal corpus and once again did not occur in the media corpus. *Harsh comments* was another found phrase that occurred twice in the whole legal corpus and did not occur in the media corpus. Next was the phrase *offending comments* which occurred 0.1 times in the legal corpus and 0.5 times in the media corpus per 1,000 words. The phrase *unlawful comments* occurred 0.4 times in the legal corpus and 0.8 times in the media corpus per 1,000 words.

Another phrase that was found was *inappropriate comments* which occurred 4 times in the whole legal corpus and did not occur in the media corpus at all. *Degrading comments* was yet another phrase that occurred once in the legal corpus and did not occur in the media corpus. *Harmful comments* occurred once in the whole media corpus but, did not occur in the legal corpus. The phrase *insulting comments* occurred 0.03 times in the legal corpus and 0.2 times in the media corpus per 1,000 words. Finally, the phrase *vulgar comments* occurred once in the whole media corpus and did not occur at all in the legal corpus. Interestingly, it can be seen that the occurrence rate of these phrases is usually greater in the media corpus when they are present in both corpora. However, the number of different synonymous adjectives to the adjective *offensive* is greater in the legal corpora. There were 11 different synonyms used in the legal corpus, while there were only 7 in the media corpus. Adjectives like, *derogatory*, *sensitive*, *racist*, *harsh*, *inappropriate* and *degrading* were not used in the media corpus when discussing the comments left on the Delfi news portal. On the other hand, adjectives like, *harmful* and *vulgar* were not used in the legal corpus, when discussing the comments. It can be seen that a greater variety of adjectives in this context was used in the official case report rather than in the media articles. It should be mentioned that news articles usually do have scope limitations, while the case report needs to be as descriptive as possible. Because of the scope limitations, perhaps the writers do not need to use that many synonymous adjectives and can get away with using *offensive*, *defamatory* and *unlawful* which were the most used adjectives in the media corpus when it comes to describing the comments left on the Delfi website.

3.1.2 Georgia v Russia Case

Once again, just like with the word frequency lists for the Delfi AS versus Estonia case the top 5 most occurring words in the case of Georgia versus Russia both media and legal corpora are *the*, *of*, *to*, *and*, *in*, not necessarily in that order. But once again since these words are necessary for sentence structure and they do not provide us with any informative value for this case study, they will not be analyzed.

To start with, the word *Georgia* was looked into, in the media corpus word frequency list, it took up the 19th place and it occurred 6.6 times per 1,000 words. While in the legal corpus word frequency list the same word was found in the 76th position and the occurrence rate was 1.7 times per 1,000 words. In the media corpus, the word *Georgia* occurred almost 4 times more often than in the legal corpus. This discrepancy, at least partially, can be attributed to the phrase *the applicant government* which was often times used in the case official report of the European Court of Human Rights instead of *Georgia*. The word *applicant* itself occurred

1 time per 1,000 words in the legal corpus. However, in the media corpus *applicant* occurred 0.8 times per 1,000 words. So the word *applicant*, did occur less often in the media corpus however, the discrepancy is not as large as it was expected to be. As we can see, similarly as with the Delfi AS versus Estonia case, the word *applicant* did appear less often in news articles, but the discrepancy was not as large as in the Delfi AS versus Estonia case. The word itself definitely represents the legal language and is part of the legal vocabulary. The word *Russia* should also be taken into account. In the media corpus word frequency list *Russia* takes up the 15th place and occurs 7.5 times per 1,000 words. In the legal corpus, the word takes up the 77th place and occurs 1.7 times per 1,000 words. It can be seen that the word *Russia* appears over 4 times more often in the media corpus than in the legal corpus. Then again, the discrepancy can also be partially explained by the phrase *the respondent government* which was often used in the official case report instead of *Russia*. *Respondent* the word itself, in the legal corpus, appears 1.4 times per 1,000 words. In the media corpus, the same word occurs 0.4 times per 1,000 words. It has been noticed that *Respondent* appeared over 3 times more often in the legal corpus than in the media corpus. Similarly, as with *applicant*, the word *respondent* can appear in legal language quite often and is also, definitely part of the legal vocabulary. However, in the journalistic discourse, in our current example, the word *respondent* does not seem to be very prominent. Even though, the words *Russia* and *Georgia* appeared more often than the words *applicant* and *respondent* in both corpora, the discrepancy in the legal corpus is not that large. In the legal corpus, *Georgia* appeared 1.7 times per 1,000 words while the word *applicant* appeared 1 time per 1,000 words. The word *Russia* appeared 1.7 times per 1,000 words, while the word *respondent* appeared 1.4 times. In the media corpus, the discrepancy is quite a bit larger in the usage of these words. The word *Georgia* appears 6.6 times per 1,000 words, but the word *applicant* appeared only 0.8 times per 1,000 words. The occurrence rate of *applicant* is over 8 times less than the occurrence rate of *Georgia*. While the word *Russia* appeared 7.5 times per 1,000 words in the media corpus, but *respondent* appeared only 0.4 times per 1,000 words. It can be seen that the word *respondent* occurs over 18 times less often than the word *Russia*. Based on these statistics it is possible to state that referring to the *applicant* and the *respondent* of the case as *Georgia* and *Russia* is much more preferable in journalistic discourse than in legal discourse since the terms of *applicant* and *respondent* can be considered to be quite legal terms. The avoidance of using *applicant* and *respondent* could be a sign of simplification, since the layman, might possibly get confused among these terms, and just using the country names when talking about the case should make the text more clear to the general public.

3.2 Lexical Bundle Analysis

In the second part of the study, 4-word lexical bundles were analyzed. As mentioned previously, lexical bundles are defined as the most frequent lexical sequences in a register (Biber et al. 1999). Since legal discourse and media discourse are analyzed in this study in terms of simplicity vs complexity of expression to the layman, lexical bundle analysis was expected to provide useful insights, especially since legal discourse is considered to be much more formal, while media discourse is more informal. Because of the formality factor, it is expected that the lists of the most often occurring 4-word lexical bundles will be quite different in the two corpora. For each corpus, 200 of the most often occurring 4-word lexical bundles were extracted. Specifically, 4-word lexical bundles were chosen, because it is believed that this would be the most optimal number of words to see a clear picture as it has been evidenced by other studies (Biber et al. 2004, Hyland 2008).

3.2.1 Structural Analysis

For this part of the study, 200 most frequently occurring lexical bundles were categorized for each corpus and for each case. The lexical bundles were categorized into the following categories, *noun phrased based*, *verb phrase-based*, *dependent clause based*, *prepositional phrase-based*, *adjective phrase-based* and *adverb phrase-based*. In terms of categorization, a linear approach from left to right was chosen, meaning that since 4-word lexical bundles can include multiple phrase classes in one bundle, the first phrase class that is encountered in the bundle determined which phrase-based bundle it is considered to be.

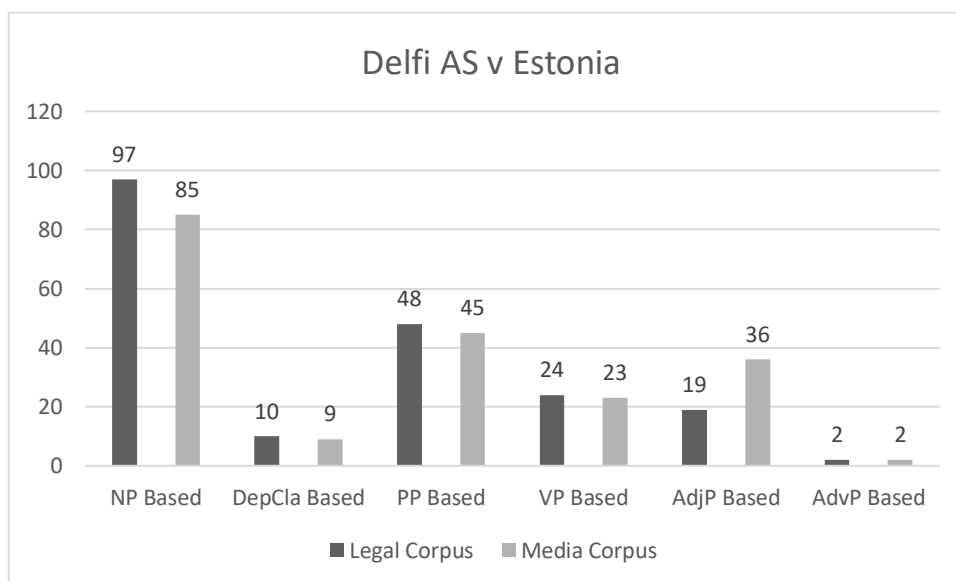


Figure 1.1 Case Delfi AS v Estonia lexical bundle structure distribution

When taking a look at the Delfi AS v Estonia case lexical bundles collected from both the legal and the media corpora, we can see that the most frequently occurring lexical bundles in terms of their structure are distributed quite evenly between the legal and the media corpus. In Figure 1.1, it can be seen that out of the most often occurring lexical bundles the *noun phrase-based* bundles were the most popular in both corpora which is quite the norm in the English language. In the legal corpus, noun phrase-based lexical bundles took up 48,5% of all analyzed bundles in that corpus, while in the media corpus 42.5% of the lexical bundles were noun phrase-based. Next, were the *prepositional phrase-based* lexical bundles, in the legal corpus, there were 24% of them, while in the media corpus they consisted of 22.5% of all analyzed bundles in that corpus. The *adjective phrase-based* bundles took up 9.5% in the legal corpus and in the media corpus they took up 18% which is quite a significant difference since there was almost double the amount of *adjective phrase-based* lexical bundles in the media corpus when compared to the legal corpus. Then there were the *verb phrase-based* bundles in the legal corpus they took up 12% of the analyzed bundles and in the media corpus they took up 11.5%. The *dependent clause-based* lexical bundles took up 5% of the legal corpus lexical bundles and 4.5% of the media corpus lexical bundles. Lastly, the *adverb phrase-based* lexical bundles took up 1% in each of the corpora.

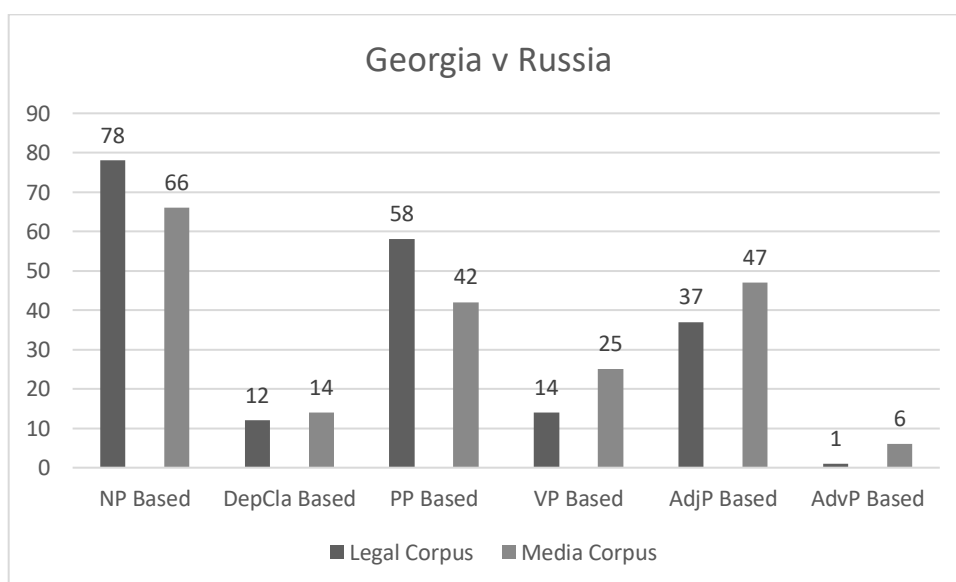


Figure 1.2 Case Georgia v Russia lexical bundle structure distribution

Even though in the Georgia v Russia lexical bundle lists, the distribution of categories is a bit more varied than in the Delfi AS v Estonia lexical bundle lists, the most popular categories still remain the same. As seen in Figure 1.2, noun phrase-based lexical bundles being the

most popular and taking up 39% of the lexical bundle analyzed in the legal corpus and 33% in the media corpus. Then prepositional phrase-based bundles took up 29% of the analyzed bundles in the legal corpus and 21% in the media corpus. It should be noted that there is an 8% discrepancy here when comparing the two corpora. Next, the adjective phrase-based lexical bundles took up 18.5% of the analyzed bundles in the legal corpus and 23.5% in the media corpus. The verb phrase-based lexical bundles occupied 7% of the analyzed bundles in the legal corpus and 12.5% in the media corpus. Dependent clause-based bundles take up 6% of the analyzed lexical bundles in the legal corpus and 7% in the media corpus. Finally, adverb phrase-based lexical bundles took up only 0.5% of the analyzed bundles in the legal corpus and 3% in the media corpus.

It should be noted, that in the media corpus of the Georgia v Russia case there were 11 more verb phrase-based lexical bundles and 5 more adverb phrase lexical bundles than in the legal corpus. While the legal corpus contained 12 more noun phrase lexical bundles and 10 more adjective phrase lexical bundles than the media corpus. In the Delfi AS v Estonia case corpora, most of the results were quite even except for the legal corpus having 12 more noun phrase-based bundles than in the media corpus and the media corpus having 17 more adjective phrase-based bundles than the legal corpus. The discrepancy of the adjective phrase lexical bundle could perhaps be explained by the more informal nature of media discourse that allows the journalists to be more descriptive and in turn, they would use more adjective phrase-based lexical bundles. While the official case report is more formal and it could perhaps be stated that it is more detail-oriented and it covers every important and relevant detail of the case, that is perhaps why the official case reports tend to have more noun phrase-based lexical bundles.

3.2.2 Functional Analysis

In this part of the study, lexical bundles were analysed in terms of their functions drawing on classification proposed in Biber's et al. (2004) study. There were three categories that the lexical bundles were categorized to, *referential expressions*, *discourse organizers* and *stance expressions*. Stance bundles express attitudes or assessments of certainty, discourse organizers reflect the relationship between prior and coming discourse and referential bundles make direct reference to physical or abstract entities, or to the textual context itself (Biber et al. 2004). Since the same lexical bundle appears multiple times in a corpus and in some cases may perform different functions, the concordance was checked of each bundle and their category was determined by taking into account which function the lexical bundles performed the most frequently.

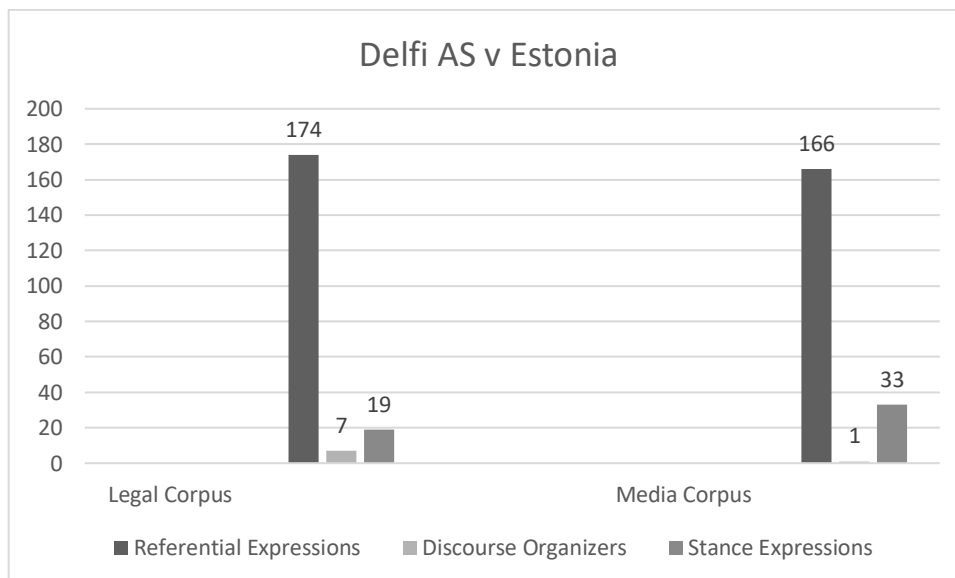


Figure 2.1 Case Delfi AS v Estonia lexical bundle distribution according to their function

Taking a look at the lexical bundles from the Delfi AS v Estonia case legal and media corpora referential expressions were the most prominent in both corpora as seen in Figure 2.1. While there were 14 more stance expressions in the media corpus and 6 more discourse organizers in the legal corpus.

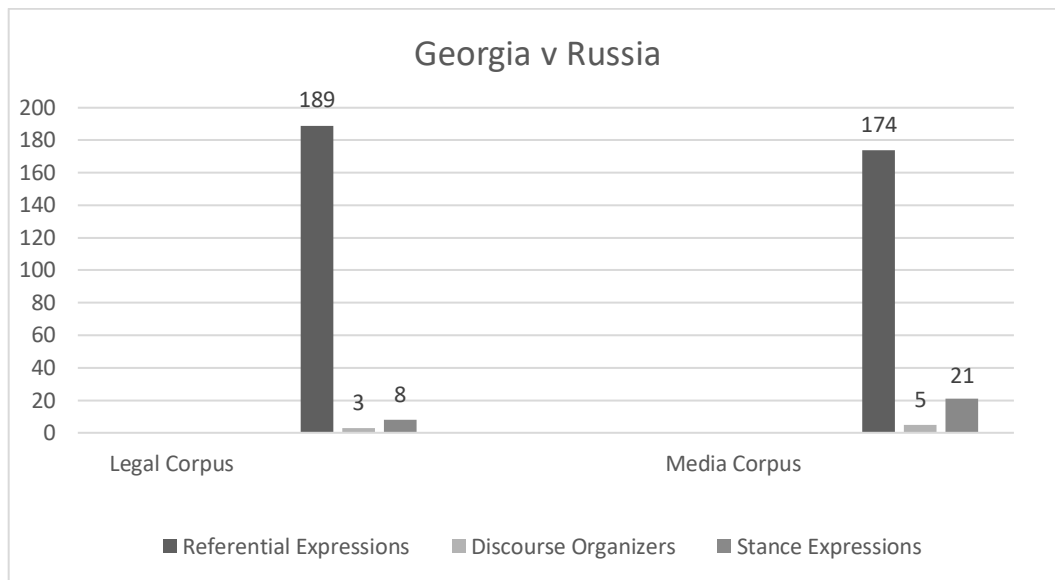


Figure 2.2 Case Georgia v Russia lexical bundle distribution according to their function

As for the functional analysis of the Georgia v Russia case lexical bundles in both media and legal corpora, once again referential expressions were the most common by far. In the media corpus, there were 13 more stance expressions and 2 more discourse organizers than in the legal corpus. Even though the differences are not very significant, it can be noted that in both of the cases the media corpus did contain more stance expressions. Journalists do tend to explain what sort of implications a certain judgment of the case might have for the future and that is perhaps why more stance expressions were found in media corpora. It also should be noted that theoretically, the media should be unbiased, perhaps that is why there are not that many stance expressions overall in the media corpora when compared to referential expressions for example.

The abundance of referential expressions in both cases and in all 4 corpora has to be addressed. To describe a case, quite a lot of details and context has to be provided for it to make proper sense and while providing context and explaining the details the official case report and the news articles use a great deal of referential expressions. The cases that reach the European Court of Human Rights are usually quite complex since otherwise they would have been resolved in other instance courts, because of the complexity of the cases the amount of details that needs to be provided is greater as well. Referential expressions can be either topic-specific and related to the case or they can be general and do not include any words that relate to the case at all, or those words are just entirely neutral. In the legal corpus of the case *Delfi AS v Estonia*, 18.4% of the referential expressions were general, while 81.6% of the expressions were topic-specific. Since the legal corpus consists of the official case report, it is not very surprising that most of the referential expressions are topic-specific

and relating to the case itself in one way or another. Here are some examples of the general referential expressions that were present in the corpus:

13) *At the same time, because of the nature of Internet media, it cannot reasonably be required of a portal operator to edit comments before publishing them in the same manner as applies for a printed media publication.*

(DvE-LC)

14) *In such cases, unlawfulness shall be established on the basis of a comparative assessment of the different legal rights and interests protected by law.*

(DvE-LC)

15) *Where an information society service is provided that consists of the transmission in a communication network of information provided by a recipient of the service...*

(DvE-LC)

Here are some examples of topic-specific referential expressions as well:

16) *The part for reading comments left by others was a separate area which could be accessed by clicking on the “read comments” button. The comments were uploaded automatically and were, as such, not edited or moderated by the applicant company.*

(DvE-LC)

17) *The legal dispute between the parties relates to whether the defendant as an entrepreneur is the publisher within the meaning of the Obligations Act, whether what was published (the content of comments) is unlawful, and whether the defendant is liable for the publication of comments with unlawful content.*

(DvE-LC)

Surprisingly, in the media corpus, of the same case, only 5.4% of referential expressions were general and 94.6% were topic-specific. Even though the case is the main topic of the news articles it was expected that there would be a more significant portion of general referential expressions present in the media corpus. Here is an example of the general referential bundles in the media corpus:

18) *Thirdly, the steps taken by Delfi to prevent or remove without delay the defamatory comments once published had been insufficient.*

(DvE-MC)

Also, here are some of the examples of topic-specific referential expressions taken from the media corpus:

19) *The Grand Chamber of the European Court of Human Rights ruled last week that the Estonian news site, Delfi, can be held liable for defamatory comments made by its users...*

(DvE-MC)

20) *After the latter decision, many organisations including Article 19 spoke out about the threat to freedom of expression this ruling posed.*

(DvE-MC)

When looking at the legal corpus of the case Georgia v Russia, it can be seen that 12.7% of the referential expressions were general and 87.3% of the expressions were topic-specific. The results are quite similar to the results of the legal corpus of the Delfi AS v Estonia case. Here are some examples of the general referential expressions that were found in this legal corpus:

21) *They added that as a result of the conflict, and the campaign of ethnic cleansing that followed, many thousands of ethnic Georgians had fled from South Ossetia and adjacent territories.*

(GvR-LC)

22) *At the end of the day, it seems that occupation does not automatically trigger spatial jurisdiction in the occupied territory, let alone personal jurisdiction, because all belligerent occupiers are equal in the eyes of the Court, but some are more equal than others.*

(GvR-LC)

Some of the topic-specific referential expressions that were found in the legal corpus look like this:

23) *Finally, Article 53 of the Convention requires a Contracting Party occupying the territory of another Contracting Party to respect the rules of international humanitarian law*

(GvR-LC)

24) *The violations of the Convention fall within the jurisdiction of the Russian Federation under Article 1 of the Convention...*

(GvR-LC)

Taking a look at the media corpus of the Georgia v Russia case, 15.5% of the referential expressions were general and 84.5% of the referential expressions were topic-specific. When comparing the results to the legal corpus results of the same case, the media corpus does have a bit larger percentage of general expressions, but the difference is not very significant. If compared to the media corpus of Delfi AS v Estonia case, it can be seen that the difference in the number of general expressions is a bit more substantial. Here are some of the general referential expressions that were present in the media corpus of the Georgia v Russia case:

25) *The Court therefore attached decisive weight to the fact that the very reality of armed confrontation and fighting between enemy military forces seeking to establish control over an area in a context of chaos not only meant that there was no “effective control” over that area, but also excluded any form of “State agent authority and control” over individuals.*

(GvR-MC)

26) *Moreover, as the dissenting judges note, it is at odds with the fact that the decision to go to war and its implementation are themselves the classic exercise of sovereign power...*

(GvR-MC)

Also here are some of the referential expressions that were topic-specific, taken from the media corpus of the Georgia v Russia case:

27) *This included the torture of Georgian prisoners of war and the expulsion of Georgian villagers from their homes in South Ossetia.*

(GvR-MC)

28) *That the majority judges in Georgia v Russia (II) fail to refer to the recent seminal General Comment on the right to life published by the UN Human Rights Committee is a surprising and significant omission.*

(GvR-MC)

All in all, it can be seen, that in all of the corpora, from the analyzed referential expressions most of them are topic-specific and are related to a specific case. It is surprising, that the media corpora did not have a larger number of general referential expressions, but since only the top 200 of the lexical bundles were analyzed, perhaps more of them could be seen further down on the list since either they did not occur that often or there might have been just a high variety of different general referential expressions.

3.2.3 Delfi AS v Estonia Case Qualitative Analysis

In the Delfi AS v Estonia case legal corpus, there were a total of 37,716 4-word lexical bundles found, including repetitions, while in the media corpus, there were a total of 13,297 4-word lexical bundles found including repetitions. Because of this discrepancy and because the corpora are of different sizes, the frequency will be normalized to how often the lexical bundles occur per 1,000,000 words. This normalization level was chosen for technical reasons in order to obtain more reasonable values and avoid small decimals. Due to scope limitations, not all of the lexical bundles could be discussed in detail. The lexical bundles that were chosen to be discussed either had large enough discrepancies among the corpora or they carried high informational value based on using the concordance tool on AntConc (Anthony 2020).

The most frequent 4-word lexical bundle in the legal corpus was *Delfi AS v Estonia* and it had the occurrence rate of 2,200.6 times per 1 million words. In the media corpus, the same lexical bundle took up the 10th place in the most frequent 4-word lexical bundle list and occurred 676.7 times per 1 million words. Even though this lexical bundle is over 3 times more frequent in the legal corpus than in the media corpus, it can be stated that the bundle is still fairly prominent in both corpora. This result is not surprising considering, that the case itself is the main subject matter of the official European Court of Human Rights case report as well as the articles that the media corpus consists of. In the media corpus, the lexical bundle occurs in sentences that usually refer to the judgment of the case. Here are some examples of it:

29) *On 16 June 2015, the European Court of Human Rights delivered its final judgment in Delfi AS v. Estonia.*

(DvE-MC)

30) *In June 2015 the ECtHR handed down its ruling in the case of Delfi AS v Estonia.*

(DvE-MC)

31) *On 16 June 2015 the Grand Chamber of the European Court of Human Rights has delivered the long awaited final judgment in the case of Delfi AS v. Estonia...*

(DvE-MC)

There were also two instances when the name of the case was included in the title of the articles:

32) *Delfi AS v. Estonia: Grand Chamber confirms liability of online news portal for offensive comments posted by its readers.*

(DvE-MC)

33) *Delfi AS v. Estonia: a blow to free expression online.*

(DvE-MC)

As can be seen from the examples, the journalists do use the name of the case in certain cases, which is understandable, since the name indicates what sort of parties are involved in the case, which makes it easier for the reader to understand who is suing who. However, in the legal corpus, most instances of *Delfi AS v Estonia* occurred as a title of each page of the case report. With the full title of the page being either, *DELFI AS v. ESTONIA JUDGMENT* or *DELFI AS v. ESTONIA JUDGMENT – SEPARATE OPINIONS*. The lexical bundle occurred more often in terms of formatting of the case report rather than in actual sentences or legal discourse. In terms of occurrence in actual sentences, the lexical bundle occurs more often in the media discourse. It does make sense because the case in itself is described in the case report, so the name of the case is rarely needed because the text refers to itself using a different phrase which is *in the present case*. The lexical bundle *in the present case* takes up the 6th place in the legal corpus lexical bundle list and has the occurrence rate of 673.1 per 1 million words. Here are some examples of how this lexical bundle is used in the official case report:

34) *In the opinion of the Chamber, in the present case both the defendant and the authors of the comments are the publishers of the comments within the meaning of the Obligations Act.*

(DvE-LC)

35) *The article in issue in the present case attracted 185 comments, apparently well above average.*

(DvE-LC)

36) *In the present case the parties' opinions differed as to whether the interference with the applicant company's freedom of expression was "prescribed by law".*

(DvE-LC)

Using *the present case* instead of *Delfi AS v Estonia* in the official case report makes much more sense since otherwise, it would just make the text awkward. In the media corpus, the lexical bundle *in the present case* has the occurrence rate of 225.6, which is almost 3 times less than in the legal corpus. Based on these results it can be stated that the prepositional phrase *in the present case* is more common in legal discourse than in media discourse.

In the legal corpus 4-word lexical bundle frequency list, it was noticed that the phrase *the applicant company* was appearing very often throughout the list. In the third place on the list, the bundle *the applicant company's* was present with the occurrence rate of 1,139.1. In the fourth position on the list the lexical bundle *that the applicant company* was present and had the occurrence rate of 880.2. Then, in the 10th place on the list, the lexical bundle *the applicant company was* appears and the occurrence rate for it is 569.6. Next, in the 11th place, there is the lexical bundle *by the applicant company* with the occurrence rate of 543.7. In the 14th position, there was also the phrase *the applicant company had* with the occurrence rate of 466. Also, in the 34th place in the list, the lexical bundle *to the applicant company* is present and the occurrence rate for it is 310.7. In the 38th position, the lexical bundle *of the applicant company* appears and has the occurrence rate of 284.8. Then, in the 45th place of the list, the phrase *for the applicant company* is there and its occurrence rate is 258.9. Finally, in the 65th position, the lexical bundle *on the applicant company* appears and has the occurrence rate of 233. As one can see, the phrase *the applicant company* appears very often in the list, oftentimes surrounded by different prepositions. In fact, out of the top 70 most often occurring 4-word lexical bundles in the legal corpus, lexical bundles that contain the noun phrase *the applicant company* take up 9 positions, which is 12.85% of the top 70 positions and that is quite a significant portion. *The applicant company* in the official European Court of Human Rights case report refers to Delfi itself as mentioned in the word frequency analysis. Also, it is fair to mention that in the word frequency analysis the word *applicant* was also considered to be very significant and the lexical bundle analysis shows the same result. Here are some examples of the noun phrase *the applicant company* in the legal corpus, as it appears in sentences, surrounded by different prepositions, pronouns and verbs:

37) *The Court of Appeal rejected the applicant company's argument that its liability was excluded under the Information Society Services Act.*

(DvE-LC)

38) *It observed that the applicant company had indicated on its website that comments were not edited...*

(DvE-LC)

39) *The applicant company was represented by Mr V. Otsmann and Ms K. Turk, lawyers practising in Tallinn.*

(DvE-LC)

40) *The comments were uploaded automatically and were, as such, not edited or moderated by the applicant company.*

(DvE-LC)

41) *The applicant company had made efforts to advise users that the comments did not reflect its own opinion and that the authors of comments were responsible for their content.*

(DvE-LC)

42) *The comments in question were removed by the applicant company some six weeks after they were uploaded on the website, upon notification by the injured person's lawyers to the applicant company (see paragraphs 17-19 above).*

(DvE-LC)

43) *At the hearing of 28 May 2007, the representatives of the applicant company submitted, inter alia, that in cases like that concerning the "Bronze Night" ...*

(DvE-LC)

44) *Fourthly, as regards the consequences of the domestic proceedings for the applicant company...*

(DvE-LC)

45) *There was no law imposing an obligation on the applicant company to proactively monitor user comments.*

(DvE-LC)

As can be seen from the examples, the aforementioned lexical bundles, in the official case report, are used to identify what has happened to Delfi and what the company did or did not do. Interestingly, in the whole media corpus, there are only 2 sentences that contain the words *applicant company*. Also, while looking at the 4-word lexical bundles that contain the word *Delfi* in the media corpus, there are 189 different 4-word lexical bundles with the word *Delfi* in them, while in the legal corpus there are 166 different 4-word lexical bundles that contain the word *Delfi*, even though, the legal corpus is roughly 3 times larger than the media corpus. In terms of 4-word lexical bundles that contain the word *applicant*, in the media corpus, there are only 5 different lexical bundles, while in the legal corpus 144 different lexical bundles. In light of these statistics, it can still be stated that it seems that journalists at least when discussing the case of Delfi AS v Estonia do seem to be avoiding the term *applicant* and *applicant company* and are instead just using *Delfi* for the sake of clarity for the reader. The same finding was found during the word frequency analysis and the lexical bundle analysis seems to support it as well.

Another 4-word lexical bundle that needs to be addressed is *of the obligations act* it takes up the 7th place in the legal corpus lexical bundle list and has the frequency of 647.2 in a million words. Here are some examples of how the lexical bundle looks in the official case report:

46) *The Chancellor of Justice referred to the legal provisions designed to ensure freedom of expression as well as the protection of everyone's honour and good name, including sections 1043 and 1046 of the Obligations Act (Võlaõigusseadus).*

(DvE-LC)

47) *However, the Supreme Court found that because of their "economic interest in the publication of comments, both a publisher of printed media and an Internet portal operator [were] publishers/disclosers" for the purposes of section 1047 of the Obligations Act.*

(DvE-LC)

48) *Thus, the Court confines itself to examining whether the Supreme Court's application of the general provisions of the Obligations Act to the applicant company's situation was foreseeable for the purposes of Article 10 § 2 of the Convention.*

(DvE-LC)

As can be seen from the examples, the lexical bundle is used to refer to different sections and provisions of the obligations act, which is quite important in the case report, because the report itself is very detail orientated and has to include every relevant act and provision. Since *of the obligations act* occurs quite often in the legal corpus, the act itself seems to be of high importance to the case. In terms of the media corpus, the lexical bundle *of the obligations act* could not be found, however, there was a 4-word lexical bundle containing the words *the obligations act*, the lexical bundle was *and the obligations act*. The occurrence rate of the lexical bundle in the media corpus is 225.6 per million words. The occurrence rate is almost 3 times less than the occurrence rate of *of the obligations act* in the legal corpus. Here are some examples of *and the obligations act* in the media corpus:

49) *On the basis of this classification, the Estonian domestic courts applied the Civil Code and the Obligations Act, rather than the E-Commerce Directive and the Estonian Information Society Services Act.*

(DvE-MC)

50) *Delfi argued in particular that the national courts had erred in applying the general provisions of the domestic law – in particular the Constitution, the Civil Code (General Principles) Act and the Obligations Act – to the facts of the case as they should have relied upon the domestic and European legislation on Internet service providers.*

(DvE-MC)

51) *The Supreme Court therefore held Delfi liable under the relevant domestic law, notably the Constitution, the Civil Code (General Principles) Act and the Obligations Act, finding that the portal had not only failed to prevent the publication of comments which degraded human dignity, contained threats and were thus clearly unlawful but also to remove the comments from its portal on its own initiative.*

(DvE-MC)

As seen from the examples, in the media corpus the lexical bundle is used just to mention the obligations act among other relevant codes, acts and directives. But no specific sections or provisions of the obligations act were mentioned in the media corpus. This could also be a sign of simplification since even though the obligations act is important to the case it is not really expanded upon in the articles. Since journalists, cannot include every detail of the case, because their text has to be concise perhaps specific sections of the obligations act are not

relevant enough to be included and even if they would be included, the numbers of the different sections of the act would not carry much informational value to a layman reader since additional research would need to be done, not to mention that the act itself was originally written in Estonian so a translation of it would need to be found. And even if a layman reader would find the relevant section of the act, it is not certain that they will be able to properly understand it without having any legal education. Considering all of these obstacles it does seem that it would not be really efficient to include specific sections of a specific act in a news article, without explaining them in the news article itself. It would seem likely that a conscious decision was made to just mention that the obligations act is relevant to the case, but to not expand much on it. A conscientious decision of omitting specific details of the case does simplify the text and makes it more reader-friendly.

Interestingly, in the media corpus, the most frequent 4-word lexical bundle is *court of human rights* which refers to the name of the court, the full name being the European Court of Human Rights. The normalized frequency of this lexical bundle is 1578,9 occurrences. While in the official case report, the full name of the court is only mentioned once at the very beginning of the report:

52) In the case of Delfi AS v. Estonia, The European Court of Human Rights, sitting as a Grand Chamber composed of...

(DvE-LC)

The full name of the court is mentioned along with the case, and what judges the Grand Chamber consisted of. In the media articles, since the focus of the articles is the judgment of the case, the court is often mentioned when the articles describe what sort of judgment was reached:

53) The Grand Chamber of the European Court of Human Rights ruled last week that the Estonian news site, Delfi, can be held liable for defamatory comments made by its users...

(DvE-MC)

54) On 16 June 2015, the European Court of Human Rights delivered its final judgment in Delfi AS v. Estonia.

(DvE-MC)

55) *But the European Court of Human Rights ruled Tuesday that Delfi, an Estonian news website, can be liable for the anonymous defamatory comments readers post on its articles.*

(DvE-MC)

Also in the media corpus, the full name of the court was used two of the names of the different articles relating to the case:

56) *European Court of Human Rights rules that news site Delfi is liable for user's comments.*

(DvE-MC)

57) *The European Court of Human Rights rules again on liability for third party comments.*

(DvE-MC)

In news articles, it is important to state which court ruled what for clarity's sake. Also, it is fair to mention, that since the media corpus consists of multiple news articles each news article should mention the court at least once, that is why the occurrence rate of the lexical bundle is so high, however, it did appear 21 times in the whole corpus and the corpus consists of 8 news articles, so the name did appear multiple times throughout different news articles.

While in the official case report, the court name is stated at the top of it, and the reader already knows where the case report came from. Further on, in the case report, the court refers to itself just using the phrase *the court*, and it occurs quite often as can be seen from the lexical bundles' list of the legal corpus. It should also be noted that the court refers to itself the most in the second half of the case report, where the judgment is described. The court needs to justify their judgment and explain their reasoning, that is why multiple different phrases are used that include the noun phrase *the court*. The lexical bundle *in the court's* has the occurrence rate of 258.9 per 1 million words, in the legal corpus lexical bundle frequency list it took up the 47th position. The lexical bundles *the court notes that* and *the court observes that* also have the same occurrence rate of 258.9. Then there is also the lexical bundle *the court considers that* which has the occurrence rate of 233 times per 1 million words. There are also lexical bundles like *the court has found* and *the court sees no* that have the occurrence rate of 103.5. Finally, some more lexical bundles that included the noun phrase *the court* and were quite frequently occurring, *the court did not*, *the court found that*, *the court has held*, *the court has to*, *the court reiterates that*, *the court should have* and *the court's*

assessment. The occurrence rate of these lexical bundles in the legal corpus is 77.6. There were also more phrases like that, but their occurrence rate was not considered to be very frequent. Here are some of the examples of how these phrases look in the case report itself:

58) *...this system can in the Court's view function in many cases as an appropriate tool for balancing the rights and interests of all those involved.*

(DvE-LC)

59) *The Court notes that as a consequence of this failure of the filtering mechanism these clearly unlawful comments remained online for six weeks (see paragraphs 18-19 above).*

(DvE-LC)

60) *The Court observes that the difference in the parties' opinions as regards the law to be applied stems from their diverging views on the issue...*

(DvE-LC)

61) *Once the Court has found that the reasons given are relevant and sufficient, the proportionality analysis begins rather than ends.*

(DvE-LC)

62) *The Court sees no reason to call into question the above distinction made by the Supreme Court.*

(DvE-LC)

63) *The Court did not include this part of the established case-law in its analysis of journalistic responsibilities in paragraph 132 of the present judgment, where it mentions...*

(DvE-LC)

64) *The Court found that the applicant company itself was to be considered the publisher of the comments, and it could not avoid responsibility by publishing a disclaimer stating that it was not liable for the content of the comments.*

(DvE-LC)

65) *Moreover, the Court has held that speech that is incompatible with the values proclaimed and guaranteed by the Convention is not protected by Article 10...*

(DvE-LC)

66) ...what the Court has to do is to look at the interference complained of in the light of the case as a whole...

(DvE-LC)

67) ...the Court reiterates that it is also mindful of the risk of harm posed by content and communications on the Internet...

(DvE-LC)

68) Lastly, the Court observes that the applicant company has argued (see paragraph 78 above) that the Court should have due regard to the notice-and-take-down system that it had introduced.

(DvE-LC)

69) However, we would like to clarify our position as regards two issues: (a) the Court's reading of the Supreme Court's judgment, and (b) the principles underlying the Court's assessment of the complaint.

(DvE-LC)

It is interesting that the court tends to use the third person when referring to itself, however, it could be considered to be a part of the formal register. In the media corpus, there are also some lexical bundles that include the noun phrase *the court*, however, the occurrence rates and the number of different variations of the phrase are quite a bit lower when compared to the phrases contained in the legal corpus. For example, the lexical bundle *in the court's* appears twice in the whole media corpus, however, in both occurrences, the official case report was quoted and that is why those phrases appeared in the articles. There is also the lexical bundle *in which the court* which appears in the media corpus with the occurrence rate of 150.3 per 1 million words. Here is an example of it:

70) This was the first case in which the Court had been called upon to examine a complaint about liability for user-generated comments on an Internet news portal.

(DvE-MC)

It should be mentioned that there were more lexical bundles in the media corpus that were not discussed, however, they only appeared once in the whole corpus. It was noticed, that the legal corpus contained a higher variety of constantly occurring different lexical bundles that

included the noun phrase *the court* when compared to the media corpus lexical bundles. As mentioned before, the reason for the high occurrence of these bundles in the official case report is that when the court is explaining their judgment it refers to itself in the third person and is providing every single decision and rationale behind their ruling. While in news articles, that is not needed and only the main factors of the decision should be mentioned.

Even though the European Court of Human Rights was only mentioned once in the official case report, other courts appear a lot more often in the 4-word lexical bundle list. In the 20th position of the legal corpus lexical bundle list, the phrase *the supreme court's* appears. The occurrence rate for the lexical bundle is 440.1. In the official case report, the lexical bundle is usually used to refer to the supreme court's judgment or reasoning:

71) *The Court has decided to read the Supreme Court's judgment in the second sense...*

(DvE-LC)

72) *In the light of the Supreme Court's reasoning, the Court must, according to its consistent case-law, examine...*

(DvE-LC)

In the media corpus, the same lexical bundle also appears, with the occurrence rate of 150.3, which is significantly lower than the occurrence rate of the same bundle in the legal discourse. Here are some examples of the lexical bundle's usage in media discourse:

73) *Next, the Grand Chamber noted the Supreme Court's characterisation of the comments posted on Delfi's portal as unlawful.*

(DvE-MC)

74) *...whether the Supreme Court's application of the domestic law to Delfi's situation had been foreseeable.*

(DvE-MC)

Before a case reaches the European Court of Human Rights, usually it goes through other instances first that is why in the official case report, all of the positions, judgments and reasonings of the other courts need to be described and identified. While in the media articles these sort of details often times can be omitted so that the text would be concise. The lexical bundle *the court of appeal* appears in the 26th position of the legal corpus lexical bundle list

and has the occurrence rate of 362.5. Similarly, as with the previously discussed lexical bundle, *the court of appeal* is used in sentences that discuss the court's judgment and actions:

75) *The Court of Appeal considered that the measures taken by the applicant company were insufficient...*

(DvE-LC)

76) *The conclusion reached in the Court of Appeal's judgment is correct, but the legal reasoning of the judgment must be amended and supplemented...*

(DvE-LC)

In the media corpus, *the court of appeal* lexical bundle has the occurrence rate of 150.3. Once again, the lexical bundle occurs more than two times less often in the media corpus when compared to the legal corpus. In the media corpus, the lexical bundle occurs only in one article, that discussed the background of the case:

77) *The Court of Appeal found that hosting protection did not apply, as Delfi had "integrated the comment environment into its news portal and invited users to post comments". The Court of Appeal held that Delfi was a provider of content services, rather than an internet intermediary.*

(DvE-MC)

The final court that was constantly mentioned in both corpora is the Court of Justice of the European Union. The lexical bundle that appeared on the lists was *the Court of Justice* which referred to the aforementioned court. Interestingly, the occurrence rate in the legal corpus is 77.7 while the occurrence rate in the media corpus is 375.9 which is significantly more than the occurrence rate in the official case report. The lexical bundle was used in several news articles, but not all of them, and it was mostly used in the context of what sort of aftereffects could occur with the European Court's of Human Rights ruling of the present case. Multiple articles mentioned, that if the case would have gone to the Court of Justice of the European Union, the ruling would have probably been completely opposite because a precedent was set by the Court of Justice of the European Union that protects websites from the type of liability that Delfi faced. The main objective of any news article should be to inform the public and generally that is what the news articles try to do, in this case, some news articles try to interpret the judgment, and inform the public on the impact that the judgment may carry in the future. The journalists used the Court of Justice of the European Union as an example of how the case could have gone a different way. It should still be emphasized that the other courts

that were mentioned occurred in the legal corpus significantly more often when compared to the occurrence rate of the media corpus, the Court of Justice of the European Union being the exception. As mentioned before, the European Court of Human Rights has to investigate describe and understand fully the judgment and actions that were made by other instances that were taking care of the case. Since the Court of Justice of the European Union was not taking care of this case, it was not mentioned too often in the case report. News articles cannot include all of the background information of the case since news articles do have their scope limits so that is why it would seem that the other courts were mentioned significantly less often in the media corpus when compared to the legal corpus.

Another prominent lexical bundle in the media corpus is *the e-commerce directive* it is present in the 5th position and has the occurrence rate of 977.4. Interestingly, in the legal corpus, there is no exact equivalent for this lexical bundle however, the closest to it would be *directive on electronic commerce*. This lexical bundle is in the 15th position of the legal corpus lexical bundle list and has the occurrence rate of 440.1. The media corpus also contains the *directive on electronic commerce* lexical bundle, but the occurrence rate for it is only 150.4. As can be seen from the occurrence rate, journalists prefer using the shortened *the e-commerce directive* over the *directive on electronic commerce* lexical bundle. While in legal discourse the term *e-commerce* was not used at all and the directive was always referred to as the *directive on electronic commerce*. Because of the formal nature of legal discourse, it is understandable why the full name of the directive is always used, also since the official case report has to be as accurate as possible, introducing the shortened name of the directive could possibly cause confusion as one might think that perhaps it is a different directive altogether. Media discourse is quite a bit less formal than legal discourse so shortening the name of a directive is quite appropriate.

3.2.4 Georgia v Russia Case Qualitative Analysis

The lexical bundle analysis was also done on the media and legal corpora of the Georgia v Russia case. Once again, the most often occurring 200 4-word lexical bundles were looked at from each corpus. In the legal corpus, a total of 116,892 4-word lexical bundles were found while in the media corpus the total number of 4-word lexical bundles was 15,546. The total word count for the legal corpus was 116,895, while the word count for the media corpus was 15,549. Because the corpora differ in size, the occurrence rate of the lexical bundles will be normalized to how often they occur per 1 million words.

To start with the most often occurring 4-word lexical bundle in the legal corpus for this case was *Article # of the convention*. Since Antconc skips numbers when analyzing lexical bundles, the hashtag in the middle of the lexical bundle represents a number, since this lexical bundle in the official case report was used to refer to specific articles of the European Convention on Human Rights. For example:

78) There has therefore been a violation of Article 3 of the Convention, and the Russian Federation is responsible for that violation.

(GvR-LC)

79) In general, it may be observed that the obligation to carry out an effective investigation under Article 2 of the Convention is broader than the corresponding obligation in international humanitarian law (see paragraph 318 above).

(GvR-LC)

80) It thus concludes that there was an administrative practice contrary to Article 5 of the Convention as regards the arbitrary detention of Georgian civilians in August 2008.

(GvR-LC)

The occurrence rate in the legal corpus for this lexical bundle is 1,659.6 per 1 million words. In the media corpus, this lexical bundle is also quite prominent as it takes up the 3rd position in the lexical bundle list and has the occurrence rate of 1,350.6. In both, media discourse and legal discourse different articles of the convention were referred to, since they are necessary for the context of the case. The articles let the reader know what sort of human rights violations is Russia being accused of in this case. Interestingly, in the media corpus, only 6 different articles of the convention are mentioned, while in the legal corpus 21 different articles were mentioned. In the official case report, every single detail needs to be included, and every violation of the convention needs to be accounted for. The journalists, however, cannot include every relevant article of the convention into their news story since they are trying to summarize the case and only provide you with the details that are necessary in understanding just the essentials of the case. Some of the news stories that were included in the media corpus did not even mention any articles of the convention at all.

Another 4-word lexical bundle that was prominent in the legal corpus lexical bundle list was *in the present case*. This lexical bundle took up the 4th position and has the occurrence rate of 701.5. This phrase is a recurrent one, as it was also prominent in the *Delfi AS v Estonia* official case report. However, similarly, as with the *Delfi AS v Estonia* case, this lexical

bundle is not as prominent in the media discourse and the occurrence rate for it in the media corpus is 321.6 per 1 million words. In the legal corpus the lexical bundle is used to refer to the case itself:

81) *In the present case, the Court has relied on, inter alia, the observations of the parties and the many documents submitted by them.*

(GvR-LC)

82) *In the present case the Court considers that a distinction needs to be made between the military operations carried out during the active phase of hostilities and the other events...*

(GvR-LC)

83) *This is also true in the present case, given that the majority of the fighting took place in areas which were previously under Georgian control (see paragraph 111 above).*

(GvR-LC)

Since the case itself is on the larger side it is not surprising that in the official case report the case is being constantly referred to. While in the media corpus, the name of the case seems to be used more often than the phrase *in the present case*. In the media corpus, the lexical bundle *Georgia v Russia II* is in the 28th position and has the occurrence rate of 707.4. This lexical bundle occurred more than twice as often than *in the present case* in the media corpus. The roman numeral two is present with the lexical bundle because that is the exact name of the case. The first case occurred several years ago. Here are some of the examples of how this lexical bundle appears in the media corpus:

84) *The European Court of Human Rights' recent Grand Chamber judgment in the case of Georgia v Russia (II) has already been the subject of strong criticism, both from within the Human Rights Building and outside.*

(GvR-MC)

85) *A 'direct and immediate cause' test reflects the notion of proximity of effects, which was referred to in Georgia v Russia (II)...*

(GvR-MC)

86) *That the majority judges in Georgia v Russia (II) fail to refer to the recent seminal General Comment on the right to life published by the UN Human Rights Committee is a surprising and significant omission.*

(GvR-MC)

In the legal corpus, this exact lexical bundle can be found as well, but it is not very prominent and has the occurrence rate of 42.8 per 1 million words. It does seem that the preferred way to refer to the case itself in the official case report is using the phrase *in the present case* while in the media corpus the preferred way would be to just refer to the case by name *Georgia v Russia II*. It should be fair to also mention that in both of the corpora, the lexical bundle *Georgia v Russia I* is also present. It refers to the first case, the occurrence rate for it is 128.6 in the media corpus and 111.2 in the legal corpus. The occurrence rate is quite similar in both corpora however, the lexical bundle appeared only in one media article. The article was the official press release of the European Court of Human Rights that means that no journalist unrelated to the court has mentioned the previous case by its name. Once again, it could have been a detail that was deemed to be too specific perhaps to mention in a news article. The previous case does provide some background information, but for a layman reader the amount of information could be considered to be overwhelming and focusing on the present case in a news article is most likely for the better in terms of readability and reader-friendliness.

In the media corpus, the most often appearing lexical bundle was *the active phase of* with the occurrence rate of 1,672.1. The same 4-word bundle can be found in the legal corpus lexical bundle list, it occupies the 10th position and has the occurrence rate of 470.5. The full phrase that this lexical bundle refers to is *the active phase of hostilities*. The phrase was so prominent in the media articles is because the court had decided that the activities of the Russian Federation should be split into phases, one of the phases being the active phase of hostilities while the other phase was the occupation. It seems that the journalists found this distinction important as it was an important detail in the case and in most of the news articles *the active phase of hostilities* was mentioned:

87) *However, the Court drew a distinction between this occupation phase and the prior period of military operations carried out during the active phase of hostilities.*

(GvR-MC)

88) *The cause of concern has been the Court's majority finding that 'military operations' during 'the active phase of hostilities' in an international armed conflict (IAC) are beyond the 'jurisdiction' of the state...*

(GvR-MC)

89) *The Court found it appropriate to examine the military operations carried out during the active phase of hostilities separately from the other events occurring after the cessation of the active phase of hostilities.*

(GvR-MC)

In the legal corpus, the lexical bundle was used to refer to the phase during which the violations were carried out.

Another interesting case is with the lexical bundle *of international humanitarian law* which takes up the 5th position in the legal corpus lexical bundle list and has the occurrence rate of 701.5. However, in the media corpus, this lexical bundle appears only once in the whole corpus and there is also the lexical bundle *under international humanitarian law* which also occurs only once. In both of those cases, the international humanitarian law is mentioned in the official press release of the European Court of Human Rights, which means that this law is not mentioned at all by journalists not associated with the court. It could also be assumed that the official press release of the case includes more details than news reports in news portals, which would make sense. The lexical bundle in the official case report is used when specific rules and provisions are mentioned of the international humanitarian law:

90) *The Court reiterates in this connection that this does not mean that States can act outside any legal framework; as indicated above, they are obliged to comply with the very detailed rules of international humanitarian law in such a context.*

(GvR-LC)

91) *This is a serious violation of international humanitarian law.*

(GvR-LC)

92) *Having regard to the complaints raised in the present case, there is no conflict between Article 2 of Protocol No. 1 and the relevant provisions of international humanitarian law concerning a situation of occupation.*

(GvR-LC)

It can be seen in the case report that the Russian Federation has violated certain provisions of the international humanitarian law, but in the media corpus, the international humanitarian law is barely mentioned. However, the acronym IHL that stands for international humanitarian law can be found in multiple articles and in some of them, it is used multiple times. However in most of the articles contained in the media corpus this piece of law is not mentioned. With that in mind, it would seem that in media discourse this specific law was mostly excluded and if included it was preferably used as an acronym instead of using the full name of it. So it could be stated that the text was being simplified and shortened by using an acronym or omitting the law entirely.

In the media corpus lexical bundles list, in the 7th position *court of human rights* is present with the occurrence rate of 1,157.6. The lexical bundle of course refers to the full name of the court, the European Court of Human Rights. Similarly, as with the *Delfi AS v Estonia* case, the full name of the court used in the text is more prominent in the media corpus than in the legal corpus. It is understandable because for each news article the name of the court should be mentioned at least once so that the reader would know which court was taking care of the case. In the official case report, the same lexical bundle appears as well, but the occurrence rate is only 59.9 per 1 million words. However, in the official case report, the court usually refers to itself simply as *the court*, the same example was present in the *Delfi AS v Estonia* case as well. There is a variety of lexical bundles in the legal corpus that includes the noun phrase *the court* and have a higher occurrence rate than *court of human rights*. For example, *the court considers that* has the occurrence rate of 179.6, almost three times larger than the occurrence rate of *court of human rights* in the legal corpus. Then there is *the court held that* and *the court observes that* with occurrence rates of 68.4. And there are also other numerous phrases that have the same occurrence rate as the *court of human rights* lexical bundle or lower. In the media corpus, there are also lexical bundles like, *court concluded that there*, *court concluded that the*, *court considered that it*, however, these phrases had a lower occurrence rate when compared to *court of human rights*.

4. Conclusions

After performing two case studies, it can be stated that certain similarities and differences between legal and media discourse and features of simplification were identified. Word frequency analysis revealed that certain legal terms like *applicant*, *judgment*, *respondent* are either avoided, used less often or are replaced with less formal alternatives in media discourse. It was also found that verbs like *may*, *shall* and *must* occurred quite frequently in legal discourse, but in media discourse, those verbs did not occur as often. In contrast, more frequent modals were included, such verbs as *can* and *will*. Media discourse also included fewer references to specific laws, acts and law articles than legal discourse even though the same specific case was being discussed. Also, in the media discourse the laws, acts and articles were sometimes generalized, while in legal discourse all of the relevant laws, acts and articles were explicitly referred to in detail.

The analysis of lexical bundles in both case studies further supported the findings of the word frequency analysis. It was found that in media discourse the cases were summarized and certain details of the cases were not as prominent as in legal discourse or omitted entirely. There were also cases where certain terms were shortened in media discourse like, *the e-commerce directive* or *IHL* and the shortened version of the term was more frequently used than the full terms *the electronic commerce directive* and *International Humanitarian Law*.

The structural analysis of the lexical bundles showed that the noun phrase-based lexical bundles were the most prominent across all four corpora. The distribution of different types of phrases was quite similar across legal and media discourse with some minor differences. The functional analysis revealed that the most frequent lexical bundles were mostly referential in all four of the corpora and most of the referential expressions were topic-specific. The abundance of noun phrase-based lexical bundles and referential expressions could be explained through the complexity of the cases. NP-based lexical bundles have been shown by numerous studies to be typical of written rather than spoken discourse, so this finding is not surprising. These cases usually go through other instances first and if the resolution of the other courts is not satisfactory then they reach the European Court of Human Rights. The other instance courts are perfectly capable of resolving simple cases, but the more a case is complex, the higher is the chance of it reaching the European Court of Human Rights. Because of the complexity of the cases, they usually contain much background information and an abundance of important details that need to be mentioned, not to mention the different laws, articles and acts that need to be adhered to. As a result of all of that, a substantial number of details need to be mentioned in the case reports and in the news articles and there

are plenty of concepts and different items to refer to, that is why noun phrase-based lexical bundles and referential expressions were so prominent in the lexical bundles' lists of all four corpora.

The features of simplification that were identified in this study include omission of certain legal terminology and, replacement of legal terminology with a less complex alternative. Because the cases in the news articles had to be summarized, certain details of the cases like, specific articles, laws and acts were either omitted or generalized. As mentioned before, certain terms were shortened in the media discourse as well. It can be seen that in media discourse certain aspects of the case are simplified so that a person that does not have a legal education may have a better chance at understanding the case and certain implications of specific judgments.

It was expected, however, that the structural and functional analyses would have shown a larger number of differences rather than similarities between the discourses. For a future study perhaps a larger sample of lexical bundles should be analyzed, or completely different data sets should be chosen to represent legal discourse and media discourse.

Summary in Lithuanian

Atviros visuomenės principas per pastaruosius dešimtmečius gerokai priartino teisinės kalbą ir terminologiją prie kasdienės kalbos vartosenos. Tačiau suprasti teisinę kalbą neturint teisinio išsilavinimo daugeliui yra tikras iššūkis. Todėl internetiniai naujienų portalai, aprašantys įvairias teises bylas, neišvengiamai turi supaprastinti savo straipsnių kalbą aprašinėdami teisinius procesus, kad jų straipsniai būtų suprantami kuo didesnei skaitytojų auditorijai.

Šiame tyrime analizuojant teisinį ir žiniasklaidos diskursus ir yra bandoma identifikuoti, kokiais būdais žiniasklaidoje įvyksta teisinio diskurso supaprastinimas, priartinant jį prie internetinių portalų skaitytojų poreikių. Tyrimui pasirinktos dvi atskiros aprašytos bylos. Iš jų medžiagos sukaupti keturi tekstynai, po du tekstynus kiekvienai bylai. Vienas tekstynas reprezentavo teisinį diskursą, kitas – žiniasklaidos kalbą. Teisinis tekstynas sudarytas iš oficialaus bylos aprašo, kuris buvo parengtas Europos žmogaus teisių teismo. Žiniasklaidos tekstynai buvo sudaryti iš straipsnių, kurie buvo publikuoti internetiniuose portaluose, aprašiusiuose šias bylas. Naudojantis tekstynų analizės programa *AntConc* (Anthony 2020) buvo sudaryti žodžių dažnumo sąrašai ir keturžodžių leksinių samplaikų sąrašai kiekvienam tekstynui. Žodžių dažnumo sąrašuose buvo analizuojami 100 dažniausių leksinių vienetų. Iš keturžodžių leksinių samplaikų sąrašų kokybinei analizei paimta 200 dažniausių leksinių samplaikų iš kiekvieno tekstyno.

Struktūrinėje analizėje leksinės samplaikos buvo kategorizuojamos pagal tai, kokia kalbos dalimi prasideda samplaika. Buvo naudojamos tokios kategorijos: samplaikos, turinčios daiktavardinių junginių fragmentų, samplaikos, turinčios veiksmažodinių junginių fragmentų, samplaikos, turinčios šalutinių sakinių fragmentų, samplaikos, turinčios prielinksnių junginių fragmentų, samplaikos, turinčios būdvardinių junginių fragmentų ir samplaikos, turinčiosrieveksminių junginių fragmentų. Funkcinėje analizėje samplaikos buvo klasifikuojamos į šias kategorijas: referentinės samplaikos, požiūrio raiškos samplaikos ir diskurso organizavimo samplaikos. Šios klasifikacijos buvo naudojamos remiantis Biber et al. (1999), Biber et al. (2004) autorių atliktais tyrimais ir juose nustatytais kategorijomis.

Tyrimo rezultatai parodė, kad žiniasklaidos diskurse teisinis tekstas yra išties supaprastintas. Kai kurių teisinių terminų yra iš viso vengiama, jie nevartojami arba pakeičiami koku nors paprastesniu, lengviau suprantamu terminu. Taip pat žiniasklaidos diskurse praleidžiamos ir neįtraukiamos kai kurios bylų detalės, specifiniai teisiniai aktai ir direktyvų straipsniai. Struktūrinė analizė taip pat parodė, kad ir žiniasklaidos, ir teisiniame diskurse dažniausios yra

samplaikos, turinčios daiktavardinių junginių fragmentų. Atliekant funkcinę analizę nustatyta, kad dažniausiai randamos buvo referentinės samplaikos abiejuose diskursuose.

Tyrimas turi tam tikrų ribotumą. Visų pirma, čia analizuoti tik dviejų bylų aprašymo atvejai. Norint surinkti patikimesnių duomenų apie teisinio diskurso supaprastinimo strategijas, ateityje būtų prasminga išplėsti tyrimo medžiagą įtraukiant daugiau atvejų. Taip pat šiame tyrime nebuvo bandoma analizuoti sintaksinių teisės ir žiniasklaidos diskurso ypatumų, pavyzdžiui, sakinio dėmenų ar sakinių ilgio, nors šie parametrai taip pat prisideda prie raiškos sudėtingumo. Tačiau norėtusi tikėtis, kad preliminarios šio darbo išvalgos paskatins kitus tyrėjus toliau gilintis į teisės diskurso raiškos subtilybes.

References

Primary Sources

Access Now Team. 2015. *Delfi AS v. Estonia: a blow to free expression online*. Available at: <https://www.accessnow.org/delfi-as-v-estonia-a-blow-to-free-expression-online/>, accessed on the 25th of May 2021.

Agenda.ge. 2021. *Georgia wins August war case against Russia in ECHR*. Available at: <https://agenda.ge/en/news/2021/167>, accessed on the 25th of May 2021.

Bodrogi, B. 2016. *The European Court of Human Rights rules again on liability for third party comments*. Available at: <https://blogs.lse.ac.uk/mediase/2016/02/19/the-european-court-of-human-rights-rules-again-on-liability-for-third-party-comments/#:~:text=February%2019th%2C%202016-.The%20European%20Court%20of%20Human%20Rights%20rules,liability%20for%20third%20party%20comments&text=In%20a%20recent%20ruling%2C%20the,comments%20posted%20by%20its%20readers.>, accessed on the 25th of May 2021.

Civil.ge. 2021. *European Court Verdict into Georgia vs. Russia Case over 2008 War*. Available at: <https://civil.ge/archives/392226>, accessed on the 25th of May 2021.

Delfi AS v Estonia (Application no. 64569/09). 2015. European Court of Human Rights. Strasbourg. Available at: <http://hudoc.echr.coe.int/eng/?i=001-155105>, accessed on the 25th of May 2021.

Duffy, H. 2021. *Georgia v. Russia: Jurisdiction, Chaos and Conflict at the European Court of Human Rights*. Available at: <https://www.justsecurity.org/74465/georgia-v-russia-jurisdiction-chaos-and-conflict-at-the-european-court-of-human-rights/>, accessed on the 25th of May 2021.

European Court of Human Rights Press Release. 2015. *Commercially-run Internet news portal was liable for the offensive online comments of its readers*. Available at: [https://hudoc.echr.coe.int/eng-press#{%22fulltext%22:\[%2264569/09%22\]}](https://hudoc.echr.coe.int/eng-press#{%22fulltext%22:[%2264569/09%22]}), accessed on the 25th of May 2021.

European Court of Human Rights Press Release. 2021. *Judgment in the case concerning the armed conflict between Georgia and the Russian Federation in August 2008 and its consequences*. Available at: <https://hudoc.echr.coe.int/eng-press#%7B%22fulltext%22:%5B%2238263%2F08%22%5D%7D> , accessed on the 25th of May 2021.

Gavron, J. 2021. *Damage control after Georgia v Russia (II) – holding states responsible for human rights violations during armed conflict*. Available at: <https://strasbourgobservers.com/2021/02/08/damage-control-after-georgia-v-russia-ii-holding-states-responsible-for-human-rights-violations-during-armed-conflict/>, accessed on the 25th of May 2021.

Georgia v Russia (II) (Application no. 38263/08). 2021. European Court of Human Rights. Strasbourg. Available at: <http://hudoc.echr.coe.int/eng?i=001-207757>, accessed on the 25th of May 2021.

Harding, L. 2021. *Russia committed human rights violation in Georgia war, ECHR rules*. Available at: <https://www.theguardian.com/world/2021/jan/21/russia-human-rights-violation-georgia-war-echr-ruling>, accessed on the 25th of May 2021.

Kuczerawy, A & P. J. Ombelet. 2015. *Not so different after all? Reconciling Delfi vs. Estonia with EU rules on intermediary liability*. Available at: <https://www.law.kuleuven.be/citip/blog/not-so-different-after-all-reconciling-delfi-vs-estonia-with-eu-rules-on-intermediary-liability/>, accessed on the 25th of May 2021.

Lambert, A & E. Parris. 2015. *European Court of Human Rights rules that news site Delfi is liable for user's comments*. Available at: <https://www.fieldfisher.com/en/services/technology-outsourcing-and-privacy/technology-and-outsourcing-blog/european-court-of-human-rights-rules-that-news-site-delfi-is-liable-for-users-comments>, accessed on the 25th of May 2021.

Megrelidze, S. 2021. *Europe's court condemns Russia over 2008 war with Georgia*. Available at: <https://apnews.com/article/europe-georgia-salome-zurabishvili-moscow-russia-854c37a67adc9cda0920a00e931efc09>, accessed on the 25th of May 2021.

Moody, G. 2015. *Shock European court decision: Websites are liable for users' comments*. Available at: <https://arstechnica.com/tech-policy/2015/06/shock-european-court-decision-websites-are-liable-for-users-comments/>, accessed on the 25th of May 2021.

Newman, L. H. 2015. *EU Court Unexpectedly Rules Estonian Website Is Responsible for User Comments*. Available at: <https://slate.com/technology/2015/06/european-court-of->

[human-rights-rules-estonian-site-delfi-is-liable-for-comments.html](https://www.eurojust.europa.eu/en/press-releases/2021/05/human-rights-rules-estonian-site-delfi-is-liable-for-comments.html), accessed on the 25th of May 2021.

Pinto, T. 2015. *Delfi v Estonia - what is the real impact?*. Available at: <https://sites-taylor-wessing.vuturvevx.com/12/778/landing-pages/radar-2015-07-delfi.asp>, accessed on the 25th of May 2021.

Radio Free Europe / Radio Liberty journalists. 2021. *European Court Says Russia Committed Rights Violations After War With Georgia*. Available at: <https://www.rferl.org/a/european-court--russia-rights-violations-war-georgia/31061859.html>, accessed on the 25th of May 2021.

Seskuria, N. 2021. *Georgia Wins a Legal War Against Russia*. Available at: <https://rusi.org/commentary/georgia-wins-legal-war-against-russia>, accessed on the 25th of May 2021.

Voorhoof, D. 2015. *Delfi AS v. Estonia: Grand Chamber confirms liability of online news portal for offensive comments posted by its readers*. Available at: <https://strasbourgobservers.com/2015/06/18/delfi-as-v-estonia-grand-chamber-confirms-liability-of-online-news-portal-for-offensive-comments-posted-by-its-readers/>, accessed on the 25th of May 2021.

Secondary Sources

Ädel, A., & B. Erman. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.

Allen, D. 2009. Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 1, 105-127.

Anthony, L. 2020. *AntConc* (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Available at <http://www.laurenceanthony.net/software>, accessed on 25th of May 2021

Atkinson, M., Smith, K., & S. Kirby. 2018. Adult Learning and Language Simplification. *Cognitive Science*, 42, 2818 - 2854.

Baker, M. 1996. Corpus-based translation studies: the challenges that lie ahead. In H. L. Somers (ed.), *Terminology, LSP and Translation. Studies in Language Engineering in Honour of Juan C. Sager*. Philadelphia: John Benjamins, 175-186.

Bhatia, V. 1983. Simplification v. Easification-The Case of Legal Texts. *Applied Linguistics – APPL Linguist* 4: 42-54.

- Biber, D., & Barbieri, F. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., S. Conrad & V. Cortes. 2004. If you look at...: Lexical bundles in university teaching and textbooks, *Applied Linguistics* 25(3): 371-405.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Breeze, R. 2013. Lexical bundles across four legal genres. *International Journal of Corpus Linguistics* 18(2): 229-253.
- Byrd, P. & A. Coxhead. 2010. On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL* 5(5): 31-64.
- Charrow, R. & V. Charrow. 1979. Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions. *Columbia Law Review* 79(7):1306-1374.
- Chen, Y. H. & P. Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14(2): 30-49.
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Cortes, V. 2006. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education* 17(4): 391-406.
- Goodman, J.D. & E. Greene. 1989. The use of paraphrase analysis in the simplification of jury instructions. *Journal of Social Behavior and Personality* 4:237-251.
- Google Dictionary by Oxford Languages. Available on www.google.com
- Hartley, J. 2000. Legal Ease and “Legalese”, *Psychology Crime and Law* 6(1): 1-25.
- Hyatt, S. 2018. *Legal Glossary*. Available at <https://www.law.du.edu/index.php/law-school-learning-aids/legal-language/>?, accessed on 25th of May 2021.
- Hyland, K. 2008. As can be seen: Lexical bundles and disciplinary variation, *English for Specific Purposes* 27(1): 4-21.
- Institute for the Languages of Finland*. Available at: https://www.kotus.fi/en/on_language/plain_language, accessed on 25th of May 2021.
- Jablonkai, R. 2009. “In the light of”: A corpus-based analysis of lexical bundles in two EU-related registers. *WoPaLP* 3: 1-27.
- Jablonkai, R. 2010. English in the context of European integration: A corpus-driven analysis of lexical bundles in English EU documents. *English for Specific Purposes* 29(4): 253-267.

- Juknevičienė, R. 2009. Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra*, 61(61), 61-72.
- Leow, R. 2003. Simplification and Second Language Acquisition. *World Englishes*, 16, 291-296.
- Masson, M. E. J. & M. A. Waldron. 1994. Comprehension of legal contracts by non-experts: Effectiveness of plain language redrafting. *Applied Cognitive Psychology* 8(1):67–85.
- Pan, F., Reppen, R., & D. Biber. 2016. Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in Telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Petersen, S.E. & M. Ostendorf. 2007. Text simplification for language learners: a corpus analysis. *SLaTE*.
- Randall, J. 2014. Tackling “Legalese”: How Linguistics Can Simplify Legal Language and Increase Access to Justice, *Language Use and Linguistic Structure*. Olomouc: Olomouc Modern Language Series. 3: 231-246.
- Redish, J. 1979. *How to draft more understandable legal documents*. Washington D.C.: Document Design Center, American Institutes for Research.
- Ritter, P.A. 1981. Simplification of Legal Language and The Bank of Nova Scotia Plain Language Mortgage Form. *University of Toronto Faculty of Law Review* 39(2):170-179.
- Tiersma, P.M. 1999. *Legal Language*. Chicago: The University of Chicago Press.
- Tománková, V. 2016. Lexical bundles in legal texts corpora—Selection, classification and pedagogical implications. *Discourse and Interaction*, 9(2), 75-94.
- Trosborg, A. 1997. *Rhetorical Strategies in Legal Language*. *Discourse Analysis of Statutes and Contracts*. Tübingen: Narr.
- Zarcadoolas, C. 2011. The simplicity complex: exploring simplified health messages in a complex world, *Health Promotion International*, 26/3: 338-350.

Appendix

1 Legal corpus containing the official case report of Delfi AS v Estonia Case.

1 Legal corpus containing the official case report of Georgia v Russia (II) Case.

1 Media corpus containing 8 news articles relating to the case Delfi AS v Estonia and containing the official press release for the case from the European Court of Human Rights.

1 Media corpus containing 8 news articles relating to the case Georgia v Russia (II) and containing the official press release for the case from the European Court of Human Rights.

1 Word frequency list containing top 100 most frequent words in Legal and Media corpora of Delfi AS v Estonia case.

1 Word frequency list containing top 100 most frequent words in Legal and Media corpora of Georgia v Russia (II) case.

1 Lexical Bundle list containing top 200 most frequent lexical bundles in Legal and Media corpora of Delfi AS v Estonia case.

1 Lexical Bundle list containing top 200 most frequent lexical bundles in Legal and Media corpora of Georgia v Russia (II) case.