LATVIJAS UNIVERSITĀTE

MAĢISTRA DARBS

RĪGA 2024

UNIVERSITY OF LATVIA

FACULTY OF HUMANITIES

DEPARTMENT OF ENGLISH STUDIES

LATVIAN COUNTRY IMAGE IN THE CORPUS OF GOVERNMENTAL WEBSITES

LATVIJAS TĒLS VALSTS IESTĀŽU TĪMEKĻA VIETŅU TEKSTU KORPUSĀ

MASTER THESIS

Kārlis Zujs

Matriculation card No. Kz18036

Advisor: assoc. prof. Zigrīda Vinčela

RĪGA 2024

Acknowledgements

I would like to express my gratitude to Assoc. Prof. Zigrīda Vinčela for her assistance, insightful feedback, and the continuous support that she provided throughout the development of my master thesis. Her expertise and comments have been instrumental in shaping this work, for which I am deeply appreciative.

Anotācija

Iepriekšējie valsts tēla pētījumi pievērsa uzmanību ideoloģijai, un tajos tika izmantota neliela atlase. Tomēr lingvistisko pētījumu, kuros aplūkots lielāks tekstu apjoms par Latviju, ir maz. Šī korpusā balstītā pētījuma mērķis bija noskaidrot Latvijas tēla atspoguļojumu valsts iestāžu tīmekļa vietnēs. Pētījuma gaitā, kurā tika izmantota korpusā balstīta diskursa analīze, tika izveidots Latvijas valdības tīmekļa vietņu tekstu korpuss. Korpusa izpētē tika izmantota atslēgvārdu un konkordances analīze. Atslēgvārdu analīze atklāja korpusa galvenās semantiskās kategorijas, savukārt konkordanču analīze atklāja atslēgvārdu lietojumu kontekstā. Iegūtie rezultāti liecina, ka Latvijas tēls ir daudzšķautnains un pozitīvi raksturots – tēmas ietver diplomātisko un militāro sadarbību, investīciju piesaisti, atbalstu Ukrainai un opozīciju Krievijai.

Atslēgas vārdi: Latvija, valsts tēls, korpuss, korpusā balstīts, diskursa analīze

Abstract

Previous studies on country images have focused on ideology and used a small sample size. However, linguistic research that examines a larger volume of texts on Latvia is scarce. The aim of this corpus-based study was to identify the country image of Latvia on Latvian governmental websites. The study, which used Corpus-Based Discourse Analysis, built a specialised corpus of texts from Latvian governmental websites. The corpus was examined using keyword and concordance analyses. Keyword analysis revealed the main semantic categories of the corpus, whilst concordance analysis revealed the use of keywords in context. The findings suggest that the image of Latvia is multifaceted and positively described – the themes include diplomatic and military cooperation, attracting investment, support to Ukraine, and opposition to Russia.

Key words: Latvia, country image, corpus, corpus-based, discourse analysis

Contents

Introduction

The popularity of the internet has led to the emergence of contemporary digital media, where an abundance of information is available in virtually every language. A large portion of media available on the internet is run by different media companies or different independent journalists. However, a portion of the websites are managed and influenced by governments, where they leverage the power of the internet to communicate the information that they find necessary to their citizens, visitors, and outsiders. The materials produced by the government have the potential to portray a country's image, where perceptions of the country and their policies can be shaped to local and international audiences. In a general sense, online dictionaries define image as 'the way that something or someone is thought of by other people' (Cambridge Advanced Learner's Dictionary & Thesaurus, Online 1) and more specifically, as 'a popular conception (as of a person, institution, or nation) projected especially through the mass media' (Merriam-Webster.com Dictionary, Online 2). Country image can therefore refer to the overall perception that people have about a specific nation, varying from the perception, reputation, and impression aspects.

This phenomenon of country images has been analysed by scholars in different fields. For example, scholars have approached analysing country images from an economical perspective, such as analysing country images and country brands (Cotirlea, 2015), country image in the eyes of international students (Zhao, You and Lin, 2022), or from the perspective of marketing (Iversen, Kleppe and Stensaker, 1998), or international public relations (Buhmann and Ingenhoff, 2014). In linguistics, studies of representations or images has been of interest to scholars as well, such as the self-representation of African countries on the internet (Fürsich and Robins, 2002), self-representation of entrepreneurs in Belarus (Miazhevich, 2007) and the theoretical perspectives to studying the image of a country (Tarasheva, 2014).

The prior studies indicate that whilst there is an interest in analysing country image, analyses of governmental websites are quite scarce. The same can be said about analyses of the country image of Latvia, with little to no research available that would study the image of Latvia. Thus, this study aims to investigate the image of Latvia in Latvian governmental websites, namely, how the country is described and positioned, what information prevails, and the similarities and differences in content and its meaning across different websites.

The **goal** of the thesis is to identify the country image of Latvia established on Latvian governmental websites by applying Corpus Assisted Discourse Studies (CADS) to identify

the quantitative statistical data and the qualitative findings that contribute to the portrayal of Latvia.

To achieve the goal, **research questions** are proposed:

1. What are the notions and applications of country images in academic research?
2. How are corpus and linguistic theories applicable to study country images?
3. What insights do the analysis of keywords and concordance lines reveal about the image of Latvia?
4. What do the semantic categories identified reveal about the country image of Latvia?

To answer the research questions, the following **research objectives** have been set:

1. To collect and study theoretical material on country images, discourse analysis, and corpus assisted discourse studies;
2. To collect texts from different Latvian governmental websites, to compile them into a corpus;
3. To perform quantitative analysis of the corpus by frequency and keywords;
4. To perform qualitative analysis of the corpus by semantic annotation and concordances;
5. To analyse and discuss the most common thematic threads regarding Latvia;
6. To compare the quantitative results with the qualitative results;
7. To draw conclusions based on the findings.

The study consists of a theoretical and empirical part. The theoretical research comprises **comparative analysis** of different theoretical sources to establish the theoretical background of the study; previous research on country images (Fürsich and Robins, 2002; Wilk, 2014; Zhang and Shu, 2023), discourse analysis (Schiffrin et al, 2001; van Dijk, 1997; Wodak, 2001), Corpus Assisted Discourse Studies (Baker, 2023), corpus design (Weisser, 2022; Koester, 2022; Gatto, 2014), frequency analysis and keyword analysis (Brezina, 2018; Pojanapunya and Todd, 2016; Kilgarriff, 2009; Mautner, 2022), and concordance analysis (Anthony, 2022; Bianchi, 2012; Gillings and Mautner, 2024). The empirical research utilises **Corpus-Based Discourse Analysis**, combining quantitative and qualitative methods to analyse the image of Latvia through keywords, semantic annotations, and concordances. Keyword analysis is used to obtain statistical data, whereas semantic annotations and concordance analysis provide qualitative results.

Outline of the structure of the paper

The study consists of an introduction, four chapters, 41 subchapters, as well as the conclusions, theses, and 2 appendices.

Chapter 1 examines the notions and applications of country image analysis in different academic areas, paying particular attention to the field of linguistics.

Chapter 2 establishes the notions of discourse and discourse analysis, focusing on the theoretical background that enables Corpus Assisted Discourse Analysis.

Chapter 3 establishes Corpus Assisted Discourse Analysis as the method of the study, examining the quantitative and qualitative approaches of frequency lists, keyword analysis, and concordance analysis to describe country images.

Chapter 4 presents the methodology and procedure of the research, as well as the discussion of results. The results discussion involves a frequency analysis of the whole corpus, followed by keyword analysis of the whole corpus and its corresponding subcorpora, and concordance analysis of each subcorpus separately. Lastly, the main statements or theses are derived from the results to form an overarching country image of Latvia.

The conclusions chapter summarises the findings of the theoretical and empirical analyses. Theses, references, and appendices provide the main statements of the study, sources, and additional information.

1. Notions and Applications of Country Images in Academic Research

The chapter focuses on establishing the notion of a country image, the different fields in which this phenomenon is examined, as well as different studies and their points of interest. The writings of researchers from different fields are explored, and their findings and approaches are explored as motivation for conducting the analysis.

1.1. Notions of country images and their points of analysis in different research fields

The topic of country images, alongside with linguistics, has been a topic of interest in various research fields, being topical due to the insights that the topic provides on different aspects of society, such as on language, politics, marketing, and economy. The examination of country images across different research fields may present a more comprehensive understanding of the term, as well as the interplay that the components have towards building a country image. Thus, this chapter examines the way that the notion of country image is viewed in linguistics, as well as from the angle of the fields of economics, marketing, and international relations.

Scholars from the field of marketing, such as Marinao-Artigas and Barajas-Portas, refer to country images as a combination of cognitive experiences and affective experiences through familiarity and reputation (2021: 2). In the field of marketing, it is a tool for implementing marketing strategies towards potential tourists and visitors (ibid.). Another field, economics, also explore the notions of country images, as it is a significant area of international business research. Country image within the scope of economics, for example, refers to "the picture, the reputation, the stereotype that businessmen and consumers attach to products of a specific country. The image is created by such variables as representative products, national characteristics, economic and political background, history, and traditions." (Nagashima, 1970 in Li et al, 2009: 623). Research in this field primarily focuses on the influence of country image on purchase intention of products (ibid.: 621; Wang et al. (2012)). Country images are also of interest to researchers in the field of political science, particularly in diplomacy and international relations. The notion of country image within the scope of political research reflects the effects that it has on people's behaviour, as well as the influence that it has in political and economic interactions with other countries (Buhmann, 2016: 15-17). For example, research in this field include the relationship between country image, diplomacy, and soft power (Kenzhalina et al, 2016: 4240).

Finally, a research field in which the notion appears is linguistics. The notion of country image (mentioned also as national image by Tarasheva (2014a)) is defined as followed; 'the cognitive representation that a person holds of a given country and its people, what a person

believes to be true about a nation', with the focus of the research being 'how images are produced and broadcast to the world (Tarasheva, 2014a: 424). Tarasheva (ibid.) examines different linguistic studies devoted to the research of country images, listing three approaches that have been taken; media studies, where the scholars examine media coverage, cultural studies, where issues regarding the boundaries in relation to 'a significant Other', and Critical Discourse Analysis, where images are construed from texts and bear some sort of impact on social structures, particularly from the perspective of the oppressed and dominated (ibid.:424-430). Research on country images include Tarasheva's analysis of Bulgaria's country image (Tarasheva, 2014b), the self-representation of African countries on the internet (Fürsich and Robins, 2002), and the self-representation of entrepreneurs in Belarus (Miazhevich, 2007).

As this research is grounded within the field of linguistics, a further examination of previous research can be useful for determining the most appropriate approach for analysis. The following subchapter examines previous research on country images, their methods, and findings.

1.2. Studies of country images in the field of linguistics

This subchapter examines several studies that analyse country images, identifying the similarities and differences that can be helpful for determining the most appropriate methodological solutions.

One such research has been done in the context of African countries, where Elfriede Fürsich and Melinda Robins examine the self-representation of African countries on the internet. This paper takes the approach of textual analysis, employing close reading to see how African countries represent themselves on the internet. The findings show that the governments of African countries address the Western society more than their own citizens, focusing on catering to Western interests instead of domestic concerns (Fürsich and Robins, 2002: 190). Examples include Senegal and Ghana, where both note the close links between tourist developments and historical slave shipping points (ibid.: 201). The reasoning behind such actions can be explained by the countries' focus on tourism and investment, allowing to earn more money by attracting Western tourists via the internet. The authors also conclude that instead of 'integration based on multi-ethnic equality', these media activities instead build an image that fits the Western ideas, neglecting 'the ideal of nation building and its anti-colonial and autarchic motives' instead reflecting recolonisation and international dependence (ibid.: 206).

Another research focuses on official media discourse in Belarus, where the self-representation of entrepreneurs in Belarussian media is examined. Miazhevich finds that two

prominent themes are used by the media; anti-globalisation and 'Sovietisation' (Miazhevich, 2007: 1334). The findings show that the stories in the newspaper 'Sovetskaya Belarussia' were more negative and that the entrepreneurs in Belarus are described as nonbeneficial to the people of Belarus, and an opposition to the 'narod' (ibid.: 1339). The findings also illustrate the Belarus' state ideology and the out-grouping of entrepreneurs from political and business communities.

The third research also focuses on press discourse, where Przyemyslaw Wilk examines stance markers about Poles and Poland in news articles of The Guardian. The paper is grounded in the theoretical framework of Critical Discourse Analysis, examining how stance relates to 'ideology reproduction in discourse' (Wilk, 2014: 80). The two stances established by the newspaper are outcasting and inclusion. Prior to Poland joining the EU, the country is construed as an outgroup, whereas after joining the EU, Poland receives a more positive stance (ibid.: 102). Another point worth mentioning is in regard to the categories that represent Poland; it is firstly categorised as being an eastern European country (prior to joining the EU), but central European country after becoming a member state (ibid.: 102-103). The author also notes that the positive representation of Poland matches the political views of the newspaper and their stance towards the EU and its enlargement process (ibid.).

The final article examined images of China in the TIME newspaper, focusing on the changes to China's image over the paper's 100-year history. Bing Zhang and Zhenzhu Shu also take a Corpus-Based Critical Discourse Analysis approach, analysing how the newspaper portrays China and what ideology is implied behind the content (2023: 34). The authors find that the reports of China vary greatly across the 100-year history of the paper, therefore the image differs between different historical periods. For example, Wartime China's image is more positive and objective, Communist China is described negatively for Western readers, whilst the more recent time periods represent China more positively and objectively once more as the country shifts away from the Communist regime. Lastly, the authors conclude that the image of China mainly depends on the newspaper, namely, their ideology, journalistic values, cultural values, and worldview differences (ibid.: 43).

The aforementioned studies show that topic of country images is often analysed from the perspective of media studies, and that the most commonly used texts are taken from different news outlets (websites, news articles, printed newspapers, etcetera). Ideology is a point that is mentioned in all of the studies, indicating that the texts are shaped by ideology, and that by analysing the language, the hidden ideological aspects can be revealed. Fürsich's and Robins' research reveals that the governments of Africa focus on the Western audience, adopting the ideology of modernisation. In Maizhevich's research, ideology is resisted by entrepreneurs,

whilst Wilk's research reveals the liberal ideology of The Guardian's discourse. Zhang's and Shu's research, meanwhile, finds that the image of China changes across time, indicating that ideology and worldviews influence the perception of the country at different periods of time.

Thus, the findings highlight the importance of ideology in language. Likewise, discourse analysis is the cornerstone of many of country image analyses, therefore these concepts must be examined in depth. The following chapter examines the notions of discourse, discourse analysis, and critical discourse analysis.

2. Overview of Discourse and Discourse Analysis

The examination of the concepts of discourse, discourse analysis, and critical discourse analysis are important for this study, as this study is based on the theoretical background of Corpus Assisted Discourse Studies. Thus, the study draws on theoretical material written by Schiffrin et al (2001), Brown and Yule (1983), van Dijk (1997), Wodak (2013), and others.

2.1. Notion of discourse

The notion of discourse varies between different scholars and disciplines. This subchapter provides comparative analysis of the different definitions that scholars have developed.

The term does not have a single definition, instead having multiple definitions that differ from one another because of 'disciplinary diversity' (Schiffrin et al, 2001: 1), and, therefore, the notion is 'essentially fuzzy' (van Dijk, 1997: 1). In a broader sense, the notion of discourse refers to language and its usage, but more specifically, particularly in linguistics, as a connected speech or writing that is longer than a conventional sentence (McArthur, 1992: 316). Discourse analysts, such as Brown and Yule (1983), van Dijk (1997), and Baker (2023) define discourse as language in use, investigating what language is used for. Discourse also refers to the different types of language topics that are specific to their field, such as political discourse, colonial discourse, media discourse, environmental discourse, among others (Baker, 2023: 4). Scholars expand to assert that discourse refers to "language in action" (Bloomaert, 2005: 2 in ibid.) and is categorised as a bundle of "meanings, metaphors, representations, images, stories, statements and so on that in some way together produce a particular version of events […]' (Burr, 1995: 48 in ibid.). This categorisation ties in together with governmental discourse and how their publications shape the perception of Latvia, and how, in turn, this might form an image of the country, the point of interest in this study. As for how discourses are defined, Baker (2023) mentions that they depend on how each person views and constructs their day-to-day experiences, thus, they are 'constantly changing, interacting with each other, breaking off and merging' (ibid.). This also means that the way that discourse is constructed and defined depends on the individual, thus, being interpretive, as one individual's views of what a particular discourse is might vary with another individual; 'where I see a discourse, you may see a different discourse, or no discourse' (ibid.). Van Dijk elaborates and establishes three main dimensions of discourse: firstly, language use, secondly, the communication of beliefs, and thirdly, the interaction in social situations (van Dijk, 1997: 1). Schiffrin et al also provide their own categorisation of discourse: 'anything beyond the sentence, language use, and a broader range of social practice that includes nonlinguistic and

nonspecific instances of language' (Schiffrin et al, 2001: 1). Nielsen and Nørreklit, meanwhile, refer to discourse as a social practice that constructs 'social identities, social relations and the knowledge and meaning systems of the social world' (Nielsen and Nørreklit, 2009: 204). Thus, the notion of discourse in the discipline of discourse analysis can be understood as the language use that goes beyond the sentence and refers to a wider range of social practices, such as how language is used to serve different interests of social groups and shape their image.

Having established the general understanding of the term, it is also important to reflect on the approach itself. The following subchapter examines the notions of discourse analysis.

2.2. Notion of discourse analysis

Discourse analysis is a term that encompasses a variety of different scholarly approaches. For example, *The Handbook of Discourse Analysis* contains more than forty collected articles that differ in the type of data collected and in the types of context considered (Schiffrin et al, 2001: 5). This indicates that the field of discourse analysis is too large to be examined completely, therefore only the major approaches are examined in depth.

In a more general sense, discourse analysis examines language in context. For example, *Oxford Companion to the English Language* defines discourse analysis as 'the analysis of connected speech and writing, and their relationship to the contexts in which they are used', and that 'discourse analysts study written texts, conversation, institutionalised forms of talk, communicative events in general, and aspects of electronic text-processing' (1992: 348). David Crystal states that discourse analysis takes 'interest in analysing the way sentences work in sequence to produce coherent stretches of language' (1997: 116), whilst several other authors, such as Brown and Yule (1983), McCarthy (1991), Hodges et al. (2008), Schiffrin (2006) share a similar idea of what discourse analysis is, namely, that it is concerned with analysing the relationship between language in use and the context alongside it.

Discourse analysis is also divided into several approaches. For example, Crystal divides the analysis into two groups; discourse analysis and text analysis, where discourse analysis focuses 'on the structure of naturally occurring spoken language, […] conversations, interviews, commentaries, and speeches', but text analysis focuses on 'the structure of written language, […] in such 'texts' as essays, notices, road signs, and chapters' (Crystal, 1997: 116). Hodges et al (2008), on the other hand, provide another classification of approaches, grouping them into three groups; formal linguistic discourse analysis, empirical discourse analysis, and critical discourse analysis. The first approach takes a structured analysis of a text

to find ‘general underlying rules of linguistic or communicative function behind the text’ (2008: 570). Empirical discourse analysis, meanwhile, use samples of written or spoken language and their use in social settings, analysing how ‘language and/or texts construct social practices’ (ibid.: 571). The last approach, critical discourse analysis, examines ‘an even wider sphere that includes all of the social practices, individuals, and institutions that make it possible or legitimate to understand phenomena in a particular way, and to make certain statements about what is ‘true” (ibid.: 570).

Critical discourse analysis (CDA) is an approach to analysing texts, primarily investigating the way that power relations and ideology is manifested in discourse, as well as the interrelationship between language and the social and political context (van Dijk, 2001; Fairclough, 1989). The aims of the approach include the exploration of links between language use and social practice, as well as the investigation of social inequality in discourse (Jørgensen and Phillips, 2002; Wodak, 2001). The three main concepts that linguists within CDA investigate include critique, ideology, and power. The first challenges surface meanings, interpreting them according to the socio-political context, and contributing to social change (Wodak, 2013: 23-25). The second concept, ideology, focuses on ‘the latent type of everyday beliefs’, and how these beliefs guide individuals’ evaluations (ibid.: 26-27). The third concept, power, focuses on how discourse ‘power abuse by one group over the others, and how dominated groups may discursively resist such abuse’ (ibid.). Thus, critical discourse analysis often analyses inequalities from ‘the perspective of those who suffer’, analysing ‘the language of those in power, who are responsible for the existence of inequalities and who also have the means and opportunity to improve conditions’ (Wodak, 2001: 10). To sum up, the difference between discourse analysis and critical discourse analysis is that CDA focuses on social problems in the society, criticising the way that power and ideology is used in discourse to create social inequality.

To sum up, the concept of discourse analysis mirrors the point established in the previous subchapter, that language does not consist of purely fixed meaning, but depends on different social factors that help to shape the images and discourses according to the ideologies and perceptions of the authors. Discourse analysis can be divided into general discourse analysis, and a critical approach. In the context of this research, analysing discourse is important. However, discourse is only one component of this study. As this study aims to analyse discourse in a corpus, a method that combines the approaches of corpus linguistics and discourse analysis is needed. Thus, the following chapter introduces theoretical material on Corpus Assisted Discourse Studies.

3. Corpus Assisted Discourse Studies as Field to Analyse Country Images

This chapter focuses on the theoretical background of Corpus Assisted Discourse Studies and the approaches that this field presents to analyse data. The approaches include frequency lists, keywords, and concordances. The study reflects on the theoretical considerations established by authors such as Baker (2023), Gries (2009; 2021), Gatto (2014), and others.

3.1. Corpus Assisted Discourse Studies (CADS)

Paul Baker's Using Corpora in Discourse Analysis is used as one of the main frameworks behind the analysis of the CADS approach. Thus, this subchapter examines his approach towards analysing discourse with the help of corpus linguistics.

The Corpus Assisted Discourse Studies (CADS) introduces an approach to analyse language through corpora or large language data set stored on computers, and corpus processes to reveal language patterns that enable researchers to identify how language is used 'in the construction of discourses (or ways of constructing reality)' (Baker, 2023: 1). Thus, in a general sense, the CADS integrates corpora and discourse to investigate how language is used in specific contexts and works as a bridge between the two different disciplines. Baker writes that the corpus-based methods have been widely used by scholars since the nineteenth century but have become more widespread and popular with the availability of personal computers (ibid.: 3). However, in the field of CADS, the purpose of the corpus is to 'enable the analysis of discourses' (ibid.). Thus, the theoretical material observed reveals the importance of bringing together corpus-based analysis and discourse analysis. The use of corpora enables the analysis of discourse, providing the researcher with authentic language data whilst allowing to identify patterns and implied meaning that can also potentially reveal country images.

3.1.1. Emergence of the CADS approach

According to Baker (2023), the emergence of CADS can be traced to the 1990s and early 2000s, where he lists different scholars who applied corpora of authentic texts together with discourse analysis (Caldas-Coutlhard (1993); Hardt-Mautner (1995); Stubbs (1996), Flowerdew (1997), McEnery et al. (2000), Holmes (2001), and others). A point to consider is that many of these studies employed a critical view of the language; however, the emerging approach, CADS, differed from the other approaches by de-emphasising the critical aspect of the research and taking a bigger interest at exploring topics where 'they may not have pre-existing hypotheses or strong beliefs about author bias or the existence of power inequalities'

(ibid.: 7). Thus, the CADS approach does not focus on power inequalities or abuses manifested through language specifically, but on identifying the more general aspects of language use. The field gained recognition and became more widely adopted in the 2010s, receiving attention in scientific journals, being presented at conferences, and being used in scientific studies, indicating its acceptance as a method of study (ibid.: 12). The focus of the Corpus-Based Discourse Analysis is important to the approach that this study takes, namely, not focusing on critical views, instead, examining the topics that are presented without the focus on ideology and power to examine the prevalent themes.

3.1.2. Advantages and limitations of CADS approach

It is also important to examine the advantages and limitations of taking the CADS approach. The main advantages listed by the author include the reduction of researcher bias, the incremental effect of discourse, resistant and changing discourses, and triangulation.

According to Baker, objectivity in research is difficult to achieve because of the personal circumstances and the cognitive biases that every researcher has, therefore it is quite difficult to be objective. Thus, an acknowledgement of one's own perspectives should be an important first step towards carrying out research (ibid.: 13). The use of a corpus allows to place restrictions on the personal views and cognitive biases by having data that is not predetermined according to some pre-existing views and ideas, reducing the ability to be selective, and, in turn, allowing for more balanced patterns and trends to show through (ibid.: 14).

The second positive point of using corpora to analyse discourse is the incremental effect of discourse, which refer to the crucial role that language plays in shapes 'underlying discourses' (ibid.: 15). The use of corpora provides the ability to determine how discourses are constructed, particularly when looking at patterns of language use and word co-occurrence in natural language texts. Examples include journalism and their ability to produce and reshape discourses, how language patterns reveal evaluative meanings and how they trigger cultural stereotypes and representation of disabled people through discourse (Becker, 1972; Stubbs, 2001 in Baker, 2023: 16). This is useful in the context of the present study, as the governmental websites are instrumental in shaping the image of the country.

The final advantage of the CADS approach is its use of triangulation, where several key points are mentioned. The first point reflects the idea that binary arguments should not exclude the use of corpus-based research, instead recommending that the logic of 'either/or' binary oppositions be disregarded in favour of 'a logic of both/and' (Derrida, 1978 and 1981 in ibid.: 19). Likewise, this leads to the point that the role of corpora or corpus linguistics

within ‘post-structuralist’ research is as an additional tool to aid research, not being the sole approach to the analysis. This, in turn, leads to the point that having triangulation in research is important, and that by using multiple methods of analysis, the different approaches work to complement each other. The author refers to Layder and her arguments; that ‘it facilitates validity checks of hypotheses, it anchors findings in more robust interpretations and explanations, and it allows researchers to respond flexibly to unforeseen problems and aspects of their research’ (Layder, 1993: 128 in ibid.: 19). Thus, by applying multiple methods of analysis, the validity of the research can be established.

Baker also mentions several disadvantages or ‘concerns’ of applying CADS approach to research.

The first mentions that the corpus analysis provides a lot of results that reflect ideas and data that researchers already know; however, the author reiterates that CADS-based research projects often produce ‘a mixture of obvious and non-obvious findings’, and that even the ‘obvious’ findings are often useful, as they provide statistical detail and reveal the extent of the findings. (ibid.: 21). Thus, the task of the researcher is to produce and incorporate both ‘obvious’ and ‘non-obvious’ findings.

Another important point refers to the interpretation of texts; who produced them, for what purposes, who read them, and so forth. These questions cannot be answered ‘by traditional corpus-based techniques, and therefore require knowledge and analysis of how a text exists within the context of society’ (ibid.). Thus, CADS can reveal language patterns, but it lies within the researcher to note what is observed and what interpretations are made from those observations.

The examination of the CADS research reveals that whilst the approach presents an objective way to uncover hidden language patterns through the use of corpora, it also poses challenges that need to be addressed and followed, such as the reduction of researcher’s bias, and the necessity for human interpretation of the data. This means that the interpretation of the data plays a large role in CADS research, and that limitations can be mitigated by the researcher having a well formulated interpretation of the results produced. Likewise, the findings reveal that corpus-based discourse analysis reduces selectivity and allows to examine both quantitative and qualitative data.

Seeing as CADS is heavily dependent on a well-built corpus, the following subchapter aims to examine and describe the different methods and approaches of classifying corpora that can be used to form the basis behind the construction of the corpus for this study.

3.2. Corpora creation and classification approaches

To establish the approach to the construction and compilation of the corpus for this research, theoretical material on corpus building and classification of corpora should be examined. The value of utilising a corpus can be summarised as follows:

> text corpora provide large databases of naturally occurring discourse, enabling empirical analyses of the actual patterns of use in a language, and, when coupled with (semi-) automatic computational tools, the corpus-based approach enables analyses of a scope not otherwise feasible. (Biber et al, 1994: 169 in Partington, 2003: 6)

The passage indicates that utilising a corpus allows to collect a large dataset that enables the researcher to examine authentic language data that is difficult to research qualitatively. Likewise, it 'redefines interpretation' (Hardt-Mauther, 1995: 22 in ibid.) by adding an 'empirical dimension to introspection and marries quantitative and qualitative research methodologies' (Haarman et al, 2002 in ibid.). This applies to the present study, as the corpus forms the main point of the analysis, but it is analysed according to qualitative and quantitative methodology. Corpora also tends to be a 'balanced, carefully thought-out collection of texts that are representative of a language variety or genre' (Baker, 2023: 57), indicating that a good or successful corpus needs to be built according to carefully selected texts, but that the choices can also open up new, previously unencountered results and thoughts.

The classification of different types of corpora is also important to mention, as the grouping of corpora depends on the features and properties that they have. Thus, the different design concepts, as well as the criteria for the compilation of corpora must be examined. For one, corpora are classified according to the time period they represent. Corpora can be synchronic, encompassing texts that represent language during a specific period of time, or diachronic, encompassing texts that represent language from different periods of time (Sketch Engine, Online 3), allowing researchers to view and track linguistic changes within it (Baker, 2023: 59). Corpora are also classified according to the language used. A monolingual corpus contains texts in one language, whilst a multilingual corpus contains two or more languages. A parallel corpus, meanwhile, is an important component in translation studies and language teaching (CLARIN, Online 4), containing language data in multiple languages that is made from translated texts, meaning that the texts align with one another (Sketch Engine, Online 5). Lastly, compilation of corpora is divided between general corpora and specialised corpora, where a general corpus contains texts that reflect balanced data about the general language, whilst a specialised corpus refers to a collection of texts where specific variety or genre of language is used to analyse particular points of interest. (Gries, 2009: 9; Baker, 2023: 56).

General corpora are also sometimes referred to as reference corpora (Weisser, 2022: 93). Their purpose is to represent language use in its entirety, 'both in terms of reception and production' (ibid.). Likewise, it is also used to extract keywords and terms, particularly when compared with the focus corpus (Online 6). A specialised corpus considers only texts that cover a single domain and adhere to specific research purposes and criteria for analysis (ibid.: 96; Baker, 2023: 56). Examples include 'language of newspapers, or the language used in academic essays, or in spoken conversations between men and women' (ibid.). Lastly, specialised corpora are described as being much smaller than reference corpora, particularly because their aim is not to cover all aspects of a given language, but to provide representativeness of data (Weisser, 2022: 96). As this study aims to compile and analyse texts taken from Latvian governmental websites, the potential corpus of this study is smaller and more specialised. However, it is important to define what the features of a specialised corpus are, and how these features would be applicable towards this study. Thus, theoretical findings on specialised corpora, their size, design, and advantages are examined in more detail to reveal the most appropriate course of action for building a corpus for the purposes of this study.

Koester (2022) writes that a smaller, specialised corpus is more advantageous because it allows a close link between the corpus and the contexts 'in which the texts in the corpus were produced', giving insights into how language is used in particular settings (2022: 49). Likewise, a small, specialised corpus presents the opportunity to balance quantitative findings of the corpus analysis with more qualitative results that take context into account (Flowerdew, 2004 in ibid.). Thus, the link between language patterns and contexts of use that specialised corpora provide are more advantageous when compared to large corpora. However, the extent of how *small* a specialised corpus is and how a corpus is specialised is not clear, therefore it is important to also define the size and the specialisation parameters of specialised corpora.

The opinions regarding what is considered a *big* and a *small* corpus varies between scholars. For example, O'Keeffe lists any written corpora under 5 million words as being small (O'Keeffe et al, 2007: 4, discussed in ibid.), whilst Flowerdew notes that small corpora contain up to 250,000 words. However, Koester states that there is no ideal size for a corpus, instead, it depends on what the corpus contains, what it investigates, and how representative it is. Thus, it can be concluded that a *small* corpus is not that small by any means, instead, the corpus size is relative and depends on the purpose of the research.

The specialisation aspect of a corpus depends on the purpose of the research as well. There are a number of parameters according to which a corpus can be specialised: specific purpose for compilation (to investigate a particular grammatical or lexical item),

contextualisation (setting, participants, communicative purpose), genre, type of discourse, subject topic, and language variety (Flowerdew, 2004: 21, discussed in ibid.: 50). Thus, having a carefully targeted corpus allows for a more reliable representation of register or genre when compared to general corpora (ibid.: 51).

Koester also writes that specialised corpora often require background information about the data, as it might be difficult to make sense of the discourse without some background knowledge (ibid.: 53). Thus, the description of the material included in the corpus is also useful, allowing the reader to familiarise themselves with the content.

The fundamental elements of a successful corpus design include authenticity, representativeness, balance, sampling, as well as size and composition (Gatto, 2014: 9; Baker, 2023: 57). The first element, authenticity, refers to the 'observation of authentic data' (Gatto, 2014: 9), which once more refers to corpora consisting of real examples of natural language use. The second important element, representativeness, refers to the fact that a corpus is not simply a collection of texts, but a representation of a specific language type. Thus, the design of the corpus depends on 'what it is meant to represent. The representativeness of the corpus, in turn, determines the kinds of research questions that can be addressed and the generalizability of the results of the research' (Biber et al, 1998: 246 in ibid.). Two types of variability are established within representativeness as well: *situational* and *linguistic*, where situational variability refers a range of text types included in the corpus, whilst the linguistic variability refers to the range of linguistic distributions (Koester, 2022: 51). Next are the terms of balance and sampling. A balanced corpus allows to cover 'a wide range of text categories considered to be representative of the language or variety under scrutiny' (Gatto, 2014: 12). Baker, meanwhile, mentions that the technique of sampling is utilised to 'ensure that the corpus is not skewed by the presence of a few very large single texts' (Baker, 2023: 57), therefore it is important to ensure that samples are collected from a range of different texts (Koester, 2022: 52). Lastly, Gatto refers to the importance of corpus size. For the author, corpus is 'an object of finite size, from which precise quantitative data concerning the language items it contains can be derived' (Gatto, 2014: 14). This means that it is up to the researcher to choose the size of the corpus, but it must be large enough for the purposes of the research, whilst also being finite. Thus, it is important to follow these steps when designing and compiling a specialised corpus.

To conclude, a specialised corpus focuses on specific topics and consists of texts limited to those domains. Specialised corpora are described as being small, but the size is relative and depends on the purposes of the research. The advantage of a specialised corpus is that it presents far less de-contextualised language, which can aid in the interpretation of data

(Koester, 2022: 55). Likewise, the context can 'tell us something about the social and cultural context from which the data were taken' (ibid.). Koester writes that specialised corpora is often analysed with the help of keyword lists and concordances, which can reveal patterns that are linked to specific contexts, showing when certain words and expressions appear and who uses them the most (ibid.). Thus, the use of the CADS approach and the design of a specialised corpus enables the analysis of a large data set, but also provides the opportunity to analyse words and their context.

Having established the theoretical foundations on corpus building and classification, it is important to examine the main tools that can be used to examine the data. Koester emphasises the usefulness of keywords and concordances when analysing a specialised corpus, whilst Mautner (2022: 253) lists frequency lists (recurring words), keywords (words that occur more frequently in one corpus than another), as well as concordances (words in their surrounding co-text) as the main tools that contribute to the study of discourse. Thus, the following subchapters examine the notions of frequency lists, keywords, and concordances.

3.3. Concept of frequency lists and keywords

The concept of frequency lists is one of the most important components of CADS research, allowing researchers to find a focus for their analyses and illustrate interesting phenomena that can then be analysed further (Baker, 2023). Gries and Rayson state that frequency lists are used when one wants to know how often words occur in a corpus, and in some cases, how widely the items are dispersed across multiple selections of a corpus (Gries, 2009: 12; Rayson, 2015: 41).

> Frequency lists allow
>
> > to direct the researcher to investigate various parts of a corpus, how measures of dispersion and distribution can reveal trends across texts and how, with the right corpus, frequency data can help to give the user a sociological profile of a given word or phrase enabling greater understanding of its use in particular contexts (ibid.: 81).

Likewise, frequency lists of lexical items present an idea of what the main themes in the corpus are (Mautner, 2022: 254). Frequency lists have a wide list of applications, ranging from language learning and teaching, diachronic variations, differences in speech acts, and others (ibid.: 78). Thus, frequency lists play an important role in uncovering the different language patterns observed in real life use. The examination of frequency lists also provides researchers with the opportunity to investigate the reasons as to why a particular word appears so frequently in a corpus, revealing the presence of hidden discourses (Baker, 2023: 161). However, frequency lists also have their limitations. Whilst they can be helpful in determining the focus of the texts, as well as the most common patterns of language, it still does not reveal

the ways in which the words are used (ibid.: 107; Hyland, 2015: 300). Frequency lists are one of the four practices from which information relevant to research is extracted. One particular use of frequency lists is that of keywords, therefore the following paragraphs examine the role of keywords in data analysis.

The notion of keywords refers to the identification of words within a set of texts that is overrepresented in one (target) corpus as compared to another, larger reference corpus (Gries, 2016: 14 in Gries, 2021: 2; Culpeper and Demmen, 2015: 90). From frequency or word lists, where simple frequency of each word can be observed, keywords can also be determined, which refer to words in the text that are more frequent and important than others. Keyword analysis is used in a wide 'range of sub-disciplines of applied linguistics from genre analyses to critically-oriented studies for different purposes ranging from producing a general characterization of a genre to identifying text-specific ideological issues' (Pojanapunya and Todd, 2016: 1). Likewise, keywords help to identify key themes and concepts in discourses, as well as typical vocabulary in a language variety, or lexical development over time (Brezina, 2018: 80). One of the first to introduce the term and popularise the idea behind the term was Raymond Williams, who found that certain words 'are "key" because they capture the essence of particular social, cultural, or political themes, thoughts or discourses' (Culpeper and Demmen, 2015: 90.). According to the scholars, keywords have 'a quantitative basis', being 'less subject to the vagaries of subjective judgements of cultural importance' and being less reliant on 'researchers selecting items that might be important and then establishing their importance', instead revealing 'items that researchers did not know to be important in the first place' (ibid.). At the same time, they argue that qualitatively defined and quantitatively defined keywords are completely separate, emphasising that 'quantitatively defined keywords can indeed be of social, cultural, or political significance if they are characteristic of social, cultural, or political texts', and that this significance can be established by 'conducting qualitative analyses of the keywords identified through initial quantitative analysis' (ibid.). Normally, keyword analysis is performed by the use of software that helps to compile and analyse the data. In the case of keywords, the software identifies and *flags* up lexical items as *key* if the difference between the focus corpus and the reference corpus is statistically significant (Mautner, 2022: 256). Thus, the identification of keywords allows researchers to analyse the similarities and differences between different texts, and, in turn, help to identify the socio-cultural contexts in which they were produced (ibid.). However, keyword analysis is not simply a set of words that characterise a corpus, but a process that requires multiple decisions such as the selection of a reference corpus, and the application of a particular statistic (Brezina, 2018: 86). Thus, it is crucial to establish an appropriate procedure to

analyse the keywords. The following paragraphs examine the theoretical background of different procedures for keyword analysis and establishes the main procedure for this study.

The procedure is an important component when analysing keywords, as the sorting of data depends on the statistical approach that is undertaken.

Firstly, the keyword procedure begins with corpora themselves. Two corpora are required; the first is a corpus of interest (also abbreviated as C) (Brezina, 2018: 80), which is also referred to as a 'focus corpus' (Kilgarriff, 2012, discussed in ibid.) and as 'node corpus' (Scott, 1997, discussed in ibid.). The terms 'focus corpus' and 'reference corpus' established by Kilgarriff (2012) are further used in the rest of the study, as they present the concepts in the most concise manner. The focus corpus is compared with a reference corpus (abbreviated as R). A statistical measure is then used to identify the words that appear in both corpora. If a word is used more often in C than in R, it is classified as a positive keyword (abbreviated as +). Likewise, if a word is used less often in C than in R, then it is classified as a negative keyword (abbreviated as -), and if a word appears more or less equally, it is classified as a lockword (abbreviated as 0). However, in terms of practical application, the author mentions that this process is done automatically by corpus software, appearing as two wordlists, where one shows the findings of C and the other shows the findings of R (Brezina, 2018: 80). Paquot and Bestgen describe a similar process that consists of five stages, where wordlists are generated for the focus and reference corpora, minimum threshold is set, a statistical test is applied, words are filtered out, and results are produced (Scott and Tribble, 2006: 58-60, discussed in Paquot and Bestgen, 2009: 3). This shows that keyword analysis has an established procedure that is relatively straightforward, but that the application of statistical measure is the more complicated aspect. This leads to the second point, the statistical measure.

There are several statistical measures for comparing corpora, with each having their own calculations for analysing the data, such as log-likelihood (LL) statistic, odds ratio (OR) statistic, simple maths parameter statistic, and chi-square tests (Pojanapunya and Todd, 2018: 1; Paquot and Bestgen, 2009: 1). According to Brezina, the most traditional statistical measure is the LL statistic. It is used to identify whether the differences between the focus corpus and the reference corpus are statistically significant (ibid.: 83-84). The cut-off point for the LL statistic is 3.84, which allows to determine that the occurrence is statistically significant and is not simply happening by chance. Lastly, the LL statistic normally identifies a large number of keywords, therefore this approach normally filters out the top keywords for further analysis (ibid.: 85).

The second major statistical measure is the simple maths parameter (SMP), where the ratio between relative frequencies of words in C and R is identified, and through which the identification and sorting of keywords can be performed. In this method, relative frequency of the focus corpus is divided with the relative frequency of the reference corpus, but a constant *k* is added as well, being any positive number, such as 1, 10, 100, or 1,000 (ibid.). The constant *k* that is added serves as a filter that allows to focus on words 'above certain relative frequencies in the corpus' (ibid.). This means that by using a higher number as a constant, the words below that threshold would be filtered out, which is convenient as it allows to set a threshold of the frequency for the corpus built for this study. Meanwhile, relative frequency refers to the 'number of occurrences of an item per million tokens, also called i.p.m. (instances per million)', and is used 'to compare frequencies between corpora (or datasets) of different sizes' (Online 7). It should also be noted that the normalisation basis can differ and be set according to the purposes of the research. The formula of relative frequency is as follows: number of hits divided by corpus size in millions of tokens equals frequency per million (ibid.).

The formula of the simple maths parameter is presented in the figure below:

Figure 3.1 **Formula of the simple maths parameter**

$$\text{simple maths parameter} = \frac{\text{relative frequency of } w \text{ in C} + k}{\text{relative frequency of } w \text{ in R} + k}$$

The simple maths parameter formula provides the keyness score (Online 8), which can then be used to identify the most common keywords.

To compare, the LL method helps to prove that that certain frequencies of the same word differ between the focus corpus and the reference corpus and are statistically significant (ibid.: 83-84). However, the LL method is described as being 'prone to identifying far too many keywords', and that 'relatively small frequency differences between C and R can reach statistical significance' in larger corpora (ibid.: 84-85). Likewise, the LL statistic is more sophisticated and more difficult to interpret. On the other hand, the SMP measure is more convenient to use and provides a parameter which 'allows the user to specify whether they want to focus on higher-frequency or lower-frequency words' (Kilgarriff, 2009). Thus, for the purposes of this study, the SMP measure is applied to determine and analyse the keywords, as this method allows to mitigate the limitations that can potentially dominate the results and allows to select the minimal frequency threshold.

3.3.1. Summary

To sum up, the theoretical material indicates that the keyword procedure depends on parameters such as the selection of the reference corpus, the implementation of minimum frequency cut-off points and the choice of the statistical measure (Brezina, 2018: 86). The theoretical background indicates that the SMP statistical measure is less sophisticated and presents the option to focus on higher-frequency and lower-frequency words, being more useful and convenient for the study. The theoretical material also shows that the link between quantitatively identified keywords and qualitative analysis is established through research that uses complementary methods to analyse the data qualitatively**,** namely, semantic annotation of keywords, as well as concordances of the words. Thus, an additional approach that combines quantitative and qualitative analysis is necessary. The following subchapter examines concordances as an equally important tool for the CADS approach.

3.4. Concordance analysis

In CADS, Baker lists concordance analysis as being one of the most effective techniques for carrying out a close examination of language and its context (Baker, 2023: 107). A concordance (also known as a concordance line) refers to a list where the occurrences of a particular word are shown in the context that they take place, having a few words to the left and right of the search term (ibid.). This procedure is also known as *key word in context* (KWIC). The use of concordances allows to identify the associations and connections that words have, and how 'they take on specific meanings for particular individuals and in given communities' (Hyland, 2015: 301). Corpus software tools that have a concordance 'are designed to find all the occurrences of a search term in a corpus and display these in an ordered fashion together with the words that surround them' (Anthony, 2022: 115). This is useful because it allows to view particular occurrences in close contexts (words that surround the keyword), as well as wider contexts (seeing the word within the context of the whole text). Likewise, it is useful in the sense that it enables analysis of large datasets. Some authors emphasise concordance lines as being 'a powerful shortcut for identifying patterns of usage', allowing researchers 'to draw up specific profiles of lexical items and gauge the role that they play in certain discourses' (Gillings and Mautner, 2024: 34-35), and allows for researchers to examine the keyword and its accompanying words without distractions (Bianchi, 2012: 49). CADS approach, meanwhile, has specific differences that are highlighted, as this study takes this particular approach to analysing data.

The first includes CADS approach beginning with a social question, whilst other approaches, such as syntactic and lexicographic, tend to begin with a linguistic question

(Gillings and Mautner, 2024: 35). The reason behind the motivation is linked to the research interests; the CADS researcher is interested in the themes and perspectives present in a specific 'social domain', how actors and events are being framed, and what kind of evaluations are attached to said actors (ibid.).

The second difference lies in the patterns that can be observed. Where the grammarians and lexicographers focus on linguistic patterns, the CADS approach examines language patterns for 'social attitudes, values and behaviour – telling us, for example, what certain groups of people are typically shown to be doing or what characteristics are attributed over to them' (ibid.).

The last difference between the CADS approach and the approaches of lexicography and grammar is in regard to the amount of co-text that researchers draw on. Whilst grammarians are interested in clauses or sentences, the CADS approach draws on a much wider range of co-text. Depending on the material, collocates can draw from 'the immediate left or right of the node' to receive the relevant information, whilst in other cases, the information can be further away (ibid.). Thus, concordance analysis can be much larger in the case of CADS and reveals the knowledge of discourse. Thus, it can be concluded that the aforementioned characteristics are important to this research as well, as the focus of the research is more social, it focuses on language patterns instead of linguistic patterns, and it examines concordances not only with its accompanying words, but how it functions within the text as a whole.

3.5. Summary

The examination of theoretical material on Corpus Based Discourse Studies reveals that a corpus-based approach to analyse discourse is the most applicable when analysing a specialised corpus. Likewise, the most useful and applicable methodological techniques to analyse the image of Latvia include frequencies, keywords, semantic annotation and concordances. Lastly, the theoretical classification of corpora and corpus building reveals the most important strategies, elements and classifications that need to be undertaken when building a corpus, providing an outline for designing and building appropriate corpora for the purposes of this study. The next chapter introduces the research methodology, corpora, and findings of the analysis.

4. Analysis of Image of Latvia

This chapter contains the empirical part of the research. First, the methodology, research procedure, research tools, and corpora are described. Afterwards, the discussion of results is presented. The results consist of frequency and keyword analysis, as well as concordance analysis that contribute towards the image of Latvia.

4.1. Methodology and procedure

The research method employed in the empirical part of the study is the Corpus-Assisted Discourse Analysis methodology, which allows to employ both quantitative and qualitative methodologies and was described in Chapter 3. Corpus-Assisted Discourse Analysis combines corpus linguistics with discourse analysis, allowing to use large amounts of textual data to view 'micro-linguistic choices within a much wider discoursal and societal context' (Gillings et al, 2024: 4).

The theoretical resources of keyword analysis, semantic annotation and concordance analysis outlined in chapters 3.3 and 3.4 were employed (Gries, 2021; Brezina, 2018; Baker, 2023). The former generates quantitative statistical data about the Latvian Governmental Web Texts corpus and its subcorpora, whilst the latter two reveal qualitative data on the themes that keywords reveal and a further analysis of the keywords in context. From there, it was possible to derive the main thematic threads to form statements that present the image of Latvia.

To answer the proposed research questions, the procedure was set as follows. Firstly, the data for the creation of the corpora was taken from several Latvian governmental websites and compiled into five subcorpora (described in the following subchapters), where each subcorpus consists of texts taken from a single website. This provides a clear distinction between the most common occurrences in each website, as well as the similarities and differences that appear. Secondly, after the texts were grouped, the data was examined through keyword analysis. Keywords were sorted according to the simple maths parameter, where the keywords that have a higher keyness score ranked higher. As lexical words were the primary focus of the analysis, grammatical words were omitted from the analysis. Thirdly, concordance analysis was applied. The concordance window was set as default on Sketch Engine, providing additional data of the KWIC. From there, the main themes of each subcorpus were identified and described, which were then compared with other subcorpora and all together. Lastly, the findings were formulated into statements that present the country image of Latvia as established by the governmental websites.

4.2. Selected sources

The data selected for the corpus is important, as each source is selected to represent a particular area of Latvian governmental discourse. In this study, the data was taken from several of Latvia's official government websites, thus, it is important to provide more context about each of these websites.

4.2.1. Ministry of Foreign Affairs of the Republic of Latvia

The Ministry of Foreign Affairs website serves as the official online platform for the ministry, producing and sharing information of Latvia and its participation in foreign relations and diplomatic activities.

There is a variety of content that can be found on the website. For example, information on the country's foreign policies, political, economic, and military alliances and other activities are provided in separate sections of the website. Likewise, there is also general information on Latvia's diplomatic missions in other countries, as well as their services and contact information is provided. It also provides a section of *news*, where news articles, press releases and official updates related to Latvia's external relations and foreign policy are provided.

To sum up, this website presents Latvia's official relations with other governments and positions themselves on the global stage.

4.2.2. Investment and Development Agency of Latvia

The Investment and Development Agency of Latvia's website is the official platform for the investment promotion agency, which is under the management of the Ministry of Economics of Latvia.

The purpose of this website is to promote Latvia for potential investment and economic development by providing assistance and resources to potential investors. The website encourages various investment opportunities by presenting industry sectors where investment opportunities are possible, as well as incentives for investing in Latvia. Likewise, they emphasise Latvian companies and their expansion of export capabilities, development of tourism in Latvia, as well as research and development activities. The website consists of different tools and pages created to attract and help investors build their businesses in Latvia, also providing statistics and information on trade and industries. As with the previous website, this website also provides a detailed coverage of events in this field through their news articles.

To sum up, this website presents Latvia's efforts to bring investment, trade, and tourism to Latvia, raising the profile of Latvia to the Western audience and presents itself as a good location for investment.

4.2.3. Constitutional Court of the Republic of Latvia

This website serves as an official platform for the Constitutional Court of the Republic of Latvia.

It is an important institution as it provides resources, services, and descriptions of the country's judicial system. Likewise, it presents rulings and events surrounding the court and its decisions in safeguarding of the Latvian Constitution. The website consists of the history, structure, and other general information about the court, the international cooperation that they have with other law institutions, among other things. This website also has a news section, where news updates about the events involving the court, legal processes and decisions are described.

This website is important as it provides an insight into the latest developments in Latvian jurisprudence, and how Latvia upholds law and its constitution.

4.2.4. Latvia.EU (missionLatvia)

This website serves as an official platform for the Investment and Development Agency of Latvia and the Ministry of Economics.

The purpose of this website is to promote Latvia, its culture, tourism, and economy to a global audience. Thus, this website is more representative of Latvia, appealing towards potential foreign tourists and investors. The website showcases a wide range of content, consisting of information about various aspects of Latvia, such as the history and identity of Latvia, key facts about the country, arts and culture, traditions, as well as information on business and innovation in Latvia. A sizeable portion of the content focuses on nature, highlighting the landscapes in Latvia, as well as the traditional holidays that are rooted in nature. Lastly, the website also highlights Latvia's achievements and landmark events, focusing on achievements in sports, culture, and art.

To sum up, this website is important as it explicitly describes Latvia to the global audience, serving as a resource for potential tourists and investors looking to learn about Latvia, and the potential to invest in it.

4.2.5. Ministry of Defence of the Republic of Latvia

This website serves as the official platform for the Ministry of Defence.

It provides a variety of materials regarding the military capabilities of Latvia, a wide range of information about the defence and security policies of Latvia, the international cooperation between Latvia and its military allies, such as NATO, its involvement in the support of Ukraine, as well as more information about topics such as military objects, state defence service, recruitment, among others. Lastly, the website also presents Latvia's efforts in international collaboration, such as the involvement and contribution of Latvian Armed Forces in regional and global security efforts. The website consists of different sections that cover the aforementioned information, but also includes a news section which offers information surrounding the events related to the Ministry of Defence, the Latvian Army, as well as the collaboration between the international institutions.

To sum up, this website presents Latvia and its approach to the prevention of military threats and safeguarding of Latvia's independence, and the collaboration with its allies.

4.3. Description of the corpora

This subchapter describes the focus corpus alongside its subcorpora, as well as the reference corpus.

The study built a specialised corpus by applying Baroni and Bernardini's (2006) notion of the web as a corpus shop, where 'researchers using the web as a corpus shop select and download texts retrieved by search engines to create 'disposable corpora' either manually' or semi-automatically (Baroni and Bernardini, 2006: 10-14, discussed in Gatto, 2014: 37). The data for the corpus was collected semi-automatically. The texts from navigation bars of the websites were collected first, then all the hyperlinks of indexed news articles (or press releases) were gathered and compiled manually into a separate document. Afterwards, all of the texts (manually collected texts from navigation bars and hyperlinks of news articles) were compiled through Sketch Engine's *Create Corpus*. The starting point for selecting the news articles was 2023 (with the exception of texts taken from navigation bars). Thus, the texts compiled refer to those that had been produced since that time period. Balance and representativeness were ensured by careful selection of the data; multiple websites were used to reveal different aspects of content produced by the Latvian government; the website of Foreign Affairs as representativeness of events related to Latvia's foreign policy, Investment and Development Agency focusing on the business and investment in Latvia, Constitutional Court of the Republic of Latvia showing synergy of law between Latvia and its

allies, missionLatvia as a page that is used as a platform to invite and intrigue potential investors to build their businesses in Latvia, and the Ministry of Defence as a page that describes the military and security aspects of Latvia.

Combining the classifications of corpus design and the selected sources established in chapters 3.2 and 4.2, the focus corpus (titled Latvian Governmental Web Texts Corpus) can be described as follows: it is a specialised corpus, as it contains a collection of texts that is made to represent a specific variety of discourse, namely, language about Latvia and the events surrounding Latvia that is produced and presented by its own government. The corpus is synchronic, as it consists of written texts that represent language during a brief period of time (2023 to 2024, with a few exceptions), and the corpus is monolingual, as all of the texts collected are in English. Lastly, the corpus is unannotated, as it has not been edited to include additional linguistic information.

In total, Latvian Governmental Web Texts corpus consists of 392,695 tokens and 346,364 words. The following table presents the focus corpus and its subcorpora alongside their size.

Table 4.1 **Corpus size**

Name	**Tokens**	**Words**
Latvian Governmental Web Texts	392,695	346,364
Ministry of Foreign Affairs	193,661	171,823
Investment and Development Agency	70,910	61,328
Constitutional Court of Latvia	59,017	52,645
missionLatvia	21,437	18,406
Ministry of Defence	47,670	42,162

The texts selected for the corpus range from January 2023 to April 2024. The time period was selected to identify the image of Latvia in this particular time period – to examine the patterns and trends that have developed in the past year and a half. The data of the texts compiled are not structured. However, as the texts themselves are all recent and describe similar events and information that take place during the timeframe, a further classification is not necessary.

The reference corpus selected for this study is TenTen Corpus Family's English Web 2021 (enTenTen21) corpus. The corpus is made up of texts collected from the Internet. The corpus consists of 61,585,997,113 tokens and 52,268,286,493 words, and covers a large variety of different genres, topics, and web sources. This corpus was selected because of its relevance, consisting of recent English written texts that have been compiled from the web. During the compilation of the focus corpus, it was also observed that different governmental

institutions use different varieties of English, or a mixture between British and American English. As a result, a reference corpus that includes both is useful for this study. It is also important to have a reference corpus that is much larger than the focus corpus, and in the case of the English Web 2021 corpus, the sample is much larger than the sample of the focus corpus.

4.4. Tools

The main tool for this study is the Sketch Engine online software, which offers the possibility to create corpora, and provides the option to perform frequency, keyword, and concordance analysis. The tool allows for an easy access to the compiled texts, and immediately allows to interpret the data by having file analysis tools embedded within the software. Likewise, Sketch Engine contains the reference corpus which is already incorporated in the tool itself, which makes it feasible to compare the data between the reference corpus and the focus corpus. With the University of Latvia receiving its Sketch Engine licence, it was deemed the most appropriate tool for the purposes of this study.

To perform the semantic annotation of the keywords, the USAS (UCREL Semantic Analysis System) online English tagger and its semantic tagset were used (Online 9). The tagset consists of 21 semantic categories that expand into 232 subcategories (Rayson et al, 2004), and is available online (Online 10). The tagset and the tagger were used semi-automatically, where the tagger was used to group the words into categories automatically, but the data was compiled and grouped manually.

4.5. Frequency of the whole corpus

Before proceeding to analyse the data in subcorpora, it is important to examine the base frequency of the Latvian Governmental Web Texts corpus. The examination of the frequency list allows to identify the most commonly used words, note the significance of frequently occurring words, and serve as a guide for the rest of the analysis.

Table 4.2 shows the top 60 most frequently used lexical words of the entire corpus, providing an insight into the most commonly used words in Latvian governmental discourse.

Table 4.2 **Frequency list of the Latvian Governmental Web Texts corpus**

Rank	Word	Frequency	Rank	Word	Frequency
1	latvia	3,324	31	nato	514
2	foreign	2,165	32	new	496
3	ukraine	1,519	33	europe	465
4	affairs	1,429	34	rights	454
5	latvian	1,424	35	riga	447

6	cooperation	1,306	36	united	425
7	minister	1,288	37	russia	425
8	european	1,220	38	secretary	416
9	defence	1,161	39	business	416
10	support	1,097	40	year	413
11	court	1,093	41	region	411
12	international	1,009	42	part	396
13	security	954	43	political	396
14	constitutional	948	44	ambassador	392
15	state	933	45	work	390
16	ministry	889	46	russia's	380
17	national	845	47	member	379
18	development	837	48	kariņš	376
19	law	777	49	policy	368
20	council	758	50	conference	365
21	countries	749	51	including	364
22	eu	722	52	krišjānis	360
23	latvia's	696	53	economic	358
24	states	660	54	case	355
25	meeting	628	55	world	351
26	baltic	613	56	republic	348
27	ministers	560	57	war	348
28	union	555	58	visit	345
29	military	535	59	relations	345
30	information	521	60	legal	343

The table indicates several interesting themes. For one, several of the most frequently recurring words include *Latvia, Latvian, Latvia's* (first, fifth, and twenty third respectively), which make sense and emphasise that the corpus focuses on discourse about Latvia and Latvians. However, other countries and alliances appear prominently in the corpus as well. The third most frequently recurring word is Ukraine, which heavily implies topics of regional and international relations, as well as the topics of security and defence. Words *defence* (9), *security* (13), *law* (19), *military* (29), and *NATO* (31), *war* (57) explicitly refer to security and defence as well. The same can be said about words *Russia, Russia's* (37 and 46) which reflect the involvement of Russia in the discourse about Latvia's security and defence. Then there are words that also illustrate the alliances that Latvia is a part of, such as *European (*8), *council* (20), *Baltic* (26), *Union* (28), *member* (47). The topic of government, law and politics is also prominent, emphasised by words such as *minister* (7), *court* (11), *constitutional* (14), *state* (15), *national* (17), *secretary* (38), *political* (43), *ambassador* (44), *policy* (49), *economic* (53), *case* (54), *relations* (59), and *legal* (60). Lastly, the frequently used words of *Kariņš* (53) and *Krišjānis* (57) are mentioned often, indicating the reflection of political leadership and a recognition of key political figures in the discourses about Latvia.

Overall, the frequency list reveals preliminary results of words that can be grouped into different thematic categories, a pattern that will be followed in the following subchapters. The results indicate a multifaceted engagement by Latvia on topics of national issues, as evidenced by words such as Latvia and Latvian, the war in Ukraine and its impact on security and defence of Latvia, as well as topics and issues surrounding law, politics, and political relations with other alliances and states. The next step would be the keyword analysis of the whole corpus, therefore the following subchapter examines the keyword list by applying the reference corpus to identify the image of Latvia.

4.6. Keyword analysis

This subchapter illustrates the quantitative and qualitative analyses of the data derived from the corpus. First, the keywords are grouped and analysed according to their keyness scores, revealing statistical data. Afterwards, semantic tagging of the keywords is performed, and the keywords are grouped according to the semantic categories established in chapter 4.4. The whole corpus is examined first, followed by separate examination of each subcorpus.

4.6.1. Keyword analysis of the whole corpus

The first keyword list was generated by selecting the Latvian Governmental Web Texts as the focus corpus and the TenTen Corpus Family's English Web 2021 (enTenTen21) as the reference corpus. The list was sorted by the keyness score by using the simple maths method which is built-in into the Sketch Engine software. The top 150 keywords with a minimum frequency of 200 were taken and compiled into a table and can be examined in Appendix 1. Once more, the page tags that were incorporated in the page text were manually removed. Table 4.3 shows the top 20 keywords to provide a general insight into the results.

Table 4.3 **Top 20 keywords of the Latvian Governmental Web Texts corpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	4,021	142,310	10,239.50	2.31	3,093.1
2	latvian	1,424	73,965	3,626.22	1.20	1,648.0
3	kariņš	376	191	957.49	< 0.01	955.5
4	krišjānis	360	220	916.74	< 0.01	914.5
5	rinkēvičs	287	170	730.85	< 0.01	729.8
6	edgars	268	2,256	682.46	0.04	659.3
7	riga	459	72,183	1,168.85	1.17	538.6
8	liaa	203	421	516.94	< 0.01	514.4
9	andris	213	5,805	542.41	0.09	496.6
10	baltic	613	220,674	1,561.01	3.58	340.8
11	ukraine	1,675	994,839	4,265.40	16.15	248.7
12	ministers	480	273,128	1,222.32	4.43	225.1

13	contested	241	113,180	613.71	1.84	216.6
14	affairs	1,379	1,289,452	3,511.63	20.94	160.1
15	nato	595	665,102	1,515.17	10.80	128.5
16	cooperation	1,306	1,715,762	3,325.74	27.86	115.3
17	constitutional	948	1,268,581	2,414.09	20.60	111.8
18	aggression	338	423,427	860.72	6.88	109.4
19	defence	1,167	1,686,029	2,971.77	27.38	104.8
20	bilateral	278	366,949	707.93	5.96	101.9

Table 4.3 reveals that the words with the highest keyness are *Latvia and Latvian* (3,093.1 and 1,648.0 keyness score), with the keywords of Latvian politicians, the capital of Latvia, and the abbreviation of the Investment and Development Agency also having significantly higher keyness scores than the rest of the keywords (*Kariņš* (955.5), *Krišjānis* (914.5), *Rinkēvičs* (729.8), *Edgars* (659.3), *Riga* (538.6), *LIAA* (514.4), *Andris* (496.6)). The closest keyness score that follows is 340.8, which is attributed to *Baltic*, a term that is used when referring to the Baltic region and the three Baltic countries of Estonia, Latvia, and Lithuania.

Table 4.3 also shows that the top keywords refer to similar topics when compared with those that occur at the top of the frequency list. The keywords refer to Latvia and Latvians themselves (*Latvia, Latvian, Riga)*, the war in Ukraine *(Ukraine, aggression, defence)*, alliances *(Baltic, NATO)*, governments, politics and institutions *(LIAA, ministers, contested, affairs, cooperation, constitutional, bilateral)*, and specific individuals *(Kariņš, Krišjānis, Rinkēvičs, Edgars, Andris)*. Some of the keywords, such as personal names, locations, and businesses indicate a clear distinction for the Latvian Governmental Web Texts corpus, whilst some words, such as *ministers, contested, affairs, cooperation, defence* are less distinctive and can be encountered in more general texts. Thus, the top keywords indicate that they are content-distinctive and can be considered legitimate when examining and characterising the image of Latvia. At the same time, the keywords provide a considerably broad image about the discourses of Latvia, therefore the examination of all of the keywords in the Latvian Governmental Web Texts corpus as well as the examination of keywords in each subcorpus is vital, as the findings might indicate different topics and themes that can then be analysed and compared.

To see whether the themes revealed by the top keywords can be generalised to the entire corpus, the 150 keywords with a minimum frequency of 200 were grouped according to semantic categories semi-automatically. The words were tagged automatically, but compiled and grouped manually.

Figure 4.1 shows the main semantic categories of the top 150 keywords. In total, 13 semantic categories were identified according to semantic categorisation.

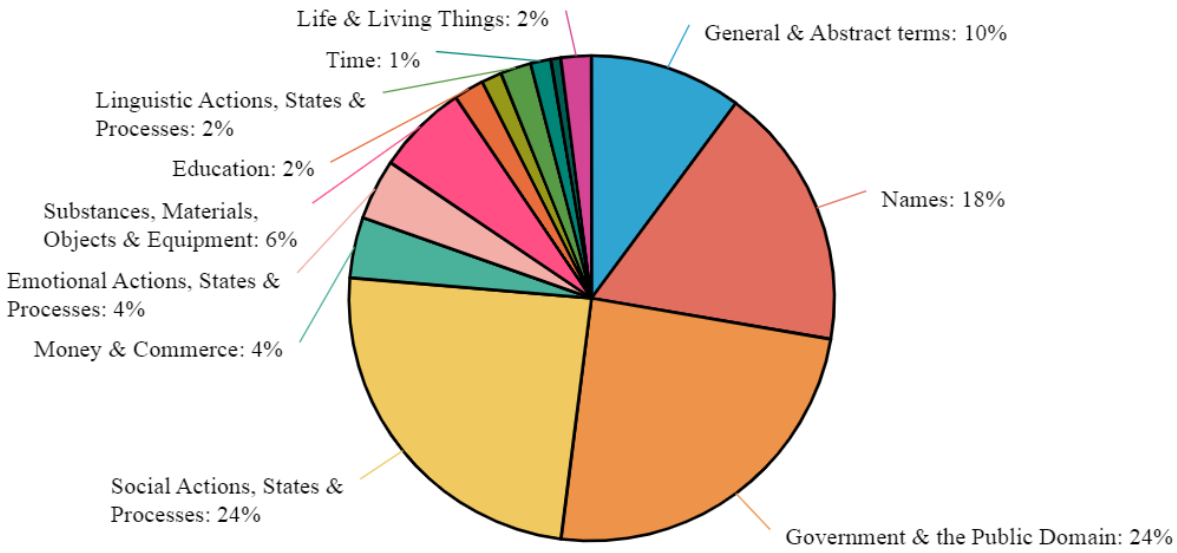

Figure 4.1 **Semantic categories of the whole corpus**

The graph indicates that the prominent two categories (24% each) are social actions, states and processes, and government and the public domain. The social actions category includes keywords and subcategories such as helping/hindering (*cooperation, promote, support, protection, ensure, implementation*), participation (*meeting, meet, visit*), toughness (*strengthen*), power (*order, hold*), relationships (*partner, establish, continue, include, provide, relation(s), union, member, united*), among others. The government and the public domain includes keywords and belonging to such thematic categories as governmental institutions and roles (*minister, ministers, constitutional, foreign, affairs, ambassador, council, nations, official, state, government, political, committee, policy, secretary, republic, agency*), law and order (*court, representative, law, legal, sanction, force*), and warfare, defence and the army (*defence, armed, military, security*), among others.

The third largest category is names (18%), which includes subcategories and keywords such as personal names (*Kariņš, Krišjānis, Rinkēvičs, Edgars, Andris, President, Director*), geographical names (*Latvia, Latvian, Riga, Baltic, Ukraine, Ukrainian, EU, Russia, Russian, Europe, European, Country*), other proper names (*LIAA, NATO, Ministry, UN, February*), among others. The rest of the categories consist of general and abstract terms (10%) that contain words such as (*priority, important, value, case, event, view, general, rule, head, global, current, potential, interest, right*), substances, materials, objects and equipment (6%) (*programme, provision, mission, technology, energy, product, article*), money and

commerce (4%) (*industry, market, business, company, export, innovation*), emotional actions, states and processes (4%) *(contested, aggression)*, education (2%) (*conference, training, project*), linguistic actions, states and processes (2%) (*organisation, development, language, discuss, discussion,*), life and living things (2%) (*society, human, person)*, time (1%) (*future, during)*, movement, location, travel and transport (1%) (*between, part)*, and numbers and measurement (1%) (*million*).

The findings show that the applied tagset categorisation reveals umbrella categories that hold more narrow themes under the surface. For example, it is clear that the prominence of the keywords *Ukraine*, *Russia, aggression, defence, security*, and other similar keywords refer to the Russia's war in Ukraine and the security and defence concerns that affect Latvia. This, however, is not clearly reflected through semantic categories. Thus, it can be interpreted that a prominent theme that the corpus reveals is the military conflict in Ukraine, which would include the aforementioned keywords.

To sum up, the keyword analysis of the whole corpus reveals that the discourse about Latvia is multifaceted; it consists of many different semantic categories that reveal the overall topics of the texts. Social actions, states, and processes category that reveals the themes of helping and hindering, participation, and relationships, the government and the public domain category (government, governmental institutions and roles, law and order, warfare, defence and security) reveals the political processes revolving Latvia, namely, the work of the many governmental institutions, the preservation of law, and an emphasis on security and defence of Latvia. The names category (personal names of governmental officials, geographical locations, other proper names of governmental institutions) establish the fact that the participants occupy a significant portion of the overall discourse about Latvia and appear as social actors that shape the events themselves. The other semantic categories appear less, referring to topics such as commerce, education, and emotional processes. The findings also show that the categories are interrelated with one another, receiving multiple tags, and therefore are overlapping, indicating that the themes are interrelated. Having examined the corpus as a single entity, it is important to examine each subcorpus on its own, as each subcorpus might provide a different outlook of what the categories and thus the themes are when examined on its own. The following subchapters examine the keywords of each subcorpus separately.

4.6.2. Keyword analysis of the Ministry of Foreign Affairs subcorpus

The Ministry of Foreign Affairs (MFA) subcorpus is the biggest of the five subcorpora compiled, consisting of 193,661 tokens, amounting for almost half of the corpus. As a result,

the minimal frequency threshold was set at 100, yielding 178 keywords. The table below shows the top 20 keywords of the MFA subcorpus.

Table 4.4 **Top 20 keywords of the MFA subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	2,367	142,310	12,222.39	2.31	3,692.0
2	kariņš	375	191	1,936.37	< 0.01	1,931.4
3	krišjānis	359	220	1,853.75	< 0.01	1,848.2
4	latvian	715	73,965	3,692.02	1.20	1,677.9
5	rinkēvičs	284	170	1,466.48	< 0.01	1,463.4
6	edgars	264	2,256	1,363.21	0.04	1,316.0
7	pelšs	168	27	867.50	< 0.01	868.1
8	andris	182	5,805	939.79	0.09	859.7
9	nb8	127	369	655.79	< 0.01	652.9
10	riga	233	72,183	1,203.13	1.17	554.4
11	baltic	421	220,674	2,173.90	3.58	474.5
12	ukraine	1,424	994,839	7,353.06	16.15	428.7
13	ministers	414	273,128	2,137.76	4.43	393.5
14	moldova	162	90,317	836.51	1.47	339.6
15	affairs	1,308	1,289,452	6,754.07	20.94	307.9
16	aggression	317	423,427	1,636.88	6.88	208.0
17	estonia	126	158,220	650.62	2.57	182.6
18	underline	222	331,356	1,146.33	5.38	179.8
19	bilateral	231	366,949	1,192.81	5.96	171.6
20	cooperation	937	1,715,762	4,838.35	27.86	167.7

The top 20 keywords in the MFA subcorpus, for the most part, are almost exactly the same as in the overall corpus. This can be explained by the large amount data that this subcorpus holds, therefore many of the keywords recur. The top keywords almost exclusively refer to names (*Latvia, Kariņš, Krišjānis, Latvian, Rinkēvičs, Edgars, Pelšs, Andris, NB8, Riga, Baltic, Ukraine, Moldova, Estonia*), highlighting the fact that many of the Latvian names are not prominently featured in the reference corpus and therefore appear as highly key. The second most important group appearing in the top keywords is related to social actions (*bilateral, cooperation, underline*), governmental topics (*minsters, affairs)* and the defence aspect (*aggression, Ukraine*). These already reflect the previously established themes in the whole corpus.

The MFA subcorpus was semantically tagged using the same approach as in chapter 4.6.1. and yielded 15 categories. The following results can be observed in Figure 4.2.

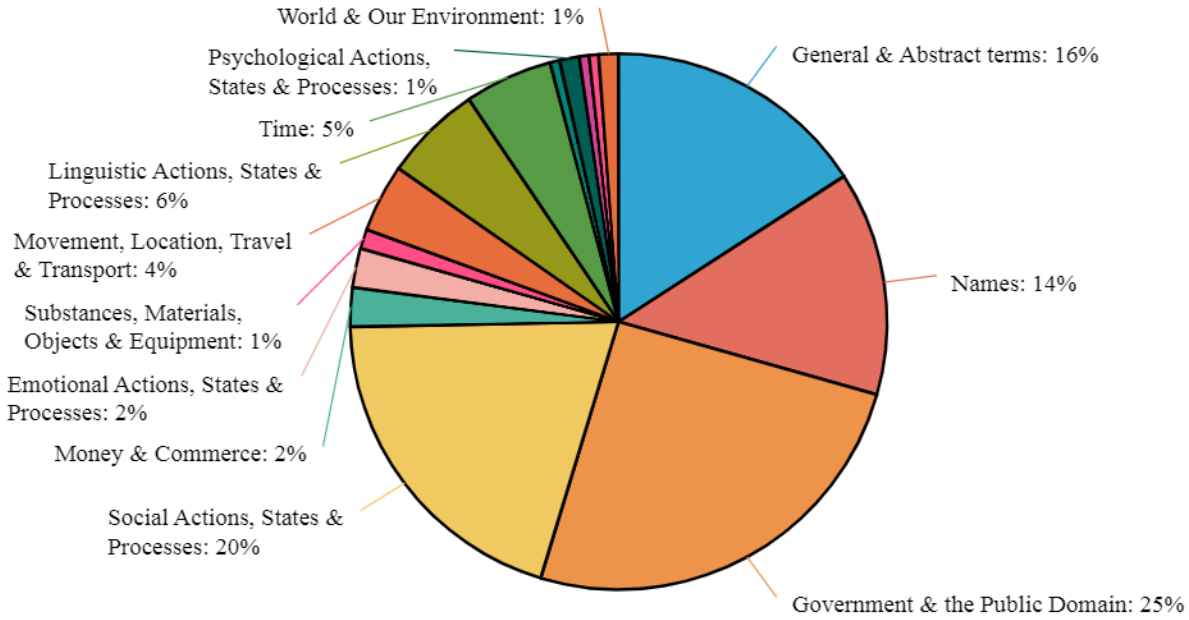

Figure 4.2 **Semantic categories of the MFA subcorpus**

The four major semantic categories have remained roughly the same when compared to the categories of the whole corpus. The most prominent one is the government and the public domain category, which has been identified in 25% of the keywords. The category consists of topics such as governmental institutions and roles (*ministry, parliamentary, minister, ambassador, presidency, delegation, secretary, republic, council, summit, representative, official, political, country, committee, expert, commission, director, president, general, head, government, union*), law and order (*reform, crime, policy, law, rule, call, official, sector, sanction*), as well as defence and security (*security, military, border, mission, war*). The second largest group is the category of social actions, states and processes (20%). Subcategories and examples include helping (*reconstruction, assistance, support, welcome, provide, take, work, strengthen, promote*), participation (*meeting, meet, participant, member, society, public, consultation, visit, part, role*), relationships (*bilateral, cooperation, partner, united, joint*), and power (*challenge, hold, decision, order, close, plan, take, right*).

The third largest group is the general and abstract terms category (16%), which contains subcategories and examples such as general actions (*make, include, ensure, develop, implementation, integration, process, situation*), chance and luck (*opportunity*), inclusion and exclusion (*include*), affect (*change, develop, action, implementation, integration, effort*), classification (*particular*), comparing (*contrast*), getting and giving (*exchange*), and importance (*importance, priority, high, important*). The names category follows right after with 14% of the words. Examples include personal names (*Kariņš, Krišjānis, Rinkēvičs, Edgars, Pelšs, Andris*), geographical names (*Latvia, Latvian, Baltic, Ukraine, Moldova, Estonia, Lithuania, Ukrainian, Russia, European, Russian, Europe*), and other proper names

(*Riga, NB8, NATO, UN, EU*). The rest of the semantic categories include linguistic actions, states and processes (6%) (*dialogue, agenda, discuss, discussion, address, language, media, view, note, project*), time (5%) (*February, current, November, year, continue, March, future, present, time*), and movement, location, transportation and travel (4%) (*international, region, regional, sea, global, national, place*), among other smaller ones.

The keyword analysis of this subcorpus reveals a similar pattern to the whole corpus, indicating that the content is roughly the same. The semantic category of government occurred the most, which makes sense considering the purpose and the content of this subcorpus. Likewise, the recurring emphasis on security and defence, as well as the prominence of the keywords *Ukraine* and *Ukrainian,* and *Russia, Russian* indicate that the discourse is heavily focused on the security and defence topics regarding Latvia and Ukraine. Interestingly, this corpus contains a higher number of general and abstract terms when compared to the whole corpus, indicating more abstract language produced by the MFA, which can potentially illustrate that the language is more abstracted and consists of concepts that are intangible. Thus, the image of Latvia in the MFA subcorpus is that of a nation engaged in regional and international cooperation with its allies, with a focus security and defence, and on relations between Ukraine and Russia.

4.6.3. Keyword analysis of Investment and Development Agency subcorpus

The Investment and Development Agency subcorpus consists of 70,910 tokens, therefore the minimum frequency threshold was determined as 35. The keyword list generated 190 keywords with a frequency higher than that. Table 4.5 presents the top 20 keywords of this subcorpus.

Table 4.5 **Top 20 keywords of the Investment and Development Agency subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	liaa	187	421	2,637.15	< 0.01	2,620.2
2	latvia	603	142,310	8,503.74	2.31	2,568.8
3	latvian	323	73,965	4,555.07	1.20	2,070.0
4	riga	96	72,183	1,353.83	1.17	623.8
5	elwind	39	65	549.99	< 0.01	550.4
6	sia	46	55,465	648.71	0.90	341.8
7	michelin	40	107,292	564.10	1.74	206.1
8	fintech	52	175,938	733.32	2.86	190.4
9	eur	82	330,375	1,156.40	5.36	181.9
10	baltic	56	220,674	789.73	3.58	172.5
11	norwegian	91	538,187	1,283.32	8.74	131.9
12	norway	101	721,265	1,424.34	11.71	112.1

13	entrepreneur	104	1,132,995	1,466.65	18.40	75.7
14	cooperation	150	1,715,762	2,115.36	27.86	73.3
15	startup	87	1,017,995	1,226.91	16.53	70.0
16	export	220	2,741,793	3,102.52	44.52	68.2
17	offshore	47	728,353	662.81	11.83	51.8
18	innovation	157	3,618,808	2,214.07	58.76	37.1
19	investment	274	6,845,641	3,864.05	111.16	34.5
20	netherlands	35	925,003	493.58	15.02	30.9

The top 20 keywords of this subcorpus reveal an entirely different range of keywords when compared to the whole corpus. Besides the words *Latvia, Latvian, Riga, Baltic, LIAA* and *cooperation*, the rest of the keywords are highly specific towards the discourse that the Investment and Development Agency produces, namely, content about investment and business in Latvia. New words, such as *Elwind, SIA, Michelin* refer to names of companies and awards, *Norwegian, Norway*, and *Netherlands* appear quite prominently as well and indicate the key business partners of Latvia, and, lastly, words that refer to money and commerce are highly key as well (*entrepreneur, startup, export, offshore, investment*).

The keywords were semantically tagged and yielded 17 categories. Figure 4.3 displays the semantic categories of the keywords.

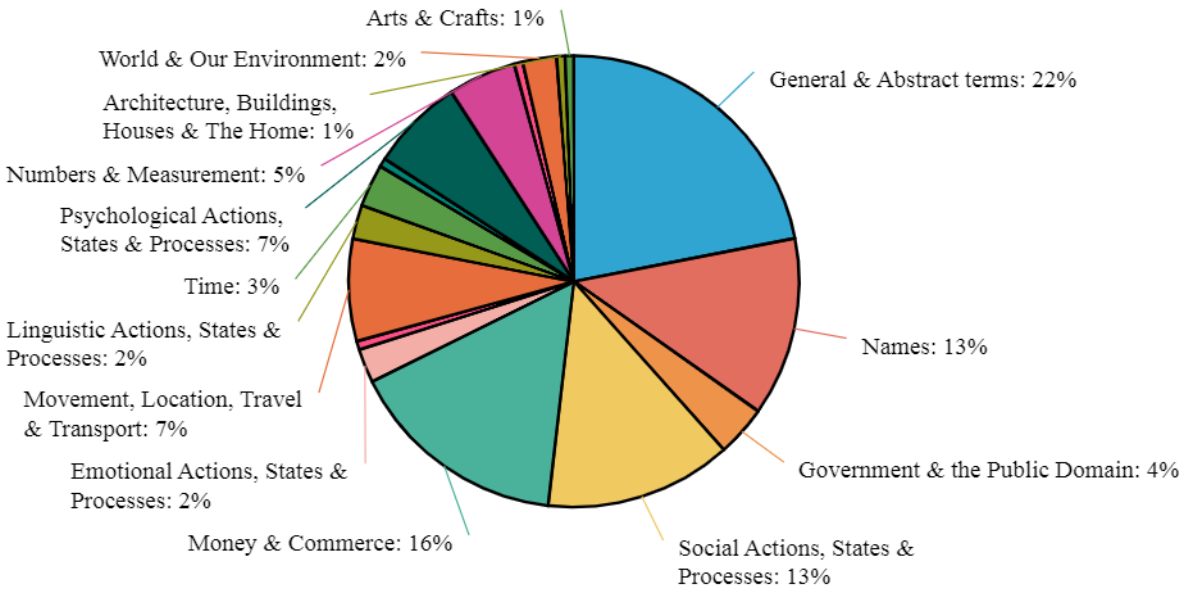

Figure 4.3 **Semantic categories of the Investment and Development Agency subcorpus**

The chart reveals that the semantic categories of this subcorpus are a more varied and evenly distributed than in the previous two subcorpora. The largest proportion of keywords this time was the general and abstract terms category, consisting of 22% of the keywords. The terms and subcategories include general actions (*develop, work, add, process, provide, offer, receive, include, use, get, find*), chance and luck (*opportunity, award*), affect (*contrast, different, increase, grow, aim, result, production, construction*), state of being (*content, just*),

evaluation (*good, great, quality, significant, important, value*), and comparing (*different, like, various*). The second large proportion is the money and commerce category (16%), consisting of keywords such as *entrepreneur, EUR, startup, export, offshore, investment, investor, import, market, product, company, industry, trade, sale, financial, fund, account*, among others. Social actions, states and processes category follows closely behind, with 13%. Subcategories include relationships (*cooperation, partner, relationship, share, group, member, part, team, conference*), helping (*guide, support, help, accessible, visit, service*), and power (*implement, attract, establish, lead*). The names category also contains 13% of the keywords. The names were those of countries and unions (*Latvia, Latvian, Baltic, Norwegian, Norway, Netherlands, Israel, Germany, Japan, Australia, UK, EU, European*), as well as other proper names of companies and locations (*Riga, Elwind, Fintech, Euro, Europe*). The rest of the categories include movement, location, travel and transport (7%) (*tourism, centre, south, region, park, international, step, local, place, area, come, go*), psychological actions, states and processes (7%) (*idea, interest, reach, expect, plan, take, point, see, system, think, know*), numbers and measurement (5%) (*size, million, total, growth, large, high, small, two*), time (3%) (*currently, year, new, start, time*), linguistic actions, states and processes (2%) (*say, story, note, information*), world and our environment (2%) (*global, green, environment, world*), and two newcomers; architecture, buildings, houses and the home (1%) (*build*), as well as arts and crafts (1%) (*culture*).

The keyword analysis of the Investment and Development Agency subcorpus presents an entirely different image of Latvia when compared to the previous subcorpus. The main topics of this subcorpus include business and finance, which is not surprising when the aim of the website is considered; its aim is to encourage investors to start businesses in Latvia, and for tourists to come and visit the country. Likewise, the list of countries mentioned in the discourse of the website allows to identify potential business partners of Latvia. Norway is the most prominent country name in the list, which implies that Latvia has significant business ties with the country and their investors. However, the mention of the Netherlands, Israel, Germany, Japan, Australia, and the UK also shows that Latvia has attracted investment from many of the Western countries, and that its potential aim is to attract even more potential investors from other similar countries. The abstract terms category implies positive aspects in terms of country image, namely, that Latvia is developing, that it provides opportunity, gets investment, is aiming to grow, expand and increase the business landscape of Latvia. The social category reiterates the notions that Latvia is a partner, that it cooperates and has partners, and is a country that should be worked with and invested in.

4.6.4. Keyword analysis of the Constitutional Court of Latvia subcorpus

The Constitutional Court of Latvia subcorpus consists of 59,017 tokens, therefore the minimum frequency threshold was determined as 30. The keyword list generated 183 keywords with a frequency list higher than 30. Table 4.6 presents the top 20 keywords of this subcorpus.

Table 4.6 **Top 20 keywords of the Constitutional Court of Latvia subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	334	142,310	5,659.39	2.31	1,709.7
2	contested	241	113,180	4,083.57	1.84	1,439.4
3	laviņš	66	5	1,118.32	< 0.01	1,119.2
4	saeima	67	1,993	1,135.27	0.03	1,100.6
5	aldis	66	6,691	1,118.32	0.11	1,009.6
6	constitutional	943	1,268,581	15,978.45	20.60	739.8
7	latvian	84	73,965	1,423.32	1.20	647.1
8	lithuania	34	201,362	576.11	3.27	135.2
9	norm	145	1,065,245	2,456.92	17.30	134.3
10	legislator	68	500,513	1,152.21	8.13	126.4
11	constitution	198	2,133,150	3,354.97	34.64	94.2
12	judgement	56	575,877	948.88	9.35	91.8
13	gambling	51	574,381	864.16	9.33	83.8
14	court	1,224	15,645,613	20,739.79	254.04	81.3
15	paragraph	117	1,645,854	1,982.48	26.72	71.5
16	applicant	142	2,098,941	2,406.09	34.08	68.6
17	republic	130	2,254,584	2,202.76	36.61	58.6
18	provision	193	3,466,406	3,270.24	56.29	57.1
19	initiate	81	1,509,821	1,372.49	24.52	53.8
20	proceeding	78	1,519,627	1,321.65	24.67	51.5

As seen in Table 4.6, the top 20 keywords of this subcorpus reveal that the discourse is heavily focused on court proceedings and law (*contested, constitutional, norm, legislator, constitution, judgement, court, paragraph, applicant, provision, proceeding)*. Likewise, names are among the most prominent keywords as well (*Latvia, Laviņš, Saeima, Aldis, Latvian, Lithuania*).

The keywords were semantically tagged and grouped according to the semantic categories. In total, 14 categories were identified. Figure 4.4 illustrates the semantic categories of this subcorpus.

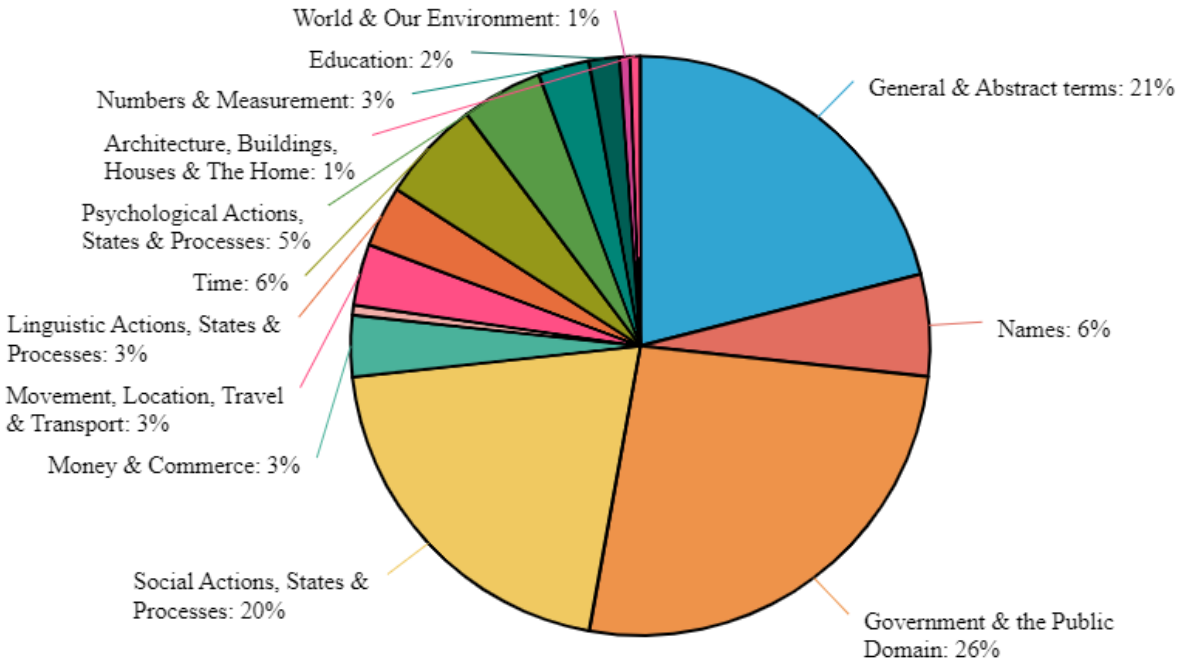

Figure 4.4 **Semantic categories of the Constitutional Court of Latvia subcorpus**

Figure 4.4 reinforces the previously stated observation that the primary focus of this subcorpus' discourse is on legal proceedings and law. The most prominent semantic category is the government and the public domain category (26%). The keywords assigned to this group include different legal terms related to the court (*constitution, constitutional, legislator, judicial, judge, court, legitimate, binding, legal, sentence, justice, regulation, law, procedure, framework, amendment, case, article, rule, application, applicant, application, right, fair, authority, force*), as well as more general terms related to governance (*republic, democracy, cabinet, democratic, president, citizen, official, council, commission, state, states, government, office, party*). General and abstract terms are ranked second, at 21%. The subcategories and keywords include general actions (*work, use, make, implement, hold, submit, execute, adopt, ensure, proceeding, prepare, provision, execution, establish, assess*), use and usefulness (*apply, value*), constraint (*restrict*), inclusion (*include),* affect (*develop, give, take, fact, basis, own*), evaluation (*importance, important, fundamental, fact*) and classification (*certain*). The category of social actions (20%) contains words about participation (*cooperation, participate, meeting, follow, role*), relationships (*society, member, public, social*), and helping (*provide, share, support, strengthen*), however the most abundant of the subcategories is the power relationship category (*initiate, rights, comply, obligation, restriction, accordance, compliance, permit, order, allow, power, govern*). These categories illustrate the judicial and social power that the court has, and it is reflected in the discourse as well, having more keywords that emphasise this power relationship. The names category is

less prominent in this subcorpus, tagged only to 6% of the keywords. The keywords assigned to this category include countries and unions (*Latvia, Latvian, Lithuania, European, Netherlands, Union, Europe*), names of people (*Laviņš, Aldis*), and the Latvian parliament (*Saeima*). The rest of the categories include time (6%) (*conclude, permanent, February, September, January, period, date, continue, year, time*), psychological actions, states and processes (5%) (*judgement, recognise, aim, decision, decide, account, accord, review*), money and commerce (3%) (*pension, income, fee, grant, tax, pay*), linguistic actions, states and processes (3%) (*paragraph, dialogue, language, request, discuss, discussion*), movement, location, travel and transport (3%) (*territory, international, land, port, national, local*), numbers and measurement (3%) (*high, three, part, two, new*), education (2%) (*conference, education, teacher, study)*, world and environment (1%) (*world*), architecture, buildings and houses (1%) (*property*).

The keyword analysis of this subcorpus reveals several predominant themes. The texts in this subcorpus focus on the court and the topics surrounding it, namely, that the discourse is centred around the constitutional court itself, the implementation of rule and law in Latvia, as well as dealing with pending cases in the context of Latvia and the European Union. The general and abstract terms include more law implementation actions activities, such as *work, implement, make, use*, but nothing that would indicate removal or deletion, indicating that the focus is on making sure that things are assessed, prepared, added, adopted, and executed. Social and general actions include both positives and negatives, which stands in contrast with previous examined data, where primarily positives were encountered. This implies the freedom, and the helping aspect of the court, whilst also emphasising rule and law, responsibility, and potential consequences. In terms of names, it can be concluded that the countries and unions that are mentioned refer to the allies and partners of the Latvian court. The primarily cooperation can be expected with the European Union, but interestingly, Lithuania and the Netherlands appear as well. One personal name is prominently featured in the keywords; Aldis Laviņš. He is the president of the Constitutional Court of Latvia, therefore it makes sense that he is mentioned more often than others. Thus, the use of names (or lack thereof) indicates that the discourse of the court is less direct in terms of social actors, emphasising the court itself rather than someone named explicitly because the court itself holds power and acts as the law. Thus, the image of Latvia is that of a country that upholds the law, ensures fair legal processes, and focuses on building regional partnerships.

4.6.5. Keyword analysis of the missionLatvia subcorpus

The missionLatvia subcorpus consists of 18,406 tokens, being the smallest of the five subcorpora. The minimum frequency threshold was set at 10, providing 50 keywords that had more than the minimum frequency hits. Table 4.7 lists the top 20 keywords of this subcorpus.

Table 4.7 **Top 20 keywords of missionLatvia subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvian	166	73,965	7,743.62	1.20	3,518.7
2	latvia	228	142,310	10,635.82	2.31	3,212.8
3	latvians	49	7,265	2,285.77	0.12	2,045.5
4	riga	76	72,183	3,545.27	1.17	1,632.7
5	spirulina	22	21,602	1,026.26	0.35	760.5
6	kokle	16	223	746.37	< 0.01	744.7
7	printful	16	773	746.37	0.01	738.1
8	rainis	14	421	653.08	< 0.01	649.6
9	betriton	11	0	513.13	0.00	514.1
10	biomedicine	15	23,489	699.72	0.38	507.3
11	baltics	12	12,437	559.78	0.20	466.6
12	nouveau	12	68,697	559.78	1.12	265.1
13	baltic	17	220,674	793.02	3.58	173.2
14	eur	19	330,375	886.32	5.36	139.4
15	startup	52	1,017,995	2,425.71	16.53	138.4
16	bog	12	189,743	559.78	3.08	137.4
17	unicorn	12	196,140	559.78	3.18	134.0
18	ict	16	315,156	746.37	5.12	122.2
19	choir	19	675,504	886.32	10.97	74.1
20	easter	15	715,847	699.72	11.62	55.5

Table 4.7 shows that the top 20 keywords are quite diverse in their content, focusing on Latvia, names of companies (*Printful, Betriton, Spirulina*), mentions of architecture and culture (*nouveau, kokle, choir*), and mentions of business and innovation (*biomedicine, startup, ICT*).

The keywords were also semantically tagged and grouped according to the semantic categories established. In total, 13 categories were identified. Figure 4.5 displays the semantic categories.

The findings show that for this subcorpus, the most prominent category is the names category (26%), which implies that the discourse within this subcorpus focuses on topics that include specific persons, locations, and businesses, most probably in the context of Latvia. The keywords that fall under this category include geographical names (*Latvian, Latvia, Latvians, Riga, Baltics, Baltic*), names of companies (Spirulina, *Printful, Betriton*), and people (*Rainis*). The second on the list is the social actions, states and processes category

(11%), which presents keywords that prominently fall into traditions subcategory (*Easter, tradition, celebration, heritage*), and positive traits (*unicorn, smart*). The general and abstract terms category follows next, featuring keywords such as *showcase, diverse, blend, symbol,* and *exhibition*.

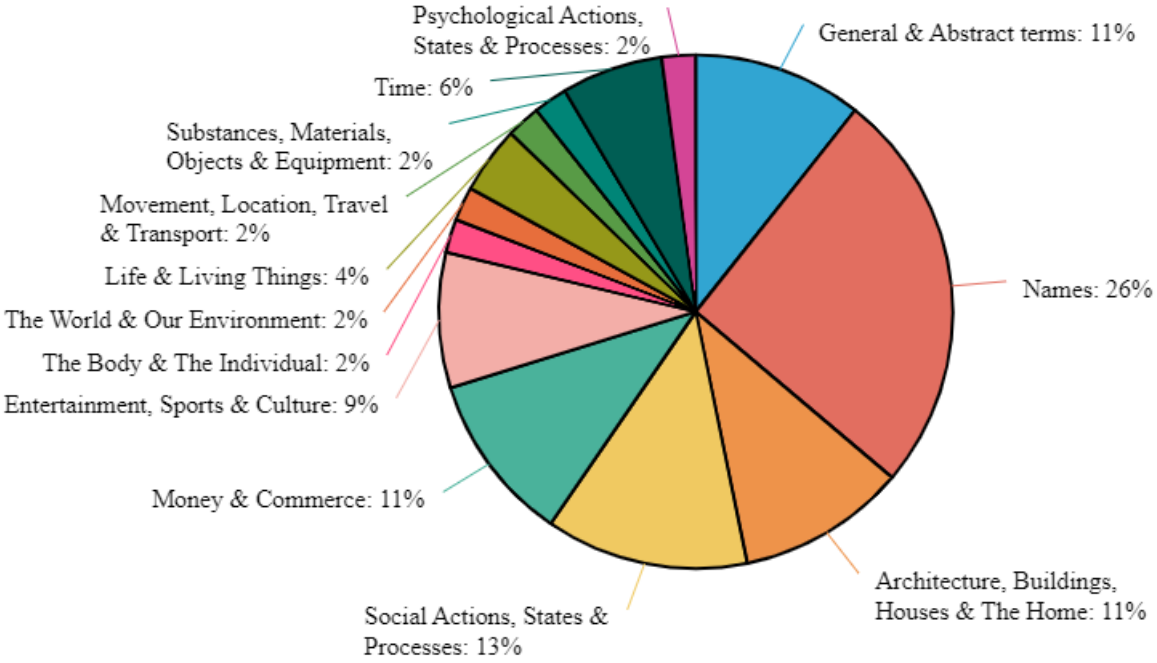

Figure 4.5 **Semantic categories of missionLatvia subcorpus**

The money and commerce category (11%) yields keywords that refer to investment (*startup, EUR, innovation, invest, innovative*). Other categories include entertainment, sports and culture (9%) (*choir, opera, cultural, folk*), time (6%) (*contemporary, ancient, winter*), life and living things (4%) (*mushroom, egg*), with the rest making up a very small percentage.

The keyword analysis of this subcorpus reveals several curiosities. Initially, the assumption or prediction was that the discourse would be centred towards describing and representing Latvia itself, namely, its facts, achievements, people, traditions, and so on. Plenty of the keywords also reflect this aspect (*talent, architecture, cultural, folk, tradition, heritage, kokle*). However, the majority of the keywords and semantic categories represent something else. The keywords contain heavy emphasis on businesses and investment, as can be seen by the keywords *Spirulina, Printful, Betriton, startup, innovation, invest, innovative,* and *biomedicine* and their high keyness scores. These reflect the heavy emphasis on presenting business achievements and business fields where potential investors could invest their capital in. Latvia is represented as being something new and innovative, for example, in biomedicine and startups. Likewise, emphasis on popular companies helps to present Latvia as a success story and allows Latvia to sell these positives to potential investors. The focus of this discourse can be attributed to the Investment and Development Agency, as the website is also

managed by them, indicating that this might influence the direction that this website takes when presenting Latvia.

4.6.6. Keyword analysis of the Ministry of Defence subcorpus

The Ministry of Defence subcorpus consists of 47,670 tokens, therefore the minimum frequency threshold was set at 24. The keyword list generated 230 keywords that matched this threshold. Table 4.8 shows the top 20 keywords of this subcorpus.

Table 4.8 **Top 20 keywords of Ministry of Defence subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	489	142,310	10,258.02	2.31	3,098.7
2	mūrniece	82	22	1,720.16	< 0.01	1,720.5
3	latvian	136	73,965	2,852.95	1.20	1,296.7
4	ināra	43	226	902.03	< 0.01	899.7
5	sprūds	37	14	776.17	< 0.01	777.0
6	naf	49	20,739	1,027.90	0.34	769.7
7	defence	938	1,686,029	19,676.95	27.38	693.5
8	saeima	32	1,993	671.28	0.03	651.2
9	skraučs	27	0	566.39	0.00	567.4
10	andris	27	5,805	566.39	0.09	518.5
11	baltic	99	220,674	2,076.78	3.58	453.3
12	nato	252	665,102	5,286.34	10.80	448.1
13	forces	150	481,386	3,146.63	7.82	357.0
14	ukraine	227	994,839	4,761.90	16.15	277.7
15	riga	28	72,183	587.37	1.17	270.9
16	cybersecurity	84	412,983	1,762.11	6.71	228.8
17	deterrence	25	83,029	524.44	1.35	223.8
18	procurement	101	661,834	2,118.73	10.75	180.5
19	armed	178	1,299,441	3,734.00	21.10	169.0
20	estonia	28	158,220	587.37	2.57	164.9

Table 4.8 indicates that top keywords mostly refer to the defence and security subcategory, as some of the most prominent keywords include *defence, forces, Ukraine, cybersecurity, deterrence, procurement,* and *armed*. Likewise, names are also highly key to the discourse in this subcorpus (*Latvia, Mūrniece, Latvian, Ināra, Sprūds, NAF, Saeima, Baltic, NATO, Riga, Estonia*), referring to the ministers that are leading the ministry, governmental institutions of Latvia, as well as alliances and other countries.

The keywords were also semantically tagged and grouped according to the semantic categories established earlier. In total, 11 categories were identified. Figure 4.6 shows the semantic categories and their percentages.

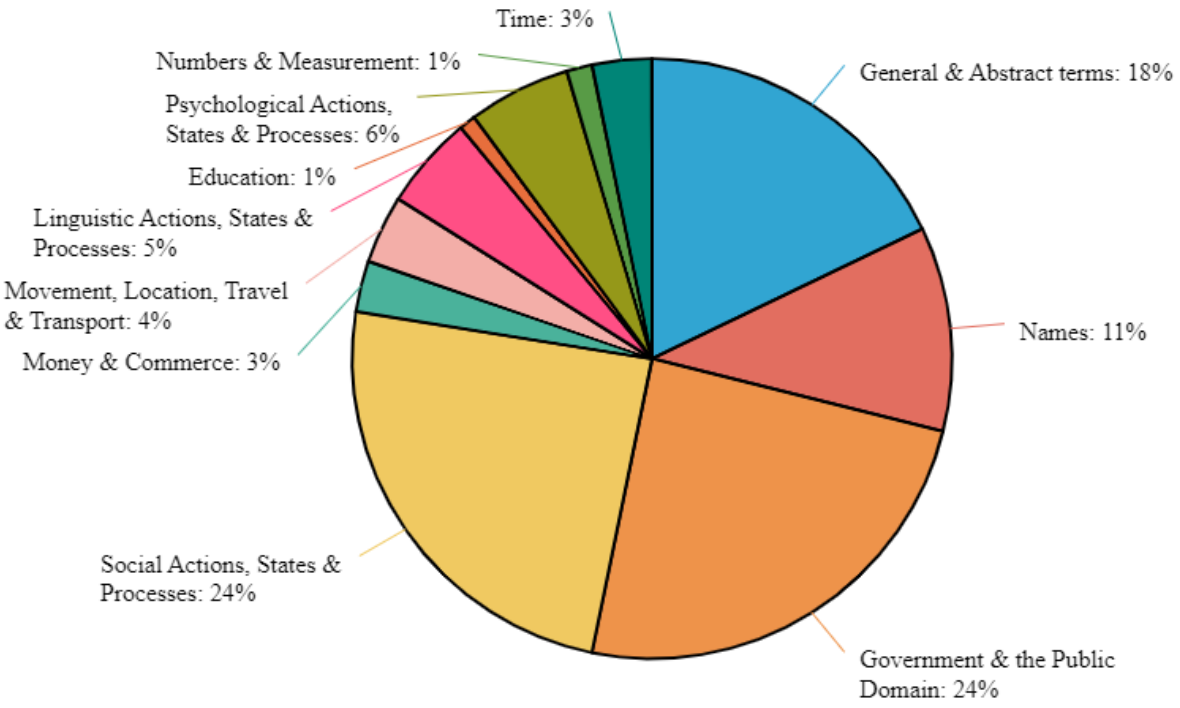

Figure 4.6 **Semantic categories of the Ministry of Defence subcorpus**

Figure 4.6 shows that the four major categories are government and the public domain (24%), social actions, states and processes (24%), general and abstract terms (18%), as well as names (11%). The rest of the categories appear significantly less.

The government and the public domain category reveals subcategories and keywords of security and defence (*defence, forces, cybersecurity, armed, ammunition, drone, artillery, military, brigade, invasion, battalion, missile, troop, guard, combat, strategic, security, defend, soldier, air, equipment, personnel, border, draft, volunteer, vehicle, unit, army, war, company, operation*), governmental institutions and roles (*ministers, ministry, minister, cabinet, institution, sector, centre, department, industry, agency, government, committee, union, programme, management*), as well as law and order (*draft, force, policy, law*).

The social actions category reveals subcategories and keywords of helping (*protection, support, promote, join, improve, agree, provide, help, contact*), hindering (*deterrence, resilience*), participation (*cooperation, coalition, collective, joint, meeting, agreement, join, agree, member, meet, group, part, community*) , relationships (*relations, allied, ally, alliance, partner, united*), and power (*capability, approve, command, announce, provide, strong, require, order, control, allow, take, use*).

The general and abstract category reveals subcategories and keywords of general actions (*affairs, implementation, action, make*), affect (*development, develop, contribute*), being (*case, become*), classification (*various, main*), getting and giving (*procurement, supply*), importance (*priority, important, critical*), danger (*threat, crisis*), among others.

The names category reveals names of countries and alliances related to the defensive interests of Latvia (*Latvia, Latvian, NAF, Baltic, NATO, Estonia, MOD, EU, European, Canada, Europe*), names related to the war in Ukraine (*Ukraine, Ukrainian, Russia, Russian*), as well as personal names (*Mūrniece, Ināra, Sprūds, Skraučs, Andris, Roberts*).

The rest of the categories include psychological actions, states and processes (6%) (*framework, aim, concept, information, system, plan, situation, establish, strategy, skill, focus, work*), linguistic actions, states and processes (5%) (*media, communication, phone, discuss, response, address, decision, report, say*), locations (*national, regional, international, region, base, area, local, land*), money and commerce (3%) (*funding, purchase, investment, million, resource, product*), and time (3%) (*continue, current, year, start, new, time*).

The analysis of the keywords shows that heavy emphasis is placed on Latvia's national security and defence, particularly in the context of the war in Ukraine. The MoD discourse focuses on military operations, acquisition of equipment, as well as the planning of strategic defences. The social actions category complements this theme by emphasising the improvement of their military capabilities, as well as hindrance of threats in case of attack. Alliances and partnerships are also a crucial component of MoD's discourse, focusing on continued cooperation and partnership with their military and political allies (such as NATO and the EU), and other strategic partners of neighbouring countries. Many of the keywords reflect this idea, which shows that Latvia seeks to continue their cooperation with their allies and take steps to improve the relationships as well. The mention of governmental institutions and roles helps to solidify these relationships, seek improvements, and provide solutions for security threats, whilst the keywords within the names category help to illustrate the main participants involved in the security landscape of Latvia.

4.6.7. Discussion of results

The keyword analysis was performed by analysing the Latvian Governmental Web Texts corpus first, and then the same procedure was applied for each individual subcorpora (5 in total). The results of the keyword analysis reveal the most prominent semantic categories in each respective subcorpora.

The keyword analysis of the subcorpora revealed a similar pattern of top keywords. For the most part, they were geographical names of Latvia and other countries, the personal names of public figures, governmental institutions, and names of the institutions from which the texts were gathered. This pattern indicates the possibility that the keyness score is higher for words that are Latvian and therefore are less likely to feature in the reference corpus (such as the names of Latvian companies, Latvian governmental institutions (such as *Saeima*) and

Latvian personal names). Likewise, names of countries that are not as prominent in the texts of the reference corpus might appear higher and receive a higher keyness score. Besides this curiosity, the top keywords identified differed across subcorpora according to their prominence in those particular texts.

The keyword analysis of the Latvian Governmental Web Texts corpus and its subcorpora also revealed the main semantic categories in the corpus that form the image of Latvia. The total number of categories in each subcorpus varied, but almost all of the subcorpora had a similar distribution of the most prominent semantic categories: government and the public domain, general and abstract terms, social actions, states and process, and names. These semantic categories indicate the overarching focus of the content that the Latvian governmental websites produce. One being the government and their involvement in political and diplomatic events, the execution of law and order, as well as the discussion of security and defence matters (government and the public domain category) and the inclusion of social participants (geographic locations, proper names, and personal names) that shape the events revolving around Latvia, another being the use of words and expressions that can be used across various subjects, and terms that relate to abstract ideas and concepts (general and abstract terms category), and another referring to the social interactions between participants (social actions, states and processes category). Some subcorpora, however, had different main categories, in particular, the Investment and Development Agency subcorpus and the missionLatvia subcorpus. These subcorpora had a more evenly distributed list of semantic categories, with the money and commerce category appearing more prominently, which indicate the different points of focus that these discourses have. However, the three main semantic categories were still quite prominent, therefore it can be assumed that the prominence of the aforementioned categories is more noticeable when compared to the rest of the semantic categories.

The findings also revealed that the discourse mostly features more positive attributes and actions than negative ones. Only when the enactment of law and security is discussed do the actions and attributions become more negative, indicating that the discourse around Latvia focuses on establishing good relations with others, whilst also protecting their own interests in terms of national security and defence, as well as the rule of law.

In terms of the semantic tags themselves, the findings showed that the selected tagset reveals general categories; this can be addressed, but the findings must be interpreted and dissected by taking context into account.

It is also important to mention that several of the keywords overlapped in terms of their categories, as the semantic tagger assigned multiple semantic tags to many of the

keywords. For example, many of the words assigned to the social actions category also were tagged in the general and abstract terms category (*implementation, relation, meet, hold*, etcetera). To have numerically accurate results, only one category was assigned to each word, therefore many of the keywords that could have appeared in multiple categories appeared only in one. This illustrates the complex nature of keywords and their semantic categories, and highlights that the tools available often provide multiple categories. At the same time, it should be highlighted that the semantic tagging nevertheless provides a good overview of the main categories of discourses about Latvia and allows to establish the main themes of the content.

To sum up, the keyword analysis and the semantic tagging of keywords allowed to identify the main semantic categories of the corpus. However, the findings indicated that the categories are somewhat general, and did not reveal how the keywords are used in context. To solve this issue, concordance analysis of the keywords that have the highest keyness score is carried out; firstly, to compare whether the established categories match, but secondly, to specify the themes established, and see how words are used in context. The next chapter describes the concordance analysis of the subcorpora.

4.7. Concordance analysis of the keywords

The following subchapter performs the qualitative analysis of the study, examining the keywords in context. The concordance analysis was carried out using the Sketch Engine software, taking the keywords and viewing them in the KWIC mode. As analysing concordances of every keyword is difficult due to size and time constraints, five keywords from each subcorpus were examined and grouped according to themes that can be identified.

4.7.1. Concordance analysis of the MFA subcorpus

The concordance analysis of the MFA subcorpus was performed for the keywords *Latvia, Latvian, Baltic, aggression,* and *cooperation*. The results of the analysis are described below.

For the keyword *LATVIA*, a wide range of different usage appears. For one, the actions that Latvia undertakes can also be identified when examining the keywords. Examples include *National action plan of Latvia*; *Latvia has joined an informal OSCE Group of Friends of Environment*; *Latvia participates in NATO operations*; and *Latvia's diplomatic missions*. Latvia's support for Ukraine is highlighted through concordances such as *Latvia's current support to Ukraine*; *Latvia commemorates the Ukrainian people*; *Latvia support Ukraine*; *Latvia is happy to support*; *Latvia's military support to Ukraine*; *the state and people of Latvia are steadfastly supporting Ukraine and will continue to provide every kind of*

assistance. These show the role that Latvia plays in Latvian-Ukrainian relations, namely, that it offers mental, humanitarian, and military support for Ukraine. Latvia and their partners are highlighted through concordances *Latvia relies on Mexico's support*; *Latvia and Kazakhstan*; *Latvia and Australia*; *Latvia and the Netherlands*; *Latvia and Lithuania*; *Latvia and Mongolia*; *Latvia and Ethiopia*; *Latvia and Japan*; and *Latvia, together with its foreign partners*. These indicate that Latvia has a wide range of countries that they have diplomatic relations with, and that they are continuing to expand by building relations with more countries (such as countries from Africa). Another interesting topic is in regard to Latvia and the UNSC (United Nations Security Council). The concordances examined reveal Latvia's plans to run for a seat in the council (*For the first time, Latvia will take part in the UNSC election in June 2025*; *Latvia's candidacy for the elected seat of the UN Security Council in 2026-2027*; *Latvia's UNSC campaign*). These show the aspirations that Latvia has on the global stage, namely, that it wants to gain higher political positions in alliances to further their political goals, such as security. Lastly, the relations between Latvia and Russia are revealed through the concordances *the following categories of Russian citizens will be permitted to enter Latvia*; *Period of restrictions on the entry of Russian citizens into Latvia*; *The Russian diplomat must leave Latvia by the end of the day this coming 10 April*. These reveal the worsening relations between the two countries, having laws and diplomatic actions that affect the relations negatively.

For the keyword *LATVIAN*, the names of more different organisations and entities of Latvia appear when the keyword is examined. Examples include *Republic of Latvia*, *Latvian Transatlantic Organisation (LATO)*; *Latvian Cultural Alliance*; *Latvian National Guard*; *Latvian Foreign Service*; *Latvian non-governmental organisations (NGOs)*; *Latvian Platform for Development Cooperation (LAPAS)*. The extensive use of different names indicates the organisations that help to further Latvia's national and foreign objectives on the international stage. Likewise, the elements of Latvia's foreign affairs are prominent, reflecting the different roles and participants that are related to Latvia's foreign relations (*Latvian foreign minister*; *Latvian diplomats*; *Latvian presidency*; *Latvian nationals*; *Latvian officials*; *Latvian delegation*; *Latvian civilian experts*; *Latvian businessmen*; *Latvian businesses*). These are also linked to the furthering of Latvia's interests.

The keyword *BALTIC* and its concordances reveal topics that focus on Latvia's security, and the NB8 alliance. Examples include (*EU Strategy for the Baltic Sea Region*; *strengthening of security in the Baltic States*; *the strengthening of security of Latvia and the Baltic Sea region*; *Cooperation between the Baltic and Nordic countries*; *The NB8 (Nordic–Baltic Eight) partnership*; *Baltic and Nordic Directors*; *Nordic and Baltic countries*; *The*

Baltic Council). The constant references to the events that involve the NB8 partnership is significant, indicating that Latvia places great effort into working with Nordic countries and considers it an important alliance, particularly for reasons of security.

The keyword *AGGRESSION* is exclusively used when referencing to the war in Ukraine. Examples of the concordances include (*Russian aggression against Ukraine*; *Russia's aggression*; *War of aggression*; *Crime of aggression*; *Russia's military aggression against Ukraine*; *Russian aggression on European economy*; *Aggression unleashed by Russia on Ukraine*; *Russia's aggression and atrocities*). The volume and the consistency of the same statements indicate a clear position that Latvia takes when Russia is concerned; it is positioned as the instigator and the aggressor in the conflict, therefore the evaluations of their activities are negative.

The keyword *UNDERLINE* is primarily used when specific politicians and institutions are expressing their political views. Topics include the Russian involvement in the war (*LATO underlines that the Russian military aggression was not launched just a year ago*; *The Parliamentary Secretary of the Foreign Ministry underlines Russia's accountability for casualties*; *Reinis Brusbārdis underlined Russia's accountability for killing Ukrainian people*; *He underlined that not only was the Russian aggression against Ukraine a threat*; *In general, to establish, highlight and further the established relationships that Latvia has*; *The Foreign Minister underlined the need to counter the disinformation spread by Russia*; *The Latvian Foreign Minister underlined the importance of setting up an international tribunal to try Russia's crimes*), cooperation between countries and alliances (*Željko Komšić underlined his interest in continuing cooperation on transatlantic security matters*; *the Ambassador underlined the friendly and close bilateral relations between the two countries*; *Edgars Rinkēvičs underlined Latvia's commitment to intensify dialogue and deepen cooperation with African countries*), and support of Ukraine (*Foreign Minister Krišjānis Kariņš underlines long-term support for Ukraine*; *Foreign Minister Kariņš underlined the need to agree on a sustainable financing solution for assistance to Ukraine*; *Andris Pelšs, who underlined that assistance to Ukraine, especially military and financial, must continue*; *Edgars Rinkēvičs underlines Latvia's unwavering support to Ukraine in the war against Russia*; *Edgars Rinkēvičs underlined the need to do whatever it takes to provide military support to Ukraine*).

The concordance analysis of the MFA subcorpus shows that the themes match with the ones established through keyword analysis. The discourse mainly focuses on the war in Ukraine and how that affects Latvia's own defence and security interests, but other topics, such as strengthening the existing relations with Latvia's partners, developing new alliances with other countries and regions, are prominent. Lastly, the discourse also emphasises

Latvia's continued interest in allying themselves with the Nordic countries, which was not observed in the initial analysis.

4.7.2. Concordance analysis of the Investment and Development Agency subcorpus

The concordance analysis of the Investment and Development Agency subcorpus was performed for the keywords *LIAA, Latvia, Latvian, Norway,* and *entrepreneur*. The results of the analysis are described below.

The keyword *LIAA* is used when emphasising its role in advancing investment in Latvia. Examples include *together with LIAA, which plays an important role in attracting foreign investment*; *In 2022, with the support of LIAA, 40 investment projects were implemented*; *In 2023 LIAA implemented a total of 46 investment projects*; *That the company is welcome and will receive the necessary support, not only from LIAA but also from the municipality*; *A number of important projects have been identified with the assistance of LIAA's economic representations*; *In total, LIAA has economic representations in 20 countries*; *The LIAA strategy aims to attract € 2 billion of investment between 2023 and 2029*; *In 2024, LIAA's target for attracting foreign investment is € 660 million*; *With LIAA supporting a variety of businesses and initiatives, including those related to startups*. These show the significant role that the Investment and Development Agency has in facilitating investment, showing that more and more investment projects are taking place in Latvia, Latvia has a bigger representation in other countries, and that it aims to attract even more earnings in the following years.

The keyword *LATVIA* primarily reveals discourses that focus on the positives of Latvia, showing it as being attractive for potential investors. Examples include *Latvia's economic transformation*; *So that entrepreneurs find Latvia attractive as a country to invest in high added value manufacturing, innovation and human capital*; *Foreign companies are still eager to expand into Latvia*; *The sector is developing and this is also influencing investors' interest in Latvia*; *Latvia's energy independence and enabling it to become an electricity exporter*; *Latvia's goal in participating in EXPO 2025 Osaka is to increase exports to Japan*; *Latvia's position for the development of biotechnology is quite strong*; *Fibenol is looking to set up its production facility in Latvia, thus appreciating our business environment and the opportunities for government support*; *Latvia's solutions for smarter cities*; *Latvia, on the other hand, has more experience with digitalisation*. These examples show that Latvia is positioned as having a good and successful business environment, having experience and knowledge in different fields, therefore the businesses and sectors are finding investors and are thus considered successful.

The keyword *LATVIAN* is used when referring to different areas of business, such as *Latvian products*; *Latvian food products*; *Latvian exports*; *Latvian imports*; *Latvian entrepreneurs*; *Latvian researchers*; *Latvian startup*; *Latvian consistency*; *Latvian solutions*. The concordances show that there are different components to Latvia's business model and that they are being used to promote Latvia, its import and export capabilities, as well as the products that they produce. As with the other keywords, the discourse is focused only on the positive aspects, which can be attributed towards the fact that Latvia's primary goal is to gain more investment and present its business landscape positively.

The keyword *NORWAY* illustrates the close relations that the two countries have, as well as Latvia's dependency on Norway and its investment. Examples include *Norway is among the 10 largest investors in Latvia*; *Norway and Germany were the leading countries in the number of investment projects attracted to Latvia*; *The cooperation between Latvia and Norway has become much closer and more diverse*; *The total amount of exports to Norway was estimated at around 631.9 million euros*; *Economic ties between Latvia and Norway are not insignificant and growing stronger*; *Latvia and Norway already have successful space cooperation*. These show that Latvia greatly values the economic and political ties that they have with Norway, and that their relations are growing stronger in the recent years. Interestingly, Norway is also mentioned in regard to the topic of war in Ukraine, namely, the line *Norway has welcomed approximately 70 000 Ukrainian refugees*, indicating its role in the support of Ukraine.

Lastly, the keyword *ENTREPRENEUR* is used in the context of potential investment. Examples of usage include lines such as *The Large Investments Programme offers entrepreneurs the opportunity to implement large investment projects*; *Entrepreneurs who plan to invest at least EUR 10 million in the start-up*; *Scientists and entrepreneurs that turn into successful start-ups*; *We again see passionate entrepreneurs whose business ideas can reach unprecedented heights*; *We are committed to continuing to provide our entrepreneurs with the support*; *Many entrepreneurs have already expressed great interest in the opportunity*. These show that entrepreneurs are also mentioned positively; they are successful, passionate, interested in the potential to invest. Likewise, entrepreneurs are relied upon to invest into Latvia, and the discourses show that great hope is placed in them choosing Latvia for investment.

The concordance analysis of the Investment and Development Agency subcorpus reveals that their discourse focuses on highlighting the established business relations that Latvia has, as well as their ambitions for further investment into Latvia. Latvia in the business sector is described positively; Latvia has developed energy, biotechnology, smart city and

digitalisation sectors, whilst new investors are described as being interested in building their businesses in Latvia. The context around the word *LATVIAN* reveals that Latvian products and knowledge is described positively, whilst the context around the word *ENTREPRENEUR* emphasises that Latvia is ready to welcome investors who contribute financially, providing them with support. Lastly, the relations between Latvia and Norway are prevalent in the Investment Agency's discourse; their relations are continuous and very close. Likewise, Norway is described as being one of Latvia's main financial contributors, opening businesses and importing Latvian products, highlighting the importance that Norway has in relation to Latvian business and economic development. Lastly, it is important to emphasise the fact that once more, the discourses are structured to only show positive information, which indicates that the discourses serve as a means to attract more investment and might not be completely reliable in terms of informational value.

4.7.3. Concordance analysis of the Constitutional Court of Latvia subcorpus

The concordance analysis of the Constitutional Court of Latvia subcorpus was performed for the keywords *Latvia, Constitution, Latvian, Lithuania,* and *gambling*. The results of the analysis are described below.

The analysis of the keyword *LATVIA* reveals multiple contexts that can be separated from one another. Firstly, the title of the court appears prominently (*The Constitutional Court of Latvia*), which is logical, as much of the content is centred around the court itself. Latvia as a member of the international community appear in the following examples (*Constitutional Court of Latvia as a member of the Bureau of the World Conference*; *Latvia's Presidency of the Committee of Ministers of the Council of Europe*). These refer to the fact that Latvia participates in conferences, which can help to improve their own execution of law and exchange their own experiences with other governments. Likewise, the second concordance line shows that Latvia has taken over presidency in a European organisation, which indicates that Latvia is respected on the international stage and is given responsibility within the lawmaking process in Europe. The context around the people of Latvia was also identified (*people of Latvia*; *non-citizen of Latvia*), which refers to the complex system of citizenship that Latvia has, namely, that it is divided between citizens and non-citizens. Latvia and the context around its history can also be observed in numerous passages, such as *Latvia's historical experience and the current geopolitical context*; *20th anniversary since Latvia became a Member State of the European Union*; *When the independent and democratic State of Latvia was restored*. These show that the Constitutional Court highlight their historical turning points in terms of independence, rule of law, and the European political course that

Latvia has undertaken. Lastly, the keyword appears in the context of the Latvian Constitution (*Reinforcing democracy and the rule of law in Latvia*; *Court, being the guardian of the Satversme, defends the fundamental values and freedoms of all individuals in Latvia*). These portray the Constitutional Court as the guardians of Constitution, therefore ensuring that law and order in Latvia is maintained.

The concordance analysis of the keyword *CONSTITUTION* reinforces the same idea; that the Court ensures compliance with Latvia's Constitution, the interpretation of what the Constitution means, and the protection of the guarantees that the Constitution offers. Examples include *Comply with the Constitution*; *Safeguard the Constitution*; *Must comply with the provisions of the Constitution*; *Guaranteed by the Constitution*; *Compatible with the Constitution*; *Being incompatible with the Constitution*; *Does not comply with the Constitution*.

The keyword *LATVIAN* reveals prevalent topics on the language situation in Latvia (*Latvian state language*; *Latvian and minority languages*; *The Latvian state has a duty to develop and protect the only official language, Latvian*; *A significant proportion of Latvian nationals do not know Latvian at a sufficient level to be able to fully integrate themselves into the society*), Latvian citizenship (*had been a Latvian citizen or a non-citizen of Latvia*; *Revocation of Latvian citizenship*; *Latvian citizenship*), and generally, members of the state (*Latvian society*; *Latvian nationals*). The context examined reveals that the issue of state and minority languages, and legal conflicts regarding Latvian citizenship are brought to the attention of the Court.

The keyword *GAMBLING* appears in the contexts of the gambling ban in Latvia (*prohibited to organise gambling and to provide the corresponding gambling services in casinos, gambling and bingo parlours, as well as betting shops across the entire administrative territory of Riga*), and conflict between the government and a local municipality regarding the ban of gambling (*Suspension of the ban on gambling by the Ķekava Municipality local government complies with the regulatory framework*). This shows that Latvian policy regarding restriction of gambling has increased.

Lastly, the keyword *LITHUANIA* and its context show the close relations that the two courts have (*long tradition of bilateral cooperation between the Constitutional Courts of Latvia and Lithuania in strengthening the rule of law*; *the cooperation between the constitutional courts of Latvia and Lithuania started in 1996*; *The lawyers of the Constitutional Courts of Lithuania and Latvia shared their experience*). These show that the two courts have a very close relationship with one another, therefore it can be concluded that Latvia's main partner or ally in the context of the Constitution is Lithuania.

To sum up, the concordance analysis of this subcorpus reveals the themes of participation in international legal communities, the safeguarding of Latvian constitutional and democratic values, the close bond that the legal communities of Latvia and Lithuania have, and several court proceedings related to Latvian language, Latvian citizenship, and gambling.

4.7.4. Concordance analysis of the missionLatvia subcorpus

The concordance analysis of the missionLatvia subcorpus was performed for the keywords *LATVIAN, LATVIA, LATVIANS, RIGA,* and *STARTUP*. The findings of the analysis are described below.

The analysis of the word *LATVIAN* reveals contexts of the Latvian landscape that the discourse aims to portray. One group can be defined as being connected to the business and academic sectors (*Latvian startup ecosystem*; *Latvian and Baltic startups*; *Latvian companies, Latvian universities*; *Latvian students*; *Latvian scientists*; *Latvian researchers*; *Latvian immunologist*), which highlights the positives that Latvia offers. Another group includes the following lines: *Latvian art*; *Latvian nation*; *Latvian identity*; *Latvian folk songs*; *Latvian hockey fans*; *Latvian language*; *Latvian culture and traditions*; *Latvian artists*. These cover the cultural aspects of the Latvians, emphasising that the Latvian people are often associated for their cultural presence.

The analysis of the word *LATVIA* reveals that it promotes different aspects of Latvia, ranging from Latvia having ancient traditions, having an established international filmmaking scene, and having a rich history. However, having examined the word in context, it can be observed that many passages also refer to Latvia and startups (*Latvia was named the most startup-friendly country*; *Latvia is the ideal place for startups*; *Ten promising startups from Latvia to keep an eye on*; *There are currently over 700 startups in Latvia*). These show that even though the website is used to promote Latvia and present its values and increase its international profile, the content is still tailored towards investors, indicating that one of the main goals of this website is to seek investors as well.

The analysis of the word *LATVIANS* shows more general and descriptive attributes of the people of Latvia. Examples include *Latvians have been a self-sufficient agricultural nation*; *Latvians have conquered the global academic music scene*; *92% of Latvians attend cultural events*; *Latvians have a strong relationship with their cultural heritage and traditions*; *Latvians still retain an utmost respect for nature*; *Latvians highly value that which has come from nature*. These construct an identity for Latvians, namely, their embeddedness within culture and nature, and their valuation and respect for agriculture and self-sufficiency.

The analysis of the word *RIGA* shows that the capital is mostly mentioned when it comes to architecture and landscapes (*Arena Riga*; *Riga Castle*; *Riga Central Market*; *Art Noveau buildings in Riga*; *Urban delights of Riga's Old Town*; *Riga, known for its rich history*; *Riga, a UNESCO World Heritage site*). These lines show that Riga is often used to promote landmarks, and, in turn, also promote tourism and Latvia's cultural heritage.

The word *STARTUP* continues to highlight Latvia's established course for attracting investment through startups by providing positive evaluations (*Local startup community*; *Startup visas for foreign founders*; *Close-knit startup community*; *Our international startup events*).

The concordance analysis of the missionLatvia subcorpus shows that the discourse presents the Latvian values, traditions and landmarks. However, it is also heavily focused on the topic of investment, which indicates that the focus of this page is not to simply present the country and its people, but to attract investment. Many of the facts and people described are related to business, science, and research, however, cultural aspects are also covered extensively.

4.7.5. Concordance analysis of the Ministry of Defence subcorpus

The concordance analysis for this subcorpus was performed for the keywords *LATVIA, LATVIAN, DEFENCE, NATO,* and *UKRAINE*. The findings are described below.

The keyword *LATVIA* reveals various contexts in which it is used. For one, there is the reference to Latvia's historical military events, as can be seen in the examples *Latvian soldiers who died fighting for Latvia's national independence and freedom*; *Commemorating the War of Independence all over Latvia*; *Lāčplēsis Day marks the victory of independent Latvia's armed forces*. Secondly, the improvement of Latvia's military capabilities is highlighted as well (*Latvia intends to spend at least 10 million euro over the next 12 months*; *The first two UH-60M Black Hawk helicopters delivered to Latvia*; *Latvia develop rocket artillery, command and control, and drone capabilities*; *Latvia joined NATO's Virtual Cyber Incident Support Capability*). Discourses around support for Ukraine is consistent as well (*Latvia continues to provide military support to Ukrainian Armed Forces*; *Latvia has committed funding in excess of 1% of its GDP to help Ukraine*; *Latvia among front-runners in terms of Ukraine's supporters*; *Latvia also provides Ukraine an extensive military training programme*). The prominence of these lines shows that the Ministry of Defence and Latvia itself is heavily interested in supporting Ukraine militarily. Lastly, the emphasis on military relations between Latvia and the US is emphasised, as can be seen in the following examples (*International military exercise in Latvia and the US*; *US has been Latvia's strategic partner*

and ally for many years; *US is ready to continue to provide long-term military presence in Latvia and support further enhancement of our military capabilities*). The prominence of these lines show that Latvia greatly values and depends on the US and its military support in particular.

The keyword *LATVIAN* refers to different aspects of Latvian defence, such as participants in the defence of Latvia (*Latvian nationals*; *Latvian army*; *Latvian Armed Forces*; *Latvian troops*), the defence sector itself (*Latvian defence sector*; *Latvian defence policy*), protection of Latvia (*Defend every inch of Latvian soil*), education (*Patriotic education of Latvian youth*), and the possibility of Latvians supporting Ukraine militarily (*allowing Latvian citizens voluntarily join the Ukrainian Armed Forces*).

The keyword *DEFENCE* is primarily used when referring to different aspects of defence, such as officials in charge of defence (*Defence Minister*; *Defence experts*), defence infrastructure (*Defence service*; *Defence policy*; *Baltic defence line*; *Defence framework*; *Defence cooperation*; *Defence equipment*; *Defence capabilities*), and budget allocated to the defence sector (*Defence spending*; *Defence Budget*; *Defence funding law*). These show that Latvia has a defence strategy, and that it has resources and leaders who shape and implement defensive strategies.

The keyword *NATO* illustrates how crucial the alliance is to its defence plans. The discourse focuses on having more protection by the alliance (*Baltics is at the core of NATO plans*; *presence of NATO national forces in Latvia*; *Host the NATO Battlegroup*), the military capabilities of the alliance (*higher combat readiness across NATO*; *NATO's air defence systems in the Baltic region*), and the integration of Ukraine into the alliance (*Ukraine's NATO integration*; *Ukraine's NATO membership*).

The keyword *UKRAINE* reveals the themes of Latvia providing military support (*Military support for Ukraine*; *To provide Ukraine with one million drones*; *A new shipment of military aid to Ukraine*; *Our assistance to Ukraine will reach €370 million*), the Russian invasion (*Russian invasion of Ukraine*; *Russia's brutal war against Ukraine*; *If Russia is not fully defeated in Ukraine*; *By prolonging the war, Russia will be able to wear down Ukraine and its supporters*), and allies who support Ukraine alongside Latvia (*NATO's multi-year Ukraine assistance programme*; *Ammunition procurement for Ukraine*; *Latvia supports Czech ammunition initiative for artillery shells to Ukraine*). These show that Ukraine is massively supported by Latvia and its allies, whilst Russia and its aggression is described negatively.

The concordance analysis of this subcorpus shows that it is heavily centred around discourses of military, but more specific themes derived from the contexts include the defence

of Latvia, the support from NATO and the US, the war in Ukraine, and the military support that Latvia and its allies provide to Ukraine.

4.7.6. Summary of concordance analysis

The results of the concordance analysis shows that the semantic categories established in the keyword analysis reveal similar themes as when the words are viewed in context. This shows that the two approaches are useful for determining the main themes in the governmental discourses, but that the concordance analysis provides more clarity through context.

The main themes of the discourse were largely the same and include Latvia building and being part of alliances, Latvia and its political and military support for Ukraine, the security and defence of Latvia, and the condemnation of Russia. These appear in several of the subcorpora and indicate that their topicality is not observed only in Foreign Affairs and Ministry of Defence, but in the rest of the governmental institutions too. Likewise, the attractivity of Latvia in terms of its cultural and business landscape is often emphasised as well. Lastly, Latvia is often described as reinforcing freedom and democratic values, which is another prominent topic of the corpus.

At the same time, more specific themes were brought to attention by viewing the words in context. These include relations between Latvia and Norway in the financial sector, Latvia and Lithuania in the legal sector, and Latvia and the US in the military sector. Other previously unencountered themes include gambling and its restriction in Latvia, as well as the startup environment in Latvia. Lastly, the context allowed to reveal that most attributes assigned to Latvia are positive; the only negatives are attributed towards Russia, indicating that Latvia does not produce content that can be detrimental to their image, whilst also emphasising negatives to their political and military enemies.

Having established the findings of both keyword and concordance analysis, the findings can be formulated into ‘theses’ that describe Latvia and its image. The following chapter describes the findings and formulates theses about the image of Latvia.

4.8. Image of Latvia

The theses are defined according to the findings expressed in the previous chapters. The theses are defined within the context of their domain, meaning that for each subchapter, multiple theses are provided. Afterwards, a few overarching theses are established to present an overall image of Latvia.

The findings of the MFA subcorpus showed that the discourse describes political and diplomatic events. Thus, the following theses can be set. Firstly, the cooperation, support,

participation, and relation aspect are highly emphasised, therefore the theses are set as follows:

1. Latvian Foreign Relations are furthered and managed by its institutions and representatives;
2. Latvia cooperates with other countries and alliances;
3. Latvia is a member of many political alliances;
4. Latvia supports, assists, and promotes relations with other states;
5. Latvia is an active participant in national, regional, and international events;
6. Latvia discusses policies and improvements with their allies.

The high frequency and keyness of words that refer to Ukraine and Russia is also key; thus, the following theses are defined:

7. Latvia supports Ukraine;
8. Latvia opposes Russian aggression;
9. Latvia restricts the movement of Russian citizens into Latvia.

Lastly, Latvia and its candidacy for high international positions should be defined separately:

10. Latvia is running for the UN Security Council seat.

These theses show that the image of Latvia is defined by its positive diplomatic relations with other nations and alliances, the support for Ukraine, the condemnation of Russian aggression, and Latvia's ambitions on the international stage.

The findings of the Investment and Development Agency subcorpus showed that the discourses mainly revolve around attracting and maintaining business in Latvia. Thus, the following theses can be defined:

11. Latvia is increasing their efforts in attracting investment;
12. Latvia is developing and working to improve their business landscape;
13. Latvia is an attractive location for startups;
14. Latvia cooperates and builds relationships with their business partners;
15. Latvia is a part of a team;
16. Latvia offers business and financial support to investors;
17. Latvia has many interesting products of import and export;
18. Latvia is significant and interesting.

The relationship between Latvia and Norway must also be emphasised separately, as this is a topic that is more prevalent when compared to other countries:

19. Latvia has significant economic ties with Norway.

These theses characterise the image of Latvia in the economic landscape; Latvia positions itself as an attractive partner that can provide the framework and support for businesses,

particularly startups. Latvia's main economic partner in the recent years is Norway, having invested heavily in Latvia.

The findings of the Constitutional Court of Latvia subcorpus revealed that the discourse revolves around Latvia and its engagement in the international legal community, the close relations between Latvia and Lithuania, and the Latvian Court and its commitment to uphold the laws of Latvia, particularly through reviews and decisions of cases. Thus, the theses established are as follows:

20. Latvia participates in the international legal community;
21. Latvia is learning from the experiences of others;
22. The people of Latvia are divided into citizens and non-citizens;
23. Latvia safeguards its Constitution, democratic values and the rule of law;
24. Latvia protects its people and its language;
25. Latvia bans and restricts gambling;
26. Latvia and Lithuania have a long and continuous relationship in the context of the law.

These theses highlight Latvia as being a country that upholds law and democracy. At the same time, the division of people into the categories leaves a negative connotation in regard to the image of Latvia.

The findings of the missionLatvia subcorpus revealed that the discourse varies between presenting positive information about Latvia, their people, the traditions, as well as the business landscape. Thus, the findings can be formulated into the following theses:

27. Latvia and its culture are unique.
28. Latvia has a close relationship with nature and traditions.
29. Riga has many urban landmarks.
30. Latvia has an attractive business and startup landscape.

These theses show that the image of Latvia is shaped by the positive attributes of Latvia itself but is also influenced by the search for investment.

The findings of the Ministry of Defence subcorpus revealed that the discourse is centred around the military sector. However, the inclusion of historical events and the war in Ukraine provides some more layers the content that is produced. Thus, the image of Latvia can be defined by the following theses:

31. Latvia remembers and takes pride in its historical military events.
32. Latvia improves its military capabilities.
33. Latvia provides military support for Ukraine.
34. Latvia condemns Russia's aggression in Ukraine.
35. Latvia relies on NATO and the US for military cooperation and assistance.

The theses present the image of Latvia for their respective discourses, thus, the last step in the process of finalisation would be to produce the overarching themes of the whole discourse. Firstly, Latvia is highlighted as being an active participant and ally in many different settings, therefore it can be stated that LATVIA IS A PROACTIVE PARTICIPANT IN NATIONAL, REGIONAL, AND GLOBAL RELATIONS AND EVENTS. Secondly, the economic and business development is highly key in the discourses of Latvia, thus, it can be stated that LATVIA IS DEDICATED TO ENHANCE ITS BUSINESS LANDSCAPE BY ATTRACTING INVESTMENT. Thirdly, the discourses manifest Latvia's legal integrity, therefore it can be stated that LATVIA SAFEGUARDS ITS CONSTITUTION, DEMOCRATIC VALUES, AND RULE OF LAW. Fourthly, the findings show that Latvia is concerned with its security, therefore it can be stated that LATVIA IS COMMITED TO ENHANCE ITS POLITICAL SECURITY. Fifthly, one of the main topics encountered in all of the subcorpora refer to the war in Ukraine, as well as Latvia's involvement. Thus, it can be stated that LATVIA SUPPORTS UKRAINE AND OPPOSES RUSSIA IN THE WAR. Lastly, the positive attributes related to Latvia and its culture is a key point as well, therefore it can be stated that LATVIA VALUES ITS UNIQUE CULTURE, NATURE, TRADITIONS, AND URBAN HERITAGE. These points illustrate that the image of Latvia is that of a nation deeply committed to international cooperation, economic growth, and cultural representation. Latvia also takes a strong stance against aggression, condemning Russia as an aggressor and helping Ukraine to fight and survive. Lastly, Latvia is described as being an attractive destination for potential business and investment.

Conclusions

The goal of this study was to examine the country image of Latvia on Latvian governmental websites. To do that, the study aimed to compile and analyse a specialised corpus, titled Latvian Governmental Web Texts Corpus, and analyse it according to the tenets of Corpus Assisted Discourse Studies (CADS), namely, keywords and concordances, which allowed to identify semantic categories and derive the main statements regarding image of Latvia.

The examination of previous studies showed that the topic of country image is examined through the lens of discourse analysis, which analyses language use beyond a single sentence, focusing on social practices, ideology, and power use in language.

The theoretical material on CADS revealed its usefulness for analysing large amounts of data whilst also taking context into account, merging the approaches of corpus-based linguistics and discourse analysis. Theoretical considerations of researchers were also examined on the resources employed in the study, namely, frequency lists, keyword analysis, and concordance analysis. It was concluded that corpus analysis is multifaceted and that scholars do not share a united approach, however, the application of the CADS and its tenets, keyword analysis and concordance analysis reveal the most key words in discourse, providing quantitative results, and the use of words in context, which provides qualitative results.

The examination of theoretical material about the classification of corpora and corpus building design allowed to present a framework that was used to create a specialised corpus for the purposes of this study, as well as find and utilise a reference corpus to compare the data.

The results of the empirical research revealed the most prominent words in the corpus, as well as the key ideas in each subcorpora. The keyword analysis also allowed to identify the semantic categories of the words, which formed the initial layer of findings. The most prominent categories included government and the public domain, social actions, states and processes, general and abstract terms, names, and money and commerce. The concordance analysis of the subcorpora showed that the semantic categories revealed by keywords matched with the themes when keywords were further explored in context, but also revealed additional themes, such as particular relations that Latvia has with other countries (Norway in business, Lithuania in law, the US in military defence), and Latvia's political involvement in the war between Ukraine and Russia.

The main statements derived from the analysis revealed that the image of Latvia is that of a proactive participant in national, regional, and global relations and events, that Latvia is dedicated to enhance its business landscape by attracting investment, that Latvia safeguards

its constitution, democratic values, and rule of law, that Latvia is committed to enhance its political security, that Latvia supports Ukraine and opposes Russia in the war, and lastly, that Latvia values its unique culture, nature, traditions, and urban heritage. The statements also indicate that Latvia is primarily given positive attributes, as very few negative attributes were identified, highlighting the genre of texts that were produced, and the fact that the government wanted to focus on the positive aspects of Latvia. This leads to the conclusion that the image of Latvia can also be described as positive.

Thus, the findings of the study are sufficient to establish the image of Latvia as presented in the texts of 2023 and 2024, revealing the key words and themes of the discourse, and the context in which these words appear in.

Choices of scope and tools can be listed as the limitations of the study. Due to size and time constraints, a wider coverage of concordances and governmental sources was limited. A wider coverage of concordance analysis could have revealed more contexts in which words were used, whilst a wider coverage of governmental sources could have revealed more areas of governmental discourse. The study addresses this issue by gathering balanced and representative results from multiple different sources that can be attributed to the overall image of Latvia. Secondly, much work on the compilation of the corpus and the semantic tagging was done manually; thus, the possibility of human error can be described as one of the limitations.

In terms of further research, this study provides a base point for further analysis, as the corpus and its findings can be applied in future studies. For example, the findings about the country image of Latvia derived from governmental websites can be compared to texts produced by foreign media or governments, looking to compare whether these statements hold true, or that there are different thematic threads that prevail. Likewise, comparative analyses can be done with other Latvian governmental websites, or with the same websites in different time periods. Such research could highlight the similarities and differences between different websites and different time periods. Thus, there are numerous possibilities to further the research on this topic, but this study provides a solid starting point for future research.

Theses

1. Previous studies reveal that country images are primarily examined through the perspective of discourse analysis and critical discourse analysis.
2. Discourse analysis examines language use in social settings, whilst critical discourse analysis focuses on social problems, and the use of power and ideology in discourse.
3. Corpus Assisted Discourse Studies (CADS) is the most suitable approach to analyse discourse in large amounts of texts, as it combines corpus-based analysis and discourse analysis, allowing to examine large amounts of texts whilst taking context into account.
4. Relevantly designed specialised corpora are the most appropriate for generating both quantitative and qualitative data in CADS.
5. Useful tools for quantitative and qualitative analyses of country images in CADS include keyword and concordance analyses.
6. The keyword analysis reveals the main themes in the corpus of Latvian governmental discourse: government and the public domain, social actions, states and processes, general and abstract terms, names, as well as money and commerce.
7. Concordance analysis corroborates the findings of the keyword analysis by providing context and thus an in-depth insight into the themes.
8. The themes derived from concordance analysis uncover support for Ukraine, opposition of Russia, and close relations between Latvia, Lithuania, Norway, and the US.
9. The data extracted from keyword analysis, semantic category analysis and concordance mining uncovers statements about Latvia that can be used to define its country image.
10. The image of Latvia is that of a proactive participant in national, regional, and global relations and events, that Latvia is committed to enhance its business landscape by attracting investment, that Latvia safeguards its constitution, democratic values, and rule of law, that Latvia is committed to improve its political security, and lastly, that Latvia supports Ukraine and opposes Russia in the war between the two.

References

1. Anthony, L. (2022) What can corpus software do? In A. O'Keeffe and M. J. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, 2nd ed. (pp. 250-262). Abingdon: Routledge.
2. Baker, P. (2023) *Using Corpora in Discourse Analysis*, 2nd ed. London: Bloomsbury Academic. Perlego edition.
3. Bianchi, F. (2012) *Culture, Corpora and Semantics*. Lecce: Universita di Salento. Available from http://siba-ese.unisalento.it/index.php/culturecorpora/article/view/12434/11073 [Accessed 19 March 2024].
4. Brezina, V. (2018) *Statistics in Corpus Linguistics.* Cambridge: Cambridge University Press.
5. Brown, G. and Yule G. (1983) *Discourse Analysis.* Cambridge: Cambridge University Press. Available from https://books.google.lv/books/about/Discourse_Analysis.html?id=9gUrbzov9x0C&redir_esc=y [Accessed on 4 April 2024].
6. Buhmann, A. (2016) *Measuring country image. Theory, method, and effects.* Wiesbaden: Springer VS. Available from https://www.academia.edu/27715438/Measuring_country_image_Theory_method_and_effects [Accessed on 5 March 2024].
7. Buhmann, A. and Ingenhoff, D. (2014) The 4D Model of the country image: An integrative approach from the perspective of communication management. *International Communication Gazette*, 77 (1): 102-124. Available from https://www.researchgate.net/publication/273311529_The_4D_Model_of_the_country_image_An_integrative_approach_from_the_perspective_of_communication_management [Accessed on 22 January 2024].
8. Cotirlea, D. (2015) Country Image vs. Country Brand: Differences and Similarities. *Ecoforum*, 4 (1): 165-171. Available from https://www.researchgate.net/publication/280742426_COUNTRY_IMAGE_VS_COUNTRY_BRAND_DIFFERENCES_AND_SIMILARITIES [Accessed on 22 January 2024].
9. Crystal, D. (1997) *The Cambridge Encyclopedia of Language*, 2nd ed. The UK: Cambridge University Press. Available from https://archive.org/details/cambridgeencycl000crys [Accessed on 27 March 2024].
10. Culpeper, J. and Demmen, J. (2015) Keywords. In D. Biber and R. Reppen (eds.) *The Cambridge Handbook of English Corpus Linguistics* (pp. 90-105). Cambridge: Cambridge

University Press. Available from https://archive.org/details/cambridgehandboo0000unse_e0h3 [Accessed on 20 March 2024].

11. Fairclough, N. (1989). *Language and Power*. London: Longman. Available from https://www.okehamptoncollege.devon.sch.uk/uploads/6/1/4/4/61443371/ebooksclub.org__language_and_power__language_in_social_life_.pdf [Accessed on 5 April 2024].
12. Fürsich, E. and Robins, M. (2002) Africa.com: the self-representation of sub-Saharan nations on the World Wide Web. *Critical Studies in Media Communication*, 19 (2): 190-211. Available from https://www.cambridgescholars.com/resources/pdfs/978-1-4438-5901-1-sample.pdf [Accessed on 18 March 2024].
13. Gatto, M. (2014) *The Web As Corpus: Theory and Practice.* London: Bloomsbury Publishing. Perlego edition.
14. Gillings, M. and Mautner, G. (2022) Concordancing for CADS. Practical challenges and theoretical implications. *International Journal of Corpus Linguistics*, 29 (1): 34-58. Available from https://www.jbe-platform.com/content/journals/10.1075/ijcl.21168.gil [Accessed 19 March 2024].
15. Gillings, M., Learmonth, M. and Mautner, G. (2024) Taking the Road Less Travelled: How Corpus-Assisted Discourse Studies Can Enrich Qualitative Explorations of Large Textual Datasets. *British Journal of Management*, 0: 1-13. Available from https://onlinelibrary.wiley.com/doi/epdf/10.1111/1467-8551.12816 [Accessed 22 May 2024].
16. Gries, S. Th. (2009) *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge. Perlego edition.
17. Gries, S. Th. (2021) A new approach to (key) keywords analysis: Using frequency, and now also dispersion. *Research in Corpus Linguistics*, 9 (2): 1-33. Available from https://www.researchgate.net/publication/349580151_A_new_approach_to_key_keywords_analysis_Using_frequency_and_now_also_dispersion [Accessed on 25 February 2024].
18. Hodges, B. D., Kuper, A. and Reeves, S. (2008) Qualitative Research: Discourse Analysis. *British Medical Journal*, 337 (7669): 570-572. Available from https://www.jstor.org/stable/20510756 [Accessed on 5 April 2024].
19. Hyland, K. (2015) Corpora and written academic English. In D. Biber and R. Reppen (eds.) *The Cambridge Handbook of English Corpus Linguistics* (pp. 292-308). Cambridge: Cambridge University Press. Available from https://ueaeprints.uea.ac.uk/id/eprint/65844/1/Written_academic_English_Hyland_AAM.pdf [Accessed on 2 April 2024].

20. Jørgensen, M. and Phillips, L. (2002) *Discourse Analysis as Theory and Method*. London: SAGE Publications Ltd. Available from https://methods.sagepub.com/book/discourse-analysis-as-theory-and-method [Accessed on 3 April 2024].
21. Kenzhalina, G. Zh., Dzhunusova, Z. Kh., Aitmagambetov, D. R., Turuntayeva, A. A. and Bolysbekova, M. K. (2016) Country Image Formation in Terms of the Concepts of Public Diplomacy and Soft Power. *The Social Sciences*, 11 (17): 4240-4244. Available from https://www.researchgate.net/publication/311698482_Country_image_formation_in_terms_of_the_concepts_of_public_diplomacy_and_soft_power [Accessed on 19 March 2024].
22. Kilgarriff, A. (2009) Simple maths for keywords. In Mahlberg, M., González-Díaz, V. and Smith, C (eds.) *Proceedings of Corpus Linguistics Conference CL2009*. University of Liverpool. Available from https://www.sketchengine.eu/wp-content/uploads/2015/04/2009-Simple-maths-for-keywords.pdf [Accessed on 25 May 2024].
23. Kleppe, A. I., Iversen, N. M. and Stensaker, I. (2002) Country image in marketing strategies: Conceptual issues and an empirical Asian illustration. *Journal of Brand Management*, 10 (1): 61-74. Available from https://www.researchgate.net/publication/233568549_Country_image_in_marketing_strategies_Conceptual_issues_and_an_empirical_Asian_illustration [Accessed 4 March 2024].
24. Koester, A. (2022) Building small specialised corpora. In A. O'Keeffe and M. J. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, 2nd ed. (pp. 48-61). Abingdon: Routledge.
25. Li, D., Ahn, J., Zhou, R. and Wu, B. (2009) A Study on the influence of country image on purchase intention of Chinese consumers based on Fishbein's model of reasoned action: Focused on USA, Germany, Japan and South Korea. *Frontiers of Business Research in China*, 3: 621-646. Available from https://fbr.springeropen.com/counter/pdf/10.1007/s11782-009-0030-2.pdf [Accessed on 19 March 2024].
26. Marinao-Artigas, E. and Barajas-Portas, K. (2021) A Cross-Destination Analysis of Country Image: A Key Factor of Tourism Marketing. *Sustainability*, 13 (7): 9529. Available from https://www.mdpi.com/2071-1050/13/17/9529 [Accessed on 19 March 2024].
27. Mautner, G. (2022) What can a corpus tell us about discourse? In A. O'Keeffe and M. J. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, 2nd ed. (pp. 250-262). Abingdon: Routledge.

28. McArthur, T. (1992) *The Oxford Companion to the English Language.* New York: Oxford University Press.
29. McCarthy, M. (1991) *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press. Available from https://sacunslc.wordpress.com/wp-content/uploads/2015/03/michael-mccarthy-discourse-analysis-for-language-teachers-cambridge-language-teaching-library-1991.pdf [Accessed on 4 April 2024].
30. Miazhevich, G. (2007) Official media discourse and the self-representation of entrepreneurs in Belarus. *Europe-Asia Studies*, 59 (8): 1331-1348. Available from https://www.jstor.org/stable/20451456 [Accessed on 18 March 2024].
31. Nielsen, A. E. and Nørreklit (2009) A discourse analysis of the disciplinary power of management coaching. *Society and Business Review*, 4 (1): 202-214. https://www.researchgate.net/publication/40242085_A_Discourse_Analysis_of_the_Disciplinary_Power_of_Management_Coaching
32. Paquot, M. and Bestgen, Y. (2009) Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction. *Language and Computers*, 68: 247-269. Available from https://www.researchgate.net/publication/233624839_Distinctive_words_in_academic_writing_A_comparison_of_three_statistical_tests_for_keyword_extraction [Accessed on 25 May 2024].
33. Partington, A. (2003) *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge. Available from https://ebookcentral-proquest-com.datubazes.lanet.lv/lib/lulv/reader.action?docID=171356 [Accessed on 5 April 2024].
34. Pojanapunya, P. and Todd, R. W. (2016) Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, 14 (1): 1-21. Available from https://www.researchgate.net/publication/301537550_Log-likelihood_and_odds_ratio_Keyness_statistics_for_different_purposes_of_keyword_analysis [Accessed on 4 April 2024].
35. Rayson, P. (2015) Computational tools and methods for corpus compilation and analysis. In D. Biber and R. Reppen (eds.) *The Cambridge Handbook of English Corpus Linguistics* (pp. 32-49). Cambridge: Cambridge University Press. Available from https://archive.org/details/cambridgehandboo0000unse_e0h3 [Accessed on 20 March 2024].
36. Rayson, P., Archer, D., Piao, S. and McEnery, T. (2004) The UCREL semantic analysis system. Available from

https://www.researchgate.net/publication/228881331_The_UCREL_semantic_analysis_system [Accessed 22 May 2024].

37. Schiffrin, D. (2006) Discourse. In Fassod, R. W. and Connor-Linton, J. (eds.) *An Introduction to Language and Linguistics* (pp. 169-204). Cambridge: Cambridge University Press. Available from https://repository.bbg.ac.id/bitstream/531/1/An_Introduction_to_Language_and_Linguistics.pdf [Accessed on 3 April 2024].
38. Schiffrin, D., Tannen, D. and H. E. Hamilton (eds.) (2001) *The Handbook of Discourse Analysis.* Oxford: Blackwell Publishers. Available from https://edisciplinas.usp.br/pluginfile.php/4966500/mod_resource/content/1/Deborah%20Schiffrin%2C%20Deborah%20Tannen%2C%20Heidi%20E.%20Hamilton%20The%20Handbook%20of%20Discourse%20Analysis.pdf [Accessed on 3 April 2024].
39. Tarasheva (2014) *The Image of a Country created by International Media: The Case of Bulgaria*. Newcastle: Cambridge Scholars Publishing. Available from https://www.cambridgescholars.com/resources/pdfs/978-1-4438-5901-1-sample.pdf [Accessed on 18 March 2024].
40. Tarasheva, E. (2014) Studying the image of a country - three theoretical perspectives. In D. Yankova (ed.) *Cross-linguistic Interaction: Translation, Contrastive and Cognitive Studies* (pp. 422-439). Sofia: St. Kliment Ohridski University Press. Available from https://www.researchgate.net/publication/265729372_Studying_the_image_of_a_country_-_three_theoretical_perspectives [Accessed on 18 March 2024].
41. van Dijk, T. A. (1988) *News as Discourse*. New Jersey: Lawrence Erlbaum Associates. Available from https://discourses.org/wp-content/uploads/2022/06/Teun-A.-van-Dijk-1988-News-As-Discourse.pdf [Accessed on 2 April 2024].
42. van Dijk, T. A. (2001) Critical Discourse Analysis. In Schiffrin, D., Tannen, D. and H. E. Hamilton (eds.) *The Handbook of Discourse Analysis* (352-371). Oxford: Blackwell Publishers. https://edisciplinas.usp.br/pluginfile.php/4966500/mod_resource/content/1/Deborah%20Schiffrin%2C%20Deborah%20Tannen%2C%20Heidi%20E.%20Hamilton%20The%20Handbook%20of%20Discourse%20Analysis.pdf [Accessed on 2 April 2024].
43. van Dijk, T. A. (ed.), (1997) *Discourse as Structure and Process: Discourse Studies: A Multidisciplinary Introduction.* London: SAGE Publications. Available from https://search.worldcat.org/title/Discourse-as-structure-and-process/oclc/464712973 [Accessed on 2 April 2024].

44. Wang, C. L., Li D., Barnes, B. R. and Ahn, J. (2012) Country image, product image and consumer purchase intention: Evidence from an emerging economy. *International Business Review*, 21 (6): 1041-1051. Available from https://www.sciencedirect.com/science/article/abs/pii/S096959311100196X [Accessed on 19 March 2024].
45. Weisser, M. (2022). What Corpora are available? In A. O'Keeffe and M. J. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics*, 2nd ed. (pp. 89-102). Abingdon: Routledge.
46. Wilk, P. (2014) *Stance in Press Discourse: The Case of Poles and Poland in the Guardian.* Opole: Wydawnictwo Uniwersytetu Opolskiego. Available from https://repo.uni.opole.pl/docstore/download/UO8f4d50afa28b4549944479cbfbcc1417/WilkP-Stancein.pdf [Accessed on 19 March 2024].
47. Wodak, R. (2001) What CDA is about - a summary of its history, important concepts and its developments. In R. Wodak and M. Meyer (eds.) *Methods of Critical Discourse Analysis* (pp. 1-13). London: SAGE. Available from https://methods.sagepub.com/book/methods-of-critical-discourse-analysis [Accessed on 2 April 2024].
48. Wodak, R. (2013) Editor's Introduction: Critical Discourse Analysis. In R. Wodak (ed.) *Critical Discourse Analysis Volume One: Histories, Concepts and Interdisciplinarity* (19-39). London: SAGE. Available from https://www.academia.edu/2544299/Wodak_Ruth_Ed_Critical_Discourse_Analysis_Four_Volumes [Accessed on 2 April 2024].
49. Zhang, B. and Shu, Z. (2023) The Images of China in TIME: A Corpus-Based Critical Discourse Analysis. *International Journal of English Linguistics*, 13 (5): 34-44. Available from https://ccsenet.org/journal/index.php/ijel/article/view/0/49219 [Accessed on 29 March 2024].
50. Zhao, X., You, X. and Lin, S. (2022) China's country image in the eyes of international students from central Asian countries. *Front Psychol*, 13: 569789. Available from https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.569789/full [Accessed on 22 January 2024].

Online sources

1. [Online 1] Cambridge Advanced Learner's Dictionary & Thesaurus. Available from https://dictionary.cambridge.org/dictionary/english/image [Accessed on 22 January 2024].
2. [Online 2] Merriam-Webster.com Dictionary. Available from https://www.merriam-webster.com/dictionary/image [Accessed on 22 January 2024].

3. [Online 3] Sketch Engine. Available from https://www.sketchengine.eu/corpora-and-languages/corpus-types/
4. [Online 4] CLARIN. Available from https://www.clarin.eu/resource-families/parallel-corpora [Accessed on 5 April 2024].
5. [Online 5] Sketch Engine. Available from https://www.sketchengine.eu/corpora-and-languages/parallel-corpora/ [Accessed on 5 April 2024].
6. [Online 6] Sketch Engine. Available from https://www.sketchengine.eu/my_keywords/reference-corpus/ [Accessed on 5 April 2024].
7. [Online 7] Sketch Engine. Available from https://www.sketchengine.eu/my_keywords/freqmill/ [Accessed on 26 April 2024].
8. [Online 8] Sketch Engine. Available from https://www.sketchengine.eu/documentation/simple-maths/ [Accessed on 26 April 2024].
9. [Online 9] Available from https://ucrel.lancs.ac.uk/usas/USASSemanticTagset.pdf [Accessed on 9 May 2024].
10. [Online 10] Available from https://ucrel-api.lancaster.ac.uk/usas/tagger.html [Accessed on 9 May 2024].

24490 words

Appendix 1. Tables

Table 1. **Top 150 keywords of the whole corpus, minimum frequency of 200**

	Absolute Frequency			Frequency per million		Score
	Word	Focus	Reference	Focus	Reference	
1	latvia	4,021	142,310	10,239.50	2.31	3,093.1
2	latvian	1,424	73,965	3,626.22	1.20	1,648.0
3	kariņš	376	191	957.49	< 0.01	955.5
4	krišjānis	360	220	916.74	< 0.01	914.5
5	rinkēvičs	287	170	730.85	< 0.01	729.8
6	edgars	268	2,256	682.46	0.04	659.3
7	riga	459	72,183	1,168.85	1.17	538.6
8	liaa	203	421	516.94	< 0.01	514.4
9	andris	213	5,805	542.41	0.09	496.6
10	baltic	613	220,674	1,561.01	3.58	340.8
11	ukraine	1,675	994,839	4,265.40	16.15	248.7
12	ministers	480	273,128	1,222.32	4.43	225.1
13	contested	241	113,180	613.71	1.84	216.6
14	affairs	1,379	1,289,452	3,511.63	20.94	160.1
15	nato	595	665,102	1,515.17	10.80	128.5
16	cooperation	1,306	1,715,762	3,325.74	27.86	115.3
17	constitutional	948	1,268,581	2,414.09	20.60	111.8
18	aggression	338	423,427	860.72	6.88	109.4
19	defence	1,167	1,686,029	2,971.77	27.38	104.8
20	bilateral	278	366,949	707.93	5.96	101.9
21	underline	242	331,356	616.25	5.38	96.7
22	ambassador	468	1,108,936	1,191.76	18.01	62.8
23	foreign	2,166	5,422,647	5,515.73	88.05	62.0
24	ukrainian	233	558,095	593.34	9.06	59.1
25	ministry	1,017	3,346,624	2,589.80	54.34	46.8
26	eu	836	2,876,621	2,128.88	46.71	44.6
27	russia	805	3,259,815	2,049.94	52.93	38.0
28	minister	1,411	6,246,399	3,593.12	101.43	35.1
29	strengthen	399	2,032,221	1,016.06	33.00	29.9
30	european	1,222	6,430,293	3,111.83	104.41	29.5
31	sanction	237	1,224,990	603.52	19.89	28.9
32	armed	241	1,299,441	613.71	21.10	27.8
33	republic	348	2,254,584	886.18	36.61	23.6
34	secretary	432	3,351,765	1,100.09	54.42	19.9
35	nations	200	1,523,601	509.30	24.74	19.8
36	un	244	1,971,635	621.35	32.01	18.9
37	export	282	2,741,793	718.11	44.52	15.8
38	organisation	351	3,526,623	893.82	57.26	15.4
39	russian	344	3,476,705	876.00	56.45	15.3
40	constitution	211	2,133,150	537.31	34.64	15.1
41	europe	530	5,902,881	1,349.65	95.85	13.9
42	union	605	6,807,305	1,540.64	110.53	13.8
43	council	774	8,818,038	1,971.00	143.18	13.7
44	representative	337	3,829,222	858.17	62.18	13.6

45	court	1,292	15,645,613	3,290.09	254.04	12.9
46	capability	254	3,052,561	646.81	49.57	12.8
47	programme	390	4,877,532	993.14		
48	security	954	12,076,799	2,429.37	196.10	12.3
49	priority	234	2,955,948	595.88	48.00	12.2
50	military	535	7,212,570	1,362.38	117.11	11.5
51	meeting	714	10,408,510	1,818.20	169.01	10.7
52	provision	228	3,466,406	580.60	56.29	10.2
53	implementation	223	3,401,342	567.87	55.23	10.1
54	international	1,009	16,077,934	2,569.42	261.06	9.8
55	relation	265	4,213,184	674.82	68.41	9.7
56	innovation	225	3,618,808	572.96	58.76	9.6
57	discuss	483	8,488,191	1,229.96	137.83	8.9
58	investment	390	6,845,641	993.14	111.16	8.9
59	region	514	9,777,203	1,308.90	158.76	8.2
60	states	510	9,755,332	1,298.72	158.40	8.2
61	sector	315	6,017,674	802.15	97.71	8.1
62	institution	262	5,166,456	667.18	83.89	7.9
63	conference	395	8,287,398	1,005.87	134.57	7.4
64	partner	376	7,964,176	957.49	129.32	7.4
65	development	970	20,912,084	2,470.11	339.56	7.3
66	legal	343	7,411,431	873.45	120.34	7.2
67	country	1,114	24,593,575	2,836.81	399.34	7.1
68	official	437	9,873,593	1,112.82	160.32	6.9
69	promote	272	6,201,295	692.65	100.69	6.8
70	discussion	308	7,048,110	784.32	114.44	6.8
71	national	895	20,976,449	2,279.12	340.60	6.7
72	aim	261	6,212,881	664.64	100.88	6.5
73	economic	358	8,635,910	911.65	140.23	6.5
74	law	795	19,701,231	2,024.47	319.90	6.3
75	implement	213	5,430,081	542.41	88.17	6.1
76	political	396	10,576,884	1,008.42	171.74	5.8
77	protection	226	6,054,300	575.51	98.31	5.8
78	support	1,263	34,671,852	3,216.24	562.98	5.7
79	mission	205	5,853,432	522.03	95.04	5.4
80	situation	297	8,638,546	756.31	140.27	5.4
81	express	201	5,888,528	511.85	95.61	5.3
82	visit	479	14,263,341	1,219.78	231.60	5.2
83	united	391	12,884,936	995.68	209.22	4.7
84	ensure	288	9,735,401	733.39	158.08	4.6
85	february	202	7,020,515	514.39	114.00	4.5
86	establish	252	8,793,649	641.72	142.79	4.5
87	policy	399	14,779,796	1,016.06	239.99	4.2
88	agency	215	8,056,434	547.50	130.82	4.2
89	language	287	10,855,724	730.85	176.27	4.1
90	committee	222	8,484,842	565.32	137.77	4.1
91	state	1,124	44,001,914	2,862.27	714.48	4.0
92	meet	416	16,574,764	1,059.35	269.13	3.9
93	president	342	13,875,749	870.90	225.31	3.9
94	technology	393	15,972,296	1,000.78	259.35	3.8

95	address	303	12,357,429	771.59	200.65	3.8
96	war	351	15,138,626	893.82	245.81	3.6
97	potential	204	8,799,265	519.49	142.88	3.6
98	global	230	10,440,638	585.70	169.53	3.4
99	current	314	14,517,543	799.60	235.73	3.4
100	opportunity	280	13,262,098	713.02	215.34	3.3
101	challenge	241	11,492,203	613.71	186.60	3.3
102	director	237	11,380,631	603.52	184.79	3.3
103	energy	277	13,427,889	705.38	218.03	3.2
104	decision	228	11,057,843	580.60	179.55	3.2
105	member	535	26,509,604	1,362.38	430.45	3.2
106	education	284	14,175,926	723.21	230.18	3.1
107	training	222	11,233,902	565.32	182.41	3.1
108	project	526	26,893,646	1,339.46	436.68	3.1
109	against	483	24,758,720	1,229.96	402.02	3.1
110	develop	350	18,272,032	891.28	296.69	3.0
111	society	202	10,773,581	514.39	174.94	2.9
112	continue	378	20,527,710	962.58	333.32	2.9
113	between	655	35,702,508	1,667.96	579.72	2.9
114	section	251	13,706,973	639.17	222.57	2.9
115	rule	216	12,005,312	550.05	194.94	2.8
116	important	333	18,637,375	847.99	302.62	2.8
117	solution	205	11,647,461	522.03	189.13	2.8
118	interest	235	13,384,277	598.43	217.33	2.7
119	market	356	20,354,951	906.56	330.51	2.7
120	human	295	17,134,010	751.22	278.21	2.7
121	business	506	30,245,087	1,288.53	491.10	2.6
122	person	285	17,106,414	725.75	277.76	2.6
123	product	339	20,651,075	863.27	335.32	2.6
124	force	272	16,613,252	692.65	269.76	2.6
125	company	548	33,803,038	1,395.48	548.88	2.5
126	hold	365	22,548,789	929.47	366.13	2.5
127	industry	220	13,621,608	560.23	221.18	2.5
128	information	521	32,550,269	1,326.73	528.53	2.5
129	event	343	21,829,403	873.45	354.45	2.5
130	value	273	18,029,398	695.20	292.75	2.4
131	case	449	29,768,759	1,143.38	483.37	2.4
132	future	219	14,774,200	557.68	239.90	2.3
133	during	474	33,072,058	1,207.04	537.01	2.2
134	provide	615	43,074,594	1,566.10	699.42	2.2
135	order	346	24,802,917	881.09	402.74	2.2
136	general	220	16,101,873	560.23	261.45	2.1
137	public	324	23,882,320	825.07	387.79	2.1
138	together	211	15,612,426	537.31	253.51	2.1
139	right	545	40,490,919	1,387.85	657.47	2.1
140	million	201	15,597,494	511.85	253.26	2.0
141	plan	298	23,873,334	758.86	387.64	2.0
142	issue	341	27,581,421	868.36	447.85	1.9
143	article	212	17,506,660	539.86	284.26	1.9
144	head	222	19,022,912	565.32	308.88	1.8

145	view	203	18,045,402	516.94	293.01	1.8
146	government	264	24,282,531	672.28	394.29	1.7
147	part	419	39,215,936	1,066.99	637.77	1.7
148	local	217	20,438,710	552.59	331.87	1.7
149	increase	218	21,742,990	555.14	353.05	1.6
150	Include	591	60,125,985	1,504.98	976.29	1.5

Table 2 **Keywords of the Ministry of Foreign Affairs subcorpus**

	Absolute Frequency			Frequency per million		Score
	Word	Focus	Reference	Focus	Reference	
1	latvia	2,367	142,310	12,222.39	2.31	3,692.0
2	kariņš	375	191	1,936.37	< 0.01	1,931.4
3	krišjānis	359	220	1,853.75	< 0.01	1,848.2
4	latvian	715	73,965	3,692.02	1.20	1,677.9
5	rinkēvičs	284	170	1,466.48	< 0.01	1,463.4
6	edgars	264	2,256	1,363.21	0.04	1,316.0
7	pelšs	168	27	867.50	< 0.01	868.1
8	andris	182	5,805	939.79	0.09	859.7
9	nb8	127	369	655.79	< 0.01	652.9
10	riga	233	72,183	1,203.13	1.17	554.4
11	baltic	421	220,674	2,173.90	3.58	474.5
12	ukraine	1,424	994,839	7,353.06	16.15	428.7
13	ministers	414	273,128	2,137.76	4.43	393.5
14	moldova	162	90,317	836.51	1.47	339.6
15	affairs	1,308	1,289,452	6,754.07	20.94	307.9
16	aggression	317	423,427	1,636.88	6.88	208.0
17	estonia	126	158,220	650.62	2.57	182.6
18	underline	222	331,356	1,146.33	5.38	179.8
19	bilateral	231	366,949	1,192.81	5.96	171.6
20	cooperation	937	1,715,762	4,838.35	27.86	167.7
21	lithuania	123	201,362	635.13	3.27	149.0
22	nato	339	665,102	1,750.48	10.80	148.4
23	font	454	1,088,730	2,344.30	17.68	125.6
24	ambassador	445	1,108,936	2,297.83	18.01	121.0
25	foreign	2,054	5,422,647	10,606.16	88.05	119.1
26	presidency	171	586,331	882.99	9.52	84.0
27	ukrainian	163	558,095	841.68	9.06	83.7
28	eu	721	2,876,621	3,723.00	46.71	78.1
29	ministry	829	3,346,624	4,280.68	54.34	77.4
30	russia	749	3,259,815	3,867.58	52.93	71.7
31	parliamentary	153	656,118	790.04	10.65	67.9
32	delegation	136	580,987	702.26	9.43	67.4
33	accessible	417	2,032,502	2,153.25	33.00	63.4
34	minister	1,163	6,246,399	6,005.34	101.43	58.6
35	sanction	234	1,224,990	1,208.30	19.89	57.9
36	contrast	482	2,735,788	2,488.89	44.42	54.8
37	reconstruction	113	635,460	583.49	10.32	51.6
38	strengthen	272	2,032,221	1,404.52	33.00	41.3

39	nations	197	1,523,601	1,017.24	24.74	39.6
40	secretary	408	3,351,765	2,106.77	54.42	38.0
41	defence	207	1,686,029	1,068.88	27.38	37.7
42	un	235	1,971,635	1,213.46	32.01	36.8
43	european	736	6,430,293	3,800.46	104.41	36.1
44	consultation	146	1,505,610	753.89	24.45	29.7
45	organisation	296	3,526,623	1,528.44	57.26	26.3
46	republic	190	2,254,584	981.10	36.61	26.1
47	russian	279	3,476,705	1,440.66	56.45	25.1
48	council	678	8,818,038	3,500.96	143.18	24.3
49	dialogue	127	1,777,233	655.79	28.86	22.0
50	summit	114	1,629,941	588.66	26.47	21.5
51	europe	392	5,902,881	2,024.16	95.85	20.9
52	union	419	6,807,305	2,163.57	110.53	19.4
53	security	733	12,076,799	3,784.96	196.10	19.2
54	representative	234	3,829,222	1,208.30	62.18	19.1
55	priority	170	2,955,948	877.82	48.00	17.9
56	meeting	568	10,408,510	2,932.96	169.01	17.3
57	integration	118	2,116,797	609.31	34.37	17.3
58	relation	231	4,213,184	1,192.81	68.41	17.2
59	agenda	108	1,987,394	557.68	32.27	16.8
60	discuss	396	8,488,191	2,044.81	137.83	14.7
61	international	710	16,077,934	3,666.20	261.06	14.0
62	size	458	10,751,756	2,364.96	174.58	13.5
63	region	408	9,777,203	2,106.77	158.76	13.2
64	assistance	144	3,481,849	743.57	56.54	12.9
65	states	393	9,755,332	2,029.32	158.40	12.7
66	reform	124	3,120,437	640.29	50.67	12.4
67	content	421	10,990,782	2,173.90	178.46	12.1
68	implementation	125	3,401,342	645.46	55.23	11.5
69	importance	128	3,558,156	660.95	57.78	11.3
70	discussion	246	7,048,110	1,270.26	114.44	11.0
71	programme	168	4,877,532	867.50	79.20	10.8
72	contribution	128	3,829,378	660.95	62.18	10.5
73	official	327	9,873,593	1,688.52	160.32	10.5
74	regional	145	4,349,882	748.73	70.63	10.5
75	political	340	10,576,884	1,755.65	171.74	10.2
76	joint	114	3,639,152	588.66	59.09	9.8
77	exchange	148	4,988,413	764.22	81.00	9.3
78	economic	255	8,635,910	1,316.73	140.23	9.3
79	country	705	24,593,575	3,640.38	399.34	9.1
80	military	203	7,212,570	1,048.22	117.11	8.9
81	express	166	5,888,528	857.17	95.61	8.9
82	climate	152	5,410,401	784.88	87.85	8.8
83	united	343	12,884,936	1,771.14	209.22	8.4
84	visit	379	14,263,341	1,957.03	231.60	8.4
85	situation	229	8,638,546	1,182.48	140.27	8.4
86	promote	163	6,201,295	841.68	100.69	8.3
87	participant	115	4,464,618	593.82	72.49	8.1
88	support	883	34,671,852	4,559.51	562.98	8.1

89	border	100	3,893,232	516.37	63.22	8.1
90	threat	104	4,054,357	537.02	65.83	8.1
91	peace	124	4,853,521	640.29	78.81	8.0
92	crime	118	4,712,747	609.31	76.52	7.9
93	sector	145	6,017,674	748.73	97.71	7.6
94	development	483	20,912,084	2,494.05	339.56	7.3
95	conference	190	8,287,398	981.10	134.57	7.2
96	mission	129	5,853,432	666.11	95.04	6.9
97	welcome	138	6,395,285	712.59	103.84	6.8
98	partner	166	7,964,176	857.17	129.32	6.6
99	february	145	7,020,515	748.73	114.00	6.5
100	policy	299	14,779,796	1,543.94	239.99	6.4
101	committee	172	8,484,842	888.15	137.77	6.4
102	format	105	5,229,009	542.18	84.91	6.3
103	expert	118	6,109,151	609.31	99.20	6.1
104	meet	316	16,574,764	1,631.72	269.13	6.0
105	war	276	15,138,626	1,425.17	245.81	5.8
106	address	223	12,357,429	1,151.50	200.65	5.7
107	sea	105	5,977,706	542.18	97.06	5.5
108	aim	109	6,212,881	562.84	100.88	5.5
109	session	109	6,370,158	562.84	103.44	5.4
110	current	247	14,517,543	1,275.42	235.73	5.4
111	trade	136	8,646,270	702.26	140.39	5.0
112	commission	103	6,610,193	531.86	107.33	4.9
113	global	147	10,440,638	759.06	169.53	4.5
114	state	617	44,001,914	3,185.98	714.48	4.5
115	law	274	19,701,231	1,414.84	319.90	4.4
116	director	158	11,380,631	815.86	184.79	4.4
117	challenge	157	11,492,203	810.70	186.60	4.3
118	member	344	26,509,604	1,776.30	430.45	4.1
119	president	178	13,875,749	919.13	225.31	4.1
120	particular	107	8,535,368	552.51	138.59	4.0
121	energy	166	13,427,889	857.17	218.03	3.9
122	november	103	8,456,211	531.86	137.31	3.9
123	language	130	10,855,724	671.28	176.27	3.8
124	hold	269	22,548,789	1,389.03	366.13	3.8
125	continue	244	20,527,710	1,259.93	333.32	3.8
126	national	243	20,976,449	1,254.77	340.60	3.7
127	effort	118	10,477,401	609.31	170.13	3.6
128	decision	124	11,057,843	640.29	179.55	3.6
129	ensure	107	9,735,401	552.51	158.08	3.5
130	opportunity	144	13,262,098	743.57	215.34	3.4
131	medium	124	11,872,350	640.29	192.78	3.3
132	human	167	17,134,010	862.33	278.21	3.1
133	march	113	11,639,513	583.49	189.00	3.1
134	society	104	10,773,581	537.02	174.94	3.1
135	view	170	18,045,402	877.82	293.01	3.0
136	interest	122	13,384,277	629.97	217.33	2.9
137	order	215	24,802,917	1,110.19	402.74	2.8
138	event	188	21,829,403	970.77	354.45	2.7

139	future	124	14,774,200	640.29	239.90	2.7
140	rule	100	12,005,312	516.37	194.94	2.6
141	important	152	18,637,375	784.88	302.62	2.6
142	information	263	32,550,269	1,358.04	528.53	2.6
143	role	100	12,645,221	516.37	205.33	2.5
144	close	139	17,637,380	717.75	286.39	2.5
145	issue	210	27,581,421	1,084.37	447.85	2.4
146	field	106	14,291,622	547.35	232.06	2.4
147	technology	118	15,972,296	609.31	259.35	2.3
148	general	113	16,101,873	583.49	261.45	2.2
149	head	129	19,022,912	666.11	308.88	2.2
150	part	264	39,215,936	1,363.21	636.77	2.1
151	note	133	19,748,491	686.77	320.67	2.1
152	provide	285	43,074,594	1,471.64	699.42	2.1
153	present	125	18,953,155	645.46	307.75	2.1
154	action	103	15,744,069	531.86	255.64	2.1
155	project	167	26,893,646	862.33	436.68	2.0
156	force	101	16,613,252	521.53	269.76	1.9
157	plan	139	23,873,334	717.75	387.64	1.8
158	develop	106	18,272,032	547.35	296.69	1.8
159	include	346	60,125,985	1,786.63	976.29	1.8
160	public	131	23,882,320	676.44	387.79	1.7
161	process	125	22,948,846	645.46	372.63	1.7
162	business	161	30,245,087	831.35	491.10	1.7
163	government	122	24,282,531	629.97	394.29	1.6
164	take	398	80,304,577	2,055.14	1,303.94	1.6
165	call	159	38,076,642	821.02	618.27	1.3
166	work	363	87,639,526	1,874.41	1,423.04	1.3
167	right	164	40,490,919	846.84	657.47	1.3
168	place	140	37,839,255	722.91	614.41	1.2
169	world	143	40,141,538	738.40	651.80	1.1
170	change	126	38,751,448	650.62	629.22	1.0
171	year	298	91,960,199	1,538.77	1,493.20	1.0
172	high	108	39,567,260	557.68	642.47	0.9
173	service	104	38,562,742	537.02	626.16	0.9
174	two	138	59,839,401	712.59	971.64	0.7
175	people	139	68,244,095	717.75	1,108.11	0.6
176	new	178	95,531,720	919.13	1,551.19	0.6
177	make	183	118,771,095	944.95	1,928.54	0.5
178	time	125	101,215,485	645.46	1,643.48	0.4

Table 3 **Keywords of the Investment and Development Agency subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	liaa	187	421	2,637.15	< 0.01	2,620.2
2	latvia	603	142,310	8,503.74	2.31	2,568.8
3	latvian	323	73,965	4,555.07	1.20	2,070.0
4	riga	96	72,183	1,353.83	1.17	623.8

5	elwind	39	65	549.99	< 0.01	550.4
6	sia	46	55,465	648.71	0.90	341.8
7	michelin	40	107,292	564.10	1.74	206.1
8	fintech	52	175,938	733.32	2.86	190.4
9	eur	82	330,375	1,156.40	5.36	181.9
10	baltic	56	220,674	789.73	3.58	172.5
11	norwegian	91	538,187	1,283.32	8.74	131.9
12	norway	101	721,265	1,424.34	11.71	112.1
13	entrepreneur	104	1,132,995	1,466.65	18.40	75.7
14	cooperation	150	1,715,762	2,115.36	27.86	73.3
15	startup	87	1,017,995	1,226.91	16.53	70.0
16	export	220	2,741,793	3,102.52	44.52	68.2
17	offshore	47	728,353	662.81	11.83	51.8
18	innovation	157	3,618,808	2,214.07	58.76	37.1
19	investment	274	6,845,641	3,864.05	111.16	34.5
20	netherlands	35	925,003	493.58	15.02	30.9
21	euro	41	1,151,370	578.20	18.70	29.4
22	accessible	61	2,032,502	860.25	33.00	25.3
23	tourism	45	1,512,251	634.61	24.56	24.9
24	contrast	72	2,735,788	1,015.37	44.42	22.4
25	korea	47	1,801,543	662.81	29.25	21.9
26	attract	69	2,710,298	973.06	44.01	21.6
27	organise	38	1,472,052	535.89	23.90	21.6
28	israel	95	4,170,747	1,339.73	67.72	19.5
29	programme	108	4,877,532	1,523.06	79.20	19.0
30	partner	172	7,964,176	2,425.61	129.32	18.6
31	representative	74	3,829,222	1,043.58	62.18	16.5
32	investor	54	3,002,984	761.53	48.76	15.3
33	sector	103	6,017,674	1,452.55	97.71	14.7
34	germany	60	3,702,577	846.14	60.12	13.9
35	japan	57	3,538,276	803.84	57.45	13.8
36	tech	42	2,625,902	592.30	42.64	13.6
37	wind	81	5,339,635	1,142.29	86.70	13.0
38	agency	119	8,056,434	1,678.18	130.82	12.7
39	innovative	35	2,419,566	493.58	39.29	12.3
40	forum	64	4,510,001	902.55	73.23	12.2
41	eu	41	2,876,621	578.20	46.71	12.1
42	import	38	2,694,964	535.89	43.76	12.0
43	european	87	6,430,293	1,226.91	104.41	11.6
44	market	265	20,354,951	3,737.13	330.51	11.3
45	development	268	20,912,084	3,779.44	339.56	11.1
46	product	264	20,651,075	3,723.03	335.32	11.1
47	company	415	33,803,038	5,852.49	548.88	10.6
48	smart	44	3,736,818	620.50	60.68	10.1
49	australia	60	5,124,747	846.14	83.21	10.1
50	potential	97	8,799,265	1,367.93	142.88	9.5
51	wood	47	4,361,982	662.81	70.83	9.2
52	technology	169	15,972,296	2,383.30	259.35	9.2
53	manufacturer	36	3,361,422	507.69	54.58	9.2
54	funding	46	4,334,252	648.71	70.38	9.1

55	foreign	55	5,422,647	775.63	88.05	8.7
56	restaurant	50	4,957,388	705.12	80.50	8.7
57	business	301	30,245,087	4,244.82	491.10	8.6
58	project	248	26,893,646	3,497.39	436.68	8.0
59	competition	45	4,984,447	634.61	80.93	7.8
60	economic	76	8,635,910	1,071.78	140.23	7.6
61	economy	50	5,799,107	705.12	94.16	7.4
62	promote	51	6,201,295	719.22	100.69	7.1
63	aim	51	6,212,881	719.22	100.88	7.1
64	country	197	24,593,575	2,778.17	399.34	6.9
65	industry	107	13,621,608	1,508.95	221.18	6.8
66	implement	42	5,430,081	592.30	88.17	6.7
67	opportunity	100	13,262,098	1,410.24	215.34	6.5
68	size	78	10,751,756	1,099.99	174.58	6.3
69	europe	41	5,902,881	578.20	95.85	6.0
70	construction	44	6,439,108	620.50	104.55	5.9
71	total	71	10,442,417	1,001.27	169.56	5.9
72	million	105	15,597,494	1,480.75	253.26	5.8
73	expert	41	6,109,151	578.20	99.20	5.8
74	participate	35	5,232,469	493.58	84.96	5.8
75	develop	117	18,272,032	1,649.98	296.69	5.5
76	production	64	10,192,423	902.55	165.50	5.4
77	guide	47	7,599,518	662.81	123.40	5.3
78	energy	82	13,427,889	1,156.40	218.03	5.3
79	solution	71	11,647,461	1,001.27	189.13	5.3
80	centre	44	7,217,285	620.50	117.19	5.3
81	conference	49	8,287,398	691.02	134.57	5.1
82	content	64	10,990,782	902.55	178.46	5.0
83	growth	50	8,612,632	705.12	139.85	5.0
84	important	102	18,637,375	1,438.44	302.62	4.7
85	trade	45	8,646,270	634.61	140.39	4.5
86	director	59	11,380,631	832.04	184.79	4.5
87	significant	39	7,750,137	549.99	125.84	4.3
88	global	52	10,440,638	733.32	169.53	4.3
89	culture	40	8,111,514	564.10	131.71	4.3
90	south	65	13,502,963	916.65	219.25	4.2
91	region	47	9,777,203	662.81	158.76	4.2
92	grant	35	7,446,677	493.58	120.92	4.1
93	support	158	34,671,852	2,228.18	562.98	4.0
94	uk	36	7,940,413	507.69	128.93	3.9
95	sale	49	11,067,797	691.02	179.71	3.8
96	park	48	11,154,605	676.91	181.12	3.7
97	relationship	40	9,326,811	564.10	151.44	3.7
98	visit	61	14,263,341	860.25	231.60	3.7
99	idea	71	16,844,832	1,001.27	273.52	3.7
100	establish	37	8,793,649	521.79	142.79	3.6
101	international	67	16,077,934	944.86	261.06	3.6
102	currently	37	9,182,494	521.79	149.10	3.5
103	green	37	9,257,865	521.79	150.32	3.5
104	financial	39	9,783,959	549.99	158.87	3.4

105	environment	37	9,557,665	521.79	155.19	3.3
106	office	58	15,168,855	817.94	246.30	3.3
107	event	83	21,829,403	1,170.50	354.45	3.3
108	various	43	11,316,108	606.40	183.74	3.3
109	customer	48	12,664,924	676.91	205.65	3.3
110	interest	49	13,384,277	691.02	217.33	3.2
111	fund	35	10,137,291	493.58	164.60	3.0
112	grow	52	15,314,358	733.32	248.67	2.9
113	award	39	11,553,137	549.99	187.59	2.9
114	offer	87	26,302,161	1,226.91	427.08	2.9
115	receive	61	19,261,647	860.25	312.76	2.7
116	quality	44	13,961,107	620.50	226.69	2.7
117	large	80	25,777,889	1,128.19	418.57	2.7
118	reach	37	11,940,404	521.79	193.88	2.7
119	account	35	11,362,774	493.58	184.50	2.7
120	material	40	13,128,071	564.10	213.17	2.6
121	step	40	13,137,208	564.10	213.31	2.6
122	value	53	18,029,398	747.43	292.75	2.5
123	expect	39	13,334,961	549.99	216.53	2.5
124	future	41	14,774,200	578.20	239.90	2.4
125	build	57	21,174,641	803.84	343.82	2.3
126	local	55	20,438,710	775.63	331.87	2.3
127	increase	58	21,742,990	817.94	353.05	2.3
128	research	65	24,378,741	916.65	395.85	2.3
129	world	106	40,141,538	1,494.85	651.80	2.3
130	plan	62	23,873,334	874.35	387.64	2.3
131	say	266	102,641,670	3,751.23	1,666.64	2.3
132	food	43	16,674,047	606.40	270.74	2.2
133	activity	37	14,426,839	521.79	234.26	2.2
134	example	49	19,869,926	691.02	322.64	2.1
135	national	51	20,976,449	719.22	340.60	2.1
136	create	69	29,154,971	973.06	473.40	2.1
137	open	59	25,084,583	832.04	407.31	2.0
138	experience	62	26,578,943	874.35	431.57	2.0
139	meet	37	16,574,764	521.79	269.13	1.9
140	add	58	26,095,501	817.94	423.72	1.9
141	head	42	19,022,912	592.30	308.88	1.9
142	process	49	22,948,846	691.02	372.63	1.9
143	share	38	18,015,702	535.89	292.53	1.8
144	provide	89	43,074,594	1,255.11	699.42	1.8
145	present	39	18,953,155	549.99	307.75	1.8
146	year	188	91,960,199	2,651.25	1,493.20	1.8
147	story	40	19,879,278	564.10	322.79	1.7
148	work	162	87,639,526	2,284.59	1,423.04	1.6
149	service	71	38,562,742	1,001.27	626.16	1.6
150	high	72	39,567,260	1,015.37	642.47	1.6
151	note	35	19,748,491	493.58	320.67	1.5
152	new	166	95,531,720	2,341.00	1,551.19	1.5
153	information	56	32,550,269	789.73	528.53	1.5
154	available	36	21,082,219	507.69	342.32	1.5

155	start	62	38,049,026	874.35	617.82	1.4
156	university	38	23,325,583	535.89	378.75	1.4
157	team	45	28,469,255	634.61	462.27	1.4
158	lead	44	28,107,655	620.50	456.40	1.4
159	great	63	40,775,470	888.45	662.09	1.3
160	place	58	37,839,255	817.94	614.41	1.3
161	area	47	31,212,115	662.81	506.81	1.3
162	good	132	87,990,830	1,861.51	1,428.75	1.3
163	different	41	27,743,725	578.20	450.49	1.3
164	help	65	44,180,801	916.65	717.38	1.3
165	result	39	26,661,586	549.99	432.92	1.3
166	small	36	24,784,935	507.69	402.44	1.3
167	take	109	80,304,577	1,537.16	1,303.94	1.2
168	group	42	31,805,881	592.30	516.45	1.1
169	member	35	26,509,604	493.58	430.45	1.1
170	state	58	44,001,914	817.94	714.48	1.1
171	point	40	32,536,832	564.10	528.32	1.1
172	part	47	39,215,936	662.81	636.77	1.0
173	include	72	60,125,985	1,015.37	976.29	1.0
174	time	107	101,215,485	1,508.95	1,643.48	0.9
175	thing	40	39,761,502	564.10	645.63	0.9
176	see	77	78,270,677	1,085.88	1,270.92	0.9
177	people	67	68,244,095	944.86	1,108.11	0.9
178	make	116	118,771,095	1,635.88	1,928.54	0.8
179	show	41	43,849,901	578.20	712.01	0.8
180	system	35	38,403,657	493.58	623.58	0.8
181	come	47	59,588,537	662.81	967.57	0.7
182	think	41	52,200,595	578.20	847.60	0.7
183	two	47	59,839,401	662.81	971.64	0.7
184	know	51	66,908,353	719.22	1,086.42	0.7
185	use	91	124,155,925	1,283.32	2,015.98	0.6
186	go	54	85,415,797	761.53	1,386.94	0.5
187	get	61	100,628,873	860.25	1,633.96	0.5
188	find	35	57,839,721	493.58	939.17	0.5
189	like	43	84,722,531	606.40	1,375.68	0.4
190	just	38	75,329,968	535.89	1,223.17	0.4

Table 4 **Keywords of the Constitutional Court of Latvia subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	334	142,310	5,659.39	2.31	1,709.7
2	contested	241	113,180	4,083.57	1.84	1,439.4
3	laviņš	66	5	1,118.32	< 0.01	1,119.2
4	saeima	67	1,993	1,135.27	0.03	1,100.6
5	aldis	66	6,691	1,118.32	0.11	1,009.6
6	constitutional	943	1,268,581	15,978.45	20.60	739.8
7	latvian	84	73,965	1,423.32	1.20	647.1
8	lithuania	34	201,362	576.11	3.27	135.2
9	norm	145	1,065,245	2,456.92	17.30	134.3

10	legislator	68	500,513	1,152.21	8.13	126.4
11	constitution	198	2,133,150	3,354.97	34.64	94.2
12	judgement	56	575,877	948.88	9.35	91.8
13	gambling	51	574,381	864.16	9.33	83.8
14	court	1,224	15,645,613	20,739.79	254.04	81.3
15	paragraph	117	1,645,854	1,982.48	26.72	71.5
16	applicant	142	2,098,941	2,406.09	34.08	68.6
17	republic	130	2,254,584	2,202.76	36.61	58.6
18	provision	193	3,466,406	3,270.24	56.29	57.1
19	initiate	81	1,509,821	1,372.49	24.52	53.8
20	proceeding	78	1,519,627	1,321.65	24.67	51.5
21	judicial	53	1,042,891	898.05	16.93	50.1
22	european	308	6,430,293	5,218.84	104.41	49.5
23	rights	77	1,586,319	1,304.71	25.76	48.8
24	fundamental	102	2,302,230	1,728.32	37.38	45.1
25	citizenship	33	730,499	559.16	11.86	43.6
26	legitimate	44	1,021,600	745.55	16.59	42.4
27	recognise	67	1,676,008	1,135.27	27.21	40.3
28	supreme	96	2,481,028	1,626.65	40.29	39.4
29	binding	32	826,904	542.22	13.43	37.7
30	consequently	33	855,908	559.16	13.90	37.6
31	cooperation	62	1,715,762	1,050.54	27.86	36.4
32	netherlands	34	925,003	576.11	15.02	36.0
33	legal	252	7,411,431	4,269.96	120.34	35.2
34	comply	45	1,312,706	762.49	21.32	34.2
35	sentence	97	3,306,938	1,643.59	53.70	30.1
36	ruling	41	1,366,844	694.72	22.19	30.0
37	obligation	51	1,742,181	864.16	28.29	29.5
38	judge	168	6,029,478	2,846.64	97.90	28.8
39	administrative	55	1,938,948	931.93	31.48	28.7
40	restriction	63	2,250,846	1,067.49	36.55	28.5
41	dialogue	50	1,777,233	847.21	28.86	28.4
42	execution	40	1,448,938	677.77	23.53	27.7
43	clause	32	1,204,155	542.22	19.55	26.4
44	democracy	51	2,013,064	864.16	32.69	25.7
45	accordance	31	1,253,651	525.27	20.36	24.6
46	restrict	36	1,471,780	609.99	23.90	24.5
47	govern	36	1,477,627	609.99	23.99	24.4
48	pension	34	1,405,801	576.11	22.83	24.2
49	conclude	70	2,966,661	1,186.10	48.17	24.1
50	justice	130	5,665,933	2,202.76	92.00	23.7
51	compliance	48	2,056,184	813.32	33.39	23.7
52	law	447	19,701,231	7,574.09	319.90	23.6
53	electricity	43	1,880,243	728.60	30.53	23.1
54	regulation	93	4,422,268	1,575.82	71.81	21.7
55	cabinet	34	1,596,844	576.11	25.93	21.4
56	principle	90	4,547,697	1,524.98	73.84	20.4
57	union	134	6,807,305	2,270.53	110.53	20.4
58	institution	99	5,166,456	1,677.48	83.89	19.8
59	procedure	93	4,930,889	1,575.82	80.07	19.5

60	identity	64	3,489,992	1,084.43	56.67	18.8
61	strengthen	37	2,032,221	626.94	33.00	18.5
62	conference	144	8,287,398	2,439.97	134.57	18.0
63	protection	104	6,054,300	1,762.20	98.31	17.8
64	territory	42	2,477,870	711.66	40.23	17.3
65	framework	49	3,192,381	830.27	51.84	15.7
66	assess	45	2,974,590	762.49	48.30	15.5
67	programme	69	4,877,532	1,169.15	79.20	14.6
68	amendment	32	2,361,428	542.22	38.34	13.8
69	permanent	31	2,298,530	525.27	37.32	13.7
70	democratic	37	2,772,832	626.94	45.02	13.6
71	case	373	29,768,759	6,320.21	483.37	13.1
72	submit	63	5,088,184	1,067.49	82.62	12.8
73	section	167	13,706,973	2,829.69	222.57	12.7
74	basis	64	5,303,026	1,084.43	86.11	12.5
75	language	126	10,855,724	2,134.98	176.27	12.0
76	establish	97	8,793,649	1,643.59	142.79	11.4
77	circumstance	33	3,027,310	559.16	49.16	11.2
78	person	181	17,106,414	3,066.91	277.76	11.0
79	adopt	41	3,895,621	694.72	63.25	10.8
80	article	176	17,506,660	2,982.19	284.26	10.5
81	president	135	13,875,749	2,287.48	225.31	10.1
82	importance	33	3,558,156	559.16	57.78	9.5
83	right	357	40,490,919	6,049.10	657.47	9.2
84	ensure	85	9,735,401	1,440.26	158.08	9.1
85	aim	54	6,212,881	914.99	100.88	9.0
86	education	121	14,175,926	2,050.26	230.18	8.9
87	criminal	32	3,756,044	542.22	60.99	8.8
88	rule	101	12,005,312	1,711.37	194.94	8.7
89	permit	33	3,929,051	559.16	63.80	8.6
90	europe	49	5,902,881	830.27	95.85	8.6
91	citizen	36	4,431,885	609.99	71.96	8.4
92	regard	78	9,706,253	1,321.6	157.60	8.3
93	implement	42	5,430,081	711.66	88.17	8.0
94	panel	39	5,084,164	660.83	82.55	7.9
95	official	73	9,873,593	1,236.93	160.32	7.7
96	fair	35	4,877,893	593.05	79.20	7.4
97	participate	37	5,232,469	626.94	84.96	7.3
98	freedom	30	4,239,196	508.33	68.83	7.3
99	income	35	4,973,948	593.05	80.76	7.3
100	human	117	17,134,010	1,982.48	278.21	7.1
101	private	60	8,975,789	1,016.66	145.74	6.9
102	international	107	16,077,934	1,813.04	261.06	6.9
103	fee	34	5,138,548	576.11	83.44	6.8
104	port	30	4,543,572	508.33	73.78	6.8
105	council	55	8,818,038	931.93	143.18	6.5
106	protect	47	7,848,094	796.38	127.43	6.2
107	application	90	15,625,895	1,524.98	253.72	6.0
108	commission	38	6,610,193	643.88	107.33	6.0
109	value	102	18,029,398	1,728.32	292.75	5.9

110	property	70	12,383,010	1,186.10	201.07	5.9
111	state	234	44,001,914	3,964.96	714.48	5.5
112	grant	39	7,446,677	660.83	120.92	5.4
113	decision	57	11,057,843	965.82	179.55	5.4
114	national	107	20,976,449	1,813.04	340.60	5.3
115	prepare	37	7,496,205	626.94	121.72	5.1
116	request	42	8,752,050	711.66	142.11	5.0
117	discuss	39	8,488,191	660.83	137.83	4.8
118	society	49	10,773,581	830.27	174.94	4.7
119	february	32	7,020,515	542.22	114.00	4.7
120	discussion	32	7,048,110	542.22	114.44	4.7
121	states	44	9,755,332	745.55	158.40	4.7
122	authority	32	7,454,371	542.22	121.04	4.5
123	tax	43	10,130,594	728.60	164.50	4.4
124	land	53	12,626,447	898.05	205.02	4.4
125	relate	36	8,680,412	609.99	140.95	4.3
126	meeting	41	10,408,510	694.72	169.01	4.1
127	teacher	30	7,633,671	508.33	123.95	4.1
128	decide	42	10,891,968	711.66	176.86	4.0
129	september	35	9,082,639	593.05	147.48	4.0
130	member	97	26,509,604	1,643.59	430.45	3.8
131	january	31	8,503,090	525.27	138.07	3.8
132	apply	42	12,027,221	711.66	195.29	3.6
133	issue	95	27,581,421	1,609.71	447.85	3.6
134	local	70	20,438,710	1,186.10	331.87	3.6
135	role	43	12,645,221	728.60	205.33	3.5
136	period	39	11,558,150	660.83	187.67	3.5
137	interest	44	13,384,277	745.55	217.33	3.4
138	certain	30	9,177,649	508.33	149.02	3.4
139	government	77	24,282,531	1,304.71	394.29	3.3
140	general	51	16,101,873	864.16	261.45	3.3
141	office	48	15,168,855	813.32	246.30	3.3
142	force	52	16,613,252	881.10	269.76	3.3
143	security	37	12,076,799	626.94	196.10	3.2
144	study	75	25,042,023	1,270.82	406.62	3.1
145	fact	52	17,571,900	881.10	285.32	3.1
146	account	33	11,362,774	559.16	184.50	3.0
147	accord	40	14,164,414	677.77	229.99	2.9
148	activity	40	14,426,839	677.77	234.26	2.9
149	provide	116	43,074,594	1,965.54	699.42	2.8
150	development	53	20,912,084	898.05	339.56	2.6
151	high	95	39,567,260	1,609.71	642.47	2.5
152	date	33	14,023,045	559.16	227.70	2.4
153	order	58	24,802,917	982.77	402.74	2.4
154	share	41	18,015,702	694.72	292.53	2.4
155	important	42	18,637,375	711.66	302.62	2.3
156	public	53	23,882,320	898.05	387.79	2.3
157	hold	50	22,548,789	847.21	366.13	2.3
158	review	31	15,230,691	525.27	247.31	2.1
159	social	33	17,158,297	559.16	278.61	2.0

160	party	31	16,169,201	525.27	262.55	2.0
161	country	43	24,593,575	728.60	399.34	1.8
162	pay	30	18,433,125	508.33	299.31	1.7
163	three	43	29,084,971	728.60	472.27	1.5
164	continue	30	20,527,710	508.33	333.32	1.5
165	allow	34	23,484,330	576.11	381.33	1.5
166	power	34	23,587,415	576.11	383.00	1.5
167	system	52	38,403,657	881.10	623.58	1.4
168	follow	39	31,170,319	660.83	506.13	1.3
169	include	71	60,125,985	1,203.04	976.29	1.2
170	part	43	39,215,936	728.60	636.77	1.1
171	year	99	91,960,199	1,677.48	1,493.20	1.1
172	work	94	87,639,526	1,592.76	1,423.04	1.1
173	world	43	40,141,538	728.60	651.80	1.1
174	own	34	33,100,746	576.11	537.47	1.1
175	give	50	53,722,522	847.21	872.32	1.0
176	support	31	34,671,852	525.27	562.98	0.9
177	take	71	80,304,577	1,203.04	1,303.94	0.9
178	place	33	37,839,255	559.16	614.41	0.9
179	time	79	101,215,485	1,338.60	1,643.48	0.8
180	two	41	59,839,401	694.72	971.64	0.7
181	use	83	124,155,925	1,406.37	2,015.98	0.7
182	new	46	95,531,720	779.44	1,551.19	0.5
183	make	35	118,771,095	593.05	1,928.54	0.3

Table 5 **Keywords of the missionLatvia subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvian	166	73,965	7,743.62	1.20	3,518.7
2	latvia	228	142,310	10,635.82	2.31	3,212.8
3	latvians	49	7,265	2,285.77	0.12	2,045.5
4	riga	76	72,183	3,545.27	1.17	1,632.7
5	spirulina	22	21,602	1,026.26	0.35	760.5
6	kokle	16	223	746.37	< 0.01	744.7
7	printful	16	773	746.37	0.01	738.1
8	rainis	14	421	653.08	< 0.01	649.6
9	betriton	11	0	513.13	0.00	514.1
10	biomedicine	15	23,489	699.72	0.38	507.3
11	baltics	12	12,437	559.78	0.20	466.6
12	nouveau	12	68,697	559.78	1.12	265.1
13	baltic	17	220,674	793.02	3.58	173.2
14	eur	19	330,375	886.32	5.36	139.4
15	startup	52	1,017,995	2,425.71	16.53	138.4
16	bog	12	189,743	559.78	3.08	137.4
17	unicorn	12	196,140	559.78	3.18	134.0
18	ict	16	315,156	746.37	5.12	122.2
19	choir	19	675,504	886.32	10.97	74.1
20	easter	15	715,847	699.72	11.62	55.5
21	mushroom	13	624,912	606.43	10.15	54.5

22	egg	40	2,632,516	1,865.93	42.75	42.7
23	architectural	12	862,471	559.78	14.00	37.4
24	globally	11	825,853	513.13	13.41	35.7
25	celebration	21	1,881,154	979.61	30.55	31.1
26	tradition	32	3,201,111	1,492.75	51.98	28.2
27	opera	13	1,343,853	606.43	21.82	26.6
28	smart	35	3,736,818	1,632.69	60.68	26.5
29	innovation	33	3,618,808	1,539.39	58.76	25.8
30	heritage	19	2,240,372	886.32	36.38	23.7
31	export	22	2,741,793	1,026.26	44.52	22.6
32	palace	11	1,348,267	513.13	21.89	22.5
33	wooden	10	1,256,984	466.48	20.41	21.8
34	architecture	20	2,679,882	932.97	43.51	21.0
35	investor	22	3,002,984	1,026.26	48.76	20.6
36	cultural	31	4,633,119	1,446.10	75.23	19.0
37	contemporary	19	2,857,492	886.32	46.40	18.7
38	talent	19	2,900,886	886.32	47.10	18.4
39	invest	20	3,141,055	932.97	51.00	18.0
40	castle	12	1,914,914	559.78	31.09	17.5
41	showcase	10	1,615,605	466.48	26.23	17.2
42	innovative	14	2,419,566	653.08	39.29	16.2
43	diverse	13	2,261,172	606.43	36.72	16.1
44	ancient	18	3,284,550	839.67	53.33	15.5
45	unique	37	6,914,602	1,725.99	112.28	15.2
46	folk	18	3,364,210	839.67	54.63	15.1
47	blend	10	1,866,937	466.48	30.31	14.9
48	symbol	11	2,140,624	513.13	34.76	14.4
49	exhibition	13	2,584,325	606.43	41.96	14.1
50	winter	23	4,646,809	1,072.91	75.45	14.0

Table 6 **Keywords of the Ministry of Defence subcorpus**

Word		Absolute Frequency		Frequency per million		Score
		Focus	Reference	Focus	Reference	
1	latvia	489	142,310	10,258.02	2.31	3,098.7
2	mūrniece	82	22	1,720.16	< 0.01	1,720.5
3	latvian	136	73,965	2,852.95	1.20	1,296.7
4	ināra	43	226	902.03	< 0.01	899.7
5	sprūds	37	14	776.17	< 0.01	777.0
6	naf	49	20,739	1,027.90	0.34	769.7
7	defence	938	1,686,029	19,676.95	27.38	693.5
8	saeima	32	1,993	671.28	0.03	651.2
9	skraučs	27	0	566.39	0.00	567.4
10	andris	27	5,805	566.39	0.09	518.5
11	baltic	99	220,674	2,076.78	3.58	453.3
12	nato	252	665,102	5,286.34	10.80	448.1
13	forces	150	481,386	3,146.63	7.82	357.0
14	ukraine	227	994,839	4,761.90	16.15	277.7
15	riga	28	72,183	587.37	1.17	270.9
16	cybersecurity	84	412,983	1,762.11	6.71	228.8

17	deterrence	25	83,029	524.44	1.35	223.8
18	procurement	101	661,834	2,118.73	10.75	180.5
19	armed	178	1,299,441	3,734.00	21.10	169.0
20	estonia	28	158,220	587.37	2.57	164.9
21	ministers	38	273,128	797.15	4.43	146.9
22	relations	69	626,425	1,447.45	10.17	129.7
23	ukrainian	58	558,095	1,216.70	9.06	121.0
24	cooperation	156	1,715,762	3,272.50	27.86	113.4
25	ammunition	35	401,827	734.21	6.52	97.7
26	drone	65	912,535	1,363.54	14.82	86.3
27	allied	36	518,183	755.19	8.41	80.3
28	resilience	43	669,914	902.03	10.88	76.0
29	capability	178	3,052,561	3,734.00	49.57	73.9
30	tel	30	538,792	629.33	8.75	64.7
31	artillery	29	525,531	608.35	8.53	63.9
32	affairs	65	1,289,452	1,363.54	20.94	62.2
33	ministry	159	3,346,624	3,335.43	54.34	60.3
34	military	320	7,212,570	6,712.82	117.11	56.8
35	e-mail	68	1,723,243	1,426.47	27.98	49.3
36	roberts	27	696,967	566.39	11.32	46.1
37	mod	39	1,120,161	818.12	18.19	42.7
38	minister	207	6,246,399	4,342.35	101.43	42.4
39	brigade	24	688,278	503.46	11.18	41.4
40	coalition	47	1,413,384	985.95	22.95	41.2
41	invasion	34	1,012,928	713.24	16.45	40.9
42	battalion	24	707,254	503.46	11.48	40.4
43	strengthen	56	2,032,221	1,174.74	33.00	34.6
44	missile	30	1,139,802	629.33	18.51	32.3
45	troop	57	2,229,092	1,195.72	36.19	32.2
46	infrastructure	84	3,566,583	1,762.11	57.91	29.9
47	eu	64	2,876,621	1,342.56	46.71	28.2
48	national	458	20,976,449	9,607.72	340.60	28.1
49	media	68	3,245,117	1,426.47	52.69	26.6
50	crisis	74	3,998,188	1,552.34	64.92	23.6
51	threat	74	4,054,357	1,552.34	65.83	23.2
52	comprehensive	46	2,679,485	964.97	43.51	21.7
53	guard	60	3,516,718	1,258.65	57.10	21.7
54	ally	27	1,601,303	566.39	26.00	21.0
55	combat	36	2,224,611	755.19	36.12	20.4
56	priority	47	2,955,948	985.95	48.00	20.1
57	strategic	43	2,759,455	902.03	44.81	19.7
58	cabinet	25	1,596,844	524.44	25.93	19.5
59	collective	26	1,743,642	545.42	28.31	18.6
60	alliance	30	2,042,155	629.33	33.16	18.5
61	security	173	12,076,799	3,629.12	196.10	18.4
62	implementation	48	3,401,342	1,006.92	55.23	17.9
63	training	153	11,233,902	3,209.57	182.41	17.5
64	framework	44	3,192,381	923.01	51.84	17.5
65	russia	44	3,259,815	923.01	52.93	17.1
66	participation	30	2,234,489	629.33	36.28	16.9

67	defend	35	2,626,511	734.21	42.65	16.8
68	european	81	6,430,293	1,699.18	104.41	16.1
69	personnel	25	1,957,704	524.44	31.79	16.0
70	coordinate	25	2,072,484	524.44	33.65	15.2
71	soldier	36	3,228,293	755.19	52.42	14.2
72	joint	39	3,639,152	818.12	59.09	13.6
73	russian	36	3,476,705	755.19	56.45	13.2
74	approve	37	4,022,310	776.17	65.31	11.7
75	prepare	68	7,496,205	1,426.47	121.72	11.6
76	air	107	12,379,919	2,244.60	201.02	11.1
77	enhance	30	3,463,142	629.33	56.23	11.0
78	command	44	5,331,096	923.01	86.56	10.6
79	implement	44	5,430,081	923.01	88.17	10.4
80	necessary	58	7,183,868	1,216.70	116.65	10.4
81	presence	33	4,071,406	692.26	66.11	10.3
82	equipment	49	6,094,951	1,027.90	98.97	10.3
83	sector	48	6,017,674	1,006.92	97.71	10.2
84	border	31	3,893,232	650.30	63.22	10.1
85	capacity	38	4,832,201	797.15	78.46	10.0
86	funding	34	4,334,252	713.24	70.38	10.0
87	regional	33	4,349,882	692.26	70.63	9.7
88	institution	39	5,166,456	818.12	83.89	9.6
89	programme	36	4,877,532	755.19	79.20	9.4
90	development	153	20,912,084	3,209.57	339.56	9.4
91	draft	25	3,401,281	524.44	55.23	9.3
92	protection	44	6,054,300	923.01	98.31	9.3
93	ensure	70	9,735,401	1,468.43	158.08	9.2
94	construction	46	6,439,108	964.97	104.55	9.2
95	meeting	72	10,408,510	1,510.38	169.01	8.9
96	aim	42	6,212,881	881.06	100.88	8.7
97	force	110	16,613,252	2,307.53	269.76	8.5
98	concept	37	6,028,236	776.17	97.88	7.9
99	contact	71	11,626,945	1,489.41	188.79	7.9
100	states	59	9,755,332	1,237.68	158.40	7.8
101	international	97	16,077,934	2,034.82	261.06	7.8
102	volunteer	25	4,205,109	524.44	68.28	7.6
103	agreement	37	6,377,020	776.17	103.55	7.4
104	contribute	29	4,996,685	608.35	81.13	7.4
105	operation	52	9,099,707	1,090.83	147.76	7.3
106	communication	38	6,640,331	797.15	107.82	7.3
107	initiative	24	4,280,016	503.46	69.50	7.2
108	policy	82	14,779,796	1,720.16	239.99	7.1
109	citizen	24	4,431,885	503.46	71.96	6.9
110	information	174	32,550,269	3,650.09	528.53	6.9
111	critical	29	5,403,401	608.35	87.74	6.9
112	section	72	13,706,973	1,510.38	222.57	6.8
113	support	178	34,671,852	3,734.00	562.98	6.6
114	various	58	11,316,108	1,216.70	183.74	6.6
115	centre	37	7,217,285	776.17	117.19	6.6
116	department	64	12,822,871	1,342.56	208.21	6.4

117	canada	30	6,195,349	629.33	100.60	6.2
118	promote	30	6,201,295	629.33	100.69	6.2
119	vehicle	37	7,692,469	776.17	124.91	6.2
120	mission	28	5,853,432	587.37	95.04	6.1
121	country	116	24,593,575	2,433.40	399.34	6.1
122	battle	26	5,716,229	545.42	92.82	5.8
123	public	107	23,882,320	2,244.60	387.79	5.8
124	purchase	36	8,078,865	755.19	131.18	5.7
125	state	195	44,001,914	4,090.62	714.48	5.7
126	unit	45	10,132,245	943.99	164.52	5.7
127	additional	38	8,612,389	797.15	139.84	5.7
128	army	25	5,662,765	524.44	91.95	5.7
129	investment	30	6,845,641	629.33	111.16	5.6
130	system	165	38,403,657	3,461.30	623.58	5.5
131	committee	36	8,484,842	755.19	137.77	5.4
132	technology	65	15,972,296	1,363.54	259.35	5.2
133	develop	74	18,272,032	1,552.34	296.69	5.2
134	europe	24	5,902,881	503.46	95.85	5.2
135	reach	48	11,940,404	1,006.92	193.88	5.2
136	region	39	9,777,203	818.12	158.76	5.1
137	potential	35	8,799,265	734.21	142.88	5.1
138	task	25	6,432,204	524.44	104.44	5.0
139	council	34	8,818,038	713.24	143.18	5.0
140	union	26	6,807,305	545.42	110.53	4.9
141	train	27	7,073,361	566.39	114.85	4.9
142	join	44	11,647,148	923.01	189.12	4.9
143	society	40	10,773,581	839.10	174.94	4.8
144	war	55	15,138,626	1,153.77	245.81	4.7
145	plan	86	23,873,334	1,804.07	387.64	4.6
146	situation	31	8,638,546	650.30	140.27	4.6
147	project	95	26,893,646	1,992.87	436.68	4.6
148	establish	30	8,793,649	629.33	142.79	4.4
149	january	29	8,503,090	608.35	138.07	4.4
150	supply	30	8,836,245	629.33	143.48	4.4
151	partner	27	7,964,176	566.39	129.32	4.4
152	service	127	38,562,742	2,664.15	626.16	4.3
153	continue	67	20,527,710	1,405.50	333.32	4.2
154	sign	40	12,317,711	839.10	200.01	4.2
155	strategy	26	8,030,529	545.42	130.40	4.2
156	announce	26	8,033,811	545.42	130.45	4.2
157	phone	27	8,471,221	566.39	137.55	4.1
158	education	44	14,175,926	923.01	230.18	4.0
159	law	61	19,701,231	1,279.63	319.90	4.0
160	industry	42	13,621,608	881.06	221.18	4.0
161	million	47	15,597,494	985.95	253.26	3.9
162	agency	24	8,056,434	503.46	130.82	3.8
163	improve	36	12,141,325	755.19	197.14	3.8
164	medium	35	11,872,350	734.21	192.78	3.8
165	agree	26	8,927,457	545.42	144.96	3.7
166	accord	40	14,164,414	839.10	229.99	3.6

167	discuss	24	8,488,191	503.46	137.83	3.6
168	response	27	9,677,542	566.39	157.14	3.6
169	regard	27	9,706,253	566.39	157.60	3.6
170	address	34	12,357,429	713.24	200.65	3.5
171	provide	114	43,074,594	2,391.44	699.42	3.4
172	decision	29	11,057,843	608.35	179.55	3.4
173	environment	25	9,557,665	524.44	155.19	3.4
174	strong	30	11,619,702	629.33	188.67	3.3
175	management	37	14,354,911	776.17	233.09	3.3
176	skill	24	9,320,791	503.46	151.35	3.3
177	common	27	10,704,354	566.39	173.81	3.2
178	base	63	25,160,088	1,321.59	408.54	3.2
179	resource	27	10,796,199	566.39	175.30	3.2
180	focus	36	14,945,835	755.19	242.68	3.1
181	area	75	31,212,115	1,573.32	506.81	3.1
182	united	31	12,884,936	650.30	209.22	3.1
183	main	28	12,255,358	587.37	199.00	2.9
184	increase	47	21,742,990	985.95	353.05	2.8
185	member	57	26,509,604	1,195.72	430.45	2.8
186	activity	31	14,426,839	650.30	234.26	2.8
187	require	44	20,747,024	923.01	336.88	2.7
188	meet	34	16,574,764	713.24	269.13	2.6
189	individual	29	14,270,724	608.35	231.72	2.6
190	group	63	31,805,881	1,321.59	516.45	2.6
191	local	40	20,438,710	839.10	331.87	2.5
192	level	45	23,632,169	943.99	383.73	2.5
193	land	24	12,626,447	503.46	205.02	2.4
194	role	24	12,645,221	503.46	205.33	2.4
195	research	46	24,378,741	964.97	395.85	2.4
196	field	26	14,291,622	545.42	232.06	2.3
197	action	28	15,744,069	587.37	255.64	2.3
198	important	33	18,637,375	692.26	302.62	2.3
199	receive	34	19,261,647	713.24	312.76	2.3
200	order	43	24,802,917	902.03	402.74	2.2
201	government	41	24,282,531	860.08	394.29	2.2
202	current	24	14,517,543	503.46	235.73	2.1
203	head	31	19,022,912	650.30	308.88	2.1
204	event	34	21,829,403	713.24	354.45	2.0
205	general	25	16,101,873	524.44	261.45	2.0
206	available	32	21,082,219	671.28	342.32	2.0
207	create	44	29,154,971	923.01	473.40	1.9
208	number	40	31,769,841	839.10	515.86	1.6
209	part	47	39,215,936	985.95	636.77	1.5
210	include	72	60,125,985	1,510.38	976.29	1.5
211	control	24	20,476,459	503.46	332.49	1.5
212	product	24	20,651,075	503.46	335.32	1.5
213	allow	26	23,484,330	545.42	381.33	1.4
214	process	25	22,948,846	524.44	372.63	1.4
215	community	25	24,685,531	524.44	400.83	1.3
216	company	32	33,803,038	671.28	548.88	1.2

217	year	85	91,960,199	1,783.09	1,493.20	1.2
218	start	35	38,049,026	734.21	617.82	1.2
219	report	26	28,503,999	545.42	462.83	1.2
220	case	25	29,768,759	524.44	483.37	1.1
221	work	73	87,639,526	1,531.36	1,423.04	1.1
222	new	75	95,531,720	1,573.32	1,551.19	1.0
223	become	24	32,738,303	503.46	531.59	0.9
224	help	31	44,180,801	650.30	717.38	0.9
225	take	54	80,304,577	1,132.79	1,303.94	0.9
226	people	40	68,244,095	839.10	1,108.11	0.8
227	use	59	124,155,925	1,237.68	2,015.98	0.6
228	time	37	101,215,485	776.17	1,643.48	0.5
229	make	40	118,771,095	839.10	1,928.54	0.4
230	say	33	102,641,670	692.26	1,666.64	0.4

Appendix 2. The corpus

The corpus has been made available to the University of Latvia on the Sketch Engine software. The corpus is available on https://app.sketchengine.eu/. The titles of the corpora are as follows:

1. Latvian Governmental Web Texts Corpus (the whole corpus)
2. Ministry of Foreign Affairs subcorpus
3. Investment and Development Agency subcorpus
4. Constitutional Court of the Republic of Latvia subcorpus
5. missionLatvia subcorpus
6. Ministry of Defence subcorpus

Appendix 3. Dokumentārā lapa

Maģistra darbs „Latvian Country Image in the Corpus of Governmental Websites” (Latvijas tēls valsts iestāžu tīmekļa vietņu tekstu korpusā) izstrādāts LU Humanitāro zinātņu fakultātē.

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

Autors: Kārlis Zujs 31. 05. 2024.

Rekomendēju/nerekomendēju darbu aizstāvēšanai

Vadītāja: Dr. philol., asoc. prof. Zigrīda Vinčela

Recenzents: prof. Inguna Skadiņa

Studiju metodiķe: Jeļena Sevastjanova

Darbs iesniegts Anglistikas nodaļā 31. 05. 2024.

Darbu pieņēma:

Darbs aizstāvēts maģistra gala pārbaudījuma komisijas sēdē

2024. gada…… jūnijā, prot. Nr. ….., vērtējums ……..

Komisijas sekretāre: