

‘RETURN TO THE INTERNATIONAL FAMILY OF DEMOCRACIES’: KEYNESS FACTOR IN THE INTERNATIONAL SPEECHES OF THE BALTIC PRESIDENTS

LĪGA ROMĀNE-KALNIŅA

University of Latvia, Latvia

Abstract. Presidential speeches as a type of political discourse are aimed not only at the negotiation and construction of the national identity of a nation-state at a local level but also at the representation and shaping of the national identity internationally. The presidents of the Baltic States have represented their individual, collective and regional identities in the international gatherings of world leaders since the restoration of independence of Estonia, Latvia, and Lithuania from the Soviet Union. The current study displays an analysis of how the keyness factor of particular lexical items used in 142 speeches given by the presidents of the Baltic States internationally from 1991 until 2021 helps to identify the tendencies of identity construction and representation, which can then be investigated in detail via a critical analysis of the discursive strategies and linguistic means applied in the speeches. Moreover, the analysis of keyword tendencies across speeches marked by different criteria shows how the process of identity construction as marked by lexical change varies across time and states. The keyness factor points to multiple identities being constructed in the international speeches, where the national identities are constructed most frequently, followed by the common European identity, Baltic regional identity, and global identity. It is also concluded that a common political past is one of the main elements of national and Baltic identities, while shared values such as democracy and cooperation are the main elements of supra-national identities.

Key words: presidential speeches, Baltic States, corpus linguistics, critical discourse analysis, identity, keyness

INTRODUCTION

In a multi-layered and complex system of social realities, drawing and understating boundaries is seen as a way to cope with the ‘chaotic and unstructured’ everyday stimuli that ‘threaten to overwhelm’ the human cognitive system, thus ‘ascribing identities’ to self and others is seen as ‘a natural function of the brain’ (Mole, 2007: 3).

However, social realities can be constructed in and via everyday social practices. One of the notable social practices that is both able and willing to construct and transform social reality is political discourse. Due to the ability of political leaders to address masses of people in order to unite, divide, persuade or direct them to a certain way of thinking or action, political discourse as one of the types of the discourses of power has been the main object of critical research in the last decades.

Moreover, one of the functions of the presidents of the Baltic States as parliamentary republics is to address the local and international audiences to represent and construct an overarching image of the nation-state. While the construction of national identity in the local speeches of the presidents of the Baltic States has been found to aim at the integration and unification of the people of Estonia, Latvia, and Lithuania, as well as at the emphasis on history as an important element of national identities of the states (Romāne-Kalniņa, 2020), presidential rhetoric at the international level has not been discussed in detail.

Thus, the current study provides an overview of the results of a corpus-assisted critical discourse analysis of the speeches by the presidents of the Baltic States (in English) given in international assemblies, conferences, and meetings of the leaders of the European Union, NATO, and United Nations from 1991 to 2021. The study focuses on how keywords and multiword constructions in the speeches point to the construction and representation of group identities – national, regional, and supra-national within various thematic areas being introduced across time and space.

CORPUS-ASSISTED CRITICAL DISCOURSE ANALYSIS

Corpus Linguistics (CL) is defined as ‘a method of analysis which involves collecting large amounts of [authentic] language data in computerised format’ and ‘using computer programmes which can sort, count, and perform statistical tests’ on the collected data in a relatively short period of time and ‘accurately identify patterns that would be difficult for the human eye to spot alone’ (Baker in Hart and Cap, 2014: 211). Critical Discourse Studies (CDS), however, is seen as a multidisciplinary methodological movement; a research paradigm that allows for the combination of various approaches to the critical analysis of the discourses of power, including institutionalised discourses, political discourse, and media discourse. The paradigm is thus defined as ‘a movement which seeks to raise critical consciousness about the discursive dimensions of social problems involving discrimination, disadvantage, and dominance with the aim of contributing to broader emancipatory projects’ (Flowerdew and Richardson, 2017: 372). Although, as the title and the definition suggest, the paradigm is primarily aimed at providing a critical perspective on dominating discourses, all of its types are likewise ‘explicitly or implicitly conducted against a vision of ideal human relations with other humans’ (ibid.: 503). The object of CDS is thus discourse as social practice – a representation of social reality in written, spoken, or multimodal form that ‘assumes a dialectical relationship between particular discursive acts and the situations, institutions

and social structures in which they are embedded' (Wodak, de Cillia, Reisigl and Liebhart, 2009: 9). Discourses are shaped by social realities, and discursive practices allow for shaping these realities in turn – a reciprocal process that demands a detailed multidisciplinary analysis.

The useful synergy of methods of CL and CDS has been established by numerous scholars across recent decades, and 'since mid-2000s, [...] the hybrid form of analysis allows qualitative investigation of quantitative results and give the analysts a much firmer grip on their data' (Baker in Hart and Cap, 2014: 211). This methodological solution has hence been entitled Corpus Assisted Discourse Analysis (CADS) (*ibid.*). The combination is particularly notable in the academic writings of recent years (Mulderriq, 2008; Baker in Hart and Cap, 2014; Haider, 2016; Haider, 2017; Române-Kalniņa, 2020; Kitishat, Kayed and Al-Ajalein, 2020; Kelly, 2020; Shah, 2021; Maglie and Centonze, 2021; Matthews, 2021).

The necessity for this combination of methodologies has arisen on the one hand from the severe critique (Stubbs, 1997; Hammersley, 1997; Widdowson, 1998; Slembrouck, 2001; Billig, 2002; Žagar, 2010; Breeze, 2011) received by Critical Discourse Analysis (CDA), later specified as CDS and its approaches, and on the other hand from the opportunities provided by the development of artificial intelligence tools that allow for computer-based, automated text analysis. Thus, the major criticisms received by the methods of CDS, including subjectivity and researcher bias, lack of reliability of the qualitative analysis, focus on negativity, aim to necessarily find power abuse and mind manipulation in every piece of discourse (Barlett, 2012: 5), may be reduced by providing objective data retrieved from online tools and analysed by the newest solution of digital technology that provide for the 'self-awareness and agency' of linguistic research (Baker, 2006: 11).

It may be noted that while word frequencies allow identification of the prominence and dispersion of a particular word in a given discourse sample, 'collocations statistically identify adjacent words and concordances allow [...] to view the keywords in context in series of concordance lines' (Matthews, 2021: 208). When looking at the significance of specific lexical items – words, phrases and expressions in textual analysis and discourse analysis, it seems reasonable to agree that keywords are 'markers of the aboutness and the style of the text' (Bondi and Scott, 2010: 1). Keywords may be seen as both 'searching tools in text mining and classification' and as 'analytic tools in text interpretation and discourse analysis' (*ibid.*). Keyness, however, refers to the comparatively 'high frequency of words or cluster of words in one corpus when compared with a reference corpus' (Haider, 2016: 64). It is suggested that three types of keywords may be found in a particular corpus of texts, namely, keywords that indicate the aboutness of the text (key words that carry the main themes of the discourse), 'high frequency words which may be the indicators of style' and proper nouns (*ibid.*).

Furthermore, the critical analysis of the text proceeds from the principles of the Discourse-Historical Approach (DHA), analysing the thematic areas in the speeches as indicated by the identified keywords, followed by a more detailed analysis of the discursive strategies used to construct these thematic areas and finally

the linguistic means of realisation of the particular strategies (Wodak et al., 2009). It is argued that five main content-related thematic areas (construction of common political past, construction of common political present and future, construction of *Homo Nationalis* (national spirit, patriotism), construction of common culture and construction of common national body) are found in the political discourse aimed at the construction of national identities (ibid.: 30). Within the thematic areas, the speaker typically exercises ‘a more or less automated’ plans of discursive action (strategies) that may be subdivided into macro and micro levels (ibid.). At the macro level, the strategies function to construct, preserve, transform or de-construct (dismantle) identities, while at the micro level, such functions as ‘singularisation, preservation, autonomisation, assimilation, dissimilation, inclusion and exclusion’ are applied (ibid.: 34). Following the top-down strategy, the final step is to identify and analyse which particular linguistic means are applied in the realisation of these discursive action plans and why, bearing in mind the reciprocal relation between the text and context. Moreover, the DHA allows for the detailed investigation not only of the immediate contextual relations but also for a more detailed analysis of historical context, which is particularly significant in national identity research.

Thus, the current study uses two online corpus tools (Sketch Engine and Voyant Tools) and a corpus analysis software (AntConc) to store, analyse and visualise word and keywords frequencies in a corpus of 142 international speeches by the presidents of the Baltic States from 1991 until 2021 and well as to analyse and visualise the distribution of identified lexical units across various speeches. The corpus has been compiled by downloading the available presidential speeches from the web pages of the presidents of Estonia, Latvia and Lithuania and online archives of the United Nations web page (Online 1–4), marking the files with the criteria for speech classification (state, president, year, occasion, gender), converting the downloaded files into Microsoft Word and Microsoft OneNote documents and uploading them into corpus analysis online tools and AntConc software where the sub-corpora classifying the corpus according to the marked criteria have been created accordingly. The analysis of the identified keywords and phrases is followed by a detailed qualitative analysis of the use of discursive strategies and linguistic means of realisation of these strategies as potentially influenced by extralinguistic factors that constitute the context of the speeches (discussed in Wodak et al., 2009, and Wodak and Mayer, 2016) in order to provide a proper understanding of the circular process of social reality construction (how the discourse is shaped by the discourse agents and situational context and how it aims to shape the social situation in turn).

REPRESENTATIVE FUNCTION IN PRESIDENTIAL SPEECHES OF THE BALTIC NATIONS

It has been noted that the common Baltic identity as a concrete or abstract unity of common values does not exist, but rather that ‘situated on the Baltic Eastern

Seaboard, Estonia, Latvia and Lithuania share the same geopolitical fate' (Duvold, Berglund and Ekman, 2020: 1). While the Baltic States have been ruled by various powers throughout history, the common fate narrative is most often built on the common political experiences in the 20th century, namely, the German and Soviet occupations and subsequent repressions and deportations of the ethnic population of Estonia, Latvia, and Lithuania (Mole, 2012). Having regained independence from the Soviet Union in the early 1990s, the Baltic States joined the EU and NATO in 2004 to a large extent because the political leaders of Estonia, Latvia, and Lithuania 'committed themselves to the EU membership at the time when the Soviet and Russian legacy was felt to be redundant and unwanted' (Duvold et al., 2020: 145). This achievement, however, seems to be based on the shared will of the political leaders and citizens of these states alike after regaining the political status of independent nation-states to 'sever the ties with Russia and to seek close ties with the West where the majority of the Baltic citizens feel they rightly belong' (ibid.). Thus, the simultaneous task of the presidents of the Baltic States across the thirty years of restored independence has been to construct both national and supra-national identities to represent national identity abroad and the supra-national identities at home.

Lucas emphasizes the role of the presidents of the Baltic States who originally represented diaspora naming Valdas Adamkus, Toomas Hendrik Ilves and Vaira Vīķe-Freiberga 'a serendipitous assortment of unlikely leaders' that was one of the reasons for the success of the Baltic States after the collapse of the Soviet Union (2009: 77). Nevertheless, it seems reasonable to state the main responsibility of persuading the international audience to accept the Baltic States in their community was held by the first presidents of the newly democratic Baltic Republics, namely, Lennart Meri, Guntis Ulmanis and Algirdas Brazauskas (see Tables 6–8 in Appendix 2). The responsibility of these presidents has been, as mentioned above, to represent the common supranational values as shared by the Baltic people in the international speeches at the same time protecting the national values, one of which, arguably the most important – was the state language (Tabuns, 1999; Subrenat, 2004). Moreover, it has been observed in public surveys even in the second decade of the 21st century, namely, that 'the countries have different foreign phobias [...] – anti-Semitism in Latvia and Lithuania and Russia and Russification in Estonia and Latvia' (ibid.: 78). Lucas further observes the differences between the Baltic identities and notes that 'Estonia's Nordic-style thrift, openness and careful planning' have led the state to be seen as the wealthiest of the Baltic states, while the identity of Latvia is defined as 'diffuse' (Lucas, 2009: 77).

As to the common Baltic identity, it has already been noted that it is very difficult to define such a common collective identity, but the joining elements are the Baltic Sea, common political history (the role of the victim) and the common membership in the EU and NATO. Hackman even notes that 'there have been different levels of collective identity in the Baltic Sea region [...] – political, social, historical, religious, linguistic and territorial identities, which can be interfering' (Hackman, 1996: 14). Nevertheless, when taking the perspective of international

speeches, the common Baltic identity seems to be a useful construction in order to both share responsibility and the role of victimhood (the common fate).

KEYNESS ANALYSIS IN INTERNATIONAL SPEECHES

As noted above, the corpus consists of 142 international speeches (197,204 words). The first step of the keyword analysis was to extract the general list of keywords and multiword constructions. The selected available reference corpus in the Sketch Engine online corpus tool was the OPUS2 corpus of the English language. The list of keywords with the statistics includes raw frequency (actual frequency of the word in the focus corpus), relative frequency (frequency per million words), highest keyness score (occurrence of a keyword when compared to the reference corpus) and the list of most frequently co-occurring collocations (occurrence +/-3 words, minimum frequency 1) is displayed in Table 1 below:

Table 1 Keywords with the highest keyness Score in the corpus

Keyword	Raw frequency	Relative frequency	Keyness Score	Collocations
Baltic	317	1448.71	307.4	Sea, States, region, Way, three, states, shores, strategy, the, countries
Latvia	912	4167.89	221.2	has, is, supports, in, will, independence, ready, its, people, Estonia
cyber	77	351.89	171	hygiene, attacks, defence, threats, space, warfare, cyber, hybrid, Efforts, domain
Latvian	104	475.29	147.4	language, companies, government, poet, delegation, Rainis, universities, Indian, IT, company
Latvians	30	137.1	128	non-ethnic, deported, involuntary, oppressed, professions, apt, emigration
Estonia	522	2385.57	126.6	has, is, supports, Latvia, in, been, Republic, firmly, believes
Estonians	27	123.39	112	Finns, Slovenian-speaking, younger, 2.24, euro-enthusiasm, russification, Spanish-speaking, Swedes, perished, electronically
Lithuania	483	2207.33	104	has, supports, is, will, Latvia, Poland, in, European, strongly, relations, Russia, Estonia

Keyword	Raw frequency	Relative frequency	Keyness Score	Collocations
thanks	21	95.97	94.6	to, sincere, expressing, full-format, nineteenth, Lalumière, domestic, free-market, visits
Crimea	25	114.25	85.8	annexation, Sevastopol, illegal, annexing, affecting, illegally, annexed, occupation
Estonian	78	356.46	84.8	language, defence, presidency, flag, Riigikogu, guard, border, parliament, people
Lithuanian	59	269.63	80.5	prosecutors, judges, business, ambassadors, Polish, Estonian, while, people
Vilnius	32	146.24	73	Riga, Conference, Chernomyrdin's, symbolized, ease, Senate, rang
transatlantic	42	191.94	70.5	link, partnership, structures, strengthening, perspective, integration,
EFP	15	68.55	69	Battle Groups, numerical, tripwire, equip, contained
peace-keeping	21	95.97	67.9	peacemaking, missions, preventive, operations, guise, diplomacy, strictly, battalion
Covid-19	14	63.98	65	pandemic, Response, amplified, swept, Multilateral
Tallinn	24	109.68	64.8	gather, reveal, Riga, chain, September, small, visited
Soviet	139	635.24	61.1	occupation, Union, former, regime, Nazi, collapse, empire, under, fifty, then, totalitarian
Lithuanians	15	68.55	60.9	Poles, sheltered, centuries-long, coexistence, love, regret, ages

As the table above suggests, the most frequently used keywords not only indicate the theme and the author of the speech, be it common political history, present or future of the Baltic States within international organisations, but also together with the list of collocations point to the type of group identity being constructed with relatively high frequency. Thus, it seems that Latvia is being referenced comparatively more frequently (raw and relative frequencies) than Estonia and Lithuania. The list of collocations to the words 'Latvia, Latvian and Latvians' show the construction of a common political past (deported, oppressed, emigrated, non-ethnic Latvians), common culture (Latvian language, Latvian poet, Rainis), as

well as common political present and future and economic situation (IT, company, delegation). Moreover, the list of collocations that are found most frequently with the word Latvia (has, is, will, supports, ready) points to a frequent use of metonymy, where the state name stands not for the geographical location of the state and not for the total population of the state, but rather for the political leadership of Latvia, for instance, 'Latvia is seriously concerned about the potential use of stockpiles of Syria's chemical weapons' (Andris Bērziņš, 2012). Discursively, metonymic reference is seen as the strategy of reference and nomination aimed at the construction of ingroups and out-groups (Wodak and Koller, 2008: 302).

Similar use of the strategy is found in the speeches by the presidents of Estonia and Lithuania. However, the notable difference is the frequent attribution of additional elements to the referencing of the state name (usually expressed by the use of nouns, adjectives, adverbs or pronouns), rather than only showing the topic of the speech also point to a continuous and repetitive tendency to emphasize particular elements of the representative values of the national or supra-national identities of the Baltic States. It is notable that in the case of Latvia, these elements would be *people*, *language*, and *independence*, as well as the historical link with Estonia (*The last Russian military units departed Latvia and Estonia at the end of August. Thus, the Second World War has ended for the Baltic States* (Guntis Ulmanis, 1994)). In the case of Estonia, similarly to Latvia, the common link that is referenced is the prolonged presence of the Soviet military troops in the countries after the restoration of independence, thus these word combinations are more frequent in the speeches by the first presidents of the states after restored independence.

Moreover, the Estonian identity and its elements seem to be emphasized more frequently than those of the other Baltic States as the collocations attributed to *Estonia*, *Estonian* and *Estonians* characterise the national identity as Scandinavian related (a reference to Finns and Swedes), inclusive part of the common supra-national the EU identity (multilingual and euro-enthusiastic), based on the common political past and the role of victimhood (russification, perished) and emphasis on such elements of national identity as elite (or banal) identity as language, flag, government (Riigikogu), parliament, but also pointing to security as one of the main elements (border, defence, guard), for instance,

On our side of the border there is a positive development, on the Russian side we find a deepening economic and social chaos. The pressure on the Estonian border will thus be increased by economic refugees, organized crime, smuggling of drugs and weapons. (Lennart Meri, 1992)

The aforementioned example illustrates the construction of a common political present and future (Wodak et al., 2009: 31) via the use of topos of threat and topos of comparison and the strategy of positive self and negative other presentation and the strategy of 'portrayal in black and white' (ibid.: 39) that is emphasized with such linguistic means as the use of pronouns (*our* – Estonian, *we* – Estonians, the international community, except Russia), semantic elements and referential

assimilation and dissimilation emphasizing our good things (positive development) and their bad things (deepening economic and social chaos, economic refugees, organized crime, smuggling of drugs and weapons). The aforementioned topoi, specifically the topos of threat, are typically used to encourage united action against the named threat. Moreover, it has been noted in the scholarly literature that the frequent use of historical references (construction of common political past, topos of history) by the Baltic leaders in international speeches is aimed at the 'western sympathy' because 'the willingness of the EC to open up the possibility of economic and political convergence was driven by collective guilt and responsibility for the consequences of the Munich pact, the Molotov-Ribbentrop pact, and the Yalta agreement' (Lasas, 2008a: 367). Thus, since the 'guilt factor on part of the Western states for the consequences of the Second World War in the Baltic States' was both recognised and used by the Baltic political leaders to 'leverage their unique historical and geopolitical position in order to accelerate their integration westward' (ibid.: 366), the use of the aforementioned linguistic techniques and argumentation strategies (topoi) may be considered as linguistic manipulation.

It is likewise observed that while the presidents of Estonia and Latvia emphasize Russia's threatening presence in the first years of restoration of their independence, the presidents of Lithuania seem to euphemize the common political history and the victim role, constructing an inclusive bilateral cooperation, for instance, '*A fresh page is being turned in the relations between our two countries*' (Algirdas Brazauskas, 1993). The example illustrates how the application of one of the most frequently occurring word combinations, namely, an inclusive pronominal reference (*our countries*) in the corpora together with a metaphorical expression FORGETTING AND FORGIVING THE PAST IS TURNING A FRESH PAGE that is part of a conceptual metaphor HISTORY IS BOOK is used in the construction of inclusive bilateral relationship via an emphasis on positive national continuation and cooperation. It seems interesting to note that while the bilateral relationship with Russia seems to be portrayed in a positive light both linguistically and discursively, the use of the same word combination, however, in singular form (*our country*) is used to portray negative associations with Russia in the speeches by the presidents of Latvia, especially in the first years of the restored independence, for instance,

In the light of the *continued presence of these forces*, foreign investment in *Latvia*, which is necessary for the *development of our small country*, has been discouraged. In addition, the *military forces of the Russian Federation* have done *substantial damage to Latvia's environment*. (Guntis Ulmanis, 1993)

Thus, the example displays the use of the word combination *our country* referring exclusively to the country of Latvians, together with another frequently used word, *small*, referring both to the geographical size of the country and the population (frequently referenced in the speeches by the presidents of Estonia and Lithuania as well) to illustrate the aforementioned victim role of Latvia as a small versus Russian

Federation as a large country whose threatening presence in Latvia damages its development in numerous areas. The statement invites international action on the one hand and displays the use of the strategy of justification on the other, as the references to the Russian military forces in the country are used to justify the underdeveloped state of Latvian businesses and environmental issues.

It is noted that the presidents of Estonia use the reference (our country/ies) comparatively less frequently (6 occurrences, 84,11 per million words) than the presidents of Latvia (32 occurrences, 392,79 per million words) and Lithuania (27 occurrences, 452,93 per million words); however, it does occur when the presidents seem to be emphasizing particular points of interest in the speeches (that have also been marked in lists of keywords), for instance, security:

Estonia's history has been turbulent, marked by hopes raised and hopes shattered. Now that our independence has been restored, our principal ambition is to obtain for *our country, for our people, for our identity, the security* of which we were deprived in the past. Security has many facets, all equally important. (Lennart Meri, 1997)

The example illustrates the use of exclusive we that refers to the citizens of Estonia, as it is combined with such key words (words carrying the meaning of the particular speech) and keywords (frequently occurring words when compared to a general English language corpus) as *people, country, independence, security* and the use of the passive voice to indirectly refer to the existing threatening other that has 'shattered hopes', 'deprived Estonians of security' as well as the people who have helped to restore Estonia's independence.

While the passive voice seems to be a frequently used form in political speeches in general, to avoid naming the agent of the action, it is likewise notable that in the speeches by the presidents of the Baltic States, the passive voice is often used when referring to the common political past. The active voice, on the other hand, is used to characterise the present and future as well as (within topos of comparison) to compare particular attributes or elements of national uniqueness, for instance, 'If we look at *where we were* in 2004 and *where we are* today, then the *numbers speak for themselves*: Estonians are 2,24 times richer than when *we joined*' (Kaljulaid, 2017). The example illustrates not only the use of exclusive we (a reference to Estonians and Estonia metonymically) and inclusive we (*if we all* in this meeting room *look*) as well as the personification metaphor *numbers speak for themselves*, but also elements of comparison: the difference between past and present (*where we were, where we are*), numerical reference (topos of numbers) and comparative adjective (*richer*) to achieve the effect of emphasis on positive national uniqueness and continuation (discursive strategy of perpetuation).

Furthermore, another corpus tool useful in the investigation of referential tendencies in the speeches is the list of N-grams (the most frequently occurring word combinations), as it more explicitly illustrates the thematic areas (even references to identities) constructed in the speeches. Table 2 below displays the list of N-grams in the corpus of the current research:

Table 2 N-grams in the corpus

N-gram	Raw frequency	Relative Frequency	Keyness
Ladies and Gentlemen	137	626.1	391.106
the Baltic Sea	61	278.77	202.52
the Baltic States	48	219.36	164.129
Human Rights Council	39	178.23	143.502
of the UN	108	493.57	123.268
Baltic Sea region	28	127.96	121.862
the European Union's	22	100.54	101.541
Millennium Development Goals	30	137.1	98.29
the UN Charter	24	109.68	97.854
European Union and NATO	22	100.54	93.8
people of Latvia	20	91.4	91.45
Sustainable Development Goals	19	86.83	87.831
the Council of Europe	80	365.6	77.678
of Latvia and	17	77.69	76.982
Thank you for your attention	19	86.83	76.249
the Border Agreement	16	73.12	74.121
the Baltic Way	16	73.12	74.121
the Soviet Union	41	187.37	71.742
of the world's	15	68.55	69.551
Latvia is ready	15	68.55	69.551
the three Baltic	15	68.55	66.906
future of Europe	23	105.11	66.265

The list of word combinations occurring with relatively high frequency in the corpus displays first of all the following factors: the type of texts is public speeches (the use of address form *ladies and gentlemen*, ending of the speech *thank you for your attention*), the speakers are representatives of the Baltic States (references to the Baltic Sea region, Baltic States, Baltic Way), the speeches are given to international audience (references to international organisations and international laws) and finally that the presidents of Latvia seem to reference Latvia's national identity more frequently and explicitly in these speeches, and that the common political past (the Soviet occupation of the Baltic States) is one of the most important elements in constructing not only the common political past and the victimhood of the Baltic States, but also common political present (security issues, justification of issues that are portrayed as caused by the historical

deprivation of independence) and common political future within the international family of democracies united under the rule of law, for instance,

One might ask *what Latvia expects from Europe?* Within the fold of the European Union, *Latvians see the opportunity to irreversibly reinforce their sovereignty, and to maintain their identity, culture and language. The European Union is a large family of European, democratic nations, where the interests of each and every nation are respected, and where solidarity is the cornerstone of co-operation. The people of Latvia see security and stability as the pillars of their development.* (Vaira Viķe-Freiberga, 2003)

The example displays the use of the discursive strategy of argumentation (discussed in Wodak and Keller, 2008; Wodak, De Cillia, Reisigl, and Liebhart, 2009 and Wodak and Mayer, 2016), specifically the use of the topos of definition (defining Europe and Latvia's place within Europe) incorporating such linguistic means as a rhetorical question, metaphors (THE EUROPEAN UNION IS FAMILY OF DEMOCRATIC NATIONS, building metaphor) that intensify the intended effect of the argument.

As stated above, when analysing the speeches in detail, it is concluded that the identified keywords and expressions are present in all the speeches, specifically in the arguments that aim to discursively construct, perpetuate or justify national identities at the same time constructing common European identity as an overarching supra-national identity uniting the family of democratic nation-states, where established and redefined values (democracy, the rule of law, security, inclusiveness) rather than geographical, economic or historical elements are seen as the uniting elements. Another example illustrating the aforementioned conclusion is found in the speech by the second president of Lithuania after the restoration of independence:

I would like to wish you all the best of luck *building our common European family* that will *unite Europe's East and West, Europe's rich history and democratic values into one creative soul.* Ladies and Gentlemen, Last, but not least, today we celebrate one small anniversary – two years ago *Lithuania returned to the European family of democracies.* (Valdas Adamkus, 2006)

The example illustrates not only the emphasis on European values but also the application of the expression 'return to the family of democracies' (the use of the family metaphor) that has become a part of the common narrative of the Baltic States, not only in the international speeches but also in the national speeches (Romāne-Kalniņa, 2020), where the factor of continuation is emphasized via the verb and noun form 'return'. While the president of Lithuania uses the argument of building a common democratic European family, the inclusive/exclusive function of the argument seems to shift slightly as Vaira Viķe-Freiberga's construction of the family of democracies references the European Union (thus implicitly

excluding the East). At the same time, Valdas Adamkus explicitly emphasizes the inclusion of both East (Russia implied) and West (the European Union implied). Moreover, the use of the personal pronoun 'I' points to the use of the strategy of perspectivization, thus expressing the personal interest and involvement in positioning the point of view of the president.

Furthermore, the aforementioned example is interesting with the use of conceptual metaphor that personifies the notion of the family of democracies into a single abstract notion of the human soul that is characterised by the shared values listed above. The family metaphor is also used to refer to broader supra-national bodies, such as the United Nations Organisation, in an attempt to construct a global identity:

Dear Members of the *UN family, Brothers and sisters, French author Albert Camus has written: "There have been as many plagues in the world as there have been wars, yet plagues and wars always find people equally unprepared."* (Kersti Kaljulaid, 2020)

The construction of this shared identity is exercised via the use of the family metaphor, further extended via the reference to the members of the UN as brothers and sisters and the use of intertextuality in reference to a well-known author (expressing shared knowledge in order to create unity). The example illustrates the use of argumentation strategy via topos of authority that introduces an argument aiming to persuade the members of the now constructed shared global (political) identity to act on particular global issues that prevent the world from dealing with crises caused by war or the Covid-19 pandemic.

In addition, when looking at the list of most frequent keywords distributed across the sub-corpora based on separate criteria (state, time period, male-female speakers), it is noted that several keywords are repetitive (state name, reference to the capital of the state, reference to the Baltic Sea region), while some keywords seem to be criteria-specific (see Table 4 illustrating the keywords in the sub-corpora in Appendix 1). The keywords seem to emphasize that of all the three Baltic States, the politicians of Latvia (specifically the male presidents) emphasize the Baltic identity more frequently. Additionally, while the presidents of Estonia and Latvia seem to refer to the Soviet Union more frequently (specifically the male presidents in the period from 1991–2004), the presidents of Lithuania (again the male presidents from 1991–2004) refer to Russia more frequently. It is also noted that the keywords in the speeches from 2005–2021 and in the speeches by the female presidents differ slightly from those in the other speeches as they seem to focus additionally on global issues and construction of global identity as well as emphasize time-specific topical matters such as cybercrimes, technology, and political conflicts.

As to the list of most frequently used words (based on the relative frequency rather than on the keyness factor), the definite and indefinite articles *the, a* have been found to be used more frequently, followed by prepositions such as *of, to, in*, linking word *and*, pronouns *we, our, it*, verbs *be, is, are, have, will, has*, and nouns

and adjectives *European, Latvia, Europe, all, international, nations, security, states, countries, world, Estonia, new, development and people*. The lists of most frequently occurring words according to the parts of speech are displayed in Table 5 in Appendix 1, and the visualisation of the repetitive words across the corpus that emphasizes the main themes is seen in Figure 1 below:



Figure 1 Thematic areas as highlighted by frequently repeated words in the corpus from Voyant Tools

The data displayed in the figure above and in Tables 4 and 5 (in Appendix 1) point to the conclusion that the supra-national and global identities are emphasized more frequently in terms of common responsibility and duties to tackle several problems in the world that are both time-period dependent (disinformation, cyber-attacks, technology and pandemic) and continuous (peace, security, development, climate). The national identities of the states, specifically Latvia, and the regional Baltic identity, however, are positioned at the centre of speeches to construct and represent these identities as a part (the main element) of the wider overarching European and global identities. In other words, it is observed that multiple identities are co-constructed in the speeches. The national identities that are based on history, culture, geography (borders) and politics are discursively positioned at the centre and are expanding to regional (Baltic) identity based on geographical location (Baltic Sea region, border with Russia) and a common history (Soviet occupation), and to wider supra-national and global identities that are based on

shared values such as democracy, human rights and law as well as common duties and responsibilities in tackling the problems in the world.

Another figure that displays the dispersion of the most frequently occurring words across corpus (across presidents) that simultaneously displays the topicality and importance of particular identities and themes across time is drawn from the Voyant Tools online (see Figure 2 and Table 3 in Appendix 1). The data display that lemma *Latvia* (including all words with this root) is referenced most frequently, specifically in the speeches by Guntis Ulmanis. Lemmas *Estonia* (most frequent reference in the speeches by Lennart Meri and Arnold Rüütel) and *Lithuania* are referenced less frequently (frequent reference in the speeches by Algirdas Brazauskas and Rolandas Paksas). The second most frequent reference is to *Europe* and the European Union, which is most dispersed in the speeches by Rolandas Paksas and Arturas Paulauskas, Lennart Meri, Valdas Adamkus and Valdis Zatlers (see the list of presidents Appendix 2). Interestingly, the presidents of Latvia seem to emphasize national, Baltic and global identity more frequently than the European supra-national identity, while the presidents of Lithuania seem to do the exact opposite. It is also notable that *security* as a keyword that has occurred with various frequencies across the whole period seems to be most emphasized in the speeches by Andris Bērziņš, Raimonds Vējonis and Valdis Zatlers, Dalia Grybauskaitė and Toomas Hendrik Ilves, whose presidential terms were in the period from 2007 until 2016 when several military conflicts involving Russia have taken place. *Russia*, however, is not referenced directly as frequently as indirect references to its potential aggression. Nevertheless, it is notable that frequent references to Russia as well as to the Soviet Union have been used in the speeches by the first presidents of the Baltic States after the restoration of independence (period from 1991–2004).

CONCLUSIONS

It has been concluded that the keywords and wordlist of most frequently occurring lexical items point to the main thematic areas and the lexical change that marks the various themes in the speeches across time, across states and presidents, namely, *common political past, national culture, common political present and future, military security, cooperation, digital development, cyber security Covid-19 pandemic*. It is noted that the frequency of reference to several thematic areas, such as Russia's threatening presence and reference to the historical victim role, shifted slightly after the Baltic States joined the EU. This thematic area is replaced in frequency by other global issues such as climate change, cybercrimes, global pandemics, fake-news, military conflicts, and terrorism; however, the implication of potential victimhood of the states reappears in the use of keywords and discursive strategies from 2015 (after military conflicts between Russia and Georgia, Ukraine, and Belarus).

It is also concluded that the presidential speeches display the discursive construction and representation of multiple identities (national, regional, supra-

national and global) via various discursive strategies (assimilation, nomination, dissimilation, inclusion, and exclusion) and linguistic means such as metaphor, metonymy, pronominal references, and rhetorical questions. While the construction of national and European identities seems to be more frequent, a difference is notable in both quantitative and qualitative data displaying the discursive construct of these identities, namely, Estonia is constructed as sharing Scandinavian regional identity, Latvia is emphasized as sharing Baltic regional identity, while Lithuania's national identity seems to be embedded in and based on its membership in the European Union.

Consequently, it is concluded that the use of discursive strategies and key lexical items also point to membership categorisation, thus creating in-groups (those who belong, for instance, to the EU, to the global family of democratic nations, or to the Baltic Sea region, and those who do not belong, typically Eastern countries and specifically Russia). Nevertheless, the results also indicate that the first presidents of Lithuania after the restoration of independence seem to use more inclusive references to Russia and the East compared to the presidents of Latvia and Estonia expressing implicit hope for cooperation. Nevertheless, after 2015 the common fear and potential victim role seems to be equally displayed in the speeches by the presidents of all three Baltic States. Thus, the keyness factor points to the simultaneous construction of multiple identities and group categorisation via frequent repetition and emphasis on various key lexical items and via the appliance of distinct discursive strategies and linguistic means.

REFERENCES

- Baker, P. (2006) *Using Corpora in Discourse Analysis*. New York and London: Continuum Publishing.
- Bartlett, T. (2012). *Hybrid voices and collaborative change: Contextualising positive discourse analysis*. New York: Routledge.
- Billig, M. (2002) Critical discourse analysis and the rhetoric of critique. In G. Weiss and R. Wodak (eds.) *Critical Discourse Analysis: Theory and Interdisciplinarity* (pp. 35–46). London: Palgrave Macmillan.
- Bondi, M. and Scott, M. (2010) *Keyness in Texts*. Amsterdam: John Benjamins B.V.
- Breeze, R. (2011) Critical discourse analysis and its critics. *Pragmatics*, 21 (4): 493–525.
- Duvold, K., Berglund, S. and Ekman, J. (2020) *Political Culture in the Baltic States: Between National and European Integration*. Johannesburg: Palgrave Macmillan.
- Flowerdew, J. and Richardson, J. E. (2017) *The Routledge Handbook of Critical Discourse Studies*. New York: Routledge.
- Hackman, J. (1996) The Baltic world and the power of history. *Anthropological Journal on European Cultures*, 5 (2): 9–33.
- Haider, A. S. (2017) *Using Corpus Linguistic Techniques in Critical Discourse Studies: Some Comments on the Combination*. Christchurch: The University of Canterbury. Available from https://www.researchgate.net/publication/317045137_Using_Corpus_Linguistic_Techniques_in_Critical_Discourse_Studies_Some_Comments_on_the_Combination [Accessed on 15 October 2019].

- Haider, A. S. (2016) *A Corpus-assisted Critical Discourse Analysis of the Arab Uprisings*. Unpublished Doctoral Thesis. Christchurch: University of Canterbury. Available from https://ir.canterbury.ac.nz/bitstream/handle/10092/13008/Haider%2c%20Ahmad%20Final_%20PhD%20Thesis%20Linguistics%2056442034%20%28Final%29.pdf?sequence=1&isAllowed=y Accessed on 7 September 2020].
- Hammersley, M. (1997) On the foundations of critical discourse analysis. *Language and Communication*, 17: 237–248.
- Hart, C. and Cap, P. (2014) *Contemporary Critical Discourse Studies*. New York, New Delhi, Sydney: Bloomsbury.
- Kelly, S. (2020) *Who Deserves Compassion? A Corpus-Assisted Critical Discourse Analysis of The Discursive Construction of Syrian Refugees, Asylum Seekers, Immigrants and Migrants in UK Newspaper Reports Published Between October 2014 and September 2016*. Unpublished PhD thesis. Nottingham: University of Nottingham. Available from <http://eprints.nottingham.ac.uk/id/eprint/60949> [Accessed on 1 September 2021]
- Kitishat, A. R., Al Kayed, M. and Al-Ajalein, M. (2020) A corpus-assisted critical discourse analysis of the Syrian refugee crisis in Jordanian newspapers. *International Journal of English Linguistics*, 10 (6): 195–201.
- Lasas, A. (2008) Restituting victims: EU and NATO enlargements through the lenses of collective guilt. *Journal of European Public Policy*, 15 (1): 98–114.
- Lasas, A. (2008a) Guilt, sympathy, and cooperation: EU—Baltic relations in the early 1990s. *East European Politics and Societies and Cultures*, 22 (2): 347–372.
- Lucas, E. (2009) The fall and rise and fall again of the Baltic States. *Foreign Policy*, 173: 72–79.
- Maglie, R. B. and Centonze, L. (2021) Reframing language, disrupting aging: a corpus-assisted multimodal critical discourse study. *Working with Older People*, 25 (3): 253–264. Available from <https://doi.org/10.1108/WWOP-06-2021-0032> [Accessed on 20 September 2021].
- Matthews, A. (2021) Sociotechnical imaginaries in the present and future university: a corpus-assisted discourse analysis of UK higher education texts. *Learning, Media and Technology*, 46 (2): 204–217.
- Miller, A. (2018) Text mining digital humanities projects: assessing content analysis capabilities of Voyant Tools. *Journal of Web Librarianship*, 12 (3): 169–197.
- Mole, R. C. M. (2007) *Discursive Constructions of Identity in European Politics*. New York: Palgrave Macmillan.
- Mole, R. C. M. (2012) *The Baltic States from the Soviet Union to the European Union: Identity, Discourse and Power in the Post-communist Transition of Estonia, Latvia and Lithuania*. London, New York: Routledge.
- Mulderrig, J. (2008) Using keywords analysis in CDA: discourses in the knowledge economy in education. In B. Jessop, N. Fairclough and R. Wodak (eds.) *Education and the Knowledge-based Economy in Europe. Educational Futures: Rethinking Theory and Practice, Sense* (pp. 149–169). Rotterdam: Sense Publishers.
- Române-Kalniņa, L. (2020) Keyword analysis in presidential rhetoric throughout hundred years of independence of Estonia, Latvia and Lithuania: corpus linguistic tools in critical discourse analysis. In Karpinska, L., Farneste, M., Placinska, A. and Vladimirska, E. (eds.) *Language for International Communication: Linking Interdisciplinary Perspectives* (pp. 95–109). Riga: University of Latvia Press.

- Shah, N. (2021) Employability of persuasive and rhetorical strategies: a corpus-assisted critical discourse analysis of prime minister Imran Khan's selected address. *University of Chitral Journal of Linguistics & Literature*, 4 (2): 56–78.
- Slembrouck, S. (2001) Explanation, interpretation and critique in the analysis of discourse. *Critique of Anthropology*, 21: 33–57.
- Stubbs, M. (1997) Whorf's children: Critical comments on critical discourse analysis. In Ryan, A. and Wray, A. (eds.) *Evolving models of language* (pp. 100–116). Clevedon: Multilingual Matters.
- Subrenat, J. J. (2004) *Estonia: Identity and Independence*. Amsterdam, New York: Rodopi B. V.
- Tabuns, A. (1999) *Changing National, States and Regime Identities in Latvia*. Prague: Research Support Scheme.
- Widdowson, H. (1998) The theory and practice of Critical Discourse Analysis. *Applied Linguistics*, 19 (1): 136–151.
- Wodak, R. and Koller, V. (2008) *Handbook of Communication in the Public Sphere*. Berlin, New York: Mouton de Gruyter.
- Wodak, R. and Meyer, M. (2016) *Methods of Critical Discourse Analysis*. London: Sage Publications Ltd.
- Wodak, R., de Cillia, R., Reisigl, M. and Liebhart, K. (2009) *Discursive Construction of National Identity*. London, New York: Routledge.
- Žagar, I. Ž. (2010) Topoi in critical discourse analysis. *Lodz Papers in Pragmatics*, 6 (1): 3–27.

CORPUS TOOLS

- Anthony, L. (2012) *AntConc* (Version 3.3.2) [Windows 3.5.7]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/> [Accessed on 30 September 2018].
- Sketch Engine Tool Online*. Available from <https://app.sketchengine.eu> [Accessed on 30 September 2018].
- Voyant Tools* (2.0. digital version) Available from <https://voyant-tools.org> [Accessed on 21 June 2021].

SOURCES ANALYSED

- [Online 1] Speeches by the President of Estonia retrieved from the Homepage of the President of Estonia. Available from <https://president.ee/et/> [Accessed on 12 September 2018].
- [Online 2] Speeches by the President of Latvia retrieved from the Homepage of the President of Latvia. Available from <https://www.president.lv/lv> [Accessed on 12 September 2018].
- [Online 3] Speeches by the President of Lithuania retrieved from the Homepage of the President of Lithuania. Available from <https://www.lrp.lt/lt> [Accessed on 12 September 2018].
- [Online 4] Speeches by the Presidents of the Baltic States given at the United Nations General Assembly since 1991 retrieved from the United Nations Digital Library Archive. Available from <https://digitallibrary.un.org/search?!n=en&cc=Speeches> [Accessed on 1 December 2020].

APPENDIX 1

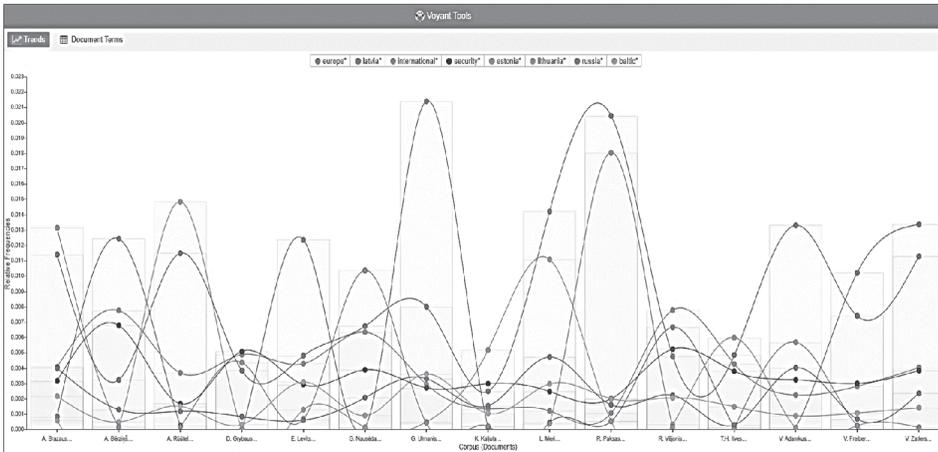


Figure 2 Word distribution across speeches in the corpus from Voyant Tools

Table 3 Relative frequency of key words (thematic areas) in the speeches as displayed in Figure 2

Order in Figure 2	President/word relative frequency	Estonia*	Latvia*	Lithuania*	Baltic*	Europe*	Russia*	International*	security*
1	Andris Bērziņš	161	12433	161	484	3229	1292	7751	6782
2	Algirdas Brazauskas	582	831	13137	2162	11391	3991	4074	3160
3	Arnold Rüütel	14832	251	168	1508	11480	1173	3687	1676
4	Dalia Grybauskaitė	0	0	4353	311	3835	829	4871	5078
5	Egils Levits	1286	12345	686	3086	4801	500	4286	2915
6	Gitanas Nausėda	130	130	10360	0	6734	2072	6346	3885
7	Guntis Ulmanis	481	21387	421	3604	7990	3304	2944	2703
8	Kersti Kaljulaid	5168	149	199	1044	2485	1541	1342	2982
9	Lennart Meri	11081	439	384	2962	14207	4717	1207	2468
10	Rolandas Paksas	534	534	18031	2003	20435	1603	1069	2003
	Arturas Paulauskas								
11	Raimonds Vējonis	317	6660	159	2062	4757	2220	7770	5233
12	Toomas Hendrik Ilves	5972	118	118	1478	4849	296	4258	3785
13	Valdas Adamkus	104	104	5685	887	13299	4016	2243	3234
14	Vaira Vīķe Freiberga	242	10213	242	1065	7406	678	2807	3001
15	Valdis Zatlers	134	13363	67	1410	11281	2350	4029	3828

Table 4 Dispersion of keywords in the sub-corpora

Sub-corpus (year) 1991–2004	Sub-corpus (year) 2005–2021	Sub-corpus (state) Estonia	Sub-corpus (state) Latvia	Sub-corpus (state) Lithuania	Sub-corpus (gender) Male	Sub-corpus (gender) Female
Baltic	cyber	cyber	Latvia	Lithuania	Baltic	cyber
Latvia	Baltic	Estonia	Latvian	Lithuanian	Latvia	EFP
Latvians	Latvia	Estonian	Baltic	Vilnius	Latvians	Zapad
Estonian	Crimea	Baltic	Latvians	Lithuanians	Latvian	Latvia
Latvians	Covid-19	Estonians	thanks	Baltic	Estonia	Baltic
Estonia	EFP	EFP	Riga	transatlan- tic	Estonians	deterrence
Lithuania	Latvian	Tallin	statehood	Kaliningrad	Lithuania	cyber- attack
Lithuanian	cyberspace	deterrence	Saeima	peace- keeping	cyber	Latvian
transatlan- tic	disinfor- mation	e-gover- nance	post-2015	Crimea	Lithuanian	Estonia
Europe	Estonia	Zapad	Soviet	EU-Russia	Estonian	digital
Estonians	Zapad	digital	Covid-19	Seimas	transatlan- tic	globally
Soviet	digital	Soviet	totalitarian	Russia	Soviet	UN's
Russia	pandemic	cyber- attack	disinfor- mation	geopolitical	Russia	Crimea

Table 5 Wordlist by parts of speech in the corpus

Nouns	Verbs	Adjectives	Pronouns	Adverbs	Conjunctions	Prepositions
country	be	interna- tional	we	not	and	of
Latvia	have	new	our	also	but	in
Europe	do	European	it	as	or	to
security	make	other	I	only	both	for
world	need	global	its	more	nor	that
state	take	human	their	well	yet	on
year	become	economic	us	even	either	with
develop- ment	continue	political	you	so	neither	as
union	achieve	important	they	together	et	by
UN	support	good	my	therefore	plus	at
Estonia	work	common	your	most	versus	from
united	believe	many	me	already		between

Nouns	Verbs	Adjectives	Pronouns	Adverbs	Conjunctions	Prepositions
Lithuania	like	demo- cratic	them	still		if
member	remain	first	his	now		into
nations	create	great	itself	ago		than
people	see	such	ourselves	all		about
council	provide	more	her	however		through
European	develop	Baltic	he	here		during
right	use	strong	them- selves	very		against
EU	come	small	one	just		among

APPENDIX 2

Table 6 Presidents of Estonia, speeches in the corpus

President	Years of service	Political affiliation	Number of speeches in the corpus	Words	Average words
Lennart Meri	1992–2001	National Coalition	11	18012	1637,45
Arnold Rüütel	2001–2006	People's Union – conservative	11	11763	1069,36
Toomas Hendrik Ilves	2006–2016	Social Democrats	10	16710	1671
Kersti Kaljulaid	2016–present	Independent	11	19850	1804,55

Table 7 Presidents of Latvia, speeches in the corpus

President	Years of service	Political affiliation	Number of speeches in the corpus	Words	Average words
Guntis Ulmanis	1993–1999	Farmer's Union	9	16429	1825,44
Vaira Vīķe-Freiberga	1999–2007	Independent	12	20386	1698,83
Valdis Zatlers	2007–2011	Independent	8	14751	1843,87
Andris Bērziņš	2011–2015	Union of Greens and Farmers	5	6092	1218,4
Raimonds Vējonis	2015–2019	Union of Greens and Farmers	5	6215	1243
Egils Levits	2019–present	Independent	11	11461	1041,90

Table 8 Presidents of Lithuania, speeches in the corpus

President	Years of service	Political affiliation	Number of speeches in the corpus	Words	Average
Algirdas Brazauskas	1993–1998	Social Democrats	6	11890	1981,67
Valdas Adamkus	1998–2001; 2004–2009	Independent	14	19035	1359,64
Rolandas Paksas	2003–2004	Order and Justice national party	5	5486	1097,2
Arturas Paulauskas	2004	Labour Party	1	1921	1921
Dalia Grybauskaitė	2009–2019	Independent	17	9575	563,24
Gitanas Nausėda	2019–present	Independent	6	7632	1272

Līga Romāne-Kalniņa (Mg. Philol.) is a PhD candidate of linguistics at the University of Latvia, currently working with technology transfer (science commercialisation) at the Investment and Development Agency of Latvia. Her research interests include language of innovation, political and media discourse, critical discourse analysis and corpus linguistics.

ORCID ID <https://orcid.org/0000-0003-0451-7149>.

Email: ligaromane@inbox.lv