

Georgia's Path to Europe - A Corpus Linguistic Analysis of Speeches by President Salome Zourabichvili and Prime Minister Irakli Garibashvili

Nino Guliashvili, PhD

School of Arts and Sciences,
English Philology Ilia State University, Tbilisi, Georgia

Doi: 10.19044/esipreprint.10.2025.p220

Approved: 08 October 2025
Posted: 10 October 2025

Copyright 2025 Author(s)
Under Creative Commons CC-BY 4.0
OPEN ACCESS

Cite As:

Guliashvili, N. (2025). *Georgia's Path to Europe - A Corpus Linguistic Analysis of Speeches by President Salome Zourabichvili and Prime Minister Irakli Garibashvili*. ESI Preprints. <https://doi.org/10.19044/esipreprint.10.2025.p220>

Abstract

Social and political phenomena come into existence through a discursive dimension. The linguistic construction of reality is pervasive as well as predetermining social and political environment. The present paper undertakes a linguistic scrutiny of Georgia's endeavor to become a member state of the European Union. Joining the EU (accession) exists linguistically in a discursive dimension before its political establishment. The accession officially started on March 3, 2022. Therefore, President Salome Zourabichvili's and former Prime Minister Irakli Garibashvili's speeches/addresses stored at the official websites represent the discursive formation of Georgia's path to the European Union. This article seeks to study the perceptions of the EU accession through 97 speeches delivered by the president and the prime minister of Georgia with the help of corpus linguistics software tools #LancsBox® and Wmatrix5® alongside corpus linguistics (CL) and discourse analysis (DA), which is referred to as corpus-assisted discourse studies (CADS). Moreover, the application of CADS to Georgian political discourse analysis is a novel endeavor itself. The paper scrutinizes lexical frequencies (a quantitative aspect) as well as identifying semantic/notional domains (a qualitative aspect) that the political figures most frequently resort to. The paper attempts to shed light on the discursive choices the politicians make regarding the European path, which is essential

in the current political unrest and polarization. Based on Keyword and T-score analysis of essential collocates along with stance-taking peculiarities, the findings show that the speakers convey their intent to the listeners through their own idiolects.

Keywords: EU accession, corpus linguistics, political discourse, discourse analysis, CADS

Introduction

The political discourse encompassing Georgia's European Union (EU) accession has gained noteworthy attention. Special emphasis could be placed on the recent developments, such as the European Commission's recommendation to grant Georgia a candidate status in December 2023. The potential of Georgia to become a member of EU may reflect broader geopolitical shifts in the region and represents an essential object of scrutiny from a variety of perspectives. The discourse analysis of this crucial topic is a timely and relevant endeavor, as Georgia's European aspirations get intertwined with ongoing political discussions and issues of national identity within the scope of Eastern European integration. By linguistic examination of Georgian political leaders' verbal repertoires, it would be possible to understand how the politicians frame the process in the ongoing polarized contexts.

The EU accession of Georgia as a political process, which officially gained momentum on March 3, 2022¹. Therefore, the present paper offers a possible way to better apprehend the discursive construction of the above-mentioned political phenomena in the speeches and addresses of the president and the prime minister who have recently been accusing each other of political irrelevancies. Language as a primary means of meaning-making reflects our conceptual system and the way we apprehend the world (Lakoff & Johnson 2003; Langacker, 1987). By studying the language that we apply to the specific context to make sense of reality, it is quite salient that the cognitive basis of complex abstract conceptions will be revealed. Bearing this in mind, the present research adopts a corpus linguistic, discourse-based "mixed-methods approach" (Brown & Mondon, 2020, 3) referred to as corpus-assisted discourse studies (CADS). The combination of CDS (Critical Discourse Studies) and DS (Discourse Studies) with CL (Corpus Linguistics) has been highlighted as "a mutually beneficial fusion" (Nikisianis et al., 2019, 280; Brown & Mondon 2020, p. 4). Moreover, the corpus-based analysis (Baker 2006; Tognini-Bonelli, 2001) employed in this study is

¹ The Georgian Dream government has put a stop to it in November 2024 due to certain political processes.

based on the two corpora that serve as the evidence to either bolster or refute the understandings in connection with the discursive formation of the political positions and stances in relation to EU integration. What is more, the application of CADS to the scrutiny of the above-mentioned case is a novel endeavor with regard to the research conducted in the Georgian socio-linguistic milieu.

The linguistic analysis in this paper focuses on how the president, Salome Zourabichvili, and the former prime minister of Georgia, Irakli Garibashvili (PM), perceive and discursively represent their stances on the country's EU integration. More precisely, the present paper attempts to answer the question of whether these political figures discursively construct similar or varied views on the current political phenomenon. The linguistic studies being performed for this research embrace *lexicometric* analysis (i.e., analysis of the lexical frequencies of keywords in context) along with *textometric* analysis (i.e., a computer-assisted textual analysis that links statistical processing of lexical items with the contextual specificities). In the first place, the focus is placed on the lexical discrepancies between the president's and the prime minister's speeches; the lexical frequencies were processed via the software tool #LancsBox®: Essential keywords, individual and shared collocates for the keywords *Europe*, *European*, and *EU* were identified and compared. Then, a more qualitative approach was carried out by focusing on the semantic domains (also known as notional domains) that are prevalent in the Salome Zourabichvili and Irakli Garibashvili corpora. For the qualitative scrutiny of the discourse Wmatrix5® with integrated Semantic Analysis System - UCREL was employed. The current study addresses the following research questions:

1. What do the keywords reveal about the politicians' attitudes towards EU accession?
2. From a linguistic perspective, do frequently used semantic domains in the President's and Prime Minister's speeches constitute different narratives?
3. How do *lexicometric* and *textometric* analyses relate to the political stance taking in the politicians' speeches/addresses?

By analyzing how the politicians picture Georgia's unity with the European Union through the language they use, the paper attempts to uncover how their discursive visualizations may differ and how potential discrepancies may be concomitant to their individual standpoints. Therefore, addressing the above-mentioned research questions promotes the pervasiveness of the essential 'how' of discourse analysis (Mautner, 2019,.10).

Literature Review

In recent years, discourses of European political issues and polarizing matters have been extensively investigated by Corpus-Assisted Discourse Studies (CADS). Since CADS methodology integrates quantitative corpus linguistic techniques with qualitative discourse analysis, it has been widely applied to explore how political actors frame a variety of controversial topics such as European integration, national identity, and sovereignty.

Wodak et al (2009) examine the discourse of EU institutions, highlighting how language can construct a unified European identity in the context of enlargement policies. Similarly, Baker (2010) applies CADS to the issue of Euroscepticism in the UK. Thus, he reveals that pro-EU and Eurosceptic rhetoric strategically construct opposing stances on sovereignty and integration. The discourse analysis of the above-mentioned cases identifies duality, which is pertinent to the EU-related debates.

Another topic is migration within the EU. Fitzgerald's (2014) application of CADS to British political parties' rhetoric has revealed that the framing of migration could have a dichotomic representation, seen as either a threat to national identity or an economic benefit. Furthermore, Kolb's (2018) analysis of Eastern European political discourses highlights how political elites find the balance between economic gains and possible threats to the sovereignty of the country.

In a broader context of polarizing political issues, Richardson (2006) demonstrates how discourses *of the War on Terror* in the US and UK construct "us" (the West) vs. "them" (terrorists) dichotomy via language justifying political decisions and military interventions. As it appears, divisive rhetoric can legitimize political stances. Finally, polarized rhetorical strategies identified in the parliamentary discourse on EU policies (Nelson, 2016) reflect tensions between unity and autonomy, a dynamic that can also be observed in Georgian political discourse.

These studies underscore the effectiveness of CADS in investigating and understanding the complex workings of language in shaping and framing political stances, positions, and identities within the context of polarized political discourses.

Brief Historical Background on the EU accession and the Political Friction between President and Prime Minister

The accession of Georgia to the European Union is officially regarded as a current agenda item for EU enlargement. Georgia applied for EU membership in March 2022, which was followed by the EU acknowledgment to view the country as an eligible candidate for becoming a member of the Union. November 8, 2023, marks the day of an official recommendation issued by the European Commission to grant a candidate

status to Georgia, which was ratified on December 14, 2023. Georgia has become one of 9 current candidate countries together with Albania, Bosnia and Herzegovina, Moldova, Montenegro, North Macedonia, Serbia, Turkey and Ukraine.

Nevertheless, the relations between Georgia and the European Union started in 1992 as a consequence of an agreement between the former European Community and the newly independent Georgia. In 2006 Georgia implemented a five-year “Action Plan” of rapprochement in the context of the European Neighbourhood Policy (ENP), which was further upgraded in 2009 under the auspices of the Eastern Partnership. The Comprehensive Association Agreement between the EU and Georgia in 2016 granted the country free visa travel to the EU. These political processes have been supported by the governing Georgian Dream and the main oppositional party, United National Movement, along with the other small parties, which are pro-Western in orientation.

President Salome Zourabichvili is a French-Georgian politician and former diplomat, who is currently serving as the fifth and the last popularly elected president of Georgia (President’s term ended in December 2024). She was endorsed by the ruling Georgian Dream party during the period of the presidential campaign. Nevertheless, a political friction between Salome Zourabichvili and Georgian Dream was incited in 2023 due to the President’s travel to Europe, so as to hold meetings with the European leaders to lobby for Georgia’s EU candidacy despite not being granted the authorization from the government to do so (Politico, 2023). Moreover, the president was accused of “strengthening the false campaign” inspired by the opposition party, United National Movement (Agenda.ge, 2023). In effect, the competition for the eligible European path is set in motion. We have seen a myriad of political evaluations and assessments. Hence, it would be interesting to scrutinize the discursive formation of the European path along with the politicians’ attitudes framed by the linguistic repertoire, which could be rendered as a novel endeavor of a linguistic scrutiny of the political discourse in Georgia.

Modals and Stance Taking

In this paper, stance is operationalized via *modal verbs* and *attitudinal language* that tend to reflect varying degrees of certainty and commitment to a position. For example, epistemic modals such as *will* indicate certainty, while modals such as *may* or *might* suggest uncertainty. Attitudes, beliefs, or evaluations toward a particular event or proposition can be expressed by means of stance-taking modal auxiliaries, which can signal the speaker’s degree of certainty, obligation, necessity, or possibility (Biber et al., 1999). Therefore, modality is not simply a grammatical feature but an

essential means of reflecting the speaker's position in relation to the truth of a proposition or social dynamics ascribed to the interaction (Palmer, 2001). Contextually, *epistemic modals* – such as *might*, *may*, *must*, *could*, and *will* – express certainty or uncertainty about the truth of a statement. For instance, the use of *must* signals an evidence or reasoning-based high level of certainty about the speaker's belief (Biber et al., 1999). Conversely, the modal *might* reflect a lesser degree of certainty, implying uncertainty or openness to alternative possibilities. These epistemic modals make it possible for the speaker to adjust the degree of commitment in relation to the truth of their claims (Hunston and Thompson, 2000).

Central lingual figures of social stance-taking are *deontic modals* (i.e., *must*, *should*) expressing obligation, permission, or necessity. The functional side of a modal *must* convey not only a strong obligation but also the speaker's authority or expectation regarding the listener's behavior (Palmer, 2001). Similarly, *should* reflects the speaker's evaluative stance via recommendation or advice on the behavior deemed appropriate or desirable. In this way, modals frame the interaction by signaling the speaker's views of the necessity or permissibility of the actions, which can influence listeners' response (Hunston and Thompson, 2000).

In comparison with the above-mentioned types of modalities, *root modality* (expressed by *can* and *could*) expresses ability or possibility by highlighting what could be possible or feasible within a given context (Biber et al., 1999). By this, the speaker can focus on the limits or potentials of a situation (Palmer, 2001).

The strategic use of modals is associated with negotiating power, persuasion, and agreement in the discourse. Thus, politicians are able to assert their positions of national importance by shifting their tone of communication from confidence and certainty to possibilities and imposing orders. Therefore, deontic, epistemic, and root modality are grammatical classifications of modal verbs based on their intrinsic features, while *attitudinal* and *affective stance-taking* characterize pragmatic roles played by the modal verbs in the process of expressing the speaker's evaluations or emotions.

Attitudinal stance conveys the speaker's judgmental view on the desirability, necessity, or appropriateness of a communicative situation. For instance, the interplay of *should* and *must* alters recommendation or expectation into strong obligation or necessity. *Affective stance*, on the other hand, underscores the speaker's emotional tone. Modals like *may* and *might* promote tentativeness and doubt, while *would* and *will* could show willingness or hope (Palmer, 2001).

Apart from the above-mentioned stance-taking strategies, there is *stylistic stance* as well, which refers to the way an individual conveys their

attitude, perspective, or positioning through stylistic choices in communication. For example, adjectives could play an essential role in expressing judgments, or perspectives while adverbs could modify the intensity or certainty of these judgments (e.g. *absolutely brilliant*).

Overall, by analyzing the means of stance taking, we can gain insight into how politicians position themselves within the political discourse via applying modality or modal phrases to their speeches as an essential tool for shaping communication dynamics.

Corpora

The present study gathers all the speeches/addresses from the President and Prime Minister that are stored in English on the official websites <https://president.ge/index.php?m=206> and <https://garibashvili.ge/en/about>. The on-record speeches for the European audience are delivered and stored in English, while the speeches designed for the Georgian audience are officially translated by the presidential and prime minister's communication services.

All the transcripts were retrieved from the official websites, and the two sets of corpora were compiled²: the Salome Zourabichvili Corpus (Corpus 1 - 69 speeches, 64 321 tokens) and the Irakli Garibashvili Corpus (Corpus 2 – 28 speeches, 29 559 tokens). Although these two sets of corpora are not of the same size, normalized absolute frequencies are used for the relevant statistical analysis.

Overall, the dataset for this study consists of 97 speeches delivered by President Salome Zourabichvili and former Prime Minister Irakli Garibashvili, obtained from official governmental websites between March 2022 and December 2023. Although the corpus is relatively small, it includes high relevance texts, which discursively frame Georgia's EU accession. The speeches under linguistic scrutiny feature a direct engagement with the European integration theme that renders the political narratives highly pertinent for the analysis. As the process of EU integration is underway, future studies could incorporate more extensive material to comprehensively investigate Georgia's EU journey.

Methodology

The present study is contingent on a mixed-methods approach employing methods of corpus-assisted discourse studies (CADS). There have

² Report [#LancsBox] This research report was automatically produced by #LancsBox (Brezina et al., 2015, 2018, 2020), a corpus analysis tool developed at Lancaster University. It uses cutting-edge technology and statistical sophistication (Brezina 2018) to analyze and visualize corpus data. For more information and tips on research report writing see the Research Report Guide.

been numerous studies conducted on EU integration in different contexts. Nevertheless, applying a corpus-assisted discourse studies (CADS) approach to Georgia's political discourse is a novel endeavor, which offers a detailed linguistic analysis of how Georgian political figures conceptualize the European path. The research explores and highlights nuanced differences in how EU accession is framed linguistically by focusing on key themes, such as national unity, European values, and political sovereignty. By applying both lexicometric and textometric analytical tools, the present paper appears to be a significant contribution to the field of political discourse studies, especially in the context of post-Soviet countries with aspirations to join EU.

The analysis of the corpora combines quantitative and qualitative approaches. #LancsBox® (Brezina et al., 2015, 2018, 2020) was used to conduct quantitative analysis with regard to identifying frequent keywords and collocates, while Wmatrix5® (Rayson, Archer & Wilson 2002, 2) facilitated semantic domain analysis. The duality of analytical approach ensures a comprehensive understanding of how key terms like *Georgia* and *EU* function within political discourse. The quantitative data serves as a starting point for close reading of concordance lines, where the contextual meaning of the terms in focus is examined. The differences are supposed to be reflected in the speech patterns of their idiolects.

Quantitative corpus analytic methods have made it possible to analyze large datasets, so as to identify patterns of language use that could not be detected via reading the textual data (Partington & Marchi, 2015; Baker, 2006; Partington, Duguid, & Taylor, 2013). Therefore, the use of corpus-analytic methods and tools led to a qualitative scrutiny of the relevant passages to gain more insightful understanding of the language patterns, which have been used in a certain context to accommodate language users' beliefs and behaviors (Lischinsky, 2018, 62).

The lexical analysis involved the elaboration of the keyword frequency lists. #LancsBox® software package was used to analyze the data. The tool of #LancsBox -*Words*, was used for in-depth analysis of frequencies of types, lemmas, and POS categories as well as comparison of corpora using the keywords technique. Keywords are characterized as lexical items that occur significantly more often in a given research corpus than in a referential wordlist (BE06 referential set of words/referential corpus is used for comparison, which is a balanced corpus of modern British English compiled to represent contemporary written English usage). **Corpus 1** (the Salome Zourabichvili Corpus) is represented with the 50 most frequently used words (an output of #LancsBox®)³ and **Corpus 2** (the Irakli Garibashvili Corpus)

³ These lists show all keywords identified in both corpora, including their statistical significance.

is also represented with the 50 most frequently used words (an output of #LancsBox®) (See Appendix). The keyword lists were filtered down to the 25 lexical types – *nouns, verbs, adjectives, modal verbs, and numerals*, as they are the major carriers of notions and concepts in the present study. As for the other frequently used words presented in the outputs, such as articles, particles, prepositions, pronouns, and conjunctions, these were removed as they are not the carriers of thematic notions.

The next step focused on certain lexical items regarding the two corpora, more specifically, Europe-related keywords - *Europe, European*, and *EU* as well as Georgia-related keywords – *Georgia, Georgian*. The software tool produced individual and shared collocates for the target lexical units.

For the qualitative *textometric* analysis, Semantic Domain Analyzer Wmatrix5® was used. It is developed by a group of researchers at Lancaster University and features a semantic tagger USAS, which groups together the words pertinent to the same semantic fields via pre-existing semantic categories that are integrated in the matrix:

A general and abstract terms	B the body and the individual	C arts and crafts	E Emotional actions, States and process
F Food and farming	G Government and public domain	H Architecture, building and housing	I Money and commerce in industry
K Entertainment, sports and games	L Life and living things	M Movement, location, transport and travel	N Numbers and measurements
O Substances, materials, objects and equipment	P Education	Q Linguistic actions, language and communication	S Social actions, states and processes
T Time	W World and environment	X Psychological actions, states and processes	Y Science and technology
N Names and grammar			

For transparency and replicability of the research, all the files of Corpus 1 and Corpus 2 have been published on Harvard Dataverse along with #LancsBox Report and Wmatrix outputs for Corpus 1 and Corpus 2 Semantic Domains Analysis (Anonymous, Anonymous 2024).

Results and Discussion

Lexical Frequencies

To highlight the prominence of the lexical units used in speeches, the study explored the lexical frequencies of notional parts of speech.

#LancsBox WORD function produced the keyword frequency lists (in reference to BE06 referential corpus), providing both absolute and relative frequencies. The latter is essential for the study, as the two corpora are not of the same size and the relative frequency represents normalized occurrences of the target words levelling each corpus to 1 000 000 tokens/word types.

Figures 1 and 2 below show the most frequent words for each corpus prepared on the basis of filtering the keywords down to the *twenty-five* most essential notional parts of speech along with *modal auxiliaries*, which tend to reveal speaker's stances towards the specific events and phenomena (particles, articles, pronouns, conjunctions and other auxiliary verbs were excluded from the list as they are not the carriers of the thematic notions).

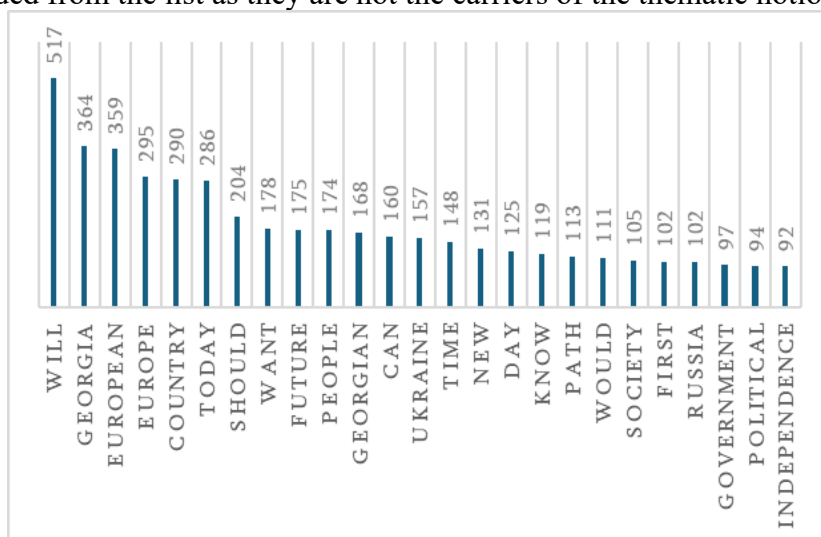


Figure 1. Absolute Keyword Frequency for the Salome Zourabichvili Corpus (Corpus 1)

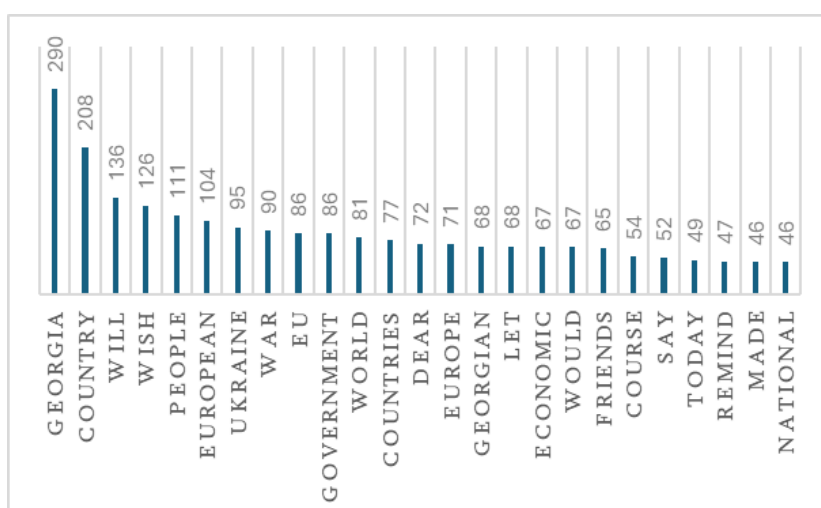


Figure 2. Absolute Keyword Frequency for the Irakli Garibashvili Corpus (Corpus 2)

Interestingly, the most frequent word in the Salome Zourabichili Corpus is the modal auxiliary *will* expressing epistemic modality – certainty about the diplomatic actions, and a united plan of national accord. Excerpts (1) and (2) instantiate the recurrence of the target lingual unit.

1. Still, I hope very much that our diplomacy will meet its standard and accomplish everything that this country needs. We all know that today, what this country needs the most is the path of Europe, to succeed on this European path.
2. It is necessary to prepare and present a united plan of national accord that will not be subject to annulment.

Another modal (excerpt 3) in the president's discourse *can be* used with the function of root modality, which highlights the possibility /feasibility of the aspirations Georgia may have on its path to becoming a member state of the EU.

1. We can see very well, and from the outside perspective, it can be seen even better, what Georgia wants, what its people want, what its society wants, and what its youth wants.

The second word in the frequency list of the President's corpus is *Georgia*, which appears to be a leading one in the Irakli Garibashvili Corpus, stressing the achievements of the country throughout the period of the Georgian Dream's rule (excerpt 4). The PM's 25 essential keywords also include modal auxiliaries *will* and *would* while discussing the polarization in the country in relation to the process of granting the status (excerpt 5), along with explaining the effect of economic sanctions against Russia (excerpt 6):

- 1) As for the fight against corruption, Trace International ranks Georgia among 20 Best Countries of Europe and 30 Best Countries of the World, being the first in a sub-component of Low Risk Degree of Corruption in Dealing with Authorities. According to the Global Corruption Index, Georgia advanced by 9 position points in 2021 and was ranked 41st in the world. Our country is ahead of such member states of the EU and NATO as Slovakia, Croatia, Bulgaria, Hungary, Romania, North Macedonia, Montenegro, Albania and Turkey.
- 2) if Europe, MEPs and our European colleagues genuinely wish to see polarization ended in Georgia, they should grant us the status of an EU Candidate Country. It will remove all the questions and speculations. It will end their destructive and radical actions and no real reason will remain for our extremist opposition to blame our governance for anything.
- 3) Against this background, you are especially well aware that imposition of sanctions at the national level, enforcement of national

economic sanctions towards the Russian Federation would translate into sanctions being imposed on our own citizens there and into sanctioning our country. Thus, irrespective of our solidarity towards Ukraine, our partners, we said that we would be guided with our national interests only and nobody will ever make us change our decision as you may know.

Figure 3 below shows the overlapping keywords that the two corpora share. These words are *Georgia*, *Georgian*, *Europe*, *European*, *Country*, *People*, *Ukraine*, *Today*, *Government* and *will /would*. The president and the PM, both share the lingual units that frame their attitudes towards the European vector. The keywords indicate the similar themes that are pertinent to their discourses: interest in the integration in the EU, people's involvement, development of the country, government's stability, and support for Ukraine, which is crucial in the current political turmoil.

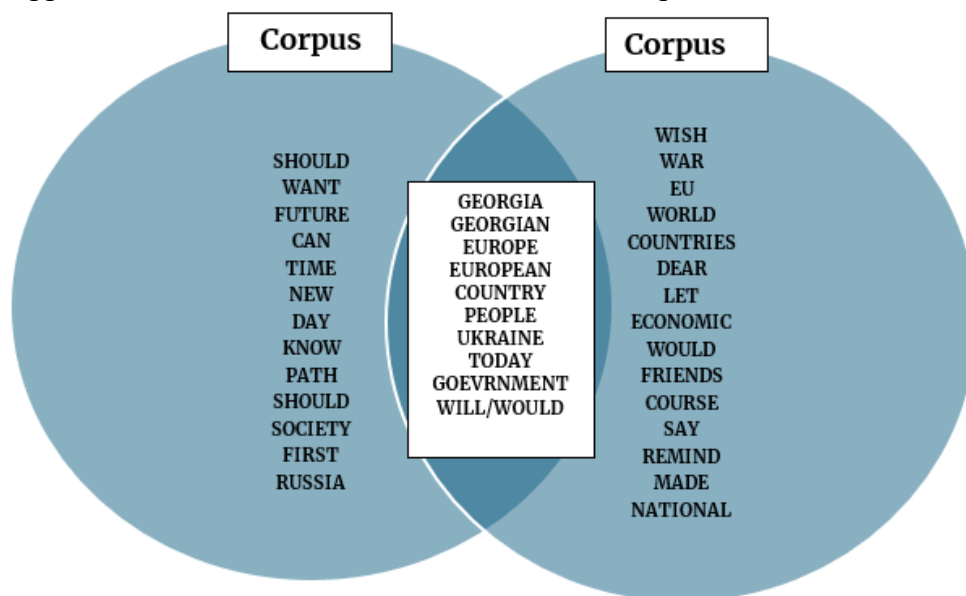


Figure 3. Shared Keywords between the two corpora

The keywords, such as *Georgia/n* and *Europe/an*, are essential to understanding how political figures frame the country's EU accession. For this reason, specific relative frequencies for Europe-related words were visualized in the diagrams below (Figures 4 and 5). The President uses the words *Europe* and *European* more often than the PM (Figure 4). General relative frequencies sum up the specific relative frequencies, showing the prevalence of EU-related words in the President's discourse.

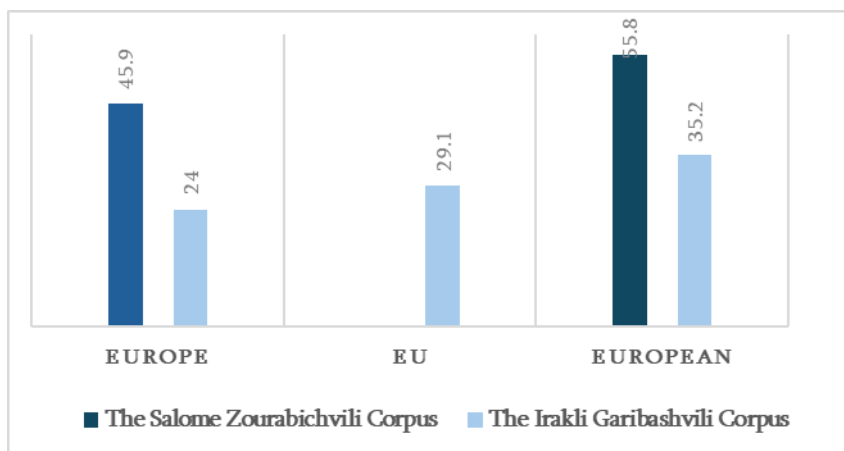


Figure 4. Specific Relative Frequencies for Europe-Related Words in both corpora

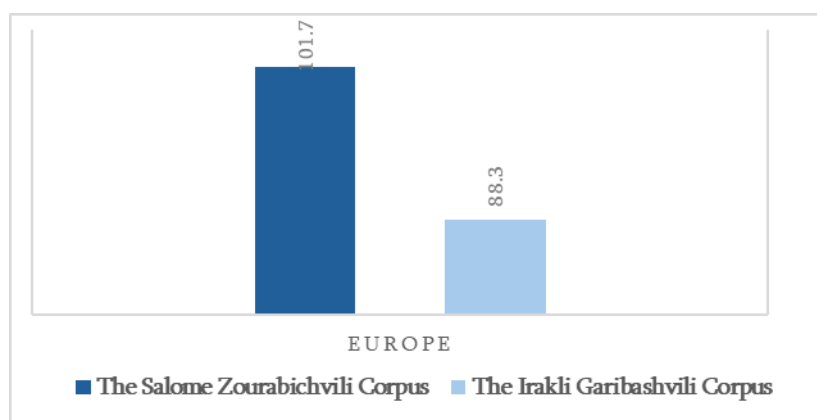


Figure 5. General Relative Frequencies for Europe-Related Words in both corpora

According to Figure 4, *EU*, as a frequently used keyword appears only in the Prime Minister's corpus, reflecting the political emphasis placed on EU membership. Interestingly, both politicians state the importance of shared European values. The PM stresses the historical connection between Europe and Georgia, which contributed to the establishment of shared European values (excerpt 7), while the president strives to promote the intrinsic reciprocity of Georgian and European values (excerpt 8):

- 1) The **Georgian people**, taking great pride in their ancient culture and history dating back thousands of years, have always been oriented towards Europe and remained loyal to its values. At the dawn of the 20th century, **Georgian** people created a most democratic, European state based on **European values**, while its essence and Constitution were compatible with current criteria of Europe with high precision.
- 2) We now just have to prove that we are driven today by our **Georgian values**, which are also **European values**. If we can prove that these

values are within us, we will overcome the disappointment and 2023 will truly be the Year of Europe.

To identify whether the difference between the frequencies was statistically significant, a significance test was conducted via an online software developed by Lancaster University research group (Rayson & Garside, 2000). This software compares absolute frequencies of the target lexical units by indicating the sizes of the corpora to balance the data and calculate statistical significance. According to the test⁴, LL (Log-likelihood) = 3.79 i.e. *p value* is not less than 0.05, which means that the difference is not significant. From the discourse perspective, we cannot claim that the president's use of EU-related words is significantly different from the PM's discourse. Their aspiration to join the EU is discursively alike.

Figure 6 shows a comparison of the specific relative frequencies for Georgia-related words in both corpora. The diagram explicates the prevalence of the target lexical item in the Prime Minister's speeches. The general relative frequencies (Figure 7) summarize the prevalence of the target lingual units in the MP's speech.

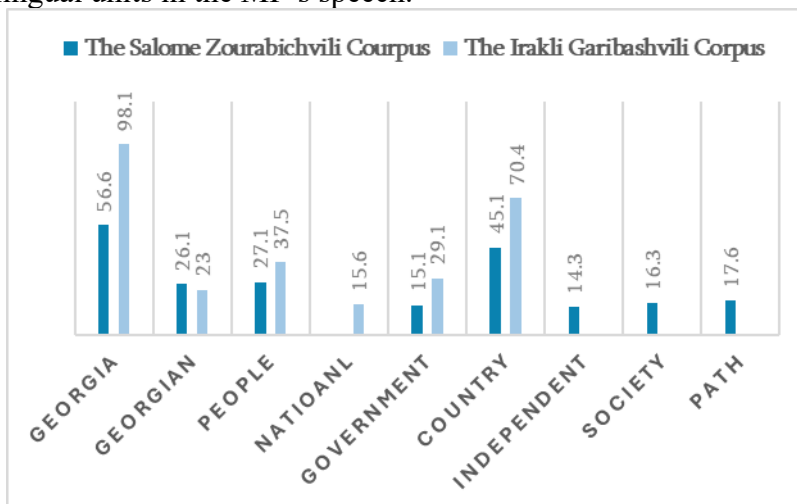


Figure 6. Specific Relative Frequencies for Georgia-Related Words in both corpora

⁴ Log-likelihood calculator results:

<https://ucrel-api.lancaster.ac.uk/cgi-bin/llsimple.pl?f1=654&f2=261&t1=64.321&t2=29.559>

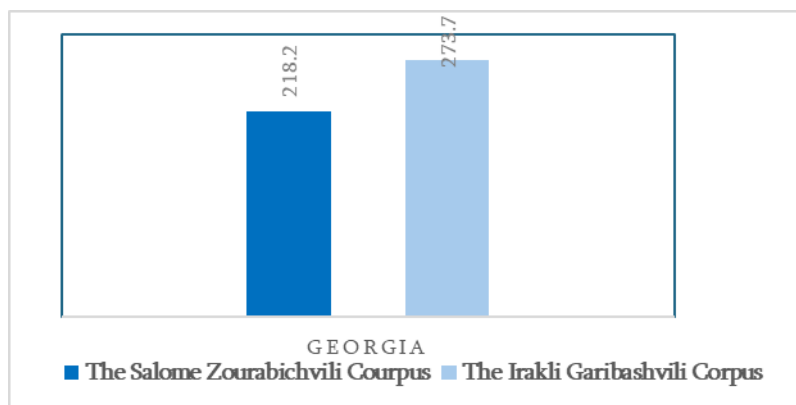


Figure 7. General Relative Frequencies for Georgia-Related Words in both corpora

To substantiate whether the above-mentioned difference was significant, the test was run, yielding the following result.⁵ $LL=25.85$ and $p<0.05$, which means that in this case the difference in use of the words related to *Georgia* is statistically significant. From a discourse perspective, the PM mentions Georgia in association with *Europe* expressively more than the President.

To go even deeper into the verbal repertoire of the politicians, the GraphColl tool in #LancsBox was applied to explore the collocates of *Europe*, *European* and *EU*. The integrated T-score statistical tool builds the visualization of the most frequent collocates, which could assemble a coherent ‘story’ or ‘narrative’ of a political figure in relation to the specific political occurrence. The narrative on its own can depict and discursively structure certain events from the perspective of a politician (Shenhav 2006). Studying collocates by means of corpus linguistic tools enables researchers to detect the meaning of the target words being developed in interaction with other ones. In other words, “you shall know a word by the company it keeps” (Firth 1957). Therefore, a T-score comes across as a statistical measure that indicates whether a word appears significantly more frequently in one corpus as opposed to a reference corpus. In the present study, LancsBox pairs T-score (statistic value is set at 5 in the parameters of LancsBox to ensure the significance) with the collocates of the target keywords. According to Figures 8 and 9, the keywords *Europe* and *European* have more collocations in the President’s discourse, while the PM’s major collocates are linked with the keyword *EU* and *European*. Salome Zourabichvili’s narrative about the European path embraces the following collocates: *European: standards, perspective, family, commission, integration, partners, family, Georgia’s, union, council* and *countries*. The following collocates prevail in the PM’s

⁵ Log-likelihood calculator results: <https://ucrel-api.lancaster.ac.uk/cgi-bin/lisimple.pl?f1=1403&f2=809&t1=64.321&t2=29.559>

discourse: *European (EU): commission, council, family, friends, parliament, integration, association, perspective, candidate, and membership.*

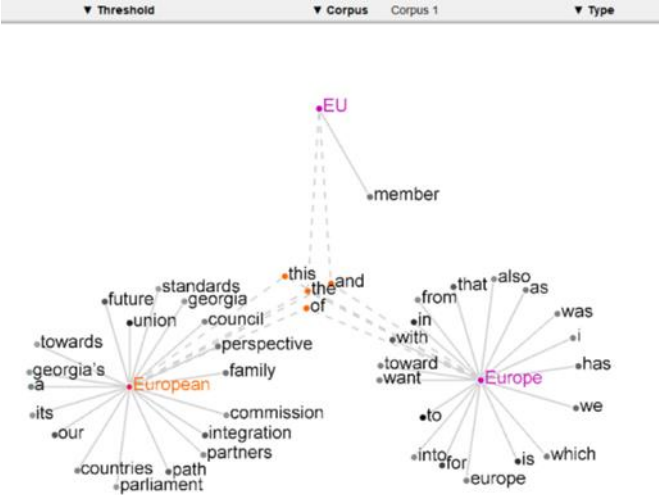


Figure 8. T-score for EU-related Words in the Salome Zourabichvili Corpus (Corpus 1)

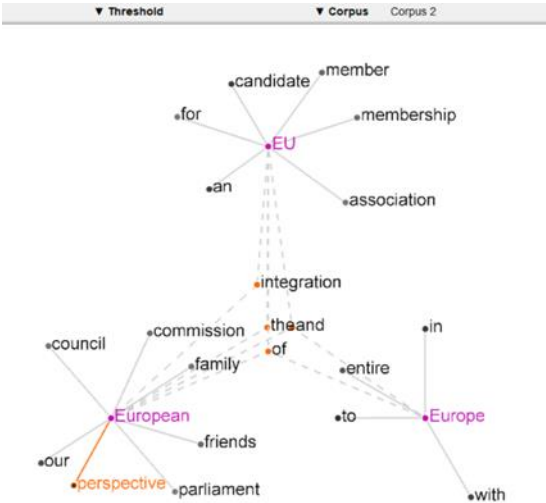


Figure 9. T-score for EU-related Words in the Irakli Garibashvili Corpus (Corpus 2)

The significant collocates of the keywords, such as *EU*, *European*, *Europe* in both corpora provide a useful insight into the political discourse of the political figures.

Semantic Analysis

A qualitative approach to discourse analysis can include the identification of the semantic domains. Therefore, to uncover the most recurrent semantic domains in the target corpora, the UCREL Semantic

Analysis System (USAS), which is integrated in Wmatrix5® was used. The tagging procedure yielded the statistically more frequent semantic fields occurring across the data. The tagging procedure is accurate in terms of assigning certain codes to the lingual units, which are available in the matrix. Table 4 below shows the distribution of 10 most significant semantic domains pertinent to the corpora (shared semantic domains are highlighted).

Table 4. Semantic domains across the two corpora

Tag & Frequencies	Salome Zourabichvili Semantic Domains – Corpus 1	Tag & Frequencies	Irakli Garibashvili Semantic Domains – Corpus 2
Z5 18758	Grammatical bin	Z5 8456	Grammatical bin
Z8 7473	Pronouns	Z8 2964	Pronouns
A3+ 2182	Existing	Z2 882	Geographical names
Z2 1585	Geographical names	G1.1 717	Government
M6 1518	Location and direction	A3+ 701	Existing
Z99 1034	Unmatched	Z99 570	Unmatched
G1.1 955	Government	A1.1.1 457	General actions / making
Z6 901	Negative	M6 382	Location and direction
T1.1.3 743	Time: Future	X7+ 367	Wanted
S5+ 740	Belonging to a group	Q2.2 333	Speech acts
S6+ 704	Strong obligation or necessity	M7 315	Places
A7+ 689	Likely	N1 305	Numbers
A1.1.1 665	General actions / making	S5+ 304	Belonging to a group
X7+ 616	Wanted	Z3 293	Other proper names
Q2.2 592	Speech acts	N5.1+ 254	Entire; maximum

The first striking observation with regard to the semantic domains is that *Pronouns*, *Existing*, *Geographical Names*, *Location and Direction*, *Government*, *Belonging to a group*, *Speech acts*, and *General Actions* are the shared ones. Nevertheless, the president puts more emphasis on *Strong Obligation/necessity* or *Probability*, and *Future time*, while the PM accentuates *Places*, *Numbers*, and *Other Proper Names*.

The semantic domain of *Geographical Names* is the leading notional lexical/semantic field for both corpora. Given the context of the study, it is obvious that the discourse is rich in the lingual units of this semantic domain. However, there is another factor to be considered – the context of use, which accentuates the individual differences. *Corpus_1>Viewing Word-SEM frequency profile* features 1585 hits of *Geographical Names* domain, out of which five most significant lemmas are GEORGIA (369), EUROPE (285), UKRAINE (161), RUSSIA (110), and BLACK SEA (23). As regards *Corpus_2>Viewing Word-SEM frequency profile*, it features 882 hits of the same semantic domain, out of which the five most frequent occurrences are GEORGIA (281), UKRAINE (94), EUROPE (65), RUSSIA (35), FRANCE

(14). The discursive use of *Geographical Names* is similar, for instance, in the excerpt below, President Zourabichvili stresses her personal endeavours to convince the European leaders in considering Georgia worthy of granting the candidacy.

The Location and Direction semantic domain is particularly recurrent across the corpora. This domain includes prepositions such as *ahead*, *towards*, *forward*, *within*, as well as verbs such as *stand*, *direct*, *lead*, *face*, *transfer*, and nouns such as *end*, *destination*, *course*, *direction*, *vector*, etc. The semantic load of this domain leads us to the conceptual metaphors (Lakof and Johnson 1980) framing Georgia's EU accession as a journey. According to the conceptual metaphor theory, the semantic domain of *journey* is used to conceptualize an abstract notion of accession or political integration, which requires metaphoric framing to be digested mentally and apprehended more easily. Georgia's EU accession as a journey is evident in both corpora. The president's speeches feature a considerable amount of journey metaphor:

- 1) Today, we **stand** together! We'll **stand** together tomorrow and **lead** our country into Europe. I make this promise to you as I have made the same promise to myself a long time ago.
- 2) Let us all come to our senses, let 's get into this new time together and show the **country direction** and **path**.
- 3) Negotiation, which is **the way to direct and immediate accession**. I want to take advantage of today 's day and wish Ukraine victory because , without Ukraine's victory, there will be no progress and no freedom.

Prime Minister Garibashvili also frames Georgia's aspiration via *journey* metaphor. Therefore, the political process for the country to become a member of the EU is conceptualized by both politicians alike:

- 1) By granting the European Perspective to Georgia, every speculation and assumption of where Georgia **stands**, has vanished. Europe has vocally said that Georgia belongs to Europe and the country will **accede to** the EU. **In this journey**, of course, the next stage is the status of an EU Candidate Country, which is a merely symbolic decision. Nevertheless, it is a significant one. That is why our governance did everything **towards this end** and continues doing its best.

The third semantic domain to focus on is *General Actions*, which could lead to the conceptualization of the European path as a dynamic process, something evolving or developing. The USAS semantic tagger, the General Actions field, is defined as “general/abstract terms relating to an

activity or action” (Rayson, Archer & Wilson 2002). From the perspective of cognitive grammar, energy and motion are realized through verbs that denote actions; hence, they imply the dynamic of the circumstances (Langacker 1987). The lemmas that belong to this semantic domain are frequent occurrences in both corpora: MAKE, PROCESS, COMMISSION, COMMIT, CREATE, IMPLEMENT, ACT, ENGAGE, SPREAD, SAFEGUARD, PREPARE, etc.

The next semantic domain to consider is *Belonging to a Group*. In the President’s speeches, the frequency of the lemmas residing within this domain is 740. The five most frequent lemmas from this semantic field are: SOCIETY, UNITY, TOGETHER, UNITED, and SOLIDARITY. For the Prime Minister the frequency of lemmas amounts to 304 with the five most reoccurring lemmas: PUBLIC, MEMBER, TOGETHER, TEAM, and SOCIETY.

As opposed to the President corpus, the PM’s discourse features the semantic domain of *Numbers*, which could be explained by his interest in financial and economic moments that a financially stable country would be more appealing to the European Union as a partner. The PM often stresses the importance of the economic growth Georgia has had for the past few years.

Out of the 15 most recurring semantic domains, *Strong Obligation or Necessity* (704) is more common for the President’s corpus. Modal auxiliaries *should*, *must*, *need* and the nouns e.g., *responsibility* and *promise* stand for the president’s obligation to be involved in building a better future for the country as well as stressing the need for a collective endeavor:

1. All of us **should** be united and act around one plan and one purpose. We **should** be working on a package of reforms that is mandatory if we seriously see out future in Europe.

In the Methodology section, I explained the focus on the notional words as they are the carriers of content in the narratives of the politicians. Nevertheless, I consider *Pronouns* worthy of attention too, as they promote the sense of collective or individual representation. Interestingly enough, *Pronouns* semantic domain is shared by both corpora with the absolute frequencies 7473 for Corpus 1 and 2964 for Corpus 2 (Table 4). I would single out the use of the collective pronoun *We* and a personal, egocentric pronoun *I*. Viewing Word _ SEM Frequency Profile in Wmatrix5 supplies relative frequencies as well.⁶ The relative frequency of Collective *We* in the

⁶ #LancsBox Report and Wmatrix outputs for Corpus 1 and Corpus 2 Semantic Domains Analysis are represented on Harvard Dataverse on the final page.
<https://doi.org/10.7910/DVN/Q2WQMD>

President's and Prime Minister's speeches is 1.58 and 1.33, respectively. Regarding the pronoun *I*, relative frequency 1.26 is related to the President's discourse and 1.25 to the Prime Minister's discourse. Seemingly, there is no significant difference between the uses of the *We-I* opposition. The President and the Prime Minister opt for *We-I* pronouns to stress either collective or individual, personal endeavors.

The Effects of Stance Taking

CADS tends to be effective in identifying trends regarding stance-taking. Both lexicometric and textometric analyses could supply quantitative as well as qualitative data in relation to the linguistic strategies of positioning stances on the specific object of investigation. In the methodological section, the importance of exploring the modal verbs along with other notional parts of speech is highlighted, as they could be the indexicals of stance-taking in the discourse. More precisely, epistemic stance could be marked via modal verbs: *will/would, may, might*.

Table 5. Relative and absolute frequencies of modal verbs indexing epistemic stance

	Corpus 1		Corpus 2	
	Absolute freq.	Relative freq.	Absolute freq.	Relative freq.
<i>Will</i>	468	0.76	133	0.47
<i>Would</i>	109	0.18	65	0.23
<i>May</i>	38	0.06	34	0.12
<i>Might</i>	24	0.04	1	0.00
Total	639		233	

Table 5 shows the distribution of modal verbs indexing stance in the discourse. The President positions her stances more effectively than the prime minister⁷ (the difference is significant $p<0.05$). Epistemic stance taking promotes discursive efficacy in the case of the Presidential addresses, as it deals with positioning facts, expectations, knowledge, or beliefs towards the EU accession with a certain degree of certainty concerning the object of discussion (Chindamo et.al. 2012, p.23). Excerpt 14 below is an interesting example from the President's addresses featuring all three types of stance-taking. The President evaluates recommendations as external ones that Georgia has to implement. This aspect is indexed via a modal auxiliary marker *may* and coupled with affective/ attitudinal modal *must* and *should*. The whole passage is bolstered through stylistic stance indicators, such as *importantly* and *ideally*.

- 2) **Importantly**, this process **should** unfold organically and without coercion. [...] We are aware of receiving recommendations, and at

⁷<https://ucrel-api.lancaster.ac.uk/cgi-bin/llsimple.pl?f1=639&f2=233&t1=64.321&t2=29.559>

times, there is a subtle sense that these suggestions **may** seem external. **Ideally**, our aspirations **should** align with these recommendations, as we strive to build a more developed, stronger, and equitable tomorrow driven by an intrinsic desire for knowledge and protection, independent of external influences. [...] We **must** strive for this, free from limitations or external pressures. True freedom is the ability to shape your own tomorrow.

Attitudinal or affective stance is also prominent in the President's speeches ($p < 0.05$)⁸, it stresses affect, including feeling, mood, disposition and attitude.

Table 6. Relative and absolute frequencies of modal auxiliaries indexing attitudinal stance

	Corpus 1		Corpus 2	
	Absolute freq.	Relative freq.	Absolute freq.	Relative freq.
<i>should</i>	204	0.33	21	0.07
<i>must</i>	91	0.15	18	0.06
<i>need</i>	75	0.12	14	0.05
Total	370		53	

The attitudinal stance is coupled with the collocate *European path* that Georgia has to follow to ensure a bright future. Excerpt 15 is the instantiation of psycho-social and emotive connections:

- 1) We all know that today, what this country **needs** the most is the path of Europe, to succeed on this **European path**. Everyone's role here is very important; everyone **should** do their part.

We can now examine the PM's stance-taking with regard to European journey. The excerpt (16) below is the passage from the PM's addresses on the EU's candidate perspective, which he evaluates as a positive achievement. Nevertheless, he stresses the importance of gaining the status of an EU Candidate Country, which could end polarization in the country. The commitment that needs to be delivered is proposed as a joint endeavor of the government, authorities, parliament and opposition, as the European perspective is a credit to his team:

- 1) We did! I already listed all these genuine achievements. European Perspective is I **would** say, a historic decision. Prior to that, you **may** recall a lot of skepticism existed in the capital cities of Europe. Some referred to the geographic distance, others criticized and pointed out many other issues, including the Russian occupation, and so on and so forth. By granting the European Perspective to Georgia, every speculation and assumption of where Georgia stands has vanished.

⁸ <https://ucrel-api.lancaster.ac.uk/cgi-bin/lisimple.pl?f1=370&f2=53&t1=64.321&t2=29.559>

Europe has vocally said that Georgia belongs to Europe and the country **will** accede to the EU. **In this journey**, of course, the next stage is the status of an EU Candidate Country, which is a merely symbolic decision. Nevertheless, it is a significant one.

This passage is a discursive framing of EU accession through epistemic stance-taking combined with a journey metaphor that cognitively reinforces position. As for attitudinal/affective stance-taking, it is realized through modal verbs *should* as well as notional verbs such as *hope* and *believe*:

- 2) Georgia knows that it belongs to Europe. Now, Europe **should** draw conclusions. It is essential for Georgia to get established within Europe, while the western nations - need to admit Georgia into their family ". These words as never before apply to the current reality. We **hope** and strongly **believe** that a historic process is getting its foundation laid today, which **will** lead to the accession of Georgia into the **European family**.

Overall, the PM's narrative blends epistemic modals and attitudinal stances with conceptual metaphors associating EU accession to a *journey* and the EU to a *family*. Taking quantitative aspect of analysis along with semantic instantiations into consideration, the term *path* in the President's speeches frames Georgia's EU accession as a journey. In addition to thus, the semantic examination suggests that this metaphor is used to highlight hope, progress, and European values. Conversely, the Prime Minister applies *path* more pragmatically, linking it to economic development and political stability within the context of EU membership.

A key finding in this study is the distinction between epistemic and deontic stance expressed via modal verbs. For instance, the use of *will* in the President's speeches indicates certainty and a forward-looking, optimistic stance on EU accession, while the Prime Minister frequently uses *may*, reflecting a more cautious and tentative position on the issue.

Conclusion

The aim of the present paper was to scrutinize how the two political figures, the President and the Prime Minister, discursively construct their disposition towards Georgia's EU accession, regardless of the political dissonance between them. CADS made it possible to reveal the tendencies, which are reflected in the idiolects of the politicians. Lexicometric analysis unveils the shared key verbal choices that are represented in the politicians' addresses: *Georgia, Georgian, Europe, European, Country, People, Ukraine* and *Government*. Nevertheless, the keyword *Georgia* appears to be

significantly prevalent in the PM Garibashvili's narrative. The concordance lines of the keywords shed light on the framing characteristics of the discourse. Therefore, the T-score analysis of EU-related words displays similar significant collocates; however, the President's narrative is more diverse in terms of varied types of collocates. The most significant difference is linked to the stance-taking in the narratives of the politicians. The President's positioning with regard to the path to EU membership is more significant through an effective use of epistemic, attitudinal and stylistic stance-taking, while the PM's narrative features the combination of epistemic and attitudinal stance-taking with a *journey* and a *family* metaphor. For Irakli Garibashvili sovereignty and economic strength of Georgia is intertwined with the process of EU accession. In effect, the discursive constructions of the EU accession in relation to politicians' positions and stances appear, on the evidence presented here, more alike irrespective of the alleged political discord.

Conflict of Interest: The author reported no conflict of interest.

Data Availability: All data are included in the content of the paper.

Funding Statement: The author did not obtain any funding for this research.

References:

1. Agenda.ge. (2023). "Ruling party official claims President 'strengthened false campaign' of UNM opposition in Parliament address." March 31. <https://agenda.ge/en/news/2023/1301>
2. Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum. doi:10.5040/9781350933996
3. Baker, P. (2010). *Corpus Methods in Linguistics: A Practical Introduction*. Cambridge University Press.
4. Biber, D., Conrad, S., & Reppen, R. (1999). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
5. Brezina, V., Weill-Tessier, P., & McEnery, T. (2020). #LancsBox 5.x and 6.x [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
6. Brezina, V., Timperley, M., & McEnery, T. (2018). #LancsBox 4.x [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
7. Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), pp. 139-173.
8. Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge: Cambridge University Press.

9. Brown, K., and A. Mondon. (2021). "Populism, the Media, and the Mainstreaming of the far Right: The Guardian's Coverage of Populism as a Case study." *Politics* 41 (3), pp. 279–295. <https://doi.org/10.1177/0263395720955036>.
10. Chindamo, M., Allwood, J., and Ahlsén, E. (2012). "Some Suggestions for the Study of Stance in Communication," 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, Amsterdam, Netherlands, pp. 617- 622, doi: 10.1109/SocialCom-PASSAT.2012.89.
11. Firth J. R. (1957). *Papers in linguistics 1934-1951*. Oxford University Press.
12. Fitzgerald, J. (2014). *Discourse and the Politics of Migration: A Corpus-Assisted Study of British Political Rhetoric*. Routledge.
13. Anonymous, Anonymous, 2024, "Corpus 1 - Salome Zourabichvili's Speeches & Corpus 2 - Irakli Garibashvili's Speeches", <https://doi.org/10.7910/DVN/Q2WQMD>, Harvard Dataverse, V1
14. Hunston, S., & Thompson, G. (2000). *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Routledge.
15. Kolb, I. (2018). *Framing the European Union: Political Discourse in the Context of EU Enlargement*. Palgrave Macmillan.
16. Lakoff, G., and Johnson, M. (2003) *Metaphors We Live By*, Chicago: University of Chicago Press.
17. Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford: Stanford University Press.
18. Lischinsky, A. (2018). Overlooked text types: From fictional to real-world discourses. In C. Taylor, & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (1st ed.), pp.60-82. Routledge. <https://doi.org/10.4324/9781315179346>
19. Mautner, G. (2019). A research note on corpora and discourse: Points to ponder in research design. *Journal of Corpora and Discourse Studies*, 2, 1–13. DOI: 10.18573/jcads.32
20. Nelson, M. B. (2016). *The Political Discourse of the European Parliament: A Corpus-Assisted Comparative Study*. John Benjamins.
21. Nikisianis, N., Siomos, T., Dimitroulia, T., Stavarakakis, Y., & Markou, G. (2019) In *Discourse, Culture and Organization: Inquiries into Relational Structures of Power*, pp.267–295. Palgrave Macmillan, London.
22. Palmer, F. R. (2001). *Modality and the English Modals* (2nd ed.). Longman

23. Partington, A., & Marchi, A. (2015). Using corpora in discourse analysis. In D. Biber, & A. Marchi (Eds.), *The Cambridge Handbook of English Corpus Linguistics*, pp.216-234. Cambridge: CUP
24. Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. Amsterdam/Philadelphia: John Benjamins
25. Politico. (2023). "Georgian parliament fails to impeach president over unsanctioned EU visits." October 18.
<https://www.politico.eu/article/georgia-parliament-fail-impeach-president-salome-zourabichvili-unsanctioned-eu-visits/#:~:text=Georgia's%20ruling%20party%20failed%20in,the%20government%20to%20do%20so.>
26. Rayson, P., Archer, D., and Wilson, A. (2002). *Introduction to the USAS category system*.
27. Rayson, P., and Garside, R. (2000). "Comparing Corpora using Frequency Profiling." In *The Workshop on Comparing Corpora*, pp. 1–6, Hong Kong, China. Association for Computational Linguistics.
28. Richardson, J. E. (2006). *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Palgrave Macmillan
29. Shenhav, S. R. (2006). Political Narratives and Political Reality. *International Political Science Review /Revue Internationale de Science Politique*, 27(3), 245–262.
<http://www.jstor.org/stable/20445054>
30. Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam and Philadelphia: John Benjamins.
31. Wodak, R., De Cillia, R., Reisigl, M., & Liebhart, K. (2009). *The Discursive Construction of National Identity*. Edinburgh University Press.

Appendix

Lancsbox Output

Table 1. Corpora

Name	Language	Texts	Tokens	Additional information
Corpus 1	English	69	64,321	Types: 6,025 Lemmas: 5,648
Corpus 2	English	28	29,559	Types: 3,974 Lemmas: 3,489

Table 2. Types in Corpus 1

ID	Type	Absolute frequency (Relative frequency)	Dispersion (CV)
1	the	3632 (564.668)	0.247
2	and	2471 (384.167)	0.227
3	of	2033 (316.071)	0.353
4	to	1977 (307.365)	0.268
5	is	1306 (203.044)	0.385
6	that	1270 (197.447)	0.490
7	in	1146 (178.169)	0.363
8	this	1118 (173.816)	0.462
9	a	1035 (160.912)	0.406
10	we	973 (151.273)	0.570
11	i	788 (122.511)	0.799
12	our	751 (116.758)	0.575
13	for	748 (116.292)	0.411
14	it	722 (112.249)	0.509
15	not	621 (96.547)	0.597
16	be	565 (87.841)	0.622
17	will	517 (80.378)	0.787
18	are	445 (69.184)	0.538
19	you	443 (68.873)	0.871
20	have	439 (68.251)	0.683
21	on	439 (68.251)	0.556
22	as	390 (60.633)	0.627
23	all	389 (60.478)	0.623
24	with	366 (56.902)	0.553
25	georgia	364 (56.591)	0.663
26	european	359 (55.814)	0.973
27	has	342 (53.171)	0.864
28	us	304 (47.263)	0.665
29	europe	295 (45.864)	1.289
30	but	294 (45.708)	0.748
31	country	290 (45.086)	0.807
32	today	286 (44.464)	0.723
33	its	277 (43.065)	1.062
34	by	262 (40.733)	0.639
35	what	258 (40.111)	1.164
36	from	230 (35.758)	0.822
37	or	226 (35.136)	0.736
38	very	221 (34.359)	1.029
39	one	218 (33.893)	0.858

ID	Type	Absolute frequency (Relative frequency)	Dispersion (CV)
40	also	205 (31.871)	0.750
41	should	204 (31.716)	1.035
42	which	198 (30.783)	0.798
43	their	197 (30.628)	0.956
44	there	191 (29.695)	1.044
45	was	191 (29.695)	1.099
46	more	185 (28.762)	1.011
47	no	185 (28.762)	0.876
48	at	181 (28.140)	0.781
49	want	178 (27.674)	1.263
50	future	175 (27.207)	1.052

Table 3. Types in Corpus 2

ID	Type	Absolute frequency (Relative frequency)	Dispersion (CV)
1	the	1658 (560.912)	0.281
2	and	1144 (387.023)	0.230
3	of	1142 (386.346)	0.441
4	to	936 (316.655)	0.376
5	our	572 (193.511)	0.518
6	in	569 (192.496)	0.319
7	a	510 (172.536)	0.567
8	is	464 (156.974)	0.655
9	that	406 (137.352)	0.524
10	we	375 (126.865)	0.593
11	i	361 (122.129)	0.619
12	georgia	290 (98.109)	0.896
13	it	276 (93.373)	0.843
14	you	257 (86.945)	0.838
15	for	255 (86.268)	0.703
16	this	229 (77.472)	0.650
17	country	208 (70.368)	0.884
18	with	206 (69.691)	0.602
19	as	200 (67.661)	0.521
20	by	193 (65.293)	0.653
21	are	190 (64.278)	0.630
22	on	182 (61.572)	0.772
23	have	148 (50.069)	0.915
24	was	148 (50.069)	1.195
25	all	136 (46.010)	1.075
26	will	135 (45.671)	0.534
27	be	128 (43.303)	0.880
28	wish	126 (42.627)	1.308
29	not	124 (41.950)	1.037
30	people	111 (37.552)	1.080
31	european	104 (35.184)	1.937
32	has	104 (35.184)	0.796
33	which	103 (34.846)	1.356

ID	Type	Absolute frequency (Relative frequency)	Dispersion (CV)
34	an	99 (33.492)	0.901
35	ukraine	95 (32.139)	1.938
36	also	91 (30.786)	1.321
37	their	90 (30.448)	1.503
38	war	90 (30.448)	1.470
39	at	88 (29.771)	1.094
40	eu	86 (29.094)	1.870
41	government	86 (29.094)	0.889
42	from	83 (28.079)	0.843
43	they	81 (27.403)	2.046
44	world	81 (27.403)	1.548
45	countries	77 (26.050)	1.109
46	my	77 (26.050)	0.997
47	well	77 (26.050)	1.213
48	about	74 (25.035)	1.504
49	me	73 (24.696)	1.013
50	dear	72 (24.358)	0.995