

To cite text:

Katona, Eszter, Renáta Németh, and Anna Unger. 2025. "The Evolution of the 'Carpathian Basin' Discourse in the Hungarian Parliament (1998–2020)." *Philosophy and Society* 36 (2): 437–460

Eszter Katona
Renáta Németh
Anna Unger

THE EVOLUTION OF THE 'CARPATHIAN BASIN' DISCOURSE IN THE HUNGARIAN PARLIAMENT (1998-2020)

ABSTRACT

We explore the use of the term "Carpathian Basin" in the Hungarian Parliament 1998-2020. The "Carpathian Basin" is a term of Hungarian geography, historically used to justify Hungary's territorial claims during the interwar period. While it was absent from official discourse for decades, it has recently gained traction among Hungary's politicians. By processing 1525 speeches, we examine changes in the discourse of three major political blocs (right-wing nationalist, liberal/left, and Fidesz) to capture the linguistic representation of the dynamics of political polarization, and to identify changes in politically driven identity patterns and framing differences.

Our paper has both methodological and substantive relevance. The methodological novelty is that we apply methods that allow automated processing of large text corpora without reading them, in a field where previously mainly qualitative approaches were used. We show that it is possible to detect changes in framing in an automated way without human coding. From a substantive point of view, our study focuses on the linguistic features of an important concept that differ from one political ideology to another.

We employ both supervised and unsupervised modeling approaches. The supervised classification was used to examine changes in the polarization of discourse, while the unsupervised tool (Structural Topic Model) supported a more nuanced, qualitative interpretation of the results. According to our results, the political ideology of the speakers of the speeches can be predicted more effectively, i.e. a kind of polarization-growth can be detected, while at the same time the deeper analysis shows that parallels can be detected in the changing discourse of different ideological sides. One such common feature is a more concentrated focus on the Hungarian nation, as opposed to neighboring peoples and the European Union. We also found discourse traits of both the left's rapprochement with the right (as an imprint of the left's opening to Hungarians beyond the borders after 2010) and the moderation of the far right.

KEYWORDS

natural language processing, Carpathian Basin, Hungarian Parliament, ideological divides

Eszter Katona: Assistant Professor, Faculty of Social Sciences, Eötvös University; katona.eszter@tatk.elte.hu
Renáta Németh: Professor, Faculty of Social Sciences, Eötvös Loránd University; nemethr.web.elte.hu
Anna Unger: Associate Professor, Faculty of Social Sciences, Eötvös University; unger.anna@tatk.elte.hu



“So when we talk about Hungarians beyond the borders, we should always put this adjective in brackets; let’s talk about Hungarians, let’s talk about Hungarians living in the Carpathian Basin and in the world, let’s talk about our annexed or separated brothers and sisters, but I think that we cannot go in the direction of what, among others, the policy of abandoning national interests has embodied [...]”

Dr Tamás Gaudi-Nagy, parliamentary speech
far-right party Jobbik, 2013

1. Introduction

The quote from the motto shows the stakes of our research, the concept of Carpathian Basin is not a neutral term, as it is loaded with geopolitical intentions, and actors in political discourse are conscious of their choice of words (see the proposal of ‘Hungarian living in Carpathian Basin’ instead of the neutral ‘Hungarians beyond the border’). As Megoran (2006) defined, a political geographical imagination is ‘the way in which people experience, conceive of, or desire a particular configuration of the relationship between space, ethnicity, nation, and political community.’

The research is at the intersection of sociology and regional studies: the use of the term Carpathian Basin in public policy discourses in the Hungarian Parliament in the period 1998-2020. In our analysis, we use databases of the ELTE Research Center for Computational Social Science (rc2s2.elte.hu/en), on which the most important preprocessing and cleaning tasks have already been carried out by the research group. The aim of our research is to identify changes in (geo)political intentions through the changing use of the term, and, in this context, the changing relationship between ideological sides.

The Carpathian Basin as a landscape unit is bordered by the Alps (in the west), the Carpathians (in the north, east and south), the Dinaric Mountains and the Šumadija Mountains (in the south). It includes Hungary and Slovakia, as well as parts of Croatia, the Czech Republic, Romania, Serbia and Ukraine; the country shares borders with seven different countries within this area (Gaudenyi, Mihajlović 2022).

The ‘Carpathian Basin’, the most important Hungarian geographical concept, is linked to the historical process of (self-)recognition and (self-)definition of Hungary (Balogh, 2021). Although the concept of the Carpathian Basin was rarely used between 1960 and 1981, it has since been increasingly back on the political agenda, becoming a central element of the governmental narrative. Based on Balogh (2021) the concept played a vital role in demarcating Hungary during disputes over its boundaries and legitimizing territorial claims lost during the interwar period. While the term fell out of use between 1960 and 1981, it resurfaced after political changes, fueled by Hungary’s interest in the fate of ethnic relatives in neighboring countries. Counter-discourses, both

foreign and domestic, emerged, with the latter addressing overlooked issues in mainstream Hungarian geography and politics. Foreign counter-narratives mirrored the interests of their respective countries.

Nowadays, there is a noticeable resurgence in the usage of the term “Carpathian Basin” within the government narrative as “an unquestionably cohesive political geographical entity”. This strategic use of language is underscored by the presence of a dedicated program on public television titled “Carpathian Basin News”, in which the news is about “Hungary, the Carpathian Basin and the world at large”, and even the weather forecast is for the whole of the Carpathian Basin.

The concept of the Carpathian Basin appeared several times in Hungarian parliamentary politics and legislation since the 1989 regime change: in a nature conservation (environment protection) context, in a historical context, and also in connection with responsibility towards Hungarians living beyond the borders (Scott, Hajdú 2022). The significance of this concept in Hungarian domestic politics and thus in party competition is difficult to understand without the development of the issue of national politics, a special concept that refers to the responsibility and actions of the Hungarian state and government towards the Hungarians who live beyond the borders. Therefore, in order to understand why the Carpathian Basin became a political symbol in Hungarian politics and how it has evolved, we briefly review how Hungarian national politics changed after the regime change in 1989.

The first prime minister of the new-born Hungarian democracy, József Antall, declared after his election that he wanted to be the prime minister of fifteen million Hungarians in spirit. This sentence perfectly symbolizes the challenges of Hungarian national politics and the significance of the Carpathian Basin as a political and geographical concept. The fifteen million Hungarians naturally included the community of millions of people living outside Hungary who were not Hungarian citizens but were still of Hungarian nationality, for whom the Hungarian state and its leaders, as leaders of the national political community, were responsible. However, the phrase “in spirit” reflected the unique situation that this responsibility was primarily ideological and symbolic, as there was no real legal relationship between Hungarians living beyond the borders and the Hungarian state. Although the Hungarian opposition parties at the time considered this declaration by the prime minister to be worrying, the issue of Hungarians living beyond the borders was not fundamentally a subject of domestic political debate (Pogonyi 2017). An important part of this was a series of state treaties concluded with certain neighbouring countries, in which the Hungarian state reaffirmed the decisions laid down in the peace treaties of 1920 and 1947 that ended the world wars, thereby renouncing its territorial revisionist aspirations. Although some right-wing politicians criticized the governments that concluded these treaties, a broad consensus emerged among the parliamentary parties that the country had no territorial revisionist claims, as these would not only jeopardize the international order and the principles of international law, but also the security of Hungarians living beyond the borders,

at a time when anti-Hungarian sentiment was already an important factor in Romanian and Slovak domestic politics (Mák 2000).

The consensus among mainstream parties on the situation of Hungarians living beyond the border, their relationship with Hungary, and the political and legal scope of action available to the Hungarian state was essentially broken by the status law adopted in 2001. This legislation did not grant dual citizenship, but it did grant special legal status to Hungarians living beyond the border, which, among other things, made it easier for them to find work and entitled them to certain social benefits (Udrea 2013). The potential budgetary and social spending implications of this legislation were strongly criticized by the Socialists and Liberals during the 2002 election campaign, which can be seen as a turning point which broke the previous consensus and highlighted the fundamental differences between the left-liberal and right-wing political camps on the issue of Hungarians living beyond the border. This confrontation was reinforced by the 2004 referendum on dual citizenship: the right wing, led by Fidesz, campaigned for easier citizenship for Hungarians living beyond the border, while part of the social-liberal camp supported abstention from voting and another part supported a no vote, with only a few left-wing and liberal politicians endorsing the yes vote (Németh et al, 2024). The invalid referendum left a serious mark on Hungarian politics: from then on, the right wing, which called itself the national side, presented itself as the main representative of the interests of Hungarians in the Carpathian Basin. Fidesz, which came to power in 2010, also expressed this by healing the wounds caused by the failed 2004 referendum and creating the possibility of dual citizenship. Contrary to earlier promises, the parliament also extended voting rights to dual citizens (in practice, the requirement of residence in Hungary for voting rights was abolished). It is also from this moment that national politics became a prominent element of governance; while in the years before 2010, the agencies and support programs dealing with Hungarian minorities were usually referred to “Hungarians beyond the border,” in the period after 2010, national politics was usually linked to Hungarians of the Carpathian Basin, and references to living beyond the border almost completely disappeared. To sum up the period after 1989, it seems that while in the first decade after the change of regime there was basically political consensus among the parliamentary parties on the situation of Hungarians living beyond the border and the Hungarian state’s responsibility towards them, after the 2001 status law, the Hungarians of the Carpathian Basin and the national policy related to them increasingly became the subject of domestic political struggles, and after the political turnaround of 2010, both national policy as a concept and instrument and the Hungarians of the Carpathian Basin as a cross-border political community transcending national borders.

Our paper has both methodological and substantive relevance. The methodological novelty is that we apply methods that allow automated processing of large text corpora without reading them, in a field where previously mainly qualitative approaches were used. We show that it is possible to detect change in framing in an automated way without human coding. From a substantive

point of view, the significance of this research lies in the fact that while the research of other authors on the topic, e.g. Balogh (2021), is based on the texts of geographical authors, this analysis examines the official political public sphere, and thus directly the geopolitical self-image of the country. The emergence of the term Carpathian Basin in Hungarian politics is significant not only for the political geographical approach, but also for political discourse and policy making. What Balogh (2021) presents in relation to the history of the development and political use of the term is significant not only because it summarizes the meanings and knowledge that have been accumulated on the term, but also because we can see that the Carpathian Basin as a term, incorporating these layers of meaning, becomes a symbol in everyday Hungarian life. Not only in politics, not only in other cultural or entertainment arenas important to society, but even in education, or, as mentioned above, in everyday phenomena such as weather reports or other map representations of territorial issues.

In the world of politics, the role of language and symbols is particularly important. These symbols, born from political action, bring otherwise distant institutions closer to individuals, endow politics—and the specific phenomena, institutions, or actors they represent—with a kind of mythical significance. (Edelman 2004). According to Cohen, “It is the very essence and potency of symbols that they are ambiguous, referring to different meanings, and are not given to precise definition” (Cohen 1979: 87).

A political symbol can be anything “that represents a political institution, hierarchy, movement, belief, or ideology”, i.e. anything can serve as a political symbol if people associate it with some aspect of political reality (Theiler 2017). A political symbol has three main functions: first, it provides a simplified and condensed meaning of certain political procedures or principles, making them easier to internalize. Second, it allows institutions and emotions to be linked, and this affective symbol is also necessary for citizens to relate to politics. Third, there is a cumulative-representative function, in which a symbol embodies, represents and links views, values and political expectations. Symbols can help form larger social groups into a common political camp, even if there is no universal agreement among them. It is important to note that for this grouping element to be effective, the symbol must remain multi-layered and somewhat ambiguous. This vagueness allows the political community it unites to be as broad as possible (Theiler 2017).

In the case of the Carpathian Basin, all three functions are clearly outlined in the Hungarian politics. Its meaning is not only a concrete geographical unit, but as a symbol of the heroic historical past, of the once united (now mutilated) nation, of the Hungarian nationality as an organically (geographically) cohesive political community. The symbol, which always has an ambiguous meaning, allows us to widen the scope of specific, legally very well-defined political action and institutions, to expand their legal-institutional boundaries, or to make these boundaries fade away or become meaningless.

In the previous phase of our research (Németh, Katona et al, 2025) we investigated the political determination of the discourse in the Hungarian

Parliament. Our results showed a resurgence in the use of the term 'Carpathian Basin'. The discourse serves as a battleground for different political ideologies in the Hungarian Parliament. Left-liberals use it neutrally to describe a diverse region, focusing on institutions and interests. In contrast, right-wing narratives often portray the Carpathian Basin as a unified geographic entity, with some questioning territorial legitimacy, while others emphasize values, emotions, and culture, offering a collective identity through its use.

In the present study we will focus on the change of this discourse over time (1998-2020). After the war, the term "Carpathian Basin" became a taboo term until the early 1980s (Balogh, 2021). Since then the term has gradually come to the fore again in recent decades (Hajdú, 2018) and now plays a role in everyday political discourse and in the construction of national identity (Scott & Hajdú, 2022). Our study focuses on the last two decades of this process and the parliamentary discourse as a whole, and focuses on the detection of change.

We use computational tools to process the speeches containing the phrase from all the parliamentary speeches of the two decades under study. From an epistemological point of view, the empirical social research trend based on text also has a history of several decades. One of these antecedents is the narrative turn in the social sciences (Goodson, Gill, 2011), which primarily sought to provide a methodological alternative to positivist research, seeing in the text the possibility of a self-reflexive approach to social phenomena. Several other, not unrelated, discipline-specific antecedents of the emergence of discursive approaches in social research include the science of language and politics, where language is seen as an indispensable tool for political action (Wodak, Forchtner 2017; Müller, 2008), or the science of language and history, where the discursive construction of national history (Wodak, 2010) or national identity (Wodak et al., 2009) is focused on. In our approach, the text is not presented as an object to be read and understood, but rather as an input to automated methods without actually being read by anyone. The use of NLP in social research is thus related to this recently popular 'text as data' approach, where text is the input to computer algorithms as an ordered, well-structured, numerical database.

2. Research questions

The research divides parliamentary speeches from 1998 to 2020 by the date of the 2010 election, thus comparing two time periods. Our aim is to investigate the change in speech styles of the three major political entities (right-wing nationalist, liberal/left and Fidesz) in relation to the discourse of the Carpathian Basin:

RQ1. Can we capture a change in the linguistic representation of political polarization? We address this research question using machine learning classification at first. We train our models on a subsample and evaluate their performance on another set of data over time. Our focus is on the results in terms

of confusion, where we observe the model's capability to distinguish between parties at two different time points.

RQ2. What changes in content can be detected across ideologies? (Changes in party discourse) Here we use Structural Topic Modelling (STM) to observe thematic changes through the word usage of the three different ideological blocks.

RQ3. What changes can be detected in the discourse of each political-ideological bloc? Is there divergence or homogenization? (Dynamics of differences between parties) In answering this research question, we will draw on the STM to examine thematic shift / changing framing, measured by word choice, in a comparison of political blocs. There are two more general background processes that are related to this research question.

1. *"Two-way mainstreaming"* means the radicalization of the traditional right and the strategic refinement of the extreme. This leads to a blurring of the boundaries between the classical far right and the centre right, and extreme narratives become part of the public discourse and even the norm (Hunyadi, 2022). A good example of the strategic refinement of extremism is the moderation of Jobbik¹ which has been observed since 2013: the party's campaign messages have been significantly moderated, statements have been made distancing themselves from their racist and anti-Semitic manifestations, and the declared goal has become to become a "real people's party" (Kovarek, Farkas 2017).

2. *The left-wing's turn in its attitude to national identity.* Since the regime change, the left has been constantly on the defensive against the right in the discourse on national identity. The socialists began to focus on renewing their attitude towards national identity after the 2010 election defeat, which brought a radical turnaround in the assessment of the attitude towards Hungarians beyond the borders. This consciously prepared turnaround is based on the MSZP's² attempt to get out of the box into which Fidesz and Jobbik tried to force the socialists. The policy of reaching out to Hungarians beyond the border can be seen as the symbolic beginning of this shift (Friedrich Ebert Stiftung-Policy Solutions, 2013).

Beyond answering the research questions, the paper highlights the relevance and innovative nature of using the Structural Topic Model (STM) in political discourse analysis. The STM is an extension of the Latent Dirichlet Allocation (LDA) topic model, developed by social researchers for their own purposes (Roberts et al., 2014, 2019). Unlike LDA, the model allows us to add meta-variables to the model in the form of categorical variables that explain the frequency and content of topics. The relevance of the method to social science is illustrated by the fact that its creators, political scientist and sociologist

1 Jobbik is a Hungarian political party that has experienced ideological shifts, initially known for its right-wing and nationalist stance, but later undergoing a moderation process to present a more centrist and moderate image.

2 The Hungarian Socialist Party (Magyar Szocialista Párt, MSZP) is a center-left political party in Hungary, historically associated with socialist and social democratic principles.

Roberts and colleagues (2014), applied the tool to open-ended questionnaire responses to immigration questions to show, for example, the effect of party preference on the prevalence and framing of immigration fears topics. The innovation has thus transformed the topic model from an exploratory tool into a method that allows testing associations with important characteristics like the party preference above. Within the meta-variables, we can specify which ones influence the frequency (prevalence/population) and which ones influence the content (framing) of the topics. A further difference with the LDA topical model is that it removes its restriction on the uncorrelatedness of topics.

3. Data and methods

A total of 1,525 speeches were analyzed. We only examined the parliamentary speeches that include the term 'kárpát_medenc*', the preprocessed, standardized Hungarian form of the Carpathian Basin.

In our analysis, we explore the temporal dimension through two categories, using the 2010 national elections as a dividing point. Fidesz changed the government in that election year, and from then on, the government radically transformed the public law system and declared its system-building intentions from the beginning. Since 2010, there has been a hegemonic party system in Hungary, where the governing party, Fidesz won $\frac{2}{3}$ majority in the parliament in four consecutive elections. The electoral system was changed fundamentally in 2011 and 2012, and since then, several minor changes like fine tunings have been implemented to guarantee the victory for Fidesz. 2010 was a turning point also from that perspective that, as mentioned before, Fidesz-led parliament implemented dual citizenship with voting rights for ethnic Hungarians, which not only extended the Hungarian electorate but put the topic of national politics, including financial supports for kin-Hungarians outside of Hungary into the agenda of the Hungarian parliament.

Besides the temporal dimension three groups of parties active during the observed period were identified: Fidesz; right-wing nationalist (Jobbik, MDF, FKGP, KDNP, MIÉP); and left-wing liberal (SZDSZ, MSZP, DK, LMP, PM, Együtt.). The last two political groups, especially the parties in the left-liberal bloc, have diverse ideologies. They don't have a clear ideological stance; instead, they are used in relation to Fidesz, the dominant party. Table 1. shows the number of speeches made by each of the three blocks at each of the two time points.

	25.06.1998 11.04.2010	12.04.2010. - 23.11.2020
Fidesz	153	397
right-wing nationalist	161	490
left-wing/liberal	161	163

Table 1. The number of speeches by ideological block and date

In our analysis, we consider parliamentary speeches as the unit of analysis. Our research group Research Center for Computational Social Science developed a convention-based cleaning and preprocessing pipeline for the Hungarian language, adjusting to the research questions. The different stages of this process were: character standardization, filtering of the texts according to certain aspects, word cleaning (deletion of non-linguistic signs, irrelevant parts), defining and removing stop words (as those words are too common), word lemmatization (deleting conjunctions) and unification (name recognition, essentially the unification of proper names). We used preprocessed data from our research group.

Natural Language Processing (NLP) is a Machine Learning technique used for understanding, interpreting, and also generating texts. In the context of our research, NLP plays an important role in extracting insights from parliamentary speeches, by using different models and algorithms.

Our research questions are addressed using two approaches: supervised learning and unsupervised learning. In supervised learning, we use algorithms to classify speeches into one of three ideological blocks. This approach involves training algorithms on our labeled corpus, where each speech is associated with one of the three political blocks. To train the models, we divide our corpus into two different datasets: the training data (80% of the texts) and the test data (20% of the texts). The two datasets will be used to create the model and then to evaluate the model's performance. The algorithm used for classification learns textual patterns and correlations in the training data to make predictions about the ideological affiliation of new, previously unseen speeches from the test data. As we know the real categories, we can compare the predicted labels and the true labels for each speech. With various measures, we can describe the efficiency of the model.

On the other hand, we also employed an unsupervised learning technique, the Structural Topic Model (STM), to analyze the topics of parliamentary speeches. In contrast to supervised learning, unsupervised learning does not need labeled data. These models identify patterns and latent structures in the texts without any prior knowledge of categories, themes. STM is used to discover latent topics within the speeches and to uncover how these topics are related to each other, to the time dimension and to the ideological blocks.

The key difference between the two approaches lies in the nature of the learning process. Supervised learning requires labeled data and aims to classify new instances into predefined categories, while unsupervised learning aims to uncover hidden patterns in the data. By employing both supervised and unsupervised learning techniques, our research provides a comprehensive analysis of parliamentary speeches, shedding light on ideological differences through thematic content.

4. Results

4. 1. Multinomial logistic regression

During classification we also focus on the reasons behind classification. The interpretation of classification models is extremely important: on the one hand, it helps us to understand how the model makes its decisions, how it classifies a given speech into different categories, and on the other hand, it provides us with deeper information about the framing of the Carpathian Basin in the Hungarian Parliament.

We used multinomial logistic regression for the classification, with default settings of the model. Logistic regression (Eisenstein 2019) is used for binary classification tasks, where the outcome variable has two possible outcomes. The model can be extended to handle multi-class classification problems. This allows us to use the model on our task: classifying parliamentary speeches into three different ideological blocks. Logistic regression models the probability that a given observation belongs to a particular class. In our research, each ideological block represents a class, and logistic regression gives an estimation of the probabilities of a speech belonging to each block. An important advantage of this model is that it provides easily interpretable results, allowing us to understand the influence of individual textual parts on the classification outcome. As mentioned before, interpretability is important for our research, as we would like to understand which terms are driving the classification of speeches into different ideological categories.

Our original sample is highly unbalanced, with Fidesz and the right-wing nationalist group being over-represented at the second time point. To deal with the imbalance, these categories were randomly undersampled (Table 2.). Random undersampling removes data points from the over-represented class, potentially leading to the loss of valuable patterns and relationships, weakening the model's ability to generalize, and reducing its capacity to learn meaningful patterns for both dominant and minority classes. However, in our case, it is a useful and unavoidable approach to address the imbalance, though it may sometimes compromise the robustness of the model.

	25.06.1998 - 11.04.2010		12.04.2010. - 23.11.2020	
	original	undersampled	original	undersampled
Fidesz	153	153	397	160
right-wing nationalist	161	161	490	160
left-wing/liberal	161	161	163	163

Table 2. The number of speeches by ideological block and date after undersampling

Outcome variable was the ideological block, and we used Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer on the preprocessed texts of the speeches. TF-IDF is a technique which reflects the importance of a word

in a document relative to all documents in the corpus. By using TF-IDF, we could underweight the influence of common words that may not carry high discriminatory power. We used ‘unigrams’ and ‘bigrams’ as features for prediction, allowing our model to capture both individual words and pairs of adjacent words.

To predict the three categories, we divided the corpus for the two time points and fitted separate models for the two time points. We used 80% of the data for training the model and 20% of the data for testing the classification performance. We use the confusion matrix to analyze the performance of the classification model. The matrix displays the counts of true positive, true negative, false positive, and false negative predictions. This allows for a comprehensive assessment of the model’s ability to correctly classify speeches. We measured the model performance accuracy. Accuracy measures the proportion of correctly classified instances among all instances in the dataset. The model performs no better or not much better than random classification. However, the model performance improved for the second time point compared to the first. Based on this improvement in the model, we can infer an increasing polarization in the framing of the Carpathian Basin in the parliamentary speeches.

time period	confusion matrix				accuracy	
25.06.1998 - 11.04.2010	Predicted category				0.49	
	Actual category		Fidesz	Left-liberal		Right-wing nationalist
		Fidesz	6	11		19
		Left-liberal	3	21		9
		Right-wing nationalist	2	4		20
12.04.2010. - 23.11.2020	Predicted category				0.58	
	Actual category		Fidesz	Left-liberal		Right-wing nationalist
		Fidesz	18	12		10
		Left-liberal	1	21		6
		Right-wing nationalist	1	10		17

Table 3. Results of the classification

Based on the confusion matrix (Table 3.), looking at the first time point, we can see that the model fails to identify speeches related to Fidesz politicians, sorting speeches to the right or left more often than correctly. It categorizes the left and right correctly, classifying relatively few wrongly. We can conclude that the model tends to classify speeches held by left-liberals as right-wing

nationalist speeches. This is also because the left-liberal bloc contains a relatively diverse party palette.

In the second period, the recognition of Fidesz politicians' speeches improves, but the classifier still assigns a lot of speeches to both as left-liberal and right-wing nationalist. The model classifies left-liberal speeches well. However, it assigns many of the right-wing nationalist speeches to the left. The model's incorrectness in this category is consistent with the moderation of right-wing parties discussed earlier.

4.2. STM

Topic models are used to discover hidden thematic structures in a collection of documents. One of the most used topic models is the Latent Dirichlet Allocation (LDA). Topic models assume that 1) each document in the corpus is generated by a mixture of underlying topics that reflect the thematic structure of the document; 2) each topic is represented as a distribution of words in the vocabulary. This distribution identifies the words that are most likely to occur in documents addressing a given topic. The aim is to find the best way to explain how these topics and words are distributed throughout the documents. This helps us to understand what the main concepts are and how important they are throughout the collection.

The Structural Topic Model (Roberts et al. 2019) is a subtype of topic models. STM allows the researcher to estimate the topics' relationship to document metadata, which can influence both the content (the words used within a topic) and prevalence of topics. Moreover, it allows correlation between topics, i.e. it does not make the unrealistic assumption that the appearance of topics in individual speeches is independent of each other. The advantages of the STM include its ability to analyze changes over time, thus we can observe the evolution of topics in documents with the inclusion of the three ideological blocs in the analysis. This opens up the possibility of exploring different ways of speaking and framing. In our analysis the structural topic model is used for validation and interpretation: to support the classification models. The inclusion of the three ideological blocs in the analysis not only enriches our understanding of topic evolution, but also provides deeper insights into how different perspectives shape discourse over time. The STM enables a more detailed examination of ideological shifts and allows us to identify subtle variations in language use and framing strategies across different ideological affiliations.

To determine the number of topics, we opted for a topic number of 7 using a qualitative and quantitative (see Fig. 1) approach.

The model has two important covariates: prevalence and content. Prevalence helps account for variations in the occurrence of topics based on external factors, which are in our analysis date and political bloc. The combination of the same variables was also used as content variable. This covariate provides insights into how external factors shape the content and language style associated with a particular topic.

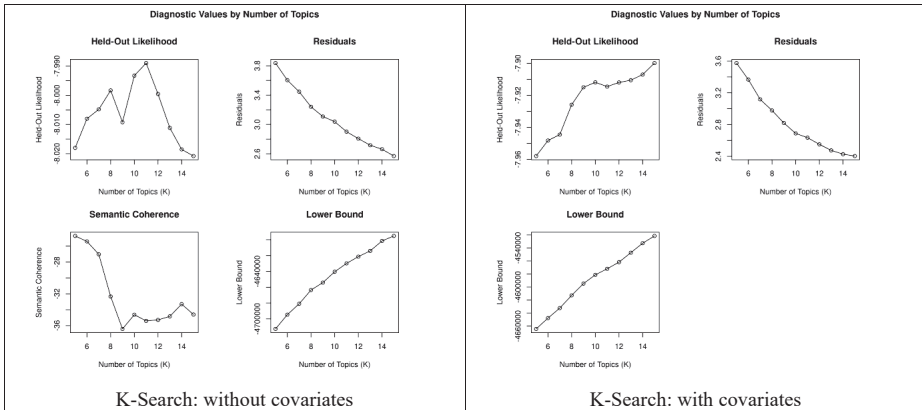


Figure 1. Diagnostic values by number of topics

The final 7-topic model is presented in Table 4. The table presents the topic labels obtained after qualitative and quantitative interpretation, the topics' prevalence, and their most relevant terms.

Instead of frequency the most relevant terms are defined by FREX, not the most frequent terms within the topic but the most distinctive ones (Roberts et al. 2019). The correlation between the topics was not found to be of a considerable size (absolute value below 0.25), this means there is no relation between their occurrence in speeches. This suggests a high degree of thematic independence among the topics³.

We can analyze the topics by examining their key terms and representative speeches. Topic 1 primarily revolves around discussions concerning agriculture and its related sectors, including water management, environmental and climate policies, and sustainable energy. The Carpathian Basin is depicted as a transboundary ecosystem within this context.

Topic 2 encompasses cultural subjects spanning the entirety of the Carpathian Basin, such as the 'hungarikum' designation, which denotes locally produced high-quality goods.

Topic 3 addresses public administration and regulatory matters affecting Hungarian communities beyond national borders, focusing on issues like voting rights.

Topic 4 revolves around speeches related to memory politics, particularly commemorations such as the 1100th anniversary of the founding of the state. Here, the Carpathian Basin is typically portrayed as the historical homeland of Hungarians for a millennium.

Topics 5 and 7 tackle various public policy domains (including church policies, cultural heritage), economy and financial matters. These topics encompass diverse subjects, with the Carpathian Basin context being just one aspect among broader economic topics such as budget discussions and development programs spanning the entire region.

3 According to the authors' experience, correlations between topics are rare.

Regarding Topic 7, it can be identified as national and neighborhood policies. These stances typically address issues pertaining to the political, cultural, and religious organizations of Hungarians beyond the borders.

ID	Label	Prevalence (1998-2010, 2010-2020)	Most relevant terms
T1	Agriculture	0.19	víz (water), élelmiszer (food), föld (land), mezőgazdaság (agriculture), növény (plant), mezőgazdasági (agricultural), termőföld (farmland)
T2	Culture	0.15	palinka (spirits), hungarikum (unique Hungarian product), székely ⁴ , emlékév (commemorative year), Duna Televízió (Duna Television ⁵), szórvány (scattered ⁶), János ⁷
T3	Public administration, regulation	0.16	ön (you), Jobbik (Jobbik, right-wing political party), Fidesz (Fidesz, currently governing party), párt (party), szavaz (vote), választás (election)
T4	Memory politics	0.16	nép (people), Alkotmány (Constitution), történelem (history), Alaptörvény (Fundamental Law ⁸), szabadság (freedom), hős (hero), csata (battle)
T5	Public Policies (monument protection, family policy, media regulation etc.)	0.10	stratégia (strategy), civil (civil), műemlék (monument), egyetem (university), változás (change), éghajlatváltozás (climate change), Duna (Danube)
T6	Economy & Finance	0.11	nemzetiségi (related to nationality), milliárd_forint (billion forint), összeg (amount), millió_forint (million forint), költségvetés (budget), nemzetiség (nationality), egyházi (related to church)
T7	National politics	0.14	tábor (camp), Erzsébet ⁹ , nemzetpolitika (national policy), nemzetpolitikai (national political), diaszpóra (diaspora), külföldi magyar (Hungarian beyond the border), értekezlet (meeting)

Table 4. Topics of the final model, with their label, prevalence, and most important terms.

4 The Székely people are Hungarians with their own ethnic identity living in Transylvania, Romania. They had autonomy between 1950-1968.

5 Hungarian television channel, one of the main tasks of which is to provide Hungarians living beyond the country's borders with programmes in their mother tongue.

6 "Szórvány": Hungarian minority communities living in a scattered area, far from the majority of the Hungarian minorities. The term is intended to distinguish them from the Hungarian minority population, who live in a relatively contiguous area and form a majority in that place (they are referred as "többségi").

7 Hungarian male first name.

8 The "Fundamental Law" is the Hungarian constitution, adopted in 2011 after the landslide victory of the Fidesz. It has been the subject of controversy both at home and abroad. Critics claim that it was adopted without sufficient input from the opposition and society at large, and that it reflects the ideology of the ruling party.

9 The "Erzsébet Camps" program offers holidays for children, including those from beyond the border.

4.2.1. Different political framing, temporal changes: focusing on topics 4 and 7

In the following, we will focus on Topic 4 and Topic 7, because these topics are the ones that show the most politically relevant use of the concept of the Carpathian Basin: memory politics (Topic 4) and national politics (Topic 7). These are the topics where the divergent political narratives to construct the perspectives of people are likely to be most detectable.

4.2.1.1. Changes in party discourse

The change in the framing of the national policy topic on the left (Figure 2) does not show a characteristic ideological change, we can only see how the support provided to Hungarians living beyond the borders has changed. E.g. the ‘Szülőföld Alap’ (Birthplace Fund) established in 2005 is replaced in the second period by the mention of the ‘Erzsébet tábor’ (Camp Erzsébet) introduced in 2020. In other words, the left-liberal discourse on national politics in both periods primarily focuses on regulations and laws.

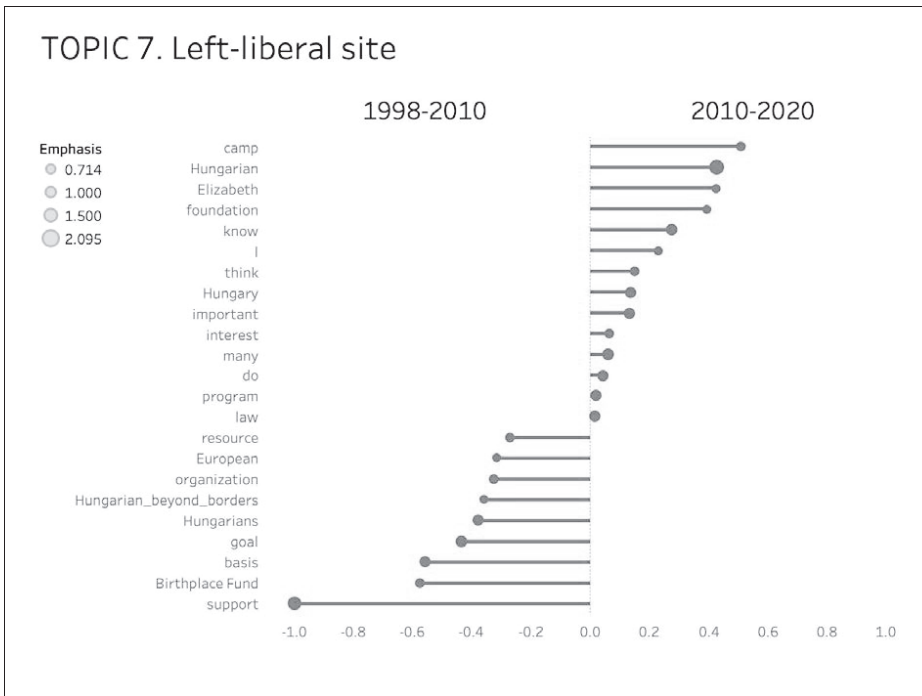


Figure 2. Change in the national politics topic on the left-liberal side.

However, a characteristic change can be detected in the left-liberal discourse on memory politics (Figure 3): terms referring to neighboring peoples other than Hungarians (‘Slovak (szlovák)’, ‘common (közös)’) are more typical of pre-2010 discourses, and their use has decreased after 2010 (Fig. 2).

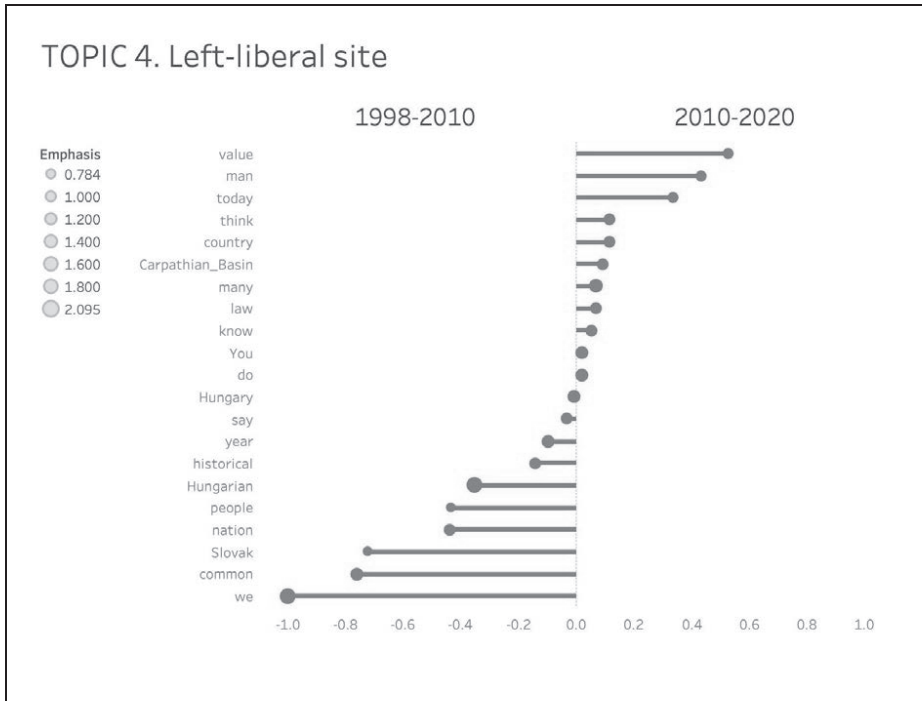


Figure 3. Change in the memory politics topic on the left-liberal side.

There is a similar ‘turn to ourselves’ in Fidesz speeches. See below the decrease in the frequency of the term ‘European Union (Európai Unió)’ and the increased use of terms referring to their own nation (‘Hungarian (magyar)’, ‘Hungarians (magyarság)’, ‘national (nemzeti)’, ‘nation (nemzet)’) after 2010. This is observed in both Topics 4 and 7 (Fig. 4.). A similar trend can be observed on the right-wing nationalist side in relation to Topic 4, see the decrease of ‘Europe (Európa)’ and the increase of ‘Hungarian (magyar)’, ‘Hungary (Magyarország)’, ‘national (nemzeti)’. In addition, the trend towards institutionalizing national policy in law is typical of Fidesz, see the rise of the terms ‘Constitution (Alkotmány)’ and ‘Fundamental Law (Alaptörvény)’ (Topic 4). Finally, on the right-wing nationalist side, in relation to national policy, a declining radicalization and a turn towards policy solutions is suggested by the decreasing use of the word ‘territory’ and the increasing use of the words ‘programme (program)’ and ‘government (kormány)’.

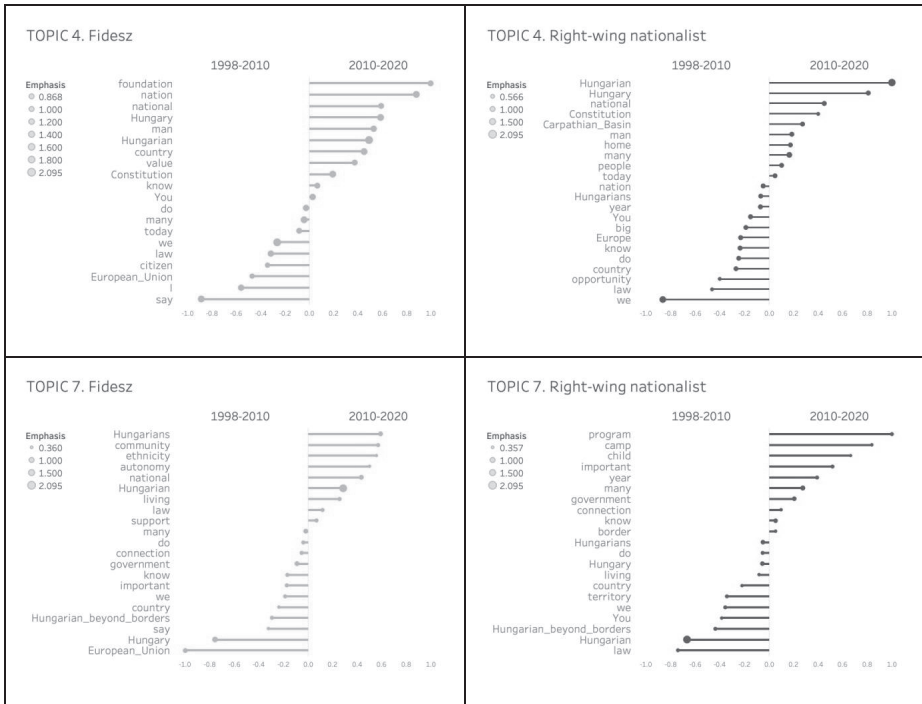


Figure 4. Change in the memory politics (Topic 4) and national politics (Topic 7) topic in the case of Fidesz and the right-wing nationalist side

4.2.1.2. Changes in the divergence of ideological narratives

There are several features to be observed when comparing political sites pairwise regarding their speeches on Topic 4 and 7. Some of the related plots are presented in Appendix.

In case of Topic 4, memory politics, the most striking change in the divergence of the left-liberal vs. right-wing narratives (Figure A1 and A2 in Appendix) is that, in the second half of the period under study the left-liberal bloc is lagging behind the right in using the national narrative (overrepresented terms on the right side: ‘Hungarian (magyar)’, ‘Hungary (Magyarország)’, ‘nation (nemzet)’, ‘national (nemzeti)’, ‘we (mi)’).

In the comparison between Fidesz and the right-wing nationalist bloc (Figure A3 in the Appendix), there is no systematic change in their divergence. Both periods are characterized by Fidesz using more terms that presumably bring memory politics within a legal framework (‘constitution (Alkotmány)’, ‘Fundamental Law (Alaptörvény)’).

Turning to national politics, Topic 7: when comparing the left-liberal bloc and Fidesz (Figure A4 in the Appendix), the trend already seen in topic 4 is emerging: by the second half of the period under study the left-liberal bloc is lagging behind the Fidesz in using the national narrative (Fidesz uses more

frequently the terms ‘Hungarian (magyar)’, ‘national (nemzeti)’, ‘Hungarians (magyarság)’. While in this second period the left-liberal bloc tends to argue within a rational legal framework (‘law (törvény)’, ‘foundation (alapítvány)’, ‘programme (program)’, ‘Camp Erzsébet (Erzsébet tábor)’), Fidesz moves beyond patriotism towards virtual community building, using words like ‘community (közösség)’ and ‘autonomy (autonómia)’. In the comparison between the left-liberal and the right-wing nationalist blocs (Figure A5 in the Appendix), the left-liberal’s backlog in terms of national rhetoric is striking in both periods under review. The biggest difference is the use of the term Hungarian (magyar).

In the comparison between Fidesz and the right-wing nationalist bloc (Figure 5), after 2010, the earlier characteristic of Fidesz using more institutional terms (‘European Union (Európai Unió)’, ‘Hungary (Magyarország)’), while the other right-wing nationalist parties used more ethnic terms (‘Hungarian (magyar)’, ‘Hungarians (magyarság)’, ‘territory (terület)’), and even seem to reverse the role reversal, with the former referring more to ethnic terms (‘Hungarian (magyar)’, ‘hungarians (magyarság)’, ‘community (közösség)’, ‘autonomy (autonómia)’) and the latter referring more to policy solutions like ‘program (program)’ and ‘government (kormány)’, see the two comparisons on figure 5.

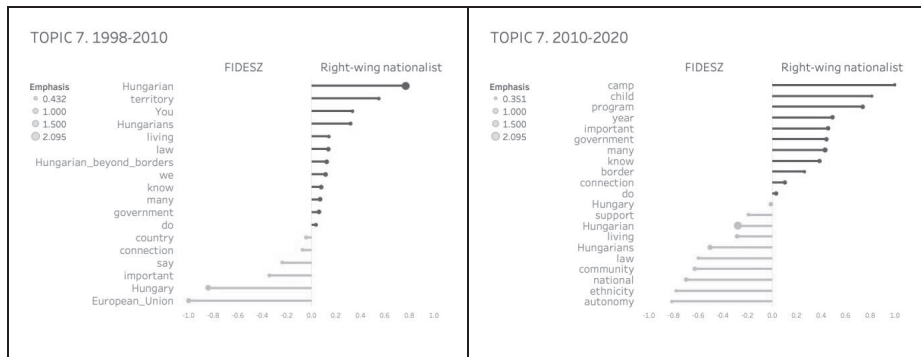


Figure 5. Changes in the divergence of framing between Fidesz and the right-wing nationalist bloc, Topic 7

4. Limitations

Our analysis also has its limitations. We only analysed speeches that contained a specific keyword, a decision justified by methodological reasons. Speeches were analysed as individual utterances, whereas they were part of conversations with dynamic and interactive patterns. In order to overcome this limitation, our future plan is to study networks of speeches in relation to the themes explored in this study.

A further limitation of the analysis is the imbalance among the categories. To overcome this problem we used undersampling. This approach could potentially cause the loss of valuable patterns and relationships, weakening the

model's ability to generalize, and reducing its capacity to learn meaningful patterns for both dominant and minority classes. Additionally, the grouping of parties, particularly within the left-wing liberal group, which consists of a diverse range of parties, may complicate the analysis from the model's perspective, making it more difficult to accurately capture patterns within this side. This variability further limits the model's ability to perform robustly.

Finally, as there is no standardised procedure for determining the optimal number of topics, we conducted robustness tests and combined quantitative and qualitative approaches before selecting the final model. We do not claim that 7 is the 'right' number of themes, but this model includes all the important themes that emerged in the other models and does not include any unrelated themes. Overall, it provides useful insights without being overly clustered.

5. Conclusion

The corpus of our analysis consisted of the Hungarian parliamentary speeches containing the term 'Carpathian Basin', delivered between 1998 and 2020. Our results have shown an increasing polarization in the Carpathian Basin discourse in general (classification efficiency increases).

Using both quantitative and qualitative approaches, we identified 7 topics in the Hungarian Parliament's discourse related to the Carpathian Basin. These topics were found to be well-defined in terms of their subject matter: speeches on agriculture, culture, public administration, memory politics, public policies, economy & finance, and national politics constituted a separate topic. The change in discourse was achieved by splitting the period under study into two (1998-2010, 2010-2020). By comparing the most relevant terms used when discussing the topics, we examined (1) the change in the political blocs' discourse and (2) the change in the blocs' divergence.

Parallels can be detected in the changing discourse of different ideological sides. When focusing on topics of memory politics and national politics, one such common feature is a more concentrated focus on the Hungarian nation, as opposed to neighboring peoples and the European Union. Even though this shift has also taken place on the left-liberal side, their disadvantage in this respect remained unchanged or in some cases increased, i.e. the national narrative was underrepresented in both halves of the period under review in the left-liberal speeches compared to Fidesz and the right-wing nationalist bloc.

We have found evidence of the left-wing's turn in its attitude to national identity, that is, discourse traits of the left's rapprochement with the right (as an imprint of the left's opening to Hungarians beyond the borders after 2010). Furthermore, "two-way mainstreaming" has also been detected, that is, moderation of the far right: the distinctive traits of Fidesz and the right-wing nationalist side swapping places.

The Carpathian Basin as a political symbol creates an opportunity for the "owner" of this symbol to expand the boundaries of Hungarian politics without changing the territorial integrity of the country in any way. The term was

even included in the 2011 Fundamental Law, which states that “We undertake to cherish and protect our heritage, our unique language, Hungarian culture, the language and culture of the nationalities of Hungary, and the natural and man-made values of the Carpathian Basin.” According to the text, the responsibility of the Hungarian state is thus to protect the values of the Carpathian Basin, which, whatever it may mean in practice, extends the scope and responsibility of the Hungarian state as a sovereign actor beyond the borders of the state. This use of symbolism therefore has such a political or competence-expanding effect on the one hand, and therefore presumably also has another effect in another direction: it is the “domestic politicization” of issues relating to Hungarian nationalities beyond the border. In other words, the issue of Hungarian minorities will be transformed from being a matter of primarily international human rights and minority rights, or inter-state issues, and, in sum, of foreign relations and foreign policy, into a domestic issue.

The introduction of dual citizenship as an instrument of nation-building has a considerable literature, so we will not discuss the lessons to be drawn from this, neither from the general international nor from the regional nor from the specific Hungarian developments. It is worth noting, however, that the introduction of dual citizenship and the intensification of the Carpathian Basin vocabulary occur simultaneously. This means that, contrary to the assessments that predict the draining effect of dual citizenship and the weakening of diaspora Hungarians, there is also an opposite effect: the extension of the public policy-making limits of the Hungarian state, of classical domestic policy issues, beyond the borders.

References

- Balogh, Péter. 2021. “The Concept of the Carpathian Basin: Its Evolution, Counternarratives, and Geopolitical Implications.” *Journal of Historical Geography* 71: 51–62.
- Blei, David M., and John D. Lafferty. 2007. “A Correlated Topic Model of Science.” *The Annals of Applied Statistics* 1 (1): 17–35.
- Cohen, Abner. 1979. “Political Symbolism.” *Annual Review of Anthropology* 8: 87–113.
- Eisenstein, Jacob. 2019. *Natural Language Processing*. Cambridge, MA: MIT Press.
- Evans, James A., and Paul Aceves. 2016. “Machine Translation: Mining Text for Social Theory.” *Annual Review of Sociology* 42 (1): 21–50.
- Friedrich Ebert Stiftung–Policy Solutions. 2013. “Nemzeti identitás – egység vagy sokszínűség? Politikai Megbékélés Beszélgetéssorozat II.” URL: https://www.policysolutions.hu/userfiles/elemzes/18/nemzeti_identitas_egyseg_vagy_sokszinuseg.pdf (last accessed: April 30, 2025).
- Gaudenyi, Tivadar, and Milan Mihajlović. 2022. “The Carpathian Basin: Denomination and Delineation.” *European Journal of Environment and Earth Sciences* 3 (2): 1–6.
- Goodson, Ivor F., and Scherto R. Gill. 2011. “The Narrative Turn in Social Research.” *Counterpoints* 386: 17–33.
- Hajdú, Zoltán. 2018. “The Rebirth of the Concept of the Carpathian Basin in Hungarian Political Language after 1988.” In: Laine, Jussi, Ilkka Liikanen,

- and James W. Scott, eds. *Post-Cold War Borders: Reframing Political Space in Eastern Europe*. London: Routledge, pp.: 207–227.
- Hunyadi, Bulcsú. 2022. “Hogyan váltotta fel a populista veszély a szélsőjobb jelentette fenyegetést?” URL: <https://policycommons.net/artifacts/3378759/hogyan-valtotta-fel-a-populista-veszely-a-szelsojobb-jelentette-fenyegetest/4177645/> (last accessed: April 30, 2025).
- Kovarek, Dániel, and Attila Farkas. 2017. “A Jobbik Mérséklődése Az Egyéni Képviselő-Jelöltek Vizsgálatának Tükrében.” *Politikatudományi Szemle* 26 (1): 31–54. URL: <http://real.mtak.hu/136127/1/kf.pdf> (last accessed: April 30, 2025).
- Mák, Ferenc. 2000. “Az új nemzeti politika és a Határon Túli Magyarok Hivatala (1989–1999).” *Magyar Kisebbség* 6 (3).
- Megoran, Nick. 2006. “For Ethnography in Political Geography: Experiencing and Re-Imagining Ferghana Valley Boundary Closures.” *Political Geography* 25: 623–640.
- Müller, Martin. 2008. “Reconsidering the Concept of Discourse for the Field of Critical Geopolitics: Towards Discourse as Language and Practice.” *Political Geography* 27: 322–338.
- Németh, Renáta et al. 2025. “What Else Comes with a Geographical Concept beyond Geography? Discourses Related to the Carpathian Basin in the Hungarian Parliament.” *Intersections. East European Journal of Society and Politics* 2025 (1).
- Németh, Renáta, and Júlia Koltai. 2021. “The Potential of Automated Text Analytics in Social Knowledge Building.” In: Rudas, Tamás, and Gábor Péli, eds. *Pathways Between Social Science and Computational Social Science: Theories, Methods, and Interpretations*. Cham: Springer International Publishing, pp.: 49–70.
- Pogonyi, Szabolcs. 2017. *Extra-Territorial Ethnic Politics, Discourses and Identities in Hungary*. New York, NY: Palgrave Macmillan.
- Roberts, Margaret E. et al. 2014. “Structural Topic Models for Open-Ended Survey Responses.” *American Journal of Political Science* 58 (4): 1064–1082.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2019. “stm: An R Package for Structural Topic Models.” *Journal of Statistical Software* 91: 1–40.
- Scott, James W., and Zoltán Hajdú. 2022. “The Carpathian Basin as a ‘Hungarian Neighbourhood’: Imaginative Geographies of Regional Cooperation and National Exceptionalism.” *Eurasian Geography and Economics* 63 (6): 753–778. URL: <https://doi.org/10.1080/15387216.2022.2082995> (last accessed: April 30, 2025).
- Theiler, Tobias. 2017. “Political Symbolism.” In: Moghaddam, Fathali M., ed. *The SAGE Encyclopedia of Political Behavior*. Thousand Oaks, CA: SAGE Publications, pp.: 633–634. URL: <http://dx.doi.org/10.4135/9781483391144.n284> (last accessed: April 30, 2025).
- Udrea, Andreea. 2013. “A Kin-State’s Responsibility: Cultural Identity, Recognition, and the Hungarian Status Law.” *Ethnicities* 14 (2): 324–346.
- Wodak, Ruth, and Bernhard Forchtner, eds. 2017. *The Routledge Handbook of Language and Politics*. London: Routledge.
- Wodak, Ruth. 2010. “The Discursive Construction of History – Brief Considerations.” *Mots. Les Langues du Politique* 94: 57–65.

Appendix

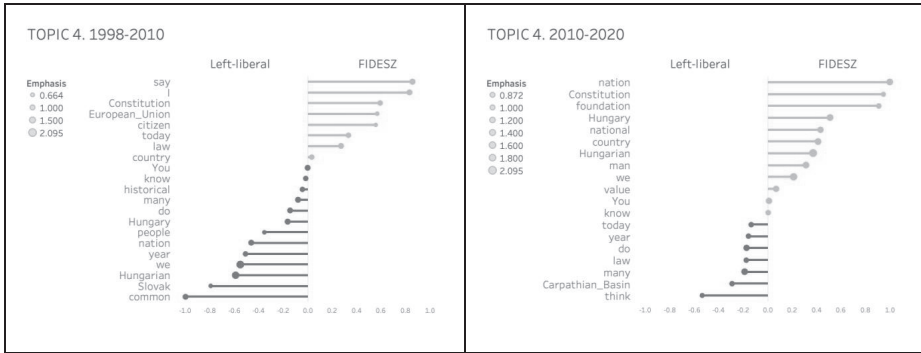


Figure A1. Changes in the divergence of framing of the left-liberal bloc and the Fidesz, Topic 4

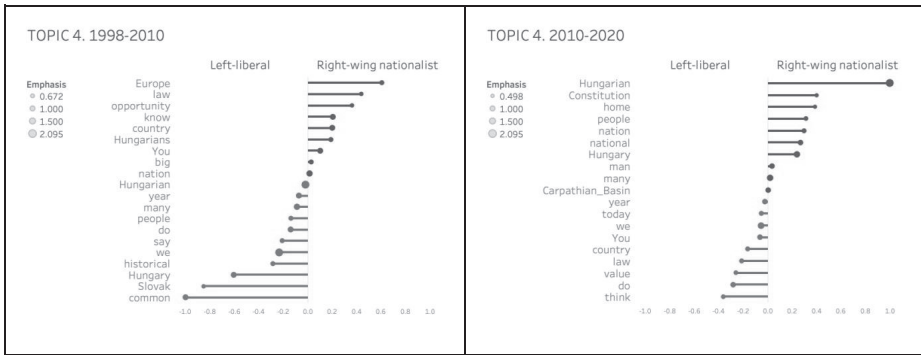


Figure A2. Changes in the divergence of framing of the left-liberal and the right-wing nationalist blocs, Topic 4

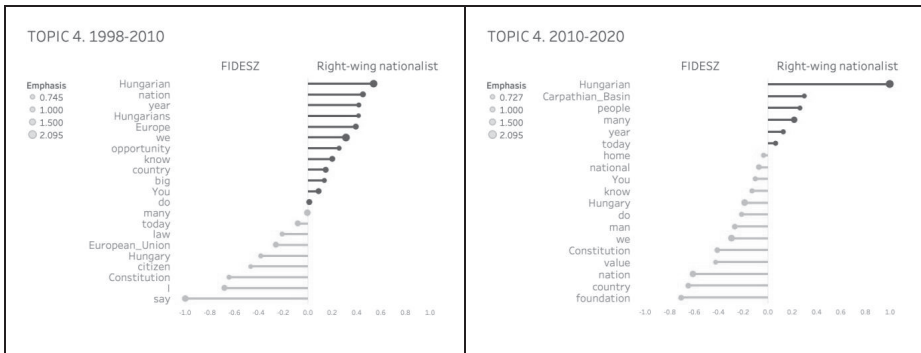


Figure A3. Changes in the divergence of framing of the Fidesz and the right-wing nationalist bloc, Topic 4

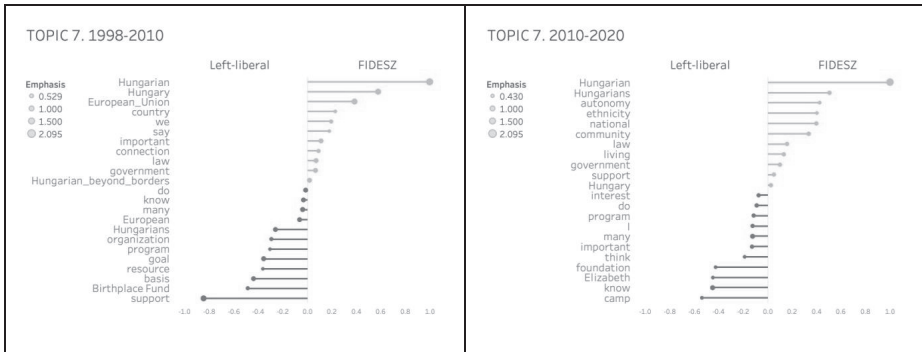


Figure A4. Changes in the divergence of framing of the left-liberal bloc and Fidesz, Topic 7

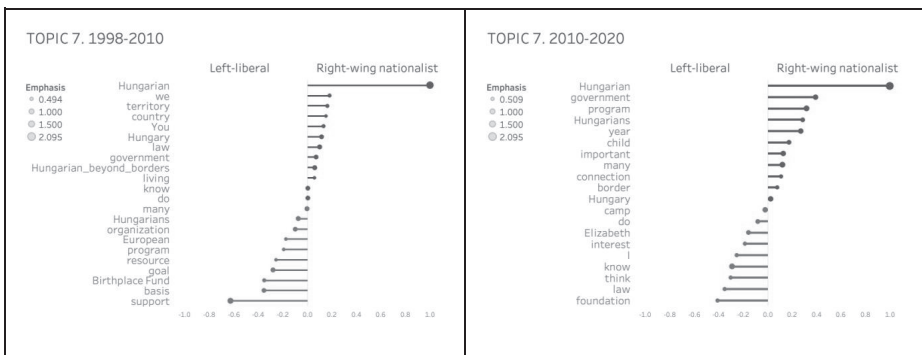


Figure A5. Changes in the divergence of framing of the left-liberal and the right-wing nationalist blocs, Topic 7

Ester Katona, Renata Nemet i Ana Unger

Evucija diskursa o „Karpatskom basenu” u mađarskom parlamentu (1998–2020)

Apstrakt

U ovom radu istražujemo upotrebu izraza „Karpatski basen” u mađarskom parlamentu u periodu od 1998. do 2020. godine. „Karpatski basen” je pojam iz mađarske geografije koji je istorijski korišćen za opravdavanje teritorijalnih pretenzija Mađarske u međuratnom periodu. Iako je decenijama bio odsutan iz zvaničnog diskursa, u poslednje vreme doživljava povratak među mađarskim političarima. Analizom 1525 govora proučavamo promene u diskursu tri glavna politička bloka (desničarsko-nacionalističkog, liberalno/levičarskog i Fidesza) kako bismo identifikovali jezičke obrasce političke polarizacije i promene u obrascima identiteta i načinu uokviravanja pojma.

Naš rad ima i metodološku i sadržinsku važnost. Metodološka novina ogleda se u primeni tehnika koje omogućavaju automatsku obradu velikih tekstualnih korpusa bez potrebe za njihovim ručnim čitanjem, u oblasti gde su ranije dominirali uglavnom kvalitativni pristupi. Pokazujemo da je moguće otkriti promene u načinu uokviravanja bez ljudskog kodiranja.

Sadržinski, fokusiramo se na jezičke karakteristike jednog značajnog pojma koji se različito koristi u zavisnosti od političke ideologije.

Koristimo i nadgledane (supervised) i nenadgledane (unsupervised) metode modelovanja. Nadglednu metoodu koristimo za ispitivanje promena u polarizaciji diskursa, dok nenadgledana metoda (Strukturalni model tema) omogućava nijansiraniju, kvalitativnu interpretaciju rezultata. Naši nalazi pokazuju da se ideološka pripadnost govornika sve lakše može predvideti – što ukazuje na porast polarizacije – dok dublja analiza otkriva i određene sličnosti u razvoju diskursa među različitim ideološkim stranama. Jedna takva zajednička crta jeste sve veći fokus na mađarsku naciju, za razliku od ranijeg naglaska na susedne narode i Evropsku uniju. Takođe identifikujemo tragove približavanja levice desnici (kao posledica otvaranja levice prema Mađarima van granica nakon 2010) i umerenijeg diskursa krajnje desnice.

Ključne reči: obrada prirodnog jezika, Karpatski basen, mađarski parlament, ideološke podele