ANALYZING LITHUANIAN PARLIAMENTARY DISCOURSE ON UKRAINIAN REFUGEES BY APPLYING TEXT ANALYSIS TOOLS

AUŠRA URBANAVIČIŪTĖ

Vytautas Magnus University, Lithuania

SUMMARY. The article examines how members of the Lithuanian Seimas refer to Ukrainian refugees by investigating the linguistic patterns used by Lithuanian politicians in Lithuanian Seimas' parliamentary sessions. The research applies the Natural Language Toolkit (NLTK), a Python library commonly employed in computational linguistics, and LitLatBERT, complemented by manual analysis. The study explores how Lithuanian politicians frame Ukrainian refugees linguistically by focusing on lexical aspects, mainly definitions of and references to refugees and modality. The article's broader aim is to showcase how NLTK libraries and the transformer-based deep learning model LitLatBERT can be combined with methods of Critical Discourse Analysis (CDA) to analyze Lithuanian parliamentary discourse.
KEYWORDS: parliamentary debates, refugee and migrant discourse, NLTK, LitLatBERT.

INTRODUCTION

With the ongoing Russo-Ukrainian war in Ukraine's territory, Europe has experienced an influx of Ukrainian refugees. This mass displacement of people has had far-reaching political, economic, and social implications for neighboring countries. Among these, Lithuania stands out due to its shared historical sentiments with Ukraine. Both nations share an anti-Soviet stance shaped by their experiences under Soviet rule, fostering a sense of solidarity and mutual support. Throughout the conflict, Lithuania has positioned itself as a steadfast ally to Ukraine, providing humanitarian aid, advocating for Ukraine's sovereignty on international platforms, and welcoming Ukrainian refugees into its borders. Despite Lithuania's optimistic outlook regarding Ukraine and its people, the influx of refugees presents challenges. An increased population often brings varying degrees of economic, social, and political instability to a country, even when refugees are met with a welcoming attitude. In Lithuania, this contrast manifests as a balancing act between

demonstrating support for Ukraine and ensuring the country's self-preservation. These dynamics can be seen in the debates within the Lithuanian Seimas, where its members discuss the accommodation, living conditions, and integration of Ukrainian refugees. As politicians construct narratives that shape public opinion and influence policy decisions through language, analyzing parliamentary debates can show how Lithuanian politicians perceive Ukrainian refugees. Narratives constructed by politicians are crucial to examine, as they often reflect the ideological power dynamics in political debates.

The primary aim of this research is to explore how Lithuanian politicians use language to frame Ukrainian refugees and how ideological power is reflected in Lithuanian parliamentary discourse. This study employs the theoretical framework of Critical Discourse Analysis (CDA) to examine the linguistic features of debates, combining manual and computational techniques to identify patterns of linguistic framing and modal expressions. Although studies that apply CDA and computational means to investigating parliamentary discourse have been conducted on languages such as English, Spanish, and Turkish (Abuelwafa 2021; Charteris-Black 2006; Gupta 2023; Islentyeva 2021), there is still a lack of computational means being applied for studying Lithuanian parliamentary discourse (Kapočiūtė-Dzikienė et al. 2014; Morkevičius 2005; Ramonaitė and Vaiginytė 2020). By using the Natural Language Toolkit (NLTK) (Bird, Klein, Loper 2009), SpaCy (Honnibal, Montani 2015), and Bidirectional encoder representations from transformers (BERT) (Devlin, Chang, Lee, Toutanova 2019), the article demonstrates how computational tools can enhance the linguistic analysis of Lithuanian parliamentary discourse. Moreover, this research shows how CDA methods can be combined with computational tools like NLTK, SpaCy, and LitLatBERT to intensify and enhance such an analysis. The main guiding questions for this research are:

1. How do Lithuanian Seimas members refer to Ukrainian refugees?
2. How do modal verbs and expressions of necessity, obligation, or possibility reveal ideological power in Lithuanian parliamentary discourse?
3. How can manual and computational methods, including CDA, NLTK-based analysis, and transformer-based deep learning models, be combined to provide a more comprehensive understanding of framing in parliamentary discourse?

THEORETICAL BACKGROUND

Language in parliamentary discourse can be a powerful medium in constructing and perpetuating ideologies. Political scientists, sociologists and linguists often

describe ideology as a system that helps establish, legitimize, and maintain power relations within discourse. Linguists, however, tend to use the term "ideological power". Scholars such as Fairclough (1989, 2010), van Dijk (1998, 2006), Verschueren (2012), and Reisigl and Wodak (2001) have emphasized that language can both reflect and reinforce ideological power, often by framing ideas as "common sense". Fairclough's (1989: 33) concept of ideological power, which he defines as the ability to project practices as universally valid, highlights the role of language in normalizing particular viewpoints while marginalizing others. This view aligns with Thompson's (1990: 146) perspective that language actively helps to structure social reality, stressing the link between discourse and societal structures. Ideological power is often encoded in language with linguistic choices that influence audience perception (Fairclough 1989, 2010; Freeden 2003; Reisigl and Wodak 2001; Van Dijk 1998, 2006; Van Leeuwen 2008; Verschueren 2012). Ideology is linked with all levels of discourse, from the smallest micro-level of language to the meso-level and influences the macro-level of language through various discourse practices. Researchers who investigate the linguistic aspects of ideology must consider the driving elements and the creative forces of ideology. Therefore, the main item of interest in this paper is ideological power, whose working definition in this paper is "means by which certain practices and ideas are projected as universal and 'common sense' " (Fairclough 1989, 2010; Van Dijk 1998, 2006; Van Leeuwen 2008). In this paper, only the micro-level of language will be analyzed.

Several studies that examine how immigration is portrayed and referred to in media and political discourses, particularly in English, note that the keywords "refugee", "asylum seeker", "immigrant", and "migrant" appeared most frequently when referring to migration (Baker 2007, 2008; Taylor 2014; Islentyeva 2021) and thus the acronym RASIM (refugee, asylum seeker, immigrant, migrant) came into use (first introduced in Baker 2007: 44). The terms found in the acronym RASIM have become more numerous in Lithuania's political and media discourse, with the migrant crisis of 2019 in Lithuania, and the Russo-Ukrainian war, which began in early 2022. RASIM, as Islentyeva (2021: 53) states, are "one of the least powerful groups in society", with their image and identity almost exclusively being constructed by more powerful groups. Works by Fowler (1991) and van Dijk (1996) explore discourses surrounding refugees and migrants and the implications of a social group not having control over their representation to the public. Parliamentary discourse has evident power in constructing how RASIM are viewed in society. This applies to how RASIM are presented in the media, thus creating their image in society in a narrative sense. Moreover, parliamentary discourse has legislative power, with the most popular opinion governing the most ideological power and most likely resulting in new laws and legislation. Therefore, concerning RASIM,

it is important to study parliamentary discourse as it can help understand how the narrative surrounding this social group spreads to other discourses and within society.

Various works investigating ideological power in discourse employ Critical Discourse Analysis (CDA) as their main approach, especially when examining power dynamics and bias in political contexts. CDA scholars (Fairclough 1989; Reisigl & Wodak 2008; van Dijk 1998, 2006; van Leeuwen 2008) contend that ideological power manifests in discourse through lexical choices, syntactic structures, and pragmatic strategies that help create narrative advantageous to dominant groups. The relationship between language and ideology operates throughout all levels of language. This includes micro-level language features like modality, declarative and imperative speech acts, nominalizations, inclusive or exclusive pronouns, rewording or overlexicalization, implicature, and metaphors. Meso-level features of ideologically loaded language refer to strategies of argumentation and legitimation, and macro-level features encompass aspects of recontextualizing discourse and topics in a specific discourse. Van Dijk's (1998) multidisciplinary approach highlights the interplay between discourse, cognition, and society. Reisigl and Wodak's (2008) Discourse Historical Approach (DHA) incorporates historical context, while Van Leeuwen's (2008) framework emphasizes the recontextualization of discourse in varied social settings. All these frameworks focus on linguistic devices such as sentence structure, active/passive voice, and discourse markers to reveal hidden ideological meanings. By combining micro-, meso-, and macro-level analyses, researchers can uncover the layered ways language legitimizes power and shapes social realities. This multidimensional approach remains essential for understanding the subtle mechanisms through which ideological power is constructed and operates in language. Therefore, CDA allows for a comprehensive study of Lithuanian parliamentary discourse, revealing how language choices reinforce ideological positions and influence public perceptions of Ukrainian refugees.

As a linguistic feature, modality conveys a speaker's attitude toward the necessity, probability, or desirability of an action or event. As such, it signals the speaker's stance regarding the discourse they are participating in. Therefore, analyzing modality can help identify which opinions are favored in a selected discourse. Modality is defined as a grammatical category reflecting the speaker's relationship to reality and is categorized into three main types: epistemic, deontic, and dynamic (Šolienė, 2013; Usonienė 2004, 2006; Ruskan 2017). Modal meanings can be expressed through verbs, parentheticals, and intonation. Research distinguishes between *primary* and *secondary modality*. Primary modality (reality, necessity, possibility) includes forms like indicative verbs for reality or conditional "can" (Lit. *galėti*) for possibility. Secondary modality encompasses categorical reliability, problematic

reliability, and evidentiality. Categorical reliability refers to linguistic items that denote assuredness with phrases like "it is true" (Lit. *tiesa*). Problematic reliability, on the other hand, refers to terms of unsureness and is expressed with linguistic items like "it seems" (Lit. *atrodo*) (Akelaitis, 1986: 5). Evidentiality, which is categorized as direct (e.g. I know (Lit. *aš žinau*)) or indirect (e.g. "according to" (Lit. *anot, pasak*)), indicates the source of information (Ruskan, 2010). Direct evidentiality centers the speaker as the source of information, whereas indirect evidentiality indicates third persons or vague entities as information sources. Modality is informative when examining particular aspects of discourse, such as how confident a speaker is of their spoken content or the use of language, which signifies necessity or possibility. In this study, only lexical expressions of modality were analyzed, and extralinguistic items were not. This was done because the material for enquiry consists of transcribed debates of the Lithuanian Seimas, in which extralinguistic information is not included.

METHODOLOGY

The dataset is a 59,912-word corpus compiled from seven plenary sessions of the Seimas on Ukrainian refugees. The corpus consists of plenary sessions between February 24th, 2022, and June 15th, 2023. This timeframe corresponds to the first plenary session that involved discussions about Ukrainian refugees following the beginning of Russia's invasion of Ukraine in 2022, while the last plenary session was dedicated to the refugee status in Lithuania. The corpus was assembled manually, so the text was not annotated. The multilingual corpus ParlaMint II (Erjavec et al., 2020) was considered for this project; however, the parliamentary texts in that corpus only date to 2020. Since an important criterion for text selection was the discussion of Ukrainian refugees, the selected texts display specific parts of plenary sessions on this topic. Before preprocessing, the corpus comprised 72,594 words; later, the corpus was cleaned of speaker names while retaining the names of represented parties. Additionally, the names of sections and formal remarks by the Seimas' Chair that indicated the beginning, end, or voting parts of the plenary sessions were removed. The final variant of the corpus consists of 59,912 words.

To achieve the study's primary aim – to investigate how Ukrainian refugees are referred to in Lithuanian parliamentary discourse – linguistic elements in the corpus, such as adjectives, verbs, adverbs, and nouns related to Ukrainian refugees, were examined. Moreover, units of secondary modality (categorical and problematic reliability and evidentiality) were studied in sentences which contained references to Ukrainian refugees, the step chosen to collect more information on how the

speakers viewed their spoken content. This analysis of modality provided insights into how Lithuanian politicians positioned themselves on the issue of Ukrainian refugees. The use of language may vary depending on the speaker and reveal differing views. However, the aim of this research was to analyze the joint stance of Lithuanian Seimas members on Ukrainian refugees in Lithuania. Therefore, exploring each speaker's language separately was unnecessary for this research.

The inquiry of how Ukrainian refugees are defined within the corpus and the types of modality used was conducted in three ways to determine how and to what degree computational methods can be implemented in analyzing parliamentary discourse and ideological power. The three approaches were a manual analysis, a semi-automated study using part-of-speech (POS) tagging with the NLTK and SpaCy libraries, and a semi-automated exploration of text using the trilingual language model Lithuanian–Latvian Bidirectional Encoder Representations from Transformers (LitLatBERT). The corpus analysis program AntConc (Anthony 2024) was applied to perform a part of the investigation. The keywords for the acronym RASIM were queried, and their context was manually checked to compile a list of references to Ukrainian refugees.

The second and third parts were supplemented by manual verification of the results. The semi-automated approach employed a script to add POS tags to the pre-processed corpus automatically. First, the corpus was tokenized, which means it was split into individual words. Afterwards, POS tagging labelled each word with its grammatical category, such as noun, verb, or adjective. The script then printed out sentences with the keywords from the acronym RASIM, along with nouns, verbs and adjectives found in the sentence. This helped to examine printed-out sentences systematically and detect what nouns, verbs, and adjectives were often used with words like "refugee", "immigrant", "asylum seeker", and "migrant". The third approach involved the LitLatBERT language model to identify how embedded words related to RASIM are defined within the text using cosine similarity measures. Word embeddings made by LitLatBERT for the Lithuanian language are like maps, where words are turned into a list of numbers (vectors) that capture their meaning based on the context in which they appear. The LitLatBERT's cosine similarity measures to what extent two words, phrases, or sentences are similar by comparing their positions in its "language map". For example, words like "migrantai" (Eng. *migrants*) and "imigrantai" (Eng. *immigrants*) are closer together in vector values than words like "karas" (Eng. *war*) and "taika" (Eng. *peace*). Moreover, LitLatBERT can assign these values to whole sentences, which helps to analyze discourse. The script divided the corpus into sentences, and long sentences exceeding LitLatBERT's 512-token size were subdivided to meet the model's input criteria. The 512-token size constraint can affect the results but is unavoidable when using

LitLatBERT. This layered approach offered a broader perspective on how Ukrainian refugees were represented in the Lithuanian parliamentary discourse.

RESULTS AND DISCUSSION

The first part of the analysis consisted of a quantitative overview of how frequently the keywords of the acronym RASIM (*refugee, asylum seeker, immigrant, migrant*) were used in the corpus. Since the research aimed to determine how Ukrainian refugees were referred to, linguistic units denoting people fleeing Ukraine or arriving in Lithuania were also counted. It was noticed that these terms could be categorized as people moving or fleeing from Ukraine and people coming or fleeing to Lithuania. These two categories are represented in Table 1, which shows how many times lemmas of words "pabėgėlis" (Eng. *refugee*), "prieglobsčio ieškotojas" (Eng. *asylum seeker*), "imigrantas" (Eng. *immigrant*), "migrantas" (Eng. *migrant*) occurred, along with their normalized frequencies per 10,000 words.

Table 1. Total mentions of RASIM keywords and their normalized frequencies in the analyzed corpus

Keyword (lemma)	Total mentions	Normalized frequency (per 10k words)
Pabėgėlis (Eng. Refugee)	49	8.18
Prieglobsčio ieškotojas (Eng. *Asylum seeker*)	3	0.50
Imigrantas (Eng. *Immigrant*)	0	0
Migrantas (Eng. *Migrant*)	41	6.84
People fleeing to Lithuania	3	0.50
People fleeing from Ukraine/war	4	0.67
Total	97	16.19

The table indicates that in the corpus, "refugees" and "migrants" appeared far more frequently than "immigrants" and "asylum seekers". Among RASIM keywords, "migrants" and "refugees" were the two most frequently used words, predominantly in the plural. The plural form "migrants" occurred 40 times (97.56%), and "refugees" with 45 mentions make up 91.83%. Although not all instances of the counted RASIM keywords involved Ukraine, it is interesting to note that the tendency to refer to migrants or asylum seekers in the plural is also noticed in studies done on English discourses (Islentyeva 2021; Van Leeuwen 2008). The framing of social groups as a collective body can lead to impersonalization, which

generalizes aspects of the subject (Van Leeuwen 2008: 46), and individuals belonging to these groups may be perceived as either negative or positive. A more detailed conclusion of whether plural forms indicate impersonalization may be achieved by examining linguistic elements surrounding the occurrences of "refugees" and "migrants".

The sentence context of the marked RASIM keywords showed that Ukrainian refugees were often mentioned with nouns denoting people. As seen in example 1, Ukrainian war refugees are referred to as people who will leave Ukraine and seek refuge. Example 2 shows another tendency to refer to Ukrainian refugees using words denoting humans, like "children" or "women". (All Lithuanian example sentences are translated by the author of the article.)

1. <…> būti pasirengusius priimti žmones, **kurie išvyks iš Ukrainos ir ieškos prieglobsčio**.[1]
2. Suteikime laisvę, prieglobstį ir atverkime Lietuvos namus **Ukrainos vaikams, mamoms ir nukentėjusiems kariams**.[2]

Furthermore, it was noticed that when in a sentence "refugees" implied Ukrainians or war refugees, they were used with verbs like "support" or "accept", as shown in example 3. Moreover, migrants were mainly referred to by also employing adjectives like "political" or "illegal". This contrast can be seen in example 4.

3. <…> tiek mobilizuojant partnerius **teikti paramą**, tiek **priimant** karo **pabėgėlius**.[3]
4. Tai buvo kovido situacijos administravimas, ar **nelegalios migracijos**, ar **karo pabėgėlių** <…>.[4]

Although such contrasts between naming migrants and refugees may indicate genre specifics of parliamentary discourse, it is an interesting occurrence. People fleeing the war in Ukraine are most frequently referred to as "refugees" or "asylum seekers". Though the study results are inconclusive, the verbs and adjectives appearing alongside verbs denoting Ukrainian refugees seem to carry supportive connotations.

Next, the semi-automated analyses were conducted using two different scripts. One script used the Natural Language Toolkit's part-of-speech tagging, while the other employed the LitLatBERT model. Both scripts made use of the Lithuanian WordNet (Garabik, Pileckytė 2013) library, an extensive lexical database of

---

[1]   <…> to be prepared to accept people who will depart from Ukraine and seek refuge.

[2]   Let us provide freedom and refuge, and let us open Lithuanian homes to Ukrainian children, wives, and wounded soldiers.

[3]   <…> both, while mobilizing partners to provide support and accepting war refugees.

[4]   It was the administration of the COVID situation or illegal migration or war refugees <…>.

Lithuanian. This database was chosen for the research because it provided lemmas, part-of-speech tags, and distinguished semantic relationships, e.g., synonyms, antonyms, and hypernyms. The outputs from both methods revealed differences in their analytical depth and effectiveness in identifying and categorizing linguistic units related to Ukrainian refugees. The first script focused on basic linguistic categorization using SpaCy's POS tagging and regex matching. Regex, short for *regular expressions*, is a tool to find patterns in texts, like searching for all words starting with "migrant". The first script identified adjectives, verbs, and adverbs and assigned synonymous nouns to RASIM keywords within the text. While this approach is straightforward and efficient, it lacks semantic depth. The output provides surface-level linguistic elements without capturing nuanced meanings or relationships between terms. Example 5 shows an instance where the script which used POS tags noted the sentence as relevant because of the lemma "pabėg*" (Eng. *refuge**); however, the script which employed LitLatBERT correctly identified that the phrase "pabėgėliai nuo karo" (Eng. *war refugees*) refers to Ukrainian refugees.

1. Todėl dar kartelį raginu Vyriausybę atkreipti dėmesį į Darbo partijos pasiūlymą dėl geresnės koordinavimo sistemos ir geresnės padėties užtikrinimo **pabėgėliams nuo karo** Lietuvoje.[5]

The results from the first script with POS tagging were consistent while identifying single-word terms like "pabėgėliai" (Eng. *refugees*) or "migrantai" (Eng. *migrants*) along with their immediate grammatical context. The script does not account for the broader semantic associations or nuanced usage of these terms within the text, which limits its analytical depth. The script with LitLatBERT and integrated SpaCy could identify terms not only through POS tagging but also through semantic embeddings. Semantic embeddings are employed in natural language processing tasks and refer to a way for a script to assign meaning to words by turning them into numbers. The assigned numbers signal word contexts and their relationship to one another in an embedding (numeral) space. This hybrid approach marked similar words related to the terms through cosine similarity and ensured a richer text analysis. Example 6 shows a phrase, which LitLatBERT identified as corresponding to the queried keyword "Ukrainian refugees": "žmonės, kurie išvyks iš Ukrainos ir ieškos prieglobsčio" (Eng. *People who will depart from Ukraine and will seek refuge*). This exemplifies how LitLatBERT can identify similar multi-word phrases.

---

[5] That is why I am once again urging the government to pay attention to the Labour Party's ("Darbo partija") proposal on a better system of communication and ensuring better conditions for war refugees in Lithuania.

1. Mes turime būti pasirengę ir raginti kitus būti pasirengusius priimti žmones, kurie išvyks iš Ukrainos ir ieškos prieglobsčio. [6]

In contrast, the script, which used only POS tagging and SpaCy, only marked the lemma for "refuge" (Lit. *prieglobstis*). The POS tagging method entails a more extensive manual output analysis than the script with LitLatBERT. The words most frequently found alongside the keyword "refugees" were "Ukrainian" (Lit. *Ukrainos*), "war" (Lit. *karo*), and "people" (Lit. žmonės). "People" and words related to "people", such as "children" (Lit. *vaikai*) and "women" (Lit. *moterys*), were also commonly mentioned when talking about refugees. Using human-denoting nouns indicates an alignment of Ukrainian refugees with humanitarian sentiments. However, the script with LitLatBERT sometimes conflated references to RASIM from other countries with Ukrainian refugees. Example 7 revolves around Belarussians who fled from Belarus after the most recent presidential election in the country. Mentions such as these were few in the corpus. However, they did appear since some parliamentary discussions in the corpus were on migration and refugee rights in a more general sense rather than specifically on Ukrainian refugees.

1. Dar labai svarbus vienas dalykas, kad mes neturime užmiršti tų žmonių, **kurie bėgdami nuo režimo** atsirado čia, Lietuvoje.[7]

This result was identified four times in contexts where people from Belarus were defined in terms of running away from the Belarussian regime. Lexical units surrounding words for "migrants" included "illegal" (Lit. *neteisėti*), "crisis" (Lit. *krizė*), and "camp" (Lit. *stovykla*). Example 8 shows that migrants, even though not explicitly stated to be illegal, are indirectly referred to as illegal by specifying that their freedom of mobility will be decided upon later. Example 9 demonstrates a frequent collocation related to migrants, referring to the 2019 migrant crisis in Lithuania when Belarus allegedly sent migrants from the Middle East to Lithuania.

1. Migracijos departamentas ir VSAT'as dėl kiekvieno **migranto** apgyvendinimo ir jo judėjimo laisvės ribojimo spręs individualiai.[8]

2. O kalbant apie atsakomybę dėl **migrantų krizės**, vienintelis, iš kurio turime reikalauti atsakomybės už **migrantų krizę** ir nusikaltimus žmogiškumui, yra A. Lukašenka ir jo režimas.[9]

---

[6] We have to be prepared to encourage others to be prepared to accept people who will depart from Ukraine and seek refuge.

[7] One more very important thing is that we must not forget the people who, fleeing the regime, reached Lithuania.

[8] The Department of Migration and VSAT (State Border Guard Service) will make separate decisions about each migrant's accommodation and freedom of movement.

[9] And talking about the responsibility for the migrant crisis, the only one of whom we must demand responsibility for the migrant crisis and crimes against humanity is A. Lukashenka and his regime.

The differences in words near "migrants" and "refugees" reveal specific differences in perception of the two RASIM keywords. It is difficult to determine whether they signify an ideological stance or refer to legislative terms. According to the State Commission of the Lithuanian Language, "a migrant" is an umbrella term for people or animals who change locations, while "a refugee" refers to people displaced by war. However, in the studied corpus, "migrants" were most often associated with illegal means of crossing borders and "refugees" were not referred to in terms of migration. Example 10 shows how migrants were sometimes mentioned in metaphoric expressions involving water. The word "užtvindyti" (Eng. *to flood*) referred to A. Lukashenka's plan to hinder the European Union's functions by increasing the number of migrants in the countries neighbouring Belarus.

1. Jeigu A. Lukašenkos planas **užtvindyti Europos Sąjungą migrantais** būtų pavykęs, šiandien kalbėtume jau ne apie teisinio reguliavimo tobulinimą, o apie rimtas pasekmes mūsų šalies ir visuomenės saugumui.[10]

Although not many references of this kind were found in the corpus, it is noteworthy that Ukrainian refugees and Belarussian people fleeing from Lukashenka's regime were not mentioned in terms of flooding or related to illegal actions. Such references were used in the context of the 2019 migrant crisis, and they referred to people being sent from the Middle East through Belarus to the European Union.

Modal words were analyzed using a script that identified keyword occurrences in the corpus based on the predefined categories of categorical reliability, problematic reliability, direct and indirect evidentiality, possibility, and necessity. The output provided insights into how often and by which lexical means the predefined categories of modality appeared in the corpus. The results revealed that secondary modality was frequent in the discourse surrounding Ukrainian refugees. However, the script assigned some sentences to more than one modality type. For example, some sentences were marked as conveying problematic reliability and possibility. Example 11 shows how the expression "in reality" (Lit. *iš tikrųjų*) was marked four times: three times for problematic reliability as the script first detected an exact match "iš tikrųjų", then "it is known / certainly" (Lit. *žinoma*), and "understandable" (Lit. *suprantama*) were identified as synonyms; and, finally, the phrase was marked as expressing direct evidentiality since the script marked "man atrodo" (Eng. *It seems to me*") as an item similar in meaning.

1. Kas lankėsi pabėgėlių priėmimo centruose – **iš tikrųjų** darbas vyksta gana organizuotai.[11]

---

[10] If A. Lukashenka's plan to flood the European Union were successful, today we would speak not about improving legal regulations but about serious consequences to the safety of our county and society.

[11] Those who visited refugee centers – truly, work is being done in an organized manner.

A more precise quantitative corpus analysis would require fine-tuning the script to mark each phrase as one category. Looking at the distribution of modality and evidentiality categories prescribed by the script using SpaCy, both primary and secondary modalities were relatively frequent in sentences describing refugees and migrants. Necessity was quite common, expressed by the verb "turime" (Eng. *we must*) and verbs preceded by the adverb "tikrai" (Eng. *really, certainly*). This can be seen in examples 12 and 13.

Mes **turime būti pasirengę** ir raginti kitus būti pasirengusius priimti žmones, kurie išvyks iš Ukrainos ir ieškos prieglobsčio. Žinau, kad tiek valstybės institucijos, tiek savivaldybės jau yra padariusios čia namų darbus, tačiau to negana.[12]

1. Mes **tikrai palaikysime** įstatymo „Dėl užsieniečių teisinės padėties" ir kitų projektų paketą, siekdami sustiprinti ir paremti karo pabėgėlius.[13]

As for evidentiality (which indicates the speaker's source of information), direct evidentiality occurred more frequently than indirect, suggesting that speakers were more inclined to mark themselves as the source of information. Direct evidentiality was expressed with words like "žinau" (Eng. *I know*) (example 12), sometimes used in the plural "žinome" (Eng. *we know*) (example 14).

1. **Mes su jumis žinome,** galbūt kol kas nelabai detaliai, bet Vyriausybė yra skyrusi nemažai dėmesio ir priėmusi sprendimus ir šiuo klausimu, kad mes ir Lietuvoje galėtume priimti pabėgėlius iš Ukrainos.[14]

Indirect evidentiality, expressed by impersonal verbs (e.g. "manoma" (Eng. *It is thought*), "kalbama" (Eng. *It is spoken*) or the infinitive form was not found. This may indicate that speakers in the studied corpus wanted to express their personal involvement in the topic, and the frequent instances of necessity and direct evidentiality may convey the joint authoritative stance of the Seimas.

CONCLUSION

The analysis of debates on Ukrainian refugees in the Lithuanian Seimas provided insights into how linguistic elements shape the framing of refugees and migrants.

---

[12] We have to be prepared to encourage others to be prepared to accept people who will depart from Ukraine and seek refuge. I know that government institutions and municipalities have done their homework here, but that is not enough.

[13] We will certainly support the package of the Law on the Legal Status of Aliens and other proposals to strengthen and help war refugees.

[14] You and I know, perhaps not in great detail yet, but the Government has paid a lot of attention and made decisions on this issue, too, so that we could accept refugees from Ukraine in Lithuania.

The study revealed that two RASIM keywords, "pabėgėlis" (Eng. *refugee*) and "migrantas" (Eng. *migrant*), were mentioned significantly more often than terms like "imigrantas" (Eng. *immigrant*) and "prieglobsčio ieškotojas" (Eng. *asylum seeker*). The two most frequent terms mainly appeared in the plural, pointing to a broader tendency to employ collective terms when referring to Ukrainian refugees and migrants, which could contribute to the depersonalization of the social groups involved.

The semantic analysis revealed that Ukrainian refugees were often mentioned with words denoting humans and accompanied by verbs like "help" and "support". However, when discussing migrants, terms like "neteisėti" (Eng. *illegal*) and "krizė" (Eng. *crisis*) were prevalent, reflecting a different attitude. A qualitative study of modal words showed that when talking about Ukrainian refugees, members of the Seimas tend to choose linguistic items of assuredness (categorical reliability), such as "aišku" (Eng. *It is clear*) or "suprantama" (Eng. *It is understandable*). This tendency may suggest a consensus about the need to support refugees. The analysis of modal words in the keyword context of "refugee" and "migrant" revealed a shared stance among Lithuanian politicians regarding the necessity of supporting Ukrainian refugees. However, the use of language expressing uncertainty and caution when discussing migrants from other regions indicates that the framing of these two groups is ideologically distinct. The frequent choice of linguistic items that indicate necessity in discussions about refugees implies a moral or legal obligation to assist them. In contrast, the more cautious language surrounding migrants reflects a more transactional or conditional approach. These findings suggest the need for policies focusing on humanitarian assistance and integration while addressing the potential risks of negative stereotyping in public discourse.

While the hybrid approach combining POS tagging and LitLatBERT allowed for a deeper understanding of the discourse, there were some limitations. The POS tagging script provided surface-level insights into the grammatical context of refugee-related terms. It did not capture the deeper meanings or relationships between terms, which could lead to incomplete or misleading interpretations. The LitLatBERT model took context into account but sometimes conflated references to refugees from different countries (e.g., Belarusian refugees) with Ukrainian refugees, leading to occasional noise in the results. This highlights the need for fine-tuning and further refinement of the model to ensure more precise results, as well as the need for manual validation of the output. In future research, there is room for refining the tools used for analysis, particularly the LitLatBERT model, to improve accuracy and mitigate issues with semantic noise. Moreover, a broader comparison of how different groups are framed across multiple European countries could provide insights into transnational discourse on refugees and migration.

REFERENCES

Akelaitis, Gintautas. *Dabartinės lietuvių kalbos įterptiniai vienetai (semantika, struktūra, paskirtis diskurse)*. Humanitarinių mokslų daktaro disertacija. Vilnius: Vilnius University, 1992.

Anthony, Lawrence. *AntConc (version 4.3.1)* [Computer Software]. Tokyo: Waseda University. Internet access: https://www.laurenceanthony.net/software [retrieved 2024 11 05].

Baker, Paul. Discourses of Refugees and Asylum Seekers in the UK Press, 1996– 2006: Full Research Report. *ESRC End of Award Report, RES-000-22-1381*. Swindon: ESRC, 2007.

Baker, Mona, and Gabriela Saldanha. Ideology. In *Routledge Encyclopaedia of Translation Studies*. New York: Routledge, 2020, 252–256.

Bird, Steven, Loper, Edward, and Ewan Klein. *Natural Language Processing with Python*. Newton: O'Reilly Media Inc., 2009.

Charteris-Black, Jonathan. Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse and Society*, 2006, nr. 17(6), 563–582.

Devlin, Jacob, Chang, Ming–Wei, Lee, Kenton, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Minneapolis: Association for Computational Linguistics, 2019, 4171–4186.

Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya, Ljubešić, Nikola Simov, Kiril, Pančur, Andrej, Rudolf, Michał, Kopp, Matyáš, Barkarson, Starkaður, Steingrímsson, Steinþór, Çöltekin, Çağrı, de Does, Jesse, Depuydt, Katrien, Agnoloni, Tommaso, Venturi, Giulia, Pérez, María Calzada, de Macedo, Luciana D., Navarretta, Costanza, Luxardo, Giancarlo, Coole, Matthew, Rayson, Paul, Morkevičius, Vaidas, Krilavičius, Tomas, Darģis, Roberts, Ring, Orsolya, van Heusden, Ruben, Marx, Maarten, and Darja Fišer. *The ParlaMint Corpora of Parliamentary Proceedings*. *Language Resources and Evaluation*, 2022. https://doi.org/10.1007/s10579-021-09574-0 [retrieved 2024 10 25].

Fairclough, Norman. *Language and power*. London: Longman, 1989.

Fairclough, Norman. *Critical Discourse Analysis: The Critical Study of Language*. London: Routledge, 2010.

Freeden, Martin. *Ideology: A Very Short Introduction*. Oxford: Oxford University Press, 2003.

Garabik, Radovan, and Indrė Pileckytė. *From Multilingual Dictionary to Lithuanian WordNet*. 2013. Internet access https://api.semanticscholar.org/CorpusID:6416578 [retrieved 2024 11 08]

Gupta, Ayan. *BERT for Discourse Analysis: A Pragmatist Approach to Governmentality*. Doctoral Thesis. Bristol: University of Bristol, 2023.

Honnibal, Matthew, and Ines Montani. *SpaCy 2: Natural Language Understanding with Bloom Embeddings, Convolutional Neural Networks and Incremental Parsing*. 2017. Internet access https://spacy.io/ [retrieved 2024 11 07].

Islentyeva, Anna. *Corpus-Based Analysis of Ideological Bias. Migration in the British Press*. London: Routledge, 2021.

Kapočiūtė-Dzikienė, Jurgita, Šarkutė, Ligita, and Andrius Utka. Seimo posėdžių stenogramų tekstynas autorystės nustatymo bei autoriaus profilio sudarymo tyrimams. *Kalbotyra*, 2014, vol. 66, 27–45. https://hdl.handle.net/20.500.12259/35097 [retrieved 2024 11 07].

Morkevičius, Vaidas. *Socialinių vertybių raiška politiniame diskurse: Lietuvos Seimo debatų turinio analizė (1992–2004 m.)*. Kaunas, 2005.

Ramonaitė, Ainė., and Adelė Vaiginytė. Tendencingumo paieškos politikų valdomoje žiniasklaidoje: Lietuvos regioninės spaudos analizė taikant kompiuterinės turinio analizės metodus. *Politologija*, 2020, nr. 100, 34–68. Internet access https://doi.org/10.15388/PoLit.2020.100.2 [retrieved 2024 11 15].

Reisigl, Martin, and Ruth Wodak. *Discourse and Discrimination: Rhetorics of Racism and Anti-Semitism*. London: Routledge, 2001.

Reisigl, Martin, and Ruth Wodak. The Discourse-Historical Approach. In *Methods of Critical Discourse analysis*. Eds. Martin Reisigl and Ruth Wodak. London: Sage, 2008, 87–121.

Ruskan, Anna. Evidencialumo raiškos priemonės lietuvių kalboje. *Lietuvių kalba*, 2010, nr. 4, 1–10. doi: 10.15388/LK.2010.22857 [retrieved 2024 11 28].

Ruskan, Anna. Evidential and epistemic adverbials in Lithuanian: evidence from intra-linguistic and cross-linguistic analysis. *Kalbotyra*, 2017, t. 70, 127–152. Internet access https://doi.org/10.15388/Klbt.2017.11197 [retrieved 2024 11 29].

Šolienė, Audronė. *Episteminio modalumo ekvivalentiškumo parametrai anglų ir lietuvių kalbose*. Vilnius: Vilnius University, 2013.

Thompson, John Brookshire. *Studies in Theory of Ideology*. USA: University of California Press, 1990.

Usonienė, Aurelija. Modalumas anglų ir lietuvių kalbose: forma ir reikšmė. Vilnius: Vilnius University, 2004.

Usonienė, Aurelija. Episteminio modalumo raiška. Ekvivalentiškumo bruožai anglų ir lietuvių kalbose. *Darbai ir dienos*, 2006, nr. 45, 97–108.

Van Dijk, Teun Adrianus. *Ideology. A Multidisciplinary Approach*. London: Sage, 1998.

Van Dijk, Teun Adrianus. Ideology and discourse analysis. *Political Ideology*, 2006, nr. 11(2), 115–140. Internet access https://doi.org/10.1080/13569310600687908 [retrieved 2024 11 05].

Van Leeuwen, Theo. *Discourse and Practice. New Tools for Critical Discourse Analysis*. Oxford: Oxford University press, 2008.

Verschueren, Jef. *Ideology in Language Use. Pragmatic Guidelines for Empirical Research*. Antwerp: University of Antwerp, 2012.

Wodak, Ruth. What CDA Is about – A Summary of Its History, Important Concepts, and Its Developments. In *Methods of Critical Discourse Analysis*. Eds. Ruth Wodak and Michael Meyer. London: Sage Publications, 2001, 1–13.

Aušra Urbanavičiūtė
Vytauto Didžiojo universitetas, Lietuva

LIETUVOS SEIMO DISKURSO ANALIZĖ PRIEGLOBSČIO IEŠKANČIŲ UKRAINIEČIŲ TEMATIKA TAIKANT TEKSTO ANALIZĖS ĮRANKIUS

SANTRAUKA. Straipsnyje nagrinėjamas Lietuvos Seimo debatų diskursas Ukrainos pabėgėlių tematika. Tyrime pasitelkta „Natural Language Toolkit" (NLTK) „Python" kalbos biblioteka, kuri yra dažnai naudojama atliekant kompiuterinės lingvistikos tyrimus. Taip pat naudotas dvikalbis LitLatBERT transformeriais grįstas mašininio kalbos mokymosi modelis. Pusiau automatizuotu būdu analizuojami leksiniai vienetai, kuriais apibūdinami pabėgėliai ir migrantai. Taip pat atlikta pusiau automatizuota modalumo raiškos analizė, siekiant nustatyti Seimo narių požiūrį į pabėgėlius ir migrantus. Straipsnio tikslas – pusiau automatizuotais būdais išanalizuoti, kaip Lietuvos Seimo nariai per parlamento debatus įvardija Lietuvoje prieglobsčio ieškančius ukrainiečius ir migrantus.

RAKTAŽODŽIAI: Seimo debatai, pabėgėlių ir migrantų diskurso analizė, NLTK, LitLatBERT.