

WHEN IS A CRISIS REALLY A CRISIS?
USING NLP AND CORPUS LINGUISTIC METHODS TO REVEAL
DIFFERENCES IN MIGRATION DISCOURSE ACROSS CZECH MEDIA

ONDŘEJ PEKÁČEK¹ – IRENE ELMEROT²

¹ Department of Sociology, Faculty of Social Sciences, Charles University,
Prague, Czech Republic

² Department of Slavic and Baltic Studies, Finnish, Dutch and German,
Faculty of Humanities, Stockholm University, Stockholm, Sweden

PEKÁČEK, Ondřej – ELMEROT, Irene: When Is a Crisis Really a Crisis? Using NLP and Corpus Linguistic Methods to Reveal Differences in Migration Discourse across Czech Media. *Journal of Linguistics*, 2023, Vol. 74, No 1, pp. 369 – 380.

Abstract: This article presents an interdisciplinary analysis of discourses on refugees, asylum seekers, immigrants, and migrants (RASIM) in mainstream and alternative media in the Czech Republic. Using techniques from corpus linguistics (CL) and natural language processing (NLP) and drawing on insights from media sociology, we demonstrate the value of an interdisciplinary approach for conducting robust research that can inform policymakers and media practitioners. Our analysis of nearly one million documents from January 2015 to February 2023 reveals distinctive terms and phrases used by alternative media, highlighting the growing divergence between the mainstream and alternative media discourse and its intensity over different periods. These findings have implications for understanding the mobilization of anti-systemic groups, particularly those on the far right.

Keywords: European refugee crisis, Czech media, alternative journalism, corpus linguistics, media sociology, natural language processing, RASIM

1 BACKGROUND

Since 2015, Europe has experienced a significant influx of refugees, mainly from war-torn countries such as Syria, Iraq, and Afghanistan. The arrival of Ukrainian refugees following the Russian invasion in February 2022 prompted reflection on the lessons learned from the previous “crisis” and led to calls for comparative studies. At the same time, factors such as the prominence given to migration by different types of media and the nature of the discourse can influence public attitudes and whether migration is perceived as a crisis, which in turn affects support for populist right-wing parties (Strömbäck et al. 2021; Lecheler et al. 2019; Klawier et al. 2022). Previous studies on discourses surrounding refugees, asylum seekers, immigrants, and migrants (RASIM) focus on a few European countries, e.g. Germany and the United Kingdom (Eberl et al. 2018; Baker et al. 2008). However, there is a limited amount of scholarly attention on this topic in the Czech context,

which, as we argue, presents several compelling research opportunities. For example, the Czech media paid considerable attention to migration in and after 2015, despite the marginal number of asylum applications (Jelínková 2019). The situation changed significantly in February 2022, and by the end of that year, the country hosted the highest number of Ukrainian refugees per capita (Plevák 2023).

The Czech media landscape experienced significant changes in ownership concentration, political pressure, and technology in the 2000s. Online alternative¹ journalism has emerged, but often takes a hostile stance toward marginalized communities and liberal democracy (Holt et al. 2019; Štětka et al. 2021) and may spread disinformation about migration (Gregor – Mlejnková 2021). With only 34% of the Czech population trusting media (Newman et al. 2022), results may differ from Western European findings on media consumption and immigration attitudes (Kondor et al. 2022).

Scholars studying RASIM or ethnicity-based media coverage face challenges due to the large data volume; therefore, qualitative studies predominate (Seo – Kavakli 2022). In the Czech context, most studies either use qualitative methods (Průchová Hružová 2021) or small data sets from limited outlets (Kluknavská 2021), with an overreliance on manual content analysis (Esser et al. 2019). Using computational methods, such as clustering (Urbániková – Tkaczyk 2020) or corpus-assisted discourse analysis (Elmerot 2021, 2022), is rare. Natural language processing (NLP) tools have gained popularity in interdisciplinary studies, with landmark Törnbergs' (2016) combination of topic modeling and discourse analysis, and some Corpus Linguistics (CL) scholars advocate the utility of adding close reading to reveal the meaning behind automated results (Brookes – McEnery, 2019).

This study aims to analyse parts of the Czech media discourse on international migration in mainstream and alternative media, covering eight years from January 2015 to February 2023, focusing on three methods. The aim is to show when different types of media consider migration a crisis and to shed light on the role of the media in shaping migration representations and public opinion in the Czech Republic.

2 RESEARCH DESIGN

2.1 Background to the research questions

Firstly, we focus on the linguistic distinction between forced and voluntary migration in the Czech mainstream and alternative media. Word choice is crucial to understanding social group othering (Elmerot 2022), and word frequency analysis is

¹ We follow the definition by Holt et al. (2019, p. 862), which argues: “Alternative news media represent a proclaimed and/or (self-) perceived corrective, opposing the overall tendency of public discourse emanating from what is perceived as the dominant mainstream media in a given system.”

an effective tool (Brouwer et al. 2017). We define ‘refugee’ as a label for people forced to leave their country for fear of persecution or harm, contrasting it with ‘migrant,’ who moves to improve their status (Douglas et al. 2019). Our first research question is:

RQ1: Between January 2015 and February 2023, what differences in usage of terms for voluntary versus forced migration are there between mainstream and alternative media?

Secondly, the power of individuals and organizations to shape media messages on migration is often emphasized in political communication literature, but studies rarely analyze their presence in the discourse (Boomgaarden – Vliegenthart 2009). Named Entity Recognition (NER) is a suitable method to find individuals and organizations, but rarely used with a migration focus (Nemes – Kiss 2021). Our second research question is:

RQ2: What are the differences in the presence of actors in the RASIM news coverage between mainstream and alternative media during periods of large refugee influxes in 2015 and 2022–2023?

Thirdly, collocation analysis is a CL tool that can provide insights into the linguistic patterns of discourse. Collocates frequently occur near the target word and can reveal significant associations between concepts (Stubbs 1995). Collocations have been widely used in RASIM studies, notably by Baker et al. (2008), who found eight consistent collocational categories. More recently, Zawadska-Palucktau (2023) analysed the portrayal of Ukrainian refugees in the Polish mainstream press during the first week of March 2022. She found that Ukrainians were more welcome in Poland than refugees from the Middle East and were more frequently referred to as war refugees. To better understand the framing (or “semantic preference”) of crucial migration terms in the Czech media discourse, we examine seasonal collocations across media types and periods (Baker et al. 2008, pp. 278–286). Our final research question is:

RQ3: In the periods 2015–2016 vs. 2022–2023, what are the dominant collocates of the terms “refugee,” “migrant,” “immigrant,” and “asylum seeker” in Czech mainstream and alternative media?

2.2 Corpus of Czech Media News on Migration

To obtain a comprehensive corpus of Czech migration-related news, we used the Newton Media Archive API to access full-length documents (including text articles and audio transcripts) with a Czech Boolean search adapted from Esser et al.

(2019).² The search was limited from 1st January 2015 to 28th February 2023 to capture at least 12 months of media coverage of each migration “crisis.”³

The search parameters yield nearly one million documents from over four thousand Czech online and offline media (see Tab. 1). The total published content of all these media is 41 million documents. Our migration corpus thus represents a small fraction of the total media output. However, during the peak of interest in March 2022, migration was a topic in about 10 percent of all documents. Migration-related documents appear in all media sections, sometimes even in sports and technology.

Documents	Media	Sentences	Tokens
998,740	4,166	48,400,000	800,000,000

Tab. 1. Czech Migration News corpus, January 2015 to February 2023

2.3 Dataset with media type labels

To investigate the differences in migration discourse between different media types, we use the labels from the ONLINE2_NOW corpus of the Czech National Corpus (Cvrček et al. 2022). To adapt it to our research, we extended it by adding offline news according to the same key and then merged the categories “anti-system” and “political tabloid” (Cvrček – Fidler 2022, p. 268) into a broader category of alternative media, as some prominent media in the latter category (such as *Parlamentní Listy*) position themselves antagonistically to the mainstream (Štětka et al. 2021).

The media type and article datasets were merged, resulting in 2,735 labeled media. The resulting dataset of 971,000 documents with an identified media type includes approximately 189,000 alternative media documents (Tab. 2) and 342,000 mainstream media documents (Tab. 3).

Period	Documents	Media	Sentences	Tokens
whole	189,783	109	9.2M	156M
Jan.–Dec. 2015	14,016	27	0.7M	12.3M
Jan. 2016–22	161,220	90	7.6M	129.4M
Feb. 2022–2023	14,547	75	0.8M	14.2M

Tab. 2. Alternative Media sub-corpus, split by periods. M = million.

² The string: *běženec** OR *běženk** OR *imigrant** OR *migra** OR *imigra** OR *pristěhoval** OR *uprchl** OR *utečen** OR *azylant**. (The English translation of *běženec*, the female form *běženkyně*, *uprchlík/uprchlice* and *utečenec*, are all ‘refugee’. Both *imigrant* and *pristěhovalce* mean ‘immigrant’ in English, and *azylant / azylantka* ‘male/female asylum seeker.’) The only difference from the original string is the exclusion of stems related to integration, assimilation, and deportation, which yielded many irrelevant documents for our study.

³ Our analysis workflow and additional visualizations are available in an open-source GitHub repository: https://github.com/opop999/media_discourse_research. Larger raw data files are available via the OSF repository: <https://osf.io/j28v3>.

Period	Documents	Media	Sentences	Tokens
whole	342,383	226	17M	276M
Jan.–Dec. 2015	48,145	109	1.8M	27.6M
Jan. 2016–22	222,858	222	9.6M	156.2M
Feb. 2022–2023	71,380	193	5.6M	92M

Tab. 3. Mainstream Media sub-corpus, split by periods. M = million.

Fig. 1 shows a distinct difference between the proportion of overall coverage devoted to migration in the mainstream ($M = 4.3\%$, $SD = 0.1\%$) and alternative media ($M = 13.9\%$, $SD = 2.5\%$). This contrast is most apparent in mid-2017, with alternative media focusing on migration in over 70% of their documents.

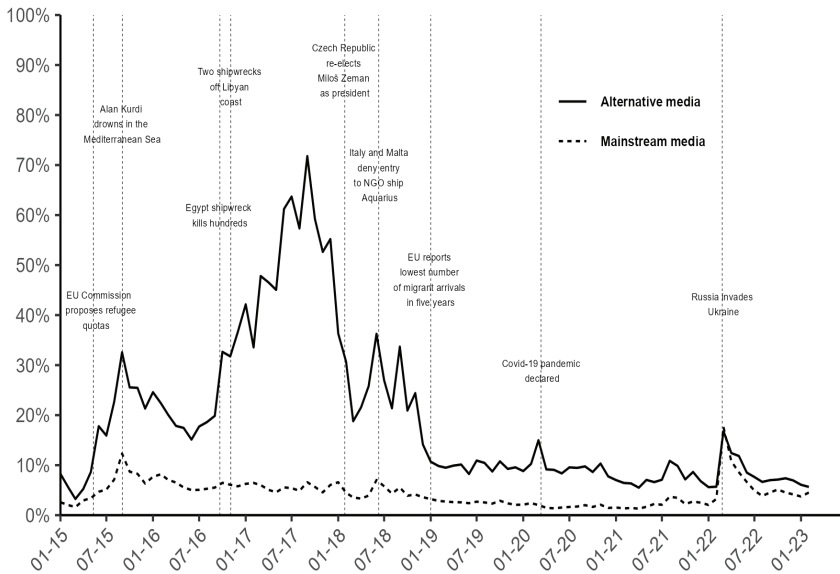


Fig. 1. The proportion of migration content across Czech media types

2.4 Data pre-processing

Czech is a highly inflected language with 13 word forms, making it challenging for automated natural language processing (Lenc – Hercig 2016). We apply several preprocessing steps to prepare the corpus for more computationally intensive analysis, including text cleaning and tokenization using R’s UDPipe library (Wijffels 2023) and a pre-trained model based on the Czech PDT UD treebank 2.6 (Straka 2018). Lemmatization reduces words to their base forms, and a deep learning model generates embeddings to increase the accuracy of this process (Straka et al. 2019). Finally, part-of-speech tagging (POS) identifies the part of speech of each token.

3 ANALYSES AND FINDINGS

3.1 Word-frequency analysis of migration labels

To operationalize our voluntary versus forced migration study, we again used our migration string (see 2.2) and applied it to the lemmatized dataset. We then removed matching lemmas that appeared less than ten times in the entire corpus, resulting in 352 terms. Next, both authors manually reviewed each of these terms and categorized them as either “refugee term” (98 types), “migration term” (198 types), or “unknown” (56 types). We used this lexicon to examine their relative monthly frequency over the entire period, stratified by alternative media and mainstream categories.

Regarding RQ1, refugee terms were more frequently used in the mainstream media between February 2015 and May 2016, with a peak around the finding of the drowned boy Alan Kurdi (see Fig. 2). After that, migrant terms were used more frequently, with a peak around the start of the invasion of Ukraine in 2022. On the other hand, alternative media had a small peak of more refugee terms at the time of Alan Kurdi and again from March 2022 to mid-2022. Mainstream media used refugee terms for longer in both 2015 and 2022, while alternative media returned to using migrant terms more quickly.

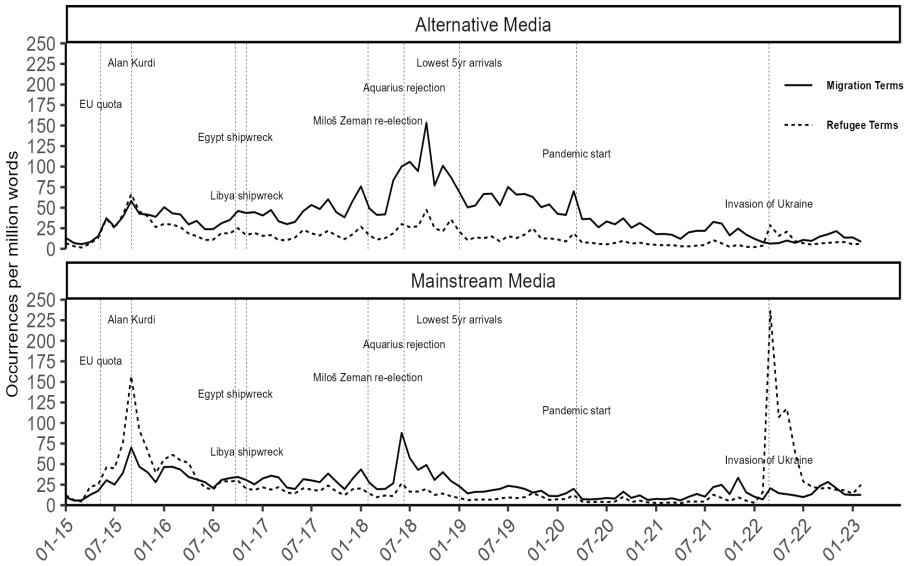


Fig. 2. Relative monthly frequencies of terms by media types

3.2 Named Entities in migration news coverage

We use Named Entity Recognition (NER) to identify and classify named individuals and organizations in our corpus, to the context influencing Czech migration discourse. The NameTag 2 model (Straková et al. 2019) identifies multiword entities from eight “parent” and forty-six “child” categories. Due to the inflectional nature of the Czech language, we use non-lemmatized text and apply stemming after analysis to obtain interpretable summaries of top entities in mainstream and alternative media. Our summary focuses on geographic locations, personal names, media organizations, and other institutions.

Regarding RQ2, the named entities of places and persons differ between media types. Five of the top thirty most frequent entities in alternative media in 2015 (see Fig. 3) are absent from mainstream media: “the West,” NATO, China, the USA, and the Czech conservative party ODS. This result supports Cvrček and Fidler’s (2022) conclusions that a Czech NATO exit was a part of the anti-systemic media discourse. Several entities have a higher frequency ranking in alternative media, such as Russia(n), ISIS, Ukraine, the USA, and Africa. Most entities unique to mainstream media are geographical. In the intermediate period (2016–2022), unique entities appear in the alternative media: Ukraine, NATO, “the West,” and Africa. One unique entity in the mainstream media is Donald Trump, while the others are primarily geographical.

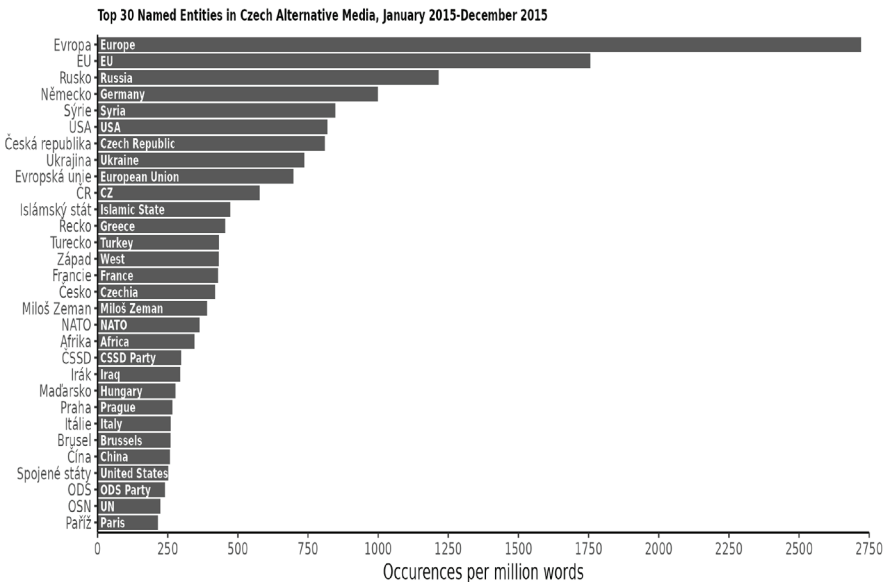


Fig. 3. Top named entities in alternative media, January to December 2015

In the final period (February 2022–2023), NATO is no longer unique in the alternative media. However, more unique entities in each media type may reveal a growing split between the two. Mainstream media refer to cities like Kharkiv and Mariupol, while alternative media refer to countries like Sweden and Turkey.

In summary, “the West” is one of alternative media’s most frequent geographical and personal entities. Furthermore, Ukraine was among the top thirty most mentioned entities in alternative media already in 2015. A future study could thus further examine this pre-2022 invasion discourse better to understand the messaging of the prominent Kremlin-linked alternative media, similar to Cvrček and Fidler (2022). Overall, there are about twice as many unique entities in both media types during the last period (four to five in periods 1 and 2, eight and nine in period 3), pointing to a growing gap between the mainstream and alternative media types.

3.3 Collocations

To address RQ3, we compared the first period, 2015, with the third period, 2022–2023, using the collocates⁴ of four lemmas: ‘migrant’, ‘immigrant’, ‘refugee’, and ‘asylum seeker’. Here, we focus on the Provenance/transit/destination, Number, and Legality collocate groups, identified by Baker et al. (2008).

Provenance/transit/destination

In 2015, all four migration terms were associated with, e.g. the lemmata ‘country,’ ‘Europe,’ and ‘state’ in both media types. ‘Migrant’ and ‘immigrant’ were associated with ‘border,’ while ‘refugee’ was associated with ‘Syria.’ ‘Migrant’ was also associated with ‘territory.’ In alternative media, ‘refugee,’ ‘migrant,’ and ‘immigrant’ co-occurred with ‘Africa.’ In 2022–2023, only ‘country’ or ‘land’ (*země*) remained within the top thirty collocations for all four nouns in both media types. We observe a shift towards Ukraine and smaller geographical entities such as ‘town’ and ‘region.’ Europe now has a much looser association than in the first period. Africa appeared only in alternative media as a collocation of ‘immigrant’ and ‘migrant.’ Sweden emerged as a prominent geographical name in the same media type, especially for ‘asylum seeker,’ just as in the NER analysis.

Numbers and numerical lemmata

During the first period, the term ‘million’ co-occurs with ‘refugee’ in both media types, with mainstream collocations also including ‘thousand.’ However, neither numerical lemma is in the top thirty for ‘asylum seeker.’ Instead, we find ‘number,’ ‘quota,’ and ‘contingent’ in both media types. In the third period, ‘quota’ and ‘contingent’ are absent. ‘Million’ appears in alternative media with ‘refugee.’ In

⁴ Like the korpus.cz’s KonText tool, we use the logDice association measure, here with a context window of three words on each side.

contrast, ‘number’ appears in both media types around ‘migrant’ and ‘immigrant,’ but only in the mainstream media for ‘asylum seeker’ (see Fig. 4).

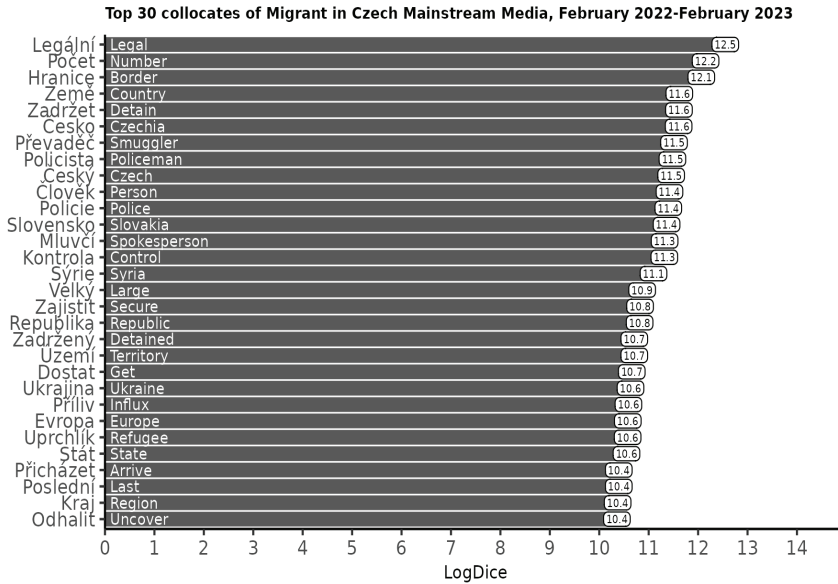


Fig. 4. Collocates with ‘migrant’, mainstream media, February 2022 to February 2023

Legality

In the first period, ‘illegal’ and ‘legal’ are found with ‘immigrant.’ At the same time, with ‘migrant,’ both collocations appear in alternative media but ‘legal’ only in the mainstream media. However, the non-lemmatized text may negate this as ‘illegal.’⁵ In the third period, ‘illegal’ is the top collocate with ‘immigrant’ in alternative media and number 13 in mainstream media, but absent with ‘migrant.’

In response to RQ3, a radical shift in discourse concerning all four migration terms is visible between February 2022 and February 2023, with increased co-occurrence of Czech geographical names and neighboring countries. For example, the word ‘quota’ disappears from prominent discourses when Ukrainian refugees and asylum seekers arrive.

4 CONCLUSION

Our article examines differences in RASIM crisis discourse in nearly one million Czech alternative and mainstream media documents over eight years. Using CL and

⁵ Czech negations may use the prefix *ne-*, which is removed when lemmatizing.

NLP methods informed by media sociology, we conclude that alternative media often use voluntary flight terminology and frequently refer to ‘the West’ as an actor.

In addition, Ukraine is continuously mentioned in a RASIM context in alternative media during 2015–2023. The gap between media types grows as more refugees arrive from Ukraine after February 24, 2022. Mainstream media then focus on Ukrainian cities affected by the war and Czech cities providing aid. At the same time, alternative media discuss international actors, e.g. ‘the West,’ Sweden, and Turkey. Language usage also changes, with ‘illegal’ collocating with ‘immigrant’ continuously but not with ‘migrant’ in 2022–2023.

Moreover, the 2022 crisis is portrayed as a reality, with increased mentions of Czech geographical names and neighboring countries in both media types compared to the more distant RASIM coverage of 2015–2016. These findings have implications for understanding far-right media communication and anti-immigrant sentiment and aim to inform policymakers and media practitioners in promoting responsible and informed migration reporting.

Our study highlights the need for further research on Czech migration discourse. For instance, verbal aspects connected to voluntary or forced migration could vary significantly between media types. Finally, linking social media data with news data and examining sources and co-occurrence origins in both media types could also shed light on how particular ideas spread in the public sphere.

ACKNOWLEDGEMENTS

The research has been supported by the *Gunvor och Josef Anérs Stiftelse* foundation, application number FB22-0088, and by the Specific University Research (SVV) grant no. 260 728.

References

- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., and Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), pages 273–306.
- Boomgaarden, H. G., and Vliegthart, R. (2009). How news content influences anti-immigration attitudes: Germany, 1993–2005. *European Journal of Political Research*, 48(4), pages 516–542.
- Brookes, G., and McEnery, T. (2019). The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 21(1), pages 3–21.
- Brouwer, J., van der Woude, M., and Leun, van der, J. (2017). Framing migration and the process of crimmigration: A systematic analysis of the media representation of unauthorized immigrants in the Netherlands. *European Journal of Criminology*, 14(1), pages 100–119.

Cvrček, V., and Fidler, M. (2022). No Keyword is an Island: In search of covert associations. *Corpora*, 17(2), pages 259–290.

Cvrček, V., Jeziorský, T., and Henyš, J. (2022). ONLINE2_NOW: monitoring corpus of online Czech. Ústav Českého národního korpusu FF UK, Praha. Accessible at: <http://www.korpus.cz>.

Douglas, P., Cetron, M., and Spiegel, P. (2019). Definitions matter: Migrants, immigrants, asylum seekers and refugees. *Journal of Travel Medicine*, 26(2). Accessible at: <https://doi.org/10.1093/jtm/taz005>.

Eberl, J.-M., Meltzer, C. E., Heidenreich, T., Herrero, B., Theorin, N., Lind, F., Berganza, R., Boomgaarden, H. G., Schemer, C., and Strömbäck, J. (2018). The European media discourse on immigration and its effects: A literature review. *Annals of the International Communication Association*, 42(3), pages 207–223.

Elmerot, I. (2021). Income, nationality and subjectivity in media text. *Jazykovedný časopis*, 72(2), pages 667–678.

Elmerot, I. (2022). Constructing “Us” and “Them” through Conflicts – Muslims and Arabs in the News 1990–2018. In L. Filardo-Llamas – E. Morales-López – A. Floyd (eds.): *Discursive Approaches to Sociopolitical Polarization and Conflict* (1st ed.), pages 122–136. Routledge.

Esser, F., Stepińska, A., Pekáček, O., Seddone, A., Papathanassopoulos, S., Peicheva, D., Milojevic, A., Blassnig, S., and Engesser, S. (2019). Event-, Politics-, and Audience-Driven News: A Comparison of Populism in European Media Coverage in 2016 and 2017. In C. Reinemann – J. Stanyer – T. Aalberg – F. Esser – C. de Vreese (eds.), *Communicating populism: Comparing actor perceptions, media coverage, and effects on citizens in Europe*, pages 123–140. Routledge: Taylor & Francis Group.

Gregor, M., and Mlejnková, P. (2021). Facing Disinformation: Narratives and Manipulative Techniques Deployed in the Czech Republic. *Politics in Central Europe*, 17(3), pages 541–564.

Holt, K., Figenschou, T. U., and Frischlich, L. (2019). Key dimensions of alternative news media. *Digital Journalism*, 7(7), pages 860–869.

Jelínková, M. (2019). A Refugee Crisis Without Refugees: Policy and media discourse on refugees in the Czech Republic and its implications. *Central European Journal of Public Policy*, 13(1), pages 33–45.

Klawier, T., Prochazka, F., and Schweiger, W. (2022). Comparing Frame Repertoires of Mainstream and Right-Wing Alternative Media. *Digital Journalism*, 10(8), pages 1387–1408.

Kondor, K., Mihelj, S., Štětka, V., and Tóth, F. (2022). News consumption and immigration attitudes: A mixed methods approach. *Journal of Ethnic and Migration Studies*, 48(17), pages 4129–4148.

Lecheler, S., Matthes, J., and Boomgaarden, H. (2019). Setting the Agenda for Research on Media and Migration: State-of-the-Art and Directions for Future Research. *Mass Communication and Society*, 22(6), pages 691–707.

Nemes, L., and Kiss, A. (2021). Information Extraction and Named Entity Recognition Supported Social Media Sentiment Analysis during the COVID-19 Pandemic. *Applied Sciences*, 11(22). Accessible at: <https://doi.org/10.3390/app112211017>.

Newman, N., Fletcher, R., Robertson, C. T., Eddy, K., and Nielsen, R. K. (2022). Reuters Institute Digital News Report 2022.

Plevák, O. (2023, February 24). Czechia hosts most Ukrainian refugees per capita. <https://www.euractiv.com/section/politics/news/czechia-hosts-most-ukrainian-refugees-per-capita/>

Průchová Hružová, A. (2021). What is the image of refugees in Central European media? *European Journal of Cultural Studies*, 24(1), pages 240–258.

Seo, S., and Kavakli, S. B. (2022). Media representations of refugees, asylum seekers and immigrants: A meta-analysis of research. *Annals of the International Communication Association*, 46(3), pages 159–173.

Štětka, V., Mazák, J., and Vochocová, L. (2021). “Nobody Tells us what to Write about”: The Disinformation Media Ecosystem and its Consumers in the Czech Republic. *Javnost – The Public*, 28(1), pages 90–109.

Straka, M. (2018). UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task. *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 197–207.

Straka, M., Straková, J., and Hajič, J. (2019). Evaluating Contextualized Embeddings on 54 Languages in POS Tagging, Lemmatization and Dependency Parsing (arXiv:1908.07448). arXiv.

Straková, J., Straka, M., and Hajič, J. (2019). Neural Architectures for Nested NER through Linearization. *Proceedings of the 57th Annual Meeting of the Assoc. for Computational Linguistics*, pages 5326–5331.

Strömbäck, J., Meltzer, C. E., Eberl, J.-M., Schemer, C., and Boomgaarden, H. G. (2021). *Media and Public Attitudes Toward Migration in Europe: A Comparative Approach*. Routledge.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), pages 23–55.

Törnberg, A., and Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media*, 13, pages 132–142.

Urbániková, M., and Tkaczyk, M. (2020). Strangers ante portas: The framing of refugees and migrants in the Czech quality press. *European Journal of Communication*, 35(6), pages 580–596.

Wijffels, J. (2023). *udpipe: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the “UDPipe” “NLP” Toolkit*. Accessible at: <https://CRAN.R-project.org/package=udpipe>.

Zawadzka-Paluckta, N. (2023). Ukrainian refugees in Polish press. *Discourse & Communication*, 17(1), pages 96–111.