

PIOTR MIROCHA
Jagiellonian University in Kraków
piotr.mirocha@uj.edu.pl
ORCID: 0000-0002-5752-5890

THE PROBLEM OF MIGRATION IN POST-2007 SERBIAN MEDIA DISCOURSE ABOUT EUROPE: FROM TOPIC MODELLING TO QUALITATIVE DISCOURSE ANALYSIS¹

Keywords: Serbia, migration crisis, Critical Discourse Analysis, corpus linguistics, corpus-assisted discourse analysis

Abstract

The study uses linguistic corpus tools in order to establish in what contexts the issue of migration in Europe was addressed in the Serbian press in the period 2007–2017. About 20,000 randomly chosen texts from the dailies *Politika* and *Danas* were subjected to topic modelling, collocational profiling, and also a contextual examination in the case of the paper editions. The interpretation of the data considered the effect of the medium's ideological slant on the discourse structures used. It emerged that the selection of topics and the arrangement of texts in the liberal daily *Danas* corresponded to its pro-European profile. *Politika*, in turn, seemed to be gradually shifting towards the portrayal of Europe as being mired in crisis. In addition, the procedure used in the study enabled reflection on the usefulness of the results of topic modelling for media discourse analysis.

In 2015, at least for a brief moment, Serbia and other countries in South-eastern Europe witnessed events which were commented upon in the international arena. Thousands of refugees and migrants from the war- and poverty-stricken areas of the Middle East and North Africa started their journey from the Mediterranean

¹ The paper has been rendered into English by a professional translator (Magdalena Szczyrbak) under a POB Heritage grant from the Jagiellonian University to Studia Linguistica Universitatis Iagellonicae Cracoviensis.

Sea through the so-called Balkan route towards Western Europe. This, however, was not the first time Serbia had experienced large-scale migration, and the term “refugee” (Serb. *izbeglica*) as used in Serbian public discourses was linked to various denotations and connotations: refugees and people who were internally resettled within the former Yugoslavia, Roma people living on informal estates, and predominantly Roma asylum seekers who were forcibly resettled from Western Europe (Greenberg and Spasić 2017).

What is more, the connotations of refugee status as constructed by the media in South-eastern Europe were not stable and varied dramatically as events unfolded (Car et al. 2019). The presence of the migrants’ representation in the media was not without certain benefits. Šarić and Felberg (2019) argue that it was often the case that the goal was to show one’s own country in a good light, and the neighbouring countries in a relatively bad light, which could be easily linked to the processes of Europeanization or the European aspirations of the countries in South-eastern Europe.

The goal of the current study was primarily to use linguistic corpus tools to establish in what contexts the issue of migration in Europe was addressed in the Serbian press. Of great relevance to the investigation was the determination of the effect that the medium’s ideological slant had on the discourse structures used. The operationalization of this goal involved the use of linguistic corpus tools as well as an analysis of media contexts and a review of the keywords belonging to the topic of migration in a representative sample of texts. This procedure too enabled reflection on, and provision for, recommendations regarding the usefulness of the results of topic modelling for media discourse analysis.

The corpus

The texts were drawn from two Serbian dailies published in the period 2007–2017. The scope of the analysis was narrowed down to the so-called quality press, which however is an elusive term. The most accurate definition is based on its opposition to the tabloid press, the difference lying in more modest graphics, more complex language and the tendency to publish content which represents distanced, dematerialized general knowledge about the world of politics, social events and social movements (Pisarek 2008: 128–134). Since this kind of media is usually perceived as more authoritative, and its journalism as more balanced and non-tabloidized, an analysis of this particular media format can be justified.

The next criterion taken into account during the corpus compilation was that of representativeness. On the one hand, it was assumed that a wide variety of ideological slants should be represented, and, on the other, readership figures were also considered. Ultimately, two dailies were included in the study: the centre-right daily *Politika* and *Danas*, a newspaper whose circulation is lower, but which continues to be popular and which is the only newspaper leaning towards oppositional and liberal views.

During the subsequent stage, approximately 20,000 randomly chosen articles were excerpted from the national and international sections of the online editions of the analyzed dailies: each title was represented by 300 texts for each year in the period 2007–2011 and 700 texts for the period 2012–2017. Next, the corpus was lemmatized with the help of the ReLDI Tagger (Ljubešić et al. 2016), and the data obtained were fed into NoSketchEngine (Rychý 2007). This enabled parallel access to the lemmatized data and the original texts. Since the neglect of natural, unprocessed data was one of the objects of critique targeting the use of lemmatization in discourse analysis (Brookes and McEnery 2016: 6), the method adopted in the current study appears to be a good compromise in the analysis of highly inflectional Slavic languages, given that it enables parallel access to text words and lemmas during the whole procedure.

Finally, it should be mentioned that the time frame was adopted relatively arbitrarily: the end of the period under study coincided with the beginning of the research project undertaken by the author of the current paper, whereas the fact that the texts spanned a ten-year period enabled identification of the effect of a number of extralinguistic factors (European integration, multiple crises) on the analyzed discourse.

The method

Topic modelling is a relatively new method which is beginning to gain ground in corpus-assisted discourse analysis. In broad terms, it is defined as a “catchall term for a collection of methods and algorithms that uncover the hidden thematic structure in document collections by revealing recurring clusters of co-occurring words” (Törnberg and Törnberg 2016: 405). Based on an analysis of such co-occurrences, algorithms which model topics may be used as effective tools for empirical categorization (Törnberg and Törnberg 2016: 407). Similarly to collocational analysis, also in this case a number of algorithms exist; however, Latent Dirichlet Association (LDA) is the most common (Blei et al. 2003).

When LDA is applied in topic modelling, each text is regarded as a “bag of words”, i.e. a numerical representation of a text which ignores the grammatical properties and the order of words, and which allows identification of the distribution of words in individual documents composing the corpus. The topics identified in the analysis – modelling human intuitions about the subject matter of a given document – are in practice statistical distributions of words in collections of texts. The number of topics in a corpus is specified by the analyst: each topic contains a list of words with differently assigned probabilities, and each text is linked to a list of topics with different probabilities (Törnberg and Törnberg 2016: 405).

The application of topic modelling in corpus-assisted discourse studies has been subject to some criticism by Brookes and McEnery (2019). The first criticism levelled at this method was that it did not allow replicability, given that random distribution was used as the algorithm’s point of departure. However, replicability can be easily achieved by defining the random state before the algorithm is used.

Other criticisms stress the low degree of sensitivity to the linguistic context. Indeed, the findings obtained with the help of each tool in corpus-assisted discourse analysis must be contextualized. Such criticism can also be directed to collocational analysis. In the latter case, similarly, the solution lies in an analysis of concordance lines; in the case of topic modelling, in turn, the analyst must take into account the qualitative analysis of texts (or at least their titles) which are the most representative of a given topic. Thanks to such a procedure the need to lemmatize the data – seen as problematic by Brookes and McEnery (2019) – does not seem to impair to a significant extent the quality of the findings.

What is more, the results of topic modelling allow one to indicate texts which are most typical of the problem that is of interest to the analyst. This, in turn, makes it possible to move swiftly between the two key stages of corpus-assisted discourse analysis, as outlined by the Lancaster researchers:

3. Corpus analysis of frequencies, clusters, keywords, dispersion, etc. – identify potential sites of interest in the corpus along with possible discourses/topoi/strategies, relate to those existing in the literature.
4. Qualitative or CDA analysis of a smaller, representative set of data (e.g., concordances of certain lexical items or of a particular text or set of texts within the corpus) – identify discourses/topoi/strategies (DH approach). (Baker et al. 2008: 295)

While the first stage of the analysis (referred to in Point 3 above) involves a corpus analysis based on keywords assigned to topics and the thematic structure of the corpus, the next stage enables the researcher to use the results of the modelling to narrow down the dataset to be used in a qualitative analysis.

It is at the second stage that the subsequent contextualization of the material selected for analysis becomes especially relevant. In the Discourse-Historical Approach in CDA, four levels of context are often mentioned: 1. linguistic or textual (co-text and co-discourse), 2. intertextual and interdiscursive (relations between utterances, texts, genres, discourses), 3. the extralinguistic circumstances of the communicative situation (social and institutional), 4. the broad social and political as well as historical context (Reisigl and Wodak 2009: 93). During the qualitative analysis of a discourse sample more insights are obtained, related to practically all levels of context, and what gains particular significance are the intertextual and interdiscursive relations in the analyzed texts when viewed through the lens of a specific edition of a given medium.

New insights which may be obtained through the qualitative analysis of a sample may be characterized in relation to a narrow or broad media context, as suggested by Lisowska-Magdziarz (2006). The broad context includes the circumstances in which a given type of discourse appeared in a medium, its history of occurrence as well as interdiscursive relations. The narrow media context, in turn, subsumes such elements as the medium itself, the genre, the accompanying visual material, the position in the hierarchy of topics and discourses, the history of the discourse's occurrence in the medium, the immediate context of the material, the exposure of the material, and its effect on the importance of the discourse (Lisowska-Magdziarz 2006: 66–67). The broad media context may be successfully examined by way of

a discourse analysis relying on statistical generalizations obtained through corpus analysis. Interpretation which refers to the narrow media context – at least at this stage of the development of the analytical apparatus – requires a synergistic application of quantitative and qualitative approaches. The current paper is an attempt at obtaining such a synergy.

Findings

The topic model was generated with the use of Latent Dirichlet Allocation (LDA) and the number of resultant topics was set at 15.² The topics initially appeared in the form of sets with the keywords most typical of each set. The topic which is central to the discussion presented in the current paper was characterized by the following lexemes and phrases:

Crna Gora 'Montenegro', *granica* 'border', *izbeglica* 'refugee [n.]', *migrant* 'migrant', *Makedonija* 'Macedonia', *Mađarska* 'Hungary', *mađarski* 'Hungarian', *broj* 'number', *Albanija* 'Albania', *azil* 'asylum', *Bugarska* 'Bulgaria', *izbeglički* 'refugee [adj.]', *viza* 'visa'.

Based on the above list, it was possible to characterize the topic as being related to migration. Table 1 shows this topic against other topics in the corpus. As can be seen, in the period 2007–2017, the issue of migration did not take centre stage, although it was addressed frequently enough for it to be identified in a quantitative analysis. The topic accounted for 4.02% in the *Danas* subcorpus and 4.36% in the *Politika* subcorpus, with an average value of 4.19%.

Predictably, the distribution of the topic of migration was uneven in the period studied. Intuitively, one could suspect that the peak would be noted in 2015, a year which witnessed the culmination of the migration crisis in the Mediterranean Sea and the so-called Balkan route. The distribution of the topic is shown in Figure 1.

The local peaks, which are different for *Danas* and *Politika*, require a comment. In the first case, the peak falls on the year 2009, when the EU visa policy applicable to Serbian nationals was liberalized. In the case of the daily *Politika*, however, more attention was directed to the events in 2011, when some West European politicians began to notice the problem of Roma migration, and threatened to withdraw from the visa-free regime.

It should also be noted that a slightly lower distribution of the topic of migration in 2015 was observed in *Danas* as compared with *Politika*. It could be hypothesized that as a liberal medium, *Danas* was less likely to portray Europe as crisis-stricken. This hypothesis is corroborated also by a lower distribution of the topic described as “European crisis” in the generated model. However, to be able to establish this with greater certainty, additional evidence is needed, which will be verified in the subsequent discussion.

² The Gensim library was used, as well as the following parameters: $\alpha = 0.5$, $\beta = 0.01$, number of iterations = 50, random state = 2.

Topic	<i>Politika</i>	<i>Danas</i>	Total
Terrorism	6.21	1.63	3.92
Crisis in the Eurozone	8.07	0.51	4.29
Migration	4.36	4.02	4.19
Kosovo	9.66	12.17	10.92
Economy	9.04	4.27	6.66
History and politics	11.75	4.52	8.13
Geopolitics	12.26	5.71	8.99
Bosnia and Hercegowina	2.48	0.96	1.72
International Criminal Tribunal for former Yugoslavia	1.51	4.21	2.86
European policy	3.89	2.18	3.03
National policy	3.38	7.68	5.53
Elections	5.27	7.48	6.37
Regional diplomacy	6.87	7.45	7.16
European integration	9.48	27.17	18.32

Table 1: Distribution of topics generated through LDA modelling in a corpus of articles excerpted from the Serbian press related to European matters in the period 2007-2017 (in %)

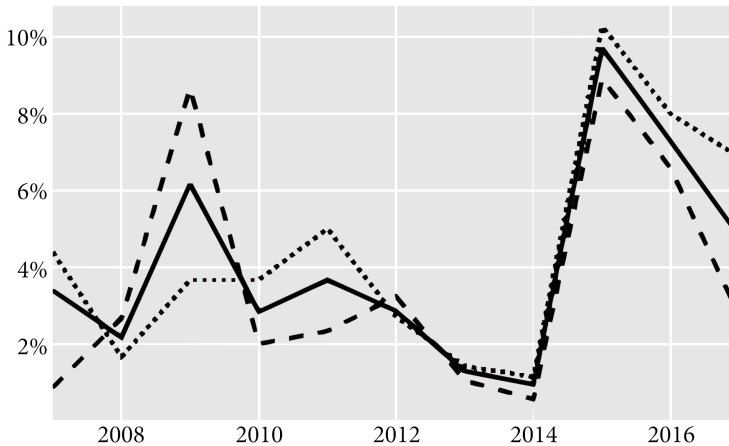


Figure 1: Chronological distribution of the articles whose dominant topic was that of migration in yearly subcorpora (solid line – average value, dotted line – *Danas*, dashed line – *Politika*)

More insights, more interpretational potential, and more contextualization potential can be gained from a qualitative discourse analysis of the sample. The generated topic model allows one to select representative texts for further analysis since the algorithm tags the articles with data showing the percentage share of a given topic. In the case of the topic of migration in *Politika*, the following proved to be three most representative texts:

1. *EK nema saznanja da se traži ukidanje bezviznog režima* 'EC has no knowledge about the intention to withdraw the visa-free regime', Beta press agency, 16 October 2012 (percentage share of the topic: 73.61%);
2. *Bugarska oklopna vozila na granici s Makedonijom* 'Bulgarian armoured vehicles at the border with Macedonia', Beta and AFP agencies, 25 August 2015 (69.44%);
3. *Oland: Šengen u opasnosti* 'Holland: Schengen in danger', Tanjug agency, 17 September 2015 (66.60%).

On the other hand, in *Danas*, the topic of migration was most visible in the following articles:

1. *Ministri o vizama* 'Ministers on visas', authored by T.N., 25 October 2012 (77.38%);
2. *Članice EU hoće zaštitu* 'EU members want protection', authored by R.D., 16 October 2012 (73.31%);
3. *Austrija diže ogradu na granici s Mađarskom?* 'Austria to build a fence on the border with Hungary?', authored by S.D., 20 September 2016 (72.02%).

Of course, it would be possible to enumerate more than three texts with a significant percentage share on the topic of migration for both dailies; however, unfortunately, there is no space for such an extensive analysis here. Meanwhile, based on the material gathered in this study, one may conclude that although – based on the automatically generated sets of keywords – the texts related to Western problems with the visa-free regime for Serbian nationals proved most typical, the two dailies devoted considerable space also to articles connected with the 2015 migration crisis.

The majority of texts are those produced by agencies or signed with initials. It should be underlined at this point that such brief news items prevail in the corpus that is based on strictly informational sections. If, however, the absence of texts signed with the author's full name is not insignificant, it could be a result of a lower volume of texts regarded as highly representative and, what follows, a greater relative saturation of keywords.

All the articles selected for a more detailed analysis were characterized by the presence of keywords assigned to the topic of migration in the generated topic model. Considering the recurrent collocations, they included the following lexemes and phrases: *azil* 'asylum', *(lažni) azilanti* '(fake) asylum seekers', *broj* 'number', *granica* 'border', *izbeglice* 'refugees', *migranti* 'migrants', *šengenska zona* 'Schengen area', *unutrašnji (poslovi)* 'internal affairs', *vizna liberalizacija* 'visa liberalization', *viza* 'visa'. On the one hand, the above are chiefly descriptions of various categories of migrants; on the other, they include names of institutions managing migration-related matters. What stands out in particular is the commonness of the topos of 'number' in relation

to migrants (*broj* ‘number’), as well as the strength of the collocation *lažni azilanti* ‘fake asylum seekers’, which was sometimes written with the adjective in inverted commas or preceded by the hedge *takozvani* ‘so-called’.

The daily *Danas*, apart from occasional keywords linked to the theme of the articles typically associated with a given topic (e.g. *Mađarska* ‘Hungary’), is characterized by lexemes and phrases related to economic migration and visa liberalization, such as *broj tražilaca azila* ‘the number of asylum seekers’ or *vizni režim* ‘visa policy’.

The daily *Politika* did not shy away from this topic either, using such descriptions as *broj lažnih azilanata* ‘the number of fake asylum seekers’ or *režim bez viza* ‘visa-free regime’. However, the difference between *Politika* and *Danas* was that the former provided much more detailed coverage of the 2015 refugee crisis with the use of such keywords as *izbeglički kamp* ‘refugee camp’, *kontrola* ‘control’, but also *preko* ‘by/through’. The last of the keywords points to the considerable presence of the linguistic modelling of space in texts related to migration and, indirectly, which is even more significant, to its transit nature within the Balkan states. This issue is contrasted with the status of Western European countries, seen as the ultimate destination of migrants and refugees.

What follows is an analysis based on texts published in the paper editions of the dailies. In principle, they could be expected either in the issues published one day later, more rarely a day earlier or on the same day (Bajić 2017). The analysis of the elements of the so-called narrow media context had to consider in particular the visual material present in the paper edition, the position in the hierarchy of topics and discourses – attested by the choice of section and the immediate context of the analyzed material, as well as the exposure of the material and its effect on the importance of the discourse – understood as the placement within a section (the hierarchy of sections in both dailies is shown in Table 2).

<i>Danas</i>	Naslovna strana → Politika → Društvo → Dijalog → Ekonomija → Globus → Beograd → Kultura → Scena → (Vojvodina Novi Sad → Hronika → Zdravlje →) Sport → Feljton/Zabava → Televizija → Periskop
	Title page → Politics → Society → Dialogue → Economy → Globe → Belgrade → Culture → Scene → (Vojvodina / Novi Sad → Chronicle → Health →) Sport → Column/Entertainment → Television → Periscope
<i>Politika</i>	Naslovna strana → Svet → Region → Događaji dana → Politika → Društvo → Hronika → Ekonomija → Kultura → Pogledi / Feljton → Srbija → Beograd → Spektar → Radio i televizija → Pogledi/Feljton
	Title page → World → Region → Current news → Politics → Society → Chronicle → Economy → Culture → Opinions/Column → Serbia → Bel- grade → Spectrum → Radio and television → Opinions/Column

Table 2: Hierarchy of sections in the dailies *Danas* and *Politika*

In the case of the texts which are the most representative of the topic of migration in the daily *Politika*, the article dated 17 September 2015, entitled: *Oland:*

Šengen u opasnosti 'Holland: Schengen in danger' could not be found in any of the paper editions. The text dated 16 October 2012, entitled: *EK nema saznanja da se traži ukidanje bezvизnog režima* 'EC has no knowledge about the intention to withdraw the visa-free regime' was the basis for the article authored by D. Vukosavljević and published in the paper edition of *Politika* on the same day, entitled: *Evropa traži hitne mere protiv lažnih azilanata* 'Europe demands immediate steps against fake asylum seekers'. The article was published in the section *Current news*, as the main text on the fifth page of the daily. On the other hand, the article dated 25 August 2015, entitled: *Bugarska oklopna vozila na granici s Makedonijom* 'Bulgarian armoured vehicles at the border with Macedonia' was published on the following day in the section *World/Region* as a brief mention in the bottom right corner.

All the texts which are the most representative of the topic of migration in the daily *Danas* were printed in the paper editions on the following day. The articles entitled: *Ministri o vizama* 'Ministers on visas', dated 25 October 2012, and *Članice EU hoće zaštitu* 'EU members want protection', dated 16 October 2012, were published, albeit briefly, in the section *Politics* as brief mentions in the right column on the second page of the daily. On the other hand, the text *Austrija diže ogradu na granici s Mađarskom?* 'Austria to build a fence on the border with Hungary?' dated 20 September 2016 was published in the section *Globe*, a section dedicated to more in-depth, structural analyses of the international situation, typically placed in the daily's centre spread, although it was in the bottom right corner, as the least important article on the page. The relation between the online and the paper editions of both dailies can be seen in Table 3.

Even a cursory glance at the gathered material reveals that the greatest saturation of keywords related to a given topic – in this case the topic of migration – can be identified in texts which are relatively short. This results, to some extent, from the specificity of the method applied in the analysis. Longer texts were less commonly dominated by one topic and in the case of the shortest, one- or two-sentence mentions, the determination of the topic may be problematic.

The placement of medium-length articles depends on the daily. The centre-right *Politika* seems to situate them somewhat farther from the title page, not only because of the arrangement of the sections. In the liberal *Danas*, on the other hand, the articles occupy a relatively privileged position in the hierarchy of texts, either on the page immediately following the title page, or in an important section situated in the daily's centre spread.

The application of a narrow media context results from the profiles of the dailies in question. In *Politika*, reports related to migration in the European context – i.e. chiefly in the context of EU institutions – did not play such an important role, competing with texts featuring news from other parts of the world. More attention was devoted to the escalation of the refugee crisis in the Mediterranean Sea and the Balkan route in 2015, most probably because of the geopolitical effect of these events, but also due to the possibility of creating and recreating the connotations of Europe and the EU as entities mired in dire crisis.

No.	% share of the topic	Date and author	Title	Comparison with the paper edition (section, publication date, page and other information)
<i>Politika</i>				
1	73.61%	16.10.2012 Beta	<i>EK nema saznanja da se traži ukidanje bezviznog režima</i> EC has no knowledge about the intention to withdraw the visa-free regime	Current news, 16.10, p. 5 Main article in the column authored by D. Vukosavljević, entitled <i>Evropa traži hitne mere protiv lažnih azilanata</i> 'Europe demands immediate steps against fake asylum seekers', based on a text authored by a news agency and included in the corpus
2	69.44%	25.08.2015	<i>Bugarska oklopna vozila na granici s Makedonijom</i> Bulgarian armoured vehicles at the border with Macedonia	World/Region, 26.08, p. 4 Short standalone note in the bottom right corner
3	66.60%	17.09.2015 Tanjug	<i>Oland: Šengen u opasnosti</i> Holland: Schengen in danger	Absent in the paper edition
<i>Danas</i>				
1	77.38%	25.10.2012 T.N.	<i>Ministri o vizama</i> Ministers on visas	Politics, 26.10, p. 2 Short article in the right column
2	73.31%	16.10.2012 R.D.	<i>Članice EU hoće zaštitu</i> EU members want protection	Politics, 17.10, p. 2 Short article in the right column entitled: <i>EU hoće zaštitu</i> 'EU wants protection'
3	72.02%	20.09.2016 S.D.	<i>Austrija diže ogradu na granici s Mađarskom?</i> Austria to build a fence on the border with Hungary?	Globe, 21.09, p. 16 Centre spread, in the bottom right corner – the least important article on the page

Table 3: The most representative articles on the topic of migration in the dailies *Politika* and *Danas* according to the LDA model and their relation to the paper editions

On the other hand, in *Danas*, as compared with *Politika*, there is much more coverage of the problems related to the economic migration from Serbia following the post-2009 EU visa liberalization. Medium-length articles featuring this topic were published regularly and placed on the second page, ranking high in the hierarchy of sections. This can be linked to the potential European integration of Serbia, much valued in the liberal Serbian discourse. Although to a lesser extent, *Danas* tended to construct the Europe-in-crisis connotation, which was the reason why reports related to the 2015 migration crisis were sporadically situated in the important, though less centrally placed, section called *Globe*.

Summary

The analysis based on topic modelling has shown that the topics of migration and Europe in the Serbian press in the period 2007–2017 had the highest frequency at two points. The first peak was noted in the period 2009–2011 in light of the economic migration from Serbia, especially that of the Roma, following the EU visa liberalization. The second peak coincided with the culmination of the refugee crisis in the Mediterranean Sea and the Balkan route in 2015.

It was also possible to recognize a number of discourse structures typical of the topic of migration, thanks to the tools applied in the analysis and word frequencies – mostly at the lexical level. These structures usually had the form of nomination strategies: a number of labels for migration actors implying varying degrees of precision (*migranti* ‘migrants’, *izbeglice* ‘refugees’, *azilanti* ‘asylum seekers’), but also lexis used to quantify them by way of the topos of ‘number’ (keyword: *broj* ‘number’).

At the same time, general observations could be made about the specific nature of discourse analysis based on topic modelling. The most typical results featured mainly medium-length articles, not necessarily those from the front pages of the dailies. Although this procedure does not allow the researcher to select for analysis more prominent texts, it directs his/her attention to what often remains unnoticed because it is unmarked, despite representing a large share of press material. This agrees with the basic tenets of Critical Discourse Analysis and the reasons for using corpus tools for the purpose of the current study.

The differences between the discourse structures dependant on the ideological slant of a given medium proved to be relatively modest. The liberal press also did not shy away from the topos of ‘number’ and other genericization devices used to portray migrants. The differences were visible above all in the amount of coverage of individual crises: for the liberal *Danas*, of greater relevance was the issue of the economic migration following the EU visa liberalization which threatened the European integration of Serbia. The centre-right *Politika*, on the other hand, dedicated considerable space to the coverage of the 2015 migration crisis, given its geopolitical significance, which allowed the possibility of constructing the connotation of Europe as being mired in crisis.

References

- Bajić P. 2017. Onlajn štampa: odnosi između sadržaja u štampanim i onlajn izdanjima dnevnih novina. – *CM: Communication and Media* XII.39: 57–82.
- Baker P., Gabrielatos C., KhosraviNik M., Krzyżanowski M., McEnery T., 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: 273–306.
- Blei D.M., Ng A.Y., Jordan M.I. 2003. Latent Dirichlet allocation. – *Journal of Machine Learning Research* 3: 993–1022.
- Brookes G., McEnery T. 2019. The utility of topic modelling for discourse studies: A critical evaluation. – *Discourse Studies* 21.1: 3–21.
- Car V., Čančar E., Bovan K. 2019. The 2015 and 2016 migration crisis in Europe: How Croatian daily newspapers represented and portrayed refugees and migrants. – *Teorija in praksa* 56.2: 681–699.
- Greenberg J., Spasić I. 2017. Beyond East and West: Solidarity politics and the absent/present state in the Balkans. – *Slavic Review* 76.2: 315–326.
- Lisowska-Magdziarz M. 2006. *Analiza tekstu w dyskursie medialnym*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
- Ljubešić N., Klubička F., Agić Ž., Jazbec I.-P. 2016. New inflectional lexicons and training corpora for improved morphosyntactic annotation of Croatian and Serbian. – Calzolari N. et al. (eds.). *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Paris: ELRA: 23–28.
- Pisarek W. 2008. *Wstęp do nauki o komunikowaniu*. Warszawa: Wydawnictwo Akademickie i Profesjonalne.
- Reisigl M., Wodak R. 2009. The discourse-historical approach (DHA). – Wodak R., Meyer M. (eds.). *Methods for Critical Discourse Analysis*. London: Sage: 87–121.
- Rychlý P. 2007. Manatee/Bonito – A modular corpus manager. – Sojka P., Horák A. (eds.). *First workshop on recent advances in Slavonic Natural Language Processing*. Brno: Masaryk University: 65–70.
- Šarić Lj., Felberg T.R. 2019. Representations of the 2015/2016 ‘migrant crisis’ on the online portals of Croatian and Serbian public broadcasters. – Viola L., Musolf A. (eds.). *Migration and media: Discourses about identities in crisis*. Amsterdam: John Benjamins: 203–238.
- Törnberg A., Törnberg P. 2016. Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. – *Discourse & Society* 27.4: 401–422.