Corpus analysis in the study of the discursive characteristics of the media image of immigrants and asylum seekers (A case study of British and American media 2018-2020)

Marina S. Matytcina – Olga N. Prokhorova – Igor V. Chekulai – Elena V. Pupynina

Abstract
The aim of this paper is to study the discursive characteristics of the media image of immigrants and asylum seekers in British and American media using the corpus analysis method. The study focuses on the published content of the News on the Web corpus dealing with immigrants and asylum seekers from the beginning of 2018 to the end of 2020. Another objective is to identify and quantify, by analytical approach, the L1 collocates and stable collocations of the studied words. The results show the media use different vocabularies in connection with each of the two groups of interest for their discursive construction.
Key words: corpus analysis, immigrants, asylum seekers, collocation, L1 collocate, text corpus

Introduction

At present, in Russian and foreign sciences, the number of studies of specific and repetitive discursive practices using the corpus analysis method is increasing. In the process of penetration of modern information and communication technologies into all spheres of science, and in particular, in philological science, the popularity of the use of linguistic corpora of texts in the study of various aspects of the language is increasing. Representativeness is one of the most important characteristics of a text corpus. Since the corpus is a kind of model of language, its representativeness determines the reliability of the data obtained on its basis. Therefore the question can be considered as "the problem of adequate reflection, adaptation or integration of large arrays of texts or some other fragments of speech activity into a significantly smaller corpus of texts" (Iordanskaja, 2007). In corpus linguistics, representativeness is understood as a balanced and proportional presentation of texts in the corpus.

Corpus linguistics reveals parallels with discourse linguistics. While discourse linguistics focuses on the integrativity, the intertextual coherence of texts, corpus linguistics focuses on the representativeness of text samples. The discourse-oriented analysis is a qualitatively oriented analysis, deep- and macro-semantic, revealing the deep bonds of statements, not necessarily lying on the surface. The corpus-oriented analysis is quantitative, creating a research focus on what, in linguistic use, provides meaningful unity. The potentially possible linguistic phenomena may exist but may not be reflected in this corpus, while the fact of the absence of a potentially possible phenomenon in the corpus is significant for the researcher. Corpus linguistics reveals parallels with discourse linguistics. The key emphasis on conducting a discourse-oriented analysis is presented in the studies of N. Fairclough, R. Wodak, T. van Dijk, E. Laclau, T. van Leuven (Dijk, 2015; Fairclough et al., 2011; Laclau, 1985; Leeuwen, 2008; Wodak, Meyer, 2009 et al.). The discourse analysis is carried out as a qualitatively oriented macrosemantic analysis that reveals deep connections between discourse and social practice: society and culture are created as discourse and at the same time generate and support a certain discourse. The use of language is socially constructible and, at the same time, socially constructive. "Language use as social practice… is socially shaped, but it is also socially shaping or constitutive. Language use is moreover constitutive in both conventional, socially reproductive ways, and

creative, socially transformative ways" (Fairclough, 1993: 134). The researchers in the field of corpus linguistics believe that corpus analysis is focused on the study of the role of language units in the construction of social reality based on the maximum correspondence of the sample characteristics to the characteristics of discourse and, thus, gives hope for objectivity (Plungyan, 2008; Chernyavskaya, 2017; Baker, 2006, Bubenhofer, 2008; Sinclair 2004 et al.). According to V.E. Chernyavskaya, "the use of the corpus is productive in connection with the problem of proving the existence of a certain phenomenon in the language, while the concepts of "being in the corpus" and "being in the language" are not identified" (Chernyavskaya, 2017: 144). The corpus-oriented analysis is quantitative, creating a research focus on what, in linguistic use, provides meaningful unity.

The novelty introduced by discourse analysis should be associated with methods of studying the summary of individual texts that already exist and are potentially possible. The advantage of discourse as a unit of operational analysis is that it allows going from the text as a relatively complete, formally limited structure to other texts. The content of the discourse is revealed not in one separate text but in the complex interaction of many texts. The works of V.I. Karasik, L.V. Balashova, N. Fairclough, and Yu. Kristeva (Karasik, 2018; Balashova, 2014; Fairclough, 2003; Kristeva, 1993 et al.) deals with the study of the dialectical connection of discourse with other discursive practices, categories, and contexts. The research focus is on understanding the text as part of the communication process as a result of its multidimensional connection with the systems of signs of other texts and external conditions relevant to the context. In turn, this presupposes a qualitatively new approach to the text – not as a relatively complete structure, but as a structure in intertextual and interdiscursive unity. In this view, the instrumental nature of discourse analysis becomes apparent. This is a descriptive practice, a kind of "anti-ontological project": the researcher studies not how it should be according to the rules, but how diversely it can be in socio-speech practice.

The combination of methodological tools of discourse analysis and corpus linguistics is considered as an opportunity to achieve the integrity of the linguistic analysis. The combination of general scientific methods of theoretical and empirical knowledge contributes to the identification of implicitly expressed attitudes and hidden effects of the influence of discourse on the system of social interactions.

Literature review
Much attention is paid to the discursive studies that allow deconstructing the schemes of building communication links between political actors, both in domestic and foreign linguistics. A special place among the works that deal with the issues of studying the interaction of language and society belongs to the research of the facts of the manifestation of power and control in discursive practice and beyond through the use of corpus analysis. The works of such linguists as V.A. Plungyan, V.E. Chernyavskaya, T. McEnery, E. Wilson, P. Baker, and other researchers of corpus linguistics deal with the study of the influence of discursive practices on the formation of social and political reality, as well as the discourse as a form of social interaction by combining general scientific methods of theoretical and empirical knowledge, critical discourse analysis and the method of corpus analysis for the study of specific and repetitive discursive practices (Plungyan 2008; Chernyavskaya 2017; McEnery, Hardie 2011; McEnery, Wilson 2001; Baker et al., 2013). The attention of the researchers in the field of corpus analysis is aimed at identifying, comparing, and interpreting collocations, keywords, concordances, and their usefulness through sample analysis of corpus (Evert, 2008; Granger, Bestgen, 2014; Handl, 2008), the study of technological processes related to the design, selection and processing of language material based on text corpora (Zubov, Zubova, 2004; Finegan, 2004; Gries,

2015), the analysis of corpus linguistics' current contributions to research in linguistics and the digital humanities, including comparing corpus techniques against other digital methods that are currently more popular amongst humanities researchers (Brown, 2020), the critical analysis of the discursive reproduction of racial stereotypes, biases and prejudice against people of other nationalities (Baker et al, 2013; Baker et.al, 2008; Mautner, 2009).

Summarizing the research on this issue, we note that the interests of linguists are mainly concentrated around the collocation possibilities of words, concordance in order to determine the stylistic, semantic, and cultural shades of their meanings, which are caused by the specifics and ideological attitudes of society. The researchers who make their research on the basis of combining critical discourse analysis and the method of corpus analysis emphasize that the study of discursive practices should focus not on a simple description of discursive structures from the point of view of their functioning in social processes but on proving the existence of a certain phenomenon in language, taking into account the accurate statistical data (Baker et.al., 2008; Brookes, McEnery, 2020). Thus, the main attention is paid to the analysis of the possibilities and advantages of using the corpus to obtain objective information about the considered language phenomenon. The corpus analysis allows us to implement the thesis put forward by cognitive linguistics about the need to study the models of language fixed in the practice of real use.

Despite the relevance of the research topic, there are still theoretical and methodological gaps in understanding the possibilities of using corpus analysis when considering the discursive characteristics of the media image of immigrants and asylum seekers in British and American media, linguistic mechanisms for the implementation of the discourse structures of and its role in the construction of social reality. In our opinion, these gaps cannot be remained free, since the study of the social functioning of language and the social conditions of its use in the formation of the media discourse of immigration is of increased interest not only among researchers but also among governments and ordinary citizens.

Research methods

The corpus analysis approach was not chosen by chance as a methodological basis for researching the discursive characteristics of the media image of immigrants and asylum seekers. Firstly, the corpus analysis is an effective method for establishing accurate statistical data, since the use of quantitative methods for solving linguistic problems allows collecting and processing more data to study the qualitative composition of the vocabulary, to obtain objective information about the frequency of language units, the degree of connectivity of parts of a phrase and the semantic prosody. The discourse research using quantitative practices is a reliable, controlled, result-oriented generalized way of analyzing a range of linguistic phenomena. Secondly, the use of state-of-the-art information technologies eliminates the possibility of human factor interference in the research when data is analyzed selectively to confirm certain hypotheses, which casts doubt on the reliability of previously obtained data and the integrity of the linguistic analysis carried out by a researcher (Baker, 2006). Thirdly, the construction of "matrices reflecting the occurrence of standard and non-standard vocabulary and structures" facilitates the search for the motivational base of those positions, attitudes, goals, and interests of the media elite that define discourse as a power resource (Ibid.).

The use of the corpus analysis method to conduct this research allows supplementing the semantic analysis and the pragmatic features of utterances, as well as identifying the most recurrent linguistic mechanisms for the implementation of discourse structures and discourse strategies involved in the representation of immigrants and asylum seekers (Gabrielatos, Baker, 2006, 2008; Baker et al., 2008; KhosraviNik, 2009). In this paper, the subject of consideration is the peculiarities of constructing the

image of immigrants and asylum seekers in media discourse, since, despite the widespread occurrence of such studies, some issues in the study of the specifics of the formation of such a discourse in the media are considered, in our opinion, one-sidedly.

This study was based on news articles from British and American broadsheets, mid-markets, tabloids, and news agencies of the *News on the Web* corpus from 2018 to 2020. The selected information sources have been divided into separate categories suggested by the *Audit Bureau of Circulations* (ABC) and are presented in Table 1.

Table 1: British and American broadsheets, mid-markets, tabloids, and news agencies, included in the study

Broadsheets	Mid-market	Tabloids	News agencies
The Times The Sunday Times	The Daily Mail	The Daily Mirror	CNN
The Daily Telegraph	The Sunday Express The Daily Express	The Sun	BBC News
The Guardian		The New York Post	FOX News
The New York Times			ABC News
The Wall Street Journal			Reuters
The Washington Post			
The Washington Times			

The NOW corpus (*News on the Web)* grows by about 180-200 million words of data each month (from about 300,000 new articles), or about two billion words each year. In our case, the analysis of corpus data allowed us to select and analyze a large number of news texts using target words: *IMMIGRANTS* and *ASYLUM SEEKERS*. The amount of information obtained from the selected sources is, in our opinion, sufficient for a comprehensive understanding and explanation of the discursive characteristics of the image of immigrants and asylum seekers in newspapers and news agencies. Predictably, the number of articles on immigration published in the media between 2018 and 2020 without restrictions on type, size, and ideological position was impressive.

The use of broad selection criteria made it possible to avoid the problem of "skimming the cream", that is, a selective approach in the choice of examples for the analysis (Santa Ana et al., 2007). A large, representative sample was vital to get an objective picture of the rhetoric of immigrants used in publications. The electronic corpus made it possible to identify and quantify frequently to recur stable phrases in which the semantically main component is the target words. Besides, in comparison with traditional linguistics, which is descriptive and introspective in nature, the quantitative data obtained using the corpus is not the subjective interpretation of the situational model by the communication agents.

Findings and discussion

To identify the discursive characteristics of the media image of immigrants and asylum seekers, we set several tasks. Firstly, to determine the collocations in which the lexical units *IMMIGRANTS* and *ASYLUM SEEKERS* would be the semantically main component and to rank the identified collocations according to the degree of stability. Secondly, to find the most common L1 collocations of the target words *IMMIGRANTS* and *ASYLUM SEEKERS*. Thirdly, to identify stable collocations of the target words *IMMIGRANTS* and *ASYLUM SEEKERS.*

In modern linguistics, there are different definitions of the concept "collocation". Most definitions in modern linguistics, there are different definitions of the concept of "collocation" are based on the phenomenon of the co-occurrence of two or more words that appear together in the text (Sinclair, 1991; Stubbs, 1995). The term "collocation" was first used by W. Fowler in his fundamental work "English Grammar" (1857) when studying the internal structure of a group of words. In the Russian-language scientific literature, it appeared in the dictionary of linguistic terms by O.S. Akhmanova, who defines collocations as "lexico-phraseological conventional combination of words in speech for the implementation of their polysemy" (Ahmanova, 1966). E.G. Borisova recognizes a collocation as a non-free non-idiomatic phrase, in which the free component does not change its meaning, and the meaning of the non-free component differs from that which is realized in free word combinations (Borisova, 1995). The study of collocation patterns in a large corpus of texts is "a way of understanding meanings and associations between words which are otherwise difficult to ascertain from a small-scale analysis of a single text" (Baker, 2006: 96). At present, the term collocation is widely used in corpus linguistics, where the concept of collocation is substantially simplified in comparison with traditional linguistics. Researchers designate this approach as statistical (Sysoev, 2010). Collocations in corpus linguistics are defined as statistically stable word combinations. Moreover, such a statistically stable combination can be both phraseological and free. It should be noted that collocation is not identical to phraseological units since the latter is traditionally understood as idioms. Collocations are usually formed in the process of changing the meanings of words as a result of re-thinking of the individual components of the combination. The main feature of collocation is the retention of the lexical meaning by one of its components with the stability of the phrase as a whole and the absence or diminution of the meaning of figurality and emotional-value load. The importance of studying collections is beyond doubt since they are present in all languages and represent the main feature of the language. "Collocation clarifies and concretizes the meaning of the words that form it. The exact meaning in any context is determined by the environment of a given word – its distributor, or collocate that are united around it and form a collocation. Collocation is fundamental to all linguistic use" (Rjazanova, 2012).

Statistical studies show not only the frequency of occurrence of words in the corpus but also how the lexical meanings of words change when the word combination is stable. This allows researchers to analyze the information extracted from the corpus from both the qualitative and quantitative perspectives to assess the strength of the relationship between words (Hunston, 2007; McEnery, Hardie, 2011).

The extensive linguistic material of the corpus shows that if the semantically main component in a phrase is the lexical unit *IMMIGRANTS*, then the auxiliary component of the phrase or the collocate is the word that is most often combined with the main component of the phrase in this corpus. In this case, the relevant units are analyzed that are at a given distance from the studied word. In the present study, these are words spaced at a distance of no more than five words (a five-word window) either in the preposition or in the postposition to the main component of the phrase without taking into account the function words. The search for collocations with so many words seems to be the most optimal since a larger number can lead to language

distortions (Baker et al., 2013). Typical repetitive combinations of the word *IMMIGRANTS*, extracted from the corpus, provide additional insight into the actual use of these patterns in the language. The collocation results of the word *IMMIGRANTS* in the studied corpus are presented in Table 2. According to this table, the phrase *ILLEGAL IMMIGRANTS* occurs in almost 27% of articles.

Table 2: Quantitative characteristics of collocates of word *IMMIGRANTS*, 2018-2020

Collocate	Quantity	Percentage share
illegal	14,777	27.01%
undocumented	9,012	16.47%
new	2,340	4.28%
young	1,739	3.18%
Mexican	1,366	2.50%
Chinese	1,353	2.47%
Indian	1,050	1.92%
unauthorized	1,043	1.91%
African	998	1,82%
other	986	1.80%
TOTAL	54,714	

The second stage of the analysis consisted of finding L1 collocates of the word under study. The words appearing in the range shown are collocates of the target (investigated) word, which in Figure 1 is denoted by the word TARGET. Figure 1 shows the arrangement of collocates on both sides of the target word—the lexical unit standing immediately before the target word is defined as L1 collocate. The analysis of the corpus data shows that L1 collocate is a defining word expressed by different parts of speech and explaining the features of the target word. L1 collocates represent speech patterns and indicate the consistency of using words to describe immigrants and asylum seekers, while L2, L3 collocates, etc. – these are words that collocate with the target word but are less common in such a pair.

| L5 | L4 | L3 | L2 | L1 | TARGET | R1 | R2 | R3 | R4 | R5 |

Figure 1: Collocate positions in relation to a target word

The analysis of the corpus indicates that the most frequent L1 collocate of the noun *IMMIGRANTS* is the adjective *ILLEGAL*, as shown in Figure 2.

Figure 2: Top L1 Collocates of *IMMIGRANTS*

The research has shown that broadsheets and news agencies use the evaluative adjective *ILLEGAL* as the L1 collocate of the noun *IMMIGRANTS* much more often than mid-markets and tabloids. According to Figure 3, 4.5% of broadsheets articles and 3.5% of news agency reports prefer the phrase *ILLEGAL IMMIGRANTS*, while broadsheets and tabloids define immigrants as illegal at 1.3% and 0.7%, respectively. In other words, all types of publications emphasize the violation of laws by immigrants and construct this circle of persons as a threat to the European population.
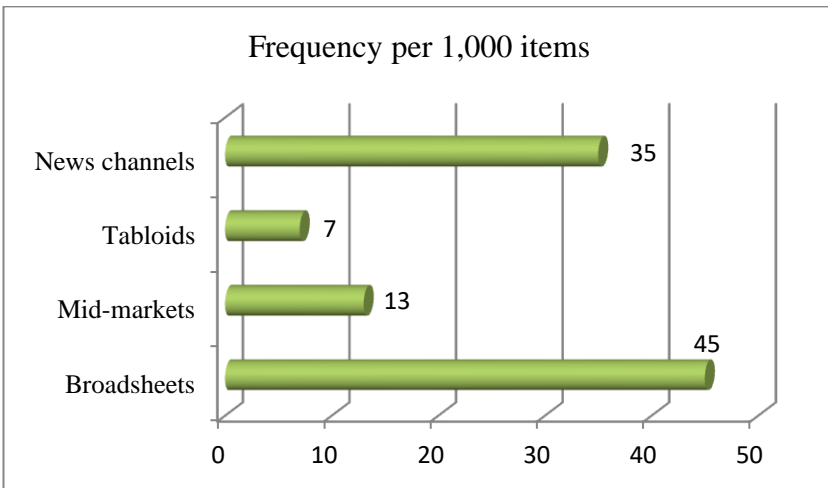
Figure 3: *ILLEGAL* as TOP L1 collocate of *IMMIGRANTS* by publication type

The allocation of the word *ILLEGAL* as L1 collocate has become the key definition in relation to criminal actions on the side of immigrants and serves as a vivid example of the use by the power structures of the media to reproduce stereotypes concerning this social group in the mind of an individual. The individual mental representations of discourse formed as a result of such an impact correlate the structures of discourse

with communicative situations, reflecting how a participant in a communicative event interprets the action taking place, and at the same time, influencing existing social relations. By describing immigrants as illegal, the English-language press and news agencies create a special concept of immigration that, firstly, highlights the problem of legal status and, secondly, portrays immigrants as lawbreakers. It is clear that the media is modifying existing social relationships in the interests of policymakers, creating and strengthening public opinion about what immigration is and who immigrants are. The public is convinced that there is a serious problem with ensuring security, and the idea of establishing need strict control is being held. Thus, both for the public and the media, illegality is equated with the concept of immigration. Attention is drawn to the fact that publications establish hypo-hyperonymic relationships between the concepts of *IMMIGRANTS* and *ILLEGAL*. That is, immigrants are classified as lawbreakers, and their migration is a crime and not a forced measure. The presentation of information through a recurrent sequence of lexical units *ILLEGAL IMMIGRANTS* in the English-language media is a mechanism for generating the necessary implication in relation to immigrants among the indigenous population.

Here are some examples from the corpus, illustrating the formation and maintenance of public opinion by the media regarding *ILLEGAL IMMIGRANTS* as a problem:

A total of 27 **illegal** immigrants and one boat skipper were detained during Ops Benteng, while eight land vehicles were seized (Petaling, 2020).

Macgregor has referred to immigrants to Europe as "Muslim invaders", advocated shooting **illegal** immigrants on the US border, and promoted a range of white nationalists conspiracy theories (Borger, 2020).

Another L1 collocate that forms a highly stable combination with the *IMMIGRANTS* lexical unit, according to this study, is the adjective *UNDOCUMENTED*. The word *UNDOCUMENTED*, sometimes used as an alternative to the adjective *ILLEGAL*, and in most cases, used in the context of discussing immigration issues in the US, not the UK. The priority topic of the publications is the coverage of the debate on immigration reform carried out by the US government in relation to foreigners who are in the country without the necessary documents:

Padilla, who graduated from the Massachusetts Institute of Technology with an engineering degree, says he became politically engaged in 1994 – as many Latinos did – when Californians passed Proposition 187, the initiative aimed at cutting public spending on **undocumented** immigrants, including public schooling and nursing home care (Morain, 2020).

As the mayor of reliably Democratic Miami-Dade County in Florida, Carlos A. Gimenez, a Republican, six days after the inauguration in 2017, became the first big-city leader in the country to reverse the county's de facto status as a "sanctuary" for **undocumented** immigrants (Shear, 2020).

The next stage of the analysis was the identification of stable collocations or "c-collocates" – the most frequently encountered words located at a distance of five words in the preposition and in the postposition without taking into account the function words to the word *IMMIGRANTS* in the studied three-year period. In accordance with Figure 4, *ILLEGAL* maintains a leading position in all types of newspapers and news agencies surveyed.
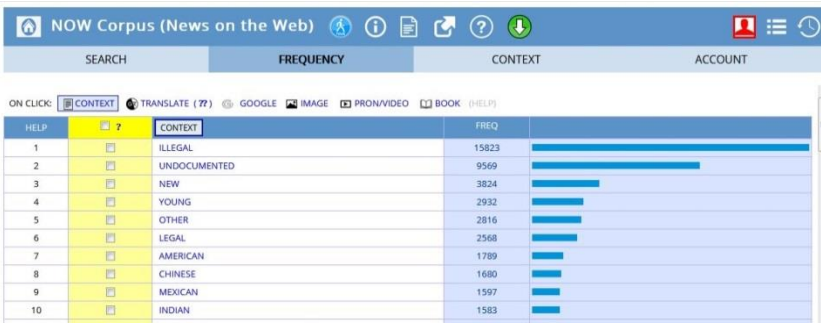
Figure 4: Top C-collocates of *IMMIGRANTS*

The quantitative characteristics of the stable collocation of IMMIGRANTS lexical unit with the ILLEGAL collocation in different types of the studied newspapers and news agencies is presented in Figure 5, according to which the broadsheets columnists in the studied period characterized immigrants using the adjective *ILLEGAL* in 7.2% of cases. For news agencies and mid-markets, this figure is 4.3% and 1.1%, respectively, and the tabloids used this phrase only in 0.5% of cases.
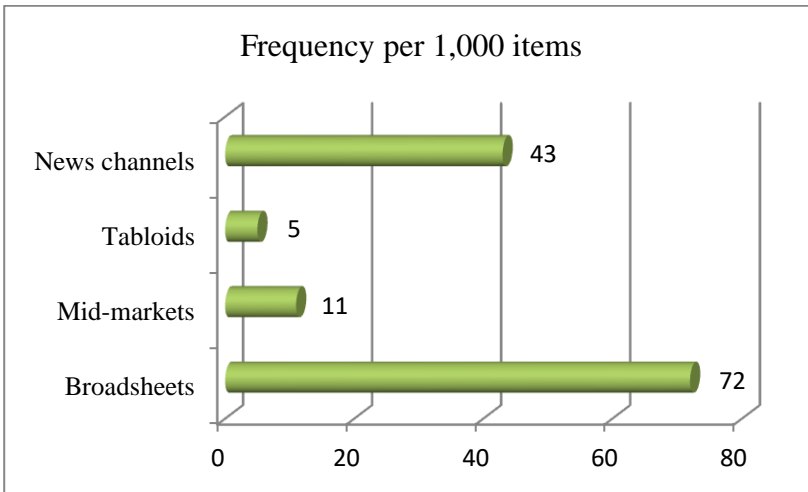
Figure 5: *ILLEGAL* as Top C-Collocate of *IMMIGRANTS* by publication type

The information obtained creates a holistic picture of the current stage of political communication, helps to better understand the discursive mechanisms of forming attitudes towards immigrants used by various publications (Gabrielatos, Baker, 2008). All stages of collocation analysis give a clear idea of those speech models that are used by the English-speaking mass media to describe immigrants.

The newspapers reflect the contextual models shared by the power structures in the communicative spaces of Great Britain and the United States, not only in relation to "illegal" immigrants but also in relation to asylum seekers, i.e. those who have already left their country, have entered another country and are applying for asylum with the authorities (Blinder, 2015). The results of the co-occurrence of the word

ASYLUM SEEKER in the studied corpus are presented in Table 3. According to this table, the phrase *AMERICAN ASYLUM SEEKERS* occurs in almost 8% of articles.
.

**Table 3: Quantitative characteristic of collocates of word *ASYLUM SEEKER*, 2018-2020**

Collocate	Quantity	Percentage share
American	252	7.95%
failed	192	6.06%
African	190	6.00%
other	129	4.07%
Honduran	103	3.25%
rejected	95	3.00%
new	81	2.56%
would-be	73	2.30%
vulnerable	71	2.24%
Iranian	70	2.21%
TOTAL	3,168	

In Bad News for Refugees paper, the researchers of the Glasgow Media Group raise the issue of addressing media coverage of hostility and aggression towards asylum seekers (Philo et al., 2013). Analyzing news reports, the authors explore ways of creating and maintaining stereotypes about immigrants in the mind of an individual, creating a discourse in which economic migrants, refugees, and asylum seekers are united into one "group" that becomes the subject of political divisions. They explain the negative reaction of indigenous people to new arrivals by describing cases of fraud in receiving benefits, theft at work, crime, which form the basis of the corpus of news texts in relation to immigrants. In the course of the study, scientists identify various semantic, pragmatic, and socio-political strategies for the formation of anti-immigrant ideas, views, and attitudes in the dominant discourse of power groups, using, inter alia, such formulations as "bogus asylum seekers" and "illegal asylum seekers". Such phrases are meaningless from a political point of view since the search for asylum in itself cannot be illegal. However, asylum seekers whose applications and appeals have been rejected may remain in the country illegally. Hence, it is not surprising that *FAILED* acts as one of the most frequent L1 collocates of the *ASYLUM SEEKER* phrase. Figure 6 shows L1 collocates of the lexical unit *ASYLUM SEEKER* in the studied period.

Figure 6: Top L1 collocates of *ASYLUM SEEKERS*

As shown in Figure 7 below, the use of *AMERICAN* as L1 collocate of the *ASYLUM SEEKER* lexical unit proves that the issue of immigration for the United States is especially sensitive. 24.5% of news articles from news agencies and 14% of broadsheet articles used the phrase *AMERICAN ASYLUM SEEKERS*, while mid-market publications mentioned this combination only in 4%, and the tabloids did not use it at all.
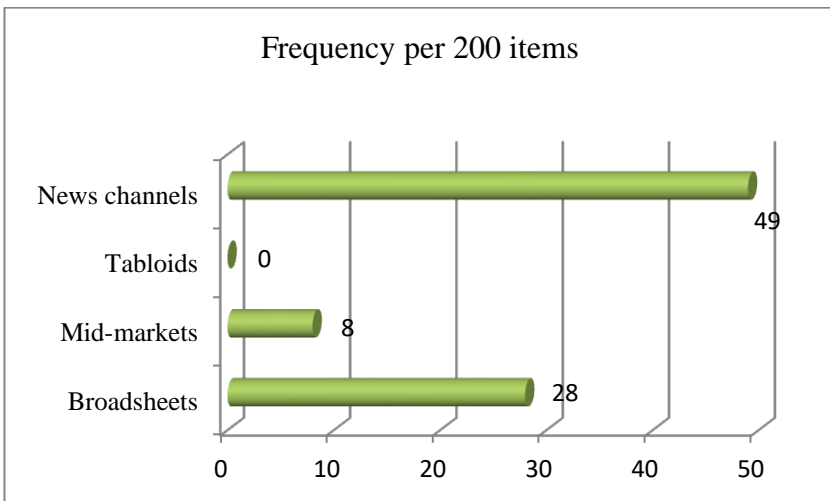
Figure 7: *AMERICAN* as Top L1 collocate of *ASYLUM SEEKERS* by publication type

Here are some examples from the corpus illustrating that asylum seekers are becoming an increasingly serious problem for the United States, which requires tightening border controls and deportation of illegal immigrants:

Central American asylum seekers detained in McAllen, Tex. on June 12 were sent to a processing center for possible separation (Leonhardt, 2018).

Central American asylum seekers wait for transport while being detained by US border patrol agents near the US-Mexico border in McAllen (Powell, 2018).

The relocation seeks to disband the group of some 1,400 Central American asylum seekers and prevent potential mass attempts to rush for the border fence to cross to the U.S., as happened in Tijuana late last year (Montes, 2019).

With regard to *c-collocates*, as shown in Figure 8, the most frequently encountered words that stand at a distance of five words in the pre-position and the postposition

without taking into account the service words of *ASYLUM SEEKERS* in the three-year period under study are the words *OTHER, CENTRAL, NEW, AMERICAN*.
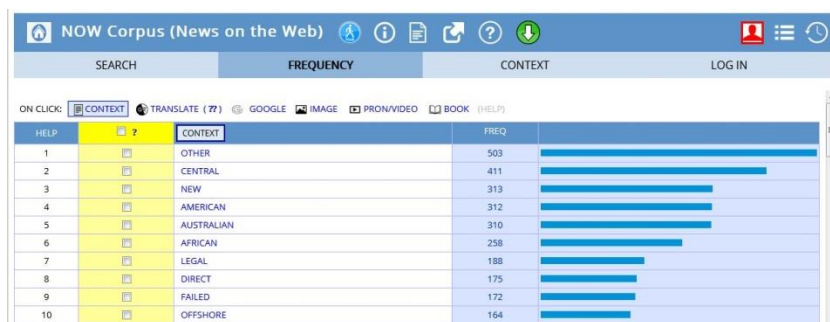
Figure 8: Top C-collocates of *ASYLUM SEEKERS*

As shown in Figure 9, as an alternative to the lexical units *AMERICAN* and *FAILED*, news agencies described asylum seekers using the word *OTHER* in 22.8% of cases, broadsheets – in 14.6%, mid-markets – in 3.8%, and tabloids in only 1% of articles.
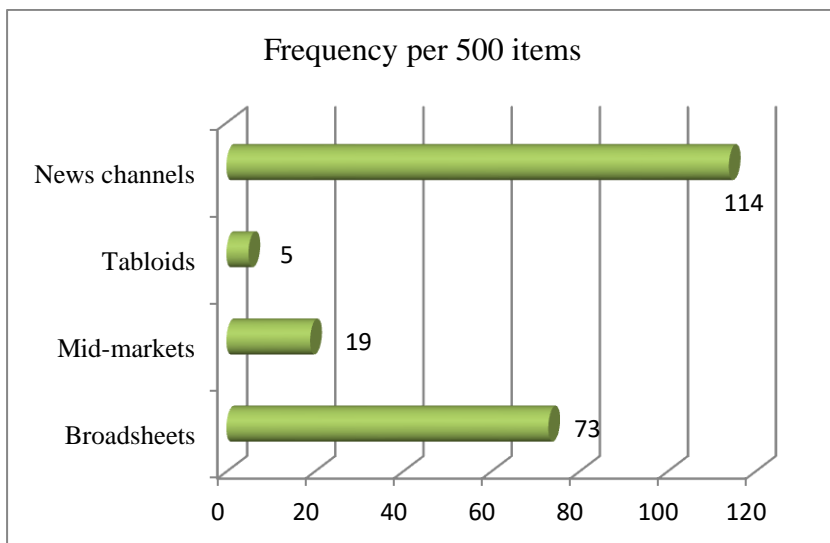
Figure 9: *OTHER* as Top C-collocate of *ASYLUM SEEKERS* by publication type

Conclusion
Analysis of the data obtained as a result of our corpus study of the target words *IMMIGRANTS* and *ASYLUM SEEKERS* reveals the following patterns in the use of those lexical units that are involved in the discursive construction of the media image of immigrants and asylum seekers. The analysis of the corpus data shows that the most commonly used combinations are the word *IMMIGRANTS* with the adjectives *ILLEGAL* and *UNDOCUMENTED*, and the same adjectives are also the L1 collocates of the *IMMIGRANTS* lexical unit, which indicates that *ILLEGAL* and *UNDOCUMENTED* are becoming key definitions in relation to criminal actions on

the part of immigrants that require government intervention. The adjective *ILLEGAL* retains leadership position in use with the word *IMMIGRANTS* and forms a stable collocation with this lexical unit in the studied three-year period.

The use of *AMERICAN* as a c-collocate and L1 collocate of the *ASYLUM SEEKER* lexical unit proves the fact that the problem of immigration is not considered by the English-speaking press as an exclusively national problem of a particular country of the European Union, but is a part of the general socio-cultural and linguistic context of the entire world community. In particular, the issue of immigration for the United States is especially sensitive. Thus, *ILLEGAL* retained the leading position in the formation of a stable combination with the target word *IMMIGRANTS* in all types of newspapers and news agencies under study, while news agencies, as an alternative to the lexical units *AMERICAN* and *FAILED*, described asylum seekers using the word *OTHER*. The collocations of the lexical units under study revealed as a result of the analysis indicate that immigrants and asylum seekers are portrayed as a threat to European security, which causes a feeling of fear among the reading audience. Based on this representation of the discourse, a conclusion is drawn in the following political context: to recognize the gravity of the situation and to take urgent measures regarding the regulation of immigration flows and the procedure for granting asylum.

The study of the discourse characteristics of the media image of immigrants and asylum seekers on the basis of corpus analysis contributes to the creation of a comprehensive statistical database on the linguistic content of the discourse about immigrants in newspaper publications and reports of news agencies. Thus, as a result of this study, a "corpus" of articles on immigration was created for three years, with a total volume of about 20 thousand articles. The use of the corpus analysis method made it possible to analyze such a large number of texts and identify patterns in the use of the studied words and the relationships between them. In other words: to shed light on the press coverage of immigration by identifying patterns in the words that appear consistently in association with immigrants and asylum seekers.

Bibliographic references

AHMANOVA, O.S. 1966. Slovar' lingvisticheskih terminov. Moscow: Sovetskaja Jenciklopedija, 606 p.

BORISOVA, E.G. 1995. Chto takoe kollokacii i kak ih izuchat'. Moscow: Philologilz, 49 p.

BAKER, P. 2006. Using Corpora in Discourse Analysis. London: Bloomsbury Academic, 198 p.

BAKER, P. – KHOSRAVINIK, M. – KRZYZANOWSKI, 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. In: Discourse & Society, vol. 19(3), pp. 273-306. DOI: 10.1177/0957926508088962

BAKER, P. – GABRIELATOS, C. – MCENERY, T. 2013. Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press. Cambridge: Cambridge University Press, 280 p.

BALASHOVA, L.V. 2014. Realizacija konceptov «svoj-chuzhoj» v rossijskom politicheskom diskurse nachala XXI v. In: Politicheskaja lingvistika, vol. 1(47), pp. 40-50

BLINDER, S. 2015. Imagined Immigration: The Impact of Different Meanings of 'Immigrants' in Public Opinion and Policy Debates in Britain. In: Political Studies, vol. 63(1), pp. 80-100

BORGER, J. 2020. Downright dangerous: Democrats' alarm as Trump stacks Pentagon with loyalists. Available online: https://www.msn.com/en-gb/news/world/downright-dangerous-democrat-alarm-as-trump-stacks-pentagon-with-loyalists/ar-

BB1aUsS2?fbclid=IwAR0BOfY8tolEcZnPO_TiJK1RnBIjvchQpjAzd1_4XIlTI7OSS
5wokP48Lno

BROOKES, G. – MCENERY, A. 2020 Correlation, collocation and cohesion: A corpus-based critical analysis of violent jihadist discourse. In: Discourse and Society, vol. 31(4), pp. 351-373. DOI: 10.1177/0957926520903528

BROWN, B. 2020. English language and digital health humanities. The Routledge Handbook of English Language and Digital Humanities. Routledge, 628 p. DOI: 10.4324/9781003031758

BUBENHOFER, N. 2008. Diskurse berechnen? Wege zu einer korpuslinguistischer Diskursanalyse. Warnke I., Spitzmüller J. Methoden der Diskurslinguistik. Berlin, New York: Walter de Gruyter, pp. 407-434

CHERNYAVSKAYA, V.E. 2017. Metodologicheskie vozmozhnosti diskursivnogo analiza v korpusnoj lingvistike.   In: Vestnik Tomskogo gosudarstvennogo universiteta. Filologija, vol. 50, pp. 135-148. DOI: 10.17223/19986645/50/9

DAVIES, M. 2016. Corpus of News on the Web (NOW): 10 billion words from 20 countries, updated every day. Available online: https://www.english-corpora.org/now/

DIJK, T. A. van. 2015. Critical Discourse Analysis. Handbook of Discourse Analysis. Second Edition. Chichester: Wiley Blackwell, vol. 1, pp. 466-485

EVERT, S. 2008. Corpora and collocations. In: Corpus linguistics. An international handbook.   Berlin,   Germany: Mouton   de   Gruyter,   pp.   1212-1248. DOI:10.1515/9783110213881.2.1212

FAIRCLOUGH, N. 1993. Critical discourse analysis and the marketization of public discourse. In: Discourse and Society, 4(2), pp. 133-168

FAIRCLOUGH, N. 2003. Textual analysis for social research. New York: Routledge, 210 p.

FAIRCLOUGH, N, – MULDERRIG, J. – WODAK, R. 2011, Critical Discourse Analysis. In: Discourse Studies. A Multidisciplinary Introduction. London: Sage, pp. 357-378. DOI: 10.4135/9781446289068.n17

FINEGAN, E. 2004. Language: its structure and use. N.Y.: Harcourt Brace College Publishers, 656 p.

GABRIELATOS, C. – BAKER, P. 2006. Representation of Refugees and Asylum Seekers in UK Newspapers: Towards a Corpus-Based Comparison of the Stance of Tabloids and Broadsheets. First International Conference: Critical Approaches to Discourse Analysis Across Disciplines. University of East Anglia, Norwich, UK

GABRIELATOS, C. – BAKER, P. 2008. Fleeing, Sneaking, Flooding. A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. In: Journal of English Linguistics, vol. 36(1), pp. 5-38. DOI: 10.1177/0075424207311247

GRANGER, S. – BESTGEN, Y. 2014. The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. In: International Review of Applied Linguistics in Language Teaching, vol. 52, pp. 229-252. DOI: 10.1515/iral-2014-0011

GRIES, S. Th. 2015. Some current quantitative problems in corpus linguistics and a sketch of some solutions. In: Language and Linguistics, vol. 16, pp. 93-117. DOI: 10.1177/1606822X14556606

HANDL, S. 2008. Essential collocations for learners of English: The role of collocational direction and weight. In: Phraseology: An interdisciplinary perspective. Amsterdam: John Benjamins, pp. 43-66. DOI: 10.1075/z.139

HUNSTON, S. 2007. Semantic Prosody Revisited. In: International Journal of Corpus Linguistics, vol. 12(2), pp. 249-268

IORDANSKAJA, L.N. – MEL'CHUK, I.A. 2007. Smysl i sochetaemost' v slovare. Moscow, 672 p. ISBN: 5-9551-0181-0

KARASIK, V.I. 2018. Adresatnaja specializacija v publichnom politicheskom diskurse. In: Vestnik Rossijskogo universiteta druzhby narodov. Serija: Teorija jazyka. Semiotika. Semantika, vol. 9(1), pp. 32-49. DOI: 10.22363/2313-2299-2018-9-1-32-49

KHOSRAVINIK, M. 2009. The Representation of Refugees, Asylum Seekers and Immigrants in British Newspapers During the Balkan Conflict (1999) and the British General Election (2005). In: Discourse & Society, vol. 20(4), pp. 477-498. DOI: 10.1177/0957926509104024

KRISTEVA, Ju.V. 1993. Slovo, dialog, roman. Moscow: Vysshaja shkola, 400 p.

LACLAU, E. 1985. Hegemony and socialist strategy: Towards a radical democratic politics. London: Verso, 240 p.

LEEUWEN, T. van. 2008. Discourse and practice. New tools for critical discourse analysis. Oxford: OUP, 184 p.

LEONHARDT, D. 2018. Team Trump's 3 Bizarre Defenses of Child Separations. Available online: https://www.nytimes.com/2018/06/18/opinion/trump-family-separation-immigrant-border.html

MAUTNER, G. 2009. Corpora and critical discourse analysis. In: Contemporary Corpus Linguistics. Continuum, pp. 32-46

MCENERY, A.M. – WILSON, A. 2001. Corpus linguistics: an introduction. Edinburgh: Edinburgh University Press, 235 p. ISBN: 0748611657

MCENERY, T. – HARDIE, A. 2011. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 292 p.

MONTES, J. 2019. Mexican Authorities Relocate Hundreds of Migrants From Border City. Available online: https://www.wsj.com/articles/mexican-authorities-relocate-hundreds-of-migrants-from-border-city-11550262962

MORAIN, D. 2020. There's a front-runner for Kamala Harris's Senate seat – but Gavin Newsom could surprise us all Opinion. Available online: https://www.washingtonpost.com/opinions/2020/11/23/gavin-newsom-replace-kamala-harris-senate-seat/

PETALING, J. 2020. No element of racism in SOPs, says Ismail Sabri. Available online: https://www.thesundaily.my/home/no-element-of-racism-in-sops-says-ismail-sabri-AJ5139828

PHILO, G. – BRIANT, E. – DONALD, P. 2013. Bad News for Refugees. London: Pluto Press, 224 p. ISBN: 9780745334325

PLUNGYAN, V.A. 2008. Korpus kak instrument i kak ideologija: o nekotoryh urokah sovremennoj korpusnoj lingvistiki. In: Russkij jazyk v nauchnom osveshhenii, vol. 2(16), pp. 7-20

POWELL, J. 2018. The 20 photographs of the week. Available online: https://www.theguardian.com/artanddesign/gallery/2018/jun/23/the-20-photographs-of-the-week

RJAZANOVA, E.A. 2012. Formirovanie grammaticheskih navykov rechi v celjah obuchenija inostrannomu jazyku v jazykovom vuze. In: Vestnik Tambov-skogo universiteta. Serija Gumanitarnye nauki. Tambov, vol. 9(113), pp. 153-156

SANTA ANA, O. – TREVINO, S. – BAILEY, M. – BODOSSIAN,K. – DE NECOCHEA, A. 2007. A May to Remember: Adversarial Images of Immigrants in U.S. Newspapers During the 2006 Policy Debate. In: Du Bois Review: Social Science Research on Race, vol. 4(1), pp. 207-232

SHEAR, M.D. 2020. Transition Highlights: Biden Plans to Name Blinken, Sullivan and Thomas-Greenfield to Cabinet. Available online: https://www.nytimes.com/live/2020/11/22/us/joe-biden-trump

SINCLAIR, J. M. 1991. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 179 p. ISBN: 0194371441

SINCLAIR, J. 2004. Trust the Text. Language, Corpus and Discourse. London, New York: Ruthledge, 224 p.

STUBBS, M. 1995. Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies. In: Functions of Language, vol. 2(1), pp. 23-55. DOI: 10.1075/fol.2.1.03st

SYSOEV, P.V. 2010. Lingvisticheskij korpus, korpusnaja lingvistika i metodika obuchenija inostrannym jazykam. In: Inostrannye jazyki v shkole, n. 5, pp. 12-21

WODAK, R. 2009. The Discourse-Historical Approach. In: Methods of Critical Discourse Analysis. London: Sage, pp. 63-94

ZUBOV, A.V. – ZUBOVA I.I. 2004. Informacionnye tehnologii v lingvistike: Ucheb. posobie. Moscow: Izdatel'skij centr "Akademija", 208 p.

Prof. Marina S. Matytcina, Doctor of Philology
Foreign Languages Department
Lipetsk State Technical University
Moskovskaya Street, 30
398055 Lipetsk
Russia
matutsina_ms@stu.lipetsk.ru

Prof. Olga N. Prokhorova, Doctor of Philology
Department of English Philology and Cross-Cultural Communication
Belgorod State National Research University
Studencheskaya Street, 14
308015 Belgorod
Russia
prokhorova@bsu.edu.ru

Prof. Igor V. Chekulai, Doctor of Philology
Department of English Philology and Cross-Cultural Communication
Belgorod State National Research University
Studencheskaya Street, 14
308015 Belgorod
Russia
chekulai@bsu.edu.ru

Ass. Prof. Elena V. Pupynina, PhD
Department of English Philology and Cross-Cultural Communication
Belgorod State National Research University
Studencheskaya Street, 14
308015 Belgorod
Russia
pupynina@bsu.edu.ru