



REPORT

Bulgarians & Romanians in the British National Press: 1 December 2012 – 1 December 2013

Published: 18/08/2014



www.migrationobservatory.ox.ac.uk

Executive summary

On 1 January 2014 the British government – along with all other EU governments – was required to lift the temporary restrictions that had been placed on Romanian and Bulgarian (A2) citizens' rights to work in the UK. These transitional controls were introduced after Romania and Bulgaria joined the EU in 2007 to reduce the likelihood of a sudden increase in immigration from those countries to the UK.

From 1 December 2012 to 1 December 2013, an important period leading up to the lifting of these transitional labour market controls, Britain's 19 main national newspapers (see list below) published more than 4,000 articles, letters, comment pieces and other items mentioning Romanians or Bulgarians. In total this amounted to more than 2.8 million words. This report uses a quantitative 'big-data' methodology from the field of corpus linguistics to provide a systematic and comprehensive analysis of the language used by these newspapers to discuss Romanians and Bulgarians during this period.

While some major news stories – such as the horsemeat scandal in early 2013, several sports stories and a case in which a Bulgarian Roma family in Greece was investigated for abduction after a distinctive blonde child was found in their care – were unrelated to migration, the focus of most of the coverage was related to migration and migrants.

The analysis provides several interesting findings which include:

- Language used by tabloid newspapers to describe and discuss Romanians as a single group was often focused on crime and anti-social behavior (**gang, criminal, beggar, thief, squatter**). This was less prevalent in broadsheet newspapers.
- Language used to describe and discuss Bulgarians as a single group, in both tabloid and broadsheet newspapers, did not consistently relate to any single social issue.
- Where Romanians and Bulgarians were discussed together this was consistently in the context of immigration, across both tabloid and broadsheet newspapers.
- References to Romanians and Bulgarians together were frequently associated with specific numbers across both tabloid and broadsheet newspapers. The most common specific numbers were **29 million** – the approximate combined populations of Romania and Bulgaria – and **50,000** – a prediction from pressure group MigrationWatch, which campaigns for reduced migration to the UK, of how many A2 migrants will add to the UK population each year for five years following the end of transitional controls.
- Verbs used to describe or discuss Romanians and Bulgarians together, across both broadsheets and tabloids were frequently related to travel (**come, arrive, move, travel, head**) and in tabloids these included metaphors related to scale (**flood, flock**).
- Words appearing before mentions of Romanians and Bulgarians as a unit in tabloid and broadsheet newspapers were frequently related to prevention of movement (**stop, control, block** in tabloids; **deter, restrict, dissuade** in broadsheets).
- Romanians and Bulgarians were regularly associated with travelling to the UK for work across both tabloid and broadsheet publications.
- Those words that are consistently used to describe Roma or Gypsies in both tabloid and broadsheet publications are generally related to either crime and antisocial behaviour, persecution or settlement.

The report builds on the Migration Observatory's previous work in quantitative analysis of media coverage of migration which can be found at www.migrationobservatory.ox.ac.uk/projects/media.

Newspapers analysed were: *The Daily Mail, The Mail on Sunday, The Times, The Sunday Times, The Sun, The Sun on Sunday, The Daily Telegraph, The Sunday Telegraph, The Express, The Sunday Express, The Guardian, The Observer, Daily Mirror, Sunday Mirror, The Independent, Independent on Sunday, Daily Star, Daily Star Sunday, The Financial Times*.

1. Introduction

International migration is a contentious and complex issue in many European countries. Debates about the scale and composition of migration to Britain continue to occur in a range of policy, public opinion, and media spheres. One of the most visible places in which these debates occur is the UK's national newspapers. Whether intentionally or not, members of the public encounter different portrayals of immigrants through the press. By presenting information about a variety of issues, newspapers can inform, shape or guide readers' understandings to certain conclusions.

This report examines a key issue in the contemporary story of British immigration: the removal of transitional labour market controls on Bulgarians and Romanians in the UK. On 1 January 2014, restrictions on the ability of people from these countries to access the UK labour market were lifted. Prior to this date, most Romanians and Bulgarians who wished to undertake paid employment in the UK had to obtain a valid form of work authorisation, while their access to benefits was restricted. In the run up to the January 2014 expiry of transitional controls, several sections of the British press discussed the potential magnitude and impact of migration after the limitations were lifted. Data measuring the scale of movement from Bulgaria and Romania since those controls were lifted are only beginning to become available (see the Migration Observatory's commentary "It's too early to know whether the number of Romanians and Bulgarians will rise or fall in 2014").

But what kinds of issues did the press raise about Bulgarians and Romanians – both as separate groups and together? Given the political attention given to media representation of Romanians and Bulgarians, this report aims systematically to document how these groups were portrayed in the national UK press in the period leading up to the removal of transitional controls. Specifically, it seeks to ascertain:

1. What issues were raised alongside mentions of both groups when they appeared together.
2. What language was used in connection with mentions of each group when they appeared separately.

Using quantitative methods to analyse newsprint coverage of these groups in the time prior to the lifting of transitional controls, the Migration Observatory found that when the national press mentioned both groups together, it was more often in the context of immigration than any other issue. Meanwhile, people or objects described as solely 'Bulgarian' tended to refer to particular sports figures or a specific story about a Bulgarian girl living with a Roma couple in Greece. This was in contrast to people or objects specifically described as 'Romanian', which included references to gangs, criminals, and the 2013 horse meat scandal that involved Romanian abattoirs. Also, the Roma were frequently described as Gypsies and, when mentioned on their own, were associated with crime and homelessness.

2. Data and strategy for analysis

This report draws upon a collection of news articles and other newspaper items, called a 'corpus', which includes over 4,000 items made up of about 2.8 million words. The articles were drawn from ten of the most widely-read UK publications and their Sunday editions from 1 December 2012 to 1 December 2013, grouped into two broad categories: tabloids/midmarkets and broadsheets. Publications were grouped together since this project was interested in examining the kinds of language generally used to describe Bulgarians and Romanians across major sections of the press rather than by individual publications. Furthermore, the dataset did not include the month immediately prior to 1 January 2014, as this busy period might unintentionally skew the results. As far as possible, the corpus captures every item in coverage that mentions the terms BULGARIA, BULGARIAN, BULGARIANS, or their counterparts ROMANIA, ROMANIAN, or ROMANIANS.

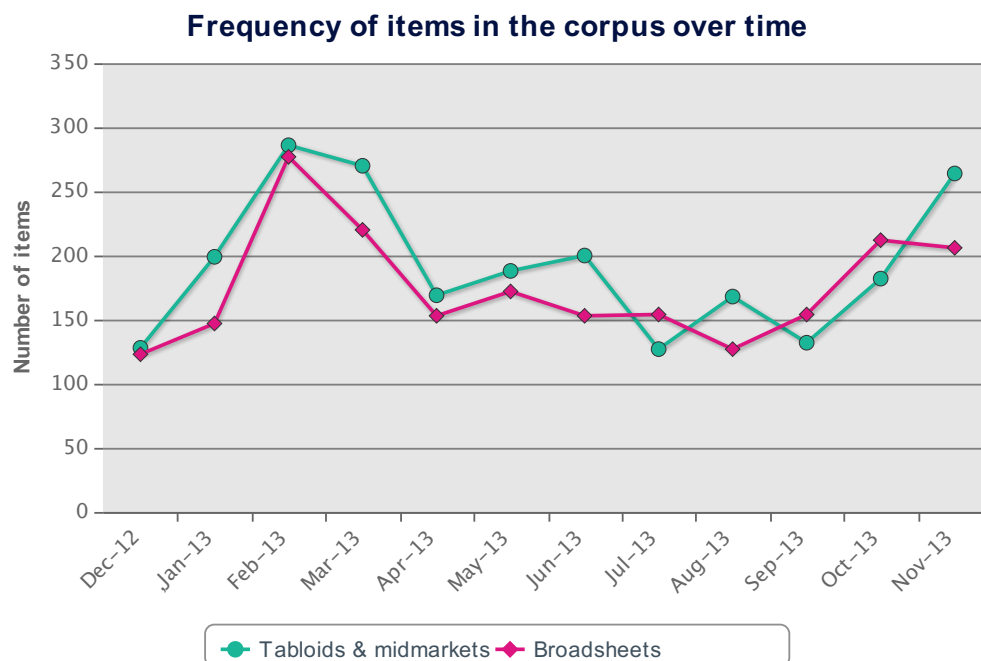
Table 1 - National UK press publications included in the study

Tabloids & Midmarkets	Broadsheets
The Daily Mail The Mail on Sunday	The Times The Sunday Times
The Sun The Sun on Sunday	The Daily Telegraph The Sunday Telegraph
The Express The Sunday Express	The Guardian The Observer
Daily Mirror Sunday Mirror	The Independent Independent on Sunday
Daily Star Daily Star Sunday	The Financial Times

The Migration Observatory A2 Corpus was built by downloading newspaper items from NexisUK, a database service that archives periodicals and similar publications. The search was intentionally kept wide, capturing all kinds of items including sports, arts and culture reviews, letters to the editors, and travel advice columns. Since this project was interested in showing how A2 countries, as well as people from those countries were portrayed in the British press, these kinds of items were relevant because they contributed to the variety of information available to readers. Results were initially filtered to exclude the majority of duplicates using NexisUK, and then manually searched to catch any remaining duplicates. Full details of these filtering steps, the number of items subsequently excluded and the overall characteristics of the corpus are presented in Appendix 1 and 2.

Figure 1 displays the frequency of items mentioning Bulgaria, Bulgarians, Romania, or Romanians in the corpus. It shows an increase in the number of items from December 2012 to February 2013, with the highest monthly number of items in tabloids and broadsheets over the entire study period appearing in February 2013. Then there was a decrease in tabloid and broadsheet coverage until March 2013. Over these 12 months, tabloids published more items than broadsheets, mentioning the key terms in all but three of the months (July, September, and October).

Figure 1



Next, to examine how these newspapers discussed Bulgarians and Romanians, a corpus linguistic approach was used. This approach involves the statistical analysis of large, computer-based datasets of texts in order to detect particular patterns of language across the corpus (McEnery and Hardie 2011). This report focuses on one kind of pattern called ‘collocation’, or the observation that certain words appear together more often than would be expected by chance, and in doing so communicate particular meanings (Sinclair 1991). For example, the Observatory’s previous report on migration in the media found that the descriptor ILLEGAL appeared more often with IMMIGRANTS than with any other word in press coverage between 2010 and 2012 (Allen and Blinder 2013). When these techniques are scaled up to large amounts of newspaper text, researchers can show how groups like immigrants are typically described in a way that minimises researcher subjectivity: “it becomes less easy to be selective about a single newspaper article when we are looking at hundreds of articles” (Baker 2006: 12). This report focuses on the range of words that appear before and after one of the targeted words, as seen in Figure 2.

Figure 2 - Positions of collocates in relation to a target word

L5	L4	L3	L2	L1	TARGET WORD	R1	R2	R3	R4	R5
----	----	----	----	----	-------------	----	----	----	----	----

Focusing on words appearing before a target word in the L1-L5 range tends to identify descriptors of the target word, whereas R1-R5 collocates tend to identify what sorts of objects are being described. For more information on collocation, as well as the statistical tests used to establish a significant link, please see Appendix 2 as well as the previous Migration Observatory report which uses identical collocation techniques.

While these tendencies in position are useful for the majority of text, some variations may still exist due to the complexity of language. In everyday language, the word ‘Romanian’ might be used either as a noun, to refer to a person (“A Romanian came to the country”), or as an adjective to describe something (‘Romanian homes’, ‘Bulgarian television’). Equally, adjectives may not appear directly before the objects they modify, especially when more than one are used (‘Bulgarian national team’, or ‘young, Romanian children’). To differentiate among these uses, corpus linguists can attach extra information called ‘metadata’ to parts of the corpus. One key type of metadata is the part of speech an individual word acquires in context – for instance, whether it is used as a noun, adjective, or verb. This kind of grammatical information is helpful for identifying patterns of portrayals because it provides a set of ‘rules’ that can reliably identify how a word is being used. For example, take the following sentences which form a very small example corpus of 16 words:

- The **Romanian** footballer played well.
- A **Bulgarian** state official spoke today.
- The police arrested the **Romanian**.

Tagging the major words according to their basic use would result in:

- The Romanian-[adjective] footballer-[noun] played-[verb] well-[adverb]
- A Bulgarian-[adjective] state-[adjective] official-[noun] spoke-[verb] today-[adverb].
- The police-[noun] arrested-[verb] the Romanian-[noun]

To answer the question “what kinds of things are typically described as Romanian?”, researchers could examine the corpus for all instances of the word ‘Romanian’ when it is used as an adjective rather than a noun. Using the corpus above, they would find that a ‘footballer’ is described as ‘Romanian’. Equally, looking for things described as Bulgarian would show a result for ‘Bulgarian official’ even though the descriptor does not immediately precede its object. This is a significant advantage over simply looking for all of the words in the R1 position, which would not necessarily pick up this collocation example. Similarly, to answer the question “what do Romanians do?”, analysts could identify

which verbs appear with the words 'Romanian' or 'Romanians'. This might reveal that 'Romanian migrants ARRIVE', or that 'Romanians will GAIN the right to work'. However, this kind of tagging does not reveal what kinds of broader claims are being made when words are used together, or whether these uses are negative or positive. This would require different kinds of analysis.

The Migration Observatory used web-based software called Sketch Engine to apply these grammatical metadata to the entire corpus, then analyse the resulting collocation patterns. Sketch Engine uses established sets of rules to determine how each word is used within a corpus, attaches the appropriate metadata, then employs statistical tests to identify collocations. It can also organise the results of these tests into one-page summaries called 'word sketches'. These sketches provide a snapshot of how a selected word is used by displaying its main descriptors or the actions it performs. While this procedure can greatly facilitate the interpretation of collocations, it is only available when the search term is a single word. It cannot accommodate a phrase like 'Bulgarians and Romanians', nor can it detect instances when the same word is spelled differently, such as 'Gypsy' or 'Gipsy'. Consequently, in order to overcome this methodological challenge, the collocation results for every key term were manually inspected based on whether they were nouns, verbs, or numerals. More detail about the Sketch Engine is provided in Appendix 2.

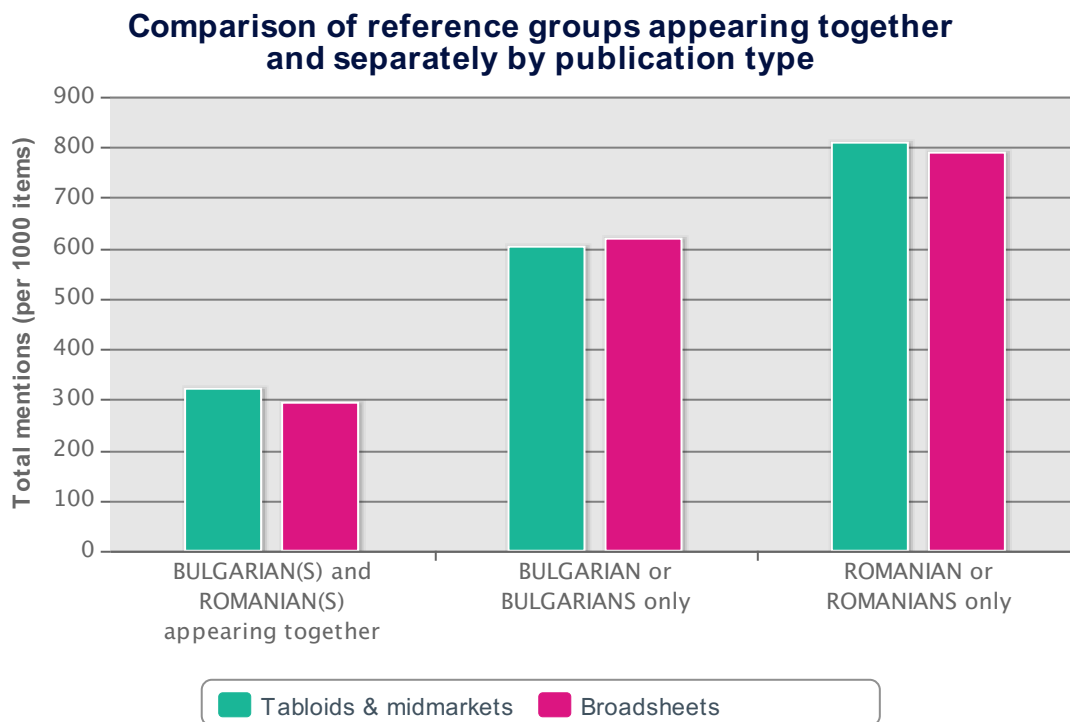
The following sections show results from the word sketches and collocation analysis. Words in ALL CAPITALS refer to examples from the texts. Frequencies of the patterns are given both in raw observations and 'normalised' counts per 1,000 items. Typically, corpus linguists will calculate how many times a word or collocation occurs per thousands or millions of words in a process called 'normalisation' (McEnery and Hardie 2011). This enables comparison across differently sized datasets: it would be misleading to report that a given word appears more often in broadsheets compared to tabloids without accounting for the fact that the broadsheet sample was larger anyway. However, strictly relying upon occurrences per million words would overstate the frequency of a collocation in tabloids while understating it in broadsheets, because tabloid items tend to be significantly shorter than broadsheet items (Gabrielatos and Baker 2008). Therefore, this report also normalised the results per 1,000 items. This effectively means that one mention of BULGARIANS in a longer broadsheet item counts as the same as one mention in a shorter tabloid item (Allen and Blinder 2013).

Finally, in order to illustrate the collocation findings, this report includes examples from the corpus as bulleted items after each relevant table along with the types of publication in which they appeared. These sentences, called 'concordance lines', were selected using an algorithm called GDEX, which was developed to identify good example sentences for dictionaries. Built into the Sketch Engine, it ranks sentences based on several criteria such as complexity and overall length. Further details are in Appendix 2, as well as in the work of Kilgarriff and his colleagues (2008). Crucially, these examples are intended only as illustrations of the quantitative findings.

3. Results: What do Bulgarian and Romanian describe in the UK press

Figure 3 shows the overall frequency of ROMANIAN and BULGARIAN when they appear together compared to when either group (ROMANIA/ROMANIANS and BULGARIA/ BULGARIANS) appears by itself. It reveals that mentions of ROMANIAN and ROMANIANS by themselves were the most frequent in the corpus, followed by mentions of BULGARIAN and BULGARIANS. Mentions of either group by themselves were more prevalent in both tabloids and broadsheets, suggesting that newspapers' coverage tended to focus on one group or the other rather than as a unit. Meanwhile, tabloids made greater reference to the two groups together compared to broadsheets. Compared to broadsheets, tabloids made about 9% more references to the two groups together, and about 3% more references to ROMANIAN or ROMANIANS separately, in relation to their total number of items published. Meanwhile, broadsheets made about 2% more references to BULGARIAN or BULGARIANS separately compared to tabloids, in relation to the total number of broadsheet items.

Figure 3



This section focuses on the words that were used in connection with the two groups when they appeared together. It presents findings from the grammatical analysis of words linked with mentions of BULGARIAN and ROMANIAN together as adjectives, i.e., as words used to describe something or someone else. Since the corpus was built to include all topics which might

appear in newspapers in relation to Bulgaria or Romania, this analysis reveals what issues the British press raised when it mentioned both of these groups in the same instance.

The first step was to examine what nouns followed either the phrases 'ROMANIAN AND BULGARIAN' or 'BULGARIAN AND ROMANIAN'. This would indicate the kinds of issues that were raised alongside both of these groups in press coverage. Table 2, showing the words that followed these phrases, reveals that migration issues were most often associated with mentions of Bulgarians and Romanians. MIGRANT, IMMIGRANT, and IMMIGRATION all feature in tabloid and broadsheet publications' discussion.

Table 2 - Nouns described by the adjectives ROMANIAN AND BULGARIAN By publication type

Tabloids			Broadsheets		
Noun	Raw frequency	Normalised	Noun	Raw frequency	Normalised
migrant	35	15.05	migrant	48	22.70
immigrant	34	14.62	worker	32	15.13
worker	26	11.18	immigrant	31	14.66
immigration	21	9.03	citizen	23	10.87
citizen	16	6.88	immigration	11	5.20
student	4	1.72	migration	4	1.89
migration	4	1.72	economic	4	1.89
national	4	1.72			
January	4	1.72			
seasonal	3	1.29			
influx	3	1.29			

As illustrated by example sentences below, it is clear how mentions of Romanians and Bulgarians also involved migration issues:

- But what we don't have are any measures on **Romanian and Bulgarian migrants**, which is what people mention if you ask them about immigration. (Tabloid)
- From Jan 1, 2014 **Romanian and Bulgarian migrants** will have free access to Britain's labour market. (Broadsheet)
- He knows full well that the Government is playing with fire by proposing to lift controls on **Romanian and Bulgarian immigration**. (Tabloid)

Next, examination of the verbs that regularly appeared after mentions of these groups, as seen in Table 3 below, show several groups of related words. One set involves inward movement: for example, COME, ARRIVE, and SETTLE. The sample concordance lines which follow the table illustrate how these words were used.

Table 3 - Top 20 verbs appearing after BULGARIANS and ROMANIANS by publication type

Tabloids			Broadsheets		
Verb	Raw frequency	Normalised	Verb	Raw frequency	Normalised
will	115	49.44	come	71	33.57
come	88	37.83	will	69	32.62
arrive	39	16.77	work	47	22.22
live	30	12.90	live	27	12.77
move	27	11.61	could	23	10.87
access	23	9.89	gain	18	8.51
lift	20	8.60	arrive	18	8.51
expect	19	8.17	move	14	6.62
gain	17	7.31	travel	11	5.20
head	16	6.88	lift	9	4.26
flock	15	6.45	allow	8	3.78
allow	12	5.16	head	8	3.78
flood	11	4.73	settle	6	2.84
settle	11	4.73	enter	6	2.84
enter	8	3.44	expect	6	2.84
travel	8	3.44	strain	5	2.36
relax	7	3.01	seek	4	1.89
prepare	5	2.15	crop	3	1.42
poise	4	1.72	curb	3	1.42
emigrate	4	1.72	blame	3	1.42

- The change in rules will allow **Romanians and Bulgarians** to come to the UK to seek work without any barriers in place. (Broadsheet)

- One week it's fears that tens, even hundreds, of thousands of **Bulgarians and Romanians** could **arrive** on our shores thanks to the freedom of movement allowed under EU rules. (Tabloid)

Another set of words centred on these groups' changing status. As the examples below illustrate, verbs like GAIN, LIFT, and ALLOW were used to signal the new rights Romanians and Bulgarians would acquire starting the 1st of January 2014.

- Next year 29 million **Bulgarians and Romanians** will **gain** the right to live and work unrestricted in Britain under European 'freedom of movement' rules yet ministers refuse to disclose how many may come here. (Tabloid)
- Mr Cameron has now set up a ministerial Cabinet committee to examine the rules on migrants' access to benefits before the **Romanians and Bulgarians** are **allowed** to move to Britain for work when limits expire on December 31. (Broadsheet)
- There are fears of a huge influx of immigrants after restrictions on **Bulgarians and Romanians** are **lifted** at the end of the year. (Tabloid)

Finally, it is also interesting to note the presence of some words linked with metaphors of water and quantity particularly in the tabloids (FLOOD and FLOCK). In its previous study, the Observatory found similar patterns.

- It is as good as given that when the borders open in 2014 **Bulgarians and Romanians** will **flood** in. (Tabloid)
- He is spoiling for a fight, particularly over fears that **Romanians and Bulgarians** will **flood** to Britain when visa restrictions are removed next year. (Broadsheet)
- Mr Duncan Smith said that he was seeking an urgent meeting of European ministers in the next fortnight to address fears that tens of thousands of **Romanians and Bulgarians** would **flock** to the UK when transition arrangements end on January 1. (Broadsheet)

Looking for verbs among the words appearing before mentions of Romanians and Bulgarians can provide a sense of what kinds of actions the press reported were done to these groups. As Table 4 shows, verbs like ESTIMATE, EXPECT, PREDICT, or FORECAST indicate a theme of anticipation, while words such as STOP, CURB, DETER, DISSUADE or BLOCK suggest processes of limiting movement. Closer inspection of RESTRICT and CONTROL revealed that these were actually referring to restrictions and controls placed upon Romanians and Bulgarians

Table 4 - Words appearing before mentions of ROMANIANS and BULGARIANS by publication type

Tabloids			Broadsheets		
Word	Raw frequency	Normalised	Word	Raw frequency	Normalised
estimate	34	14.62	control	21	9.93
stop	26	11.18	estimate	17	8.04
open	18	7.74	deter	13	6.15
expect	17	7.31	lift	11	5.20
lift	14	6.02	allow	10	4.73
flood	13	5.59	open	10	4.73
control	13	5.59	curb	9	4.26
allow	12	5.16	impose	9	4.26
curb	9	3.87	suggest	8	3.78
predict	6	2.58	try	8	3.78
publish	6	2.58	restrict	6	2.84
surge	5	2.15	rule	5	2.36
exist	5	2.15	forecast	4	1.89
impose	5	2.15	fill	4	1.89
block	5	2.15	predict	4	1.89
impact	5	2.15	warn	4	1.89
throw	5	2.15	dissuade	3	1.42
attempt	5	2.15	aim	3	1.42

- The Home Office is aiming to **deter Romanians and Bulgarians** from arriving next year. (Broadsheet)
- Ministers have made no attempt to **estimate** the number of **Romanians and Bulgarians** who will come to Britain when border controls are scrapped, it emerged last night. (Tabloid)
- The Seasonal Agricultural Workers Scheme **allows 21,250 Romanians and Bulgarians** a year to come to Britain to pick crops for up to six months. (Broadsheet)
- The vast majority in a poll want working restrictions left in place to stop an expected stampede of **Romanians and Bulgarians** to Britain in January. (Tabloid)

Finally, the analysis showed UK publications used a wide range of numerical figures to characterise these groups, whether expected or actual. Words that referenced quantities such as THOUSAND and MILLION were frequent, as were a variety of other numerals. Table 5 displays the instances when numerical references were associated with Bulgarians and Romanians. The 50,000 figure was particularly associated with a report published by MigrationWatch UK, but a range of other numbers was also mentioned. The raw and normalised frequency figures show that these kinds of references were made somewhat more often in the tabloids than in the broadsheets.

Table 5 - Numerals and specific figures associated with mentions of BULGARIANS and ROMANIANS

Tabloids			Broadsheets		
Figure	Raw frequency	Normalised	Figure	Raw frequency	Normalised
million	69	29.66	million	24	11.35
29	50	21.50	29	13	6.15
thousands	29	12.47	50,000	12	5.67
50,000	20	8.60	21,250	5	2.36
millions	13	5.59	250,000	5	2.36
350,000	8	3.44	26,000	4	1.89
250,000	8	3.44	80,000	4	1.89
70,000	6	2.58	millions	4	1.89
35,000	4	1.72	35,000	3	1.42
13,000	4	1.72	350,000	3	1.42
30,000	3	1.29			
400	3	1.29			

- MigrationWatch UK has estimated that about **50,000 Romanians and Bulgarians** a year will come to Britain when curbs on them entering the jobs market are lifted. (Broadsheet)
- Next year **29 million Bulgarians and Romanians** will gain the right to live and work unrestricted in Britain under European “freedom of movement” rules yet ministers refuse to disclose how many may come here. (Tabloid)
- And in a toughening of existing rules aimed specifically at the prospect of **tens of thousands of Bulgarians and Romanians** arriving after January 1, migrants caught begging or sleeping rough will be deported and banned from returning to Britain for a year. (Broadsheet)

4. Results: Portrayals of Romanians and Bulgarians as separate groups

Although a great deal of press coverage treated Bulgarians and Romanians together as a unit, newspapers actually made more frequent reference to either group by themselves, as seen in Figure 3. Turning attention to text that mentions only one group can show differences and similarities in the kinds of issues raised. To do this, we analysed the R1 collocates of ROMANIAN or BULGARIAN when they were adjectives and appeared in the absence of the other group.

Table 6 displays the top twenty nouns described as ROMANIAN. Several interesting findings emerge. First, references to the horse meat scandal of January 2012 prominently appear in both tabloids and broadsheets, as evidenced by mentions of words such as **ABBATOIR**, **SLAUGHTERHOUSE**, and **HORSE** which are highlighted in blue:

- The two **Romanian abattoirs** suspected of supplying the horsemeat yesterday insisted that it had been labelled properly. (Broadsheet)
- Horse found in UK supermarket foods has been traced to a **Romanian slaughterhouse**. (Tabloid)

Table 6 – Top 20 nouns described as ROMANIAN by publication type

Tabloids			Broadsheets		
Noun	Raw frequency	Normalised	Noun	Raw frequency	Normalised
side	21	9.03	abattoir	26	12.29
gang	19	8.17	migrant	16	7.57
abattoir	17	7.31	ambassador	13	6.15
criminal	13	5.59	gang	13	6.15
woman	13	5.59	immigrant	13	6.15
immigrant	13	5.59	authority	12	5.67
passport	11	4.73	woman	11	5.20
gypsy	11	4.73	beggar	9	4.26
beggar	10	4.30	horse	9	4.26
defender	10	4.30	orphanage	8	3.78
national	10	4.30	capital	8	3.78
slaughterhouse	8	3.44	supplier	7	3.31
ambassador	8	3.44	horsemeat	7	3.31
thief	8	3.44	citizen	7	3.31
authority	8	3.44	leu	6	2.84
minnow	7	3.01	girl	6	2.84
squatter	7	3.01	dictator	5	2.36
capital	7	3.01	parliament	5	2.36
keeper	7	3.01	soprano	4	1.89
stopper	5	2.15	slaughterhouse	4	1.89

Another group of related words centred on issues of crime, highlighted in pink. Words like GANG and BEGGAR were prominent across the publication types, while tabloids also mentioned Romanians alongside THIEF and SQUATTER.

- **Romanian gangs** are already thought to be responsible for most cashpoint fraud as well as a lot of pickpocketing and shoplifting. (Tabloids)
- The Home Office insists that paying for flights and coaches is a cheap way of ridding London’s streets of **Romanian beggars**. (Broadsheet)
- A furious judge blasted a gang of **Romanian thieves** who began shoplifting thousands of pounds of goods just weeks after arriving in Britain. (Tabloid)

Also, several references to sporting topics appeared in the tabloid subcorpus as seen by the group of words highlighted in green. SIDE, DEFENDER, and KEEPER all illustrate this theme. The word MINNOW was used in the context of the Romanian football club Cluj to signal their relative status in comparison to other clubs.

- Sadly for the **Romanian side** their group rivals Galatasaray were fighting back in Braga to lead and book their place in the next round with United. (Tabloid)
- The United manager claims his side’s defensive worries remain after they conceded their 33rd goal this season to lose at Old Trafford to **Romanian minnows Cluj**.

Finally, the word NATIONAL, which was observed in the tabloid subcorpus, was frequently used to signal the origins of a group or individual, as in the phrase ROMANIAN NATIONAL or NATIONALS. Closer examination of these instances reveals that the topic of alleged criminal behaviour, and to a lesser extent infectious diseases, is often raised alongside these phrases.

- The **Romanian national** was given an eight-hour rest period and officers had until the early hours of this morning to decide whether they intended to charge or release him. (Tabloid)
- The City of London Police’s dedicated cheque and plastic crime unit calculates that 92 per cent of all cash machine fraud is committed by **Romanian nationals**. (Tabloid)
- However, the report did warn an influx could lead to pressure on already scarce primary school places, and of high levels of diseases such as measles among **Romanian nationals**.

Meanwhile, the kinds of actions that the press reported Romanians taking were similar in both newspaper groups. Table 7 shows how verbs like COME, LIVE, and MOVE were most frequently used in connection with ROMANIAN in both publication types. Also, both publication types mentioned the verb ARREST in connection with mentions of Romanians – although in this case the action was being done to Romanians – confirming the link with criminality revealed by the noun collocations in Table 6.

Table 7 - Verb collocates of ROMANIAN across publication types

Tabloids			Broadsheets		
Verb	Raw frequency	Normalised	Verb	Raw frequency	Normalised
live	16	6.88	come	28	13.24
move	11	4.73	live	14	6.62
arrest	8	3.44	seek	8	3.78
sleep	5	2.15	arrest	7	3.31
rank	3	1.29	dress	3	1.42
			list	3	1.42
			stay	3	1.42

- A **Romanian arrested** for shoplifting was shown to be wanted for more than 60 serious offences including robbery, grievous bodily harm and kidnapping in Eastern Europe. (Broadsheet)
- The police figures published yesterday also show **Romanians arrested** for suspicion of involvement in 10 murders, 142 rapes and thousands of other serious crimes. (Tabloid)

Next, these tests were replicated among those texts only mentioning Bulgarians. Table 8 reports the top 20 noun collocates of BULGARIAN when it was used as an adjective. These are the nouns that the press typically described as Bulgarian:

Table 8 – Noun R1 collocates of BULGARIAN across publication types

Tabloids			Broadsheets		
Noun	Raw frequency	Normalised	Noun	Raw frequency	Normalised
woman	13	5.59	woman	17	8.04
midfielder	7	3.01	town	10	4.73
capital	7	3.01	Roma	10	4.73
couple	6	2.58	mother	9	4.26
star	5	2.15	couple	9	4.26
immigrant	5	2.15	nurse	8	3.78
dissident	4	1.72	capital	8	3.78
ambassador	4	1.72	ambassador	7	3.31
ace	4	1.72	driver	6	2.84
winger	4	1.72	immigrant	6	2.84
resort	4	1.72	politics	5	2.36
Gypsy	4	1.72	authority	5	2.36
striker	4	1.72	soprano	4	1.89
wine	4	1.72	interior	4	1.89
official	4	1.72	investigation	4	1.89
foreign	4	1.72	wine	4	1.89
open	4	1.72	citizen	4	1.89
millionaire	3	1.29	dissident	3	1.42
drug	3	1.29	actress	3	1.42
international	3	1.29	singer	3	1.42

One theme of press coverage involved a story of a Bulgarian girl who was found to be living with a Roma family in Greece as indicated by words like WOMAN, COUPLE, and MOTHER highlighted in blue. Also, sporting references highlighted in green appear in the tabloids as evidenced by the nouns TONEV, STAR, and MIDFIELDER, although the latter two words were used exclusively in a series of items about the Bulgarian player Stiliyan Petrov’s retirement in May 2013:

- They insist they were given the child by a **Bulgarian woman**, who was unable to look after her. (Tabloid)
- The real parents of a girl known only as Maria were finally identified last night as an impoverished **Bulgarian couple** who abandoned her in Greece, triggering a legal row over the child’s future custody. (Broadsheet)
- The **Bulgarian midfielder** retired as a player earlier this month after a successful battle against leukaemia. (Tabloid)

Turning attention to the actions that the press reported Bulgarians doing, Table 9 displays the verbs associated with BULGARIAN. It shows that tabloids tended to associate this group with actions indicating future movement, including MOVE, PLAN, and HEAD.

Table 9 - Verb collocates of BULGARIAN across publication types

Tabloids			Broadsheets		
Verb	Raw frequency	Normalised	Verb	Raw frequency	Normalised
move	10	4.30	kill	4	1.89
sell	7	3.01	deny	3	1.42
plan	7	3.01	receive	3	1.42
consider	6	2.58			
arrive	6	2.58			
head	5	2.15			
emigrate	3	1.29			
migrate	3	1.29			
flood	3	1.29			
pull	3	1.29			
accuse	3	1.29			

- Around 400,000 Bulgarians plan to move West when controls are lifted in less than six months - and the UK is now top of their list. (Tabloid)
- More than 80,000 Bulgarians are likely to move to Britain in a new wave of large-scale migration. (Tabloid)
- Even without new EU rules to make it easy for them, many Bulgarians are already heading overseas. (Tabloid)

When compared, these findings show that Bulgarians and Romanians tended to be collectively referred to in the context of immigration, while Romanians as a separate group tended to appear more often in connection with criminality and economic poverty - topics which were largely absent in the language surrounding mentions of Bulgarians only.

5. Results: Analysis of Roma and Gypsy

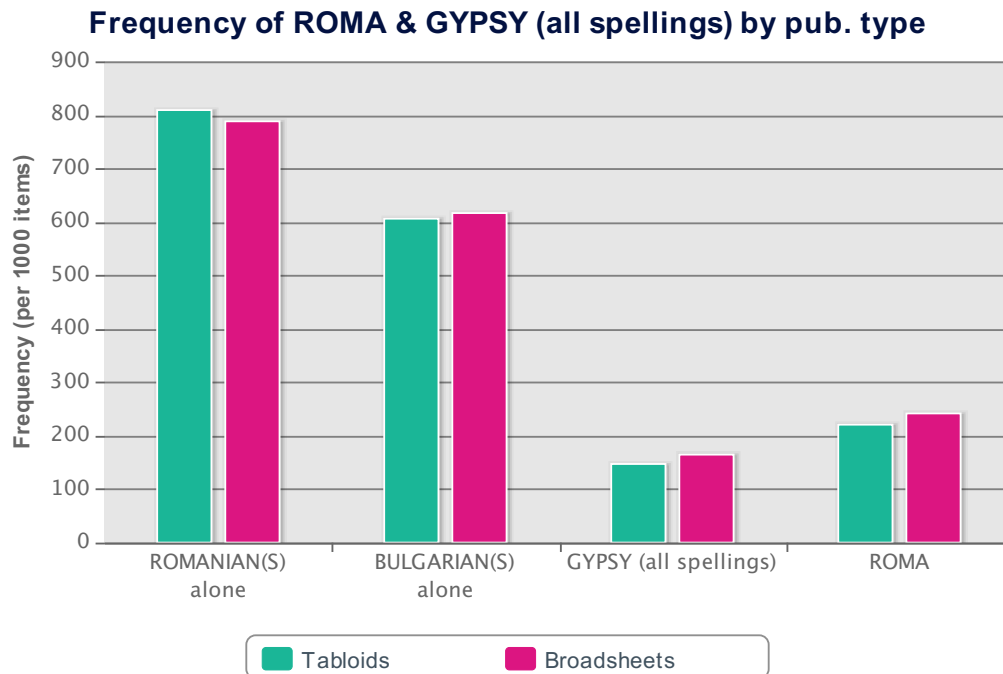
Finally, analysing the words associated with each separate group revealed an interesting finding. As seen in Tables 6 and 8, mentions of ROMA, GYPSY, and its variant GIPSY are collocated with references to each group. For example, closer review of all the statistically significant results for the adjective ROMANIAN reveals that it was also used to describe the Roma, who are alternatively referred to as GYPSY, GIPSY, and TRAVELLERS. While this variety of spellings and denominations prevents this association from being highlighted automatically, a more careful scrutiny reveals a strong co-occurrence of Romanian and Roma.

- Romanian gypsies are stealing metal worth millions of pounds to build palaces in their homeland. (Tabloid)
- Romanian travellers who were sent home earlier this year on taxpayer-funded coaches were yesterday back sleeping rough in some of the capital’s most exclusive postcodes. (Broadsheet)
- The Home Office was able in its document to identify only one case of ‘fraudulent benefit claims’: the 2012 conviction and jailing of Lavinia Olmazu, an EU national from Romania, who helped a gang funnel £2.9 million in false benefits claims to 170 Romanian gypsies. (Broadsheet)

Following this observation, a separate analysis was conducted for the words ROMA and GYPSY as well as their variations. The use of ‘?’ indicates that alternative spellings were included in the search: G?PS? would return GIPSY

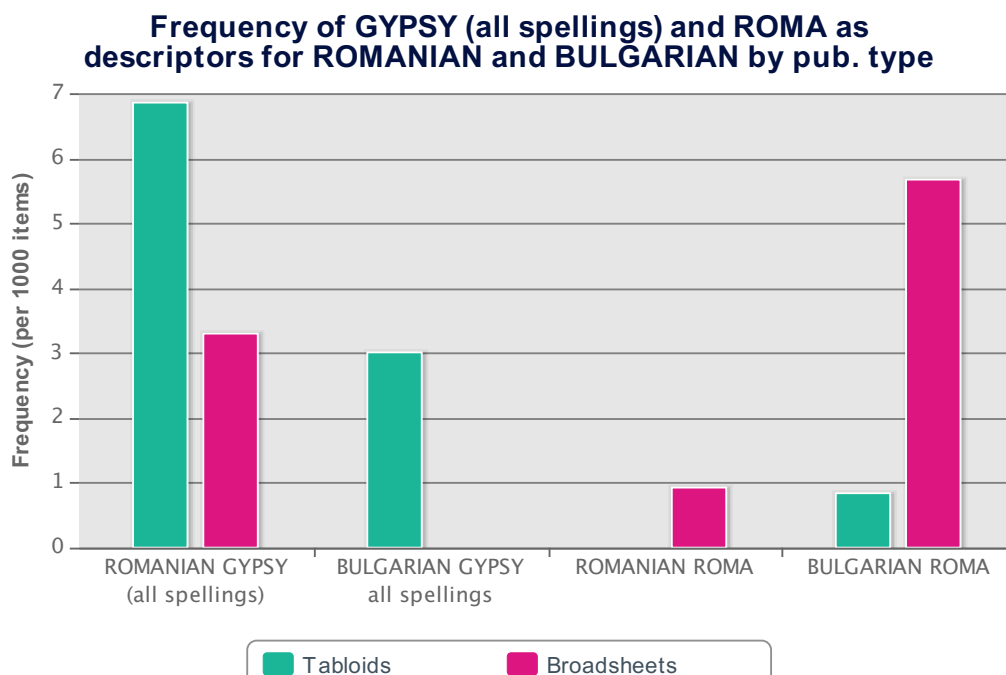
as well as GYPSY, GIPSIES, or GYPSIES. Both of these terms are frequent in the corpus as seen in Figure 4, which displays their frequencies per 1,000 items in relation to mentions of the main two target groups.

Figure 4



While most of the time these mentions refer to separate groups, they are sometimes used to describe Romanians or Bulgarians. Interestingly, as seen in Figure 5, variations of GYPSY are used twice as frequently to describe Romanians than Bulgarians in the tabloid press, and exclusively used to describe Romanians in broadsheets. The word ROMA, on the other hand, is almost always used to describe Bulgarians, although usually in reference to the particular story of a kidnapped Bulgarian girl. In contrast, the phrase ROMANIAN ROMA only appears twice in the broadsheet sample. However, these results should be interpreted cautiously due to the relatively small number of instances.

Figure 5



Closer examination of R1 collocations for G?PS? highlights a theme of settlement, expressed in words like COMMUNITY and SETTLEMENT. These were more frequently seen in the tabloid press, as Table 10 shows. Broadsheet coverage also involved the previously mentioned story of a kidnapped Bulgarian girl. Words like FAMILY and COUPLE illustrate this case.

Table 10 - R1 noun collocations for GYPSY (all spellings) by publication type

Tabloids			Broadsheets		
Word	Raw frequency	Normalised	Word	Raw frequency	Normalised
camp	34	14.62	camp	6	2.84
couple	11	4.73	family	6	2.84
community	10	4.30	couple	4	1.89
settlement	7	3.01	community	4	1.89
Sasha	5	2.15			
girl	4	1.72			
gang	4	1.72			
town	3	1.29			

- The girl was spotted by a police officer who was suspicious because her looks stood out in the **gypsy camp**. (Tabloid)
- I guess it shouldn't be surprising, as so few people know anything about the **Gypsy community**, it's so secretive and tight-knit. (Broadsheet)
- The **gypsy couple** claim the girl came into their care through a Bulgarian middle-man and Maria's biological mother, also Bulgarian, in early 2009. (Broadsheet)

An analysis of verbs that collocate with G?PS?, shown in Table 11, reveals that the actions they are reported to do, or have done to them, tend to revolve around crime. This is particularly prominent in the tabloid press through words such as CLAIM, STEAL, ABDUCT, and ARREST which tended to be used in reference to particular stories involving alleged kidnappings.

Table 11 - Verbs appearing after GYPSY (all spellings) by publication type

Tabloids			Broadsheets		
Verb	Raw frequency	Normalised	Verb	Raw frequency	Normalised
claim	11	4.73	claim	4	1.89
steal	10	4.30	steal	4	1.42
accuse	9	3.87			
abduct	7	3.01			
arrest	7	3.01			
vanish	6	2.58			
sell	6	2.58			

- Meanwhile, as the blonde child removed from a Roma family in Ireland has turned out to be biologically theirs, one doubts whether the many, age-old tales of gypsies stealing children to order are nothing more than apocryphal. (Tabloid)
- A gypsy claiming to be the real mum of Maria, the ‘blonde angel’ rescued from a travellers’ camp in Greece, was questioned by police yesterday. (Tabloid)

This is also seen within the most significantly frequent verbs associated with ROMA, shown in Table 12. Words like ACCUSE and STEAL suggest a link relating to criminality, while LIVE, SLEEP, and SETTLE express a settlement dimension.

Table 12 - Verbs appearing after ROMA by publication type

Tabloids			Broadsheets		
Verb	Raw frequency	Normalised	Verb	Raw frequency	Normalised
live	19	8.17	charge	12	5.67
sleep	5	2.15	arrive	11	5.20
accuse	5	2.15	involve	7	3.31
steal	5	2.15	accuse	6	2.84
cause	5	2.15	gather	5	2.36
settle	4	1.72	beg	5	2.36
abduct	3	1.29	search	4	1.89
			suffer	4	1.89
			evict	3	1.42
			persecute	3	1.42
			demolish	3	1.42
			abduct	3	1.42
			hang	3	1.42
			photograph	3	1.42
			settle	3	1.42
			avoid	3	1.42

These themes are echoed by an analysis of R1 collocations for the word ROMA, displayed in Table 13 below. Settlement and kidnapping are again visible, as well as a relatively frequent association in the tabloids with GYPSY and its alternative version GIPSY. The analysis also reveals that mentions of ROMA rather than G?PS? tend to be associated with migration issues as indicated in Table 12 by the references to ROMA MIGRANTS and IMMIGRANTS.

Table 13 - R1 noun collocations for ROMA by publication type

Tabloids			Broadsheets		
Word	Raw frequency	Normalised	Word	Raw frequency	Normalised
Gypsy	41	17.63	couple	64	30.26
community	28	12.04	community	58	27.42
gipsy	27	11.61	family	51	24.11
family	24	10.32	camp	36	17.02
camp	21	9.03	child	32	15.13
couple	18	7.74	woman	23	10.87
migrant	18	7.74	migrant	22	10.40
girl	10	4.30	population	21	9.93
woman	10	4.30	immigrant	16	7.57
population	8	3.44	settlement	15	7.09
group	7	3.01	parent	11	5.20
gang	5	2.15	girl	10	4.73
Slovakians	4	1.72	dialect	5	2.36
parent	4	1.72	Rights	5	2.36
association	3	1.29	boy	5	2.36
mum	3	1.29	gipsy	4	1.89
			representative	4	1.89
			beggar	4	1.89
			youth	4	1.89
			Gypsy	4	1.89

- It is understandable why Roma gypsies would want to escape to a country where even the lowest wages are 10 times those they can earn at home. (Tabloid)
- One of the issues on the agenda is whether Roma families could attempt to escape poverty and discrimination in Romania by heading to the UK. (Broadsheet)
- The Deputy Prime Minister described tensions between Roma immigrants and established communities in Slough and Sheffield as a real dilemma, adding that people find some of their behaviour ‘offensive’ and ‘difficult to accept’. (Broadsheet)

Overall, the similarity in collocations between ROMA and GYPSY suggests that they are used interchangeably, particularly in contexts related to crime and settlement.

6. Conclusions

This report aimed to show how the UK national press talked about the A2 countries of Bulgaria and Romania - as well as the people from those countries - during an important period leading up to the removal of transitional controls. Specifically, it examined what issues were raised in connection with mentions of BULGARIA and ROMANIA

(including their variants), and then explored how these groups were characterised. Using corpus linguistics techniques, the report revealed that the joint portrayal of Romanians and Bulgarians differed from their portrayal separately.

When the two groups appeared together, they were most often described in relation to migration issues. Furthermore, tabloid and broadsheet coverage exhibited anticipation of a certain magnitude of migration from these countries, as well as the need to curb future migration. The variety of figures circulated was particularly striking, given the relative lack of data regarding the actual scale of migration from Romania and Bulgaria since January 2014.

References to Romanians by themselves revealed a strong association with crime, while references to Bulgarians, although less common in the corpus, were concentrated around a particular story of a young girl suspected by authorities of having been kidnapped.

Mentions of the ROMA or GYPSIES were also frequent, and particularly associated with issues of crime and settlement. Interestingly, while the similar findings obtained for these groups suggest that the two terms are used interchangeably, Gypsy appears to be more associated with Romanians, while Roma is a descriptor for Bulgarians. However, this difference may be accounted by conventions of writing style: preference for phrases other than 'Romanian Roma', for example, might be driving this result.

Several findings emerged as particularly key:

Across tabloid and broadsheet publications, mentions of Bulgarians and Romanians appearing together were usually characterised by migration issues. Table 2 shows that when newspapers used the phrases BULGARIAN AND ROMANIAN or ROMANIAN AND BULGARIAN, they were usually describing things or groups of people related to migration. This is particularly striking given that the search strategy did not intentionally seek out any particular issue area.

There was significant discussion about people from these groups coming and working in the UK. As seen in Tables 3 and 4, language related to arrival and employment was linked to mentions of Bulgarians and Romanians. Both tabloid and broadsheet publications reported these anticipated arrivals using a wide range of numbers and language related to quantities as shown in Table 5.

When mentioned by themselves, Romanians were more frequently linked to criminality and economic poverty compared to Bulgaria and Bulgarians. Tables 6 and 7 show that references to gangs, crime, and economic hardship such as sleeping rough were associated with mentions of Romanians. This is in contrast to the language around Bulgarians when they were mentioned by themselves as seen in Tables 8 and 9, which tended to focus on stories involving sport or a particular event involving a Bulgarian girl found to be living with a Roma family in Greece. Both tabloids and broadsheets made more references to Romania and Romanians by themselves compared to Bulgaria and Bulgarians, as seen in Figure 3.

The Roma and Gypsy groups tend to be associated with crime and settlement. Tables 10 to 13 show that words related to criminality as well as a continued presence in the UK appear with mentions of both groups, although migration issues tend to be linked with the Roma rather than Gypsy groups. As shown in Figure 5, when ROMA is used in the context of either specific group, it tends to describe Bulgarians and not Romanians. Meanwhile, GYPSY and its variant GIPSY are more likely to be used to describe Romanians rather than Bulgarians.

References

- Allen, William and Scott Blinder. "Portrayals of Immigrants, Migrants, Asylum Seekers, and Refugees in National British Newspapers, 2010–2012." Migration Observatory Report, University of Oxford, 2013.
- Baker, Paul. *Using Corpora in Discourse Analysis*. London: Bloomsbury Academic, 2006.
- McEnery, Tony and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2011.
- Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

Related material

- Migration Observatory commentary - It's too early to know whether the number of Romanians and Bulgarians will rise or fall in 2014
- Migration Observatory report - Migration in the News: Portrayals of Immigrants, Migrants, Asylum Seekers, and Refugees in National British Newspapers, 2010–2012

Appendix 1: Characteristics of the corpus

Source	Items analysed	Items removed	Total items in Corpus	Total words
Tabloids			2,326	1,143,341
Daily Mail	405	0		
Daily Mirror	356	108		
The Express	483	15		
Daily Star	245	7		
The Sun	837	21		
Broadsheets			2,115	1,703,220
The Guardian	281	1		
The Independent	246	1		
Financial Times	248	0		
The Telegraph	637	9		
The Times	703	13		
			4,411	2,846,561

Appendix 2: Data and methods used

Data Selection

The sample was compiled by accessing the online news archive Nexis and by searching articles published in the five most read tabloids and broadsheets from the 1st of December 2012 to the 1st of December 2013. The search term used was 'ROMANIA! OR BULGARIA!', which enabled us to retrieve articles that contained any of the terms 'Romania', 'Romanian', 'Romanians', or 'Bulgaria', 'Bulgarian', 'Bulgarians' in any part of the text. Naturally, this wide search term retrieved many items from topics as varied as news, business and sport bulletins, or travel advice. This breadth of the sample enabled us to view whether immigration was a salient topic in relation to mentions of these groups. It also helped surface any differences in the way the two groups of people are represented.

The resulting sample, or corpus, contains 4,441 articles totalling 2,846,561 million words. Although duplicates were initially removed by Nexis, some duplicates were still retrieved. They were identified and subsequently removed by pasting all the headlines into Excel and automatically highlighting the cells with identical content that were published on the same date as well as on the same page of the newspaper. Also, an initial analysis revealed that there was a series of 68 articles titled "Your Holiday £" that appeared in The Mirror which reported currency exchange rates. This meant that the key term BULGARIAN collocated most frequently with LEV, the Bulgarian currency, as well as other international currencies. In order to avoid this skewing of findings, these articles were removed from the corpus.

Collocation and Sketch Engine

Sketch Engine is a web-based piece of software designed for lexicographers who are building dictionaries and need to identify how words have different senses or meanings across large amounts of language. It does this by creating Word Sketches, or snapshots of how a given word functions as well as with which other words it tends to appear.

Part of this understanding comes from analysing collocations, or the greater likelihood of two words appearing together more than by chance, through statistical measurement. In this report, two statistical tests were used: Mutual Information (MI) Score and log-likelihood (LL). Results were determined to be significant if they had a MI score of at least 5.0 and log-likelihood score of at least 6.67 (Allen and Blinder 2013). Another key part of this understanding comes from grammar, or the rules that govern how words relate to each other in a sentence. The Sketch Engine enables users to use both of these techniques by attaching a word's grammatical function to itself. This identifies whether it is a noun, verb, adjective, numeral, etc., in a process called 'part of speech (POS) tagging'. Therefore, using the Sketch Engine, a user can efficiently identify all of the words that modify a target word because they will be POS-tagged as 'adjectives'. Furthermore, the user can display all of these instances using a concordance view, which shows how a given collocation appears in the original corpus by displaying some of the surrounding text.

However, although this Word Sketch function works with single words, it does not allow for phrases such as BULGARIAN AND ROMANIAN. In the course of completing this analysis, the Migration Observatory found that simply sketching 'Romanian' as an adjective was not necessarily answering the question "what things or objects are described as only Romanian?" This is because some of the collocation results were derived from examples in the corpus where the phrase 'Bulgarian and Romanian' had actually occurred - rather than simply 'Romanian'. Therefore, when analysing how these groups were portrayed when they appeared together in news coverage, the Sketch Engine filtered the corpus to display all instances where ROMANIAN and BULGARIAN appeared within ten words of each other. In the parts of the report where one group is analysed separately from the other, the Sketch Engine filtered the sample to show only those instances where one group was mentioned while the other was not present within a ten word window. Then, the collocation results were manually examined using concordance analysis to check which part of speech they contained.

Finally, the Sketch Engine enables users to sort concordance lines by how well they exemplify a 'typical' collocation. It does this by using a ranking system called GDEX that emphasises three broad qualities in candidate sentences: (1) how well it reflects frequent usage that is observed in the corpus; (2) how well it informs a reader; and (3) how intelligible and readable it is for a user without a large amount of surrounding context. Full details of how these characteristics are operationalised in the ranking system contained with the Sketch Engine can be found in Kilgarriff et al. (2008). This function, originally aimed at dictionary-writers who need to search their corpus for sentences to illustrate a definition, is also helpful for choosing examples of the quantitative collocation results. Although the concordance examples in this report are highly ranked by GDEX, they are only provided as illustrations of the statistical and grammatical collocation results presented in the tables.



The Migration Observatory

Based at the Centre on Migration, Policy and Society (COMPAS) at the University of Oxford, the Migration Observatory provides independent, authoritative, evidence-based analysis of data on migration and migrants in the UK, to inform media, public and policy debates, and to generate high quality research on international migration and public policy issues. The Observatory's analysis involves experts from a wide range of disciplines and departments at the University of Oxford.



COMPAS

The Migration Observatory is based at the ESRC Centre on Migration, Policy and Society (COMPAS) at the University of Oxford. The mission of COMPAS is to conduct high quality research in order to develop theory and knowledge, inform policy-making and public debate, and engage users of research within the field of migration.

www.compas.ox.ac.uk

About the authors

William Allen
Research Officer, Migration
Observatory
william.allen@compas.ox.ac.uk

Dora-Olivia Vicol
PRS Doctoral Researcher, School of
Anthropology and Museum Ethnography
dora-olivia.vicol@sant.ox.ac.uk

Press contact

Rob McNeil
Head of Media and Communications
robert.mcneil@compas.ox.ac.uk
+ 44 (0)1865 274568
+ 44 (0)7500 970081

Suggested citation

Vicol, Dora-Olivia and William Allen. "Bulgarians and Romanians in the British National Press: 1 Dec 2012 - 1 Dec 2013." Migration Observatory report, COMPAS, University of Oxford, August 2014.

