REPORT

Migration in the News:

Portrayals of Immigrants, Migrants, Asylum Seekers and Refugees in National British Newspapers, 2010–2012

COMPAS

UNIVERSITY OF OXFORD

www.migrationobservatory.ox.ac.uk

Executive summary

Britain's national newspapers play a critical role in framing the country's discourse on immigration. Developing a clearer understanding of the language that these newspapers use to discuss migrants and migration can therefore provide us with important insights into the nature of this debate and the role of newspapers in it.

This report capitalises on new methods for analysing what is commonly known as 'big data' and provides a quantitative analysis of the language used by all 20 of Britain's main national daily and Sunday newspapers. It covers all news stories, letters and other published content dealing with migrants and migration over the last three years – from the beginning of 2010 to the end of 2012. This analysis has involved computer-aided analysis of a 'corpus' of some 58,000 news stories and other newspaper items, made-up of more than 43 million words, from stories which include key terms such as MIGRANTS, IMMIGRANTS, REFUGEES, ASYLUM SEEKERS or variations of those words.

The study has endeavored to remove human bias from the analysis – as far as possible – by using modern computing techniques rather than human readers and coders to identify significant patterns. The analysis focuses on describing the findings rather than drawing conclusions about debates on immigration, politics or media practices.

The report is not designed to highlight good or bad practice by specific news outlets or journalists, and for this reason divides the newspapers into groups rather than identifying individual newspapers.

- The broadsheet group covers *The Times* and *The Sunday Times*; *The Daily Telegraph* and *The Sunday Telegraph*; *The Independent* and *The Independent on Sunday*; *The Guardian* and *The Observer* and *The Financial Times*.
- The mid-market group covers *The Daily Mail* and *The Mail on Sunday* and *The Daily Express* and *The Sunday Express.*
- The Tabloid group covers *The Sun* and *The Sun on Sunday* (*The News of the World* is not included for methodological reasons), *The Daily Mirror* and *The Sunday Mirror, The People*, *The Daily Star* and *Daily Star Sunday*.

The report looks at newspapers' language of migration by examining the words that are particularly likely to appear in close proximity to the four key target words listed above: MIGRANTS, IMMIGRANTS, REFUGEES and ASYLUM SEEKERS. The study examines two types of relationships:

- Descriptors or "L1 collocates" - the word located immediately to the left of the target word.
- Consistent collocates or "C-collocates" – words that regularly appear within five words (to the left or right) of the target word, consistently over the three-year study period.

Some key results refer to the L1 collocates, or words used to describe different migrant groups. ILLEGAL was the most common modifier of 'immigrants' throughout the 43 million word corpus, for example, while FAILED was the most common modifier of 'asylum seekers'. These associations may not be surprising, but the methods used in this report document not only that these uses exist but how frequent they are.

Other results come from broader sets of associations. Coverage of migration and asylum includes the vocabulary of numbers (with words like THOUSANDS and even MILLION), discourses of security or legality (words like TERRORIST and SUSPECTED) and language of vulnerability (CHILD, DESTITUTE, VULERNABLE). Further details on these and other findings are presented below.

The results also show that newspapers use different, though overlapping, vocabularies in connection with each of the four groups of interest. For example, migrants, rather than immigrants, are more frequently associated

with economic words (JOBS, BENEFITS, ECONOMIC). Refugees seem to attract a separate, varied and heavily international set of terms (FLEEING, CAMP, BORDER); the language around asylum seekers has more in common with discussions of immigrants and migrants than with refugees.

Key findings:

- The most common descriptor for the word IMMIGRANTS across all newspaper types is ILLEGAL, which was used in 10% of mid market stories, 6.6% of tabloid stories and 5% of broadsheet stories.
- Other descriptors of immigrants refer to their place of origin, with EU and Eastern Europe featuring in all three types of newspapers.
- Consistent collocates for IMMIGRANTS reveal a focus on numbers, with words like MILLION and THOUSANDS appearing across all newspaper types.
- Other consistent collocates for IMMIGRANTS in tabloids include words referring to movement such as INTO, STAY and STOP and also include words which indicate concerns around security or legality such as TERRORIST, SUSPECTED and SHAM.
- Consistent collocates for IMMIGRANTS in broadsheets include family words such as SON (also a mid-market c-collocate), CHILDREN and DAUGHTER and also religion with JEWISH and MUSLIM both featured.
- Consistent collocates for MIGRANTS across all newspaper types include ECONOMIC. Other words that relate to the employment and unemployment such as JOBS and BENEFITS appeared in tabloids and midmarkets. JOBS also appears in the broadsheets, but less frequently. This contrasts with the collocates of IMMIGRANTS, which included very few words relating to economics and work.
- Words suggesting water as a metaphor for migration, such as FLOOD, INFLUX and WAVE are c-collocates of both MIGRANTS and IMMIGRANTS. INFLUX was most widely used, but WAVE appeared as a c-collocate for IMMIGRANTS in both tabloids and broadsheets and tabloids also used FLOOD in conjunction with MIGRANTS.
- FAILED is the most common descriptor for ASYLUM SEEKERS across all newspaper types. ILLEGAL is also a descriptor in both mid-market and broadsheet newspapers. Other descriptors in broadsheets suggest vulnerability – such as CHILD, DESTITUTE and VULNERABLE.
- Some c-collocates for ASYLUM SEEKERS in midmarket newspapers focused on illegality and permanence, including ILLEGAL, CRIMINALS and STAY. Broadsheets also consistently used ILLEGAL and CRIMINALS, albeit at a lower frequency and among a larger set of c-collocates.
- Language around REFUGEES was strikingly different to other target words, even ASYLUM SEEKERS. Both descriptors and consistent collocates of REFUGEES focused on conflict and fleeing and on refugees' nationalities. Broadsheets had considerably more descriptors and consistent collocates for REFUGEES than mid-market or tabloid newspapers.

This report is the first output from the Migration Observatory's media analysis project and as such is not intended to deal with every question about how British national newspapers have covered the topic of migration. Further study of the corpus - and updated versions of it – should help to create a more detailed understanding of the nature of newspaper coverage of migration, which in turn, may shed more light on the complicated nexus of media, policy and public opinion.

The report is, however, intended to provide both useful insights into the language used and the ways that different types of newspaper approach the subject of immigration, and to provide a bed of evidence for further social science investigations into the subject of migration in the media.

Introduction

How do you know what you know about immigration? In Britain, where 13% of the population was born abroad, immigration and migrants are part of day-to-day life for many. As a prominent issue in political debate, migration is often in the news, and is discussed in countless neighbourhoods and workplaces as well.

Immigration is a large-scale and complex phenomenon and a similarly complex array of factors may shape our impressions of it. As much as we might learn from direct experience and from our own everyday interactions, no one can possibly take in the phenomenon of 'immigration' as a whole in this way. Beyond everyday experiences, many of us get information about immigration and asylum in a more indirect manner through various forms of media. Some seek out information by reading news stories and opinion pieces about immigration, or deliberately finding information on television, radio, or the internet. Others may encounter information in a more haphazard way by leafing through a newspaper or surfing TV channels or websites—perhaps even by reading the sports pages or book reviews and chancing upon a mention of a foreign-born athlete winning a gold medal for Britain or a new film telling a story of migrants from distant places or of World War II refugees.

Portrayals of migrants and refugees are plentiful in today's media-rich environment, whether encountered accidentally or deliberately. Either way, these portrayals may well have a powerful role in shaping how members of the British public understand migration and asylum. This is, of course, not unique to immigration: members of the public rely at least in part on newspapers and other media to inform them about all sorts of issues, from international conflicts or the national economy to football results and celebrity gossip. Britain's national newspapers in particular often set the agenda for the country's political discourse, both informing their readers about issues of the day and at times guiding them toward certain ways of thinking about these issues.

This report takes one of those issues—immigration—and scrutinises the language used by Britain's national newspapers in covering the topic. To do this, the Migration Observatory has created a comprehensive collection, or 'corpus', of all national newspaper articles dealing with migration over three years, adding up to some 43 million words. Until recently, such a volume of texts would have been impossible to analyse as a whole; however, with the rise of so-called Big Data techniques, the analysis of full data sets with millions of elements has become possible. This project has made use of modern computer-based techniques to conduct a quantitative analysis of all 43 million words used in this newly-assembled database of news articles, identifying patterns in the use of these words and the relationships among them.

The study does not evaluate claims about media influence on attitudes or perceptions. It does, however, set the stage for research to test such claims, by providing rigorous analysis of the language that the British press has actually used to discuss migration, rather than relying on the potentially partial impressions of newspapers that individual readers, including researchers, may develop. Before one can examine the impact of media coverage of immigration, it is essential to get a better grip on what coverage actually looks like.

The goal of this study, instead, is to systematically and comprehensively examine how British national newspapers have portrayed immigrants, migrants, asylum seekers and refugees in recent years. When immigrants are mentioned, how are they likely to be described? Which words tend to be used in conjunction with asylum seekers, and how do these words differ from the language used to discuss refugees? When British newspapers discuss the origins of immigrants, are they likely to focus on Europe, Asia, or elsewhere? Most generally, what are the recurrent words and themes that come up when migrants are mentioned in newspaper coverage?

Crucially, this study addresses these questions in a comprehensive rather than selective way. Rather than examining coverage of prominent, high-profile events or focusing on particular newspapers, the University of Oxford Migration News Corpus includes, as far as possible, all national newspaper articles that mention immigrants, migrants, refugees and asylum seekers over the past three years (2010-2012). Then, the study's quantitative

analysis treats the corpus as a large database. Using computer software to sift through the entire corpus, the study systematically searches for patterns in the words that newspapers use alongside mentions of these four groups. These methods identify regularly occurring patterns by statistical tests rather than by human perception, which can introduce errors or biases. Systematic assessment of the media environment around immigration at this scale has rarely been attempted, but a notable exception—the work conducted by linguists at Lancaster University on press coverage from 1996 to 2005 (Baker et al. 2008)—shows that this type of comprehensive study is achievable.

The results reveal patterns in the language that newspaper use when talking about immigrants, migrants, asylum seekers and refugees. These patterns emerge from the data through the use of corpus linguistic techniques detailed below (and some basic decisions about procedures and statistical tests). They were not based in any way on researchers' prior expectations or theories. Some of the notable results refer to the words used to describe immigrants, migrants, asylum seekers and refugees. It may not be surprising that ILLEGAL was the most common modifier of "immigrants" throughout the 43 million word corpus, or that FAILED was the most common modifier of "asylum seekers," but the methods used in this report document not only that these uses exist but how frequent they are.

Other results come from broader sets of associations – for example, coverage of migration and asylum includes the vocabulary of numbers (with words like THOUSANDS and even MILLION), discourses of security or legality (words like TERRORIST, SUSPECTED and SHAM) and language of vulnerability (CHILD, DESTITUTE, VULNERNABLE). The results also showed that newspapers use different, though overlapping, vocabularies in connection with each of the four groups of interest. For example, migrants, rather than immigrants, are more frequently associated with economic words (JOBS, BENEFITS, ECONOMIC). Refugees seem to attract a separate, varied and heavily international set of terms (FLEEING, CAMP, BORDER); the language around asylum seekers has more in common with discussions of immigrants and migrants than with refugees.

Importantly, the research was designed to assess the overall news environment, not to single out particular newspapers for praise or criticism. The results show separate trends for the main types of newspapers (tabloid, mid-market and broadsheet), but not for individual newspapers. Also, the findings are strictly descriptive; the analysis does not try to assess news reporting against some notion of how journalists should be covering migration. The report does not even assess whether coverage is favourable or unfavourable toward migration. Rather, the results refer to patterns in the way words are used—more precisely, the word choices that tend to regularly accompany mentions of migrant groups. The goals are to catalogue and describe the language used in British national newspapers to talk about migration and asylum and as far as possible to allow the patterns that emerge from the data to speak for themselves.

Data: The corpus

The report is based on an analysis of a collection (or 'corpus') of news articles and other newspaper items assembled specifically for this research project. The University of Oxford Migration News Corpus is made up of over 58,000 newspaper items containing about 43 million words pieces appearing in 20 national UK publications from 1 January 2010 to 31 December 2012. The corpus captures, as far as possible, every mention of the terms IMMIGRANTS, MIGRANTS, REFGUEES and ASYLUM, as well as the related words and phrases DEPORTATION and ILLEGAL ALIEN, in the selected newspapers' coverage.

Table 1. National UK Tabloid and broadsheet titles included in the study

Tabloids	Mid-markets	Broadsheets
The Sun, The Sun on Sunday	The Express The Sunday Express	The Times The Sunday Times
Daily Mirror Sunday Mirror	The Daily Mail The Mail on Sunday	The Guardian The Observer
Daily Star Daily Star Sunday		The Independent The Independent on Sunday
The People		The Daily Telegraph The Sunday Telegraph
		The Financial Times

The newspapers selected appear in Table 1 according to categories based on the model of 'popular, mid-market, quality' publications as used by the Audit Bureau of Circulations (ABC). The list includes only national British publications. It does not include *The News of the World* but does include the *The Sun on Sunday*, for practical reasons explained in the next section. Table A1 in Appendix A shows details on the size of each section of the corpus.

The corpus of newspaper items was created by retrieving articles and other newspaper items through NexisUK, a service that archives newspapers and periodicals. The search included all types of newspaper items—not only news articles but also sports articles, theatre reviews, editorials and letters to the editor. All of these contribute to the information available to readers about immigrants, migrants, asylum seekers and refugees and therefore were considered relevant to the study. Items were included in the corpus if they contained any of the following key words:IMMIGRANTS, MIGRANTS, REFUGEES, ASYLUM, DEPORTATION, ILLEGAL ALIEN. Variations of these words were also included (e.g. 'immigration', 'deported'). Some exclusions were made to avoid confusion with non-migration uses of these terms: for example, 'deportment' refers to etiquette, 'Deportivo' is a Spanish football club; 'migration' often refers to animal movement. The results were filtered to exclude duplicates ('highly similar items' as determined by NexisUK). This approach follows previous work conducted at Lancaster University that investigated how migrant groups were portrayed in British media from 1996-2005 (e.g. Baker et al. 2008, Gabrielatos and Baker 2008).

Strategy for analysis: Which words appear with 'immigrants'?

To identify how newspapers have consistently described migrant groups, this study uses a 'corpus linguistic' approach. Corpus linguists use software with automated means for analysing texts and empirically documenting characteristics of words—for example, how often they appear (McEnery and Hardie 2012). This study uses corpus linguistic methods to examine which words were most commonly used in conjunction with the words IMMIGRANTS, MIGRANTS, ASYLUM SEEKERS, and REFUGEES. These four words were used as the 'target' words; the analysis finds collocates of each of these four.

Two related reasons led to this choice of approach. First, corpus methods can analyse very large amounts of text; second, they minimise (and make transparent and replicable) the input from researchers' judgements. Instead of enlisting people to read and interpret ('hand-code') a large number of items, corpus linguistic techniques use computer-assisted techniques to scour large sets of texts and detect patterns in language that might be unnoticed in even close readings of texts by humans. Computerised methods can count and classify enormous numbers of words. Moreover, they do so consistently, without introducing human error or fatigue. Researchers need to make decisions only about procedures to follow, and do not make judgements about the meaning or tone of particular words, phrases, sentences, or articles.

One such researcher decision was to focus the analysis on 'collocations': words that are particularly likely to appear together. In other words, collocates appear near one another across a number of texts, more often than would happen by mere chance (Stubbs 1995, Sinclair 1991). By looking at the words that are especially likely to appear with a target word, researchers can detect patterns of language use that convey how a word of interest is used. When a target word is MIGRANTS, for example, collocations indicate the words and concepts that are associated with migration in newspapers' writings. Collocation thus offers "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).

Sometimes, collocations show how a given word is used, suggesting something about the word's meaning or likely context. For instance, in British English the noun CHEESE often appears with words that indicate different *types* of cheese, such as CHEDDAR and PARMESAN (McEnery and Hardie 2012). Collocates convey additional information about the word CHEESE – that it can come in many types. Sometimes, collocations can reveal less obvious information, such as the phrase TO BREAK OUT. This phrase TO BREAK OUT is often collocated with words like INFECTION or DISEASE (Sinclair, 1991: 70), suggesting a common use that is not obvious from the literal meaning of BREAK.

To ensure that collocations are real, consistent patterns, rather than coming from a particular researcher's intuition or from an atypical usage, corpus linguistics assesses them in a large body of text, using statistical tests for significant associations. This process allows researchers to quantify the strength of a relationship between two words (Hunston 2007, McEnery and Hardie 2012).

How do these statistical tests work? Begin from the assumption that a corpus of texts is a collection of randomly distributed words – sometimes called a 'bag of words' approach. In a truly random array, like words jumbled together in a sack, there would be no observable patterns of co-occurrence: any word could appear alongside any other word. However, language in the real world does not work this way. Rather, words often appear together in recognisable patterns in order to convey meaning. Statistical tests identify whether words appear together more often than one would expect by sheer chance as if they were being pulled out randomly from a bag of words. The precise tests used for this report appear in Appendix B.

A further step is to focus on consistent collocates, or 'c-collocates', to identify only collocations that appeared in each separate year of coverage. In this study, collocates were considered consistent only if they were associated with one or more of the four target words (IMMIGRANTS, MIGRANTS, ASYLUM SEEKERS, REFUGEES) in each of the three years taken separately. A word collocated in 2010 and 2011 would not be labelled as a c-collocate. Taking this step prevents single events from distorting the overall picture from the drumbeat of steady coverage (Gabrielatos and Baker 2008). Short-term changes in coverage might be interesting in their own right, such as coverage of nationality and ethnicity in relation to the 2012 London Olympics. C-collocates, however, better show how different newspapers portrayed migrant groups consistently across the entire 2010-2012 period. For this reason, the News of the World was not included in the data set, since it stopped publishing during this period. It could not be combined with the Sun on Sunday because the latter is maintained in the NexisUK database as a part of The Sun rather than as a separate title.

This study shows the results of analysing collocations at two different levels. The first examines words appearing within five words of the target word in either direction. This level of analysis is suggestive of a general association between words. In the case of this report, collocations with IMMIGRANTS, MIGRANTS, REFUGEES, and ASYLUM SEEKERS indicate the language that consistently appears when these groups are mentioned. Prior work suggests an optimum window of ten words, for a balance between capturing enough words for meaningful analysis and limiting the results to reflect meaningful relationships (Baker, Gabrielatos, and McEnery 2013). Of course, the choice of five words on either side is arbitrary: an analysis could legitimately use a window of four or six words on either side.

Figure 1. Collocate positions in relation to a target word

L5	L4	L3	L2	L1	TARGET	R1	R2	R3	R4	R5

The second level focuses exclusively on the 'L1 collocate', the word appearing immediately before the word of interest. Figure 1 shows how these two forms of analysis relate to one another. The ten word window considers any word appearing in the depicted range as a collocate of the target word, which in this case is TARGET. Analysis of the L1 collocate, meanwhile, is confined to words appearing in the position immediately before the target word, depicted as L1 in Figure 1. Focusing on the L1 position hones in on words used to directly describe immigrants, migrants, refugees and asylum seekers. As the analysis below demonstrates, the L1 collocate will often be an adjective modifying the word of interest, as in '*economic* migrants' or '*Iraqi* refugees'. In a large corpus with millions of words, collocates in the L1 position can show patterns in the words used to describe immigrants, migrants, refugees and asylum seekers. Taken together, these two levels of collocation analysis provide considerable indication of the language that British newspapers have used in the last three years in conjunction with immigrants, migrants, refugees and asylum seekers.

Of course, sole reliance on quantitative identification of c-collocates in the ten-word window may still miss the context of the larger sentence. Equally, analysis of collocations in this study does not account for negation (the use of 'not' to reverse the meaning of a sentence) as well as the attribution of a claim (whether a claim is merely reported, or actually stated by an author). For example, take the following set of hypothetical sentences:

> We should **deport** all **immigrants**.
>
> Some people say we should **deport** all **immigrants**.
>
> We should not **deport** all **immigrants**.
>
> Some people say we should not **deport** all **immigrants**.

Clearly, these sentences convey entirely different meanings. In an analysis like the one conducted in this study, the word DEPORT would be collocated with IMMIGRANTS given this particular set of text, indicating some sort of association between these two concepts. However, this finding would not reveal what sorts of claims are being made when these words are used together. To say more about the different ways these words can be used together, and the broader meanings generated, would require further analysis of a different type.

The following sections show results from the analysis of collocations, using both the ten-word window and the narrower L1 collocate list. Only plural forms of these target words were considered (e.g., 'immigrants' and not 'immigrant'), allowing a focus on the portrayal of groups rather than individuals. Throughout, words displayed in ALL CAPITALS refer to an example from the texts. The normalised count is the number of appearances per 1000 items. The normalised count, as opposed to the raw number of mentions of a word, allows for a better comparison across tabloids, mid-markets and broadsheets.

Usually, when documenting how many times a word or collocation occurs in a body of texts, corpus linguists will normalise their results by computing the number of mentions per thousand or per million words of text (McEnery and Hardie 2012). This allows comparison between 'bags of words' containing different amounts of text. However, since mid-markets and broadsheets tend to contain longer articles than tabloids, focusing on appearances per million words would be misleading in another sense by exaggerating the frequency of words in tabloids relative to broadsheets. More details on this decision are available in Appendix B.

To illustrate some of the main findings, examples from 2012 coverage were then selected. To select as impartially as possible, these examples were chosen from among the top scorers generated by an algorithm, called GDEX, used for choosing example sentences for dictionaries. GDEX (short for 'good example') ranks sentences using criteria such as the length of the sentence, the rarity of the other words in the sentence and whether the target word is the main subject of the sentence. More information on this system appears in Appendix B as well as in work by Kilgarriff and colleagues (2008). Critically, these examples are used only for illustration. They should not be taken as representative of uses in the larger corpus as a whole, and the choice of examples does not affect the quantitative results of the research.

Results: IMMIGRANTS

The following sections present results from the analysis of collocations of four key target words, each of which marks people crossing international borders: IMMIGRANTS, MIGRANTS, ASYLUM SEEKERS, and REFUGEES. Each word (or phrase) appears many times in the University of Oxford Migration News Corpus—not surprisingly, since these were among the terms that marked articles for inclusion in the corpus. The analysis reveals numerous, complex patterns in the words used in conjunction with each of these target words, with both similarities and differences across the four words, and across the different categories of newspapers represented in the corpus.

The analysis begins with the c-collocates of IMMIGRANTS. First, focusing on L1 collocates—the words appearing consistently just before the word IMMIGRANTS—presents an extremely clear finding. The most common L1 collocate by far was ILLEGAL, as shown in Table 2. This was true across all three types of newspapers, but most frequent in the mid-markets. The normalised rate of 98.96 means that the phrase ILLEGAL IMMIGRANTS appears in almost 10% of mid-market items in the corpus (99 of every 1000), compared with 5% in broadsheets and 6.6% in tabloids.

Examples of this usage rated as 'good examples' by GDEX include portrayals of 'illegal immigrants' as problems, whether in general or in reference to particular areas of enforcement or detection. These examples come from tabloids; (see below for examples from mid-markets and broadsheets):

Table 2. L1 collocates of IMMIGRANTS, 2010–2012

Tabloids		Mid-markets		Broadsheets	
L1 collocate	Normalised	L1 collocate	Normalised	L1 collocate	Normalised
ILLEGAL	66.46	ILLEGAL	98.96	ILLEGAL	49.59
EASTERN EUROPEAN	1.40	EASTERN EUROPEAN	3.59	JEWISH	5.92
EU	0.70	IRISH	3.40	EUROPEAN	4.38
		EU	2.36	AFRICAN	2.67
		NON-EU	2.17	IRISH	2.48
		AFRICAN	1.89	MUSLIM	2.29
		POLISH	1.70	UNDOCUMENTED	2.26
				RECENT	1.87
				ITALIAN	1.87
				SKILLED	1.41
				ASIAN	1.21
				RUSSIAN	1.16
				NON-EU	1.16
				POLISH	1.16
				TURKISH	1.07
				EASTERN EUROPEAN	1.02
				MEXICAN	0.96
				LEGAL	0.88
				CARIBBEAN	0.74
				POOR	0.63

> The issue of **illegal** immigrants has been a problem for years. (Tabloid)
>
> About 20 tenants squeezed into outbuildings were found to be **illegal** immigrants. (Tabloid)

Within tabloids in the sample, the only other consistent L1 collocates referred to the European context for immigration. These were EASTERN EUROPEAN and EU:

> Some people are pretty upset about the flood of **Eastern European** immigrants into this Country. (Tabloid)
>
> Raising enough cash for a bed is now even harder for homeless Scots, following the massive rise in **Eastern European** immigrants. (Tabloid)

In a notable pattern, mid-market and broadsheet newspapers shared most or all of the tabloids' collocates, but added others as well. ILLEGAL remained at the top of the list for mid-markets and broadsheets. Examples from these publication types include references to difficulties in detection and enforcement, as well as to court rulings:

> Yet the same agency then just writes off 80,000 cases of **illegal** immigrants who never made an application to come here at all. (Mid-market)
>
> Judges and lawyers display great ingenuity in discovering reasons why **illegal** immigrants should be allowed to stay. (Mid-market)
>
> It follows concerns that foreign criminals and other **illegal** immigrants are able to use the right too easily to fight deportation. (Broadsheet)

Broadsheets' use of ILLEGAL IMMIGRANTS also included coverage of the political issues surrounding immigration reform efforts in the US. Furthermore, mid-markets shared the tabloids' other L1 collocates as well, while making additional reference to places or nationalities: IRISH, NON-EU, AFRICAN, and POLISH. Broadsheets also included AFRICAN and IRISH like mid-markets, but also added ITALIAN and ASIAN in addition to religious references like JEWISH and MUSLIM. Broadsheets also included the descriptors SKILLED and RECENT, as well as UNDOCUMENTED.

The word UNDOCUMENTED, sometimes proposed as an alternative to ILLEGAL, arises in the context of US rather than UK migration, as revealed by closer manual inspection of its appearance in the corpus. Broadsheets included significant coverage of debates over immigration reform in the US, where discussion centres on policies toward that country's estimated 11-12 million illegal or undocumented immigrants:

> Latinos, who make up about 80% of America's 11 million **undocumented** immigrants, are an increasingly important section of the electorate. (Broadsheet)
>
> There may be more than 12m **undocumented** immigrants in the US, though the number fell in the recession. (Broadsheet)

This points to several likely reasons for the wider set of terms found in broadsheets: first, these newspapers are larger in size. Relatedly, broadsheets on the whole may include more coverage of international news. In addition a larger number of broadsheets are included in the corpus which may add to the greater variety of broadsheet coverage observed. Prior work has also found a wider variety of terms used in broadsheets' coverage of immigration and asylum in earlier (1996-2005) coverage (Baker et al. 2008).

The contrast between ILLEGAL and UNDOCUMENTED also points to the controversial, disputed nature of the language used to describe immigrants who lack legal authorisation to remain in the country in which they reside. The Associated Press recently changed its official house style to recommend against using the phrase 'illegal immigrant',

even after reaffirming its use as recently as October 2012, proposing that 'illegal' should be used to modify actions (i.e. 'illegal immigration') but not people (i.e. 'illegal immigrants'). (The AP does not recommend 'undocumented' as a suitable alternative.)

However, this report examines language as it is actually used, rather than normative debates over what language should be used, and so simply reports on patterns in language use that actually occur. Regardless of recommendations from the AP or elsewhere, it would be impossible to accurately document British national newspapers' coverage of immigration without heavy use of the phrase ILLEGAL IMMIGRANTS.

Table 3 shows the broader view by identifying collocates appearing within the ten word window on either side of IMMIGRANTS rather than just the L1 collocates. ILLEGAL remains at the top of the list for all three types of newspapers. The normalised rates show that the word ILLEGAL appears within five words of IMMIGRANTS in about 10% of mid-market items in the corpus (100 of every 1000). For broadsheets and tabloids, this is closer to 5% and 7% respectively, but still considerably more often than any other c-collocate. The tabloid list, however, also includes the word SHAM, which does not appear as a c-collocate for mid-markets or broadsheets.

Table 3. C-collocates of IMMIGRANTS, 2010–2012

Tabloids		Mid-markets		Broadsheets	
C-collocate	Normalised	C-collocate	Normalised	C-collocate	Normalised
ILLEGAL	67.25	ILLEGAL	100.57	ILLEGAL	51.71
INTO	11.86	BRITAIN	23.42	BRITAIN	7.82
MILLION	7.15	NUMBER	15.20	SON	7.25
NUMBER	6.98	MANY	14.92	CHILDREN	6.58
STAY	6.11	EU	12.65	JEWISH	6.56
EU	5.76	AMNESTY	10.58	NUMBER	6.03
THOUSANDS	5.76	MILLION	10.01	AMNESTY	4.79
COMING	5.23	EASTERN	8.12	EUROPEAN	4.38
STOP	4.27	THOUSANDS	8.12	GENERATION	3.80
SEEKERS	3.14	EUROPE	7.84	EASTERN	3.66
EASTERN	2.70	BENEFITS	6.99	AFRICAN	3.55
TERRORISTS	2.70	SEEKERS	6.70	MILLION	3.53
WAVE	2.35	JOBS	6.52	THOUSANDS	3.31
SUSPECTED	2.27	INFLUX	5.57	BORN	3.28
ARRIVED	2.09	NUMBERS	5.57	IRISH	3.25
HOUSES	2.09	SON	4.91	RECENT	2.76
INFLUX	2.09	COUNTRIES	4.72	MUSLIM	2.73
HOUSING	1.57	NON	3.78	JOBS	2.73
SHAM	1.57	ARRIVED	3.68	DAUGHTER	2.64
		IRISH	3.68	RUSSIAN	2.53

Figure 2, drawn from the online tool accompanying this report, compares the most frequent c-collocates across the paper types. This comparison reveals new similarities. Evidence of the focus on numbers in the migration debate appear across all three types: tabloids and mid-markets tended to use MILLION, NUMBER and THOUSANDS, while broadsheets also used these words in slightly lesser degrees:

> The Commons Home Affairs Select Committee found a 'shocking' lack of supervision led to controls being relaxed too frequently, letting in **thousands** of immigrants unchecked. (Tabloid)

> If you are one of the hundreds of **thousands** of illegal immigrants who has managed to enter Britain, however, you will know the truth: that the UK border is just a facade. (Mid-market)

> Security agencies believe **thousands** of illegal immigrants are using the ID cards or passports, which the gangs say are rented from European Union citizens. (Broadsheet)

Europe remained prominent in all three newspaper types. The tabloid list is also suggestive of a distinct discussion of movement, both in terms of allowing as well as preventing movement, as evidenced by words like INTO, STAY, COMING, and STOP:

> Between 2004 and 2007 were the years when our economy peaked, with 300,000 immigrants **coming** into the country. (Tabloid)

> I see that the Government are to allow 60,000 skilled immigrants **into** the country each year. (Tabloid)
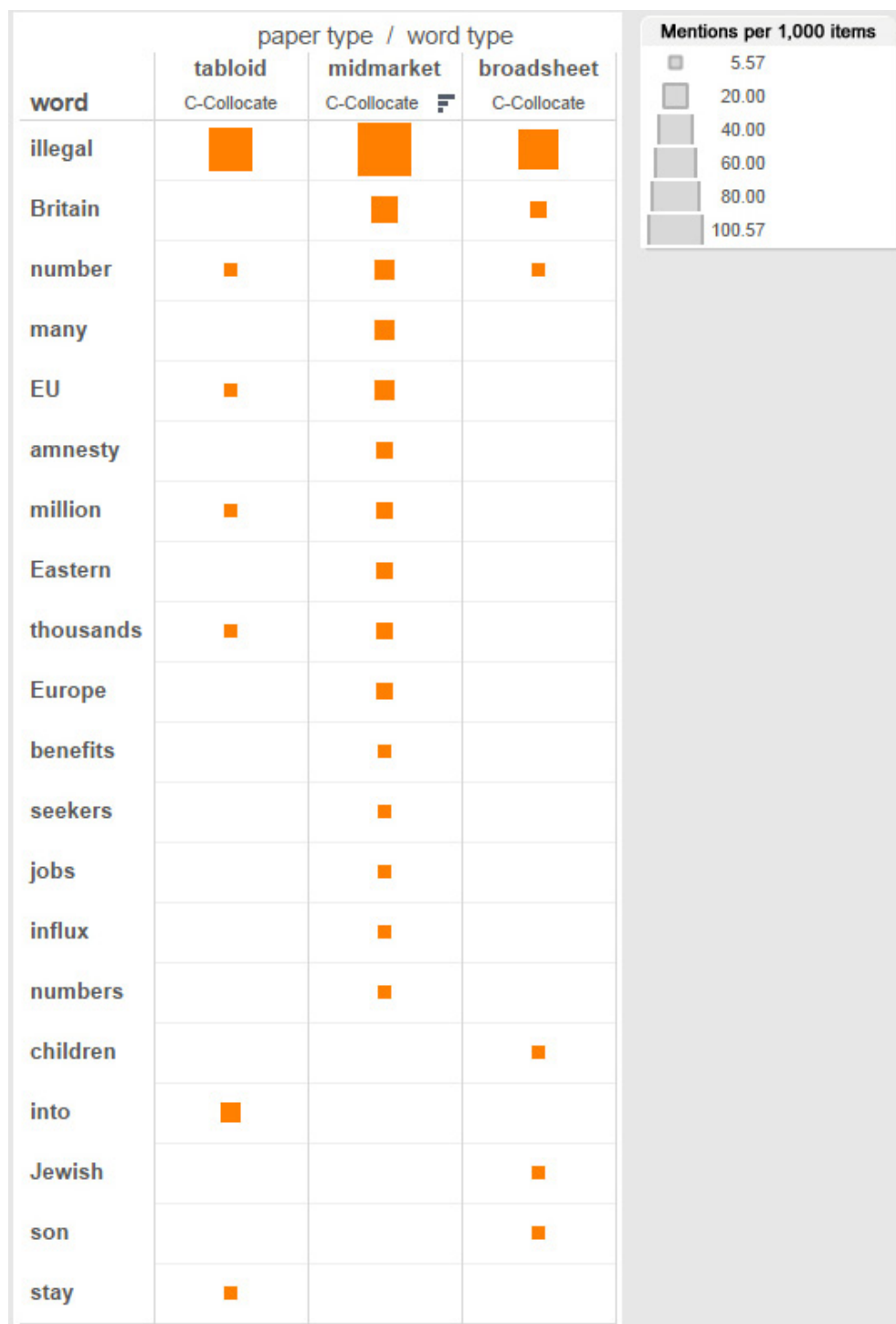
Meanwhile, broadsheets—and to a lesser extent mid-markets—made reference to **family** members of immigrants, particularly when discussing the SON, DAUGHTER, or CHILDREN of an immigrant:

The Labour leader spoke of his own background as the son of immigrants who fled the Nazis and said he was incredibly proud of the multi-ethnic diverse Britain. (Mid-market)

> Rosenberg was the **son** of poor Lithuanian Jewish immigrants, and grew up in the East End. (Broadsheet)

> No sensible official would house the **children** of Nigerian immigrants with a known racist. (Broadsheet)

Figure 2. Top 20 C-collocates of IMMIGRANTS, 2010-2012

word	tabloid C-Collocate	midmarket C-Collocate	broadsheet C-Collocate
illegal	■	■	■
Britain		■	■
number	■	■	■
many		■	
EU	■	■	
amnesty		■	
million	■	■	
Eastern		■	
thousands	■	■	
Europe		■	
benefits		■	
seekers		■	
jobs		■	
influx		■	
numbers		■	
children			■
into	■		
Jewish			■
son			■
stay	■		

Mentions per 1,000 items

□	5.57
■	20.00
■	40.00
■	60.00
■	80.00
■	100.57

Tabloids also mentioned TERRORISTS in the same context as IMMIGRANTS; this was also observed in the mid-markets and broadsheets although it appeared outside of the top twenty collocations at number 26 out of 33 c-collocates in the mid-markets, and 58 out of 78 c-collocates in the broadsheets (see Supporting Information online). Finally, words related to 'water' metaphors of mass entry—such as INFLUX and WAVE—appear in tabloid and mid-market collocations (see Gabrielatos and Baker 2008, Charteris-Black 2006). INFLUX appeared in the broadsheets ranked at number 26, while WAVE appeared at number 38 (again, see Supporting Information).

Other notable words related immigration with particular impacts or policy contexts, including BENEFITS (mid-markets); HOUSES and HOUSING (tabloids); and AMNESTY (mid-markets and broadsheets):

How much immigrants can claim in **benefits** depends on how they came to be in Britain, where they are from and how long they have been here. (Mid-market)

Yesterday, we revealed how jobless immigrants get half of council **houses** in some parts of Britain. (Tabloid)

Johnson, who in 2008 campaigned for an **amnesty** on illegal immigrants and repeated the call during his first term, changed tack as he highlighted the need to 'haul up the drawbridge against illegals' as part of the solution - comments likely to resonate with the Tory grassroots (Broadsheet)

Results: MIGRANTS

As Table 4 shows, the most common L1 collocates for MIGRANTS have a great deal in common with the list for IMMIGRANTS. ILLEGAL is at or near the top for each publication type. Similar to the findings for IMMIGRANTS, mid-market publications used ILLEGAL before MIGRANTS the most often, with tabloids in this case using the phrase the least often.

Table 4. L1 collocates of MIGRANTS, 2010-2012

Tabloids		Mid-markets		Broadsheets	
L1 collocate	*Normalised*	*L1 collocate*	*Normalised*	*L1 collocate*	*Normalised*
NON-EU	4.62	ILLEGAL	15.49	ECONOMIC	7.69
ILLEGAL	4.54	ECONOMIC	13.98	ILLEGAL	6.83
ECONOMIC	3.58	EU	6.80	SKILLED	5.18
EU	2.79	NON-EU	6.52	AFRICAN	3.42
EASTERN EUROPEAN	1.31	SKILLED	4.25	NON-EU	3.14
EUROPEAN	0.52	AFRICAN	3.59	EU	1.49
		EASTERN EUROPEAN	3.49	EASTERN EUROPEAN	1.38
				POLISH	0.77
				UNDOCUMENTED	0.58
				EUROPEAN	0.50
				IRREGULAR	0.47
				ASIAN	0.33
				TEMPORARY	0.19
				URBAN	0.17

Notably, all of the publication types used ILLEGAL IMMIGRANTS more often than ILLEGAL MIGRANTS. In fact, comparing Table 3 and Table 4 reveals that the former was about 14 times more common than the latter in the case of tabloids, six times more common with mid-market publications and seven times more common in broadsheets.

Meanwhile, in addition to the tabloid L1 collocates, the mid-markets and broadsheets also mentioned different kinds of migrants, including ECONOMIC, SKILLED, TEMPORARY and URBAN. The appearance of economic and work-related terms is notable here, as it suggests a different context for MIGRANTS than for IMMIGRANTS, where economic terms did not appear as collocates. Inspection of example sentences shows that 'economic migrants' can

be used in a variety of ways. In mid-markets, some usages differentiate economic migrants as less worthy of entry than other categories such as asylum seekers or students:

> They are **economic** migrants not asylum seekers. (Mid-market)

> In recent years however, it's clear a substantial number – probably tens of thousands a year are **economic** migrants seeking a back door into the UK. (Mid-market)

Other portrayals of economic and skilled migrants link these groups with valued and needed talents:

> They make an enormous contribution to society and, in the case of talented **economic** migrants, bring skills which will be essential in restoring the country to financial health. (Mid-market)

> Both temporary **skilled** migrants and genuine students are vital for the UK economy, but the UK Borders Agency needs to be better resourced to track them coming in, keep track of them while they're here and, importantly, record when they leave. (Broadsheet)

Broadsheet coverage also referenced economic migrants alongside challenges for immigration control, particularly the issue of ascertaining the reason for migration:

> The designation was introduced to weed out bogus colleges set up solely to sponsor **economic** migrants. (Broadsheet)

> This resulted in a huge influx of **economic** migrants, many claiming to be political refugees, initially settling in Germany but eventually in the UK. (Broadsheet)

Broadsheets' collocates also referred to more geographic origins of migrants, including AFRICAN, POLISH, and ASIAN. Finally, IRREGULAR as an alternative to ILLEGAL appears only in broadsheet coverage as a consistent L1 collocate of MIGRANTS. However, even there it appears less than once per 2000 items (0.47 mentions per 1000 items).

Expanding to the ten-word window shows prominent mentions of the British or European context (as evidenced by words like EU, EUROPEAN, UK and BRITAIN) alongside references to MIGRANTS across the three publication types, as shown in Figure 3. Also, broadsheets and mid-markets reference quantities of MIGRANTS, for example by using NUMBER and THOUSANDS, with broadsheets using THOUSANDS slightly less frequently.

Figure 3. Top 20 C-collocates of MIGRANTS, 2010-2012

Table 5 further shows that tabloids and mid-markets also mention JOBS and BENEFITS in the context of MIGRANTS:

> Home Secretary Theresa May has binned the original version drawn up under Labour, which taught migrants about **benefits** and human rights laws. (Tabloid)

> There will be fresh rows about Britons losing **jobs** to migrants – and more pressure on services. (Mid-market)

Like the other publication types, broadsheets also mention JOBS although at a lower frequency. CAP also appears alongside broadsheet mentions of MIGRANTS. Manual inspection of those instances reveals that CAP consistently occurred in discussion of policy proposals to place a numerical limit on non-European migrants coming as skilled workers with job offers.

A **cap** on skilled migrants from non-EU countries has not yet bitten in the sluggish economic climate. (Broadsheet)

Ms May is expected to reassure vice-chancellors in the speech today that there will be no **cap** on student migrants. (Broadsheet)

Table 5. C-collocates of MIGRANTS, 2010-2012

Taboids		Mid-markets		Broadsheets	
C-collocate	*Normalised*	*C-collocate*	*Normalised*	*C-collocate*	*Normalised*
EU	10.82	EU	25.31	EU	8.90
BRITAIN	7.24	BRITAIN	22.10	ECONOMIC	8.35
NON	6.54	ILLEGAL	17.00	ILLEGAL	7.58
NUMBER	6.28	UK	16.53	SKILLED	6.47
ILLEGAL	5.15	ECONOMIC	14.83	BRITAIN	5.92
JOBS	4.36	NUMBER	12.46	NUMBER	5.37
HERE	4.10	JOBS	10.20	NON	5.26
ECONOMIC	3.92	BENEFITS	9.63	EUROPEAN	4.27
EUROPEAN	3.05	NON	9.25	AFRICAN	3.94
BENEFITS	2.97	THOUSANDS	9.16	EUROPE	3.91
BRITS	2.88	EUROPE	7.74	EASTERN	3.17
MILLION	2.62	EASTERN	7.27	JOBS	3.03
EASTERN	2.35	COUNTRIES	5.76	OUTSIDE	2.98
NUMBERS	2.09	SKILLED	5.19	THOUSANDS	2.95
FLOOD	2.01	INFLUX	5.19	CAP	2.67
INFLUX	1.66	NUMBERS	4.91	INFLUX	2.51
COUNTRIES	1.40	COMING	4.91	ASYLUM	2.45
		AFRICAN	4.91	SEEKERS	2.34
		STOP	4.72	HIGHLY	2.26
		ALLOWED	4.63	AFRICA	1.85

* Includes 20 most frequent c-collocates for mid-market and broadsheets; full results available in Supporting Information

Finally, similar to the c-collocates of IMMIGRANTS, water metaphor terms such as FLOOD (in tabloids) and INFLUX (in all three types) appear as c-collocates of MIGRANTS as shown in Table 5. These terms tend to be used as metaphors for the arrival of migrants in the UK:

MPs called on the PM to deploy 'all necessary steps' to stop the **flood** of migrants and their families, currently enough to fill 'eight new cities'. (Tabloid)

The **influx** of migrants has put more pressure on public transport and led to more congestion. (Tabloid)

Transitional restrictions were imposed to prevent the kind of **influx** of migrants seen when Poland joined the EU. (Mid-market)

Many of the fastest-growing areas are those that have seen an **influx** of migrants in recent years, and are now seeing higher birth rates. (Broadsheet)

Results: ASYLUM SEEKERS

ASYLUM SEEKERS were most consistently described as FAILED across all three publication types, as shown in Table 6. This was the only significant L1 description used in the tabloids throughout 2010-2012, and one of two observed in the mid-market publications, where FAILED was most frequently used as measured by appearances per 1000 items in the corpus. Mid-market items were about three times as likely as broadsheets and tabloids likely to include a usage of FAILED ASYLUM SEEKERS. Tabloid coverage used this phrase at a slightly lower rate than broadsheets, again as measured by rate per 1000 items in the corpus.

Table 6. L1 collocates of ASYLUM SEEKERS, 2010-2012

Tabloids		Mid-markets		Broadsheets	
L1 collocate	Normalised	L1 collocate	Normalised	L1 collocate	Normalised
FAILED	6.72	FAILED	20.68	FAILED	7.38
		ILLEGAL	0.09	CHILD	1.46
				REFUSED	0.58
				DESTITUTE	0.36
				ILLEGAL	0.22
				VULNERABLE	0.17

Examples of this phrase included the following sentences focused on enforcement and deportation:

The UK Border Agency needs to deal with a raft of missing foreign criminals, **failed** asylum seekers and illegal immigrants. (Tabloid)

The report finds students whose visas have expired are regarded as a 'low priority' by the agency compared to illegal immigrants and **failed** asylum seekers. (Mid-market)

It is one of 13 secure centres set up to hold foreign national prisoners, **failed** asylum seekers and migrants who overstay. (Broadsheet)

Meanwhile, the broadsheets included the terms used in tabloids and mid-markets as part of a longer list of modifying words, although fewer than for IMMIGRANTS and MIGRANTS. Broadsheets used the modifiers DESTITUTE and VULNERABLE and REFUSED as an alternative to FAILED. However, these three modifiers all occurred in less than 1% of broadsheet items. Interestingly, mid-markets and broadsheets used ILLEGAL to describe asylum seekers, although this usage was not very common, occurring in less than 1 in 1000 mid-market items and in about 2.2% of items in broadsheets. As Baker and colleagues (2008) highlighted, this phrase is actually meaningless from a policy perspective: it is not illegal to seek asylum, although failed or refused asylum seekers may stay in Britain illegally after their applications and/or appeals have been rejected.

In the wider set of c-collocates, meanwhile, all three newspaper types refer to other migrant groups (IMMIGRANTS, MIGRANTS, REFUGEES) in the context of the term ASYLUM SEEKERS, as shown in Table 7.

Table 7. C-collocates of ASYLUM SEEKERS, 2010-2012

Tabloids		Mid-markets		Broadsheets	
C-collocate	*Normalised*	*C-collocate*	*Normalised*	*C-collocate*	*Normalised*
FAILED	7.06	FAILED	21.34	FAILED	8.21
IMMIGRANTS	2.27	IMMIGRANTS	6.70	REFUGEES	3.28
		ILLEGAL	6.04	MIGRANTS	2.37
		CRIMINALS	5.57	IMMIGRANTS	2.01
		MIGRANTS	3.87	DETENTION	1.71
		NUMBER	3.21	CHILD	1.60
		STAY	3.12	THOUSANDS	1.52
				ILLEGAL	1.24
				TREATMENT	1.21
				DEPORTATION	1.05
				HOUSING	0.96
				AUSTRALIA	0.91
				REFUSED	0.88
				COUNTRIES	0.83
				CRIMINALS	0.72
				DESTITUTE	0.66
				WHOM	0.55
				VULNERABLE	0.52
				TENS	0.52
				PRISONERS	0.47

* Includes 20 most frequent c-collocates for broadsheets; full results available in Supporting Information

Figure 4 also shows how mid-markets and broadsheets mention CRIMINALS alongside ASYLUM SEEKERS, whereas broadsheets make greater reference to words related to the process of seeking asylum and government enforcement of elements of the process, including DETENTION and DEPORTATION:

> Shock details revealed by MPs show that 275,000 missing foreign **criminals**, failed asylum seekers and others have not yet been dealt with. (Mid-market)

> Legislation to enable the **deportation** of asylum seekers would be introduced when Parliament resumes today, the prime minister said. (Broadsheet)

> Concern about the **detention** of all child asylum seekers—young children detained with their parents and under-18s detained in adult facilities—has grown in recent years. (Broadsheet)

Figure 4. Top 20 C-collocates of ASYLUM SEEKERS, 2010-2012

word	tabloid C-Collocate	midmarket C-Collocate	broadsheet C-Collocate
failed	▮	▮	▮
immigrants	▪	▮	▪
illegal		▮	▪
criminals		▮	▪
migrants		▪	▪
number		▪	
stay		▪	
Australia			▪
child			▪
countries			▪
deportation			▪
destitute			▪
detention			▪
housing			▪
refugees			▪
refused			▪
tens			▪
thousands			▪
treatment			▪
vulnerable			▪

Mentions per 1,000 items: 0.52, 5.00, 10.00, 15.00, 21.34

Results: REFUGEES

No single consistent L1 collocate stood out as a modifier for REFUGEES across the three publication types, as Figure 5 shows. Across the 2010-2012 period, the only modifier used consistently by tabloids to describe refugees was FLEEING, but this usage was uncommon relative to L1 collocates of other target words, with only .35 appearances per item, similar to the rate in broadsheets. Meanwhile, broadsheets made much greater reference to the origins of refugees (AFGHAN, IRAQI, SOMALI, AFRICAN, BURMESE). The greater range of international news in broadsheets could account for this observation, with relatively frequent mentions of PALESTINIAN REFUGEES in coverage of events in the Middle East, for example.

Figure 5. All L1 collocates of REFUGEES, 2010–2012
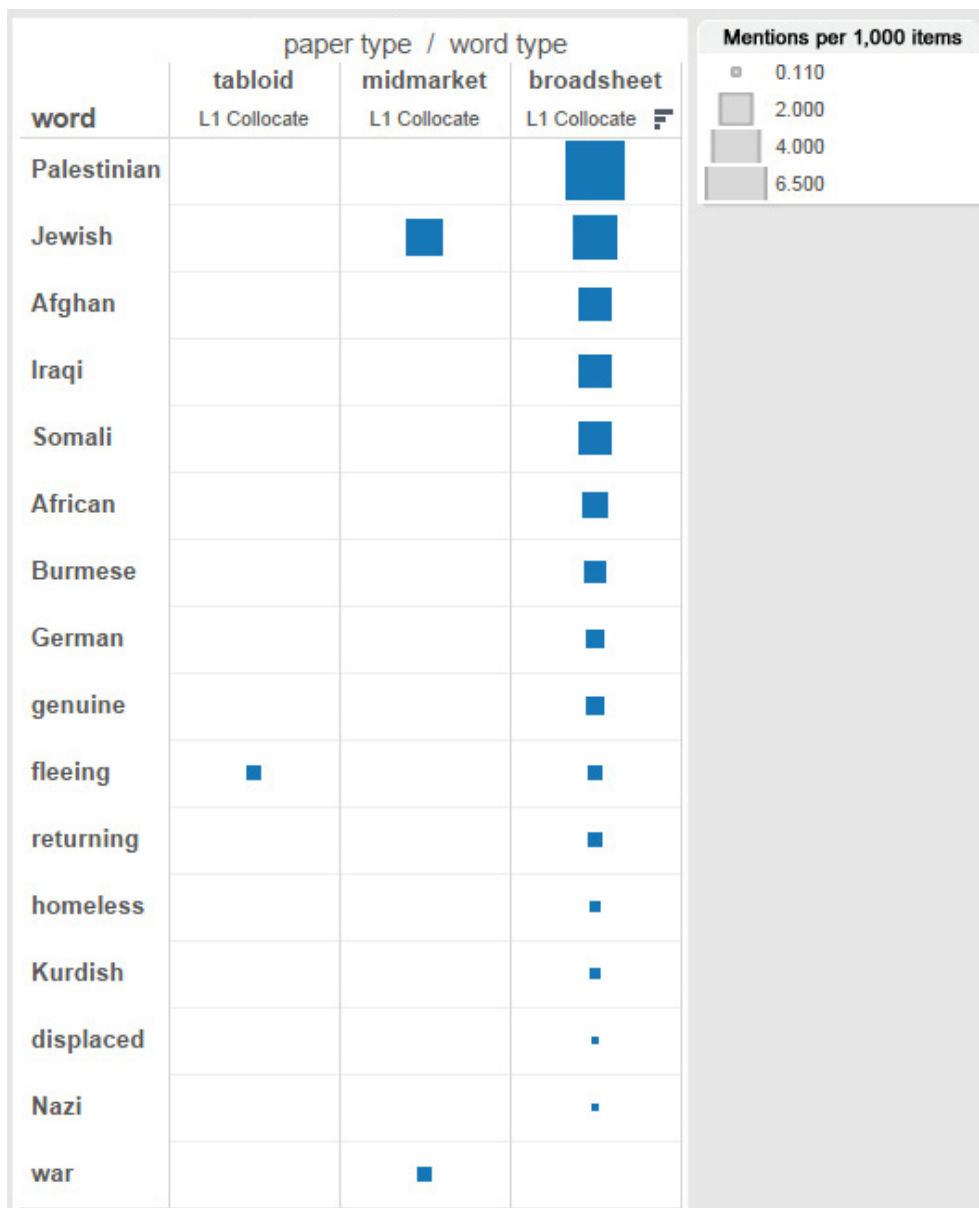
In contrast, JEWISH REFUGEES, occurring in both mid-markets and broadsheets, seems to refer to UK residents whose ancestors fled from the Nazis, or from other instances of persecution, as shown in examples from mid-markets:

> My parents came to Britain as immigrants, **Jewish** refugees from the Nazis. (Tabloid, Mid-market, and Broadsheet)

> This simple statement goes to the heart of what drives his remarkable generosity: his origins as the son of **Jewish** refugees from Russia. (Mid-markets)

This phrase spiked in usage when Labour leader Ed Miliband used it in a major speech on immigration that was widely covered in the press – his speech is the source of the first example sentence listed above. However, JEWISH appeared with REFUGEES in other instances in the corpus, otherwise it would not be a consistent collocate from 2010 to 2012.

Table 8. L1 collocates of REFUGEES, 2010-2012

Tabloids		Mid-markets		Broadsheets	
L1 collocate	Normalised	L1 collocate	Normalised	L1 collocate	Normalised
FLEEING	0.35	JEWISH	2.46	PALESTINIAN	6.50
		WAR	0.38	JEWISH	3.77
				AFGHAN	1.90
				IRAQI	1.82
				SOMALI	1.79
				AFRICAN	1.32
				BURMESE	0.83
				GERMAN	0.58
				GENUINE	0.52
				FLEEING	0.36

Looking at the broader window surrounding mentions of REFUGEES adds additional collocates, as shown in Table 9. This includes mentions of organisations dealing with refugees directly such as the UN. Analysis of uses of HIGH, as found in the broadsheets, shows it occurring in mentions of the UN High Commissioner for Refugees. Also, mid-market and broadsheets' use of JEWISH tended to occur in the context of Jews escaping the Nazis, or other historical references to Jewish experiences. Quantity terms such as THOUSANDS, as well as location words like BORDER, also appeared in the tabloids and broadsheets, but were not as prominent in mid-market coverage of refugee stories. CAMP appears in the list for tabloids and broadsheets, but not in mid-markets, which atypically had an even shorter list than tabloids in this case.

Table 9. C-collocates of REFUGEES, 2010-2012

Tabloids		Mid-markets		Broadsheets	
C-collocate	Normalised	C-collocate	Normalised	C-collocate	Normalised
FLEEING	3.75	FROM	12.09	THOUSANDS	7.88
CAMP	2.97	ASYLUM	3.40	BORDER	7.44
FLED	2.70	JEWISH	2.93	PALESTINIAN	7.11
ASYLUM	2.53	UN	1.61	FLEEING	6.03
THOUSANDS	2.53	WAR	1.61	FLED	5.32
HELP	2.44			HIGH	5.15
BORDER	1.83			RETURN	4.77
				UN	4.49
				ASYLUM	4.30
				ACROSS	4.16
				JEWISH	4.13
				TURKEY	3.88
				NUMBER	3.77
				COMMISSIONER	3.64
				SEEKERS	3.33
				MILLION	3.25
				CAMPS	3.06
				CAMP	2.95
				INFLUX	2.78
				HELP	2.73

* Includes 20 most frequent c-collocates for broadsheets; full results available in Supporting Information

Changes in 2012 coverage

The above results show consistent patterns in coverage for the 2010-2012 period. This is a useful time period to examine, including both the general election campaign of 2010 and the era of active immigration policy-making under the current coalition government. But coverage of migration may shift from year to year or month to month. The analysis above deliberately blocks out those changes in order to look at consistent patterns. However, changes from year to year may be of interest as well.

A brief comparison of 2012 coverage to the prior two years, however, reveals some of the more recent changes in newspaper coverage. For each target word (IMMIGRANTS, MIGRANTS, ASYLUM SEEKERS, REFUGEES), this comparison sought, first, words that were collocates in 2012 but 2010-2011 and, second, collocates from 2010 and 2011 that dropped off the list for 2012. Taken together, these two sets of words suggest changes in coverage.

Some of the notable results of this analysis were as follows. First, both tabloids and mid-market publications mentioned STUDENTS in the presence of IMMIGRANTS more in 2012 than in 2010 or 2011. This included discussion about student visas and the revoking of London Metropolitan University's 'highly trusted status' by the UK Border Agency:

> Still more immigrants signed up as **students** without any intention of going to college. (Tabloids)

> London Metropolitan University was condemned as very seriously deficient at ensuring its international **students** do not become illegal immigrants. (Mid-market)

Second, BULGARIAN and ROMANIAN appear as tabloid collocates of IMMIGRANTS in 2012; similarly, BULGARIA and ROMANIA are collocates of MIGRANTS in 2012 but not in 2010 or 2011 in broadsheets. These findings echo ongoing public debates surrounding A2 migration:

> The countryside is under threat from a wave of **Romanian** and **Bulgarian** immigrants, MPs were warned yesterday. (Tabloid)

> Migrants from **Romania** and **Bulgaria** must also pass a Habitual Residence Test to claim benefits, unlike the citizens of other EU countries. (Broadsheet)

Third, use of water metaphors referring to mass entry also changed within the three year period. FLOOD and FLOW were significantly collocated with IMMIGRANTS in 2010 and 2011 mid-market coverage, but not in the 2012. In broadsheets FLOOD did not appear near IMMIGRANTS as strongly in 2012, although SURGE arose as a related term:

> A shake-up of the legal system to stop 'trivial' judicial reviews is planned after a **surge** in the actions by immigrants facing deportation. (Broadsheet)

Meanwhile, in mid-market coverage, FLOW and WAVE were significantly collocated with MIGRANTS in 2010 and 2011 but not in 2012. Similarly, for broadsheets, WAVES was not significantly collocated with MIGRANTS in 2012 coverage, unlike the previous two years.

Finally, the word BOGUS showed some interesting patterns: it was a significant collocate of IMMIGRANTS in 2012 broadsheet coverage, though not in 2010 or 2011. Meanwhile, BOGUS had been a significant collocate of ASYLUM SEEKERS in mid-markets in 2010 and 2011 but dropped off the list in 2012. Closer examination of broadsheet mentions in 2012 shows BOGUS as a modifier of STUDENTS, in the context of other immigrant groups:

> Keith Vaz, chairman of the Commons home affairs committee, said the agency was 'unable to focus on its key task of tracking and removing illegal immigrants, overstayers or **bogus** students from the country' despite continued pressure to improve during the past six years. (Broadsheet)

This represents a change in the use of BOGUS in the migration context. BOGUS was often used to describe asylum seekers in the news in the late 1990s and early 2000s, despite objections that this usage was illegitimate: since asylum seekers have not yet had the validity of their claims assessed, they could not be clearly defined as bogus (Gabrielatos and Baker, 2008). Although fluctuations in the use of BOGUS in 2012 among the three publication types indicate some shifting, it remains in use in at least some national newspapers, some of the time. It remains to be seen how the shift in the use of BOGUS, and decline in 2012 in mid-markets, will last over the long term.

Implications

The purpose of this report is to shed light on British press coverage of migration by revealing patterns in the words that appear consistently in association with immigrants, migrants, refugees and asylum seekers. The potential implications, therefore, adhere as closely as possible to the patterns that emerged from the analysis of collocations. Of course, some degree of selection is inevitable when picking out which of the raw statistical results have significant broader implications. But the choices here are led by the results, as the implications centre on the most prominent patterns that emerged from the analysis.

Three of the most noteworthy patterns to emerge are the following:

1. the prominence of the phrase ILLEGAL IMMIGRANTS,
2. the featuring of the EU and EASTERN EUROPE as the primary geographical reference point for coverage of IMMIGRANTS and MIGRANTS, and
3. and the distinct—though in some cases overlapping—vocabularies surrounding the words IMMIGRANTS, MIGRANTS, REFUGEES, and ASYLUM SEEKERS. These findings have been detailed above; in this section, we will suggest some broader implications of these results.

The predominance of ILLEGAL as a modifier of IMMIGRANTS, and to a lesser extent MIGRANTS clearly stands out. ILLEGAL far outnumbered any other modifier of immigrants, across all three publication types. Aside from the political controversies around this very phrase, it is worth noting that immigrants with legal status far outnumber those without it, according to the best estimates of the size of both types of migrant populations. Of course, this fact does not carry any automatic implications for news coverage, as newspapers are not obligated to reflect migration statistics in the way they design coverage.

It may be tempting here to jump to conclusions linking media coverage with public attitudes, and in turn policy-makers' attention. It is true that public opinion shows a very widespread concern with 'illegal immigration' (German Marshall Fund 2011; Migration Observatory 2011 - Thinking Behind the Numbers). At this point, however, we can only speculate about whether extensive media coverage is driving this concern. Media coverage might instead, or in addition, take its cues from policy makers (see Bennet 1990 on the theory of 'indexing') and/or segments of the public itself.

The second key finding is the emergence of the EU and Eastern Europe as the primary geographic reference point for discussions of immigrants and migrants, especially in tabloid coverage. This is not entirely new (see Gabrielatos and Baker 2008), but the prevalence of Europe over other geographical areas is striking, particularly in tabloid coverage. Mid-markets, and especially broadsheets, included a wider variety of geographic terms, but in tabloids the EU and Eastern Europe were the only geographic terms that consistently collocated with IMMIGRANTS and MIGRANTS. Again, this balance in coverage does not reflect the prevalence of EU and non-EU migrants in official statistics, as non-EU nationals continue to clearly outnumber EU nationals in the UK's migrant populations and in

new arrivals (see our briefing on "Migration Flows of A8 and Other EU Migrants to and from the UK"). On the other hand, 'news' by definition emphasises events that are new and therefore presumably interesting to news consumers. The dramatic change in migration to Britain from Eastern Europe after the 2004 EU expansion was and continues to be particularly newsworthy by that standard, relative to older patterns of migration. Again, there is no obligation for media coverage to reflect migration statistics proportionally.

Third, refugees and asylum seekers are described using different sets of terms. Asylum seekers, but not refugees, are described as FAILED, more often than by any other single collocated word. Asylum seekers, but not refugees, are also associated with IMMIGRANTS in all three publication types, and with ILLEGAL (in the mid-markets) as well as being DESTITUTE and VULNERABLE (in the broadsheets). The terms surrounding REFUGEES, especially in tabloids and mid-markets, are relatively distinct. The discourse around the word REFUGEES is much more international in nature, with c-collocates including CAMPS, the UN, WAR and a number of specific geographical terms. REFUGEES are not described frequently as FAILED: this would be nonsensical, as refugee status implies success in the process of seeking asylum. Rather, refugees are depicted as FLEEING and, in the broadsheets, are associated with numerous countries of origin. Thus, the word REFUGEES rather than ASYLUM SEEKERS appears most associated with international crises. Meanwhile, ASYLUM SEEKERS are more likely to be associated with people in Britain seeking refugee status, and often with being unable to attain that status.

Of course, this reflects a real distinction between asylum seekers and refugees: if the terms are used according to their technical definitions, asylum seekers are applying for international protection and thus aiming to become refugees, while refugees have already been determined worthy of protection. On the other hand, the same people might be depicted in one way when FLEEING war-torn areas or residing in CAMPS far afield (when identified in the news as refugees), and then very differently when arriving in Britain (and identified as asylum seekers).

In addition, the distinct vocabulary surrounding refugees represents a change since the late 1990's and early 2000's. Baker and colleagues found significant overlap in the consistent collocates of all four target words, IMMIGRANTS, MIGRANTS, ASYLUM SEEKERS and REFUGEES. In 2010-2012, by contrast, REFUGEES shared fewer c-collocates with the other three terms. This may reflect increasing consciousness of the distinctions between refugees and asylum seekers, and also between refugees/asylum seekers on one hand and immigrants/migrants on the other hand. This is certainly a grey area, as refugees and asylum seekers do figure into UK immigration statistics and fit official definitions of migration to the UK if they stay in the country for at least a year. However, most immigrants and migrants are not asylum seekers or refugees. Increasingly, British press coverage seems to reflect this by using more distinct vocabularies to describe each, though this is truer for refugees than asylum seekers.

Limitations of the data and analysis

This study reports on patterns in the words used by national UK newspapers in connection with immigrants, migrants, asylum seekers and refugees. However, textual data—particularly of newspaper content—presents some methodological limitations. As reliable and comprehensive sources for national and even regional publications, newspaper databases like NexisUK are often used as sources for linguistic studies. However, several constraints and limitations are worth mentioning.

First, the exact content of these databases and the precise methods by which texts are uploaded (and duplicate articles removed) are not fully disclosed, due to the commercial nature of the services. Second, a newspaper-only, text-only database does not capture many aspects of media coverage. It omits pictures and videos, as well as visual impact of large banner headlines. Also, it focuses on newspapers to the exclusion of broadcast media and on-line publications. Efforts were made to include BBC political programming, but transcripts are not produced or available, and manual transcription is prohibitively costly. Even if a full corpus of spoken-word transcripts on political news in Britain were available, it would introduce new challenges for comparison to the findings from a printed-word corpus.

Meanwhile, online texts—whether through the BBC and other media outlets' websites, or social media such as blogs and Twitter—are easier to analyse, but collection of these texts systematically poses difficulties as well. While these are not insurmountable for later stages of data collection, they were not attempted in this project.

More broadly, this study contributes to the building of a comprehensive base of evidence about how British newspapers have portrayed different migrant groups. It reports on patterns in the words used in conjunction with IMMIGRANTS, MIGRANTS, REFGUEES and ASYLUM SEEKERS. As such, it can answer questions about what sorts of language appears in media coverage of migration, but cannot answer questions about the impact of these patterns on readers of newspapers and more broadly on public understandings of and attitudes toward migration. Such aims would require future research of a different kind. Media criticism often assumes that media portrayals strongly influence public attitudes and perceptions of migrants. Analysing media coverage does not establish that link, although it lays the groundwork for future work that might explore that question more closely.

Finally, this report does not attempt to define a right or wrong way of covering migration, or to evaluate the approaches taken by national newspapers (Leveson 2012). Instead it describes and quantifies the language used in British national newspapers as they provide news about migration to their various segments of the public that read them.

References

- Atkins, BT Sue, and Michael Rundell. *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press, 2008.

- Baker, Paul. *Using Corpora in Discourse Analysis.* London: Bloomsbury Academic, 2006.

- Baker, Paul, Costas Gabrielatos, Majid Khosravinik, Michal Krzyzanowski, Tony Mcenery, and Ruth Wodak. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society 19*, no 3 (2008): 273–306.

- Baker, Paul, Costas Gabrielatos and Tony McEnery. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press, 2013.

- Bennet, W. Lance. "Toward a Theory of Press-State Relations in the United States." J*ournal of Communication 40*, no. 2 (1990): 103–127.

- Charteris-Black, Jonathan. "Britain as a Container: Immigration Metaphors in the 2005 Election Campaign." *Discourse and Society 17*, no. 5 (2006): 563-581.

- Gabrielatos, Costas and Paul Baker. "Fleeing, Sneaking, Flooding a Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005." *Journal of English Linguistics 36*, no.1 (2008): 5-38.

- Hunston, Susan. "Semantic Prosody Revisited." *International Journal of Corpus Linguistics 12*, no. 2 (2007): 249-268.

- Kilgarriff, Adam, Milos Husák, Katy McAdam, Michael Rundell and Pavel Rychlý. "GDEX: Automatically Finding Good Dictionary Examples in a Corpus." Paper presented at the European Association of Lexicography, Barcelona, Spain, 15-19 July 2008.

- Leveson, The Right Honourable Lord Justice. "An inquiry into culture, practice and the ethics of the Press." *Leveson report, Vol II, Section 8.* The Stationery Office, London, 2012.

- Lindquist, Hans. *Corpus Linguistics and the Description of English.* Edinburgh: Edinburgh University Press, 2009.

- McEnery, Tony and Andrew Hardie. *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press, 2011.

- Pollach, Irene. "Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis." *Organizational Research Methods 15*, no. 2 (2012): 263-287.

- Sinclair, John. *Corpus, Concordance, Collocation.* Oxford: Oxford University Press, 1991.

- Smith, Michael. *WordSmith Tools.* Liverpool: Lexical Analysis Software, 2012.

- Xiao, Richard and Tony McEnery. "Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective." *Applied linguistics 27*, no. 1 (2006): 103-129.

Related material

- AP Stylebook: 'Illegal immigrant' no more - http://blog.ap.org/2013/04/02/illegal-immigrant-no-more/

- Transatlantic Trends: Immigraion Survey 2011 - http://trends.gmfus.org/immigration/about/

- Migration Observatory report: Thinking Behind the Numbers - www.migrationobservatory.ox.ac.uk/reports/thinking-behind-numbers-understanding-public-opinion-immigration-britain

- Migration Observatory briefing: Migration Flows of A8 and Other EU Migrants to and from the UK - www.migrationobservatory.ox.ac.uk/briefings/migration-flows-a8-and-other-eu-migrants-and-uk

Appendix A: Characteristics of the Corpus

Table A1. Number of items by publication title and year

Tabloids	2010	2011	2012	Total	Percent*
The Sun and The Sun on Sunday	1,689	1,590	2,011	5,290	9.1%
Daily Mirror and Sunday Mirror	1,179	1,142	1,254	3,575	6.1%
Daily Star	1,013	597	438	2,048	3.5%
The People	82	106	130	318	0.5%
Daily Star Sunday	63	71	101	235	0.4%
TOTAL	4,026	3,505	3,943	11,465	19.6%

Mid-markets	2010	2011	2012	Total	Percent*
The Express	2,242	2,078	1,225	5,745	9.8%
Daily Mail and Mail on Sunday	1,397	1,370	1,268	4,035	6.9%
Sunday Express	422	362	226	1,010	1.7%
TOTAL	4,061	3,810	2,719	10,590	18.1%

Broadsheets	2010	2011	2012	Total	Percent*
The Guardian	2,615	2,395	2,094	7,104	12.2%
The Times	2,503	2,183	2,393	7,079	12.1%
Daily Telegraph	2,012	1,792	2,228	6,032	10.3%
Financial Times	1,569	1,418	1,551	4,538	7.8%
Sunday Times	1,084	1,178	1,332	3,594	6.2%
The Independent	1,224	994	1,320	3,538	6.1%
The Observer	843	754	546	2,143	3.7%
Sunday Telegraph	446	486	528	1,460	2.5%
Independent on Sunday	264	287	257	808	1.4%
TOTAL	12,560	11,487	12,249	36,296	62.2%

Entire corpus	2010	2011	2012	Total	Percent*
	20,647	18,802	18,911	58,351	99.9%

*Totals do not add to 100% due to rounding.

Table A2. Number of words in each annual subcorpus by publication type

Type	2010	2011	2012	Total	Percent
Tabloids	1,798,886	1,553,144	1,689,662	5,041,692	11.5%
Mid-markets	2,749,754	2,507,580	1,912,388	7,169,722	16.3%
Broadsheets	10,933,800	10,078,718	10,742,940	31,755,458	72.2%
TOTAL	15,482,440	14,139,442	14,344,990	43,966,872	100.0%

Appendix B: Notes about methods

Selection of texts

Items were found by searching the NexisUK newspaper database using the following query string: refugee! OR asylum! OR deport! OR immigr! OR emigr! OR migrant! OR illegal alien! OR illegal entry OR leave to remain AND NOT deportivo AND NOT deportment. The '!' character indicates that words beginning with those letters are also captured in the search. The exclusions were needed because 'Deportivo' is a Spanish football club, while 'deportment' refers to etiquette. Excluding these terms filtered out numerous items that would have had nothing to do with immigration or refugees. Also, the search avoids explicitly searching for 'migration' because of the risk of retrieving unrelated articles involving environmental affairs (e.g., such as animal migration) or technology (e.g., data migration). 'Migrants', on the other hand, is not a relevant term in either of these topics. Highly-similar items as determined by NexisUK were filtered out of the results. This approach relied on the precedent and experience of Gabrielatos and colleagues (2008).

Collocation analysis

Using Wordsmith Tools (Smith 2012), we performed an analysis of collocationsof each target group to see which words regularly appeared in proximity to these mentions. The span used for this analysis included any words located within five words to the left or right of references to the migrant group being analysed, resulting in a ten-word window. Two statistical criteria were used on words appearing within this window. First, to be judged as an annual collocate, a pair of words had to display a Mutual Information (MI) Score of at least 5.0 within that year's subcorpus. Second, it had to have a log-likelihood score of at least 6.63 in the same period. For the precise mathematical details of each test, see Pollach (2011).

It is known that MI can overvalue associations involving infrequent or rare words, making these collocations seem more important than they actually are (Xiao and McEnery, 2006). Meanwhile, the log-likelihood test shows the statistical significance of a possible collocation. Generally, corpus linguists agree that using a combination of measures to assess the strength of collocations minimises the risk of relying entirely on one technique (Lindquist, 2009). However, since these annual results could be susceptible to fluctuations around specific events, and we were also interested in identifying how migrant groups were regularly described over a three-year period, we reported consistent collocates (Gabrielatos and Baker, 2008). These are words that met the above criteria, and also were statistically significant in every separate year of the corpus.

Normalisation of collocation results

When comparing frequencies of collocations as they appear in two different contexts, such as tabloid versus broadsheet coverage, it is misleading to report the raw figures if the two sets of coverage have different amounts of text. This is especially apparent in our study, as Table A1 shows that the broadsheet sample contains more than three times as many items as either the tabloid or mid-market samples. Therefore, it is important to normalise the collocation frequencies to enable comparison across differently sized datasets. Typically, corpus linguists would normalise their results in terms of occurrences per 1,000 or 1,000,000 words. However, since tabloid articles are considerably shorter than broadsheet articles, using this convention would understate the relative frequency with which a word appears in broadsheets, and overstate it in tabloids (Gabrielatos and Baker 2008). Therefore, we chose to normalise the collocation results per 1000 items, for each publication type dividing the number of instances of a given collocate by the number of articles and then multiplying by 1000. The resulting figure, which is reported alongside the raw frequency, tells us how many times a given collocate occurs per 1000 items.

For example, if both a tabloid and a broadsheet ran a story using the word DEPORTATION once, but the broadsheet story was twice as long as the tabloid's, then the typical 'number per million words' measure would suggest that tabloids use DEPORTATION twice as often as broadsheets. But on a given day the readers of each newspaper's migration story would see each word mentioned once. Thus, arguably, normalising by number of items rather

than words is more appropriate here: it avoids overstating how often a c-collocate appears in a tabloid publication relative to its overall coverage (Gabrielatos and Baker 2008). A single appearance of DEPORTATION in a broadsheet article thus counts the same as a single appearance of DEPORTATION in a shorter tabloid article (or in an even shorter letter-to-the-editor), even if the broadsheet article has more words. The results presented below therefore normalise per 1000 newspaper items rather than per million words.

Selection of sentence examples

To illustrate the key findings in this report, we included examples of collocation patterns that were drawn from the corpus. However, one of the challenges associated with this task involves selecting sentences that are typical for a particular collocation. To accomplish this, we used GDEX, a system that scores collocations based on how typical, informative and intelligible they are. Originally developed to automatically separate out 'good examples' of how a word is used which would be most appropriate for dictionaries, it applies a series of rules and weights to rank sentences that feature a collocation. These examine a variety of characteristics, including the length of the sentence, the rarity of the words in the sentence and whether the collocation in question is the main subject. The precise details of the rules are contained in Kilgarriff et al. (2008), while a fuller discussion of what 'good' dictionary examples look like can be found in Atkins and Rundell (2008). GDEX proved useful in the context of this report for its ability to quickly identify typical and informative examples of major collocation patterns in our corpus. The examples provided in this report were all highly ranked by the GDEX formula. However, their presence is only to serve as illustrative examples of results from the analysis of collocations.

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

Suggested citation

Allen, William and Scott Blinder. "Migration in the News: Portrayals of Immigrants, Migrants, Asylum Seekers and Refugees in National British Newspapers, 2010 to 2012." Migration Observatory report, COMPAS, University of Oxford, August 2013.