

The Immigration Issue in the European Electoral Campaign in the UK: Text-Mining Public Debate from Newspapers and Social Media*

Paul Nulty

paul.nulty@gmail.com

Monica Poletti

m.poletti@lse.ac.uk

June 22, 2014

Abstract

In recent years, the issue of immigration has become increasingly salient in the UK political and media debate. Moreover, with the development and persistence of the economic and financial crisis within the EU, immigration has been linked to growing opposition and criticism towards the European Union. In a country in which Euroscepticism has historically been high compared to countries in continental Europe, EU immigration-related statements connected to EU free-border agreements became more widespread. For this reason, we expect immigration to be a prominent issue in the electoral campaign of the upcoming 2014 European Parliament elections in the media. By covering (potential) EU immigrants and EU immigration issues in a certain way, media tend to promote or restrain certain ideas of immigration, that might eventually affect public's views. In fact, we know from previous studies that immigration, particularly in times of economic crisis, is a challenge for society that can be framed not only in positive or negative terms, but also in economic or cultural terms.

By looking at the news coverage of several newspapers, selected according to their political orientations and including both broadsheets and tabloids, this study first considers the salience of coverage of EU immigrants and EU immigration issues in UK newspapers in the three months preceding the EU elections of May 2014. It further explores whether news coverage of different newspapers is framed in terms of economic (e.g. jobs) or cultural (e.g. identity) terms. In addition, we mine information from social media sources (twitter and wikipedia) to discover how the immigration debate is framed by politically engaged members of the public on these platforms. Although inferring public opinion from social media can be problematic, it is a rewarding approach given the volume of public text available, and the ability to identify political affiliations through network connections. Understanding representation of immigration and its connections with public opinion is crucial not only in order to contribute to the scholarly debate on anti-immigration and

*Author addresses: Department of Methodology, London School of Economics and Political Science, London WC2A 2AE. Paper prepared for presentation at the 'Text Mining and Public Policy' workshop at LSE, 27th June 2014. This paper is a work in progress. The latest versions of the code, the manuscript, and supplementary materials are available at: <https://github.com/pnulty/LSEtext>. This research was supported by the European Research Council grant ERC-2011-StG 283794-QUANTESS.

Euroscepticism attitudes, but also to better inform public policy decisions. In addition to these substantive contributions, this study employs advanced text-mining methods for data collection and data analysis. Text from politically engaged twitter users is collected using a novel iterative keyword-learning approach that allows high precision and recall when the aim is to discover only tweets relevant to a particular issue. This text, and the text from newspaper articles, is analysed using topic modelling to investigate the salience of sub-issues in the immigration debate. In addition, weighted log-odds ratios are used to measure linguistic differences across newspapers and social media users from different sides of the ideological spectrum in the UK.

1 Introduction

In recent years, immigration policy-related issues have become increasingly salient in the UK public debate. This is partly connected with the development and persistence of the initially worldwide and then specifically European economic and financial crisis. Despite European countries have been affected in different ways by the crisis, unemployment has been increasing virtually everywhere in Europe, including the UK. In such a context, in a country in which Euroscepticism has historically been quite high compared to other countries in continental Europe, it was relatively easy to establish a connection in the British public debate between restrictive immigration policy positions and growing opposition and criticism towards the European Union. EU immigration-related statements connected to reconsider EU free-border agreements to avoid other Europeans to come and work in the UK have lately become more and more widespread, particularly targeting immigrants coming from historically poorer economies such as Eastern European countries, or from countries badly hit by the crisis such as the Southern European ones. This somewhat “economic” debate added up and strengthened a pre-existing “cultural” debate about the strains and failures of UK multi-culturalism and the impact of different ethnic and religious minorities values or practices (e.g. Islam) on British society.

Connection with EU issues seems to have allowed strongly anti-immigration policy statements to break through the barriers of right wing extremism and to become mainstream, together with an increasing stress in the public debate on defence of British jobs and British values. Evidence of the fact that restrictive immigration policy preferences became mainstream not only in the public debate but also among the public emerged quite palpably from the results of the 2014 European Parliament elections in the UK. The success (XX%) of the previously minority right-wing populist UK Independence Party (UKIP), that based its whole electoral campaign on promoting more restrictive immigration policies (also within the EU) and possibly on leaving the EU, was unprecedented. “26 million people in Europe are looking for work. And whose job are they after” was undoubtedly one of UKIP most famous poster of the electoral campaign. Although more culturally-oriented statements did not fail to be heard ¹.

¹ A famous statement by Nigel Farage has been for instance: “I got the train the other night, it was rush hour [...]. It wasn't until after we got past Grove Park that I could actually hear English being audibly spoken in the carriage. Does that make me feel slightly awkward? Yes.” (XXX)

Having turned into a particularly prominent issue in the media electoral debate of the EP elections, we choose the immigration-policy related debate as focus of our study. In any given policy area it is key to understand how the public debate is shaped and what its dynamics are. With the goal of gaining support for their policy propositions, politicians selectively emphasize features of an issue in ways that most favour their position (Riker, 1986; Schattschneider, 1960). They for instance highlight likely effects of a specific choice or its connection to important values (e.g. Jacoby, 2000, p.75). By reporting policy debates, media allows the frames of issues made by political actors to reach a wide public (Entman, 2004), and this allow citizens to make up their mind among competing arguments. Media, however, are also using frames. In covering specific issues in a certain way and in the modality of reporting statements from different politicians, different media tend to promote or restrain different ideas and policy proposals, that might also eventually affect public's views and behaviour. This is why it is important to investigate which frames are used more regularly in the political and media debate and in which media outlets they tend to be predominant.

Immigration is particularly a challenge for society in times of crisis, and the framing of immigration policy-related issues debate in the public discourse has more immediate consequences at elections time. By looking at the news coverage of several newspapers, selected according to their political orientations and including both broadsheets and tabloids, this study first considers the salience of coverage of different immigration policy-related issues in UK newspapers in the three months preceding the EU elections of May 2014. It further explores whether news coverage of different newspapers is framed in terms of economic (e.g. jobs) or cultural (e.g. identity) terms.

In addition to these substantive contributions, this study employs advanced text-mining methods for data collection and data analysis in a field in which these have been rarely used. Text from newspaper articles is analysed using topic modelling to investigate the salience of sub-issues in the immigration debate. In addition, weighted log-odds ratios are used to measure linguistic differences across newspapers and social media users from different sides of the ideological spectrum in the UK.

Mass media play an important role in how the public understands and reason about con-

tentious social and political issues, such as immigration. They reflect and contribute to policy-issues related debate in several ways, and their role is even more important at elections time, when what issues and how the media decides to report them is particularly influential on public opinion. Agenda-setting and framing effects might be crucial in this respect. Agenda-setting effects refers to the ability to influence how salient a topic is at a given point in time (McCombs Shaw, 1972). If an issue is covered frequently and presented as a prominent news, the public will tend to regard the issue as more important. Agenda-setting is the relative prominence given to issues/subjects (first level) or attributes of issues (second level) in the media (McCombs, 2005; Ghanem, 1997).

Framing is similar to agenda-setting in the sense that they both focus on public policy issues in the news. In addition to agenda-setting, however, framing analysis also looks at how this specific issue is talked about in the news. Any specific topic can be viewed in several ways and can be portrayed by a variety of perspectives. Stressing a particular interpretation or a particular feature of an issue makes alternative interpretations or features less prominent, and a news frame is precisely “an emphasis in salience of certain aspects of a topic” (De Vreese, 2002). Framing the issue in certain ways might not only inadvertently bias the public exposed to the frame toward a specific stance about immigration or towards certain immigration policy preferences (Bennett, 2001; Liu Kirkwood, 2006), but it could also eventually affect peoples political behaviour in the Election Day (Iyengar, Peters, Krosnic, 1984).

Strong frames are not necessarily “true” arguments or intellectually better ones. They can be based (and often are) on exaggerations and lies, and they can feed publics prejudices and fears. In the same way as some of the most common sources of anti-immigrant attitudes among people are related to material interests and national identity (e.g., Citrin et al., 1997; ORourke and Sinnott, 2006; Hainmueller and Hiscox, 2007; Lahav, 2004), some of the most common ways to portray anti-immigration discourse in the media seem to be by focusing either on material/economic interests or on national/cultural identity. Immigrants are therefore often portrayed either as a threat to the material world (“immigrants take British people job“, ”immigrants are a burden to the national social welfare system“, ”immigrants cost us too much money“, etc.) or as a threat to their symbolic world (“immigrants take control of the UK”,

”traditional British values are getting lost due to immigration”, ”immigrants cannot even speak English properly“). The same economic and cultural dimensions are used for presenting them as a resource. Immigrants can be portrayed as an economic advantage for the country (”immigrants do the jobs that British do not want to do“, ”immigrants makes the NHS more sustainable because they pay taxes“) or as a cultural enrichment for the country (”immigrants enrich our culture“, etc.).

Although framing effects might depend on frequency of exposure to a specific frame, perception of credibility of their sources (e.g. Druckman, 2001) and from individual characteristics of the receiver such pre-existing values (e.g. Druckman, 2001; Barker, 2005) or knowledge (e.g. Kinder Sanders, 1999; Miller Krosnick, 2005), in general framing effects literature suggests that if immigrants were predominantly portrayed by the media elite as a threat to the UK, than many British people will be more likely to support negative attitudes towards immigrants. Consequently they will tend to vote for anti-immigration party as well as to support policies that seek to limit the amount of immigrants entering the UK and that either exclude them from British society or that force them to adapt their values to those of the host country. Different choice and preferences would arise if immigrants were instead portrayed as a resource to the UK society (XXX).

2 Merging Different Approaches

The literature related to framing and framing effects is wide (e.g. Edelman, 1993; Goffman, 1974; Iyengar, 1991; Zaller 1992; Semetko and Valkenburg, 2000; Gross D'Ambrosio, 2004; Vliegenthart, Schuck, Boomgarden and De Vreese, 2008; Chong Druckman, 2007; Entman, 2007). The implication of such a vast literature is that different types of frames have been identified (De Vreese, 2005: 53). Frames could for instance be generic, when they are general and applicable to different issues and in different contexts, or issue-specific, strictly related to a specific issue and to the context taken into account. Moreover, different strategies have been developed for identifying frames in the news (De Vreese, 2005). The first way is inductive, that is no predefined frames are used. The approach is explorative and it looks for the array

of frames that could be found in the news without any apriori ideas (e.g. Gamson, 1992). It is usually very labour-intensive, therefore often based on small samples and quite difficult to replicate (Hertog McLeod, 2001). The second way is deductive, namely texts are analysed using predefined frames that are searched for in the news. This approach requires to have well-defined a priori ideas of the type of frames that will be found in the news and well-defined operationalization of them. Although it is easier to replicate, and allows for larger sample to be dealt with, it still requires long time if carried out through manual content analysis.

Traditionally, in political science and in political communication, framing and saliency of specific issues have been investigated and detected either through interpretative textual analysis or through more systematic manual content analysis (XXXX). More rarely, automated approaches using computer-aided dictionary search have been adopted. Although the latter have the advantage of being a much faster and reliable (i.e. replicable obtaining exactly the same results) approach than manual content analysis, they tended to have the disadvantage of being a more superficial and less flexible method. In computational and corpus linguistics, however, progress have been made both in agenda-setting and in framing analysis, using more complex algorithm than the past that ends up in giving more valid way to quickly identify frames within very large corpus of text.

The aim of the study is to look at the electoral discourse on immigration during the 2014 European Elections campaign in the UK using these latter type of techniques. We use both a deductive and inductive approach to capture frames and issue-saliency, that in automated text-mining techniques are called supervised and unsupervised approach respectively (Laver, Benoit Garry, 2003). First of all, in a deductive way, we look at whether immigration is portrayed using an economic or cultural frame in terms of the consequence for society. Frames could be either positive or negative. Thus, we also look at whether these frames are more focused on portraying immigrants as a threat or as a resource.

The type of frames presented by specific media outlet could be influenced by many factors. Given that one of the main role of media outlets is to report news, one of this factor is certainly the type of political discourse carried out by the political elites and the type of frames used by politicians during their electoral campaign. Given the predominance in the media by

UKIP views on immigration, we expect that immigrants will be mostly reported as a threat to the UK society. We want to explore, however, whether it is actually the economic frame that is predominant (as UKIP leader stresses in order to avoid “racism” labelling of its party) or the cultural frame. Another relevant factor is Political Conservatism, which is associated with psychological management of public fears and uncertainty (Jost, Glaser, Kruglanski, Sulloway, 2003). Thus, in general, we expect conservative right-wing newspapers will tend to represent immigration issues in different ways than left-wing Liberal newspapers, with the former being more likely to frame immigration as a threat to the country. Another important factor is the type of audience the media outlet targets. We then expect that tabloids, which tend to be more sensationalistic than broadsheet, will favour the framing of immigrants as a threat.

Subsequently, we want to verify what the salience of these frames in the immigration discourse is in an inductive way. Although they are not identical processes, and although framing include a broader range of cognitive reasoning than agenda-setting, similarities can be found between the two processes, particularly between second-level of agenda setting and framing. Both focus indeed on the most salient aspects or sub-issues of the issue of interest (McCombs, 1997: 37; Weaver, 2007). In this study, we therefore use the second-level of agenda setting as a validation of the relevance of the deductive frames analyzed before.

For doing this, we use what we can call a specific application of the automated-version of the agenda-setting approach. Instead of looking at the salience of immigration within the electoral campaign issues, which we already know, we look at the salience of different sub-issues of immigration within news about immigration. How is the immigration discourse portrayed in the media? What topics shape the public conversation on it? The relevance of sub-issues within the immigration discourse somehow relates to the way news are framed. By identifying which issues these are, and how they evolved over time, we could validate whether the two deductive-based frames used in the first part of the analysis, are actually the sub-topics that mostly shape the public conversation about immigration. This has the advantage of putting the frames into a wider picture and get a deeper sense of how the specific electoral debate has evolved over time.

3 Immigration Frames

We therefore consider two/four news frame to be of particular relevance:

- Framing immigration in terms of economic benefits
- Framing immigration in terms of economic threat
- Framing immigration in terms of cultural enrichment
- Framing immigration in terms of cultural conflict

In addition to framing of immigration in one of these frames, it is important to consider also the visibility of these frames compared to different portrayals of immigration in the news. In order to detect the prevalence of these frames, we need to decide upon an appropriate set of words that indicate when this frame is being applied.

4 Data Selection

We selected newspapers according to their circulation, including both broadsheets and tabloids, if available electronically via Factiva for the period we have studied. We also included newspapers of different ideological orientation.

Table 1 displays the names of the newspapers included in our study, along with the type of newspaper, their political orientation, and the number of articles talking about immigration that are included in the corpus of data from each newspaper.

[table goes here]

5 Text-Mining Techniques

We describe the use of two quantitative text analysis techniques: Topic modelling with Latent Dirichelet Allocation (LDA), and frame analysis a model of word meanings in a semantic space.

5.1 Topic Modelling

Latent Dirichelet Analysis (LDA) is a widely used method for identifying clusters of words which represent the topics present in a set of documents. LDA is a probabilistic model which treats each document as a mixture of topics, where each word in the document belongs to exactly one topic. In this model, topic structures are hidden variables which can be inferred from the observed variables — the distribution of words in the documents.

[...]

An accessible overview of the model can be found in [Blei \(2012\)](#). Topic modelling with LDA is a technique from computational linguistics which has been widely applied in other disciplines. To apply LDA to our collection of documents, we experimented with three freely available topic modelling software packages: *gensim* a semantic analysis package in python; *topicmodels*, a topic modelling package in R, and *MALLET*, a collection of statistical natural language methods implemented in Java. The results presented here were generated using MALLET.

5.2 Frame Analysis

While LDA has been widely applied, there is less consensus on which method to use to discover how issues are framed in text. In this paper, we model our methodology on the approach taken by [Sagi, Diermeier and Kaufmann \(2013\)](#), who quantify how Republicans and Democrats in US Senate debates frame issues such as abortion and terrorism. In their analysis, [Sagi, Diermeier and Kaufmann \(2013\)](#) choose specific terms which indicate particular way of framing the issue — for example, the terms ‘choice’ and ‘life’ reliably indicate two different attitudes to the abortion debate. Similarly, the terms ‘crime’ and ‘war’ are representative of two different views of the issue of terrorism. Following their approach, we construct co-occurrence vectors for words we believe might be associated with particular ways of framing the immigration issue.

Once these indicative words have been chosen, an association between the issue of interest (in this case immigration) and each of the framing terms must be measured. We measure this association by calculating the similarity of the contexts in which the terms occur. Follow-

ing [Sagi, Diermeier and Kaufmann \(2013\)](#), we use the Java software package *semanticvectors* to do this. First, we extract all of the contexts in which the target term or one of the framing terms occurs in the collection of articles — for each word this is a context window of length 50: 25 words preceding the term of interest and 25 words after.

For each term, we then have a set of context vectors. These are large, sparse matrices, where the columns are all of the words in the vocabulary of the corpus, and the row indicates how many times each of these words has occurred in the current context vector. The context vectors are summed and normalized, and the dimensionality of the resulting large matrix is reduced by singular value decomposition. The cosine similarity between the reduced, combined context vectors can then be used as a measure of association between the two words in the document collection. This approach falls under the umbrella of Latent Semantic Analysis, a model of word meaning whereby a word is considered to be a point in a semantic space constructed from the contexts in which it occurs. The context may be an entire document (which enables a kind of word clustering not dissimilar to LDA), or a smaller window of words, as we have described in this section.

To compare framing across different variables, the document collection can be separated into two sub-corpora according to the variable of interest. This was political party in the case of [Sagi, Diermeier and Kaufmann \(2013\)](#); for this paper we are interested in how the issue of migration is framed by different newspapers. Therefore, we divided the corpus according to newspaper source and computed different context vectors for words in each of these sub-corpora.

6 Data Analysis

We used Factiva to download every document in the UK press from the 12 weeks preceding the European Elections that contained one of the following terms: *immigrant*, *immigrants*, *immigration*, or *immigrate*². Factiva’s bulk download allows 100 articles to be downloaded at once, which are retrieved as single .rtf files — each file contains 100 articles, with a header containing metadata such as source and date before each article.

²Due to an error while collecting the data, no articles from The Daily Mirror were retrieved.

We used a Python script to split each article into a separate text file and label each one with its date and the newspaper in which it was published. We also created a mapping between the names of the online and offline versions of the same newspaper, to allow us to merge these. We used Factiva’s option to remove duplicates. Initially, we had 3950 articles, but we made the decision to restrict our analysis to national newspapers with large circulations. This resulted in a final set of 2833 documents.

6.1 Text pre-processing

The topic modelling and semantic vectors methods operate on term-frequency matrices of words. For a given portion of text, the data input to the algorithm is a matrix where the columns are word types and the rows indicate the frequency of the word in the portion of text. The portion of text may be a whole document (as it is for LDA) or a context window around a term of interest (as it is for the semantic vectors approach). Representing natural language text in this way is sometimes called the ‘bag-of-words’ model, as the resulting matrix does not make any use of the order in which the words occurred in the original text.

The MALLET package allows for a list of stopwords to be excluded from the analysis. A standard list of stopwords includes semantically light function words such as conjunctions and common prepositions. To this list, we appended some terms specific to our corpus — for example we needed to exclude the titles of the newspapers, the phrases ‘copyright’ and ‘all rights reserved’, and the search terms.

A common pre-processing step is to apply a stemmer to the bag of words so that words differing only in morphological stem are combined — for example, the counts for *earn*, *earns*, *earning*, and *earned* would be summed into a total for the stem *earn**. We did not apply this step for the LDA modelling as it makes the word clusters a little more difficult to read, and in any case words with the same stem tended to be assigned to the same topic.

6.2 Topic modelling analysis

The LDA process requires that the number of topics to be discovered is specified in advance. If the number of topics is small, the topics will be general (for example, there might be a single

No.	Label	Prominence	Top five words
0	-	0.04626	war book world books benn
1	-	0.09065	eu european europe britain cameron
2	-	0.08557	home office uk asylum case
3	-	0.17853	cameron minister tory david prime
4	-	0.06529	clegg farage debate nick nigel
5	Benefits	0.15959	tax benefits work pay money
6	-	0.0775	london local council east south
7	-	0.05039	film music theatre play story
8	-	0.02568	wine food world huffington fish
9	-	0.38159	time dont back good day
10	Women/Religion	0.03672	women church muslim christian country
11	-	0.04242	bbc radio tv show series
12	-	0.09385	news newspapers thesun english
13	-	0.09246	court year police home human
14	-	0.05685	love sex men women book
15	-	0.09175	family children mother home parents
16	-	0.02815	gardiner mp labour asked facebook
17	-	0.0582	british workers minister brokenshire labour
18	-	0.08713	britain country british world english
19	-	0.0692	mail damonl newspapers daim
20	Business	0.07714	business uk market prices house
21	-	0.04717	police man men french found
22	UKIP Campaign	0.12362	ukip farage party racist nigel
23	Schools	0.05901	school english schools students children
24	-	0.12003	care health report public government
25	Minorities	0.0748	ethnic cent white minority groups black
26	-	0.25435	uk migration year eu britain
27	-	0.39536	public political media social left
28	-	0.05902	scotland scottish border uk independence
29	-	0.18485	ukip party labour voters vote

topic for economic words), but these topics might be indistinct and too similar to each other to be useful. We began by experimenting with different numbers of topics until the resulting word groups seemed to indicate distinct, fine-grained topics relating to the immigration debate. We found that specifying 30 topics gave the best results.

The words in each topic are indicated in the table below. For topics that we chose for further analysis, the first column is a general label that we chose (after the fact) to refer to this topic. The second column indicates the portion of the entire corpus that each topic covered. The words in each topic are ordered by how much they contribute to the topic.

From the overall list of 30 topics, we chose a subset to examine in more detail. Some of the topics seem to be indistinct mixtures of function words or words with particular relevance

to a single story, and we did not examine these further. Instead, we compared topics relevant to cultural and economic framing of the immigration debate, both across the corpus as a whole and within particular newspapers and particular dates of publication.

Some LDA implementations require that each topic is of equal prevalence across the corpus as a whole. However, we use the hyperparameter optimization option in MALLET which allows for different topics to have different prominences across the corpus. Therefore, we can measure the prominence of interesting topics for the complete set of articles, shown in Figure 1. From the six topics that we chose to focus on, it seems as though economic issues were more prominent in the debate during this time than cultural issues.

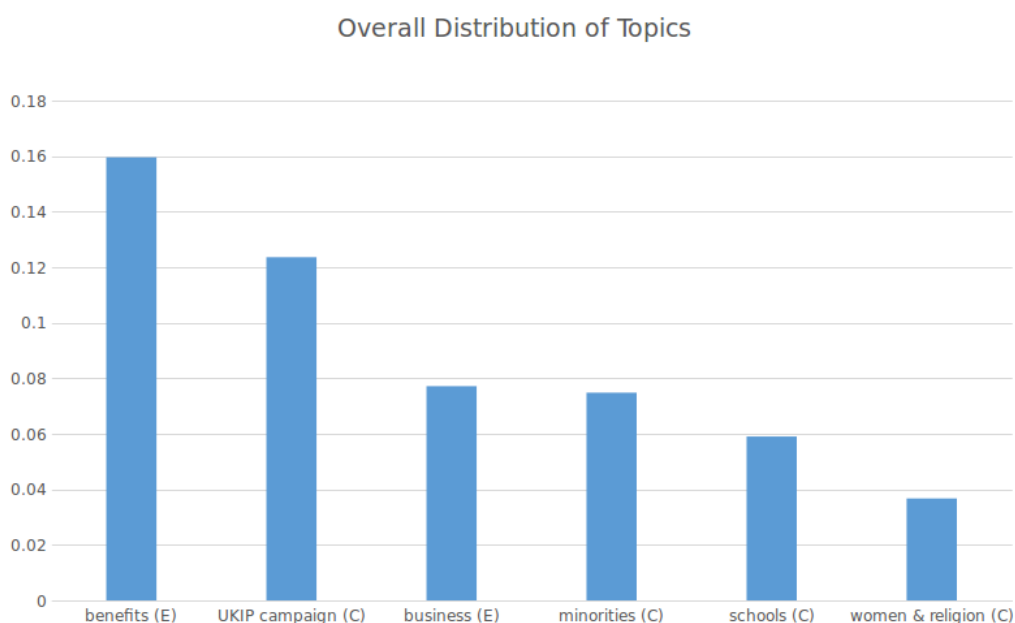


Figure 1: The Dirichelet prior for each topic across the whole corpus. This number is roughly proportional to the overall portion of the collection assigned to each topic (McCallum, 2002), but, unlike the topic-document scores, is not an exact fraction.

We then measured the average prominence of each of the topics in each of the newspapers. In the figures below, the y-axis indicates the mean proportion of a document the newspaper indicated on the x-axis that is composed of words assigned to the topic of interest. The x-axis shows 7 major daily UK newspapers, approximately listed in order of their broad ideological stance.

6.3 Economic Topics

Figure 2 shows the distribution of the ‘Benefits and Welfare’ topic across seven newspapers. Although it is possible for the same word to occur in the most prominent words for multiple topics, for a particular document, each word is assigned to only one topic. This means that for a given document, we can say what proportion of the document is assigned to each topic. These charts show the mean topic proportion for each newspaper. The sum of the topic proportions across all 30 topics is 1. That is, on average, 8.6% an article from *The Daily Express* is composed of words assigned to the ‘Benefits and Welfare’ topic, compared to 3.6% for an article from *The Guardian*.

Figure 3 shows the topic proportions for the ‘Business’ topic. The papers traditionally considered in the ‘broadsheet’ category have higher proportions of text about this topic.

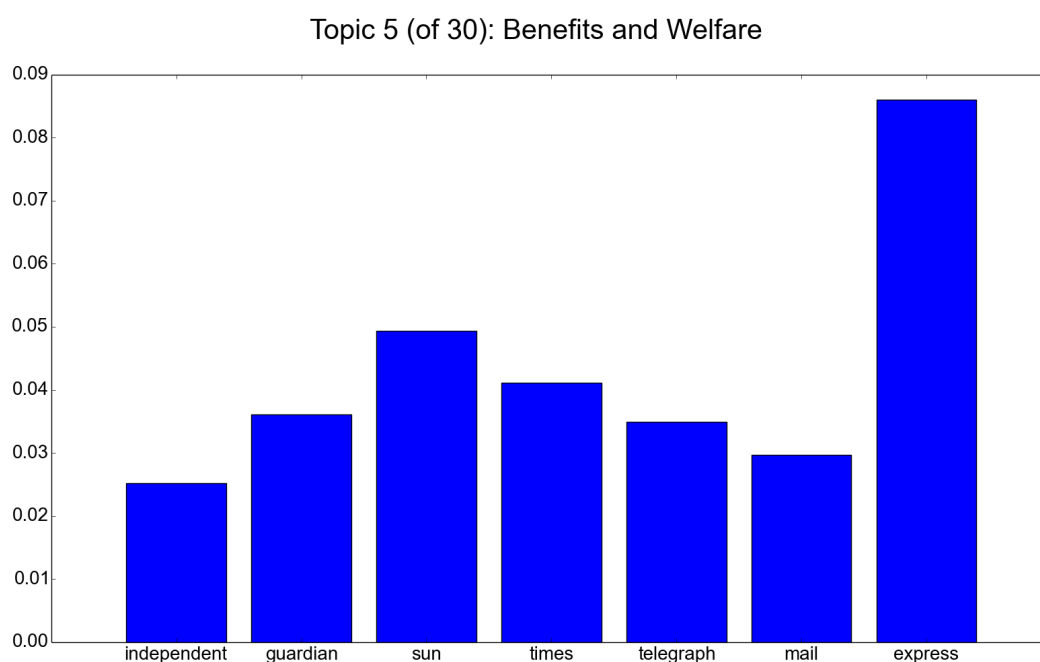


Figure 2: The benefits and welfare topic by newspaper source. The top twenty words in this topic were: *tax, benefits, work, pay, money, million, benefit, government, jobs, welfare, income, cost, taxes, state, public, budget, system, paid, pounds, billion*.

We initially included the Financial Times in the article collection, but the results indicated that it was largely composed of the Business topic, as in figure 4.

In a way this is a useful kind of face-validation — we expect that the FT will cover business issues more than other sources. In addition, the distribution of the other papers in

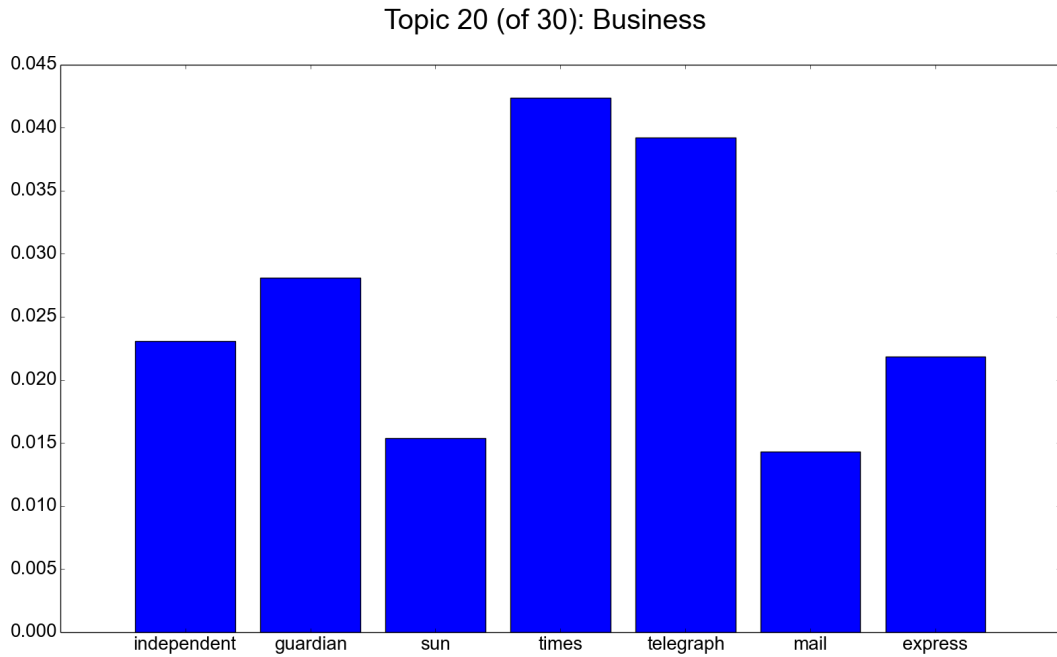


Figure 3: The business by newspaper source. The top twenty words in this topic were: *business uk market prices house year company economy cent growth million london investment businesses housing bank companies capital years high*

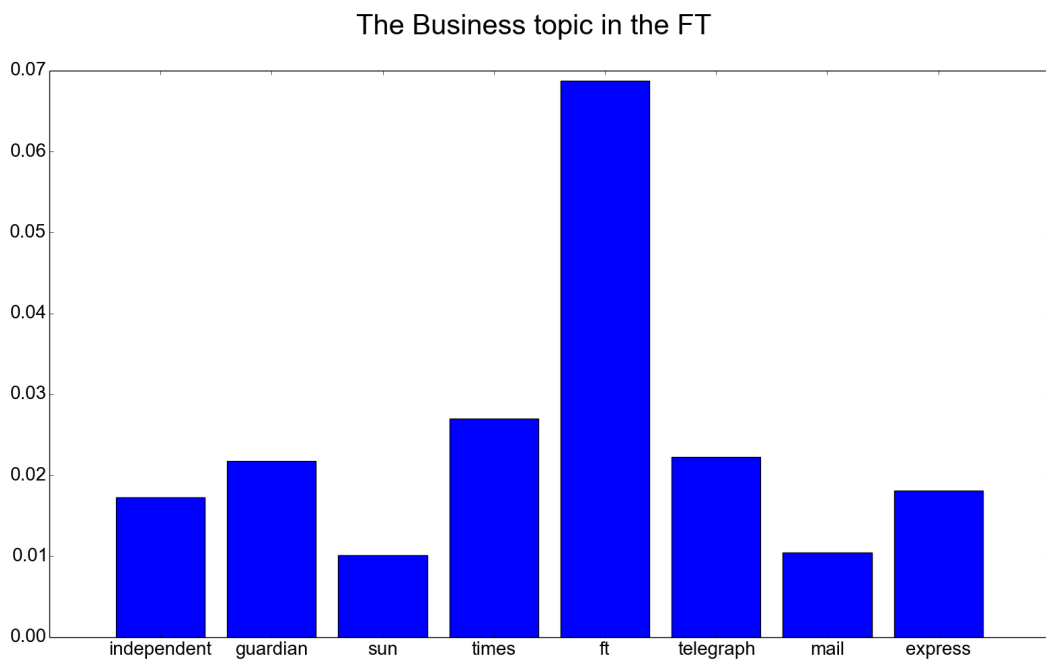


Figure 4: The *business* topic by newspaper source including the Financial Times. The top twenty words in this topic were: *business uk market prices house year company economy cent growth million london investment businesses housing bank companies capital years high*

this model is very similar to their distribution in 3, despite the fact that it is a different run of the LDA model, where the topics will be assigned different numbers and composed slightly differently due to the extra articles and different values for the random initialization of the topic prior.

6.4 Cultural Topics

Figures 5 and shows the topic proportions for the ‘Minorities’ and ‘Women and Religion’ topic.

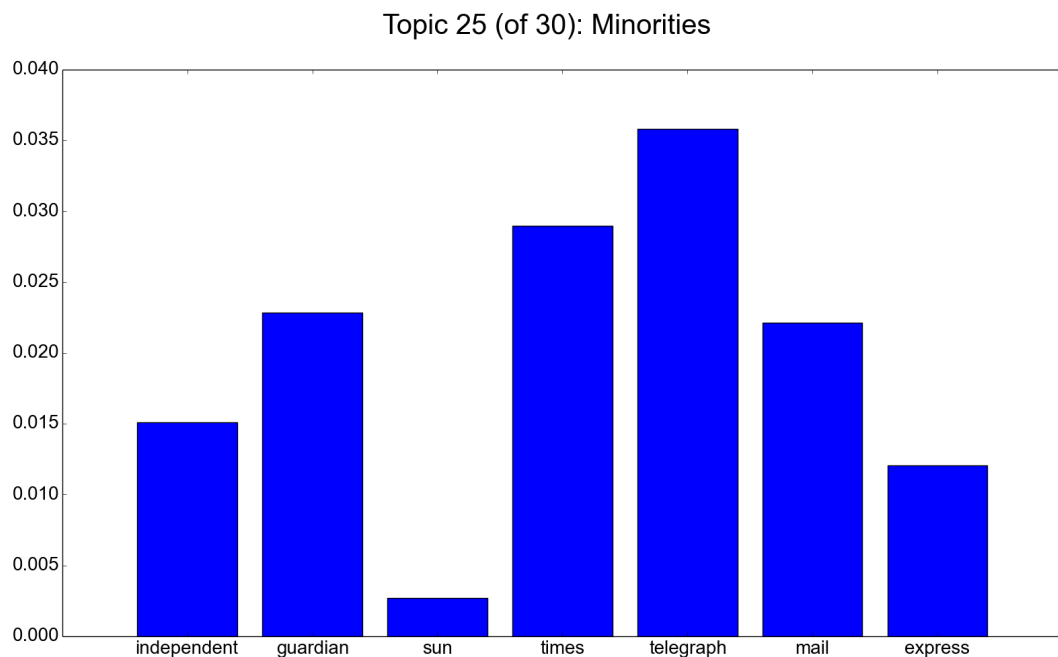


Figure 5: The *minorities* topic by newspaper source. The top twenty words in this topic were: *ethnic cent white minority groups black population study britain british minorities english university report research communities class found born online*

We also observed how the prominence of topics varied over time. We divided the articles according to the week in which they were published, which allows us to see the average proportion of a document that is composed of a given topic for in a give week. The x axis corresponds to the week of the year in which the article was published (we collected articles from the 22nd February to 22nd May 2014). The results for this time-series analysis show strong spikes at particular points, indicating that large amounts of the text written about a particular topic was driven by a single story or closely related group of stories.

[time series charts]

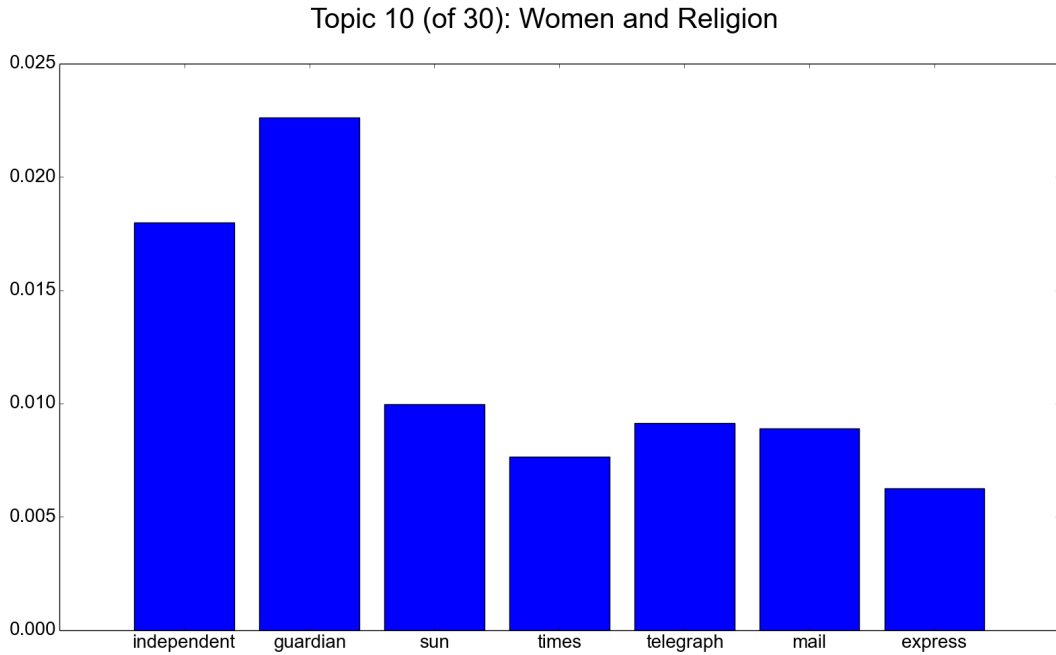


Figure 6: The *minorities* topic by newspaper source. The top twenty words in this topic were: *women church muslim christian country gay violence female faith muslims girls religious archbishop uk culture catholic britain england jewish countries*

6.5 Frame Association Analysis

It is difficult to select terms that precisely delineate the framing of immigration issues in the same way that ‘choice’ and ‘life’ define the abortion debate. Nonetheless, we compiled a list of words to capture aspects of the economic and cultural debate around immigration in the UK. We then measured the association score for each of these words with the word stem *immigr**. For this analysis, we again removed stopwords from the term-frequency matrix, and the semantic vectors package also performs word stemming. We compared the association scores across three broadsheet, three tabloid, and a group of three broadly right/centre-right newspapers. The charts below show the results.

These results are difficult to interpret. It is not appropriate to compare the overall association score for words across all documents, as the general frequency of the word will effect its association — work will have a higher frequency than Bulgarian simply because it is a more frequent term overall. Comparing lexical relations among terms with asymmetric overall frequencies is an ongoing problem in natural language processing.

We might expect that right-wing newspapers which are generally perceived to be opposed

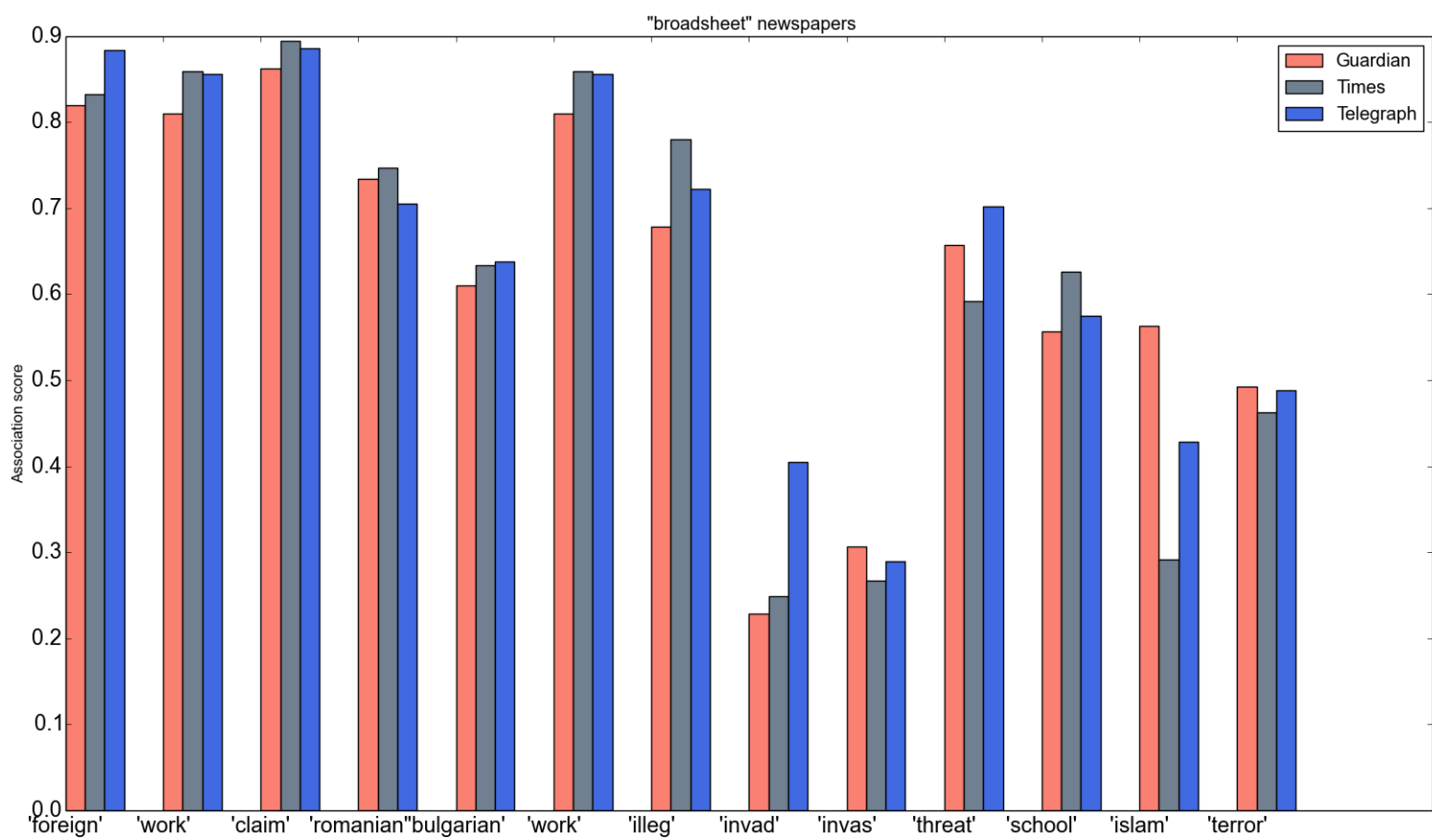


Figure 7: Association scores indicating framing of the immigration debate in broadsheet newspapers.

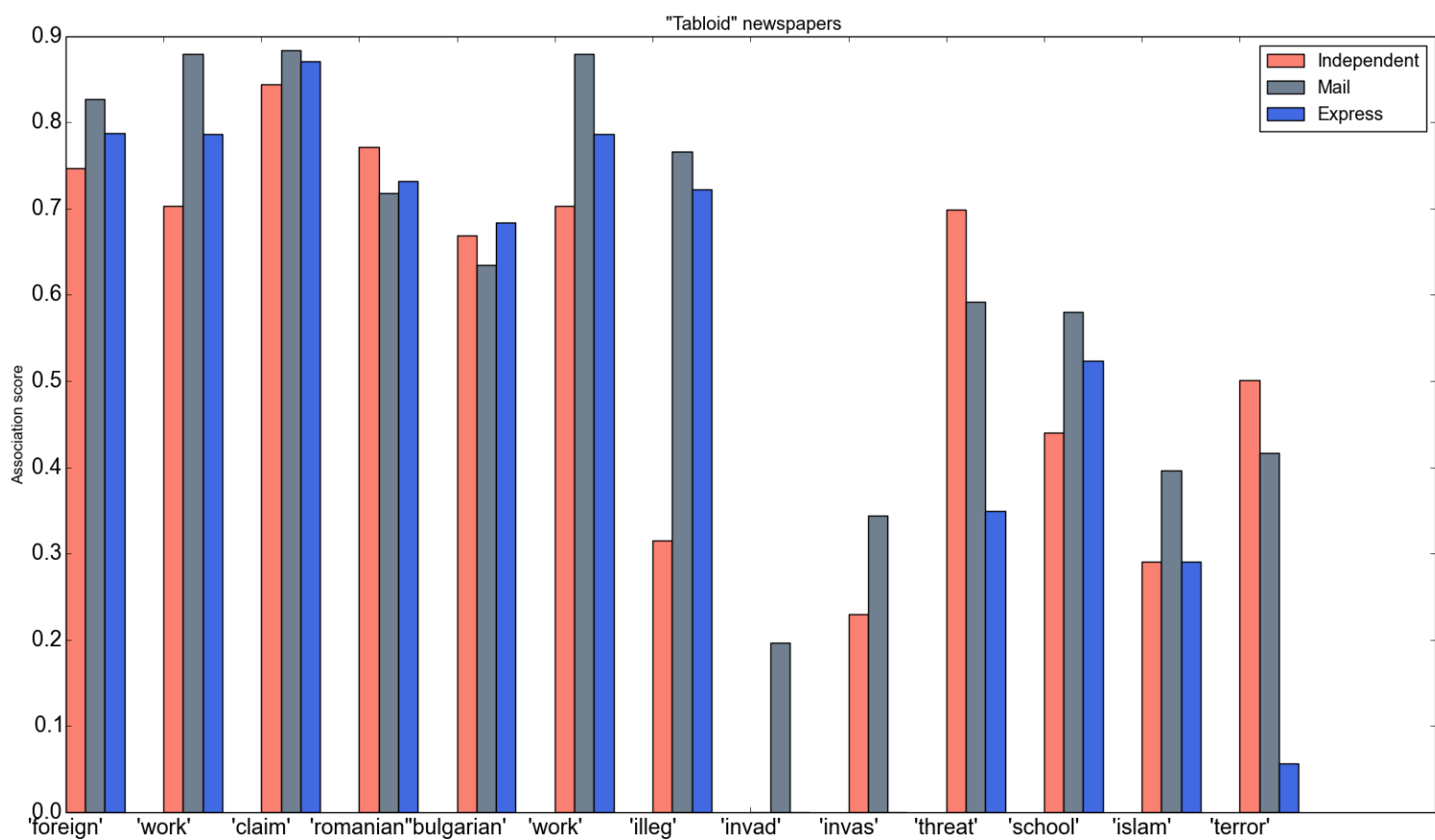


Figure 8: Association scores indicating framing of the immigration debate in tabloid newspapers.

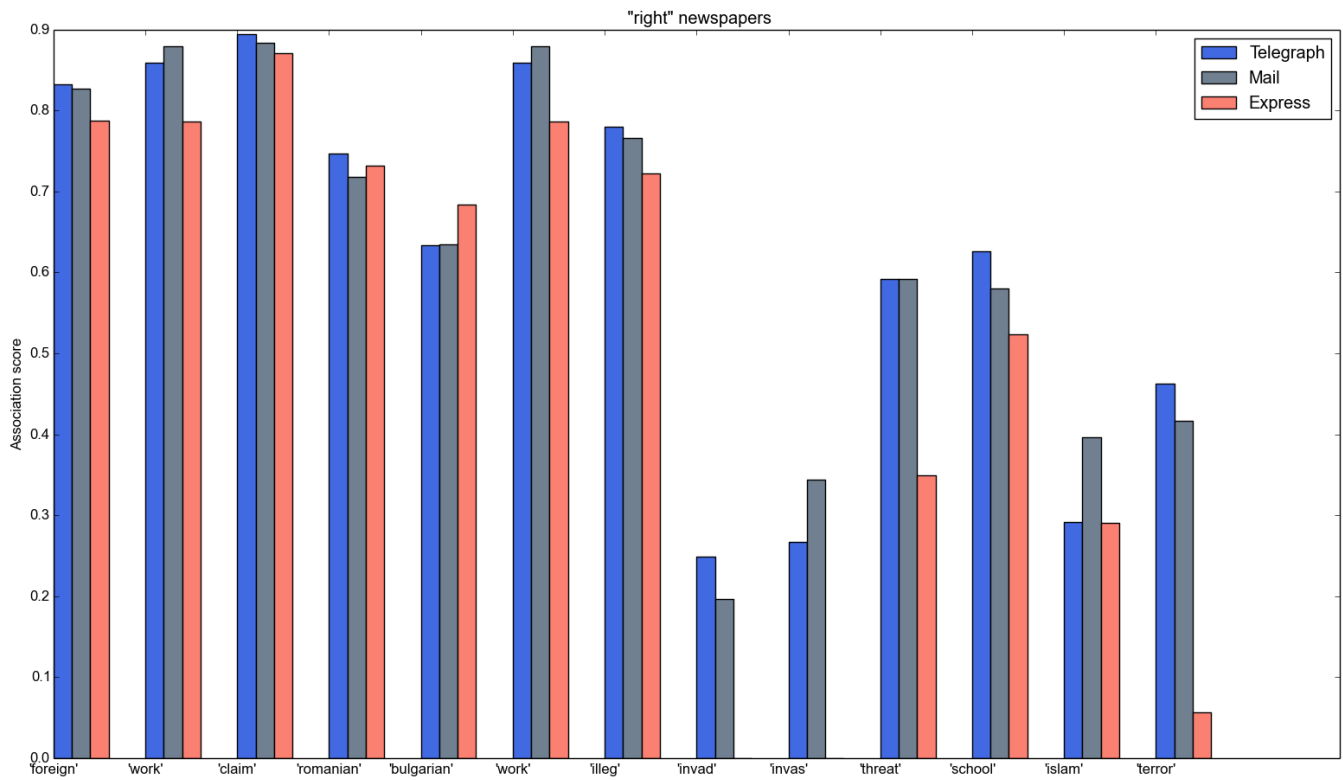
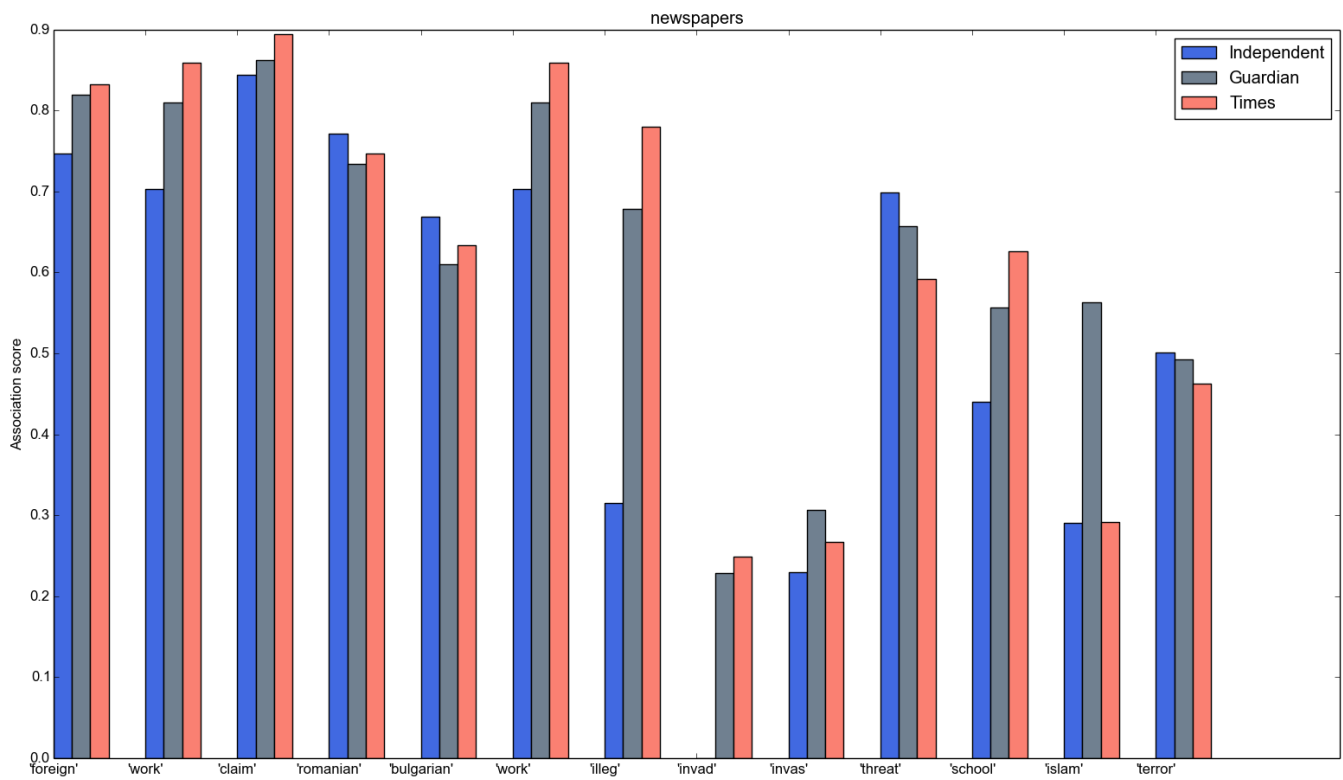


Figure 9: Association scores indicating framing of the immigration debate.

to immigration would be more likely to frame immigration as *invasion*, yet the results indicate that *The Daily Express* does not use the stem *invad** or *invas** within 50 words of *immigr**. We confirmed this result with a simple text search of the 352 Daily Express articles, and found that the terms invade or invasion did not occur in the documents. Furthermore, when the term did occur in other newspapers, we found that it was often used in a sarcastic manner to caricature the language of opponents to immigration, e.g.:

predictions of a mass invasion of the uk by foreigners were as accurate as saying
five million scots will invade orpington if edinburgh splits from the uk — *Daily Mail*, 5th April 2014

Closer inspection of articles in The Guardian containing the stem terror* again highlight the difficulty of selecting specific terms to identify particular ways of framing the debate:

she spoke of the terror she feels at the prospect of going back to her home country
— *Guardian*, 24th March 2014

7 Discussion and Conclusion

References

- Blei, David M. 2012. “Probabilistic topic models.” *Communications of the ACM* 55(4):77–84.
- McCallum, Andrew Kachites. 2002. “MALLET: A Machine Learning for Language Toolkit.” <http://mallet.cs.umass.edu>.
- Sagi, Eyal, Daniel Diermeier and Stefan Kaufmann. 2013. “Identifying Issue Frames in Text.” *PLoS one* 8(7):e69185.