

To appear in: *International Journal of Corpus Linguistics* in 2022.

## **Keywords through time: Tracking changes in press discourses of Islam**

Isobelle Clarke

Gavin Brookes

Tony McEnery

This paper applies a new approach to the identification of discourses, based on the use of Multiple Correspondence Analysis (MCA), to the study of discourse variation over time. The MCA approach to keywords deals with a major issue with the use of keywords to identify discourses – that is, the allocation of individual keywords to multiple discourses. Yet, as this paper demonstrates, the approach also allows us to observe variation in the prevalence of discourses over time. The MCA approach to keywords allows the allocation of individual texts to multiple discourses based on patterns of keyword co-occurrence. Metadata in the corpus data analysed – in this study, UK newspaper articles about Islam – can then be used to map those discourses over time, resulting in a clear view of how the discourses vary relative to one another as time progresses. The paper argues that the drivers for these fluctuations are language external, i.e. the real world events being reported on in the newspapers studied.

### **Keywords**

Keyword analysis, Multiple Correspondence Analysis, Newspaper discourse, Islam, Corpus-assisted discourse studies

### **1. Introduction**

In corpus linguistics, keywords can be understood as words which occur with a significantly higher frequency in the corpus under analysis in comparison with another corpus. Keywords can offer possible analytical signposts to the linguistic traces of discourses (Baker, 2004), which we understand to be ‘set[s] of meanings, metaphors, representations, images, stories, statements and so on that in some way together produce a particular version of events’ (Burr, 2015: 74-75).

Keywords are *signposts* because they do not necessarily map onto discourses straightforwardly but have to be analysed and interpreted within their wider textual settings to determine whether and how their presence contributes to a discourse.

As a technique, keywords is still relatively new, with work using it only beginning on any scale with the development of *Wordsmith Tools* in the late 1990s (Scott, 1996). While there is little doubt that the technique has proved helpful in corpus studies of discourse, there is still

To appear in: *International Journal of Corpus Linguistics* in 2022.

scope for its development. Work on measures for assessing keyness (e.g. Gabrielatos, 2018), their use in controlling collocation networks (McEnery, 2005), and the role that dispersion may play in improving the technique (Egbert and Biber, 2019) are some proposed refinements of the technique. This paper is concerned with one aspect of keyword analysis that has been relatively neglected to date – change over time. We explore this here in the context of corpus-assisted discourse studies (CADS) which is ‘a subset of corpus linguistics: “that set of studies into the form and/or function of language as communicative discourse which incorporate the use of computerised corpora in their analyses”.’ (Partington et al., 2013: 10). We explore what value the analysis of keywords can bring to the study of temporal variation in discourses by focusing on a topic that has received sustained attention in CADS: the representation of Islam and Muslims in the UK press. In part, our choice of topic is intended to promote critical reflection on our technique for looking at keywords over time in this paper – studies such as Baker et al. (2013) and Baker and McEnery (2019) has already provided a sense of how this representation has shifted between 1998 and 2014. Yet the analysis of keywords over time in these studies was achieved at a very coarse level of granularity through the contrast of two time periods, e.g. 1998-2009 (Baker et al., 2013) and 2010-2014 (Baker and McEnery, 2019), and is relatively subjective, in the sense that it was left to the skill of analysts to group the keywords and disaggregate the discourses. In looking at our technique for deriving keywords and focusing on the question of change over time, we wish to show the potential of the technique presented here to control, guide and provide a more fine-grained, keyword-driven analysis of discourse through time.

In pursuit of these goals, we will firstly briefly review previous work on the representation of Islam and Muslims in the UK Press. We then explore existing approaches taken to time using keywords and briefly present the approach to keywords taken in this article (after Clarke et al, 2021) based upon Multiple Correspondence Analysis (MCA), and explain how it has been used to derive discourses from our corpus. We then turn to the question of how those discourses, when mapped through time, allow the MCA of keywords to reveal temporal variation in discourses. In exploring those discourses, we will look at events as reported in the press which may explain observed fluctuations.

To appear in: *International Journal of Corpus Linguistics* in 2022.

## **2. Islam in the press: Change over time**

British press representations of Islam and Muslims have been widely researched using qualitative, critical (e.g., Richardson 2004) and quantitative, corpus linguistic approaches. Due to space limitations, we will not review this extensive body of work here (see Baker et al. (2013) for an overview). Instead, we focus on the studies by Baker et al. (2013) and Baker and McEnery (2019), since these used keywords as a primary driver of their analyses. When considered as a pair, these studies also present a commentary on change over time.

Baker et al. (2013) conducted a study of representational discourses around Islam and Muslims in a 147-million-word corpus of UK national newspaper articles published between 1998 and 2009 (inclusive). To explore temporal variation in their data, they employed a range of techniques and approaches, including analysing keywords and the collocates of the most frequent words referring to Islam across all years (*Muslim, Muslims, Islam* and *Islamic*), close-reading and mapping the fluctuating frequency of words over time (for details see Gabrielatos et al. (2012)). They reported, *inter alia*, a gradual move towards more personalising stories about Muslims (as opposed to the abstract concept of Islam) and a gradually increasing focus on stories about Muslims in a UK context (as opposed to other countries). Some concepts were prevalent across the corpus, like extremism, while others, such as veiling and references to tolerance/hatred, were more restricted to particular time periods. Changes in discussions of homosexuality were also noted; in the earlier period the press described the common ground shared by Muslims and gay people as oppressed groups within society, while from 2006 the discussion was more critical of Islam, linking it to homophobia and contributing to a discourse in which Muslims were represented as not meeting ‘British’ standards for tolerance.

In a follow-up study, Baker and McEnery (2019) explored representational press discourses around Islam between 2010 and 2014. They used the keywords technique to compare two corpora, with the 1998-2009 corpus serving as a reference corpus to find words which were key in the latter period (2010-2014). They noted four broad areas of relative stability across the two periods. First, the relative frequency of *islamic* remained fairly stable, and it retained its association with extremism. Secondly, Islam and Muslims remained strongly associated with conflict. Thirdly, a ‘horror discourse’ around veil-wearing was persistent, with shrouds and veiling women evaluated negatively. Finally, the phrase ‘devout Muslim’ retained its negative associations, with Muslims described as ‘devout’ represented as cheating, becoming radicalised,

To appear in: *International Journal of Corpus Linguistics* in 2022.

engaging in extremist activity, or clashing with ‘western’ cultural values. Yet even these areas of relative stability exhibited slight changes – for example, while the association of *islamic* and extremism remained, it had also intensified, as had the focus on violence.

Overall, however, stability was the exception. In terms of change, Baker and McEnery (2019) found that keywords indicated changes in two broad discourse topics: (i.) countries and conflict and (ii.) extremism. With countries and conflict, focus shifted away from conflict in Iraq, Palestine, and America, which dominated the 1998-2009 data, towards Syria, Libya, Iran and Egypt. Conflict words such as *war*, *terrorist*, *suicide*, *raids*, and *hijackers* fell in relative frequency, while words like *islamist(s)*, *rebels*, *crisis*, *revolution*, *protesters*, and *jihadi(st)* rose. Baker and McEnery (2019) argued that this pattern reflected changes in the types of global events in which Islam and Muslims were reported; the initial period focused on large-scale orchestrated attacks, while the latter period focused on civil forms of conflict, producing keywords such as *crisis*, *revolution*, *protests*, and *uprising*.

### **3. Keywords and time**

Studies using keywords to examine variation over time have tended to rely, as Baker and McEnery (2019) did, on temporally structured corpora and sub-corpora to identify keywords that reveal discourses and types of language use that are statistically salient at one time period relative to another. For example, to study how the language and ideology of the British Labour Party changed during the ‘New Labour’, Fairclough (2000) obtained keywords from a corpus of documents, speeches and newspaper articles from this period by comparing it against one consisting of older Labour texts. Other studies have tracked the development of phenomena over a continuous time period by comparing multiple sub-corpora representing consecutive years. For example, Baker et al. (2020) analysed changes in press discourses around obesity between 2008 and 2017 by dividing their corpus into yearly sub-corpora and calculating keywords for each year in comparison to all other years, combined. Brookes and Baker (2021) extended this analysis by using keywords to study discourses that were characteristic of obesity coverage at particular points in the year by comparing each month against the rest (e.g., comparing coverage in January across all years against all other months, across all years, combined).

As may be inferred from the examples so far, dividing a corpus in different ways for a keyword analysis will produce different sets of keywords and hence, in this case, different views

To appear in: *International Journal of Corpus Linguistics* in 2022.

of change through time (Marchi, 2018). There is no intrinsically optimal approach; the approach taken to structuring time in the corpus, or the method for comparing time periods, will depend on the research question. Fairclough (2000) divided time by a transition of a political party from one state to another. Partington (2010) chose specific sample points to compare a non-continuous sequence of points in time in his study comparing newspaper language in 1993 and 2005. Brookes and Baker (2021) use a continuous sequence of data and impose a top-down segmentation of it by year and month. All of those studies then take a different approach to establishing keywords – two long periods of time contrasted with one another (Fairclough, 2000), pairwise comparison of years of interest (Partington, 2010) and one segment of time against all remaining segments of time (Brookes and Baker, 2021).

Yet like other analytical methods, keyword analysis is not flawless and many of its limitations, by now well-documented (see Baker, 2004), are visible in studies of diachronic variation, such as its emphasis on presence over absence, its tendency to highlight difference at the expense of similarity, and the impact of corpus segmentation on the representation of time. The technique also often aggregates discourses – the keywords provided may be associated with multiple discourses present in the corpus. Prising apart the discourses indexed by the keywords is a matter for the analyst. To make analysis more manageable, and to aid the recognition of patterns beyond the level of single keywords, analysts often group their keywords into thematic or semantic categories. Different approaches have been taken to this. For example, Baker et al. (2020) categorised keywords thematically through manual analysis of their uses in context before comparing the categories' relative frequencies to identify which ones had risen, fallen, or stayed relatively stable in frequency over time. Partington et al. (2013), meanwhile, used dictionary definitions as the basis for identifying 'sets' of evaluative lexical keywords in their study of changes in newspaper language. All approaches have limitations, though, and the creation of meaningful categories and the assignment of keywords to those categories often involves some element of compromise, especially as keywords may contribute to more than one discourse. Determining where this happens is, again, a matter for the analyst.

While assigning keywords to categories can have the advantage of illuminating broader trends, it can also result in keywords being mischaracterised and the generalisations extrapolated from frequency trends being, accordingly, inaccurate. This challenge arguably becomes more pronounced the more that texts in our corpus vary. For corpora representing long time periods,

To appear in: *International Journal of Corpus Linguistics* in 2022.

we also need to be mindful of the potential for words which are keywords or collocate with keywords to change usage over time (Partington (2010: 94-95)).

Developing categories that truly reflect the potentially wide variety of uses any given keyword can be put to within the context of a corpus can be labour-intensive and may result in a proliferation of categories which defy analysis in any depth. In the following section we address such issues by proposing a technique that: i.) is more sensitive to such variation; ii.) assists the analyst in disaggregating discourses within keyword lists to facilitate large scale data analysis, and iii.) provides a more accurate and refined allocation of keywords to discourses, including the assignment of a given keyword to multiple discourses, where appropriate.

#### **4. Data and approach**

##### 4.1. The corpus

We constructed a corpus to extend the previous studies of Islam and Muslims in the UK press by keeping our frame of reference as close to the two studies (Baker et al. 2013 and Baker and McEnery, 2019) as possible. This entailed i.) using the same search terms<sup>1</sup> as those used by the two studies, and ii.) searching all UK national newspapers<sup>2</sup> available in the news consolidator *LexisNexis* from 1<sup>st</sup> January 2015 to 31<sup>st</sup> December 2019. While the titles covered are generally the same, there are some changes; for example, the *News of the World* ceased publication in 2011 – so while it is in part of the corpus built by Baker and McEnery (2019), it is entirely absent from our corpus. Our corpus is also larger than the previous corpora, totalling 382,946 texts (325,084,672 words), which is double the size of the corpus in Baker et al., (2013) and over three times larger than that built by Baker and McEnery (2019). This increase in size may be due to an increased number of newsworthy events involving Muslims, the general increase in news

---

<sup>1</sup> Alah OR Allah OR ayatollah! OR burka! OR burqa! OR chador! OR fatwa! OR hejab! OR imam! OR islam! OR Koran OR Mecca OR Medina OR Mohammedan! OR Moslem! OR Muslim! OR mosque! OR mufti! OR mujaheddin! OR mujahedin! OR mullah! OR muslim! OR Prophet Mohammed OR Q'uran OR rupoush OR rupush OR sharia OR shari'a OR shia! OR shi-ite! OR Shi'ite! Or sunni! OR the Prophet OR wahabi OR yashmak! AND NOT Islamabad AND NOT shiatsu AND NOT sunnily.

<sup>2</sup> Newspapers included *Daily Star Online*, *Daily Star Sunday*, *Daily Star*, *Express Online*, *Guardian.com*, *i-Independent Print Ltd*, *Independent Magazine*, *Independent Traveller*, *Independent.co.uk*, *MailOnline*, *mirror.co.uk*, *Sunday Express*, *Sunday Mirror*, *telegraph.co.uk*, *The Daily Mail*, *The Daily Mirror*, *The Daily Telegraph*, *The Express*, *The Guardian*, *The Independent on Sunday*, *The Independent*, *The Mail On Sunday*, *The Mirror*, *The News of the World*, *The Observer*, *The People*, *The Sun*, *The Sunday Telegraph*, *The Sunday Times*, *The Times*.

To appear in: *International Journal of Corpus Linguistics* in 2022.

production, and the inclusion of online versions, which can be continuously edited, producing a new text for each edit. Accepting such differences, we then combined our corpus with the corpus from Baker and McEnery (2019) and, using bespoke Perl scripts, we cleaned the texts, putting the metadata and text in xml format. Our resulting corpus covers press reportage of Islam and Muslims between 1<sup>st</sup> January 2010 and 31<sup>st</sup> December 2019 (497,523 articles, 395, 930,045 words).

#### 4.2. The MCA Approach to Keywords

We ran MCA keyword analysis on our corpus, as reported in Clarke et al. (2021). This previous study had two aims – firstly to see if the MCA would confirm or challenge the findings of Baker and McEnery (2019) and then, using our corpus, to explore what had happened in the later part of the decade. In what follows, we briefly explain the MCA approach, detail the discourses we discovered using the technique, compare the findings with those of Baker and McEnery (2019) and, finally, note how the representation of Muslims and Islam developed in the latter half of the decade. More detail on the MCA approach to deriving keywords can be found in Clarke et al. (2021).

MCA is a technique which identifies relationships between individuals and categorical variables (Benzécri, 1979). In our case, we conceive of the individuals as the texts within a corpus and the variables associated with them as the keywords of that corpus. There were multiple ways in which we could have calculated our keywords and each of these choices would have likely produced a different keyword list to some degree. One approach could have been to collect a reference corpus of general news from a similar time period to retrieve an entire corpus keyword list. Another approach could have involved segmenting the corpus into temporal sub-corpora and comparing the various sub-corpora, either with the rest of the corpus (e.g. Baker et al., 2020), or individual sub-corpora. Because we were interested in gradual change over time, we calculated keywords by comparing one year with the previous year in the dataset, beginning with 2011 as our target corpus and 2010 as our reference corpus. Keeping in line with Baker et al. (2013) and Baker and McEnery (2019), we used log-likelihood (Dunning, 1993) as our statistic. A keyword list was not computed for the 2010 sub-corpus as the target corpus, meaning that the keywords reveal how things have changed since 2010. We reduced our nine keyword lists by discarding keywords that did not (i.) have a log-likelihood value of 3.84 or above, and

To appear in: *International Journal of Corpus Linguistics* in 2022.

(ii.) occur in at least 5% of texts in the target corpus. These keyword lists were then consolidated into a single list of 567 keywords and a matrix was built which recorded which of the keywords were present or absent in each newspaper article. Each row in the matrix represented a specific newspaper article, each column represented a keyword in the consolidated list and each cell had one of two values P (Present) or A (Absent). Using the package ‘FactoMineR’ (Husson et al., 2020) in R, MCA was used to generate an ordered series of dimensions of weakening strength comprising the most common patterns of co-occurring keywords across the texts of the corpus. Each dimension is a continuum of keyword variation, with a positive and a negative pole, which we interpret as representing a pair of competing or oppositional discourses – one on the negative side of the continuum and one on the positive. Interpretation is driven by close reading – by looking at the texts associated with each pole of any given dimension, and how the keywords associated with the dimension are used in those texts, we can develop a sense of what drives the distinction between the two poles of the dimension in question. The analysis proceeds from the first ranked dimension, downwards, until we reach a dimension which we are not able to interpret.

The hypothesis that each side of a dimension of co-occurring keywords represents a discourse is at the root of our approach to keywords and consequently needs careful consideration. Clarke et al. (2021) did just that and provided a corroboration for the hypothesis – the MCA grouped keywords based on their co-occurrence in texts, and these groups effectively disaggregated the discourses that were indicated by the keywords (see Table 1).

In their corpus as a whole, Clarke et al (2021) revealed 9 dimensions of keyword variation. Except for Dimension 1, which opposes short texts with long texts<sup>3</sup>, each dimension does indeed represent a pair of competing or oppositional discourses. For instance, in Dimension 6, they found that keywords on the positive side (e.g. *oil, economy, agreement, deal, cut, large, major, business*) co-occurred in articles that referred to globalisation, such as a particular country’s role and influence in the global economy, various trade deals and agreements, and the

---

<sup>3</sup> Dimension 1, or D1, opposes the presences of keywords on the positive side with the absences of the keywords on the negative side. Texts strongly associated with positive D1 are considerably longer than the texts associated with negative D1. Thus, Clarke et al. (2021) suggest that D1 largely reflects text length. As explained in Clarke (2019), this is a consequence of analysing the presence or absence of features, as the length of the text is the greatest influence on the presence of features. Shorter texts are more likely to have the absence of features.

To appear in: *International Journal of Corpus Linguistics* in 2022.

successes/failures of international business, whereas the keywords on the negative side (e.g. *western, americans, muslim, middle, east, war*) co-occurred in articles that distinguished between different groups and identities and often positioned them as being in opposition with each other, such as an article in *The Independent* citing Barack Obama's speech on winning the Nobel Peace prize, in which Obama positions America and 42 other countries as engaged in a war against ISIL/ISIS. This contrast in the patterns of co-occurring keywords in Dimension 6 was subsequently interpreted as opposing a globalisation discourse, wherein countries come together in order to participate on the global stage and in the global market and economy, with a tribalism discourse, wherein groups are opposed to each other in an 'Us versus Them' dichotomy.

Table 1: Dimensions and discourses identified by Clarke et al. (2021).

Dimension	Discourse on positive pole	Discourse on negative pole
1	Texts which are short	Texts which are long
2	War, conflict and terrorism	Everyday life and events (often in opinion articles)
3	Reporting of foreign affairs	Local and domestic affairs
4	Western political conflict	Overseas armed conflict
5	UK policy	US policy
6	Globalization	Us (non-Muslims) vs. Them (Muslims)
7	Corruption and human rights	Terror attacks
8	The rise of the far right	The radicalisation of British Muslims
9	Political processes and elections	Political processes and security threats

While we showed that MCA could guide the grouping of keywords into discourses, another level of aggregation remained in that the discourses identified were presented as

monolithic and unchanging. The next section explains how MCA may also allow us to disaggregate discourses over time.

#### 4.3. Using keyword co-occurrence to explore changing discourse through time

In our study, MCA derives dimensions based on the frequent co-occurrence of keywords across articles and assigns coordinates and contributions to the categories of keywords (e.g. *presence/absence* of keyword) and to articles for each dimension – articles are assigned a coordinate indicating their association with either the positive or negative side of a dimension based on how the keywords associated with a dimension co-occur in the articles. Specifically, the coordinate  $y^i$  of a particular article point  $M^i$  for a particular dimension is the simple mean of the coordinates  $y^k$  of the  $V$  keyword category points  $K^i$  in article  $i$  (i.e. the set of  $V$  keyword categories in article  $i$ ), divided by the square root of the eigenvalue of the dimension  $\sqrt{\lambda}$  (Le Roux and Rouanet, 2010: 41):

$$y^i = \frac{1}{\sqrt{\lambda}} \sum_{k \in K_i} \frac{y^k}{V}$$

So, articles with coordinates close to zero are not represented by the patterns of keyword variation that the dimension represents, while articles with high positive or high negative coordinates are strongly associated with it. Given that each article is time-stamped, we can then assess the association of any article to each dimension over time. Hence, our approach to mapping the individual dimensions over time, based on keywords, disaggregates our view of the discourses in two senses – we can explore the separate dimensions identified and, because our results are sensitive to change at the level of the basic structural unit in the corpus, the time-stamped article, we can look at those disaggregated discourses through time. To achieve this, we draw upon another output of the MCA used in Clarke et al. (2021) – the coordinates assigned to each article for every dimension. Using that we can establish how associated each text is to each dimension of keyword variation. For instance, Example 1 is a headline from an article in the corpus and Table 2 presents the coordinates of this article for each dimension.

- (1) French police launch series of anti-terror raids in cities across the country with at least three people held in Toulouse (*Daily Mail*, 16.11.2015)

Table 2: Dimension coordinates for Example 1.

<b>Example</b>	<b>Dim.2</b>	<b>Dim.3</b>	<b>Dim.4</b>	<b>Dim.5</b>	<b>Dim.6</b>	<b>Dim.7</b>	<b>Dim.8</b>	<b>Dim.9</b>
1	0.73	-0.31	-0.05	0.03	0.19	-0.1	0.07	0.02

Table 2 shows that Example 1 is strongly positively associated with Dimension 2, moderately associated with negative Dimension 3 and slightly positively associated with Dimension 6. It is not strongly associated with the other dimensions. Because, as noted, we know when each text was produced, it is possible to plot these coordinates over time. We can plot these coordinates over time in order to observe and assess at what points in time the texts were most strongly associated with each of the dimensions and if there are substantial shifts in discourse, gradual change or stability over time.

Accordingly, we plotted each article's dimension coordinates over the days represented by the corpus. We also plotted a 60-day moving average by taking the mean coordinates of all the articles published in that 60-day period on the respective dimension. Thus, the trend line indicates, on average, how associated the articles in the 60-day period are to the particular pole of the dimension. This is important as it shows us, for the two discourses in the dimension, how associated the articles are to them at any particular time – if the plot shifts towards the negative it shows a growing influence of the negative pole discourse and vice versa. We are seeing the see-sawing of the influence of discourses relative to one another.

The 60-day moving average facilitates the identification of broad trends in the corpus. Other windows can be used, depending on the interests of the analyst. If a researcher is interested in short-term impacts in reportage, then a shorter duration can be selected (e.g. 30-days), while a longer duration (e.g. 120-days) can be selected for long-term impacts, although clear trends could be obscured by taking larger windows. As we were interested in broad trends, we took a 60-day moving average. Subsequently, we used the trend line as a guide, exploring texts that were strongly associated with the peaks or troughs seeking explanations in the events that were reported on at that point.

## 5. Results

We will set aside Dimension 1 as it relates solely to text length, as discussed. Of more interest are the 8 Dimensions which allow us to look at discourse. In what follows a trough denotes a shift to the negative pole of a dimension while a peak denotes a shift to the positive pole of a dimension.

### 5.1 Dimension 2: War, Terrorism and Conflict vs. Everyday Life

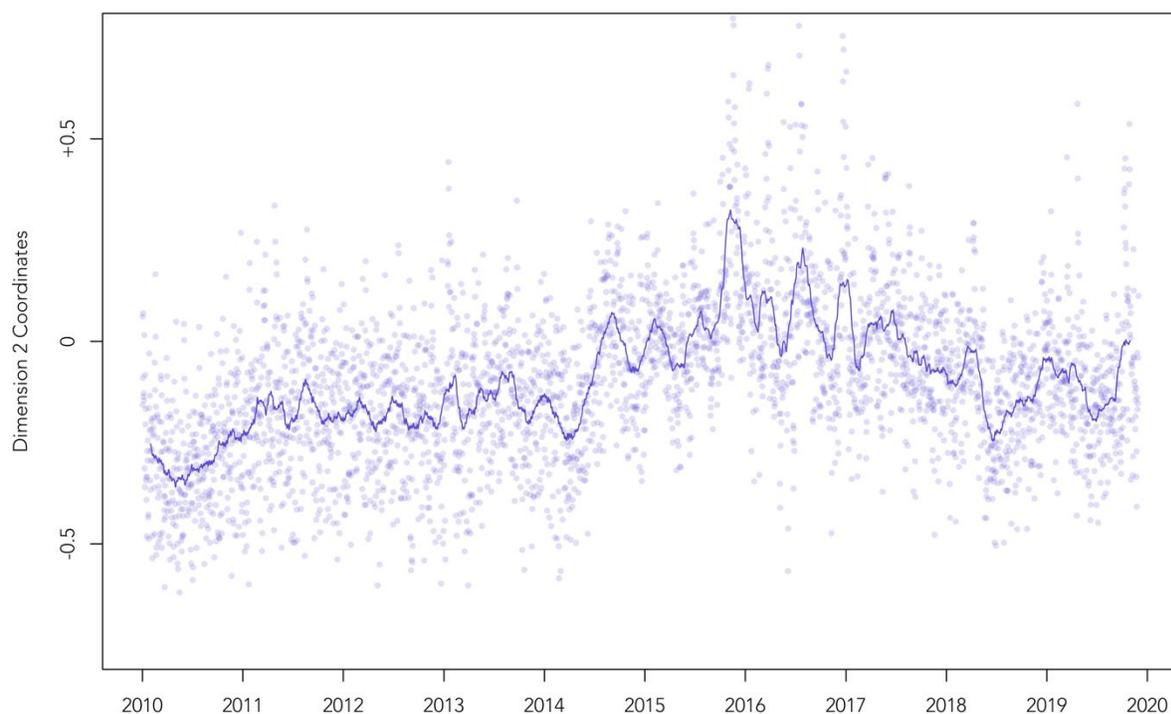


Figure 1: Plotting Dimension 2 through time: War, Terrorism & Conflict (positive) vs. Everyday Life (negative).

Figure 1 shows that whilst there was a general strengthening in the association of articles to war, terrorism and conflict from early 2010, driven by events such as the Arab Spring protests and revolutions, prior to 2014, articles in the corpus reporting on war, terrorism and conflict were, on average, less dominant than those that reported in links between Muslims, Islam and everyday

To appear in: *International Journal of Corpus Linguistics* in 2022.

life. Notably, Figure 1 shows that the war, terrorism and conflict discourse strengthens in June 2014, corresponding with the rise of ISIS and the Syrian Civil War (ISIS take control of Tikrit and Mosul), widely reported public beheadings carried out by ISIS in July and August, and the organization's genocide of Yazidis. Another peak of the war and conflict discourse occurs in early 2015 with the Charlie Hebdo and Hypercacher Jewish supermarket terror attacks in France, Boko Haram's massacre of civilians in Nigeria, and the murder of Jordanian fighter pilot Moaz al-Kasasbeh by ISIS. Another peak occurs in the middle of 2015 and this aligns with "Bloody Friday", where simultaneous attacks occurred in France, Kuwait, Syria, Somalia and Tunisia.

The largest peak in Figure 1 shows that, on this dimension, reportage at the end of 2015 in the UK national Islam press corpus was dominated by war, terrorism and conflict with few articles associated with the opposite pole. This particular peak is aligned with Opération Chammal, which refers to the French airstrikes on ISIS in Syria and Iraq, and the 13<sup>th</sup> November 2015 Paris terrorist attacks by ISIS. The rest of the figure shows that reportage concerning war, terrorism and conflict in the corpus is more dominant than reportage on everyday Muslim life when particular terror attacks have occurred, especially those in the West, for example in Brussels (March, 2016), Nice and Munich (July, 2016), Berlin (December, 2016), Westminster (March, 2017), Manchester (May, 2017), Paris (June, 2017), Barcelona (August, 2017), Münster (April, 2018), and Paris and Liège (May, 2018), and London (November, 2019).

So the ground truth we get from close reading matches our expectations from the graph; overall, the trend line on Figure 1 indicates, we would argue, that when terror attacks that are newsworthy (from a UK press perspective) take place, reportage becomes more strongly associated with war, terrorism and conflict in general, and to the extent that these topics are often more common than reportage around Islam and Muslims in any other context during these periods. Whilst reportage associated with war, terrorism and conflict rises when there is conflict in the Middle East, reportage in the corpus is not so dominated by stories directly linked to that conflict – the effect appears to be indirect. Thus, it may be possible to hypothesise that the UK national press increases its focus on reportage on war, terrorism and conflict when there are terror attacks in Western, particularly European, countries.

## 5.2 Dimension 3: Foreign vs. Domestic Affairs

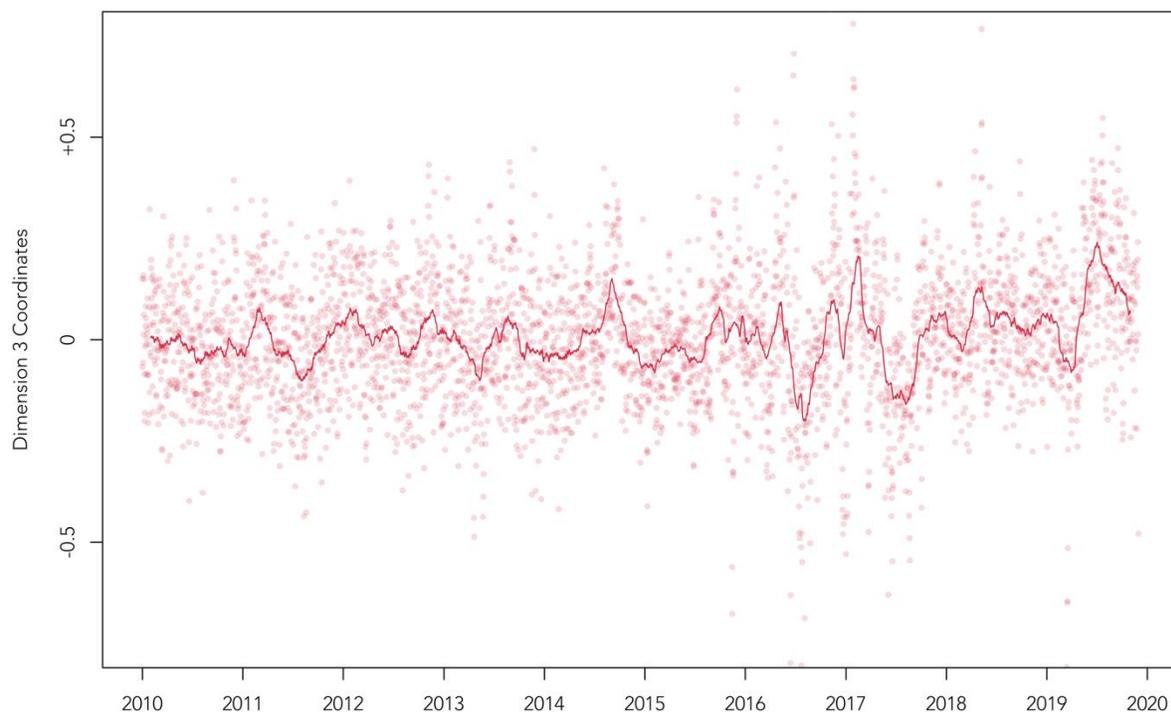


Figure 2: Plotting Dimension 3 through time: Foreign affairs (positive) vs. domestic affairs (negative).

Figure 2 shows how reportage in the corpus changed over time with respect to foreign affairs and domestic affairs. Before 2014, there was an even spread between articles concerning foreign and domestic affairs with minor fluctuations sparked by particular events. This graph shows a fairly steady see-sawing of attention between the two discourses on the dimension. These are driven by world events.

For example, with the rise of ISIS in the Syrian Civil war and violence in Gaza in later 2014, reporting becomes much more strongly associated with foreign affairs. The reporting then shifts back to domestic affairs in early 2015 following the attack at Charlie Hebdo and the Hypercacher Jewish supermarket, after which it was found that Hayat Boumeddiene, the wife of the supermarket attacker, had a sister living in Britain. There is then a further shift in reportage from domestic to foreign affairs around July 2015, when the Joint Comprehensive plan of action

To appear in: *International Journal of Corpus Linguistics* in 2022.

was signed (Iran Nuclear Deal Framework) and France launched airstrikes against ISIS in Syria (September 2015). Reportage focuses on domestic affairs in the run up to the UK referendum on European Union membership. But then shifts to focus on foreign affairs during the 2016 U.S. presidential election up to the Inauguration of Donald Trump.

Overall, Figure 2 indicates that reportage concerning foreign and domestic affairs is greatly influenced by major political and current events. Articles about domestic affairs are highly focused on British nationals, often criminals or victims, and their political decisions (e.g. Brexit), whereas articles reporting on foreign affairs are influenced by overseas conflict, elections, and political decisions.

### 5.3 Dimension 4: Western Political Conflict vs. Overseas Conflict

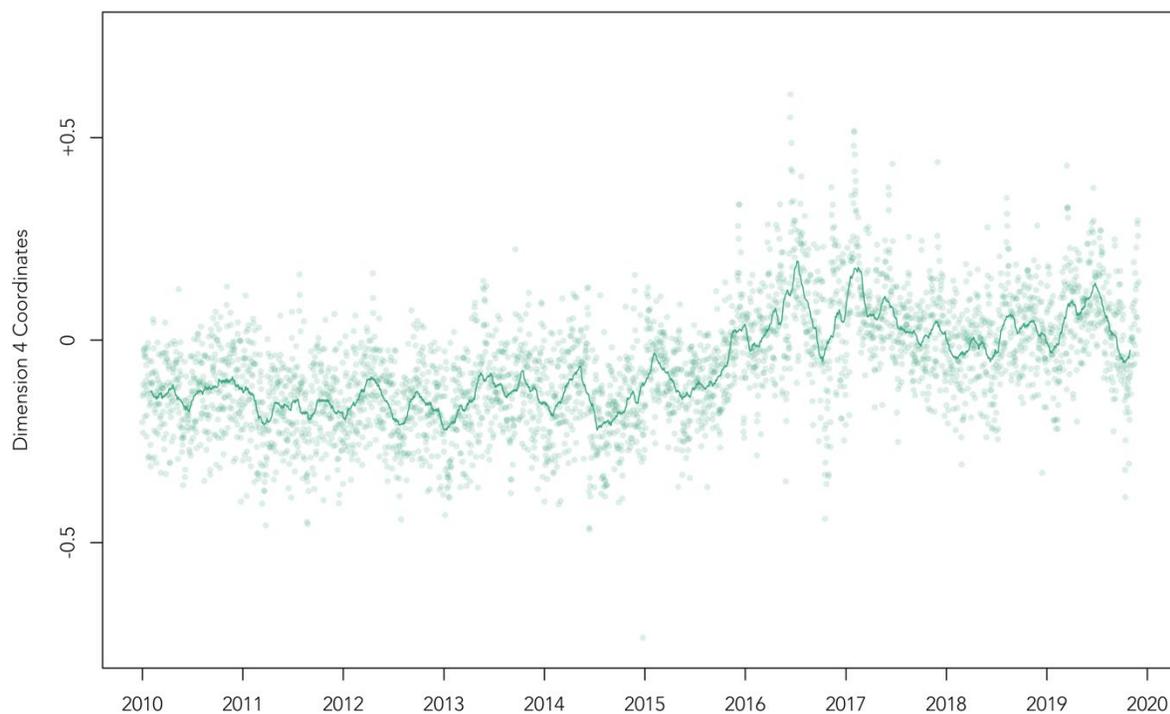


Figure 3: Plotting Dimension 4 through time: Western political conflict (positive) vs. overseas armed conflict.

To appear in: *International Journal of Corpus Linguistics* in 2022.

Figure 3 shows how reportage in the corpus changed over time with respect to Western political conflict or overseas conflict, primarily in the middle and near east. Figure 3 shows that reportage before 2016 predominantly focused on overseas conflict, aligning with the Arab Springs protest and events in the Syrian civil war. However, from the beginning of 2016 the focus shifts to Western political conflicts, aligning in particular with the UK referendum on EU membership and the social and political fallout that resulted from the ‘leave’ verdict. The focus switches back to overseas conflict later in 2016 to align with the Battle for Mosul (October, 2016), but shifts back to a Western conflict when reporting on the U.S. election and Trump’s controversial executive order suspending entry to the U.S. for immigrants from seven Muslim-majority countries. Following this, the reportage maintains its overall association with western political conflict, reporting on the formation of the Brexit party, as well as numerous stories accusing politicians and political parties of discrimination, including links between the British Conservative Party and Islamophobia (especially during the leadership election), and the Labour Party and Anti-Semitism, although there are shifts where reportage concerns overseas conflict, such as in early 2018 when Russia, Turkey and Syria are in conflict and in mid-2018 after Trump withdraws from the Iran Nuclear deal leading to clashes between Iran and Israel.

Overall, Figure 3 shows that prior to ISIS’s rise to prominence, reporting of conflict in the corpus focused mostly on that which occurred overseas. However, as the consequences of the rise of ISIS become apparent, political conflict in the West – especially with respect to a rise in right-wing views and anti-Muslim discrimination – becomes much more dominant in the data. One possible reason for this is because the civil wars in Syria and Libya led to the European migrant crisis and immigration is a contentious issue for right-wing political actors and groups. Thus it could be argued that Western political conflict, in terms of increased support for right-wing groups and anti-immigrant and anti-Muslim sentiment expressed through support for politicians like Donald Trump in the U.S. and political movements like the ‘Vote Leave’ campaign in the UK, has received increased focus in our data partially due to increased levels of migration to Western states resulting from civilians fleeing the civil wars in Libya and Syria.

#### 5.4 Dimension 5: UK vs. US policy

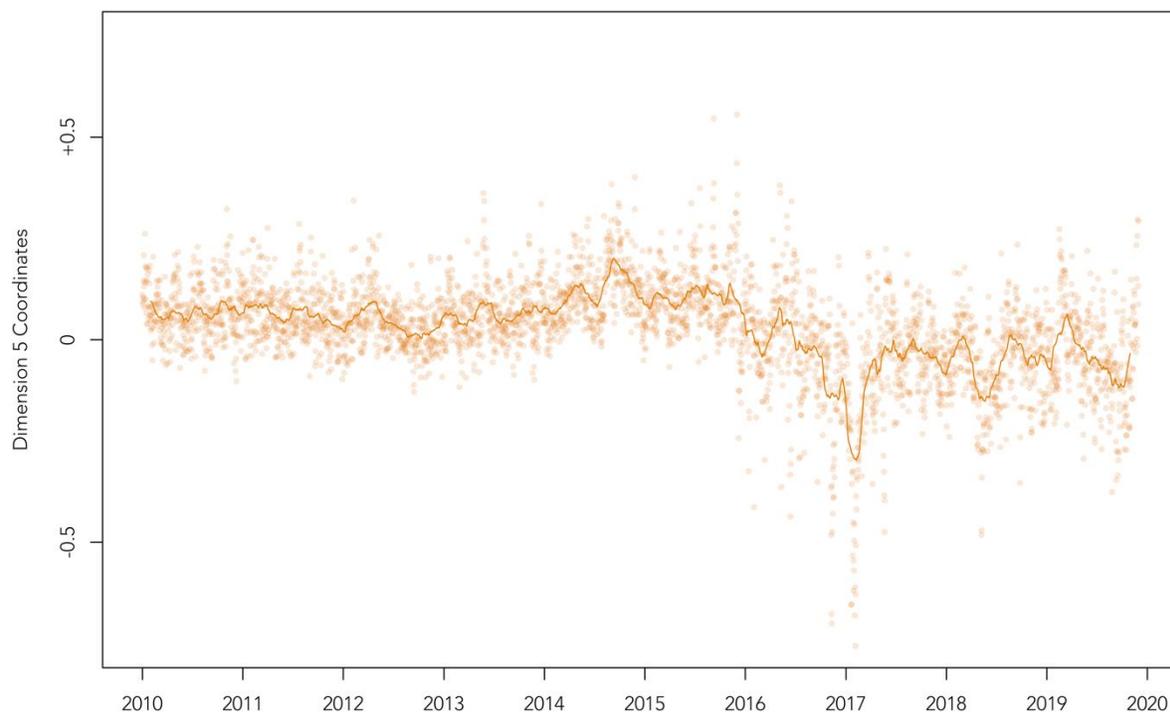


Figure 4: Plotting Dimension 5 through time: UK (positive) vs. U.S. (negative) policy.

Figure 4 shows changes with respect to reporting on UK and U.S. policy. It indicates that, before early 2016, reportage predominantly concerned UK policy. One of the first peaks occurs in early 2012 where articles focus on the consultation on equal civil marriage in England and Wales (March 2012). Such articles discuss the views of different religions, including Islam, towards same-sex marriage, an interesting echo of Baker and McEnery's (2019) findings. The highest peak occurs in the second half of 2014, following the rise of ISIS with articles discussing UK military intervention in the Syrian Civil war. However, at the end of 2015, focus switches to U.S. policy, which aligns with the beginning of the 2016 U.S. presidential election, specifically the emergence of the candidates and the Iowa caucuses in February. Following the Caucuses, focus switches again to UK policy and the EU referendum (and its associated campaigns in the run-up). Following the year of the referendum, we see the biggest trough in 2017, which coincides with Trump's inauguration, his aforementioned Muslim travel ban and immigration policies,

To appear in: *International Journal of Corpus Linguistics* in 2022.

Trump's public recognition of Jerusalem as the capital city of Israel (December 2017), and the U.S. pulling out of the Iran Nuclear Deal (May 2018). Focus then shifts to the UK context in early 2019 following an interview with Shamima Begum, a British girl who left the UK in 2015 to join ISIS, with many stories discussing the UK Home Secretary's decision to strip Begum of her British citizenship. Additionally, in early 2019 reportage referred to then-Prime Minister Theresa May's announcement that she would resign if parliament approved her Brexit deal, encouraging stories about possible future leaders of the Conservative party, bringing to the fore claims of Islamophobia in the Conservative party. The second half of 2019 is characterised by increased focus on U.S. policy, particularly laws around gun control in light of events such as the El Paso shooting, thwarted white supremacist attacks against Muslims, and Trump's decision to withdraw from Syria.

Overall, Figure 4 shows that prior to the 2016 U.S. presidential election, UK press reportage around Islam tended to focus on domestic policy, especially by distinguishing UK law and policies with sharia law. However, since Trump's election, the press's focus had shifted to concern predominantly U.S. policy. Notably, this interest seems to centre on anti-Muslim comments and policies attributed to the then-President.

### 5.5 Dimension 6: Globalisation vs. Tribalism

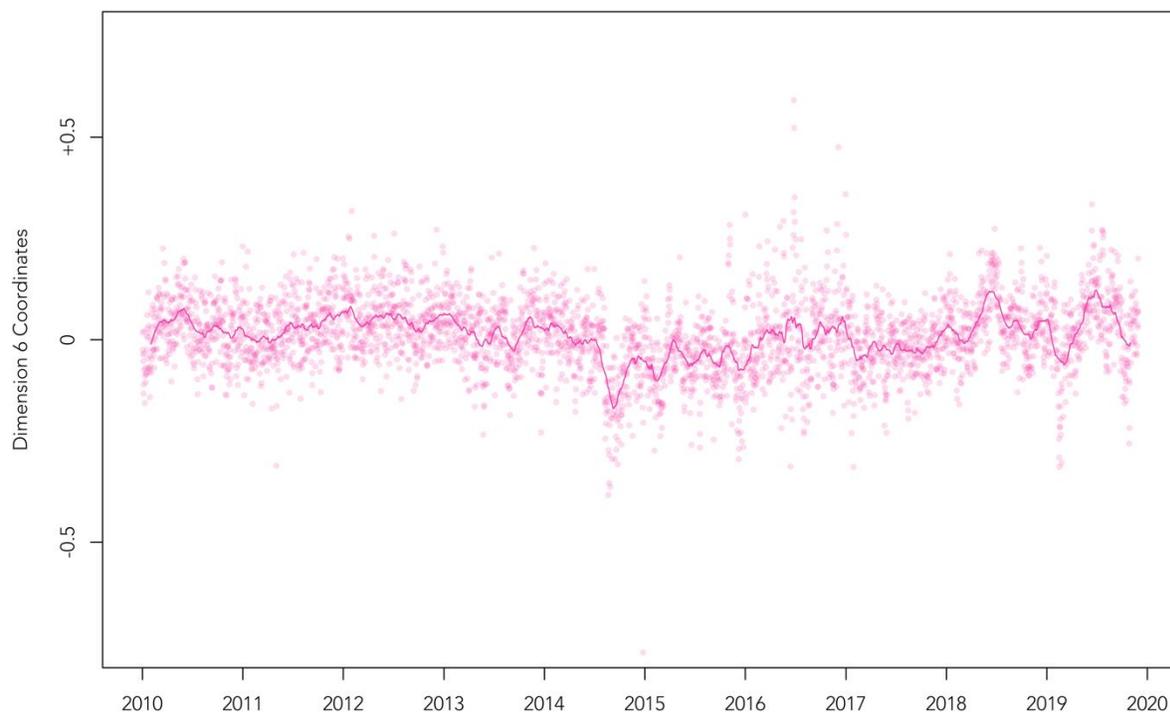


Figure 5: Plotting Dimension 6 through time: Globalism (positive) vs. Tribalism (negative).

Figure 5 shows changes to discourses around globalisation and tribalism in the corpus. In general, reportage prior to 2014 concerns globalisation, as opposed to tribalism, although there are a few exceptions, such as following the Ghouta chemical attack where reports drew sharp distinctions between the inhumane Syrian dictatorship of President Bashar al-Assad and Western democracies. Many of the pre-2014 articles associated with globalisation discuss the Eurozone crisis, including the impact of the EU embargo on Iranian oil in Greece, Britain's economy, and global travel, trade and sporting events.

To give a sense of how issues causes the focus of the balance between the poles of this dimension to fluctuate, consider the period 2016-2018. From 2016 to 2018, reportage is generally evenly divided between tribalism and globalisation, although there are peaks where reportage concerns UK inflation rates and the impact of Brexit on the economy, and troughs where articles position Trump in opposition to Clinton in the U.S. presidential election race.

To appear in: *International Journal of Corpus Linguistics* in 2022.

However, by the beginning of 2018, reportage shifts to focus on globalisation. Many of the articles strongly associated with globalisation in 2018 report on the impact of the Carillion Crisis, Donald Trump's tariffs on steel and aluminium imports and the Brexit bill on the global and British economy.

Overall, Figure 5 shows that press reporting of Islam involves a balance between reportage of, on the one hand, an Us vs. Them dichotomy (Tribalism) and, on the other hand, reporting on the global economy, communication and trade (Globalisation). This is particularly clear during the polarising EU referendum campaigns in the UK and the Presidential election in the U.S. The Vote Leave and Trump campaigns invoked Tribalistic rhetoric to demonise migrants and other groups, while also highlighting the economic case of voting leave (in the UK) or for Trump (in the US).

#### 5.6 Dimension 7: Corruption and Human Rights vs. The Aftermath of Terror Attacks

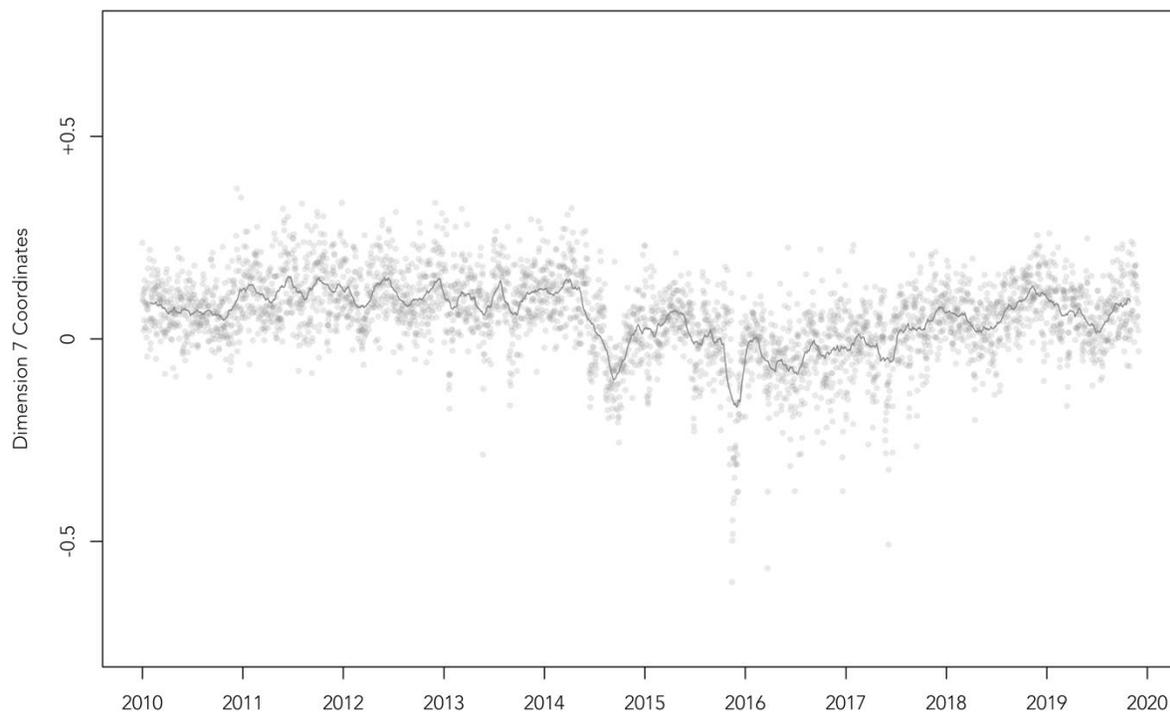


Figure 6: Plotting Dimension 7 through time: Human rights (positive) vs. Aftermath of terrorist attacks (negative).

To appear in: *International Journal of Corpus Linguistics* in 2022.

Figure 6 shows how reportage changed over time with respect to corruption and human rights or the aftermath of terror attacks. It shows that before 2014, reporting was generally dominated by articles associated with corruption and human rights issues. For example, at the end of 2010 (the first peak) articles concern Sakineh Mohammadi Ashtiani – an Iranian woman sentenced to death by stoning for alleged adultery, which made international headlines as a human rights issue, as it was viewed that the crime did not warrant such punishment. Additionally, in the first half of 2012, articles concern Libya's refusal to hand over Gaddafi's son Saif al-Islam Gaddafi, who was indicted with crimes against humanity.

By June 2014, there is a shift to reporting the aftermath of terror attacks, and this aligns with the rise of ISIS, public beheadings carried out by the group, the genocide of Yazidis, and attacks in Ottawa and Paris. Other troughs occurring after 2015 can also be connected to terror attacks that the press reports on. Shifts towards human rights in post-2015 articles often occur as the result of the deaths of influential people who have either campaigned for human rights, such as Pierre Bergé who campaigned for gay rights and died in 2017, or who have acted against human rights, such as the Uzbek President Islam Karimov who died in 2016.

Overall, Figure 6 is to a degree a mirror of Figure 1 (Dimension 2), where the larger peaks on Figure 1 are the larger troughs on Figure 6 (Dimension 7), as they align with western terror attacks. Figure 6 shows that the conflict occurring during the Arab Spring protests and the Syrian civil war is largely framed by the UK national press as a human rights and corruption issue, as opposed to a terror attack. This contributes to a broader discourse which opposes Western progressiveness, liberality and humanity with Middle Eastern medieval brutality, inhumanity and corruption.

### 5.7 Dimension 8: The Rise of the Far-Right vs. The Radicalisation of Muslims

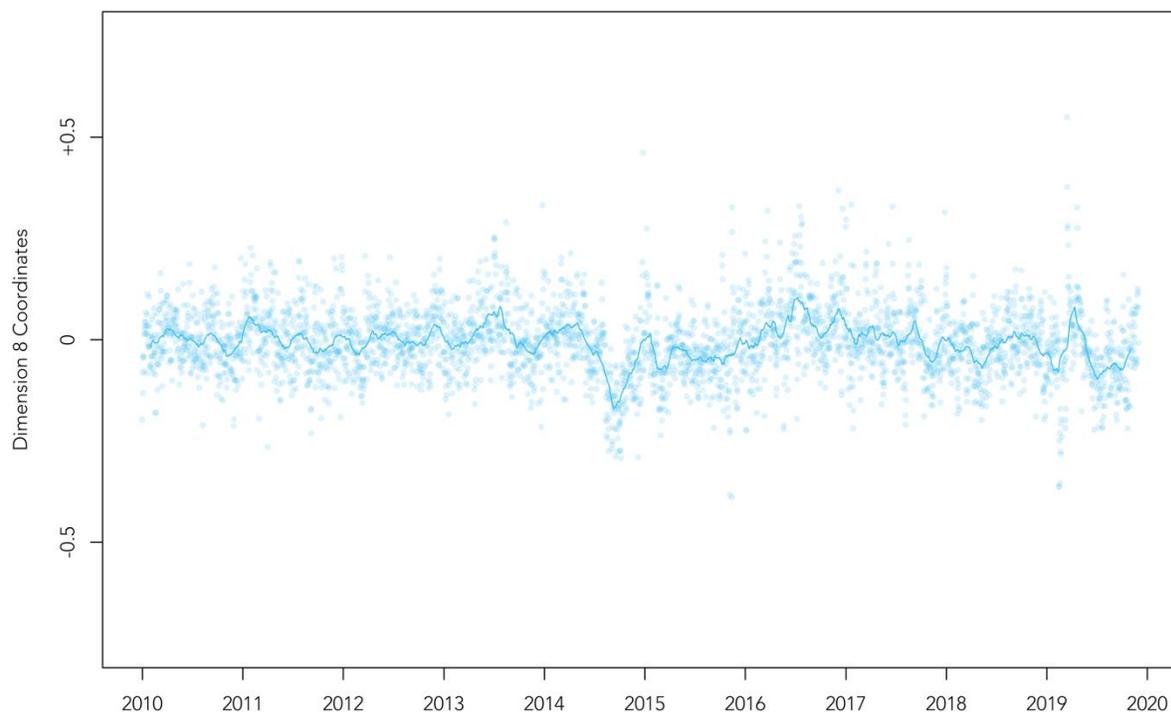


Figure 7: Plotting Dimension 8 through time: Rise of the far-right vs. The radicalisation of Muslims.

Figure 7 shows changes concerning the rise of the far right or the radicalisation of British Muslims. The figure shows that, in general, there is a balance over time between these two positions. There are a few small peaks where reportage is more associated with the rise of the far right following particular events, such as the rise of sectarianism in Egypt during the country's parliamentary elections (December 2011), and the rise of far-right terrorism linked to the violence committed by Anders Breivik, a self-declared fascist and Nazi who killed 77 people in Oslo (July 2011), and Brenton Tarrant, a white supremacist terrorist who murdered 51 people in two mosques in Christchurch, New Zealand (March, 2019). There are also peaks where reportage concerns the rise of the far-right political parties and movements, such as the electoral gains made by right-wing parties like UKIP in the UK (May 2013) and Alternative für Deutschland in Germany (September 2017). Other relevant events include the victory of the Vote Leave

To appear in: *International Journal of Corpus Linguistics* in 2022.

campaign in the EU referendum (June 2016), the election of Donald Trump as U.S. President (November 2016), and electoral success for the far-right Freedom Party who entered into a coalition government with the conservatives in Austria (December 2017) and the election of Narendra Modi, the far-right Hindu nationalist party leader, as Prime Minister (May 2014 and May 2019). Finally, there are also peaks where right-wing and racist sentiments and comments are discussed and critiqued in the press, such as racism in UKIP (May 2013 and April 2014), Islamophobia in the Conservative Party, and Boris Johnson's derisory comments on the burka (August, 2018).

The largest trough where reporting shifts to predominantly concern the radicalisation of British Muslims follows the rise of ISIS. Notably, there are troughs where the articles report on the beheadings of James Foley (August 2014), and David Haines (September 2014), blaming, among others, radicalised British Muslim Mohammed Emwazi (Jihadi John) (February 2015).

There are other troughs where the underlying narrative concerns radicalised British Muslims who have fled the UK to Syria to join ISIS, such as with Mohammed Nahin Ahmed and Yusuf Zubair Sarwar, who pleaded guilty to preparing for acts of terrorism after returning to the UK (July 2014) and British Muslim Shabazz Suleman, who joined ISIS in 2014 and was captured in Syria but disappeared in February 2019. Other troughs include articles that refer to individuals who have fled the UK to join ISIS, subsequently escaped the group or were captured, and wish to return to the UK, such as Jack Letts (Jihadi Jack) (October, 2017), Shukee Begum (October 2015), and Shamima Begum (February, 2019).

Other articles strongly associated with troughs report on British Muslims that have committed crimes, such as the sentencing of Sabrina Kouider and Ouissem Medouni for murdering their nanny, Sophie Lionnet (May, 2018), and Jamshed Javeed, who was jailed for six years for fleeing the UK to join ISIS (March 2015). Prior to the rise of ISIS, reportage in the Islam corpus referring to radicalised Muslims includes interviews with victims from the Rochdale child sex abuse ring, and other crimes such as kidnapping (e.g. of Faris Heeney, October 2013).

Overall, Figure 7 indicates that for the most part there is a balance between reportage that critiques the rise of the far right with that which presents radicalised Muslims, especially British Muslims, as criminals and threats. It could be argued that the balance over time is a domino

effect. Reportage that presents Muslims as criminals and threats instils fear of Muslims, which feeds right-wing thinking and narratives, thus driving support to far-right parties, which are subsequently reported on, and in many cases critiqued, by the press.

### 5.8 Dimension 9: Political Processes Related to Elections vs. Security Threats

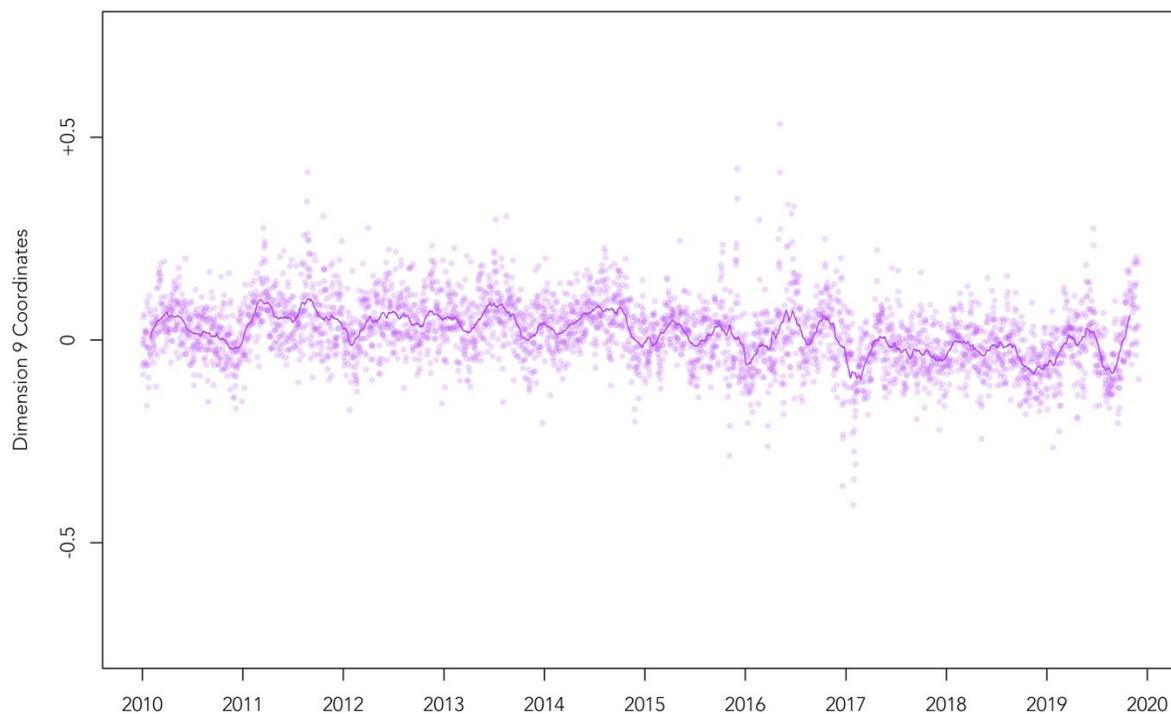


Figure 8: Plotting Dimension 9 through time: Processes related to elections (positive) vs. Security threats (negative).

Figure 8 shows how coverage of political processes related to elections vs. reporting of security threats fluctuated in reportage of Islam over time. When there is an election, political unrest, and/or a government is overthrown, there is an increase in articles concerning political processes related to elections. For example, there is a peak around February 2011 when Hosni Mubarak resigned as President of Egypt following 18 days of demonstrations in the Egyptian revolution. Other peaks are associated with reports on rebels seizing Zliten and Tripoli (August 2011), which meant that Muammar Gaddafi lost power, the 2013 Egyptian coup d'état (July 2013),

To appear in: *International Journal of Corpus Linguistics* in 2022.

where the president of Egypt, Mohammed Morsi was removed from power, India's election in May 2014, as well as reports of who has control and power during the Syrian civil war, such as ISIS taking control of Mosul and Tikrit and the Kurds taking control of Kirkuk.

Post 2014, the peaks tend to result from articles concerning political processes related to elections in the West, predominantly the UK, such as the 2015 UK General Election (May 2015), the Labour Party leadership election (August-September 2015), the 2016 UK EU membership referendum (April-June 2016), the 2016 U.S. presidential election (November 2016), the UK snap general election, announced in April 2017 and held in June 2017, UK local elections (May 2018), Brexit negotiations and deals that are voted on in parliament (February-March, 2019), Theresa May's announcement that she will not continue as PM in the next stage of Brexit negotiations (March 2019), her official announcement of her resignation in June, and then the Conservative party's leadership election, where Boris Johnson wins, becomes Prime Minister (June-July 2019), and then calls for a snap election following his failure to induce parliament to approve an EU withdrawal agreement for the end of October.

In addition to the peaks, there are many troughs on Figure 8 indicating that reportage in the corpus predominantly concerns political processes related to security at these particular times. Notably, these troughs follow terror attacks, suspected terror plots, and leaks of classified information. For example, there is a trough in October 2010, following the discovery of a package containing explosives on a US-bound cargo plane in the UK and Dubai, which bore the hallmarks of al-Qaida. Additionally, another trough can be seen around June 2013, following the leaks of classified information by Edward Snowden – an act which was perceived as a security threat. Following the rise of ISIS, the identification of Jihadi John as British citizen Mohammed Emwazi, and stories of the three Bethnal Green girls who fled to Syria to join ISIS, reportage concerned new security measures to identify radicalised Britons likely to turn to terrorism (February, 2015). Furthermore, there are troughs following particular attacks or threats of an attack, such as the fear of a terror attack on New Year's celebrations, leading to increased police in the UK and cancelled fireworks displays in Paris and Brussels (January 2016), the murder of a priest in Normandy (July, 2016), the Bastille Day terror attack in Nice (July 2016), the murder of British embassy worker Rebecca Dykes in Beirut (December 2017), the threat of cyber-attacks (December 2017), and the El Paso attack (August 2019). These articles often discuss UK terror threat levels, which either remain the same or rise. There are also troughs where reportage

To appear in: *International Journal of Corpus Linguistics* in 2022.

discusses new government policies, acts and orders on security, such as new anti-terrorism laws (June 2015), Trump's travel ban on citizens from Muslim-majority countries (January, 2017), Trump withdrawing from the Iran Nuclear deal, which led to the Iranian nuclear weapons threat coming back to the fore (May 2018), and relaxed visa regulations between Belgrade and Tehran, leading to a spike in Iranian migrants crossing the English channel (December 2018). Another trough occurs in December 2018 when the U.S arrested Meng Wanzhou, the chief financial officer of Huawei. These articles mention that the biggest threat to security is the China-U.S. trade war because it could lead to economic slumps, which lead to populism, nationalism and extremism. Finally, some of the troughs also occur where articles refer to government and security services failures to keep the country secure, such as in December 2017, where it was revealed that the security services failed to act on intelligence concerning the Manchester arena bombing, and in January 2019 where the poor planning of Brexit has left Britain at risk of both far-right- and ISIS-inspired terrorism.

Overall, Figure 8 shows that there is generally a balance between articles concerning political processes related to elections and those relating to security threats. The plot shows that there is an increase in articles concerning security threats after terror attacks, suspected terror plots, new legislation related to security, and leaks of classified information, whereas there is an increase in articles concerning elections when there are elections or democratic votes, or when power is lost and governments are overthrown. The plot shows that political processes related to security threats do not predominate in the corpus until the rise of ISIS, and rise further with the election of Donald Trump. In terms of ISIS-related security threats, the narrative depicts the threat as radicalised western citizens who have fled their home countries to join ISIS and who may be a threat upon return to their home countries, while Trump related security threats are in part due to Trump's rhetoric, where he frames people from Muslim-majority countries, among others (e.g. Mexicans) as threats, and in part due to his actions, such as withdrawing from the Iran Nuclear deal, withdrawing U.S. troops from Syria, and his arrest of Chinese nationals, which are often framed by the UK press as threatening global security.

## 5.9 Broad Trends Influencing Multiple Dimensions



Figure 9: Plotting all Dimensions through time.

Figure 9 is a plot of the moving averages for all 8 dimensions plotted on top of each other. From this plot, we can observe if there are any points in time where reportage suddenly shifts along multiple dimensions, and if any particular events align with such shifts.

Notably, Figure 9 shows that the reportage concerning Islam prior to 2014 did not shift dramatically along all dimensions. This broadly corroborates Baker and McEney's (2019) findings. Nevertheless, as that previous study suggested, there were some changes. There is a particular point, specifically around early 2011, where a change in reportage can be seen to influence many of the dimensions, albeit not dramatically. In particular, at the beginning of 2011, there is a small increase in reportage concerning war, terrorism and conflict (+D2), foreign affairs (+D3), overseas conflict (-D4), corruption and human rights (+D7), the rise of the far right (+D8), and political processes related to elections (+D9). Reportage associated with this time and along these dimensions concern the South Sudan Independence referendum, Egyptian

To appear in: *International Journal of Corpus Linguistics* in 2022.

revolution against President Mubarak, and protests in Libya against Gaddafi. Connecting all these events is not only region, but also widespread political change, as a result of discontent with the current establishment.

The first major shift across multiple dimensions occurs at the beginning of 2014, where reportage becomes increasingly associated with war, terrorism and conflict, foreign affairs, overseas conflict, UK policy (+D5), the aftermath of terror attacks (-D7), and the radicalisation of Muslims (-D8). There is also a slight increase in reportage associated with elections (D9+) and tribalism (-D6). Reportage associated with this time along these dimensions concern numerous attacks against or by terrorist groups overseas in the Middle East, including Yemeni and U.S. forces attacks against al-Qaeda in the Arabian Peninsula, the suspected Islamic Front attack on the Carlton Citadel Hotel in Aleppo, the capturing of Tikrit and Mosul by ISIS, violence in Gaza, the beheadings of numerous soldiers, journalists and aid workers by ISIS, and the Genocide of Yazidis by ISIS. Whilst this shift covers numerous groups, it predominantly marks the rise of ISIS, with reportage focusing on their brutality, and the threat of British Muslims being persuaded to flee the UK to join them.

Reportage shifts by the end of 2014, where it becomes more associated with domestic affairs (-D3), western political conflict (+D4), U.S. policy (-D5), globalisation (+D6), human rights and corruption, the rise of the far right, and security threats (-D9). Reportage during this time concerns debates around immigration, including UKIP's rise in local and European Parliamentary elections, Obama's decision to give up to 5 million immigrants work permits, rise in support for anti-immigration parties across Norway, Denmark and Sweden, and Australia's approach towards pushing asylum seekers' boats back out to sea and sending them to Papua New Guinea and Nauru. This shift may therefore mark the impact of the migrant crisis.

It is interesting to note that the growing turbulence in discourse towards the end of the period covered by Baker and McEnery (2019) was not directly identified by them. We would assume that this was because of the aggregation of the discourses in that study. As the discourses are clearly disaggregated here, it is easier for us to see the trends within a dimension and across a dimension at a relatively high level of resolution through the 60-day rolling average.

Moving to the period not covered by any previous study, Figure 9 shows that another dramatic shift in reportage occurs around the second half of 2015, where it becomes more

To appear in: *International Journal of Corpus Linguistics* in 2022.

associated with war, terrorism and conflict, foreign affairs, western political conflict, and the aftermath of terror attacks. Reportage during this time concerns the French airstrikes on ISIS in Syria and Iraq, and the 13<sup>th</sup> November 2015 Paris attacks by ISIS. France responded to the Paris terror attacks with another airstrike a couple of days after the attack. This shift indicates the force, aggressiveness, orchestration and determination of ISIS and arguably serves as a justification for international military intervention.

Another shift occurs around April 2016 until August 2016, where reportage is more associated with war, terrorism and conflict, domestic affairs, western political conflict, the aftermath of terror attacks, the rise of the far-right and elections. Reportage around this time concerns the UK EU referendum (and its associated campaigns), and terrorist attacks in Nice, Würzburg, Ansbach, and Russel Square. This shift marks a political change and also portrays the imminent threat of ISIS.

By late 2016, the focus of the reporting shifts to foreign affairs, western political conflict, U.S. policy, elections, and then security threats. Articles concern the U.S. presidential election and Donald Trump's aforementioned travel ban for citizens of seven Muslim-majority countries which, for political purposes, is framed as a national security measure.

Figure 9 also shows a shift in reportage around March 2017, where reportage is more associated with war, terrorism and conflict, domestic affairs, UK policy and elections. Reportage concerns the French presidential election, and local terror attacks, including the Westminster attack, Champs-Élysées shooting attack, Manchester Arena bombing and the Champs-Élysées car ramming attack, carried out by ISIS.

Another shift occurs around early-to-mid 2018, where reporting concerns foreign affairs, U.S. policy, globalisation, and the aftermath of terror attacks. During this time, reportage concerns Trump's decision to pull out of the Iran Nuclear Deal and numerous terrorist attacks, including the terrorist attack in Münster, the Paris knife attack, and the shooting of two policewomen in Liege.

Figure 9 shows a shift in reportage around mid-2018, where it becomes more associated with domestic affairs, western political conflict, UK policy, tribalism, and corruption and human rights. This aligns with Boris Johnson's opinion piece in *The Telegraph*, in which he compared Muslim women who wear the veil to 'letterboxes' and 'bank robbers'. This piece led to a

To appear in: *International Journal of Corpus Linguistics* in 2022.

multitude of stories not only criticising his comments as Islamophobic and calling for new policy on Islamophobia, but it also led to stories that came to his defence, praising him for his appraisal of the veil.

Starting at the end of 2018 and up to the start of 2019, the reporting shifts and is more associated with war, terrorism and conflict, security threats, overseas conflict, foreign affairs and the radicalisation of Muslims, which aligns with conflict in Gaza and the terror attacks in Melbourne and Nairobi. Then reportage shifts again, becoming less associated with war, terrorism and conflict and more associated with elections, the aftermath of terror attacks, and the rise of the far right, which aligns with the far-right terrorist attack in Christchurch and the Sri Lanka Easter Sunday bombings.

## **6. Conclusion**

This article introduced a new approach to studying temporal variation in discourse across a corpus. This approach builds on the MCA-based approach to keyword analysis introduced in Clarke et al. (2021) by using the grouping of frequently co-occurring keywords as a starting point for the identification of discourses. While the previous study by Clarke et al. (2021) demonstrated the utility of the MCA-based approach for grouping keywords into categories that can be interpreted as representing opposing sets of discourses, this paper has demonstrated that when the corpus is temporally structured, these co-occurring keywords and the discourses they signal can be successfully plotted on a timeline to indicate meaningful fluctuations in prominence. This fluctuation is meaningful in the sense that the discourses signalled by the sets of keywords typically gain prominence in response to real-world events to which Islam and Muslims are linked. Yet the selection of discourses is also a balancing act – as a real-life event leads to one pole of a dimension being foregrounded, the other pole becomes backgrounded, at least for the time-being. The interaction between the dimensions shown in Figure 9 is even more complex, with individual dimensions or bundles of dimensions appearing to respond to real-world events. This approach has thus provided a useful starting point for our CADS analysis, illuminating how the construction of Islam and Muslims as newsworthy in the press relies on various, often competing/opposing discourses, and how the discourse that is selected will largely

To appear in: *International Journal of Corpus Linguistics* in 2022.

depend on the nature of the real-life social and political events being reported on and in which the press seeks to implicate the religion and/or its adherents.

## References

- Baker, P. (2004). Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis. *Journal of English Linguistics*, 32(4), 346–359.
- Baker, P., Brookes, G., Atanasova, D., & Flint, S. (2020). Changing frames of obesity in the UK press 2008–2017. *Social Science and Medicine*, Online First.
- Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge University Press.
- Baker, P., & McEnery, T. (2019). ‘The Value of Revisiting and Extending Previous Studies: The Case of Islam in the UK Press’. In: R. Scholz (Ed.), *Quantifying Approaches to Discourse for Social Scientists* (pp. 215–249). Palgrave Macmillan.
- Benzécri, J.P. (1979). Sur le calcul des taux d’inertie dans l’analyse d’un questionnaire. *Cahiers de l’Analyse des Données*, 4, 377–378.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Brookes, G., & Baker, P. (2021). *Obesity in the News: Language and Representation in the Press*. Cambridge University Press.
- Burr, V. (2015). *Social Constructionism* (third edition). Routledge.
- Clarke, I. (2019). Functional linguistic variation in Twitter trolling. *International Journal of Speech Language and the Law*, 26(1), 57–84.
- Clarke, I., McEnery, T., & Brookes, G. (2021). Multiple Correspondence Analysis, Newspaper Discourse and Subregister: A Case Study of Discourses of Islam in the British Press. *Register Studies*, forthcoming.
- Duguid, A. (2010). Newspaper discourse informalisation: a diachronic comparison from keywords. *Corpora*, 5(2), 109–138.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence, *Computational Linguistics*, 19(1), 61–74.
- Fairclough, N. (2000). *New Labour, New Language?* Routledge.

To appear in: *International Journal of Corpus Linguistics* in 2022.

Gabrielatos, C. (2018). 'Keyness analysis: nature, metrics and techniques'. In: C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 225-258). Routledge.

Gabrielatos, C., McEnery, T., Diggie, P., & Baker, P. (2012). The peaks and troughs of corpus-based contextual analysis. *International Journal of Corpus Linguistics*, 17(2), 151–175.

Husson, F., Josse, J., Le, S. and Mazet, J. (2020). 'FactoMineR'. <https://cran.r-project.org/web/packages/FactoMineR/FactoMineR.pdf>

McEnery, T. (2005). *Bad Language, Purity and Power from 1586 to the Present*. Routledge.

Partington, A. (2010). Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: an overview of the project. *Corpora*, 5(2), 83–108.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies (CADS)*. John Benjamins.

Pertwee, E. (2020). Donald Trump, the anti-Muslim far right and the new conservative revolution. *Ethnic and Racial Studies*, 43(16), 211–230.

Richardson, J.E. (2004). *(Mis)Representing Islam: the racism and rhetoric of British Broadsheet newspapers*. John Benjamins.

Scott, M. (1996). *Wordsmith Tools*. Oxford University Press.