

## **Global responsibility: An exploratory corpus assisted discourse analysis of the Rohingya crisis in online media**

**Jasper Roe**

Jasro2@alumni.uv.es

University of Valencia, Spain

### ***Abstract***

This study conducts an exploratory corpus-assisted discourse analysis of the representation of the Rohingya minority group across online news media in the United Kingdom. The purpose of the study is to identify and interpret the discursive patterns employed in popular online news media when depicting the Rohingya minority and associated crises affecting the group in Myanmar and worldwide. Through the use of a combination of frequency, collocation, and concordance analysis, a synchronic study was undertaken using data collected from fifteen major online news media producers in the United Kingdom. The data was collected over a period from January 2017 – August 2020 through freely accessible digital archives. The research study found that particular discourses of security, internationalization, and power are commonly employed when reporting on the Rohingya, while equally a sympathetic viewpoint is often adopted which focuses specifically on global responsibility and failures of international society. The findings offer insight into socio-political processes of representation and discourse in the ‘new social location’ (Scholz, 2019) of online news media, while offering relevant insight into the discourses of urgent and pressing humanitarian issues.

**Keywords:** Discourse, Corpora, Linguistics, Rohingya, Myanmar

## **1. Introduction**

In Rakhine State, Myanmar, a protracted series of crises affecting the ethnic Rohingya minority group have been taking place for decades. The United Nations identify that the Rohingya are suffering from a ‘long history of discrimination and persecution which could amount to crimes against humanity’ (OHCHR, 2017) and rights groups and UN leaders suspect acts of genocide have taken place against them (Albert & Chatzky, 2019). In December 2020 these issues reached the international media on a scale not before seen, as a case was brought against the Government of Myanmar and its security forces at the International Court of Justice, which prescribed steps to reduce the risk of actions that could amount to genocide (ICJ, 2020). The World Health Organization (2019) argue that for the Rohingya the status quo is ‘not a tenable option – politically, socially, and economically’.

The reasons for these crises have been described as ‘overwhelmingly complex’ and often over simplified (Dussich, 2018). Bowser (2019) contends that the emergence of discrimination against the Rohingya worsened in the 1930s, as a result of the policies of the British colonial regime, and Pugh argues that the vestiges of these colonial structures which were intended to ‘divide and rule’ have resulted in an ongoing process of ethnic and religious discrimination causing the Rohingya to be viewed as an enemy of today’s state of Myanmar (Pugh, 2013). Cheesman (2017) argues that the creation of a national races superstructure has been key in creating the Rohingya as a ‘other’, existing outside the ‘national races’ discourse, and this has led to the Rohingya being excluded from the ‘trinity’ of the state, nation, and territory. Formal legislation was enacted in both 1974 and 1982, which had the effect of ‘stripping the Rohingyas of their nationality, rendering them foreigners in their own land’ (MSF, 2019), and since then, actions that could amount to genocide and crimes against humanity have continued to be undertaken against the Rohingya by the government of Myanmar (OHCHR, 2017).

Despite the urgency and seriousness of these long-running issues and their reporting on by international organisations, non-governmental organisations, and the media, there has been little scholarly interest focused on the Rohingya (Farzana, 2017) and even less so on the representations of such issues in mainstream media. This study synchronically analyses the representation of the Rohingya crises through a Corpus-Assisted Discourse Studies (CADS) approach. The primary research question of this exploratory study is to investigate how the Rohingya crises are discursively represented in online news media producers, with a focus on those in the United

Kingdom. To conduct this research, a small, specialized corpus was compiled using digital archives and analysed using Sketch Engine, a web-based corpus and text analysis application.

## **2. Background**

The long history of persecution of the Rohingya (OHCHR, 2017) is complex and globally unique in relation to other refugee crises given its protracted nature. This said, given the relative lack of scholarly enquiry into the Rohingya minority (Farzana, 2017) it is useful to consider existing literature relating to the representation of refugee issues in the press. Several research projects have investigated the representation of refugee issues and humanitarian crises in the media in the United Kingdom and elsewhere. One of the most seminal of these was the RASIM project, in which Baker and Gabrielatos (2008) found that general categories of discourse tended to occur when describing refugee groups, and discourses could often be broadly categorized into concepts including ‘economic problem’, ‘provenance/destination/transit’, ‘legality’ and ‘plight’. Studies have equally previously asserted that refugees and specifically Muslim refugees are often depicted negatively in the press (Baker and McEnery 2005; Baker and Gabrielatos 2008, KhosraviNik, 2014; Daghigh and Rahim 2020).

On the broader scale, there have been various research projects which investigate the representation of Muslim groups in Western media. Daghigh and Rahim (2020) for example, have researched the discourse of political leaders in Muslim minority countries, while identifying that this can contribute to alleviating the negative representations often found in Western media. However, no research studies as of yet explore the representation of the crises affecting the Rohingya in the online news media of the United Kingdom and the associated implications of such discourse, although Nahar (2019) has focused on the representation of the Rohingya in the Daily Star in Bangladesh, while Afzal (2016) has conducted a similar investigation in media in Pakistan and the USA.

Brooten (2015) has also researched the reporting on violence in Rakhine state in mainstream global media, finding that media representations tended towards employing notions of human rights only when they drew on the cultural identity of norms that the interested groups would want to universalize (Brooten, 2015). As an example, Brooten (2015) found that the identification of liberal democracy in global media representations was used to frame an international hierarchy in which ‘whites are the models and saviors of nonwhites’, a position

developed from Said's (1978) Orientalism. Further to this, Brooten and Verbuggen (2017) conducted fieldwork with reporters working within Myanmar and found that a shared set of journalistic practices led to polarization of reporting, and that a new generation of 'parachute journalists' and freelancers can lead to the production of heavily influenced or biased discourses, which in this case places focus on the identification of misery and suffering (Brooten and Verbuggen, 2017). Afzal (2016) conducted the only study to date which explores the discursive representation of Rohingya issues in British mainstream newspapers, while also comparing the findings to Pakistani and American articles through frame analysis. Afzal (2016) found that across all the collected data there was a broadly sympathetic role for the suffering of the Rohingya. Further to this, Siddiquee (2019) analyzed state media within Myanmar, arguing that in an era of post-truth politics, information is often oversimplified when describing violence, a term that Venkatamaran (2018) labels 'discursive simplification'.

There is a need to further investigate and add to the body of literature on this topic, as newspaper texts can be a useful area for studying the production and reproduction of discourses (Baker, 2006), and have a cumulative power which can act as a barometer for social change (Fairclough, 1992). This is particularly true of online news media, which is a rapidly growing platform. This said, no studies to date have aimed to use a small, specialized corpus to investigate the representation of Rohingya issues in online news media, despite the format being on course to overtake other forms of media production in the near future. This is evidenced by data from the UK Office of Communications (Ofcom, 2019), which found that while television is still the most common format for news consumption in the United Kingdom, access via the Internet is now the second most common method, with 66% of respondents in an annual survey finding news media in this manner. Simultaneously, the data shows that print newspapers are in a steady state of decline (Ofcom, 2019). Aside from this growing popularity, online news media also has several unique features which make it a new frontier in text and discourse analysis. To exemplify, the rise of the internet has been described as creating particularized discourse communities and new social spaces (Scholz, 2019) in which discourse occurs and media is consumed, recycled, and shared. The lack of current scholarly research (Farzana, 2017) and the growth in online news media's consumption (Ofcom, 2019) then provides a backdrop for the importance and utility of this research study into the 'new social spaces' found online.

### **3. Data and Analytic Methods**

In order to explore the discursive representation of the Rohingya in online news media in the United Kingdom, A Corpus Assisted Discourse Studies (CADS) approach was adopted. CADS methods take advantage of the synergistic nature of combining quantitative technique such as frequency analysis with in-depth analysis of concordances, providing counterbalance and triangulation (Freake, Gentil & Sheyholislami, 2010). Through these forms of analysis, it is possible to find patterns which can illuminate the traces of discourse that can be carried through language (Baker, 2006).

As pilot studies are recommended to identify the size of a potential corpus in corpus-assisted discourse approaches (Baker, 2006), in this case a pilot study was undertaken with the aim of identifying whether the use of freely available online archives of UK media producers would provide enough relevant data for the creation of a specialized corpus. An initial search of archives of three major online media brands (The BBC, Sky News, and The Guardian) were chosen through convenience sampling and investigated, searching for archived and current articles which were related to the subject of the Rohingya refugee crises and appeared within a time span of five years, from 2015 to 2020. For possible inclusion, the article was required to feature the node word, 'Rohingya' in the headline. Initial searches yielded between ten and fifteen articles per outlet, leading to the setting of a target of collecting a minimum of ten full text articles for each media outlet included in the study. Texts collected were required to be full articles rather than summaries. Thirty texts were then sampled, cleaned, and processed for upload to a corpus compilation and analysis program, Sketch Engine, for an initial analysis. They were then automatically tagged at the sentence and paragraph level by the software. This allowed for familiarization with the data and the research tools as advocated for by Baker (2006).

The pilot study revealed that in order to develop a suitable representative corpus, the timespan needed to be narrowed to a period of three years, as five years included few relevant articles, but did provide many hundreds of irrelevant articles with embedded hyperlinks to more recent, and more relevant articles. Additionally, the pilot study revealed that a target sample of 10 articles from each online media provider was likely to be achievable across the major online news media outlets digital archives. Given major events that took place in the context of the Rohingya crisis from 2017 – 2020, this was selected as the timespan for data collection. Finally, initial

familiarization with the online archives revealed that multiple searches needed to be undertaken with both capitalized and lower-case search terms, and that results were best organized in reverse-chronological order. Articles could then be collected beginning from their most recent appearance, moving backwards until the cut-off date was reached or a quota of ten articles was obtained.

On completing the pilot study, representative online news media sites were then selected. Online news media sites based in the United Kingdom were ranked by their readership and popularity. As readers are not passive and meaning can be introduced through the interaction between a text and its reader (McIlvenny, 1996) a greater readership may suggest a larger contribution to the discourse order, or what Gusfield (1986) terms the conceptual orderliness of the public arena. In order to identify the online news media archives from which data was to be collected, two sources were consulted. The first of these was the Publishers Audience Measurement Company (PAMCo), and the United Kingdom Office of Communications (Ofcom). PAMCo (2020) is the body responsible for measuring audience engagement for published media in the United Kingdom. Ofcom (2019) meanwhile, is the UK's telecommunications regulations body. These two sources provide accurate, realistic and reliable information on audience numbers, although their separate methodologies made direct comparison of readership challenging.

Through analysis of data provided by Ofcom (2019) and PAMCo (2020) a list of 15 online news media publications were selected. The challenge in comparing the two sources of data was that PAMCo offers a metric of 'monthly brand reach' based on an unspecified methodology, while Ofcom (2019) provided clear rankings of readership in terms of popularity but did not provide exact numbers of readers. Ofcom (2019) highlighted the BBC and Sky News as having the largest readership, but these publications were not included in PAMCo's 'total brand reach' metric. For this reason, self-reported readership figures available on the news brands' websites were selected. A further challenge that occurred was that these self-reported figures for the BBC and Sky News refer to global readership, not just domestic audiences, which are unpublished or not collected. Despite this, both are major online news brands based in the United Kingdom, and thus were included in the corpus. This led to the total identification of the following fifteen news brands for inclusion in the study: BBC News, Sky News, The Sun, The Mirror, The Mail, The Metro, The Guardian, The Independent, The Express, The Telegraph, The Evening Standard, The Times, The Daily Star, The Daily Record, and The I. Websites consulted included all editions of the standard titles, including Sunday editions.

Once data collection began, a pattern began to emerge which demonstrated the specific time periods around which articles on the Rohingya and associated crises were clustered. Throughout this process, 142 articles were collected in total for inclusion in the corpus. Overall, 24 articles were collected in 2017, 30 were collected in 2018, 59 were collected in 2019, and 29 were collected in the period to September 2020 when data collection was completed. It is likely that the pattern of availability of articles coincided with major events which were considered ‘newsworthy’ such as the court case brought to the International Court of Justice by The Gambia, mass-migration movements, ‘clearance operations’ or significant rises in violence and tensions. In the period since data collection ceased, additional crises in camps in Bangladesh, and political crises in Myanmar, suggest that a further increase in media representation will have taken place.

On completion of data collection, the 142 articles were checked, cleaned, and additional information such as advertisements and embedded links were manually removed. This led to a total word count of 93,218 words, which were batch uploaded to the Sketch Engine platform. Given the highly specialized nature of this study, this was deemed an adequate size and representativeness for investigation based on principles of small, specialized corpus creation outlined by Koester (2010). The small size of the corpus equally has some unique advantages, including helping to build a closer relationship between the corpus and the context, and less decontextualization in comparison to large reference corpora (Koester, 2010). Following upload, basic analysis was taken to identify the range between article size (10,154 words), the mean article size for each media producer (6,014 words), and the median article size (6,032 words). This was used to confirm that the data demonstrates a fairly good degree of consistency and representativeness across the news brands sampled, as visible in Table 1.

Table 1: UK online news media producers ranked by readership and associated data

Online news media producer	Self-Reported Readership	Articles collected	Word count of articles collected
BBC News	347 million / weekly (BBC, 2019)	10	10,734
Sky News	8 million / month (Sky Media, 2019)	10	5,260
	<b>Monthly brand reach (millions) (PAMCo, 2019)</b>		
Sun	33,918	10	5,830
Mirror	27,760	10	6,741
Mail	26,530	10	6,014
Metro	25,188	10	4,701
Guardian	24,042	10	9,344
Independent	21,903	10	9,055
Express	21,689	10	6,361
Telegraph	20,873	10	6,141
Evening Standard	14,962	10	3,876
Times	9,197	10	7,005
Daily Star	6,864	2	580
Daily Record	5,857	10	5,821
I	7,348	10	5,755
<b>Total</b>		<b>142</b>	<b>93,218</b>

The tools chosen to analyse the data in answering the research questions included frequency, collocation, and concordance analysis. Frequency analysis is a suitable starting point which can illuminate interesting phenomena (Baker, 2006) prior to more qualitative analysis, despite appearing to be a simple tallying exercise (McEnery and Hardie, 2012). To analyse frequencies, a list was generated of most common tokens in the corpus, and stop words were discarded. On removing stop words, the tokens which appeared most frequently in the corpus were tabulated, with both their raw frequencies and percentage of the corpus as a whole. As the Word Sketch function in Sketch Engine requires a minimum of a ‘few hundred’ (Rychly, 2008) occurrences to find meaningful data, the tokens which appeared over 200 times in the corpus were selected for further investigation, while others were discarded. Following the generation of a frequency list, the second step was collocation analysis, conducted through the Word Sketch function of Sketch Engine. Collocation analysis was chosen as collocations can exploit a topos



‘without the need for any explicit argumentation’ and therefore can suggest unconscious associations which are ways that discourse can be maintained (Baker and Gabrielatos, 2008).

The Word Sketch function locates collocations in relation to parts of speech, and also includes the statistical measure of Log Dice in the return of results. Log Dice is similar in interpretation to the Mutual Information score, in that it measures the strength of association between two terms (Evert, 2009) and has been described as scaling well on different corpus sizes (Rychly, 2008). Log Dice has a maximum value of 14, indicating a very strong association. Although there is no established cutoff point for collocations specified in the Log Dice score, Hunston (2010) proposes that in a similar score, Mutual Information, any value above 3 can be considered a collocate, while research from Durrant and Doherty (2010) suggests that a score of above 6 can be considered to trigger ‘mental priming’, in which encountering one word leads to thinking of the other (Baker, 2014). These associations can reveal the most overt lexical patterns, which can help the researcher obtain a number of different discourses (Baker, 2006). For the search in this study, a collocation span of +/- 3 was set, having been chosen during the pilot study. Word Sketches were produced using the settings to automatically locate collocations in all parts-of-speech with no minimum score, in an effort to generate a broad picture of the discursive patterns across the small, specialized corpus.

Following this, a concordance analysis was conducted to help validity-check hypotheses, as well as anchor, and triangulate findings (Baker and Egbert, 2016). Concordances can serve as a text by themselves (Sinclair, 1991), while also representing traces of discourse (Baker, 2006). Including concordance analysis also serves to combine quantitative and qualitative methods, serving as a more productive approach than relying on the former only (Baker, 2006).

Concordances were sampled to ensure a manageable amount of data (Tribble, 2010) and were identified using Sketch Engine’s randomized sampling feature. A set number is recommended for this process (Baker, 2018) and so of the tokens later selected for collocation analysis, three concordance lines were sampled for each, leading to a total of 12 concordances for close qualitative analysis. Concordances were analyzed for semantic preference, defined as ‘the relation between a word or lemma and a set of semantically related words’ (Stubbs, 200, 65), as well as through researcher-led introspection and interpretation of the concord (Baker, 2006).

## **4. Results**

### **4.1 Frequency Analysis**

Within the corpus, the most frequent terms related to the Rohingya, Myanmar, and Bangladesh. Moving away from these expected items however, there is equally a clear focus on the human elements of the crises, with ‘people’ and ‘refugees’ as the fourth and fifth most commonly occurring terms. A further point of interest is a focus on Aung San Suu Kyi, considering that both ‘Suu’ and ‘Kyi’ occur frequently throughout the corpus. In this sense, it seems that when the Rohingya crises are mentioned, it is often in relation to Aung San Suu Kyi as acting head of state, indicating an important power dynamic. In relation to semantic preference and the metaphor of semantic field (Morley, 2009), the crises seem to occur in conjunction with terms fundamentally related to security, law, military, and government. This is evidenced through terms 5, 11, 13, 14, and 15 in Table 2 below. Similar studies, such as those within the RASIM project (Baker and Gabrielatos, 2008) equally noted semantic fields of legality. From this introductory frequency analysis, it can be said that the Rohingya crises are often discursively centered around both the individual, such as Suu Kyi, and the collective, such as ‘people’, or ‘Rohingya’. On the other hand, there is a discourse of legality, signified by terms including ‘rights’, ‘court’, ‘government’, and ‘state’.

Table 2: Most Frequent Tokens in the Corpus

Number of terms	Token	Raw Frequency	Percentage of Corpus
1	Rohingya	1,092	1.2%
2	Myanmar	799	0.9%
3	Bangladesh	386	0.4%
4	People	355	0.4%
5	Military	305	0.3%
6	Refugees	277	0.3%
7	Suu	270	0.3%
8	International	255	0.3%
9	Genocide	255	0.3%
10	Kyi	234	0.3%
11	Rights	218	0.2%
12	Rakhine	213	0.2%
13	Court	210	0.2%
14	Government	206	0.2%
15	State	205	0.2%

## 4.2 Collocation Analysis

The results of the collocation analysis revealed further insight into discourses at work. The two terms with the highest frequency in the corpus were chosen for initial investigation, including ‘Rohingya’ (N=1,092) and ‘Myanmar’ (N=799), and two other terms were chosen for comparison based on the researcher-judged likelihood of revealing relevant information. The two terms selected were ‘Refugee’ (N=277) and ‘Genocide’ (N=255).

Beginning with ‘Rohingya’, the strongest collocates when ‘Rohingya’ is in the subject position in the corpus include ‘flee’, with a frequency of 40 and Log Dice score of 11.9, along with ‘protect’, which occurs 13 times and has a log dice score of 11.24. This suggests a strong associative relationship. On the other hand, ‘destroy’ is the second strongest collocate, with a frequency of 7 and Log Dice score of 10.43. Conversely, when examining Rohingya as the grammatical subject, the most common collocate is that of ‘flee’ with a high level of frequency (40) and an equally high Log Dice of 11.9. Here there is a similarity that can be drawn between those identified by Baker and Gabrielatos (2008) identification of categories related to RASIM representation, as there appears to be a clear set of collocates relating to provenance, travel, and

transit (flee, remain, arrive). Overall, there is a clear pattern in the set of collocations here which includes the subject fleeing, and the object in need of protection from destruction.

The second term to be analysed for collocations using Sketch Engine was ‘Myanmar’. The strongest collocations noted when the token occupied the subject position were ‘deny’ and ‘reject’, both with a Log Dice score of 10.5 and 10.73. These relatively strong relationships cast ‘Myanmar’ at the state level as having significant agency. Other collocates often included those related to armed forces, including ‘military’ and ‘soldier’ with Log Dice scores of 12.03 and 11 respectively. When Myanmar was in the object position, then the most common collocate again is ‘flee’, occurring as ‘flee Myanmar’, with a frequency of 34 and Log Dice score of 12.19. From this brief analysis, it is clear that ‘Myanmar’ functions as both a location and as an active agent in the construction of crises. Myanmar is noted metonymically as having power to withdraw or disallow.

Thirdly, ‘Refugee’ corresponded highly with ‘Rohingya’ (LD 9.54) demonstrating that commonly, the Rohingya are depicted first and foremost as refugees. A second strong collocate as a modifier of ‘refugee’ is ‘desperate’ (LD 9.53). When looking at nouns modified by ‘refugee’, ‘camp’ occurs 71 times in the corpus and has a LD score of 12.84, while ‘crisis’ has a frequency of 21 and LD of 11.48. Refugees as subject occurred frequently again with ‘flee’ (LD 11.56), and other verbs related to provenance also occurred, echoing the findings of Baker and Gabrielatos (2008) who found keywords relevant to concepts of ‘plight’ frequently occurred when describing refugees in newspapers in the United Kingdom.

The final token chosen for the collocation analysis was ‘genocide’. The most common collocations were ‘commit’ (LD 10.44) and ‘prevent’ (LD 11.63). Other terms of note that occurred at low frequency of 3 or lower included ‘possible’, ‘ongoing’, ‘alleged’ and ‘full blown’. Overall, the creation of a dichotomous discourse is consistent with the findings for ‘Rohingya’ and ‘Myanmar’. The patterns of ‘commit’ and ‘prevent’ genocide are coherent with the terms ‘protect’ and ‘destroy’ and to a lesser extent ‘deny’ and ‘reject’. These categories set up a relatively simple frame of events, in a manner that could be described as discursive simplification (Venkatamaran, 2019). In coming to a summary of this section, the collocation analysis demonstrates that there is often a construction of binary categories which are attributed to the tokens in question, creating a simplified discourse. There are equally considerable overlaps with findings of Baker and

Gabrielatos (2008) including common categories of ‘plight’ in terms such as ‘desperate’ and ‘fleeing’.

### 4.3 Concordance Analysis

In order to achieve a balance of comparative statistical techniques with researcher inferencing from data (Partington, 2018), a concordance analysis was then carried out on the four terms from the collocation analysis, including ‘Rohingya’, ‘Myanmar’, and ‘Refugee’ and ‘Genocide’. Using Sketch Engine’s inbuilt tools, a sample of every 10<sup>th</sup> concordance line was selected to give three concordances for each term, resulting in 12 concordance lines for overall qualitative-interpretive analysis. Inspecting the first set of concordances for ‘Rohingya’, the patterns identified by the collocation analysis of themes of protection were further confirmed in the first two sampled lines.

1. *There are calls to protect the Rohingya against further atrocities.*
2. *PROTECT THE ROHINGYA.*

The third concordance line sampled describes an individual’s experience with visiting camps in Bangladesh:

3. *She spoke of the "depths of suffering" she saw at the Bangladesh camps and said the world had "failed the Rohingya".*

In the above, there is mention of the world ‘failing’ the Rohingya people. This demonstrates an international outlook and sense of global responsibility. This is coupled with the emotive ‘depths’ of suffering, which contributes to the general picture of the severity of the crises. This concordance suggests the importance of internationality, and references to other actors on a global scale as having moral responsibility. Broadly speaking, this brief concordance analysis suggests that as with the collocation analysis, the Rohingya crises are at times presented in an internationalized context, with shared global responsibility for protection against horrific consequences, in the face of destructive and catastrophic consequences.

The second token for sampling of concordances is ‘Myanmar’. The sampled concordances demonstrate that ‘Myanmar’ is used metonymically at times to refer to the state. These instances tend to reveal more telling traces of discourse:

*4. Will Omar get justice for his murdered family? Myanmar has long denied carrying out genocide and says it is carrying out its own investigations into the events of 2017.*

In this concordance, ‘Myanmar’ is used metonymically and is linked to the state’s denial of genocide. There is an implicit judgement through the rhetorical question that justice should be sought, thus, in this case, support is shown for the Rohingya people while the Government of Myanmar is depicted as in a state of denial. Coming to the following line, again a theme of internationalization emerges.

*5. In Myanmar it is often their very communities, even families ... we need to ensure this issue is brought on at the highest world stage.*

Among the previous calls for protection, and the world ‘failing’ the Rohingya, this concordance argues that the Rohingya crises must be brought out for all to see on a global scale but stops short of indicating who must witness it at the ‘world stage’. From this, it can again be argued that there is a tendency towards describing the crises as needing intervention from a ‘global’ level, without specific calls for action or identification of those required. The next concordance sampled describes Aung San Suu Kyi and the ICJ case brought against Myanmar. The dispersion of articles suggests that this has been a major topic when discussing the Rohingya in online news media, given the high density of articles collected in 2019. This sampled concordance depicts the leader as standing in defense of alleged genocidal actions conducted by the state. Although this concordance is primarily descriptive, the frame in which it is contained is again one that references the supra-national, relating to the International Court of Justice. It could be argued that this again represents a discourse of internationalization of the Rohingya crises.

*6. Myanmar's leader Aung San Suu Kyi has defended her country against allegations of genocide at the UN International Court of Justice (ICJ).*

The final term for which concordances were sampled is ‘refugee’. Even from this limited sample, it is clear for the term ‘refugee’, that the media acknowledges the agency of the Government of Myanmar as denying, refusing, or rejecting. There is also an implicit call for a path to citizenship

as necessary in solving the crises, although academic research has suggested this is not necessarily a solution (Farzana, 2017).

*7. The government's refusal to allow a credible investigation, to offer freedom of movement or a clear path to citizenship, will deter most refugees from returning.*

The following sample refers to refugees at the camps in Bangladesh, in Cox's Bazar. This concordance offers reported speech but no direct quotation of the voice of the refugees in question, although it highlights frustration at a lack of input from the international aid community; the agencies or governments included under this heading are unidentified. Overall there is a sense that there is a failure on behalf of the international community to provide support for the crises.

*8. Refugees in Cox's Bazar complain the international aid community does not utilise their experience and say the lack of education risks creating a 'lost generation'.*

The final sample for the token 'refugee' resulted in the below concordance:

*9. The desperate journey of Rohingya refugees forced to leave their homeland amid 'human rights nightmare.'*

This ninth concordance relates again to the categories highlighted by Baker and Gabrielatos (2008) that common categories for the depiction of refugees in UK newspapers include provenance, transit, and plight. In this concordance, 'desperate' and 'nightmare' may relate to 'plight', while 'journey' and 'forced to leave their homeland' can be seen as broadly representing transit and destination.

The final set of concordances sampled were for the token 'genocide'. The first concordance relates directly to the case of the International Court of Justice and refers to a quotation which appeals to a general conception of humankind and a commitment to the end of genocide. This is an interesting concordance as again, there are links to a strong, interconnected action similar to that of the 'international' aspects identified in prior concordances.

*10. It is indeed sad for our generation that 75 years after humankind committed itself to the words "never again", another genocide is unfolding right before our eyes,' Gambian Justice Minister Aboubacarr Tambadou told the court.*

The following concordance again refers to the ICJ case, and this appears to be a key driver of the reporting on the Rohingya crises. This concordance uses reported speech and describes the lack of action that the 'world' has taken in protecting the Rohingya or stopping genocide from occurring:

*11. Abubacarr Marie Tambadou, the Gambia's attorney general and justice minister, opened the case against Myanmar. He declared the world had done "nothing" to stop the genocide as it unfolded.*

The final concordance appears to describe a general description of the crises in Myanmar as a whole, along with accusations of genocide. This concordance again references Aung San Suu Kyi as a key figure, and one with responsibility for the issues being described.

*12. Political unrest has dogged the country in recent years, with accusations of genocide and ethnic cleansing hanging over premier Aung San Suu Kyi.*

In summary, the concordance samples for these four tokens provide telling insights into the discursive representation of the Rohingya crises in online news media in the United Kingdom, and a number of similar patterns can be observed between those developed in the collocation and concordance analysis.

## **5. Discussion and Conclusion**

This research study aimed to provide a preliminary investigation into the representation of the Rohingya minority group and the humanitarian crises facing them in online news media in the United Kingdom, through a CADS approach, using a small, specialized corpus. Although this is a small-scale study, it has demonstrated that the corpus is a useable tool and can be used to identify some traces of discourse (Baker, 2006), through the examination of connected, mutually supporting statements (Koteyko, 2014).



The frequency analysis showed a semantic field focused around legality, law, security, and government. This was an interesting point to begin further investigation. On further analysis of the collocative patterns for the terms ‘Rohingya’, ‘Myanmar’, ‘Refugee’, and ‘Genocide’, several other discursive features became clear. As with Baker and Gabrielatos (2008), common terms included ‘fleeing’ relating to the semantic category of ‘plight’, while when ‘Rohingya’ was in the object position, the token coincided with ‘protect’ or ‘destroy’. ‘Myanmar’, on the other hand, was depicted as having agency when in the subject position, with terms such as ‘deny’, or ‘reject’ commonly co-occurring, along with ‘military’, marking the role of Myanmar’s military within construction of the crises.

Finally, when triangulating the findings using close readings of sampled concordances, it is clear that in addition to the semantic fields of legality, security, government, and power (in terms of denial, protection, and rejection) there is a clear discourse of internationalization. Most concordances sampled characterized the Rohingya crises in terms of global responsibility and its inability to provide protection from destruction or from genocidal acts, along with reference to the international aid community, the ‘world stage’ and the failure of the world to act. These findings demonstrate powerful support for the suffering of the Rohingya people, and do not demonstrate the same ‘exclusionary rhetoric’ that has been found in other studies of refugee representation (KhosraviNik, 2014; Baker and Gabrielatos, 2008), but do confirm the findings of Afzal (2016) that the Rohingya minority are generally depicted sympathetically in media discourse.

This study has therefore identified several findings relevant to the representation of the Rohingya and the Rohingya crises affecting the minority group in Myanmar, Bangladesh and around the world. The findings demonstrate unique discourses that are employed in online news media in the United Kingdom, and strengthen findings of other authors (Afzal, 2016) that sympathetic discourses are often present, in contrast to findings of other studies. Furthermore, specific dimensions of international responsibility, legality, and security are often present, while verbs that are used in conjunction with the term ‘Rohingya’ often include those that apply power and agency, for example ‘destroy’ and ‘protect’. This study has equally demonstrated the ability of a CADS approach to examine online news media data from UK-based providers effectively through freely available digital archives. As further crises occur in Myanmar, additional research needs to be undertaken to ensure visibility and action is taken to reduce unnecessary suffering and provide solutions to such pivotal humanitarian crises.

Research projects in this field could aim to explore broader ranges of media data from cross-cultural perspectives, especially those which contrast discursive representations in Eastern and Western media. Further investigations may combine the CADS methods with a range of other techniques, for example qualitative interviews, in order to generate clearer comparisons between the reality of life for the Rohingya people and the representations found in mainstream media. These potential projects could contribute to finding suitable solutions and generating further understanding of the role of media representations in such crises.

## **References**

- Afzal (2018). Elements of pathos and media framing as scientific discourse: A newspaper perspective on Rohingya crisis. *International Journal of Advanced and Applied Sciences*. 3, 89-99.
- Albert, E. & Chatzky, A (2018). The Rohingya crisis. *Council on Foreign Relations*. Retrieved from <https://www.cfr.org/background/rohingya-crisis>.
- Baker, P. & Egbert, J (2016). *Triangulating methodological approaches in corpus linguistic research*. Routledge.
- Baker, P. & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*. 4, 197-226.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Bowser (2019). "I don't want to see them in this country": colonialism, capitalism, and the origins of Burmese Islamophobia 1930-1938. In B. Ahmed (Ed.). *International Conference on the Rohingya Crisis in Comparative Perspective: Book of Abstracts*. London: University College London.
- Brooten, L. & Verbruggen, Y. (2017). Producing the news: Reporting on Myanmar's Rohingya crisis. *Journal of Contemporary Asia*, 47, 440-460.
- Brooten, L. (2015). Blind spots in human rights coverage: Framing violence against the Rohingya in Myanmar/Burma. *Popular Communication*, 13, 132-144.
- Cheesman, N. (2017). How in Myanmar "National Races" came to surpass citizenship and exclude Rohingya. *Journal of Contemporary Asia*, 47(3), 461-483.

- Durrant, P. & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6, 125-155.
- Daghighi, A., & Rahim, A H (2020). Representation of Muslim minorities in politicians' discourse: Jacinda Arden vs Donald Trump. *Journal of Muslim Minority Affairs*, 5(2), 179-195  
DOI: 10.1080/13602004.2020.1773099
- Dussich, J. (2018). The ongoing genocidal crisis of the Rohingya minority in Myanmar. *Journal of Victimology and Victim Justice*, 1(1), 4-24
- Fairclough, N. (1992). Discourse and text: Linguistic and intertextual analysis within discourse analysis. *Discourse & Society*, 3, 193-217.
- Farzana, K. (2017). *Memories of Burmese Rohingya refugees: Contested identity and belonging*. Palgrave Macmillan.
- Freake, R., Gentil, G. & Sheyholislami, J (2010). A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society*, 22, 21-47.
- Fullbrook, R. (2019). Preface. In B. Ahmed (Ed.) *International Conference on the Rohingya Crisis in Comparative Perspective*. Book of Abstracts. London: University College London.
- Gabrielatos, C. & Baker, P. (2008). Fleeing, sneaking, flooding. *Journal of English Linguistics*, 36, 5-38.
- Garland, D. (2014). What is a "history of the present"? On Foucault's genealogies and their critical preconditions. *Punishment & Society*, 16(4), 365-384.
- Gusfield, J. (1986). *The culture of public problems*. Univ. of Chicago Press.
- Hunston, S., (2010). *Corpora in applied linguistics*. Cambridge University Press.
- ICJ (2020). *Application of the convention on the prevention and punishment of the crime of genocide (The Gambia v. Myanmar)*. ICJ. Retrieved from <https://www.icj-cij.org/files/case-related/178/178-20200123-PRE-01-00-EN.pdf>.
- KhosraviNik, M. (2014). Dynamics of world events and immigration representations in the British press. In C. Hart & P. Cap (Eds.). *Contemporary critical discourse studies*. Bloomsbury, London, pp.501-519.
- Koester, A. (2010). Building small specialised corpora. In A. O'Keefe & M. McCarthy (Eds.). *The Routledge handbook of corpus linguistics*. (pp.66-89). Routledge.
- Koetyko, N. (2014). *Language and politics in post-Soviet Russia*. Palgrave Macmillan.

- McEnery, T., Hardie, A. (2012). *Corpus Linguistics: Method, theory and practice*. Cambridge University Press.
- McIlvenny, P. (1996). Heckling in Hyde Park: Verbal audience participation in popular public discourse. *Language in Society*, 2, 27-60.
- Morley, J., & Partington, A., (2009). A few frequently asked questions about semantic or evaluative prosody. *International Journal of Corpus Linguistics*, 14, 139-158.
- MSF (2002). *Ten years for the Rohingya refugees: Past, present, and future*. Retrieved from <https://www.msf.org/ten-years-rohingya-refugeespast-present-and-future-report-summary>.
- MSF (2019). *Timeline: A visual history of the Rohingya refugee crisis*. Retrieved from <https://www.doctorswithoutborders.org/what-we-do/news-stories/news/timeline-visual-history-rohingya-refugee-crisis>.
- Nahar (2019). The 2017 Rohingya crisis: A corpus-assisted discourse analysis of representation in The Daily Star. *International Journal of Science and Humanity*, 9 (4).
- Ofcom (2019). *News consumption in the UK: 2019*. Retrieved from [https://www.ofcom.org.uk/\\_\\_data/assets/pdf\\_file/0027/157914/uk-news-consumption-2019-report.pdf](https://www.ofcom.org.uk/__data/assets/pdf_file/0027/157914/uk-news-consumption-2019-report.pdf).
- OHCHR (2014). Expert raises alarm on Rakhine State. *OHCHR*. Retrieved from <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=14476&LangID=E>.
- PAMCo (2020). About us. *PAMCo*. Retrieved from <https://pamco.co.uk/about-us/>
- Pugh (2013). *Is citizenship the answer? Constructions of belonging and exclusion for the stateless Rohingya of Burma*. Centre on Migration, Policy and Society Working Paper No 107.
- Rychly, P. (2008). A lexicographer-friendly association score. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*, 6-9.
- Said, E. (1978). *Orientalism*. Pantheon Books.
- Scholz, R. (2019). *Quantifying approaches to discourse for social scientists*. Springer International Publishing.
- Siddiquee, M. (2019). The portrayal of the Rohingya genocide and refugee crisis in the age of post-truth politics. *Asian Journal of Comparative Politics*, 5, 89-103.
- Sinclair, J. (1991). *Corpus concordance collocation*. Oxford University Press.

Jasper Roe

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell.

Tribble, C. (2010). What are concordances and how are they used? In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. (pp.3 – 13). Routledge.

Venkataraman, N. (2018). What's not in a frame? Analysis of media representations of the environmental refugee. In M. Schroter & C. Taylor (Eds.). *Exploring silence and absence in discourse – Empirical approaches*. (pp.241-279) Palgrave.

WHO (2019). *Invisible: The Rohingyas, the crisis, the people and their health*. World Health Organization.