

Register, Belief and Violence: A multi-dimensional approach

Tony McEnery and Gavin Brookes
Lancaster University

1. Introduction

This chapter examines the linguistic features of texts promoting extremist violence. According to the Government of the United Kingdom – the context in which our work is situated – *extremism* can be defined as ‘the vocal or active opposition to our fundamental values, including democracy, the rule of law, individual liberty and the mutual respect and tolerance of different faiths and beliefs.’ (HM Government 2015: 9). The Government also notes that it ‘regard[s] calls for the death of members of our armed forces as extremist’ (ibid.). As this definition indicates, in legal terms at least, extremism is a broad concept capable of encapsulating a wide range of actions, including linguistic (i.e. ‘vocal’) behaviours, which could be deemed to threaten any of the Country’s major institutions or to impinge upon the rights or safety of its citizens. In this chapter, we examine texts which promote ideologically motivated extremist violence, focusing in particular on texts found in the possession of convicted violent jihadists. Note that, as a result, our use of quotations from the texts in question will be sparing – we have no wish to further the promotion of violence ourselves. Accordingly, where we do, for illustrative purposes, provide brief examples, we will not produce any which explicitly promote violence.

Conceptualisations of the relationship between extremism and Islam are, as Baker and Vessey (2018: 257) point out, both ‘complex and contested’. A relevant distinction here being that between the adjectives ‘Islamic’ and ‘Islamist’. Where ‘the former has been popularly used in western media (and implies a form of extremism connected to Islam generally) and the latter refers to extremism connected more specifically to politically motivated Islam’ (ibid.). The related term, ‘Islamism’, they contend, refers to ‘the desire to impose a version of Islam over society while other terms like “militant Islam”, “radical Islam” or “fundamentalist Islam” complicate definitions further and are also often found in western news media, implying that Islam is an intolerant and insensitive religion’ (see also: Kramer 2003; Baker, Gabrielatos and McEnery 2013; Esposito 2014). As well as providing a definition of extremism, the British Government also offers a definition of Islamist extremism specifically, which it describes as ‘any form of Islam that opposes democracy, the rule of law, individual liberty and mutual respect and tolerance of different faiths and beliefs’ (HM Government 2013: 1). However, given the clear and understandable sensitivities regarding the use of terms such as Islamism and Islamist, in this chapter we will refer instead to *violent jihadism* when discussing the motivations of the people from whom these texts were acquired.

In this chapter we analyse the register features of violent jihadist texts by drawing on techniques from corpus linguistics. Corpus linguistics refers to a group of methods that use specialist computer programs to study language in large bodies of machine-readable text (see McEnery and Hardie 2012). Such a body of data is known as a *corpus* (pl. *corpora*) – the Latin word for ‘body’. Corpora are usually assembled in a principled manner with the aim of representing a language or particular linguistic variety on a broad scale (Biber, Conrad and

Reppen, 1996). Based on a corpus containing texts collected by violent jihadists, we apply an analytical framework known as Multi-Dimensional Analysis (introduced in the next section) to ascertain the extent to which the texts in our corpus constitute a single register or whether they form a range of registers. Furthermore, we also investigate whether the register features of these texts differ according to the degrees of extremism that they are judged to encode and propagate. As will become clear, the answers to these questions are not only linguistically interesting, but also raise considerations for those working within organizations, such as counter-terrorism forces, that are responsible for identifying such texts both quickly and reliably.

We should note at this point that while we are focusing on violent jihadists here, we do not perceive Islam to be a religion that is in any way characteristic of extremism or particularly likely to motivate violence. As such, we could have studied extremist language in texts produced by and for members affiliating with other religions and ideological groups such as fascists (as studied by Richardson, 2017) or white supremacists (as studied by Brindle, 2017). We study violent jihadists here because we were provided with a privileged insight into the texts collected by members of this ideological group, as will be discussed in Section 3.. Before considering that data, however, the next section provides a theoretical and methodological backdrop to this study by introducing the central concept of register and the aforementioned Multi-Dimensional Analysis approach to register.

2. Register and Multi-Dimensional Analysis

The perspective of register combines the analysis of linguistic characteristics that are frequent within a text variety with the analysis of the real-world situations in which that variety is used. The assumption underlying analyses of register is thus that linguistic features are functional, with particular features bearing an association with texts' communicative purposes and situational contexts. Registers can therefore be considered groupings of texts that are defined by factors that are external to the texts themselves, such as the social or situational conditions of their medium, their contexts of production, or their purpose. Biber and Conrad (2009: 6) point out that the description of a register covers three major components: 'the situational context, the linguistic features, and the functional relationships between the first two components' (ibid.). They elaborate:

Registers are described for their typical lexical and grammatical characteristics: their linguistic features. But registers are also described for their situational contexts, for example whether they are produced in speech or writing, whether they are interactive, and what their primary communicative purposes are. [...] [L]inguistic features are always functional when considered from a register perspective. That is, linguistic features tend to occur in a register because they are particularly well suited to the purposes and situational context of the register. Thus, the third component of any register description is the functional analysis.

(Biber and Conrad 2009: 6)

As noted, to examine the register features of the extremist texts in our data, we subject those texts to a Multi-Dimensional Analysis (MDA). MDA is an approach to register analysis

developed by the linguist Douglas Biber during the 1980s (Biber 1984, 1988) as a way of identifying the major linguistic parameters along which textual registers vary in English. MDA is driven by a lexico-grammatical account of register variation, since Biber's argument is that registers are formed by distinct combinations of words and grammatical categories. MDA rests upon the use of factor analysis to identify the co-occurrence of particular linguistic features in a text or group of texts. Biber (1988) uses 67 of these, which are grouped into 16 broader categories including, *inter alia*, tense and aspect markers, place and time adverbials and pronouns and pro-verbs (see also: Conrad and Biber 2001: 18-19). These 67 features form the basis for the investigation presented here.¹

To give an example of how these features help to identify groups of linguistic features relevant to identifying a register, Biber (1988) observed a pattern whereby texts with a high frequency of, *inter alia*, private verbs (e.g. *believe*, *think*) are also likely to exhibit a high frequency of *that*-deletion and contractions, as well as the lower frequency of such features as nouns, prepositions and attributive adjectives. These features combine in different ways to form different 'dimensions', or 'sets of syntactic and lexical features that co-occur frequently in texts' (Biber 1989: 5), along which registers place themselves – so the features bundle to create a number of dimensions and the registers place themselves in distinct configurations along those dimensions.

The dimensions involved in MDA are interpreted and labelled in terms of their perceived functions. Biber (1988) proposed six major dimensions of variation in English, to which he assigns the following functional labels: Dimension 1: Involved v. Informational, Dimension 2: Narrative v. Non-Narrative, Dimension 3: Elaborated v. Situation-Dependent, Dimension 4: Overt Expression of Argumentation, Dimension 5: Abstract v. Non-Abstract, and Dimension 6: Online Informational Elaboration. In MDA, each text comprising the data can be situated along each dimension in accordance with the dimension score assigned to it. The mean dimension score assigned to a group of texts can then be used to characterise its discourse. For example, Biber's Dimension 1, 'Involved versus Informational Production', comprises 25 features with high scores, including, for example, the use of private verbs, *that*-deletion, contractions, present tense verbs, and second-person pronouns, meanwhile features with low negative loadings include nouns, word length, prepositions, and attributive adjectives (Berber Sardinha 2018: 127). In this case, the frequent cooccurrence of the features along the positive pole results in the texts in question being interpreted as having an 'involved production' communicative function. The cooccurrence of features along the negative pole, on the other hand, indicates a shared communicative function of 'informational production'. Because the features along the positive and negative poles tend not to occur with similar frequency within the same texts, the presence of the features of one (e.g. 'involved production') usually indicates that the features of the other (e.g. 'informational production') are largely absent. Other dimensions, such as 2 and 4, have a single pole, meaning these dimensions are characterised by the frequency or absence of a single set of linguistic features. For example, for Dimension 2, texts can be characterised either as containing narrative or

¹ Readers interested in finding out more about the features can find a useful crib sheet here: <http://corpora.lancs.ac.uk/stats/docs/multidimensional.pdf>.

non-narrative features, and Dimension 4 as containing either overt or non-overt persuasion (Berber Sardinha 2018: 128).

When carrying out MDA, each text in a corpus is simultaneously scored for each dimension using standardised counts of the relevant features. This means that each text will be assigned a score for each dimension. Analysts typically proceed by calculating the mean dimension scores for each of the registers represented in their data, leading to the characterisation of registers in terms of the aforementioned dimensions. For example, on the basis of these dimensions, Biber (1988) identified the following registers: General narrative exposition; Imaginative narrative; Involved persuasion; Learned exposition; and Scientific exposition. The characteristics of individual registers become more salient, and are rendered more apparent, when their mean dimension scores are compared against each other, essentially illuminating the most salient linguistic characteristics of each.

Since the development of the MDA approach, research carried out by Biber, his colleagues and others has focused on extending this method and/or on applying it to new areas of inquiry. Indeed, the MDA framework has been applied to the analysis of an impressive and growing range of registers and discourse domains, as well as to an increasing number of languages, where the patterns of register variation originally put forward by Biber (1988) have proven to be a useful starting point for analyses. Although language-wide studies have been carried out, the majority of MDA research focuses on language used in specific contexts, producing accounts of the registers that are characteristic of domains as diverse as university writing (Biber 2006), televised dialogue (Quaglio 2009), call centre interactions (Friginal 2008), pop song lyrics (Bértoli-Dutra 2014), medical encounters (Staples 2015) and political tweets (Clarke and Grieve 2019), to name just a few (see Biber 2019 for a comprehensive overview). This chapter presents a study of a relatively under-researched type of discourse which has yet to be analysed using the MDA framework; extremist language.

Ours is not the first linguistic study concerned with the discourse of Islamist extremism. For example, Droogan and Peattie (2016) examined shifts in the themes in Al Qaeda's *Inspire* magazine using a modified form of thematic network analysis. Wignell, Tay and O'Halloran (2017) utilised a multimodal approach to discourse analysis in order to explore the use of image and text in the magazine of the so-called Islamic State of Iraq and Syria (ISIS), *Dabiq*, as well as in online materials produced by an affiliated British group, Rayat al Tawheed. These authors were particularly interested in how such multimodal ensembles functioned to rally their assumed reader-supporters. Also adopting a multimodal perspective, Ingram (2017) compared both of the aforementioned publications, *Inspire* and *Dabiq*, in a study which analysed how each publication used language and imagery in their attempts to appeal to and radicalise their readerships. A small body of work in this area has also employed more quantitative, including corpus linguistic, methods in order to study the language associated with Islamist extremism. Prentice et al. (2011) utilised the corpus analysis tool *Wmatrix* (Rayson 2008) to analyse the persuasive strategies emergent from texts that they claimed incited violence in the context of the Gaza conflict. In a later study, Prentice et al. (2012) utilised this same tool in order to compare the key semantic domains emerging from corpora representing pro- and counter-extremist texts. More recently, Conoscenti (2016) used the corpus technique of collocation analysis in order to analyse the communicative strategies of *Dabiq*, while Baker and Vessey (2018) combined quantitative corpus methods

with more qualitative, manual discourse analysis in their comparative study of the discursive themes and linguistic strategies employed in the aforementioned English-speaking *Inspire* and *Dabiq* magazines and ISIS’s French-speaking magazine, *Dar al Islam*. Corpus methods have also been adopted in studies examining texts that can be viewed as more indirectly relating to or contributing towards extremist discourse, for example McEnery, McGlashan and Love’s (2015) corpus-assisted discourse analysis comparing press and social media representations of the murder of Lee Rigby by two violent jihadists in London in 2013.

Despite a seeming surge in linguistic interest in the language associated with violent jihadism in recent years, to our knowledge such texts have yet to be subjected to a register analysis. As such, we are currently unfamiliar with the register features of such texts and cannot be sure, in empirical terms at least, of the extent to which texts produced to incite jihadist extremist violence constitute a register distinct from those used in more general writing about Islam. Accordingly, we do not know the extent to which there are linguistically quantifiable differences between moderate, fringe and extreme texts about Islam which may break them into distinct registers. In this chapter, we set out to answer these questions and the next section outlines the data and methodological approach that we use to do it. It also presents the definition of *moderate*, *fringe* and *extreme* used in this chapter.

3. Methodology

The texts analysed in this chapter derive from an ongoing project examining jihadist discourse (see: Baker, McEnery and Vessey, forthcoming; Brookes and McEnery, 2020). The titles of texts associated with 11 successful terrorist prosecutions were supplied to us by contacts working in counter-extremism at the British Home Office or the London Metropolitan Police. The texts had been found on the hard drive of violent jihadists successfully prosecuted for terrorist offences in the UK. All the texts were in some way centrally concerned with Islam. Based on the data provided to us, we were able to find the texts in question and build the corpus used in this study.

So far, we have referred to the texts in our data simply as ‘extremist texts’. However, we can be more specific; experts in counter-terrorism research categorised these texts and made their categorizations available to us (see Holbrook (2015) for details of the coding of the texts). Based on their expert close reading of the texts, they classified each as either ‘moderate’, ‘fringe’ or ‘extreme’. Moderate texts do not promote social isolation or violence, fringe texts do not overtly promote violence but they do promote alienation from mainstream society while extremist texts openly advocate and facilitate violence. Table 1 provides a breakdown of the number of texts and words belonging to each of these categories.

Table 1. Corpus composition, by ideological category.

Corpus	Texts	Words
Extreme	170	1,775,340
Fringe	54	486,650
Moderate	51	1,721,442

The Extreme sub-corpus contains numerous articles from the aforementioned ISIS magazines, *Inspire* and *Dabiq*, as well as a variety of other texts including transcripts of interviews and lectures, biographies, political treatises, statements released by groups like ISIS and Al Qaeda, how-to guides which contain advice on topics like computer encryption, bomb-making and engaging in combat, and articles written in the style of news reports. The fringe texts contain more ambiguous messages which could be interpreted as advocating violence, but which do so implicitly and as such also have potential for non-violent readings. Such texts advocate withdrawal or segregation from civil society. Thus, while the Extreme texts were likely crafted with the intention of advocating and enabling violence, the purpose of the Fringe texts is less clear, but they could operate as ‘gateway’ texts designed to subtly nudge readers towards more extreme positions. Alternatively, we could interpret the Fringe texts as having been written strategically to advocate violence but in a way that evades the attention of legal authorities. Texts categorised as ‘Moderate’ were those which referred to Islam in some way, but which were not categorised as ‘Extreme’ or ‘Fringe’. Such texts typically involved scholarly discussion of religious topics in books and other sources. Some of these texts were written by people associated with terrorism or incitement to violence in other contexts but did not do so in the texts themselves, neither implicitly nor explicitly. The texts were judged by the experts rating them to advocate a more tolerant view of Islam based on co-operation and/or more peaceful co-existence with non-Muslims. Other texts in this category were written by non-Muslims, although they all had Islam as a central theme. Given that all of the owners of these texts were native speakers of English, texts belonging to this final category are potentially useful for identifying aspects of language which are likely to be familiar to English-speaking Muslims, though they are not necessarily associated with extremist discourse. Nonetheless, we had to consider the possibility that such aspects might also be present in the ‘Extreme’ texts too, perhaps because they are well-known to all Muslims but could also function as a means of making extremist language more familiar and persuasive for their intended Muslim readerships.

These categorisations are useful for our analytical purposes, as they help us to frame the texts in terms of ideology, given that central to each of the categories is a perceived attempt, on behalf of the author, to propagate a particular worldview. For these purposes, we use the term *ideology* to describe ‘the way in which what we say and think interacts with society’, whereby ideology ‘derives from the taken-for-granted assumptions, beliefs and value systems which are shared collectively by social groups’ (Simpson 1993: 5). Similarly, at the level of reception, these texts were also perceived to have contributed to the formation of the worldviews of the individuals who had been radicalised to the extent that they had committed or planned to commit terrorist crimes in pursuit of the ideologies propagated by at least some of these texts. Although this categorisation was not carried out by linguists, and as such was based on non-linguistic characterisations of ideology, we are of the view that language is nevertheless likely to be significant to the ways in which those ideologies are expressed (*ibid.*). Our analysis therefore sets out to explore the extent to which these categories, of ‘moderate’, ‘fringe’ and ‘extreme’, are *linguistically* meaningful, and to find out whether distinct forms of language might be associated with each of them in such a way that these categories align with distinct registers.

When assessing a corpus, a key consideration is representativeness. According to Biber (1993: 244), the representativeness of a corpus is determined by ‘the extent to which a sample includes the full range of variability in a population’. This dataset can be described as representing language associated with violent jihad, since in all 11 cases the text possessors were convicted terrorists. Beyond this, a principled approach to corpus compilation might involve the use of a sampling frame, obtaining extremist texts which provide a balanced view of a range of known terrorist organisations, ideologies and different countries and languages. Our corpus does not follow such a sampling frame. Instead all available texts from the context studied were included in it. The advantage of working with the texts that we have is that we know that they have been acquired, and presumably read, by people who have been legally judged to have become violent jihadists. The texts were written almost exclusively in English, reflecting the first language of the people who were convicted who had thus sourced and read the materials, although they also contained some evidence of code-switching to Arabic. As such, the corpus is maximally representative of the context studied – texts collected by and found in the possession of 11 convicted extremists.

Finally, due to its unusual nature, there are certain details about this corpus which cannot be shared with readers. Although some extremist texts, like the *Inspire* and *Dabiq* magazines, have been reported on by mainstream media, others are less widely known. To prevent other such publications from gaining public prominence, we will not provide a list of the names of the texts in our corpus, thus the usual claims pertaining to replicability in corpus linguistics research (see: Leech 1992) cannot apply in this case. Further details on corpus compilation and cleaning, including discussion of the legal and ethical issues attending to the use of extremist materials in linguistic research, can be found in Baker, McEnery and Vessey (forthcoming).

Having introduced MDA in some depth in the previous section, we will not go too deeply into the workings of the technique here. However, on a practical level, we analysed each of the corpora set out in Table 1 (respectively containing the ‘extreme’, ‘fringe’ and ‘moderate’ texts) using MDA. We began by using the ‘Analyze Corpus’ function of the *CQPweb* tool (Hardie 2012) to generate the frequency information necessary to undertake MDA. This frequency information then formed the bases of the MDA which was carried out using the freely available online *Lancaster Stats Tools*.² This analysis is reported in the next section.

4. Analysis

4.1. Initial findings: Registers

The MDA approach is useful in determining the extent to which the texts within each of the three classifications expressed some degree of linguistic homogeneity with regard to register. If we look at the three classifications and, on the basis of the MDA undertaken, ask the question, ‘What registers do the texts in the corpus most closely resemble?’, the answer should reveal differences and similarities between the classifications. Table 2 below shows,

² See <http://corpora.lancs.ac.uk/stats/toolbox.php>. This website is best used in concert with Brezina (2018).

for each classification, which register is the closest match for the texts in that section of the corpus, using the aforementioned registers established by Biber (1988) for reference.

Table 2. Correspondence of each sub-corpus with the registers established by Biber (1988).

Register	Moderate		Fringe		Extreme	
	Count	Percentage	Count	Percentage	Count	Percentage
General narrative exposition	28	54.90	27	50	118	67.82
Imaginative narrative	-	-	1	1.85	2	1.15
Involved persuasion	17	33.33	14	25.93	33	18.97
Learned exposition	6	11.76	6	11.11	10	5.75
Scientific exposition	-	-	6	11.11	11	6.32

Of note in this table is the overall difference between the moderate texts and the rest; Moderate texts are most similar to general narratives, involved persuasion and learned exposition. Only the fringe and extreme texts approximate to imaginative narrative and scientific exposition. Similarly, as we move across the table another difference is apparent – the relative proportion of texts which are most similar to general narrative increase markedly in the extreme category. Overall, narrative is clearly important to all three categories – 54.90% of moderate, 51.85% of fringe and 68.97% of extreme texts are most similar to either general or imaginative narrative.

Unless readers are familiar with the system of analysis used by Biber, the results of it can appear to be somewhat abstract. However, the labels produced by the system are typically good guides to the content of the texts thus classified. Below is a brief example, drawn from the corpus of general narrative exposition, by way of illustration. It is an extract from the book *My Life with the Taliban* (Zaeef, 2010).

We were still waiting by the road when I saw the tanks coming, firing flares into the sky. Burning debris fell all around us, hitting cars here and there. They pointed their guns at the cars along the road, screaming at people like animals.

This was the first time I had seen a convoy in Kandahar. It was very strange, and worrying. I asked my friend whether it was always this bad. ‘Today was a good day’, he said. ‘This is our daily routine, and many times lives are lost when they pass through the city’.”

The work in question is indeed a general narrative – an autobiographical account of life in the Taliban. The short passage quoted shows some of the key features of narrative within the Biber model that places this text into the narrative category – for example, the use of third person pronouns (e.g. *he*), past tenses (*were, saw, fell, was*) and perfect aspects (*pointed, asked*). Across this text, such features appear in a preponderance such that the factor analysis used to carry out our analysis determines that the set of such features is so relatively frequent in the text that it is placed, on a continuum between non-narrative and narrative, on the narrative part of the continuum.

4.2. Initial Findings: Authors

A possibility this raises is that, if we consider the author, publisher or organization linked to the text in question, we may find a distinctive style emerge – some authors who prefer a narrative style to learned exposition, for example. The dataset we are using allows us to approach that problem – while the majority of authors and organizations represented in the text typically provide but one text, we can look at 10 cases in the corpus where there are five or more texts from a single author or organization. The results of that analysis are shown in Table 3. In the Table, the number of texts produced in the register that an author produces most frequently is emboldened.

Table 3. Correspondence of texts produced by each author with the registers established by Biber (1988).

Register	ISIS	Awlaki	Bin Laden	Maqdis	Yahya	Hizb	Azzam	Oadah	Muhajiroun	Deedat
General narrative exposition	29	3	19	4	2	6	6	2	2	5
Scientific exposition			3	2					3	
Imaginative narrative			1							
Involved persuasion		6	1	3	6	1	2	4		
Learned exposition	1	1			1	1			1	
Total	30	10	24	9	9	8	8	6	6	5

While the volume of data in this table is not substantial enough to allow meaningful statistical analysis, several points can be made, informally, about it. Firstly, General Narrative Exposition, the dominant register in Table 2, is the only register produced by all authors in Table 3. For most authors (6 in total), it is also the register that they produce most frequently. Of the authors who do not produce this register most frequently, three (Awlaki, Yahya and

Oadah) produce Involved Persuasion most frequently while Muhajiroun produce Scientific Exposition most frequently. ISIS stands at the opposite extreme to authors like Awlaki in that their documents are almost exclusively General Narrative Exposition. So, without examining at this point *why* different authors and organizations seem to make choices about which registers to use, we can certainly see some evidence in our dataset that while General Narrative Exposition is indeed a very dominant register in the corpus in general and in extremist writing in particular, there are authors whose writing is represented in other styles and some of those are more prevalent for them in the corpus than General Narrative Exposition. It is very tempting, at this point, to ascribe choice to the author in this. However, while our data can allow us to say that for a specific document an author did draw upon a particular register, we cannot, for example, in the case of Awlaki argue that he prefers an authorial style based upon Involved Persuasion. This is because we are studying what was collected by the violent jihadists, not what, in general, Awlaki wrote. If we were studying a comprehensive corpus of Awlaki's writings and found that he had an overwhelming preference for Involved Persuasion, we might begin to make a claim about what style he prefers. On the basis of the evidence we have we might make this inference, but similarly we might make the inference that the violent jihadists, for some reason, prefer his writing in this register and that they downloaded it accordingly, ignoring other registers that he wrote in.

4.3. Initial findings: Mode of Production

A quite different way of approaching the data would be to consider the different modes of production in the text; in other words, was the data written or a transcription of discourse that was originally spoken? This may explain some of the findings presented so far. For example, if we find that the Extremist material has within it more speech than writing and subsequently discover that the spoken data falls more commonly into the General Narrative category, we have an explanation, based on choice of mode of production, which would explain the apparent preference of the Extremist data for this register.

In order to categorise our data, we were able to rely both on notes made by the analysts who had provided us with the titles that formed our corpus and our own readings of the text to determine when we were dealing with what claimed to be transcribed speech. Of course, we were not able to determine whether the speech was spontaneous or if it was text which had been spoken, but this is precisely the type of distinction that we could expect the MDA to identify, should it be present in the data, as shown in previous studies (Biber 1984). Applying the perspective of mode of production split the data into 237 written texts and 38 transcribed spoken texts. The initial coding of the texts as written and spoken revealed a marked skew in the data, as Table 4 shows.

Table 4. Number of texts in the corpora, categorised by mode (percentages in brackets).

	Written	Spoken
Moderate	51 (100%)	0 (0%)
Fringe	49 (90.74%)	5 (9.26%)
Extreme	137 (80.58%)	33 (19.42%)

So, the hypothesis that the data may, in fact, be prone to skew depending on mode of production does seem worth investigating. In order to do so, the data was subject to MDA once again, this time with the texts categorised as speech and writing alone – if the spoken language across all categories (moderate, fringe, extreme) distributes differently to the written language, this would give some initial indication that any difference in register preferences of the three categories (moderate, fringe, extreme) may be a consequence of there being different proportions of speech and writing across the categories. The results of placing the spoken and written data on Biber’s six dimensions of variation are given in Table 5 below.

Table 5. Variation in the corpus along Biber’s dimensions of variation.

Dimension	Results ³
1. Involved v. Informational	F = 2.2; p = 0.1353808, r ² = 0.8%
2. Narrative v. Non-Narrative	F = 5; p = 0.02569213, r ² = 1.8%
3. Elaborated v. Situation-Dependent	F = 1.5; p = 0.2286787, r ² = 0.5%
4. Overt Expression of Argumentation	F = 2.9; p = 0.09232212, r ² = 1%
5. Abstract v. Non-Abstract	F = 10.4; p = 0.001425511, r ² = 3.7%
6. Online Informational Elaboration	F = 2.1; p = 0.1524262, r ² = 0.7%

These results are not encouraging for this hypothesis. Only one of the results, Abstract v. Non-Abstract, has scores which would lead us to have a reasonable degree of confidence that we are seeing a difference worth reporting. The result certainly looks significant in statistical terms, but in descriptive terms it is not significant – the r² score suggests that what we are looking at along this dimension accounts for only 3.7% of variation between the spoken and written texts. So, if we have found a difference, it is a small one. Otherwise, speech and writing in the corpus appear very similar. This is suggestive of the speech in the corpus being scripted or, perhaps in the case of interviews in newspapers, heavily edited before publication. The example below, a part of a transcription of an interview with Ayman Zawahiri, is a good example of this:

INTERVIEWER: We are happy to interview you four years after the New York and Washington raids [of 9/11].

DR. AYMAN AL-ZAWAHIRI: I, too, am happy to address our Muslim umma through you during this critical stage of its history. I would like to take this opportunity to thank you, and pray that Allah Almighty will reward you for publicizing the word of truth in the midst of the Crusader campaign and global war that is being waged against Islam and the Muslims.

³ In the results column of this table, and each table following, F is a measure of variation between means for values taken from the texts analysed, the higher the value, the more scattered from the mean the actual values analysed are. The significance value p gives the confidence we may have of the result not being the product of chance (where a value of 0.05 is at the 95% confidence level, 0.01 is at the 99% confidence level, for example) and r² is a measure that shows how much of the variation observed may be accounted for by the dimension in question.

INTERVIEWER: Dr. Ayman, how do you view the Crusader campaign, four years after it began?

DR. AYMAN AL-ZAWAHIRI: The new Crusader campaign is failing, just like the previous ones, by the grace of Allah. America and its Crusader allies have not accomplished a single thing-except for throwing their armies into the battlefield to take blows on a daily basis, to have their soldiers killed on a daily basis, and to have their economies bled on a daily basis.

What did they accomplish in Afghanistan? They evicted the Taliban government from Kabul, but it centered itself in the villages and mountains-where the real power of Afghanistan lies. Northern Afghanistan and Kabul have become a scene of chaos, pillaging, looting, defiling [women's] honor, and drug trafficking, which have flourished and thrived under the American occupation. Then they held elections, which resembled a masquerade more than anything else- since the country's periphery is controlled by highway bandits and warlords, and since the international committees monitoring the elections - or rather, those who bear false witness - could not (even if they really wanted to) cover more than ten voting districts; and since transferring the ballot boxes takes fifteen days, under the control of the warlords and highway bandits, and then under the control of the occupation forces; and since any resistance, or anything resembling resistance or opposition, is met with bombardment, missiles, the burning of villages, and the killing of hundreds.

Then, after all this, they obtained the false testimony of the U.N. [United Nations], which saw nothing for it to bear witness about-except for a few theatrics in some [voting] districts in the cities. This is but one of many examples of the U.N:s hypocrisy, which they claim to be the symbol of their international legitimacy.

Brief as this example is, it is notable in lacking any of the markers of spoken interaction that one might intuitively expect – there are no markers of hesitation, no repairs and the syntax is somewhat complex (consider the coordinated infinitival clauses in the final sentence of the example). In terms of Biber's analytical scheme, we might expect informational interaction to have positive scores on the involved vs. informational discourse dimension (for example, Nini, 2019). Yet the features we would expect for the text to be placed at the involved part of the continuum are largely absent – for example that deletion ('pray that Allah'), present tenses (this example starts with present tenses, but shifts to the past and, in the fuller text, the past tenses are dominant) and second person pronouns (the first and third person predominates in the pronouns used in this quote).

4.4. Exploring Combinations and New Dimensions

It may, of course, be the case that the way in which speech and writing is used across moderate, fringe and extreme texts is not smoothly distributed, as the previous analysis presumed. So, what would happen if we looked at the six categories created by combining speech and writing with moderate, fringe and extreme? The results of undertaking that analysis according to Biber's six dimensions are given in Table 6 below.

Table 6. An MDA of moderate, fringe and extremist speech and writing using Biber's six dimensions.

Dimension	Results
1. Involved v. Informational	F = 1; p = 0.4325968, r ² = 1.4%
2. Narrative v. Non-Narrative	F = 5; p = 0.0006705594, r ² = 6.9%
3. Elaborated v. Situation-Dependent	F = 0.8; p = 0.553918, r ² = 1.1%
4. Overt Expression of Argumentation	F = 1.5; p = 0.2010924, r ² = 2.2%
5. Abstract v. Non-Abstract	F = 2.7; p = 0.02957478, r ² = 3.9%
6. Online Informational Elaboration	F = 2.1; p = 0.0875549, r ² = 2.9%

Once again, the results seem to suggest that variation within the dataset is low. Only one of these results is statistically significant at the 99% level (Narrative v. Non-Narrative) but once again the volume of variation explained by this factor is low (6.9%), as we might expect given that we know there is a preponderance of general narrative material in all categories in our data. Abstract v. Non-Abstract is significant at the 95% level but explains only 3.9% of variation. Figure 1 shows the results for the Narrative v. Non-Narrative dimension compared with the findings of Biber (1988).

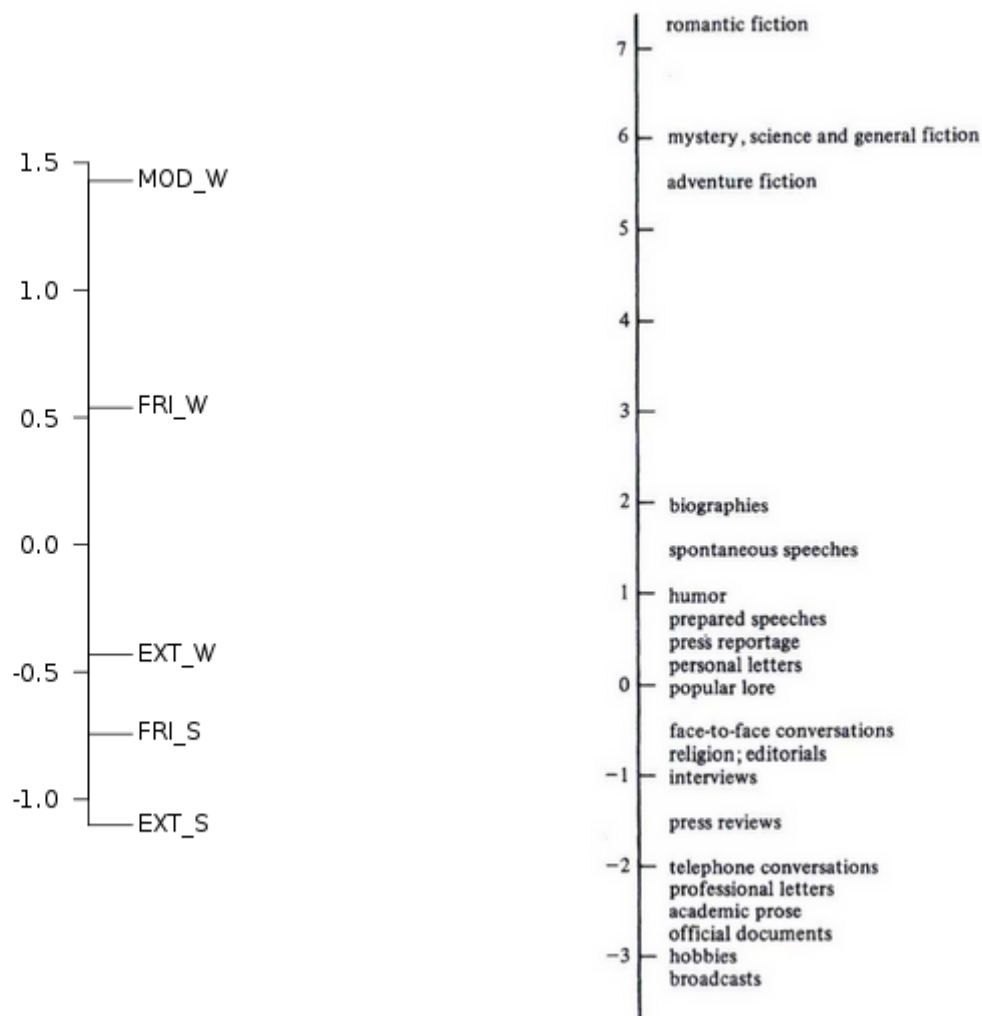


Figure 1. Moderate, fringe and extremist speech and writing on the Narrative v. Non-Narrative dimension (left) compared to Biber's findings for this dimension (right).

The dimensions are very densely grouped, relative to the scale of variation that Biber observed, and the data in our corpus displays a much lower degree of variation on both of these dimensions, clustering around 0. For Narrative, we may say that fringe and moderate writing exhibit a slight preference for this relative to extremist writing, fringe speech and extremist speech. However, relative to some of the forms of speech and writing that Biber studied, none of the data explored has as strong a preference for Narrative as Biographies or Fiction. Similarly, none of them are as non-narratively oriented as he found telephone conversations, professional letters, academic prose, official documents, writing about hobbies and broadcasts to be.

One further way in which we may seek some principled difference between the files in the data that might align with register variation would be to look at the type of texts in the corpus – might they vary linguistically from one another in some systematic way? The notes provided to us by the analysts with each file suggested a categorization scheme for the corpus texts, as did the texts themselves. The analysts would often give labels to a text such as 'statement' or 'poem' that would lead one to assume that, if used as a way of categorizing the

texts, some register variation may be visible. Similarly, the texts themselves very often self-labelled, for example as a poem or a magazine (article). Working from these notes, and based on a supporting close reading of the texts, we allocated each text to one of the categories given in Table 7. Texts were allocated to a category based on either their form (e.g. book, magazine, poetry) or dominant function. Note that the Book and Magazine categories are ones which are functionally mixed.

Table 7. A functional classification of the corpus texts.

Category	Description	Number of texts in this category
Argument	A disputation in which a proposition is argued through leading to a conclusion.	83
Book	Lengthy prose in the form of a book. This may be a composite of a number of other categories.	53
Forum	An exchange in an online forum.	3
Interview	A Q&A Interview – these appear to be transcriptions of real interviews as opposed to the use of a Q&A structure as a rhetorical device in an Argument.	11
Lecture	Transcribed lecture/sermon.	10
Magazine	Prose in the form of a magazine. This may be a composite of a number of other categories.	34
Poem	A text composed almost exclusively of a poem or poetry.	2
Statement	A statement of what the author claims to be i.) fact or ii.) an authoritative point of view. Such statements can cover narrative prose claiming to be a true account of a particular set of events.	79

The classification in Table 7 is, of course, preliminary. However, given that such features were largely derived from the work of the analysts who were familiar with the texts and that a categorization of the texts using these categories is possible, using the framework seems warranted. Our aim in undertaking the classification was to see whether such an approach to the texts was fruitful. If it was, then this could call for a closer exploration of the categories and, in particular, for a disaggregation of the texts in the Book and Magazine categories into functional units. Table 8 gives the results of the MDA of these categories.

Table 8. An MDA of the texts, categorized by function.

Dimension	Results
1. Involved v. Informational	F = 1.7; p = 0.1116541, r ² = 4.2%
2. Narrative v. Non-Narrative	F = 5.3; p = 1.168323E-5, r ² = 12.1%
3. Elaborated v. Situation-Dependent	F = 0.7; p = 0.7117613, r ² = 1.7%
4. Overt Expression of Argumentation	F = 1.9; p = 0.07288708, r ² = 4.7%
5. Abstract v. Non-Abstract	F = 3; p = 0.004809807, r ² = 7.3%
6. Online Informational Elaboration	F = 3.1; p = 0.003701417, r ² = 7.5%

At first glance, these results seem much more promising – the last three dimensions seem to have significant results which explain between 4.7 and 7.5% of the variation observed. However, as Figure 2 shows, the poetry (Figures 2, 3 and 4) and forum discussions (Figures 3 and 4) seem to account for these results. Given the very low counts of poems and forum discussions in the corpus, it would be hard to conclude that these are useful results – they are certainly not very helpful in terms of characterizing the dataset overall given that there are only five texts in the corpus in these categories.

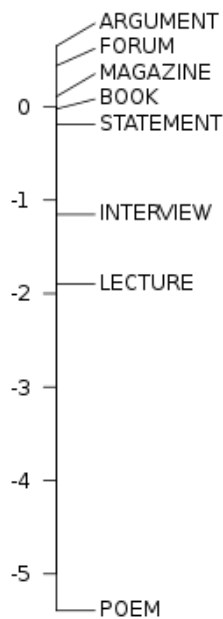


Figure 2. Dimension 4, text types only.

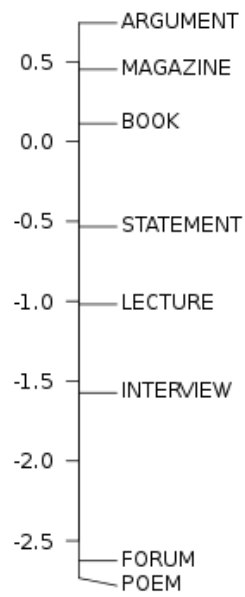


Figure 3. Dimension 5, text types only.

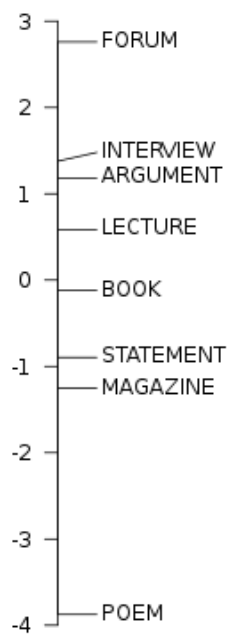


Figure 4. Dimension 6, text types only.

Might it be that by combining the mode of production categorization with the text category we may be able to gain some insight into variation in the texts that has eluded our analysis so far? The results of the MDA this produces are given in Table 9.

Table 9. An MDA of text types categorized by mode of production

Dimension	Results
1. Involved v. Informational	F = 2.1; p = 0.0295375, r ² = 6.7%
2. Narrative v. Non-Narrative	F = 4.5; p = 1.444902E-5, r ² = 13.3%
3. Elaborated v. Situation-Dependent	F = 0.8; p = 0.6315774, r ² = 2.6%
4. Overt Expression of Argumentation	F = 1.5; p = 0.1631546, r ² = 4.7%
5. Abstract v. Non-Abstract	F = 2.6; p = 0.0075517, r ² = 8%
6. Online Informational Elaboration	F = 3.2; p = 0.001162657, r ² = 9.7%

Once again, Abstract v. Non-Abstract and Online Informational Elaboration seem to give results worthy of investigation, this time with higher r² values than observed previously. However, as Figures 5 and 6 show, data sparsity once again accounts for the results. Poetry accounts for the outlier in both figures, while Arguments produced in speech represent another outlier. However, the corpus has only two texts in this category.

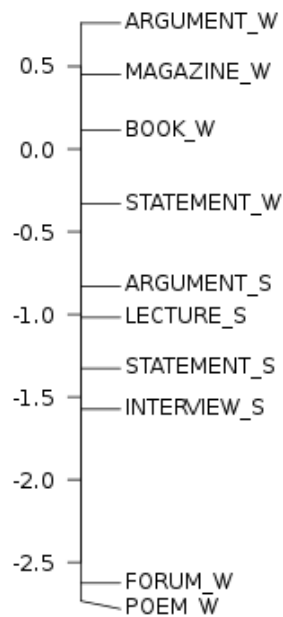


Figure 5. Dimension 5, text category plus mode combined.

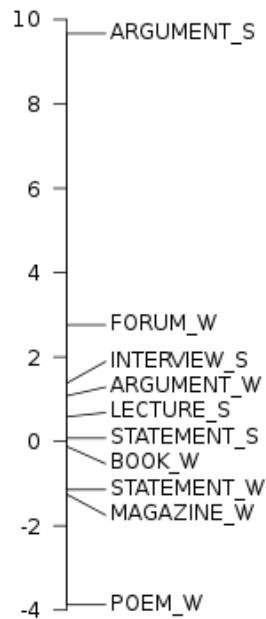


Figure 6. Dimension 6, text category plus mode combined.

Of course, the possibility exists that Biber's (1988) dimensions are not the ones that would best characterise the differences in this dataset. Might it be that, within the dataset we are observing, the values identified by Biber configure in a novel way meaning that we need to build dimensions of variation, afresh, from the bottom up? To explore this, we started with a scree plot of Eigen values of principal factors (for details, see Brezina (2018)) as a way of determining the number of clusters we might search for (see Figure 7). We then calculated 7 factors on the basis of this analysis.

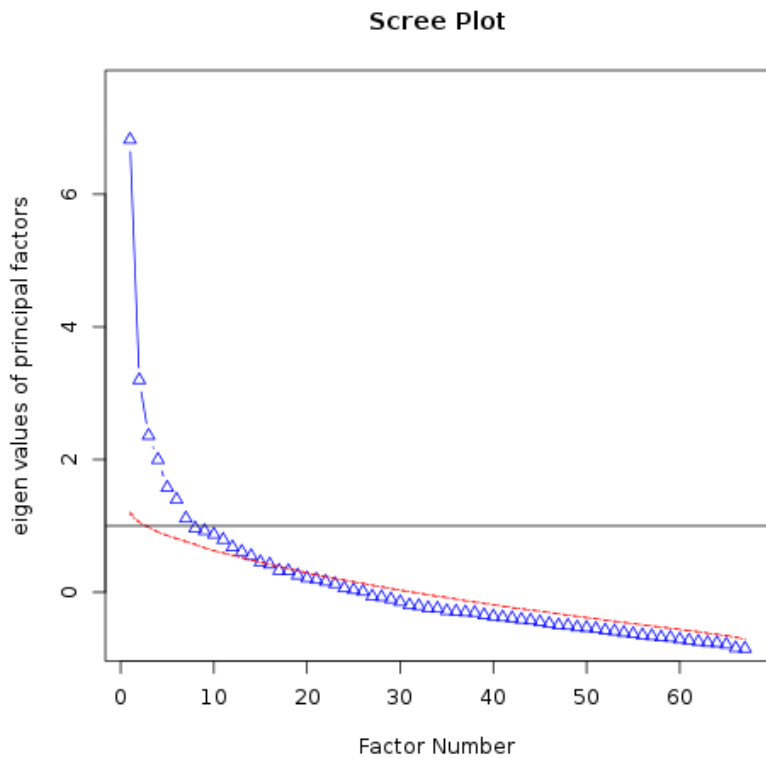


Figure 7. A plot of Eigen values of principal factors comparing the spoken and written sections of the corpus.

Once again, however, the factors do not reveal a substantial difference between the speech and writing, as Table 10 shows. None of the results are statistically significant, even at the 95% level, and the values of r^2 are very low indeed.

Table 10. A seven factor MDA analysis

Dimension	Results
1	$F(df = 1, 273) = 1.07; p = 0.3008447, r^2 = 0.39\%$
2	$F(df = 1, 273) = 1.45; p = 0.2289638, r^2 = 0.53\%$
3	$F(df = 1, 273) = 0.77; p = 0.3805662, r^2 = 0.28\%$
4	$F(df = 1, 273) = 5.34; p = 0.02154485, r^2 = 1.92\%$
5	$F(df = 1, 273) = 2.38; p = 0.1240777, r^2 = 0.86\%$
6	$F(df = 1, 273) = 0.04; p = 0.8472702, r^2 = 0.01\%$
7	$F(df = 1, 273) = 0.67; p = 0.4136613, r^2 = 0.24\%$

5. Conclusions

Our MDA strongly suggests the following about our corpus of violent Jihadist texts. Firstly, at least as measured by the features underlying Biber’s original study, the texts in our corpus are generally quite homogenous linguistically. When factors are combined and the texts are looked at from different analytical perspectives, they tend to remain stubbornly grouped. Secondly, there are limits on how far the perspectives on the texts can be pushed. In

principle, given the categorizations that we have for the texts, we could combine mode of production (2 categories) with analysts' ratings (3 categories) and text types (eight categories) to produce 48 categories to which we could assign our texts in order to explore variation. However, as has already been shown, the text collection does not support this analysis. Even when we combine mode and analyst rating, we produce an empty category – there is no moderate spoken material in the dataset. Even where there are members of a category, these may be few and they may lead to results which are unhelpful and do not characterise the data in general as we saw with Forum, Spoken Arguments, Discussion and Poems. Finally, however, we should be mindful of the useful, though admittedly weak, signals that the analyses have produced: i.) the nature of the texts in the corpus is generally narrative; ii.) the type of narrative may vary by analysts' category; iii.) narrative v. non-narrative may be worth exploring on the basis of the results presented in Tables 4, 6, 8; and iv.) abstract v. non-abstract may be worth investigating based on the results presented in tables 4, 6, 8 and 9. These all represent potentially interesting avenues for future study.

The results of the MDA investigations give us confidence that our corpus represents a linguistically coherent body of texts in the sense that the texts seem to be similar. Not only were they based on a similar topic and produced and read by people with similar goals and who, in all likelihood, shared a similar world-view, they are also linguistically similar to the extent that we can say that the corpus is composed of what appears to be a single register of what we might call Islamic discourse which arches across the categories of moderate, fringe and extreme. On the face of it, this lack of variation could be viewed as something of a non-result. However, the insights into the homogeneous nature of these texts have implications for researchers interested in studying such texts in the future. For such researchers, this finding suggests that significant linguistic variation identified within sets of extremist texts similar to our own is unlikely to result from differences at the level of register. Indeed, our findings provide a warrant for regarding such texts as a single register for analytical purposes. If variation is to be sought, it is more likely to reside in factors other than those studied here, for example, the time period or variety of English in which the texts were composed.

On a methodological note, our findings also suggest the utility of MDA for studying similarity across texts. By and large, MDA has tended to be applied in previous research to identify variation within and across datasets, with emergent differences often forming the focus of the study and constituting headline findings. However, the ability of MDA to reliably reveal similarities, as highlighted by this study, points to an application of the method in studies seeking to study overlaps in, *inter alia*, certain ideologies, values, and world-views that are propagated by the language used in and across sets of texts. Indeed, what binds the homogeneous set of texts in our data is just this: they index a register which seems to be related to Islam. Such an approach could be applied in corpus-assisted discourse studies (see: Baker et al. 2008) as an entry point for grouping texts according to their ideological stances and for down-sampling texts and linguistic features for more qualitative analysis of the linguistic manifestations of those ideologies.

A limitation of our findings, suggested above, pertains to the issue of representativeness. As noted previously, the corpus analysed in this chapter was not assembled to provide a general assessment of writing about Islam. Our texts were associated with a subset of Muslims who engaged in terrorist acts. A full assessment of the

representativeness of a corpus depends on a prior full definition of the ‘population’ that the sample comprising the corpus is designed to represent (Biber 1993). The types of texts that we have worked with here thus present a challenge on this front, as the wider ‘population’ of texts about Islam is both ill-defined and unknown. This makes a full assessment of the representativeness of this corpus in terms of general texts about Islam impossible. As such, we cannot be sure of the extent to which the texts of which we were availed represent the full set of texts about Islam ‘out there’ in the world. That notwithstanding, the general homogeneity of the texts that we have analysed could provide some evidence that this is a relatively restricted linguistic register, and provides a starting point for future research of such texts. This point is important in two respects. Firstly, it is useful to advise people where not to look for difference – approaching these texts to look at register differences, based upon the Biber features, is not productive. The differences between the texts is not rooted in register. However, this brings us to a second way in which these results are useful – they indicate the importance of the register used to those who produced the texts in our corpus. This point becomes more important when we consider other work which we have undertaken where we have found more differences between the texts. For example, in Brookes and McEnery (2020) our argument is that many of the changes that are observed are attempts at challenging the doxa of Islam – they are part of an ideological struggle for the meaning of the religion. An important feature of that struggle is what is *not* fought over – the register in which the arguments are presented. This constitutes an important form of symbolic capital (Bourdieu, 1989) which is indispensable to all of the authors in the corpus. In some ways, therefore, we might argue that the inflexibility of the register used in the corpus examined in this chapter is a signal of the importance that is attached to this register as a form of symbolic capital that allows the arguments presented by writers in the Fringe and Extreme categories to be accepted. In this sense, the findings presented here are critical as they reveal the importance of adherence to register even by those wishing otherwise to subvert and revise belief.

References

- Anon (2013). Restricted Report. Home Office, UK.
- Baker, P. and Vessey, R. (2018). A corpus-driven comparison of English and French Islamist extremist texts. *International Journal of Corpus Linguistics*, 23(3), 255–278.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T. and Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306.
- Baker, P., Gabrielatos, C. and McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Baker, P., Vessey, R. and McEnery, T. (Forthcoming). *The Language of Violent Jihad*. Cambridge: Cambridge University Press.
- Berber Sardinha, T. (2018). Dimensions of variation across Internet registers. *International Journal of Corpus Linguistics*, 23(2), 125–157.

- Bértoli-Dutra, P. (2014). 'Multi-Dimensional analysis of pop songs'. In: T. Berber Sardinha and M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis, 25 years on: A tribute to Douglas Biber*. Amsterdam: John Benjamins, pp. 149–176.
- Biber, D. (1984). A model of textual relations within the written and spoken modes. Ph.D. Dissertation, University of Southern California.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62, 384–414.
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Biber, D. (2019). 'Multi-dimensional analysis: a historical synopsis'. In: T. Berber Sardinha and M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues*. London: Bloomsbury, pp. 11–26.
- Biber, D. and Conrad, S. (2009). *Register, Genre and Style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S. and Reppen, R. (1996). Corpus-based investigations of language use. *Annual Review of Applied Linguistics*, 16, 115–136.
- Bourdieu, P. (1989). Social space and symbolic power. *Sociological Theory*, 7(1), 14–25
- Brezina, V. (2018). *Statistics in Corpus Linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brindle, A. (2017). *The Language of Hate: A Corpus Linguistic Analysis of White Supremacist Language*. London and New York: Routledge.
- Brookes, G. and McEnery, T. (2020). Correlation, collocation and cohesion: A corpus-based critical analysis of violent jihadist discourse. *Discourse & Society*, 31(4), 351–373.
- Clarke, I. and Grieve, J. (2019). Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018. PLOS ONE, Online. Available here: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0222062>.
- Conoscenti, M. (2016). 'ISIS' Dabiqcommunicative strategies, NATO and Europe. Who is learning from whom?' In M. Ceretta and B. Curli (Eds.), *Discourses and Counter discourses on Europe: From the Enlightenment to the EU*. London and New York: Routledge, pp. 215–244.
- Conrad, S., and Biber, D. (2001). 'Multidimensional methodology and the dimensions of register variation in English'. In: S. Conrad and D. Biber, (Eds.), *Variation in English: Multidimensional studies*. Pearson Education: Harlow, pp. 13–42.
- Droogan, J. and Peattie, S. (2018). Reading jihad: Mapping the shifting themes of Inspire magazine. *Terrorism and Political Violence*, 30(4), 684–717.
- Esposito, J. L. (2014). *The Oxford Dictionary of Islam*. Oxford: Oxford University Press.
- Friginal, E. (2008). *The Language of Outsourced Call Centers: A corpus-based study of cross-cultural interaction*. Amsterdam: John Benjamins.

- Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, 17(3), 380–409.
- HM Government. (2013). Tackling extremism in the UK. London: Cabinet Office. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/263181/ETF_FINAL.pdf (last accessed August 2018).
- HM Government (2015) *Counter-Extremism Strategy*. London: Cabinet Office. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/470088/51859_Cm9148_Accessible.pdf.
- Holbrook, D. (2015). Designing and applying an extremist media index. *Perspectives on Terrorism*, 9(5). Online. URL: <http://www.terrorismanalysts.com/pt/index.php/pot/article/view/461/html>.
- Ingram, H. J. (2017). An Analysis of Inspire and Dabiq: Lessons from AQAP and Islamic State's Propaganda War. *Studies in Conflict & Terrorism*, 40(5), 357–375.
- Kramer, M. (2003). Coming to terms, Fundamentalists or Islamists? *Middle East Quarterly*, Spring 2003, 65–77.
- Leech, G. (1992). 'Corpora and theories of linguistic performance'. In: J. Svartvik (ed.), *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991*. Berlin: Walter de Gruyter, pp. 105–122.
- McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Nini, A. (2019). 'The Multi-Dimensional Analysis Tagger'. In: T. Berber Sardinha and M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research Methods and Current Issues*. London and New York: Bloomsbury, pp. 67–94.
- Prentice, S., Taylor, P., Rayson, P., Hoskins, A. and O'Loughlin, B. (2011). Analyzing the semantic content and persuasive composition of extremist media: A case study of texts produced during the Gaza conflict. *Information Systems Frontiers*, 13(1), 61–73.
- Prentice, S., Taylor, P., Rayson, R. and Giebels, E. (2012). Differentiating act from ideology: Evidence from messages for and against violent extremism. *Negotiation and Conflict Management Research*, 5(3), 289–306.
- Quaglio, P. (2009). *Television Dialogue: The sitcom Friends vs. natural conversation*. Amsterdam: John Benjamins.
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), 519–549.
- Richardson, J. E. (2017). *British Fascism: A Discourse-Historic Analysis*. Hannover: ibidem Verlag.
- Simpson, P. (1993). *Language, Ideology and Point of View*. London and New York: Routledge.
- Staples, S. (2015). *The discourse of nurse-patient interactions: Contrasting the communicative styles of U.S. and international nurses*. Amsterdam: John Benjamins.
- Wignell, P., Tan, S. and O'Halloran, K. L. (2017). Under the shade of AK47s: a multimodal approach to violent extremist recruitment strategies for foreign fighters. *Critical Studies on Terrorism*, 10(3), 429–452.
- Zaeef, A.S. (2010). *My Life with the Taliban*. New York: Columbia University Press.