

Correlation, collocation and cohesion: A corpus-based critical analysis of violent jihadist discourse

Gavin Brookes and Tony McEnery
Lancaster University

Abstract

This paper explores the language of violent jihad, focussing in upon lexis encoding concepts from Islam. Through the use of correlation statistics, the paper demonstrates that the words encoding such concepts distribute in dependent relationships across different types of texts. The correlation between the words cannot be simply explained in terms of collocation, rather the correlation is evidence of other forms of cohesion at work in the texts. The variation in patterns of cohesion across a spectrum of texts from those advocating violence to those which do not promote violence demonstrates how these concepts are contested and redefined by violent jihadists and the role that collocation and other forms of cohesion can play in the process. The paper concludes that the terms, and their redefinition, are a key part of the symbolic capital used by groups to create identities which licence violence.

1. Introduction

Following Firth's (1957: 6) dictum that 'you shall know a word by the company it keeps', (critical) discourse analysts employing corpus approaches have sought to examine the representations of, and discourses surrounding, particular words and concepts by analysing their collocates. Recent examples include Baker et al. (2013) who utilised collocation as an entry point for examining the representation of Muslims in the British press, Wright and Brookes (2018) who used the technique to identify groups depicted as not being able to speak English in right-leaning newspapers, and Taylor (2018) who used collocation to explore changes in the metaphorical representations of the Windrush Generation in UK public discourse. In this paper, we build on such collocation-based approaches to discourse by incorporating consideration of correlation in the (critical) study of discourse. Correlation is a technique which measures whether two variables are related by considering the extent to which they tend to occur within the same text; 'we are looking at whether, if one variable increases, the other variable increases, decreases or stays the same' (Brezina, 2018: 141). Using a corpus of violent jihadist¹ texts, in this article we consider the extent to which a series of researcher-identified terms correlate with one another across these texts, with observed correlational relationships interpreted in terms of ideology and textual cohesion.

For the purpose of this study, 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-

¹ Conceptualisations of the relationship between extremism and Islam are, as Baker and Vessey (2018: 257) point out, both 'complex and contested'. In this article we will refer to violent jihadism when discussing the motivations of the people from whom the texts discussed here were acquired. We acknowledge that this term has the potential to imply a fraught distinction between legitimate and illegitimate violence (see Lukin, 2019) for a discussion of the specialisation of violence). It is not our intention to perpetuate such a distinction, rather we decided to use this term as it is the best available (to our knowledge). We should also 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 extremist rhetoric.

granted assumptions, beliefs and value systems which are shared collectively by social groups' (Simpson, 1993: 5). Ideologies offer particular representations not only of how things in the world are but also of how they ought to be (from the perspective of a text creator, organisation or other social group) (Fairclough, 2003). From this view, Language can be viewed as ideological when text creators' lexical and grammatical choices serve to promote a particular perspective. Hart (2014: 4-5) points out that '[d]iscourses are inherently ideological in so far as Discourses in the same domain can exist in competition with one another'. He elaborates, '[w]hich competing Discourses establish themselves as dominant depends on various factors including the discourse of powerful actors [...]. On this account, discourse, and the discourse of powerful speakers/ institutions in particular, represents a site for the (re)articulation of ideology and the legitimation of (sometimes harmful) social action' (ibid.; see also: Fairclough, 2005). In this paper, we are interested in how linguistic choices in the jihadist texts under analysis support particular ideologies and serve to, amongst other things, legitimise harmful social action in the form of jihadist violence.

Ideology is central to the activities and associated discourses of terrorist groups. As Malešević observes, as well as requiring power at an organisational level, terrorism 'entails the presence of a relatively distinct set of normative principles. Precisely because terrorist activities are generally regarded as being illegitimate or senseless, terrorist organisations devote a great deal of attention to ideology' (2017: 269), whereby 'distinct ideological narratives [can be used to] justify the use of violence' but also to enact 'a structural change whereby such ideas and practices take root among wider sectors of population' (ibid: 270). 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 been convicted of committing terrorist crimes in pursuit of the ideologies propagated by the texts. The role of ideology in jihadism has been written about at length by, amongst others, Hellmich (2008), Mitchell (2008), Lohlker and Abu-Hamdeh (2013) and Manne (2017). Such treatise have tended to adopt critical perspectives on jihadist ideology, for example by pointing out its inconsistencies and hypocrisies and by elucidating the ways in which it undermines and contradicts the teachings of Islam. We will return at the end of this section to linguistic and discourse-based studies of jihadist ideology. For now, we turn our attention to lexical cohesion.

In addition to the ideologies of which they are constitutive and by which they are constituted, we also interpret the correlational relationships in these texts as contributing to their cohesion. Cohesion can be defined as the 'set of resources for constructing relations in discourse that transcend grammatical structure' (Martin, 2015: 61). This sense of studying relationships within (and across) texts beyond merely their grammatical parts is also captured by Mahlberg (2006): 363, who defines textual cohesion as 'the property of connectedness that characterises a text in contrast to a mere sequence of words' (Mahlberg, 2006: 363). Cohesion is a by now well-established concept in linguistics and (Critical) Discourse Analysis, with major interest in it initiated by Halliday and Hasan's work in Systemic Functional Linguistics (SFL) (Halliday, 1964, 1973; Hasan, 1968; Halliday and Hasan, 1976, 1995). Halliday (1973) modelled cohesion according to non-structural relations between words above the level of the sentence, under the rubric of the 'textual metafunction' (1973: 141). In their account, Halliday and Hasan (1976) organised their inventory of cohesive

resources in terms reference, ellipsis, substitution, conjunction and lexical cohesion (this framework was elaborated on by Gutwinski (1976) who, influenced by Gleason (1968), introduced the concept of grammatical parallelism) (see Martin (2015) for a useful discussion).

The concept of cohesion, particularly as defined by and studied approached within the work of Halliday and Hasan, has a long-lasting legacy in linguistic and discourse analytical research and can, for instance, be seen in more recent work in the SFL tradition (e.g. Martin 2010, 2012) (for recent reviews, see: Halliday and Webster, 2009; Hasan et al. 2005, 2007). Cohesion has also received relatively recent attention from corpus linguists. In terms of collocation specifically, in their aforementioned account Halliday and Hasan (1976) only touch briefly on the role of collocation in cohesion, though more detailed, persuasive accounts of the utility of collocation for exploring cohesion at the lexical, grammatical and semantic levels have subsequently been provided, amongst others, Kjellmer (1991), Sinclair (1996) and Stubbs (2015). Of particular relevance to this study, Morley (2009) demonstrates how an analysis based on this approach can help to account for the ways in which lexical cohesion contributes towards what he refers to as ‘rhetorical movement of the discourse’ (2009: 5). The approach set out in this study elaborates on such collocation-based approaches by introducing consideration of correlation. Yet, at the same time, this study also contributes to the study of cohesion more generally; writing in 2015, JR Martin called for research on cohesion to address its relation to not only linguistic but also social modules in order to enhance our understanding of the ways in which ‘the social motivates patterns of cohesion’ (see also: Bednarek and Martin, 2010; Martin and Wodak, 2003). By studying the role of cohesion within and across texts written by different authors designed to persuade and unite individuals around particular sets of ideologies, the study reported in this article contributes towards the need for cohesion to be situated within the social.

This is not the first linguistic study concerned with the discourse and ideologies of violent jihadism, as a large body of research has explored these topics. To offer a couple of recent examples, Holbrook (2014) explored the framing and evolution of the public discourse of Al-Qaeda based on analysis of 250 statements made by the organisation’s two leaders, Ayman Al-Zawahiri and Osama Bin Laden, over a period of twenty years. This study focused in particular on the ways in which the leaders ‘diagnosed’ problems, offered solutions, and advocated escalating violence. The analysis also showed how the leaders developed a more critical stance towards Islam following the 9/11 attacks and tailored messages for particular audiences, revealing tensions between the respective leaders’ discourse. More recently, Al-Rikaby and Mahadi (2018) examined the discursive strategies in so-called ‘calls to jihad’ by Al-Zawahiri in 2006 and the former leader of ISIS, Abu Bakr Al-Baghdadi, in 2015. Employing various concepts from Critical Discourse Analysis, this study showed how both leaders focused on in-group/out-group representations and the impacts that these have had on Muslim societies (for further examples of studies of jihadist ideological discourse, see Droogan and Peattie, 2016; Wignell et al., 2017; Ingram, 2017; Spier, 2018). A small body of work in this area has also employed more quantitative, including corpus linguistic, methods. For example, Prentice et al. (2011, 2012) utilised the *Wmatrix* tool (Rayson 2008) to analyse the persuasive strategies emergent from texts inciting violence in the context of the Gaza conflict. Conoscenti (2016) adopted a collocation-based perspective in an analysis of

the communicative strategies of *Dabiq*, while Baker and Vessey (2018) combined corpus methods with more qualitative discourse analysis in their study of the discursive themes and linguistic strategies employed in the English-speaking *Inspire* and *Dabiq* magazines and in ISIS's French-speaking *Dar al Islam*. However, to our knowledge the present study is the first to examine the contribution of textual cohesion, explored through the perspectives of collocation and correlation, to violent jihadist discourse.

The remainder of this article is divided into four sections. Following this introduction, Section 2 outlines our data and analytical approach, before we present our analysis of the correlations in our corpus in Section 3. In Section 4, we present a discussion in which we interpret the patterns of textual cohesion and the rhetorical strategies identified in terms of the symbolic capital that is invested in them. Finally, Section 5 concludes the article by considering the implications of our findings and methodological approach for others applying corpus methods to critical discourse studies.

2. Data and approach

The texts analysed in this paper derive from an ongoing project examining jihadist discourse (see: Baker et al. 2020). The texts have all been associated with people convicted in the UK of terrorist offences and have been coded, through close reading, by experts (see Holbrook, 2019), into three categories within an 'Extremist Media Index' (Holbrook, 2015): 'Moderate', 'Fringe' or 'Extreme' (see Table 1).

Table 1. Corpus composition, by ideological category.

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

All texts in the corpus are about Islam. The Extreme sub-corpus contains jihadist texts which provide an 'Endorsement/glorification of violence in contemporary context and/or stark dehumanization' (Holbrook, 2015:60). Fringe texts are 'religiously or ideologically conservative and isolationist, politically radical and confrontational, but without any justifications conveyed for violence' (Holbrook, 2015:66). Moderate texts are in no way linked to an encouragement to violence, isolationism or radical beliefs. The texts were written mostly in English, reflecting the first language of the people who were convicted and who had thus sourced and read the materials, although they also contained some elements of code-switching to Arabic, for example when quoting from the hadith or when referencing concepts from Islam. The categorisations in Table 1 help us to frame the texts in terms of ideology, given that central to each text in the corpus is a perceived attempt, on behalf of the author, to propagate a particular worldview.

Returning to the categorisations in Table 1, although this grouping of the texts was not carried out by linguists, and as such was based on non-linguistic categorisations of ideology, we are of the view that language is nevertheless likely to be significant to the ways in which those ideologies are expressed (Simpson, 1993: 5). 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 and which forms of language might be associated with each of them.

Given our access to expert coding of the texts in our corpus, the starting point for our analysis is that subject expertise. We had access to a report (Anon, 2013) which identified a series of key-concepts, which were often directly lexical, that experts in the study of terrorist communications, and key members of the UK Muslim community, believed were important in allowing the analysts to place the texts into the different categories. After using Anon (2013), close reading and their subject expertise to classify our corpus texts, analysts also produced a document containing a set of terms that *they* thought were salient in the texts to the degree that they helped guide their classifications. Also, they noted some spelling variants of these. We used these terms to explore the corpus to find further variations in the use of these terms through concordancing (i.e. close reading of extended samples of their use in context). This raised the possibility that terminology specific to Islam may be a good way to approach the data through a linguistic analysis. Figure 1 shows the canonical form of terms compiled through this approach, though our analysis covered both these and any spelling variants of them.

ahadeeth, ahadith, akhira, anasheed, aya, ayah, ayn, bara, batil, bayah, bidah, caliphate, dar al harb, dawa, deen, dunya, fiqh, fitna, fitnah, hadith, ijihad, ilm, jahanaam, jahiliyyah, jannah, jihad, jihaad, jizya, jizyah, khutba, kifaya, kufaar, kufar, kuffar, kuffaar, kuffars, kuffarun, kufir, munafiq, mushrikeen, mushrikun, nasheed, qadr, qital, riddah, sahabah, shahada, shahadah, shaheed, shahid, shariah, shirk, sira, sunna, sunnah, taaghoot, tafsir, taghoot, taghut, takfir, taqleed, taqlid, taqwa, tawhid, ummah

Figure 1. Terms noted by the text analysts as being important in categorizing texts.²

These terms have a combined frequency of 15,106 in the whole corpus, occurring in 163 of the 174 Extreme texts with a frequency of 461.7 per hundred thousand words (per 100,000 words – hereafter referred to as PHTW), 45 of the 54 Fringe texts (440.5 PHTW) and 42 of the 51 moderate texts (145.07 PHTW). While these terms do not deal with all of the issues identified by Anon (2013) as distinguishing the texts apart, they do at least give a fuller account of the concepts from Islam which might contribute to this, so we used these terms as the basis of our correlation analysis.

As a reminder, correlation analysis measures the extent to which two variables tend to occur within the same text. When an increase in the values of the first variable also means an increase in the values of the second variable, we can say that the two items are positively correlated. On the other hand, if the first variable increases as the second decreases, then the two items are negatively correlated. If the frequencies of the two variables appear to have no bearing on one another (i.e. an increase in the first variable does not seem to militate in favour of either an increase or decrease in the second variable), then we can conclude that

² While we do not have space to define all of the terms in this figure, we direct readers seeking definitions to the *Glossary of Islam* compiled by Sourdel and Sourdel-Thomine (2007).

there is virtually no correlation between these items and thus that there is not a statistically meaningful relationship between them (Brezina, 2018: 141-142).

We begin this study by using correlation to establish any relationships amongst the words introduced in Figure 1. Variant forms of each word were considered separately. As we were dealing with a scale linguistic variable (word frequency), we used a parametric test of correlation, Spearman’s Correlation (Brezina, 2018: 142). The distribution, by file, of each word in each category was calculated. The result of the correlation is an effect size, *r*. This is conventionally split into four broad bands (after Cohen, 1988: 79-80): 0 (no effect), +/- 0.1 to +/- 0.29 (small effect), +/- 0.3 to +/- 0.49 (medium effect) and greater than +/- 0.5 (large effect). Secondly, we consider the actual interaction of correlating words, including through collocation. Following this, we consider the rhetorical functions of the correlational and collocational pairings by assigning them to semantic categories based on close reading of examples in context. As part of this step in our analysis, we also consider self-correlation of semantic categories; that is, cases whereby words belonging to the same semantic category have a proclivity to occur alongside each other within texts. Finally, we consider semantic categories which do *not* self-correlate, but which tend to correlate with other categories.

To understand the rhetorical purposes of any correlations found, we also devise a scheme of rhetorical goals that the correlations highlighted in our analysis serve. This is based on close analysis of the ways in which the relevant words are used in their original contexts – a perspective facilitated by concordance. The structure of the next section, in which we report our findings, reflects the analytical sequence mapped out above.

3. Analysis

3.1. Correlation

Our analysis began by looking, in each category of texts, at the words in Figure 1 and ascertained, through correlation, whether any of them co-occurred with each other within the same text. Tables 2 and 3 show how many pairs of words correlate with one another with a medium or large effect. For the sake of brevity, we do not list the correlations here, but we will introduce and discuss specific correlations in the course of the discussion.

While these tables show the scale of co-occurrence of these terms, they do not directly demonstrate that any given correlating pair is directly associated within the text (in other words, the correlation of A and B is not in itself direct evidence that A is being used, for example, to justify B in the text). What we seek through this test are indications of what we might look for in a (critical) discourse analysis of these texts.

Table 2. Terms correlated with a medium effect.

Present in Categories	Count
Extreme Only	51
Fringe Only	71
Moderate Only	26

Extreme and Fringe	3
Extreme and Moderate	4
Fringe and Moderate	4
All Three	0

Table 3. Terms correlated with a large effect.

Present in Categories	Count
Extreme Only	19
Fringe Only	112
Moderate Only	45
Extreme and Fringe	1
Extreme and Moderate	1
Fringe and Moderate	6
All Three	0

Some observations can be made on the basis of these tables. Firstly, there is a positive correlation between some of these concepts but there are no negative moderate or large correlations found in the corpus at all. This indicates that there is a degree of connectedness between the concepts which is what we should expect from terms arising from a belief system, i.e. they are deployed in concert with rhetorical purpose.

Secondly, the different text types exhibit interesting similarities and differences, with the accent on difference. In the moderate effect part of the correlation scale, no correlated pair is shared across all three text categories (i.e. Moderate, Fringe and Extreme), three correlated pairs are shared across the Fringe and Extreme categories, four across Moderate and Extreme, and four across Moderate and Fringe. The remaining 149 of the total 159 correlated pairs occur in one text category only. In the large effect part of the correlation scale, no correlated pairs occur in all three text categories, one correlated pair is shared across the Fringe and Extreme categories, one across Moderate and Extreme, and six across Moderate and Fringe. The remaining 176 of the total 184 correlated pairs occur in one text category only.

This evidence suggests that the three types of texts draw upon a repertoire of concepts from the belief system that they relate to, and that they configure those in ways which do not always lead to sharp disjunctions between them, at least in terms of lexis. With that said, it should also be noted that the shared correlations drawn upon do represent a clear minority of the correlations both in the medium and large effect part of the correlation scale. Also, we have yet to consider frequency – it may be that large frequency effects minimise or amplify the importance of the correlations observed. There is an obvious intersection between correlation and frequency – in its simplest form, words which do not occur cannot correlate. This serves as a helpful reminder that in each text not every word that was explored in this analysis engaged in a correlation, as Table 4 shows.

Table 4. Terms correlated with a small or no effect. Results for all forms, including variant forms, shown.

Category	Words with no or low correlation	Words not engaging in correlation because they have a frequency of 0 in the category
Moderate	<i>dawa, shahid, takfir</i>	<i>anasheed, aya, bara, bayah, bidah, dar-al-harb, fitna, jahanaam, jizyah, nasheed, shaheed, taqlid</i>
Fringe	<i>deen, shahada</i>	<i>ahadeeth, anasheed, ayn, dawa, qital, shahid, taqleed</i>
Extreme	<i>anasheed, ayn, batil, bidah, dawa</i>	<i>aya, jahanaam, khutba, kifaya, sira</i>

This view on correlation is interesting as it shows concepts which are not drawn on by some categories, while also showing words which, when they are drawn upon, do not correlate notably with other terms. The decision not to normalise the terms explored at this point to standard orthographic forms also reveals some interesting patterns of affinity – consider *jizya* and *jizyah*. Moderate texts barely mention the concept, but when they do, they use the form *jizya* (seven occurrences) and never *jizyah*. Fringe texts use both *jizya* and *jizyah* equally (each word occurs thirteen times in this category). Extreme texts, however, strongly favour *jizyah* (106 occurrences) over *jizya* (22 occurrences). Orthography clearly matters in these texts.

Overall a picture emerges once again of terms being i.) selected within the categories; ii.) brought together for use, and iii.) avoided altogether in some cases. Also, some words more actively correlate with others; the word *qadr* in the large effect part of the correlation scale in the Fringe texts engages in a large number of correlations, being involved in 10 of the 112 correlations in this category. By contrast, *jizyah* correlates with only one term. In analysing a set of texts and terms such as these, it is clearly the case that the words should not be studied in isolation – they are linked to one another within a belief system. This suggests that an approach where we look at words in connection with one another, and where we explore words which seem salient in the texts, may be fruitful.

3.2. Interaction

What is the nature of the interaction of the correlating words? To explore whether the correlations are simply collocations, the collocates of each term were calculated in each category. The statistic used was the Dice co-efficient (see Gablasova et al., 2017), with a collocational span of 5 words either side of the node (-5 to +5). Searches were performed using #LancsBox (Brezina et al., 2018).

This shows that the correlations are generally not collocations, with there being only one exception in each category: *kufr/shirk* (Extreme), *kufr/shirk* (Moderate) and *fiqh/hadith* (Moderate). It is striking that *kufr/shirk* is a pair which a.) correlates in both Extreme and Fringe with a large effect and b.) is composed of words which collocate with each other, suggesting that this pair is quite distinct in its behaviour and markedly so. Another observation arising from the exploration of the collocates was even more striking. Many of the terms collocate with themselves; an indication that the words exhibit what might be called ‘burstiness’ (Church and Gale, 1995) – once they have been used they are likely to be used

again but in this case in close proximity to the original mention. This is apparent in 5 words in the Moderate texts (*jihad, kufr, shirk, sunnah, taqleed*), 9 in the Fringe texts (*deen, hadith, jahiliyyah, jihad, kuffar, kufr, sunnah, tafsir, ummah*) and 10 in Extreme texts (*ayah, fiqh, fitnah, hadith, jihaad, jihad, kufr, shirk, tawhid, ummah*). In all three categories, *jihad* and *kufr* self-collocate. The behaviour of the words is of note for a number of reasons. Firstly, not all words examined exhibit this behaviour – in each of the categories only a small fraction self-located. Secondly, self-collocation is not simply a function of frequency. Some frequent words do not self-collocate (e.g. *sunnah* which occurs 407 times in in the Extreme category) while some infrequent words do (e.g. *tafsir* which occurs just 33 times in the Fringe category). Thirdly, *jihad*, while self-collocating in all three categories, does so with weakening strength as we move from Extreme through Fringe to Moderate, *kufr*; on the other hand, does quite the reverse.

All of the above sets parameters within which our analysis must operate. Firstly, when looking at the correlating words we need to consider a broader context than collocation will permit, using close reading to look for explanations of the correlations. Secondly, when looking at some collocates, we need to be aware that they tend to self-collocate and to explain, where possible, the effect and purpose of this. Thirdly, as was shown with the discussion of *kufr* and *jihad*, while measurement may at times seem to make the categories seem similar, a close inspection of the words in question may reveal that they differ across the categories. Let us begin this exploration with the self-collocating words shared by all categories, *kufr* and *jihad*.

Jihad collocates with itself 142 times in the Extreme category. Other than avoiding the +1 and -1 positions, it collocates relatively evenly in all the remaining slots (15, 15, 20 and 21 times in slots 2, 3, 4 and 5, symmetrically). This pattern is not the result of a single text. As the waning strength of the collocation might lead us to expect, the self-collocation of *jihad* in Fringe is less frequent, occurring only 8 times and in the Moderate category only 6 times. In each case, the pattern of distribution is similar to that in the Extreme category (+1/-1 avoided but the remaining slots evenly populated with examples).

What is the cause of the burstiness of *jihad* and why is it strongest in terms of collocation strength and frequency in the Extreme data? The burstiness is a function not of the word, but of the discursive purpose to which it is put in the texts, in particular those in the Extreme category. The concept of *jihad* is central to the discourse in numerous texts and its burstiness is produced by its repetition in the context of definition and explanation, as the following example demonstrates:

“Jihad with your wealth The financial Jihad has preceded the physical Jihad in every verse except one. This is to point out to us the importance of the Jihad of wealth because Jihad depends on it. In other words, no money no Jihad, and Jihad needs lots of it.”

(*44 Ways to Support Jihad* by Anwar Al-Awlaki)

In the Moderate texts, this strategy is not necessary – they are relying on a given understanding of the word and the concept is not the central focus of the discourse. Accordingly, while the word *jihad* occurs 518 times in 17 of the 51 Moderate texts (26.24

times PHTW), it self-collocates only six times. By contrast, the word *jihad* occurs 3,695 times in 130 of the 174 texts in the Extreme category (174.76 times PHTW), self-collocating 142 times. In the Extreme category, compared to the Moderate category, *jihad* i.) is more frequent; ii.) appears in a greater proportion of texts and iii.) engages in self-collocation more frequently. It is the need to introduce the concept, persuade readers of an interpretation of it and to differentiate the sub-types of *jihad* that causes the word's burstiness in the Extreme category. Burstiness here is an indication of an oppositional discourse resting upon the redefinition of a term with a given meaning. Clearly, more can be said about this word (for a fuller analysis of it, see Baker et al., 2020). But for now, we can see that discourse drives the burstiness of words.

Kufr is quite distinct in many ways from *jihad*. The decreasing strength of *jihad* from Extreme to Moderate might lead us to expect that the word will become less frequent as we shift across categories; this is not always the case. It is possible that less frequent occurrences of two words, but more exclusive co-occurrence of them, will lead to an increase in collocation strength. This is what we find with *kufr*, which collocates with itself 42 times in the Extreme category. Other than avoiding the +1 and -1 positions, it collocates relatively evenly in all the remaining slots (8, 4, 2 and 7 times in slots 2, 3, 4 and 5 symmetrically). This pattern is not the result of a single text – the collocates at positions 2 through 5 are produced by 5, 2, 2 and 4 texts respectively. The self-collocation of *kufr* in Fringe is less frequent, occurring only 6 times in 2 texts, and just 4 times in one text in Moderate. Overall, the mention of *kufr* declines from Extreme (38.73 cases PHTW, 59 out of 174 texts), through Fringe (26.82 cases PHTW, 16 out of 54 texts) to Moderate (2.38 cases PHTW, 9 out of 51 texts).

So, the increased strength of collocation is a sign of a greater exclusivity between the examples of *kufr* self-collocating in the Moderate as opposed to Extreme, for example. *Jihad*, a concept that was mentioned relatively frequently in all three categories, was the subject of increased self-collocation in the Extreme category where an oppositional interpretation of *jihad* seemed to be given. In the case of *kufr*, the word is used frequently in the Extreme texts and self-collocates with the use of self-collocation being similar in this category to *jihad*. In the Moderate texts, however, *kufr* exhibits a different pattern – the word is generally avoided but when it is mentioned it is more likely that it will self-collocate than in Extreme texts. In this case, the Moderate texts seem to avoid the concept, but in one case the word is used and carefully defined, with the different types of *kufr* being outlined (*50 Questions and Answers on Islamic Monotheism*, Anonymous). This text accounts for 11 of the 47 uses of *kufr* in the Moderate category. While this might lead one to set aside the example, it turns out to be typical of the category. If we examine the remaining examples, we find that all mentions of *kufr* are either used as part of definitions of the term (23 examples), without definition (18 examples), or with an associated definition (6 examples). Although the term *kufr* is thus used in a range of ways in the Moderate data, our analysis provides evidence that it tends to exhibit a strongly metalinguistic character (Jakobson, 1985), in particular when used as part of or alongside definitions. However, as noted, the term could also be used without definition. There could be a number of reasons for this. One explanation is that the texts' creators assume knowledge of the concept on the part of their target or 'imagined' readers (Bell, 1984) a hypothesis supported by the use of the term in the Extreme texts; in a random sample

of 100 uses of *kufr* in the Extreme category, only 11 were definitions of the term, with three further examples mixing use with definition.

Returning to the use of *kufr* in the Extreme texts, the largest use of the word in this category was either a mention of the word in the form of some personification or embodiment of *kufr* (50 examples – “thereafter kufr and its tyranny will be destroyed”, *Dabiq*, issue 7) or a direct claim that a person, act or organization has committed kufr (36 examples – “you then went overboard in your kufr”, *Inspire*, issue 14). So the overall picture emerging from *kufr* confirms the hypothesis that self-collocation may be a signal that definition is occurring, though the relationship between self-collocation and frequency is worth exploring, because what may be interpreted from the frequent yet defined (*jihad* in Extreme) is different to a degree from what may be inferred from the infrequent yet defined (*kufr* in Moderate).

Let us return to the correlating pairs which are also collocates of each other. In Extreme, the *kufr/shirk* pair is made up of two words which both self-collocate and mutually collocate. When we look at *shirk* we find self-collocation is, once again, a sign that an elaboration of meaning is being provided in the text, as in the following example “... is a very important book which talks about Tawheed and warns against Shirk, including Shirk of the graves and Shirk of the palaces.” (*Jihad Recollections*, Issue 1). However, the definitions do not occur in the context of the definitions of *kufr*. What brings the two words together is coordination – in examples such as “Would you want to live in a society where schools teach your kids Kufr, Shirk and sex education at the tender age of 8” (*Advice for Those Doing Hijarah*). *Kufr* and *shirk* collocate 39 times and on 35 occasions the two are coordinated together. The use of collocates in coordination has been observed before (McEnery, 2005) and is a key to establishing a spiral of signification in which concepts not necessarily linked are forced together, in this case sex education, kufr and shirk. This compounds any oppositional definitions as the definitions and the spirals of signification can work together to reinforce negativity by association and to draw concepts not specific to Islam, in this case sex education, into a frame of reference defined by the author in which the concepts are, in this case, condemned.

Self-collocation of *shirk* in the Fringe category is once again related to elaboration; “They should remove all influences of Jahilliyyah which make this concept impure and which may have the slightest element of hidden Shirk, such as Shirk in relation to homeland, or in relation to race or nation, or in relation to lineage or material interests” (*Milestones*, Sayyid Qutb). As with the Extreme examples, coordination explains why the terms correlate and collocate – all 27 examples of *kufr* and *shirk* collocating in the Fringe data are coordinated, though in no cases are non-Islamic concepts included as part of that coordination, unlike the Extreme cases. Nonetheless, *kufr/shirk* in Fringe discourse is working in a way that is very similar to the way that it operates in the Extreme category. This stands in sharp contrast to the Moderate data, in which this pairing of words neither correlates nor collocates.

The exploration of the collocated pairs so far has begun to suggest rhetorical processes that may be associated with these pairs. To examine these we will begin by placing each of the words examined into broad semantic fields. Ideally, we may want to do this with a readily available automated semantic tagger, but this is not possible as none are developed for the domain we are exploring. Accordingly, a provisional ontology was developed which

is broad enough to capture key differences between the terms explored while giving what we think is a reasonable level of specificity in the ontology itself. This ontology is given in Table 5 below. In the system we use sub-categories sparingly and only in cases where there are terms which are so distinct that they clearly warrant a sub-category.

Table 5. Semantic fields of words in Figure 1.

Category	Sub-category (if applicable)	Definition	Example
Adherence		A key element of adherence to the faith	<i>Deen</i>
Arts		Art forms related to the religion	<i>Nasheed</i>
Authority		Authorities which may be cited to justify an argument within the religion	<i>Hadith</i>
Conflict		Spiritual or physical struggle or unrest	<i>Fitna</i>
Negative		An act evaluated as negative within the religion	<i>Riddah</i>
Positive		An act evaluated as positive within the religion	<i>Dawa</i>
Spiritual	Negative Mixed	A reference to a spiritual or supernatural entity or place, real world or otherwise.	<i>Janah</i> (<i>jahanaam</i> negative, <i>dunya</i> mixed)
State		Reference to a geo-political entity	<i>Caliphate</i>
Them		Non-Muslims	<i>Kuffaar</i>
Us	Negative	Muslims	<i>Ummah</i> (<i>munafiq</i> negative)
Us and Them		Muslims in relation to non-Muslims	<i>Al-wala-wal-bara</i>

The elements of the correlated pairs in Tables 2 and 3 were assigned to each of these categories. The use of the sub-categories of the Spiritual field is to prise apart the temporal (mixed), the infernal (negative) and the heavenly (unmarked). Similarly, in the Us category those who are identified as being an insider outgroup were identified as negative. By contrast, other categories, such as Them, were relatively flat, though of course in context a series of sub-categories may emerge. What this categorization aimed to do was simply distinguish the terms investigated by their accepted meaning.

The grouping of the correlated pairs by concepts reduced the number of unique pairs. For medium correlation, the semantic fields produce 29 distinct pairs for Extreme (previously 58), 30 for Fringe (previously 78) and 20 for Moderate (previously 34). For large correlation, the semantic fields produce 15 distinct pairs for Extreme (previously 21), 37 for Fringe (previously 113) and 18 for Moderate (previously 52).

3.3. Self-Correlation

One thing that the analysis does quite quickly is to identify where the correlations are, in fact, from the same broad semantic field. Table 6 summarises these pairs which, in the data, self-correlate in terms of semantic field.

Table 6. Semantic fields self-correlating in each category of text.

Moderate Correlation	Extreme	Fringe	Moderate
Authority	4	6	6
Negative	1	1	1
Them	4	0	0
Strong Correlation	Extreme	Fringe	Moderate
Authority	0	0	11
Adherence	0	0	2
Negative	2	3	0
Them	1	4	0
Spiritual	0	1	0

Given that different words may occur in the same semantic field, this means that concepts, as well as words, can display burstiness. There is a good linguistic explanation for this – cohesion by reiteration, produced when ‘a word that is in some way associated with another word in the preceding text, because it is a direct repetition of it, or is in some sense synonymous with it’ recurs (Halliday and Hasan, 1976: 319). When referring to a group, for example, it is possible to set up textual cohesion by using the same word to refer to that group repeatedly and/or to use broadly equivalent words to talk about them. Similarly, we may use a singular and a plural form as we move through a text when referring to the group as the example from *Jihad* (by an unknown author) involving MUSLIM and *kuffar* below (underlined for ease of presentation) illustrates. This extract also provides an example of noncanonical opposition (Davies, 2012) as the lemma MUSLIM, though referring ostensibly to the same group, is used to construct an oppositional relationship between ‘Muslims allying themselves with the Kuffar’, who are placed *against* those who do not, who are referred to simply as ‘Muslims’.

“(Re: the Kuffar and the Muslims, allying themselves with the Kuffar against the Muslims). ‘Kill them even if they are praying in the mosque’ Even if a Muslim gives water to the Kuffar who are fighting against Muslims they are Kafir.”

Note that in the example neither *kuffar* or MUSLIM collocates with itself even once, but the words exhibit burstiness. While collocation is the second driver of lexical cohesion (lexical cohesion may be formed by a chain of words which tend to occur ‘in the same lexical environment’, Halliday and Hasan, 1976: 319), it is not key here. In this example we are seeing two chains of lexical cohesion, linked by narrative, which are consequently co-dependent, formed by lexical cohesion through repetition. The correlation is a signal that this may be happening.

Yet repetition is not occurring in every category. Of the eleven categories, only 3 produce this effect through medium correlation and only five produce it through a large correlation. This is, in part, a further indication of topicality (van Dijk, 1977). Some categories are well populated and do not seem to produce these interwoven chains in the way that the categories in Table 5 do – notably, words belonging to the Us category do not produce this effect, even though these words are frequent in the corpus (1,589 occurrences across 130 texts). Also, of those categories which do produce these self-correlating fields, those fields are not the same in each category. For example, fields self-correlating in Moderate (Authority and Adherence) do not self-correlate at all in the Extreme and Fringe categories and Adherence only self-correlates in the Moderate texts, and not at all in the Extreme or Fringe data, either as a medium- or large-strength correlation. Let us look at the theme of Adherence in the Moderate texts to exemplify what is happening. Consider *deen* and *taqleed*, one of the correlated pairs which gives rise to Adherence self-correlating with a large effect for the Moderate texts. In *The Nature of Taqleed* (by Taqi Uthmani), *taqleed* is a central topic and as a consequence lexical cohesion through repetition of *taqleed* is common, occurring 205 times. It appears in the book in clusters and produces numerous cohesive ties. As *taqleed* is discussed, it causes Uthmani to discuss it in relation to *deen* to emphasise that the two are related yet different, as in the following excerpt:

“As far as the Islamic rules are concerned, there are of two types. The first are those which are known by necessity to be part of the Deen of the Prophet sallallahu alaihi wa sallam like the five prayers, Zakaat, fasting in Ramadhan, Hajj; the prohibition of adultery, wine and so on. Taqleed is not allowed in these issues since they are such that everyone should know and understand.”

Hence dimensions of adherence are discussed and related to one another, causing the correlation.

As mentioned, lexical cohesion can link to rhetorical structure and in the examples given we can see this. In one, two concepts produced lexical cohesion and the chains of cohesion were co-dependent because the discussion of the concepts were linked (*kuffar* and MUSLIM). In the second example, the need to distinguish apart two concepts led to them being brought together, chaining in a co-dependent way (*taqleed* and *deen*). So, the need of the discourse to introduce, link or differentiate concepts is realised, or at least visible, through lexical cohesion which, in turn, is observable to a degree through correlation. We say to a degree because, even though we are dealing here with relatively specific terminology that is

fairly widespread across the corpus, some false positives do occur. For example, in *The Nature of Taqleed*, *deen* also appears as part of a name producing examples which have to be identified and set aside. So, while correlation is a useful guide, it is not an infallible one and all of the results of it must be examined manually.

3.4. Beyond Self-Correlation

To understand the rhetorical purpose of categories which do not self-correlate, but which do correlate with other categories, we devised a scheme of rhetorical goals that these correlations appear to serve. Our claim here is not that these correlations serve this category all of the time. Rather, these categories are a plausible interpretation of the rhetorical purpose of these correlations, derived from a qualitative investigation of relevant concordance lines. Our goal in each case was to categorise on the basis of what was the predominant rhetorical function of each pairing. Table 7 below gives the categories, while Table 8 shows the distribution of these categories across the corpus. The self-correlating examples are included in this table in the category Cohesion. While principally focussed on Rhetoric, we have included Cohesion as a category here as it has a role to play in the texts which is, arguably, rhetorical, as the examples involving coordination in section 3.2 showed.

Table 7. Rhetorical elements and examples³

Rhetorical Category	Description	Example
Afterlife	Linking actions on Earth to the afterlife, or vice versa	‘whoever calls to the call of jahiliyyah then he is from the gatherings of jahannam ’ ⁴
Association	Linking a group with an action or evaluation, or vice versa	‘this is more severe in kufr and inclination towards the mushrikeen ’ ⁵
Cohesion	The category self-correlates	See previous section
Consequences	Linking an action/state to its outcome/origin, or vice versa	‘ Taqwa from the might of kufr ’ ⁶
Evaluation	Linking an action with an action of evaluation, or vice versa	‘the shirk of putting off Jihad ’ ⁷
Scriptural warrant	Where something is justified with reference to accepted written authorities	‘Shaykh az-Zarqawi Abu Mus'ab (rahimahullah) anticipated the expansion of the blessed jihad from

³ It should be noted that the examples given have been selected as demonstrating a point in as little space as possible, hence these examples are cohesion by collocation rather than repetition.

⁴ From *Dabiq*, Issue 1.

⁵ From *The Evidence for the Ruling Regarding Alliance with the Infidels and Matters Related to it* by Shaykh Al-Islam.

⁶ From *Sharpening the Sinan for the Fighting the Government of Pakistan and its Army* by Abu Yahya Al-Liby.

⁷ From *Clarifying the Obligation of Migration from the Lands of Disbelief to the Lands of Islam* by Abd Al Aziz bin Salih al-Jarbu.

		Iraq into Sham and linked it to this hadith ' ⁸
--	--	---

Table 8. The number of pairs in categories in each section of the corpus (figures in parentheses give the rank and percentage of examples of the pairs in this rhetorical and text category).

Moderate Correlation	Extreme	Fringe	Moderate
Afterlife	8 (4, 13.79%)	23 (2, 29.49%)	3 (4, 8.82%)
Association	16 (2, 27.59%)	12 (3, 15.38%)	8 (2, 23.53%)
Cohesion	11 (3, 18.97%)	7 (4, 8.97%)	7 (3, 20.59%)
Consequence	0 (6, 0%)	2 (=5, 2.56%)	0 (6, 0%)
Evaluation	1 (5, 1.72%)	2 (=5, 2.56%)	2 (5, 5.88%)
Scriptural warrant	22 (1, 37.93%)	32 (1, 41.03%)	14 (1, 41.18%)
Strong Correlation	Extreme	Fringe	Moderate
Afterlife	2 (4, 9.52%)	20 (3, 17.7%)	3 (4, 5.77%)
Association	7 (2, 33.33%)	34 (2, 30.09%)	15 (2, 0%)
Cohesion	3 (3, 14.29%)	15 (4, 13.27%)	13 (3, 28.85%)
Consequence	0 (6, 0%)	1 (5, 0.88%)	0 (=5, 0%)
Evaluation	1 (5, 4.76%)	0 (6, 0%)	0 (=5, 0%)
Scriptural Warrant	8 (1, 38.10%)	43 (1, 38.05%)	21 (1, 40.38%)

Let us illustrate how correlation and rhetoric may intersect with such words with regard to one whole text example, by an anonymous author, which contains two correlating pairs: *sunnah/jihad* and *hadith/jihad*. In the system of rhetorical categorization, each are in the Scriptural Warrant category. In the example, words in the first pair are underlined and those in the second emboldened. Words examined thus far which do not participate in a correlating pair in this text are italicized.

The Role of the Mujahidah

There are two extreme viewpoints of **Jihad**

Jihad of the Nufs – an internal struggle to improve your own self

Death to the *Kuffar* by whatever means necessary

According to Qur'an and Sunnah

Jihad is to fight without rest, to exhaust all your efforts

60% of the Qur'an refers to **Jihad**

Defensive **Jihad**

To protect your lands, honour, life, family and wealth.

Offensive **Jihad**

To conquer lands and give mercy to others (to bring Islam to others).

It is to remove the oppression of man made law and is launched by the Islamic State.

⁸ From *Dabiq*, Issue 1.

In the UK and in all non occupied countries it is Fard for men to train for **Jihad**

In the lands of occupation, e.g. Palestine and Kashmir, **Jihad** is Fard for men, women and children.

“March forth, whether you are light (being healthy, young and wealthy) or heavy (being ill, old and poor), and strive hard with your wealth and your lives in the Cause of Allah. This is better for you, if you but know”.

[At-Taubah 9:41]

Sisters encourage your husbands to train and go for **Jihad**

Raise your children as Mujahideen and Scholars

Do not introduce them to culture and nationalism

Hadith:

‘Fight to make Allah’s *Deen* the highest’

Child birth is the biggest **Jihad** for a woman

In the text both of the correlations are indeed focused on providing a scriptural warrant for jihad. Jihad provides the main cohesive link in the text – it does not self-collocate. Lexical cohesion is formed through repetition. Other concepts, such as *deen* and *kuffar* are mentioned. In the case of *kuffar* it is arguable that the word forms a further correlating pair with *jihad* as the variant *kuffaar* in the Extreme category does correlate with *jihad* producing a pair with medium strength correlation which would be counted as Association rhetorically. The example of *deen* is different – it does not correlate and is only mentioned as part of the hadith which is the critical Scriptural Warrant for the waging of *jihad*. So in this text we can see how correlation and rhetoric can work, how correlation and topic can be linked by lexical cohesion, but also how, in the context of the rhetorical function of correlation, the frequency of each part of the correlated pair may be different in any given text in which it occurs. Through those correlations we see the Scriptural Warrant for *jihad* (through correlation of *jihad* with *sunnah* and *hadith*) and the object of *jihad* introduced through the Association of *kuffar* with *jihad*. Note that, as the correlations are manifested in the text, yet are observable at the level of the category of texts, we are seeing in action in this text correlated pairs which occur in other texts also, allowing us to see trends across the data manifested in individual texts and vice versa.

The Fringe texts appeal to the Afterlife more than the other texts (for a further analysis of the Afterlife category, see Baker et al., 2020). The Extreme texts, on the other hand, rely more on Association than the Fringe texts. While the Moderate texts rely on a similar proportion of pairs in this category, Extreme has a larger number of such pairs. Similarly, Extreme relies more on the Cohesion category than the other two categories. However, all three categories rely heavily on Scriptural Warrants. This gives us an idea, at a high level of abstraction, of how ideology and text interact in the three categories in the corpus.

4. Violent jihad and symbolic capital

Our investigation indicates that the analysts were right to identify the words examined as important; they, doubtlessly along with others, do indeed have a role to play in the

organization of discourse. Yet another sense in which the words looked at here are important is in the symbolic capital (Bourdieu, 1989:17) that they convey. These are Arabic loanwords associated with meanings that the authors claim derive from texts associated with the divine. For adherents of the religion, these words have symbolic capital – value that is ‘recognized as legitimate’ (Bourdieu, 1989:17) and whose ‘perception and appreciation [...] express the state of relations of symbolic power’ (Bourdieu, 1989: 20). The terms, and their source, represent a form of symbolic power. These words, because of their association with the divine, become important sources of prestige in themselves. Such social power, through which the terms used constitute a ‘consecration or revelation’, represents ‘political power par excellence’ (Bourdieu, 1989:23). Symbolic power of this sort is a key resource that may be employed in a discursive struggle to form perceptions of the social world. By discursive struggle, we refer to the ‘dialogical struggle (or struggles) as reflected in the privileging of a particular discourse and the marginalization of others’ (Hardy and Phillips, 1999: 3). This struggle has both objective and subjective dimensions. At the objective level, the persuasion can be to action – what Bourdieu terms group demonstrations which ‘manifest the size, strength, and cohesiveness’ of the group through action (Allan 2011:430).

Religion is, classically, a symbolic system and hence an approach to understanding it through symbolic capital seems warranted (see also: Swartz, 1996). Within these systems, ‘symbolic power [...] finds expression in everyday classifications, labels, meanings and classifications that subtly implement a social as well as symbolic logic of inclusion and exclusion’ (Swartz, 2013: 39). This explains the connectedness of the terminology explored in this analysis – it comes together in repeated patterns because it is a manifestation of a system ordered by a logic. That logic focuses on acts that are good and bad (Positive and Negative in the semantic fields used here), insiders (Us), outsiders (both Them and the negative Us group) and identifies different forms of struggle and conflict that can result from the logic of this system (the Conflict category). The authors, in writing about these topics, are asserting that their symbolic capital affords them ‘the social authority to impose symbolic meanings and classifications’ (ibid) within the system. They do this through ‘public recognition of their capital holdings and positions occupied in social hierarchies’ (ibid).

The differences between the Moderate, Fringe and Extreme texts link to the struggle to acquire symbolic capital through which to influence action. Textual authorities derived either directly or indirectly from the divine are key for all groups as a base of symbolic capital which may be drawn down but also, as we saw with the definitional preoccupation in the corpus, cast afresh, redefined. The divine is a potentially powerful source of the doxa – the natural order – of social order. Religion as a form of social organization can draw upon this apparently infallible guide to these norms. Yet the doxa in these texts is also the fulcrum of change across the categories; while the Moderate texts may be relying on a doxa to the extent that they do not engage in definition as much as in the other categories, the definitional focus of the Fringe and Moderate texts, focused on a set of terms which form part of the symbolic capital of the religion, is a clear sign that they are seeking to redefine the doxa. As individuals are persuaded, a group of those convinced forms, reminding us that the ‘second major mode of Bourdieu’s political sociology’ is the power of ‘authoritative nomination and the symbolic fabrication of collectives’ (Wacquant, 2004: 6). Symbolic capital is used for naming and definition; naming and definition form symbolic power; symbolic power, when

accepted, forms new groups who abide by that new doxa and in doing so may legitimate social action, for example of Us on Them; ‘the categories of perception, the systems of classification, that is, essentially, the words, the names which construct social reality as much as they express it, are the crucial stakes of the political struggle, which is a struggle to impose [...] vision and division’ (Bourdieu, 1990: 134).⁹

The struggle in the corpus, represented by the divisions within it, is a struggle for symbolic power through the accumulation of symbolic capital. In that struggle, different systems, representative of one or another version of such power, are in competition. The competition is played out using, and distorting, a complex system based on the definition of words and their relationship to one another and to actions – it is this that explains much we have seen in this analysis.

5. Conclusion

This article has demonstrated how the study of correlation, and in particular the repetition of correlating words, can shed light on the workings of discourse in and across ideologically loaded texts. In this case, we have interpreted the correlational patterns occurring across a set of violent jihadist texts as evidence of an ideological struggle for symbolic power, acquired through the accumulation of symbolic capital. This struggle is a significant one, given that the success of any terrorist organisation in recruiting and mobilising individuals to its cause is contingent not only upon how receptive those individuals are towards the ideological import of its messages (including those circulated through the (mass) media), but also on the ability of the organisation in question to establish itself as the ‘dominant or only legitimate conveyor of these ideas’ (Malešević, 2017: 271). Of course, analysis of these texts alone cannot account for all of the influences and processes that led to their possessors holding the particular extremist views that resulted in their arrests. However, we believe that this approach has been fruitful in uncovering the types of rhetorical strategies employed in these (and likely other) texts to attempt to persuade their readers to a particular (extremist) ideological viewpoint and, potentially, a course of violent action. Given that the individuals found in possession of these texts were also found guilty of terror-related offenses, it is possible that some or all of these strategies were effective in these cases.

If this is the case, we must bear in mind that these texts will have likely played only a partial role in the formation of the worldviews that led to their possessors being convicted of terror-related offenses. As Malešević (2017: 270-271) points out, ‘[a]lthough terrorist outfits deploy intense ideological rhetoric and justify their violent acts through direct references to central principles of their respective ideological doctrines, one should not take such pronouncements at face value. Since all social organisations have to reconcile their doctrinal principles with the bureaucratic models of management, ideological messages are almost never popularly absorbed as they are presented’. Moreover, the reception of doctrinal ideas is not straightforward but can be subject to, amongst other things, contradictions, misinterpretations and resistant readings (Billig et al. 1988). Hence, Malešević, argues that

⁹ The analysis could extend here to include symbolic violence and, in the context of the Conflict category, of how that might in some sections of the corpus be a pathway to licencing physical violence. However, the interpretation of the data in the context of symbolic power will stop short of a consideration of symbolic violence here.

the type of ideologisation we have found in our texts to be ‘regularly contingent, uneven and contested process that remains dependent on the coercive prop of the organisational shell’ (2017: 271). For example, Malešević also contends that although ISIS invests great emphasis and huge financial resources into its Takfirist/Salafist ideology, such ideas would not have attracted anything near the attention they did were it not for the group’s violent displays of its organisational might, including on the battlefield and in its ‘macabre beheadings of Westerners’ (ibid.). The ideological force and consequences of the texts studied here must thus be viewed in concert with the power and broader activities of their associated organisations.

Violent jihadist discourse has provided a convenient test case for this study. However, the approach used in this paper could theoretically be applied to any corpus or set of texts, including but not just in analyses of claims to symbolic power between competing texts. Future work could test the flexibility of this approach by applying it, for example, to texts produced by fascists or white supremacists. Whatever the focus of future work may be, the introduction of non-linguist expert perspectives in the selection of search terms, as in this study, can render that selection process more robust and grant the resultant terms greater real-world correspondence.

Acknowledgements

The research reported in this paper was supported by the UK’s Economic and Social Research Council, grant number ES/R008906/1. The authors wish to thank Annabelle Lukin and Charlotte Taylor for reading an earlier draft of this work.

References

- Al-Rikaby, A and Mahadi, T (2018) The significance of the discursive strategies in al Baghdadi's and al-Zawahiri's hortatory speeches: a multidisciplinary approach. *Journal of Language and Politics*, 17(6), 769–788.
- Allan, K (2011) *The Social Lens*. London: Sage.
- 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, 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 (2020) *The Language of Violent Jihad*. Cambridge: Cambridge University Press.
- Bell, A (1984) Language Style as Audience Design. *Language in Society*, 13(2), 145–204.
- Billig, M, Condor, S, Edwards, D, Gane, M, Middleton, D and Radley, A (1988) *Ideological Dilemmas: A Social Psychology of Everyday Thinking*. London: Sage.
- Bourdieu, P (1989) Social space and symbolic power. *Sociological Theory*, 71(1): 14–25.
- Bourdieu, P (1990) *In Other Words: Essays towards a reflexive Sociology*. Stanford, CA: Stanford University Press.
- Bourdieu, P (1991) *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.

- Brezina, V (2018) *Statistics in Corpus Linguistics: A practical guide*. Cambridge: Cambridge University Press.
- Brezina, V, Timperley, M and McEnery, T (2018) #LancsBox [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.
- Church, KW and Gale, WA (1995) Poisson mixtures. *Natural Language Engineering* 1(2): 163–190.
- Cohen, J (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conoscenti, M (2016) ISIS' Dabiq communicative strategies, NATO and Europe. Who is learning from whom?' In Ceretta, M and Curli, B (eds), *Discourses and Counter discourses on Europe: From the Enlightenment to the EU*. London: Routledge, pp. 215–244.
- Davies, M (2012) A New Approach to Oppositions in Discourse: The Role of Syntactic Frames in the Triggering of Non-Canonical Oppositions. *Journal of English Linguistics*, 40(1), 47–73.
- Droogan, J and Peattie, S (2018) Reading jihad: Mapping the shifting themes of Inspire magazine. *Terrorism and Political Violence* 30(4): 684–717.
- Fairclough, N (2003) *Analyzing Discourse: Textual Analysis for Social Research*. London: Routledge.
- Fairclough, N (2005) 'Critical discourse analysis in transdisciplinary research'. In Wodak, R and Chilton, P (eds), *A New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity*. Amsterdam: John Benjamins, pp. 53–70.
- Firth, JR (1957) *Papers in Linguistics 1934–1951*. Oxford: Oxford University Press.
- Gablasova, D, Brezina, V and McEnery, T (2017) Collocations in corpus-based language learning research: identifying, comparing and interpreting the evidence. *Language Learning* 67(1): 155–179.
- Gleason, HA (1968) *Contrastive Analysis in Discourse Structure*. Washington, DC: Georgetown University Institute of Languages and Linguistics.
- Gutwinski, W (1976) *Cohesion in Literary Texts: A Study of Some Grammatical and Lexical Features of English Discourse*. The Hague: Mouton de Gruyter.
- Halliday, MAK (1964) *The linguistic study of literary texts*. In H. G. Lunt (ed), *Proceedings of the Ninth International Congress of Linguists*. The Hague: Mouton de Gruyter, pp. 302–307.
- Halliday, MAK (1973) *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, MAK and Hasan, R (1976) *Cohesion in English*. Longman, London.
- Halliday, MAK and Hasan, R (1995) *Language, Context and Text*. Oxford: Oxford University Press.
- Hardy, C and Phillips, N (1999) No joking matter: discursive struggle in the Canadian refugee system. *Organization Studies* 20(1): 1–24.
- Hart, C (2014) *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury.
- Hellmich, C (2008) Creating the ideology of Al Qaeda: From hypocrites to Salafi-Jihadists. *Studies in Conflict & Terrorism*, 31(2), 111–124.

- Holbrook, BD (2014) *The Al-Qaeda doctrine. The framing and evolution of the leadership's public discourse*. London: Bloomsbury.
- Holbrook, D (2019) 'The terrorism information environment: analysing Terrorists' selection of ideological and facilitative media', *Terrorism and Political Violence*, published online, <https://doi.org/10.1080/09546553.2019.1583216>.
- Ingram, HJ (2017) An Analysis of Inspire and Dabiq: Lessons from AQAP and Islamic State's Propaganda War. *Studies in Conflict & Terrorism* 40(5): 357–375.
- Jakobson, R (1985) Metalanguage as a linguistic problem. In: Rudy, S (eds) *Selected Writings, VII*. Berlin: Mouton, pp. 113–121.
- Kjellmer, G (1991) 'A mint of phrases'. In K. Aijmer and B. Altenberg (eds), *English Corpus Linguistics*. London: Longman, pp. 111–127.
- Lohlker, R and Abu-Hamdeh, T (2013) *Jihadi thought and ideology*. Berlin: Logos Verlag.
- Lukin, A (2019) *War and Its Ideologies: A Social-Semiotic Theory and Description*. London: Sage.
- Mahlberg, M (2006) Lexical Cohesion: Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics* 11(3): 363–383.
- Malešević, S (2017) *The Rise of Organised Brutality: A Historical Sociology of Violence*. Cambridge: Cambridge University Press.
- Manne, R (2017) *The mind of the Islamic State. ISIS and the ideology of the caliphate*. Amherst, New York: Prometheus Books.
- Martin, JR (2012) *Collected Works, vol. 5: Discourse Analysis* (ed. by W. Zhenhua). Shanghai: Shanghai Jiao Tong University Press.
- Martin, JR (2015) 'Cohesion and Texture'. In D. Tannen, H.E. Hamilton and D. Schiffrin (eds), *The Handbook of Discourse Analysis* (Second Edition). Oxford: Blackwell, pp. 61–86.
- McEnery, T (2005) *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. London: Routledge.
- Mitchell, J (2008). The contradictory effects of ideology on jihadist war-fighting: The Bosnia precedent. *Studies in Conflict & Terrorism*, 31(9), 808–828.
- Morley, J (2009) Lexical cohesion and rhetorical structure. In: Flowerdew, J and Mahlberg, M (eds) *Lexical Cohesion and Corpus Linguistics*. Amsterdam: John Benjamins, pp. 5–22.
- 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.
- Rayson, P (2008) From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519–549.
- Simpson, P (1993) *Language, Ideology and Point of View*. London: Routledge.
- Sinclair, J McH (1996) The search for units of meaning. *Textus*, 9, 75–106.
- Sourdél, D and Sourdél-Thomine, J (2007) *A Glossary of Islam*. Edinburgh: Edinburgh University Press.

- Spier, TE (2018) Extremist propaganda and Qur'anic scripture: A 'radical' corpus-based study of the Dabiq. *Discourse & Society* 29(5): 553–567.
- Stubbs, M (2015) 'Computer-Assisted Methods of Analyzing Textual and Intertextual Competence'. In D. Tannen, H.E. Hamilton and D. Schiffrin (eds), *The Handbook of Discourse Analysis* (Second Edition). Oxford: Blackwell, pp. 486–504
- Swartz, DL (1996) Bridging the Study of Culture and Religion: Pierre Bourdieu's Political Economy of Symbolic Power. *Sociology of Religion* 57(1): 71–85.
- Swartz, DL (2013) *Symbolic Power, Politics and Intellectuals*. Chicago, IL: The University of Chicago Press.
- Taylor, C (2018) Representing the Windrush generation: metaphor in discourses then and now. *Critical Discourse Studies*, Online first: <https://www.tandfonline.com/doi/full/10.1080/17405904.2018.1554535>.
- van Dijk, TA (1977) Sentence topic and discourse topic. In: Stolz, BA (ed) *Papers in Slavic Philology 1: In Honour of James Ferrell*. Ann Arbor, MI: Department of Slavic Languages and Literature, University of Michigan, pp. 49–61.
- Wacquant, L (2004) Pointers of Pierre Bourdieu and democratic politics. *Constellations* 11(1): 3–15.
- Wignell, P, Tan, S and O'Halloran, KL (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.
- Wright, D and Brookes, G (2018) 'This is England, speak English!': a corpus-assisted critical study of language ideologies in the right-leaning British press. *Critical Discourse Studies* 16(1): 56–83.