

This is not the publisher's copy. That is available here: doi 10.1075/ijcl.17.2.05pre

Title: The language of Islamic extremism: towards an automated identification of beliefs, motivations and justifications

Authors: Sheryl Prentice (Department of Psychology), Paul Rayson (Computing Department) and Paul Taylor (Department of Psychology)

Affiliation: Lancaster University

Contact details:

s.prentice@lancaster.ac.uk

paul@comp.lancs.ac.uk

p.j.taylor@lancaster.ac.uk

Correspondence address:

Department of Psychology

Fylde College

Lancaster University

Lancaster

LA1 4YF

UK

The language of Islamic extremism: towards an automated identification of beliefs, motivations and justifications.

Sheryl Prentice, Paul Rayson, & Paul Taylor

Lancaster University

Abstract

Several recent studies have sought to understand individuals' motivations for terrorism through the content of terrorist material. To date, however, these studies have not capitalised on automated language analysis techniques, particularly advances made in corpus linguistics. In this paper, we demonstrate how applying three corpus-linguistic techniques to extremist statements can provide insights into the ideologies that underpin terrorist media. Our data consisted of a corpus of 250 statements (~500,000 words) promoting violence in support of a terrorist cause, each written in first person, and published between 1996 and 2009. Using the online software tool WMatrix, we submitted these data to frequency count, key word and key concept, and concordance analyses. Results showed that authors centre their rhetoric on the themes of morality, social proof, inspiration and appeals to religion. Authors also referred to the world via contrasting concepts (e.g. truth versus falsehood), suggesting a polarised way of

thinking when compared to a general population usage. Finally, we show how the technique of collocation can aid the establishment of networks between people and places. We discuss how the examination of language can provide further insights into our understanding of the terrorist mindset, as well as how it might support the formulation of evidence based counter-terrorism strategies.

Key words: terrorism, ideology, rhetoric, corpus linguistics, forensics

1. Introduction

Previous research into extremist ideology has been qualitative in nature and therefore of limited scale. In addition, only a few studies have focussed on the language of extremists themselves. This paper considers how automated corpus-linguistic techniques may be used to facilitate the process of identifying the ideology expressed in extremist material. The techniques we explore identify patterns in the usage of words and word collocations that are unique to the corpus under examination. Interpreting these patterns affords insights into an author's ideology by showing what ideas, beliefs and motivations he or she expresses, and, on further examination of examples, what attempts he or she makes to justify or enforce these ideas, beliefs and motivations.

In the next section we give an overview of linguistic studies into extremism and suggest how our work builds on this existing research. In sections 3 and 4, we provide a brief description of the specific corpus linguistic techniques and outline how they were

employed in an analysis of a corpus of extremist texts. We discuss the results arising from this application in section 5 and conclude with a summary of our findings, some caveats, and suggestions for future research.

2. Literature Review

There is a large body of work examining how public authorities construct terrorism and extremism in their publications. For example, Lazar and Lazar (2007) show how America uses the language of law to justify violence in the post-Cold War era, while Becker (2007) suggests that pronouns, questioning strategies, and foregrounding allowed German politician Schroder to construct an ‘us-them’ dichotomy to forge support for the Iraq war. Common to these analyses is a demonstration of how linguistic presentation can provide insights into Western understanding of extremism. For example, Stoltz (2007) highlights the discriminatory discourse surrounding Arabs in the American press, observing that Arabs are routinely constructed in a vague manner ‘through shifting references’ and through contrasts to all that is ‘Western and free’. This vague presentation of the Arab identity is argued by some to contribute to an environment of fear. As Stenvall (2007) notes, by constructing terrorism fears in the form of agentless passives, such fears become “powerful, free-floating entities with little or no visible connection to those who fear.”

The vague use of terms extends to what some authors describe as ‘collateral language’ (Collins & Glover 2002) or ‘language of war’ (Smith 2007), particularly with words

such as 'jihad', 'terrorism' and 'fundamentalism'. Renold (2002) argues that the American press use the term 'fundamentalism' so frequently in association with Islam that it has become inappropriately synonymous with the religion. Such misappropriation can lead to misunderstandings or intolerance between religious/ethnic groups. In addition, various studies have looked at the Western media (Baker 2010; Baker et al. 2008; Baker & McEnery 2005; Flowerdew, Li & Tran 2002; Johnson & Suhr 2003; Martin & Phelan 2002) and politicians' negative representation of Muslims and other 'minority' groups that they perceive as a threat to social order (Every & Augoustinos 2007; Lazar & Lazar 2004; Van der Valk 2003). As the authors of these studies suggest, the negative treatment of such groups can lead to a feeling of being side-lined within society, thus breeding dissatisfaction, which has long been identified as a key component in the emergence of extremism (Berman 2006; Iannaccone & Ransford 1968; Sandbrook & Romano 2004).

This large body of work on public construction of terrorism and extremism is extended by a smaller set of studies on the language that extremists themselves use. There have been a number of studies on the discourse of extreme right-wing political parties, particularly those present in Europe (Bowden 2008; Reisigl & Wodak 2001; Simon-Vandenberg 2008; Testa & Armstrong 2009). For example, in their book on anti-semitism, Reisigl and Wodak identify five types of discursive strategy that appear in racial, national or ethnic discourse: i) referential or nomination strategies that construct or identify social actors as in-group or out-group; ii) predicational strategies that assert stereotypical, evaluative attributions; iii) argumentation strategies that are used to justify the positive or negative attributions made; iv) perspectivation, which is the framing or

discourse representation of how authors express their involvement or position their point of view in the discourse; and, v) intensifying and mitigation strategies, which help to qualify or modify the epistemic status of a proposition.

Another limited area of work on extremist discourse is the study of how online hate groups use the Internet to communicate their ideology. Brindle (2009) used corpus techniques to study the language of a white supremacist forum and found that ‘homophobia, racism and sexism are inseparably interlinked’ in their rhetoric. Similarly, Duffy (2003) studied four different hate group websites (white Nationalists, Neo-Nazis, the Ku Klux Klan, and black separatists) and identified two prominent themes: The Plea for Fairness and Justice theme in which authors use ideas of fairness, justices and morality to justify claims and beliefs, and The Natural Order and Resurrection of the People theme that advocates a new world in which things are restored to their “original order”.

Although these studies explore a number of extreme groups, linguistic research into the ideology of Islamic extremists remains scarce. A notable exception is Chertoff (2008), who observed that the language of extremists is relatively unchanged, as “it mimics the radical rhetoric of the last century”, using words such as “vanguard” and “revolution” to describe themselves and “imperialist,” and “establishment” to describe the enemy. Chertoff (2008) argues that even when using Muslim terminology, extremist rhetoric “evokes a set of norms and tactics that depart from a traditional understanding of Islam in crucial ways”, including, for example, their distortion of ‘jihad’ to mean solely armed struggle, just as was observed in relation to the Western media. In this way, Chertoff

(2008) claims, the worldview expressed within extremist texts is similar to that of 20th century communism and fascism in its desire to radically reorder the society of the world, using violence against civilians as a way to overthrow existing powers.

Another exception to this relatively less researched area is the work of Stout (2009), who studied the language of both captured documents and open source jihadist material written by Al-Qa'ida affiliates and found that such texts were not simply ideologues, pure propaganda or technical descriptions, but also contained elements of strategic thinking. Such thinking, Stout (2009:876) states, is 'grounded in mainstream global thought on revolutionary warfare' and is concerned with 'the extent to which the foot soldiers ignore their prescriptions, engaging in disjointed, counterproductive operations.' Stout suggests these findings highlight Al-Qa'ida's ability to adapt to new strategic approaches, which could potentially increase the group's chances of success. Practically, what Chertoff's and Stout's analyses show is that it is possible to interpret the presentation within extremist texts by carefully examining the language used.

In this paper we focus on building an understanding of Islamic extremist ideology using corpus linguistic techniques. The aforementioned studies have been qualitative in nature (with the exception of the corpus based studies of Baker 2010; Baker et al. 2008; Baker & McEnery 2005; Brindle 2009), suggesting that the full range of quantitative techniques available to study the language have not yet been fully utilised in this domain. Further, it appears that corpus linguistic techniques have not been utilised in the study of Islamic extremist ideology. Our analysis addresses this by making use of automated quantitative approaches—specifically, frequency counts, key words,

concordances, and key concepts—to examine how the language used by extremist authors reflects their ways of thinking and feeling (see Pennebaker 2002, Taylor & Donald 2004, for successful examples of this approach in other areas).

This approach proved useful in a prior study (Prentice et al. forthcoming) in which we analyzed 50 extremist texts surrounding the Gaza conflict using both corpus and content analyses and found a predominance of moral and social proof arguments, together with a large number of arguments that sought to put pressure on and highlight the commitments of their audience as the conflict heightened. However, as shall be seen from the discussion below, specific techniques (particularly collocations) could also be utilised for the purpose of establishing networks or connections between people and/or places in extremist literature. To the best of our knowledge, the collocation technique has not been used for such a purpose before.

3. Corpus Compilation

We examined a corpus of 250 extremist texts downloaded from open-source websites available in the UK. We focused our analysis on online material because of the significant qualitative differences that are known to exist in literature taken from different media (Gregory & Carroll 1978). Such differences would render conclusions about the corpus difficult because there is no way of determining whether the observed trends are the result of differences in media. The Internet also plays an important role in the continued functioning of extremist groups, as Chen et al. (2004) have argued:

“Terrorists across different jurisdictions heavily utilize modern transportation and communication systems for relocation, propaganda, recruitment, and communication purposes. Thus, addressing issues such as how to trace the dynamic evolution, communication, and movement of terrorist groups across different jurisdictions and how to analyze and predict terrorists’ activities, associations, and threats becomes an urgent and challenging issue.”

(Chen et al. 2004:333).

To be included in our corpus, a text had to conform to three criteria. First, a text had to be written in the English language. The corpus software used in our analysis is not currently capable of processing Arabic language data. Although this is a limitation of our analytical techniques, it is also a conceptually useful criteria, since these texts have presumably been considered important enough to be either translated or written in English in order to reach a wider audience. Second, the text had to explicitly advocate the use of violence. We use this criterion to avoid confounding results with texts from authors who seek only to advocate a strict version of their beliefs. Third, the text had to be written in the first person, therefore avoiding possible third-hand recounting of narratives.

The final corpus of 250 texts contains a total of 453,496 words (a mean of 1,814.0 words per text, Range = 107 – 19,450 words). The texts were written between 1996 and 2009. We included texts over this time span to provide a firm basis on which to track

changes over time. The authors of the texts range from well-known groups and organizations (e.g., Al-Qa’ida, Hamas, and Hezbollah) to individual authors who appear to have no particular group affiliation. There are 15 groups featured in the data set and 90 unaffiliated authors (see Table 1). There were no significant differences in text length across the groups (A Fisher’s Analysis of Variance test with group as the fixed factor and text length as the Dependent Variable was not significant, $F_{(15,249)} = 1.09, ns$).

As can be seen from Table 1, the corpus was designed to contain a range of material that may serve different purposes for our analysis. For example, we have within the corpus texts that are written by authors who report to be ‘core’ Al-Qa’ida, and others who arguably hold a similar religious fundamentalist ideology (Donohue & Taylor 2007). We also have statements from groups with different motivations (e.g., Hezbollah) and statements from groups who are not typically associated with the promotion of violence (e.g., Jama’a Islamia and Supporters of Sharia). We sought this range in our data to help increase generalisability of our findings and reduce the bias that might come from studying a few groups intensely. The overall size of our corpus represents a compromise between the wish to have a corpus large enough for automatic quantitative analysis and the requirements imposed by our three criteria above.

Table 1.
Number of Extremist Texts within the Corpora as a
Function of Group Association

Group/Organization Name	Number of Texts
Abu Sayyaf Group	1
Al-Qa’ida	77
Chechen	6
Fatah	1

Hamas	21
Hezbollah	28
Islamic Army in Iraq	1
Islamic Djamaat of Dagestan Shariat	1
Islamic Jihad Union	1
Islamic State of Iraq	1
Jama'a Islamia	1
National and Islamic Forces	1
Palestinian Authority	16
Taliban	2
Supporters of Shareeah	2
Total for groups	160
Unaffiliated Authors	90
Total	250

4. Methodology

Before we could undertake the analysis of the 250 extremist texts described in section 3, these texts were cleansed of all extraneous information, such as headers and footers, introductions, details on forum posters, and data sources. They were then combined into a single document and uploaded to the corpus software tool WMatrix (Rayson 2008). Once uploaded, the document was automatically passed through the USAS semantic tagger (Rayson et al. 2004), which uses an internal lexicon (Piao et al. 2005) and several word sense disambiguation techniques (Rayson et al. 2004) to assign items (words and multi-word units) to one or more semantic field categories. Figure 1 shows the USAS tagset categories, which are intended to capture a general interpretation of the world.

A General and abstract terms	B The body and the individual	C Arts and crafts	E Emotional actions, states and processes
F Food and farming	G Government and the public domain	H Architecture, buildings, houses and the home	I Money and commerce
K Entertainment, sports and games	L Life and living things	M Movement, location, travel and transport	N Numbers and measurement
O Substances, materials, objects and equipment	P Education	Q Linguistic actions, states and processes	S Social actions, states and processes
T Time	W The world and our environment	X Psychological actions, states and processes	Y Science and technology
Z Names and grammatical words			

Figure 1. The USAS tagset top-level domains

The tagset is arranged into 21 discourse fields, with a further 232 subdivisions (see <http://ucrel.lancs.ac.uk/usas/> for the full hierarchy). For example, the word Islam is assigned to the ‘Social actions, states and processes’ field under the sub-category of ‘Religion and the supernatural’.

While undergoing this process, USAS will inevitably come across items that it does not recognise. These ‘unmatched’ items can be downloaded in list form from WMatrix and manually assigned to appropriate categories by the user. Due to the large number of Arabic terms in the data set, the list of unmatched items in this case was substantial, totalling 11,568 items. We manually matched items occurring 3 times or more (a total of 7,771 items) on the basis that this provided a good coverage of the vocabulary in the dataset (see Piao et al. 2004 for a discussion on lexicon coverage). Using this criterion left 3,797 items unmatched. The matching of items that occurred 3 times or more was a significant undertaking, and has made ready the USAS tagger for future analyses of extremist texts.

Having prepared the data, we used WMatrix to obtain summaries of word frequency, key words, concordances, key concepts and collocations. The word frequency summary in WMatrix lists the ten most frequent word and semantic patterns present in the data set, along with the five most frequently occurring items in those categories. This provides a useful starting point for the analysis. However, in order to establish whether language items were a feature particular to extremist literature, we compared our corpus with two corpora of general English language use, namely, the written and spoken elements of the British National Corpus (BNC) Sampler. Each of these corpora contains approximately 1 million words.

Comparisons were made with both the written and spoken components of this corpus based on our assessment that extremist texts are often written to be read aloud, and thus contain features of both spoken and written discourse. Items of language that emerge as significant in comparisons with both the written and spoken corpora would arguably constitute particularly salient features of Islamic extremist discourse.

Our comparisons produce a list of 'key' words or concepts that characterise the extremist texts. Keyness comparisons produce a list of words or groups of semantically related terms (concepts) that occur significantly more in one corpus relative to another. The results can be shown in visual form as 'clouds', which display up to 100 key items that result from a comparison (see Figures 2, 3, 7 and 8 below). Items are displayed in alphabetical order and 'ranked' by font size, with a larger font size associated with greater significance. The 'clouds' provide a useful summary of the main features of the

texts. In the user interface, these ‘clouds’ are live, meaning that specific items of interest can be clicked on to obtain a list of concordance examples.

The final corpus technique used in this paper is collocation. The use of collocates can confirm observations made from concordance examples. However, we also use this technique to demonstrate how particular people and/or places of interest can be seen to be linked to one another, thus identifying the establishment of networks.

As a further development of this, WMatrix also allows the user to obtain semantic collocations (i.e., to find the concepts that are commonly associated with a specific term or vice versa). This feature, which exploits the semantic tags assigned by the USAS tagger, is used here to provide information on how authors think or feel about particular people or places. This potentially allows the user to identify people and places that may be of interest for further investigation, either because authors see them as influential, or because they are seen as a potential target.

5. Analyses and Discussion

5.1. Word frequency summary

Table 2 presents the most frequent words in our corpus of extremist texts as a function of selected language patterns. Specifically, our analysis identified words with negative connotations (those listed as ‘Negative’), words relating to existence (‘Positive’), words

relating to warfare ('G3' – the tag for the Warfare, defence and the army; weapons concept), and words containing the fragments 'ed', 'ing', 'rabi', 'ila'.

Table 2. The most frequent terms in the extremist literature corpus as a function of WMatrix

Common word/string patterns	WMatrix label	Examples
Terms with negative connotations	Negative	enemy, fight, killed, fighting, other
Terms denoting existence	Positive	is, are, be, all, was
Terms with the word string pattern 'ing'	ing	fighting, killing, during, saying, being
Terms denoting warfare, defence and the army	G3	jihad, war, army, military, forces
Terms with the word string pattern 'ed'	ed	killed, need, asked, indeed, started
Terms with the word string pattern 'rabi'	rabi	arabian, Saudi Arabia, arabic, arabism, Arabia
Terms with the word string pattern 'ila'	ila	similar, Khilafah, similarly, available, annihilate

From the initial summary shown in Table 2, it is interesting to note the lack of frequent content words with positive connotations, whereas there appears to be no shortage of negative items. One possible interpretation of this finding is that authors are focussing their narratives on negative aspects of context and thus have a negatively orientated style of writing. However, that is not to say that authors do not view some of these negatively classified items (such as 'fighting') in a positive way; something that will become apparent through further analysis below.

Although many of the terms in Table 2 relate to the 'war and conflict' context in which the extremist texts are situated, others appear to be identity-related (e.g., 'arabian', 'arabic', 'arabism', 'Arabia' and 'Khilafah'), or otherwise concerned with drawing

parallels (as suggested by terms such as ‘similar’ and ‘similarly’). To investigate these initial observations in more detail, and to see what can be further learned from a corpus analysis, we employed techniques of statistical keyness, concordances and collocations. These techniques allow for data-driven investigations of the language patterns within the corpus of interest.

5.2. Key word clouds

By comparing the corpus of extremist literature with the two BNC Sampler subcorpora, it was possible to establish statistical similarities and differences across the results. This should give an indication of the key words that are particular to extremist literature (i.e., if particular words still appear despite changing the reference corpus then we can say more reliably that they are a feature of extremist narrative). Figures 2 and 3 show the results of comparing the extremist literature corpus with the BNC Written Sampler (Figure 2) and the BNC Spoken Sampler (Figure 3). As stated above, the larger a word appears on these Figures, the greater the differences in its occurrence between our corpus of extremist texts and the BNC corpora. WMatrix uses the log-likelihood statistic to calculate significance or keyness¹. In this analysis, items with a log-likelihood value of 6.63 ($p < 0.01$) or above were deemed to be key items. The ‘clouds’ in Figures 2 and 3 display the top 100 of these key items.

Afghanistan **against** aggression **Allah** alh almighty **America** american americans **and** arab are army banner battle
because believers bin **blessings** blood **brothers** cause companions crusader crusaders enemies enemy evil faith
fight fighting forces Gaza **God** hamas hearts **him** holy honor infidels **Iraq** iraqi **is islam islamic**
Israel **jews** **Jihad** killed killing land martyr martyrdom martyrs **messenger** Muhammad
Mujahideen muslim muslims nation not o occupation of **OUR** Palestine
palestinian paradise path **peace** people praise **prophet** religion resistance rulers sake scholars sheikh swt **that the**
their them they this those truth **Ummah** upon us victory wa war **we who** will world
you zionist

Figure 2. Top 100 keywords of the extremist literature corpus when compared to the
BNC Written Sampler

Afghanistan **against** aggression **Allah** almighty **America** american americans among
and arab are army battle believers bin **blessings** blood **brothers** by cause companions crusaders death
enemies enemy evil fear **fight** fighting forces from Gaza **God** hamas has **him** **his** holy in
in_order_to infidels **Iraq** is **islam islamic** Israel **its** jews **Jihad** killed killing
land leaders Lebanon martyrdom martyrs media **messenger** military Muhammad **Mujahideen**
muslim muslims nation nor occupation **of our** Palestine
palestinian paradise path **peace** people praise **prophet** regime **religion** resistance rulers scholars
soldiers state **the their** them those to **Ummah** upon us victory war which **who** will
with **world** zionist

Figure 3. Top 100 keywords of the extremist literature corpus when compared to the
BNC Spoken Sampler

What is encouraging about these key word clouds is the amount of matching between the results. There are only isolated differences between the two clouds, and those that do exist may be explained by the nature of the corpora being examined. For example, Figure 4 gives the pronouns ‘they’, ‘we’ and ‘you’ as key, whereas Figure 5 does not contain these. This is likely due to the fact that the extremist literature corpus contains texts that are written for oral dissemination. They therefore contain features typical of both written and spoken discourse. One feature of spoken discourse is frequent use of pronouns, as the spoken word tends to be more relational than written discourse (Biber et al. 1999). Thus, when compared with a large corpus of written texts, the relational nature of extremist texts comes to the fore.

Likewise, the words ‘leaders’, ‘Lebanon’ and ‘media’ appear in the top key words when the texts are compared with the spoken BNC Sampler, and although also key when compared with the written element, the same terms do not appear in the top 100 items. This is probably due to the inclusion of media reporting in the written element of the BNC Sampler. These omissions are somewhat important if we view concordance examples of ‘media’ and ‘leaders’, for example. These are shown on Figures 4 and 5.

```

    &lsqb; and &rsqb; those among our  leaders  who sold themselves , and I am not
nd a clearer revolution and better  leaders  who want to defend the oppressed an
> , oh millions of Arabs , oh Arab  leaders  ? What are you going to do in light
ing and watching . One of the Arab  leaders  claims that he opposes the attack c
have been to get rid of the local  leaders  of heresy namely the rulers of Riya

```

Figure 4. Concordance examples of the term ‘leaders’

```

ised questions ! Intelligent use of  media  channels could be more effective tha

```

advantage means trying to use other media channels to convey our message even
emised upon its ability to create a media storm . Kidnappings of journalists ,
sources to own their own successful media devices . Therefore , such groups ha
. For example we could contact the media agency 's staff to convey our viewpo

Figure 5. Concordance examples of the term 'media'

osnia , Kashmir , Chechnya , and Afghanistan . We Have the Right to Kill 4 Mil
the aggressions against Palestine , Iraq , and Afghanistan will be our target
s Everywhere Must Advance to Defend Iraq Another quality of strength is that
sight the mothers of Palestine and Gaza , and we shall avenge them harshly .
the exchange for those killed in Palestine , Somalia , Sudan , the Philippine

Figure 6. Concordance examples of key geographical names

There are at least two things to observe from the concordances in Figures 4 and 5. First, the concordances of 'leaders' suggest a feeling of dissatisfaction, which has been noted as one of the key components driving individuals to violence (Iannaccone & Berman 2006; Ransford 1968; Sandbrook & Romano 2004). Second, the concordances of 'media' give some indication of the extent to which media is perceived as important to authors as a means of furthering their 'cause'. Note that both of these differences emerge only because we used more than one corpus for comparison.

In terms of the similarities between the comparisons, both bring up more or less the same people and places: Afghanistan, America, Americans, Gaza, Iraq, Israel, Palestine, and also, arguably, Lebanon. This suggests that the authors of the texts we have collected hold common 'enemies' in America and Israel. Similarly, they have common causes in Afghanistan, Iraq, Gaza and Palestine. These causes are used by authors as a

means to ‘justify’ violent extremism against their common enemies, as is evident in the concordance examples in Figure 6.

5.3. Key concept clouds

Having considered language at a word level, we now move to examine our texts at the semantic level (i.e., at the level of concepts or themes). Figures 7 and 8 give the top 100 key concepts present in the extremist literature corpus when compared with both the BNC Written Sampler (Figure 7) and the BNC Spoken Sampler (Figure 8). As with the key word analysis, concepts were considered significant when the log-likelihood difference between the two corpora was greater than 6.63 ($p < .01$).



Figure 7. Top 100 key concepts of the extremist literature corpus when compared with the BNC Written Sampler

Table 3. Showing positive/negative semantic pair concepts for the extremist literature dataset

Semantic pair concepts	
darkness and light	those who help and those who hinder
success and failure	those who are intelligent and those with inability/unintelligence
serenity (calm) and violence	those who are strong and those who are weak
life and death	those in power and those who have no power
warfare versus anti-war	those who are affluent and those with no money
contentedness and sadness	what is allowed/permitted and not allowed/permitted
those who are respected and those who have no respect	what is lawful/ethical and what is unethical
those who are selfish and unselfish	what people have an obligation to do and what they are not obligated to do
those who are trying hard and those who are inattentive	what is true and what is false
those who are religious and those who are non-religious	

As Table 3 shows, the comparisons are pervasive. The Islamic extremists featured in this data set appear to view the world in contrasts, and thus could be said to possess a polarised way of thinking.

A second observation is that there is evidence to support the presence of the tactics identified through a review of the literature, specifically those studies concerned with the thematic analysis of extremist messages (i.e. Duffy 2003; Prentice et al. forthcoming). In particular, the theme of Moral Proof/Fairness and Justice (i.e. use of moral arguments or comparisons to justify claims and beliefs) established in such studies continues to suggest itself as the pervasive tactic used by authors of extremist literature. The majority of concepts found in Figures 7 and 8 fit into this category:

Allowed, Anti-war, Crime, Damaging and Destroying, Deserving, Ethical, Helping, Hindering, Inability/unintelligence, Inattentive, Knowledgeable, Law and order, Lawful, No constraint, Non-religious, Not allowed, Personality Traits, Unethical, Violent/Angry, Warfare. The contrasting morality of the concepts listed here (e.g., ‘Ethical’ versus ‘Unethical’) are an indication of authors seeking to reassert a positive in-group morality versus a negative out-group morality, which was a key finding in the Prentice et al. study.

5.4. Collocations of people and places

We now turn to an analysis of collocations (i.e. the most commonly co-occurring terms) in the extremist literature corpus. Our focus here is to view the collocations of people, places or other key items of interest in the extremist literature. For instance, as indicated by both key concept clouds in Figures 7 and 8, Geographical names, Geographical terms and Personal names are significant in the extremist literature corpus. Using the key word clouds, it is also possible to establish specific places and people who are continually referred to.

Table 4 demonstrates such an analysis by showing collocates of the places ‘Afghanistan’, ‘America/n/ns’, ‘Iraq/i/is’, and ‘Israel/i/is’ within the extremist texts. The scores shown alongside each word pair are mutual information scores, which are a measure of the strength of association between two terms. A score of 3 or above is

considered significant (Hunston 2002)². The list in Table 4 is rank-ordered, such that collocates toward the top of the list have a stronger association than those at the bottom.

Table 4. Showing up to the top 50 collocates of ‘Israel/i/is’, ‘America/n/ns’, ‘Afghanistan’, ‘Iraq/i/is’.

15.82	'Greater	Israel'	8.69	American	solders	6.59	Emirate	Afghanistan	8.3	Iraqi	Minister
9.84	Israeli	aircrafts	8.69	recession	American	6.07	Taliban	Afghanistan	7.78	Iraqi	neutral
9.68	Children	Israel	7.88	America	Saud	5.68	Pakistan	Afghanistan	6.95	O	Iraqis
9.21	recognizing	Israel	7.52	America's	defeat	5.59	happened	Afghanistan	6.49	war	'Iraq
8.69	Bani	Israel	7.19	American	economy	5.17	throughout	Afghanistan	6.29	Iraq	link
8.48	Greater	Israel	6.97	America	dream	5.08	campaign	Afghanistan	6.29	invade	Iraq
8.43	recognized	Israel	6.87	American	administration	5.01	problems	Afghanistan	6.08	Iraqi	peoples
8.09	recognition	Israel	6.72	American	bombing	4.96	Iraq	Afghanistan	6.03	invasion	Iraq
8.07	normalization	Israel	6.61	American	goods	4.91	attacks	Afghanistan	5.95	Iraqi	nation
7.46	Israeli	soldier	6.6	Americans	allies	4.86	occupation	Afghanistan	5.85	State	Iraq
7.1	Israeli	tanks	6.6	American	planes	4.84	thousands	Afghanistan	5.72	withdrawal	Iraq
7.01	State	Israel	6.47	personnel	American	4.83	Afghanistan	Iraq	5.59	Iraqi	people
6.72	Israeli	existence	6.38	American	soldier	4.76	Afghanistan	general	5.51	Iraqi	forces
6.48	Israel	withdraw	6.34	American	British	4.64	troops	Afghanistan	5.2	policies	Iraq
6.47	Israeli	alliance	6.33	rear	American	3.62	jihad	Afghanistan	5.06	especially	Iraq
6.28	Israel	continues	6.32	led	America	3.52	today	Afghanistan	5.03	Sunnis	Iraq
6	huge	Israel	6.3	American	policy	3.22	brothers	Afghanistan	4.96	Iraq	Afghanistan
5.83	Israeli	government	6.13	America	Zionism	3.13	Mujahideen	Afghanistan	4.83	Afghanistan	Iraq
5.82	Israeli	soldiers	6.02	overcome	America	3.12	America	Afghanistan	4.71	occupation	Iraq
5.68	Israeli	aggression	6.01	Americans	Afghan	3.1	American	Afghanistan	4.71	Iraqi	political
5.57	inside	Israel	5.95	Americans	British	3.02	Muslims	Afghanistan	4.71	brethren	Iraq
5.43	Israeli	army	5.88	America	collapse	2.96	war	Afghanistan	4.69	Crusader	Iraq
5.28	Israeli	came	5.86	American	policies	2.65	only	Afghanistan	4.68	British	Iraq
4.83	security	Israel	5.86	American	collapse	2.39	Jihad	Afghanistan	4.55	Iraq	liberate
4.73	state	Israel	5.81	American	president	2.24	Muslim	Afghanistan	4.55	raised	Iraq
4.58	Israel	United	5.81	former	American	2.22	against	Afghanistan	4.5	faithful	Iraq
4.58	United	Israel	5.77	American	begun	2.2	Islamic	Afghanistan	4.43	Iraqi	army
4.55	support	Israel	5.77	service	American	1.4	The	Afghanistan	4.38	attack	Iraq
4.29	American	Israeli	5.74	America	Britain				4.26	aggression	Iraq
4.1	Israel	Gaza	5.68	American	forces				4.25	sons	Iraq
3.9	Israeli	Palestine	5.65	American	intelligence				4.19	fighters	Iraq
3.14	Israeli	against	5.64	America	allies				4.06	American	Iraq
2.91	The	Israeli	5.64	American	alliance				3.91	The	Iraqi
2.61	Israel	its	5.64	American	interests				3.73	defeat	Iraq
1.92	Israel	will	5.63	extent	American				3.7	brothers	Iraq
1.69	The	Israel	5.6	American	supporting				3.68	Palestine	Iraq
			5.6	America	House				3.32	people	Iraq
			5.58	American	soldiers				3.2	children	Iraq
			5.57	American	targets				3.18	war	Iraq
			5.54	American	tanks				3.18	What	Iraq
			5.34	alliance	America				3.06	against	Iraq
			5.33	America	itself				3.02	forces	Iraq
			5.27	American	Jewish				3	support	Iraq
			5.24	States	America				3	Americans	Iraq
			5.23	American	foreign				2.91	land	Iraq
			5.19	American	ruler				2.89	America	Iraq
			5.18	strike	America				2.87	Iraq	American
			5.16	kill	Americans				2.84	Islamic	Iraq
			5.14	agents	America				2.81	great	Iraq
			5.14	American	allies				2.73	Iraq	support

It is notable from Table 4 that particular places collocate with one another. For example, Palestine collocates with Iraq and Afghanistan, where parallels are being drawn between the situations in each of the countries. Similarly, the terms frequently co-occurring with ‘Israel’ show a concern with the issue of others’ recognition and support of Israel as a country, for example, recognising, recognized, recognition, normalization and alliance (also note the inverted comma usage around ‘Greater Israel’, suggesting it

is seen by authors as a theoretical concept, rather than a concrete entity). It is also interesting to note the appearance of ‘Children’ as a collocate in relation to ‘Iraq’, which is potentially being used to provoke an emotional response with the audience. Finally, America’s economic situation appears to be a talking point for the authors of extremist literature, with the terms ‘recession’ and ‘economy’ occurring toward the top of the list. There is a sense that authors pinpoint this issue as a sign of weakness, possibly to undermine America’s authority as a world power, in order to convince others that it is possible to defeat it (note also the occurrence of ‘collapse’ as another collocate). Such observations would of course need to be verified through additional analyses.

Another way of approaching collocation is to view the semantic tag collocates of items. Semantic tag collocations give a view of the terms most commonly associated with particular concepts or vice versa. Tables 5-7 show the top 50 significant collocates (with a mutual information score of 3 or above) for the USAS semantic categories Personal Names, Geographical Names and Places, and Groups and Other Proper Nouns.

Table 5: Top 50 significant collocates for the semantic category Personal Names

MI Score	Collocate	MI Score	Collocate
8.46	son	6.58	hazara
7.84	bush	6.58	hayat
7.83	faheem	6.58	haroon
7.07	master	6.58	fath
7.02	mr.	6.58	essay
6.74	prophet	6.58	bir
6.68	illa	6.58	assalam
6.68	s'ad	6.58	aslan
6.68	overcame	6.58	ameer-ul-mumineen
6.68	ilaha	6.58	agency
6.58	tora	6.52	banu
6.58	maulawi	6.5	babri
6.58	qaida	6.49	mullah

6.58	yasser	6.49	tmq
6.58	qari	6.49	ezzedeen
6.58	kofi	6.46	acquiring
6.58	sayyed	6.43	abdur
6.58	salman	6.42	insha
6.58	lal	6.39	surat
6.58	khoja	6.37	concerns
6.58	hector	6.36	ezz
6.58	zaad	6.33	salaah
6.58	yaum	6.32	anticipation
6.58	wali	6.32	abu
6.58	muhi	6.32	shaikh

The significant collocates of Personal Names (Table 5), for example, include other personal names - indicating connections between particular individuals, which could aid the establishment of personal networks. The collocates also include the term ‘overcame’, which is further support for our previous observation (see Prentice et al. forthcoming) that particular individuals are portrayed as positive examples to the audience, while others, as suggested by terms such as ‘agency’, are not. The names associated with threatening or negative terms could potentially help to identify individuals at risk, or those who could be used as effective counter message bases, while names associated with positive terms identify those who are held in high esteem in terrorist circles. Further collocates such as ‘bir’ and ‘surat’ show connections being made between people and places, a point that is further established in the discussion of Table 6 below.

Table 6: Top 50 significant collocates for the semantic categories Places and Geographical Names

Places		Geographical Names	
8.48	enter	6.07	invaded
8.07	stationing	6.07	disappear
8.07	rural	6.07	buried
9.65	western	5.19	northern
9.02	united	5.17	ongoing
6.02	tourism	5.14	north

7.66	web	5.96	domestic	6.02	safavid	5.02	boycotting
7.49	neighbouring	5.93	never	6.02	inseparable	5.02	soviets
7.34	resting	5.83	cities	6.02	ignite	5.02	spark
7.34	holiest	5.66	intend	6.02	four-fifths	5.02	sa'ud
7.34	dictatorial	5.57	sees	5.8	assad	5.02	recognizing
7.21	occupy	5.49	crackdown	5.7	extorted	5.02	mastery
7.07	rear	5.49	christian	5.64	liberating	5.02	killings
7.07	fortified	5.41	surrounding	5.61	jihad.	5.02	governed
					fundamental		
7.07	residential	5.34	internet	5.61	ly	5.02	alleged
6.9	visits	5.31	outside	5.53	nile	4.88	islamia
6.9	nearby	5.27	moved	5.43	mujahed	4.85	steal
6.81	eastern	5.17	safe	5.43	recession	4.85	pilgrimage
6.81	leave	5.16	regimes	5.43	copts	4.8	slaying
6.73	european	5.14	targeting	5.43	collusion	4.8	paved
6.66	rescue	5.11	gulf	5.4	southern	4.8	mash'al
6.49	imposing	5.09	two	5.28	inch	4.8	inability
6.48	holy	5.07	stayed	5.28	idf	4.8	dragged
							commander
6.34	intervention	5.01	retreat	5.28	diyala	4.8	-in-chief
6.2	tribal	4.99	several	5.28	devoured	4.8	annihilated
6.2	streets	4.99	withdraw	5.28	befalling	4.77	bani
6.2	qualified	4.96	preparing	5.28	al-nafisi	4.76	usurped
6.2	interference	4.95	begun	5.22	south	4.76	ambassador

Table 6 shows that a number of collocates used in association with Geographical names are personal names (e.g. al-nafisi), a factor that may assist in establishing terrorists' areas of interest or their location. In addition, the term 'jihad' is a high scoring collocate in relation to Geographical Names, which may suggest potential or existing terrorist targets or training grounds. There is also an indication of the strategic associations extremist authors make with regard to particular places, i.e. terms suggesting terrorist movements (for instance, 'enter', 'stationing', and 'moved'). Other notable collocates are those that provide a positive or negative evaluation of a place or geographical location, for instance, 'dictatorial'. This may be used to identify enemy from in-group locations. Negative or violent associations could also highlight potential targets, as they indicate places towards which contempt is felt by extremist authors.

Whilst potentially identifying the movement, location and evaluation of individuals and places within extremist discourse is of interest, terrorists usually function within groups, and therefore group associations are also important. With this in mind, Table 7 shows the collocates for Groups and Other Proper Nouns. Other Proper Nouns are included here as they contain organizational names, and over a third of the occurrences in this semantic category are made up by the term ‘mujahideen’ (i.e. guerrilla fighters), which accounts for 566 out of 1,594 items. This is important when we consider that many of the collocates are individuals’ names, suggesting that particular individuals are being connected to terrorist groups and organizations.

Table 7: Top 50 significant collocates for the semantic categories Other Proper Nouns and Groups

Other proper nouns				Groups			
8.76	abubakar	8.06	abdul	10.02	bring	6.78	pious
8.76	qari	8.03	al-din	9.71	posted	6.71	local
8.76	tahir	7.76	mulla	9.14	hezbi- wahdat	6.7	tied
8.76	lal	7.76	rahman	8.82	damned	6.7	axis
8.76	krishan	7.76	abdur	8.8	al-jihad	6.64	malicious
8.76	sulaiman	7.63	persian	8.65	qaida	6.51	crusader- zionist
8.76	soubhanahu	7.59	maulawi	8.28	zio- american	6.5	often
8.76	rahmatu	7.54	abd	8.28	rotary	6.41	usurping
8.76	muhi	7.54	srinagar	8.28	bravest	6.41	tribal
8.76	ignite	7.54	rahmatullah	8.02	jihadic	6.4	largest
8.76	essay	7.54	professor	7.96	inspection	6.4	foundations
8.76	dokka	7.33	zakah	7.92	educational	6.36	link
8.76	assalam	7.18	affan	7.78	disunity	6.34	individuals
8.76	ameer-ul- mumineen	7.09	shaykh	7.7	crusader- zionistic	6.22	small
8.76	alaikom	7.08	northern	7.66	develop	6.21	jall
8.53	soviet	7.03	abdullah	7.54	hundreds	6.18	international

8.44	el-dean	6.89	zainab	7.44	section	6.09	corrupt
8.35	sayyed	6.89	hasan	7.41	zionist- crusader	6.06	pledge
8.35	hector	6.89	hafiz	7.32	al-qaida	6.01	social
8.28	ezz	6.87	member	7.28	swearing	6	advise
8.18	shaikh	6.8	umar	7.28	satanic	5.82	striking
8.18	saif-ur- rahman	6.76	mus'ab	7.26	shia	5.82	russian
8.18	ben	6.76	ambassador	6.97	family	5.78	guerilla
8.11	mustafa	6.71	heroic	6.92	belonging	5.78	assure
8.09	tmq	6.68	sayyid	6.9	awareness	5.75	facing

In addition, terrorist groups also appear to be connected with one another (note the occurrence of 'al-qaida' as a group collocate), which may aid the establishment of group networks. However, this result may be a feature of the fact that the majority of the texts featured in the corpus are written by Al-Qa'ida members and therefore such members may be keen to draw comparisons between themselves and other groups in order to gain support from individuals sympathetic towards those groups.

An additional finding from the collocates listed in Table 7 is that Groups/the mujahideen are described as being 'heroic', 'pious' and the 'bravest'. Consistent with observations made in a previous study (see Prentice et al. forthcoming), this suggests that they are held up in extremist rhetoric as examples to others. Building on this, the presence of the collocates 'educational' and 'family', which are arguably words that point to the positive moral evaluations made by authors when describing their group, indicate that it is not only individuals who are presented to others as a positive influence, but also the groups with which those individuals are involved.

The final collocates that are of interest in Table 7 are those that suggest group location, i.e. 'northern'; identity, i.e. 'persian' and 'shia'; communications, i.e. 'essay' and

‘posted’; or targets i.e. ‘srinagar’ (referring to the Srinagar airport attack). As with the other observations made in this section, both at the word and semantic level, such connections may be of importance to those investigating groups identified as posing a potential threat or risk to others.

5.5. Summary

The analysis and discussion in this paper has shown that the combination of data-driven corpus techniques is essential in order to produce a well-rounded overall view of the data set. Indeed, by looking at the data through a variety of techniques, we have achieved consistent results, not only throughout the analyses in this paper, but also when compared with the results of related studies, which demonstrates the robustness of our findings. However, our discussion has highlighted the importance of not taking initial results at face value. Although the word frequency counts suggested that extremist literature is generally negative, a thorough analysis revealed that it is in fact more bi-polar in nature – a finding that may add to the evidence base on which effective counter-terrorism strategies are built.

6. Conclusion

The techniques demonstrated in this report show only a fraction of how corpus linguistic analyses could reveal something of the beliefs and motivations behind authors

of extremist texts. Other methods, such as range and dispersion (used to view the extent to which features are distributed across texts), stylistic analysis (investigating the extent to which authors' writing styles differ), and text re-use (examining the repeated use of text fragments) could also prove useful in the evaluation of extremist material and would be interesting to trial in future work. In addition, if a sufficient amount of data were available, it should be possible to extend the key word and concept analyses by performing cross-group comparisons, which may allow more refined conclusions.

With these findings aside, it should be noted that the methodology used in this paper is not without its limitations. The corpus tools only present the patterns in the data, leaving interpretation ultimately in the hands of the user. Although subjective judgements introduce the issue of researcher bias, the automated approaches used here mean that the initial results can be replicated, and therefore the likelihood that similar observations will be made is increased.

In terms of the taggers, generally speaking USAS has a low rate of error: 8.95%. However, this will have been somewhat increased due to the nature of the extremist material, i.e. its high prevalence of Arabic terms that the tagger will be unfamiliar with, hence the system's production of a large number of unmatched items. Although unmatched items were assigned to appropriate concepts to reduce the error rate, this was a somewhat time consuming process. Despite matching the majority of the unmatched items, we were still left with nearly 4,000 words unmatched, which may have affected the results. We can be sure that everything we found was there to find, but not that everything that could be found was in fact found.

Overall the analysis presented in this paper highlights the value gained from combining quantitative techniques with more qualitative approaches, such as the examination of concordance examples. The combination of frequency counts, key word and concept analyses and collocation techniques has allowed us to take a varied approach to the data, and provided us with some consistent observations, with the corpus analysis confirming the presence of all the influence tactics identified in previous manual thematic studies of extremist literature. This opens the possibility of a scalable technique which can be applied to a much larger dataset without losing the advantages of qualitative analysis. One other significant point to draw from the results was the observation of a ‘rhetoric-of-antithesis’ in the texts, with authors presenting a somewhat polarised account to encourage the use of violence. Such findings may inform future analyses of extremist material.

Acknowledgements

We would like to thank Andrew Hoskins (Department of Sociology, University of Warwick) and Ben O’Loughlin (New Political Communication Unit, Royal Holloway) who helpfully supplied some of the data used in this study.

Notes

¹ The log-likelihood statistic measures significance using the observed relative frequency of occurrence of an item in two corpora, and the item's expected frequency of occurrence across the corpora. For further details see

<http://ucrel.lancs.ac.uk/llwizard.html>

² Mutual information is established by dividing the observed frequency by the expected frequency, 'converted to a base-2 logarithm' (Hunston 2002:70-71).

References

Baker, P. 2010. "Representations of Islam in British broadsheet and tabloid newspapers 1999-2005". *Journal of Language and Politics*, forthcoming.

Baker, P. and T. McEnery. 2005. "A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts". *Journal of Language and Politics*, 4 (2), 197-226.

Baker, P., R. Wodak, G. Gabrielatos, M. Khosravinik, M. Krzyzanowski, & T. McEnery. 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 and Society*, 19 (3), 273-306.

Barber, S. 2002. "The formation of cultural attitudes: The example of the three kingdoms in the 1650s". In A. I. Macinnes & J. Ohlmeyer (Eds.), *The Stuart Kingdoms in the Seventeenth Century*. Dublin: Four Courts Press, 169-185.

Becker, A. 2007. "Between "us" and "them": Two interviews with German chancellor Gerhard Schroder in the run-up to the Iraq war". In A. Hodges & C. Nilep (Eds.), *Discourse, War and Terrorism*. Amsterdam and Philadelphia: John Benjamins, 161-184.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *The Longman grammar of spoken and written English*. London: Longman

Bowden, Z. A. 2008. "Poriadok & Bardak (Order and Chaos): The neo-fascist project of articulating a Russian "People"". *Journal of Language and Politics*, 7 (2), 231-347.

Brindle, A. 2009. "Just keep your pants on and in the closet and you'll be fine." The corpus analysis of a white supremacist web forum". Presentation given at the Corpus Research Seminar, Lancaster University. Available at <http://www.ling.lancs.ac.uk/groups/crg/archive.html>.

Chertoff, M. 2008. "The Ideology of Terrorism: Radicalism Revisited". *Brown Journal of World Affairs*, 15 (1), 11-20.

Collins, J. & Glover, R. 2002. *Collateral Language*. New York and London: New York University Press.

Chen, H., Wang, F-Y. & Zeng, D. 2004. "Intelligence and Security Informatics for Homeland Security: Information, Communication, and Transportation". *IEEE Transactions on Intelligent Transportation Systems*, 5 (4), 329-341.

Donohue, W. A., & Taylor, P. J. 2007. "Role effects in negotiation: The one-down phenomenon". *Negotiation Journal*, 23, 307-331.

Duffy, M. E. 2003. "Web of Hate: a Fantasy Theme Analysis of Rhetorical Vision of Hate Groups Online". *Journal of Communication Inquiry*, 27 (3), 291-312.

Every, D. & Augoustinos, M. 2007. "Construction of racism in the Australian parliamentary debates on asylum seekers". *Discourse & Society*, 18 (4), 411-436.

Flowerdew, J., Li, D. C. S. & Tran, S. 2002. "Discriminatory news discourse: some Hong Kong data". *Discourse & Society*, 13 (3), 319-345.

Frey, H. 2007. "Paul Serant and the extreme right's rhetoric of antithesis". *Journal of European Studies*, 37 (4), 373-389.

Geden, O. 2005. "The discursive representation of masculinity in the Freedom Party of Austria (FPO)". *Journal of Language and Politics*, 4 (3), 397-420.

Gregory, M. & Carroll, S. 1978. *Language and situation: Language varieties and their social contexts*. London: Routledge.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Iannaccone, L. R. & Berman, E. 2006. "Religious extremism: The good, the bad, and the deadly". *Public Choice*, 128, 109-129.

Johnson, S. & Suhr, S. 2003. "From 'Political Correctness' to 'Politische Korrektheit': Discourses of 'PC' in the German Newspaper, Die Welt". *Discourse & Society*, 14 (1), 49-68.

Lazar, A. & Lazar, M. M. 2004. "The discourse of the New World Order: 'out-casting' the double face of threat". *Discourse & Society*, 15 (2-3), 223-242.

Lazar, A. & Lazar, M. M. 2007. "Enforcing justice, justifying force: America's justification of violence in the New World Order". In A. Hodges & C. Nilep (Eds.), *Discourse, War and Terrorism*. Amsterdam and Philadelphia: John Benjamins, 45-66.

Martin, P. & Phelan, S. 2002. "Representing Islam in the Wake of September 11: A Comparison of US Television and CNN Online Message board Discourses". *Prometheus*, 20 (3), 263-269.

Pennebaker, J. 2002. "What our words can say about us: Towards a broader language psychology". *Psychological Science Agenda*, 15, 8-9.

Piao, Scott S. L., Rayson, P., Archer, D. & McEnery, T. 2004. "Evaluating Lexical Resources for A Semantic Tagger". *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 2, 499-502. Available at: http://www.comp.lancs.ac.uk/computing/users/paul/publications/pram_lrec04.pdf

Piao, S. L., Archer, D., Mudraya, O., Rayson, P. Garside, R., McEnery, T., & Wilson, A. 2005. "A large semantic lexicon for corpus annotation". *Proceedings from the Corpus Linguistics Conference Series, 1*, ISSN 1747-9398.

Prentice, S., Taylor, P. J., Rayson, P., Hoskins, A. & O'Loughlin, B. forthcoming. "Analyzing the Semantic Content and Persuasive Composition of Extremist Media: A Case Study of Texts Produced during the Gaza Conflict". *Information System Frontiers*, Special Issue on Terrorism Informatics.

Ransford, H. E. 1968. "Isolation, Powerlessness, and Violence: A Study of Attitudes and Participation in the Watts Riot". *The American Journal of Sociology*, 73, 581-591.

Rayson, P. 2008. "From key words to key semantic domains". *International Journal of Corpus Linguistics*, 13 (4), 519-549.

Rayson, P, Archer, D, Piao, S. & McEnery, T. 2004. "The UCREL semantic analysis system". *Proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 7-12.

Reisigl, M. & Wodak, R. 2001. *Discourse and Discrimination: Rhetorics of racism and anti-Semitism*. London and New York: Routledge.

Renold, K. 2002. "Fundamentalism". In J. Collins & R. Glover (Eds.), *Collateral Language*. New York and London: New York University Press, 94-108.

Sandbrook, R. & Romano, D. 2004. "Globalisation, extremism and violence in poor countries". *Third World Quarterly*, 25 (6), 1007-1030.

Simon-Vandenberghe, A-M. 2008. "Those Are Only Slogans": A Linguistic Analysis of Argumentation in Debates With Extremist Political Speakers". *Journal of Language and Social Psychology*, 27 (4), 345-358.

Smith, A. 2007. "Words Make Worlds: Terrorism and Language". *FBI Law Enforcement Bulletin*, 12-18.

Stenvall, M. 2007. "'Fear of terror attack persists': Constructing fear in reports on terrorism by international news agencies". In A. Hodges & C. Nilep (Eds.), *Discourse, War and Terrorism*. Amsterdam and Philadelphia: John Benjamins, 205-222.

Stoltz, G. I. 2007. "Arabs in the morning paper: A case of shifting identity". In A. Hodges & C. Nilep (Eds.), *Discourse, War and Terrorism*. Amsterdam and Philadelphia: John Benjamins, 105-122.

Stout, M. 2009. In Search of Salafi Jihadist Strategic Thought: Mining the Words of the Terrorists. *Studies in Conflict Terrorism*, 32:10, 876-892.

Taylor, P. J., & Donald, I. J. (2004). The structure of communication behavior in simulated and actual crisis negotiations. *Human Communication Research*, 30, 443-478.

Testa, A. & Armstrong, G. 2009. "Words and actions: Italian ultras and neo-fascism". *Social Identities*, 14 (4), 473-490.

Van der Valk, I. 2003. "Right-wing parliamentary discourse on immigration in France". *Discourse & Society*, 14 (3), 309-348.