From keywords to discursive legitimation: Representing 'kuffar' in jihadist propaganda magazines

Anina L. Kinzel
College of Arts and Humanities
Swansea University
Singleton Park
Swansea, Wales
SA2 8PP
795737@swansea.ac.uk

Abstract

This paper explores the integration of Corpus Linguistics and Discourse Analysis methodologies, specifically the use of keyword analysis to pinpoint salient discourse representations of given social identities. The particular identity examined here is that of 'kuffar', which is a derogatory term used by jihadist-ideology groups to describe people who do not share their faith. Two measures (Log Likelihood and %DIFF) are used to determine the keyness value of the term kuffar in the propaganda magazines produced by two such groups: Al Qaeda (*Inspire*) and Islamic State (*Dabiq*). Although they yield different keyness values, both confirm the salience of this social identity within jihadist propaganda. The results of a Key Word in Context concordance analysis of the term kuffar show how negative representations thereof are mainly legitimized on impersonal authority grounds (Van Leeuwen, 2007). The results also reveal a trend towards reminding Muslims of their "duty" to fight and kill kuffar individuals. This is supported by positive expressions to describe Muslims who perform this duty, and negative expressions and traits associated with the non-believers. This case study supports the view that CL and Discourse Analysis can offer a 'useful synergy' (Baker et al. 2008, 2015) to research into, amongst other areas, representation and ideology in language.

1. Introduction

The research in this paper was carried out as part of a research project funded by Swansea University and directed by professors Stuart Macdonald and Nuria Lorenzo-Dus. The project examined jihadist propaganda from different disciplinary perspectives, including Terrorism Studies and Linguistics. All available online publications of five jihadist magazines released between January 2009 and the end of June 2015 were collected. These publications are *Jihad Recollections*, *Gaidi Mtaani*, *Inspire*, *Dabiq* and *Azan*. To be included in the data set, publications had to be written in English and meet the project's definition of a magazine (Macdonald et al., 2015). The reason this paper focuses on *Dabiq* and *Inspire*, produced by IS and Al-Qaeda, respectively, is that these two groups currently pose the biggest terrorist threat to the Western world. Terrorist groups are also often regarded as acting the same, especially if they follow a jihadist ideology. However, they might target different social groups and compete for recruits in fighting a common enemy – usually an out-group who does not share their ideology and faith. Comparative research within these groups is therefore most useful to understand their similarities and differences. The aim of this paper is twofold: (1) to contribute to the current academic debate about one of the main analytic tools in Corpus Assisted Discourse Studies (CADS): keywords; and (2) to show how IS and Al-Qaeda discursively construct "kuffar" in their propaganda magazines and legitimize this as a negative social identity. "Kuffar" is a derogatory term, which describes individuals on religious grounds, or lack thereof. Analyzing this term provides insights into the role of religion in jihadist ideologies, which is still heavily debated.

1.1 CADS

Corpus-Assisted Discourse Studies (CADS) is the application of Corpus Linguistics (CL) tools such as keywords, wordlist, and concordance line analysis in Discourse Studies. The field has been evolving since the 1990s, when new Corpus Linguistics tools were developed. Until 2004 not many studies used a CADS approach (Partington, 2004). Partington describes CADS as 'the uncovering, in the discourse type under study, of what we might call *non-obvious meaning*' (Partington, 2008, p. 191). It combines the statistical techniques of CL, such as keyword and frequency lists and concordancing, with the qualitative tools of Discourse Analysis, namely close-readings and reflection (Partington, 2008) to uncover non-obvious meaning. Baker 2015 argues

that using a CADS approach, especially with a large corpus, removes some of the bias a DA researcher may have. It also gives DA researchers a starting point: a CADS methodology may drive the analysis in 'unexpected directions' (Baker, 2015, p. 144). In a 2008 journal article, Baker et al. conclude that using Corpus Linguistics tool and Critical Discourse Analysis is 'a useful methodological synergy' (Baker, et al., 2008, p. 273). Some examples of work that uses a CADS approach to examine how Muslims are constructed by various media are as follows: Baker et al. use Sketch Engine, a corpus analysis tool for grammatically tagging items, and other CL techniques to examine how collocates of the word 'Muslim' is constructed in British Newspapers articles between 1998 and 2008 (Baker et al., 2013). In their analysis they focus on the two most frequent immediate right-hand collocates "world" and "community" (Baker et al., 2013). Their findings include that Muslims are often associated with negative aspects, they are also portrayed as being easily offended and in conflict with non-Muslim communities (Baker et al., 2013). Muslim world was often referring to different branches of Islam as one while being constructed on language or geographical grounds, rather than religious grounds (Baker et al., 2013). McEnery et al. also look at the word "Muslim" as part of their analysis of media reactions to the murder of Private Lee Rigby by Michael Adebolajo and Michael Adebowale, who converted to Islam (McEnery et al., 2015). The findings showed that the word is associated with the murderers and the victim, but is used to distance this action from other Muslims. Islam as a keyword is associated with negative aspects such as betrayal, radicalization and extremism (McEnery et al., 2015).

1.2 Keyword analysis

Keyword analysis is a much used technique of CL (Gabrielatos and Marchi, 2011). Bondi notes that '[T]he study of keywords has become central in corpus linguistics, especially through the development of techniques for the analysis of the meaning of words in context' (Bondi, 2010, p. 3). In the broadest sense, keywords are words that are important in a given text (Stubbs, 2010). They mirror what the text is about (Scott and Tribble, 2006), which is why they are an important tool of CL to help identify a subset of textual items to analyze (Baker, 2006a). Keyword lists in CL tools such as AntConc (Anthony, 2016) and WordSmith (Scott, 2016a) are based on statistical significance compared to a usually larger reference corpus. This is done by first compiling two wordlists of the two corpora. Words with a high 'keyness' appear in the given text more frequently than expected. Although there are different statistical techniques to determine the 'keyness' of words, log-likelihood (Baker et al 2006b) is arguably the standard one. It is also a technique that has received some criticism, leading to the CL / CADS research community seeking to develop alternative techniques. One such technique, which is compared in this paper to Log Likelihood, is called %DIFF. Introduced by Gabrielatos and Marchi (2011, 2012), its main difference lies in that it is based on effect size

1.2.1 Log-Likelihood

Log-likelihood is a test that shows statistical significance (Baker et al., 2006b). It assigns a *p*-value to every word and measures how significant the word is in comparison to a reference corpus. According to Biber et al. the *p*-value 'represents the probability that this keyness is accidental' (Biber, et al., 2007, p. 138). Log-Likelihood does not take into consideration the size of the study corpus, which means it does not reflect the definition of keywords (Gabrielatos and Marchi, 2011). If an English language magazine, which also features some Arabic words was compared to an English language corpus, such as the British National Corpus (BNC) (British National Corpus, 2007), all Arabic words would be assigned a low *p* –value, because they appear much more frequently in this magazine. However, this does not mean that all of these words are especially *key* in this corpus or indicate aboutness, these words are just different compared to the BNC.

1.2.2 Effect size (%DIFF)

Effect size measures 'the difference of the frequency of a word in a study corpus when compared to that in the reference corpus' (Gabrielatos and Marchi, 2011). Contrasting to log-likelihood, the size of the two corpora is also taken into account, which is especially useful when comparing corpora of different sizes. In this case, the reference corpus does not have to be larger than the study corpus.

1.3 Structure

This paper will first compare the results of a keyword analysis of the corpus (see 3.1), focusing

on the term 'kuffar'[1], by applying log-likelihood and %DIFF measures. It will then report the results of a KWIC analysis for "kuffar" that shows how its negative discourse representation is legitimized by the magazine authors.

3. Methodology

3.1 Data

The corpus for this study consist of 22 issues of *Dabiq* (185,951 words) and *Inspire (304, 347 words)*. 13 of these issues are from *Inspire* and were published between the summer of 2010 and the spring of 2015. The remaining nine issues were published by IS between the summer of 2014 and the spring of 2015. The digital versions of the issues were collected in the summer of 2015, coinciding with the onset of the project.

3.2 Procedure / Framework

Aim (1) entailed calculating keywords using Log-likelihood and %DIFF (see section 4 and Appendix A and B for details). In order to further our understanding of how "kuffar" is used and constructed as a social identity for aim (2), it was necessary to carry out a Key Word in Context (KWIC) analysis, which requires manual examination of concordance lines generated by AntConc (Anthony, 2016) containing those KWIC in *Dabiq* and *Inspire*. As aim (2) was specifically concerned with the discursive justification for the negative construction of 'kuffar' identity in jihadist ideology, an established discourse analytic framework for the analysis of legitimation was followed, namely Van Leeuwen's (2007) legitimation framework. This framework consists of four key categories: Authorization, moral evaluation, rationalization and mythopoesis. Authorization, which this analysis will focus on for space reasons, is subdivided into six self-explanatory sub-categories: These are personal authority, expert authority, role model authority, impersonal authority, the authority of tradition and the authority of conformity. The reason this analysis focuses on impersonal authority is that it is the most frequently used strategy to legitimize the term "kuffar" in *Dabiq* and *Inspire*[2]. Two coders, the author of this paper and one of the principal investigators of the project, individually coded each of the concordance lines. Differences in coding were resolved through discussion. 12% of the concordance lines in both *Dabiq* and *Inspire* could not be assigned to any legitimation category: they were general instances of "kuffar" in which the derogatory term was not used alongside any legitimating behavior, such as rulings on fighting against and killing "kuffar".

The corpus was not lemmatized because one of the project's aims was to examine stylistic variation, including at the spelling level, between the different magazines, especially regarding use of Arabic terms.

4. Results

In CADS, the analysis usually starts with a keyword analysis. As can be seen in Tables 1 and 2, the log-likelihood for "kuffār" is the same in both tables, even though in the first *Dabiq* is the study corpus, and in the second *Inspire* is the study corpus.

word	frequencies		LL	%DIFF
	SC	RC		
kuffār	28	125	123.45	-86.31
kuffar	47	2	31.99	1335.81

Table 1: Log-Likelihood and %DIFF for "kuffar" and "kuffar" with Inspire as study corpus

The highest %DIFF with *Inspire* as the study corpus was the word inspire with a %DIFF of 28738.42, however since this is the title of the magazine as well as a verb, this is not surprising and does not reflect what the text is about. It appeared once in *Dabiq* probably as a verb. The word "kuffar" is 30th in the list with a %DIFF of 1335.81. The word is only used twice in *Dabiq*, which suggests that this is the preferred spelling in *Inspire*. This is an interesting finding, considering kuffār is only used 28 times in *Inspire* and 125 times in *Dabiq*. Its %DIFF is also not very high, compared to the 100th highest, which still has a %DIFF of 342.96.

Note that if two or more words have the same %DIFF, they are all the top Xth word, for example the words "road", "verily", "al-bukhari", "depend", "dry", "perspective", "proceed", "race" and "reveal" all have the same %DIFF of 358.24 and are therefore all the top 95th word. These words also happen to have a Log-likelihood of 5.87. According to this measure, they are not very relevant to the corpus. Similarly, "kuffar" would

[1] This term is the Arabic equivalent to "disbelievers" and is used to describe people in the West who do not share the same faith as IS and Al-Qaeda

[2] For further details see a forthcoming paper by the author

not have been in the top 100 keywords, if the keywords had been measured by log-likelihood; it has a LL of 31.99 as can be seen in Table 1.

In table 2 the %DIFF and Log-Likelihood of the words "kuffār" and "kāfir"[3] are shown. "kuffār" is not in the Top 100 keywords, it has a %DIFF of 630.67, but it is not significant in *Dabiq*, since it is also used in *Inspire*. The top keyword in *Dabiq* is "shām", which is a region in Syria near the city of Damascus, with a %DIFF of 28542.34. This is logical, because IS are in Syria fighting a war for territory among other things. This word is only used once in *Inspire*, which could be because it is not that important to Al-Qaeda, or because there are other spellings they use.

word	frequencies		LL	%DIFF
	SC	RC		
kuffār	125	28	123.45	630.67
kāfir	53	3	82.23	2791.51

Table 2: Log-Likelihood and %DIFF for "kuffar" and "kāfir" with Dabiq as study corpus

"kāfir" is the 30th keyword, with a %DIFF of 2791.51 and a log-likelihood of 82.23. It probably would not have been in the top 30 keywords if they had been compiled using log-likelihood.

Two words with a very high log-likelihood are "wa" with 707.35 and "abū" with 641.25. Abū means "father of", and is often used as part of a name. "Wa" is part of the phrase "salla llāhu ʿalayhi wa-alehe wa-sallam", which translates to "peace be upon him" and is used immediately after mentioning the prophet Mohammad. Both terms are therefore not what the two texts are about, they are frequently used as part of or following names. They cannot be considered keywords, although they would be likely at the top of the list according to Log-Likelihood.

Keywords indicate what the text is about. Despite the different keyness values, "kuffar" emerges as a salient keyword using both methods. Other keywords mentioned in this section did not occur as keywords using both methods. It is worth pursuing other case studies to determine advantages and disadvantages of both methods. This is especially relevant to CADS Researchers, who usually only analyze the top keywords. It is important to ensure that these words are indeed the most important ones in the given corpus (Gabrielatos and Marchi, 2011).

The paper will now move on to the aforementioned KWIC analysis. In the context of terrorist magazines, impersonal authority is defined as follows: Legitimation through laws, rules and regulations (such as laws of Islam, the Koran, scriptures) or unwritten guidelines made by the terrorist groups. The following are two typical examples of impersonal authority as legitimation in *Inspire*:

> Every Muslim who carries the creed of pure Tawheed, alliance to Muslims and disavowal from kuffār and lives in the West has a duty to find a suitable method to severely kill the kuffār and waylay[4] them at every place of ambush until they cease killing Muslim women and their children. (*Inspire*, issue 13)

> We are proud that every muslim [sic] believes in the Islamic unity and waging jihad against kuffar who occupy our lands. (*Inspire*, issue 10)

This example is a general appeal to Muslims living in the West, who identify with a jihadist ideology such as the one Al-Qaeda puts forward, to carry out their duty to kill non-believers (kuffar). This example shows that *Inspire* uses the words "duty" to remind Muslims what they have to do. The example uses the oneness of Allah (Tawheed) as a reminder and it might want to inflict guilt on Muslims who do not follow this "duty". In fact, it attempts to do this by explicitly stating that in order to belong to the terrorist group, one must not associate with non-believers. The impersonal authority in this case comes from Al-Qaeda, who have ruled that Muslims living in the West have to kill kuffar.

The second example shows the unity and sense of community one can expect when joining Al-Qaeda. It expresses pride in waging war (jihad) against the non-believers) and makes joining the fight sound quite appealing to Muslims. This is a softer approach to reminding Muslims of their duty, playing on the fact that Muslims who share Al-Qaeda's ideology might feel isolated and discriminated against and long to belong to a group of like-minded individuals. This example portrays the terrorist very positively and the kuffar as negatively, especially by describing them as "occupy[ing]" their lands. The impersonal authority is again expressed by Al-Qaeda based on their ideology.

[3] "kāfir" is another spelling for the term "kuffar"

[4] stop

The analysis will now focus on *Dabiq*. These are typical examples of legitimation through impersonal authority in *Dabiq*:

> The Muslims in the West will quickly find themselves between one of two choices, they either apostatize and adopt the kufrī religion propagated by Bush, Obama, Blair, Cameron, Sarkozy, and Hollande in the name of Islam so as to live amongst the kuffār without hardship, or they perform hijrah to the Islamic State and thereby escape persecution from the crusader governments and citizens. (*Dabiq*, issue 7)

> A halt of war between the Muslims and the kuffār can never be permanent, as war against the kuffār is the default obligation upon the Muslims only to be temporarily halted by truce for a greater shar'ī interest, as in the offer of truce from the Prophet (sallallāhu 'alayhi wa sallam) to the mushrikīn of Makkah in Hudaybiyah. (*Dabiq*, issue 8)

> But due to the deviance and arrogance of the enemies of Allah, they plot against His religion and His allies. Their plots almost cause the mountains to collapse out of shock that the kuffār dare to oppose the Lord of the heavens and the earth. (*Dabiq*, issue 9)

The first example describes the two choices Muslims living in the West have. They can adopt Islam the way non-believers spread it and not face any problems from the government. The other option is to join the Islamic State and flee pursuit from Western governments. The words "apostatize" and "kufrī religion" imply that this is not the option Muslims should choose. The impersonal authority is expressed by IS, which is explicitly mentioned in the text.

The second example describes the fact that according to *Dabiq*, Muslims are always at war with the non-believers (kuffar). It emphasizes their obligation to fight the kuffar. A cease-fire can only be temporary, which happened in a historic situation that is outlined in the example, involving the Prophet, which indicates a second layer of legitimation. This example supports the call to join IS in their fight, which is outlined in the previous example.

In the third example, kuffar are associated with being deviant and arrogant and it is explicitly stated that they oppose Allah, which might inflict fear in Muslims who are living in the West or have not joined IS's jihad. All three examples show that IS portray their jihad as the only true way to obey Allah and fight the non-believers, who oppose him.

4. Conclusion

This paper has compared %DIFF with the traditional technique Log-Likelihood and pointed out that kuffar emerges as a salient term using both methods. In this case, the method did not change the salience of the keyword in question. This paper did not seek to prescribe to use one method over the other, it simply presented %DIFF as an appropriate technique to select keywords for further analysis, especially to compare corpora of different sizes, since it takes into account the size of both corpora, unlike Log-Likelihood. Log-Likelihood will likely continue to be the predominantly used method. However, the newest version of WordSmith (WordSmith 7.0) includes %DIFF as an option of "statistical tests" to calculate keywords (Scott, 2016b). Choosing which statistics to use is an important decision a researcher has to make in the process of an analysis. More research using %DIFF as a method of identifying keywords is needed as it is important to validate new emerging methods.

The paper then analysed the term "kuffar" in the terrorist propaganda magazines *Inspire* and *Dabiq*. The findings include a trend towards reminding Muslims of their "duty" or "obligation" to fight and kill non-believers (kuffar). This is supported by positive expressions and notions to describe Muslims who perform this duty, and negative expressions and traits, such as being deviant, arrogant and occupying lands, associated with the non-believers. Al-Qaeda portrayed making hijra (travelling to the territory of war) and joining them as something to take pride in. They also imply a sense of belonging in their propaganda, which may speak to individuals who feel isolated. Their propaganda seems welcoming. Contrastingly, IS focus on the negative traits associated with non-believers and emphasise that kuffar and Muslims can never be at peace. They draw on historical examples of the Messenger and Allah to legitimise their discourse and evoke fear and guilt in Muslims who have not joined their jihad.

To conclude, Corpus linguistics techniques are useful tools to apply in discourse analysis to narrow the text down to a few areas of interest. The CADS approach combines quantitative and qualitative analyses as outlined above, which ensures that the findings are reliable.

Acknowledgments

Many thanks go to Nuria Lorenzo-Dus for the support, expertise and comments that greatly improved this paper. I'd also like to thank the Cyberterrorism Project and Swansea University. Thanks also go to Matteo Di Christofaro for insights into effect size (%DIFF).

References

Alan Partington. (2004). Corpora and discourse, a most congruous beast. In Alan Partington, John Morley and Louann Haarman (Eds.), *Corpora and Discourse* (pp. 9-18). Bern: Peter Lang.

Alan Partington, (2008). The armchair and the machine: Corpus-Assisted Discourse Studies. In Carol T. Torsello, Katherine Ackerley and Eric Castello (Eds.), *Corpora for University Language Teachers* (pp. 189 - 213). Oxford: Peter Lang.

British National Corpus. (2007). *What is the BNC?* Retrieved from http://www.natcorp.ox.ac.uk/: http://www.natcorp.ox.ac.uk/corpus/index.xml

Costas Gabrielatos and Anna Marchi. (2011). Keyness: Matching metrics to definitions. *Theoretical-methodological challenges in corpus approaches to discourse studies - and some ways of addressing them.* Portsmouth: 5th November. [conference slides]

Costas Gabrielatos and Anna Marchi. (2012). Keyness: Appropriate metrics and practical issues. *CADS International Conference.* Bologna, Italy. [conference slides]

Douglas Biber, Ulla Connor, Thomas A. Upton, Molly Anthony and Kostyantyn Gladkov. (2007). Rhetorical appeals in fundraising. In Douglas Biber, Ulla Connor and Anthony Upton, *Discourse on the Move: Using corpus analysis to describe discourse structure* (pp. 121-151). Amsterdam: John Benjamin Publishing Company.

Laurence Anthony. (2016). *AntConc Homepage*. Retrieved April 14, 2016, from Laurence Anthony's website: http://www.laurenceanthony.net/software/antconc/

Marina Bondi. (2010). Perspectives on keywords and keyness: an introduction. In Marina Bondi and Mike Scott (Eds.), *Keyness in texts* (pp. 1-20). Amsterdam: John Benjamins Publishing Company.

Michael Stubbs. (2010). Three concepts of keywords. In Marina Bondi and Mike Scott (Eds.), *Keyness in Texts* (pp. 21-42). Amsterdam : John Benjamin Publishing Company.

Mike Scott and Christopher Tribble. (2006). *Textual Patterns - Key words and corpus analysis in language education.* Amsterdam: John Benjamin Publishing Company.

Mike Scott. (2016a). *Downloads*. Retrieved April 14, 2016, from WordSmith Tools: http://www.lexically.net/wordsmith/downloads/

Mike Scott. (2016b). *Get started guides for WordSmith.* Retrieved from WordSmith Tools: http://www.lexically.net/downloads/version7/HTML/index.html?keywords_calculate_info.htm

Paul Baker. (2006a). *Usind Corpora in Discourse Analysis.* London: Continuum.

Paul Baker, Andrew Hardie and Tony McEnery. (2006b). *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Paul Baker, Costas Gabrielatos, Majid Khosravinik, Michal Krzyzanowsk, Tony McEnery and Ruth Wodak. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 273-307.

Paul Baker, Costas Gabrielatos and Tony McEnery. (2013). Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998–2009. *Applied Linguistics, 34*(3), 255–278.

Paul Baker. (2015). Introduction to the Special Issue. *Discourse & Communication, 9*(2), 143-147.

Stuart Macdonald, Nuria Lorenzo-Dus,

Matteo di Cristofaro, David Mair, Anina Kinzel, Saffron Lee, Nyasha Maravanyika, David Nezri, Jodie Parker, Elliot Parry, Kate Thomas and Luke Walker. (2015). *Online Terrorist Magazines*. Cyberterrorism Project Research Report (No. 5). Available via: www.cyberterrorism-project.org

Theo Van Leeuwen. (2007). Legitimation in discourse and communication. *Discourse & Communication*, 1(1), 91-112.

Tony McEnery, Mark McGlashan and Robbie Love. (2015). Press and social media reaction to ideologically inspired murder: The case of Lee Rigby. Discourse & Communication, 9 (2), 237–259.

Appendix A: %DIFF in *Inspire*

Table 1:Top 30 keywords according to DIFF% in Inspire (study corpus) compared to Dabiq (reference corpus)

Word	Observed frequencies		Expected frequencies		Over/under-use in SC	LL	Normalised frequencies		%DIFF
	SC	RC	SC	RC			SC	RC	
inspire	472	1	293.60946	179.39054	+	437.76	0.001550861	5.37776E-06	28738.42
sheikh	154	1	96.214516	58.785484	+	136.73	0.000506001	5.37776E-06	9309.15
mujahid	141	1	88.144912	53.855088	+	124.50	0.000463287	5.37776E-06	8514.87
almighty	123	1	76.971613	47.028387	+	107.61	0.000404144	5.37776E-06	7415.10
anwar	119	1	74.488658	45.511342	+	103.86	0.000391001	5.37776E-06	7170.70
figure	113	1	70.764225	43.235775	+	98.24	0.000371287	5.37776E-06	6804.11
lone	96	1	60.211665	36.788335	+	82.35	0.000315429	5.37776E-06	5765.44
lamp	60	1	37.865068	23.134932	+	48.96	0.000197143	5.37776E-06	3565.90
wires	48	1	30.416202	18.583798	+	37.95	0.000157715	5.37776E-06	2832.72
summer	46	1	29.174724	17.825276	+	36.13	0.000151143	5.37776E-06	2710.52
explosion	42	1	26.691769	16.308231	+	32.50	0.000138	5.37776E-06	2466.13
irāq	40	1	25.450291	15.549709	+	30.68	0.000131429	5.37776E-06	2343.93
materials	40	1	25.450291	15.549709	+	30.68	0.000131429	5.37776E-06	2343.93
legitimate	39	1	24.829553	15.170447	+	29.78	0.000128143	5.37776E-06	2282.84
allāh	581	15	369.96033	226.03967	+	443.09	0.001909005	8.06664E-05	2266.54
jihadi	76	2	48.417628	29.582372	+	57.76	0.000249715	1.07555E-05	2221.74
abyan	38	1	24.208814	14.791186	+	28.88	0.000124857	5.37776E-06	2221.74
cooking	36	1	22.967336	14.032664	+	27.08	0.000118286	5.37776E-06	2099.54
experiences	34	1	21.725859	13.274141	+	25.28	0.000111715	5.37776E-06	1977.34
measures	33	1	21.10512	12.89488	+	24.39	0.000108429	5.37776E-06	1916.25
fbi	32	1	20.484381	12.515619	+	23.49	0.000105143	5.37776E-06	1855.15
ahl	31	1	19.863642	12.136358	+	22.60	0.000101857	5.37776E-06	1794.05
bin	146	5	93.731561	57.268439	+	105.02	0.000479716	2.68888E-05	1684.07
capability	29	1	18.622164	11.377836	+	20.83	9.5286E-05	5.37776E-06	1671.85
dār	29	1	18.622164	11.377836	+	20.83	9.5286E-05	5.37776E-06	1671.85
add	28	1	18.001426	10.998574	+	19.94	9.20002E-05	5.37776E-06	1610.75
bars	28	1	18.001426	10.998574	+	19.94	9.20002E-05	5.37776E-06	1610.75
hassan	26	1	16.759948	10.240052	+	18.18	8.54288E-05	5.37776E-06	1488.56
shari	127	5	81.937524	50.062476	+	88.27	0.000417287	2.68888E-05	1451.90
build	25	1	16.139209	9.8607908	+	17.30	8.21431E-05	5.37776E-06	1427.46
humanity	25	1	16.139209	9.8607908	+	17.30	8.21431E-05	5.37776E-06	1427.46
operational	25	1	16.139209	9.8607908	+	17.30	8.21431E-05	5.37776E-06	1427.46
press	25	1	16.139209	9.8607908	+	17.30	8.21431E-05	5.37776E-06	1427.46
jihad	618	25	399.13506	243.86494	+	426.48	0.002030577	0.000134444	1410.35
entitled	24	1	15.51847	9.4815296	+	16.43	7.88574E-05	5.37776E-06	1366.36
israel	95	4	61.453143	37.546857	+	64.85	0.000312144	2.1511E-05	1351.09
kuffar	47	2	30.416202	18.583798	+	31.99	0.000154429	1.07555E-05	1335.81

Appendix B: %DIFF in *Dabiq*

Table 2:Top 30 keywords according to DIFF% in Dabiq (study corpus) compared to Inspire (reference corpus)

Word	Observed frequencies		Expected frequencies		Over/under-use in SC	LL	Normalised frequencies		%DIFF
	SC	RC	SC	RC			SC	RC	
shām	175	1	66.74997	109.2500316	+	327.95	0.000941	3.28572E-06	28542.34
tandhīm	115	1	43.9943	72.00570265	+	212.45	0.000618	3.28572E-06	18722.11
anh	85	1	32.61646	53.38353817	+	154.88	0.000457	3.28572E-06	13812.00
ribāt	85	1	32.61646	53.38353817	+	154.88	0.000457	3.28572E-06	13812.00
khalīfah	60	1	23.13493	37.86506778	+	107.09	0.000323	3.28572E-06	9720.23
shar	54	1	20.85937	34.14063488	+	95.67	0.00029	3.28572E-06	8738.21
nusayrī	50	1	19.34232	31.65767962	+	88.06	0.000269	3.28572E-06	8083.53
rahimahullāh	199	4	76.99002	126.0099796	+	350.35	0.00107	1.31429E-05	8042.61
sahābah	44	1	17.06675	27.93324672	+	76.68	0.000237	3.28572E-06	7101.50
tawāghīt	80	2	31.09942	50.90058291	+	138.23	0.00043	6.57145E-06	6446.82
bid	34	1	13.27414	21.72585856	+	57.80	0.000183	3.28572E-06	5464.80
hadīth	64	2	25.03124	40.96876186	+	108.08	0.000344	6.57145E-06	5137.46
shuhadā	31	1	12.13636	19.86364211	+	52.16	0.000167	3.28572E-06	4973.79
expanding	30	1	11.7571	19.2429033	+	50.29	0.000161	3.28572E-06	4810.12
murji	54	2	21.23863	34.7613737	+	89.36	0.00029	6.57145E-06	4319.10
mushrikīn	52	2	20.4801	33.51989606	+	85.63	0.00028	6.57145E-06	4155.43
wilāyāt	26	1	10.24005	16.75994803	+	42.82	0.00014	3.28572E-06	4155.43
al-islām	24	1	9.48153	15.5184704	+	39.09	0.000129	3.28572E-06	3828.09
consolidation	24	1	9.48153	15.5184704	+	39.09	0.000129	3.28572E-06	3828.09
deviance	24	1	9.48153	15.5184704	+	39.09	0.000129	3.28572E-06	3828.09
mosul	24	1	9.48153	15.5184704	+	39.09	0.000129	3.28572E-06	3828.09
treachery	22	1	8.723007	14.27699277	+	35.39	0.000118	3.28572E-06	3500.75
tāghūt	86	4	34.13351	55.86649344	+	137.85	0.000462	1.31429E-05	3418.92
zakāh	42	2	16.68749	27.3125079	+	67.08	0.000226	6.57145E-06	3337.08
prophethood	21	1	8.343746	13.65625395	+	33.54	0.000113	3.28572E-06	3337.08
factions	163	8	64.85366	106.1463375	+	259.08	0.000877	2.62858E-05	3234.79
abū	404	20	160.8067	263.193258	+	641.25	0.002173	6.57145E-05	3206.14
al-kubrā	20	1	7.964485	13.03551514	+	31.69	0.000108	3.28572E-06	3173.41
dajjāl	19	1	7.585224	12.41477632	+	29.86	0.000102	3.28572E-06	3009.74
wa	451	24	180.1491	294.8509376	+	707.35	0.002425	7.88574E-05	2975.64
al-baqarah	36	2	14.41192	23.58807501	+	56.04	0.000194	6.57145E-06	2846.07
azza	36	2	14.41192	23.58807501	+	56.04	0.000194	6.57145E-06	2846.07
jall	36	2	14.41192	23.58807501	+	56.04	0.000194	6.57145E-06	2846.07
al-hākim	18	1	7.205962	11.7940375	+	28.02	9.68E-05	3.28572E-06	2846.07
entails	18	1	7.205962	11.7940375	+	28.02	9.68E-05	3.28572E-06	2846.07
ibnul-qayyim	18	1	7.205962	11.7940375	+	28.02	9.68E-05	3.28572E-06	2846.07
kāfir	53	3	21.23863	34.7613737	+	82.23	0.000285	9.85717E-06	2791.51
sinai	35	2	14.03266	22.96733619	+	54.21	0.000188	6.57145E-06	2764.23
khilafah	34	2	13.6534	22.34659738	+	52.39	0.000183	6.57145E-06	2682.40
banners	17	1	6.826701	11.17329869	+	26.19	9.14E-05	3.28572E-06	2682.40
currency	17	1	6.826701	11.17329869	+	26.19	9.14E-05	3.28572E-06	2682.40
nusra	17	1	6.826701	11.17329869	+	26.19	9.14E-05	3.28572E-06	2682.40
īd	17	1	6.826701	11.17329869	+	26.19	9.14E-05	3.28572E-06	2682.40
feature	67	4	26.92754	44.07245593	+	102.95	0.00036	1.31429E-05	2641.48
airstrikes	50	3	20.10084	32.89915725	+	76.76	0.000269	9.85717E-06	2627.84