

**The Representation of the Arab Spring Narrative
in English and Arabic News Media**

Najat Hashem Mohammad Alian

Submitted in accordance with the requirements
for the degree of

Doctor of Philosophy

The University of Leeds

School of English

December 2016

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2014 The University of Leeds and Najat Hashem Mohammad Alian. The right of Najat Hashem Mohammad Alian to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

Acknowledgment

الحمد لله رب العالمين
al-ḥamdu li-llāhi rabbi l-‘ālamīn

This work would not have been possible without the expert, careful, and patient comments, corrections and advice provided by Dr Alison Johnson, my main supervisor. She has been a tremendous mentor for me, in academia and in life. She was always there when in need, continuously had faith in me through hardships, her unique help, constant encouragement and her trust in my abilities right from the start are very much appreciated. I would also like to thank my second supervisor Professor James Dickins for his expert and invaluable comments and suggestions, especially in terms of the Arabic data. I am also grateful to my PhD committee for reading my thesis and making constructive comments and feedback: Dr Fiona Douglas and Dr Charlotte Taylor.

I would also like to express my gratitude to those whom I consider as my partners and friends who have accompanied me on this journey, Mahmud Diesy, my twin sister and best friend Hayat, and her husband Bassam Saleh for their love, constant support and encouragement. With love, I dedicate this work to them.

I would like to thank David Woolls, for his help in creating the Arabic Wikipedia reference corpus. Many thanks also to Martin Callaghan for his technical support and help in the concatenation of files.

I would like to thank my friends and colleagues at the School of English, Ashraf Abdullah, David Wright, Emmanuel Sarfo, Camille Laporte, Neveen Al-Saeed, and Nurshafawati Ahamd for the good times we spent together, whether in Leeds or abroad.

Finally, I am so grateful to all my family members and friends. I am also grateful to my father and mother, whose love and prayers were as always very important to me during the writing of my thesis, my brothers and sisters for their love and emotional support

Abstract

Since its emergence in late December 2010, the Arab Spring narrative has sparked many controversies among researchers, commentators, analysts, and scholars from different disciplines around the world in terms of the causes and the reasons behind it and even its name: ‘Arab Spring’. This study explores the Arab Spring narrative from its emergence to its continuing dénouement in the English and Arabic mainstream news media from corpus-linguistic and critical discourse analytic perspectives. The Arab Spring bilingual corpus consists of two main sub-corpora, English and Arabic, compiled from *LexisNexis* and other news websites. Totalling 15,088 articles and 11,522,846 words, the English sub-corpus consists of 7,018 texts with total of 5,901,416 words, while the Arabic sub-corpus comprises 8,070 news texts and a total of 5,621,430 words. Taken from prominent news media outlets from Western, Arab and Islamic countries, and divided into two major text types (news and editorials and opinions) with date range coverage from 15 June 2010 until 31 August 2013, it allows us to diachronically and synchronically examine the discursive construction of the Arab Spring narrative. Combining quantitative and qualitative methods associated with Corpus Linguistics (CL) and Critical Discourse Analysis (CDA), the current study explores the key topics associated with the Arab Spring at both the linguistic as well as the semantic levels by means of frequency, keyword (KKW list function in my case), collocation list functions and concordance. Analysis also identifies the main news actors and news

values. Actors and events are represented, negatively and positively by means of lexical choice, and the different presentation strategies indicate that many of the Arab Spring news stories are politically, socially, and ideologically polarized. The contrasting themes/concepts within the resulting semantic categories (by means of pairs of items with positive/negative connotations) are also prevalent. For example, at the lexical level the following contrasting pairs are revealed: *democracy/dictatorship; religious, sectarian/secular; peace/ violence; government/regime; allies/enemy; corruption/ reform, opposition/ support*. Similarly, at the grammatical level items, such as *pro/anti, is/non-* and *not*, also indicate the contrastive as well as the polarizing nature of the Arab Spring narrative.

Contents

1.	The Arab Spring: An Introduction	1
1.1	What Sparked the Arab Spring?	1
1.2	Arab Spring Geography	7
1.3	Arab Spring: A Controversial Term	10
1.3.1	Arab Spring: Definition and History	12
1.4	Aims and Objectives	15
1.5	Structure of the Study	20
2.	Methods and Approaches	23
2.1	Introduction	23
2.2	Data Collection and Corpus Design	24
2.2.1	Description of the Arab Spring Bilingual Corpus	25
2.2.2	Arabic Data: Transcription Scheme	39
2.2.3	Differences Between English and Arabic	42
2.2.4	Some Challenges in Cross-Linguistic Corpus-Based Analysis	53
2.2.5	Search Terms: Arab Spring and the ‘Protest Vocabulary’	56
2.2.6	Using Reference Corpora	62
2.2.7	Ethical Considerations	68
2.3	Theoretical and Methodological Frameworks	70
2.3.1	Topical Analysis of the News and Editorials Texts	77
2.3.2	Criticism of Keyness: Definitions and Statistical Measures	92
2.3.3	Methodological Concerns: Making the Keywords Database	101
2.3.4	Key Keyword Categorization	107
2.3.5	Advantages and Limitations of Corpus-Based CDA	116
2.4	Summary	121

3.	Who is Being Reported? Representation Strategies of News Actors	124
3.1	Introduction	124
3.2	Representation of News Actors: A Sociosemantic Inventory	125
3.2.1	Exclusion	127
3.2.2	Inclusion	129
3.3	News Actors in the Arab Spring Narrative: The Elite	132
3.4	Locating News Actors	137
3.5	Representing News Actors in the Arab Spring Narrative	143
3.5.1	Functionalization	143
3.5.2	Classification	169
3.5.3	Assimilation	174
3.5.4	Objectivation	179
3.6	So, Whose Voice? What Role?	182
3.7	Summary	185
4.	Unlocking the Arab Spring: Key Topics	188
4.1	Introduction	188
4.2	KKW Function for Identifying Topics in the Arab Spring Corpus	190
4.3	Arab Spring: A Diachronic Perspective on Topics and Events	240
4.4	Geography: Drawing the Map	250
4.5	Summary	268
5.	The Newsworthiness of the Arab Spring	270
5.1	Introduction	270
5.2	What are News Values?	271
5.2.1	Negativity	275
5.2.2	Recency	276
5.2.3	Proximity	280
5.2.4	Relevance	283

5.2.5 Eliteness and Source and Voice	284
5.2.6 Agenda	287
5.2.7 Consonance	289
5.2.8 Personalization	291
5.3 Newsworthiness of the Arab Spring: Quantitative Perspective	293
5.4 Tracking News Values: Selected Peak Months	313
5.4.1 News Values Before and After the Arab Spring: Up to December 2010 and Afterwards and Selected Peaks	313
5.5 Summary	324
6. Conclusion	326
6.1 Summary of Conclusions in Relation to Results	327
6.2 Arab Spring; A Narrative of the Powerful	332
6.3 What is the Arab Spring?	334
6.4 What is Newsworthy about the Arab Spring?	336
6.5 Contribution to Knowledge	338
6.6 Implications and Further Research	339
References	341
Appendix A	360

List of Tables and Figures

Tables

Table 2.1 Statistics of the Arab Spring bilingual corpus	27
Table 2.2 Breakdown of the Arab Spring sub-corpora	36
Table 2.3 DIN31635 standard for the transliteration of Arabic alphabet	40
Table 2.4 Names and Shapes of Arabic Letters	44
Table 2.5 Arabic diacritics with examples	46
Table 2.6 Nunation sounds in Arabic	47
Table 2.7 Sounds and pronunciation of Hamza	49
Table 2.8 Sun and Moon letters in Arabic	51
Table 2.9 Illustration of the three major macro-rules of news topic structure	81
Table 2.10 Illustration of Situation Triad, (S)Td, Structure in the English editorials and opinions	88
Table 2.11 Illustration of Situation Triad, (S)Td, structure in the Arabic editorials and opinions	89
Table 2.12 Examples of the top and lowest ten KKWs along with their statistics in the British news KKW list	103
Table 2.13 Minimum number of texts for KKWs and their % of each sub-corpus	104
Table 2.14 Distribution of shared/unshared KKWs across the English and Arabic sub-corpora	106
Table 2.15 Number of KKWs in the top ten semantic categories in the English and Arabic sub-corpora	111
Table 2.16 Example of plotting results in pivot table of Excel taken from the English sub-corpus	113
Table 3.1 KKW items of the name-pn category with their overall Avg. % of the English sub-corpus as a whole	133
Table 3.2 KKW items of the Arabic name-pn category with their Avg. % of the Arabic sub-corpus as a whole	134

Table 3.3 Functionalization by KKWs of the government and security category with their % of each English text type sub-corpus	145
Table 3.4 Functionalization by KKWs of the government and security category with their % of each Arabic text type sub-corpus	150
Table 3.5 KKW functionalization items of the protest vocabulary category with their % of each of the English text type sub-corpora	154
Table 3.6 KKW Functionalization items of the protest vocabulary category with their % of each of the Arabic text type sub-corpora	156
Table 3.7 KKW Functionalization items of the social, states, and process category with their % of each English text type sub-corpus	158
Table 3.8 Lexical collocates of dictators of the English editorials and opinions sub-corpora	161
Table 3.9 Functionalization KKWs of social actions, states and processes category with their % of each Arabic text type sub-corpus	166
Table 4.1 Top ten KKW-semantic categories ordered by overall number of words with their Avg. % of English and Arabic text type sub-corpora	197
Table 4.2 Actual totals of KKW items of the top ten categories in the English and Arabic sub-corpora	198
Table 4.3 KKW items referring to political systems with their % of each English text type sub-corpus	207
Table 4.4 KKWs referring to political systems with their % the Arabic sub-corpora	210
Table 4.5 KKW items of the emotion category with their % of each of English and Arabic text type sub-corpora	228
Table 4.6 Key geographical groups and lexical sets with overall Avg. % of the English sub-corpus	251
Table 4.7 Key geographical groups and lexical sets with their overall Avg. % of the Arabic sub-corpus	252
Table 4.8 Top ten R1 collocates of both items Arab and Western of the English sub-corpus	267
Table 5.1 Breakdown of news values of the GEOGRAPHY category of English and Arabic sub-corpora	294

Table 5.2 KKW news values ordered by total number of words with their average percentage of each English and Arabic sub-corpus	296
Table 5.3 Actual number of KKW items realizing news values in the English and Arabic sub-corpora	297
Table 5.4 Number of news values in English and Arabic KKW semantic categories	312
Table 5.5 News Values in December 2010	314
Table 5.6 News values during peak months in the English and Arabic after the Arab Spring	324

Figures

Figure 2.1 Geographical distribution of the Arab Spring sub-corpora	32
Figure 2.2 Proportion of texts and words in the Arab Spring sub-corpora	33
Figure 2.3 Search terms for collecting English and Arabic data	59
Figure 2.4 Pyramidal model of a news report schematic structure	79
Figure 3.1 Adaptation of van Leeuwen's (2008) analytical framework of news actors in the Arab Spring	126
Figure 3.2 Geographical distribution of KKW news actors of the name-pn category of the English and Arabic sub-corpora	139
Figure 3.3 Concordance examples of the KKW president from the British news sub-corpus	148
Figure 3.4 Top ten 3-word clusters of president in the GV news sub-corpus	149
Figure 3.5 Examples of the top frequent 3-word clusters of وزير of the Arabic news sub-corpus	152
Figure 3.6 Concordance examples of rebels of the British editorials and opinion sub-corpus	155
Figure 3.7 Top ten 3-word clusters of sheikh referring to news actors in the British news	160

Figure 3.8 Instances of dictators with context items support, backed and pro of the three main English editorials and opinions sub-corpora	162
Figure 3.9 Concordance examples of Western leaders of the American and the British news sub-corpora	164
Figure 3.10 Lexical collocates of الطغاة<the-tyrants> of the Arabic editorials and opinions sub-corpus	168
Figure 3.11 Examples of the most frequent 3-word clusters of Islamist of the American news sub-corpus	171
Figure 3.12 Examples of the top frequent 3-word clusters of الإسلامية of the Arabic news sub-corpus	173
Figure 3.13 Concordance examples of is, has and will with countries as news actors of the English news during May 2011	180
Figure 3.14 Concordance examples of said showing instances of utterance autonomization of the English news sub-corpus	181
Figure 3.15 Examples of 3-word clusters of وقال<and-said> of the Arabic news sub-corpus	181
Figure 4.1 Top ten 2-3-word clusters of الإعلام in the Arabic corpus	196
Figure 4.2 Concordance lines of the Arab Spring with key geographical items in the BR-news sub-corpus	202
Figure 4.3 Shared KKW items of the English government and security	204
Figure 4.4 Concordance examples of regime/s of the British news sub-corpus	208
Figure 4.5 Top 3-word-clusters of ضد<against> in the Arabic editorials and opinion sub-corpus	220
Figure 4.6 Examples of most frequent 2-5-word clusters of انتفاضة<uprising> forms in the Arabic editorials and opinions sub-corpus	227
Figure 4.7 Concordance examples of peace from the British news sub-corpus	229
Figure 4.8 Examples of concordance lines of terror in the British news sub-corpus	233
Figure 4.9 Concordance examples of KKW items of the English emotions category	235

Figure 4.10 Peak events before the Arab Spring (15 June-31 Dec 2010)	241
Figure 4.11 concordance lines of the KKW items Christians from the British news in December 2010	242
Figure 4.12 Frequency of the term Arab Spring per month in the English and Arabic sub-corpora (Jan-2011 and Aug-2013)	244
Figure 4.13 KKW geographical groups and their Avg. % of each of the English and Arabic text type sub-corpora	257
Figure 4.14 Concordance examples of countries as news actors from the American news (Aug-2013)	261
Figure 4.15 Concordance examples of countries as news actors from the Arabic news	262
Figure 5.1 All instances of said in the CNNe-11-81	286
Figure 5.2 Top 3-word clusters of وقالت<and.said.fem> in the Arabic news	302
Figure 5.3 Top collocates to the L1 of said of the English news	304

1. The Arab Spring: An Introduction

1.1 What Sparked the Arab Spring?

On 17 December 2010, young Tunisian street vendor Mohammad Bouazizi set himself on fire, as a symbolic act of protest, after police confiscated his cart, and banned him from selling fruit to earn a living. Mohammad Bouazizi sparked the Tunisian uprising that toppled the Tunisian President Zine al-Abidine Ben Ali, who fled the country to Saudi Arabia on 14 January 2011, ending his 23 years in power. The self-immolation of Mohammad Bouazizi catalysed the *Arab Spring*, or الربيع العربي <Ar-rabī‘ Al-‘arabī> narrative in the news, engendering media celebration, speculation and controversy. The headlines (1, 2, 3, 6 and 7) shown in Example 1, as well as the extracts in Examples 2 and 3 (lines 2 and 3) are examples of how Mohammad Bouazizi’s story has been received in the English as well as the Arabic news media. His death is represented as the incident that ignited the popular protests in Tunisia, and الثورات في العالم العربي <*the revolutions in the Arab world*>.

1. HEADLINES

1 Tunisian unemployment sparks unrest
(FTe-10-123/ news: 28-Dec)

2 Tunisia suicide protester **Mohammed Bouazizi** dies
(BBCe-11-01/ news: 05-Jan)

- 3 Inspiration [Bouazizi] for violent protests in Tunisia
dies (CNNe-11-01/ news: 05-Jan)
- 4 Middle East unrest: From Tripoli to Tehran, **Arab spring** sprouts new wave of fury: Violent clashes reported around the region as anti-government movements gain momentum: Libya
(GRDe-11-10/news: 17-Feb)
- 5 Rebellion spreads as '**Arab Spring**' takes hold
(TMSe-11-08/news: Feb-19)
- 6 هل كان البوعزيزي يعلم انه سيسقط بن علي حين أشعل النار في نفسه؟
Did Al-Bouazizi know that Ben Ali will fall when he set himself on fire? (QAa-11-18/ news: 20-Jan)
- 7 كيف أشعل بائع فاكهة تونسي فتيل الثورات في العالم العربي؟
*How a **Tunisian fruit seller** ignited the **revolutions** in the Arab world?* (WSTa-11-30/ news: 28-March)
2. The death of Tunisian vegetable seller **Muhammad Bouazizi** helped spark protests that led to the toppling of Tunisian President Zine al-Abidine Ben Ali a week later, and is considered as one of the catalysts for the regional **Arab Spring**. (BBCe-13-61/ news: 17-May)
3. 1 **Tunisia**, you may recall, is where the **Arab Spring**
2 began, when a hapless fruit peddler, driven to his wit's
3 end by corruption and official harassment, set himself
4 on fire. That sparked a popular uprising, which sent
5 Tunisia's dictator fleeing, emboldened protesters from
6 **Libya to Egypt to Yemen to Syria**, and put **Tunisia**
7 itself - a North African nation of about 10 million
8 people - on the path to democracy.
(WPe-11-188/ editorial: 01-Dec)

Since then, the Arab Spring, as indicated by Example 3 (lines 5, 6 and 7), spread beyond Tunisia in numerous protest movements and revolutions to other countries across the Middle East, such as *Egypt, Libya, Yemen, Bahrain, Syria*, and finally *Turkey* in June 2013. The major slogan of the protesters and demonstrators in the Arab world was إسقاط النظام يريد الشعب <aš-ša‘b yurīd ‘isqāṭ an-niḏām – *the people want the fall of the regime*>; a slogan that also attracted the international as well as the Arab news media, as indicated in Examples 4 (lines 2 and 3) and 5 (in line 2 for the Arabic, and in lines 5 and 6 for English translation):

4. 1 When young Arabs marched this year chanting
 2 **"the people want the fall of the regime"**, the
 3 slogan that has come to **define the Arab spring**,
 4 they demonstrated an extraordinary resolve to
 5 destroy the autocratic order in the Middle East.
 (FTe-11-137/ opinion: 27-Sep)

5. 1 وقال الشاهد ان بعض المشيعين [في درعا – سوريا] ردّدوا
 2 'الشعب يريد اسقاط النظام' وهو الشعار الذي تردّد في الثورتين
 3 التونسية والمصرية.
 4 *the witness also said that some mourners [in*
 5 *Daraa – Syria] chanted 'the people want the fall*
 6 *of the regime'* and it is the slogan that was
 7 *chanted in the Tunisian and Egyptian revolutions*
 (QAa-11-58/ news: 04-April)

It has become a phenomenon of interest and research for commentators, analysts and scholars from different academic and social disciplines and domains in an

attempt to explain and clarify what has really happened, and what is still happening, in the countries across the Middle East. Moreover, the Arab as well as the international reactions towards the Arab Spring and its events have been, and still are, disparate and are generated from different perspectives and provide different interpretations. For example, some scholars arguing for the role of the social and digital media in protest movements indicated that some voices, 'have been so bold as to label them as the "Twitter Revolutions" or "Facebook Revolutions" (Cottle, 2011, p. 647). Many events were documented and communicated by ordinary people using different means of social and digital media (Haenska-Ahy and Shapour, 2013; Cottle, 2011) that went '*beyond sending and receiving messages*' (Bennett and Segerberg, 2012, p. 739, original italics). Those scholars, however, argued that new social and digital media might not have produced the *Arab Spring*, but rather were either part of it, or played a significant role in shaping many of its events that led to political change in countries throughout the Middle East (Haenska-Ahy and Shapour, 2013; Barton, 2012; Cottle, 2011; Howard and Hussain, 2011).

What is more, despite the fact that they have become more affected by the interactions and interrelations with the new social and digital media, the mainstream media remain the most important in reporting and framing the news, especially that of the protest and social movements that need to mobilize support and get their message across to the wider public (Haenska-Ahy and Shapour, 2013; McCurdy, 2012; Cottle, 2008, 2011; Aday et al., 2013). Other interpretations are made by

scholars from different disciplines other than CDA or CL studies. For instance, some scholars and observers in Middle East studies expressed different as well as contradictory interpretations of the *Arab Spring* in terms of its economic motivations. ‘Some have interpreted the uprisings as a demand *for* free-market economies, while many others have read them as protests *against* neoliberalism’ (Baron and Pursley, 2011, p. 380, original italics).

From a political science perspective, the *Arab Spring* has been investigated in terms of the causes and reasons behind this phenomenon. Scholars tried to clarify and explain why such a phenomenon took many social and political experts and scientists by surprise (Kurzman, 2012a; Dupont and Passy, 2011; Goodwin, 2011). They tried to explain the phenomenon of the Arab Spring in terms of its being a series of ‘*unpredictable* protest events’ that began in Tunisia and continued to sweep throughout many countries in the Middle East (Dupont and Passy, 2011, p. 447, original italics). In the same context, the Arab Spring, as Goodwin (2011, p. 452) points out, represents an ‘egregious yet illuminating failure’ of many think-tanks to predict such a phenomenon despite the fact of their setting many demographic, military, economic and political indicators for providing warnings of political unrest in countries around the world. Such indicators were unable to identify the psychological and mental aspects of mass uprisings in terms of predicting when and how the individuals might join a collective action of protest against a particular political regime (Goodwin, 2011; Kuran, 1995) indicating that

reasons, such as ‘economic downturn’ might not be sufficient triggers for revolutions (Kurzman, 2012a, p. 378).

Having taken these perspectives into consideration, there is no doubt that the protests of the *Arab Spring* reflect a clamour for freedom, human rights and equality, and represent an outcry against corruption and despotic regimes. Thus, the Arab Spring, as some scholars describe it, is a movement that ‘demands political freedom, an end to the monopoly of power, and an end to corruption and plutocracy’ (Amanat, 2012, p. 148). It represents ‘in some way’ a replicate of ‘the Iranian Revolution of 1979’ as both events were successful in toppling the dictatorships (Kurzman, 2012b, p. 163). However, other dissenting views consider it either as a struggle for power that ‘does not fit into the simplistic liberal democratic narrative aka Arab Spring’ (Hadar, 2011, online), or as part of a conspiracy that is ideologically and politically motivated and led by foreign agendas against the region (Elhousseini, 2013).

My research investigates a custom-built bilingual corpus of over 11 million words and 15,088 news texts in English and in Arabic, taken from news sources around the world between 2010 and 2013, in order to ask what this narrative tells us about the now unravelling Arab Spring narrative in the news and what it tells us about the nature of news itself.

1.2 Arab Spring Geography

Many definitions associate the *Arab Spring* with the term ‘Middle East’ or the ‘Arab World’ or ‘countries across the Middle East and North Africa (abbreviated as MENA)’. However, even the term ‘Middle East’ is somehow controversial. Therefore, it is necessary to explain what the term ‘Middle East’ means, and where it starts and ends in terms of the regions or countries it includes, at least for the purposes of this study. From a historical perspective, ‘the Middle East’ is relatively a new term. The use of this ‘label’ has been commonly used during ‘the Second World War when the British military established a Middle East Command in the area under the authority of the War Office’ (Milton-Edwards, 2011, p. 9). Before this term got its wider usage throughout the world, the area was referred to as ‘the Near East’, an older label that used to be more flexible covering all histories, languages and cultures of the region (Choueiri, 2005, p. 1). With the fall of the Ottoman empire and the ‘emergence of the Arab world as a political block of states grouped under the umbrella of the Arab League [1945], the designation widened to include North Africa as a whole’ (Choueiri, 2005, p. 1). Hence, with this wider designation, the ‘Middle East’ becomes a geographical region that embraces the Arab countries in North Africa, Iran, Turkey, Greater Syria and Iraq (Choueiri, 2005; Milton-Edwards, 2011). Therefore, this wider designation will be adopted throughout this study. Moreover, the current research is conducted at the macro level of the *Arab Spring* in the Middle East rather than a single case study of one of its countries, for example *Tunisia*, *Bahrain* or *Syria*, that witnessed the protest

movements. One of the reasons is that most of the events that broke throughout the Middle East region have been framed into the Arab Spring narrative (Tyner and Rice, 2012, p. 131). Such framing, however, has caused considerable controversy around the use of the term *Arab Spring*, especially considering that not all events can be described, in terms of what the term metaphorically and implicitly means, as democratic birth, political change, peaceful transition of power, and/or movement for freedom. For example, in Syria, no sooner had the popular and peaceful uprising inspired by the Arab Spring erupted in March 2011 than violence escalated between the government and the opposition, and developed into a bloody civil war, still taking place at the time of writing my thesis, drawing in world powers and resulting in a refugee crisis. Similarly, in Libya and Yemen, the overthrow of the old regimes in the two countries by October 2011 and February 2012 respectively, resulted in more violence, leading to civil wars and more humanitarian crises. However, this does not mean that the term *Arab Spring* stopped being used by the different English and Arabic news outlets in the contexts of later events (of 2013 for example), which were, directly or indirectly, connected/linked to the Arab Spring, as indicated in Examples 6 (lines 1 and 5):

6. 1 Protests in **Syria**, inspired by the **Arab Spring**
 2 which saw changes of government in Tunisia,
 3 Egypt and Libya, were brutally suppressed by
 4 security forces in 2011. The stand-off has since
 5 escalated into a **civil war** which has claimed tens
 6 of thousands of lives so far, according to UN
 7 estimates. (BBCe-13-99 / news – 10-July)

Nevertheless, although the term *Arab Spring* was the main query term for my corpus, and kept being used by all English and most of the Arabic news outlets throughout the period specified for this study (June 2010 to August 2013), it did not mean that it was *the favoured label* for all news outlets examined in this study, especially in the Arabic sub-corpus (discussed in section 2.5). Example 7 (lines 2, 5, 9 and 13) shows how the term <الربيع العربي> *the-Arab Spring* is used, and replaced by the item <الصحوّة الإسلامية> *Islamic awakening* in the Iranian *Mehr* (MHRa) news sub-corpus, to refer to the events that swept across the Middle East countries. Example 7 reflects one out of only three instances of the term *Arab Spring* in a total of 525 news articles in the Arabic MHRa news sub-corpus, which indicates rejection of the term *Arab Spring* in the Arabic corpus articles.

7. 1 واوضح [علي أكبر ولايني، سياسي إيراني] انه منذ بدء الموجة
2 الجديدة للصحوّة الإسلامية في الدول العربية بالشرق الأوسط وشمال
3 افريقيا، بذلت القوى الغربية مساعي حثيثة لمصادرة هذه الحركة،
4 وصرح: أن القوى الغربية ومن خلال استخدام مصطلحات من قبيل
5 "الربيع العربي" سعت الى التغطية على المظاهر والجوانب
6 الإسلامية لهذه النهضة الشعبية،

7 And he [Iranian politician Ali Akbar Velayati]
8 explained that since the start of the new wave of the
9 **Islamic awakening** in the Arab countries in the
10 Middle East and North Africa, the Western powers
11 have made unrelenting efforts to eliminate this
12 movement, and he said that the Western powers and
13 by the use of terms such as "**the-Arab Spring**"
14 sought to cover up the Islamic manifestations and
15 aspects of this popular renaissance,

(MHRa-13-48/ news: 29-April)

The following section continues with the controversies provoked by the term *Arab Spring* and the reasons behind such controversies.

1.3 Arab Spring: A Controversial Term

The term *Arab Spring* <الرّبيع العربي – Ar-rabī‘ Al-‘arabī> has provoked many controversies not only over its meaning and representations, but also over its name Arab Spring. The term became more controversial than other terms such as the *Arab uprisings* or the *Arab awakening*. Despite its occurrences within the contexts of concepts such as *protest* <احتجاج - ‘iḥtijāj>; *uprising* <انتفاضة - ‘intifāda>; *awakening* <صحوة - ṣaḥwa>; or *revolution* <ثورة - ṭawra>; it led some intellectuals and scholars, in the West and the Middle East, to argue against the term either because it has negative connotations or gives ‘a false impression’ that all recent protests and events in the Middle East countries are homogenous (Dupont and Passy, 2011, p. 450; Tyner and Rice, 2012). The term is allusive; it has a clear allusion to the 1968 Prague Spring, which was brutally put down by the Warsaw Pact troops led by the Soviet Union and did not lead to any major long lasting result or development (Khouri, 2011). What is more, for those who argued against the term, the media’s ongoing representation of the term *Arab Spring* reinforces the orientalist construction of the Arab world and Middle East (Khouri, 2011; Tyner and Rice, 2012). Other allusions come from the whole term, that is *Arab Spring*. For example, the ‘spring’ metaphor symbolizes that it is something that occurs at the level of a season or a few months and that it might not have a long lasting impact (Khouri,

2011). A more optimistic perspective, however, considers the notion of ‘spring’ as ‘a universal symbol standing for a fresh beginning, new growth, and new life’ (Torlakova, 2014, p. 6). In its political sense, the ‘spring’ metaphor also stands for optimistic periods of political change and transformation, and throughout history, ‘the social movements given the spring label have shared a hope for liberalization in the face of oppressive regimes’ (Zimmer, 2011, online). On the other hand, the term ‘Arab’ connotes ‘that those countries experiencing the “Arab Spring” are homogenously Arab’, [and that in turn] diminishes the role of Berbers, Touaregs, Kurds, Bedouins’ (Tyner and Rice, 2012, p. 131), and other ethnic groups in the Middle East region. Therefore, the term *Arab Spring* becomes more of an ambiguous phenomenon, from a CDA perspective, than other terms such as *protest* or *revolution* that might denotatively and respectively indicate disagreement with or change of the political system by which a country is governed, and thus could be looked at as a phenomenon that is ideologically or politically motivated. Finally, there is no doubt that there are terms other than ‘Arab Spring’ (e.g. *Arab revolutions/uprisings* or *Islamic Awakening*) that have been used whether by different mainstream news media or other political and social entities throughout the world. Nevertheless, the *Arab Spring* is still the most commonly used term referring to the events that broke out across many countries in the Middle East region between 2010 and 2013.

1.3.1 Arab Spring: Definition and History

In September 2013, the *OED* (2013, online) added a new entry for the term and defined it simply as ‘[a] process or period of political or cultural liberalization in the Arab world; (now) spec. a series of anti-government or pro-democratic uprisings and demonstrations in various countries in North Africa and the Middle East, beginning in Tunisia in December 2010’. In this sense, the *OED* also gives historical citations showing that the term can be traced back to 1975. The 1975 citation states that ‘this stance [that does not prioritize cultural and political creativity to overcome cultural backwardness] will be hard to accept for those who have been impatiently waiting for the advent of an “Arab Spring”’. Searching for the source provided by the *OED*, and from which the citation has been taken, it appeared to be a book review written by the Orientalist, Detlev Khalid, and published by the *Islamic Research Institute* in Islamabad in 1975. It is concerned with a book in French, published in 1973 and entitled *La crise des intellectuels arabes: Traditionalisme ou historicisme/ The Crisis of the Arab Intellectual*, written by Abdallah Laroui, a Moroccan historian and novelist. The book deals with the problems of Arab intellectuals, and defines such problems in terms of history, tradition and modernity (Khalid, 1975). However, it is worth mentioning that, in this sense also, the term *Arab Spring* actually predates this citation. Based on an online search, the term refers to a title of a book, *Un Printemps Arabe* (or in English, *An Arab Spring*), which has been written by the French journalist and historian, Jacques Benoist-Méchin. According to Al-Husseini (2014, online), the term *Arab*

Spring was used for the first time by Benoist-Méchin in his book in which he describes the Arab revolts that took place in the Middle East region during the 1950s, and compares them to the revolutions of 1848 in Europe, known as the ‘Springtime of Nations’. Within the same context, Galtung (2012, p. 19) considers the *Arab Spring* as ‘the third Arab revolt in less than a century’, stating that the first one, in 1916-1918 was against the Ottoman Empire, the second, in 1952-69 against Western imperialism, and that the third revolt, in 2011 was against the U.S.-Israeli empire (Galtung, 2012, p. 20). Referring back to the *OED* citations, another one shows reference to *an Arab Spring* in the 2003 *New York Times Magazine* article ‘*Dreaming of Democracy*’, which is mostly concerned with the war in Iraq and which states that ‘the war, which is vastly unpopular in the Arab world, is far more likely to improve the fortunes of the Islamists, [...] than to ventilate the region with an Arab spring’. Finally, the *Boston Globe*’s 2005 article ‘*The Arab Spring*’ is the last historical citation in the *OED* that is concerned with the term before it has been used again in 2011. It states that the political changes unfolding in the Middle East region, especially the Arab world, are ‘being called an “Arab Spring”’, and Bush’s critics are right to give him credit for helping to bring it about. What his allies need to bear in mind is that cracks in the ice of tyranny and misrule don’t always lead to liberation’ (Jacoby, 2005, online).

In this context, commentators used the term in 2005 to refer to some events that occurred in the Arab world after ‘the American invasion of Iraq in 2003 and President George W. Bush’s “freedom agenda”’ (Gelvin, 2012, p. 32). Some of

these events include the Cedar revolution in Lebanon and the withdrawal of the Syrian troops from it, the pledge of Mubarak to have free presidential elections in Egypt, the Saudi municipal elections, as well as the right for women in Kuwait to vote (Gelvin, 2012, p. 32). The historical citations given by the *OED* are believed to present the term *Arab Spring* as an event that is long-awaited not only by those in the Middle East region, but also by many observers around the world. It is a revolutionary event that would bring about *change, reform, freedom* and *democracy*. Given that the literature on the origin of the term *Arab Spring* is inadequate, as it fails to take account of the nature of contexts in which it was previously used, and based on the information available to this study, the term seems to be used in different, yet related, contexts of a political, social, and revolutionary nature. This shows that the term *Arab Spring* <الربيع العربي> - Ar-rabi‘ Al-‘arabī> did not suddenly come into widespread use immediately after the suicide of the Tunisian vendor, Bouazizi, in late December 2010 as it had existed for quite a long time before and has simply been resuscitated by the Western media, becoming a popular term to refer to the events that swept the Middle East at the start of the second decade of the 21st century.

The first reference to the term *Arab Spring* that denotes the events which started in Tunisia can be seen in a citation given by the *OED* taken from Lynch’s 2011 ‘Obama’s Arab Spring’ in the *Foreign Policy Magazine* stating that ‘if these protests continue to spread, both inside of countries and across to other Arab countries, then we really could talk about this being Obama’s ‘Arab Spring’

(Lynch, 2011, online). Historical citations in the *OED* give an indication of the nature of the contexts in which a word or an item tends to occur, although they might not provide a full account of its (negative or positive) prosodies (Stubbs, 1995)

1.4 Aims and Objectives

Against this background, and since few studies in CL and CDA have attempted to explore the representation of the Arab Spring in the news media bilingually or by using corpus linguistic methods, the aim of this study is to investigate the representation of the Arab Spring narrative in the English and Arabic news media. In terms of corpus linguistics, this study combines corpus-based and corpus-driven approaches (Baker et al., 2013a, p. 259; Tognini-Bonelli, 2001). In the former the corpus is seen as a repository of examples to test, exemplify, or back pre-existing theories, whereas in the latter ‘the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence’ (Tognini-Bonelli, 2001, p. 84). Using the different *WordSmith Tools* (Scott, 2015) techniques and functions, such as frequency/cluster, keyword, or the key keyword (KKW) in my case, collocation list functions, and concordances, the study aims to answer the following questions:

- 1- Who are the main news actors, and what strategies are employed in the representation of the main news actors?

- 2- What are the main topics and themes associated with Arab Spring, and how they are presented in the different news media outlets in terms of language use?
- 3- What are the news values that have made the Arab Spring a newsworthy story and how KKW function can be used in identifying the news values prevalent in the Arab Spring narrative?

One of the questions that is based on the analysis of KKW is to examine the potential changes over time in the way *the Arab Spring*, or الربيع العربي <Ar-rabī‘ Al-‘arabī> is represented in terms of its associated topics/events before and after its emergence on 17 December 2010. That is, the time between 15 June 2010 and 31 August 2013. Using the different CL techniques, it will be possible to examine such changes over time in two different ways. Synchronically, they enable researchers to explore certain language variation or a linguistic phenomenon occurring within different (sub)corpora at any given point in history (e.g. Gabrielatos and Baker, 2008). Diachronic analysis, on the other hand, allows for investigating as well as comparing the development of a language variety or a linguistic phenomenon at two or more different times in history (e.g. Partington, 2010, 2012). For example, dividing both the English (with its three main English varieties, the American, the British and the Global Variety) as well as the Arabic Arab Spring sub-corpora into 39 monthly sub-corpora each, it is possible to examine change overtime of the main topics/events and actors before and after the emergence of the Arab Spring on 17 December 2010. Initial analysis shows, that both the English as well as the Arabic

sub-corpora share several peak topics/events that appear to be newsworthy. Before the Arab Spring, that is the period between 15 June and 31 December 2010, there has been a total of 13 peak events distributed across the months of July, September and December 2010. After the Arab Spring, which, on the other hand, extends between 1 January 2011 and 31 August 2013, there have been 21 peak events distributed across May, October and December 2011; May and September 2012, and March and July 2013. The following is a simple presentation of the timeline of the peak events during the before (that is, 15 June – 31 December 2010) and after (that is 1 January 2011 – 31 August 2013) the emergence of the Arab Spring on 17 December 2010 (for more details see Chapter 4: section 4.3):

1- Before the Arab Spring (July, September and December 2010):

In July 2010, there are five peak events of which four are shared between the English as well as the Arabic sub-corpora. These are the Israeli-Palestinian peace talks and Gaza blockade; the death of the Lebanese Shi'ite cleric Fadlallah; The Israeli Prime Minister Benjamin Netanyahu's visit to Washington and his meeting with Obama; The US sanctions against Iran. The last peak event regarding Syria, its alliance with Iran as well as its support to militant Islamic groups Hezbollah and Hamas was only prominent in the British sub-corpus.

In September 2010, there are two main peak events: The issue of settlements between the Israelis and the Palestinians (moratorium) as well as the event of the Holy Quran burning (9/11).

In December 2010, there are six main peak events of which four are shared between the four main (the American, the British, Global Variety English as well as the Arabic) Arab Spring sub-corpora. These are the Israeli-Palestinian peace talks; the issue of WikiLeaks documents, Lebanon-Hezbollah and the Scud missiles; and the elections in Egypt. The issue of the Christians in Iraq is only prominent in the British sub-corpus, whereas the main event that ignited the Arab Spring, that is the protests in Tunisia and the story of Bouazizi, is a peak event in the Arabic sub-corpus.

2- After the Arab Spring, there are 21 peak events (distributed across May, October and December 2011; May and September 2012, and March and August 2013):

- 2011 (May, October and December):

In May, there are five peak events of which four are shared between the two main English and Arabic sub-corpora. The shared peak events include Bin Laden's death; Obama's speech (US); Israeli-Palestinian peace negotiations; and the US sanctions against Syria. The G8 summit is peak event in the Arabic sub-corpus.

In October, the two peak events that are shared between the two main English and Arabic sub-corpora are Gaddafi's death (20 October), and Tunisia's election (23 October). The release of the Israeli soldier Shalit by Hamas on 18 October is peak event in the English sub-corpus, whereas the situation in Syria is peak event and more salient in the Arabic sub-corpus.

In December, elections in Syria and Egypt are shared peak events, whereas the Arab League talks to allow observers into Syria is mostly salient peak event in the English (mainly the American and the British sub-corpora).

- 2012 (May and September):

In May, there are three peak events of which two, Syrian violence (NU) and elections (Egypt and Algeria) are shared. The Saudi-Bahrain union is mainly peak event in the Arabic sub-corpus.

Benghazi attack and the anti-Islam film are peak events in September 2012.

- 2013 (March and July)

In March, there are four main events of which two are shared between the English as well as the Arabic sub-corpora. These are Obama's Mideast visit (Israel, the West Bank and Jordan between 20-23 March), and the conflict in Syria as the opposition in Syria takes Syria seat at the Arab summit (27 March). The Arab summit and the King Abdullah II of Jordan interview are two peak events that are more salient in the Arabic sub-corpus only.

In July, the most peak event in the Arab Spring corpus is the overthrow of the Egyptian President Mohamed Morsi (3 July 2013).

As can be seen from the summary timeline of the main events across the Arab Spring English and Arabic sub-corpora, applying corpus linguistic techniques allows for conducting diachronic as well as synchronic comparisons. By using

corpus linguistic techniques that facilitate collecting larger data volumes as well as analysing them at quantitative and qualitative levels by means of frequency lists, keyness (KKW lists in my case) and concordances, my study therefore contributes to academic research at two levels. First, I provide more evidence of the potential of corpus linguistic methodology in conducting critical discourse analysis. Secondly, and given the important role the media play in constructing social reality, this study contributes to our understanding of the language use of the news media in representing the Arab Spring narrative from the perspectives of different languages and cultures.

1.5 Structure of the Study

The main aim of this study is to investigate the Arab Spring in English and Arabic news media from CL and CDA perspectives. The current study consists of six chapters: introduction, methods and approaches, three analysis chapters, and a conclusion. Each chapter has its own introduction, main body/analysis and concluding section.

In this chapter I have introduced: The Arab Spring narrative in terms of its events; the geographical aspects of the Arab Spring; the Arab Spring in terms of its geographical location; the controversies around the term ‘Arab Spring’ and how it is defined by the *OED* (2013, online) in the context of historical citations; the research questions.

Chapter 2 is concerned with methods and approaches adopted in this study. It mainly focuses on the corpus linguistic techniques in terms of their potential, and the main areas of concern with regard to critical discourse analysis. It also discusses some frameworks in CDA in regard to news media, narrative analysis and media discourse structure. It also highlights the theoretical and methodological frameworks applied in this research. From a CDA perspective it aims to conduct lexical analysis in terms of news actors, topics, and news values, drawing on theoretical frameworks and Bell's (1991) approach to news values in news events. It also discusses the corpus linguistic techniques utilized in collecting data and designing the corpus.

Chapters 3, 4 and 5 are analytical chapters in which each chapter addresses one of the main questions stated in this study. Chapter 3 looks into the representation strategies of the main news actors in the English and Arabic sub-corpora in order to see *who* is represented in the story. Drawing on van Leeuwen's (2008, 1996) taxonomy of Social Actor Network of which the strategies of *personalization* and *impersonalization* are the most relevant, Chapter 3 examines strategies of *nomination* (as an aspect of *personalization*) and *objectivation* (as an aspect of *impersonalization*, which mainly deals with strategies of *spatialization* and utterance *autonomization*, as well as inclusion and exclusion strategies. Chapter 4 investigates the main topics associated with the Arab Spring: in other words, *what* is represented and *where* they are located. It diachronically as well as synchronically investigates similarities and differences between the different

English and Arabic sub-corpora, in terms of topics, languages (English and Arabic), as well as text types (news and editorials and opinion texts). Identifying topics allows for better understanding the nature of the Arab Spring narrative in terms of its definition, attitudes and the way it is linguistically as well as semantically represented. Chapter 5 examines the Arab Spring narrative in terms of its newsworthiness: in other words, *why* the stories are important for the media and therefore the reading public. Drawing on Bell's (1991, p. 155-8) approach to news values in news events and actors, it investigates how the Arab Spring has attracted, and is still attracting, constant international news media attention in terms of news values, such as *Eliteness*, *Proximity*, *Negativity*, *Relevance*, *Personalization* and *political agendas*. Such news values are considered as the criteria or factors 'by which one "fact" is judged more newsworthy than another' (Bell, 1991, p. 155)

Chapter 6 is the conclusion of the study. It discusses the results of the analytical chapters (3,4, and 5) and the implications of the current research in terms of the study of language in the news from the perspectives of corpus linguistics and critical discourse analysis, as well as the technical implications for translation and equivalence of meanings across languages. The chapter concludes with suggestions and recommendations for further research.

2. Methods and Approaches

2.1 Introduction

This chapter introduces the methods used to investigate the representation of the Arab Spring narrative in English and Arabic news media in terms of its key news actors and topics (KKW-semantic macrostructures), and its newsworthiness. It is divided into two main parts. The first part is concerned with data collection and corpus design. It describes the methodological procedures used for building the Arab Spring corpus in terms of the sources, text types as well as the size of the English and Arabic data. The translation as well as the transliteration scheme of the Arabic data is also introduced in this part, followed by a discussion of the reference corpora used in the study. The final section highlights some ethical issues regarding copyright and permissions concerning data collection and corpus building. The second part introduces the methodological as well as the theoretical frameworks within which this study is conducted, including the use of corpus software, *WordSmith Tools* (Scott, 2015). It also discusses the use of manual as well as automated CL analytical methods such as *Wmatrix* (Rayson, 2008, 2003) that are used for investigating the main topics (i.e. semantic macrostructures) that tend to be associated with the Arab Spring across the English as well as the Arabic sub-corpora. Section 2.9 discusses topical analysis in terms of the schematic structure of news articles and editorial texts. This is followed by discussing some of the methodological concerns in Section 2.10, and then highlighting the advantages and

limitations of the corpus approach to critical discourse analysis (CDA). Finally, the chapter concludes with a summary of what has been discussed.

2.2 Data Collection and Corpus Design

In modern linguistics a corpus ‘can be defined as a collection of [...] (1) *machine readable* (2) *authentic texts* (including transcripts of spoken data) which is (3) *sampled* to be (4) *representative* – of a particular language or language variety’ (McEnery et al., 2006, pp. 4-5, original italics). Hence, the Arab Spring bilingual corpus is a specialized corpus that has been designed to explore the Arab Spring narrative in the English and Arabic mainstream news media over a specific period of time. It is comprised of news and editorials texts taken from different Western, Arabic and Islamic broadsheet newspapers and news websites, and published over a period of 39 months, between 15 June 2010 and 31 August 2013: the period which comprises the period leading up to the protests starting in Tunisia in 2010 until the events in Turkey in the summer of 2013. This allows us to examine the temporal aspects of the discursive construction of the Arab Spring narrative. It also enables us, by combining quantitative and qualitative methods associated with corpus linguistics (CL) and critical discourse analysis (CDA), to explore the representation of the Arab Spring narrative in the news within contexts related to the *protest vocabulary* such as *protest*, *uprising*, *revolution*, *awakening*, *rebellion*, *riot* and *violence*. A corpus-based approach allows also for comparisons of how particular themes, topics and news actors across the different sub-corpora are presented

(negatively or positively) before and after the emergence of the Arab Spring narrative. The date which I have identified as the start of the narrative is 17 December 2010, the date when the Tunisian street vendor, Mohammad Bouazizi, set himself on fire as a symbolic act of protest against police mistreatment, but I have begun my corpus in June 2010 in order to see the emergence of this event in terms of media reporting of social disquiet that might have alerted Western governments to the Arab Spring.

2.2.1 Description of the Arab Spring Bilingual Corpus

The Arab Spring corpus consists of two main sub-corpora, English and Arabic, compiled mainly from the online news archive, *LexisNexis*, and other news websites. The resources were considered as elite, high-profile international news media outlets that are influential and have a level of internationally acclaimed credibility. Thus, the source of the news is one important criterion which one should consider when building a corpus. According to Bell (1991, p. 18), typical criteria for selecting the source or news outlet could be summarized in three points. The first criterion is the geographical area. Second is the audience type or size and last is the time of news production. In terms of this study, the sources selected sought to represent different areas, countries and cultures across the world, despite the fact that geographical boundaries have become a less significant issue with the advent of the Internet and other electronic media (Mautner, 2008). Selection decisions or criteria for data collection and corpus building may also differ based on the nature

of the research question. As Mautner (2008, p. 37) points out, if a certain event is to be investigated, then time (before and after the event) could be one of the criteria for selection.

The Arab Spring corpus covers news six months before the emergence of the Arab Spring narrative on 17 December 2010 and continues throughout all the major events in countries such as Tunisia, Egypt, Libya, Bahrain, Yemen, Syria, Jordan, ending with the events in Turkey in the summer of 2013. Events are still unfolding but in a counter-direction, resulting in civil wars, such as those in Syria and Libya, regional and economic instabilities and a return to the oppressive regimes that the Arab Spring suggested might be coming to an end. In addition, if the issue is polarized, data should be drawn from different media resources with different political or ideological alignments, and finally if controversy surrounds an ideologically or politically ‘loaded keyword’, the first selection criterion is then the occurrence of that word in the texts to be collected (Mautner, 2008, p. 37). Based on these major selection criteria, the Arab Spring corpus has been designed to investigate the range of views and perspectives on the Arab Spring narrative in the English and Arabic news media by means of identifying and analysing the different key concepts and topics associated with the events reported within the context of the Arab Spring narrative. The Arab Spring corpus therefore consists of two main sub-corpora, English and Arabic in order to capture a range of discourses and both Eastern and Western perspectives. The English Arab Spring sub-corpus (henceforth, ENASC) consists of 7,018 texts with a total of 5,898,333 words, while

Table 2.1 Statistics of the Arab Spring bilingual corpus

Code	Media Name	Media Type	Region	No. of Texts	Word Count
ENASC sub-corpus					
British English: BR-ENASC					
BBCe	<i>The British Broadcasting Corporation</i>	Broadcasting	UK	955	615,313
DTLe	<i>The Daily Telegraph</i>	Newspaper	UK	435	309,803
FTe	<i>The Financial Times</i>	Newspaper	UK	488	332,540
GRDe	<i>The Guardian</i>	Newspaper	UK	485	375,010
NDPe	<i>The Independent</i>	Newspaper	UK	514	394,560
TMSe	<i>The Times</i>	Newspaper	UK	417	301,580
American English: AM-ENASC					
CNNe	<i>Cable News Network</i>	Website	USA	707	587,100
IHTe	<i>The International Herald Tribune</i>	Newspaper	USA	540	559,589
NYTe	<i>The New York Times</i>	Newspaper	USA	809	755,183
WPe	<i>The Washington Post</i>	Newspaper	USA	560	541,511
Global Variety English: GV-ENASC					
JPTe	<i>The Jerusalem Post</i>	Newspaper	Israel	856	893,957
HRTe	<i>Hurriyet Daily News</i>	Newspaper	Turkey	103	62,933
ZMNe	<i>Today's Zaman</i>	Newspaper	Turkey	149	172,337
Sub-Total				7,018	5,901,416
Modern Standard Arabic: MSA-ARASC					
HRMa	<i>Al-Ahram</i>	Newspaper	Egypt	398	279,958
ALMa	<i>Al-Alam News</i>	Website	Iran	752	383,440
MHRa	<i>Mehr News Agency</i>	Website	Iran	595	214,461
BWBa	<i>Al-Bawaba</i>	Website	Jordan	627	332,040
DUSa	<i>Ad-Dustour</i>	Newspaper	Jordan	927	546,543
SFRa	<i>As-Safir</i>	Newspaper	Lebanon	885	759,782
QAa	<i>Al-Quds Al-Arabi</i>	Newspaper	Palestine	1,017	900,791
JAZa	<i>Al-Jazeera</i>	Website	Qatar	1,420	1,132,068
RBYa	<i>Al-Arabiya news</i>	Website	Saudi Arabia	552	297,471
WSTa	<i>Asharq Al-Awsat</i>	Newspaper	Saudi Arabia	897	774,876
Sub-Total				8,070	5,621,430
Grand Column Total				15,088	11,522,846

the Arabic Arab Spring sub-corpus (henceforth, ARASC) comprises 8,070 texts and a total of 5,621,430 words totalling altogether 15,088 news texts and 11,519,763 words (Table 2.1).

To aid the comparative aspects of the study, the corpus has been divided into a number of sub-corpora in terms of language: English and Arabic (13 English sub-corpora and 10 Arabic sub-corpora) and world variety (6 sub-corpora for British English, 4 sub-corpora for American English, 3 sub-corpora for global variety of English, and one sub-corpus for the Arabic/ MSA- Modern Standard Arabic). The corpus has also been subdivided into text types (news and editorial & opinion), date of editorial & opinion) and date of publication (39 monthly and 4 yearly sub-corpora). As Table 2.1 shows, the Arab Spring bilingual corpus consists of a total of 23 sub-corpora (13 English and 10 Arabic). Each sub-corpus is given a reference code to indicate the name of the media outlet from which news articles have been collected. The capital letters represent the media name (e.g. BBC) whereas the small letters, *e* and *a*, indicate the language of the text: *e* for English and *a* for Arabic. What is more, as texts were taken within a time span from 15 June 2010 to 31 August 2013, each text file was given a unique name that shows the source from which it was taken, language, year of publication and serial number. For example, the British English file BBCe-10-01 indicates that the article is taken from the *British Broadcasting Corporation (BBC)*, its language is English (*e*), and is published in 2010 (-10) and is the first one in the folder (-01). Similarly, in Arabic, the Arabic file JAZa-10-01 indicates that the file is from *Al-Jazeera Satellite*

Channel (JAZ), its language is Arabic (*a*), published in 2010 (-10), and it is number one (-01) in the JAZ sub-corpus. This procedure has been applied to all text files in the Arab Spring sub-corpora. Furthermore, records of all sub-corpora were kept in a separate spreadsheet, including all metadata such as reference code, title, date of publication, media type, text type, author or resource, word count, website address if available, country of origin, and in some cases images and captions that came with the texts. An example of the metadata is shown in Appendix A. In addition, extratextual information was also kept in the text files and marked up (to ignore) through *WordSmith Tools* mark-up setting <*> which allows for cutting ‘out all wording starting at each < symbol and ending at the next > symbol up to 200 characters apart’ (Scott, 2015, p. 116). Mark-up is useful in enabling the analyst to relate the text to its original context (Mautner, 2009a; McEnery et al., 2006). This was applied to the English sub-corpora but not the Arabic as it was too time-consuming and a limitation that has to be lived with. Table 2.1 also shows that there are two different media types (news websites and newspapers) which represent two different languages (English and Arabic), and come from different Western, Arab and Islamic countries and cultures. However, it is worth mentioning that two pan-Arab newspapers, *Al-Quds Al-Arabi* (QAa) and *AShraq Al-Awsat* (WSTa) are published in London, but owned by Palestinian expatriates and a member of the Saudi royal family respectively. In addition, although most of the Arab Spring corpora (English and Arabic) were collected from *LexisNexis*, some of them were manually collected from their online news archives. There were a couple of reasons

for this; either the media outlet archive was not available through *LexisNexis* as in the case of the *BBC*, *Al-Jazeera* and *Al-Quds Al-Arabi*, or data were not available until mid-2011, as was the case with *Asharq Al-Awsat*, *Al-Bawaba*, *Al-Ahram*, *Mehr* and *Al-Alam*. Data were obtained for all the news media outlets from 15 June 2010 until 31 August 2013 except for the following news media outlets: *Al-Ahram* (HRMa) of Egypt and *Al-Alam* (ALMa) of Iran, as data between June 2010 and mid-April 2011 were not available online. However, inclusion of these two corpora from Egypt and Iran was considered as important, as they represent countries which are seen worldwide as being influential in the Middle East and the rest of the world. Additionally, two English sub-corpora, *Hurriyet Daily News* (HRTe) and *Today's Zaman* (ZMNe) of Turkey, and the *Jerusalem Post* (JPTe) of Israel, were collected from their English versions online and hence representative of different cultures as well as forming a global variety of English (GV). While the JPTe sub-corpus covers news from 15 June 2010 to 31 August 2013, the Turkish sub-corpora only cover the Arab Spring news between 1 May 2013 and 31 August 2013, since this was an event that occurred after I thought I had completed my data collection, but which I thought important to include. The main purpose of collecting data from Turkish news media is to explore the representation of the Arab Spring narrative during the unrest in Turkey that broke out during May 2013.

Ultimately, there is no doubt that an element of subjective judgement is always involved in the process of data selection, which cannot be completely avoided but might be justified and counterbalanced by means of transparency, accountability

and the use of established approaches to data selection, sampling (Mautner, 2008, p. 37) and sampling frames such as size, time and text type or subject (Biber, 1993; Baker et al., 2008). What is more, given that Arabic data were collected from different Arab and Islamic (non-Arabic-speaking such as Iran) countries in the Middle East, data from the Iranian news media outlets (ALMa and MHRa) were collected from their Arabic versions available on their websites as well as *LexisNexis*. In this context, it should be highlighted that Arabic has three major forms. These are the formal Arabic language or Classical Arabic, which is the language of the Quran, and is considered to be the base of the syntactic and grammatical norms of the Arabic language, Modern Standard Arabic (MSA) *الفصحى* <al-fuṣḥā> which is widely used by news media and modern literature, and finally colloquial Arabic, ‘which is the form of Arabic used in every day oral communication’ (Al-Sulaiti and Atwell, 2006, p. 148; Alansary and Nagi, 2014, p. 8). The Arabic data in the Arab Spring corpus are considered as representative of Modern Standard Arabic (MSA- referred to as MSA-ARASC) as this form of Arabic is the official form of language that is widely used by news media and other official and formal institutions throughout the Arab and Islamic world (Ryding, 2005, p. 5). Regarding English data, Table 2.1 also shows that the English sub-corpus is representative of three English forms: British, American and a global variety of English. The ENASC sub-corpus has been accordingly subdivided into three sub-corpora: BR-ENASC, AM-ENASC and GV-ENASC as they are

representative of the British, the American as well as global varieties of English sub-corpora respectively.

Figure 2.1 shows the geographical distribution of the Arab Spring corpus and its composite sub-corpora. The countries to which the ENASC as well as the ARASC sub-corpora belong are highlighted in red. The ENASC sub-corpora are presented in blue shapes, whereas their ARASC counterparts are presented in yellow. In terms of the text numbers as well as the word count, which account for the whole corpus size, Table 2.1 also shows that the ARASC sub-corpus is slightly above 1000 more texts than that of ENASC. However, it is the ENASC sub-corpus that is higher in terms of word count and the number of its subdivisions, totalling 13 sub-corpora, as shown in Figure 2.1:

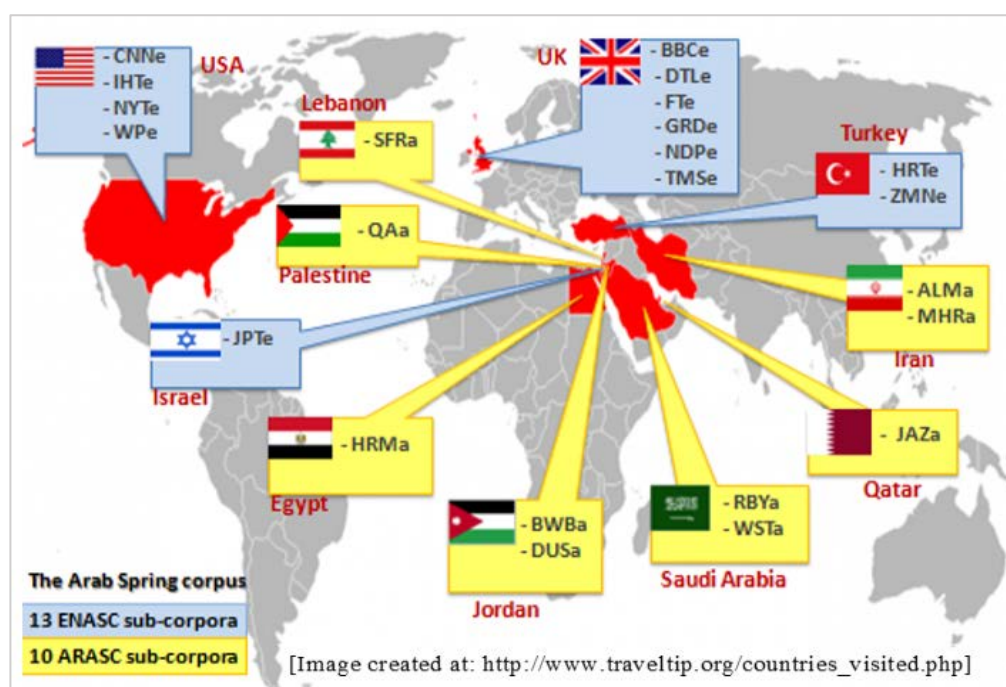


Figure 2.1 Geographical distribution of the Arab Spring sub-corpora

In the same context, imbalance in the sizes of the sub-corpora can be attributed, as Hardt-Mautner (1995) points out, to the amount of news coverage devoted to the political events or topics under investigation. Another reason for the imbalance in the sizes of the different sub-corpora is that some newspapers produce more, or longer articles than others, for example the difference between news reports and opinion articles. Another reason is availability of or access to data online. Moreover, compiling a specialized corpus requires being selective when choosing data for the purposes of studying a particular subject or language variety, and the issue of data quality or content ‘takes [therefore] equal or more precedence over the issues of quantity’ (Baker, 2006, p. 29). Figure 2.2 shows the overall proportion of texts and words in the four main Arab Spring sub-corpora.

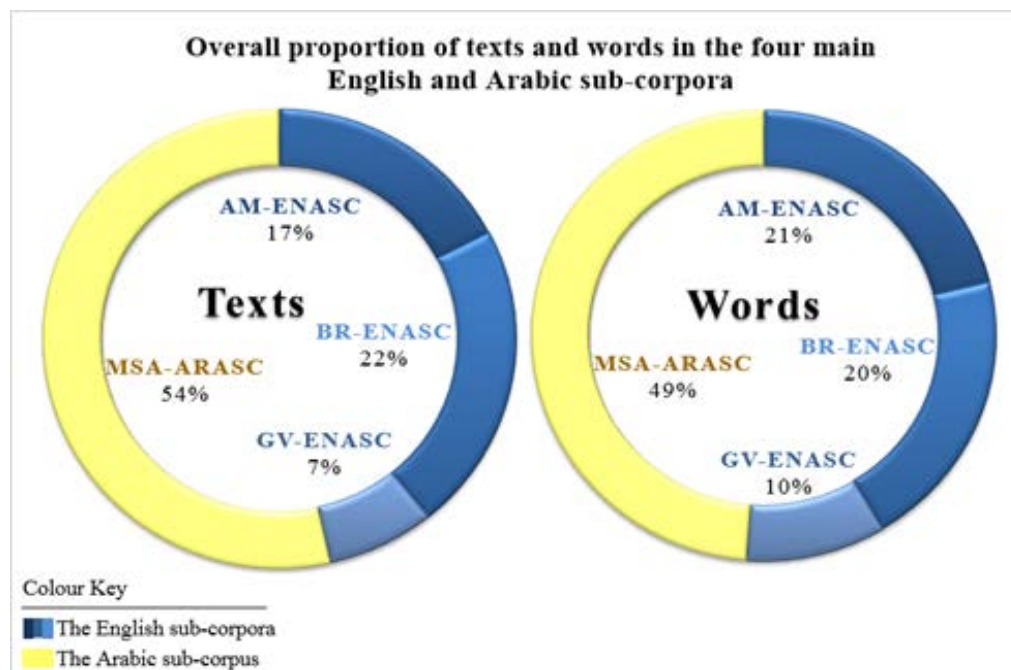


Figure 2.2 Proportion of texts and words in the Arab Spring sub-corpora

Nevertheless, this slight difference, as Figure 2.2 shows, is not an issue, since the two sub-corpora are comprised of a reasonably large number of texts that are representative of their sources (English and Arabic news media), media type (newspapers and news websites), topic (Arab Spring), texts types (news and editorial and opinion, and time span (15 June 2010-31 August 2013), which can enable investigating and analysing language patterns. Additionally, the size of samples was not defined, as the aim was to include whole texts rather than samples of a specific word count. Thus, although representativeness and balance are considered as crucial in corpus building (McEnery et al., 2006, p. 21; Biber, 1993; Meyer, 2002; Biber et al., 1998), they ‘remain largely heuristic notions’ (Baker et al., 2008, p. 10) which are not based on reliable scientific measures, but rather rely ‘heavily on intuition and best estimates’ (McEnery et al., 2006, p. 16; Hunston, 2002) or the informed and intuitive judgement of the corpus compilers (Leech, 2007; Sinclair, 2005). Judgement is especially exercised in terms of sampling techniques or frames that are concerned with issues such as: sample size (full texts or sized samples), classification of text types, categories or genres and time. By the same token, representativeness and balance might also become problematic and could not fully be achieved, especially when other aspects or considerations, such as comparability, are sought in corpus design (cf. Leech, 2007, pp. 141-3), which might also contain sub-corpora compiled for comparative purposes either between different forms of a given language (e.g. American and British English) or more than one language such as English and Arabic (Baker et al., 2008). In other words,

a specialized corpus could be designed to represent a particular text type or subject (Baker, 2006; Hunston, 2002). For example, Johnson et al. (2003, p. 29) compiled a corpus of articles from three broadsheet newspapers in the UK between 1994 and 1999 to explore the discourse of ‘political correctness’. Only articles that contained the search terms *political correctness*, *political incorrectness*, *politically correct*, *politically incorrect* and *PC* (Johnson et al., 2003, p. 31) were included. Similarly, Taylor (2010, p. 221, original italics) analysed the ‘changing rhetorical role of *science* in UK broadsheet newspapers from 1993 and 2005’ in the SiBol corpora. To aid investigation of the research question, Taylor (2010, p. 225) compiled two additional sets of corpora consisting of all articles which contained the search term *scien** from 1993 and 2005 in other two major UK broadsheet newspaper the *Independent* and the *Daily Mail*. Hence, the size of the corpus is perhaps not the key issue when building a specialized corpus ‘to investigate the discursive construction of a particular subject, but how often we would expect to find that subject mentioned within it’ (Baker, 2006, p. 28). From that perspective, the Arab Spring corpus is a specialized and comparable corpus that contains two main sub-corpora, English and Arabic. Both corpora are designed to be representative of two text types (news as well as editorials and opinions), covering a period of 39 months between 15 June 2010 and 31 August 2013.

Table 2.2 Breakdown of the Arab Spring sub-corpora

Sub-corpus	News		Editorials and opinions		Grand Row Total	
	Texts	Words	Texts	Words	Text	Words
AM-ENASC	1,803	1,689,432	813	753,951	2,616	2,443,383
BR-ENASC	2,234	1,456,063	1,060	872,743	3,294	2,328,806
GV-ENASC	428	363,307	680	765,920	1,108	1,129,227
MSA-ARASC	5,309	3,150,909	2,761	2,470,521	8,070	5,621,430
Grand Column Total	9,774	6,659,711	5,314	4,863,135	15,088	11,522,846

The English sub-corpus, is also representative of three English forms: British, American and global variety of English as shown in Table 2.2. Table 2.2 also shows that the total number of news texts (9,773) accounts for about 65% of all the texts, whereas the editorial and opinion texts (5, 315) represent 35% of the whole corpus. The classification of text types was mainly based on either the extra textual information provided by *LexisNexis* or on topics as they were presented in the newspapers and websites from which data were collected. World news and news reports, covering relevant events and topics on the Middle East before and after the emergence of the Arab Spring during the specified time frame, were classified under News, whereas editorials, comments, opinions, leaders and feature articles were classified as Editorials and opinion.

The overall number of texts reflects the high level of newsworthiness of the Arab Spring and its related stories as well as the greater attention it has received from global news media. The BR-ENASC, for example, is the highest in terms of the total number of texts which amounts to 3,294, whereas the GV-ENASC has the lowest figures in the ENASC corpus, because this sub-corpus was collected at the end of the data collection period, directly as a result of the events unfolding in Turkey at the time. MSA-ARASC, on the other hand, is the highest of them all in terms of the total number of texts and words, reflecting the prominence of this news in the Arabic language media. Data are therefore considered as sufficient to reflect the level of newsworthiness of the Arab Spring in each media source and figures in Table 2.2 offer an insight into the levels of information which readers might receive on the Arab Spring and its related stories. For instance, the number of news texts in the GV-ENASC (428 News texts) as well as the BR-ENASC (2,234 News texts) shows that the latter offers more information about the Arab Spring news if it is compared with GV-ENASC. The fact that the number of news texts of the GV-ENASC sub-corpus is lower than its other English varieties (the American and the British) counterparts is because that it contains two newspapers sub-corpora (that is, the two Turkish sub-corpora, ZMNe and HRTe: Table 2.1) that contain texts covering only the later time period, that is between the months of 1 May and 31 August 2013. Only the *Jerusalem Post Newspaper* (JPTe) sub-corpus of the GV-ENASC covers the whole period (between 15 June 2010 and 31 August 2013)

specified for this study. In other words, the search terms (see section 2.2.3) were not used in collecting data from the two Turkish sub-corpora mentioned above to cover the earlier time period (between 15 June 2010 and 31 April 2013), as events in Turkey relating to the Arab Spring started to unfold by late May 2013, and hence can be said that the GV-ENASC might not be fully representative of the time period specified for this study.

However, there are other considerations in terms of *news values* which might help in exploring how certain events are selected to become news, such as relevance, type of readership and political or ideological agendas of such news media and the powers behind them (Harcup and O'Neill, 2001; O'Neill and Harcup, 2009; Galtung and Ruge, 1965; Bell, 1991; Fowler, 1991; van Dijk, 1991; Bednarek and Caple, 2014). In addition, for critical discourse analysis, a corpus might not reflect 'the "power" of individual texts' (Baker, 2006, p. 19, original emphasis) or other semiotic aspects or visuals such as images, layout or graphics, which are considered by CDA as crucial 'to the creation of meaning' (Mautner, 2009a, p. 34; van Leeuwen, 2008; Wodak and Meyer, 2009; Hardt-Mautner, 1995). Taking into consideration the different sources and languages of the Arab Spring news data, the following section discusses the transcription scheme for the presentation of the Arabic data.

2.2.2 Arabic Data: Transcription Scheme

One of the main issues encountered in developing a query to collect data was the difference between the two languages under investigation. Hence, translation and transliteration of Arabic have been conducted to enable the English reader to understand and compare between the two languages in terms of the information and examples provided in this study. Given the amount of Arabic data, transliteration was conducted by *Intellibe*, an Arabic-to-Latin text transcriber: software program available online [<http://www.intellaren.com/Intellibe>]. The program automatically converts Arabic scripts, as diacritics or vowels are added, to Latin (or English) script transliteration. The transliteration follows the *Deutsches Institut für Normung* or the ‘DIN 31635’ standard for the transliteration of Arabic alphabet (Wikipedia, 2013) as shown in Table 2.3. According to the *Deutsches Institut für Normung* website (<http://www.din.de/en>), the standard for the new edition of DIN 31635, specifies the transliteration of the letters of the Arabic alphabet into Latin letters. The DIN 31635 standard is largely based on the recommendations accepted by the International Orientalist Congress 1936 with regard to several languages, including Arabic:

Table 2.3 DIN31635 standard for the transliteration of Arabic alphabet

Names of the letters		Independent form	DIN 31635
'alif	ألف	ا	'/ā
bā'	باء	ب	b
tā'	تاء	ت	t
ṭā'	ثاء	ث	ṭ
jīm	جيم	ج	ǧ
ḥā'	حاء	ح	ḥ
ḵā'	خاء	خ	ḵ
dāl	دال	د	d
ḏāl	ذال	ذ	ḏ
rā'	راء	ر	r
zayn	زين	ز	z
sīn	شين	س	s
šīn	سين	ش	š
ṣād	صاد	ص	ṣ
ḏād	ضاد	ض	ḏ
ṭā'	طاء	ط	ṭ
ḏā'	ظاء	ظ	ḏ
'ayn	عين	ع	'
ḡayn	غين	غ	ḡ
fā'	فاء	ف	f
qāf	قاف	ق	q
kāf	كاف	ك	k
lām	لام	ل	l
mīm	ميم	م	m
nūn	نون	ن	n
hā'	هاء	ه	h
wāw	واو	و	w / ū
yā'	ياء	ي	y / ī

The Arabic data and examples are presented in the form of interlinear glosses following the Leipzig Glossing Rules (Comrie et al., 2008). Diacritics shown in Table 2.3 are explained in sub-section 2.2.3 below. As Example 6 shows, each line of the interlinear gloss shows a different kind of information. The first line represents a short chunk of the Arabic extract shown in Example 6. The second line is a transliteration of the Arabic text shown in the first line, whereas the third line is a literal word and morpheme ordered translation of Arabic into English (Abdullah, 2014, p. 43). The last italicized line is an idiomatic English translation of the Arabic text shown in the first line. As Example 8 (parts 1 and 2) shows, words and morphemes are separated by hyphens (e.g. al-jihadiyy-ūn – the-jihadist-s), which gives the reader a better understanding of the way Arabic words are constructed, and how suffixes can be added to roots (Abdullah, 2014, p. 44). Grammatical category labels for words and morphemes, are sometimes added and printed in small capitals and separated by periods in the literal translation as shown in the third lines of the gloss.

8. [HEADLINE] بريطانيا: الجهاديون في سوريا يشكّلون تهديداً لأوروبا
Britain: Jihadists in Syria pose a threat to Europe
- 1 بريطانيا: الجهاديون في سوريا
bariṭānyā: al-jihadiyy-ūn fī sūryā
Britain: the-jihadist.MASC.N-S.NOM.PL in Syria
Britain: The jihadists in Syria
- 2 يشكّلون تهديداً لأوروبا
yuṣakkil-ūna tahdīd-an li-'urubbā
pose.MASC.PRS.V-**they**. threat.MASC.N-ACC to-Europe
pose a threat to Europe

(BWBa-13-22/ news: 14-Feb)

Depending on the purpose of the gloss, however, either a word or a category label from the target language, that is, English can be used. For example, the grammatical morpheme *-ūna* of the item <yušakkil-ūna>, as indicated (in bold) in Example 8 (part 2), is rendered as ‘-they’. In terms of its grammatical category labels, *-ūna* can be rendered as 3MASC.PL.NOM. This kind of glossing reflects the ‘one-to-many-correspondences’ (Comrie et al., 2008) of the Leipzig Glossing Rules. It enables the reader to better understand the relationship between the original text as shown in the first line and its idiomatic translation in the fourth line, which is printed in italics. In this context, it is worth mentioning that Arabic grammar is also informed by the Quranic Arabic Corpus (Dukes, 2009, online), ‘an annotated linguistic resource which shows the Arabic grammar, syntax and morphology for each word in the Holy Quran’. The following section briefly discusses some of the characteristic features of the Arabic language, and highlights some of the major differences between English and Arabic.

2.2.3 Differences Between English and Arabic

This sub-section provides a brief description of the Arabic language characteristic features would help in outlining the difference between the two languages. According to Ryding (2005), Arabic is a Semitic language that belongs to the Afro-Asiatic language family, and has twenty-eight (28) consonants of which two are semivowels, that is, *و* <wāw – و – w>; and *ي* <yā’ – ي – y>; these semivowels serve, based on context, either as consonants or as vowels. For example, the letter

<wāw> represents either the sound of /w/ as in the word ولد <walad – boy>; or the long vowel /ū/ as in the word نور <nūr – light>.

Arabic script has a number of distinctive features. First, it is written and read from right to left and written in cursive style rather than print-writing. Its letters have no distinction between uppercase and lowercase, but rather take variant forms, as shown in Table 2.4, depending on whether they are independent (or isolated), initial, medial or final in a word. (Ryding, 2005, p. 11; Haywood and Nahmad, 1976, p. 3). Table 2.4 gives the name of the Arabic letters along with their shapes. The cursive nature of Arabic script, as Table 2.4 shows, requires several forms for each letter (Ryding, 2005, p. 12). Most letters are linked together on both sides when they are medial, but there are six letters (و ز ر ذ د ا), as shown in bold in Table 2.4, which can only be connected to a preceding letter. These six letters are called by Ryding (2005, p. 12) ‘non-connectors’ as they cannot be attached to a following letter. Arabic words containing the non-connectors are shown in Example 9 (from lines 1 to 6) along with their transliteration and English translation:

9. 1 الاستبداد <al-’istibdād – *the-repression*>
- 2 الذي <allaḍī - *which.MASC*>
- 3 المعارضة al-mu’āraḍa – *the-opposition*>
- 3 حركة <ḥaraka – *movement*>
- 5 الحزب <al-ḥizb – *the-party*>
- 6 النووي <an-nuwawī - *the-nuclear*>

Table 2.4 Names and Shapes of Arabic Letters

Letter Name	Independent	Initial	Medial	Final
'alif	ا	ا	ا	ا
bā'	ب	ب	ب	ب
tā'	ت	ت	ت	ت
ṭā'	ث	ث	ث	ث
jīm	ج	ج	ج	ج
ḥā'	ح	ح	ح	ح
kā'	خ	خ	خ	خ
dāl	د	د	د	د
ḍāl	ذ	ذ	ذ	ذ
rā'	ر	ر	ر	ر
zayn	ز	ز	ز	ز
sīn	س	س	س	س
šīn	ش	ش	ش	ش
ṣād	ص	ص	ص	ص
ḍād	ض	ض	ض	ض
ṭā'	ط	ط	ط	ط
ḍā'	ظ	ظ	ظ	ظ
'ayn	ع	ع	ع	ع
ḡayn	غ	غ	غ	غ
fā'	ف	ف	ف	ف
qāf	ق	ق	ق	ق
kāf	ك	ك	ك	ك
lām	ل	ل	ل	ل
mīm	م	م	م	م
nūn	ن	ن	ن	ن
hā'	ه	ه	ه	ه
wāw	و	و	و	و
yā'	ي	ي	ي	ي

Furthermore, Arabic has diacritics and *sukūn* الحركات والسكون <al-ḥarakāt wa-s-sukūn>. The set of three short vowels are *fathā*, *ḍamma* and *kasra* which consist of the sounds /a/, /u/ and /i/ respectively. The diacritics are placed either above the consonant like *fathā* َ and *ḍamma* ُ or underneath it as in the case of *kasra* ِ. When these vowels are lengthened, they are followed by the letters 'alif, wāw and yā' (أ, و and ي) consisting of the sounds /ā/, /ū/ and /ī/ respectively. There are also two main diphthongs in Arabic which are formed when a short vowel and a semivowel واو or ياء <wāw - و or yā' - ي> are combined in a word (Ryding, 2005). These are /ay/ as in the word بَيْت <bayt – house.SING.MASC.N>, and /aw/ in فَوْق <fawq – above.PREP - above>; when the previous consonant has *fathā* َ, semivowels must have سُكُون ْ <*sukūn* – silence> (Haywood and Nahmad, 1976, p. 9; Ryding, 2005). سُكُون <*Sukūn* - silence> indicates the absence of a vowel. It is represented by a small circle ْ and placed directly above the letter as in the word غُرْفَة <ḡurfa – room.SING.FEMN.N - a room> (Ryding, 2005; Haywood and Nahmad, 1976). These diacritics are referred to in Arabic as حركات <ḥarakāt - movements> (Ryding, 2005). They are shown in Table 2.5 with examples. Short vowels are normally not written into a text as they are expected to be guessed by the Arabic readers, except for in the Holy Quran and children's schoolbooks, as, in the case of the Holy Quran, all diacritics must be added in order to follow the correct pronunciation and avoid any ambiguity when reading the sacred text (Ryding, 2005, p. 30).

Table 2.5 Arabic diacritics with examples

Diacritics and Word Examples			
Arabic Diacritics	written Form	Arabic	
فتحة - fatha	/a/	بَلَدٌ	balad country.SING.MASC.N <i>country</i>
	/ā/	طَالِبٌ	ṭālib student.SING.MASC.N <i>student</i>
ضمة - ḍamma	/u/	سُفُنٌ	sufun ships.PL.FEM.N <i>ships</i>
	/ū/	نُورٌ	nūr light.SING.MASC.N <i>light</i>
كسرة - kasra	/i/	مِشْمِشٌ	mišmiš apricots.PL.MASC.N <i>apricots</i>
	/ī/	سَفِينَةٌ	safīna ship.SING.FEM.N <i>ship</i>
سكون - sukūn		غُرْفَةٌ	ḡurfa room.SING.FEM.N <i>room</i>
Diphthongs	/ay/	بَيْتٌ	bayt house.SING.MASC.N <i>house</i>
	/aw/	يَوْمٌ	yawm day.SING.MASC.N <i>day</i>

In the case of nunation <تَوِين – *tanwīn*>, that is, the addition of the final nūn /n/ sound to a noun or adjective, the above vowels are written double, and pronounced with a final /n/ sound. These are tanwīn fatḥh ً, tanwīn ḍam ُ and tanwīn kasr ِ, which are pronounced with the final /n/ sounds /an/, /un/ and /in/ respectively. They are used as indefinite markers, as in the case of ‘a’ or ‘an’ in English, above nouns, adjectives and adverbs depending on their grammatical form in a sentence such as accusative, nominative or genitive (Ryding, 2005; Haywood and Nahmad, 1976). Table 2.6 shows an illustration of the Arabic indefinite noun <بيت – bayt – house.SING.MASC.N- *a house*>. Note that with the accusative form and tanwīn fatḥh ً, the letter <’alif – ا – A> is added. But if the word ends in <تاء مربوطة – tā’ marbūṭa – joined T>; the ’alif is not added, as in the word <مدرسة – madrasat-an – school.INDEF.SING.FEM.N- *a school*>.

Table 2.6 Nunation sounds in Arabic

تَوِين (tanwīn – <i>nunation</i>) sounds			Word Example
tanwīn fatḥh	ً	/an/	بَيْتًا bayt-an house-INDF.ACC <i>a house</i>
tanwīn kasr	ِ	/in/	بَيْتٍ bayt-in house-INDF.GEN <i>a house</i>
tanwīn ḍam	ُ	/un/	بَيْتٌ bayt-un house-INDF.NOM <i>a house</i>

Although diacritics are not considered as letters in Arabic, همزة (ء – Hamza – a glottal stop) often stands as a separate letter in Arabic writing. It is written either by itself /ء/ as in ماء <mā' - water> or it appears as a diacritic above or under 'alif, consisting of the sounds /'a/, /'u/ and /'i/, and above واو <wāw> and ياء <yā'>. As shown in Table 2.7, the different sounds and pronunciation of Hamza are based on its position in the word. If a *hamza* with *fathā* (أ) is followed by the long vowel 'alif (ا) with the sound /ā/, the hamza and fathā (أ) are dropped, and the symbol َ or مدّة <madda - extension>; is used above the 'alif (ا) sounds as /'ā/, as in مرآة <mir'ā – mirror.SING.FEM.N. - mirror>; and آسيا <'āsyā – Asia> (Haywood and Nahmad, 1976; Ryding, 2005). Another diacritical symbol, or حَرَكَة <ḥaraka>; is شدة <šadda> which indicates تشديد <tašdīd – gemination> or the doubling of a consonant letter (Ryding, 2005; Haywood and Nahmad, 1976). Instead of writing the letter twice, šadda (ّ) is written above the doubled consonant to show it is pronounced with double emphasis or strength (Ryding, 2005; Haywood and Nahmad, 1976), as in the word حُرِّيَّة <ḥurriyya– freedom.SING.FEM.N - freedom>, and قِصَّة <qīṣṣa – story.SING.FEM.N - story>. Another difference between Arabic and English is the definiteness and indefiniteness markers. There is no indefinite article in Arabic that is equivalent to the 'a' and 'an' in English. Indefiniteness is indicated, however, by the presence of تنوين (tanwīn – *nunation*) at the end of a word (see Table 2.6). (Haywood and Nahmad, 1976; Ryding, 2005).

Table 2.7 Sounds and pronunciation of Hamza

Hamza ٠	Initial	Medial	Final
on 'alif ا	أ / 'a/ أخ 'ak <i>brother</i>	أ سألت sa'alat <i>she asked</i>	أ بدأ bada'a <i>he started</i>
	أ / 'u/ أخت 'uḵt <i>sister</i>	أ / 'a/ رأس ra's <i>head</i>	أ اتكأ 'ittaka'a <i>he reclined</i>
	إ / 'i/ إسلام 'islām <i>Islam</i>		
on wāw و		و / 'ū/ رؤوس ru'ūs <i>heads</i>	
		و / u' / سؤال su'āl <i>question</i>	و / u' / لؤلؤ lu'lu' <i>pearl</i>
on yā' ي		أ / 'a/ هيئة hay'a <i>organization</i>	أ / 'i' / دافئ dāfi' <i>warm</i>
separate ء		ء / ' / إجراءات 'ijrā'āt <i>measures</i>	ء / ' / شيء šay' <i>thing</i>

Indefiniteness, on the other hand, is marked with the definite article ال <al or 'alif-lām – like 'the' in English>, which is prefixed and attached to either a noun or an adjective (Ryding, 2005; Haywood and Nahmad, 1976). Unlike English, the Arabic definite article ال <al> cannot be written separately. However, the lām <ل - L> in ال <al> is sometimes written and pronounced, as in the word القمر <al-qamar – the-

moon>, whereas in other cases it is written but not pronounced, as in the word الشمس *aš-šams* – *the-sun*>, based on the type of the letter with which the word begins. Therefore, the Arabic alphabet, as Table 2.8 shows, is divided into two groups, consisting of 14 letters each: الحروف الشمسية <al-ḥurūf aš-šamsiyya - *the Sun Letters*>; and الحروف القمرية <al-ḥurūf al-qamariyya - *the Moon Letters*> (Haywood and Nahmad, 1976; Ryding, 2005). When a word begins with one of الحروف الشمسية <al-ḥurūf aš-šamsiyya - *the Sun Letters*>, the lām <ل – L> in ال <al-the> assimilates to the first letter of that word, resulting in a doubled consonant, which is not written twice, but rather a šadda ّ is written over it to indicate its being doubled (Ryding, 2005; Haywood and Nahmad, 1976). For example, as in the words

الصَّحْوَة	الثَّوْرَة	التَّمْر د	الشَّرْق
aš-ṣaḥwa	aṭ-ṭawra	at-tamarrud	aš-šarq
the-awakening.FEM	the-revolution.FEM	the-revolt.MASC	the-east.MASC
<i>the awakening</i>	<i>the revolution</i>	<i>the revolt</i>	<i>the east</i>

In this case, the lām <ل – L> is written but not pronounced whereas in the case of the Moon Letters, it is written and pronounced, as in the words

الاحتِجَاج	العُنف	الغرب	المكان
al-’iḥtijāj	al-’unf	al-ḡarb	al-makān
the-protest.MASC	the-violence.MASC	the-west.MASC	the-place.MASC
<i>the protest</i>	<i>the violence</i>	<i>the west</i>	<i>the place</i>

Both cases of the Sun and Moon Letters have been considered in the English transliteration conducted in this study.

Table 2.8 Sun and Moon letters in Arabic

الحروف الشمسية	Sun letters	الحروف القمرية	Moon letters
ت	t	ا	'/ā
ث	ṭ	ب	b
د	d	ج	ǧ
ذ	ḏ	ح	ḥ
ر	r	خ	ḫ
ز	z	ع	ʿ
س	s	غ	ǧ
ش	š	ف	f
ص	ṣ	ق	q
ض	ḍ	ك	k
ط	ṭ	م	m
ظ	ẓ	ه	h
ل	l	و	w / ū
ن	n	ي	y / ī

Finally, in terms of its basic sentence structure and word order, Arabic follows the Verb-Subject-Object (VSO) pattern. Other essential principles of sentence structure are *agreement* <or مطابقة - muṭābaqa> of sentence elements or features as well as their order. Agreement refers to the matching between the verb and its subject or between noun and its dependents such as adjectives or pronouns (Ryding, 2005). For example, a singular feminine noun takes a singular feminine adjective which comes after the noun it describes as shown in Example 10 (part 2):

10. كتب محمد رسالة قصيرة

1 كتب محمد

katab-a

wrote.SING.MASC.PST.V-PFV

Mohammad wrote

Muḥammad-un

Mohammad-NOM

2 رسالة قصيرة

risālat-an

letter.SING.FEM.N-INDEF.ACC

A short letter

qaṣīrat-an.

short.SING.FEM.ADJ-INDEF.ACC

In Example 10 (part 2), the adjective <رسالة قصيرة> <qaṣīrat-an – short.SING.FEM.ADJ-INDEF.ACC> agrees with the noun <رسالة> <risālat-an-letter.SING.FEM.N-INDEF.ACC> in all its inflections: gender, number and indefiniteness. If the word <رسالة> is dual, it then becomes <رسالتين> <risālat-ayni - letter.FEM.N-s.DU.ACC - two letters>. This means that there are three number categories in Arabic: المفرد <al-mufrad – the singular>, المثنى <al-muṭannā – the dual> and الجمع <al-jam‘ - the plural> (Haywood and Nahmad, 1976; Ryding, 2005). The dual in Arabic has two forms according to its grammatical case. In the nominative case, the suffix <ان> <āni> is added to the singular from as in the word ولدان <walad-ān – boy.SING.MASC.N-s.DU.NOM – two boys>, whereas in the accusative as well as the genitive cases, the suffix <ين> <ayn> is added to the singular form as in ولدَيْن <walad-ayn – boy.SING.MASC.N-s.DU.ACC.GEN – two boys> (Ryding, 2005, p. 129; Haywood and Nahmad, 1976, p. 40). As far as translation and transliteration are concerned, not all حركات <ḥarakāt – diacritics> are provided with the Arabic data and examples given in this study; however, they were considered during the transliteration process in order to reflect the exact pronunciation as well as meaning that a word should represent.

As can be seen from the discussion above, there are fundamental differences between English and Arabic. Before getting further into discussing the search terms used in collecting the English and Arabic data, the following sub-section highlights some of the challenges in cross-linguistic corpus-based studies at the theoretical, translational as well as the methodological levels.

2.2.4 Some Challenges in Cross-Linguistic Corpus-Based Analysis

Despite the benefits of employing the various corpus linguistic (CL) techniques in discourse studies, comparative ‘analysis of two or more languages does, however, bring a range of additional challenges’ (Taylor, 2014, p. 373), especially when such languages (in my case English and Arabic) belong to different writing systems and represent different cultures. This sub-section highlights therefore some of the main challenges related to this study, and which have been previously raised and discussed by some important cross-linguistic CADS studies conducted, for example, by Freake et al. (2011), Vessey (2013) and Taylor (2014). CADS (or corpus-assisted discourse studies, see sub-section 2.2.5 for more details) approach is a sub-set of corpus linguistics which allows for the ‘investigation and comparison of features of discourse types, integrating into the analysis, where appropriate, techniques and tools developed within corpus linguistics’ (Partington, 2010).

One of the challenges in cross-linguistic corpus linguistic analysis is linguistic, yet translation-related one (Taylor, 2014). According to Vessey (2013, p. 13-4) and Taylor (2014, p. 373), building our specialized corpora using search terms (see. sub-

section 2.2.4) necessitates identifying comparable items in the different languages under investigation. For example, the difference in the writing systems between English and Arabic in which the latter is written in cursive script. That is, Arabic cannot be written in individual or separate letters as in English. Thus, the issue of using the Arabic equivalents of the English search term *Arab Spring*, along with other items, in both the definite and indefinite forms. The Arabic equivalent الربيع العربي <ar-rabī‘ al-‘arabī>, for example, literally corresponds to the English equivalent *the-Arab Spring*. Reporting findings as well as the presentation of the translated data in concordance lines are also two of the translation-related challenges (Taylor, 2014).

In the same context, comparing and contrasting the search terms and/or other words of interests and their equivalents in translation is another challenge, as words and their equivalents in translation often have different semantic prosody (Vessey, 2013, p. 13; Taylor, 2014). Semantic prosody, according to Louw (1993, p. 157), is the ‘consistent aura of meaning with which a form is imbued by its collocates’ (see section 2.3). It is an aspect of evaluative meaning which reflect the different attitudes and ideologies of the text (or discourse) producers within a given community or culture (Morley and Partington, 2009; Vessey, 2013). Additionally, collocational relations represent the different associations and connotations of the different items under investigation, and hence are much dependent on word forms (Stubbs, 1996). This in turn might be problematic in cases where analysis involves comparisons between languages, such as English, that make few distinctions in

gender (that is, feminine and masculine) with languages that much dependent on gender distinction, such as French (Vessey, 2013, p. 14) and Arabic, as it is the case in this study. Thus, ‘the lack of direct one-to-one equivalents between languages’ makes cross-linguistic statistical comparisons a challenging issue, resulting in other methodological difficulties relating to keyness analysis and the use of reference corpora (Vessey, 2013; Freaque et al., 2011; Taylor, 2014; Gabrielatos and Marchi, 2012, 2011) (see sub-sections 2.2.6 and 2.32 for more details).

Another translation-related, and hence cultural issue in cross-linguistic corpus-based analysis is the ‘compilation of corpora’ (Vessey, 2013, p. 6; Freaque et al., 2011; Taylor, 2014). For example, Vessey (2013, p. 6-10) highlights some crucial points in using multilingual and monolingual corpora, which is considered as the foundation of the CADS approach. While monolingual corpora are comprised of texts of a single language, and primarily designed for intra-lingual studies, multilingual corpora, on the other hand, involve texts of more than one language (Xiao and Yue, 2009). However, in Vessey’s (2013, p. 7) terms, ‘[m]onolingual corpora may contain texts produced by multicultural and multilingual populations in the medium of the dominant group, which are then collected to form a corpus’. For example, in her study, which is concerned with investigating a bilingual corpus (consisting of English and French text submitted to the 2007 Bouchard Taylor Commission in Quebec), Vessey (2013, p. 7), showed how ‘a relatively homogeneous population was indexed by the French language, whereas a quite

diverse population of minorities (including, for example, the Jewish community and the Italian community) was indexed by the English language'. This means that 'monolingual corpora drawn from real contexts may contain texts from a heterogeneous populace' (Vessey, 2013, p. 7), and overlooking this point may lead to misleading results, as whether certain pattern or a linguistic phenomenon 'is typical of a population or localised to a few language users or specific cases' (Baker, 2010b, p. 39; Vessey, 2013; Freake et al., 2011; Taylor, 2014). As far as this study is concerned, this issue is also a challenge as each of the English as well as the Arabic sub-corpora comprising the Arab Spring corpus consists of texts drawn from different contexts representing not only different languages or language variety (e.g. American, British and global variety English), but also different cultures. For example, and as discussed in sub-section 2.2.1, even though the global variety (GV) English sub-corpus represents one variety of English, it consists of texts that are basically drawn from two culturally different contexts (that is, the *Jerusalem Post* of Israel; the *Hurriyet Daily News* and the *Today's Zaman* newspapers of Turkey). Similarly, the Arabic sub-corpus is comprised of sub-corpora collected from different media outlets representing various Arab and Islamic (e.g. Saudi Arabia and Iran) contexts with different alignments. Thus, taking such ethnic and cultural differences into account may help in explaining, as will be shown throughout the following analytical chapters, why different results in terms of topics or focus, for example, are found in each sub-corpus (Vessey, 2013).

2.2.5 Search Terms: Arab Spring and the 'Protest Vocabulary'

Since the suicide of the Tunisian street vendor, Mohamed Bouazizi, on 17 December 2010 and the emergence of *the Arab Spring* global news media started

to use the term *Arab Spring* to describe the events that swept throughout the Middle East. Most of the time the term *Arab Spring* was associated with different terms such as: *protest* <’iḥtijāj – احتجاج>, *uprising* <انتفاضة – ’intifāḍa>, *revolution* <ثورة – ṭawra>, *revolt* <تمرد – tamarrud; or عصيان – ’iṣyān>, *rebellion* <تمرد – tamarrud or ثورة – ṭawra>, *Islamic awakening* <الصحوّة الإسلاميّة – aṣ-ṣaḥwa al-’islāmiyya> or *Arab Awakening* <الصحوّة العربيّة – aṣ-ṣaḥwa al-’arabiyya>, *conspiracy* <مؤامرة – mu’āmara>, *riot* <شغب – šaḡab>, *violence* <عنف – ’unf>. Such terms imply, whether explicitly or implicitly, expressions of disagreement with, rejection of or objection to a previous regime by means of words or action. For example, the thesaurus entry for the word *protest* (as a noun in both forms: countable and uncountable) in the *Macmillan English Dictionary for Advanced Learners* (2007, CD-ROM) shows that 60 words are closely or almost closely related to the meaning of *protest* of which the above terms (henceforth, *protest vocabulary*) such as *uprising*, *revolt*, *rebellion*, *revolution* and *riot* are closely related.

For the purpose of this study, a query has been developed as shown in Figure 2.3 to collect data from the aforementioned English and Arabic news outlets between 15 June 2010 and 31 August 2013. Before fully explaining the procedure for conducting the *LexisNexis* query, it is worth mentioning that these are not the only terms or concepts by which the Arab Spring is defined or with which it is associated in the news. The follower of the news would find other terms like

turmoil < اضطراب - 'idṭirāb>;
tension < توتر - tawattur>;
unrest < اضطراب - 'idṭirāb>;
crisis < أزمة - 'azma>;
upheaval < ثورة - ṭawra; or انقلاب - 'inqilāb>;
conflict < صراع - širā'>;
demonstration < مظاهرة - muḏāhara >;
opposition < معارضة - mu'āraḏa>;
 and even *civil war* < حرب أهلية - ḥarb 'ahliyya>.

Hence the decision to select the search terms presented in Figures 2.3 and not others was subjectively made based on the research questions and the study's focus. As shown in Figure 2.3, the English terms were used to collect the English data, while their Arabic equivalents were used to collect the Arabic. The search terms were not case-sensitive for English. This was not an issue for Arabic, as it has no upper or lower case for its letters. In addition, the 'Terms and Connectors' option (mainly AND/OR) in *LexisNexis* was used to limit the retrieval of news texts to those including the main term *Arab Spring* with at least one of the other keywords selected for the query. *Bouazizi*, < البوعزيزي - Al-bū'azīzī> was also selected as he was a main news actor. His suicide, as represented by many International and Arab news media, was considered as the catalyst for the Arab Spring.

<i>Arab Spring</i> <الربيع العربي - ar-rabī‘ al-‘arabī>	AND
<i>protest*</i> <احتجاج - al-‘iḥtijāj / الاحتجاجات - al-‘iḥtijāj-āt>	OR
<i>uprising*</i> <انتفاضة - al-‘intifāḍa / انتفاضات - al-‘intifāḍ-āt>	OR
<i>revolution*</i> <ثورة - at-tawra / الثورات - at-tawr-āt>	OR
<i>revolt*</i> OR <i>rebel*</i> <تمرد - tamarrud>	OR
<i>awakening</i> <صحوة - aṣ-ṣaḥwa>	OR
<i>conspirac*</i> <مؤامرة - al-mu‘āmara / المؤامرات - al-mu‘āmar-āt>	OR
<i>riot*</i> <الشغب - aṣ-ṣaḡab> OR <i>violence*</i> <عنف - al-‘unf>	OR
<i>Bouazizi</i> <البوعزيزي - al-bū‘azīzī>	

Figure 2.3 Search terms for collecting English and Arabic data

While many people in the Arab world have considered him as شهيد <šahīd – martyr>, some Muslim clerics have issued a *fatwa* (a directive or a religious rule) stating ‘that suicide violates Islam even when it is carried out as a social or political protest’ (Worth, 2011). Therefore, it is important to investigate how ‘Bouazizi’ has been represented before and after the actual emergence of the *Arab Spring*. Another aspect of the query is the use of the wildcard character (*), or the asterisk at the end which allows searching for variations of a word based on the specified number of characters. For example, *protest** might be retrieved in the English data as nouns, *protest* or *protests*, whether definite or indefinite, *protestors* and *protestation*, as well as verb forms: *protested*, *protesting*. Unlike English, the definiteness of the Arabic search terms was a major issue for retrieving the Arabic data, since Arabic

is written in cursive style rather than printed. Therefore, the Arabic search terms were prefixed with the definite article **الـ** <al - the> and used in two forms, definite and indefinite. For example, the Arabic equivalent for *protest** was used as indefinite **احتجاج** <'iḥtijāj* - protest*>; and definite **الاحتجاج** <al-'iḥtijāj* - the-protest*>; *awakening* as **صحو** / **صحو** <ṣaḥwa / صحو>; and **الصحو** – **الصحو** <aṣ-ṣaḥwa – الصحو>; *revolution** as **ثور** - * <ثور>; and **الثورة** – **الثورة** <al-thawra – الثورة>. Therefore, the decision to use definite and indefinite forms in the Arabic query was made because it was not possible to use the asterisk (*) at the beginning as this technique is not supported by *LexisNexis*. Therefore, the Arabic search terms were used in two forms. For example, *protest** or *the protest** would be retrieved in Arabic as the indefinite singular noun such as **احتجاج** <'iḥtijāj – a protest>, or plural such as **احتجاجات** <'iḥtijāj-āt – protests> or with the definite article **الـ** <al - the> such as **الاحتجاج** <al-'iḥtijāj – the-protest>, or **الاحتجاجات** <al-'iḥtijāj-āt – the-protests>. However, it is worth mentioning that, unlike English, the word **احتجاج** <'iḥtijāj - protest> is only a noun and cannot be a verb in Arabic. The basic verb of **احتجاج** <'iḥtijāj> in its present masculine singular form is **يَحْتَجُّ** <yaḥtajj-u – (to)protest.MASC.SING.PRS.V-NOM>. The focus of the query was mainly on the noun form for both languages though it was not possible to avoid the verb forms in some English terms such as *protest* (*protest(s)*, *protested*, *protesting*). Some Arabic equivalents such as **شغب** <šaḡab – riot>, **عنف** <'unf – violence> and **تمرد** <tamarrud - rebellion or revolt> are uncountable nouns, but can be used with or without the definite article **الـ** <al- the>. Another major issue regarding the Arabic query was the lexical gaps in the Arabic data. For instance,

using the main search term *Arab Spring* with the AND/OR technique resulted in retrieving few texts from two Arabic sub-corpora: ALMa and MHRa – news media outlets from Iran. These two sources used the term *Islamic Awakening* rather than *Arab Spring* when referring to the events that swept throughout the Middle East. Therefore, the decision to conduct a different query was made in order to fill such lexical gaps in the Arabic data by using the following search terms with the connector OR. The query that has been used to fill the gaps in the Arabic Data contains the two following terms:

الربيع العربي <ar-rabī‘ al-‘arabī – *the Arab Spring*> OR

الصَّحْوَة الإسلاميّة <aṣ-ṣaḥwa al’islāmiyya – *the-Islamic awakening*>

As far as retrieving English and Arabic data for months prior to the emergence of the Arab Spring are concerned, the term الشرق الأوسط <aṣ-ṣarq al-‘awsaṭ - *the Middle East*> was used as the main query term instead of the term *Arab Spring*. This was conducted to limit the retrieval of data to news related to the Middle East region. The purpose of collecting data between 15 June 2010 and late January 2011 was to investigate the nature of the news within the contexts of the protest movements, violent events or riots in the Middle East, and to compare their meanings and associations before and after the *actual* emergence date of the term *Arab Spring*. However, it is worth mentioning that the term *Arab Spring*, based on a *LexisNexis* search, was not widely used by news media until late January 2011 for English and early March 2011 for Arabic. Therefore, the decision to use the same search terms shown in Figure 2.3 was also made by only using the connector OR. In this instance,

irrelevant texts which were not directly related to the events in the Middle East were discarded. The following subsection highlights the English and Arabic reference corpora which are used in this study and discusses their significance in exploring hypotheses about language.

2.2.6 Using Reference Corpora

Using reference corpora is useful as it allows the researcher to explore hypotheses about language as well as to support findings and provide evidence for such hypotheses (Mautner, 2009a; Baker, 2006; Hunston, 2002). For example, the British National Corpus (BNC- with 100 million words) is a general corpus that represents modern British English as a whole, written and spoken, can be compared to a smaller corpus, or even one text file, in order to investigate which words occur in the smaller corpus, or text file, more frequently than we would normally expect them to occur by chance alone (Baker, 2006, p. 43; McEnery et al., 2006; Hunston, 2002; Bondi and Scott, 2010). Reference corpora might also be synchronic, as in the case of the Arab Spring corpus though it is not a reference corpus, reflecting a language variety at a certain time in history or diachronic, reflecting the historical development of a language variety at two different times in history (Baker, 2006; Mautner, 2009a). In that sense, reference corpora are useful when conducting CDA, although there is no straightforward answer for which reference corpus to use when carrying out discourse analysis (Baker, 2006, p. 43). However, we should try to obtain the reference corpus that best reflects ‘some aspect of the smaller corpus or

text sample we are studying' (Baker, 2006, p. 43). Since the Arab Spring corpus consists of two main sub-corpora: English and Arabic, and the English corpus consists of three main sub-corpora representing British, American, and global variety of English forms, three main reference corpora have been used. These are: The British National Corpus (BNC of 100 million words) for British English, the Corpus of Contemporary American English (COCA of 450 million words) and a Wikipedia Arabic Reference List (WA of 150 million words) that was specially created for this study by David Woolls (2013, CFL Software Limited, UK, <http://www.cflsoftware.com>). The frequency distribution list for word forms in the Wikipedia Arabic corpus (Sharoff, 2006a) was downloaded in plain text format from the Querying Arabic Corpora interface, developed also by Sharoff (2006a) from the Centre for Translation Studies at the University of Leeds, and publicly available online [<http://corpus.leeds.ac.uk/query-ar.html>]. Although there are other Arabic corpora such as, the Corpus of Contemporary Arabic (1 million words) (Al-Sulaiti and Atwell, 2006), the International Corpus of Arabic (ICA, around 100 MW) (Alansary et al., 2007; Alansary and Nagi, 2014), and the Leipzig University Arabic collection (around 42 MW) (Eckart et al., 2014), they are smaller than the Arabic Wikipedia corpus (WA) in terms of size. Perhaps the Arabic Web 2012, or the arTenTen12 corpus of 5.8-billion words (Arts et al., 2014), is one of the biggest Arabic corpora, which is freely available for all to investigate through Sketch Engine (Kilgarriff et al., 2004) interface. According to Arts et al. (2014, p. 368-9), there are some challenges facing the arTenTen12 corpus, this very large web-

crawled corpus of contemporary Arabic includes texts from many domains and genres. This means, with the Arabic language family, that many texts are likely to appear in different language varieties, such modern standard Arabic (MSA), classical Arabic, Quranic Arabic and various dialects, and that identifying the language variety of each text is a challenge in itself (Arts et al., 2014, p. 369). Another issue facing the arTenTen12 corpus preparation is the need to build ‘an interface option that allows users to use the undiacritized form while keeping the diacritized form as an option for advanced users’ (Arts et al., 2014, p. 369), who ‘would prefer that word sketches be computed on diacritized forms’ (Arts et al., 2014, p. 369). For example, it would not be possible to see a word sketch for the item العالم <al-‘ālam - *the-world*> without noise resulting from العالم <al-‘ālim – *the-scientist*> as both items are written as العالم when not diacritized (Arts et al., 2014, p. 369). Therefore, and to avoid such technical issues, it is perhaps better decided using the Wikipedia Arabic (WA) reference corpus (Woolls, 2013) for analysis purposes set for this study.

Regarding the GV-ENASC sub-corpus, a decision was made to use the COCA rather than the BNC reference corpus is the fact that the COCA reference corpus is more up-to-date and larger than the BNC reference corpus. According to Scott (2009, p. 81), both size and date of the reference corpus are two important criteria in corpus linguistics and linguistic comparisons along with the size and genre of the study corpus itself. For example, in his study that is concerned with the debates on fox hunting, Baker (2009, p. 132) points out that ‘[c]omparing a smaller corpus or

set of texts to a larger reference corpus is [...] a useful way of determining key concepts across the small corpus as a whole'. In addition, although there is a reference corpus for global varieties of English, such as the Global Web-Based English (or GloWbE) reference corpus (Davies, 2013), it does not actually represent the variety of English (which can be said Middle Eastern English of mainly Israel and Turkey) represented by the GV-ENASC corpus presented in this study. The GloWbE corpus, which is freely available online through the BYU interface, is composed of 1.9 billion words in 1.8 million web pages from 340,000 websites in 20 different English-speaking countries, covering the time period 2012-13 (Davies, 2013). Other than the United States, Canada, Great Britain, Ireland, Australia and New Zealand, the remaining countries in which the web pages and websites comprising the GloWbE corpus include India, Sri Lanka, Pakistan, Bangladesh, Singapore, Malaysia, Philippines and Hong Kong (representing East Asian countries), South Africa, Nigeria, Ghana, Kenya, Tanzania (representing countries in Africa) and Jamaica (Caribbean). From this perspective, 'it might be hard to obtain a large, perfectly-matched reference for some comparisons' (Scott, 2009, p. 81). In other words, there are no simple answers to 'which reference corpus is applicable to use' (Baker, 2006, p. 43). In his study *In Search of a Bad Reference Corpus*, Scott (2009) examines the influence on the keywords by using reference corpora (RCs) of different sizes, genres and time. In terms of size, Scott (2009, p. 82-6) used RC texts chosen from 4, 054 BNC texts (spoken and written) using a randomizing function, so that 22 different sizes (the smallest RC is based on ten

texts, whereas the biggest RC is based on 4,000 texts) of RC were selected. For comparison with the various RCs, Scott (2009, p. 82) used two source texts (that is, a short 615-word doctor-patient interview, and a lengthy 46,000-word text of business leaders). His finding suggested that using a mixed bag RC, the larger the RC the better, but not in the case of the small doctor-patient interview as a moderate size RC may suffice (Scott, 2009, p. 91; 2010, p. 51). For the research question regarding the genre and time, Scott (2009, p. 86-7) compares the leaders of commerce text from the 1990s with a reference corpus comprised of all Shakespeare's plays (the genre is drama) representing a time period of the late sixteenth and seventeenth century. He concludes that the keyword procedure is robust, and that 'keywords identified even by an obviously absurd RC can be plausible indicators of the aboutness, which reinforces the conclusion that keyword analysis is robust' (Scott, 2009, p. 91). In both cases, results show that there is no really bad reference corpus (Scott, 2009). As far as the GV-ENASC sub-corpus is concerned, initial analysis showed that no significant difference between the KKW lists that resulted from comparisons with the two reference corpora (BNC or COCA). That is, in terms of the resulting number of KKWs or their type, except in some cases where, for example, a number of grammatical items such as *not*, *don't*, *it's*, and *cannot* were KKWs when compared to COCA, whereas they were not when compared with the BNC (for a detailed discussion of keyness, see section 2.3 and sub-section 2.3.2).

According to Partington (2009, p. 12) and Partington (2010, p. 90), in a corpus-assisted discourse analysis (or CADS) approach, ‘it is only possible to both uncover and evaluate the particular features of a discourse type by comparing it with others’. CADS can be summarized as the ‘investigation and comparison of features of discourse types, integrating into the analysis, where appropriate, techniques and tools developed within corpus linguistics’ (Partington, 2010, p. 88). The aim of CADS approach is to investigate, in discourse type under investigation, non-obvious meanings – that is, meanings that might not be readily available to perusal by the naked eye (Partington, 2010, p. 88; Marchi, 2010, p. 163). It is ‘a subset of corpus linguistics: “that set of studies into the form and/or function of language as *communicative discourse* which incorporate the use of computerised corpora in their analysis”’ (Partington et al., 2013, p. 10, original italics). Thus, within CADS framework, which combines between corpus linguistics and discourse studies (Taylor, 2014, p. 372), it is possible to conduct comparisons between a specialized (or *monogeneric*) corpus and large *heterogeneric* corpora such as the BNC, to explore the behaviour of different linguistic items under study (Partington, 2009, 2004). From the CADS perspective, all discourse analysis is properly comparative, which entails either working with pre-existing corpora or compiling a specialized corpus, as in the case of the Arab Spring corpus, to synchronically or diachronically study particular linguistic features (Partington, 2010, p. 90; 2009, p. 12; Morley and Bayley, 2009). Analyses and comparisons in this study are conducted by using *WordSmith Tools* (Scott, 2015), which allows for the comparison of frequencies in

word lists and the analysis of keywords by means of collocation and concordance (Baker, 2006; Mautner, 2009a). However, researchers may employ other techniques besides the quantitative or statistical ones used in corpus linguistics. This may include, as in the CADS approach for example, the researcher's intuition or other information outside the corpus to investigate particular discourse type features under study (Partington, 2010, p. 90; 2009, p. 10). Using corpora to conduct discourse analysis, therefore, has many advantages, and the procedures followed for constructing them are important in determining their usefulness (Baker et al., 2008, p. 54; Meyer, 2002, p. 53).

The following subsection discusses some legal and ethical issues pertinent to the construction and exploitation of corpora.

2.2.7 Ethical Considerations

There are some legal and ethical issues that compilers of corpora should consider before gathering and distributing data. However, it should be noted that laws and ethics regarding data and information use in research vary between countries, as well as between different academic and research bodies within the same country, which might have their own internal guidelines and procedures for their researchers to follow (McEnery and Hardie, 2012), and hence this study. As far as legal issues are concerned, the first pertinent question is whether data intended for inclusion in the corpus are freely available or copyrighted material (McEnery and Hardie, 2012, p. 57). If texts or data of any kind, whether online or in print, are protected by

copyright laws then corpus compilers must seek the permission of the publishers and authors who own the copyright on the work, especially when creating large corpora for commercial purposes (McEnery and Hardie, 2012; Baker, 2006; Meyer, 2002). In that case, gaining permissions can be quite a lengthy process, as commercial corpora usually contain a large amount of texts (Baker, 2006). Alternative approaches include collecting data only from free websites which permit copying and redistribution of texts, collecting data without seeking permission for distribution of data, but instead making it available to other researchers through tools that prevent copyright violations, and finally redistributing only the web addresses or links from which data were collected (McEnery and Hardie, 2012, p. 59-60). However, such approaches might not be deemed ideal solutions. According to McEnery and Hardie (2012, p. 59), restricting corpus data to public domains or free websites such as Wikipedia, for instance, would 'skew its representativeness'. Similarly, other researchers might not be able to conduct further contextual analyses by means of some web-based concordancers which do not allow full text processing, and finally data might not be found due to the fact that web pages and links may change over time and therefore not be available (McEnery and Hardie, 2012). Another instance pertaining to copyright issues is that, when creating a small corpus for single use, gaining permission might then be easier, especially if there is one, or a few sources, from which researchers aim to collect their data (Baker, 2006). However, if a corpus is intended for non-profit academic research, corpus compilers should state that in any letter requesting

permission to use copyrighted material (Baker, 2006; Meyer, 2002) as publishers and authors might sometimes ask for money to use their work (Meyer, 2002, p. 62). As far as the Arab Spring corpus is concerned, archives for most of the newspapers and websites from which data were collected were accessible and available through the online news database *LexisNexis*. However, access to *LexisNexis* is only available for authorized users, who are entitled to use its online archives and services for the purposes only of academic research or study, providing professional services to clients and academic services to students (LexisNexis, 2010). The remaining data (such as data collected from the *BBC*, *Al-Jazeera* and *Al-Quds Al-Arabi*) were officially and publicly archived online and no passwords were required for the archives' access (Bruckman, 2002). Furthermore, this research is conducted in accordance with the University of Leeds' policy on Research Data Management and Good Research Practice. The following section discusses the theoretical and methodological frameworks of corpus linguistics and critical discourse analysis within which this study has been conducted, as well as the advantages and limitations of the corpus-based approaches to CDA.

2.3 Theoretical and Methodological Frameworks

This study is mainly informed by the theoretical and methodological notions and concepts of corpus linguistics (CL) and critical discourse analysis (CDA/ CADS). In its corpus-based approach, the study draws upon notions of *frequency*, *keyness* (*Key Keyword function*) as well as *collocation* (and its related notions of *semantic*

preference and *semantic prosody*) to examine the representation of the Arab Spring narrative in English and Arabic news media. Frequency is simply the total occurrences of all items or words in a given corpus along with their types (Baker, 2006; McEnery and Hardie, 2012; Hunston, 2010). It provides information on the most and least frequent uses of all words in a corpus, enables comparing different texts or a corpus of texts to reveal differences/ similarities in lexical choices and finally provides insights to the most frequent topics and discourses (Scott, 2015; Baker, 2006). In other words, frequency helps in identifying ‘patterns of use that otherwise often go unnoticed by researchers’ (Biber et al., 2004, p. 376). Likewise, frequency clusters (also known as lexical bundles or n-grams) help in identifying potential sites of interest in the corpus and how frequent words are used in their contexts (Baker, 2006; Baker et al., 2008; Baker, 2004). For example, in her analysis of the speeches of Tony Blair and George W. Bush between 2005 and June 2007, Milizia (2010, p. 134-37) first examined the word *climate*, and the 2-word-cluster (or bigram) *climate change*, and found that *climate change* is a concern in Blair’s speeches, which is very clear from the context in which the phrase is found. She also stated that *climate change* is found only in context of negativity, and her study showed that phraseological combination, in the form of word clusters, is important in revealing the ‘aboutness’ of a given text or collection of texts (Milizia, 2010). Word or frequency clusters therefore tell us a great deal about how a particular discourse type is constructed (Partington, 2009). ‘Clusters are words which are found repeatedly together in each other’s company, in sequence’ (Scott,

2015, p. 437). They are technically known as n-grams, where n represents the number of words in the cluster (Scott, 2015; Partington et al., 2013; Forchini and Murphy, 2008; Milizia, 2010; Gerbig, 2010). For example, *the Arab World* is a 3-gram (or 3-word cluster) and was, as we might expect, one of the most frequent 3-word clusters in the English (ENASC) corpus as a whole. Thus, using *WordSmith Tools* (Scott, 2015), one can specify the number of words in a cluster from two to, realistically, ten words (Partington, 2009).

Additionally, I investigate the different patterns of the *Arab Spring* and its associated topics, such as their collocates, semantic preference and semantic prosody, which ‘can be seen as the semantic extension of collocation’ (Baker et al., 2008, p. 278). According to Stubbs (1996, p. 89) ‘meanings are conveyed not only by individual words and grammatical forms, but also by the frequency of collocations and the distribution of forms across texts’. Collocations are lexical items that frequently co-occur with each other, and hence such collocational relations show the associations and connotations these items have, as well as the assumptions they embody from the contexts in which they occur (Oakey, 2010, p. 14; Baker, 2006; Stubbs, 1996). Semantic preference refers to the relation of an item that co-occurs with ‘a class of words which share some semantic feature’ (Stubbs, 2001, p. 88). Semantic prosody (or discourse prosody), on the other hand, is the ‘consistent aura of meaning with which a form is imbued by its collocates’ (Louw, 1993, p. 157). It is an aspect of evaluative meaning which reflects the attitude or stance of the text producer whether negative or positive, good or bad

(Partington, 2004; Morley and Partington, 2009; Baker et al., 2008; Mautner, 2009a; Stubbs, 2001; Hunston, 2004, 2007). In addition, although semantic preference and semantic prosody show us, as (Mautner, 2009a, p. 128) points out, the kinds of contexts in which lexical items are used as well as their associated attitudes and opinions, both notions have different operating scopes. According to Partington (2004, p. 151), semantic preference relates ‘the node item to another item from a particular semantic set ... [whereas semantic prosody] can affect wider stretches of texts’. For example, Stubbs (2001, p. 65) explains how the word *cause* ‘occurs overwhelmingly often with words for unpleasant events’ such as *death, disease, damage or trouble*. Semantic preference, in that sense, becomes tied more with the collocation phenomenon and ‘contributes powerfully to building’ prosody whereas semantic prosody ‘dictates the general environment which constrains the preferential choices of the node item’ (Partington, 2004, p. 151). Hence, one central objective of this study is to explore the discursive construction of the *Arab Spring* narrative within the contexts of its associated topics/themes, or the KKW items denoting them.

Keyness, on the other hand, points towards the ‘aboutness’ of the text (Baker et al., 2008; Scott, 2015; Baker, 2006, 2004). It ‘is defined as the statistically significantly higher frequency of particular words or clusters in the corpus under analysis in comparison with another corpus, either a general reference corpus, or a comparable specialized corpus’ (Baker et al., 2008). However, it should be highlighted that there are two kinds of keyness: positive and negative. According to Scott (2015, p.

241, original italics), ‘a word which is *positively* key occurs *more* often than would be expected by chance in comparison with the reference corpus [whereas] a word which is *negatively* key occurs *less* often’. In the same context, there are three types of key words that are expected to show in a key word list. The three types of keywords include proper nouns (e.g. names of people, places or organizations), lexical words such as nouns, verbs and adjectives, which show the ‘aboutness’ of the text or the corpus of texts and, finally, grammatical words or high-frequency words like *because* or *shall* or *already*, which may be key indicators more of style than of ‘aboutness’ (Baker, 2006; Baker et al., 2008; Scott, 2015, p. 236; Baker, 2004). A key word list is calculated by comparing the frequency of each word in the word list of the study corpus with the frequency of the same word in the word list of the reference corpus (Scott, 2015, 2009, 1997; Scott and Tribble, 2006; Bondi, 2010; Baker, 2004; Stubbs, 2010; Gabrielatos, 2007a; McEnery and Hardie, 2012). In addition, the calculation of keyness takes into consideration the frequency data as well as the size of each study corpus (Baker, 2004, 2006). Statistical tests which are usually conducted to produce key word lists include the chi-square test of significance and the log likelihood test (Scott, 2015; Baker, 2006, 2004; Dunning, 1993). These give each word a probability value (or p-value), that is a number between 0 and 1 to indicate ‘the amount of confidence that ... a word is key due to chance alone’ (Baker, 2006, p. 125). For example, setting a p-value of 0.01 means that there is 1% danger of being wrong in claiming a relationship whereas 0.05 suggests 5% danger of error, a risk that is usually considered as acceptable by

the social sciences (Scott, 2015). Researchers in corpus linguistics may often want to set a comparatively lower p-value such as 0.000001 to obtain fewer key words (Scott, 2015). For example, in his paper on the representations of Islam in British broadsheet and tabloid newspapers 1999-2005, Baker (2010a, p. 317) used the log-likelihood statistical test to carry out comparative keyword analysis; each word was assigned a 'keyness' score at which all words were key at $p < 0.00000000001$. He conducted detailed analysis on the strongest 300 lexical keywords in both types, the broadsheets and the tabloids (Baker, 2010a). The log-likelihood test, according to (Scott, 2015, p. 244), 'gives a better estimate of keyness, especially when contrasting long texts or a whole genre against your reference corpus'. As far as this study is concerned, keyness is automatically calculated by means of *WordSmith Tools 6.0* (Scott, 2015) software using the log-likelihood test with p value set at 0.000001.

However, the problem with conducting key word analysis is that words might be key, not because they are pervasive throughout the whole corpus, but rather because they have a higher frequency in some or a few of the texts. This issue could be sorted out by using the key key-word technique (Baker, 2004; Scott, 2015, 1997; Gabrielatos, 2007b), which allows the researcher to investigate if the words are high in frequency in many or a few texts and to decide whether to include or exclude such texts from the analysis. A key keyword (KKW) is a word that 'is "key" in more than one of a number of related texts. The more texts it is "key" in, the more "key key" it is' (Scott, 2015, p. 225). Moreover, using the KKW list function 'allows

for the identification of *associates*, [...], which form an alternative means of calculating collocation in the wider sense (Taylor, 2013, p. 90, original italics). Associates, which are KWs that co-occur with other KKWs in a number of texts (Scott, 2015, 1997), like collocates are ‘useful in determining different senses of individual key keywords’ (Baker, 2006, p. 148). In that sense, associates can also be useful in providing us with clues to what kind of issues and events the term is associated, that is its context, and how this could affect its semantic features. In the English KKW list, for example, two of the strongest associates of the word *protest* in the English news texts, covering the period between 15 June and 17 December 2010, are *Israel* and *Palestinians*. This implies that there were quite a number of news reports on issues that were concerned with the Palestinian-Israeli conflict and the peace process before the events of the Arab Spring broke out in the Middle East on 17 December 2010. However, it is worth mentioning that associates of a KKW are different from its collocates. A collocate of a KKW ‘would have to occur within a given distance of it, whereas an associate is “associated” by being key in the same text’ (Scott, 2015, p. 216). The KKW function is also useful for conducting topical analysis. It allows for statistically identifying the main topics across larger volumes of texts or text types, which might not be feasible via manual means (Al-Hejin, 2015). However, different text types, such as news articles as well as editorial texts, entail different analytical approaches, since both types have different schematic structures as well as functions. The following sub-section discusses these

differences along with illustrative examples of topical analysis of both text types (news and editorials).

2.3.1 Topical Analysis of the News and Editorials Texts

From a CDA perspective, the study also aims to identify the main topics (events and news actors) associated with the Arab Spring in order to examine the way they are discursively presented in culturally and ideologically different contexts. CDA is primarily concerned with ‘the way social power abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context’ (van Dijk, 2008, p. 352; 2001, p. 85). It is fundamentally interested in the dialectical relationship between language and society which considers discourse as ‘socially *constitutive* as well as socially shaped’ (Fairclough and Wodak, 1997, p. 358, original emphasis; Fairclough, 1995b, p. 55), and hence one of its paradigmatic principles is its focus ‘on *institutional environments* as key sites of research into the connections between language, power, and social processes’ (Blommaert, 2005, p. 34, original emphasis). Accordingly, mass media become a particularly interesting subject of CDA analysis for ‘their pivotal role as discourse-bearing institutions’ (Garrett and Bell, 1998, p. 6; Bell, 1995, p. 23) as well as their undeniable power in constructing and reinforcing social reality (Mautner, 2009a; van Dijk, 2008, 2001). The mainstream media therefore become the site where dominant groups seek to disseminate their ideologies, values and attitudes (Fairclough, 2001). Therefore, exploring news media discourses related to the Arab Spring narrative in culturally

and ideologically different contexts also necessitates an interdisciplinary approach, to understand the processes of news production and comprehension within these contexts.

In addition, analysing media discourse structure of news stories (more specifically, news reports/articles, and editorials and opinions) shows how they may differ in their structure from other kinds of narratives that people tell, for instance, in everyday conversations or in novels. As far as news reports are concerned, they basically consist of an attribution that is concerned with the source, time and place of the news story, the abstract that generally includes the headline and the lead, which covers the main event of the news story, and finally the story proper, which contains one or more episodes that in turn covers one or more events (Bell, 1998, 1991, 1995; van Dijk, 1991, 1988b). In that sense, the concepts of *macrostructures* (and more specifically *macropropositions* and *macro-rules* – which are basically concerned with *topics*) (Kintsch and van Dijk, 1978; van Dijk, 1998, 1991, 1988b, a) and *news values* (Bell, 1991; Bednarek and Caple, 2014; Potts et al., 2015; Bednarek, 2015; Harcup and O'Neill, 2001; O'Neill and Harcup, 2009; Harcup and O'Neill, 2016; van Dijk, 1988b; Fowler, 1991; Brighton and Foy, 2007; Richardson, 2007; Schaudt and Carpenter, 2009) are important aspects of the analysis of news stories in terms of news selection, production and comprehension as well as the analysis of news structures. For example, news values (Bell, 1991) such as *Proximity*, *Eliteness*, *Negativity*, *Relevance*, *Consonance*, *Superlativeness*, and the newspaper's own agenda, are important in understanding how/why some events or

topics in the Arab Spring stories are more newsworthy than others from the perspectives of different news media outlets representing different cultures and ideologies. Therefore, understanding how topics (the headline and the lead), which represent the higher levels of the ‘pyramid’ (Figure 2.4) in the news stories (van Dijk, 1991, p. 72; 1988a, p. 92-3), are formed is also relevant.

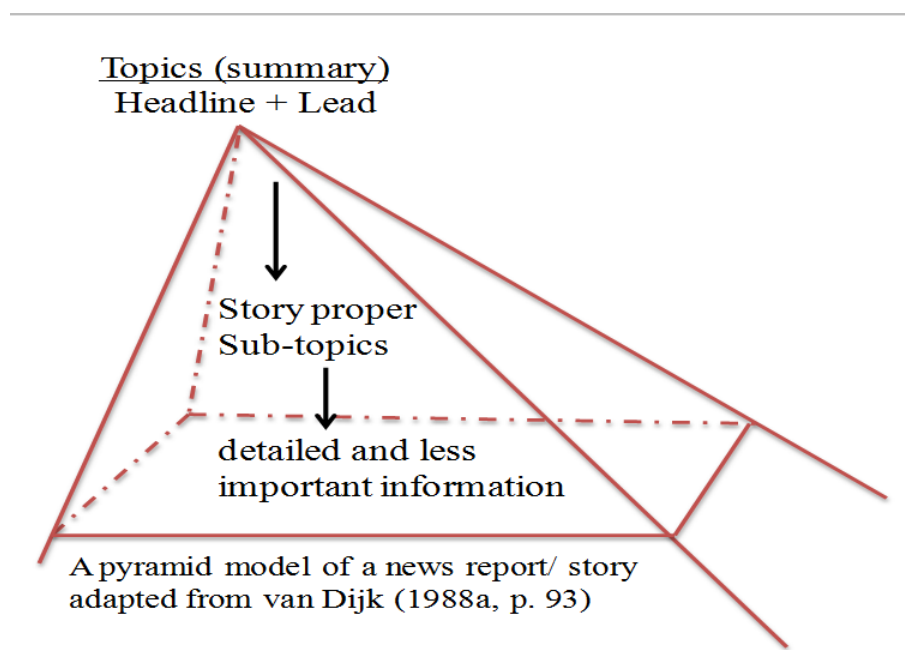


Figure 2.4 Pyramidal model of a news report schematic structure

According to van Dijk (1991, pp. 71-4), topics have important discursive functions: they provide the analyst with deeper insight into the issues under investigation, they reflect many psychological and sociological aspects of the news, they manifest the powers and ideologies behind news production, and finally they are crucial cognitive information sources, which ‘influence the representation readers

construct in their mind of specific' events or situations that they most likely to recall and use later.

Thus, topics (or macropropositions) are derived from the meanings (propositions) of the sentences of the news story by means of three major macro-rules. These are: deletion, generalization and construction (van Dijk, 1991, 1988b). Deletion is concerned with deleting any irrelevant information not referred to by a subsequent proposition. Generalization, on the other hand, allows the substitution of a sequence of propositions for one general proposition that denotes their meaning. The final macro-rule of topic structure is construction, which is similar to generalization and is concerned with verbs and actions rather than things or people; it deals with reducing a sequence of actions to lie under one term or topic. This could be best illustrated by the following extract of a news report, CNNe-10-48, from the CNNe sub-corpus, represented in Table 2.9 illustrating how topics are structured by macro-rules:

Table 2.9 Illustration of the three major macro-rules of news topic structure

HEADLINE	S1	4 Israelis shot dead in West Bank
LEAD	S2	<u>Four Israelis -- including a pregnant woman</u> -- were killed Tuesday near Hebron in the West Bank in a shooting for which the militant wing of Hamas claimed responsibility, officials said.
STORY PROPER	S3	The incident occurred near Bani Naim junction, the largely Palestinian territory where Jews have settled in places like Hebron , Israel Defense Forces spokeswoman Lt. Col. Avital Leibowitz said.
	S4	The <u>victims</u> were in a car on Route 60, the IDF website said.
	S5	Guy Gonen, a paramedic who was one of the first people on the scene, said the car was sprayed with bullets.
	S6	More than a dozen bullet holes were found in the left side of the white Subaru station wagon in which <u>they</u> were driving.
	S7	The attack comes ahead of direct talks scheduled to begin Thursday in Washington between Israeli and Palestinian leaders - the first such talks since 2008.
	S8	<u>The four killed -- two men and two women -- were from the settlement of Beit Hagai.</u>
	S9	IDF forces were searching for the attackers, the IDF website said.
<hr/>		
Bold refers to the information about the place mentioned in S1		
<u>Underlining</u> refers to the information about the news actors in S1		
Italics refers to the agents of the shooting (CNNe-10-48/ news: 31-Aug)		
<hr/>		

Considering the headline (S1), this reflects instances of deletion, generalization and construction. An instance of deletion can then be identified in S2 (Sentence 2), S3 and S4. This includes information, in bold print, about the place where the incident happened as well as the agents of the shooting. This information has been deleted and reduced to *West Bank* in S1 because there is no subsequent reference to such places or agents of shootings (militants of Hamas) in the rest of the story. The

underlined information in S2 and S8 shows the number and gender of those killed which has been replaced by *4 Israelis* in S1 and by *victims* in S4. Finally, an instance of construction that is concerned with reducing sequences of actions to one term or topic is represented in the word *shot* in S1. The word *shot* is a cover term for a sequence of actions reported between S2-S9: there were 4 Israelis in their car on Route 60 near Hebron, which was targeted and attacked with bullets, killing the four Israelis in that car. In this context, it should be noted that the formation of topics, according to van Dijk (1991, p. 73), is subjective, which means that whatever is considered as relevant or important for one journalist or reader might not necessarily be the same for the other, and hence headlines and leads of news reports and stories are not objective summaries but ‘necessarily biased by specific beliefs, attitudes and ideologies’. In that sense, minor topics or certain details can sometimes be manipulated and ‘upgraded’ to a headline or lead level. Compare, for instance, *4 Israelis* in S1, *The victims* in S4, *The four killed -- two men and two women --* in S8, with *Four Israelis -- including a pregnant woman --* in S2 (the lead); the phrase *-- including a pregnant woman --* in S2 is believed to generate a sense of sympathy with the victims while at the same time building up a more of negative view of Hamas as killers (or *attackers* as in S9) of innocent people. This is just a brief illustration of how topics and certain details are ‘foregrounded’ or ‘backgrounded’ in the news, and how producers in media discourse exercise their powers in determining what to include or exclude, and how events or topics are also represented (Fairclough, 1995b, 2001).

In the theoretical sense, doing topical analysis is very important in CDA, but technically this might not be feasible when dealing with larger volumes of data or texts. Additionally, this structural analysis of news reports cannot be applied to editorials and opinion articles since they ‘are diverse in their styles or textual strategies, and that is part of the point, to suggest a distinctive “voice” for the newspaper’ (Fowler, 1991, p. 209). In other words, editorials and opinion articles are different in both their schematic structure and function from the news items or reports to which they refer (van Dijk, 1992, 1988b, 1991, 1988a, 1993, 1996b, 1998). While news reports are meant to give new information about recent events and no explicit opinions of the individual reporter, these events are already known in editorials and opinion articles, and they involve explicit and dominant opinions, which represent the official stance or point of view of their newspaper or writer/editor on respective events or issues (van Dijk, 1988b; a, p. 124; 1993; Fowler, 1991; van Dijk, 1998). From this perspective, opinion articles, especially editorials, are ‘generally institutional, and not personal’ (van Dijk, 1996b, online) which means that they represent the opinions, values and beliefs of the institutions which produce them, and they are ‘governed by the same institutional constraints which build the ideology’ (van Dijk, 1996b; Fowler, 1991, p. 47). Editorials and opinion articles are meant not only to influence the public opinion, but also to address the influential news actors and the power elites, as well as to influence their attitudes towards specific events or actions (van Dijk, 1992, p. 244; 1988a, 1993, 1991). In this context, opinions, as “evaluative beliefs”, presuppose values, and

involve judgements about somebody or something as being, for example, good or bad (van Dijk, 1998, p. 29). For that reason, editorials and opinion articles basically have argumentative and persuasive functions which entail a different schematic structure from that of news reports (van Dijk, 1988a).

Although there might not be a conventional schematic structure for editorials or opinion articles, they may have, according to van Dijk (1996b, 1993, 1992, 1991), three functional and schematic categories or superstructures: definition and explanation of events, evaluation of events – especially of actions and actors, and finally a conclusion or moral which might include recommendation, advice or warning. Similarly, Bolívar (2001, 1994), in her scholarly work on the structural analysis of written text, including newspaper editorials and opinion articles, proposes a model based on what she calls the *triad* (Td). It is considered as the basic or minimal unit of interaction whose function ‘is to negotiate the transmission of information and evaluation in written text’ (Bolívar, 1994, p. 279). In other words, the triad can be defined ‘as a coherent segment of text with a topic and a function in the discourse’ (Bolívar, 2001, p. 137). It consists of three fundamental *turns* (Tn) of which each serves a distinct function in the written text: the *Lead* (L) initiates the topic and introduces the ‘aboutness’ of the triad as well as a posture or modality, the *Follow* (F) continues with the topic, and explains different aspects of the situation, and finally the *Valuate* (V) which closes the segment with an evaluation of the previous two turns (Bolívar, 1994, 2001). Each turn in LFV is realized by one or more *sentences* (S) conceived, according to Lyons (1994, p. 279;

in Bolívar, 2001, p. 136), as ‘the product of ordinary language behaviour’, or text sentences, rather than ‘system sentences’. In other words, these are sentences which are context-dependent rather than decontextualized ones.

Based on her findings on the structural analysis of 23 editorials selected from *The Guardian* covering the first three months of 1981, Bolívar (1994, p. 280) points out that not all triads consist of three turns, but rather can have more than three turns only if the sequence LF is repeated before the closing V turn. This means that ‘triads such as LFLFV or LFLFLFV can be found when the V turn is delayed by the writer’ (Bolívar, 1994, p. 280). The triad, LFV, can also be classified into three types: the *Situation* triad (S) initiates a reference to and an evaluation of the event, the *Development* triad (D) expands on the preceding S triad, and finally the *Recommendation* triad (R) closes the reference and evaluation introduced in the S triad (Bolívar, 1994, p. 280-1). These triads may then combine to make up a larger unit in the structure called *Movement* (Mv), which is in turn comprised of three types: A for the actual world (*that is or was*), B for the world of possibilities (*that might be*), and C for the world (*that should be*); These movements may, as well, combine to produce the largest unit in the model which is called the *artefact* (Bolívar, 1994p, 280-1).

Although three-part structures cannot be considered as universal in a way that ‘they occur in all types of discourse’, there is evidence that they are also found in other languages such as Spanish and Japanese (Bolívar, 1994, 2001). For example, in her

paper on evaluation in text, Bolívar (2001) analyses newspaper editorials as well as conference abstracts in English and Spanish, and provides evidence that the LFV structure is present in both languages as well as in the different text types and genres used in the analysis. Her findings also show that the V turn is ‘obligatory’ in texts of evaluative and argumentative nature such as editorials and opinion articles whereas it is ‘optional’ in other types such as news reports or abstracts (Bolívar, 2001, p. 136). It is worth mentioning, however, that although evaluations or opinions are obligatory elements in texts of evaluative, persuasive or argumentative nature, they may not be expressed explicitly. According to van Dijk (1996b), the level of explicitness depends on a number of factors such as the nature of opinions themselves, the situation (or context models) as well as the role and position of the writer or the producer of the text. For example, in his analysis of an editorial, published in 1993 by the *Washington Post*, about violent events which occurred in Nicaragua at that time, van Dijk (1996b, online) attributes the implicitness of opinions expressed in that editorial to a number of reasons. These are the *Washington Post*’s political position, as a conservative newspaper, as well as ‘its relation with the Congress, and hence its relation to a U.S. ally and (female) president’ (van Dijk, 1996b, online). In this perspective, evaluations or opinions in other texts, such as news reports, may be signalled or expressed in the same way by means of several processes such as the selection of events, topics or news actors as well as the way they are presented linguistically or visually (Bolívar, 2001; van Dijk, 1991; Fairclough, 1995b). As far as the Arab Spring corpus is concerned,

initial analysis shows that both English and Arabic editorials and opinion articles follow a three-part structural pattern, and hence reflects Bolivar's (1994, p. 276) triad as 'the minimal unit of interaction in written text'. Examples of the schematic structure of a complete triad (Td) in both English and Arabic extracts from editorials and opinion articles in the ARASC corpus is represented in Tables 2.10 and 2.11 respectively, followed by a brief discussion of their content. As Tables 2.10 and 2.11 show, each extract represents the first triad in the editorial from which it was taken. It is called the *situation* triad (S Td) in which the LFV structure can be identified. The S triad has the initiating function of presenting the actual event and evaluating it by means of its internal structure composed of the three LFV turns. For example, the L turn in Table 2.10, introduces the topic of the S triad, which is the first in an editorial, DTLe-12-49, published in the *Daily Telegraph* newspaper in 2012 about the death of Megrahi and his release from prison in 2009. It shows how the topic (or event) is not only initiated but also stated in evaluative terms, as seen in the use of *brings to an end* and *a very embarrassing chapter*. The L turn in Table 2.11, on the other hand, has the function of initiating the topic of the *Situation* triad (S triad). This triad is the first in an editorial AQa-13-04 from *Al-Quds Al-Arabi* newspaper published in 2013, and concerns the threats surrounding the partition of Syria. Unlike the L turn in the English extract in which the writer's own

Table 2.10 Illustration of Situation Triad, (S)Td, Structure in the English editorials and opinions

[HEADLINE] The story of Megrahi's release must now be told					
Td	Tn	S	Structure	Text (S-Triad)	Function
S	L	1	Initiates topic	The death from cancer of the Lockerbie bomber, Abdelbaset Ali Mohmed al-Megrahi, brings to an end a very embarrassing chapter for this country - but we still await the full story of his release from prison in August, 2009.	Presents situation Initial evaluation
	F	2	Continues with topic	Megrahi was the only person convicted of the murder of 270 people when Pan Am flight 103 crashed near Lockerbie in December 1988.	Expands on situation
		3	Continues with topic	Serving a life sentence in Greenock, he was released by the Scottish Justice Minister, Kenny MacAskill, on the grounds that he was likely to die of prostate cancer within three months.	Explains other aspects of the situation
	V	4	Closes topic	To the discomfiture of the British authorities - but to no one's surprise - the man found guilty of the worst peacetime atrocity in this country's history received a hero's welcome in Tripoli, where he survived for almost three more years.	Closes topic Reinforces evaluation in L (S1)
(DTLe-12-32/ Editorial: 21-May)					

Table 2.11 Illustration of Situation Triad, (S)Td, structure in the Arabic editorials and opinions

HEADLINE:					
السيد نصر الله وتقسيم سورية					
<i>As-Syyed Nasrallah and the Partition of Syria</i>					
Td	Tn	S	Structure	Text (S-triad)	Function
	L	1	Initiates topic	فاجأ السيد حسن نصر الله الكثيرين من أنصاره ومتابعيه يوم أمس عندما أكد أن سورية 'مهددة بالتقسيم أكثر من أي وقت مضى' <i>Sayyed Hassan Nasrallah surprised many of his supporters and followers yesterday when he affirmed that Syria is 'threatened by partition more than any time in the past',</i>	Presents situation/ initial evaluation/ warning
	F	2	Continues with topic	ولكنه لم يقل كيف يمكن منع هذا التقسيم، ومن هو المسؤول عن وصول البلاد الى هذا المصير المظلم. <i>But he did not say how this partition can be prevented, and who is responsible for the reaching of the countries to this dark fate</i>	Expands on situation
S	V	3	Closes topic	التقسيم الذي يواجهه سورية، ولا نبالغ إذا قلنا إن خطواته الأولى قد بدأت فعلا، لن يتوقف قطارُه عند الحدود السورية، وقد يمتدُّ إلى دول الجوار مثل لبنان والعراق والأردن ليصل إلى السعودية وربما تركيا وإيران أيضاً. <i>The partition, which faces Syria, and we do not exaggerate when we say that its first steps have already begun, will not clutch its train onto the Syrian border, and it might extend to the neighbouring countries such as Lebanon, Iraq, Jordan, Saudi Arabia and possibly Turkey and Iran also.</i>	Closes topic/ reinforces evaluation / warning in L (S1)
(QAa-13-04/ editorial: 03-Jan)					

opinion or evaluation is expressed explicitly, the L turn in the Arabic S triad is introducing the topic and evaluating it, indirectly or implicitly, by means of a quotation from the prominent news actor, *Sayyed Hasan Nasrallah* (Chief of the Lebanese Shi'ite Hezbollah). For example, the quotation 'مهدة بالتقسيم أكثر من اي وقت مضى' <'muhddadat-un bi-t-taqṣīm-i akṭar-a min ayy-i waqt-in maḍā' - *is threatened by partition more than any time in the past*> enhances the credibility and seriousness of the topic that the writer is introducing, and reflects an initial warning against the partition that Syria is encountering. According to van Dijk (1991, p. 152), quotations serve several functions that make news, arguments and opinions more persuasive. The F turn in both triads, then continue with the topic, expand on the situation and explain different aspects of the events initiated in L turns. In the English F turn, however, there are two sentences whereas in Arabic there is only one. As described by Bolívar (1994, p. 280), 'triads can exhibit more than three turns provided that the sequence L F is repeated and V is final'. In this situation, both the English and Arabic V turns closes the S triads with evaluations of the previous two turns, L and F. The English V turn closes the topic and reinforces the evaluation initiated in the L turn. The evaluation is expressed in stronger terms seen in the use of *discomfiture*, *worst* and *atrocities*. The Arabic V turn (S3), on the other hand, closes the topic and concludes with reinforcing the evaluation as well as warning presented in L (S1) that the partition threatening Syria has already begun, 'ولا نبالغ إذا قلنا أن خطواته الأولى قد بدأت فعلا' <wa-lā nubālīg 'iḍā qul-nā anna ḵuṭuwāti-hi al-'ūlā qad bada'at fi 'lan - *and we do not exaggerate when we say that its steps have*

already begun>. The Arabic V turn (S3) also concludes with a warning that التقسيم <At-taqsīm – *the partition*> will not stop on the Syrian borders, but rather it will extend to the neighbouring countries. This warning is expressed by لن يتوقف قطاره عند الحدود السورية <lan yatawaqqaf qitār-u-hu ‘ind-a al-ḥudūd-i as-sūriyya – *and its train will not stop on the Syrian borders*>, and وقد يمتد إلى دول الجوار <wa-qad yamtadd-u ‘ilā duwal-i al-jiwār – *and it might extend to the neighbouring countries*>.

This brief discussion shows that the three-part structure is present in both English and Arabic editorials and opinion articles, which will positively aid the comparative aspects of this study. However, and as mentioned earlier, this kind of detailed analysis of topics might not be feasible on a larger scale, and hence the KW/KKW list function, like other CL techniques such as collocation and concordance, becomes an effective tool for identifying the main topics across both text types. The keyness approach, however, was criticized and debated not only in terms of its definition and senses (e.g. Stubbs, 2010; Baker, 2004) but also in terms of its related metrics, that is, the different statistical measures used to calculate the level of keyness and the compatibility between such metrics and definitions of keyness found in different sources (e.g. Gabrielatos and Marchi, 2012, 2011; Gabrielatos and McEnery, 2005; Gabrielatos, 2007a; Kilgarrieff, 2001, 2012; Pojanapunya and Watson Todd, 2016; Scott, 2010; Culpeper, 2009). Applying keyness approach in cross-linguistic corpus-based studies, as mentioned earlier (see also 2.2.4), is another challenging issue (e.g. Vessey, 2013; Freake et al., 2011; Taylor, 2014, p.

375), as keywords cannot directly be compared when different languages are involved. The following sub-section discusses in more detail some of these issues and criticisms relating to the different keyness approaches to corpus-based studies in terms of keywords and keyness definitions, concepts and senses as well as the different metrics, or statistical measures used in investigating keyness, and finally discusses the challenges that keyness approach brings when conducting cross-linguistic corpus-based analyses.

2.3.2 Criticism of Keyness: Definitions and Statistical Measures

As discussed earlier (see section 2.3, pp 69-72), the keyness method identifies items of unusual frequency in comparison with a reference corpus of some suitable kind or norm (Scott, 2015, p. 229; Scott and Tribble, 2006, p. 55; Baker, 2004). Keywords (KW/KKW) hence provide a useful way to characterise a text or a genre as they are used to refer to words that are important in some way either in a text or a given culture (Scott, 2015; Stubbs, 2010; Bondi, 2010; Baker, 2009, 2004). In that sense, keyness becomes ‘a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail’ (Scott and Tribble, 2006, p. 55-6). However, the issue with investigating these important words, according to Stubbs (2010, p. 21-32) is that ‘there are several different concepts of “keywords”’, and he pointed out that they are used in three different senses (that is, cultural, statistical and phraseological) derived from quite different academic traditions. Discussing the

three senses in his study regarding the concepts of the term ‘keywords’, Stubbs (2010, p. 39-40) concludes that the problem lies in the large gap between the individual words and the social world, and that although the concept of ‘keywords’ may be productive, it cannot stand on its own. In the same context, Stubbs (2010, p. 39-40) pointed out that the term ‘keywords’ ‘assumes other concepts, such as cognitive schemata, textual collocates and semantic fields, text and text-type’, of which the latter, ‘in turn imply the concept of social institutions’.

Another related criticism levelled at the keyness approach to corpus-based analyses is that the metrics, or the statistical measures used for calculating *keyness* need to be fully consistent with the definition of *keyword* (Gabrielatos and Marchi, 2012, 2011). In their study regarding metrics of keyness and other practical issues, Gabrielatos and Marchi (2012, 2011), argued that there is a contradiction between the definitions of the two widely-used interrelated constructs in corpus linguistics, keywords and keyness, as presented in the literature and corpus software manuals. While the term *keywords* is statistically defined in relation to frequency difference (as in Scott, 2015), the definition of *keyness* on the other hand treats a word as key ‘if its frequency in the text when compared with its frequency in a reference corpus is such that the *statistical probability* as computed by an *appropriate procedure* is smaller than or equal to a p-value specified by the user (Scott, 2015). In terms of Scott’s (2015) definition, *keyness* of an item is computed mainly by considering the frequency of that item as well as the number of running words in both the study as well as the reference corpora lists and then cross-tabulated. Statistical tests used for

calculating *keyness* include the chi-square test of significance and the Dunning's (1993) Log Likelihood test (Scott, 2015, p. 245). Similarly, Biber et al. (2007, p. 138, original italics) point out that the *keyness* of an item 'represents the value of *log-likelihood* or *Chi-square* statistics; in other words it provides an indicator of a keyword's importance as a content descriptor for the appeal. The significance (p value) represents the probability that this *keyness* is accidental'. With these definitions in mind, Gabrielatos and Marchi (2012, 2011) also pointed out that if keywords are defined in relation to frequency difference, the metric for *keyness* should represent the extent of the frequency difference rather than statistical significance of that difference. Statistical significance is the 'p-value of the frequency difference, as measured by a statistical test – usually log-likelihood or Chi-square' (Gabrielatos and Marchi, 2012, 2011; Scott, 2015). Gabrielatos and Marchi (2012, 2011) also argued that tests of statistical significance are largely dependent on the sample size, and therefore different conclusions may be drawn in different studies because of the size of the samples, if such conclusions were only based on statistical significance testing. In that case, such tests do not reflect the actual significance of the relationship, whether it is weak or strong, between the variables in the study and reference corpora (Gabrielatos and Marchi, 2012, 2011).

Therefore, Gabrielatos and Marchi (2011, 2012) proposed the %DIFF (the percentage or proportion of frequency difference) measure as an alternative metric for *keyness*, as it is fully consistent with the definition of *keyword*. The %DIFF measure (also used by Gabrielatos, 2007, and Gabrielatos and McEnery 2005) is a

simple and straightforward effect-size metric for *keyness* analysis, which indicates the proportion (%) of the difference between the normalised frequencies (that is, frequencies per one thousand or one million words) of a word in two corpora, or sub-corpora (Gabrielatos and Marchi, 2012, 2011). The %DIFF metric, according to Gabrielatos and Marchi (2012, 2011), is calculated as follows:

$$\%DIFF = \frac{(NF \text{ in SC} - NF \text{ in RC}) \times 100}{NF \text{ in RC}}$$

in which NF represents normalized frequency, SC for the study corpus and RC for the reference corpus. Thus, the %DIFF is more appropriate metric for *keyness* than the log-likelihood as it is fully consistent with the definition of *keywords*, especially that the vast majority of keyword analyses in corpus studies do not examine all the keywords, but rather specify a certain number of keywords (usually the top 100 keywords), and the ranking criterion (which is usually *keyness*) hence becomes very important (Gabrielatos and Marchi, 2012, 2011). One of the advantages of the effect size metric is that it enables researchers to tell if the frequency difference or the relationship between variables in the study and reference corpora is weak or strong regardless of the sample or corpus size, and hence provide information that allows comparisons of such relationships or differences across studies (Gabrielatos and Marchi, 2012, 2011). Another advantage of the %DIFF measure is that it ‘may be very easily employed in revealing not only for differences, but also similarities ‘by identifying those items which had a very low percentage frequency difference,

because this would, in effect, flag up items which had very similar patterns of occurrence' (Taylor, 2013, p. 93).

However, there are some practical issues regarding the %DIFF metric. For example, Gabrielatos and Marchi (2012, 2011) noted that however %DIFF is large, it needs to be statistically significant; that is, frequency differences that are below the statistical significance threshold (accepted threshold is $p \leq 0.01$) adopted in the study should not be included in the analysis. Another issue is the handling of zero occurrences in the reference corpus as users, based on the %DIFF metric, cannot divide by zero (Gabrielatos and Marchi, 2012, 2011). Gabrielatos and Marchi (2012, 2011) suggested substituting all zero frequencies in the reference corpus with 1E-19 (that is, 1 divided by one quadrillion), as an approximation of zero, to allow for divisions by it. However, such division, according to Gabrielatos (2012, online), 'results in extremely high %DIFF values'. In other words, 'very large %DIFF values flag up potentially interesting differences. However, the LL score will indicate the extent to which we can trust this large %DIFF score' (Gabrielatos, 2012, online). The final issue is that current tools do not accommodate the %DIFF metric; Gabrielatos and Marchi (2012, 2011) provided 'a step-by-step guide to carrying out this procedure using WordSmith Tools and basic Microsoft Excel functions'

Other measures of effect-size for identifying keywords include *frequency ratio*, proposed by Kilgarriff (2001; 2012, p. 4), and the *odds ratio* (OR) measure which

was considered in a study conducted by Pojanapunya and Watson Todd (2016). Like the %DIFF metric, both measures are also dependent on normalized frequency. Frequency ratio is simply based on counting occurrences in each corpus, dividing each number by the number of words in that corpus, optionally multiplying by 1,000 or 1,000,000 to give frequencies per thousand or million, and finally dividing the first number by the second (or vice a versa) to give a ratio (Kilgarriff, 2012, p. 4). Again, to solve the problem of zero occurrences in the reference corpus, Kilgarriff (2012) suggested *adding one* to all zero frequencies, including those for words that were present in study corpus but absent in the reference corpus, and then there would be no zeros and that ratio can be computed for all words. However, according to Gabrielatos (2012, online), adding 1 to normalised frequencies in both corpora so that zero frequencies are treated as if they are 1 does not treat both frequencies equally, and hence this technique skews the results. In terms of the odds ratio (OR) metric, it measures relative proportions of word frequencies in the study as well as the reference corpora, and suggests how much the difference is between the word frequencies in both corpora (Pojanapunya and Watson Todd, 2016). In their study that is concerned with keyness statistics, Pojanapunya and Watson Todd (2016) compared the LL and the OR metrics through two case studies that are concerned with keyword analyses of advance fee scams against the BNC corpus and research articles in applied linguistics against a corpus of research articles from other academic disciplines. They argued that the two methods produce different keywords applicable to research focusing on different purposes (Pojanapunya and

Watson Todd, 2016). Their analyses showed that while both the LL and OR keywords concern the aboutness of the corpus, they differ in their specificity and pervasiveness across the corpus (Pojanapunya and Watson Todd, 2016). Results of their study showed that the LL method highlights words which are relatively common in general use serving genre purposes, whereas the odds ratio (OR) flags up more specialized words serving critically-oriented purposes (Pojanapunya and Watson Todd, 2016; Graham, 2014). (For more information on Log-Likelihood and effect size and other metrics of keyness see UCREL's Log-Likelihood site available at <http://ucrel.lancs.ac.uk/llwizard.html>).

As can be seen, the effect size measures discussed above, along with the LL measure, are mainly dependent on frequency. The effect size metric 'measures the size of the difference of the normalised frequency of a word in two corpora, not the statistical significance of that difference' (Gabrielatos, 2012). Thus, the main issue is that the rankings of words based on LL and other effect size measures (e.g. %DIFF and OR) are likely to be in a different sequence (Graham, 2014, online). While LL flags up words that are relatively common in general use, effect size measures are believed to highlight more specialised words which are specific to the study corpus (Gabrielatos and Marchi, 2012, 2011; Gabrielatos, 2012; Pojanapunya and Watson Todd, 2016; Graham, 2014). However, effect size metrics might not be the answer to all types of research questions in corpus linguistics, especially 'that different research purposes and contexts have different requirements' (Culpeper, 2009, p. 36). For example, although effect size measures might contribute to solving

issues regarding comparative words ranking in the corpora under investigation, such metrics, like other statistical measures, might not be the fully right answer to studies involving, for instance, ‘lemmatisation of word forms’, or when ‘contractions’ (that is, short forms of words such as *it’s* and *I’ll*) are highly (in)frequent (Culpeper, 2002, p. 27-8). Additionally, although effect size metrics highlight more specialised items characterising the study corpus (Gabrielatos, 2012), it is perhaps ‘[t]he choice of the reference corpus will affect whether you acquire keyword results that are all relevant to the particular aspect of the text(s) you are researching’ (Culpeper, 2009, 2002; Scott and Tribble, 2006). According to (Culpeper, 2009, p. 35), ‘[t]he closer the relationship between the target corpus and the reference corpus, the more likely the resultant keywords will reflect something specific to the target corpus’. For example, in his study regarding keyness in Shakespeare’s play *Romeo and Juliet*, Culpeper (2002, 2009) analysed the speech of the six main characters in the play, and the comparative reference corpus was the speech of the six characters minus the one being investigated’. The procedure which Culpeper (2002, 2009) used in his analysis produced a different set of keywords that reflect the distinctive styles of each character compared with the other characters in the same play. Culpeper’s (2009, 2002) procedure, for example, contrasts with Scott and Tribble’s (2006, p. 63) three analyses of the same play, which used three different reference corpora: tragedies by Shakespeare; the complete works of Shakespeare including poetry; and the BNC including the spoken component, which in turn showed sort of different set of keyword results.

The final point relating to the criticism of keyness approach is the difficulty in comparing keyness scores across corpora consisting of different languages, especially that each keyword list draws on a different reference corpus (Vessey, 2013, p. 18; Freake et al., 2011, p. 30). According Vessey (2013, p. 14) ‘the lack of direct one-to-one equivalents between languages makes comparisons of frequency and statistical significance’ is challenging in cross-linguistic corpus analysis. In other words, ‘when a word has a different keyness from its translation in another language, it is difficult to assess whether this is due to significant differences between the focus corpora or differences between the reference corpora’ (Freake et al., 2011, p. 30). Hence, ‘the lack of parallel reference corpora in different languages’ is another challenging issue in comparing keywords across different languages (Vessey, 2013, p. 21). To address this problem, Vessey (2013, p. 21) suggested parallel keyword analysis. Parallel keyword analysis, according to Vessey (2013, p. 16), ‘refers to the selection or compilation of similar reference corpora in two languages, each suited to the composition of its respective primary corpus, in order to produce comparable keywords’.

The following sub-section discusses some of the methodological concerns regarding the production of the final version of the KKW lists as well as the decisions taken to resolve them. It briefly discusses two main issues: the creation of the keywords database and the calculations and extraction of KWs and KKW lists of the eight main Arab Spring sub-corpora. This is followed by sub-section 2.3.4, discussing the categorization process, highlighting some of the manual as

well as the automated techniques, such as the *Wmatrix* (Rayson, 2008, 2003), and introducing the pivot tables in Excel involved in sorting the data as well as the categorization process.

2.3.3 Methodological Concerns: Making the Keywords Database

According to Scott (2015, p. 237), the point of key keyword technique is that it allows researchers to study the keywords that recur often over a number of files. A keyword is defined by Scott (1997, p. 236, original italics) ‘as a *word which occurs with unusual frequency in a given text*’, whereas a key-keyword (KKW) ‘is one which is “key” in more than one of a number of related texts. The more texts it is “key” in, the more “key key”’ (Scott, 2015, p. 240) (see section 2.3, p. 74-7 for more details on keyness). In order to build a keywords database, by means of *WordSmith Tools*, version 6. (Scott, 2015, p. 238-9), the user will need a set of keyword lists. According to Scott (1997, p. 235), the KKW listing technique involves ‘various procedures, which derive from and build on one another’. Thus, the word listing procedure enables keyword processing, which in turn allows for creating a keywords database by which key-keywords (KKWs) will be revealed (Scott, 1997, p. 235). In terms of the Arab Spring sub-corpus, the eight wordlists of the four main sub-corpora (two for each) were conducted by means of *WordSmith Tools*, version 6. (Scott, 2015), and were then compared with their reference corpora. As mentioned earlier, the British news as well as editorials and opinion wordlists were compared with the BNC reference corpus; the American and the global

variety English wordlists were compared with COCA, and the Arabic wordlists were compared with the Wikipedia Arabic (Woolls, 2013).

The KeyWords tool default settings, for all keyword lists, in *WordSmith Tools* version 6. (Scott, 2015) were as follows: minimum frequency = 3; procedure for computing keywords = Log Likelihood; and the p-value = 0.000001. According to Scott (2015, p. 255), '[t]he minimum frequency is a setting which will help to eliminate any words or clusters which are unusual but infrequent. [...]. The default setting of 3 mentions as a minimum helps reduce spurious hits here. In the case of short texts, less than 600 words long, a minimum of 2 will automatically be used'. The KKW lists were then conducted, and the default settings for the database were as follows: minimum texts per entry = 5 (that is, only keywords that appear in 5 or more texts of each sub-corpus will be used for the database); minimum keywords per text = 10 (that is, any text ended up with very few KWs will be ignored); minimum texts for associates is 3 and the statistic for associates is MI3 with minimum strength of 3. However, the default cut-off point regarding the *raw* minimum number of texts (that is, to include keywords that appear in 5 or more of the texts of each sub-corpus) was not considered, but rather the relative value of 1% or more of each sub-corpus was decided as a cut-off point for keywords to be included in the analysis. This cut-off point was applied to each one of the eight study sub-corpora. For example, in the British news sub-corpus, which consists of a total of 2,237 texts, the item *Arab* is key in 335 texts (as Table 2.12 shows), which relatively equal 27.06% of the British news sub-corpus. Table 2.12 shows examples

of the top and lowest ten KKW items along with their statistics from the KKW list of the British news sub-corpus:

Table 2.12 Examples of the top and lowest ten KKWs along with their statistics in the British news KKW list

The top and lowest ten KKWs in the British news sub-corpus			
Word	Key in Number of Texts	% of sub-corpus	Overall Frequency
Arab	335	27.06	1,827
Al	294	23.75	2,216
Syria	239	19.31	1,444
Israel	238	19.22	1,676
Mr	223	18.01	2,323
Israeli	203	16.4	1,329
Syrian	185	14.94	1,094
Egypt	177	14.3	924
Palestinian	167	13.49	1,123
protests	162	13.09	713
Syrians	13	1.05	58
refugees	13	1.05	112
strikes	13	1.05	61
blog	13	1.05	47
coup	13	1.05	76
threat	13	1.05	70
terror	13	1.05	62
Bashar	13	1.05	56
Riyadh	13	1.05	52

As Table 2.12 also shows, the minimum number of texts in the British news sub-corpus is 13, which relatively equals 1.05% of the British news sub-corpus. For example, each of the items *Syrians*, *refugees*, *coup*, *threat*, *terror*, *Bashar* and

Riyadh are KKWs in 13 texts (as Table 2.12 shows), which means that each one of them is KKW in 1.05% of the British news sub-corpus. In terms of the American news sub-corpus, as Table 2.13 shows, the minimum number of texts for the KKWs is also 13, which relatively equals 1.02% of the whole sub-corpus (which consists of a total of 1,803 texts). Similarly, in the Arabic news sub-corpus, as Table 2.13 also shows, only words that are key in 44 or more texts are included in the analysis. Table 2.13 shows the minimum number of texts for the KKWs along with the number of the resulting KKWs in each of the main four English and Arabic text type sub-corpora:

Table 2.13 Minimum number of texts for KKWs and their % of each sub-corpus

Sub-corpus	News			Editorials and opinions		
	Min. No. of Texts	% of sub-corpus	No. of KKWs	Min. No. of Texts	% of sub-corpus	No. of KKWs
English						
American	13	1.02	379	5	1.00	350
British	13	1.05	304	5	1.10	282
Global variety (COCA)	5	1.35	272	6	1.06	289
Arabic						
Arabic (MSA)	44	1.01	191	23	1.02	203

The aim of setting cut-off points, as Baker (2004, p. 352) and Culpeper (2009, p. 36) pointed out, is to have a sufficient number of items that meets the researcher's aims and objectives set his/her study (in my case, to identify the main topics

associated with the Arab Spring narrative in the English and Arabic news media), not to have an overwhelming number to analyse, and to restrict the one-off or extremely rare words being included in the final version of the KKW lists under investigation. The decision on this cut-off point (that is, 1% or more of the texts) was also aimed to ensure a level of consistency as well as representativeness (Taylor, 2013, p. 90; Baker, 2004, p. 351-2), though this issue raises concerns regarding the fact that ‘there is no popular consensus about cutoff points’ in terms of keyword analyses (Baker, 2004, p. 351). The last reason behind deciding on this cut-off point (that is, a keyword that appears in 1% or more of each sub-corpus is to be included) is to relatively ensure an equal level of saliency across the study sub-corpora because some of the Arab Spring sub-corpora are different in terms of size.

However, this decision was not completely useful as quite a number of the resulting items, especially in the English sub-corpus, were only KKWs in one or two out of its main six comprising sub-corpora, and hence could not be considered as representative of the English sub-corpus as a whole. Another decision therefore was that an item had to be a KKW that is shared at least between three out of the six main English KKW lists in order to be considered as ‘permanent’ key keyword – that is, a KKW item across the English sub-corpus as a whole (McEnery, 2009, p. 99). Thus, in McEnery’s (2009, p. 98) terms, ‘[k]ey keywords are keywords which are key in all, or the majority, of the subsections of a corpus’ The KKW function is also useful in distinguishing ‘relatively transient’ from ‘permanent’ keywords

which in turn provide a list of concepts that are representative of the study sub-corpora as a whole (Bachman, 2011). This threshold collectively produced a total of 295 out of 663 KKW's of which 108 KKW's were shared between the six English sub-corpora. In terms of the Arabic sub-corpus, all of the resulting 277 KKW items, and of which 117 KKW's are shared between its two main sub-corpora, were included in the analysis. Table 2.14 shows a general summary of the breakdown of the resulting KKW items (that is, shared/unshared) across the English as well as the Arabic sub-corpora:

Table 2.14 Distribution of shared/unshared KKW's across the English and Arabic sub-corpora

Number of sub-corpora	No. of shared/unshared KKW's	
	In ENASC	In ARASC
6	108	
5	47	
4	72	
3	68	
2	133	117
1	235	160
Total included	295	277

The following sub-section continues with discussing the categorization of the resulting KKW items of both the English as well as the Arabic sub-corpora by means of the *Wmatrix* (Rayson, 2008), and explains sorting the results by means of a pivot table in Excel.

2.3.4 Key Keyword Categorization

Analysis of the main eight KKW lists (six for English and two for Arabic) also involves grouping the resulting KKWs of both English and Arabic into semantically similar categories, which in turn requires further analysis of their concordance (Baker, 2010b) as well as comparisons of their usage in the corpus. The categorization of the six English (the American, the British and the global variety) KKW lists is based on an earlier analysis conducted by the web-based corpus software tool, *Wmatrix* (Rayson, 2008), which assigns to every item (words and multi-word units) a part-of speech label, then assigns these items to one or more semantic set or subset (Rayson, 2008, p. 519; Prentice et al., 2012, p. 265). The rationale behind using this software was only to obtain an overall picture of the representation of the Arab Spring narrative in the English data as well as to get an initial idea about the key topics and concepts across the English sub-corpus. In that sense, the approach adopted in this study ‘can be described as top-down, in that the categories are pre-defined’ (Archer et al., 2009, p. 139), as they have been identified by *Wmatrix* (Rayson, 2008, 2009), and the fact that its semantic labels rather than statistical results that have been considered in this study. As a software tool for

corpus analysis and comparison, *Wmatrix* provides a web interface to the English USAS (UCREL Semantic Analysis System) and CLAWS (the Constituent Likelihood Automatic Word-tagging System) corpus annotation tools, and standard methodologies such as frequency lists and concordances (Rayson, 2009, 2008). It also extends the keywords approach to key grammatical categories as well as key semantic domains (Rayson, 2008, 2009). Thus, by means of *Wmatrix*, users can upload their own corpus data (the English sub-corpus in my case) to the system, so that it can be automatically annotated and viewed within the web browser, and which can also be stored in a folder (Rayson, 2009). Like *WordSmith Tools* (Scott, 2015), *Wmatrix* uses the log-likelihood ($LL \geq 6.63$) and p-value (≤ 0.01) statistic to calculate keyness of items (Prentice et al., 2012, p. 282). As the log-likelihood test takes into consideration the frequency of each word in both the study as well as the reference corpora, it also considers the sizes of both the study and the reference corpus (see UCREL log-likelihood site, <http://ucrel.lancs.ac.uk/llwizard.html> for more information) (Scott, 2015). The log-likelihood test ‘gives a better estimate of keyness, especially when contrasting long texts or a whole genre against your reference corpus’ (Scott, 2015, p. 245). The *Wmatrix* method is also useful as it allows for using grammatical categories and/or semantic sets to group together lower frequency items which might, by themselves, be overlooked in the analysis (Rayson, 2008, p. 543; Prentice et al., 2012). By considering less frequent KWs, KKWs or collocates through grouping them into grammatically or semantically similar categories, different pictures may then be revealed (Baker et al., 2013a, p.

262). In terms of the Arabic sub-corpus, analysis of the resulting KKW lists and the semantic categorization of the KKW items were manually conducted as Arabic is not supported by *Wmatrix* software (Rayson, 2008). However, although it can be said that the categories are automatically pre-defined by means of the USAS system used by *Wmatrix* (Rayson, 2008), semantic sets were revised during this stage, and some minor changes were conducted on the *Wmatrix* semantic labels in the way that suits the research questions as well as the purposes set for this study. For example, most of the English and Arabic query terms, which are of protest/revolutionary nature (e.g. *protest/s*, *protester/s*, *uprising/s*, *revolution/s* and *Spring*) are grouped under the category PROTEST VOCABULARY, which USAS would assign them under G3 (that is, Warfare, defence and the army, Weapons), a sub-category of the general semantic domain (G) Government and the Public Domain. The categorization process also includes consulting dictionary definitions or other sources of information other than the study corpus (Partington et al., 2013, p. 10).

Based on the cut-off points and decisions discussed above, analysis identified 15 main semantic categories into which the resulting KKW items of both English and Arabic were grouped. These are: GEOGRAPHY (e.g. *Arab*, *Syria/n/s/s*, *Tunisia/n/s*, *US*, *America*, *country/countries*, *Western*, *bank* and *square*); GOVERNMENT AND SECURITY (e.g. *regime/s/s*, *democracy*, *elections*, *reform/s*, *activists*, *army*, *security*, *weapons*, and *militants*); NAME-PN (e.g. *Obama*, *Assad*, *Netanyahu*, *Mubarak*, *Clinton*, *NATO*, *UN*, *(Muslim)Brotherhood*, and *Bouazizi*); FUNCTION WORDS (e.g. *pro*, *anti*, *against*, *it's*, *he*, and *we*); SOCIAL ACTIONS, STATES AND

PROCESSES (e.g. *opposition, aid, power, king, sheikh, people, women, groups, movement/s and parties*) ; FAITH AND RELIGION (e.g. *Islam/ic, Muslim/s, Christians, Jews, sectarian, secular, Shiite, Sunni, and religious*) PROTEST VOCABULARY (e.g. *protest/er/s, revolution/s, uprising/s, spring, demonstration/s and rebel/s*); LANGUAGE AND COMMUNICATION (e.g. *said, talks, negotiations, media, internet, speech, and journalists*); GENERAL AND ABSTRACT TERMS (e.g. *conflict, freedom, and crisis*); EMOTIONS (e.g. *peace, violence, clashes, terror, threat and unrest*); TIME (e.g. *post, Thursday, year/'s, young, former and generation*); ECONOMY AND FINANCE (e.g. *economic, freeze, jobs, oil, and prices*); PSYCHOLOGICAL-MENTAL-PROCESSES (e.g. *stance, strategic* and in Arabic *الصحوة* <*the-awakening*>; *المشهد* <*the-scene*>; and *القضية* <*the-cause*>); LIFE/DEATH (e.g. *assassination, killing, killed, death and suicide*); and NUMBER AND MEASUREMENT (e.g. *billion, cent, and per, and أكثر* <*more*>). However, only the top ten categories, in terms of the overall total number of KKWs, were considered. Nevertheless, the KKW items in the remaining semantic categories will also be considered when necessary or required. Table 2.15 shows collective results of the top ten semantic categories in the two main English and Arabic sub-corpora ordered by the *actual* total number of KKW items in each category:

Table 2.15 Number of KKWs in the top ten semantic categories in the English and Arabic sub-corpora

KKW-Semantic Categories	Actual No. of KKWs		
	English	Arabic	Grand Total
1. GEOGRAPHY	100	72	172
2. GOVERNMENT AND SECURITY	61	43	104
3. NAME-PN	46	31	77
4. FUNCTION WORDS	18	33	51
5. SOCIAL ACTIONS, STATES AND PROCESSES	17	30	47
6. FAITH AND RELIGION	16	10	26
7. PROTEST VOCABULARY	12	12	24
8. LANGUAGE AND COMMUNICATION	10	15	25
9. GENERAL AND ABSTRACT TERMS	5	12	17
10. EMOTIONS	7	3	10
Column Totals	292	261	553

As Table 2.15 shows, for example, the first top category is GEOGRAPHY, which includes KKW items referring to geographical places and locations (e.g. *Arab*, *Syria* and *countries*). The second top semantic category is GOVERNMENT AND SECURITY, which contains KKW items referring to events, processes or entities of governmental, political, security and military nature (e.g. *regime*, *government*, *democracy*, *activists*, and *war*). Additionally, in the NAME-PN category, which is the third top semantic category in terms of the total number of words, consists of KKW

items that refer to names of high-profile politicians and leaders, organizations and other people. The PN stands for proper nouns, such as *Obama* and *Assad*, or proper names, including noun phrases such as *United Nations*, *the European Union* and *the Muslim Brotherhood*. The aforementioned categories and the eight remaining semantic categories, such as NAME-PN, FUNCTION WORDS, PROTEST VOCABULARY, FAITH AND RELIGION, are discussed in more detail in Chapter 4 that is concerned with the topics associated with the Arab Spring, but will also be considered in Chapter 3 which is concerned with the news actors. Nevertheless, where relevant, all of the remaining KKWs in the English as well as the Arabic sub-corpora will be considered, as they have also been included in the categorization process, but not included in the final version of the resulting top ten KKW semantic categories discussed in the analytical chapters. The whole procedure was conducted by means a pivot table in Excel (Partington et al., 2013, p. 312; Taylor, 2013, p. 99). A Pivot table (Table 2.16) can automatically sort, count totals or give averages of statistical data in a given table. In sorting the eight KKW lists, which have been manually combined in one Excel spreadsheet, the pivot table counts the number of the sub-corpora in which the item is a KKW. This means that the higher the value under the ‘Grand Row Label’, the higher the rank a KKW item will get. Accordingly, the semantic categories were ordered based on the grand total number of KKW items in each category, whereas the KKW items within each category were ordered by

Table 2.16 Example of plotting results in pivot table of Excel taken from the English sub-corpus

Example of plotting results in a pivot table in Excel showing KKW examples of the GOVERNMENT AND SECURITY category in the English sub-corpora																					
KKW-semantic category	Key in Number of Texts						Percentage (%) of sub-corpus						Number of KKWs						Grand Row Total in the English sub-corpus		
	News			Editorials and opinions			News			Editorials and opinions			News			Editorials and opinions					
GOVERNMENT AND SECURITY	AM	BR	GV	AM	BR	GV	AM	BR	GV	AM	BR	GV	AM	BR	GV	AM	BR	GV	No. of Texts	Avg. % of sub-corpus	No. of KKW
Regime	74	153	33	52	59	65	5.79	12.36	8.89	10.42	12.97	11.50	1	1	1	1	1	1	436	10.32	6
Democracy	64	38	20	62	35	58	5.01	3.07	5.39	12.42	7.69	10.27	1	1	1	1	1	1	277	7.31	6
Democratic		13	6	19	16	21		1.05	1.62	3.81	3.52	3.72		1	1	1	1	1	75	2.74	5
Grand Column Totals and Avg. % sub-corpus	138	204	59	133	110	144	5.40	5.49	5.30	8.88	8.06	8.50	2	3	3	3	3	3	788	7.03	17

their overall average percentage of the corpus. This can be illustrated by an example, as shown in Table 2.16, of a pivot table showing how the KKW items, *Regime*, *Democracy* and *Democratic*, from the GOVERNMENT AND SECURITY category in the English sub-corpora are sorted. As Table 2.16 shows, although both items *Regime* and *Democracy* are KKWs in all of the six English sub-corpora, the overall average percentage of the English sub-corpus in which the item *Regime* is a KKW is higher than that of the item *Democracy*. In this way, the KKW item *Regime* is ranked as number one within the GOVERNMENT AND SECURITY category. In terms of the number of KKWs within each category, a pivot table counts the number of the sub-corpora in which each item is a KKW, and hence the grand total (or aggregated) number of the three items, *Regime*, *Democracy* and *Democratic* in the GOVERNMENT AND SECURITY category is counted as 17 as shown under the Grand Row Total rather than only three (which in this example the *actual* number of KKW items in the GOVERNMENT AND SECURITY category).

Initial analysis also showed that quite a number of these ‘KKW-semantic categories’ are also collocates of the term *Arab Spring*. Investigating the collocational behaviour of the term *Arab Spring* is believed to be useful not only for uncovering the nature of key discourses and contexts in which the term tends to occur, but also for diachronically examining these topics in terms of the *OED* definition of the term. That is, the extent to which the *OED* definition represents what has actually been reported by the different English as well as Arabic news media. To this end, analysis has been conducted, by means of *WordSmith Tools 6.0*

(Scott, 2015), by examining the collocation lists for the term *Arab Spring*, ordered by joint frequency, that is, co-occurring within a span of five words on either side of the node Arab Spring, as well as the MI (mutual information) score. Although the MI measure is commonly criticized for its tendency to give more importance to relatively low frequency words, it is believed to be useful since these are usually content words such as nouns, adjectives or verbs that might be clear indications to semantic prosodies or key topics (Gabrielatos and Baker, 2008, p. 11; Baker, 2006, p. 102). However, very low frequency can be restricted, by means of *WordSmith Tools* (Scott, 2015) that allows researchers to ‘set the minimum of co-occurrence frequency of an item to be considered as collocate of a given node word’ (Xiao and McEnery, 2006, p. 105). In this context, the minimum co-occurrence frequency for an item to be considered as collocate of the term Arab Spring was set to five (5), with MI score ≥ 3.00 . The MI calculation method also allows for measuring the difference between the observed and expected frequencies, and hence the strength of the collocation (for comparative details on the different collocation statistical tests, see Baker, 2006, pp. 101-4; Scott, 2015; Mautner, 2007; Gabrielatos and Baker, 2008). Another reason for using the MI statistical measure is perhaps to maintain consistency of the results when comparing them, if comparison is required, with the results displayed by the reference corpora used in this study, as they all have the MI as a collocation score. These are the BNC for the British English, the COCA for the American English, as well as the Leeds Arabic Internet corpus (LAIC) (Sharoff, 2006b) for the Arabic data. The LAIC corpus was created

by the Centre for Translation studies at the University of Leeds, and publicly available online. Its interface, based on the information available on its website, was developed by Sharoff (2006a).

To examine the collocational behaviour of the term *Arab Spring* in the ASC corpus over time, concordances of the term *Arab Spring* have been prepared from both the ARASC (MSA) as well as the three main ENASC (BR, AM, and GV) sub-corpora. The results were collectively calculated in terms of their text types (news and editorials and opinions) as well as year of publication (2011, 2012 and 2013). In other words, a total of 24 collocation list searches have been conducted of which there are eighteen lists for the three main ENASC (BR, AM and GV) sub-corpora (6 each) as well as six lists for the ARASC (MSA) sub-corpus, two for each year. Comparisons of the resulting KKW and the collocation lists were then conducted in order to examine salient topics associated with the *Arab Spring*. Like KKW items, the resulting collocates were also grouped into semantically similar categories, such as GEOGRAPHY, GOVERNMENT AND SECURITY and NAME-PN. However, only KKW items that were also collocates of the term Arab Spring were considered. The following sub-section discusses the advantages and the limitations of the corpus-based approach to CDA analysis.

2.3.5 Advantages and Limitations of Corpus-Based CDA

As has been discussed above, topical analysis of news stories is significant in giving deeper insights into how events are presented in news reports, though it is difficult

to apply this method to a large corpus of texts. In addition, this method might not be applicable in the case of editorial and opinion articles since they have a different schematic structure from that of the news reports. In this context, corpus linguistics becomes a potential method that allows CDA to quantitatively as well as qualitatively work with much larger datasets than the traditional CDA, which mainly deals with ‘qualitative analyses of single texts’ (Baker, 2006, p. 6; Mautner, 2009b). Thus, corpus linguistics software such as *WordSmith Tools* (Scott, 2015) provides the researcher with a variety of tools for analysing corpora and studying patterns. At the quantitative level, frequency and keyword lists, for example, can help the researcher to identify potential sites of interest in the corpus (Baker, 2006; Baker et al., 2008; Mautner, 2009a). A corpus-based approach enables researchers to analyse collocational patterns of words by means of the MI (mutual information) score, which provides information about the degree by which two words or more occur together (Scott, 2015; Mautner, 2009a; Partington et al., 2013; Baker, 2010b). Collocational information, as Hunston (2002) points out, can reveal different meanings of a word under study other compared with its literal or denotative meaning, as a result of its association with other words. She provides an example showing that the collocates of the physical meaning of the word *LEAK* are *oil*, *water*, *gas* and *roof*, while other collocates such as *documents*, *information* and *letter* are associated with the metaphoric sense of the word (Hunston, 2002, p. 76). This implies that corpora allow for extracting semantic profiles from them and identifying the semantic preference and prosodies, which are less accessible to

human intuitions, of words under investigation (Louw, 1993, p. 173; Hunston, 2002; Mautner, 2009a). Furthermore, a corpus-based approach also helps in studying patterns of lexical choices, which usually express different ideological positions (Fairclough, 2001; Stubbs, 1996; van Dijk, 1998). In addition, using large corpora can help reduce researchers' bias and counteract some of the persisting criticisms levelled against CDA such as the cherry-picking of small and unrepresentative samples of texts, which are selected, according to some critics (e.g. Widdowson, 2004), in a way that best suits the researchers' political agendas (Mautner, 2009a; Gabrielatos and Baker, 2008, p. 6; Baker and McEnery, 2015b). Thus, using larger volumes of data, comparing them with reference corpora, and using other corpus-driven techniques, such as KW and KKW functions, ensure objectivity as well as authenticity of the results and findings, and 'help us to avoid over-focussing on atypical aspects of our texts' (Baker and McEnery, 2015b, p. 5). Using reference corpora also helps researchers to test their theories and hypotheses about language (Baker, 2006).

In the same context, applying manual as well as automated corpus tools, such as the *Wmatrix* (Rayson, 2008) methods in grouping the resulting KKWs and/or collocates in the sub-corpora under investigation into semantically similar categories is useful in identifying salient topics at both word as well as semantic levels. Although the *Wmatrix* method is not without limitation. For example, the software is not capable of processing data in different languages other than English, and some modern languages, such as French, Spanish, Finnish and Russian (Rayson, 2008; Prentice

et al., 2012), and hence the Arabic data was not included in the automated categorization. Another limitation of *Wmatrix* is that in many instances, the corpus tool missed distinguishing the different senses, an issue identified by Baker (2004), of the KKW items of the English data, and instances of mistagging were also identified. For example, the KKW item *spring*, was tagged by *Wmatrix* tool under O2, that is, a *general object* belonging to the broader semantic category of SUBSTANCES, MATERIALS, OBJECTS AND EQUIPMENT, which does not reflect its sense in the Arab Spring corpus. Another issue with using *Wmatrix* was the relatively small sizes (and English language forms) of the reference corpora used, compared to the British BNC and the American COCA reference corpora. Therefore, and as recommended by Rayson (2008, p. 544), careful manual analysis of concordances of KW/KKW items or categories was then required to check for mistagging and poor dispersion of high frequency items. Although manual analysis of concordances and categorization of KW/KKW items are generally considered as researcher-driven and subjective (Partington et al., 2013, p. 312), a level of objectivity is believed to be maintained, especially given that, as in the case of this study, the KWs, KKWs or collocates for which concordances are investigated are not subjectively selected, but rather they have to fulfil certain statistical criteria (Gabrielatos and Baker, 2008, p. 7).

However, there are limitations on the use of corpus-based methods from a CDA perspective. One problem with using corpora in discourse analysis is that they contain decontextualized language data which may not enable the researcher, for

example, to know the ideologies of the text producer, and which, therefore, affects the interpretation of the text (Baker, 2006; Mautner, 2009a; Hardt-Mautner, 1995; Flowerdew, 2005; Mautner, 2009b). Another limitation is the lack of contextual features which is considered as problematic, especially when analysis is concerned with pragmatic issues (Flowerdew, 2005, p. 324). Contextual elements may include visuals, colours, fonts, gestures or facial expressions, and reducing corpora to text-only format ‘contributes crucially to the creation of meaning’ (Mautner, 2009a, p. 34; Baker, 2006; Hunston, 2002). In this regard, Partington (2004, p. 11) points out that ‘specialized (or *monogeneric*) corpora make discourse study feasible since, in a collection of texts of a similar type, the interactional processes and the contexts they take place in remain reasonably constant, or at least alter in relatively predictable ways’. Moreover, decontextualized data can be a methodological advantage, since researchers are encouraged to carry out their analysis without preconceived ideas or positions towards the circumstances of text production, hence reducing the level of bias or subjectivity. Nevertheless, ‘the de-contextualization of a text into a list of words is but the first step of a corpus linguistic approach’ (Archer, 2009, p. 160; Scott and Tribble, 2006). Additionally, corpus linguist linguists who regularly use (key) word lists emphasize the importance of re-connecting those lists with their original texts from which they came, and where possible, with their context of production in order to better understand the meaning behind the language used (Archer, 2009, p. 160; Scott and Tribble, 2006). Finally, there is no doubt that the corpus-based approach significantly contributes to CDA as it helps in

triangulating hypotheses about language as well as research findings and results by means of using multiple methods of analysis (Baker, 2006; Baker et al., 2008).

2.4 Summary

In this chapter the methods used to investigate the representation of the Arab Spring narrative in English and Arabic news media were highlighted. It is divided into four main sections. Section 2.1 is an introduction; Section 2.2 is concerned with the methodological procedures for data collection and corpus design. It includes seven subsections: sub-section 2.2.1 is a description of the Arab Spring bilingual corpus in terms of its comprising English and Arabic sub-corpora, word count and text types. Sub-section 2.2.2 introduces the transcription scheme of the Arabic data used in this study, followed by subsection 2.2.3, which continued with discussing linguistic differences between the two languages, and highlighted some of the characteristic features of Arabic. Sub-section 2.2.4 highlighted some of the challenges in cross-linguistic corpus-based and CADS analyses at the translational, theoretical and methodological levels. The ‘protest vocabulary’ and the development of the query used to collect data were discussed in section 2.2.5. The reference corpora used in the analysis, and the ethical considerations regarding corpus design and compilation were also discussed in sub-sections 2.2.6 and 2.2.7 respectively.

Section 2.3 is the main third part in this chapter, which introduced the methodological and theoretical frameworks of CDA/CADS and CL within which

the study is conducted. Several theoretical notions in CL such as keyness, collocation, semantic prosody and semantic preference, as well as CDA concepts such as topics, and their attendant analytical tools such as frequency lists, frequency clusters, KW and KKW lists were also outlined. Section 2.3 had four main sub-sections: sub-section 2.3.1 was concerned with topical analysis in which the differences between the two text types (that is, news and editorials and opinion texts), in terms of their schematic structures and topical analysis were also discussed. Initial analysis showed that both the English as well as the Arabic editorials and opinion articles followed the LFV structure, or the *triad*, defined by Bolivar (1994) as the minimal unit of interaction in written texts. In sub-section 2.3.2, criticisms of the keyness approaches in terms of its definitions and their compatibility with the different statistical measures, such as log likelihood (LL) and frequency difference (%DIFF) as well as frequency ration were discussed. The challenges to cross-linguistic corpus-based studies using keyness approaches were also discussed in sub-section 2.3.2. In sub-section 2.3.3, some of the methodological issues regarding the creation of the keyword database, the production of the final version of the KKW lists and the decision taken to resolve them were also discussed. This was followed by sub-section 2.3.4, which discussed the automated and manual categorization of the resulting English and Arabic KKW items, and illustrated with an example of how pivot tables in Excel can be used to process large scale data and plot results. Sub-section 2.3.5 concluded with an outline of the advantages and limitations of the corpus-based approach to CDA,

indicating that despite limitations, corpus linguistics analytical methods, at both quantitative and qualitative levels, significantly contribute to CDA in terms of both the authenticity and objectivity of the analysis as well as results and findings.

Therefore, by using the different techniques and methods from corpus linguistics (CL) and critical discourse analysis (CDA), my study, through the three following analysis Chapters (3, 4 and 5), aims to examine how the Arab Spring narrative is represented in the English and Arabic news media in terms of the *who*, *what*, *where* and *why*. Analysis in Chapter 3 aims to identify *who* the main news actors are in the Arab Spring narrative, and how they are represented in terms of van Leeuwen's (van Leeuwen, 1996, 2008) CDA model for the representation of social/news actors. Relevant strategies include *nomination*, *categorization*, *inclusion* and *exclusion*. It mainly focuses on the NAME-PN category, and other KKW items of interest in the other semantic categories in both the English and Arabic sub-corpora. In Chapter 4, the KKW list function (Scott, 2015) is mainly used in order to identify *what* the main topics/events, or semantic macrostructures (van Dijk, 1988b) are that are prevalent across both English and Arabic sub-corpora, and *where* they are located. Chapter 5 attempts to investigate *why* the stories/topics covered are important to the news media, and hence to the reading public. It mainly draws on Bell's (1991) taxonomy of news values in news events in order to investigate the newsworthiness in the Arab Spring narrative.

3. Who is Being Reported? Representation

Strategies of News Actors

3.1 Introduction

This chapter investigates *who* the main news actors are in the Arab Spring narrative, and how they are discursively constructed in both the English as well as the Arabic sub-corpora. It mainly draws on van Leeuwen's (2008, 1996) *Social Actors Network* (see Figure 3.1), and hence examines the ways in which the main social/news actors across the resulting top ten KKW-semantic categories (see Tables 2.15 and 4.1) are linguistically realized through the different discursive strategies, such as *inclusion* (e.g. *personalisation* and *impersonalization*) and *exclusion* (e.g. *suppression* and *backgrounding*). According to Bell (1991, p. 175), the category of *news actors* is one of the main elements that make up the structure of news stories. It relates closely, along with other elements such as place, time, source, headlines and leads, to the journalist's short-list (the five W's and an H) of what should go in a news story. Thus, identifying *who* are involved in the Arab Spring narrative, how they are represented, *where* they come from, and what events/topics they are associated with is very important in understanding *why* the Arab Spring narrative is important to the different English and Arabic news media, and hence to their reading public. Analysis in this chapter starts with discussing the main news actors identified in the NAME-PN category of both the English as well as

the Arabic sub-corpora, and locating them geographically. It then continues with discussing the main representation strategies as identified in this as well as the remaining semantic categories. Before getting further into who the news actors are, and how they are represented, it is worth first discussing the main representation strategies for news actors that are of relevance in this chapter, that is inclusion and exclusion. According to van Leeuwen (2008, p. 40), social/news actors can also be represented either in terms of their unique identities or by what they do, in other words, whether they are nominated or categorized. I, therefore, use the analysis in this chapter as a starting point for my analysis, developing it in Chapters 4 and 5 by means of investigating the news events/topics and news values.

3.2 Representation of News Actors: A Sociosemantic Inventory

Adapted from van Leeuwen's (2008, p. 52) Social Actors Network, Figure 3.1 shows some of the ways in which the two main strategies, inclusion and exclusion, are realized. According to van Leeuwen (2008, p. 28; 1996, p. 38), 'representations include or exclude social actors to suit their interests and purposes in relation to the readers for whom they are intended'. In cases of inclusion, the news/social actors are said to be foregrounded and emphasized. That is, they are clearly expressed in the news, and get the attention of the news media.

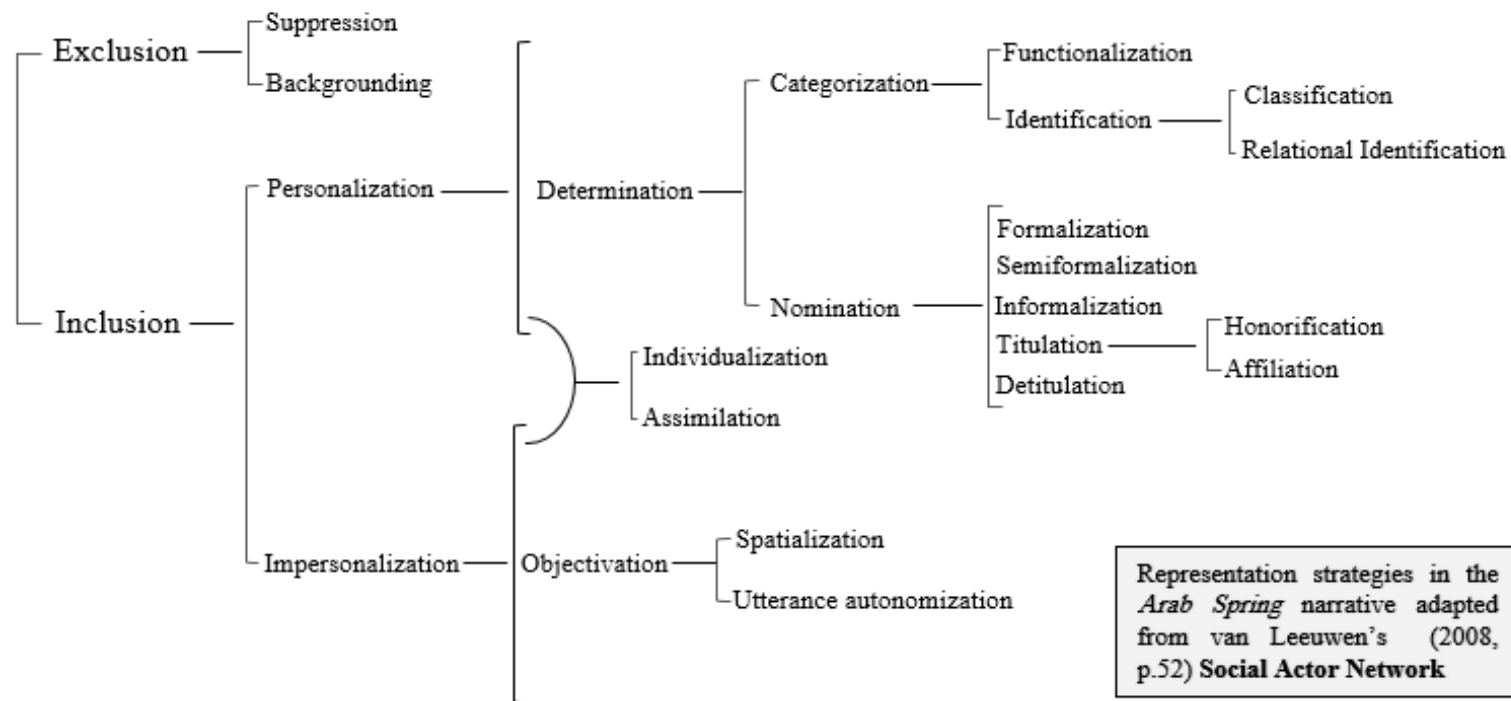


Figure 3.1 Adaptation of van Leeuwen's (2008) analytical framework of news actors in the Arab Spring

In terms of exclusion, the social/news actors are either backgrounded or suppressed (van Leeuwen, 2008). Although some exclusions may be ‘innocent’ as readers are assumed to know the details already, it nevertheless is considered as an important aspect of critical discourse analysis, especially in cases of ‘critical comparison of different representations of the same social practice’ (van Leeuwen, 2008, p. 29). These two main strategies are discussed in more detail in the two following sections.

3.2.1 Exclusion

Exclusion mainly occurs when the ‘relevant actions are included, but some or all of the actors involved in them are excluded’ (van Leeuwen, 2008, p. 29, 1996, 39). There are two ways through which exclusion can be realized: suppression and backgrounding. Suppression is when there is no reference to the news/social actor(s) anywhere in the text, whereas backgrounding is when the excluded actors are not mentioned in relation to any action, and can be traced elsewhere in the text (van Leeuwen, 2008, p. 29). Additionally, and in cases of ‘radical exclusion’, both the news/social actors and their activities are excluded. Therefore, exclusion is considered, according to van Leeuwen (2008, p. 28-9), as an important aspect of critical discourse analysis. In other words, critical discourse analysis (CDA) is not only interested in accounting for what is present in the text, but also in what is absent (Kress and van Leeuwen, 2001; van Leeuwen, 1993; Baker et al., 2008; Partington, 2014; Fairclough, 1995a, 2001, 1989). One way to identify what is absent is to conduct a contrastive analysis of different representations, texts, or

version of the same social practice or phenomenon (van Leeuwen, 2008, 1996, 1993; Fowler, 1996, p. 35). In his study about the representation of social subjects and their practices, van Leeuwen (1993, p. 207) provided examples of exclusion of the participants in one social practice, that is ‘going to school’, in different texts. While all school staff other than the teacher were suppressed in one text, secretaries and heads were not in the other, but often excluded through backgrounding (van Leeuwen, 1993, p. 208). In a similar approach, and in his study that is concerned with the role of corpus linguistics in researching absences, Partington (2014) tracks appearances and disappearances of the items *government* and *regime*, between December 2010 and the end of November 2011 in the White House press briefings on the Arab uprisings. Partington (2014, p. 129-32) also explains how keyness of items can be considered as absolute/relative presences and absences when comparing/contrasting different texts or corpora. From this perspective, the KKW lists of the English (the American, the British and the Global Variety) as well as the Arabic sub-corpora can be investigated to compare between their different varieties and text types, and examine how the main news actors are represented from the perspectives of different languages and language varieties, and hence different cultures and ideologies. According to Partington (2014, p. 129) ‘[p]erhaps the most obvious and most frequent way of identifying what is entirely or relatively absent – what is either missing or rare – in one set of texts is to compare that set with another set’. Thus, in this chapter, investigating how the main news actors involved in the Arab Spring narrative, and which quite a number of them are shared between

the English (that is, the American, the British and the Global Variety) and the Arabic sub-corpora, can be useful in identifying similarities and/or difference between them in terms of the representation strategies employed by the producers of texts.

Finally, in terms of their linguistic realization, both suppression and backgrounding can be actualized through *passive agent deletion*, *non-finite clauses*, as well as *nominalization and process nouns*. Examples 11 and 12 are strategies of passive agent deletion as well as nominalization respectively:

11. Secular leader killed in Arab Spring city
(TMSe-13-60/ news)

12. Killing of Islamists Deepens Crisis in Egypt
(NYTe-13-85/ news:)

3.2.2 Inclusion

As Figure 3.1 also shows, there are two main aspects of inclusion: Personalization and Impersonalization. *Personalization* is usually realized through *nomination* and *categorization*. According to van Leeuwen (2008, p. 41), nomination is basically realized by proper nouns, which can be formal, through the use of *surname* only (e.g. *Obama* - أوباما) with or without honorifics (e.g. *President Obama*). Another strategy of nomination is semi-formalization, which gives the *name* and *surname*, such as *Barack Obama*, as shown (in bold) in Example 13 (line 2), whereas

nomination can be informal through given name only. Example 14 (lines 1 and 4), shows how *Bashar Al-Assad of Syria* is informally represented through his given name only:

13. 1 That report was a template for the anti-Iraq war
2 positions taken by **Barack Obama**, ...

(NYTe-13-29/ news: 16-March)

14. 1 فيقول عدنان: "... ولم تتحسن الأوضاع بعد أن تولى بشار السلطة.
2 والآن تغير كل شيء."
3 and Adnan says: "... and the situation did not
4 improve after **Bashar** took power, and now
5 everything has changed."

(WSTa-12-32/ news: 18-Jan)

In terms of *categorization*, this is realized through *functionalization* and *identification*. *Functionalization* is actualized when news actors are referred to in terms of activity or something that they do (van Leeuwen, 2008, p. 42). For example, in the GOVERNMENT AND SECURITY category, *functionalization* is actualized in KKW items of both Arabic and English, such as <الرئيس>the-president>; <الوزير>the-minister>; *officials* and *activists* can also be considered as actualizations of *functionalization*. On the other hand, *identification* occurs when social/news actors are defined in terms of what they, more or less permanently are, which include age, gender, provenance, race, ethnicity, class, wealth and religion (van Leeuwen, p. 42).

The second main aspect of inclusion is *impersonalization*. It is realized through *objectivation* of which *spatialization* and *utterance autonomization* are two main aspects. *Spatialization* occurs when social/news actors are referred to in terms of place. This happens, as Example 15 (lines 2 and 6) shows, when the items, such as <the Americans> الأمريكيين; and <the Israelis> الإسرائيليين, which are seen as aspects of personalization, are substituted by the items <America> أمريكا; and <Israel> إسرائيل:

15. 1 واعتبر الحروب أن الربيع العربي هو نقطة تحول تاريخي أربك
 2 أمريكا وإسرائيل وفتح النوافذ لكي يقوم الفلسطينيون باستغلال
 3 هذا الربيع العربي لصالح القضية الفلسطينية.
 4 and Alhroub [Palestinian activist] considered the
 5 Arab Spring as a historical turning point that
 6 confused **America and Israel**, and opened the
 7 windows so that the **Palestinians** exploit this
 8 Arab Spring in favour of the Palestinian cause
 (BWBa-12-26 news/ 25-Jan)

The second aspect of objectivation is *utterance autonomization* in which social/news actors are represented in terms of their utterances (van Leeuwen, 2008, p. 46). Examples 16 (line 1) and 17 (line 1) show instances of *utterance autonomization* in which the items *report says* and *death reports* represent a kind of impersonal authority to the utterances, which are often used in connection with the utterances of high-status and official spokespersons. Nevertheless, looking at the two examples from the CDA perspective, the grammatical forms (singular and plural) have an ideological role to play.

16. HEADLINE Egypt: **Report says** Bahrain used torture in crackdown
(GRDe-11-152/ news: 5-Nov)
17. HEADLINE **Despite death reports,** Syria denies Hama campaign
(CNNe-11-112/ news:

The following section introduces the main news actors in the English as well as the Arabic NAME-PN categories.

3.3 News Actors in the Arab Spring Narrative: The Elite

Tables 3.1 and 3.2 show the resulting KKW items in the NAME-PN category of both English and Arabic sub-corpora respectively along with their average percentage of each sub-corpus. The NAME-PN category ranks as number three amongst the top ten key semantic categories (see Tables 4.1 and 4.2 in section 4.2) with an overall total number of 75 KKW items of English and Arab altogether, as Table 3.1 shows. The English sub-corpus contains a total of 46 KKW items of which only 15 (in bold) are shared between its main six component sub-corpora. The 15 shared KKW items include names of prominent people (e.g. *Al-Assad's*, *Obama's*, *Netanyahu*, *Mubarak*, *Morsi*, *Abbas*, *Erdogan*, and *Gaddafi*), and names of groups, organizations and political parties, such as *(Muslim)Brotherhood*, *Hamas*, *Al-Qaeda*, and *Hezbollah*. As Table 3.1 also shows, only six (bold and underlined) of these shared items, that is, *Al*, *Obama's*, *Mubarak*, *Gaddafi* and *Hamas*, are collocates of the term *Arab Spring*.

Table 3.1 KKW items of the NAME-PN category with their overall Avg. % of the English sub-corpus as a whole

KKW items of the NAME-PN category of the English sub-corpus			
PERSONAL NAME		4.15	
Mr	15.76	Ms	2.64
(Al) Assad	13.81	(bin) Laden	2.43
<u>Obama</u>	10.62	Morsi's	2.26
Netanyahu	6.88	Abu (Mazen)	2.14
<u>Mubarak</u>	6.25	Gaddafi's	2.12
Morsi	5.20	Mohammed	1.89
Abbas	4.70	Ahmadinejad (of Iran)	1.71
Erdogan	3.87	Putin (of Russia)	1.70
Bin (Laden)	3.85	Netanyahu's	1.66
<u>Obama's</u>	3.57	(Hilary) Clinton	1.59
Assad's	3.41	El	1.52
Gaddafi	3.37	Bashar (Al-Assad)	1.47
(Ben) Ali	3.31	Bouazizi	1.39
Abdullah	3.24	(Al) Maliki	1.34
Ben (Ali/'s)	2.79	Sarkozy	1.30
Saleh	2.71		
GROUPS AND ORGANIZATIONS		4.14	
(Muslim) Brotherhood	10.07	EU	2.78
<u>Hamas</u>	8.44	AKP (of Turkey)	2.34
UN	5.69	(United) Nations	2.32
<u>(AL)</u> Qaeda	4.53	Fatah (of Palestine)	2.13
(Security) Council	3.89	NATO	1.89
Hezbollah	2.99	(Arab) League	1.69
Ennahda (Renaissance /of Tunisia)	2.98	(Al) Qaeda's	1.45

Bold: Shared KKWs

Bold underlined: Shared KKWs and collocates of the term *Arab Spring*

Table 3.2 KKW items of the Arabic NAME-PN category with their Avg. % of the Arabic sub-corpus as a whole

KKW items of the NAME-PN category of the Arabic sub-corpus		
PERSONAL NAME		2.10
مرسي	<u>Mrosi</u> (of Egypt)	4.65
الأسد	<u>Assad</u> (of Syria)	2.79
نتنياهو	<u>Netanyahu</u> (of Israel)	2.73
أوباما	<u>Obama</u> (of United States)	2.30
مبارك	<u>Mubarak</u> (of Egypt)	2.23
بن	<i>Ben</i>	2.18
أردوغان	<i>Erdogan</i> (Turkey)	2.17
القذافي	<i>Gaddafi</i> (Libya)	2.02
بشار	<i>Bashar</i> (Al-Assad)	1.93
نجاد	<i>Nijad</i> (of Iran)	1.74
هنية	<i>Haniyeh</i> (of Hamas)	1.38
حمد	<i>Hamad</i>	1.38
الغنوشي	<i>al-Ghannushi</i> (of Tunisia)	1.28
آية	<i>Aya</i>	1.26
البوعزيزي	<i>Al-Bouazizi</i>	1.24
كلينتون	<i>Clinton</i> (of United States)	1.21
صالح	<i>Saleh</i> (of Yemen)	1.19
المرزوقي	<i>Al-Marzouqi</i> (of Tunisia)	1.12
لاريجاني	<i>Larijani</i> (of Iran)	1.08
عباس	<i>Abbas</i> (of Palestine)	1.06
groups and organizations		2.22
الإخوان	(Muslim) <u>brothers</u>	3.80
حماس	Hamas	3.16
القاعدة	Al-Qaeda	1.91
النهضة	<u>Ennahda</u> (of Tunisia)	1.86
العدالة والتنمية	<u>Justice and Development</u> (Party of Turkey)	1.74
مجلس	<i>council</i>	1.63
المجلس	<i>council</i>	1.49
القمة	<i>summit</i>	1.47
الأمم	(United) Nations	1.15
الاتحاد	(European) Union	1.03
Bold: Shared KKWs		
Bold underlined: Shared KKW and collocate of the term		
<the Arab Spring> الربيع العربي		

The remaining 31 KKW items are shared between only five, four, or three sub-corpora. The KKW items (6 items in total) that are shared by five sub-corpora include names of prominent people (e.g. *Ali*, *Abdullah*, *Mohammed* and *Ahmadinejad*), and groups or organizations (e.g. *Fatah* and *NATO*). The KKW items (totalling 14 KKW's) shared by four sub-corpora, also include personal names of either prominent people (e.g. *Mr*, *Ben (Ali)*, *Saleh*, *(Bin) Laden*, *Gaddafi's*, *Netanyahu's*, *Bashar*, and *Maliki*), ordinary people (e.g. *Mr*, and *Bouazizi*), political parties (e.g. *Ennahda* and *AKP*) as well as international organizations (e.g. *UN* and *EU*). Finally, the KKW items (10 items) that are shared by three sub-corpora contain names of prominent people (e.g. *Ms*, *Morsi's*, *Putin*, *(Hilary) Clinton*, and *Sarkozy* as well as groups and organizations (e.g. *(Security) Council*, *(United) Nations*; *(Arab) League*, and *(Al) Qaeda's*). Only two KKW items (i.e. *Clinton* and *Nations*) of this groups is collocates of the term *Arab Spring*.

On the other hand, Table 3.2 shows the Arabic NAME-PN category, which contains a total of 31 KKW items of which 15 are shared between its two main sub-corpora; the bolded items, as shown below, are also collocates of the term *الربيع العربي* <the-Arab Spring>. The 15 shared KKW items include names of prominent people and politicians, such as

مرسي < <i>Morsi</i> >;	الأسد < <i>Al-Assad</i> >;	نتنياهو < <i>Netanyahu</i> >;
أوباما < <i>Obama</i> >;	مبارك < <i>Mubarak</i> >;	أردوغان < <i>Erdogan</i> >;
القذافي < <i>Gaddafi</i> >;	بشار < <i>Bashar</i> >; and	البوعزيزي < <i>Bouazizi</i> >;

as well as names of groups, parties and organizations, which include:

(المسلمين) الإخوان <the-Brothers> (Muslim);
 حماس <Hamas>;
 القاعدة <Al-Qaeda>;
 النهضة <Ennahda> (Ennahda Party, Tunisia)
 العدالة والتنمية (حزب) <Justice and Development> (Party, Turkey).

Similarly, the remaining 16 KKW items of the Arabic NAME-PN category are all KKW items in the Arabic news sub-corpus except the KKW item عباس <Abbas>, which is only a KKW in the Arabic editorials and opinion sub-corpus. Eleven of these KKW items also refer to prominent people and politicians, of which the bolded as shown below, are collocates of the term الربيع العربي <the Arab Spring>:

بن (علي، لادن) <Ben> (e.g. Ben Ali and Ben Laden);
 نجاد (أحمدي) <Nejad> (i.e. Ahmadinejad, Iran);
 هنية (إسماعيل/ حماس) <Haniyeh> (i.e. Ismail Haniyeh, of Hams);
 حمد (قطر) <Hamad> (of Qatar);
 الغنوشي (راشد/ النهضة) <Al-Ghannoushi> (of Ennahda Party/ Tunisia);
 آية (إيران) <Aya> (e.g. Ayatollah, Iran);
 كلينتون (هيلاري) <Clinton> (i.e. Hilary Clinton);
 صالح (علي عبد الله/ اليمن) <Saleh> (e.g. Ali Abdullah Saleh, Yemen);
 المرزوقي (منصف/ تونس) <Al-Marzouki> (i.e. Moncef Al-Marzouki, Tunisia);
 لاريجاني (علي/ إيران) <Larijani> (i.e. Ali Larijani, Iran); and
 عباس (السلطة الفلسطينية) <Abbas> (e.g. the Palestinian Authority).

The remaining KKW items (5 KKWs in total) also refer to political as well as international groups and organizations, of which the bolded as shown below, are also collocates of the term الربيع العربي <the Arab Spring>:

- مجلس الأمن <**Council**> (e.g. *the-Security Council*);
 المجلس الانتقالي <*the-council*> (e.g. *the-Transitional Council*);
 القمة العربية <**the-summit**> (*the Arab Summit*);
 الأمم المتحدة <*the-Nations*> (e.g. *the-United Nations*); and
 الاتحاد الأوروبي <*the-union*> (*the-European Union*).

One of the interesting points to draw from these names in both English and Arabic is that the number of KKW items referring to names of prominent people is higher than the number of items referring to groups and organizations. Statistical analysis shows that this pattern is common across the different English (AM, BR, and GV) and Arabic (MSA) text type (i.e. news and editorials and opinions) sub-corpora. Nevertheless, most of these KKW items (of people and entities) are indicators for one significantly salient news value of *Eliteness*, indicating the importance as well as the prominence of topics associated with the Arab Spring narrative across the English as well as the Arabic text type sub-corpora.

3.4 Locating News Actors

Figure 3.2 shows the geographical distribution of the main news actors (people and groups and organizations) across the Arab Spring text type news sub-corpora. As indicated in Figure 3.2, the main news actors involved in the Arab Spring narrative

were classified into five main groups in terms of the countries or geographical locations to which they belong. Ordered by the overall average percentage of the English sub-corpus. These categories include the *Western and European*, the *Arab Nationals*, the *International*, the *Islamic*, and *Russia*. The aforementioned groups are highlighted in red, grey, blue, green, and dark blue respectively. In terms of the news actors displayed in Figure 3.2, all are shared between the main English and Arabic sub-corpora except actors highlighted in small coloured circles. The blue circle indicates that the news actor is only key in the English sub-corpus as a whole, whereas the yellow circle indicates that the news actor is only key in the Arabic sub-corpus. The news actors, as Figure 3.3 shows, represent both the PEOPLE as well as the GROUPS AND ORGANIZATIONS sub-categories of NAME-PN. In terms of the KKW items *Mr* and *Ms* in the English sub-corpus as a whole, these were not included in Figure 3.2, but considered in the analysis. As Figure 3.2 shows, the most frequent news actors in the NAME-PN category, and hence most included in the Arab Spring narrative come from countries and locations belonging to the *Western and European Group*. It includes the *United States*; *Israel*, *Europe*, and *France*. From the United States, President *Obama* and US Foreign Secretary of State (*Hillary Clinton*). In the American news, the most frequent topics associated with Clinton are Israel and Obama. From Israel, we have Prime Minister (Benjamin) *Netanyahu*, and from Europe it is the *EU* (or the *European Union*), which is only key in four sub-corpora in the British corpus and the GV-ENASC sub-corpora as well as the Arabic news.

Geographical distribution of the KKW news actors of the NAME-PN category with their Avg. % of each of the English and Arabic text type sub-corpora



Figure 3.2 Geographical distribution of KKW news actors of the NAME-PN category of the English and Arabic sub-corpora

The last news actor from the *Western and European Group* is former President *Sarkozy* from France.

In terms of the *Arab Nationals Group*, it contains twelve countries along with a general Arab World category that contains القمة العربية <the Arab summit>, and the (Arab) League (see Tables 3.1 and 3.2). The Arab countries from which the Arab news actors (18 in the English, and 12 in the Arabic sub-corpora) come are *Syria*; *Egypt*; *Palestine*; *Jordan*; *Saudi Arabia*; *Lebanon*; *Libya*; *Yemen*; *Tunisia*; *Qatar*; *Iraq*. The main news actors in this group are distributed as follows: from Syria, we have President *Bashar Al-Assad*; from Egypt ousted President Mohammed *Morsi*, former President *Mubarak* and the (Muslim) *Brotherhood*; from Palestine, current President Mahmoud *Abbas* (Abu-Mazen); *Fatah* (PLO/ Palestine Liberation Organization); Hamas leader (Ismael) *Haniyeh* and *Hamas*; from Jordan, King *Abdullah II*; from Saudi Arabia, *Al-Qaeda*. From Lebanon, the Shi'ite militant group *Hezbollah*, from Libya, *Gaddafi*; from Yemen, ousted President (Ali Abdullah) *Saleh*, and finally from Tunisia, which contains the highest number (5) of news actors, but only three that are mainly shared between the English and Arabic sub-corpora. These include, the ousted President (Ben) *Ali*, *Ennahda* (Party) and *Bouazizi*; from Qatar, Sheik Tamim Ben *Hamad* Al-Thani, which is only key in the Arabic news, whereas Prime Minister Nouri (Al) *Maliki*, from Iraq, is only key in the American and the British news as well as both of the GV sub-corpora.

The International Group includes the *UN* and (UN Security) *Council* and *NATO* (the North Atlantic Alliance). *The Islamic Group* includes two main countries, Turkey and Iran. From Turkey, we have Prime Minister (Recep Tayyeb) *Erdogan*; which is shared between all the Arab Spring sub-corpora of both languages English and Arabic; *AKP*, or العدالة والتنمية (حزب) <*Justice and Development Party*> which is only key in both of the GV-ENASC and Arabic text type sub-corpora.

Finally, the *Russia Group*, which only includes President Vladimir Putin is key only as Figure 3.2 shows in the both American sub-corpora as well as the British editorials and opinions. The keywords that are mostly associated with Putin, are *Russia*, *Mr*, *Syria*, *protesters* and *trial*, indicating the nature of contexts with which they are associated.

One of the most important observations to draw from Figure 3.2 is perhaps the presence of quite a number of names referring to groups and organizations. In the *Arab Nationals Group* the most significant news actors belong to the GROUP AND ORGANIZATION sub-category. This tells us, perhaps, that news actors from the Arab world are represented in a collective rather than individual way. For example, the keywords that are mostly associated with the item *EU* in the English sub-corpora are *Turkey*, *Iran* and *sanctions* and *aid*, but also collocate with *UN*, *said*, *membership*. Similarly, in Arabic sub-corpus, the item الاتحاد <al-'ittiḥād - the-Union.SING.MASC.N – *the Union*> mostly collocates with the item الأوروبي <al-'awrūbbī – the-European.SING.MASC.ADJ – *the-European*> in the Arabic news.

Other words are دعم <support>, and مساعدات <aid> as well as عقوبات <sanctions>. As can be seen, these are the contexts in which the term EU and الاتحاد الأوروبي <al-’ittiḥād al-’awrūbbī - *the European Union*> are used, but as Bang (2003, p. 62) points out, ‘the question is who is represented as a giver of aid and who is represented as a receiver’. According to van Leeuwen (2008, p.33), ‘the representations can endow social actors with either active or passive roles. Thus, in the context of the Arab Spring, the EU is mainly represented as the caretaker and the provider of support to the Arab Spring countries especially Egypt, and other Arab nations as the Examples from 18 to 23 show.

18. HEADLINE 1 **EU aid to Arab nations** may be linked
2 to reforms
LEAD 3 **BRITAIN IS** pressing the **European**
4 **Union** to threaten to halt aid to countries
5 in the **Arab world** who do not live up to
6 their promises on democracy, human
7 rights and economic reform.
(NDPe-11-10 / news: 16-Jun)
19. HEADLINE 1 Egypt: Egypt condemns **EU** threats to halt
2 **aid** amid mounting death toll: Minister
3 insists crisis is Egypt's internal affair 38
4 Muslim
20. HEADLINE **EU ends** arms embargo against Syrian
rebels (CNNe-13-66/ news: 27-May)
21. HEADLINE 1 Lost without trace, the £500m of **EU aid**
2 to counter **corruption in Egypt**
3 (TMSe-13-43/ news: 18-July)
22. HEADLINE 1 وزارة الخارجية: الاتحاد الأوروبي يتحمل مسؤولية
2 القرارات الطائشة وافتعال الأزمات
3 *Foreign Ministry: The European Union*
4 *takes responsibility responsible for*

5 *making reckless decisions and fabricating*
crises (MHRa-12-18/ news: 24-Jan)

23. HEADLINE 1 We'll turn our guns on Libya rebels if they
 2 attack civilians, **Nato threatens**
 3 (NDPe-11-103/ 9-June)

The themes and topics that are associated with both *EU* and *NATO* are relatively negative. These two powerful institutions from the *Western and European* as well as the *International Groups* shows the context in which these two items occur.

In the following section, analysis continues with news actors of the NAME-PN as well as the other semantic categories of both the English (American, British, GV), and the Arabic sub-corpora. It focuses on some of the representation strategies introduced by van Leeuwen (2008), such as *functionalization*, *classification*, *assimilation* and *objectivation* as forms of personalization/ impersonalization strategies, and discusses how the semantic categories, along with their KKW constituents help in identifying such strategies

3.5 Representing News Actors in the Arab Spring Narrative

3.5.1 Functionalization

According to van Leeuwen (2008, p. 42), when social/news actors are referred to in terms of their role or occupation, it then can be said that they are functionalized. *Functionalization*, which is considered as one aspect of the substitution strategy in

van Leeuwen's (2008, p. 53) Social Actor Network, can be realized through suffixes such as *-er*, *-ant*, *-ent*, and *-ian* (van Leeuwen, 2008, p. 42). Analysis showed that quite a number of news actors are functionalized by means of KKW items belonging to different semantic categories. These include, the GOVERNMENT AND SECURITY, PROTEST CATEGORY, SOCIAL ACTIONS, STATES AND PROCESSES as well as LANGUAGE AND COMMUNICATION. For example, there are 14 (Table 3.3) out of a total of 61 KKW items of the GOVERNMENT AND SECURITY category of the English sub-corpus by which functionalization of the news actor is realized. There are 7 KKW items (that is, *minister*, *president*, *officials*, *prime*, *activists*, *ambassador* and *colonel*) that refer to elite news actors. Some of these items (e.g. *minister* and *president*) are also parts of the nomination strategies. The remaining 7 KKW items (that is, *militants*, *fighters*, *extremists*, *refugees*, *civilians*, *citizens* and *voters*) refer to non-elite news actors. However, it is worth mentioning that the above KKW items, like other items of other semantic categories, can also be considered as realizations of other two strategies, that is, individualization and assimilation, which are realized by singularity and plurality respectively. Thus, it can be said from the items above that elite news actors are usually individualized, whereas the 'ordinary' are assimilated, as Table 3.3 shows. Thus, according to van Leeuwen (2008, p. 53), 'in actual discursive strategies, the choice need not always be rigidly either-or. [...], and social actors can be, for instance, both classified and functionalized'. To continue with these KKW items from the perspective of functionalization, Table

3.3 shows that there is a difference between the different English text types sub-corpora in terms of the significance of these items:

Table 3.3 Functionalization by KKWs of the GOVERNMENT AND SECURITY category with their % of each English text type sub-corpus

Functionalization of news actors of the English GOVERNMENT AND SECURITY category							
KKW		News			Editorials and pinions		
		AM	BR	GV	AM	BR	GV
elite	<i>minister</i>	9.95	2.83	11.05	3.01		8.14
	<i>president</i>	4.54	10.18	5.12	4.01	6.37	1.95
	<i>officials</i>	7.67	2.67			1.10	
	<i>prime</i>	4.07	2.58	5.39	1.40		5.49
	<i>activists</i>	7.28	5.49	2.43	2.61	1.98	1.59
	<i>ambassador</i>	2.43	1.62	2.70	1.00	1.98	
	<i>colonel</i>	1.80	1.13			1.54	
non-elite	<i>militants</i>	2.74	2.50			2.20	
	<i>fighters</i>	3.13	1.70	1.35			
	<i>extremists</i>	1.49	1.21		1.80	1.32	1.24
	<i>refugees</i>	1.02	1.05	2.16	1.20		1.59
	<i>civilians</i>	1.33	1.37			1.10	
	<i>citizens</i>				1.00	1.32	1.42
	<i>voters</i>	1.10	1.05			1.10	
Column Total Avg. %		3.73	2.72	4.31	2.00	2.00	3.06
Bold		Highest value in each sub-corpus					

For example, while the KKW item *minister* is most significant in the American and GV news, and the GV editorials sub-corpora, the KKW items *president* has the highest value of 10.18% of the British, reflecting the relatively significant levels of

inclusion of the different elite news actors (e.g. of the NAME-PN category). Looking vertically at the results shown in Table 3.3 could help in giving insights into the roles or occupations that are most relatively included/excluded or foregrounded/backgrounded in each of the English text types sub-corpora in terms of the GOVERNMENT AND SECURITY category. Thus, from the perspective of Partington's (2014, 129) notion of keyness as absolute and relative absence/presence, the most relatively present news actors filling the 'elite' role are *minister*, followed by *officials* and *activists* with highest values of (9.95%), (11.05%) and (8.14) of the American news sub-corpus respectively (see Table 3.3). In terms of the 'non-elite' roles, it can be said that the 'ordinary' news actors are functionalized in terms of roles with mostly negative meanings or connotations, such as *fighters* and *militants* with values of 3.13% and 2.74% of the American news sub-corpus respectively. In the British news sub-corpus, it is the ordinary news actors that fill the role of *militants* that are mostly included with percentage of 2.50% of the British news sub-corpus. In the GV-ENASC sub-corpus, on the other hand, it is the 'ordinary' news actors filling the role of *refugees* that are mostly included with the highest values of 2.16% and 1.59% of the GV news and editorials and opinion sub-corpora respectively.

In terms of similarities, it can be said that both the American and the British news sub-corpora are the most similar in terms of the functionalization strategy in the sense that both are sharing the same number of KKW items, though significantly different especially in the elite part, which also reflect the same roles and

occupations of news actors covered in their news. In terms of the English sub-corpus as a whole, only the roles of *president* and *activists* are shared between the six English text type sub-corpora (Table. 3.3). For example, the most frequently functionalized news actors in terms of the role realized by *president* (3, 494 concordance lines) of the American news sub-corpus are: *Barack Obama*, *Bashar Al-Assad*, *Hosni Mubarak*, *Mahmoud Abbas*, *Mohamed Morsi*, *Zein el Abidine ben Ali*, *Abdullah Gul* (of Turkey), *Ali Abdullah Saleh*. The the least frequently functionalized are *Nicolas Sarkozy* (of France) and *Vladimir Putin* (of Russia). One of the observations made from this pattern is that there is a difference in terms of the formalization (that is, surname) /semi-formalization (that is, name and surname) between these names. For example, out of 416 co-occurrences with the item *president*, *Barack Obama* is functionalized and formalized in 216 instances as *President Obama*, and in 169 instances is semi-formalized as *President Barack Obama*. On the other hand, *Bashar Al-Assad*, is formalized in 23 instances as *President Assad*, and semi-formalized as *President Bashar Al-Assad* in 393 instances out of a total of 411 co-occurrences with *president*. In terms of the British news, however, *Bashar Al-Assad* is the most frequently functionalized in terms of the role *president* than *Obama*. For example, out of 2,985 instances of *president*, there are 334 co-occurrences with *Assad*; in 38% (that is, 128) of the co-occurrences with *president*, the representation of *Bashar Al-Assad* is formalized as in *President Assad*, which is relatively higher than that of the American news sub-corpus. Figure

3.3 shows concordance lines examples of the KKW item *president* of the British news sub-corpus:

1	between a US Senate delegation and President <u>Bashar Assad</u> of Syria earlier
2	dictator Col Muammer Gaddafi and President <u>Bashar al-Assad</u> of Syria,
3	has been bolstered by what he and President <u>Obama</u> now refer to as
4	uding Colonel Gaddafi of Libya and President <u>Assad</u> of Syria were courted
5	ie Minister Binyamin Netanyahu and President <u>Obama</u> , has vowed to quit the
6	the meeting that Mr Netanyahu and President <u>Abbas</u> could make "the
7	from Mr Cameron, Mr Obama and President <u>Hollande</u> of France to force
8	Mr Cameron, President Obama and President <u>Sarkozy</u> wrote an
9	. "I think the American people and President <u>Obama</u> will understand that.
10	Russia, the party of Mr Putin and President <u>Dmitry Medvedev</u> . Mr
11	: King Juan Carlos of Spain and President <u>Assad</u> Palestinians rally in

Figure 3.3 Concordance examples of the KKW *president* from the British news sub-corpus

Moreover, in terms of the Global Variety news sub-corpus, it showed different pattern in terms of the most frequently functionalized news actors by *president*. For example, out of 863 instances of *president* of the GV news, there are 106 co-occurrences with *Morsi*; *Bashar* (86); *Obama* (78), *Assad* (76); *Mohamed* (69); *Barack* (65); *Mahmoud* (62); *Abbas* (61); *Hosni* (50) and *Mubarak* (49). The reason why *Morsi* is the most frequently functionalized by *president* could be accounted for due to the fact that most co-occurrences of *president* and *Morsi* (that is, 86 out of 346 instances of *president*) occurred in the Turkish *Zaman* (ZMNe) and *Hurriyet* (HRTe) news sub-corpora, which are known for their pro-Muslim Brotherhood stance. Only 20 co-occurrences of *president* and *Morsi* (out of 517 instances of *president*) were found in the Israeli *Jerusalem Post* (JPTe) newspaper. However,

for a wider picture, Figure 3.4 shows the top ten 3-word clusters of the most frequently functionalized news actor by *president* in the GV news sub-corpus:

N	Cluster	Freq.	Length
1	President Barack Obama	56	3
2	Us President Barack	47	3
3	President Hosni Mubarak	45	3
4	President Mohamed Morsi	45	3
5	Syrian President Bashar	42	3
6	President Bashar Assad	40	3
7	President Mahmoud Abbas	37	3
8	President Of The	37	3
9	President Bashar Al	30	3
10	President Shimon Peres	28	3

Figure 3.4 Top ten 3-word clusters of *president* in the GV news sub-corpus

As far as the Arabic sub-corpus is concerned, Table 3.4 shows that there is a similar pattern in terms of the significance of the elite functionalization in the GOVERNMENT AND SECURITY category. However, there is a difference between the English and Arabic sub-corpora in the sense that other elite actors, such as *activists* and *officials* are not relatively present in the Arabic sub-corpus as a whole. Another difference is the number of non-elite functionalization is much less in this category than that

of English, with the item *المواطن*<*the-citizen*> is only KKW in the Arabic editorials and opinion sub-corpus as Table 3.4 shows:

Table 3.4 Functionalization by KKWs of the GOVERNMENT AND SECURITY category with their % of each Arabic text type sub-corpus

Functionalization of news actors in the Arabic GOVERNMENT AND SECURITY category				
	KKW	Translation	News	Editorials and pinions
elite	الرئيس	<i>the-president</i>	2.65	4.61
	السفير	<i>the-ambassador</i>	1.17	
	الوزراء	<i>the-ministers</i>	1.21	
	رئيس	<i>prime/ president</i>	1.97	
	وزير	<i>minister</i>	1.08	
non-elite	المواطن	<i>the-citizen</i>		1.15
Column Total Avg. % of sub-corpus			1.81	2.26
Bold	Highest value in each sub-corpus			

As Table 3.4 also shows, there is a difference in terms of the significance of functionalization by the KKW item الرئيس<*the-president*> with the highest value, that is, 4.61% of the Arabic editorials and opinions sub-corpus. The most functionalized news actors by the KKW item الرئيس<*the-president*> in the Arabic editorials and opinions sub-corpus in terms of joint frequency are *مرسي*<*Morsi*> (304 out of 847 instances of وزير); *أوباما*<*Obama*> (218); and مبارك<*Mubarak*> (217). In the Arabic news sub-corpus, it is *بشار الأسد*<*Bashar Al-Assad*> (689 co-occurrences out of 5,563 instances of الرئيس); *مرسي*<*Morsi*> (504); and *أوباما*<*Obama*> (486). Like the English sub-corpus, many of the elite news actors

were also classified in terms of provenance, and in some cases, religion. Other collocates of الرئيس <the-president> in the Arabic corpus as a whole also include السابق <the-former>; المخلوع <the-ousted>; الجديد <the-new>; الحالي <the-current>; خطاب <speech>; and المنتخب <the-elected>, which are mostly shared with the English sub-corpora. Additionally, the Arabic news sub-corpus shows more significance in terms of the number of KKW items by which elite news actors are functionalized. However, there are two items, that is, the indefinite رئيس which has different functions and meanings depending on context, but generally occurs in phrases as رئيس الوزراء <the-prime minister>; or رئيس الحكومة <the-head of government>.

Similarly, the KKW item وزير <minister> is often post-modified by items referring to specific type of ministry or cabinet, as in وزير الدفاع <the-Defence Minister>; وزير الخارجية <the-Foreign Minister>. Figure 3.5 shows 15 of the most frequent 3-word clusters of وزير <minister> in the Arabic news sub-corpus: As Figure 3.5 shows, there is quite a number of elite news actors who are functionalized by the KKW item وزير <minister> in the Arabic news sub-corpus, also belong, as discussed earlier, to the *Western and European Group*, such as *America* (lines 10 and 15), *Britain* (line 9), *France* (line 4), *Germany* (line 11), and *Russia* (lines 1, 5 and 6). Additionally, the presence of the items وقال <also said> and قال <said> (Figure 3.4, lines 3 and 7) also adds to the eliteness of the sources of the news stories. According to Bell (1991, p. 192), ‘among politicians it is ministers who have overwhelming

access to the media', and hence the quality of the news sources affects its newsworthiness. Therefore, *who says?* is of one of the fundamental questions of news work (Bell, 1991, p. 190), . In this context, elite and powerful social/news actors become a favoured source for news media. The more elite they are the more newsworthy is their story (Bell, 1991). The elite are the most quoted in the news, and they are usually political figures of high profile, as Figure 3.5 (lines 3 and 7) indicates:

Examples of the top frequent 3-word clusters of وزير <minister> in the Arabic news sub-corpus

N	Cluster	Translation	Freq	Length
1	وزير الخارجية الروسي	<i>the Russian Foreign Minister</i>	49	3
2	نائب وزير الخارجية	<i>Deputy Foreign Minister</i>	37	3
3	وقال وزير الخارجية	<i>the Foreign Minister also said</i>	36	3
4	وزير الخارجية الفرنسي	<i>the French Foreign Minister</i>	35	3
5	الروسي سيرغي لافروف	<i>the Russian Sergey Lavrov</i>	31	3
6	الخارجية الروسي سيرغي	<i>the Russian Foreign () Sergey</i>	31	3
7	قال وزير الخارجية	<i>the Foreign Minister said</i>	29	3
8	وزير الخارجية المصري	<i>the Egyptian Foreign Minister</i>	28	3
9	وزير الخارجية البريطاني	<i>the British Foreign Minister</i>	27	3
10	وزير الخارجية الأمريكي	<i>the American Foreign Minister</i>	23	3
11	وزير الخارجية الألماني	<i>the German Foreign Minister</i>	21	3
12	وزير الخارجية الإيراني	<i>the Iranian Foreign Minister</i>	19	3
13	وزير الخارجية التركي	<i>the Turkish Foreign Minister</i>	19	3
14	وزير الخارجية السوري	<i>the Syrian Foreign Minister</i>	18	3
15	وزير الدفاع الأمريكي	<i>the American Defence Minister</i>	18	3

concordance collocates plot patterns clusters timeline filenames source text notes

201 entries Row 16 T S ... الخارجية البريطاني

Figure 3.5 Examples of the top frequent 3-word clusters of وزير of the Arabic news sub-corpus

Looking at the KKW items *said* in the American, British and GV, and وقال in the Arabic news sub-corpora, it is not surprising that we see elite and prominent news actors, as identified in the NAME-PN category (section 3.4, Figure 3.2), such as

Obama, Assad, Netanyahu and *(Hillary) Clinton*, who are amongst the most quoted as news sources in both the English as well as the Arabic sub-corpora. For example, of the functionalized and nominated news actors who are mostly associated with *said* (16,579 instances) in terms of joint frequency, especially to the immediate left (L1) in the American news sub-corpus are:

officials (404), *official* (321), *Obama* (176); *Clinton* (125); *analysts* (114), *activists* (81); *spokesman* (79); *Netanyahu* (59) and *minister* (57).

The KKW items *civilians*, *citizens* and *refugees* are amongst the least frequent in terms of their co-occurrence with *said*.

Concerning the functionalized news actors of the remaining semantic categories, these include the PROTEST VOCABULARY, SOCIAL ACTIONS, STATES AND PROCESSES, as well as LANGUAGE AND COMMUNICATION. As Table 3.5 shows, there are three (that is, *protesters*, *rebels*, and *demonstrators*) out of a total of 12 KKW items in the PROTEST VOCABULARY by which news actors are functionalized and assimilated (that is, in the form of plural) in the same time. The most significantly used in functionalizing the news actors of this category across the six English sub-corpora is the KKW item *protesters*, with the highest value, that is 17.54% of the American news sub-corpus. However, it is the KKW item *rebels* that was the most revealing in terms of its associated topics/events. According to Bell (1991, p. 195, original italics), ‘many commentators have pointed out the significance of how news actors

are labelled [...]. It is particularly evident in situations of conflicts: one side's *terrorist* is the other side's *freedom fighter*'.

Table 3.5 KKW functionalization items of the PROTEST VOCABULARY category with their % of each of the English text type sub-corpora

PROTEST VOCABULARY of the English text type sub-corpora						
KKW	News			Editorials and opinions		
	AM	BR	GV	AM	BR	GV
protesters	17.54	10.58	10.78	6.01	5.27	3.89
rebels	5.56	3.88	2.70	2.81	5.27	1.59
demonstrators	4.39	2.34	1.89			1.24
Column Total Avg. % of sub-corpus	9.16	5.60	5.12	4.41	5.27	2.24
Bold	Highest value in each sub-corpus					

This is one of the reasons why the KKW item *fighters* (see Table 3.3) has been categorized under the GOVERNMENT AND SECURITY category. Another reason is that *fighters* is quite often used in the sense of military planes as in *jet fighters*. Nevertheless, the contexts are so much interrelated, and one topic/actor leads to or reveals the other. On the one hand, *protesters* and *demonstrators* share quite a number of collocates across the main English text type sub-corpora, such as

police, government, anti, forces, streets, square, security, thousands, killed, killing, crackdown, violence, hundreds, gas, fire, pro, democracy, peaceful, demands, clashes, spring, Arab, street, took, unarmed, Syrian, Tahrir, Shia, Sunni, violent and angry.

With such associations, it becomes clear that the negative aspects of the items *protesters* and *demonstrators* is what make them get into the news. On the other hand, the contexts of the KKW item *rebels*, which is relatively present/included in all sub-corpora (Table 3.5), are different, though quite a number of its associated topics are similar in terms of negativity. Analysis showed, as will also be discussed in the following analysis chapters, that its most associated topics across the English sub-corpora, especially in the American and the British, are

Syrian, Libyan, support, weapons, government, arming, forces, Syria, fighting, arm, military, Muammar, Gaddafi, Assad, Obama, arms, help, Islamist, control, aid, president, NATO, Damascus, army, killed, armed, Houthi, Benghazi, regime, anti, Western, backing, Qatar, Saudi, Turkish, Erdogan, spring, city, overthrow, ground, US, council, Shia, Sunni, troops, and prison.

Figure 3.6 shows concordance lines examples of *rebels* taken from the British editorials and opinion sub-corpus and sorted by L3:

1 but it has not quite managed to paint the rebels as Western puppets and dupes.
 2 Leon Panetta wanted to supply Syria's rebels with arms, but President Obama,
 3 If Mr Cameron decides to supply Syrian rebels with British arms, he will set
 4 now the US intends to arm the rebels, and Saudi Arabia, according to
 5 and offer military aid to the Syrian rebels. A realist wing quietly laments
 6 for military intervention to protect the rebels against Colonel Muammar
 7 but to send in Special Forces to aid the rebels, and who knows how many
 8 a few light weapons to the Syrian rebels will be more effective? The
 9 pushing to supply weapons to the Syrian rebels - arguing that this is necessary
 10 villages play host to Syrian armed rebels who slip across the border at
 11 Or will Obama decide to back the rebels with enough firepower to
 12 few countries believed to be supplying rebels with money and weapons. It sees

Figure 3.6 Concordance examples of *rebels* of the British editorials and opinion sub-corpus

As indicated in Figure 3.6, there was divisions in opinions in terms of arming the Syrian (e.g. lines 2, 3 and 9) and the Libyan rebels. While the West and its allies in the region supported the Libyan rebels, they were hesitant about arming the rebels in Syria and supplying weapons to the opposition there. In terms of the Arabic sub-corpus, it showed a similar pattern to that of English in terms of the resulting functionalization KKW items (3 items), and their equivalents. The items المتظاهرين <the-demonstrators> has the highest value, that is, 2.52% of the Arabic news sub-corpus, whereas the item الثوار <the-rebels> is only relatively present/included, in Partington's (2014, 129) terms, with a value of 1.06% of the Arabic editorials and opinion sub-corpus. The two KKW items, المتظاهرين and المحتجين are relatively absent in the Arabic editorials and opinion sub-corpus, as Table 3.6 shows:

Table 3.6 KKW Functionalization items of the PROTEST VOCABULARY category with their % of each of the Arabic text type sub-corpora

PROTEST VOCABULARY			
KKW	Translation	MSA	
		News	Editorials and opinions
المتظاهرين	<i>the-demonstrators</i>	2.52	
المحتجين	<i>the-protesters</i>	1.24	
الثوار	<i>the-rebels</i>		1.06
Column Total Avg. % of sub-corpus		1.38	1.06
Bold	Highest value in each sub-corpus		

Similar to that pattern of the American and the British sub-corpora, the KKW functionalization item الثوار <the-rebels> also revealed aspects of power relation and the elite interests. For example, تسليح الثوار <arming of-the-rebels> and weapon deals with Qatar and Saudi Arabia, as well as the stance of the British government regarding the arming of the rebels in Syria. However, despite similarities, there was a difference between the Arabic editorials and opinion sub-corpus (the Arabic news was also examined) with that of English (the American, the British and the GV) sub-corpora in terms of the classification strategies. *Rebels* were mostly classified in terms of *provenance* and *religion* in most of the English sub-corpora, especially the American and the British. For example, out of 497 instances in the British news, *rebels* co-occurred with *Shia* (7); *Islamist* (7) and *Sunni* (4) of which all were to the immediate left (L1) of *rebels*. On the other hand, in the Arabic editorials and opinion sub-corpus, the KKW item الثوار <the-rebels> was mostly classified in terms of *provenance* and *age*. For example, in 508 instances of الثوار <the-rebels>, there were (14) co-occurrences with العرب <the-Arab> of which (11) were to the immediate right (R1) of the term; السوريين <the-Syrian> (9); and الشباب <the-young.PL.MASC> (7). Similarly, there was no reference to الثوار in terms of religion, but rather in terms of classification as in الثوار العرب <the-Arab rebels>; and functionalization as in قادة الثوار <leaders of the-rebels or the rebel leaders> in the Arabic news sub-corpus.

The last two semantic categories of which *functionalization* is realized are SOCIAL ACTIONS, STATES AND PROCESSES; and LANGUAGE AND COMMUNICATION. In the

SOCIAL ACTIONS, STATES AND PROCESSES category, there are four main KKW items (that is, *king*, *leaders*, *sheikh* and *dictators*) that are distributed across the English text types sub-corpora, as shown in Table 3.7, along with their percentage (%) of each sub-corpus:

Table 3.7 KKW Functionalization items of the SOCIAL, STATES, AND PROCESS category with their % of each English text type sub-corpus

Roles in the English SOCIAL ACTIONS, STATES AND PROCESSES category						
KKW	News			Editorials and opinions		
	AM	BR	GV	AM	BR	GV
king	5.01	2.99	2.70	2.40	2.20	
leaders	3.76	2.18	1.35	2.40	1.10	2.83
sheikh	1.25	3.39	2.16		1.98	
dictators				2.20	1.76	1.06
Column Total Avg. % of sub-corpus	3.34	2.85	2.07	2.34	1.76	1.95
Bold	Highest value in each sub-corpus					

As Table 3.7 shows, functionalization by the KKW item *king* is relatively the highest in terms of the American and GV news as well as the American and the British editorials and opinions sub-corpora. However, KKW items, such as *king* and *sheikh* are amongst the items by which *nomination* is realized, especially when titulated in the form of *honorification* (van Leeuwen, 2008, p. 41). In this context, it can be said that ‘nominations of this kind in fact blur the dividing line between

nomination and categorization' (van Leeuwen, 2008, p. 41). Nevertheless, the fact that there are more than one news/social actor occupying such roles of *king* or *sheikh* in the context of the Arab Spring, they were also included as functionalization items. For example, the KKW item *king* has the highest value, that is 5.01% of the American news sub-corpus (Table 3.7). The most frequent 3-word clusters of *king* (822 instances) in the American news sub-corpus, which contain names of news actors are

King Abdullah II (78) of Jordan, and as *Jordan's King Abdullah* (23);

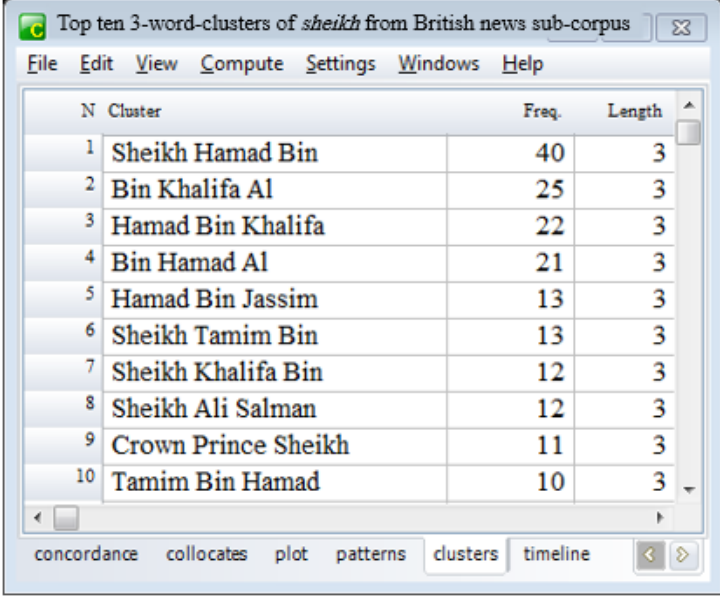
King Mohammed VI (36) of Morocco;

King Hamad Bin Isa (23) of Bahrain, and as *Bahrain's King Hamad* (5).

In terms of the British news sub-corpus, the KKW *sheikh*, with value of 3.39% of the BR news sub-corpus, is the most relatively included functionalization item by which quite a number of news actors were nominated/titulated and functionalized. Of the most top collocates of *sheikh* (in 405 instances in the British news sub-corpus) in terms of total joint frequency, and that referring to news actors (rather than to a place, such as *Sharm* (39) *el* (39) in *Sharm el Sheik* of Egypt) are:

Al (174); *Bin* (107); *Hamad* (105); *Khalifa* (77); *Thani* (51); *Tamim* (48); *Salman* (34); *minister* (28); *Emir* (28); *said* (25); *Ahmed* (23); *Sabah* (21); *leader* (21); *Qatar* (18); *Jassim* (17); *prime* (17); *Mohammed* (16); *Nasrallah* (16); *crown* (14); *prince* (13); *Hassan* (12); *Zayed* (11); and *Fadlallah* (8).

Figure 3.7 shows the top ten most frequent 3-word clusters of *sheikh* referring to news actors in the British news sub-corpus:



N	Cluster	Freq.	Length
1	Sheikh Hamad Bin	40	3
2	Bin Khalifa Al	25	3
3	Hamad Bin Khalifa	22	3
4	Bin Hamad Al	21	3
5	Hamad Bin Jassim	13	3
6	Sheikh Tamim Bin	13	3
7	Sheikh Khalifa Bin	12	3
8	Sheikh Ali Salman	12	3
9	Crown Prince Sheikh	11	3
10	Tamim Bin Hamad	10	3

Figure 3.7 Top ten 3-word clusters of *sheikh* referring to news actors in the British news

Considering the KKW items *leaders* and *dictators*, these items, as their forms (that is, plural) suggest, were used as functionalization and assimilation items in different contexts. In terms of *dictators*, it is only shared by the three English editorials and opinions sub-corpora. Although the KKW item *dictators* does not have the highest value (when compared to other KKW items) in any of the English editorials and opinions sub-corpora, it can be said that *dictators* is relatively most included with value of 2.20% of the American editorials and opinion sub-corpus when compared with the British and GV sub-corpora. Table 3.8 shows all lexical collocates of

dictators (joint frequency ≥ 5 and MI score ≥ 3) across the American, the British as well as the GV editorials and opinions sub-corpora:

Table 3.8 Lexical collocates of *dictators* of the English editorials and opinions sub-corpora

All lexical collocates of *dictators* with joint frequency ≥ 5 and MI score ≥ 3 in the American, British and GV editorials and opinions sub-corpora

American	Freq.	MI	British	Freq.	MI	Global Variety	Freq.	MI
dictators	160	12.26	dictators	150	12.52	dictators	91	13.13
Arab	26	5.72	Arab	17	5.01	Arab	16	5.71
countries	7	5.59	toppled	6	9.11	toppled	9	11.30
world	6	4.24	world's	5	7.51	Fall	6	8.40
Egypt	6	4.60	<u>Western</u>	5	4.99	<u>pro</u>	5	7.76
region	5	5.14	Tunisia	5	5.80	<u>Western</u>	5	6.28
<u>support</u>	5	5.18	swept	5	8.69	Mubarak	5	7.61
toppled	5	8.50	region	5	5.27	Four	5	7.97
			Fall	5	6.72			
			<u>backed</u>	5	7.67			

Bold: shared collocates

As shown in Table 3.8, there are two collocates (in bold), that is, *Arab* and *toppled*, that are shared between the three (American, British and GV editorials and opinions) sub-corpora, reflecting the common theme of *toppling* the (Arab) dictators (in the Arab world). The collocate items, that is, *support*, *backed*, *Western* and *pro* are also used in a wider context (see Figure 3.8) in which *dictators* are also referred to as being supported and backed by the US, the Western and other Arab

countries (e.g. Saudi Arabia) in the region. Figure 3.8 shows how *dictators* are represented in the contexts of the collocate items, *support* (lines 1-5 of the American); the item *backed* (lines 6-10 of the British) and *pro* (lines 11-16 of the GV) editorials and opinions sub-corpora:

1 power to support the removal of dictators and the beginning of
 2 the old U.S. policy -- support for dictators in the name of preserving
 3 the old U.S. policy - support for dictators in the name of preserving
 4 of uncritical support for friendly dictators who are helpful on matters
 5 America's decades-long support for dictators who accommodated its
 6 These frequently western-backed dictators had destroyed mainstream
 7 rabs are rising up against US-backed dictators in the region, they have
 8 evidence. True, after western-backed dictators swept the political spectrum
 9 ey backed the Tunisian and Egyptian dictators until the last minute. They
 10 entury of tyranny by Western-backed dictators would take a while to get o
 11 * A world in which pro-Western dictators are either out of office or
 12 n economic aid, to pro-American dictators, as was the case with Egypt
 13 US shipped tanks to pro-Western dictators in the region, today it can
 14 istration, substituted authoritarian dictators - some pro-Western - with
 15 ationships with pro-Western Arab dictators to dealing with Arab public
 16 - if until Tahrir the pro-Western dictators were the allies of the United

Figure 3.8 Instances of *dictators* with context items *support*, *backed* and *pro* of the three main English editorials and opinions sub-corpora

In terms of the KKW item *leaders* (see. Table 3.7), it is relatively the most included functionalization item as it is shared between all the English six type sub-corpora, with its highest value, that is, 3.76% of the American news sub-corpus. Analysis showed that there are a number of shared collocates, especially to the immediate left (L1) of *leaders* across the English text types sub-corpora. Most of these (lexical) collocates reflect patterns of classification in terms of *provenance* and *religion*. In

the American news sub-corpus, for example, the top twenty-five lexical collocates (with joint frequency ≥ 5 and MI ≥ 3.00) to immediate left (L1) of *leaders* (1,621 instances) include eight (in bold) and two (underlined) items indicating geographical and religious classifications respectively:

opposition (85); ***Arab*** (66); ***world*** (61); *Political* (56); ***Palestinian*** (35); *Brotherhood* (32); ***Israeli*** (32); *tribal* (28); *religious* (28); *news* (26); *military* (25); ***European*** (18); *protest* (16); ***Iranian*** (16); *Hamas* (16); *autocratic* (16); *authoritarian* (16); ***Western*** (15); *rebel* (15) *Islamist* (15); ***Iraqi*** (12); *elected* (11); *government* (10) and *party* (9).

Similarly, in the British news sub-corpus, of the top twenty-five immediate left (L1) collocates of *leaders* (987 instances), there are seven (in bold) and four items (underlined) indicating geographical as well as religious classifications of *leaders* respectively:

Arab (62); ***world*** (38); *opposition* (37); *political* (31); ***Palestinian*** (31); ***Western*** (31); *religious* (21); ***Israeli*** (16); *Hamas* (14); *new* (12); *military* (12); *tribal* (11); ***European*** (10); *rebel* (9); *Islamist* (9); *Muslim* (8); *authoritarian* (8); *settler* (7); *Christian* (7); *union* (6); *party* (6); ***international*** (6); *EU* (6); *community* (6) and ***China's*** (6);

As indicated above, there are six geographically (that is, *Arab*, *world*, *Palestinian*, *Western*, *Israeli* and *European*), and two religiously (that is, *religious* and *Islamist*) related L1 collocates that are shared between the American as well as the British

news sub-corpora. There are at least two important points that can be made from the geographical pattern of *leaders* within the Arab Spring context. These can be illustrated by the concordance examples of the item *Western leaders* of the American and British news sub-corpora as shown in Figure 3.9:

1 the transition to a renewed democratic order. *Western leaders* have made it clear that this needs
 2 U.S. President Barack Obama and other *Western leaders* will continue to push for peaceful
 3 Libya in March, President Obama and other *Western leaders* assured their war-weary publics th
 4 d Cameron, the UK prime minister, and other *western leaders* stressed at an international confere
 5 Soviet states to reform, Mr. Obama and other *Western leaders* are racing to invent something. "T
 6 assets, leading to a breakdown of services. *Western leaders* will take up the issue of releasing
 7 are numbered as, following the Arab Spring, *western leaders* apparently oppose heads of state w
 8 little to alter the horror of the events there. *Western leaders* believed the Syrian government's
 9 threatens the security of Israel and the West. *Western leaders* who backed the uprisings against
 10 vilian team on the ground within weeks, while *Western leaders* agreed to push for a UN Security
 11 Spring Benefits Israel BAGHDAD -- While *Western leaders* including President Obama called
 12 in Arab nations won widespread acclaim, with *Western leaders* welcoming the demise of Hosni

Figure 3.9 Concordance examples of *Western leaders* of the American and the British news sub-corpora

Given the great value which is placed on individuality in many spheres of the Western world, categories, such as *assimilation* and *individualization* are of primary significance in CDA (van Leeuwen, 2008, p. 37). The first point that can therefore be made from the concordance examples in Figure 3.9 is that *leaders* does not only represent aspects of linguistic functionalization and assimilation, but is also believed to function as a signal for agreement (e.g. line 10) and/or unanimity of opinion; or other forms of actions among the news actors, especially when pre-modified by geographical items, such as, *Western*, *European* and *world*, as indicated by the examples in Figure 3.9 as well as the L1 collocates (of *leaders*)

above. The second point, however, shows that there is another pattern of individualizing news actors, especially who come from elite countries, and are of high governmental and presidential ranks, especially the *United States* and *Britain*, as shown in Figure 3.9 (lines 2, 3, 4, 5 and 11). This suggest that geographical items relating to the US and the UK have different political and governmental associations, as will be seen in sub-sections 3.5.3 and 3.5.4.

As also indicated, quite a number of collocate items of *leaders* have negative meanings and connotations (e.g. *autocratic*, *authoritarian*, *rebel*, *protest*, *tribal* and *military*). Some of these collocates (e.g. *tribal*) are used in certain contexts, and only applied to certain groups and/or nationalities rather than others (Krishnamurthy, 1996). For example, in the Arab Spring context, analysis showed that the most geographically related collocates of *tribal* (501 instances) in terms of joint frequency in the English sub-corpus as whole (there are 184 and 130 instances in the American and British news sub-corpora respectively) are *areas* (41); *Pakistan's* (18); *Libya* (15); *Yemen* (13); *regional* (13); *Yemeni* (11); *Jordan* (7); *Iraq* (6); *Afghanistan* (6). *Tribal* also co-occurs with quite a number of military and warfare related items, such as *military*, *warfare*, *war*, *militia*, *fighting*, and *fights*, and is also associated with religious classification items, such as *religious*, *sectarian*, *Sunni*, *Islamists* and *Islamist*. The most frequent lexical collocate of *tribal* in the English sub-corpus as a whole is *leaders* (66), with 57 co-occurrences are to the immediate right (R1) of *tribal*. For example, Krishnamurthy (1996, p. 134) looked at the collocational profiles of three near-synonymous terms, that is,

ethnic, *racial* and *tribal*, in a 121-million-word sample corpus from the Bank of English, and found out that *tribal* clearly has pejorative connotations, co-occurring 6 times with the item *primitive*, but never with *ethnic* or *racial* despite their being more significantly frequent than *tribal* in the corpus. In terms of the geographical associations, Krishnamurthy (1996, p. 143) also noted that only *tribal* had co-occurrences with items such as, *Sahara*, *Krahn*, *Chadian Africa*; *Navajo*, *Mohawk* and *American Indians*; and *Pakistan*, *Kabul* and *Afghanistan*. He concluded that '[t]he English word 'tribal' clearly has pejorative connotations, and if we continue to use it, and apply it only to certain groups of human beings, we are merely recycling the prejudices that the English-speaking culture has developed with regard to those groups' (Krishnamurthy, 1996, p. 146-7). In terms of the Arabic sub-corpus, there is also a similar pattern to that of English in terms of KKW items equivalence or semi-equivalence. As Table 3.9 shows, there are two KKW items,

Table 3.9 Functionalization KKWs of social actions, states and processes category with their % of each Arabic text type sub-corpus

Roles of the Arabic SOCIAL ACTIONS, STATES AND PROCESSES category			
KKW	Translation	MSA	
		News	Editorials and opinions
الملك	<i>the-king</i>	1.28	
قائد	<i>leader</i>	1.10	
الطغاة	<i>the-tyrant</i>		1.06
Column Total Avg. % of sub-corpus		1.19	1.06
Bold	Highest value in each sub-corpus		

that is, الملك <the-king>; and قائد <leader> with 1.28% and 1.10% of the Arabic news sub-corpus respectively. On the other hand, the item الطغاة <the-tyrants> is only KKW, with 1.06% of the Arabic editorials and opinions sub-corpus: The most included news actors functionalized by الملك <the-king> in the Arabic news in terms of the most frequent 3-word-clusters are

الملك عبد الله الثاني <King Abdullah II> (283) of Jordan;

الملك محمد السادس <King Mohamed VI> (110) of Morocco;

الملك عبد الله بن عبد العزيز <King Abdullah Bin Abdul-Aziz> (21) of Saudi Arabia.

Other collocates of الملك <the-king> (1,661 instances in the Arabic news) with ≥ 5 of co-occurrences are: خطاب <speech> (29); الحكومة <the-government> (28); and الإصلاح <the-reform>. The KKW item قائد <leader>, which is only relatively present in the Arabic news, is mostly used to refer to قائد الثورة الإسلامية <the Supreme Leader of the Islamic Revolution> of Iran.

In terms of the KKW item, الطغاة <the-tyrants>, it is only relatively present with 1.06 of the Arabic editorials and opinions sub-corpus. Similar to that pattern of *dictators* of the English sub-corpus, the news actors who were functionalized by الطغاة <the-tyrants>; were also classified by provenance as in الطغاة العرب <the-Arab tyrants> as indicated by Figure 3.10 (line 4), which shows the most frequent lexical collocates of الطغاة (140 instances) of the Arabic editorials and opinions sub-corpus. Analysis

also showed that الطغاة <the-tyrants> is also used to refer to الحكام <the-rulers> in the region, as Figure 3.10 (line 5) also shows:

N	Word	Set	Relation	Total
1	الطغاة	the-tyrants	14.22	142
2	الشعوب	the-peoples	7.06	11
3	العربي	the-Arab (SING.MASC.ADJ)	4.06	9
4	العرب	the-Arab (PL.MASC.ADJ)	17.20	8
5	الحكام	the-rulers	8.98	8
6	الثورة	the-revolution (FEM.N)	4.30	6
7	بسقوط	with-(a)fall	10.26	5
8	العربية	the-Arab (FEM.ADJ)	16.53	5

Figure 3.10 Lexical collocates of الطغاة <the-tyrants> of the Arabic editorials and opinions sub-corpus

The last category, in terms of functionalization, is the LANGUAGE AND COMMUNICATION category. *journalists* that is most relatively present in the English sub-corpus, with the highest value 2.43% of the GV news and shared only by four sub-corpora, that is the British and the GV news, and both the American news as well as editorials and opinion sub-corpora. *Journalists* is mostly classified by *foreign*, *international* and *Western*. Examination of its contexts showed that *journalists* who were covering the Arab Spring events, *suffered killing*, *arrest*, *injury*, and was *barred* from entering *Syria* to cover the events there. No Arabic

KKW items by which functionalization is realized are relatively present of significant in this is category.

3.5.2 Classification

In the case of classification, it is one aspect of categorization by which news actors of the Arab Spring narrative are mainly classified and identified by means of *governance, religion, gender*; and to a lesser extent *age*. The main KKW-semantic categories that have the potential of classifying news actors are GEOGRAPHY (e.g. *Arab/s, Syrian/s, Israeli/s* and *American/s*). In terms of religion, KKW items (e.g. *Muslim, Islamist, Christians* and *Jewish*) of the FAITH AND RELIGION category are also frequently used in the classification/ identification strategies employed in the representation of the different news actors of the Arab Spring. Analysis also showed that religious classification is significantly higher in the English (that is, the American, the British and the GV) sub-corpora than that of the Arabic (see. Chapter 4, sub-section 4.2). However, the English as well as the Arabic sub-corpora showed similar pattern of religious classification in terms of text type. That is, both the English as well as the Arabic editorials and opinions showed higher levels of religious classification than that of their news sub-corpora. In terms of the English sub-corpus, there are fourteen out of a total of sixteen KKW items of the FAITH AND RELIGION category of which five (in bold) are shared between the main six English text types sub-corpora:

Muslim, Islamist, Shia, Islamic, Sunni, Islamists, Jewish, Muslims, sectarian, secular, Jews, Shi'ite, Christians and religious.

Some of these KKW items (that is, *Islamists, Muslims, Jews* and *Christians*) can be said that they are more personalizing of news actors than others, which also reflect other aspect of assimilation (see 3.5.3). Additionally, the two KKW items *Islamic* and *Islamist* are used interchangeably when referring to different news actors (people and groups) although the item *Islamist* is commonly meant to refer to Muslim individuals and groups/ organizations advocating Islamic political rule (Baker et al., 2013b, p. 84; Geller, 2011, online; Abaza, 2010). *Islamic* on the other hand generally appears to be part of names of countries (e.g. the *Islamic Republic of Iran, Islamic State of Iraq*); groups (e.g. the *Islamic Jihad*, the military wing of Hamas, and the *Libyan Islamic Fighting Group*); parties (e.g. the Muslim Brotherhood affiliated party the *Islamic Action Front* of Jordan); and/or organizations (e.g. *Organisation of Islamic Co-operation*), titles/names of news actors (e.g. the *Supreme Leader of Islamic Revolution* of Iran), or refer to historical events (e.g. *1979 Islamic Revolution* of Iran). Another common pattern of *Islamic* is that it is also used as a modifier of different items denoting different news actors (that is, individuals and groups/organizations) belonging to different semantic categories. For example, items referring to news actors of the GEOGRAPHY category (e.g. *world, region* and *countries*); GOVERNMENT AND SECURITY (e.g. *regime, governments, nation/s, officials, activists, extremist/s* and *militant/s*); SOCIAL ACTIONS, STATES AND PROCESSES (e.g. *parties, movement/s, group/s*, and *society*);

FAITH AND RELIGION (e.g. *fundamentalist/s*). Analysis also showed that both KKW items (that is, *Islamist* and *Islamic*) share several collocates denoting different news actors of which quite a number of these collocates have negative meanings and connotations, and hence can be said that there is no real distinction between *Islamic* and *Islamist* across the English sub-corpora as a whole. For example, both *Islamic* and *Islamist* share collocates (with joint frequency ≥ 5 and MI score ≥ 3.00) belonging to different semantic categories, such as NAME-PN (e.g. *Hamas*, *Al-Qaeda*, *Ennahda*, *Hezbollah*, *Brotherhood* and *Morsi*); GOVERNMENT AND SECURITY (e.g. *government/s*, *regime*, *militant/s*, *extremist/s*, *terrorist/s*, *activists*); EMOTIONS (e.g. *violence* and *threat*); PROTEST VOCABULARY (e.g. *protesters*) and SOCIAL ACTIONS STATES AND PROCESSES (e.g. *parties*, *movement/s*, *group/s*, *members*, *leader/s*, *supporters*, *organization/s*, *oppositions* and *forces*). Figure 3.11

Examples of the top 3-word clusters of *Islamist* of the American news sub-corpus

N	Cluster	Freq.	Length
1	Moderate Islamist Party	46	3
2	A Moderate Islamist	34	3
3	The Moderate Islamist	26	3
4	The Muslim Brotherhood	26	3
5	The Islamist Group	24	3
6	Islamist Party Ennahda	24	3
7	The Islamist Party	17	3
8	Islamist Party That	16	3
9	The Islamist Movement	12	3
10	Islamist President Mohamed	12	3
11	Islamist Led Government	12	3

concordance collocates plot patterns clusters

89 entries Row 1 T S

Figure 3.11 Examples of the most frequent 3-word clusters of *Islamist* of the American news sub-corpus

shows examples of the top 3-word clusters of *Islamist* of the American news sub-corpus.

Similarly, in the Arabic sub-corpus, there are seven out of a total of ten KKW items of which religious classification is realized. These are

الإسلامية <***the-Islamic***> (FEM.ADJ) ;
 الإسلاميين <***the-Islamists***> (MASC.PL.GEN), but also the-Islamic.PL.GEN ;
 الطائفية <*the-sectarian/ism*> (FEM.ADJ / FEM.N) ;
 المسلمين <***the-Muslim/s***> (MASC.PL.GEN) ;
 الطائفي <the-sectarian.MASC.ADJ -*the sectarian*>;
 الإسلامي <***the-Islamic***> (MASC.ADJ) ;
 العلمانية <*the-secular/ist/ism*> (FEM.ADJ/N) ;
 السلفية <*the-Salafist*> (FEM.ADJ) or <*the-Salafism*> (FEM.N) ; and
 الإسلاميون <***the-Islamists***> (MASC.PL.NOM) but also the-Islamic.PL.NOM

There are four (shown in bold above) KKW items that are shared between the two main Arabic sub-corpora, and of which the KKW item الإسلامية <***the-Islamic***> is most significant with highest value of 2.96% of the Arabic news sub-corpus. On the other hand, the KKW item الإسلاميين <***the-Islamists***> is the second significant religious classification item in the Arabic sub-corpus, with its highest value 2.57% of the Arabic editorials and opinions sub-corpus. In a similar pattern to that of the English sub-corpora, both adjective forms, the feminine الإسلامية <al-'islāmiyya - *the-Islamic*>; as well as the masculine الإسلامي <al-'islāmī - *the-Islamic*> are often used to refer to different news actors, including individuals and groups/organizations.

Figure 3.12 shows some of the most frequent 3-word clusters of the feminine form الإسلامية<the-Islamic> (2,334 instances) of the Arabic news sub-corpus:

N	Cluster	Set	Freq	Length
1	المقاومة الإسلامية حماس	<i>the-Islamic Resistance Hamas</i>	97	3
2	حركة النهضة الإسلامية	<i>the-Islamic Ennahda Movement</i>	71	3
3	حركة المقاومة الإسلامية	<i>the-Islamic Resistance Movement</i>	67	3
4	الإسلامية في إيران	<i>the-Islamic (...) in Iran</i>	51	3
5	الجمهورية الإسلامية الإيرانية	<i>the Islamic Republic of Iran</i>	47	3
6	الصحوة الإسلامية في	<i>the-Islamic awakening in</i>	29	3
7	الثورة الإسلامية في	<i>the-Islamic revolution in</i>	27	3
8	الحركات الإسلامية في	<i>the-Islamic movements in</i>	23	3
9	قائد الثورة الإسلامية	<i>Leader of the-Islamic Revolution</i>	22	3
10	لحركة المقاومة الإسلامية	<i>of the-Islamic Resistance Movement</i>	21	3
11	تطبيق الشريعة الإسلامية	<i>application of the-Islamic Sharia law</i>	20	3

Figure 3.12 Examples of the top frequent 3-word clusters of الإسلامية of the Arabic news sub-corpus

However, and as Vessey (2013, p. 14) pointed out, the fact that ‘English does not, on the whole, make distinctions in gender’, comparisons with gendered Arabic items can be problematic, especially that different word forms, can have different collocates and/or associations (Stubbs, 1996). Thus, when comparing the two Arabic forms, it can be said there is a difference between the two forms in terms of how news actors (that is, people and groups), who are identified as *Islamic/ Islamist* (feminine and/or masculine) are referred to or evaluated. For example, the most

frequent (evaluative) collocates to the immediate right (R1) of the feminine form الإسلامية <al-'islāmiyya> (2,334 instances) of the Arabic news sub-corpus include

المتطرفة <the-extremist> (11); المتشددة <the-radical> (11); المعتدلة <the-moderate> (9); الراديكالية <the-radical> (8); الحاكمة <the-ruling> (8); المسلحة <the-armed> (7); and المحظورة <the-banned> (5).

On the other hand, the most frequent collocates to the immediate right (R1) of the masculine form الإسلامي <al-'islāmī> (1,197 instances) of the same sub-corpus include المعتدل <the-moderate> (17), which mostly refer to the *Islamic/Islamist Ennahda Party* of Tunisia. Some of the immediate left (L1) collocates of الإسلامي <al-'islāmī> also include العالم <the-world> (153); and التطرف <the-extremism> (7). The remaining KKW items, such as

المسلمين <the-Muslims>, الإسلاميين <the-Islamists>(GEN); and الإسلاميون <the-Islamists>(NOM) are more personalizing items by which assimilation can also be realized. Assimilation was also frequently used in both the English as well as the Arabic sub-corpora.

3.5.3 Assimilation

According to van Leeuwen (2008, p. 37), social/news actors can be referred to as individuals or as groups. Thus, *individualization* is realized by singularity (e.g. *president Obama*), whereas *assimilation* is realized by plurality (e.g. *Arabs, activists, officials, Muslims, leaders, dictators, women, journalists and protesters*),

or by a mass noun or a noun denoting a group of people (e.g. *group, party, nation* and *government*). *Collectivization* is hence one form of *assimilation*, which can be used in both representation strategies of *personalization* as well as *impersonalization* (see Figure 3.1). As has been discussed so far, assimilation is quite frequent in both the English as well as the Arabic sub-corpora. In terms of collectivization, there are quite a number of KKW items of different semantic categories of both English and Arabic that have the potential to collectivize news actors. Analysis showed that many of these items refer to elite news actors (e.g. countries, parties, and/or organizations). The semantic categories of both English and Arabic, for example, include:

GEOGRAPHY: Examples of collectivization of news actors of the English sub-corpora include noun forms of geographical items, such as *world, country, countries, Egypt, America and Europe*. Other examples may include phrases such as, *Arab world, Western world, international (community), European nations/governments*.

Similarly, in Arabic, collectivization can be realized through the KKW items and their collocates such as *العالم العربي* <*the-Arab world*>; *الأمة العربية* <*the-Arab nation*>; *الغرب* <*the-West*>; and many geographical items in the noun form, such as *سوريا* <*Syria*>, *الأردن* <*Jordan*>; *إسرائيل* <*Israel*>; and *واشنطن* <*Washington*>.

GOVERNMENT AND SECURITY: examples of the English KKW items of which a number (in bold) are shared between the six main English text type sub-corpora,

include *regime*, *military*, *government*, *police*, *forces*, *parliament*, *regimes*, *army*, *embassy*, *administration*, *governments*, *armed*, *allies* and *intelligence*. Thus, the KKW items *government/s* and/or *administration* and *regime/s* are used to refer to different political systems of which some are associated with certain contexts rather than others (see. sub-section 4.2 for more details) The singular form *regime* is the most significant across the English text type sub-corpora with the highest value of 12.97% of the British editorials and opinions. On the other hand, the singular form *government* is the most significant with its highest value 17.23% of the American news sub-corpus. Analysis showed that collocates of the items *regime* and *regimes*, are items belonging to other semantic categories, other than the GOVERNMENT AND SECURITY, (e.g. *President*, *military*, *forces* and *Zionist*), such as NAME-PA (e.g. *Assad's*, *Bashar*, *Gaddafi*, and *Mubarak*), GEOGRAPHY (*Syrian*, *Iranian*, *Arab* and *Middle East*), FAITH AND RELIGION (e.g. *Islamic*), PROTEST VOCABULARY (e.g. *protests*, *uprising* and *rebels*); TIME (e.g. *former*, *previous*, *current* and *old*); FUNCTION WORDS (e.g. *is*, *has*, *his*, *this*, *anti* and *pro*) and GENERAL AND ABSTRACT TERMS (e.g. *change*). Other collocates with more negative meanings and connotations include *oppressive*, *authoritarian*, *autocratic*, *despotic*, *repressive*, *brutal*, and *embattled*. For example, out of a total of 1,791 (that is, 3.99 ptw) instances of the term *regime* in the British news sub-corpus, there were 79 occurrences of collocate items (with ≥ 5 co-occurrences with ≥ 3 of MI), to the immediate left (L1), with negatively evaluative meanings, such as *anti* (37); *brutal* (10); *authoritarian* (8); *corrupt* (7); *military* (6), *autocratic* (6) and *embattled* (5).

Additionally, the KKW item *administration* is most relatively included in only three sub-corpora, with its highest value 3.13% of the American news sub-corpus. The most frequent lexical collocates of *administration* (876 concordance lines) of the American news sub-corpus include items belonging to different semantic categories, such as

NAME-PN: e.g. *Obama* (346); *Bush* (34) and *Clinton* (16); *Barack* (8);

LANGUAGE AND COMMUNICATION: e.g. *said* (134); *say* (22); *called* (14);

announced (12); *criticized* (10) and *talks* (9).

GOVERNMENT AND SECURITY: e.g. *officials* (132), *official* (75); *senior* (70);

president (27); *new* (22); *military* (16); security

(14); *weapons* (8); *sanctions* (9); *policy* (8).

GEOGRAPHY: e.g. *Washington* (15); *Israel* (13); *Egypt* (12); *American* (9); and

SOCIAL ACTIONS, STATES AND PROCESSES: e.g. *support* (9).

Similarly in the Arabic GOVERNMENT AND SECURITY category, KKW items referring to collectivized news actors, which are also shared (in bold) between the two main Arabic text types sub-corpora include **النظام** <*the-regime*>; **السلطة** <*the-authority*>; الشرطة <*the-police*>; الأنظمة <*the-regimes*>; الحكومة <*the-government*>; الجمهورية <*the-republic*>; الأمة <*the-nation*>; السلطات <*the-authorities*>, **الكيان** <*the-*

entity>; القوات <forces>; البرلمان <the-parliament>; الجيش <the-army>; and السفارة <the-embassy>.

NAME-PN: examples of English include the *United Nations*, the *European Union*, *NATO*, the *Security Council*, *AKP*, *Hezbollah*, *Hamas*, and *Fatah*. In terms of the Arabic sub-corpus, such KKW items include الأمم المتحدة <the-United Nations>; مجلس الأمن <the-Security Council>; القمة العربية <the-Arab summit>; حماس <Hamas>; حزب العدالة والتنمية <Justice and Development Party>, and النهضة <Ennhada>.

SOCIAL ACTIONS, STATES AND PROCESSES: example of the English KKW items include *opposition*, *party*, *parties*, *movement*, *groups*, *people* and *coalition*. Similarly in Arabic, such KKW items of this category include الأحزاب <the-parties>; الجماعة <the-group>; الحراك <the-movement.MASC.SING.N – the movement>; الحركات <the-movements>; الحزب <the-party>; الحركة <the-movement.FEM.SING.N – the-movement>; القيادة <the-leadership>; المؤتمر <the-conference>; النخب <the-elites>; and قوى <forces>. Other KKW items that can be used in collectivizing ‘ordinary’ news actors include الجماهير <the-crowds>; الشباب <the-youth> (PL.); الشعب <the-people>; الشعوب <the-peoples>.

LANGUAGE AND COMMUNICATION: examples of English KKW items of this category include *statement/s*, *speech*, and *report/s*. Similarly, in Arabic, of these collectivising items is التقرير <the-report>.

Other than collectivization, some of the above items can also be used in other representation strategies by which news actors are objectivated. For example,

spatialization is a form of *objectivation*, which can be realized by items of the GEOGRAPHY category. On the other hand, *utterance autonomization* is the form of objectivation in which news actors are represented by means of reference to their utterances (van Leeuwen, 2008, p. 46). The following sub-section discusses two examples of objectivation.

3.5.4 Objectivation

As mentioned above, quite a number of GEOGRAPHY related KKW items ‘often express *who* rather than *where*’ (Bell, 1991, p. 200, original italics), and are quite often represented as news and or political actors rather than as physical territories (Bell, 1991, p. 200). Thus, in van Leeuwen’s (2008, p. 46) terms, *spatialization* occurs when actors are represented by means of reference to a place. This shows how GEOGRAPHY related items can be used either linguistically as pure names of places, or semantically as political entities that represent, as Bell (1991, p. 200) points out, news authorities (e.g. *the US administration* and *the Foreign Ministry*). Figure 3.13 shows concordance examples of some of the KKW grammatical items (that is, *is*, *has* and *will*) of the FUNCTION WORDS category of the English news sub-corpus, as a whole, during May 2011, ordered by R1 and showing some countries and cities (underlined) as their main subjects:

1 with Israelis. Yet the United States **is** also mindful of the limits of its
 2 an introduction to the document. "It **is** also the year when repressive
 3 promoting freedom of expression. It **is** also committed to tackling
 4 has been focused on Israel, Jordan **is** also being forced to renegotiate
 5 eds to to plug a housing gap. Iraq **is** also still rife with bribery and
 6 to mediate in that conflict. Africa **will** also be represented at the
 7 radical Islam and Al Qaeda. There **has** also been a crackdown on the
 8 in Britain and other countries, **has** also made some allies less able
 9 seceded from Khartoum. The UK **will** also establish a presence in
 10 want to settle in the UK. Britain **has** also refused to take a single
 11 the Security Council. Washington **will** also be under pressure to apply
 12 complicating factor: Saudi Arabia **is** also looking to improve its

Figure 3.13 Concordance examples of *is*, *has* and *will* with countries as news actors of the English news during May 2011

Another form of objectivation is *utterance autonomization* in which social/ news actors are represented in terms of their utterances (van Leeuwen, 2008, p. 46). For example, in the English sub-corpus, the KKW items *report/s*, *statement*, *talks*, *negotiations*, *speech* and *resolution* of the LANGUAGE AND COMMUNICATION category, have the potential, and can be used as realizations of this kind of objectivation. In *utterance autonomization*, items such as *report* and *reports*, as Figure 3.14 shows, represent a kind of impersonal authority to the utterances, which are often used, as van Leeuwen (2008, p.46) in connection with the utterances of high-status spokespersons and officials or experts (e.g. line 3). Figure 3.14 shows instances of *said* of the English news sub-corpus, as a whole, during July 2013:

1 and angry residents. The report **said** several cars were destroyed and
 2 NA reported. An earlier report **said** at least 50 people were wounde
 3 rity Committee's annual report **said** spy chiefs considered this UK's
 4 rontline capabilities' The report **said** there was a growing threat of
 5 "vary considerably", the report **said**, but suggest they include sarin,
 6 anean coast, a Reuters reporter **said** hundreds of people fought a
 7 lohamed Brahmi. Local reports **said** that the Governor's building had

Figure 3.14 Concordance examples of *said* showing instances of *utterance autonomization* of the English news sub-corpus

Similarly, in the Arabic sub-corpus, KKW items, such as <التقرير>*<the-report>*; <المحادثات>*<the-talks>*; and <المفاوضات>*<the-negotiations>* of the LANGUAGE AND COMMUNICATION category, can be used in various contexts as realizations of *utterance autonomization*. Figure 3.15 (lines 1, 4 and 7) shows examples of 3-word clusters of the KKW item وقال<*and-said*> (SING.MASC.PST.V), of the Arabic news sub-corpus by which *utterance autonomization* is realized:

Examples of 3-word clusters of وقال< <i>and-said</i> > of the Arabic news sub-corpus				
	Cluster	Translation	Freq	Length
1	وقال البيان إن	<i>and the-statement said that</i>	16	4
2	في الشرق الأوسط	<i>in the Middle East</i>	16	4
3	إن الولايات المتحدة	<i>that the United States</i>	16	4
4	وقال مصدر في	<i>and a source in (...) said that</i>	15	4
5	وقال شهود إن	<i>and witnesses said that</i>	14	4
6	وقال شاهد عيان	<i>and a witness said that</i>	14	4
7	وقال التقرير إن	<i>and the-report said that</i>	14	4

Figure 3.15 Examples of 3-word clusters of وقال<*and-said*> of the Arabic news sub-corpus

3.6 So, Whose Voice? What Role?

Again, the *utterance autonomization* can be considered as a strategy that also represents what someone says. ‘But someone does not mean just *anyone*’ (Bell (1991, p. 191, original italics). Even in such impersonalized representation, items such as *report/s*, *surveys*, *statements* or *speech* lend a kind of impersonal authority to the utterance, which is often used in connection with the utterances of high-profile, official or elite news/social actor/s or spokesperson/s (van Leeuwen, 2008, p.46). *Who says?* is one of the fundamental questions of news work, and the different representation strategies discussed in this chapter revealed the nature of the news sources and actors of which most represent the powerful and the elite in the Arab Spring narrative. In this context, the elite and powerful social/news actors become a favoured source for news media. The more elite they are the more newsworthy is their story. The elite are the most quoted and they are usually political figures and groups/organizations of high profile, and hence news becomes ‘what an authoritative source tells a journalist’ (Bell, 1991, p. 191). Throughout this chapter, we saw high profile news actors’ names, such as *Obama*, *Assad*, *Netanyahu*, *Mahmoud Abbas*, *King Abdullah II of Jordan*, and *(Hillary) Clinton* as the most relatively included elite figures in both the English and Arabic news. According to Partington (2014), if you want to compare the relative absences and disappearance of actors from texts, the best way is to compare two lists or two texts of the same story. If we look at Figure 3.3, we can see who is present and who is absent through the blue and yellow circles. Specifically, this reveals that although

the group, *Hamas*, is present, its leader, *Haniyeh*, is excluded from the English news stories; also, *Abbas*, the Palestinian President is not relatively of the most key in the Arabic news corpus, although he is key in the Arabic editorials, though he is backgrounded in terms of rank. While *Fatah*, *Hezbollah* and *Hamas* are three key news actors (that is group/organizations) and shared by most of the English sub-corpora, only *Hamas* is relatively present in the Arabic sub-corpus as a whole, showing that both *Fatah* and *Hezbollah* are relatively excluded from the narrative and not given a voice. We have seen that of the most relatively present and significantly prominent news actors (individuals and groups/organizations) are those who come from the *Western and European* groups, and are the focus of the news media as the news actors and the news source (Bell, 1991, p.194). The presence of the *EU* and *NATO* shows that this is a narrative of power and military alliance. The actors who are given the most voice, based on occurrences of *said* only, are *Obama* followed by (*Hillary*) *Clinton*, then *Netanyahu*, *Abbas*, *Cameron* and *Assad*. The relatively significant presence of items of the LANGUAGE AND COMMUNICATION category of both English and Arabic, such as *talks*; *negotiations* (المفاوضات); *speech*; *resolution*; *solution* (الحل); *report* (التقرير); and *statement* (البيان) are mostly used, as analysis showed, in connection with the utterances of high-profile and powerful news actors, including individuals (e.g. *Obama*, *Clinton*, *ministers*, *officials* and *activists*) and organizations (e.g. *UN Security Council*). Additionally, while *Obama's* presidency is related to the word *administration*, *Assad's* is related to the much more negative word *regime*. *Regime*, as discussed in

sub-section 3.5.3, is significant in the British English sub-corpus, rather than *government* or *administration*; the word *regime* is meant to be there, as a signifier of an oppressive and non-innocent entity (see also sub-section 4.2 for more details and discussion). *Netanyahu's government* is related to the word *peace* in relation to terms like *peace talks*, making him and his country a positively evaluated group. Similarly, *Abbas' Palestinian Authority* is also associated with items such as *peace negotiations* and *peace talks*, but also with issues related to *freezing settlements* and *reconciliation* (with Hamas). *Hillary Clinton* is presented as active in the context of Libya and the Syrian rebels in both the English and the Arabic news and this makes her also a relatively positive figure. These names therefore represent the polarization of the good and the bad in the Arab Spring story.

The analysis in this chapter provided an insight into how the different news actors across the English and Arabic sub-corpora are discursively represented. The semantic categories and their comprising KKW items appeared to be very useful in identifying the kind of strategies employed, especially that such KKW items can linguistically be considered as realizations of such strategies, which are provided by van Leeuwen's (2008, 1996) Social Actor Network. For example, analysis showed how the GOVERNMENT AND SECURITY category is much related to the *functionalization* strategy, showing the kind of roles that the news actors played in this category. Again, not only the name-pn category suggests that the Arab Spring is the narrative of the elite and power, but also the GOVERNMENT AND SECURITY also showed that most of the roles are of elite and high-profile nature. Thus, *presidents*,

ministers, activists, ambassadors are all elite and are the most relatively included and foregrounded. Functionalization through the GOVERNMENT AND SECURITY category showed that the ordinary news actors (e.g. *civilians, refugees* and *citizens*) do not really have a role other than providing the personalization aspects to the narrative; they are relatively backgrounded and have no voice. In the Arabic sub-corpus, ordinary news actors (only *المواطن*<*the-citizen*>) are relatively absent in the GOVERNMENT AND SECURITY, showing that the elite and the powerful are the ultimate news actors and news sources. Functionalization through the PROTEST VOCABULARY showed the suffering of the ordinary *protesters* and *demonstrators*, but also the *rebels* revealed much about the elite and the powerful, and their interests in weapon deals and arming the rebels. Functionalization through the SOCIAL ACTIONS, STATES AND PROCESSES, showed also that most of the roles belong to the elite, *kings, sheikhs, leaders*, and even *dictators*

3.7 Summary

This chapter discussed the most frequent discursive strategies employed in the representation of the most significantly prominent news actors (individuals and groups/ organizations) of the NAME-PN category as well as other semantic categories of the English as well as the Arabic sub-corpus. The van Leeuwen's (2008, 1996) Social Actor Network was employed to examine *who* and *how* the different news actors of the different semantic categories are linguistically and semantically represented. Section 3.1 introduced the frameworks for the analysis and section 3.2

discussed the main strategies of inclusion and exclusion and their related discursive strategies. Section 3.3 discussed the main news actors in the name-pn category, whereas section 3.4 located them geographically. In section 3.5, the different representation strategies of news actors were introduced and discussed. Sub-section 3.5.1 discussed functionalization; sub-section 3.5.2 discussed *classification* in terms of *provenance*, *religion*, *gender* and *age*. GEOGRAPHY, FAITH AND RELIGION as well as the SOCIAL ACTIONS, STATES AND PROCESS categories also appeared useful in revealing the way the news actors are classified and identified. In sub-section 3.5.3, the *assimilation* strategy and how it is realized through individuality (singularity) and assimilation (plurality) and collectivization (through mass nouns or nouns denoting a group of people) were also highlighted and examples of the semantic categories were also provided for both languages of English and Arabic. Analysis in this sub-section showed that individualization, assimilation and collectivization are significantly salient strategies across the English as well as the Arabic text types sub-corpora, and which can be considered from two different perspectives. According to van Leeuwen (2008, p. 38), the powerful and the elite news actors and governments are usually individualized (e.g. *US administration*), whereas their people ‘the We’ are usually collectivized not only through the first-person, but also through items denoting a group of people such as the *Western nations*. While collectivizing the news actors or people means representing them as homogeneous and in agreement, individualizing, on the other hand, allows their titles, credentials and institutional affiliations to be showcased (van Leeuwen, 2008,

p. 38). However, in contexts of conflicts and wars, which are usually accompanied with polarized opinions, it is believed that assimilation and collectivization of different news actors with polarized attitudes, usually function as ‘deflecting personal accountability’, as Bell (1991, p. 194) pointed out. The ordinary news actors are usually assimilated and quantified by means of aggregation, that is, treating them as statistics. In sub-section 3.5.4, the objectivation strategy was also discussed by providing two examples of *spatialization* and *utterance autonomization*, which also suggested the eliteness of the utterances as they are usually represent the high-profile news actors and sources (van Leeuwen, 2008, p. 46).

In this chapter I have looked at *who* is being reported and also who is absent; in the next chapter I look at *what* and *where* is reported. Rather than leaving behind the topic of *who* is being reported, we see that the *what* and *where* also involve *who* is being reported, because the topics involve important and elite actors in the Arab Spring narrative, rather than the ordinary people and events of their lives and struggles.

4. Unlocking the Arab Spring: Key Topics

4.1 Introduction

This chapter addresses the second research question that is concerned with the main topics associated with the Arab Spring corpus, and quantitatively as well as qualitatively examines the top ten key semantic categories (i.e. topics or macrostructures) which are discovered using *Wmatrix*. The top ten semantic categories, as Table 4.1 shows, are GEOGRAPHY; GOVERNMENT AND SECURITY; NAME-PN; FUNCTION WORDS; SOCIAL ACTIONS, STATES AND PROCESSES; FAITH AND RELIGION; PROTEST VOCABULARY; LANGUAGE AND COMMUNICATION; GENERAL AND ABSTRACT TERMS; and EMOTIONS. Analysis in this chapter attempts to investigate these categories in terms of differences and similarities between languages (English and Arabic), text types (news and editorials) as well as change over time. The categorization of the eight KKW lists (six KKW lists for the English sub-corpora, and two KKW lists for the Arabic) resulted in a number of semantically meaningful categories (ordered by the overall number of KKWs in each category) of which GEOGRAPHY, GOVERNMENT AND SECURITY as well as NAME-PN are the most significant in terms of the number of the resulting KKW items.

The *OED* (2013) defines the *Arab Spring*, more specifically, as ‘a series of anti-government or pro-democratic uprisings and demonstrations in various countries in North Africa and the Middle East, beginning in Tunisia in December 2010’. Such

definition suggests that the *Arab Spring* in my corpus is described within discourses of a geographical, governmental, political as well as a revolutionary nature in both the English and Arabic sub-corpora, as indicated in Example 24 (lines 3, 4 and 5).

24. 1 Is this the long-awaited "Arab Spring"?
 2 The consequences of Egypt's unrest could be
 3 great for other **Arab countries** and rulers.
 4 **Democracy** is a rare commodity in the **region**
 5 and several other **governments** could be sitting
 6 on similar political volcanoes.
 (BBCe-11-20/news: Jan-31)

However, further analysis showed that contrasting themes/concepts within these and other categories (by means of pairs of items with positive/negative connotations) are also prevalent. For example, at the lexical level the following contrasting pairs are revealed: *democracy/dictatorship*, *religious*, *sectarian/secular*, *peace/ violence*, *allies/enemy*, *corruption/ reform*, *opposition/ support*. Similarly, at the grammatical level items, such as *pro/anti*, *is/non-* and *not*, also indicate the contrastive as well as the polarizing nature of the Arab Spring narrative. The following section discusses in more detail the KKW semantic categories in the English and Arabic sub-corpora identified by using *WordSmith Tools*, and briefly describes the top ten key categories in the Arab Spring corpus with a focus on the two most significant categories: GEOGRAPHY and GOVERNMENT AND SECURITY, as categories having the potential to construct the *where* and the *what* in the Arab Spring corpus respectively.

4.2 KKW Function for Identifying Topics in the Arab Spring

Corpus

WordSmith's KKW list function (explained in Chapter 2) allows for investigating if a set of words are high in frequency in many or a few texts in the corpus as well as for deciding whether to include or exclude such texts from the analysis (Baker, 2004; Scott, 2015). A key keyword is a word that 'is "key" in more than one of a number of related texts. The more texts it is "key" in, the more "key key" it is' (Scott, 2015). Based on this definition, a KKW item in this study means a keyword that is key in a number of texts relatively equalling 1% or more of the texts comprising its respective sub-corpus. For example, the KW item *Arab* is key in 403 texts of the American news sub-corpus (which consists of a total of 1,803 texts), and hence the KW item *Arab* is said to be KKW in 31.56% of the (1,803 texts comprising) the American news sub-corpus. The overall resulting KKW items of the English sub-corpus were 663 KKW items. Only 292 KKW items, which are shared by 3, 4, 5 and 6 of its main six sub-corpora, have been included in the analysis. In terms of the Arabic sub-corpus, all of the 277 resulting KKWs were included. The resulting KKW items of the two main sub-corpora (along with their language forms: British, American and Global Variety of English and Modern Standard Arabic, as well as text types: news and editorials and opinions) were manually categorized into semantically similar categories. The categorization of the English sub-corpora was also conducted, as discussed in sub-section 2.3.4, and

hence the categories identified, by the *Wmatrix* (Rayson, 2008, p. 544) automated method (although minor changes were made on the names of the categories), which extends ‘the keyness technique from key words to key semantic domains’. In addition, it is worth mentioning that the categorization of the KKWs and/or the collocates does not rely so much on their dictionary meaning, but rather on their contexts (Baker et al., 2013a, p. 262). For example, the examination of the concordance lines (a total of 2,275 lines) of the collocates as well as associates of the KKW item *Bank* in the English corpus showed that the term is most suited to the GEOGRAPHY category (e.g. the *West Bank*: 1,718 concordance lines; *East Bank Jordanians/ of Jordan*: 30 lines) than into the ECONOMY AND FINANCE category (e.g. *Central Bank*: 115 concordance lines; *World Bank*: 108 concordance lines). The decision to include the KKW item *Bank* in the GEOGRAPHY category was then pragmatically taken, as the aforementioned category is the most salient, and the ECONOMY AND FINANCE category is not amongst the top ten categories. Likewise, the KKW item مجلس <majlis – council.SING.MASC.N - *Council*: as in مجلس الأمن <majlis–council; الأمن-al-’amn–the-security: *the Security Council*>, or مجلس التعاون الخليجي <majlis–council; التعاون - at-ta’āwun – the-cooperation.SING.MASC.N – *the Cooperation*; الخليجي - al-ḵalījī – the-gulf.SING.MASC.ADJ: *the Gulf Cooperation Council*>; and the KKW items العدالة <al-’adāla – the-justice.FEM.N – *Justice*>; والتنمية <wat-tanmiya – *and-the-development*>, which are mostly used in the Arabic corpus to refer to the Turkish political party حزب العدالة والتنمية – *The Justice and Development Party*>, have been

categorized under the NAME-PN category (most specifically of groups and organizations sub-category). names. Another example of categorizing items based on their contextual rather than dictionary meaning is the KKW item *(the)spring* - الربيع (e.g. the term *(the)Arab Spring* – الربيع العربي), which is mostly used in contexts of items that mostly belong to the PROTEST VOCABULARY, such as *uprisings*, *protests*, *revolts* and *revolutions*, rather than items from the TIME category to which *Wmatrix* (Rayson, 2008) would automatically assign it. In the same context, the Arabic item الشعبية <aš-ša‘biy-ya - the-popular.DEF.ADJ-FEM – *the popular*>, which is only KKW in the Arabic editorials and opinions sub-corpus, was categorized under the EMOTION category although it is used as a post-modifier of other KKW items belonging to the PROTEST VOCABULARY category, as in the phrases

الثورات الشعبية العربية <*the-popular Arab revolutions*>;

الثورة الشعبية <*the popular revolution*>;

الانتفاضات الشعبية <*the popular uprisings*>;

الاحتجاجات الشعبية <*the-popular protests*>

or in a few instances as part of names of political parties such as,

الجهبة الشعبية لتحرير فلسطين <*the Popular Front for the Liberation of Palestine*>.

The decision to include the KKW item الشعبية <*the popular*> in the EMOTION category is firstly based on its meaning, like its English equivalent, as an item referring to something liked and enjoyed, or beliefs shared by a large number of people. Another reason is the fact that many items in the PROTEST VOCABULARY reflect in some of their senses feelings of anger or resentment. Additionally, analysis of the

concordance lines (1,173 lines) of the KKW الشعبية <the popular> in the Arabic editorials and opinion sub-corpus, showed that a number of its collocates, especially to the immediate right (R1) and immediate left (L1), as well as associates refer to items with either emotional meanings and/or connotations. For example, in her study regarding how *corpus linguistic* has been constructed in its own literature, Taylor (2008, p. 187-8) looked at the right collocates of the word *corpus*, then argued that the term *corpus linguistics* occurred less frequently than expected because there were so many alternatives used with the word *corpus* such as *based*, *analysis*, *studies* as well as *research*. Thus, in terms of the KKW item الشعبية <the-popular>, the item الإرادة <al'irāda – the-will.FEM.SING.N – the will> is a strong associate as well as collocate to the immediate left (L1, 43 instances, with MI score 9.36) of the KKW item الشعبية <the-popular>, as in the phrase الإرادة الشعبية <the-popular will>. Other items amongst the top twenty collocates to the immediate left (L1) of the item الشعبية <the-popular> include المقاومة <the-resistance>; الحاضنة <the-incubator>; المشاركة <the-participation>; الجماهير <the-crowds>; اللجان <the-committees> and الثقافة <the-culture>. In terms of the immediate right (R1) collocate items, the emotional item السلمية <as-silmiy-ya – the-peaceful.ADJ-FEM – the-peaceful>, is the top second lexical collocate to R1 the KKW item الشعبية <the-popular> (with 23 out of a total of 28 instances, and MI score 7.44). Another collocate item that has emotional connotations related mostly to strong feelings, or fierceness is the item العارمة <al-‘ārim-a – the-fierce.ADJ-FEM – the fierce>.

As can be seen from the above examples of the categorization process, it is important, as Baker (2004, p. 353) pointed out, to conduct ‘separate analyses of individual keywords in order to note their general functions’ before combining them together in ways that make sense, though grouping of KWs/KKWs is a subjective process. In other words, the multifunctionality of language and the fact that many word forms and lemmas have different meanings and usages should be considered before assigning words or word forms under investigation to a given semantic category or a news value (Potts et al., 2015, p. 11; Baker et al., 2008, p. 278; Bednarek, 2016; Potts, 2015). According to Stubbs (1996, p. 172), one form of a word is usually more common than the others and that word forms can have quite different collocates. Collocates of a given word contribute to its meaning (Baker et al., 2008; Louw, 1993), and can give ‘us a semantic analysis of [...that] word’ (Sinclair, 1991, p. 116). In the context of the Arab Spring, however, there is the GENERAL AND ABSTRACT TERMS category which contains KKW items (e.g. *conflict*, and *crisis*), of both the English and Arabic sub-corpora, that are also collocates of the term *Arab Spring* (الربيع العربي). Collocational analysis of these general terms, such as, *conflict*, *crisis* and *freedom*, along with their Arabic equivalents, though in the definite form, <الصراع>*the-conflict*; <الأزمة>*the-crisis*; and <الحرية>*the-freedom* respectively, showed that they occurred in contexts invoking different discourses, related to geography, religion and other belief systems, language and linguistic activities politics, military and warfare as well as economy and finance. The

decision was to include these KKWs as general terms serving different contexts, hence comprising the GENERAL AND ABSTRACT TERMS category.

Similarly, in the LANGUAGE AND COMMUNICATION category, which was aimed to contain items that refer to actions, states or processes of linguistic nature, or people or organizations working in the field of information or communication. Hence, other than KKW items, such as *said* and *وقال* <and-said.SING.MASC>, names such as the English KKW items *Facebook*, *Twitter*, *WikiLeaks* (KKW item in the British sub-corpora only) and *journalists*, and the Arabic KKW item *الجزيرة* <Al-Jazeera>, were included in this category. Other items, such as *media* (الإعلام), *internet*, *talks*; *negotiations* (مفاوضات); and *speech* were also included. For example, in a total of 1,987 instances of the Arabic item *الإعلام* <al-'i'lām – the-media or the-information> in the Arabic sub-corpus, its most frequent co-occurrences were with the lexical items *وسائل* <wasā'il – ways or means> (655); *حرية* <hurriyya – freedom> (91); and *وزير* <wazīr – minister> (52). Thus, the co-occurrence of *الإعلام* with *وزير* <minister>, as in the phrase *وزير الإعلام* <Minister of Information>, means that the item *الإعلام*, like other KKW items, may fit more in this sense under the GOVERNMENT AND POLITICS category, and hence contribute to the news value of *Eliteness*. However, the decision was to categorize it under LANGUAGE AND COMMUNICATION as its general function in the Arabic corpus is believed to fit more in this category. Figure 4.1 shows the top ten 2-3-word clusters of the KKW item *الإعلام*, which are believed to provide a better understanding of how the term *الإعلام* is mostly used in the whole Arabic sub-corpus:

N	Cluster	Translation	Set	Freq	Length
1	وسائل الإعلام	<i>the media</i>		647	2
2	الإعلام في	<i>the media in</i>		145	2
3	في الإعلام	<i>in the media</i>		106	2
4	في وسائل	<i>in media</i>		90	2
5	في وسائل الإعلام	<i>in the media</i>		90	3
6	الإعلام العربي	<i>the Arabic media</i>		77	2
7	حرية الإعلام	<i>the freedom of media</i>		72	2
8	ووسائل الإعلام	<i>and the media</i>		60	2
9	وزير الإعلام	<i>the Minister of Information</i>		52	2
10	أن الإعلام	<i>that the media</i>		45	2

concordance collocates plot patterns clusters timeline filenames

281 entries Row 15 T S ... على ا

Figure 4.1 Top ten 2-3-word clusters of الإعلام in the Arabic corpus

Table 4.1 shows statistical information about the top ten significant key categories in terms of their total number of KKWs along with their overall percentage of each English and Arabic text type sub-corpus corpus. The results are also visually represented by means of coloured data bars. Sorted automatically by means of a pivot table in Excel, the categories are ordered by the grand total number of KKW items in each category. Data bars in Excel conditional formatting make it easy to visualize values in a range of cells, and hence the longer the bar the larger the value

Table 4.1 Top ten KKW-semantic categories ordered by overall number of words with their Avg. % of English and Arabic text type sub-corpora

Key semantic category	News				Editorials and opinions				Grand Row Totals
	ENASC			ARASC	ENASC			ARASC	
	AM	BR	GV	MSA	AM	BR	GV	MSA	
1 GEOGRAPHY	<div><div>93</div><div>5.00</div></div>	<div><div>80</div><div>5.06</div></div>	<div><div>69</div><div>6.52</div></div>	<div><div>74</div><div>2.23</div></div>	<div><div>89</div><div>4.76</div></div>	<div><div>90</div><div>5.12</div></div>	<div><div>72</div><div>7.16</div></div>	<div><div>78</div><div>2.94</div></div>	<div><div>645</div><div>4.83</div></div>
2 GOVERNMENT AND SECURITY	<div><div>54</div><div>3.64</div></div>	<div><div>48</div><div>2.97</div></div>	<div><div>42</div><div>3.36</div></div>	<div><div>31</div><div>1.88</div></div>	<div><div>48</div><div>2.87</div></div>	<div><div>42</div><div>2.64</div></div>	<div><div>38</div><div>3.21</div></div>	<div><div>34</div><div>2.45</div></div>	<div><div>337</div><div>2.94</div></div>
3 NAME-PN	<div><div>47</div><div>4.54</div></div>	<div><div>42</div><div>4.55</div></div>	<div><div>36</div><div>4.13</div></div>	<div><div>33</div><div>1.86</div></div>	<div><div>38</div><div>4.14</div></div>	<div><div>37</div><div>4.38</div></div>	<div><div>27</div><div>4.72</div></div>	<div><div>19</div><div>2.63</div></div>	<div><div>279</div><div>3.98</div></div>
4 FUNCTION WORDS	<div><div>11</div><div>9.11</div></div>	<div><div>9</div><div>1.81</div></div>	<div><div>11</div><div>8.06</div></div>	<div><div>28</div><div>12.92</div></div>	<div><div>16</div><div>8.04</div></div>	<div><div>11</div><div>2.76</div></div>	<div><div>16</div><div>8.98</div></div>	<div><div>31</div><div>18.62</div></div>	<div><div>133</div><div>10.88</div></div>
5 SOCIAL ACTIONS, STATES AND PROCESSES	<div><div>15</div><div>3.27</div></div>	<div><div>11</div><div>2.70</div></div>	<div><div>13</div><div>2.70</div></div>	<div><div>16</div><div>2.07</div></div>	<div><div>14</div><div>2.51</div></div>	<div><div>11</div><div>1.80</div></div>	<div><div>14</div><div>1.69</div></div>	<div><div>26</div><div>2.06</div></div>	<div><div>120</div><div>2.33</div></div>
6 FAITH AND RELIGION	<div><div>13</div><div>4.36</div></div>	<div><div>14</div><div>4.59</div></div>	<div><div>11</div><div>4.31</div></div>	<div><div>9</div><div>2.00</div></div>	<div><div>15</div><div>4.34</div></div>	<div><div>14</div><div>5.82</div></div>	<div><div>14</div><div>5.65</div></div>	<div><div>12</div><div>1.77</div></div>	<div><div>102</div><div>4.25</div></div>
7 PROTEST VOCABULARY	<div><div>11</div><div>7.00</div></div>	<div><div>10</div><div>4.73</div></div>	<div><div>10</div><div>4.50</div></div>	<div><div>8</div><div>2.92</div></div>	<div><div>11</div><div>3.35</div></div>	<div><div>10</div><div>4.46</div></div>	<div><div>11</div><div>2.69</div></div>	<div><div>9</div><div>5.48</div></div>	<div><div>80</div><div>4.41</div></div>
8 LANGUAGE AND COMMUNICATION	<div><div>9</div><div>5.31</div></div>	<div><div>9</div><div>2.39</div></div>	<div><div>8</div><div>3.71</div></div>	<div><div>14</div><div>2.31</div></div>	<div><div>8</div><div>2.43</div></div>	<div><div>6</div><div>1.61</div></div>	<div><div>6</div><div>2.54</div></div>	<div><div>2</div><div>1.49</div></div>	<div><div>62</div><div>2.84</div></div>
9 GENERAL AND ABSTRACT TERMS	<div><div>4</div><div>1.80</div></div>	<div><div>3</div><div>1.27</div></div>	<div><div>5</div><div>1.78</div></div>	<div><div>2</div><div>1.38</div></div>	<div><div>4</div><div>2.35</div></div>	<div><div>4</div><div>3.24</div></div>	<div><div>2</div><div>3.10</div></div>	<div><div>12</div><div>1.97</div></div>	<div><div>36</div><div>2.08</div></div>
10 EMOTIONS	<div><div>5</div><div>3.51</div></div>	<div><div>7</div><div>2.46</div></div>	<div><div>4</div><div>4.31</div></div>	<div><div>2</div><div>1.70</div></div>	<div><div>4</div><div>3.26</div></div>	<div><div>5</div><div>2.51</div></div>	<div><div>4</div><div>6.06</div></div>	<div><div>3</div><div>1.36</div></div>	<div><div>34</div><div>3.21</div></div>
Grand Column Totals	<div><div>262</div><div>4.70</div></div>	<div><div>233</div><div>4.03</div></div>	<div><div>209</div><div>4.84</div></div>	<div><div>215</div><div>3.49</div></div>	<div><div>247</div><div>4.16</div></div>	<div><div>230</div><div>4.11</div></div>	<div><div>204</div><div>5.33</div></div>	<div><div>226</div><div>4.83</div></div>	<div><div>1,828</div><div>4.42</div></div>
Bar Colour Key	■ Total Number of KKWs				■ Overall Avg. % of sub-corpus				

Table 4.2 Actual totals of KKW items of the top ten categories in the English and Arabic sub-corpora

Key Categories	Example KKWs of English and Arabic categories	No of KKWs		Row Totals
		ENASC	ARASC	
1 GEOGRAPHY	<i>Arab</i> (العربي); <i>Syria</i> (سوريا); <i>Egypt</i> ; <i>Israel</i> ; <i>US</i> ; <i>Russia</i> ; <i>countries</i> ; <i>world</i> ; <i>region</i> ; <i>foreign</i> ; <i>middle</i> ; <i>east</i> ; (الشارع - <i>the-street</i>)	100	72	172
2 GOVERNMENT AND SECURITY	(<i>democracy</i> - الديمقراطية); (<i>gouvernement</i> - الحكومة); (<i>regime</i> - النظام); (<i>president</i> - رئيس); (<i>security</i> - الأمن); <i>elections</i> ; <i>minister</i> ; <i>military</i> ; <i>activists</i> ; <i>war</i> ; (<i>Zionist</i> - الصهيوني); (<i>occupation</i> - الاحتلال)	61	43	104
3 NAME-PN	<i>Mr</i> ; (<i>Assad</i> - الأسد); (<i>Obama</i> - أوباما); (<i>Netanyahu</i> - نتنياهو); (<i>Abbas</i> - عباس); <i>UN</i> ; (<i>Muslim</i>) <i>Brotherhood</i> ; (<i>Security</i>) <i>Council</i> - مجلس الأمن;	46	31	77
4 FUNCTION WORDS	<i>not</i> ; <i>anti</i> ; <i>pro</i> ; (<i>against</i> - ضد); (<i>we</i> - نحن); <i>he</i> ; <i>our</i> ; <i>is</i> ; <i>has</i> ; (<i>between</i> - بين); (<i>on</i> - على); (<i>which</i> .FEM - التي); (<i>was</i> .MASC - كان);	18	33	51
5 SOCIAL ACTIONS, STATES AND PROCESSES	(<i>opposition</i> - المعارضة); <i>party</i> ; <i>leaders</i> ; <i>king</i> ; <i>women</i> (المرأة); (<i>people</i> - الشعب)	17	30	47
6 FAITH AND RELIGION	<i>Muslim/s</i> ; (<i>Islam</i> - الإسلام); <i>Jewish</i> ; <i>secular</i> ; <i>Christians</i> ; <i>religious</i> ; (<i>sectarian</i> - الطائفي); <i>Shia</i> ; <i>Sunni</i> ; <i>Jews</i>	16	10	26
7 PROTEST VOCABULARY	(<i>protests</i> - احتجاجات); <i>demonstrations</i> ; (<i>uprisings</i> - انتفاضات); (<i>rebels</i> - الثوار)	12	12	24
8 LANGUAGE AND COMMUNICATION	<i>said</i> (قال); <i>talks</i> (محادثات); <i>negotiations</i> ; <i>media</i> ; <i>speech</i> ; <i>journalists</i>	10	15	25
9 GENERAL AND ABSTRACT TERMS	(<i>conflict</i> - الصراع); (<i>intervention</i> - التدخل); (<i>crisis</i> - الأزمة); <i>freedom</i>	5	12	17
10 EMOTIONS	(<i>peace</i> - السلام); (<i>violence</i> - العنف); <i>peaceful</i> ; <i>fear</i> ; <i>terror</i> ; <i>threat</i>	7	3	10
Column Totals		292	261	553

Although keyness, as discussed earlier (see sub-section 2.3.2), cannot directly be compared across different languages, as comparison of KW items involves using different reference corpora in each language (Freake et al., 2011; Vessey, 2013; Gabrielatos and Marchi, 2012) it is believed, as Table 4.1 shows, that there still is a level of comparability, as will be discussed shortly, between the different sub-corpora in terms of the nature of the resulting semantic categories, and the number of shared KKW items, or the topics they denote, in each category, and hence some observations can be made.

As Table 4.1 also shows, the blue data bars visually represent the total number of KKWs in each category across the English and Arabic text types sub-corpora, whereas the green data bars represent the average percentage of each sub-corpus. The Grand Row Total displays the grand (i.e. plotted or aggregated) total number of KKWs, and the overall average percentage of the whole sub-corpora, whereas the Grand Column Totals display the ‘actual’ grand total number of KKWs as well as the overall average percentage of a sub-corpus. The results that are shown under the Grand Row Total in Table 4.1 are the plotted results, and are used for the sake of the analysis only. However, in order to show the actual number of the KKWs, Table 4.2 represents a pivot table in a simpler form, and shows the actual raw number of the resulting KKW items of the top ten semantic categories of both the English as well as the Arabic sub-corpora as a whole, along with English and Arabic example KKWs of each category. As can be seen, the results in Table 4.2 give an overall picture of the top ten semantic categories showing an identical pattern to

that in Table 4.1 in terms of the order of these categories, which is based on the overall total number of KKW items in each category. However, both Tables 4.1 and 4.2 show that both English and Arabic sub-corpora represent different patterns in terms of the number of the top categories and their order. The English sub-corpora show that there are three top categories, that is GEOGRAPHY, GOVERNMENT AND SECURITY and NAME-PN categories with the same order as number one, two and three respectively. On the other hand, the Arabic sub-corpus, as Table 4.2 shows, has five top categories of which the FUNCTION WORDS (with 33 KKW items) ranking as number three followed by the NAME-PN (with 31 KKW items) and the SOCIAL ACTIONS, STATES AND PROCESSES (with 30 KKW items) categories ranking as four and five respectively. The difference in the number of the top semantic categories between English and Arabic can be justified by the fact that the totals in the Arabic data represent all of the KKW items in the two main Arabic sub-corpora altogether, whereas in the English data, the totals represent the KKW items that are only shared by three (out of six) or more of the English sub-corpora. For example, the 18 KKW items of the FUNCTION WORDS category of the English corpus are the items that are shared by three (out of six) or more of the English sub-corpora (which originally total 36 KKW items). Another reason for the difference in results can be attributed to the difference between the reference corpora used for comparing the different English and Arabic word lists. However, when taken individually, it is the Arabic editorials and opinion sub-corpus, as Table 4.1 shows, in which the FUNCTION WORDS category (with 31 KKW items) ranks as number three followed by the

SOCIAL ACTIONS, STATES AND PROCESS (26 KKW items) as number four. On the other hand, the NAME-PN category (with 33 KKW items) ranks as number two in the Arabic news sub-corpus, which perhaps indicates a difference in focus between the two text type sub-corpora of both English and Arabic.

Before getting further into the resulting KKW semantic categories, it is worth first discussing, though briefly, some of the observations made from the results shown in Tables 4.1 and 4.2, and giving a brief description of the remaining categories in Table 4.1. A first observation is that the top ten resulting categories are generally shared between the different English as well as Arabic text type sub-corpora, indicating a similarity between them despite the differences in their relative keyness. The resulting categories, of which quite a number of them comprising KKW items (see examples of KKWs in Table 4.2) are collocates of the term *Arab Spring*, provide a useful summary of the main topics associated with the *Arab Spring* at both the word as well as the semantic levels. Such association is called within the *Wmatrix* approach (Prentice et al., 2012, p. 275; Rayson, 2008) the ‘semantic tag collocations’. ‘In semantic tag collocations, the node is a word ...[such as *Arab Spring*] and the collocate is a semantic category, such as ...[GEOGRAPHY], or vice versa’ (Prentice et al., 2012, p. 275). Figure 4.2 shows examples of some of the KKW items (in bold) of the GEOGRAPHY category associated with the term *Arab Spring*, taken from the British news sub-corpus.

1 Mr Obama's handling of the Arab spring and of Iran's nuclear pro
 2 ntries at the heart of the Arab spring - Libya, Egypt and Syria
 3 Tunis. Tunisians led the Arab Spring; they know the world will
 4 the very least, derail the Arab Spring. Syria could try to defle
 5 the West had provoked the Arab Spring uprisings, and cautioned
 6 le the West has backed the Arab Spring, with China it talks with
 7 : World briefing Why some Arab spring countries are doing better
 8 ding world events like the Arab Spring. Media One memo - said to
 9 world. Marginalised by the Arab Spring, which has dealt a severe
 10 d, particularly during the Arab Spring. The emir visited Gaza la

Figure 4.2 Concordance lines of the Arab Spring with key geographical items in the BR-news sub-corpus

As the most key semantic category, in terms of the overall total number of KKW items of both the English as well as the Arabic sub-corpora, the GEOGRAPHY category contains, as Table 4.2 shows, the highest ‘actual’ total number of KKW items (100 and 72 KKW items of English and Arabic respectively), indicating the lexical richness of this category as well as the newsworthiness of its related topics, the latter of which will be discussed in Chapter 5. According to Prentice et al. (2012, p. 277), GEOGRAPHY related items could refer (along with their different grammatical forms: e.g. noun, adjective, and genitive) to specific locations or geographical areas of countries (e.g. *Arab/s*, *Israel/i/s/’s* and *Syria/n/ns/’s*), cities (e.g. *Gaza*, *Cairo*, *Jerusalem* and *Damascus*), specific places (e.g. *Tahrir square*) or to a place that refers to a location (e.g. *Middle*, *East*, *region*, *countries*, *world*, *square* and *border*). Similarly, in Arabic, grammatical forms also include the in/definite and/or the masculine/feminine forms such as:

العربي <the-Arab.MASC>;
 السوري <the-Syrian.MASC>;
 مصر <Egypt>;
 إسرائيل <Israel>;
 السورية <the-Syrian.FEM>;
 تونس <Tunisia>;
 سوريا <Syria>;
 الفلسطينية <the-Palestinian.FEM>;
 تركيا <Turkey>;
 الأمريكية <the-American.FME>; and
 دول <countries.INDEF.FEM.PL>.

However, analysis shows that various English as well as Arabic KKW geographical items often pattern with verbs, and hence fill in the role of ‘political actors’ (Bell, 1991, p. 200) involved in different (positive/negative) actions rather than mere expressions of geographical places or locations. Examples of this pattern can be seen in Figure 4.1 (lines 3, 4, 5, 6, and 7) in which the bolded KKW items *world*, *Syria*, *West*, and *countries* are represented as grammatical subjects of the verbs *will*, *could try*, *had provoked* (also *has backed*) and *are doing* respectively. Other than items referring to locations in general (e.g. *world* and *countries*), the items of this category, as also indicated in Figure 4.1, refer to various ‘geographical identities’ (Taylor, 2014, p. 378) of different regions and/or cultures such as *Arab* (e.g. *Libya*, *Tunisia*, *Egypt* and *Syria*), *Western and European* (e.g. *West*), *Islamic* (e.g. *Iran*) and *East Asian* (e.g. *China*). This geographical multiplicity could potentially help in identifying other ideological patterns of alliance/rivalry that are found in various contexts, especially those related to the GOVERNMENT AND SECURITY category,

which ranks as number two in terms of the overall total number of KKW items of the Arab Spring corpus as a whole.

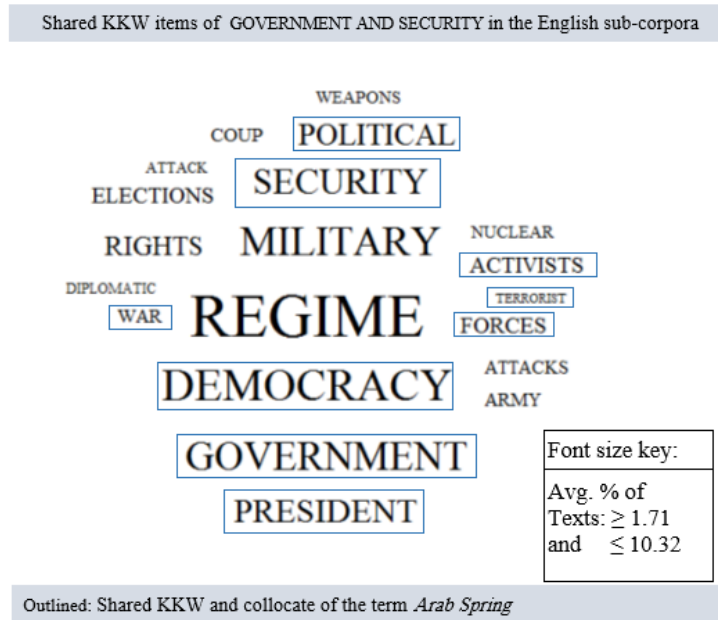


Figure 4.3 Shared KKW items of the English GOVERNMENT AND SECURITY

The GOVERNMENT AND SECURITY is the second top category in terms of the overall total number of KKW items, and includes terms that generally refer to governmental, political, military, and security topics. The English sub-corpus, as Table 4.2 shows, has an overall total of 61 KKW items of which 20 are shared between its main six sub-corpora. Presented by the overall average percentage of the sub-corpora, Figure 4.3 shows a KKW cloud of the 20 shared KKW items of the GOVERNMENT AND SECURITY category in the English sub-corpus as a whole. The KKW item *regime* has the highest value 10.32% of the English sub-corpus as a whole. The outlined KKW items, *democracy*, *government*, *security*, *president*,

political, forces, activists, war and terrorist, are also collocates of the term *Arab Spring*.

In terms of the Arabic sub-corpus, only 12 out of 43 KKW items are shared between its two main sub-corpora. Ordered by their percentage of the sub-corpus, the KKW items in the GOVERNMENT AND SECURITY include

- النظام <*the-regime*> (MASC) [e.g. النظام السوري – *the-Syrian regime*]
 السياسي <*the-political*> (MASC) [e.g. الإسلام السياسي – *political Islam*]
 الانتخابات <*the-elections*> (FEM)
 السلطة <*the-authority*> (FEM)
 الرئيس <*the-president*> (MASC) [e.g. الرئيس السوري – *the-Syrian President*]
 السياسية <*the-political*> (FEM) [e.g. القوى السياسية – *the-political powers*]
 [e.g. الأحزاب السياسية – *the-political parties*]
 الديمقراطية <*the-democracy*> (FEM.N)
 الصهيوني <*the-Zionist*> (MASC) [e.g. الكيان الصهيوني – *the Zionist entity*]
 الكيان <*the-entity*> (MASC) [e.g. الكيان الإسرائيلي – *the Israeli entity*]
 الأمن <*the-security*> (MASC) [e.g. name: مجلس الأمن – *Security Council*;
 [e.g. قوات الأمن – *the-security forces*]
 الاحتلال <*the-occupation*> (MASC)
 [e.g. الاحتلال الإسرائيلي – *the-Israeli occupation*]
 الجيش <*the-army*> (MASC) [e.g. الجيش المصري – *the-Egyptian army*]
 [e.g. الجيش السوري الحر – *the-Free Syrian Army*].

of which all, except two items الكيان<*the-entity*>; and الجيش<*the-army*>; are collocates (in bold) of the term الربيع العربي<*the Arab Spring*>. As expected, many of the items of the Arabic sub-corpus are also shared with the English sub-corpus, which indicates similarity in focus between them, for example, issues related to

<regime (change)> (تغيير) النظام; الديمقراطية <democracy>, الانتخابات <the-elections>; and الأمن <the-security>. However, the Arab-Israeli conflict, especially issues related to الاحتلال الإسرائيلي <the-Israeli occupation>, tends to be one of the main issues or topics associated with الربيع العربي <the Arab Spring> in the Arabic corpus. In fact, the English equivalent of the KKW item الاحتلال <the-occupation>, was only KKW in one (i.e. the GV editorials and opinion sub-corpus) of the English sub-corpora, and hence was not included. Nevertheless, analysis shows that there are KKW items that are shared by five (i.e. *settlement*), four (i.e. *settlements*), or three (i.e. *Flotilla* and *statehood*) of the English sub-corpora, and which directly relate to the Israeli-Palestinian conflict. The issue of الاستيطان <al-'istīfān – the-settlement> is only key in the Arabic editorials and opinion sub-corpus.

Another observation to draw from the GOVERNMENT AND SECURITY category is the presence of contrastive aspects indicated by the negative and positive semantic pair items (e.g. *government and/or administration/regime, democracy/ suppression, elections/coup, corruption/ reform, allies/ war or enemy*), which in turn might suggest the polarizing nature of Arab Spring narrative. An example of these contrastive aspects is the number of KKW items (and/or their forms) used to refer to political, governmental, or other ruling systems in general. In the English sub-corpus for instance, the KKW items *government/s and/or administration and regime/s* are used to refer to different political systems of which some are associated with certain contexts rather than others. Table 4.3 shows these KKW items along with their percentage of each of the English text types sub-corpora:

Table 4.3 KKW items referring to political systems with their % of each English text type sub-corpus

<i>Regime/s, government/s and administration of the main six text type sub-corpora</i>						
KKW	News			Editorials and opinions		
	AM	BR	GV	AM	BR	GV
regime	5.79	12.36	8.89	10.42	12.97	11.50
government	17.23	6.30	5.39	5.01	1.32	4.78
regimes				3.01	2.64	3.36
administration	3.13			1.80		2.12
governments	1.49			1.80	1.76	
Bold	Highest value of each sub-corpus					

As a general observation to make from the results shown in Table 4.3 is that the singular form of the KKW item *regime* is the most significant across the English text type sub-corpora with the highest value of 12.97% of the British editorials and opinions. On the other hand, the singular form *government* is the most significant with a percentage of 17.23 of the American news sub-corpus. Analysis showed that collocates of the items *regime* and *regimes*, are items belonging to other semantic categories, other than the GOVERNMENT AND SECURITY, (e.g. *President, military, forces* and *Zionist*), such as NAME-PA (e.g. *Assad's, Bashar, Gaddafi, and Mubarak*), GEOGRAPHY (*Syrian, Iranian, Arab* and *Middle East*), FAITH AND RELIGION (e.g. *Islamic*), PROTEST VOCABULARY (e.g. *protests, uprising* and *rebels*); TIME (e.g. *former, previous, current* and *old*); FUNCTION WORDS (e.g. *is, has, his, this, anti* and *pro*) and GENERAL AND ABSTRACT TERMS (e.g. *change*). Other collocates with more negative evaluative meanings, especially to the immediate left

(L1) to the KKW items regime/s include *oppressive*, *authoritarian*, *autocratic*, *despotic*, *repressive*, *brutal*, and *embattled*. For example, out of a total of 1,791 (that is, 3.99 ptw) instances of the term *regime* in the British news sub-corpus, there were 79 occurrences of collocate items (with ≥ 5 co-occurrences with ≥ 3 of MI), to the immediate left (L1), with negatively evaluative meanings, such as *anti* (37); *brutal* (10); *authoritarian* (8); *corrupt* (7); *military* (6), *autocratic* (6) and *embattled* (5). Figure 4.4 shows concordance examples of the items *regime/s* in the contexts of some of these negatively evaluative collocate items of the British news sub-corpus:

1 , since approving the plan 12 days ago the regime has merely intensified its brutal
 2 region where the overthrow of authoritarian regimes has been slow to usher in hoped-for
 3 the Ghouta suburb of Damascus. The Syrian regime has used military force to attack the
 4 , are concerned about the way repressive regimes have been falling in the Arab Spring.
 5 for many years with these deeply repressive regimes," he said. "The British Government
 6 a diplomat standing beside a brutal Iranian regime," he said. Mr Alizadeh said he was
 7 prepared to take action against oppressive regimes. He said the Arab Spring uprisings
 8 laws used to silence critics of a repressive regime, he was given an 18-year sentence for
 9 is not demonstrating against an authoritarian regime - her country is considered among the
 10 an effort to shake off repressive autocratic regimes. However, many argue that the unrest

Figure 4.4 Concordance examples of *regime/s* of the British news sub-corpus

As analysis in Chapter 3 showed, the items *regime/s*. like other items in other categories, are represented as news actors appearing in the agent role, especially if they are involved in negative actions, then these actions are emphasized (van Dijk, 1998), as Figure 4.4 (lines 1, 3 and 4) shows. As Figure 4.4 also shows, the geographical items *Syrian* (line 3) and *Iranian* (line 6) are used as pre-modifiers of *regime*, whereas the item *British* (line 5) pre-modifies the item *government*, which

indicates a kind of negative/positive evaluation that is actualized through the choice of lexical items (Dijk, 1998, p. 33). The KKW item *administration* is most relatively included in only three sub-corpora, as Table 4.3 shows, with its highest value of 3.13% of the American news sub-corpus. The most frequent lexical collocates of *administration* (876 concordance lines) of the American news include

Obama (346); *said* (134); *officials* (132), *official* (75); *senior* (70); *Bush* (34); *president* (27); *say* (22); *new* (22); *Clinton* (16), *military* (16); *Washington* (15); *security* (14); *called* (14); *Israel* (13); *announced* (12); *sought* (12); *Egypt* (12); *former* (12); *criticized* (10); *rights* (10), *support* (9); *sanctions* (9); *American* (9); *talks* (9); *Barack* (8); *weapons* (8); *speaking* (8) and *policy* (8).

As the above collocates suggest, *administration* is significantly associated with items related to GOVERNMENT AND SECURITY, NAMES-PN and LANGUAGE AND COMMUNICATION (e.g. *said*, *say*, *called*, *announced*, *criticized*, *talks* and *speaking*), which again confirms Bell's (1991, p. 190) notion regarding the news sources and actors, and that *who says?* is one of the most important questions of news work.

In terms of the Arabic sub-corpus, some the KKW items referring to different political systems, as Table 4.4 shows, can be seen as equivalents of those of the English sub-corpora, though in the definite forms, such as النظام <*the-regime*>, الحكومة <*the-government*>; and الأنظمة <*the-regimes*>; the Arabic KKW item الكيان <*the-entity*>; which is shared by the Arabic sub-corpus with highest value of 1.91% of the Arabic editorials and opinions sub-corpus, is almost exclusively used

in certain political and/or ideological contexts in which *Israeli* and *Zionist* are of its most frequent collocates, as discussed earlier in Chapter 3 (sub-section 3.5.3). As Table 4.4 also shows, the most included item is النظام <*the-regime*> which is shared by the two main Arabic sub-corpora with its highest value of 10.20% of the Arabic editorials and opinions sub-corpus:

Table 4.4 KKWs referring to political systems with their % the Arabic sub-corpora

KKW	Translation	Arabic sub-corpora	
		News	Editorials and opinions
النظام	<i>the-regime</i>	3.14	10.20
السلطة	<i>the-authority</i>	1.58	5.50
الأنظمة	<i>the-regimes</i>		2.86
الحكومة	<i>the-government</i>	2.36	
الكيان	<i>the-entity</i>	1.25	1.91
Bold	Highest value of each sub-corpus		

Both the KKW items النظام <*the-regime*>; and الأنظمة <*the-regimes*>, showed similar pattern to that of the English in only one sense as a system of government, or governments that have not been democratically elected. In the Arabic news sub-corpus, for example, there were 4,337 instances of the KKW item النظام <*the-regime*> of which quite a number of the top lexical collocates were used in the sense equivalent to that of the English item *the-regime*. Of these collocates, along with the total number of co-occurrences were

السوري <the-Syrian.MASC.ADJ – *the Syrian*> (663) ;
 السابق <the-former.MASC.ADJ – *the former*> (246) ;
 الشعب <*the-people*> (224) ;
 إسقاط <*overthrow*> (189) ;
 ضد <*against*> (183) ;
 سوريا <Syria> (134) ;
 يريد <want.PRST.SING.MASC.V - *want*> (120) ;
 الحاكم <the-ruling.MASC.ADJ – *the-ruling*> (101) ;
 المعارضة <the-opposition.FEM.N> (99) ; and
 الإيراني <the-Iranian.MASC.ADJ> (94).

Other senses of the KKW item النظام are equivalent to the English items *order* and *system* as in the phrases with collocates to the immediate right (R1) of النظام :

النظام السياسي <*the-political system*> (110)
 النظام العالمي الجديد <*the-news world order*> (37)
 النظام الانتخابي <*the-electoral system*> (36)
 النظام الدستوري <*the-constitutional order*> (8)
 النظام القضائي <*the-judicial system*> (12)
 النظام الرأسمالي <*the-capitalist system*> (10)

Additionally, the KKW item السلطة <*the-authority*> is also used in more than one sense. It is shared between the two main Arabic sub-corpora, with its highest value 5.50% of the Arabic editorials and opinions sub-corpus. Both the English and the Arabic sub-corpora share the sense of السلطة <*the-authority*> as a form of political system or government that mostly occurs in the context of السلطة الفلسطينية <*the-*

Palestinian Authority>; and *رئيس السلطة الفلسطينية محمود عباس* <*the Palestinian Authority President Mahmoud Abbas*>. In other contexts, *السلطة* means *the-power* as in the phrases

إلى السلطة في <*to the-power in*>;
السلطة في مصر <*the-power in Egypt*>;
الصراع على السلطة <*the-struggle for power*>; and
البقاء في القصر <*the-staying in power*>

Analysis also showed a predominance of military as well as security related issues, indicating the geostrategic significance of the Arab Spring at both the regional as well as the global levels, especially in contexts where prominent names of people (e.g. *Obama*) and/or organizations (e.g. *Security Council*) are involved. This might be one of reasons why the NAME-PN category is amongst the most key categories of the Arab Spring sub-corpora.

As so far indicated, this is not surprising since geopolitics, wars and conflicts, especially in the Middle East, are common areas of news media focus (Christensen and Christensen, 2013, p. 351; Al-Hejin, 2015, p. 27). Additionally, these three categories basically involve the most important constituents of the journalist's list, 'where', 'what' and 'who' through which many of the news values are also constructed (Bell, 1998, p. 74). What is also interesting is that the NAME-PN category is higher in rank than that of the PROTEST VOCABULARY when it comes to the total number of KKW's. This suggests that the Arab Spring is not only a narrative about

pro-democracy or *anti-government protests*, *demonstrations* or *uprisings* in various countries across the region, but also is a narrative about themes related to power and power relations with which items of the NAME-PN category are mostly associated. As Examples 25 (headline, lines 1, 5 and 7) and Example 26 (headline, lines 1 and 3) show, *Barak Obama*, *David Cameron*, *Muammar Gaddafi* and *Bashar al-Assad* (political and governmental leaders) as well as the *Security Council* (a powerful institution of the *United Nation*) and *NATO* (the *North Atlantic Treaty Organization*, which is also known as the North Atlantic Alliance) are instances of the significantly prominent names of people as well as organizations of the NAME-PN category.

25. [HEADLINE] Libya tops agenda as **Obama** plans to outline vision for Middle East
- 1 **David Cameron** will today urge **Barack Obama** to back a
 2 concerted attempt to end the stalemate in Libya over the next
 3 few weeks.
 4 During talks in Downing Street, the Prime Minister and US
 5 President are expected to agree to press other **Nato** countries to
 6 shoulder more of the burden of the military effort to drive
 7 **Muammar Gaddafi** from power.
 (NDPe-11-79 / news: 25-May)

26. [HEADLINE] **Obama** rejects Palestinians' U.N. bid
- 1 [...] He [Obama] called on the **Security Council** to
 2 follow the U.S. example and impose sanctions on the
 3 government of Syrian President **Bashar al-Assad**, whom
 4 Obama has said must step down.
 (WPe-11-124/ news: 22-Sep)

Likewise, Example 27 (headline, and line 1) shows names of powerful and elite people, such as أوباما <'ūbāmā – *Obama*>; as well as organizations: مجلس الأمن

<majlis al-'amn – *the Security Council*> that are also significantly salient in the Arabic corpus. Additionally, the three examples also contain KKW items of other key categories showing the interplay between them, and this influences the way different topics/events, as well as social actors are discursively constructed. For example, the KKW items of the GEOGRAPHY category, as indicated in Example 25 (headline and lines 2, 4, and 6), such as *Middle East, Libya, US and countries*; Example 26 (headline, lines 2, and 3), such as *U.S. and Syrian*, as well as the KKW item السوري <as-sūrī – *the-Syrian*.MASC> in Example 27 (headline, and line 3):

27. [HEADLINE] أوباما يدعو مجلس الأمن إلى «معاقبة» النظام السوري

*Obama calls on the Security Council to
“punish” the Syrian regime*

1 أوباما يدعو مجلس الأمن

'ūbāmā yad'ū majlis-a al-'amn-i

Obama calls(on) Council-ACC the-Security-GEN

Obama calls on the Security Council

2 إلى معاقبة

'ilā mu'āqabat-i

to.PREP punishment.FEM.N-GEN

to (the) “punishment” of

3 النظام السوري

an-niḏām as-sūrī

the-regime.MASC.N the-Syrian.MASC.ADJ

the Syrian regime (SFRa-11-163/ news: 22-Sep)

Similarly, the KKW items of the GOVERNMENT AND SECURITY category, such as those of Example 25 (lines 4, 5 and 6): *prime, minister, president and military*; Example 26 (line 3): *government and president*, and Example 27 (headline and line

3): النظام <an-niḏām – *the-regime*> are also used. Such KKW items, as analysis in Chapter 3 shows, are considered as linguistic indicators of the different discursive strategies, such as nomination/ referential as well as predication strategies to distinguish, define or qualify other socio-political events, processes and/or social actors (Reisigl and Wodak, 2009; Baker et al., 2008; Al-Hejin, 2015; Prentice et al., 2012).

A final observation regarding the key categories shown in Tables 4.1 and 4.2 is that different statistical measures and calculation methods are believed to ‘yield different results’ (Baker, 2006, p. 119). Looking at Table 4.1 from the perspective of the overall average percentage of the sub-corpora (visually represented in the green data bars) comprising the Arab Spring corpus as a whole, the rank of these categories changes, and hence a different picture is believed to be obtained. Showing the results in round brackets as (rank: overall average % of the sub-corpora), the FUNCTION WORDS category (1:10.88%), as expected, ranks as number one. According to Scott (2015, p. 236), function words ‘would not usually be identified by the reader as key’. They may be key indicators more of style than of "aboutness". The style of a text, according to Baker (2006, p. 127), ‘may play some role in the discourses within it’. Therefore, the fact that such grammatical words are KKW should prompt the researcher to investigate why such words have cropped up with unusual frequencies (Scott, 2015, p. 236). The FUNCTION WORDS is then followed by the GEOGRAPHY (2:4.83%), the PROTEST VOCABULARY (3:4.41%) as

well as the FAITH AND RELIGION (4:4.24%) categories that rank as number two, three and four respectively. These and other categories are briefly discussed below.

The FUNCTION WORDS category contains grammatical words such as prepositions (e.g. *against*), pronouns (e.g. *our*, *we*, *he*, *his*), determiners (e.g. *the*), conjunctions (Baker, 2006, p. 53), prefixes (e.g. *anti-*, *pro-* and *non-*), modality (e.g. *cannot*), or other grammatical items such as *its*, *will*, *has* along with ‘verb contractions and negative contractions’ such as *that’s*, *don’t* and *cannot* (Partington, 2010, p. 91). Unlike content words whose main function is to express meaning, function words, it is claimed, do ‘not always reveal much of interest, especially in terms of discourse’ (Baker, 2006, p. 100). They serve to define the grammatical relations between words within a sentence, a phrase or a cluster (Scott, 2015, p. 426). However, there is evidence, that function words can reveal much of interest about the nature of discourses and topics, especially in contexts of politically and ideologically oriented discourses, as in the case of the Arab Spring corpus. In this context, the FUNCTION WORDS category of the English sub-corpus, contains a total of 18 KKW items (*not*, *it’s*, *anti*, *don’t*, *the*, *its*, *will*, *has*, *our*, *that’s*, *cannot*, *against*, *he*, *we*, *his*, *non*, *pro-* and *is*). Only two items (i.e. *anti* and *cannot*) are shared between the main six English sub-corpora, but eight (*anti-*, *don’t*, *the*, *will*, *has*, *against*, *he* and *we*) are also collocates of the term *Arab Spring*. As mentioned in the introduction, some items (e.g. *against*, *pro-/anti-*; *is/non-* and *not*) of this category may indicate the contrastive aspects as well as the polarizing nature of the different topics and discourses of the Arab Spring narrative. Additionally, some of

these grammatical items, such as *against*, *pro-* and *anti-* can be, through their collocates, indicative of emotive or attitudinal stances (Duguid, 2010, p. 206). As their meanings also suggest, the KKW items *against* and *anti* can be indicative of discourses related to opposition, protest and disagreement. Similarly, the *pro-* prefix meaning suggests support to and/or advocacy of an idea, or other social or political activities or practices. These items (*against*, *anti* and *pro*) are therefore expected, as Duguid (2010, p. 195, 207) points out ‘to express aspects of subjectivity’, or to be indicators of stance due to the lexical choices (or collocates) that might be more or less evaluative. In terms of the Arab Spring corpus, for example, amongst the top twenty collocates with highest joint frequency (≥ 5), especially to the immediate right (R1) of each of the KKW items *anti-* and *pro-*, and average MI score ≥ 3.00 , in both the English news as well as the editorials and opinions sub-corpora, are items denoting different topics, such as the GOVERNMENT AND SECURITY:

anti-: government, corruption, regime, and terrorism, or
pro-: democracy, government, reform, regime and military;

or of the GEOGRAPHY category:

anti-: American, Israel, Western, Israeli, or
pro-: Western, Israel, Palestinian, American, Syrian and Iranian;

as well as with items of the FAITH AND RELIGION, which mainly refer to religious practices as well as other belief systems:

anti-Semitism, Semitic, Islam, Muslim, Americanism, and Islamic;

and finally items of the NAME-PN category, such as:

anti-: Assad;

pro-: Morsi, Assad, Gaddafi and Mubarak.

Similarly, in Arabic, the FUNCTION WORDS category contains an overall total number of 33 KKW grammatical items of which 12 are shared between its two main sub-corpora, and are also collocates of the term الربيع العربي <the-Arab Spring>; examples of these shared items are:

prepositions: على <on>; في <in>; عن <about>; and إلى <to>;

pronouns: هذه <this>FEM; and التي <which>FEM;

conjunctions: أو <or>; and إذا <if>; and

function verbs: كان <kāna – was> MASC.

Example 28 shows some of the most frequent phrases in which KKW items the above grammatical categories are associated with the term *Arab Spring*. the most frequent KKW item

28. 1 في دول الربيع العربي
 fī duwal arrabī' al'arabī
 in countries the-spring the-Arab
 in the countries of the Arab Spring
- 2 انتفاضات الربيع العربي التي
 'intifādāt arrabī' al'arabī **allatī**
 uprisings the-spring the-Arab **which.FEM**
 the uprisings of the Arab Spring which
- 3 أو ما يُعرف بـ
 'aw mā yu'raf bi
 or what (is) known as
 or what is known as
- 4 إذا كان الربيع العربي

idā if <i>if the Arab Spring was</i>	'kān-a was-PFV	ar-rabī' the-spring	al- 'arabī the-Arab
5 أن الربيع العربي كان	'ann-a	ar-rabī'	al- 'arabī
	that	the-spring	the-Arab
			kān-a was-PFV
			<i>that the Arab Spring was</i>

There are three KKW items that are not shared between the two main Arabic text type sub-corpora, but they are collocates of the term <الربيع العربي> *the Arab Spring*. The item <بين> *bayn - between* is a KKW item in the Arabic news, whereas the items <خصوصاً> *kuṣūṣān – especially*; and <ضد> *did - against* are KKWs in the Arabic editorials and opinion sub-corpus. Analysis shows that these items can reveal much about the nature of topics and discourses in which they occur. For example, as its meaning may indicate, the location adverb <بين> *bayn - between* is expected to express instances of contrast, difference, similarity, parallel, relation, connection, and involvement of people or things, as the headline in Example 29 indicates (in bold):

29. HEADLINE دراسة اسرائيلية: سقوط الأسد فرصة ممتازة لإعادة العلاقات الاستراتيجية بين تل أبيب وأنقرة

*Israeli study: The fall of Al-Assad is an excellent opportunity for the restoration of strategic **relations between** Tel Aviv and Ankara*

(QAa-11-129/ news: 23-October)

Similarly, the KKW item <did-against>ضد, which is only key in the Arabic editorials and opinions sub-corpus, but is also a collocate of the term الربيع العربي <the Arab Spring>, has almost the same functions as the English grammatical items *against* and *anti*. Figure 4.5 shows examples of the most and least frequent 3-word-clusters of the item <did-against>ضد taken from the Arabic editorials and opinions sub-corpus:

N	Cluster	English Translation	Freq.
1	ضد النظام السوري	against the Syrian regime	30
2	ضد نظام بشار	against Bashar's regime	14
3	ضد الشعب الفلسطيني	against the Palestinian people	12
4	ضد الجماعات الإسلامية	against the Islamist groups	10
5	الحرب ضد الارهاب	the war against terror	10
6	ضد قطاع غزة	against Gaza Strip	10
7	في الثورة ضد	in the revolution against	6
8	حرب ضد ايران	war against Iran	6
9	العنف ضد المتظاهرين	the violence against the demonstrators	6
10	العمليات العسكرية ضد	the military operations against	6
11	جرائم ضد الإنسانية	crimes against humanity	5
12	ضد المنشآت النووية	against the nuclear facilities	5
13	ضد الاستبداد والفساد	against tyranny and corruption	5
14	ضد الكيان الصهيوني	against the Zionist entity	5
15	ضد المتظاهرين السلميين	against the peaceful demonstrators	5

Figure 4.5 Top 3-word-clusters of <against>ضد in the Arabic editorials and opinion sub-corpus

Examining frequent clusters that are defined by Scott (2015, p. 448) as ‘words which are found repeatedly together in each others' company, in sequence’ may therefore be more revealing of the contexts of a particular word under analysis

(Baker, 2006, p. 68); ‘effectively, they are a kind of extended collocation’ (Partington and Morley, 2002, p. 1; Partington, 2006, p. 266). The Arabic examples of the 3-word clusters of the item **ضد**<did-against> in Figure 4.5 show a similar pattern to that of the English grammatical items *anti-* and *pro-* in terms of its associations with other items of other key categories. They also show how the different key categories might generally interact within the ASC corpus. For example, the 3-word cluster **ضد الجماعات الإسلامية**<against the Islamic groups> (see Figure 4.5, line 4) includes two items, that is **الجماعات**<the-groups>; and **الإسلامية**<the-Islamic.FEM> from two different categories. The item **الجماعات** <al-jamā‘āt - the-groups> is the plural form of the KKW item **الجماعة** <al-jamā‘a – the-group> from the SOCIAL STATES, ACTIONS AND PROCESSES category, which contains terms referring to people in general: e.g. **المرأة**<the-woman>; and **الشعب**<the-people>; groups and affiliations: e.g. **الحركة**<the-movement>; power relationship: e.g. **القوى**<the-powers>; as well as other social actions, states and practices such as:

المعارضة<the-opposition>; **الاستبداد**<the-repression>; **المقاومة**<the-resistance>; and **المصالحة**<the-reconciliation>.

The second item, **الإسلامية** <al-‘islāmiyya – the- Islamist.FEM>, is one out of the ten KKW items that comprise the FAITH AND RELIGION category of the Arabic sub-corpus. The FAITH AND RELIGION category basically refer to religions, most specifically the religion of **الإسلام** <Islam>, as well as other belief systems and ideologies. Five of these ten items (in bold as shown below) are shared between the

two main Arabic sub-corpora and are also collocates of the term *الربيع العربي* <the Arab Spring>:

- الإسلامية <the-Islamic.FEM>;
 الإسلاميين <the-Islamists.MASC.GEN>;
 الطائفية <the-sectarianism.FEM>;
 الإسلام <Islam>;
 المسلمون (الإخوان) <the-Muslims.PL.GEN> (e.g. *the Muslim Brotherhood*)
 الطائفي <the-sectarian.MASC>;
 الإسلامي <the-Islamic.MASC>;
 العلمانية <the-secularism.FEM>;
 السلفية <the Salafist.FEM>; and
 الإسلاميون <the-Islamists.MASC.NOM>.

As for the 3-word cluster *العنف ضد المتظاهرين* <violence against the demonstrators>, the two KKW items *العنف* <the-violence>; and *المتظاهرين* <the-demonstrators> belong to the EMOTIONS as well as the PROTEST VOCABULARY categories respectively:

العنف	ضد	المتظاهرين
al-‘unf	ḍid	al-mutaḍāhir-īn
the-violence.MASC.N	against	the-demonstrator.MAS.N-GEN.PL
<i>violence against the demonstrator</i>		

In this context, the PROTEST VOCABULARY category contains items that reflect expressions of strong disagreement or opposition to something, either by means of a statement or action. Looking at Table 4.2 again, the English as well as the Arabic

sub-corpora are identical in terms of the ‘actual’ total number of KKW items (12 items each), but are relatively different in terms of their overall average percentage of the sub-corpora. As Table 4.1 shows, the American news sub-corpus has the highest value of 7.00), followed by the Arabic editorials and opinions sub-corpus (5.48), whereas the English GV editorials and opinions sub-corpus has the lowest value of 2.69. The relative significance of the PROTEST VOCABULARY could be accounted for due to the fact that the resulting KKW items of this category were used in the query to collect data (see section 2.5 for more information). In the English sub-corpus, seven (e.g. *protests*, *protesters*, *revolution*, *spring*, *rebels*, *uprising* and *protest*) out of 12 KKW items are shared between its six main sub-corpora, and are also collocates of the term *Arab Spring*. The remaining items (e.g. *demonstrators*, *uprisings*, *demonstrations*, *revolutions* and *rebel*) are key either in four or three sub-corpora. The items *uprisings* and *revolutions* are also collocates of the term *Arab Spring*. Similarly, in Arabic, five out of the 12 KKW items are shared between the two main Arabic sub-corpora, and are also collocates of the term *الربيع العربي* <*the-Arab Spring*>. These are:

الربيع <*the-spring*>; *الثورة* <*the-revolution*>; *الثورات* <*the-revolutions*>;
ثورات <*revolutions*>; and *الاحتجاجات* <*the-protests*>;

along with the item *احتجاجات* <*protests*>, which is only key in the Arabic news sub-corpus that is also collocate of the term *الربيع العربي* <*the-Arab Spring*>. The remaining items are KKW's either in the editorials and opinions sub-corpus, such as

ثورة<revolution>; الانتفاضة<the-uprising>; الانتفاضات<the-uprisings>;
and الثوار<the-rebels.MASC>;

or in the Arabic news sub-corpus, such as the items

المتظاهرين<the-demonstrators.MASC.GEN>; and المحتجين<the-protesters.MASC.GEN>.

As can be seen, the resulting KKW items of the PROTEST VOCABULARY category may suggest a difference of focus in terms of the two main Arabic sub-corpora, as the two items المتظاهرين<the-demonstrators>; and المحتجين<the-protesters> are only KKWs in the Arabic news sub-corpus, and can be considered as the main news actors in the Arabic PROTEST VOCABULARY category. The item المتظاهرين<the-demonstrators> is more significant with a value of 2.52% of the Arabic news than that of the term المحتجين<the-protesters>; with a value of 1.24% of the same sub-corpus. Despite their difference in significance, collocational analysis of the two terms showed that both KKW items, المتظاهرين<the-demonstrators>; and المحتجين<the-protesters> ((in 1,112 and 574 concordance lines respectively); share quite a number of collocates. However, analysis showed that the item المتظاهرين<the-demonstrators>, has more collocates (that is, 297) than المحتجين<the-protesters> (that is 117 collocates) with joint frequency (≥ 5) and a minimum MI score of 3. This is because the KKW item المتظاهرين<the-demonstrators> occurs more frequently in the Arabic news sub-corpus. Some of the shared collocates between these two items are

الشرطة<the-police>; الأمن<the-security>; قتل<killing>; مئات<hundreds>;
 قوات<forces>; العنف<the-violence>; الغاز<the-gas>; الشوارع<the-streets>;
 الغاضبين<the-angry.MAC.PL>; and النظام<the-regime>; but not the items
 الحكومة<the-government> and السلميين<the-peaceful.MASC.PL.ADJ>; as they are only
 collocates of المتظاهرين<the-demonstrators>.

As indicated by the collocate examples above, the items with which the news actors (see sub-section 3.2.2) in the PROTEST VOCABULARY category are mostly associated belong to contexts of GOVERNMENT AND SECURITY, EMOTIONS, GEOGRAPHY, LIFE/DEATH, AND NUMBER AND MEASUREMENT categories. Examples 7 (lines 1 and 2) and 8 (line 7) represent the nature of contexts in which the two KKW items المتظاهرين<the-demonstrators> and المحتجين<the-protesters> generally occur in the Arabic news sub-corpus. For example, in Example 7 (lines 1 and 2), as it appears in a contexts which can be described as negative, indicating that KKW items of the protest vocabulary are mainly associated with negativity. the most important news values, for an event to make it into the news.

7. HEADLINE 1 جيش مصر ينفي العنف ضد المتظاهرين
 2 Egypt's **army denies violence against**
 3 **the demonstrators**
 (JAZa-11-108 / news: 23-July)

8. HEADLINE 1 25 قتيلا بسوريا في احتجاجات جديدة اجبها مقتل القذافي
 2 25 killed in Syria in protests fueled by
 3 Gaddafi's death
- LEAD 4 قال ناشطون ان مقتل الزعيم الليبي المخلوع معمر القذافي
 5 أيج مظاهرات مناهضة للحكومة في أنحاء سوريا بعد
 6 صلاة الجمعة وان قوات الامن قتلت 25 شخصا في حملة
 7 متواصلة ضد المحتجين الذين يطالبون بالإطاحة بالرئيس
 8 السوري بشار الأسد.
 9 Activists said that death of the ousted Libyan
 10 leader Muammar Qaddafi fueled anti-
 11 government **demonstrations** across Syria
 12 after Friday prayer and that the security
 13 forces killed 25 people in a continuing
 14 campaign **against the protesters** who demand
 15 the overthrow of the Syrian President Bashar
 16 Al-Assad. (BWBa-11- /news: 21-Oct)

In terms of the KKW items *الانتفاضة* <the-uprising>, and its plural form *الانتفاضات* <the-uprisings>, their keyness in the Arabic editorials and opinion sub-corpus can mainly be attributed to the difference in size between the two main text type sub-corpora. In fact, the items *الانتفاضة* <the-uprising>; and *الانتفاضات* <the-uprisings>; were also KKW in the Arabic news, but their keyness did not reach the cut-off point – that is, $\geq 1.00\%$ of each sub-corpus, set for the analysis. Nevertheless, the item *الانتفاضة* <al-'intifāda - the-uprising>, along with its other forms, is considered as a key concept in Arabic, as it mainly refers to the legitimate uprising against oppression and tyranny. Additionally, only in its indefinite singular/dual/plural forms can the item *انتفاضة* <uprising>, as Figure 4.5 shows, be used in association with the term *الربيع العربي* <the Arab Spring>.

2-5-word-clusters of انتفاضة <uprising> forms in the Arabic editorials and opinion sub-corpus

N	Cluster	Translation	Freq
1	الانتفاضات العربية	<i>the-Arab uprisings</i>	82
2	الربيع العربي	<i>the-Arab Spring</i>	53
3	الانتفاضات الشعبية	<i>the-popular uprisings</i>	53
4	الانتفاضة الشعبية	<i>the-popular uprising</i>	52
5	الانتفاضة السورية	<i>the-Syrian uprising</i>	44
6	هذه الانتفاضات	<i>these uprisings</i>	38
7	انتفاضات الربيع	<i>uprisings of the-Spring</i>	33
8	انتفاضة شعبية	<i>popular uprising</i>	32
9	انتفاضات الربيع العربي	<i>uprisings of the-Arab Spring</i>	32
10	انتفاضة الأقصى	<i>Al-Aqsa uprising</i>	30

concordance collocates plot patterns clusters timeline

200 entries Row 1 T S 5

Figure 4.6 Examples of most frequent 2-5-word clusters of انتفاضة <uprising> forms in the Arabic editorials and opinions sub-corpus

However, such items were not the only terms that have been used to refer to the various events of the Arab Spring, but items of other categories, such as the EMOTIONS category, have also been used. Table 4.5 shows the resulting KKW items in the EMOTION category along with the percentage of each of the English and Arabic sub-corpora. In the Column Total, Table 4.5 also shows the average percentage of the KKW items the EMOTIONS category of each sub-corpus:

Table 4.5 KKW items of the EMOTION category with their % of each of English and Arabic text type sub-corpora

KKW item in EMOTIONS category	KKW items of emotions category and their % of each of English and Arabic text types sub-corpora							
	News				Editorials and opinions			
	AM	BR	GV	MSA	AM	BR	GV	MSA
<i>peace</i>	6.73	6.38	8.36		8.62	4.62	19.65	
<i>violence</i>	4.46	3.39	4.04		1.60	3.08	1.77	
<i>clashes</i>	2.82	1.70	2.96					
<i>unrest</i>	2.27	2.42	1.89				1.06	
<i>peaceful</i>	1.25	1.21			1.80	2.20		
<i>terror</i>		1.05				1.54	1.77	
<i>threat</i>		1.05			1.00	1.10		
العنف< <i>the-violence</i> >				2.15				1.95
السلام< <i>the-peace</i> >				1.24				1.11
الشعبية< <i>the-popular</i> >								1.02
Colum Totals, avg. % of sub-corpus	3.51	2.46	4.31	1.70	3.26	2.51	6.06	1.36

Although as discussed earlier, in sub-section 2.3.2, that keyness cannot be compared directly, several important observations can still be made from the figures shown in Table 4.4. One observation is when looking vertically at the figures in each of the eight sub-corpora, it can be said that *peace* and peace-related topics are apparently more significant than *violence* across the English sub-corpora, with the highest scores, 19.65 and 8.62, in the Global Variety and the American editorials and opinion sub-corpora respectively. Another point is that, unlike English, both

Arabic sub-corpora show more significance in terms of the KKW item *العنف* <the-violence> than that of the item *السلام* <the-peace>, with the highest value of 2.15% of the Arabic news sub-corpus. As in other semantic categories across the English and Arabic sub-corpora, the presence of the KKW items *peace* and *violence* reflects the contrastive themes and concepts in the Arab Spring narrative. Not only are *peace* and *violence* different in terms of their meanings and connotations, but also in terms of their associated topics.

Before getting into the differences/similarities between the different corpora, it is worth discussing, though briefly, two main and general observation made by the analysis. First, almost all sub-corpora showed similar pattern in terms of the *news actors* associated with the two items *peace* (السلام), and *violence* (العنف). In terms of *peace*, it is mostly associated with high-profile and elite news actors (e.g. *Obama* (أوباما); *Netanyahu* (نتنياهو); *Abbas* (عباس); *Clinton* (كلينتون); *the Security Council* (مجلس الأمن); *US / America/ Washington* (الولايات المتحدة/ أميركا/ واشنطن); and *Israel* (إسرائيل)), who, in various contexts, are mostly constructed as promoters and supporters of *peace*, as Figure 4.7 (lines 2, 3, 4, 5 and 7) shows:

1 tween settlement activity and the path of *peace*". He added: "If (Netanyahu) decides
2 believe Prime Minister Netanyahu wants *peace*. I think he is willing to take risks for
3 was convinced that Mr Netanyahu "wants *peace*" in the Middle East and was "willing
4 told Israelis they had a "true partner" for *peace* in Mr Abbas. In a speech to students
5 working to achieve a just and lasting *peace* in the region," Clinton said. The US
6 nsurgency and posing a threat to regional *peace*. In Washington, Barack Obama, the
7 the TNC asking UN forces to keep the *peace* in urban areas until a new leadership

Figure 4.7 Concordance examples of *peace* from the British news sub-corpus

Peace almost exclusively appears in contexts related to the *peace process in the Middle East, peace talks, peace negotiations, the Arab Peace Initiative, and Peace Treaty*, but also to topics related to *security*. *Violence*, on the other hand, is mostly associated with ‘non-elite’ or ‘ordinary’ news actors. However, elite news actors (e.g. *Obama* and *Netanyahu*, *UN*, *Security Council*; *officials* and *activists*) are also associated with *violence*, but appear in contexts that either condemning *violence*, or urging other news actors (*people* or *organizations/ groups*) to renounce/stop *violence*, as Examples 9 (lines 1 and 2) and 10 (lines 1 and 2) shows:

9. 1 *US President Barack Obama* **condemned**
 2 **violence** against Tunisian citizens "peacefully
 3 voicing their opinion in Tunisia".
 (BBCe-11-07/ news: 15-January)

10. 1 On Friday, *Barack Obama* **urged** *Assad* to
 2 **stop** the "outrageous use of **violence** to quell
 3 protests". (GRDe-11-34/ news: 25-April)

Thus, amongst those ‘ordinary’ or ‘non-elite’ news actors, who are mostly found in violence-related contexts are *people* (الشعب); *civilians* (المدنيين); *citizens* (المواطنين); *protesters* (المحتجين); *demonstrators* (المتظاهرين); *women* (المرأة); and *children* (الأطفال). According to Bell (1991, p. 194), elitenss of news actors is one of the most important news values, and ‘for those who are not elite, the surest way is to be a victim of crime, accident or disaster’, to get into the news. *Violence* is then one of the major topics through which *Negativity* is actualized in the Arab Spring narrative. In terms of topics, the KKW item *violence* showed a multiplicity of topics and

contexts, belonging to different semantic categories, such as GEOGRAPHY; GOVERNMENT AND POLITICS; NAME-PN; FAITH AND RELIGION; PROTEST VOCABULARY; SOCIAL ACTIONS; STATES AND PROCESS; LANGUAGE AND COMMUNICATION and EMOTION.

As discussed above, despite similarities between the different English and Arabic sub-corpora in terms of topics and/or news actors, and the nature of such topics that are associated with either *peace* or *violence*, there was a difference of focus, especially before (that is, between 15 June and 31 December 2010) and after (that is, between 1 January 2011 and 31 August 2013) the emergence of the Arab Spring, between the different sub-corpora. However, analysis showed that difference of focus between the sub-corpora was in terms of the topics associated with *violence* rather than *peace*. For example, most of the English and Arabic sub-corpora showed a steady pattern of topics associated with *peace* before and after the Arab Spring. That is the *Middle East peace process*, *peace talks* and *peace negotiations* between the *Israelis* and the *Palestinians* and *peace*, *security* and *stability* in the region. In terms of *violence*, however, there was a significant shift of focus within and between the different sub-corpora before and after the Arab Spring. Before the Arab Spring for example, both the American and the British news sub-corpora showed that the *violence in Iraq*, and *sectarian violence*, and violence between the *Israelis* and the *Palestinians*. However, the British news also showed more focus on the elections (in Egypt), as the item election (7 co-occurrences) was amongst the top collocates of *violence* (398 instances), after *Iraq* (32) and *instability* (26), in 2010.

Another example is that the GV news sub-corpus did not show any significance in terms of topics associated with *violence*. This is partly because of the size of the GV-ENASC news (2010) as it only consists of the *Jerusalem Post* newspaper. However, the GV editorials and opinions sub-corpus showed focus on the *Palestinian violence against Israel, terrorism and peace*. Finally, the Arabic news sub-corpus also showed focus on *violence* in العراق<Iraq> (10 co-occurrences); الطائفي<the-sectarian violence>; and الانتخابات<the-elections>; but also الاحتجاجات<the-protests> (5); and أعمال العنف<acts/actions of violence> (18); and الشغب<the-riot> (10); in Tunisia were also significantly prominent as this event was the catalyst of the Arab Spring.

After the Arab Spring, some of the most associated topics with *violence* included the events in *Syria, protests and demonstrations, violence against civilians, women and sexual harassment and sectarian conflicts*, which was shared by most of the sub-corpora. For example, out of 421 instances in the American news in 2011, the first top lexical collocate was *Syria* (25 co-occurrences); *government* (15); *stop* (13); *protesters* (10), *protest* (10); *renounce* (10); *Syrian* (10); *women* (9), whereas in the American editorials and opinion sub-corpus of the same year, there was more focus on *sectarian* (5 co-occurrences out of 109 instances of *violence*) conflict in the region, but also on *violence against* (7) *Israel, protesters* and the *Syrian civilians*.

Additionally, the appearance of the two KKW items *terror* and *threat* indicates that there are security concerns, at the national, regional as well as the international levels, that are significant in the English sub-corpus. This may also be the reason behind the appearance of quite a number of military and security related KKW items, such as *weapons*, *security*, *nuclear*, *armed*, *war*, *attacks*, *terrorist*, *terrorism*, *tear*, *forces* and *intelligence*, in the GOVERNMENT AND SECURITY category of the English sub-corpora. Figure 4.8 shows examples of concordance lines of the KKW item *terror* from the British news sub-corpus, representing some of the contexts in which the term is used:

```

1 Jihadists using British website to spread terror, think tank claims MIDDLE
2 Secretary Theresa May said the UK's terror threat level would stay at
3 may have been work of 'self-starters' Terror threat level remains substantial,
4 to fight in Somalia and would pose a terror threat if they decided to return
5 Syria chemical weapons 'most worrying terror threat to UK' Al-Qaeda could
6 reports The Arab Spring poses a new terror threat to Britain by creating an
7 MI5 head; Security chief warns of new terror threat and rising cost of

```

Figure 4.8 Examples of concordance lines of *terror* in the British news sub-corpus

Looking at Table 4.5 again, the highest scores 1.77 and 1.05 for the KKW items *terror* and *threat* are in the GV-ENASC editorials and opinion, and the British news sub-corpora respectively. Analysis of the concordances of these two KKW items showed difference in terms of their associated topics. In the British news sub-corpus, for example, 23 out of 189 of the instances of the KKW item *terror* appeared in contexts related to the *war on terror* discourse, and 5 instances of September 11

terror attacks, but also in contexts related to *threats* of *terror attacks* by *Al-Qaeda*, as indicated in Figure 4.8 (line 5), and other *groups*, *organizations* or *networks* as these italicized items are also collocates (with min. frequency 5 and $MI \geq 3$) of the term *terror* in the British news sub-corpus. In the GV-ENASC editorials and opinion sub-corpus, on the other hand, there are 224 instances of the item *terror*; the top lexical collocate in terms of total co-occurrences is the item *attacks* (27), followed by *war* (21), *groups* (14), *Hamas* (14) and *Israel* (13); *organizations* (11); *violence* (11); *Gaza* (8); *Islamic* (8); *Qaeda* (7) *Palestinian* (7), *Hezbollah* (6) and *Iranian* (5), suggesting more focus, especially in the Israeli *Jerusalem Post* (JPTe) newspaper (with 221 of the instances, and 3 instances in the Turkish ZMNe news), on issues relating to *terror attacks* by *Hamas* and *Hezbollah*, *violence*, and *nuclear threat* from *Iran*. In the GV-news sub-corpus, however, there are a total of 72 instances of *terror* of which 5 main collocates, *attack* (10); *against* (10); *attacks* (7); *Hamas* (7) and *war* (7) are the most frequent collocates, which indicate similar focus on topics related to the Israeli-Hamas conflict along with topics related to the war on terror discourse.

In terms of the Arabic sub-corpus, although neither الإرهاب <al-'irhāb – (the)*terror/ism*> nor التهديد <at-tahdīd – (the)*threat*> are KKW items, the item الإرهاب <wa-l-'irhāb – *and-the-terror/terrorism*> is the most frequent collocate of العنف <the-violence>, with 45 co-occurrences out of 1,005 instance of العنف in the Arabic editorials and opinions sub-corpus. The most second frequent 3-word cluster

in the Arabic editorials and opinions sub-corpus also contains the item والإرهاب as in the phrase

أخبار العنف والإرهاب <the violence and the terror/ism news>

The item والإرهاب <and-the-terror/ism> is also collocate of the KKW item العنف <the-violence> in the Arabic news sub-corpus, with 13 co-occurrences out of 1,855 instances of العنف <the-violence>; but not of the most twenty frequent collocates of العنف, such as أعمال <acts/actions> (228); ضد <against> (127); سوريا <Syria> (89); وقف <stop/stopping> (80); and استخدام <use.N> (53).

Figure 4.9 shows examples of key items of the EMOTIONS category, taken from the English news during May 2011. It contains items denoting strong feelings or referring to other emotional states, actions and processes.

ruptions in supply tied to unrest in Libya and Bahrain for the
 an 230 injured in sectarian clashes in Cairo as fears grow that
 ad been kept up despite the violence in Bahrain. "At the very le
 f the fighting in Libya and unrest in Yemen. Daniel Benjamin, co
 down like rats". As for the unrest in Syria, he stopped short of
 as soon as you can" as the violence in Syria increased. "Her fir
 on and called for an end to violence in Syria, Libya and Yemen.

Figure 4.9 Concordance examples of KKW items of the English emotions category

Most items in this category have negative meanings, such as the English KKW items *violence*, *unrest*, *clashes*, *terror* and *threat* as well as the Arabic item العنف

<al-‘unf – *the-violence*>. It is worth mentioning, however, that the categorization of the items in the English EMOTION category was also informed by the web-based corpus software tool, *Wmatrix* (Rayson, 2008). The results were also informed by consulting the *OED* (2013, online), which showed that some of the meanings and thesaurus entries of the aforementioned items are believed to belong to the EMOTIONS category, especially that they carry in their meanings negativity as well as the lack of peacefulness. However, items with positive meanings with more of an evaluative function (e.g. *peace*, *peaceful* in the English sub-corpus; and السلام <*the-peace*> as well as الشعبية <*the-popular.FEM.ADJ*> in its Arabic counterpart) are also included, as they are used, in the form of modifiers (these are usually placed after the word/ noun phrase in Arabic) of other items (or topics). For example, in the English sub-corpus, the item *peaceful* pre-modifies, or often collocates with, other (KKW) items (e.g. *peaceful change* or *peaceful transition*) from other categories. For example, it occurs in the GENERAL AND ABSTRACT TERMS category, which refer to general ideas or concepts (e.g. *conflict*, *intervention*, *freedom* and *transition*). However, the item *peace* is mostly associated with other key terms of the categories of GEOGRAPHY (e.g. *Israel/i*, *Palestinian/s*, *Middle*, *East*, and *Arab*), the GOVERNMENT AND SECURITY (e.g. *security*), the NAME-PN (e.g. *Hamas*, *Netanyahu* and *Abbas*) as well as the LANGUAGE AND COMMUNICATION category (e.g. *talks*, *negotiations*, *said*, *resolution* and *speech*), which contains expressions referring to linguistic actions, states and processes as well as other forms/ means of communication (e.g. *media*, *internet*, and *journalists*).

Likewise, in the Arabic sub-corpus, the item *الشعبية* <the popular> post-modifies items of the PROTEST VOCABULARY category, such as in the term *الثورة الشعبية* <the-popular revolution>; and *الانتفاضة الشعبية* <the-popular uprising>; or items from the SOCIAL ACTIONS, STATES AND PROCESSES category as in the phrase *الحركة/ الحركات* <the-popular movement/s>. As for the item *السلام* <as-salām – the-peace>, analysis shows a similar pattern to that of its English equivalent, *peace*, in terms of its frequent association with KKW items of different key categories, such as

geography:	<i>الشرق الأوسط</i> <the-Middle East>; <i>إسرائيل</i> <Israel>;
GOVERNMENT AND SECURITY:	<i>الاستيطان</i> <the-settlement>; <i>الأمن</i> <the-security>;
NAME-PN:	<i>نتنياهو</i> <Netanyahu>; <i>أوباما</i> <Obama>; and finally
LANGUAGE AND COMMUNICATION:	<i>الحل</i> <the-solution>; <i>وقال</i> <and-said.MASC>; <i>الحوار</i> <the-dialogue>; <i>المفاوضات</i> <the-negotiations> and <i>التقرير</i> <the-report>.

Example 30 (in the Lead) shows how the KKW item *السلام* <the-peace>, is associated with other KKW items (in bold) of the NAME-PN category: *نتنياهو* <Netanyahu>; and GEOGRAPHY, such as the item *الشرق الأوسط* <the Middle East>:

30. HEADLINE

كلنتون ستطمئن إسرائيل بشأن مصر

Clinton to reassure Israel over Egypt

LEAD

وصلت وزيرة الخارجية الأميركية هيلاري كلينتون إلى إسرائيل مساء الأحد بعد زيارة لمصر، ومن المرتقب أن تلتقي رئيس الوزراء الإسرائيلي بنيامين نتنياهو لبحث عملية السلام في الشرق الأوسط وقضايا إقليمية أخرى،

US Secretary of State Hillary Clinton arrived in Israel Sunday evening after a visit to Egypt, and expected to meet with Israeli Prime Minister Benjamin Netanyahu to discuss the peace process in the Middle East and other regional issues.

(JAZa-12-324 /news: 25-Aug)

Having considered the overall average percentage of the sub-corpora, it can then be said that civil resistance, revolutionary change as well as ideological conflicts are themes that are significantly prominent in the Arab Spring narrative. The examples above show how the different key categories interrelate with each other, and are highlighted through the different CL techniques, such as the concordance lines, the collocates or the word clusters. In both cases, however, the results shown in Tables 4.1 and 4.2, be it the total number of KKW's, or the overall percentage of the sub-corpora, give an overview of the main topics that have mostly attracted the attention of the news media during the time specified for this study (from 15 June 2010 to 31 August 2013).

Looking at the Arab Spring from the perspective of the EMOTIONS category and its associated topics, it can be said that the Arab Spring is a narrative in which various contrastive themes and topics do exist. While some consider it as a chance for real

peace and stability in the Middle East, others see it as a threat to international *peace* and *security* (see Figure 4.8, line 7). The EMOTIONS category revealed differences in the way news actors are represented in contexts of *peace* and *violence*. It also showed, as also discussed in Chapter 3, how the elite and the powerful are represented as promoters of *peace*, whereas the ‘ordinary’ and ‘non-elite’ news actors are mostly represented as the victims (or witnesses) suffering the acts of *violence*. Finally, *peace* and *violence* showed differences between each other, not only in terms of their associated news actors, but also in the extent of topics and their nature before and after the emergence of the Arab Spring.

In order to examine potential changes, synchronically as well as diachronically in the way the Arab Spring is represented in terms of its associated topics/events and news actors across the different sub-corpora, the next step in the analysis is to compare the difference in the frequency of the term *Arab Spring* since its first appearance in January 2011 until 31 August 2013 in the English as well as the Arabic sub-corpora. In terms of the English as well as the Arabic sub-corpora covering the period (15 June – 31 December 2010) before the emergence of the Arab Spring, these were examined based on the number of articles per month in both sub-corpora.

4.3 Arab Spring: A Diachronic Perspective on Topics and Events

Taken collectively, both the English as well as the Arabic sub-corpora have been sub-divided into 39 monthly sub-corpora each, to examine change over time of the main topics before and after the emergence of the Arab Spring on 17 December 2010. The analysis begins by discussing the peak events before the emergence of the Arab Spring, and covers a period of seven months (between 15 June and 31 December 2010), while the events/topics that peaked with the term *Arab Spring* covering a period of 32 months between January 2011 and August 2013. In terms of the period between 15 June and 31 December 2010, there are 10 out of 13 peak events/topics, as Figure 4.6 shows, that are shared between both the English as well as the Arabic sub-corpora. The peak events/topics are mainly distributed across the months of July, September and December 2010.

As Figure 4.10 also shows, there are five peak events/topics during July 2010 of which four are shared between the English as well as the Arabic sub-corpora. The most salient topic during the aforementioned month was the death of the prominent Lebanese Shi'ite cleric Fadlallah (4 July 2010). What is perhaps interesting is the fact that both the years 2010 (before) and 2012 (after the Arab Spring), show a similar pattern in terms of the nature of one of their events/topics during September. Both years 2010 and 2012 witnessed anti-Islam events/topics during the month of

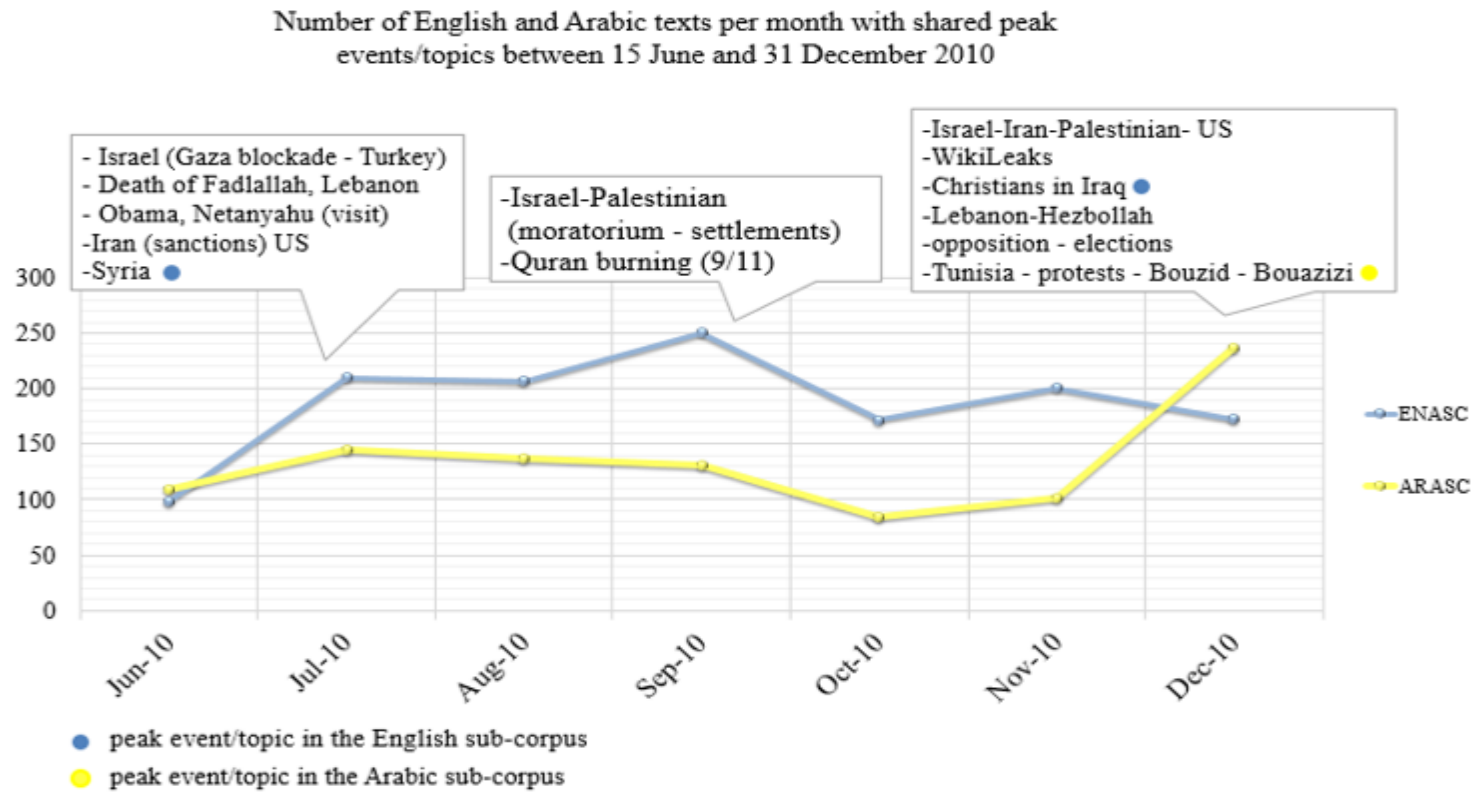


Figure 4.10 Peak events before the Arab Spring (15 June-31 Dec 2010)

concordance lines of the KKW item *Christians* in the British sub-corpus during December 2010

N	Concordance	%	File
1	Security fears mar Christmas for Iraqi <i>Christians</i> Many church services in Iraq	2%	BBCe-10-278.tx
2	Iraqi <i>Christians</i> flee Baghdad after cathedral	3%	GRDe-10-83.txt
3	Mubarak has urged Egypt's Muslims and <i>Christians</i> to stand united against terrorism	4%	BBCe-10-290.tx
4	amid fears for the safety of the country's <i>Christians</i> . In October, a suicide attack on	5%	BBCe-10-278.tx
5	. Church leaders have now warned <i>Christians</i> to keep celebrations to a	7%	BBCe-10-278.tx
6	al-Qaida drive to cause chaos Thousands of <i>Christians</i> have been forced to flee in	9%	GRDe-10-83.txt
7	two men and arresting more than 160 <i>Christians</i> who rioted after they were	10%	FTe-10-114.txt
8	of worship Freedom of worship for <i>Christians</i> varies greatly across the Middle	11%	GRDe-10-86.txt
9	to destabilise Egypt. Several hundred <i>Christians</i> later clashed nearby with	11%	BBCe-10-290.tx
10	handheld rocks upon thousands of Coptic <i>Christians</i> demonstrating below. "Imagine	12%	GRDe-10-87.txt
11	UN for help. The mass movement of Iraq's <i>Christians</i> , the remnants of which make up	24%	GRDe-10-83.txt
12	in the Lebanese parliament. In Jordan, <i>Christians</i> are free to profess their faith,	25%	GRDe-10-86.txt
13	its members. "Sectarian polarisation of <i>Christians</i> and Muslims stretches back	30%	GRDe-10-87.txt
14	than concrete barriers to persuade many <i>Christians</i> that they are still safe in Iraq.	34%	BBCe-10-278.tx
15	tensions between Muslims and Egypt's <i>Christians</i> , who make up ten per cent of	37%	TMSe-10-73.txt

concordance collocates plot patterns clusters timeline filenames source text notes

45 entries Row 1 T S

Figure 4.11 concordance lines of the KKW items Christians from the British news in December 2010

September, which coincided with the 9/11 anniversary of the terrorist attacks (2001) on the Twin Towers in New York, and the Pentagon in Washington D.C. The first event (of September 2010) was the issue of the Quran burning in the United States, whereas the second event of the same nature was the anti-Islam film that was produced in the United States, and published online in early September 2012. Another event in 2010 that is also related to the FAITH AND RELIGION category, is the issue of the *Iraqi Christians* (18 December 2010), who fled Baghdad, apparently after the siege at a Catholic cathedral, in Baghdad, back on 31 October 2010. As indicated by the blue circle in Figure 4.10 (peaks of December 2010), the issue of the *Iraqi Christians* was only key in the British sub-corpus. However, comparison of the KKW lists of the three main English (AM, BR, and GV) sub-corpora, showed that the aforementioned event/topic was only a KKW in the British sub-corpus and not in the American or GV sub-corpora. Concordance analysis of the term *Christians* showed that the *Coptic Christians*, in Egypt, were also involved in the coverage during December 2010 that immediately preceded the actual emergence of the term *Arab Spring*. Figure 4.11 shows 15 out of a total of 45 occurrences (in 8 texts) of the KKW item *Christians* from the British sub-corpus during December 2010. The saliency of religiously related topics suggests that the Arab Spring arose out of religiously polarised and catalytic events that demanded action. Another important topic that is also shared between both the English as well as the Arabic sub-corpora is the issue of WikiLeaks' release of thousands of US diplomatic cables

Frequency of the term *Arab Spring* per month in the main English and Arabic sub-corpora with peak events between January 2011 and August 2013

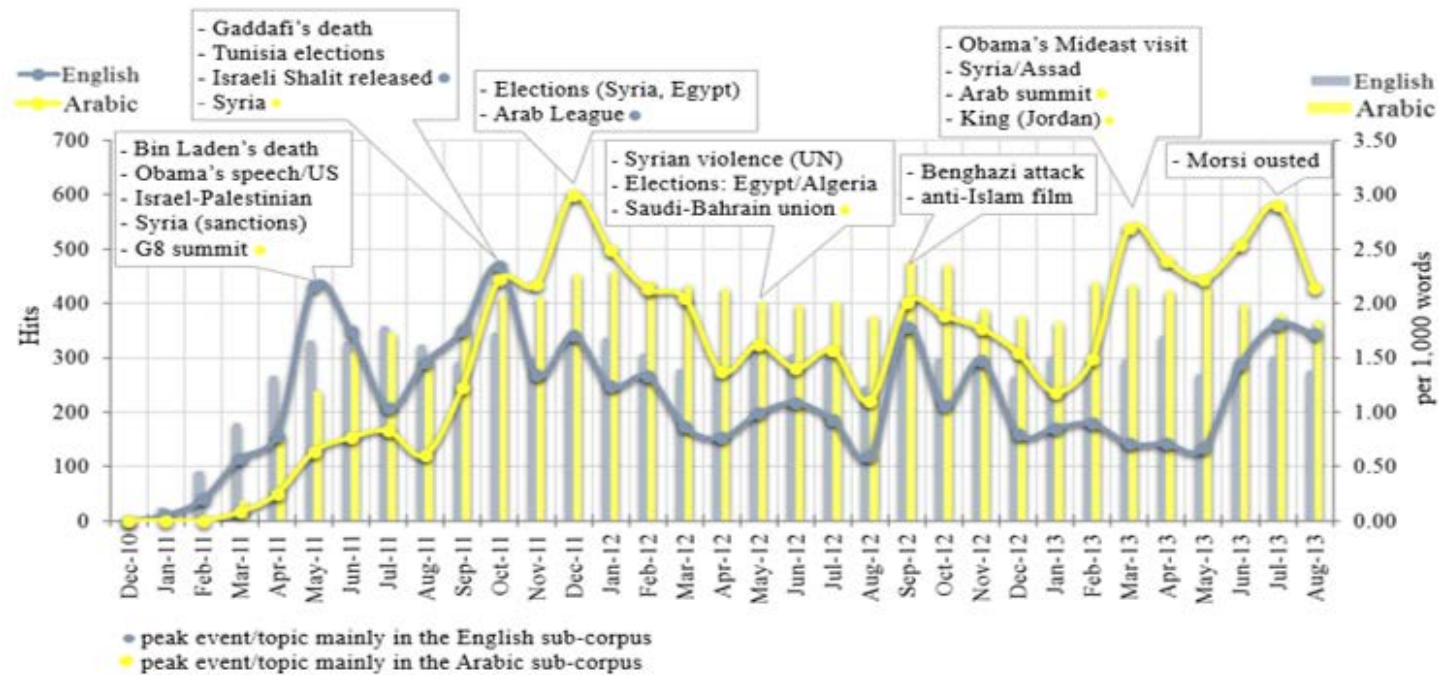


Figure 4.12 Frequency of the term *Arab Spring* per month in the English and Arabic sub-corpora (Jan-2011 and Aug-2013)

late November of the same year (5 December 2010). The WikiLeaks cables involved very important documents of which many were related to security issues. For example, leaks relating to the *Syrian* and Iranian plans to supply the *Shi'ite militant group Hezbollah* in *Lebanon* with new *weapons* systems, and SCUD missiles, which in turn would magnify the *threat* to *Israel*. Another leak was related to the stance of the *Arab Gulf countries*, especially *Saudi Arabia*, on *Iran* as the fundamental *threat* to the *Middle East region*. The most important peak event/topic during December 2010, however, was the one directly related to the *Arab Spring*, that is, Bouazizi and the protests in Tunisia. This peak event/topic is only key in the Arabic sub-corpus, as indicated in Figure 4.10 (by the yellow circle, peak events of December 2010). In terms of the period (between January 2011 and August 2013) after the emergence of the Arab Spring as Figure 4.12 shows, there was a gradual increase in the use of the term *Arab Spring* from its first appearance during January 2011 with some peaks in May, October and December of the same year, showing a continuing and increased importance in this topic.

In their corpus analysis of the representation of the refugees and asylum seekers (RASIM project) in the UK press between 1996-2005, Gabrielatos and Baker (2008, p. 17-8) discuss how the higher frequency of articles per month, which pertain to RASIM, show connections between the peaks and the nature of actual events around them, and consequently the increasing press attention to RASIM stories. Their study shows that despite the fact that some peaks have been directly connected with political events related to RASIM issues in the UK (for example,

the Asylum Bill during March-April 2004), other peaks show that different events such as ‘major wars, natural disasters, and terrorist attacks’ were some of the main causes for the increasing focus of the press on RASIM (Gabrielatos and Baker, 2008, p. 18). Similarly, Baker et al. (2013) combined corpus-based and corpus-driven approaches in order to compare the changing frequencies of the two most frequent pairs *Muslim world* and *Muslim community* over a continuous period of time between 1998 and 2005. From my perspective, the analysis of the KKW lists of the main ASC sub-corpora shows that there is a link between the rise in frequency of the term *Arab Spring* and some of the peak events or topics, such as elections, Bin Laden and Gaddafi’s deaths and Morsi’s ousting from power. All of these events are expressed in terms of reasons for optimism in the Arab Spring.

As Figure 4.12 shows, there are 21 peak events or topics of which only 14 peaks are shared between the two main English as well as the Arabic sub-corpora. For example, the KKW lists of the BR, AM and GV of the English as well as the MSA of the Arabic sub-corpora show that the Obama’s speech (19 May 2011) on the Arab Spring, US policy in the Middle East, Israel-Palestinian relations, the issue of borders, as well as the Bin Laden’s death along with his tape, which had been posthumously published praising the *Arab Spring*, were some of the key events during May 2011. However, the Arabic sub-corpus shows more focus on the event of قمة الثماني <qimmat at-tamānī - *the-G8 summit*> held in France (26 May 2011), as indicated by the small yellow circle in Figure 4.8. A Similar pattern of the difference in focus can also be seen across the different peak events/topics shown in Figure

4.8. For example, although Gaddafi's death (20 October 2011) tends to be a salient topic in the BR-ENASC, AM-ENASC and MSA-ARASC, the KKW list for the GV-ENASC, which has been created by using both the BNC and COCA reference corpora, reflects a higher saliency of topics that are related to the Israeli-Palestinian negotiations as well as the release of the Israeli soldier Gilad Shalit (18 October 2011), who had been a captive of the Palestinian militant Islamist group, Hamas, since 2006. Similarly, comparison of the KKW lists for December 2011 also shows quite a difference in the nature of events around its peak. For example, although the British, the American, as well as the Arabic sub-corpora show focus on issues that are mainly concerned with *Syria*, *elections* <الانتخابات - al-'intikāb-āt – *the-elections*>, *opposition* and *democracy*, the KKW list of the GV-ENASC sub-corpus reflects again issues that are mainly concerned with Israeli diplomacy as well as its relations with Arabs and Muslims, especially the Muslim Brotherhood represented by *Hamas*. In the same context, the British sub-corpus also showed different, yet related, results. For example, two of the top lexical KKWs (e.g. *Arab* and *League*) indicate issues that are directly related to the Arab League observers in Syria (19 December 2011).

As can be seen so far, although both the English as well as the Arabic sub-corpora tend to share many of the peak events/topics, in quite a number of instances, a difference in focus is also noticeable. This could be accounted for due to the fact difference of sizes of the Arab Spring component sub-corpora, which is believed to result in a difference in the amount of coverage of the same news events. For

instance, the Turkish HRTe and ZMNe of the GV-ENASC, include news texts that only cover the period between May and August 2013, and would consequently result in a different representation of the events that are related to the *Arab Spring* as a whole. According to Hardt-Mautner (1995, p. 29), the imbalance of the sizes of the sub-corpora used for investigating the representation of events related to the EC/EU debate in the British press between 1971 and 1994, ‘was an inevitable consequence of the differences in the amount of editorial coverage devoted to the political topic under investigation’. Another reason might be the fact that not all new media outlets, such as the Arabic version of the Iranian news outlet MHRa, use the term *Arab spring* <الربيع العربي> to refer to the various events across the Middle East countries, and hence ‘the development in the number of corpus articles [in my case containing the term *Arab Spring*] differs from newspaper to newspaper’ (Gabrielatos et al., 2012, p. 162). From a CDA perspective, the difference of focus could also be attributed to the fact that many events in different countries throughout the Middle East took place after the emergence of the *Arab Spring* in late December 2010. From this perspective, several political, economic and social factors are believed to play a role in the selection of events that are considered as newsworthy. For example, the political stance and agenda of the newspaper itself (e.g. conservative, left, right, centre or liberal), the style of the newspaper (in this case broadsheets) as well as its frequency (e.g. daily or weekly) are some of the main factors that shape the newsworthiness of events in the press (Partington, 2012; Fowler, 1991; O'Neill and Harcup, 2009; Baker et al., 2013a; Gabrielatos and

Baker, 2008; Bell, 1991). Finally, and perhaps most importantly is that language also expresses culture and there seems to be some indication that Arabic language news has different cultural foci compared with English language news. Ideologies are therefore connected with the language in which readers access the news. In that sense reading in Arabic will give a reader a different picture of the world from reading in English.

Looking at Figures 4.10 and 4.12 again, although the majority of peak events/topics are directly related to the three most key categories (i.e. GEOGRAPHY, GOVERNMENT AND SECURITY and NAME-PN), other topics relating to other top categories (e.g. FAITH AND RELIGION, PROTEST VOCABULARY, FUNCTION WORDS, and LANGUAGE AND COMMUNICATION) are also evident. Thus, in order to better understand how the Arab Spring is represented across the different English as well as the Arabic sub-corpora, (synchronically and diachronically) in terms of these interrelated and interconnected categories, it was decided to investigate some of the most significant KKW items associated with some peaks, before (Figure 4.10) and after (Figure 4.12) the first mention of the Arab Spring, with more focus on the month, that is, December 2010, that immediately preceded the emergence of the Arab Spring. Beginning with the first top category shown in Tables 4.1 and 4.2, the following section looks at how the Arab Spring is represented in terms of the GEOGRAPHY category as well as its related topics over the times specified above.

4.4 Geography: Drawing the Map

All KKW items that refer to the same country, or a geographical location were collectively put into lexical sets, and then classified into six main groups in terms of what these lexical sets have in common. Such grouping provides a useful starting point for the analysis. It also gives an overall picture of some of the common patterns between the two main English and the Arabic sub-corpora in terms of what can be considered as the ‘foregrounded nationalities’ (Taylor, 2014, p. 398), as well as other geographical locations in the ASC corpus as a whole. The geographical groups (highlighted in colours as shown in Tables 4.6 and 4.7 along with their overall average percentage of the English and Arabic sub-corpora respectively) represent the *Arab nationalities* (grey), the *Western and European* (red), the *Islamic* (green), the *geographical location* (blue), *Russia* (dark blue), and the *East Asian* (orange). The items that are in bold are shared KKW items that are also collocates of the term *Arab Spring*. For instance, in the English sub-corpus, as Table 4.6 shows, ‘the Syria set’ of the *Arab identities* group contains five KKW items (*Syria*: 22.73; *Syrian*: 15.70; *Damascus*: 4.13; *Syria's*: 4.01; *Syrians*: 1.77) of which the first two items (in bold) are shared between the six English sub-corpora, and are collocates of the term *Arab Spring*. On the other hand, ‘the سوريا – sūryā – Syria set’ of the Arabic sub-corpus (see Table 4.7) has only four KKW items

Table 4.6 Key geographical groups and lexical sets with overall Avg. % of the English sub-corpus

KKW geographical groups and lexical sets with their overall Avg. % of the English sub-corpus							
Israel	11.68	United States	4.15	Western	3.47	France	1.88
Israel	25.76	US	7.51	Western	3.47	French	1.88
Israeli	16.13	United	5.02				
Israel's	8.79	States	3.16			Europe	1.67
Israelis	4.38	America's	2.77			European	1.78
Jerusalem	3.35	American	2.58			Europe	1.54
Arab	19.46	Syria	9.94	Egypt	7.83	Palestine	6.93
Arab	35.91	Syria	22.73	Egypt	19.65	Palestinian	16.00
Arabs	3.01	Syrian	15.70	Egyptian	10.40	Palestinians	9.50
		Damascus	4.13	Egypt's	6.69	Gaza	7.83
Saudi Arabia	5.19	Syria's	4.01	Cairo	4.88	Palestine	2.82
Saudi	7.93	Syrians	1.77	Egyptians	3.39	West	2.75
Arabia	5.44	Libya	4.39	Tahrir	1.98	Bank	2.03
Saudis	1.59	Libya	9.91	Tunisia	4.36		
Iraq	3.67	Libyan	4.17	Tunisia	8.65	Bahrain	3.76
Iraq	5.87	Benghazi	3.30	Tunisian	3.91	Bahrain	5.55
Iraqi	2.23	Tripoli	2.67	Tunisia's	2.45	Bahraini	2.31
Baghdad	2.14	Libya's	2.00	Tunis	2.15	Bahrain's	2.15
Iran	6.06	Libyans	1.60			Yemen	3.06
Iran	13.26	Turkey	5.85			Yemen	4.74
Iranian	6.70	Turkey	10.72	Afghanistan	2.77	Yemeni	1.95
Iran's	4.58	Turkish	7.71	Afghanistan	2.77	Yemen's	1.37
Tehran	2.72	Turkey's	4.47			Pakistan	1.93
Iranians	1.52	Ankara	3.53			Pakistan	1.93
Middle East	5.15	Istanbul	2.03			Kurdish	1.78
Middle	5.58	foreign	3.58			Kurdish	1.78
East	4.73	foreign	3.58	region	3.56		
				region	4.73	country	3.41
gulf	2.99			regional	2.17	countries	4.36
gulf	2.99	state	2.93			country's	3.59
		state	2.93	international	2.44	country	1.78
world	2.1			international	2.44	border	2.22
world	2.1	global	1.86			border	2.37
		global	1.86	square	1.72	borders	1.99
				square	1.72	Russia	1.99
						Russian	1.46
						Russia's	1.28
						China	1.95
						China	2.23
						Chinese	1.58

colour key:

■ Arab nationalities and countries	■ Islamic
■ Western and European	■ Russia
■ Geographical location	■ East Asian

Bold: shared KKW items and collocates of the term Arab Spring

Table 4.7 Key geographical groups and lexical sets with their overall Avg. % of the Arabic sub-corpus

KKW geographical groups and lexical sets with their overall Avg. % of the Arabic sub-corpus			
عرب	5.80	إسرائيل	2.81
العربي <i>the-Arab</i> MASC	9.99	إسرائيل <i>Israel</i>	4.85
العربية <i>the-Arab</i> FEM	4.09	الإسرائيلي <i>the-Israeli</i> MASC	2.72
العرب <i>the-Arabs</i> PL N	2.53	الإسرائيلية <i>the-Israeli</i> FEM	1.95
سوريا	4.35	القدس <i>Jerusalem</i>	1.43
السوري <i>the Syrian</i> MASC	7.46	لإسرائيل <i>to-Israel</i>	1.20
السورية <i>the-Syrian</i> FEM	4.76	أمريكا	2.42
سوريا <i>Syria</i>	4.24	الأمريكية <i>the-American</i> FEM	2.98
السوريين <i>the-Syrians</i> MASC GEN PL	1.12	الأمريكي <i>the-American</i> MASC	2.82
مصر	3.85	واشنطن <i>Washington</i>	2.71
مصر <i>Egypt</i>	6.75	الولايات <i>the-states</i>	2.04
المصري <i>the-Egyptian</i> MASC	2.85	المتحدة <i>the-United</i>	1.83
المصرية <i>the-Egyptian</i> FEM	1.95	أمريكا <i>America</i>	1.56
فلسطين	3.05	الغرب <i>the-West</i>	1.69
الفلسطينية <i>the-Palestinian</i> FEM	4.19	إيران	2.53
الفلسطيني <i>the-Palestinian</i> MASC	3.28	إيران <i>Iran</i>	3.56
غزة <i>Gaza</i>	3.03	تهران <i>Tehran</i>	2.51
الفلسطينيين <i>the-Palestinians</i> MASC GEN	2.34	الإيراني <i>the-Iranian</i> MASC	2.04
فلسطين <i>Palestine</i>	1.73	الإيرانية <i>the-Iranian</i> FEM	1.63
السعودية	2.78	تركيا	2.36
السعودية <i>the-Saudi</i> FEM	2.86	تركيا <i>Turkey</i>	3.75
السعودي <i>the-Saudi</i> MASC	2.53	التركي <i>the-Turkish</i> MASC	2.11
تونس	2.54	التركية <i>the-Turkish</i> FEM	1.72
تونس <i>Tunisia</i>	4.65	أنقرة <i>Ankara</i>	1.33
بوزيد (Sidi)Bouzid	2.09	المنطقة <i>the-region</i>	2.44
التونسي <i>the-Tunisian</i> MASC	2.00	دولة	2.33
التونسية <i>the-Tunisian</i> FEM	1.65	الدولة <i>the-country</i> FEM	2.97
العراق <i>Iraq</i>	1.84	الدول <i>the-countries</i>	2.91
ليبيا	1.79	دول <i>countries</i>	2.06
ليبيا <i>Libya</i>	2.25	دولة <i>a country</i> FEM	1.06
الليبي <i>the-Libyan</i> MASC	1.34	الشرق الأوسط	1.76
البحرين	1.72	الشرق <i>the-East</i>	1.84
البحرين <i>Bahrain</i>	2.07	الأوسط <i>the-Middle</i>	1.72
البحرينية <i>the-Bahraini</i> FEM	1.18	الخارجية <i>the-foreign</i> FEM	1.55
الأردن	1.71	الشارع <i>the-street</i> MASC	1.55
الأردن <i>Jordan</i>	1.76	الخليج	1.33
الأردني <i>the-Jordanian</i> MASC	1.51	الخليج <i>the-gulf</i> MASC N	1.77
		الخليجي <i>the-gulf</i> MASC ADJ	1.20
		الخليجية <i>the-gulf</i> FEM ADJ	1.02
		البلدين <i>the- (two) countries</i>	1.11
		روسيا	1.11
		روسيا <i>Russia</i>	1.20
		الروسي <i>the-Russian</i> MASC	1.02

Colour key:

- Arab Nationals
- Western and European
- Islamic
- Geographical location
- Russia

Bold: shared KKW's and collocates of the term <الربيع العربي> *the-Arab Spring*

of which the first three items, that is السوري<the-Syrian.MASC>; السورية<the-Syrian.FEM>; and سوريا<Syria>; are shared and collocates of the term الربيع العربي <ar-rabī' al-'arabī – the Arab Spring>. In terms of the item *Damascus* <دمشق – Dimašq>; it is neither a KKW nor a collocate of the term *Arab Spring*. Additionally, both the English as well as the Arabic sub-corpora show, as indicated by Tables 4.6 and 4.7, semantic similarity in terms of their statistically significantly salient geographical lexical sets, or the groups to which they belong. In this context, it is worth mentioning that the *Arab nationals and countries* group shown in Tables 4.6 and 4.7, contain only the top ten out of totals of 16 lexical sets for each of the English as well as the Arabic sub-corpora respectively, whereas the other geographical groups contain all of the resulting KKW items). The six remaining geographical sets in the English sub-corpus include *Lebanon, Jordan, Qatar, Morocco, UAE, and Algeria*. In terms of the Arabic sub-corpus, the six remaining geographical sets also include *Lebanon, Qatar, Yemen, Algeria, Sudan and Morocco*.

From a CDA perspective, the GEOGRAPHY category of both the English as well as the Arabic sub-corpora does not only show the countries or the places where the *Arab Spring* events have occurred, but also reflects the geostrategic as well as the geopolitical interests of the Arab Spring at both the regional as well as the international levels. Examples 31 and 32 from the English sub-corpus show how some KKW items of different geographical groups (in bold) are expressed. Most of these KKW items (underlined) are also collocates of the term *Arab Spring*,

indicating the nature of the contexts in which they interconnect with other KKW items (*italicized*) of other KKW semantic categories. For example, *revolution* from the PROTEST VOCABULARY, or *democracy* and *political reform* from the GOVERNMENT AND SECURITY category, as indicated in Example 31 (lines 1 and 2 respectively). Other contexts reflect more of religious (e.g. *Sunni radicalism*) and/or emotional (e.g. *threat*) aspect, as shown in Example 32 (lines 5 and 7 respectively).

31. [HEADLINE] Democracy in **Tunisia** should be just the beginning

- 1 Whether or not **Tunisia's** *revolution* leads to a stable *democracy*
- 2 and *political reform* in the **region** now depends, in part, on how the
- 3 **US** and **Europe** respond. Recent events offer the chance both to
- 4 help **Tunisia's** reformers and renew support for democracy in the
- 5 **Arab world**.

(FTe-11-04/ opinion: 20-Jan)

32. HEADLINE Taking the White House to task for its **Middle East** blunders

- 1 [...]
- 2 Nasr [American academic] suggests that **America's** **Arab**
- 3 *allies* would rather see *war* between the **United States** and
- 4 **Iran** than a rapprochement. He insists that *sanctions* that
- 5 weaken **Iran** will open the **Middle East** to *Sunni* radicalism
- 6 and greater penetration by **China** and **Russia**. And, shades of
- 7 the Cold War, he argues that the real *threat* to **American**
- 8 interests comes not from **Iran** but from **China** and **Russia**.

(WPe-13-51/ opinion: 05-May)

As also indicated in Tables 4.6 and 4.7, the *Western and European Group* is the first and second most key geographical groups in the English as well as the Arabic

sub-corpora respectively. The presence of the KKW items belonging to the *Western and European Group*, such as *US* and *Europe* as indicated in Example 31 (line 3), and Example 32 (lines 2,3,6,7 and 8) represents instances of the involvement of these powerful and elite countries in different events of the *Arab Spring* across the *Middle East*. Similarly, in Arabic, Example 33 (lines 3,4,5 in Arabic, and 7,8 and 10 in English) shows instances of the *Arab Spring* contexts in which KKW items of the *Western and European Group* are used. As Example 10 also shows, geographical items reflect different presentation strategies employed in categorizing/classifying news actors in terms of provenance, ethnicity, or race (van Leeuwen, 2008, p. 42). For example, the KKW item <الأمريكي>the-American.MASC>, Example 11 (line 3, US: line 7), is an instance of identification or classification. Another point to make from Example 34 is that the bolded and underlined KKW items, belong to different semantic categories, and are also collocates of the term <الربيع العربي>the-Arab Spring>:

34. HEADLINE 1 أوباما يلقي خطاب "الربيع العربي"
 2 **Obama** is to deliver "the Arab Spring" Speech
- LEAD 3 يلقي الرئيس الأمريكي باراك أوباما اليوم الخميس بمقر وزارة
 4 الخارجية الأمريكية خطابا ينتظر أن يعلن فيه عن استراتيجية
 5 جديدة للولايات المتحدة في **العالم العربي** في ظل ما تشهده
 6 **المنطقة** من **ثورات شعبية**
- 7 *US President Barak Obama is to deliver today*
 8 *Thursday at the US State Department a speech in*
 9 *which he is expected to announce a new strategy*
 10 *of the **United States** in the **Arab world** in light of*
 11 *the popular **revolutions** that **the region** witnesses.*
- (JAZa-11-78/news: 19-May)

As also indicated in Tables 4.6 and 4.7, there is a difference between both the English as well as the Arabic sub-corpora in terms of the number of the key geographical groups. That is, the *East Asian Group* is only key in the English sub-corpus of which ‘the China set’ is the only comprising item. ‘The China set’, as Table 4.6 shows, consists of two KKW items: *China* and *Chinese* with values of 2.23% and 1.58% of the English sub-corpus respectively. However, further analysis showed that the different geographical groups pattern differently in terms of “‘key-keyness’” (Scott, 2015, p. 238) across the different English (AM, BR, GV) and Arabic (MSA) text type (news and editorials and opinions) sub-corpora. Figure 4.13 shows the main geographical groups with their overall percentage of each sub-corpus, along with examples of the most key geographical sets across the English as well as the Arabic text type sub-corpora. As Figure 4.9 also shows, *the Western and European* group is most key in only three sub-corpora (the British and the GV news as well as the GV editorials and opinions sub-corpora). The Western and European group has its highest value 9.51% of the GV editorials and opinions sub-corpus. On the other hand, *the East Asian Group* is only key in both the American as well as the British text types sub-corpora. It has its highest value 2.61% of the American editorials and opinions sub-corpus, whereas its lowest value 1.31% of the British news sub-corpus. In terms of the *Arab Nationals Group*, it is most key in only five sub-corpora (both texts types of the American and Arabic sub-corpora, and the British editorials and news sub-corpus), but, again, the GV-ENASC editorials and opinions sub-corpus has the highest value (7.31).

KKW geographical groups with their Avg. % of each of the English and Arabic text types sub-corpora
with examples of the top key geographical sets

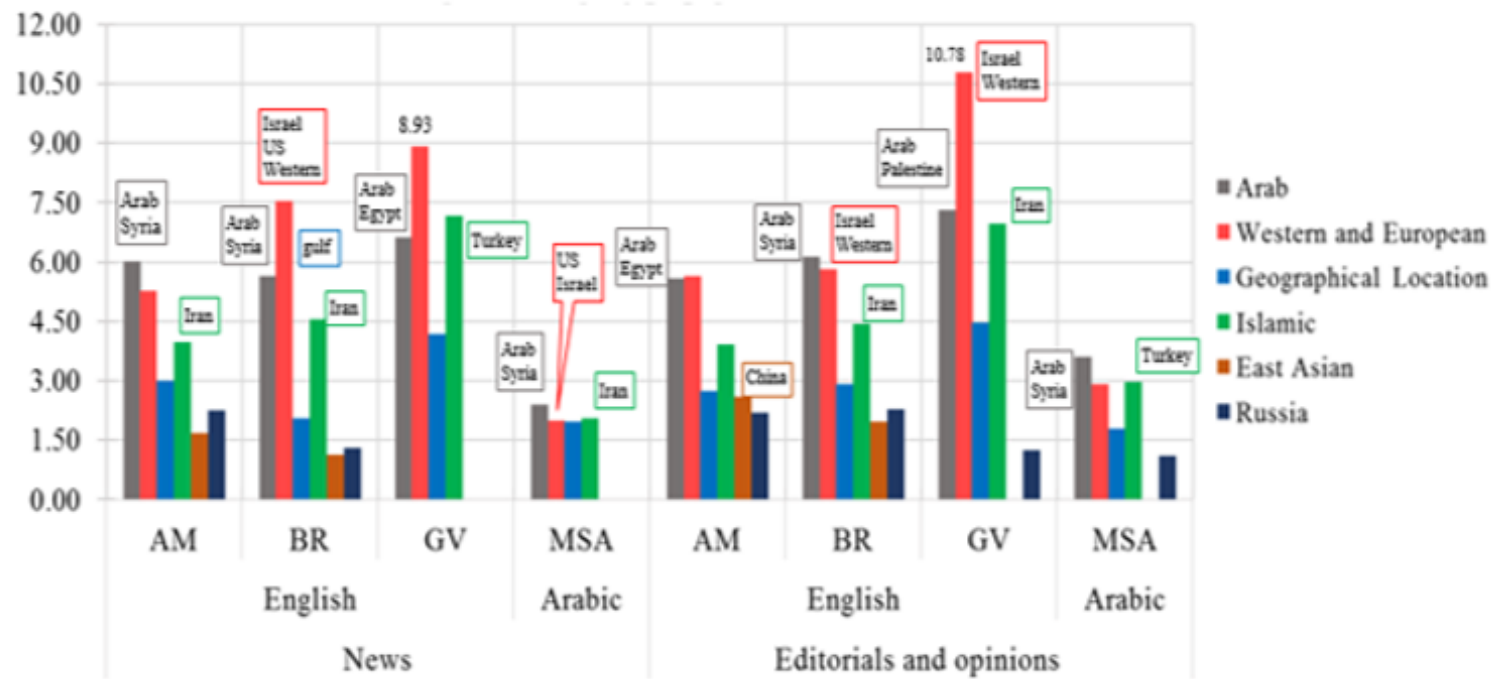


Figure 4.13 KKW geographical groups and their Avg. % of each of the English and Arabic text type sub-corpora

Nevertheless, most of the examples of the top geographical sets, as shown in Figure 4.13, show similar patterns to those shown in Tables 4.6 and 4.7. For example, ‘the Syria set’ is mostly the second significantly salient geographical set (after ‘the Arab set’) of *the Arab Nationals Group* across the English as well as the Arabic sub-corpora, followed by ‘the Egypt set’. Looking at the results from CDA and topical analytic perspectives, the results shown in Figure 4.13 can be said to ‘represent what news-makers construe to be the most important information about’ (van Dijk, 1991, p. 71) the geographically related news events/topics. Additionally, the significant saliency of the different geographical sets, shown in Figure 4.13 embodies the criteria of the journalistic decisions about the ‘newsworthiness’ of their related events (van Dijk, 1991, p. 71). The geographical groups and their related KKW items, like other items of the different semantic categories, have their significance at both the linguistic as well as the semantic levels. Linguistically, these KKW items are amongst other items that can be used in the different discursive strategies employed in the representation of news actors or events and newsworthiness. At the semantic level, the geographical groups, especially the *Western and European Group* suggest that the *Arab Spring* is also a narrative of elite and powerful nations (e.g. *US, Europe, Israel* and the *West*), and *Russia*, people (e.g. *Obama, Clinton, Netanyahu*, and *Putin*) and/or groups and institutions (e.g. *the United Nations, UN Security Council* and the *EU*). The resulting KKW items in the GEOGRAPHY category, or the topics they denote, can be linked to different news values of which *Eliteness* is the most significant. Topics are important not only in terms of their

choice, but also in how they are covered by the different news media (Herman and Chomsky, 2002). For topics are amongst the many ways, such as framing of issues and filtering information, that serve the media's purpose in inculcating and defending the social, economic and political interests and agendas of the elite groups in a given society or state (Herman and Chomsky, 2002, p. 298). Examples 12 (lines 2, 4 and 5) and 13 (lines 5, 6 and 7) show how abuse of power is represented or legitimated by text and talk of the dominant groups or institutions (van Dijk, 1996a, p. 84). The power of specific social groups may be restricted to specific domains, such as politics, the media, law and order, which in turn results in different "centres" of power and elite groups controlling such centres (van Dijk, 1996, p. 84). In that sense, the media, especially the news media, becomes one of the valued social resources and access to them is a sign of power (van Dijk, 1996, p. 85, Fairclough, 2001). Example 13 (line 3) shows, through the item *the senators said*, who is being given access to this valued social source, and hence a voice. According to Fairclough (2001, p. 53), access to public discourse is confined to those qualified, who 'come mainly from the dominant bloc', which includes the capitalist class, the middle class and the professionals. The dominant bloc, as Examples 35 (lines 2 and 4) and 36 (lines 2 and 6) clearly show, is the West, as represented through the items, *western*, *western nations*, *American* and *United States*.

35. 1 [...]

2 Accepting that **western** domination of the **Middle East**
 3 is coming to an end, however, should not be confused
 4 with saying that **western nations** *will not defend their*
 5 *interests*.

(FTe-13-44/ comment: 18-Jun)

36. 1 In a **statement** sharply critical of President
 2 Obama's decision not to suspend **American**
 3 assistance to **Cairo**, the *senators said* the events
 4 in **Egypt** and elsewhere in the **Middle East**
 5 "will directly impact the *national security*
 6 *interests* of the **United States**, and we cannot
 7 *remain disengaged*".

(TMSe-13-69/news: 17-Aug)

It is also notable that GEOGRAPHY related items 'often express *who* rather than *where*' (Bell, 1991, p. 200, original italics). Figure 4.14 shows concordance examples of how some key items (in bold) of different groups taken from the American news sub-corpus during August 2013 are represented as news/political actors rather than as physical territories (Bell, 1991, p. 200), and hence considered as the main subjects (or the agents) that perform the actions of the verbs (in bold) that follow them (e.g. *Syria collapsed/ denies, Jordan could face/ does not want, Saudi Arabia delivered* and finally *Israel directed*):

and Yemen tumbled, and Syria collapsed into civil war the
 Why do Russia, Iran and China continue to support a regime
 intervention in Syria, then Jordan could face terrorist attack
 ish Egypt's rulers. And Saudi Arabia delivered a blank che
 hing Syria, even though Syria denies using chemical weapon
 e no such evidence that Israel directed the coup, Israeli
 . Despite that meeting, Jordan does not want to be seen as
 regime change in Syria. Iran says the main objective of th
 jected to shrink as the United States withdraws forces fr

Figure 4.14 Concordance examples of countries as news actors from the American news (Aug-2013)

37. HEADLINE 1 Despite death reports, **Syria** denies Hama
 2 campaign
 LEAD 3 **Syria** denied Tuesday that a military campaign
 4 was under way against the restive city of Hama,
 5 even as human rights groups reported deaths,
 6 arrests and clashes on the city's outskirts.
 7 [...]
 8 Also Tuesday, the **United States** reiterated
 9 *support for the Syrian people and called for an*
 10 *end to the government crackdown.*
 11 "**Syria** claims it is interested in a dialogue with
 12 the opposition. Yet its actions in cities like Hama
 12 and along the Turkish border directly undermine
 14 the credibility of its words and its initiative,"
 15 State Department spokeswoman Victoria Nuland
 16 said in a statement.

(CNNe-11-112/ news: 05-July)

Similarly, the Arabic concordances in Figure 4.15 also show examples of some KKW geographical items that are also collocates of the term الربيع العربي <the Arab Spring> as the main actors performing different actions of the verbs (in bold) that follow rather than pure names of countries or places. The use of geographical items

<i>Concordance examples from the Arabic news showing countries as news actors</i>	
Arabic concordance	English Translation
قاله كيري أن "الولايات المتحدة" تدعو قادة مصر من	what Kerry said that <i>"The United States"</i> calls on Egypt's leaders of
جدير بالذكر أن مصر تعيش أزمة اقتصادية صعبة،	worth mentioning that <i>Egypt</i> lives a difficult economic crisis,
أعلنت حماس أن مصر أغلقت بشكل رسمي معبر رفح الحدودي	<i>Hamas</i> announced that <i>Egypt</i> officially closed the Rafah border crossing
مستددة على أن إيران تمتلك الوسائل التكنولوجية لتوجيه	stressing that <i>Iran</i> owns the technological means to guide
إلى سفاراتهم أن تركيا لا تتحاز إلى السنة أو الشيعة،	to their embassies that <i>Turkey</i> does not take sides with the Sunnis or the Shiites,
في مصر أن الولايات المتحدة تدعم الإخوان	in Egypt that <i>the United States</i> supports the Brotherhood,

Figure 4.15 Concordance examples of countries as news actors from the Arabic news

in the representation of news actors is called, in van Leeuwen's (2008, p. 46) terms, *spatialization* that reflects a representational choice of *impersonalization*, which occurs when actors are represented by means of reference to a place. This shows how GEOGRAPHY related items can be used either linguistically as pure names of places, or semantically as political entities that represent, as Bell (1991, p. 200) points out, news authorities (e.g. *the US administration* and *the Foreign Ministry*), or 'a set of stereotypes about a nation and its people' (Bell, 1991, p. 200). According to stereotypes Fowler (1991, p. 19) says these are 'socially constructed mental pigeon-holes into which events and individuals can be sorted, thereby making such events and individuals comprehensible'. In that sense, geographical terms, such as the *Arab world*, *Arab street*, *Western world*, *gulf states*, or *the Middle East (region)*, can be considered as 'geographical labels' or stereotypes that may negatively or positively, provoke different associations 'in the minds of some readers', and hence influence the prevailing attitudes towards a given country or its people. Such geographical terms or labels can also be considered as linguistic realizations of different assimilation and/or collectivization strategies which are quite frequently used and employed across the English as well as the Arabic sub-corpora, as discussed in Chapter 3 (section 3.5 and 3.5.3). For example, in his corpus-assisted comparative study that is concerned with the ways the *Arab world* is represented in sections of the UK and English-language Arab presses from different time periods, that is 2010 and 2013, Partington (2015, p. 240-1) found out that in all datasets examined, the term *Arab world* is very strongly primed to appear in prepositional

phrases, such as *in* and *across the Arab world*, and is also strongly primed to avoid functioning as an actor or doer of an action. His study also showed that, in the 2010 UK data, the *Arab world* is frequently represented in the role of audience, with ‘a degree of negative stereotyping since it is represented as an audience which is oversensitive, quick to take offence and needs to be placated and handled with care’ (Partington, 2015, p. 241). Comparing the item *Arab world* with the item *Western world* of the same period in terms of their grammatical profiles, Partington (2015, p. 233) found that the term *Western world*, like the item *Arab world*, is also represented as part of a prepositional phrase, preceded by *in* and *across*. However, his study also showed that the ‘*Western world* is represented as less passive, less recipient and less frequently as an audience than the *Arab world*’ (Partington, 2015, p. 233, italics added). Similarly, in this study, comparison of the occurrences of the two items *Arab world* (العالم العربي) and *Western world* (العالم الغربي) in terms of their raw and relative (ptw) frequencies in both the English as well as the Arabic sub-corpora (see sub-section 3.5.3 for more details on geographical items and collectivization strategy), showed that the item *Arab world* (2,328 occurrences, that is 1.55 ptw) is more frequent than the item *Western world* (68 occurrences, that is, 1.08 ptw) in the English sub-corpus as a whole. Likewise, the item <the-العالم العربي> *Arab world* (2,457 occurrences, that is 1.92 ptw) is also more frequent than the item <العالم الغربي> *the-Western world* (93 occurrences, that is, 1.23 ptw) in the Arabic sub-corpus as a whole. Additionally, although these two terms (that is, *Arab world* and *Western world*) might have different associations and representations,

the issue is that the Arabic terms above represent the masculine forms only, and hence comparing the English items with such Arabic gendered items (Vessey, 2013) might not reflect the whole senses. That is, the two classifying items *the-Arab* and *the-Western*, which correspond to both of their Arabic equivalents, that is, the (in/definite) masculine and feminine forms: *the-Arab* for *العربي* <al-‘arabī> and *العربية* <al-‘arabiyya>; *the-Western* for *الغربي* <al-ġarbī>; and *الغربية* <al-ġarbiyya> respectively. Hence, in Stubbs’ (1996) terms, different grammatical forms can have different collocates, and in terms of the notion of collocational priming, as Hoey (2005, p. 10) points out, it ‘is sensitive to the contexts (textual and social) in which a lexical item is encountered, and it is part of our knowledge of a lexical item that it is used in certain combinations in certain kinds of texts’. Thus, in the context of the Arab Spring, for example, the masculine form *العربي* <*the-Arab.masc*> (20,051 occurrences) in the Arabic sub-corpus as a whole is primed to post-modify (masculine) items (that is, the top ten collocates to the immediate left, L1) relating to different contexts of different geographical, social and governmental nature that include

الربيع < <i>the-spring</i> > (10,063);	العالم < <i>the-world</i> > (2,463);
الوطن < <i>the-homeland</i> > (425);	الشارع < <i>the-street</i> > (260);
الخليج < <i>the-gulf</i> > (246);	بالربيع < <i>by-the-spring</i> > (230);
للربيع < <i>to-the-spring</i> > (191);	الشباب < <i>the-youth</i> > (158);
الصراع < <i>the-conflict</i> > (1360; and	النظام < <i>the-regime</i> > (132).

On the other hand, the feminine form العربية <the-Arab.fem> (18,913 occurrences) in the Arabic sub-corpus as a whole, is primed to post-modify items (that is, the top ten L1 collocates) that are also relating to different contexts of geographical, social and governmental nature, and to names of countries and organizations, such as

الدول <the-countries> (2,135);	الثورات <the-revolutions> (1,479);
الشعوب <the-peoples> (760);	الجامعة <the-League> (757);
المنطقة <the-region> (693);	البلدان <the-countries> (436);
المملكة <the-kingdom> (327);	الأنظمة <the-regimes> (252);
المجتمعات <the-communities> (237); and	الأمة <the-nation> (189).

In terms of the masculine form الغربي <the-Western> (860 occurrences), this item is strongly primed to post-modify items (that is, L1 collocates) relating to topics of which some have negative meanings and/or connotations, such as الاستعمار الغربي <the Western colonialism>, and other items relating to contexts of military, governmental, as well as socio-political nature which include

العالم <the-world> (93);	الإعلام <the-media> (55);
التدخل <the-intervention> (38);	الموقف <the-stance> (21);
الاستعمار <the-colonialism> (19);	التحالف <the-alliance> (15);
المعسكر <the-camp> (14);	المجتمع <the-community> (13);
العسكري <the-military> (13); and	المشروع <the-project> (11).

On the other hand, the feminine form الغربية <the-Western FEM> (2,339 occurrences), which is used more frequently than that of the masculine form in the Arabic sub-corpus as a whole. The feminine form الغربية <the-Western.FEM> is primed to post-

modify items (that is, the top ten collocates to the immediate left (L1) of الغربية) denoting different topics that relate to both *Arab* (Palestinian and Moroccan) and *Western* contexts. These L1 collocates include

الضفة <the(West)-Bank> (489);	الدول <the-countries> (346);
القوى <the-powers> (100);	الصحراء <the-Sahara> (50);
والضفة <and-the-Bank> (47);	المصالح <the-interests> (46);
الحكومات <the-governments> (41);	والدول <and-the-countries> (36);
بالضفة <in-the-Bank> (29); and	الهيمنة <the-dominance> (28).

In terms of the English sub-corpus as a whole, Table 4.8 shows the top ten collocates to the immediate right (R1) of each of the items *Arab* (18,510 occurrences); and *Western* (3,817 occurrences) in the English sub-corpus:

Table 4.8 Top ten R1 collocates of both items *Arab* and *Western* of the English sub-corpus

Arab (18,510)		Western (3,817)	
R1 collocate	Freq.	R1 collocate	Freq.
spring	7,377	governments	178
world	2,328	countries	164
League	922	powers	156
countries	572	leaders	87
states	393	allies	79
Emirates	342	nations	78
uprisings	242	backed	75
leaders	227	diplomats	73
Israeli	202	media	69
world's	176	world	68

As Table 4.8 shows, the R1 collocates of both *Arab* and *Western* include the masculine and feminine Arabic equivalents, when compared to the different forms of the Arabic items and their L1 collocates. In this context, it can be said that phrases such as the *Arab world* (العالم العربي), and/or the *Western world* (العالم الغربي) are believed to have their own collocations, which are separate from those of their components, as Hoey (2005, p.10-1) pointed out.

As indicated so far, the analysis in this sub-section has given an overview of the main linguistic as well as semantic patterns of the GEOGRAPHY category across both of the English as well as the Arabic sub-corpora.

4.5 Summary

In this chapter, we have seen the usefulness of the KKWs in identifying the topics associated with the Arab Spring narrative in my corpus. We have seen that geographical regions and places are like actors. There is an almost similar pattern of the *where* and the *who*, making the non-human and human place and faces interchangeable. In his study on news media and headlines, Bell (1991) showed that place is the third most important element (after actors and time) that appears in the headline and it sometimes replaces actors. This means that actors are more important than *where* the news takes place. The GOVERNMENT AND SECURITY topic is also dominant in this narrative, but the presence of the word *elections* does not indicate democracy is being achieved. Words like *elections* and *democracy* are words in disguise; they have other functions, such as drawing the people's attention

from real issues like jobs and family. The word *democracy* hides behind it the issues of *security*, *fear* and *terrorism*. We will see in the next chapter that these topics are associated with news values of *Negativity*.

We looked at the use of function words of both English and Arabic (e.g. *pro-*, *anti*, *between* (بين), and *against* (ضد)), which show as KKWs, indicating that function words, far from being insignificant, carry important meanings that indicate conflict, difference and opposition. These words are closely related to the protest vocabulary and also indicate the emotional content of the narrative.

In the next chapter, we look at the newsworthiness and news values of these KKWs and semantic macrostructures in the corpus in order to assess *why* these words and concepts are selected by the different English and Arabic news media.

5. The Newsworthiness of the Arab Spring

5.1 Introduction

As we have seen in Chapter 4, topics, or the KKW that denote them, are significant in providing deeper insights into ‘what news-makers construe to be the most important information about a news event’ (van Dijk, 1991, p. 71). They also embody a set of news values or ‘factors which take a story into the news’ (Bell, 1998, p. 74). Topics are therefore important in explaining the selection of news items, the production as well as the formulation of news (van Dijk, 1988b, p. 119; Bell, 1991, p. 155; O'Neill and Harcup, 2009, p. 162). This chapter is therefore concerned with the newsworthiness of the Arab Spring in the English and Arabic news media before and after the emergence of the Arab Spring on 17 December 2010. It investigates the main news values prevalent in the Arab Spring corpus with focus on the three most key semantic categories (GEOGRAPHY, GOVERNMENT AND SECURITY, and NAME-PN), and how they are linguistically expressed. It then conducts a diachronic as well as a comparative analysis in order to examine the linguistic realizations as well as the semantic concerns or associations of these news values across the aforementioned categories in terms of text types (news and editorials and opinions) with a focus on some of the KKW-collocates and peak events shown in Figures 4.6 and 4.7. Specifically, the analysis in this chapter concerns the news values of *Negativity*, *Eliteness* and *Proximity*, that are related to values in the news actors and events, which in terms of Bell's (1991, p. 158)

approach make up the news content rather than those related to the values in the news process. This means that other news values that are formulated, for example, in terms of various economic conditions, such as sales, budgets for news gathering or the amount of advertising (van Dijk, 1988b, p. 120), or other factors related to the news process, such as *continuity* and *competition* (Bell, 1991, p. 158) are beyond the scope of this chapter.

It is worth discussing what news values are, and why they are important in CDA news media analysis. The following sections discuss the nature of news values from the perspectives of both journalism and Communication Studies as well as Linguistics. They also highlight some of the influential taxonomies in the three disciplines, with a focus on Bell's (1991) category of news values in news actors as well as Bednarek and Caple's (2014) and Potts et al.'s (2015) discursive approach to news values, especially the approach that is concerned with topic-associated news values.

5.2 What are News Values?

News values are generally considered as a set of factors by which an event or a story is newsworthy or not. According to Bednarek and Caple (2014, p. 136), news values can be defined from two perspectives. Firstly, from the perspective of journalism and Communications Studies, news values are typically defined as properties of events or stories, or as criteria that news-makers apply in order to select events or stories as news (Bednarek and Caple, 2014, p. 136). They determine

the way in which news stories are structured and represented (Bell, 1991, p. 155; Bednarek and Caple, 2014, p. 136). Such values include *Proximity* (geographical and cultural aspects of the stories), *Eliteness* (prominent and influential people, countries, nations or groups and organizations), *Negativity* (negative aspects of events of stories such as conflicts, wars and terrorism), *Superlativeness* (i.e. the intensity or magnitude of events), and *Personalization* (the human face of an event or items referring to ordinary people). Secondly, from a linguistic perspective, news values are expressed, constructed, highlighted, established and emphasized through language (Bell, 1991; Bednarek and Caple, 2014, 2012; Bednarek, 2015), and most copy editing serves to maximize the news values (Bell, 1991, p. 79). For some linguists, who researched news media, news values are also considered as criteria for selection that include ‘a variety of economic and news gathering aspects’ (Bednarek, 2015, p. 2). For example, van Dijk (1988b, p. 120) points out that news values reflect different ‘economic, social and ideological values in the discourse reproduction of society through media’, and points out ‘that such constraints have a cognitive representation’. Fowler (1991, p. 11) also states that ‘real events are subject to conventional processes of selection: they are not intrinsically newsworthy, but only become ‘news’ when selected for inclusion in news reports’. Similarly, news values, according to Bell (1991, p. 155) are ‘the values by which one ‘fact’ is judged more newsworthy than another’. Richardson (2007, p. 91) likewise defines news values as ‘the criteria employed by journalists to measure and therefore to judge the ‘newsworthiness’ of events’. He further describes news

values as ‘the (*imagined*) preference of the expected audience’, in the sense that they are guesswork that may result in missing the mark and misjudging the values of the audience (Richardson, 2007, p. 94). As most of these definitions imply, news values are seen by linguists as playing an important role in the news process as well as in determining what to include or exclude, and how to order and produce news stories (Bednarek, 2015, p. 1). However, such news processes also reflect the different aspects of the dialectical relationship between language and other social practices, in the sense that language ‘is always simultaneously constitutive of social identities, social relations and systems of knowledge and beliefs’ (Fairclough, 1995b, p. 54-5). These constitutive effects correspond to different functions of language (Fairclough, 1992, p. 64) through which news values are also constructed. Nevertheless, in their study that is concerned with news values, and their importance for CDA studies, Bednarek and Caple (2014, p. 138) pointed out that news values ‘have not yet been the focus of critical linguistic analyses of news discourse’ despite their ideological aspects. Inspired by Bell’s (1991, p. 156-8) category of values in news actors and events, Bednarek and Caple (2014) propose a discursive approach as well as the use of corpus linguistic techniques for the study of newsworthiness in news stories. They define news values ‘as values that are *construed* in and through discourse’ (Bednarek and Caple, 2014, p. 136, original italics).

From these perspectives, using corpus linguistic techniques, such as the KKW and collocation lists functions, and concordance analysis in studying the news values in

the Arab Spring narrative, having looked at who, what and where these stories involve and are located, I can now reveal *how* and *why* some events or topics in the Arab Spring stories are more newsworthy than others. This research evaluates newsworthiness from the viewpoint of the different news media outlets representing different cultures and ideologies. In this context, there are a number of news value taxonomies or lists provided by different scholars from journalism and media studies as well as linguists. The most influential list of this kind was offered by Galtung and Ruge (1965). They identified twelve factors for newsworthiness, such as *frequency, threshold, unambiguity, meaningfulness, negativity* and *reference to elite people, nations or persons*. However, despite its being considered as the foundational study and a widely accepted analysis of news values (Fowler, 1991; Bell, 1991), it has since been, according to Mautner (2008, p. 33), modified and critiqued by several scholars and authors such as van Dijk (1988b), Fowler (1991), Bell (1991) as well as Harcup and O'Neill (2001, 2016). For example, Bell (1991, p. 156), points out that most of Galtung and Ruge's (1965) news factors are more concerned with the nature of events and actors in the news. Bell (1991, p. 156) also states that the way these factors are linguistically represented can enhance the newsworthiness of a given story. Thus, Bell's (1991, p. 156-60) linguistic approach is concerned with three categories of news values of which the category of values in news actors and events is the one on which the analysis in this chapter mainly draws. It mainly relates to the content of the news (Bell, 1991, p. 156) . The other two categories are more concerned with the news process as well as the news quality

(Bell, 1991, p. 156). In terms of values in news actors and events, Bell (1991, p. 156) proposes twelve news values, ten of which correspond with the news values identified in this chapter. These values include *Negativity*, *Recency*, *Proximity*, *Consonance*, *Unexpectedness*, *Superlativeness*, *Relevance*, *Personalization*, *Eliteness* and *Attribution* (Bell, 1991, p. 156-8). Other news values identified in this chapter, as will be discussed shortly, are *Agenda* and *Positivity*.

5.2.1 Negativity

Negativity is considered as the most fundamental news value, and includes bad news, such as conflict between people, damage, injury or death. Example 38 shows how *Negativity* is construed through the items *embattled*, *rebels* and *ammunition* (in the headline) as well as the items *rebel fighters* and *dangerously low* and *ammunition* (in the Lead sentences).

38. HEADLINE Libya's **embattled rebels** beg for
ammunition

LEAD **Rebel fighters** in western Libya say they
are running "**dangerously low**" on
ammunition, and are now using their
reserves.

(BBCe-11-109/ news: 05-Aug)

5.2.2 Recency

Recency, which is related to Galtung and Ruge's (1965) *Frequency*, or Bednarek and Caple's (2014) *Timeliness*, is concerned with the news that 'has only just happened' (Bell, 1991, p. 156). However, although new or recent happenings are preferred, and are more likely to be reported, other aspects of time (past or historical) can also be included, especially in terms of the editorials and opinion articles. For example, Bednarek and Caple (2014, p. 155), Bednarek (2015, p. 3) and Potts et al. (2015, p. 4) define news events in terms of time, or *Timeliness*, as recent, ongoing, about to happen or otherwise relevant to the immediate situation or time – that is, current or seasonal. From that perspective, reference to historical or past events, especially in editorials and opinion texts, becomes necessary given the aim or function of such kinds of texts as persuasive or evaluative. According to van Dijk (1998, p. 60-1), 'ideological opinions [and editorials] selectively invoke and hide history' in a way that serves the current affairs or news stories that have been recently reported. In that sense, using history or referring to past events is necessary in two ways. First, it is used 'to show historical continuity, so that we learn from history' (van Dijk, 1998, p. 61). Secondly, and from a more cultural angle, using history provides editorials and opinion texts with continuity needed in presenting a given news story, or other socio-political phenomenon as good or bad, friend or enemy of our ingroup or culture to which we belong. Example 39 represents an excerpt from the British editorials and opinion sub-corpus, taken from the *Financial Times* on 19 March 2011. As the first sentence, or the Lead (L) turn

in Bolivar's (1994, p. 279; 2001) terms (see section 2.9, pp. 49-57 for more details), presents the main topic – that is, Libya's uprising as *chapter two of the Arab Spring* -- the third sentence, or the Valuate (V) turn, is completely an historical reference (in bold) to the 1991 Gulf war, which implies warning and a 'lesson from history' (van Dijk, 1998, p. 61). Additionally, invoking an historical event, such as the 1991 Gulf war in this example, and drawing 'parallels [...] between the situations in' (Prentice et al., 2012, p. 274) both Libya and Iraq (through *just as*) has an important function. It serves the writer's aim to persuade the audience (or readers) of the importance of the UN Security Council's resolution on Libya, during March 2011. The resolution, as indicated in the Lead (L) sentence of the Development Triad (D)Td, is concerned with getting a no-fly zone over Libya as well as authorising external intervention to prevent Gaddafi's regime from massacring civilians. Additionally, *Timeliness* is constructed in Example 39 through the items *week ago*, the tense and aspects in the items *advanced*, *looked* and *would be choked*, as indicated in the L sentence of the Situation Triad, or (S)Td.

39. HEADLINE Chapter two of the second Arab awakening

(S)Td L Only a **week ago** as Muammer Gaddafi's forces advanced eastwards, slaughtering Libyan rebels, it looked as though Libya's uprising would be choked by a ring of steel closing around the rebel stronghold of Benghazi.

F [...]

V Just as in Iraq **after the 1991 Gulf war**, when a defeated Saddam Hussein ruthlessly **beat back** the Shia intifada, the US and Europe **seemed to be**

looking on, arms folded, as a rebellion they had **cheered on was drowned** in blood.

(D)Td L There is, naturally, a lot of scepticism as to whether a vote at the UN Security Council can change all that.

F [...]

V The UN-mandated forces will need to act fast, neutralise Col Gaddafi's warplanes and destroy his tanks and artillery if they continue to menace Benghazi and other Libyan towns.

(FTe-11-34/ comment: 19-Mar)

Similarly, in Arabic, Example 40 represents an excerpt taken from the editorials and opinion sub-corpus of *الشرق الأوسط* <aš-šarq al-ʿawsaṭ – *the Middle East*> newspaper. Example 40 also shows how the writer initiates the topic by referring to previous events (in bold), such as *أحداث حماة* <*incidents-of-Hama*> in Syria during February 1982, and the Iranian support of the Syrian government against the armed rebellion of the Muslim Brotherhood at that time. Using history in this example also serves to draw parallels between both the current situation (the protests that broke out during March 2011) and that of Hama (on March 1982) in Syria. Examples that reflect aspects of *Timeliness* are the items *اندلعت* <broke.FEM-out – *broke out*>; *وجدت* <found.FEM – *found*>; and *كان يجب* <*had-to*>, as indicated in the L sentence of the situation triad (S)Td in Example 40:

40. HEADLINE

الأسد بين المظاهرات والعزلة

Al-Assad between the-demonstrations and the-isolation

- L حين اندلعت أحداث حماة (فبراير-1982)، وجدت القيادة الثورية في إيران نفسها أمام امتحان عسير، حيث كان يجب عليها أن تختار ما بين حزب البعث (العلماني) الحاكم، أو الانتصار لثورة الإخوان المسلمين المسلحة الذين يقتربون كثيرا من فكر وأيديولوجيا الملاكي الراديكاليين.
- (S)Td

When the incidents of Hama broke out (February 1982), the revolutionary leadership in Iran found itself facing a difficult test, where it had to choose between (the secular) ruling Ba'ath Party, or to support the armed revolution of the Muslim brotherhood who have a lot in common with the thought and ideology of the radical mullahs.

(WSTa-11-33/ opinion: 30-Mar)

In that sense, using history can be an important aspect of the news value of *Timeliness*, especially when it is used in the evaluations of the current or recent affairs reported in the news. In the same context, although news articles may use history, and give historical backgrounds to a given story, it is believed that its function is more informative than persuasive, evaluative or argumentative, as in the case of the editorials and opinion text articles. Nevertheless, according to van Dijk (1988b, p. 5), 'that distinction is notoriously problematic [... and] news articles may feature opinions, despite the ideological belief of many journalists that news only gives the facts and not opinion'. From this perspective, topics (or the KKW items denoting them) remain useful when analysing news values, not only in news articles but also in editorials and opinion texts, despite the fact that news values

were mainly devised for news articles. In fact, news values can be a useful tool in exploring the attitudes (negative or positive) and stances prevalent in such kinds of texts, which in turn emphasize their ideological nature. According to Bell (1991, p. 156), news values ‘are not neutral, but reflect ideologies and priorities held in society’.

5.2.3 Proximity

Proximity means geographical closeness or cultural similarity and familiarity of one country with another (Bell, 1991, p. 157). *Proximity* is related to Gatlung and Ruge’s (1965) factor of *Meaningfulness*. Example 41 shows the item *United Kingdom* (in bold), which reflects both geographical as well as cultural *Proximity* for both the British as well as the American audiences respectively:

41. HEADLINE Obama blends ceremony with diplomacy
in the **United Kingdom**
(CNNe-11-86 / news: 24-May)

However, *Proximity* in the news does not necessarily mean the physical or the cultural closeness to the target audience, especially in political news discourse in which ideology, power relations, and interests are important issues that underlie many of the news factors. This indicates that ‘ideological Proximity’ (van Dijk, 1988b, p. 125), which can also include political alignment, plays an important role in the way that news events or actors are represented. Examples 42 and 43 show

how items (in bold) indicating ideological aspects related to different religions and other belief systems could suggest ideological rather than geographical or cultural *Proximity*. For instance, the items *New York* (in the headline) and *American* (Sub-Head 1) suggest geographical closeness (to the American audience) or cultural familiarity (to the Western and European audience). On the other hand, the items *mosque* (although in New York) and *Muslim* (although American) suggest a different kind of *Proximity* (i.e. ideological) for (Muslim) audiences other than American, Western or European, who might be from other parts of the world:

42. HEADLINE Imam walks a tightrope over **mosque** in
New York
Sub-Head 1 **American Muslim** tries to connect 2
worlds, and has many enemies in both
(IHTE-10-18 / news: 23-Aug)

43. HEADLINE **West** must aid **liberal** rebels, says Blair
LEAD WESTERN governments should do more
to help the "**liberal democratic**" strands
of the revolutions in the Middle East,
which are badly organised compared
with their Islamist opponents, Tony Blair
said yesterday.
(DTLe-11-133/ news: 30-Dec)

Similarly, in Example 44, which shows a headline about the same subject in Example 22, both items <الولايات المتحدة> al-wilāyāt al-muttaḥida – *the-United States*> and <مسجد> masjid – *mosque*> suggest *geographical Proximity* (for the Muslim and

Arab audience in the United States). However, the same item مسجد<mosque>, as a place, where Muslims worship, has the potential for *ideological Proximity*, and hence for the (Muslim) audience of countries other than the United States:

44. HEADLINE موقف أوباما من "مسجد قرطبة" يثير جدلاً في الولايات المتحدة

Obama's stance on "Cordoba mosque" sparks controversy in the United States

(RBYa-10-135 / news: 14-Aug)

By the same token, Example 45 shows that *geographical Proximity* (for the British audience) is basically constructed by the item *Canterbury* (in the Lead sentence). However, the item *Christians* adds to the *ideological Proximity*, for the British or the audience who identify with the *Christians in the Middle East*:

45. HEADLINE Archbishop fears for **Christians in Middle East**

LEAD **Christians in the Middle East** are "more vulnerable" than they have been for centuries, the **Archbishop of Canterbury** has warned. (BBCe-11-222 / news: 09-Dec)

In addition, *Proximity*, according to Bell (1991, p. 157-8), need not be the same as the news value of *Relevance*, which is related to the effect on the target audience's lives, or is closer to their experience.

5.2.4 Relevance

Relevance is more about the way ‘events and *decisions* may affect our lives’ (van Dijk, 1988b, p. 122, italics added), and hence the *Impact*, in Bednarek and Caple’s (2014) terms, which relates to the direct effects and consequences on the target audience. Looking again at Example 39, the L sentence in the development triad (D)Td, the *decision* that was taken by the UN Security Council over Libya is relevant to Libya and the Libyan people although it has been made miles away in New York. In that sense, *Relevance*, which is related to the effects on the audience’s lives is believed to override not only ‘Cultural Proximity’, as Fowler (1991, p. 14) points out, but also the other two aspects of *Proximity*: the *geographical* as well as the *ideological*. However, *Relevance* is not always defined in terms of the reading public or audience, but rather, as in the case of the Arab Spring narrative, ‘in terms of large and powerful groups’ (van Dijk, 1988b, p. 122). Example 46 shows how *Relevance* is determined in terms of power relations (e.g. Sub-Head 1: *Gulf ally* and Lead: *fortifying* and *ally*) as well as the interests of the elite and the powerful (e.g. headline: *U.S.-Saudi jet deal*), as indicated by the items in bold, rather than the ‘ordinary’ Americans. Additionally, Example 46 represents other news values, such as the news value of *Superlativeness*, which is concerned with representing the event in terms of its intensity or high/ large scale (Bednarek and Caple, 2014; Bednarek, 2015). The items *crucial times* (in the headline), *more* and *major* (in the Sub-Head 1 and the Lead) are representations of *Superlativeness*; *Eliteness* (e.g. *U.S.*, *Obama administration* and *Royal Saudi Air Force*) that involves reference to

elite nations, news actors such as politicians, organizations or institutions of high status (Bell, 1991, p. 158). The item *new* establishes *Unexpectedness*, which closely relates to *Novelty* (Bell, 1991, p. 157; van Dijk, 1988b, p. 121). In other words, the news value of *Unexpectedness* reflects the representing the events as new, unpredictable, rare (Bell, 1991) or surprising.

5.2.5 Eliteness and Source and Voice

Another important news value that is related to *Eliteness* is that of *Attribution*, which concerns with ‘the eliteness of a story’s source’ (Bell, 1991, p. 158). According to Bell (1991, p. 191), there are two kinds of news sources: ‘suppliers of information which the journalist wants to know, or news actors whose own utterances have news value’. In Example 46, *the Obama administration* is the source of the news that is actualized by means of the items *announced* and *saying* (in the Lead sentence):

46. HEADLINE **U.S.-Saudi jet deal** comes at crucial time in Mideast;

Sub-Head 1 With more Iran tensions and Iraq on its own, **Gulf ally gets major arms lift**

LEAD **Fortifying a crucial ally in the Gulf, the Obama administration announced** a major new weapons deal with Saudi Arabia on Thursday, saying it had agreed to sell the Royal Saudi Air Force F-15 fighter jets valued at nearly \$30 billion.

(IHTE-11-217/ news: 30-Dec)

Attribution can also involve reference to ordinary people. In fact, it is important to news outlets, as it enhances the news value of *Personalization*, but ‘their status as sources is accidental rather than privileged’ (Fowler, 1991, p. 22). However, this might not always be the case, especially in news events of a polarizing nature such as the Arab Spring narrative. Citing ‘members of the public’ (Fowler, 1991, p. 22), or participants, who happen to be, for example, victims or witnesses (Bell, 1991, p. 194) reflects patterns of the negative or positive representations of the ingroup or the outgroup respectively. In their ‘Propaganda Model’, Herman and Chomsky (2002, p. 37), explain how the propaganda system consistently portrays the people abused in the ‘enemy’ state as ‘worthy’, whereas those treated with equal or more severity by its government or allies will be ‘unworthy’. From that perspective, victims and opposition sources or voices of the *outgroup* or its *allies* are worthy of inclusion and emphasis. The notion of a ‘worthy’ victim’s voice can be illustrated in in Example 47 and Figure 5.1 that represent the CNNe-11-81 news article during 22 May 2011. Example 47 shows how the victims – that is, *slain mourners in Syria*, are given voice (in bold) through the item *witness says*. Furthermore, Figure 4.1 shows how the victims (or *slain mourners*) in Syria, Iran’s most important ally, are ‘given more quotations from witnesses’ (Bazzi, 2009, p. 64), as indicated in Figure 5.1, *witness said* (lines 1, 7, and 8), *activist said* (lines 2, 4, 5 and 6) and *group said* (line 3). Consider how the ‘official source’ – that is, *state-run Syrian Arab News Agency* (also called *SANA*), as indicated in Figure 5.1 (line 10), is backgrounded, and the voice of the ‘32 “martyrs”’ (those are the *members of the Syrian security*

forces, who have been killed, according to the article, since the beginning of the events in Syria) is treated as unworthy:

47. HEADLINE Thousands bury slain mourners in Syria,
witness says
 (CNNe-11-81/ news: 22-May)

N	Concordance	%
1	by security forces, a witness said , in a stark change from	12%
2	with brute force, an activist said . The latest round of violence	17%
3	fired at demonstrators, the group said . On Saturday, violence	27%
4	of Friday's victims, an activist said . "We were chanting and	33%
5	us with their bullets," the activist said of Saturday's violence. "I saw	36%
6	to the main road," the activist said . On Sunday, the funeral	46%
7	at the funeral march, the witness said . "We were all expecting	52%
8	did not bother us," the witness said . "We do not know why this	58%
9	number of people, the witness said . Gunfire was heard	65%
10	State-run Syrian Arab News Agency said Sunday that 32 "martyrs" have	89%

Figure 5.1 All instances of *said* in the CNNe-11-81

As can be seen, *Attribution* is not only about the eliteness of sources, but also about whose voices are being reported or quoted, how they are represented, identified or integrated in discourse (Bednarek, 2016, p. 6). From a CDA perspective, quotation patterns, and the inclusion/ exclusion of voices are useful in revealing the ideologies and opinions of the news texts producers (van Dijk, 1993, p. 252; 1988b; Baker et al., 2008, p. 295), which in this case are biased and politically motivated. Quotation

patterns also reflect, as van Dijk (1995, p. 10) points out, ‘modes of access of various news actors or sources to the news media’. Such access is a reciprocal relationship between the powerful elite, and the news media (Fowler, 1991, p. 22), and it is this relationship that defines the strategies concerned with how news is produced and presented (van Dijk, 1995, p. 29). Access to discourse or communicative events takes many different forms, and is controlled by the powerful actors, who may set or select, *inter alia*, the setting (time and place), the participants or the audience, the topics and/or the agendas (van Dijk, 1995).

5.2.6 Agenda

In terms of *Agenda*, this a news value that is concerned with what defines or fits the news organisation’s own agenda in terms of the topics to be covered or discussed (Harcup and O'Neill, 2001, p. 279; 2016). However, it is generally believed that in many situations the mass media, especially the news media follow, or support, their government’s perspective and agenda, not only at the local but also at the international levels (Herman and Chomsky, 2002, p. 88; van Dijk, 1995, p. 28). For example, and within the context of the Arab Spring, a number of key topics fall under the news value of *Agenda*, especially politically oriented topics. For both the English as well as the Arabic sub-corpora key topics related to الانتخابات <the-elections>; حقوق الإنسان <human rights>; الأمن <the-security>; and السياسة <the-politics or policy> fall under the news value of *Agenda*. Difference in perspectives and representations of the different topics/ actors is then expected, given the fact that

the news outlets ‘have different patrons, subscribe to different agendas and project different images of themselves’ (Zayani and Ayish, 2006, p. 487). For example the elections of the ‘ingroup’ or its allies ‘will be found to legitimize, no matter what the facts, [while that of the ‘outgroup’ or its allies] will be found deficient, farcical, and failing to legitimize-again, irrespective of facts’ (Herman and Chomsky, 2002, p. 88). Example 26 shows how the issue of الانتخابات التشريعية <legislative elections>, in Jordan, is represented in the Lead. The extract is taken from *As-Safīr* (*The Ambassador*) newspaper in Lebanon, which is also ‘well known to be a Syrian-backed newspaper’ (Fisk, 2013, online). As the Lead in Example 48 shows (in bold), the elections in Jordan, a key ally of the United States, are represented as إلهاء <’ilhā - distraction.MASC.N – *distraction*> that is meant to prevent الأردنيين <the Jordanians> from focusing on their concerns with the rise in prices. As can also be noted, there are several observations to draw from Example 48. First, the order of news values plays an important role in the way events and actors are represented. For example, *Eliteness*, through the item الملك <al-malik – *the-king*> is backgrounded in the headline as well as the Lead in which it is represented through the item الملك عبد الله الثاني <al-malik ‘abdullah at-ṭānī – *King Abdullah II*>. Moreover, although *Proximity* (for the Arab audience) is indicated through the item الشارع الأردني <aš-šārī‘ al-’urdunī – the-Jordanian.MASC.ADJ street.MASC.N – *the Jordanian street*>, the metaphors الشارع الأردني <the Jordanian street> in the headline, and the item غضب <gḍab aš-šārī‘ - anger.MASC.N (of) the-street – *anger of the street*> in the Lead, fit more as items reflecting the news value of *Consonance*.

5.2.7 Consonance

Consonance, according to Bell (1991, p. 157) is the compatibility of a news story with the preconceptions about the social group or country from which the news actors come. In other words, *Consonance* concerns with the stereotypical representation of the news event or actors in a way ‘that is consonant with the

48. HEADLINE الشارع الأردني يفاجئ الملك: رسالة إلى الداخل والخارج

The Jordanian street surprises the King: A message for home and abroad

LEAD

لم تنجح خطة النظام الأردني هذه المرة، إذ لم يستطع الملك عبد الله الثاني إلهاء الأردنيين بانتخابات تشريعية يقول معارضون إن نتائجها ستكون معروفة مسبقاً. وبالرغم من أن الملك قد جمّد في وقت سابق قرار رفع الأسعار تفادياً لغضب الشارع، إلا أن "صندوق النقد الدولي" كان الأمر الناهي هذه المرة.. فكان انفجار الغضب في شوارع عمان وباقي المحافظات مفاجئاً، في رسالة تحذير من الشعب الأردني للداخل من جهة، ولأنظمة عربية أخرى قد تختار تفضيل مصالح المؤسسات النيوليبرالية العالمية على حساب شعوبها،

*The Jordanian regime's plan did not succeed this time, as King Abdullah II could not **distract the Jordanians by legislative elections, opponents say that their results will be known in advance.** And despite that the King had frozen earlier the decision to raise the prices to avoid the **anger of the street**, but it was 'the International Monetary Fund' that was the boss this time, and hence the burst of anger in the **streets of Amman** was surprising, in a warning message from the **Jordanian people** for home in one hand, and to other Arab regimes that might choose to favour the interests of the international neoliberal institutions at the expense of **their peoples.*** (SFRa-12-275/ news: 15-Nov)

attitudes of journalists and readers' (van Dijk, 1988b, p. 121). However, consonance in this sense does not necessarily mean that readers should share the stereotypes that journalists have about the different events or actors or that 'dissonant stories are less newsworthy' (van Dijk, 1988b, p. 122). Dissonant stories might be newsworthy in the sense that they provide the ingroup with information about the outgroup, and help in confirming the values and norms of the ingroup (van Dijk, 1988b, p. 123). In that sense, *Consonance*, that is more concerned with *stereotypes and labels* Looking again at Example 48, not only does the item الشارع الأردني <the Jordanian street> underlie the news value of *Consonance*, but it also reveals a different kind of representation strategy. There is a strategy of *Impersonalization*, through *spatialization* (van Leeuwen, 2008, 1996), in representing the main news actor الشارع الأردني <the Jordanian street> as agent, whereas الملك <the-king> is backgrounded and represented as patient (van Leeuwen, 2008, p. 23). *Spatialization* occurs when news actors are represented by means of reference to a place (van Leeuwen, 2008p. 46; 1996, p. 59). In this context, it can be said that *Impersonalization* can add negative (or positive) connotations to an activity of a social/news actor, while at the same time backgrounding his/her identity and role (van Leeuwen, 2008, p. 47; 1996, p. 60). For example, the associations invoked by the item الشارع <the street> are mostly negative. Of these associations are loitering, crime, violence, indigence, and most saliently demonstrations (Regier and Khalidi, 2009, p. 16). From that perspective, the verb يُفاجئ <yufāji'-u – surprises.MASC.PRS.V-nom – surprises>; which underlies *Unexpectedness* (or

surprise) fits well, and confirms such associations. Therefore, it can be said that *Impersonalization* in Example 48 has its effects on the news value of *Personalization* constructed by the item الأردننيين <the Jordanians> in the Lead.

5.2.8 Personalization

However, this does not mean that *Personalization* lacks its own effects on the representation of ‘ordinary’ news actors. In fact, *Personalization* is considered by most commentators on the media as dangerous (Fowler, 1991, p. 15-6). *Personalization* is also a socially constructed value, whose ‘functions are to promote straightforward feelings of identification, empathy or disapproval; to effect a metonymic simplification of a complex historical and institutional process’ (Fowler, 1991, p. 15). In the Arab Spring narrative, *Mohammad Bouazizi*, the Tunisian street vendor who set himself on fire, and whose death catalysed the Arab Spring, is constructed by the English and Arabic news media as a symbol of revolution, anger and political upheaval. Additionally, the use of persons by the media as symbols avoids serious discussion and explanation of the real causes of the different social and economic problems (Fowler, 1991, p. 16). *Personalization*, in that sense, ‘is the isolation of the person from his relevant social and institutional context’ (Hall, 1973, p. 236). For example, the angry crowds in protests and demonstrations are reported over and over again, but the real causes for such events are rarely documented (Fowler, 1991, p. 16). Therefore, it can be said that *Personalization*, like other news values, is not only socially constructed (Fowler,

1991), but also ideologically oriented, and the destination for ideology is only made possible by the human subjects – that is, people and the category to which they belong (Hall, 1973, p. 237; Fowler, 1991, p. 16). Similarly, in terms of the news value of *Voice and source*, Example 48 shows that it is the voice of the *opponents* <معارضون> that is represented, through the item

يقول معارضون

yaqūl-u

say.SING.MASC.PRS.V-NOM

opponents say

mu‘āriḍ-ūn

opponent.MASC.N-s.PL.NOM

as the source for the description and evaluation of the event being reported (van Dijk, 1988b, p. 87). There is no reference, for instance, to sources, which Fowler (1991, p. 22) describes as ‘highly privileged’, such as ‘high ranking politicians, experts or police officers’ (van Dijk, 1988b, p. 87), who can be considered as official representatives of the Jordanian authorities. Additionally, although introducing ordinary ‘speakers conveys both the human and the dramatic aspects of news events’ (van Dijk, 1988b, p. 87), and hence enforces the news value of *Personalisation* (Bell, 1991, p. 194; Fowler, 1991, p. 22), it has other ideological functions. As Example 48 indicates, the news value of *Voice and Source* can be considered as a means for accentuating and negating the viewpoints of the ‘ingroups and the outgroups’ (van Dijk, 1998, p. 57) respectively. The news value of *Voice and Source* can be employed to serve certain purposes and interests, especially if two or more conflicting ideologies or interests are involved in the reported events. From that perspective, including non-elite voices or sources might be politically or

ideologically oriented, and not only what Fowler (1991, 22) describes as ‘accidental’. News media are not a neutral, common-sense, or rational mediator of social or news events, but essentially contribute to reproducing ideologies in societies (van Dijk, 1988b, p. 11; Fairclough, 1992, p. 65). Finally, it can be said that news values, like topics, interrelate with each other and that their categorization can be a useful starting point to a more detailed analysis. The following section discusses the resulting news values identified in both the English as well as the Arabic text type sub-corpora in terms of their associated topics as well as statistical results.

5.3 Newsworthiness of the Arab Spring: Quantitative Perspective

This section is concerned with the newsworthiness of the Arab Spring narrative. It quantitatively as well as qualitatively discusses the linguistic realization as well as the semantic associations of the resulting news values associated with the top ten key categories of the English and Arabic text types sub-corpora. The categorization of the news values is generally based on the KKW-semantic categories and their constituent KKW items as well as their grammatical forms. According to Potts et al. (2015, p. 21), semantic categories or ‘semtags appear the most insightful in themselves for providing an overview of newsworthiness, and collocation analysis may be useful for identifying news values established in the co-text of topic-associated words’. However, the fact that language is multifunctional, and that many words and/or word forms may have different functions/meanings in a given

text or corpus of texts (Potts et al., 2015), and hence another rule may apply along with concordance analysis, that is, categorization can be based on the main function or sense of that given word in the corpus, as Baker, (2004, 353) pointed out. For example, the GEOGRAPHY category, which contains KKW items referring to specific countries, cities and other geographical places and locations has the potential to construct the news value of geographical or cultural *Proximity* (e.g. *Arab* and *Syria/n's*); *Eliteness* (e.g. *US*, *America/n/ns*, *Europe/an* and *international*); *Personalization* (e.g. *Arabs*, *Syrians* and *Palestinians*); and *Superlativeness* (e.g. *global* and *regional*). Thus, the actual total number of KKW items in the GEOGRAPHY category (see. Table 4.2 for more details on the actual total number of KKWs in each semantic category taken collectively) of both English (that is, 100 geographical KKW items) and Arabic (that is, 72 geographical KKW items) were categorized under four main news values as shown in Table 5.1:

Table 5.1 Breakdown of news values of the GEOGRAPHY category of English and Arabic sub-corpora

News Value in GEOGRAPHY	Number of KKWs in each sub-corpus	
	English	Arabic
<i>Proximity</i>	69	53
<i>Eliteness</i>	20	16
<i>Personalization</i>	9	3
<i>Superlativeness</i>	2	0
Column Total	100	72

Another example is that of the GOVERNMENT AND SECURITY category and its related news values in the English sub-corpus. The actual total number of KKWs of the English sub-corpus in this category is collectively 61 items. Most of these items were categorized under seven (7) resulting news values. For instance, the most significant news value in this category in terms of the number of KKW items is *Negativity* (19 items, e.g. *armed*, *arrested*, *war*, and *attack*); followed by *Eliteness* (17, e.g. *president*, *minister*, *government* and *activists*); *Agenda* (15, e.g. *democracy* and *elections*); *Personalization* (5, e.g. *refugees*, *civilians* and *citizens*); and *Consonance* (2, e.g. *regime* and *regimes*). Thus, for example, if *Eliteness* is to be counted based on both the categories of GEOGRAPHY and GOVERNMENT AND SECURITY, in the English sub-corpus, either collectively or individually for its main six sub-corpora, only the totals (that is 20 and 17) of the KKW items under *Eliteness* are counted along with their percentage of the sub-corpora. This kind of processing was conducted by means of a pivot table in Excel after manually adding the resulting KKW news values to the sub-corpora and their semantic categories worksheet (see Tables 5.3 and 5.4 for more details on the resulting KKW-news values and their related semantic categories).

Table 5.2 shows statistical information about the resulting twelve news values prevalent across the Arab Spring text types sub-corpora. The resulting KKW items were categorized based on their potential mean Ordered by the overall total number of KKW items, these news values include *Eliteness*, *Proximity*, *Negativity*, *Agenda*, *Personalization*, *Consonance*, *Source and Voice*, *Relevance*, *Timeliness*.

Table 5.2 KKW news values ordered by total number of words with their average percentage of each English and Arabic sub-corpus

No. News value	News				Editorials and opinion				Grand Total
	ENASC			ARASC	ENASC			ARASC	
	AM	BR	GV	MSA	AM	BR	GV	MSA	
1- Eliteness	83	65	60	72	74	65	58	59	536
	4.02	3.97	4.39	1.83	3.56	3.73	5.25	2.51	3.63
2- Proximity	69	67	56	64	64	68	51	58	497
	5.41	4.92	6.49	2.31	4.89	5.34	6.72	2.95	4.84
3- Negativity	33	32	28	10	29	29	27	23	211
	3.87	2.86	3.33	2.34	2.66	3.02	2.68	2.66	3.01
4- Agenda	22	20	20	16	23	19	18	21	159
	3.88	3.30	3.33	2.35	3.83	2.90	4.41	2.98	3.40
5- Personalization	25	23	16	11	23	22	16	14	150
	4.21	3.42	3.27	1.82	3.24	3.22	4.02	2.04	3.29
6- Consonance	9	10	9	4	12	11	12	8	75
	5.31	6.40	5.36	1.85	5.18	7.07	6.30	3.11	5.44
7- Source and Voice	5	6	4	12	5	3	3	1	39
	7.47	2.14	4.99	2.46	2.36	1.83	2.54	1.02	3.18
8- Relevance	4	2	3	1	3	3	4	6	26
	1.66	1.41	1.89	1.12	2.00	2.56	1.28	2.15	1.85
9- Timeliness	2	2	4	1	4	3	5	4	25
	5.76	1.58	2.90	2.64	5.11	2.86	3.68	3.54	3.62
10- Superlativeness	1		1		2	2	2	4	12
	1.25		1.62		1.40	2.09	3.27	1.67	1.92
11- Unexpectedness	1	1	1		1	1	1	1	7
	1.49	1.05	7.28		2.00	2.42	3.89	1.33	2.78
12- Positivity	1	1			1	1			4
	1.25	1.21			1.80	2.20			1.62

Bar colour key ■ Total No. of KKWs ■ Overall Avg. % of sub-corpus

Table 5.3 Actual number of KKW items realizing news values in the English and Arabic sub-corpora

News value	Example KKW items of English and Arabic	In No. of categories	No of KKW	
			EN	AR
<i>Eliteness</i>	(Obama - أوباما); (America - أمريكا); UN; activists; government; (president - رئيس); leaders; journalists	5	86	76
<i>Proximity</i>	(Arab - العربي); (Syria - سوريا); (countries - دول); (we - نحن); its; our; pro; (the-nation - الأمة); (Islam - الإسلام)	6	75	60
<i>Negativity</i>	military; (war - الحرب); attack; terrorism; anti; (against - ضد); (protests - احتجاجات); (conflict - الصراع); (sectarianism - الطائفية); (violence - العنف); threat; dictators; (tyrants - الطغاة)	7	39	26
<i>Agenda</i>	(democracy - الديمقراطية); (security - الأمن); (elections - الانتخابات); (opposition - المعارضة); (negotiations - المفاوضات); (peace - السلام). ((Arab)Spring - الربيع العربي)	6	24	21
<i>Personalization</i>	Mr; Ms; (Bouazizi - البوعزيزي); (protesters - المحتجين); (Arabs - العرب); (Islamists - الإسلاميين); Christians; Jews; (people - الشعب); women; (the-woman - المرأة); he; refugees; citizens.	7	28	16
<i>Consonance</i>	Islam; Islamist; Muslim; (secular - العلمانية); (the Zionist entity - الكيان الصهيوني); (regimes - الأنظمة);	2	12	9
<i>Source and Voice</i>	said; media; speech; (and-said.MASC - وقال); (and-added.MASC - وأضاف); (and-said.FEM - وقالت); (and-pointed.MASC - وأشار); (the report - التقرير); (adding.MASC - مضيفا); (and-stressed.MASC - وأكد)	1	6	12
<i>Relevance</i>	(the-relations - العلاقات); aids; visit; (the-change - التغيير); (intevention - التدخل); resolution; allies;	4	5	6
<i>Timeliness</i>	it's; (was - كان); will; has; that's; is; (took place - جرى); (became - باتت);	2	5	4
<i>Superlativeness</i>	regional; global; (also - أيضاً); (other - أخرى)	3	2	3
<i>Unexpectedness</i>	(coup - الانقلاب).	1	1	1
<i>Positivity</i>	peaceful	1	1	
Total number of KKWs			284	234

Superlativeness, *Unexpectedness*, and *Positivity*. The results are also visually represented in coloured data bars. Both the overall total number of KKW items, and overall average percentage the sub-corpora are represented by the gradient blue and green data bars respectively.

There are some important observations to be made from the results shown in Table 5.2. A first observation is that the resulting news values are mostly shared between the different English as well as Arabic text type sub-corpora, which indicates similarity between them despite the differences in their relative keyness. Additionally, looking at the results shown in Table 5.2, from the perspective of the Grand Row Total number of KKW items, it is *Eliteness* that has the highest value (542), followed by *Proximity* and *Negativity* with values of (497) and (211) respectively. The news value of *Positivity*, which is only key in the American as well as the British English text type sub-corpora, has the lowest value (4 KKW items) in terms of the number of KKW items. For example, in the British sub-corpus, although the KKW item *peaceful* is mostly used to modify KKW items such as *protests* and *demonstrations*, it is also frequently used to define different socio-political concepts, such as *transition*, *change*, *resolution*, and reform as indicated in Examples 49 (line 1) and 50 (in the headline).

49. 1 But by now standing ready to map out a route to
 2 **peaceful reform** in the heat of the **Arab Spring**,
 3 King Abdullah [of Jordan] is showing that
 4 sometimes in politics, there is an alternative.

(TMSe-11-74 / editorial: 14-Jun)

50. HEADLINE Obama Offers Support for **Peaceful
Transfer of Power** in Yemen
(NYTe-12-43/ news: 12-Feb)

This suggests that while *Eliteness* and *Negativity* are consistent and universal values in English and Arabic news, only the British and American news has any optimism about the Arab Spring.

For clarity of analysis, Table 5.3 collectively shows the ‘actual’ total number, along with examples, of the KKW items representing the resulting news values in both the English as well as the Arabic sub-corpora. As can be seen, both Tables 5.2 and 5.3 show identical patterns in terms of the order of the resulting news values. A further observation to draw from the results shown in Tables 5.2 and 5.3 is that both the English and Arabic sub-corpora clearly show keyness in the top five news values (that is, *Eliteness*, *Proximity*, *Negativity*, *Agenda* and *Personalization*) in terms of the number of KKW items representing them. However, the Arabic sub-corpus shows a different pattern in terms of the news value of *Source and Voice*.

Source and Voice ranks as number six, with a total of 12 KKW items of which all are focused in the Arabic news sub-corpus. This is expected since the use of verbs in (in)direct speech is common in this type of text, especially political news which, according to Mellor (2005, p. 96), is the most predominant type in the Arab news media. From that perspective, it can be said that ‘the abundance of political news in the Arab media means that politicians and officials serve as news sources’ (Mellor, 2005, p. 134). This pattern is also dominant in the three main, that is, the American,

British, and Global Variety English news sub-corpora in which *said* is the only shared KKW item between them. The KKW items *said* has its highest value of 28.82% of the American news sub-corpus. However, it is worth mentioning that other reporting verbs or verb forms, such as *says*, *noted* and *reported*, and the ‘attributive adjunct *according to*’ (Murphy, 2005, p. 132) are also KKW items, but not shared between three or more of the English sub-corpora. Therefore, they were not included in the final version of the KKW list set for the analysis. For example, the reporting verb form *says* is only KKW with value of 5.65% of the BR news sub-corpus. A concordance (2,243 lines) analysis of the KKW item *says* shows that it is most salient in the BBC news sub-corpus with 1,276 hits, with a relative frequency of 2.57 per thousand words. The most frequent 3-word-cluster *our correspondent says* occurs also in the BBC news sub-corpus with 68 occurrences. Looking back again at Table 5.2, the KKW item *وقال* <wa-qāl-a – *and-said*.MASC-PFV – *and (he/it) said*>, has the highest value of 8.39% of the Arabic news sub-corpus. Both English and Arabic use of *said* *قال* <qāla – *said*>, and its feminine form *قالت* <qāl-a-t – *said*.PFV-FEM – *(she/it) said*) indicates ‘a mere reiteration of what the source said’ (Mellor, 2005p. 135), and these are considered as neutral ‘speech-reporting verbs’ (Caldas-Coulthard, 1994, p. 306). The other Arabic KKW items in the *Source and Voice* category, such as *وأضاف* <*and-added*.MASC>; *وأشار* <*and-pointed*.MASC> can be used as variants of *وقال* <*and (he/it) said*>, ‘as they mainly add to the flow of the text’ or discourse (Mellor, 2005, p. 135-6). The use of the participle or adjective forms *مضيفا* <*adding*.MASC>; and *مشيرا* <*pointing*.MASC> also adds to the flow of the

text, but can also be more attributed to the influence of literary styles on Arab journalistic genres (Mellor, 2005, p. 135). The verb أكد<'akkad-a – stressed.MASC-PFV - (he/it) stressed>, on the other hand, indicates emphasis that adds reliability and credibility to the reported speech (Mellor, 2005, p. 136). What is perhaps interesting is that although the feminine form وقالت<and-said.FEM> is a KKW item in the Arabic news, it does not actually reflect the 'quantity of female speech reported in the [Arabic] press' (Caldas-Coulthard, 1994, p. 304), but rather the inclusion or citations from other sources. For example, analysis shows that the KKW item وقالت<and-said.FEM> mostly co-occurs (especially to the immediate right R1) with feminine items (i.e. in/definite or singular/plural nouns, or names of female actors, or pronouns) related to different semantic categories. Some items relate to LANGUAGE AND COMMUNICATION, such as مصادر<sources>; الصحيفة<the-newspaper>; وكالة<agency>; المتحدث<the-spokeswoman>; مراسلة<correspondent>, and الدراسة<the-study>. Other collocates of وقالت<and-said.FEM> belong to FUNCTION WORDS: e.g. ان<that>, في<in>, نحن<we>, and أيضا<also/too>; GOVERNMENT AND SECURITY: e.g. وزارة<ministry>; الحكومة<the-government>; السلطات<the-authorities>, المنظمة<the-organization>, mostly in reference to Amnesty International>, and الناشطة<the-activist>; NAME-PN: e.g. كلينتون – (Hilary) Clinton; كرمان – (Tawakkol) Karman; SOCIAL ACTIONS, STATES AND PROCESSES: e.g. الحركة<the-movement>; المعارضة<the-opposition> and جماعات<groups>, and finally GEOGRAPHY: e.g. الخارجية<the-foreign>, واشنطن<Washington>; and الداخلية<the-interior>. Figure 5.2 shows examples of the most frequent 3-word-clusters of

<and-said.FEM> in the Arabic news, and of which only two (lines 4 and 12) clearly refer to female news actors - that is, <the-spokeswoman> المتحدثة and هيلاري <Hilary Clinton>.

N	Cluster	Translation	Freq
1	وقالت الصحيفة إن	<i>and the newspaper said</i>	46
2	وقالت وزارة الخارجية	<i>and the Foreign Ministry said</i>	16
3	وقالت مصادر في	<i>and sources in ... said</i>	14
4	وقالت المتحدثة باسم	<i>and the spokeswoman for ... said</i>	14
5	وقالت وزارة الداخلية	<i>and the Interior Ministry said</i>	14
6	وقالت مصادر أمنية	<i>and security sources said</i>	12
7	وقالت المنظمة في	<i>and the organization said in</i>	11
8	وقالت وكالة الأنباء	<i>and the news agency said</i>	11
9	وقالت في بيان	<i>and (it/she) said in a statement</i>	11
10	وقالت المصادر إن	<i>and the sources said that</i>	9
11	وقالت للجزيرة نت	<i>and (she/it) said to Al-Jazeera net</i>	9
12	وقالت كلينتون في	<i>and (Hilary) Clinton said in</i>	9

Figure 5.2 Top 3-word clusters of <and.said.fem> in the Arabic news

In terms of the English news sub-corpora, analysis shows that the most significantly salient collocates of *said*, with minimum joint frequency 5 and MI score 3.00 or more, also belong to different key semantic categories. Figure 5.3 shows examples of the top collocates, to the immediate left (L1) of *said* of which five (in bold) are KKWs and collocates of the term *Arab Spring*. The collocates of *said* in Figure 5.3 also indicate that both the news values of *Eliteness* and *Source and Voice* mostly ‘collocate’ with each other.

Another important observation to draw from the results related to the news value of *Source and Voice* is that they reflect different strategies used in the representation of the different news actors in both English and Arabic. Some of these strategies include *nomination* (e.g. *Obama, minister, Clinton, and Netanyahu*), and *functionalization*, which occurs when actors are represented in terms of their occupation or role (van Leeuwen, 2008, p. 42), such as *officials, spokesman, activists* and *analysts*. A further strategy is that of *objectivation*, which ‘is realized by metonymic reference’ (van Leeuwen, 2008, p. 46), and includes *spatialization* (e.g. *Washington said*), *utterance automatization* (e.g. *report said*). The above examples show that the most salient news value of *Eliteness* can be realized through items belonging to five (out of the ten) key categories. These categories, also ordered by the number of KKW items, are NAME-PN, GEOGRAPHY, GOVERNMENT AND SECURITY, SOCIAL ACTIONS, STATES AND PROCESS as well as the LANGUAGE AND COMMUNICATION. The KKW items construing *Eliteness* suggest that the *Arab Spring* is a narrative of power and power relations, and that the words of the highly privileged ‘are generally put forward’ (Caldas-Coulthard, 1994, p. 304). Nevertheless, *Proximity*, also is another important, and the second most significant news value across the Arab Spring sub-corpora of which the American news sub-corpus (Table 4.1) has the highest value with 72 KKW items.

N	Word	Relation	Texts	Total	Total Left	Total Right	L1
1	he	5.32	3	6,845	5,185	1,660	4,913
2	she	5.44	3	1,435	1,096	339	1,056
3	officials	16.54	3	740	690	50	550
4	official	5.35	3	659	530	129	450
5	it	3.61	3	1,958	782	1,176	370
6	Obama	4.05	3	425	360	65	276
7	also	3.04	3	413	306	107	258
8	they	3.64	3	1,260	464	796	228
9	spokesman	5.94	3	493	367	126	182
10	ministry	5.27	3	321	278	43	175
11	group	3.86	3	370	282	88	171
12	who	3.27	3	876	395	481	161
13	minister	4.16	3	626	533	93	157
14	Clinton	4.79	3	191	175	16	155
15	analysts	4.66	3	161	151	10	137
16	government	3.09	3	701	436	265	134
17	Netanyahu	4.18	3	188	165	23	127
18	activists	3.63	3	174	157	17	125
19	statement	5.63	3	602	198	404	119
20	agency	4.39	3	168	157	11	118
21	report	4.37	3	219	158	61	117

concordance collocates plot patterns clusters timeline filenames

Bold: KKW and collocate of the term *Arab Spring*

Figure 5.3 Top collocates to the L1 of *said* of the English news

Both the GV-ENASC as well as the Arabic editorials and opinion sub-corpora have the lowest value of 55 KKW items each. However, looking again at the results from the perspective of the percentage of sub-corpora, it is the GV-ENASC editorials and opinion sub-corpus that has the highest value of (6.72), whereas the Arabic news sub-corpus has the lowest value (2.31). In addition, the example KKW items in Table 5.3 indicate that *Proximity* can be established through items belonging to

six main semantic categories of which three – that is, GEOGRAPHY, FAITH AND RELIGION and FUNCTION WORDS, are shared between the two main English as well as the Arabic sub-corpora. The remaining three categories, which only have one KKW item each, are in the Arabic editorials and opinion sub-corpus. These categories, along with their KKW items, include GOVERNMENT AND SECURITY: الأمة <al-'umma - the-nation.FEM.N – the-nation>; SOCIAL ACTIONS, STATES AND PROCESSES: المجتمع <al-mujtama' - the-society.MASC.N or community.MASC.N – the-society/ the-community>; and EMOTIONS: الشعبية <aš-ša'biy-ya - the-popular.ADJ-FEM – the popular>. As their meanings suggest, these items mainly convey *Proximity* in all of its aspects (*geographical, cultural and ideological*). However, analysis shows that other news values (e.g. *Personalization, Relevance, Negativity, Timeliness* and *Eliteness*) are also constructed by the same item (*the nation*), especially when collocates as well as the text type in which it occurs, are considered. For example 176 instances of the item الأمة <al-'umma - the-nation> – that is, 15% of a total of 1,191 concordance lines in the Arabic editorials and opinion sub-corpus, construct *Proximity*. On the other hand, 294 occurrences of the item الأمة <the-nation> - that is, 38% of a total of 780 concordance lines in Arabic news sub-corpus, construct *Eliteness*, which involves references to names of political parties and/or entities. For example, حزب الأمة القومي السوداني <National Umma Party, Sudan>; and مجلس الأمة الكويتي <National Assembly of Kuwait>. In the Arabic editorials and opinion sub-corpus, the item الأمة <the nation> constructs *Proximity* in almost 18% of 900 concordance lines through collocates, especially to the

immediate right (R1), that refer to a number of semantic categories. For example, items referring to GEOGRAPHY, as in the item الأمة العربية <the-Arab.FEM – *the Arab nation*> (117 occurrences: 13%); FAITH AND RELIGION: as in الأمة الإسلامية <the-Islamic.FEM – *the Islamic nation*> (24 occurrences); *Personalization* is also realized (in 0.055%) through collocates belonging to the SOCIAL STATES, ACTIONS AND to the immediate left (L1) as in جماهير الأمة <*crowds of-the-nation*> (20); شعوب الأمة <*peoples of-the-nation*> (15); and أبناء الأمة <*sons of-the-nation*> (14 occurrences). Additionally, *Negativity* is also constructed to the L1 of the item الأمة <*the-nation*> through collocates referring to GOVERNMENT AND SECURITY, as in أعداء الأمة <*enemies of-the-nation*>; or the FUNCTION WORDS category as with the KKW item ضد الأمة <*against the-nation*>, with 15 and 6 occurrences respectively.

In terms of the English sub-corpora, *Proximity* can be realized through KKW items (e.g. *Israel**, *U.S./ America**/ *Europe** and *Western*) belonging to the *Western and European* group of the GEOGRAPHY. The aforementioned items, however, are categorized under the news value of *Eliteness*, as they refer to elite nations and/or people. Example 51, nevertheless, shows how *Proximity* (for the Western and European audience) is constructed through geographically related KKW items (in bold), such as *Europe*, *EU*, *borders* and *Foreign*. As also shown in the Lead sentence in Example 51, the KKW item *its* of the FUNCTION WORDS category, also contributes to the construction of *Proximity*. In the headline of Example (11), however, there are at least five KKW items constructing a number of news values. *The Arab Spring*, which is categorized under *Agenda*, as it is the main event that

attracted the Arab and the international news media. The item *refugees*, which belongs to the GOVERNMENT AND SECURITY category, simultaneously constructs *Negativity* as well as *Personalization*, and these are the news values that are emphasized. Similarly, for the items *says* and *Hague*, these mainly construct both the news values of *Voice and Source* as well as *Eliteness* respectively:

51. HEADLINE *Arab Spring* refugees not welcome here, says
Hague

LEAD **Europe** is right to shut *its borders* to refugees
and migrants trying to enter the **EU** in the wake
of the *Arab Spring*, says William Hague, the
Foreign Secretary.

(NDPe-11-74/ news: 23-May)

In contexts in which the news value of *Eliteness* is emphasized, other news values, such as *Agenda* and *Relevance*, are also foregrounded. For example, in contexts of a governmental, socio-political and security nature, topics related to *democracy*, *elections*, *peace*, *freedom of speech and belief*, *human rights*, and *security* (see Table 5.2) are also emphasized. As Table 5.2 shows, the news value of *Agenda* is the fourth most key news value, in terms of the overall total number of KKW items. It has its highest value (23 KKW items) in the American editorials and opinion, whereas the lowest value (16 KKW items) is in the Arabic news sub-corpus. However, looking at the results from the perspective of the overall average percentage of each of the sub-corpora, it is the GV-ENASC editorials and opinion sub-corpus, which has the highest value of (4.41), whereas the Arabic news sub-

corpus has the lowest value of (2.35). Nevertheless, the news value of *Agenda* generally shows similar patterns across the different English as well as the Arabic text type sub-corpora in terms of the topics claiming it. Examples 52 and 53 show how *human rights* (Example 52, emboldened in the Lead), and *religious freedom* (Example 53, in the headline), are represented as important topics, and the fact that such topics are discussed by elite people or entities, intensifies their representation as high matters on *Agenda*. However, looking at Examples 52 and 53 from the perspective of CDA, it can be said that both *Eliteness* and *Agenda* related topics are politically and ideologically oriented, especially when other news values are considered. In other words, the representations of the news values in both Examples 52 and 53 reflect a polarized discourse that follows ‘the logic of Ingroup-Outgroup relations’ (van Dijk, 1998, p. 58). While Example 52 reflects a strategy of negative ‘outgroup’ (*Iranian officials*) representation, Example 53 embodies an instance of ‘ingroup’ (*allies of the United States*) representation. Neither the headline nor the Lead, in Example 53, mentions the names of the ‘allies’ involved in the repression of *religious freedom*, and that they are de-emphasized or backgrounded. Additionally, topics related to FAITH AND RELIGION, or other belief systems intensify the news value of *Consonance*, and they are well reported, especially if ‘ideological villains’, as van Dijk (1988b, p. 122) points out, are involved. For example, in his studies about the representation of Islam and Muslims in the British press, Baker (2010a, p. 310; 2012, p. 247) found that Muslims, in the tabloids, were mostly represented in contexts related to terrorism and extremism, focusing on a small

52. HEADLINE **US imposes sanctions on Iranian officials**
over abuses

LEAD **US President Barack Obama** has ordered unprecedented **sanctions against senior Iranian officials** for "sustained and severe violations of **human rights**".
(BBCe-10-149/ news: 29-Sep)

53. HEADLINE **State Dept. Report** says countries have repressed **religious freedom** with laws

LEAD Countries around the world, including **allies** of the **United States**, have used laws on blasphemy and apostasy to suppress political opponents, the **State Department** said on Monday in an annual **report** chronicling a grim decline in **religious freedom** that has resulted in rising bigotry and sectarian violence
(NYTe-13-51/ news: 21-May)

number of high-profile Muslim ‘villains’. In terms of the British broadsheets, (Baker, 2010a), Muslims were represented in a wider range of contexts, with a more restrained reporting stance. In some of the Arab Spring sub-corpora (English and Arabic), although extremism is mostly associated with Islam, the focus is more on groups/movements/parties rather than Muslims as ordinary people. For example, out of 73 instances, the only collocates of the KKW item *extremists* in the British

news, that relate to *Islam*, are the KKW forms *Islamic* and *Islamist*, with 7 and 5 co-occurrences respectively. Additionally, other KKW items of the FAITH AND RELIGION category, such as الطائفي <the-sectarian.MASC>; and الطائفية <the-sectarian.FEM/sectarianism.FEM> in Arabic; and *sectarianism* in English, are categorized as a realization of *Negativity*.

In terms of the news value of *Agenda* in the Arabic sub-corpus, there is quite a number of key topics that are shared with its English counterparts, as indicated by the example KKW items shown in Table 5.3. Example 54 represent contexts of the item السلام <the-peace>; and عملية السلام <peace process> in the Middle East, which are also considered as high matters in terms of *Agenda*.

54. HEADLINE	1	ملك الأردن: زيارة أوباما ستشجع السلام
	2	<i>Jordan's King: Obama's visit will</i>
	3	<i>promote peace</i>
LEAD	4	قال ملك الأردن عبد الله الثاني اليوم السبت إنه يأمل
	5	أن تعطي الزيارة المرتقبة للرئيس الأميركي باراك
	6	أوباما للمنطقة زخماً لعملية السلام، واصفاً ما يجري
	7	في المنطقة بأنه خطير جداً.
	8	<i>Jordan's King Abdulla II said today</i>
	9	<i>Saturday that he hopes the upcoming</i>
	10	<i>visit of the American President Barak</i>
	11	<i>Obama to the region would give impetus</i>
	12	<i>to the peace process, describing what is</i>
	13	<i>happening in the region as very</i>
		<i>dangerous.</i>

(JAZa-13-61/ news: 09-March)

This again suggests the fluidity as well as the interrelationship between the different semantic categories, and hence the news values they claim. For example, although *Eliteness*, as shown in Table 5.3, is the highest in terms of the total number of KKW items, of both English and Arabic (86 and 76 KKW items respectively), it is *Negativity* and *Personalization* that are the highest in terms of the number of semantic categories, with 7 categories each. For example, collective results show that *Negativity*, has its highest value, in terms of the overall number of KKW items of the English and Arabic sub-corpora, in four out of the seven semantic categories. These are GOVERNMENT AND SECURITY, PROTEST VOCABULARY, GENERAL AND ABSTRACT TERMS, and EMOTIONS, which are also shared (in terms of Negativity) between all the English and Arabic text type sub-corpora.

Table 5.4 Number of news values in English and Arabic KKW semantic categories

No. KKW semantic Category	Number and Names of News Values in the English and Arabic Semantic Categories		
	In ENASC	In ARASC	News value
1- GOVERNMENT AND SECURITY	7	7	<i>Negativity; Eliteness; Agenda; Consonance; Personalization; Unexpectedness; Relevance; Proximity</i>
2- SOCIAL ACTIONS, STATES AND PROCESSES	5	6	<i>Eliteness; Agenda; Personalization; Relevance; Negativity; Proximity</i>
3- GEOGRAPHY	4	3	<i>Proximity; Eliteness; Personalization; Superlativeness</i>
4- FUNCTION WORDS	4	4	<i>Timeliness; Proximity; Negativity; Personalization; Superlativeness</i>
5- FAITH AND RELIGION	4	4	<i>Consonance; Personalization; Proximity; Negativity</i>
6- LANGUAGE AND COMMUNICATION	4	3	<i>Source and Voice ; Agenda; Relevance; Relevance; Eliteness</i>
7- PROTEST VOCABULARY	3	3	<i>Negativity; Personalization; Agenda</i>
8- GENERAL AND ABSTRACT TERMS	3	5	<i>Negativity; Agenda; Relevance; Timeliness; Superlativeness</i>
9- EMOTIONS	4	2	<i>Negativity; Agenda; Positivity; Proximity</i>
10- NAME-PN	2	2	<i>Eliteness; Personalization</i>
Bold: News value shared between all English and Arabic text type sub-corpora			

5.4 Tracking News Values: Selected Peak Months

This section examines change over time of the main news values, in the English and Arabic sub-corpora, and their related topics in some of the selected peaks shown in Figures 4.4 and 4.6. Its main focus, however, is on the news value of *Eliteness*, as the most important KKW news value in the Arab Spring narrative. As discussed in Section 3.3, using the KKW technique in *WordSmith Tools* (Scott, 2015) is useful in identifying the peak months, their main topics, and hence the main news values during the period (15 June 2010 to 31 August 2013) set for the study. Analysis in this section focuses on four main months (December 2010, May 2011 and 2012, and July 2013) of which each one represents one peak and one year of the aforementioned period. Analysis begins with December 2010 as the month that immediately preceded the actual emergence of the Arab Spring, and the month that witnessed the catalyst event in Tunisia. Analysis in this section only focuses on the Headlines in which news values are mostly concentrated.

5.4.1 News Values Before and After the Arab Spring: Up to December 2010 and Afterwards and Selected Peaks

5.4.1.1 News Values Before the Arab Spring

Table 5.4 shows the most key news values that were prevalent during the peak of December 2010, along with their underlying semantic categories in both the English as well as the Arabic sub-corpus. As Table 5.4 shows, the six most key news values

during December 2010, in terms of the number of KKW items, are *Eliteness*, followed by *Proximity*, *Agenda*, *Source and Voice*, *Personalization*, and *Negativity*.

Table 5.5 News Values in December 2010

News value	No. of KKWs		Avg. % of sub-corpus	
	EN	AR	EN	AR
<i>Eliteness</i>				
GEOGRAPHY	10	9	28.12	3.58
NAME-PN	7	8	9.28	3.37
GOVERNMENT AND SECURITY	3	6	7.51	4.17
SOCIAL ACTIONS, STATES AND PROCESSES		3		4.97
FAITH AND RELIGION	1		7.04	
<i>Proximity</i>				
GEOGRAPHY	17	14	15.21	7.25
<i>Agenda</i>				
GOVERNMENT AND SECURITY	3	5	7.04	7.12
LANGUAGE AND COMMUNICATION	2	1	7.04	2.40
EMOTIONS	1		12.68	
SOCIAL ACTIONS, STATES AND PROCESSES	1		8.45	
<i>Source and Voice</i>				
LANGUAGE AND COMMUNICATION	4	7	10.92	3.64
<i>Personalization</i>				
SOCIAL ACTIONS, STATES AND PROCESSES		4		3.24
GEOGRAPHY	1	1	9.86	3.37
NAME-PN	1	1	9.86	2.88
FAITH AND RELIGION	1		11.27	
PROTEST VOCABULARY		1		4.33
<i>Negativity</i>				
PROTEST VOCABULARY		3		4.64
GOVERNMENT AND SECURITY		3		3.52
EMOTIONS		1		2.40
Grand Column Total	52	67	15.38	4.91

In terms of *Eliteness*, Table 5.4 shows that *Eliteness* has the highest number in terms of KKW items, comprising almost 40% of the resulting news values, during December 2010 in both the English as well as the Arabic sub-corpora. The English and Arabic sub-corpora also share a similar pattern in terms of *Proximity*, which relatively comprises 33% and 21% of the resulting news values during December 2010. However, the Arabic sub-corpus shows a different pattern in terms of the number and rank of the remaining news values. Although both the English and Arabic sub-corpora share the news value of *Agenda* (12% and 0.09% respectively), it is the news values of *Source and Voice*, *Personalization* and *Negativity* that simultaneously rank as number three, in terms of the number of KKW items in the Arabic sub-corpus. Nevertheless, as discussed earlier, different statistical measures yield different results. Looking collectively at Table 5.4, it is *Eliteness* and *Personalization* that have the highest number of KKW semantic categories, with 5 categories each.

There are three out five categories that are collectively shared between the English as well as the Arabic sub-corpora in terms of *Eliteness*, and of which GEOGRAPHY is the highest in terms of the overall number of KKW items. ‘The Israel set’ and ‘the United States set’, of the *Western and European Group*, are the only sets that construct *Eliteness* in terms of reference to elite countries or nations. They are shared between the English as well as the Arabic sub-corpora, except in the GV-ENASC in which ‘the United States set’ is not a KKW. In terms of their comprising KKW items in December 2010, ‘the Israel set’ contained the KKW items *Israel*,

Israeli, and *Jerusalem*, in the English sub-corpus, whereas ‘the United States set’ contained the KKW items *U*, *S*, and *US*. In the Arabic sub-corpus, the aforementioned KKW geographical sets contained the KKW items إسرائيل<*Israel*>; الإسرائيلي<*the-Israeli.MASC*>; الإسرائيلية<*the-Israeli.FEM*>; in ‘the Israel set’, whereas ‘the United States set’ includes the KKW items الأمريكي<*the-American.MASC*>; and الأمريكية<*the-American.MASC*>. However, the remaining KKW items (see. Tables 4.3 and 4.4) comprising the two sets were also included in the analysis. Comparing the two sets in terms of their presence in the English and Arabic Headlines, it was ‘the United States set’ that had the highest number of mentions, that is, 29 and 5 instances in a total of 173 and 236 headlines in the English and Arabic sub-corpora respectively. On the other hand, ‘the Israel set’ shows a similar pattern in both the English and Arabic headlines in terms of mentions, with 15 instances each. However, when taken individually, it was the British sub-corpus that had the highest mentions of ‘the Israel set’ and ‘the United States set’, with 9 and 18 instances out of 104 headline articles during December 2010.

Examples 55 and 56 show all instances of ‘the Israel set’ in the American and British headlines respectively. As Example 55 (lines 1, 2, 3 and 4) shows, the underlined KKW items associated with the news value of Eliteness, constructed by ‘the Israel set’ in the American news Headlines construct *Personalization*, through *Palestinian firefighters*; *Israeli girls* and *Arabs*; and *militant*, as indicated in Example 55 (lines 1, 2 and 4 respectively). For the Western and European audience, ‘the Israel set’ constructs *Eliteness* and *Proximity* that is mainly cultural in aspect.

Another news value that is associated with ‘the Israel set’ is that of *Negativity*, which is mainly constructed through the KKW items *army*, *kills*, and *militant* (line 4), along with other items, such as *criticized*, *barring* (line 1), and *avoid* (line 2). Simultaneously, the items *criticized* and *letter* establish the news value of *Source and Voice*, while the item *more* reflects the news value of *Superlativeness*. Looking at the Headlines in Example 55 from a CDA perspective, there are a number of presentation strategies that are employed in the representation of Israel and its related topics. The headline in Example 55 (line 1), for instance, reflects a strategy of exclusion (van Leeuwen, 2008, 1996), of which *backgrounding*, in this case, through passive agent deletion is one of its aspects. In other words, the headline in Example 55 (line 1) tells us that ‘Israel criticized’, but not who criticizes it. A closer look at the article from which the headline 1 (of Example 55) showed that the news actor, or the agent of criticizing, was not mentioned either in the headline or in the Lead sentence, which also reflects an ‘Us and Them’ pattern (van Dijk, 1998).

55. HEADLINES

- 1 **Israel** criticized for barring Palestinian firefighters
(CNNe-10-116/ news: 14-Dec)
- 2 Letter urges **Israeli** girls to avoid dating Arabs
(CNNe-10-127 / news: 29-Dec)
- 3 Gaza: **Israel** to Allow More Exports
(NYTe-10-125/ news: 09-Dec)
- 4 **Israeli** Army Kills Militant in Gaza
(NYTe-10-129/ news: 29-Dec)

In terms of the British Headlines of December 2010, there were seven out of eight headlines in which a number of KKW items of different semantic categories associated with ‘the Israel set’ appeared. These items also construe different news values. Example 56 shows seven headlines from the British sub-corpus in which ‘the Israel set’ is mentioned along with KKW items. The difference between both the American and the British headlines is that the British headlines contain more KKW items relating to NAME-PN category. For example, KKW items that construct *Eliteness*, such as *US*, *embassy*, *Hamas* and *Hezbollah*, as indicated in Example 56 (line 3), or *Britain* and *diplomatic* (headline 5), and *Obama’s* (line 7). Negativity is also associated with ‘the Israel set’, through the KKW items *anti* and *military* (line 2 and 4 respectively), but also through items such as *turn down*, *block*, *hangs*, *hangs* and *accused*, *risks* and *anger*, *collapse*, as well as the idiomatic term *in tatters* (headlines 1, 3, 4, 5, 6 and 7 respectively). KKW items, such as *peace*, *talks* and *settlements* (lines 1, and 6), make the news value of *Agenda* the third most key news value during December 2010. Finally, the KKW item *Palestinians* (line 1 and 5) also construct *Personalization*, and simultaneously reflect a strategy of *assimilation*. *Assimilation*, in van Leeuwen’s (2008, p. 37) terms, can be actualized by plurality, whereas *individualization* is constructed through singularity. The aforementioned strategies are important in CDA. While *assimilation* is usually employed in constructing ‘ordinary’ persons or countries, news media tend to individualize elite news/social actors or nation. (van Leeuwen, 2008, p. 37, 1996), as indicated in Example 56 (line 6). As can be seen, the underlined items in the

British and the American headlines, which are also KKW in December 2010, also construct different news values, shown in Table 5.4.

56. HEADLINES

- 1 Palestinians turn down **Israeli** peace proposal
(DTLe-10-44/ news: 29-Dec)
- 2 Rabbis' edict mirrors growth of anti-Arab feeling in **Israel**
(FTe-10-111/ news: 11-Dec)
- 3 The **US** embassy cables: Arms dealing: Hamas and Hezbollah: **US** used **Israel** intelligence to block arms from Iran and Syria
(GRDe-10-78/ news: 7-Dec)
- 4 Iran hangs man accused of passing military secrets to **Israel**
(NDPe-10-70 / news: 29-Dec)
- 5 Britain risks **Israeli** anger by extending diplomatic recognition to Palestinians
(NDPe-10-69/ news: 28-Dec)
- 6 Peace talks collapse as **Israel** stands ground on settlements
(TMSe-10-70/ news: 8-Dec)
- 7 Obama's peace plan in tatters after **Israel** calls his bluff on settlements; Middle East
(TMSe-10-71 / news: 9-Dec)

In the Arabic sub-corpus, on the other hand, 'the Israel set', as a realization of *Eliteness*, generally shows a similar pattern in terms of the other news values associated with it. However, *Negativity* is mainly the news values that is more associated with 'the Israel set' in the Arabic headlines in December 2010. As Example 57 shows, seven out of a total of 18 headlines in which the 'the Israel set' is mentioned. In Example 57 (headlines 3 and 5 along with English translation), the

underlined KKW items <المستوطنات>the-settlements.FEM.PL>; and <الاستيطان>of-the-settlement.MASC>as shown in Example 34 (headlines 3 and 6) can be considered as important matters in terms of *Agenda*, and belong to the GOVERNMENT AND SECURITY category.

57. HEADLINES

- 1 الحكومة تدين مؤتمرا إسرائيليا يحرض على "الوطن البديل"
*The government condemns an **Israeli** conference that incites "the Alternative Homeland"* (DUSa-10-114/ news; 07-Dec)
- 2 الآلاف يتظاهرون ضد العنصرية والتمييز في إسرائيل
*Thousands demonstrate against racism and discrimination in **Israel*** (DUSa-10-123/ news 11-Dec)
- 3 إسرائيل تثمن إسقاط أميركا للاستيطان
***Israel** praises **America's** drop of-the-settlements* (JAZa-10-107/ news: 13-Dec)
- 4 مستقبل المنطقة في قبضة التطرف الإسرائيلي
*Future of the region in the grip of the **Israeli** extremism* (SFRa-10-73/ opinion: 14-Dec)
- 5 أبو مازن [محمود عباس] لعشرات الإسرائيليين: استبدلنا ثقافة العنف.. والسلام أهم من المستوطنات
*Abu Mazen [Mahmoud Abbas] to tens of **Israelis**: We replaced the violence culture ... and peace is more important than the settlements* (WSTa-10-100/ news: 20-Dec)
- 6 قلق في لبنان.. وفي إسرائيل أيضا
*Concerns in Lebanon .. and in **Israel** too* (QAa-10-122/ opinion: 23-Dec)
- 7 الناطق باسم الحكومة الإسرائيلية يعتبر حماس دولة لها جيشها واسلحتها ويطالب بمواجهتها
***Israeli** government spokesman considers Hamas as a state that has its army and weapons, and demands its confrontation* (QAa-10-128/ news: 28-Dec)

Other underlined KKW from the GEOGRAPHY category are أمريكا<America>; and لبنان<Lebanon>, which construct *Eliteness* and *Proximity*, as indicated in Example 34 (lines 3 and 6 respectively). In terms of the NAME-PN category, the underlined item أبو مازن<Abu-Mazen (or Abbas)>; and the KKW items *Hamas*, construct *Eliteness* as indicated in Example 57 (Headlines 5 and 7 respectively). Looking at Example 57 from the perspective of the Arabic grammatical forms, that is, the masculine and feminine, of the KKW items comprising ‘the Israel set’, it can be said that, through collocates, different news values are realized. According to Stubbs (1996, p. 172), ‘different word forms can have quite different collocates’. For example, the definite masculine adjective الإسرائيلي<al-’isrā’īlī – the-Israeli.MASC.ADJ – *the Israeli*> in Example 57 (headline 4), post-modifies the item التطرف<at-taṭarruf – (*the*)-*extremism*>, and hence constructs *Negativity*. On the other hand, the definite feminine الإسرائيلية<al-’isrā’īliyya – the-Israeli.FEM.ADJ – *the Israeli*> post-modifies the item الحكومة<al-ḥukūma – the-government.FEM.N – *the government*>, which intensifies *Eliteness*, constructed by the item الناطق باسم الحكومة<an-nāṭiq – *the spokesman*>, while simultaneously constructing the news value of *Source and Voice*, through the items الناطق<an-nāṭiq – *the spokesman*>; and يعتبر<ya’tabir – considers.SING.MASC.PRS.V - *considers*>.

In terms of the news values that are directly related to the peak events, in Tunisia (see. Figure 4.7), that immediately preceded the *Arab Spring* <الربيع العربي>, these mainly include *Proximity*, *Negativity*, *Personalization*, *Eliteness*, and *Source and Voice*. In terms of *Proximity*, for example, there were 73 mentions, in 65 (61 news

and 4 opinions and editorials) out of 236 Arabic Headlines articles, of KKW items relating to ‘the Tunisia set’ (see Table 4.4), which relatively equals 32.62 PTW in the Arabic headline sub-corpus in December 2010. The KKW item تونس<Tunisia> has the highest value (33 mentions) in term of mentions in the December 2010 Arabic Headlines. This is followed by the KKW item بوزيد<Bouzid>; التونسي<the-Tunisian.MASC.ADJ- the Tunisian>; and التونسية<the-Tunisian.FEM.ADJ – the Tunisian>, with 26, 6 and 5 mentions respectively. *Negativity*, as Example 35 (headline 1) shows, is constructed through the KKW item متظاهرين<demonstrators> of the PROTEST VOCABULARY, along with the items انتحار<suicide>; عاطل<unemployed>; مواجهات<confrontations>, which are also collocates of ‘the Tunisia set’, with at least four co-occurrences and minimum of 3.00 MI score. Example 58 (headline 2) shows realizations of *Eliteness* (in bold), such as بن علي<Ben Ali>; and وزير الإعلام<Minister of Information>; along with the news value of *Source and Voice*, through the items (underlined) يعلن<declares>; and ينتقد<criticizes>.

58. HEADLINES

- 1 مواجهات بين الشرطة ومتظاهرين إثر انتحار شاب عاطل في تونس
*Confrontations between police and **demonstrators** after
suicide of an **unemployed** youth in Tunisia*
(RBYa-10-170/ news: 23-Dec)
- 2 تونس: أحداث سيدي بوزيد تطيح بوزير الإعلام والرئيس بن علي يعلن
تفهمه لها وينتقد استغلالها سياسيا
*Tunisia: Incidents of Sidi Bouzid topple the **Minister of
Information and President Ben Ali** declares his
understanding them and criticizes them being politically
exploited*
(WSTe-10-118/ news: 30-Dec)

As we have seen, the news values in the immediate months before the emergence of the Arab Spring story in the world's news, focused on *Eliteness* and *Proximity* as the top two.

5.4.1.2 News Values after the Arab Spring

After the emergence of the story and for the period of the remainder of the corpus between January 2011 and June 2013 during the peaks of the story shown in Figure 4.8, the same two news values are the most prominent (as Table 5.5 shows), but with different events and different people. For example, in May 2011 there was a peak event in the Arabic corpus which was the G8 summit, which coincided with Bin Laden's death, but in the period before the Arab Spring the Tunisian issues were reported and shown as key in the Arabic, but these were not key in the British news. These two news values (*Eliteness* and *Proximity*) are the most persistent across my corpus and this is the pattern for news more generally too, as Bell (1991) tells us. Therefore, whatever the story, it is the most powerful who are seen as most newsworthy. As the peaks indicate, for example in May 2011, it was *Obama* who was the most written about person, while in July 2013 President *Morsi* of Egypt was ousted.

Table 5.6 News values during peak months in the English and Arabic after the Arab Spring

News Value	No. of KKWS		Avg. % of sub-corpus	
	English	Arabic	English	Arabic
Eliteness	117	92	17.27	4.23
Proximity	113	81	16.96	4.43
Agenda	36	27	12.35	3.89
Negativity	31	20	10.90	4.40
Personalization	27	13	15.14	3.01
Consonance	22	6	12.63	3.94
Source and voice	9	17	11.71	4.10
Timeliness	7	8	9.77	4.20
Relevance	3	5	7.63	3.16
Unexpectedness	3	3	27.29	3.62
Superlativeness		2		1.91

5.5 Summary

In this chapter, we have seen that newsworthiness in the Arab Spring corpus is related to *Eliteness* and *Proximity*, showing that this is a narrative of power. *Proximity* is not only geographical but cultural and ideological. The massive events across the Middle East also resulted in *Negativity* being the third most significant value, where topics related to the GOVERNMENT AND SECURITY were the most important. *Personalization* is also an important news value, though it is politically oriented rather than located at the level of the ordinary individual.

The media play an important role in framing and representing whatever symbol they want, for example by representing *Bouazizi* as the spark of the revolution. The least important news value is *Positivity* through the word *peaceful*, which is only used in the American and the British news and which is relatively absent from the Arabic and GV sub-corpora. We can see that *Relevance* is also more important than geographical proximity, for example through words such as *alliance*, *relations* and *resolutions*. However, many decisions taken which affected people in the Middle East were taken by the United Nations outside the region.

6. Conclusion

This study had two main goals. The first was to contribute to the academic research interested in using corpus linguistic techniques to conduct critical discourse analysis. The use of larger datasets and their analysis through quantitative and qualitative methods allows for a better understanding of the patterns of meaning within news discourses. Secondly, given the important role the media play in constructing social reality, this study provided more understanding of language use in the news media and in particular the representation of the Arab Spring narrative from the perspectives of different languages and cultures, that is English, with its three main forms (American, British and global English), along with Arabic.

The main research questions that the study attempted to answer were

- 1- Who are the main news actors in the Arab Spring corpus, and what strategies are employed in the representation of the main news actors?
- 2- What are the main topics and themes associated with Arab Spring, and how they are presented in the different news media outlets in terms of language use?
- 3- What are the news values that have made the Arab Spring a newsworthy story and how KKW function can be used in identifying the news values prevalent in the Arab Spring narrative?

The following section provides a summary of my conclusions in relation to the results of the study.

6.1 Summary of Conclusions in Relation to Results

This section provides a summary of my observations from the three analysis chapters. Each chapter is concerned with one of the research questions. By answering each research question, my study aimed to identify *who* the main news actors are, and what their role is in the narrative. Combining van Leeuwen's (2008, 1996) Social Actor Network along with the KKW function (that is, the KKW semantic categories) as well as other corpus linguistic techniques (that is, concordance, collocation and clusters technique) revealed the different discursive strategies employed in the representation of the news actors (individuals and groups/organizations) at the linguistic as well as the socio-semantic levels. The representation strategies, that is *nomination*, *functionalization*, *classification*, *assimilation* and *objectivation*, identified in section 3.5 and its related sub-sections, confirmed the initial observation that the Arab Spring is a narrative of the powerful and the elite. For example, in sub-section 3.5.1, analysis showed that most relatively included news actors (of different semantic categories) were also functionalized through KKW items denoting high-status or high-profile roles of both English and Arabic, though the number of high-profile roles of the GOVERNMENT AND SECURITY are higher in the English than that of the Arabic. For example, although both the English and the Arabic share a set of high-status roles that are relatively significant and foregrounded, such as *(prime) minister*, *president*, *ambassador*, other elite roles, such as *officials*, *activists* and *colonel* are only relatively present and significant in either the whole or some of the English sub-corpora (see Tables. 3.3

and 3.4). Another example is that the number of ‘ordinary’ roles (that is, *militants, fighters, extremists, refugees, civilians, citizens, voters*) of the GOVERNMENT AND SECURITY of the English sub-corpus is higher than that of the Arabic sub-corpus of which المواطن <*the-citizen*> is relatively present, yet backgrounded, in the Arabic editorials and opinions. The remaining representation strategies identified in this chapter also suggest the significance of eliteness of the news actors involved in the Arab Spring narrative. In sub-section 3.5.2 for example, analysis showed how the news actors (elite and ordinary) were mostly classified and defined in terms of two main classification categories: *provenance* (that is, geographical identities) and *religion*. Those news actors who were mostly referred to in terms of *provenance* and *religion* belong to the elite (individuals and groups/organizations). In terms of the religious classification, this was more significant in the English sub-corpus in terms of the number of KKW items and their keyness than that of the Arabic. While *Islamist* is the most key and relatively included in the American news, *Shia* is the top classifying KKW item of the British news sub-corpus. The KKW items *Muslim* is the most key across the remaining English sub-corpora, that is, both of the GV text types as well as the American and the British editorials and opinions sub-corpora. Some of these KKW items (that is, *Islamists, Muslims, Jews* and *Christians*) can be said that they are more personalizing of news actors than others, which also reflect other aspect of assimilation (see 3.5.3). Additionally, the two KKW items *Islamic* and *Islamist* were used interchangeably when referring to different news actors (people and groups) although the item *Islamist* is commonly

meant to refer to Muslim individuals and groups/ organizations advocating Islamic political rule (see Figure 3.11). Analysis also showed that both forms *Islamic* and *Islamist* collocate with negative words such as *terror*, *terrorism*, *extremist/s*, *militant/s*, *violence* and *threat*, and *regime/s*. A similar pattern was identified in Arabic for one grammatical form, the feminine الإسلامية <the-Islamic/Islamist> in the sense that it was used to define or refer to news actors, mostly groups and organizations (see. Figure 3.12). The different representation strategies identified in the Arab Spring narrative helped much in revealing *who* is being reported and *who* is absent, and what roles they occupy (see 3.6 and 3.7). In the Arab Spring, the elite are the ultimate news actors and the news source. The ordinary actors are relatively backgrounded and muted, and hence can be considered as background news actors whose presence is only important to the personalizing effects of the narrative.

The second analytical chapter, that is Chapter 4, investigated *what* the main topics/events are that make up the international and Arab and Islamic news media, using corpus linguistic techniques to identify *where* the Arab Spring events took place, and what countries and nations other than the ‘Arab world’ were involved. Analysis in this chapter showed the usefulness of the KKWs in identifying the topics associated with the Arab Spring narrative in my corpus. We have seen that geographical regions and places are like actors. There is an almost similar pattern of the *where* and the *who*, making the non-human and human place and faces interchangeable. In his study on news media and headlines Bell (1991) showed that

place is the third most important element (after actors and time) that appears in the headline and it sometimes replaces actors. The pattern showing place as an actor is prevalent in both the English and Arabic sub-corpora as shown in sub-section 4.4 (Figures 4.14 and 4.15 along with Example 37). This pattern was also revealed by the objectivation strategies (sub-section 3.5.4 and Figure 3.13). This means that actors are more important than *where* the news takes place. The government topic is also dominant in this narrative, but the presence of the word *elections* does not indicate democracy is being achieved. Words like *elections* and *democracy* are words in disguise; they have other functions, such as drawing the people's attention from real issues like jobs and family. The word *democracy* hides behind it the issues of *security*, *fear* and *terrorism*. The *security aspects* of the Arab Spring were revealed in more than a context (e.g. *peace* and *violence*, *terror* and *threat*). Analysis related to the EMOTIONS category showed how *security* and *peace* are two topics that are linked together (see sub-section 4.2, Table 4.5 and Figures 4.7 and 4.8). We will see in the next chapter that these topics are associated with news values of *Negativity*. We looked at the use of the function words of both English and Arabic (e.g. *pro-*, *anti-*, *between* (بين), and *against* (ضد)), which show as KKW's, indicating that function words, far from being insignificant, carry important meanings that indicate conflict, difference and opposition. These words are closely related to the protest vocabulary and also indicate the emotional content of the narrative (see sub-section 4.2 and Figures 4.5)

In the last analytical chapter, we look at the newsworthiness and news values of these KKWs and semantic macrostructures in the corpus in order to assess *why* these words and concepts are selected by the different English and Arabic news media. In the last analytical chapter, my study looked into *why* the Arab Spring and its associated events attracted so much attention, not only for the news media, but also for researchers and commentators from different domains and disciplines. In this chapter, we have seen that newsworthiness in the Arab Spring corpus is related to *Eliteness* and *Proximity*, showing that this is a narrative of power. *Proximity* is not only geographical but cultural and ideological. The massive events across the Middle East also resulted in *Negativity* being the third most significant value (see subsection 5.2.1), where topics related to GOVERNMENT AND SECURITY were the most important. *Personalization*, as analysis in sub-section 5.2.8 shows, is also an important news value, though it is politically oriented rather than located at the level of the ordinary individual. analysis in this chapter showed how the media play an important role in framing and representing whatever symbol they want, for example by representing Bouazizi as the spark of the revolution. The least important news value is *Positivity* through the word *peaceful*, which is only used in the American and British news and which is relatively absent from the Arabic and GV sub-corpora. We can see that *Relevance* (see 5.2.4) is also more important than geographical proximity as it is more about the way events and decisions may affect our lives; for example, through words such as *alliance*, *relations* and *resolutions*. However, many decisions taken which affected people in the Middle East were

taken by the *United Nations* outside the region. Analysis of the news actors (see. 3.4.5 and Figure 3.14) in the first analysis chapter also showed evidence of the importance of the news value of Relevance,

6.2 Arab Spring; A Narrative of the Powerful

The first question sought to identify the main news actors in the study corpora: the American, the British, and the global variety of English as well as the Arabic. The categorization of the KKW items in the NAME-PN category led to the identification of five main groups of which the *Western and European Group* is the most key in the Arab Spring narrative corpora as a whole. The dominant news actors in this group are the *United States*, *Israel*, and *France*, which shows that both expected and less expected news actors form prominent roles in the story. France's position in the story is accounted for by President Sarkozy's early role in the Arab Spring narrative, though he disappeared from it after the death of Gaddafi. As well as Sarkozy, the narrative included other high-profile political leaders, such as *Obama*, *Clinton* and *Netanyahu* as well as organizations such as the *United Nations*, the *UN Security Council* and the *EU*. In other words, the most included news actors in terms of their relative keyness belong to what Fairclough (2001) calls, the dominant bloc. Comparing the KKW lists thus revealed an overview of who was relatively included and excluded in the English and Arabic sub-corpora in the four remaining categories: the *Arab Nationals*, the *International*, the *Islamic* and the *Russia Groups*. Analysis showed that inclusion and exclusion follows not only and

ideological, but also power relations, alliance and the interests of the dominant group patterns. As the results in Figure 3.3 showed, the number of news actors belonging to the sub-category GROUPS AND ORGANIZATIONS in the English (the American, the British, and the global variety) sub-corpus is higher than that of its Arabic sub-corpora. There is a difference in the pattern of news actors that represent the groups. They are mainly Muslim in the English corpus and the pattern is that we have *Muslim Brotherhood*, *Hamas*, *Hezbollah*, *Fatah* and the *Arab League*. On the other hand, though the Arabic corpus shares *Muslim Brotherhood* and *Hamas* with the English, it does not contain the *Arab League* or *Fatah*. The groups are presented collectively and generally their leaders are not given a voice. However, *Haniyeh* (leader of Hamas) was key in the Arabic sub-corpus and *Abbas* of the *Palestinian Authority* was key in almost all of the English sub-corpora.

Other aspects of power and Eliteness were shown through the different representation strategies (see sections 3.3 and 3.5) and the roles identified. For example, through *functionalization*, quite a number of KKW items of different semantic categories of both English and Arabic lend themselves as linguistic realizations of high-status and high-profile roles. Of these semantic categories is GOVERNMENT AND SECURITY (e.g. *president* (الرئيس), *minister* (وزير), *officials*, *activists*, *ambassador* (السفير); and colonel); of the SOCIAL ACTIONS, STATES AND PROCESSES, we have a number of high-status roles such as *king*, *leaders*, *sheikh* and ‘dictators’ (Table 3.7), which is higher than that of the Arabic which only has three (see Table 3.9) elite roles (الملك<the-king>; قائد<leader>; and الطغاة<the-tyrants>). of

LANGUAGE AND COMMUNICATION we only have *journalists* which are relatively present in four out of the six main English sub-corpora.

Eliteness through collectivization/objectivation was also prevalent. For example, the GOVERNMENT AND SECURITY (e.g. *government/s, regime, administration, council, army, military and intelligence*). of GEOGRAPHY there are various items relating to the *Western and European Group* (e.g. *US, America, Israel, Europe, France and Western*). Even in such impersonalized representation (see sub-section 3.54 and section 3.6), items such as *report/s, surveys, statements, speech* and *resolution* lend a kind of impersonal authority to the utterance, which is often used in connection with the utterances of high-profile, official or elite news/social actor/s or spokesperson/s. *Who says?* is one of the fundamental questions of news work, and the different representation strategies discussed in the first analysis chapter revealed the nature of the news sources and actors of which most represent the powerful and the elite in the Arab Spring narrative.

6.3 What is the Arab Spring?

In the second analysis chapter, KKW lists were compared to examine the main topics that are associated with the Arab Spring, in an attempt to have a better understanding of the political and religious context of the Arab Spring story. What I found was that for example in the FAITH AND RELIGION category there were lots of terms that were not really describing faith and religion but instead were actually hiding sectarianism and conflict (see sub-section 3.5.2 and section 4.2). Analysis of

the representation strategies of news actors (see sub-section 3.5.1) showed evidence that the Arab Spring narrative is a narrative of power and alliances and of weapon deals (see Figure 3.6) and not really a narrative about the suffering of the people and their democratization. The functionalization of news actors through KKW items of the PROTEST VOCABULARY (Table 3.5), revealed much about the issues relating to weapon deals and the controversy regarding arming of the Syrian and the Libyan *rebels* (الثوار) by the West (especially the US and the UK) and their Arab allies in the region. This was also more evident in the analysis concerning two of the top ten categories. Analysis of news values also revealed patterns of power relations and alliances through the news values of *Eliteness* and *Source and Voice* (sub-section 5.2.5 and Example 46), and *Relevance* (see 5.2.4). In September 2013, the *OED* added a new entry for the term and defined it simply as '[a] process or period of political or cultural liberalization in the Arab world; (now) *spec.* a series of anti-government or pro-democratic uprisings and demonstrations in various countries in North Africa and the Middle East, beginning in Tunisia in December 2010'. However, my analysis has shown that the Arab Spring narrative is far from being a simple process of democratization or liberalization; it is one of destruction.

Keyword analysis identified top ten semantic categories of both English and Arabic (see Table 4.1). These are GEOGRAPHY; GOVERNMENT AND SECURITY; NAME-PN; FUNCTION WORDS; SOCIAL ACTIONS, STATES AND PROCESSES; FAITH AND RELIGION; PROTEST VOCABULARY; LANGUAGE AND COMMUNICATION; GENERAL AND ABSTRACT TERMS; and EMOTIONS. Each category has its own story. For example, in

the GOVERNMENT AND SECURITY category we see words such as *democracy*, *elections*, and high profile titles such as *president*, *minister*, *government* and *regime*. These words present a picture of dominance in which ordinary people fail to have a voice or a presence. However, both the English and Arabic sub-corpora showed a difference in focus during peak events of the Arab Spring (see section 4.3, Figures 4.10 and 4.12). Further analysis also showed that contrasting themes/concepts within these and other categories (by means of pairs of items with positive/negative connotations) are also prevalent. For example, at the lexical level the following contrasting pairs (for both English and Arabic) are revealed: *government/regime*; *security/terrorism*; *democracy/dictatorship*; *religious*, *sectarian/secular*; *peace/violence*; *allies/enemy*; *corruption/reform*; *opposition/ support*, to mention only a few. Similarly, at the grammatical level items, such as *pro/anti*, *is/non-* and *not*, also indicate the contrastive as well as the polarizing nature of the Arab Spring narrative. The narrative is therefore one of continuing and persistent opposition and conflict, rather than one of liberation and resolution.

6.4 What is Newsworthy about the Arab Spring?

Analysis in Chapter 5 evaluates the newsworthiness of the Arab Spring and its associated events from the viewpoint of the different news media outlets representing different cultures and ideologies. I use the resulting KKW lists and categorize them in terms of their potential to construct one or more news values. Based on Bell's (1991, p. 92) news values in news events and actors, the

categorization of the KKW lists led to the identification of twelve main news values of which *Eliteness* (that is, of news actors and news source) has the most significance in terms of the percentage of each of the English and Arabic sub-corpora. This again suggests that the Arab Spring might be a narrative of power and eliteness rather than only a freedom or democracy movement. The involvement of the powerful and the overthrow of the powerful regimes in the Arab Spring countries attracted the attention of news media, through not only *Eliteness*, but also *Negativity*, which is the third most significant news value in the Arab Spring across the Arab Spring sub-corpora, after *Proximity*. *Proximity*, as analysis showed, is not only about geographical closeness, but also cultural and ideological place. There were many instances in which cultural and ideological *Proximity* were also significant. If we look at the category of GOVERNMENT AND SECURITY, this is predominately constructed through the news value of *Negativity*. *Personalization* is employed ideologically and is politically motivated; it is not about ordinary people unless they are victims. *Personalization*, in Chomsky's terms, highlights the people who are the victims of their governments. *Personalization* reveals so many things about the nature of dominance, power relations and the actions of the elite and powerful. *Relevance*, many of the decisions taken outside the Middle East which directly affected the people there, such as the military intervention in Libya and the bombing of Tripoli, the intervention in Syrian. However, *Relevance* is not always defined in terms of the reading public or audience, but rather, as in the case of the

Arab Spring narrative, in terms of the elite and powerful groups (see sub-section 5.2.4 and 5.2.5)

6.5 Contribution to Knowledge

This study has contributed to the body of research in corpus linguistics and critical discourse analysis in terms of its topic and methodology. Firstly, there is little research, to my knowledge, that has been done on the Arab Spring with a huge amount of data collected from the most influential news media in the English and Arabic-speaking worlds. My study conducted an interdisciplinary as well as multidisciplinary investigation to analyse the representation of the Arab Spring narrative from different perspectives and more than two cultures, which basically reflects the American, the British as well as the Arab and Islamic cultures. Secondly, my research has also contributed to the academic research interested in using corpus linguistic techniques, and conducting critical discourse analysis. The use of the different corpus linguistic techniques facilitates collecting larger data volumes as well as analysing them at quantitative and qualitative levels. My study provided more evidence of the potential of corpus linguistic methodology in conducting critical discourse analysis. Thirdly, my study provides also new techniques for processing large amounts of data through, for example, the Quick Analysis Tool in Excel, which allows for the processing of larger amounts of data.

6.6 Implications and Further Research

There are a number of important methodological implications. The first implication relates to the presentation of Arabic data and its contextualization. Translation and transliteration of Arabic data is very time consuming. This is one of the aspects of this thesis that took the most of my time. There are many decisions that have to be made in terms of translating literally or idiomatically and also representing a left to right language and right to left language on the screen in tables and figures is a daunting task. There are also issues in relation to the reference corpora. For instance, is it fair to compare each different language corpus with a different reference corpus? Each reference corpus will provide different results. The Arabic Wikipedia reference corpus was a compromise, as the language there is not representative of Modern Standard Arabic and anyone can add text. It therefore contains many Arabic dialects and even machine translation, so it is not a really satisfactory reference corpus. Even with the other Arabic reference corpora freely available for researchers to use, issues of representativeness in terms of size, language varieties and time are still some of the main challenges facing the Arabic reference corpora. Nevertheless, there are limitations to the BNC too, and there is never, perhaps, a perfect reference corpus. I also feel it is not fair to simply combine all the different news sources in a corpus, as they represent very different political and ideological positions; some are more inflammatory or exaggerated and others create stories that circulate and maintain conspiracy theories rather than news. Other news sources describe themselves as Pan-Arab, but they are far from that; they are

aligned with different sectarian groups. However, in spite of all of these and other limitations, the benefits of looking at the same story in two very different languages are great, since there are fascinating strands of meaning to be extracted from the corpus. There is much further work to do to tease out all the differences between the Arab Spring narrative in Arabic and in English and to understand the cultural differences between the two.

References

- Abaza, H. 2010. Is is Islamic or Islamist? *Newsweek*. [Online]. [Accessed 10 April 2014]. Available from: <http://www.newsweek.com/it-islamic-or-islamist-73961>.
- Abdullah, A. R. A. 2014. *An Ethnographic Sociolinguistic Study of Virtual Identity in Second Life*. PhD thesis, University of Leeds.
- Aday, S., et al. 2013 Watching From Afar: Media Consumption Patterns Around the Arab Spring. *American Behavioral Scientist*. **57**(7), pp.899-919.
- Al-Hejin, B. 2015 Covering Muslim Women: Semantic Macrostructures in BBC News. *Discourse & Communication*. **9**(1), pp.19-46.
- Al-Husseini, F. 2014. *Waves of Arab Democratisation*. [Online]. [Accessed 30 December 2014]. Available from: <http://weekly.ahram.org.eg/News/5292/21/Waves-of-Arab-democratisation.aspx>.
- Al-Sulaiti, L. and Atwell, E. S. 2006 The design of a corpus of Contemporary Arabic. *International Journal of Corpus Linguistics*. **11**(2), pp.135-171.
- Alansary, S. and Nagi, M. 2014. The International Corpus of Arabic: Compilation, Analysis and Evaluation. In: *EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP)*, Doha, Qatar. pp.8-17.
- Alansary, S., et al. 2007. Building an International Corpus of Arabic (ICA): Progress of Compilation Stage. In: *7th International Conference on Language Engineering*, Cairo, Egypt.
- Amanat, A. 2012 The Spring of Hope and Winter of Despair. *International Journal of Middle East Studies*. **44**(01), pp.147-149.
- Archer, D. 2009. Promoting the Wider Use of Word Frequency and Keyword Extraction Techniques. In: Archer, D. ed. *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*. Farnham: Ashgate, pp.159-162.

- Archer, D., et al. 2009. Love - 'A Familiar or Devil'? An Explanation of Key Domains in Shakespeare's Comedies and Tragedies. In: Archer, D. ed. *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*. Farnham: Ashgate, pp.137-157.
- Arts, T., et al. 2014 arTenTen: Arabic Corpus and Word Sketches. *Journal of King Saud University - Computer and Information Sciences*. **26**(4), pp.357–371.
- Bachman, I. 2011 Civil partnership – “Gay marriage in all but name”: A corpus-driven analysis of discourses of same-sex relationships in the UK Parliament. *Corpora*. **6**(1), pp.77-105.
- Baker, P. 2004 Querying Keywords: Questions of Difference, Frequency, and Sense in Keywords Analysis. *Journal of English Linguistics*. **32**(4), pp.346-359.
- Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. 2009. The Question Is, How Cruel Is It? Keywords, Fox Hunting and the House of Commons. In: Archer, D. ed. *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*. Farnhama: Ashgate, pp.125-136.
- Baker, P. 2010a Representations of Islam in British Broadsheet and Tabloid Newspapers 1999-2005. *Journal of Language and Politics*. **9**(2), pp.310-338.
- Baker, P. 2010b. *Sociolinguistics and Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Baker, P. 2012 Acceptable Bias? Using Corpus Linguistics Methods With Critical Discourse Analysis. *Critical Discourse Studies*. **9**(3), pp.247–256.
- Baker, P., et al. 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), pp.273-306.
- Baker, P., et al. 2013a Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998-2009. *Applied Linguistics*. **34**(3), pp.255-278.

- Baker, P., et al. 2013b. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Baker, P. and McEnery, T. eds. 2015a. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan.
- Baker, P. and McEnery, T. 2015b. Introduction. In: Baker, P. and McEnery, T. eds. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan, pp.1-19.
- Bang, M. 2003. *A Corpus Based Study of Representation of Foreign Countries in the South Korean Press*. MA thesis, University of Birmingham.
- Baron, B. and Pursley, S. 2011 EDITORIAL FOREWORD. *International Journal of Middle East Studies*. **43**(03), pp.379-382.
- Barton, G. 2012 Reflections on the 'Arab Spring': introduction. *Islam and Christian-Muslim Relations*. **23**(4), pp.415-416.
- Bazzi, S. 2009. *Arab News and Conflict: A Multidisciplinary Discourse Study*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Bednarek, M. 2015 Voices and Values in the News: News Media Talk, News Values and Attribution. *Discourse, Context & Media*.
- Bednarek, M. 2016 Investigating Evaluation and News Values in News Items that are Shared Through Social Media. *Corpora*. **11**(2), pp.227-257.
- Bednarek, M. and Caple, H. 2012 'Value added': Language, image and news values. *Discourse, Context & Media*. **1**(2-3), pp.103-113.
- Bednarek, M. and Caple, H. 2014 Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond. *Discourse & Society*. **25**(2), pp.135-158.
- Bell, A. 1991. *The Language of News Media*. Oxford: Blackwell.
- Bell, A. 1995 Language and the Media. *Annual Review of Applied Linguistics*. **15**, pp.23-41.

- Bell, A. 1998. The Discourse Structure of News Stories. In: Bell, A. and Garrett, P. eds. *Approaches to Media Discourse*. Oxford: Blackwell Publishers Ltd., pp.64-104.
- Bennett, W. L. and Segerberg, A. 2012 The Logic of Connective Action Digital Media and the Personalization of Contentious Politics. *Information Communication & Society*. **15**(5), pp.739-768.
- Biber, D. 1993 Representativeness in Corpus Design. *Literary and Linguistic Computing*. **8**(4), pp.243-257.
- Biber, D., et al. 2007. Rhetorical Appeals in Fundraising. In: Biber, D., et al. eds. *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp.121-151.
- Biber, D., et al. 2004 If you look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*. **25**(3), pp.371-405.
- Biber, D., et al. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Blommaert, J. 2005. *Discourse: A Critical Introduction*. Cambridge: Cambridge University Press.
- Bolívar, A. 1994. The structure of newspaper editorials. In: Coulthard, M. ed. *Advances in Written Text Analysis*. London and New York: Routledge, pp.276-294.
- Bolívar, A. 2001. The Negotiation of Evaluation in Written Text. In: Scott, M. and Thompson, G. eds. *Patterns of Text: In honour of Michael Hoey*. Amsterdam and Philadelphia: John Benjamins, pp.129-158.
- Bondi, M. 2010. Perspectives on Keywords and Keyness: An Introduction. In: Bondi, M. and Scott, M. eds. *Keyness in Texts*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp.1-18.
- Bondi, M. and Scott, M. eds. 2010. *Keyness in Texts*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Brighton, P. and Foy, D. 2007. *News Values*. London: Sage.

- Bruckman, A. 2002. Ethical Guidelines for Research Online. [Online]. [Accessed 25 December 2013]. Available from: <http://www.cc.gatech.edu/~asb/ethics/>.
- Caldas-Coulthard, C. R. 1994. On Reporting the Representation of Speech in Factual and Factional Narratives. In: Coulthard, M. ed. *Advances in Written Text Analysis*. London and New York: Routledge, pp.295-308.
- Choueiri, Y. M. 2005. *A Companion to the History of the Middle East*. USA: Blackwell Publishing.
- Christensen, M. and Christensen, C. 2013 The Arab Spring as meta-event and communicative spaces. *Television and New Media*. **14**(4), pp.351-364.
- Comrie, B., et al. 2008. *Leipzig Glossing Rules* [Online]. [Accessed 25 December 2013]. Available from: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>.
- Cottle, S. 2008 Reporting demonstrations: the changing media politics of dissent. *Media Culture & Society*. **30**(6), pp.853-+.
- Cottle, S. 2011 Media and the Arab uprisings of 2011: Research notes. *Journalism*. **12**(5), pp.647-659.
- Culpeper, J. 2002. Computers, Language and Characterisation: An Analysis of Six Characters in Romeo and Juliet. In: Melander-Marttala, U., et al., eds. *Conversation in Life and in Literature: Papers from the ASLA Symposium, Association Suedoise de Linguistique Appliquée (ASLA), Uppsala*. Universitetsstryckeriet, pp.11-30.
- Culpeper, J. 2009 Keyness: Words, Parts-of-Speech and Semantic Categories in the Character-Talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*. **14**(1), pp.29-59.
- Davies, M. 2013. *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE)*. Available from: <http://corpus.byu.edu/glowbe/>.
- Duguid, A. 2010 Investigating *anti* and some Reflections on Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS). *Corpora*. **5**(20), pp.191-220.

- Dukes, K. 2009. *The Quranic Arabic Corpus*. [Online]. [Accessed 10 January 2013]. Available from: <http://corpus.quran.com/>.
- Dunning, T. 1993 Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*. **19**(1), pp.61-74.
- Dupont, C. and Passy, F. 2011 The Arab Spring or How to Explain those Revolutionary Episodes? *Swiss Political Science Review*. **17**(4), pp.447-451.
- Eckart, T., et al. 2014. Large Arabic Web Corpora of High Quality: The Dimensions Time and Origin. In: *Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools, LREC 2014, Reykjavik, Iceland*.
- Elhousseini, F. 2013. *The Arab Spring: conspiracies or national will?* . [Online]. [Accessed 13 May 2013]. Available from: <http://www.middleeastmonitor.com/articles/middle-east/5992-the-arab-spring-conspiracies-or-national-will>.
- Fairclough, N. 1989. *Language and power*. London: Longman.
- Fairclough, N. 1992. *Discourse and social change*. Cambridge: Polity Press.
- Fairclough, N. 1995a. *Critical discourse analysis : the critical study of language*. London ; New York: Longman.
- Fairclough, N. 1995b. *Media discourse*. London ; New York: E. Arnold.
- Fairclough, N. 2001. *Language and power*. 2nd ed. Harlow: Longman.
- Fairclough, N. and Wodak, R. 1997. Critical Discourse Analysis. In: van Dijk, T. A. ed. *Discourse as Social Interaction*. London, Thousand Oaks, CA, New Delhi: Sage, pp.258-284.
- Fisk, R. 2013. *Beirut newspaper defies closure: Lebanese officials say left-wing daily 'endangered security of the state' with peace talks report*. [Online]. [Accessed 28 October 2016]. Available from: <http://www.independent.co.uk/news/world/beirut-newspaper-defies-closure-lebanese-officials-say-left-wing-daily-endangered-security-of-the-2322563.html>.

- Flowerdew, L. 2005 An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*. **24**(3), pp.321-332.
- Forchini, P. and Murphy, A. 2008 N-grams in Comparable Specialized Corpora: Perspectives on Phraseology, Translation, and Pedagogy. *International Journal of Corpus Linguistics*. **13**(13), pp.351–367.
- Fowler, R. 1991. *Language in the News: Discourse and Ideology in the Press*. London and New York: Routledge.
- Fowler, R. 1996. *Linguistic criticism*. 2nd ed. Oxford ; New York: Oxford University Press.
- Freake, R., et al. 2011 A Bilingual Corpus-Assisted Discourse Study of the Construction of Nationhood and Belonging in Quebec. *Discourse & Society*. **22**(1), pp.21-47.
- Gabrielatos, C. 2007a. If-Conditionals as Modal Colligations: A Corpus-Based Investigation. In: Davies, M., et al., eds. *Proceedings of the Corpus Linguistics Conference: Corpus Linguistics 2007, Birmingham*. University of Birmingham, pp.1-20.
- Gabrielatos, C. 2007b Selecting Query Terms to Build a Specialised Corpus from a Restricted-Access Database. *ICAME Journal*. **31**, pp.5-43.
- Gabrielatos, C. 2012. %DIFF-FAQ With More Details. [Online]. Available from: http://ucrel.lancs.ac.uk/ll/DIFF_FAQ.pdf.
- Gabrielatos, C. and Baker, P. 2008 Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*. **36**(1), pp.5-38.
- Gabrielatos, C. and Marchi, A. 2011. Matching Metrics to Definitions. In: *Corpus Linguistics in the South: Theoretical-Methodological Challenges in Corpus Approaches to Discourse Studies - and Some Ways of Addressing Them, 5 November 2011, University of Portsmouth*.
- Gabrielatos, C. and Marchi, A. 2012. Keyness Appropriate Metrics and Practical Issues. In: *CADS International Conference Corpus-assisted Discourse*

Studies: More than the Sum of Discourse Analysis and Computing?,
University of Bologna, 13-14 September 2012.

Gabrielatos, C. and McEnery, T. 2005. Epistemic Modality in MA Dissertations. In: Fuertes Olivera, P. A. ed. *Lengua y Sociedad: Investigaciones recientes en lingüística Aplicada*. Valladolid: Universidad de Valladolid, pp.311-331.

Gabrielatos, C., et al. 2012 The peaks and troughs of corpus-based contextual analysis. *International Journal of Corpus Linguistics*. **17**(2), pp.151-175.

Galtung, J. 2012 The Arab Revolt—What Next? *Palestine-Israel Journal of Politics, Economics & Culture*. **18**(1), pp.19-21.

Galtung, J. and Ruge, M. H. 1965 The Structure of Foreign News The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers. *Journal of peace research*. **2**(1), pp.64-90.

Garrett, P. and Bell, A. 1998. Media and Discourse: A Critical Overview. In: Bell, A. and Garrett, P. eds. *Approaches to Media Discourse*. Malden: Blackwell Publishing, pp.1-20.

Geller, P. 2011. *Islamic vs. Islamist: An Artificial Distinction*. [Online]. [Accessed 10 March 2014]. Available from: <http://pamelageller.com/2011/10/islamic-vs-islamist-an-artificial-distinction.html/>.

Gelvin, J. L. 2012. *The Arab Uprisings: What Everyone Needs to Know*. Oxford: Oxford University Press.

Gerbig, A. 2010. Key Words and Key Phrases in a Corpus of Travel Writing. In: Bondi, M. and Scott, M. eds. *Keyness in Texts*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp.147-168.

Goodwin, J. 2011 Why We Were Surprised (Again) by the Arab Spring. *Swiss Political Science Review*. **17**(4), pp.452-456.

Graham, D. 2014. *KeyBNC [Computer Software]*. [Online]. Available from: <http://crs2.kmutt.ac.th/Key-BNC/>.

Hadar, L. T. 2011. This Is a Struggle for Power, Not Arab Spring. [Online]. [Accessed 15 November 2012]. Available from:

<http://www.cato.org/publications/commentary/is-struggle-power-not-arab-spring>.

- Haenska-Ahy, M. T. and Shapour, R. 2013 Who's Reporting the Protests? Converging practices of citizen journalists and two BBC World Service newsrooms, from Iran's election protests to the Arab uprisings. *Journalism Studies*. **14**(1), pp.29-45.
- Hall, S. 1973. The Determinations of News Photographs. In: Cohen, S. and Young, J. eds. *The Manufacture of News: Deviance, Social Problems and the Mass Media*. London: Constable, pp.226-243.
- Harcup, T. and O'Neill, D. 2001 What Is News? Galtung and Ruge revisited. *Journalism Studies*. **2**(2), pp.261-280.
- Harcup, T. and O'Neill, D. 2016 What is news? *Journalism Studies*. pp.1-19.
- Hardt-Mautner. 1995. *Only Connect: Critical Discourse Analysis and Corpus Linguistics*. [Online]. [Accessed 1 November 2013]. Available from: <http://ucrel.lancs.ac.uk/papers/techpaper/vol6.pdf>.
- Haywood, J. A. and Nahmad, H. M. 1976. *A New Arabic Grammar of the Written Language*. London: Percy Lund, Humphries & Co. Ltd.
- Herman, E. S. and Chomsky, N. 2002. *Manufacturing Consent : The Political Economy of the Mass Media; with a new introduction by the authors*. New York: Pantheon Books.
- Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language*. Oxon: Routledge.
- Howard, P. N. and Hussain, M. M. 2011 The Upheavals in Egypt and Tunisia: the role of digital media. *Journal of Democracy*. **22**(3), pp.35-48.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. 2004. Counting the Uncountable: Problems of identifying evaluation in a text and in a corpus. In: Partington, A., et al. eds. *Corpora and Discourse*. Bern: Peter Lang, pp.137-188.

- Hunston, S. 2007 Semantic prosody revisited. *International Journal of Corpus Linguistics*. **12**(2), pp.249-268.
- Hunston, S. 2010. How Can a Corpus be Used to Explore Patterns? In: O’Keeffe, A. and McCarthy, M. eds. *The Routledge Handbook of Corpus Linguistics*. London and New York: Routledge, pp.152-166.
- Jacoby, J. 2005. *The Arab spring*. [Online]. [Accessed 1 November 2014]. Available from: http://www.boston.com/news/globe/editorial_opinion/oped/articles/2005/03/10/the_arab_spring/.
- Johnson, S., et al. 2003 From ‘Politically Correct Councillors’ to ‘Blairite Nonsense’: Discourses of ‘Political Correctness’ in Three British Newspapers. *Discourse & Society*. **14**(1), pp.29-47.
- Khalid, D. 1975 Book Review: La crise des intellectuels arabes: Traditionalisme ou historicisme (The Crisis of Arab Intellectuals) by Abdallah Laroui. *Islamic Studies*. **14**(1), pp.77-80.
- Khoury, R. G. 2011. Drop the Orientalist term ‘Arab Spring’. *The Daily Star* [Online]. [Accessed 8 May 2013]. Available from: <http://www.dailystar.com.lb/Opinion/Columnist/2011/Aug-17/Drop-the-Orientalist-term-Arab-Spring.ashx#axzz1clXQEC6E>.
- Kilgariff, A. 2001 Comparing Corpora. *International Journal of Corpus Linguistics*. **6**(1), pp.97-133.
- Kilgariff, A. 2012. Getting to Know Your Corpus. In: *Text, Speech, Dialogue (TSD 2012)*, *Lecture Notes in Computer Science*, Springer.
- Kilgariff, A., et al. 2004. The Sketch Engine. In: *Proceedings of EURALEX, Lorient, France* pp.105-116.
- Kintsch, W. and van Dijk, T. A. 1978 Toward a Model of Text Comprehension and Production. *Psychological Review*. **85**(5), pp.363-394.
- Kress, G. and van Leeuwen, T. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication* London: Arnold.

- Krishnamurthy, R. 1996. Ethnic, Racial and Tribal: The Language of Racism? In: Caldas-Coulthard, C. R. and Coulthard, M. eds. *Text and Practices*. London and New York: Routledge, pp.129-149.
- Kuran, T. 1995 The Inevitability of Future Revolutionary Surprises. *American Journal of Sociology*. **100**(6), pp.1528-1551.
- Kurzman, C. 2012a THE ARAB SPRING UNCOILED. *Mobilization*. **17**(4), pp.377-390.
- Kurzman, C. 2012b The Arab Spring: Ideals of the Iranian Green Movement, Methods of the Iranian Revolution. *International Journal of Middle East Studies*. **44**(1), pp.162-165.
- Macmillan English Dictionary for Advanced Learners*. 2007. [CD-ROM]. Oxford: Macmillan Publishers Limited.
- Leech, G. 2007. New resources, or just better old ones? The Holy Grail of representativeness. In: Hundt, M., et al. eds. *Corpus Linguistics and the Web*. Rodopi: Amsterdam, pp.133-149.
- LexisNexis. 2010. *Supplemental Terms for Specific Materials*. [Online]. [Accessed 27 August 2014]. Available from: <http://www.lexisnexis.com/uk/nexis/auth/retrieveMode.do>.
- Louw, B. 1993. Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In: Baker, M., et al. eds. *Text and Technology: In Honour of John Sinclair*. Philadelphia and Amsterdam: John Benjamins, pp.157-176.
- Lynch, M. 2011. *Obama's 'Arab Spring'?* . [Online]. [Accessed 8 May 2013]. Available from: http://mideast.foreignpolicy.com/posts/2011/01/06/obamas_arab_spring.
- Marchi, A. 2010 The Moral in the Story': A Diachronic Investigation of Lexicalised Morality in the UK Press. *corpora*. **5**(2), pp.161–189.
- Mautner, G. 2007 Mining Large Corpora for Social Information: The Case of *Elderly*. *Language in Society*. **36**(1), pp.51–72.

- Mautner, G. 2008. Analyzing Newspapers, Magazines and Other Print Media. In: Wodak, R. and Krzyzanowski, M. eds. *Qualitative Discourse Analysis in the Social Sciences*. Basingstoke: Palgrave Macmillan, pp.30-53.
- Mautner, G. 2009a. Checks and Balances: How Corpus Linguistics can Contribute to Critical Discourse Analysis. In: Wodak, R. and Meyer, M. eds. *Methods of Critical Discourse Analysis* 2nd ed. London: Sage Publications Ltd, pp.122-138.
- Mautner, G. 2009b. Corpora and Critical Discourse Analysis. In: Baker, P. ed. *Contemporary Corpus Linguistics*. London: Continuum, pp.32-46.
- McCurdy, P. 2012 Social Movements, Protest and Mainstream Media. *Sociology Compass*. 6(3), pp.244-255.
- McEnery, T. 2009. Keywords and Moral Panics: Mary Whitehouse and Media Censorship In: Archer, D. ed. *What's in a Word-List? Investigating Word Frequency and Keyword Extraction*. Farnham: Ashgate, pp.93-124.
- McEnery, T. and Hardie, A. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- McEnery, T., et al. 2006. *Corpus-Based Language Studies: An advanced resource book*. London and New York: Routledge.
- Mellor, N. 2005. *The Making of Arab News*. Lanham: Rowman & Littlefield Publishers, Inc.
- Meyer, C. F. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Milizia, D. 2010. Keywords and Phrases in Political Speeches. In: Bondi, M. and Scott, M. eds. *Keyness in Texts*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp.127-145.
- Milton-Edwards, B. 2011. *Contemporary Politics in the Middle East*. 3rd edn ed. Cambridge: Polity.
- Morley, J. and Bayley, P. eds. 2009. *Corpus-Assisted Discourse Studies on the Iraq Conflict: Wording the War*. New York and London: Routledge.

- Morley, J. and Partington, A. 2009 A Few Frequently Asked Questions about Semantic - or Evaluative - Prosody. *International Journal of Corpus Linguistics*. **14**(2), pp.139-158.
- Murphy, A. C. 2005 Markers of Attribution in English and Italian Opinion Articles: A Comparative Corpus-Based Study. *ICAME Journal*. **29**, pp.131-150.
- O'Neill, D. and Harcup, T. 2009. News Values and Selectivity. In: Wahl-Jorgensen, K. and Hanitzsch, T. eds. *The Handbook of Journalism Studies*. London: Routledge, pp.161-174.
- Oakey, D. 2010. English Vocabulary and Collocation. In: Hunston, S. and Oakey, D. eds. *Introducing Applied Linguistics: Concepts and Skills*. London and New York: Routledge, pp.14-23.
- OED Online*. 2013. [Accessed 10 February 2013] Available from: <http://www.oed.com>
- Partington, A. 2004 "Utterly Content in Each Other's Company": Semantic Prosody and Semantic Preference. *International Journal of Corpus Linguistics*. **9**(1), pp.131-156.
- Partington, A. 2006. Metaphors, Motifs and Similes across Discourse Types: Corpus-Assisted Discourse Studies (CADS) at Work. In: Stefanowitsch, A. and Gries, S. T. eds. *Corpus-Based Approaches to Metaphor and Metonymy*. pp.267-304.
- Partington, A. 2009 A linguistic Account of Wordplay: The lexical Grammar of Punning. *Journal of Pragmatics*. **41**(9), pp.1794-1809.
- Partington, A. 2010 Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK Newspapers: An Overview of the Project. *Corpora*. **5**(2), pp.83-108.
- Partington, A. 2012 The Changing Discourses on Antisemitism in the UK Press from 1993 to 2009: A Modern-Diachronic Corpus-Assisted Discourse Study. *Journal of Language and Politics*. **11**(1), pp.51-76.
- Partington, A. 2014 Mind the Gaps: The Role of Corpus Linguistics in Researching Absences. *International Journal of Corpus Linguistics*. **19**(1), pp.118-146.

- Partington, A. 2015. Corpus-Assisted Comparative Case Studies of Representations of the Arab World. In: Baker, P. and McEnery, T. eds. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan, pp.220-243.
- Partington, A., et al. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Partington, A. and Morley, J. 2002. *From Frequency to Ideology: Investigating Word and Cluster/Bundle: Frequency in Political Debate*.
- Pojanapunya, P. and Watson Todd, R. 2016 Log-Likelihood and Odds Ratio: Keyness Statistics for Different Purposes of Keyword Analysis. *Corpus Linguistics and Linguistic Theory*.
- Potts, A. 2015. Filtering the Flood: Semantic Tagging as a Method of Identifying Salient Discourse Topics in a Large Corpus of Hurricane Katrina Reportage. In: Baker, P. and McEnery, T. eds. *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan, pp.285-304.
- Potts, A., et al. 2015 How Can Computer-Based Methods Help Researchers to Investigate News Values in Large Datasets? A Corpus Linguistic Study of the Construction of Newsworthiness in the Reporting on Hurricane Katrina. *Discourse & Communication*. pp.1-24.
- Prentice, S., et al. 2012 The language of Islamic extremism: Towards an automated identification of beliefs, motivations and justifications. *International Journal of Corpus Linguistics*. **17**(2), pp.259-286.
- Rayson, P. 2003. *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. PhD thesis, Lancaster University.
- Rayson, P. 2008 From key words to key semantic domains. *International Journal of Corpus Linguistics*. **13**(4), pp.519-549.
- Rayson, P. 2009. *Wmatrix: A Web-Based Corpus Processing Environment*, Computer Department, Lancaster University. [Online]. [Accessed 17 November 2014]. Available from: <http://ucrel.lancs.ac.uk/wmatrix/>

- Regier, T. and Khalidi, M. A. 2009 The Arab Street: Tracking a Political Metaphor. *The Middle East Journal*. **63**(1), pp.11-29.
- Reisigl, M. and Wodak, R. 2009. The Discourse-Historical Approach (DHA). In: Wodak, R. and Meyer, M. eds. *Methods of Critical Discourse Analysis*. 2nd ed. Los Angeles and London: Sage, pp.87-121.
- Richardson, J. E. 2007. *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Basingstoke and New York: Palgrave Macmillan.
- Ryding, K. C. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Schaudt, S. and Carpenter, S. 2009 The News That's Fit to Click: An Analysis of Online News Values and Preferences Present in the Most-Viewed Stories on Azcentral.Com. *Southwestern Mass Communication Journal*. **24**(2), pp.17-26.
- Scott, M. 1997 PC Analysis of Key Words — and Key Key Words. *System*. **25**(2), pp.233-245.
- Scott, M. 2009. In Search of a Bad Reference Corpus. In: Archer, D. ed. *What's in a Word-list? Investigating Word Frequency and Keyword Extraction*. Surrey: Ashgate, pp.79-91.
- Scott, M. 2010. Problems in Investigating Keyness, or Clearing the Undergrowth and Marking Out Trails. In: Bondi, M. and Scott, M. eds. *Keyness in Texts*. Amsterdam / Philadelphia: John Benjamins Publishing Company, pp.43-57.
- Scott, M. 2015. *WordSmith Tools Manual, Version 6.0*. Liverpool: Lexical Analysis Software Ltd.
- Scott, M. and Tribble, C. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Sharoff, S. 2006a Creating General-Purpose Corpora Using Automated Search Engine Queries. *WaCky*. pp.63-98.
- Sharoff, S. 2006b. *Leeds Arabic Internet Corpus (LAIC)*. [Online]. [Accessed 1 February 2013]. Available from: <http://corpus.leeds.ac.uk/itweb/htdocs/Query.html#>.

- Sinclair, J. 2005. Corpus and Text- Basic Principles. In: Wynne, M. ed. *Developing Linguistic Corpora: a Guide to Good Practice*. [Online]. Oxford: Oxbow Books, pp.1-16. [Accessed 29 November 2012]. Available from: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>.
- Sinclair, J. M. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, M. 1995 Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*. **2**(1), pp.23-55.
- Stubbs, M. 1996. *Text and Corpus Analysis: Computer-assisted Studies of Language and Culture*. Oxford: Blackwell Publishers Ltd.
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, M. 2010. Three Concepts of Keywords. In: Bondi, M. and Scott, M. eds. *Keyness in Texts*. Amsterdam/ Philadelphia: John Benjamins Publishing Company, pp.21-42.
- Taylor, C. 2008 What is corpus linguistics? What the data says. *ICAME Journal*. **32**(-), pp.179-200.
- Taylor, C. 2010 Science in the News: A Diachronic Perspective. *Corpora*. **5**(2), pp.221-250.
- Taylor, C. 2013 Searching for similarity using corpus-assisted discourse studies. *Corpora*. **8**(1), pp.81-113.
- Taylor, C. 2014 Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics*. **19**(3), pp.368-400.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Torlakova, L. 2014 Metaphors of the Arab Spring: Figurative Construals of the Uprisings and Revolutions. *Journal of Arabic and Islamic Studies*. **14**(1), pp.1-25.

- Tyner, J. A. and Rice, S. 2012 Moving beyond the 'Arab Spring': The ethnic, temporal, and spatial bounding of a political movement. *Political Geography*. **31**(3), pp.131-132.
- van Dijk, T. A. 1988a. *News Analysis: Case Studies of International and National News in the Press*. Hillsdale, NJ: Lawrence Erlbaum.
- van Dijk, T. A. 1988b. *News as Discourse*. Hillsdale, NJ: Lawrence Erlbaum.
- van Dijk, T. A. 1991. *Racism and the Press*. London and New York: Routledge.
- van Dijk, T. A. 1992. Racism and Argumentation: "Race Riot" Rhetoric in Tabloid Editorials. In: van Eemeren, F. H., et al. eds. *Argumentation illuminated*. Dordrecht: Foris, pp.242-259.
- van Dijk, T. A. 1993. *Elite Discourse and Racism*. Newbury Park, CA: Sage.
- van Dijk, T. A. 1995. Power and the News Media. In: Paletz, D. ed. *Political Communication and Action*. Cresskill, NJ: Hampton Press, pp.9-36.
- van Dijk, T. A. 1996a. Discourse, Power and Access. In: Caldas-Coulthard, C. R. and Coulthard, M. eds. *Texts and Practices*. London and New York: Routledge, pp.84-104.
- van Dijk, T. A. 1996b. *Opinions and Ideologies in Editorials (Paper for the 4th International Symposium of Critical Discourse Analysis, Language, Social Life and Critical Thought, Athens, 14-16 December 2015, 2nd draft, March 1996)*. [Online]. [Accessed 3 September 2014]. Available from: <http://www.discursos.org/unpublished%20articles/Opinions%20and%20ideologies%20in%20editorials.htm>.
- van Dijk, T. A. 1998. Opinions and Ideologies in the Press. In: Bell, A. and Garrett, P. eds. *Approaches to Media Discourse*. Oxford, Malden: Blackwell, pp.21-63.
- van Dijk, T. A. 2001. Critical Discourse Analysis. In: Schiffrin, D., et al. eds. *The Handbook of Discourse Analysis*. Oxford: Blackwell, pp.352-371.
- van Dijk, T. A. 2008. *Discourse and Power*. Basingstoke: Palgrave Macmillan.

- van Leeuwen, T. 1993 Genre and Fields in Critical Discourse Analysis: A Synopsis. *Discourse & Society*. 4(2), pp.193–223.
- van Leeuwen, T. 1996. The Representation of Social Actors. In: Caldas-Coulthard, C. R. and Coulthard, M. eds. *Text and Practices*. London and New York: Routledge, pp.32-70.
- van Leeuwen, T. 2008. *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press.
- Vessey, R. 2013 Challenges in Cross-Linguistic Corpus-Assisted Discourse Studies. *Corpora*. 8(1), pp.1-26.
- Widdowson, H. G. 2004. *Text, Context, Pretext: Critical Issues in Discourse Analysis*. Oxford: Blackwell.
- Wikipedia. 2013. *DIN 31635*. [Online]. [Accessed 23 February 2013]. Available from: http://en.wikipedia.org/wiki/Sun_and_moon_letters.
- Wodak, R. and Meyer, M. 2009. Critical Discourse Analysis: History, Agenda, Theory and Methodology. *Methods of Critical Discourse Analysis*. pp.1-33.
- Woolls, D. 2013. *Wikipedia Arabic Corpus Reference List*. CFL Software Limited, UK <http://www.cflsoftware.com>.
- Worth, R. F. 2011. *How a Single Match Can Ignite a Revolution*. [Online]. [Accessed 22 May 2013]. Available from: <http://www.nytimes.com/2011/01/23/weekinreview/23worth.html?src=twrhp>.
- Xiao, R. and McEnery, T. 2006 Collocation, Semantic Prosody, and Near Synonymy: A Cross-Linguistic Perspective. *Applied Linguistics*. 27(1), pp.103–129.
- Xiao, R. and Yue, M. 2009. Using Corpora in Translation Studies: The State of the Art. In: Baker, P. ed. *Contemporary Corpus Linguistics*. London: Continuum, pp.237-261.
- Zayani, M. and Ayish, M. I. 2006 Arab Satellite Television and Crisis Reporting: Covering the Fall of Baghdad. *International Communication Gazette*. 68(5-6), pp.473-497.

Zimmer, B. 2011. The "Arab Spring" Has Sprung. *Visual Thesaurus*. [Online].
[Accessed 12 January 2012]. Available from:
<https://www.visualthesaurus.com/cm/wordroutes/the-arab-spring-has-sprung/>.

Appendix A: Example of metadata from the British sub-corpus

File Name	Headline/ Title	Date	Mon-Yr	Year	Text Type	News Outlet	Sub-corpus	Main sub-corpus	Byline and Dateline	URL	Region	Media Type	Word Count
BBCe-10-01	Ahmadinejad says Iran nuclear fuel deal 'still alive'	15/06/2010	Jun-10	2010	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	314
BBCe-10-01	Ahmadinejad says Iran nuclear fuel deal 'still alive'	15/06/2010	Jun-10	2010	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	314
DTLe-10-01	Britain is stuck with a war that it can't afford, and can't win;	15/06/2010	Jun-10	2010	Editorials & opinions	DTLe	BR	ENASC	Mary Riddell	Nexis	UK	Newspaper	1,147
BBCe-10-248	Egypt Muslim Brotherhood 'quits election'	01/12/2010	Dec-10	2010	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	313
BBCe-10-281	Amnesty International urges Iran to stop Kurd execution	25/12/2010	Dec-10	2010	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	264
GRDe-10-88	Crackdown threat in Tunisia after graduate protests: Two die	30/12/2010	Dec-10	2010	News	GRDe	BR	ENASC	Julian Borger l	Nexis	UK	Newspaper	603
NDPe-10-70	Iran hangs man accused of passing military secrets to Israel	29/12/2010	Dec-10	2010	News	NDPe	BR	ENASC	Donald Macin	Nexis	UK	Newspaper	776
TMSe-10-71	Obama's peace plan in tatters after Israel calls his bluff on set	09/12/2010	Dec-10	2010	News	TMSe	BR	ENASC	Giles Whittell	NEXIS.	UK	Newspaper	1,039
BBCe-11-63	G8 summit: Arab uprisings dominate Deauville agenda	26/05/2011	May-11	2011	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	713
DTLe-11-28	Arab Spring will add to extremism if we do not help, says Can	27/05/2011	May-11	2011	News	DTLe	BR	ENASC	Andrew Porte	Nexis	UK	Newspaper	555
FTE-11-69	Aid pledged to Arab spring states	27/05/2011	May-11	2011	News	FTE	BR	ENASC	SECTION: FR	Nexis	UK	Newspaper	120
GRDe-11-40	WikiLeaks and Guardian hailed as catalysts of Arab spring	13/05/2011	May-11	2011	News	GRDe	BR	ENASC	Peter Walker	Nexis	UK	Newspaper	364
NDPe-11-81	Obama fails to endorse Coalition's spending cuts	26/05/2011	May-11	2011	News	NDPe	BR	ENASC	Andrew Grice	Nexis	UK	Newspaper	556
TMSe-11-49	Assad tightens his iron grip;	07/05/2011	May-11	2011	News	TMSe	BR	ENASC	Evelyn Scott	NEXIS.	UK	Newspaper	896
BBCe-11-167	Muammar Gaddafi death: World reaction	20/10/2011	Oct-11	2011	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	960
DTLe-11-100	Arab Spring activists win freedom award;	28/10/2011	Oct-11	2011	News	DTLe	BR	ENASC	NEWS; Pg. 26	Nexis	UK	Newspaper	123
FTE-11-153	Spring harvest; Gaddafi's downfall just one milestone on road	22/10/2011	Oct-11	2011	Editorials & opinions	FTE	BR	ENASC		Nexis	UK	Newspaper	457
GRDe-11-133	Tunisia elections: The revolution is unfinished.	20/10/2011	Oct-11	2011	News	GRDe	BR	ENASC	Angelique Chr	Nexis	UK	Newspaper	1,459
NDPe-11-176	Syria slips towards sectarian war;	27/10/2011	Oct-11	2011	News	NDPe	BR	ENASC	Robert Fisk	Nexis	UK	Newspaper	819
TMSe-11-131	Gaddafi is dead. Now for a far worse threat;	27/10/2011	Oct-11	2011	Editorials & opinions	TMSe	BR	ENASC	Bronwen Mad	NEXIS.	UK	Newspaper	1,049
BBCe-12-62	Analysis: Bin Laden papers details	03/05/2012	May-12	2012	News	BBCe	BR	ENASC	Gordon Corer	http://w	UK	Broadcasting	591
BBCe-12-63	Tunisia fines TV channel owner over controversial film	03/05/2012	May-12	2012	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	313
DTLe-12-31	UN blames al-Qaeda for bombings in Syria	19/05/2012	May-12	2012	News	DTLe	BR	ENASC	Adrian Blomfi	Nexis	UK	Newspaper	577
GRDe-12-50	Egyptian presidential election: Democratic deficits:	24/05/2012	May-12	2012	News	GRDe	BR	ENASC	Simon Tisdall	Nexis	UK	Newspaper	671
NDPe-12-57	Egypt's elections leave its divisions unresolved;	29/05/2012	May-12	2012	Editorials & opinions	NDPe	BR	ENASC	LEADING AR	Nexis	UK	Newspaper	575
TMSe-12-37	Arab revolt is the real game-changer;	30/05/2012	May-12	2012	Editorials & opinions	TMSe	BR	ENASC	Roger Boyes-(NEXIS.	UK	Newspaper	313
BBCe-12-121	Brahimi has 'no illusions' about 'toughest yet' Syria mission	03/09/2012	Sep-12	2012	News	BBCe	BR	ENASC	Lyse Doucet-(http://w	UK	Broadcasting	936
BBCe-12-144	Egyptian jihadist's path to Syria	25/09/2012	Sep-12	2012	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	744
DTLe-12-56	The Arab Spring turns sour for America;	13/09/2012	Sep-12	2012	Editorials & opinions	DTLe	BR	ENASC	Con Coughlin	Nexis	UK	Newspaper	1,307
DTLe-12-62	Arab leaders should condemn recent Muslim riots over anti	18/09/2012	Sep-12	2012	Editorials & opinions	DTLe	BR	ENASC	SECTION: LE	Nexis	UK	Newspaper	1,203
FTE-12-78	Arab spring leaves trail of state frailties in its wake	13/09/2012	Sep-12	2012	News	FTE	BR	ENASC	Heba Saleh in	Nexis	UK	Newspaper	781
GRDe-12-98	Leading article: Muslim protest: Breaking the cycle	22/09/2012	Sep-12	2012	Editorials & opinions	GRDe	BR	ENASC		Nexis	UK	Newspaper	445
NDPe-12-93	The Arab Spring was never what it seemed;	13/09/2012	Sep-12	2012	Editorials & opinions	NDPe	BR	ENASC	Patrick Cockb	Nexis	UK	Newspaper	861
TMSe-12-74	Arab Autumn; Reaction to a provocative anti-Islamic film ha	17/09/2012	Sep-12	2012	Editorials & opinions	TMSe	BR	ENASC	SECTION: EI	NEXIS.	UK	Newspaper	650
TMSe-12-76	Arab Democracy; The threat of American disengagement fro	22/09/2012	Sep-12	2012	Editorials & opinions	TMSe	BR	ENASC	SECTION: EI	NEXIS.	UK	Newspaper	645
BBCe-13-38	Palestinian disappointment at Obama presidency	21/03/2013	Mar-13	2013	News	BBCe	BR	ENASC	Jon Donnison	http://w	UK	Broadcasting	1,055
BBCe-13-43	Jordan's King Abdullah swears in new government	30/03/2013	Mar-13	2013	News	BBCe	BR	ENASC		http://w	UK	Broadcasting	288