

Arab Women in Western Press: Designing News Corpora for Arab Women in British, American, and German News Media during the Arab Spring

Zahra Mustafa-Awad¹, Majdi Sawalha²,
Monika Kirner-Ludwig³, Dua'a Tabaza¹

¹ The University of Jordan,
Department of Linguistics, Faculty of Foreign Languages,
Jordan

² The University of Jordan,
Computer Information Department,
School for Information Technology,
Jordan

³ University at Albany,
State University of New York,
Department of Educational Theory & Practice,
United States

{z.awad,sawalha.majdi}@ju.edu.jo,
mkirner-ludwig@albany.edu, duaa.tabaza@gmail.com

Abstract. This paper reports on designing comparable corpora to investigate the representation of Arab women in Western press during the so-called Arab Spring. Two corpora were constructed by collecting news articles on Arab Women published by most popular news media in the USA, UK and Germany during the period 2010-2015. The first corpus is made up of articles published in English by a British newspaper, *The Guardian* and an American one, *The New York Times*. The second is composed of articles collected from five German news media: *Die Welt*, *Die ZEIT*, *Frankfurter Allgemeine Zeitung*, *Süddeutsche Zeitung*, and *Der Spiegel*. The corpora were managed using Sketch Engine. The significance of our corpora comes from its combining two topics: Arab women and the Arab Spring. Our study emphasizes the political importance of the Arab Spring and highlights the unprecedented roles that Arab women played in the uprisings that characterized it. The results show that there are clear differences between the two corpora in terms of size and emphasis on themes related to Arab women during that period. These results have important implications for researchers and educators in different disciplines including Corpus Linguistics, Discourse Analysis and Mass Communication.

Keywords: Arab women, corpora, news texts, Arab spring.

1 Introduction

News corpora are one type of specialized corpora that can be used for research purposes, natural language processing development, retrieving information, and systems of machine learning. Examples of news corpora include the Reuters' corpus which was launched in 2000, and was composed of 810,000 articles (Rose, Stevenson, & Whitehead, 2002); and Antonio Gulli's corpus which was compiled in 2004 and was made up of more than 1,000,000 news articles (Gulli, & Signorini 2005).

Also, there are the BBC two datasets which were created to cover news from 2004 to 2005, one consists of 2225 news stories published on various topics and the other is made up of 737 articles taken from their sport website. In addition, there is The New York Times Annotated Corpus (Sandhaus, 2008) which was launched in 2008 and was composed of 1.8 million news articles published between 1987 and 2007.

News corpora have become the basis for many studies on the portrayal of various social groups in media. For example, Baker, Gabrielatos, & McEnery (2013) used a four-million-word corpus to explore the linguistic patterns of representing Muslims in British newspapers. Their results indicate that the portrayal of Muslims and Islam in British press was generally negative.

Also, Al-Hejin (2012) analyzed two huge corpora to compare between the image of Muslim women, including Arab women, in news articles published during the period 2001 and 2007 by BBC and Arab News. He found that they are represented differently in the two corpora in terms of semantic macro-structure; while CNN concentrates on their depiction in contexts of crime and conflict, Arab News provides various shades of their portrayal in a wide range of situations combining their achievements and concerns.

A lot of attention has been given lately to the image of Arab women in media. This special interest is due to their active involvement in the political unrest in the Middle East referred to as the Arab Spring, which has led to dramatic consequences and even to military conflicts in the region FIDH, (2012); and to their featuring in the worldwide media coverage of those events (Pew Research Center, 2012).

However, most research on the issue has focused on content analysis, which highlighted aspects of their revolt, modernity and empowerment (Al-Ali 2012; Dastgeer and Gade 2016; Kelek 2012; Kratochwil 2012; Sjoberg and Whooley 2015). In studies concentrating on linguistic inquiry, Arab women have been dealt with primarily using corpora on Muslim women (Özcan, 2013; Al-Hejin, 2014, Samie and Sehlkoglu, 2015). There is, so far, no evidence of using comparable corpora to study the depiction of Arab women in the press during the Arab Spring, using semantic analysis.

This study describes compiling and analyzing two comparable corpora on Arab women in English and German news during the Arab Spring (2010 - 2015). Each corpus is specialized consisting of news articles collected from the most popular print and online news media in the USA, the UK, on one hand, and Germany, on the other. The purpose of creating the corpora is to analyze the representation of Arab women in Western press during that period using Corpus Linguistics (LC) and Critical Discourse Analysis (CDA) approaches.

The importance of research based on such news corpora comes from the importance of news per se: the perceptions and the attitudes people form through their exposure to media which could lead to taking actions with unfavorable consequences.

Table 1. English articles used in the EAWC according to newspaper.

	Number of articles	Percentage	Number of words	Percentage
<i>The Guardian</i>	505	61.66%	435,410	58.92%
<i>New York Times</i>	314	38.34%	303,554	41.08%
Total	819	100%	738,964	100%

2 The Design of the Arab Women Corpora

In preparation of the English and German datasets to build the corpora, any standard links non-specific or irrelevant to the individual articles were excluded from the test versions. Also, all articles were sequentially numbered and double checked for their relevance to Arab women only. All texts were uploaded to Sketch Engine on 16/2/2017.

2.1 The English Arab Women Corpus (EAWC)

The English Arab Women Corpus (EAWC) was created by searching the websites of *The Guardian* and *The New York Times* for relevant articles during the period of October 4, 2010 to June 30, 2015. This period was considered in the study although, historically, the uprisings started in December 2010 and began to fade out in mid-2012. However, protests have continued to take place in a number of Arab countries for several subsequent years and the struggle for power is still going on in some of them. Search terms include in addition to "Arab women/woman", the nationality pertaining to each Arab country followed by "women/woman", e.g., "Egyptian women", "Iraqi women", etc.

We included all the Arab countries in the search terms because the so-called Arab Spring had its effects spread across the Arab world, although the major events took place in Tunisia, Egypt, Yemen, Libya and Syria. The terms were searched in the headlines and the bylines of the articles for two reasons; the first is to make sure that Arab women were the main topic of the news story, and the second is to identify the nationality of women referred to by personal names in the headline through looking at the byline, which provides it in most cases.

For example, Tawakkul, a proper noun, was one of the frequent key words that appeared with a high significance score. It refers to a Yemeni activist who was one of the three winners of Nobel Prize for Peace in 2011; her nationality was, in many instances, mentioned in the byline of the news articles on the topic.

The total corpus consisted of 819 articles made up of 738,964 words. Table [1] shows the distribution of articles and number of words used to build up the EAWC. It is obvious from the figures in Table 1 that the British newspaper, *The Guardian*, had a higher coverage of Arab women during the Arab Spring than the American newspaper, *The New York Times*, in terms of both number of articles (505 / 314), and word counts (435,410 / 303,554), respectively.

Table 2. German articles used in the GAWC according to news medium.

Newspaper	Number of articles	Percentage	Number of words	Percentage
<i>Die Welt</i>	51	15.84%	37,380	17.02%
<i>Süddeutsche Zeitung (SZ)</i>	76	23.60%	41,895	19.07%
<i>Die ZEIT</i>	59	18.32%	57,466	26.16%
<i>Frankfurter Allgemeine Zeitung (FAZ)</i>	78	24.22%	39,738	18.09%
<i>Spiegel</i>	58	18.01%	43,155	19.65%
Total	322	100%	219,634	100%

2.2 The German Arab Women Corpus (GAWC)

The German Arab Women Corpus (GAWC) was compiled by searching the websites of *Die Welt*, *Süddeutsche Zeitung (SZ)*, *Die ZEIT*, *Frankfurter Allgemeine Zeitung (FAZ)* and *Der Spiegel*. To collect related articles, expressions searched for were *arabische Frauen (Arab women)*, *tunesische Frauen (Tunisian women)*, etc. Again, all Arab countries were included in the search and the concentration was on articles which mention Arab women of various nationalities in their headlines or bylines for the same reasons given above.

The GAWC is made up of 322 articles consisting of 219,634 words. Table [2] shows a summary of the number of articles and word counts in the GAWC. It is clear from Table 2 that there are variations in the number of articles and the word counts published by German newspapers on Arab women during that period. Although SZ and FAZ published more articles on the topic (76, 78), respectively; than *Die Welt*, *Die ZEIT* and *Der Spiegel* (51, 59, 58), respectively; *Die ZEIT*'s coverage was the most extensive in terms of the number of words (57,466).

A closer look at Tables 1 and 2 shows that the size of the two corpora is not very big in relation to other conventional linguistic corpora, as we included only such news stories that deal with Arab women as their main topic.

The two tables also indicate that there is a big difference between the two corpora with regard to size; the English corpus is three times larger than the German one. This could be due, among other reasons, to the tendency of German media to generally include Arab women in their coverage of Muslim women (cf. Mustafa-Awad & Kirner-Ludwig, 2017).

2.3 The format and Annotation of the Corpora

Our two corpora were grammatically tagged using Sketch Engine, which adopts the Penn Treebank tagset with slight modifications, with total of 58 tags. For the purpose of this paper we concentrated on the main seven parts of speech, i.e. verbs, nouns, adjectives, adverbs, conjunctions, pronouns, and prepositions.

The corpus was formatted to XML and uploaded to Sketch Engine. Provided metadata include the title of the article, newspaper name, author, date of publication, and section on the website. A sample appears below:

```

"Perceptions of Arab women have been revolutionised" the Guardian,11 March, 2011
<title>Perceptions of Arab women have been revolutionised</title>
<newspaper>the Guardian</newspaper>
<author>Soumaya Ghannoushi,</author>
<section>Arab and Middle East unrest</section>
<Arab Country> All countries </Arab Country>
<story.date>Friday 11 March 2011 20.30 GMT</story.date>
    
```

Fig. 1. A sample of XML format of the arab women corpora.

Table 3. Top 15 most frequent words of the arab women corpora.

a. English Arab Woman Corpus			b. German Arab Woman Corpus	
No	Lemma	Frequency	Lemma	Frequency
1	woman	8673	Frau 'woman'	2350
2	right	1681	Mann 'man'	655
3	Arab	1652	gegen 'against'	552
4	man	1628	arabisch 'Arab'	450
5	Saudi	1604	Mädchen 'girl'	403
6	work	1058	jung 'young'	303
7	girl	1028	Saudi-Arabien 'Saudi-Arabia'	300
8	against	1026	dürfen 'be allowed to'	281
9	Egypt	980	Ägypten 'Egypt'	258
10	force	801	Kopftuch 'headscarf'	255
11	law	800	Kind 'child'	250
12	political	689	stehen 'to stand'	246
13	rape	688	Mensch 'human being'	229
14	activist	662	tragen 'wear'	223
15	drive	654	islamisch 'Islamic'	216

3 Data Analysis

The Arab Women Corpora (AWC) were analyzed using Sketch Engine (Killgariff 2014) to generate frequency word lists, concordance, collocations, word sketches and sketch differences. For the purposes of this study, stop words were not considered in the analysis. Samples of (AWC) analyses are given below.

3.1 Frequency Lists of the Arab Women Corpora

The English Arab Women corpus consists of 738,964 words sorted in 28,198 different lemmas; the most frequent (15) are shown in Table 3 a. On the other hand, The German Arab Women Corpus is made up of 221,898 words distributed into 26,090 different lemmas; the top 15 are given in Table 3 b.

Table 4. Top 15 collocations of the Lemmas *Woman* and *Frau* from arab woman corpora.

a. English Arab Woman Corpus				b. German Arab Woman Corpus		
No	Collocates with 'woman'	Freq.	MI Score	Collocates with Frau(en)	Freq.	MI Score
1	Arab	352	10.15	Jung 'young'	113	10.45
2	right	356	10.15	Muslimisch 'Muslim'	44	9.16
3	young	160	9.13	Saudisch 'Saudi'	36	8.89
4	Allow	89	8.31	gegen 'against'	31	8.45
5	Group	39	7.05	Recht 'right'	22	8.16
6	Face	35	6.98	arabisch 'Arabic'	21	8.09
7	driver	23	6.40	verschleiert 'veiled'	19	8.00
8	appoint	20	6.22	Ägyptisch 'Egyptian'	17	7.87
9	activist	14	5.62	fahren 'drive'	15	7.66
10	veil	12	5.48	Rolle 'role'	14	7.54
11	participate	12	5.47	Aufstand 'rebellion'	12	7.35
12	prison	12	5.45	spielen 'play'	10	7.09
13	free	8	4.88	marokkanisch 'Moroccan'	9	6.59
14	Egypt	8	4.75	Unterdrückung 'oppression'	8	6.59
15	honourable	5	4.23	arbeiten 'to work'	7	6.53

A comparison between Table 3 a and b shows that 10 of the top 15 lemmas in the two sets are equivalent, e.g., *woman/Frau*, *man/Mann*, *Arab/arabisch*, *country/Land*, *against/gegen*. As for the rest of the lemmas in each set, they appear only in one corpus but not in the other.

For instance, *right* (1682), *work* (1081), *force* (801), *law* (800) *rape* (688), and *activist* (662) feature only in EAWC. On the other hand, *sagen* 'to say, speak' (850), *neu* 'new' (289), *dürfen* 'be allowed to' (281), *Kopftuch* 'headscarf' (255) and *stehen* 'to stand' (246) occur only GAWC.

These results indicate that there were discrepancies in emphasis on issues related to Arab women during that period by the news outlets under investigation. While the English language news media focus on issues like rights, violence, oppression and activism, the German ones concentrate on revolution, restriction, dress code and religion.

3.2 Collocations

When searching for collocations of the lemma *woman* in EAWC, the attribute was set as lemma in the range from 5 to 5. The top 15 collocations are presented according to their frequencies and MI (Mutual Information) scores in Table 4a. In a parallel approach, we conducted the same search for collocations of the lemma *Frau* (i.e. including *Frauen* pl.) in GAWC. The top 15 lexical collocations are shown in Table 4b.

We can clearly see from Table 4 below that although a number of the top collocates of *woman* in EAWC have equivalents in GAWC, others do not, despite the overlap between the two corpora in certain instances. Common collocates stress issues like rights and dress code. For instance, women's rights are exhibited by the collocation of

Table 1. Ethnic and national distinctions.

a. English Arab Women Corpus				b. German Arab Women Corpus		
#	Lemma	Freq	Score	Lemma	Freq	Score
1	Somalia	277	81.9	<i>saudi-arabisch</i>	81	53.6
2	Sudanese	282	69.4	<i>Qatar</i>	26	49.7
3	Lebanese	154	63.9	<i>Ägypterin</i>	35	41.1
4	Saudis	102	62.5	<i>Bahrein</i>	9	35.8
5	Yemen	370	60.7	<i>Mauretanien</i>	15	34.8
6	Tunisian	239	60.3	<i>Tunesierin</i>	18	32.7
7	Mauritania	48	50.5	<i>Oman</i>	14	28.6
8	Palestinian	212	46.7	<i>marokkanisch</i>	36	28.4
9	Libya	290	41.6	<i>Iraks</i>	7	28
10	Egyptian	651	40.2	<i>Somaly</i>	6	24.2
11	Moroccan	120	34.6	<i>Jemen</i>	81	22.9
12	Syrian	468	33.6	<i>Dubai</i>	34	21
13	Iraqi	270	33.2	<i>Palästinenserin</i>	8	15.6
14	Jordanian	148	32.7	<i>südsudanesisch</i>	6	15.3
15	Djibouti	25	29.7	<i>Jordanien</i>	26	13
16	Qatar	207	21.9	<i>libanesisch</i>	25	12.5
17	Algeria	67	18.9	<i>libysch</i>	31	11.9
18	Oman	68	14.8	<i>Sudan</i>	22	11
19	Kuwait	69	10.9	<i>Algerien</i>	19	10
20	Bahraini	49	9.9	<i>Kuwait</i>	9	9.5
21	Emirates	75	8.9	<i>Syrerin</i>	5	8.9

right (356, IM 10.15) with *woman/women* and *Recht* (22, 8.16) with *Frau(en)*. One of these rights that occurred in the two corpora is driving expressed by *driver* (23, 6.40) in EAWC and with *fahren* ‘drive’ (15, 7.66) in GAWC, respectively. As for dress code, references are made to *veil* (12, 5.48) and *verschleiert* ‘veiled’ (19, 8.00) in the latter. As for collocates of *women* that occur only in the EAWC, they focus on empowerment, activism, struggle and restriction. For example, empowerment is expressed by *appoint* (20, 6.22) and *free* (8, 4.75), and activism by *participate* (12, 5.4) and *activist* (14, 5.62).

On the other hand, collocates of *Frau* in GAWC concentrate on religion, rebellion, nationality and oppression. The frequent collocation of *muslimisch* + *Frau(en)* suggests that the attribute of being *Muslim* is inherently and prototypically tied in with the image of Arab women.

There are also a number of German collocates that explicitly and implicitly point to the uprisings, e.g., *Aufstand* ‘rebellion, uprising’ (12, 7.35). As for *oppression*, it is expressed by *Unterdrückung* ‘oppression’ (8, 6.59). Another theme that emerges through the collocation with Arab women in GAWC is nationality. This includes countries where the uprisings took place, such as *ägyptisch* ‘Egyptian’ (17, 7.87), as well others where they did not, e.g., in Morocco (*marokkanisch* ‘Moroccan’ 9, 6.95) and Saudi Arabia (*saudisch* ‘Saudi’ 36, 8.89).

Table 2. Keywords of arabic women corpora.

a. English keywords			b. German keywords		
lemma	Freq	Score	lemma	Freq	Score
Sunni	82	53.5	<i>Yeziden</i>	17	66.7
uprising	198	43.6	<i>Jungfräulichkeitstest</i> 'virginity test'	13	45.9
apostasy	54	37.8	<i>Muslimin</i> 'Muslim woman'	91	31.9
refugee	390	30.1	<i>Salafistische</i> 'Salafist'	7	28
sectarian	62	28.2	<i>Verhüllung</i> 'veil'	16	27.7
Shiite	40	28.1	<i>Hidschab</i> 'hijab'	16	27.7
demonstrator	69	25.7	<i>Frauensport</i> 'female sports'	16	26.3
displace	102	22.1	<i>Friedensnobelpreis</i> 'peace nobel prize'	44	24.8
fundamentalist	35	21.8	<i>Verheiratung</i> 'forced/planned marriage'	8	23.7
Islamic	556	20.4	<i>Religionspolizei</i> 'religious police'	19	22.5
virginity	101	20.2	<i>Geschlechtertrennung</i> 'gender segregation'	24	21
famine	25	19.1	<i>Fahrverbot</i> 'driving ban'	38	20.7
Yazidi	52	18.8	<i>Aufstand</i> 'rebellion, uprising'	60	19.3
fighter	244	18.6	<i>Demonstrantin</i> 'female demonstrator'	25	18.5
Coptic	29	18.1	<i>Kinderehe</i> 'child marriage'	6	18
Salafist	27	18	<i>Nikab</i> 'niqab'	12	17.3
revolution	491	17.9	<i>Unverschleiert</i> 'unveiled'	14	17.2
pro-democracy	18	17.7	<i>Bloggerin</i> 'female blogger'	23	17
protester	274	16.7	<i>Nichtmuslim</i> 'non-Muslim'	6	14.3
secularist	21	15.1	<i>Sklaverei</i> 'slavery'	25	13.9
hymen	11	11.1	<i>Genitalverstümmelung</i> 'genital mutilation'	29	13.8
Awakening	10	10.2	<i>Kopftuch</i> 'headscarf'	255	13.7
extremist	91	10.1	<i>Frühling</i> 'spring'	83	13.2
polygamy	27	9.9	<i>Frauenrecht</i> 'female right'	71	13.2
demonstration	177	9.9	<i>Pilotin</i> 'female pilot'	16	11.9
enslave	20	9.8	<i>Frauenrechtsgruppe</i> 'female rights group'	8	11.7
condemnation	31	9.7	<i>Christin</i> 'female Christian'	27	11.4
non-Muslim	13	9.5	<i>Unislamisch</i> 'non-islamic'	7	10.7
circumcision	23	9.5	<i>Aktivistin</i> 'female activist'	86	10.4
sit-in	22	9.4	<i>Inhaftierung</i> 'arrest'	9	9.8

a. English keywords			b. German keywords		
lemma	Freq	Score	lemma	Freq	Score
devout	19	9.3	<i>Angreiferin</i> 'female attacker'	6	9.7
niqab	39	8.9	<i>Polygamie</i> 'polygamy'	9	9.5
harasser	14	8.5	<i>Wahlrecht</i> 'right to vote'	29	9.4
Liberation	14	8.5	<i>Sunnit</i>	8	9
activist	662	8.5	<i>Erlaubnis</i> 'permission'	25	8.9
headscarf	37	8.3	<i>Massengewaltigung</i> 'mass rape'	9	8.7
blogger	64	8.3	<i>Modernisierung</i> 'modernization'	15	7.8
dissent	39	8.2	<i>Unterdrückung</i> 'oppression'	29	7.5
migrant	66	8.2	<i>Patriarchalisch</i> 'patriarchic'	12	7.4
hijab	88	8.1	<i>Drangsalieren</i> 'to bully, to harrass'	5	7.1

The results suggest that, although the news media under investigation reported on women's revolt, activism and protest in the countries where the uprisings were taking place, they also addressed the concerns of women in all Arab countries such as oppression, rights and empowerment.

The results also show that the press coverage varies in its emphasis on the various themes related to Arab women depending on many factors including place and language of publication.

3.3 Key Words

In order to get a fuller picture of the themes related to Arab women reflected in the news during the Arab Spring, we decided to investigate the key words in the two datasets. To that end, we uploaded two reference corpora, one for our English corpus and the other for the German one.

The generated key words list show high frequency and significance scores of terms referring to various Arab nationalities including ones beyond those in which the uprisings took place. This is why we display the key words indicating nationality in a separate table, i.e. Table 5.

As for other significant key words including nouns, adjectives and verbs, they are given in Table 6 below. As both datasets show, religious sectarianism is one salient topic coming up. For instance, in the English corpus we find Sunnis and Shiite with frequencies of 82 and 40 and scores of 53.5 and 28.1, respectively.

This is in addition to Islamic, which occurs in baffling high numbers, i.e. 556 times, with a score of 20.4. Similarly, in the German data, *Yeziden* scored highest with 66.7 (despite its relative low frequency at 17). Within similar ranks, we have *Muslimin* 'Muslim woman' with a frequency of 91 and a score of 31.9.

Although both datasets point to dress code as a common issue, the occurrence of words indicating or referring to this matter are more significant in the German corpus than in the English one. For instance, both *Verhüllung* and *Hidschab* occur at the same score of 27.7 in the German data, while headscarf and hijab occur only at 8.3 and 8.3,

respectively, in the English data. All in all, the German data seems to point more to issues of women's rights or their oppression.

This manifest itself in e.g., lemmas such as *Jungfräulichkeitstest* 'virginity test' with a score of 45.9, *forced marriage* (23.7), *driving ban* (20.7) and *Religionspolizei* 'religious police' with (22.5). *Genitalverstümmelung* 'genital mutilation' (13.8) also falls into this group of lemmas. This aligns with our findings on collocates.

In the English corpus, these issues do not score as high. For instance, virginity scores at 22.2, hymen at 11.1, polygamy at 9.9, slave at 9.8, and circumcision at 9.5. When it comes to women's activism, this is highlighted more in the English data than in the German one, where e.g., *Aufstand* 'rebellion, uprising' (60, 19.3) and *Demonstrantin* 'female demonstrator' (25, 18.5). On the other hand, their English equivalents "uprising" occur at (198, 43.6) and demonstrator at (69, 25.7).

4 Evaluation

Our corpora are clearly different from other comparative corpora used in previous studies in the sense that they combine two topics: Arab women and The Arab Spring. They can be of a great value for researchers and educators in various disciplines including Corpus Linguistics, Discourse Analysis and Mass Communication.

As for research, they can be used for conducting comparative studies between the images of Arab women of different nationalities in Western press, as it includes news articles about women of all Arab countries.

The sub-corpora within the EAWC can be utilized for studying the differences between the newspapers published in English. Similarly, sub-corpora constituting the GAWC can be used for studies on the different German newspapers. Our corpora can be also compared to corpora of news reporting on Arab women during the Arab Spring in other languages including Arabic.

In addition, Arab Spring researchers can find a rich amount of news for content analysis when it comes to women. Journalistic styles between English and German can be compared as well. As for educators, our corpora can be a rich resource for professors and students of various disciplines to draw from for illustration and analysis.

5 Conclusion

It can be concluded that there are clear differences in the news coverage of Arab women during the Arab Spring with regard to the language of publication, the intensity of reporting, the focus on themes associated with them, and stressing their nationalities. The results show that American and British press have covered the Arab Spring more extensively than the German news media, which is reflected in the big difference in the size of the two corpora.

The findings also suggest that there is variation in highlighting general issues related to Arab women such as rights and specific ones associated with certain nationalities like driving. These results have important implications for researchers and educators in different disciplines including Corpus Linguistics, Discourse and Analysis and Mass Communication.

References

1. Al-Ali, N.: Gendering the arab spring. *Middle East Journal of Culture and Communication*, vol. 5, no. 1, pp. 26–31 (2012) doi: 10.1163/187398612X624346
2. Al-Hejin, B.: *Covering muslim women: A corpus-based critical discourse analysis of the BBC and Arab News* (Doctoral dissertation, Lancaster University) (2012)
3. Baker, P., Gabrielatos, C., McEnery, T.: Sketching muslims: A corpus driven analysis of representations around the word ‘muslim’ in the british press 1998–2009. *Applied Linguistics*, vol. 34, no. 3, pp. 255–278 (2013) doi: doi.org/10.1093/applin/ams048
4. Bardici, M. V.: *A discourse analysis of the media representation of social media for social change-The case of Egyptian revolution and political change*. Malmö Haögskola/Kultur Och Samhälle, Digitala Vetenskapliga Arkivet (2012)
5. Greene, D., Cunningham, P.: Practical solutions to the problem of diagonal dominance in kernel document clustering. *BBC News Datasets*, vol. 148. pp. 377–384 (2006) doi: 10.1145/1143844.1143892
6. Gulli, A., Signorini, A.: The indexable web is more than 11.5 billion pages. In: *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*, ACM, pp. 902–903 (2005)
7. Dastgeer, S., Gade, P. J.: Visual framing of muslim women in the arab spring: Prominent, active, and visible. *International Communication Gazette*, vol. 78, no. 5, pp. 432–450 (2016) doi: 10.1177/17480485166402
8. *Federación Internacional por los Derechos Humanos: Women and the Arab spring: taking their place?* (2016)
9. Kelek, N.: *Die arabische revolte und die frauen - eine reise durch ägypten, tunesien und marokko*. köln: kiepenheuer & witsch (2012)
10. Kratochwil, G.: *Die neuen arabischen Frauen: Erfolgsgeschichten aus einer Welt im Aufbruch*. Zurich: Orell Füssli Verlag (2012)
11. Jaworska, S., Krishnamurthy, R.: 2012-On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990–2009. *Discourse & Society*, vol. 23, no. 4, pp. 401–431 (2012) doi: 10.1177/0957926512441113
12. Kandil, M. A.: *The Israeli-Palestinian conflict in american, arab, and british media: corpus-based critical discourse analysis* (2009)
13. Kilgarriff, A.: The Sketch Engine: ten years on. *Lexicography*, vol. 1, no. 1, pp. 1–30 (2104)
14. Mohanty, C. T.: Under western eyes: Feminist scholarship and colonial discourses. *Feminist review*, vol. 30, no. 1, pp. 61–88 (1988) doi: 10.1057/fr.1988.42
15. Özcan, E.: Lingerie, bikinis and the headscarf: visual depictions of muslim female migrants in German news media. *Feminist Media Studies*, vol. 13, no. 3, pp. 427–442 (2013) doi: 10.1080/14680777.2012.712382
16. Farooq, S., Sehlkoglou, S.: Strange, incompetent and out-of-place: Media, muslim sportswomen and London 2012 olympics. *Feminist Media Studies*, vol. 15, no. 3, pp. 363–381 (2015)
17. Rose, T., Stevenson, M., Whitehead, M.: The reuters corpus volume 1-from yesterday's news to tomorrow's language resources. In: *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, vol. 2, pp. 827–832 (2002)
18. Santaemilia, J., Maruenda, S.: The linguistic representation of gender violence in (written) media discourse: The term ‘woman’ in Spanish contemporary newspapers. *Journal of Language Aggression and Conflict*, vol. 2, no. 2, pp. 249–273 (2014)
19. *Signal Media: The signal media one-million news articles dataset* (2015)
20. Sjoberg, L., Whooley, J.: The Arab spring for women? representations of women in middle east politics in 2011. *Journal of Women, Politics & Policy*, vol. 36, no. 3, pp. 261–284 (2015) doi: 10.1080/1554477X.2015.1050902