UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue e Letterature Europee e Americane
Classe LM-37

Tesi di Laurea

How Muslims are represented in the press: a corpus-driven analysis of American and Pakistani online newspaper articles

Relatrice
Prof. Sara Gesuato

Laureando
Francesco Gottardo
n° matr.1242042/LMLLA

Anno Accademico 2021/2022

Abstract

In the present work the issue is addressed of how Muslims and Islam are linguistically represented by US and Pakistani media. Previous studies on this topic show that journalists tend to represent Islam in negative terms, talking about Muslim men mostly as terrorists, Muslim women as victims of their own culture and all Muslims as a monolithic group of 'others' (i.e. *them* contrasting with *us*). The aim of this study is to identify recurrent linguist patterns in how Muslims and Islam are talked about in online articles from the American newspaper *New York Times* and the Pakistani newspaper *The Dawn*.

The data were collected through the *Sketch Engine* platform (www.SketchEngine.eu), which also serves as corpus-building software, directly from the newspapers' websites (www.nytimes.com and www.dawn.com). By adopting a corpus-driven approach, recurrent word combinations are identified so as to determine what the texts are about and what stance or viewpoint they convey.

The research questions addressed are the following: 1) what are the most recurrent topics/notions mentioned? 2) How are Islam and Muslims talked about? 3) Do media mostly report good news or bad news about Islam and Muslims? 4) How similarly or differently are Islam and Muslims characterized in the two corpora?

The results provide partial support for the findings of previous studies. That is, on the one hand, both in the *New York Times* and *The Dawn* newspapers, Islam and Muslims were often represented in negative terms, within a discourse of violence and conflict, a distinction was often made between us (non-Muslims) and them (Muslims), and most of the collocates of the words referring to Muslims carried negative semantic prosody. On the other hand, one unexpected pattern emerged, that is, although *The Dawn* newspaper is supposed to be the Muslim League mouthpiece, the phenomenon of 'othering' was detected there, too, with regard to the collocates of the word *Muslim*. Some marked differences in line with the findings of previous studies emerged as well: in the American paper, the semantic fields of conflict and violence were more frequent than in the Pakistani paper; the Pakistani paper focused on politics more than the American paper; and the American paper focused on radical Islam, while the Pakistani paper focused on liberal Islam.

Table of contents

1. CHAPTER 1. INTRODUCTION……....5

1.1. Statement of the problem……....5

1.2. Purpose of the study……5

1.3. Overview of chapters……....6

2. CHAPTER 2. LITERATURE REVIEW…….……...9

2.1. Introduction…….…9

2.2. Analytic studies……9

2.3. Comparative studies…….……12

2.4. Conclusion……...18

3. CHAPTER 3. METHODOLOGICAL APPROACH……..19

3.1. Introduction……....19

3.2. Theoretical framework……..19

3.3. Data collection and description……28

3.4. Focus of the present study…….….29

4. CHAPTER 4. FINDINGS……..31

4.1. Introduction……....31

4.2. Findings about the *New York Times* corpus…….…31

4.3. Findings about *The Dawn* corpus……....109

4.4. Findings about the *Word Sketch Difference* of key terms…….…..186

5. CHAPTER 5. CONCLUSION……195

5.1. Introduction…….….195

5.2. Research goal and approach……....195

5.3. Main findings about the *New York Times* corpus……...196

5.4. Main findings about *The Dawn* corpus……...197

5.5. Main findings from the *Word Sketch Difference* of key terms……...198

5.6. Conclusion…….199

RIASSUNTO……..201

REFERENCES……...211

1. Introduction

1.1 Statement of the problem

In the present work the issue is addressed of how Muslims and Islam are linguistically represented by US and Pakistani media.

Previous studies on this topic have examined how Western media portray Islam. Their findings show that journalists tend to represent Islam in negative terms, framing it within a discourse of conflict or violence, conservativity or backwardness, and especially harsh contrast with the values of Western civilization. In particular, Muslim men are mostly talked about as terrorists, Muslim women as victims of their own culture, and all Muslims as a monolithic group of 'others' (i.e. *them* contrasting with *us*). All of this contributes to a biased image of Muslims and/or Islam.

1.2 Purpose of the study

The aim of the present work is to identify recurrent linguistic patterns in how Muslims and Islam are represented in online articles from the American newspaper *New York Times* and the Pakistani newspaper *The Dawn*. Adopting a corpus-driven approach, recurrent word combinations are identified so as to determine what the texts are about, and what stance or viewpoint they convey.

The specific issues addressed are the following:

1) What are the most recurrent topics/notions mentioned?

2) How are Islam and Muslims talked about?

3) Do media mostly report good news or bad news about Islam and Muslims?

4) How similarly or differently are Islam and Muslims characterized in the two corpora?

The above research questions are then operationalized as follows:

1) What are the most frequent content words, word combinations and keywords in the *New York Times* and in *The Dawn* as identifiable through the 'frequency *wordlist*', '*N-grams*' and '*keyword*' list function in Sketch Engine (www.sketchengine.eu)?

2) What is the lexical, semantic and grammatical co-text of use of the most frequent words and the top keywords in the *New York Times* and in *The Dawn* as revealed by '*concordances*' and the '*Word Sketch*' function in Sketch Engine?

3) Do the five most frequent words and the top five keywords carry positive or negative semantic prosody, considering the positive or negative meaning of their collocates?

4) Finally, what are the collocates of the words *Islam*, *Muslim* and *Islamic* in the two newspapers considered, as revealed by the '*Word Sketch Difference*' function in Sketch Engine?

The idea for this study came to me immediately after the withdrawal of the US army from Afghanistan (30th August, 2021) and the consequent occupation of this country by Taliban. Being in Italy, I was exposed to news regarding this subject only by Western media, which focused on the negative consequences of the Talibans' occupation of Afghanistan due to their religious extremism. I started to think that in the Western world, media frame the religion of Islam mostly within a discourse of bad news and I wondered whether Middle-Eastern media talked about it in a different way. Moreover, I noticed that I frequently associated terrorism to Islamic people, and so I started wondering where and how this nexus came about, and whether a different representation of Islamic people and/or events in a country with a Muslim majority may emerge or even become prominent. For these reasons I decided to compare a Western newspaper and a Middle-Eastern newspaper, namely the *New York Times* and *The Dawn*.

1.3 Overview of chapters

Chapter 2, *Literature review*, provides an overview of recent studies about how mass media verbally represent people and/or events perceived to be socially and cognitively distant form "the Western world". Most of these works focus on the portrayal of Islam and Muslims, but a few which focus on media representation of other social groups are included. Chapter 3, *Methodological approach*, presents the research approach adopted in the present study. In particular, it outlines the theoretical framework in which the study is set, the datasets investigated, and the main focus of the present work. Chapter 4, *Findings*, presents the results of the analysis conducted on the *New York Times* and *The*

Dawn corpora. Lastly, Chapter 5, *Conclusion*, discusses and draws the implications from the findings.

2. Literature review

2.1 Introduction

The following section provides an overview of recent studies about how mass media linguistically represent people and/or events perceived, or assumed to be, socially and cognitively distant from the "the Western world". Since the focus of my research is the portrayal of Islam and Muslims, most of the reported studies focus on this subject. The works summarized here were selected from academic databases, repositories and platforms (i.e. Padua University *Galileo Discovery* (https://galileodiscovery.unipd.it), Padua University *Padua@Thesis* (https://tesi.cab.unipd.it), *EBSCO* (https://ebsco.com), *Google Scholar* (https://scholar.google.com), and *Proquest* (https://proquest.com). More precisely, searches were carried with the following phrases: *linguistic representation Islam*, *linguistic representation Muslims*, *press representation Islam*, *press representation Muslims*, *media representation Islam* and *media representation Muslims* such that one or more of their component terms should appear anywhere in given bibliographic records. These searches returned hundreds of results. From these, I read those works whose titles and/or abstracts appeared to me to be directly relevant to the subject matter of my research (Islam and Muslims), and which appeared to have a linguistic slant. I also considered other studies, namely those which discussed other subject matters than Islam and Muslims, but which similarly reported on linguistic research about media representation of given social groups so as to gain possibly useful insights into their research approach. They comprise analytical studies, which examine one source of data, and comparative studies, which analyze two or more sources. The purpose of this overview is to take stock of what has already been discovered about media's representation of Islam and Muslims and to show the main methodologies used in this kind of works. Such findings will provide the background to the present study.

2.2. Analytic studies

Lemmouh (2008) analyzed recurring linguistic features that contributed to a stereotyped image of out-groups, notably Muslims, in *New York Times* articles from the years 1990, 1995 and 2000. More specifically, he examined: 1) the lexical items referring

to or co-occurring with the terms *Muslim*/*Muslims*; 2) the frequency with which terms denoting Muslims and other individuals or groups of people occurred as subjects and agents in active and passive clauses; and 3) the types of actions represented as carried out by Muslims as opposed as to those carried out by other individuals or groups of people. The findings showed that: 1) the terms *Muslim*/*Muslims* occurred in sentences representing conflicting or even violent situations; 2) the terms denoting Muslims occurred in subject position in transitive active clauses more often than terms denoting other individuals or groups of people, thus contributing to a representation of Muslims as intentional actors; and 3) terms denoting Muslims represented their referents as engaged in violent acts more often than other individuals or groups of people. The results pointed to a systematic 'othering' and stereotyping of Muslims as compared to other participants.

In Baker et al.'s article (2012), a collocational analysis of the word *Muslim* was carried out in a 143-million-word corpus of British newspaper articles published between 1998 and 2009. Findings revealed that the word *Muslim* tended to be used in adjectival form collocating most frequently with nouns, the latter categorized into the following semantic groups: 1) Conflict (e.g. *extremist*, *fanatic*, *terrorist*, *fundamentalist*); 2) Religion (e.g. *cleric*, *faith*, *festival*, *preacher*); 3) Culture (e.g. *dress*, *culture*, *teaching*, *opinion*); 4) Ethnic/national identity (e.g. *community*, *population*, *country*, *state*, *leader*, *voter*); 5) Attributes specifying class membership (e.g. *woman*, *man*, *family*, *officer*, *Briton*, *shop*); and 6) Group/organization (e.g. *group*, *organization*, *association*). By analyzing these collocates more in-depth, the authors made the following observations: 1) collocates belonging to the categories Religion, Culture and Group/organization, although they appear to be neutral terms, were used in discussions related to conflict; 2) collocates belonging to the category Ethnic/national identity were treated as interchangeable with terms referring to religious notions, for example, *Somalian community* and *Muslim community* were used interchangeably; and 3) collocates belonging to the category Attributes specifying class membership were typically used to differentiate Muslim attributes from British or Western ones. Furthermore, quantitative analysis of the frequency of the categories of use revealed that: 1) collocates indexing references to religious aspects were a minority; 2) the two most frequent categories were those presenting *Muslim* as an attribute of ethnicity or nationality or as an attribute specifying class membership. Lastly, the authors affirmed that, through the use of the

adjective *Muslim*, Muslims were presented as a homogeneous group of people characterized by negative traits because attributes associated to Muslims were contrasted with those attributed to non-Muslims and because Muslims were presented within a discourse of conflict.

Acim (2015) studied the ideological discourse of the *New York Times* Op-eds about Islam and Muslims published between 2007 and 2015. The author observed that, through the use of passive structures and nominal structures, writers reported events in which they omitted reference to agency to the benefit of those individuals or groups of people who committed reprehensible actions (e.g. *The years before the attacks have been thoroughly hashed out through the report of the 9/11 commission and by memoirs and histories*). Furthermore, the author noticed an excessive use of synonymous (including pejorative terms) for the same referents (e.g. *extreme Muslim groups*, *Sunni extremists*, *Sunni Islam*, *conservative Muslim*) through which the audience's attention was directed toward the negative aspects of Muslims and Islam. The author concluded that the presence of nominalizations, passivization and (over)lexicalization contributed to the propagation of certain perceptions about Islam and Muslims.

The work of Mahmoud and Bahareh (2017) analyzed how the US media represented Islam and Muslims, taking into consideration a 670,000-word corpus of news stories published between 2001 and 2015, which included the words *Muslim*, *Muslims*, *Islam* or *Islamic* in their titles. In particular, the authors analysed the concordances of their key terms to determine what topics were frequently mentioned in the representation of Islam and Muslims. Results showed that: 1) the most recurrent collocates of the terms *Islam* and *Muslims* were terms associated with negative concepts (e.g. *Radical*, *War*, *Isis*, *Violence*, *Jihad*, *Group*, etc.); 2) Muslims were generally described as rebellious groups of people threatening the lives of innocent people for religious reasons; 3) most of the occurrences of the word *we* excluded Muslims as potential readers or addresses; 4) the word *Government* was used to refer to the governments of the countries of Middle East, United States, and Europe, thus highlighting the involvement of diverse nations with the issues of Islam and Muslims; 5) the word *Police* had the highest frequency of occurrences with the word *Muslim*; and 6) the three top collocates of the word *women* were *Muslim*, *Rights* and *Children*. Moreover, women were principally portrayed as victims of war who

were being abused. Overall, these results reinforce the discussions of previous studies which indicate a biased representation of Islam and Muslim in media.

Abdul-Rakkz and Khalil (2020) studied the way in which the world perceived Islamophobic incidents (i.e. acts of discrimination and violence committed towards Muslims) by analyzing data of 12 news reports selected from three news agencies: the *BBC*, *Independent*, and *Fox News*. Through the linguistic framework of Appraisal Theory, the authors analyzed three aspects of news reports: 1) attitude, which has to do with evaluating feelings; 2) engagement, which deals with the linguistic resources through which writers express their involvement toward what is referred to in the text; and 3) graduation, which is about the use of intensifiers, downtoners, boosters and hedges, namely, mechanisms by which writers present themselves as more or less aligned with the information offered in the text. The authors found that: 1) most of the attitudes towards Islamophobic incidents were negative, thus condemning the behavior of the people committing these crimes; 2) the forms of engagement used by reporters were highly heteroglossic rather than monoglossic, that is, reference was made to other voices and viewpoints other than the writer's; and 3) writers adopted a neutral position most of the time. In summary, the findings revealed that reporters tended to portray Islamophobic incidents as the actions of aggressors whose negative conduct was generally condemned by societies.

2.3 Comparative studies

Saifuddin and Jörd (2016) carried out a meta-analysis of 345 studies about media representations of Muslims and Islam published between 2000 and 2015 to examine the media's role in the construction of a Muslim and Islamic identity. Their quantitative analysis revealed the following: 1) scholars have been increasingly interested in this area of research; 2) most of the studies covered Western countries, while Muslim countries and Muslim media were neglected; 3) authors from the USA were the most numerous, followed by those from UK and those from Australia; 4) most of the studies favored a quantitative as opposed to a qualitative approach; 5) more than half of the studies built their research on a theoretical framework, with the Theory of Framing being the most commonly used; 5) approximately 90% of the studies focused on analyzing media content, while the rest of the studies explored media and audience perspective; 6) half of

the studies focused on analyzing newspapers, followed by those which focused on television content, the Internet, and just one study on the radio; 7) most of the studies analyzed media content over a time period, and those that did, largely examined media content spanning less than a month; and 8) a small part of the studies incorporated Muslim audiences or pro-Muslim media institutions in their analysis.

The authors pointed out the following findings about their qualitative analysis: 1) a worldwide change in the patterns of representations of Muslims since the attacks of 11 September 2001: post 9/11 portrayals of Muslims and Islam were mostly negative, a common theme used by the US media being that Muslims are terrorists, extremists, fundamentalists, radicals, and fanatics; 2) Muslims were strongly associated with terrorism; 3) Muslim women were mostly presented as victims of their own culture and religion; 4) the US media used pro-war and anti-Muslim frames, while non-US media were anti-war in their portrayals; 5) Muslim migrants were largely presented as a threat to national cultures; 6) in non-Western countries, foreign media, rather than national media, perpetuated higher negative attitudes against Muslims; and 7) several studies showed that the representation of Islam in news media content influenced mosque-building debates, since it infused fear towards Islam. Furthermore, the authors of the meta-study highlighted a lack of comparative research and a dearth of research on online media.

A comparative-contrastive work on two editorials of the *New York Times* and *The Dawn* was carried out by Mahmood et al. (2018), who examined the portrayal of the people involved in the Army Public School, Peshawar Attack, namely the Taliban terrorists vs. the victims killed and injured. The analysis showed that: 1) the *New York Times* editorial referred to the attackers with terms denoting Taliban, like *Pakistani Taliban*, *The Taliban*, *A Taliban spokesman*, *Taliban gunman*, *The group*, while *The Dawn* editorial referred to the attackers as *militants* only; 2) in the *New York Times* editorial, the terms denoting Taliban appeared as subject agents in most of the sentences, thus emphasizing their responsibility for carnages, while *The Dawn* editorial tended to omit the responsibility of the attackers by using agentless sentences; and 3) in the *New York Times* editorial, focus was placed on the attackers, who appeared as subjects in active sentences, while in *The Dawn* editorial, focus shifted toward the victims of the attack, who appeared as subjects in passive sentences. The results revealed that in the *New York*

Times editorial, the Taliban were criticized, and the victims were represented in a way inspiring solidarity with them. The findings also showed that *The Dawn* editorial was predominantly concerned with the plight of the victims and their families, represented as part of "us", expressing only weak criticism towards "them", that is, the Taliban, who were not explicitly held responsible for the attack.

In Muhammad and Salma (2019) a comparative-contrastive analysis is offered about the representation of Islam and Muslims in the editorials of two Australian newspapers having opposite political stances: *The Age*, which is considered a left-leaning newspaper that favors multiculturalism and supports migration in Australia, and *The Australian*, which is considered a right-wing newspaper that supports anti-immigration policies. The authors collected data on *Islam* and *Muslim* as key terms during a time period January 1, 2016 to March 31, 2018. Adopting the framework of Critical Discourse Analysis, the authors focused on the lexical choices that writers made at the time of reporting facts about Islam and Muslims. The results revealed that: 1) out of total 11 selected editorials of *The Age*, seven contained supportive and positive themes regarding Islam and Muslims, that is, words like *targeted Muslims*, *asylum seeker*, *rights of these people* were used; three contained neutral contents combining positive terms such as *grievous errors*, *regrettable*, and *extraordinary refugee crisis* with negative terms such as *Islamic State fighters*, *atrocities* and *chief perpetrator*; and only one editorial contained negative themes regarding Islam and Muslims expressed by words such as *hardline Muslims*, *religious extremists*, *threat of terrorism*; 2) out of total 11 selected editorials of *The Australian*, there was not a single editorial containing positive themes regarding Islam and Muslims; none of the editorial could be categorized as neutral and all the editorials contained explicitly negative themes against Islam and Muslims, that is, words like *Islamist terrorism*, *exporters of terrorism*, *Islamic militants*, *the threat*, *Islamic extremists*, *patriarchal religion*, *sexist code*, *jihadists* were used. The results reported here suggest that *The Age* and *The Australian* portrayed Muslims in an entirely different way: *The Age* portrayed Islam and Muslims favorably, while *The Australian* constructed Islam and Muslims in a critical way. The authors concluded by highlighting the importance of a positive portrayal of Islam and Muslims by press in order to reduce the already existing dichotomy between Muslims and mainstream Australians, and help Muslims integrate in the country.

Younes et al. (2020) conducted an analysis of Islam-related terminologies in Eastern media (*The Jordan Times* and *Al-Jazeera*) and Western mass media (*BBC* and *The Guardian*) published between March 2018 and October 2019. By applying Halliday's Systemic functional linguistic theory, the authors revealed that remarkable differences exist between Eastern and Western mass media outlets in their use of Islam-related terms: 1) a total of nine different terms – including *Islamic extremist*, *Islamic militancy*, and *Islamic extremism* – appeared only in the Western mass media, while only three terms – *Islamic movement*, *Islamic incarnations*, and *Muslim attackers* – appeared only in the Eastern media; 2) the most frequently used Islam-related term was *Islamist*, which was more commonly used in Western media than in Eastern media, and was mostly used in negative contexts, co-occurring with words like *extremism*, *terrorism*, *violence*, and *militancy*; 3) a number of adjectives, such as *Islamic*, *Jihadist*, and *Islamist*, were used to describe the terrorists; and 4) the term *Allahu Akbar*, which is considered holy among the Muslims, was mostly perceived in a negative way by the non-Muslims. In summary, the Islam-related terms used in both Western media discourse and Eastern media discourse on terrorism were mostly used in negative contexts. To conclude, the authors of the study pointed out that there is a lack of comparative research in this field of study, and highlighted that only few studies focused on media in Muslim-majority nations.

Dheskali (2020) offered a corpus-based comparison of framing devices used in the representation of the Israeli-Palestine conflict by the media in American journals (*CNN International*, *New York Times*), Arab-speaking networks (*AlJazeera*, *AlArabiya*), Israeli journals (*Jerusalem Post*, *Times of Israel*) and a Palestinian online journal (*Palestine Chronicle*). By *framing devices* Dheskali means all the resources that writers use in order to focus their audience's attention towards specific themes, and thus "promote a particular problem definition, causal interpretation, moral evaluation and/or treatment recommendation for the item described" (Entman 1993: 52, quoted in Dheskali 2020: 53). Framing devices may include: values, pictures, lexical choices, syntactic structures, hedging other's sources, boosting one's own sources. The results of the analysis indicated that the Israeli and American media made opposite lexical choices compared to Palestinian and Arab-speaking networks to represent the same entities (e.g. *occupation forces* vs. *Israeli Defence Forces*, respectively). Moreover, all reporters tended to hedge others' report of sources (e.g. "*According to* Hamas officials, the soldiers were from

Sayeret Matkal" in *Times of Israel*) and instead to boost and extensively represent the news coming from their own sources (e.g. "The UN General Assembly later voted *overwhelmingly* to demand Israel to comply with the UN's highest legal body" in *AlJazeera*). The Palestinian perspective was supported by the Arab-speaking networks, while the Israeli perspective was supported by American journals. The analysis showed that there is no single truth, but rather multiple or even conflicting truths produced by the media of the different online journals.

Media's tendency to use different framing devices depending on their country's perspective was made clear also in Ming (2020). This article reported a corpus-assisted discourse study of the representations of post-colonial Hong Kong from 2000 to 2017 in one Chinese English newspaper, *China Daily*, and two American newspapers, the *New York Times*, and the *Washington Post*. The findings revealed that *China Daily* preferred to represent Hong Kong within an economic frame, thus, using words which referred to the semantic field of economics, for example *development*, *prosperity*, *products*, and so on; on the other hand, the *New York Times* tended to represent Hong Kong within a political frame, using words referring to the semantic field of politics, for example *democracy*, *demonstrations*, *liberties*, and so on. These lexical differences were attributed to the different ideological positions of these newspapers: American newspapers tended to align with the national interests of the US, thus emphasizing Hong Kong's autonomy rather than the growing integration of Hong Kong with the Chinese mainland; in contrast, *China Daily* highlighted the economic cooperation and prosperity of Hong Kong and downplayed political disputes to demonstrate the success of the integration of Hong Kong with mainland China.

Fumo (2021) conducted a comparative-contrastive study on how the Italian and Spanish press depicted immigration during the period from 2017 to 2020. In particular, she focused on the semantic prosody of the terms *inmigrante* and *migrante*. Findings revealed that: 1) the most widely used modifiers in both languages were *irregolare*, *illegale* and *indocumentado*; 2) the verbs occurring frequently in association with the above nouns were *rimpatriare*, *detenere* and *intercettare* in the Italian corpus, and *detener*, *deportar* and *retener* in the Spanish corpus; and 3) a high recurrence of digits preceded the words in question. According to Fumo, through these linguistic features, migrants were represented by the media more as a commodity than as human beings.

In Cervi et Al. (2021) a comparative-contrastive analysis on how Muslims and Islam were represented in Italian and Spanish media was carried out, taking into consideration the online version of the two most influential conservative and left-leaning newspapers in each country (respectively, *Il Corriere della Sera* and *La Repubblica* for Italy, and *El Mundo* and *El País* for Spain) from 2015 to 2020. The keywords *Islam*, *Islamic Muslim*/*Muslims* in both languages were used to build the corpus. Quantitative analysis revealed the following results: 1) the most widely used Islam-related word was *Islamic*, followed by *Islam*, while *Muslim* was the least frequently used word; 2) left-leaning newspapers (*La Repubblica* and *El País*) dedicated more space to Islam-related topics than conservative newspapers; 3) Islam-related content increased immediately after a terroristic event; 4) Islam and Muslims were almost uniquely mentioned in hard news; 5) the most recurrent topic was terrorism; 6) the Italian media mostly used the definition *Islamic terrorism* referring to terrorism, while the Spanish media used terms or expressions related to *Jihad* (e.g. *jihadism*, *jihadist*, *jihadi terrorism*); 7) terrorism was mostly discussed in international news in Italy, while in Spain it was discussed almost evenly in international and national news; 8) the second recurrent topic was radicalization, which was connected to immigration and the problem of integration; and 9) both Spanish newspapers often framed Israel within the Israeli-Palestinian conflict, while the nation was not mentioned in Italy.

The qualitative analysis revealed the following results: 1) Islamophobic content was strongly present in all the newspapers; 2) Islamic civilization was represented as a monolithic world characterized by violence, thus clashing with Western civilization; 3) the issue of foreign fighters was treated more in Italian newspapers than in Spanish newspapers; 4) left-leaning newspapers tended to make a distinction between moderate and radical Muslims, while conservative newspapers did not; 5) as for soft news, most of the content was about Islamic traditions and, in particular, in left-leaning newspapers of both countries most of the articles referred to women's headscarves (e.g. *La Repubblica* framed the female veil as oppressive towards women); 6) Islamophobic content decreased over time; 7) Islamophobia was mainly represented by the construction of the relationship between Islam and terrorism; 7) conservative newspapers of both countries openly referred to the impossibility of coexistence between Muslims and non-Muslims; and 8) headlines showed a greater amount of explicit Islamophobia as opposed to texts.

According to the authors, these results confirm previous studies which showed that Islamophobia was predominant in media portrayals of Islam and Muslims in Italy and Spain. The authors also reached the conclusion that conservative media tend to represent Islam and Muslims more negatively than left-leaning media.

2.4 Conclusion

Overall, what emerges from these studies is that media tend to frame Islam and Muslims within a discourse of conflict and violence: Muslims are mostly presented as terrorists and Muslim women are portrayed as victims of their own culture. This negative representation of Islam and Muslims is based on recurrent linguistic choices (e.g. the choice of the active vs passive voice in sentences, the choice of different terms to refer to the same referent). In addition, media contribute to a representation of Muslims as a monolithic group of people whose values clash with Western civilization, thus making it difficult for them to integrate within non-Islamic countries. All these studies show that most of the articles exclude Muslims as potential readers, making a distinction between "us" (non-Muslims) and "them" (Muslims). Similarly, those studies which discuss other subject matters than Islam or Muslims show that: 1) media tend to make a distinction between "us" (natives) and "them" (migrants); and 2) media discourse produced in countries with opposing ideologies tend to use different terms to represent the same entities and phenomena.

This section has shown that most of the scholars conducted comparative-contrastive studies, using a mixed-method (i.e. qualitative and quantitative) approach. The data usually consisted of journalistic texts, but occasionally included magazine and radio or television data. The main focus was on news produced by US media, contrasted with that coming from varied sources: Middle Eastern, UK, Italian and Spanish, Australian and Chinese news. The results of these studies broaden our understanding of Western representation of "other" people and/or events. An analysis which has not been carried out yet, though, is a comparative mixed-method study of media's representation of Islam and Muslims in the US media vs Pakistani media. This is the focus of the present work, in which I intend to investigate how newspapers from these two countries discuss Islam and Muslims.

3. Methodological approach

3.1 Introduction

The aim of this chapter is to give an overview of the methodological approach of the present study. Firstly, the theoretical framework adopted to design the research project will be outlined, including a brief description of the specific types of analysis that it supports. Secondly, the corpora investigated will be presented, specifying how they were collected and organized, and pointing out their main characteristics. Thirdly, the focus of the present work will be specified, presenting the research questions and how they were operationalized. Each section of this chapter therefore introduces key notions that set the context for the analysis whose results are reported in the following chapter.

3.2 Theoretical framework

In the present work, data will be analyzed through a *corpus-driven* approach, as outlined in Tognini-Bonelli (2001: 17): "the corpus driven approach builds up the theory step-by-step in the presence of evidence. The observation of certain patterns leads to a hypothesis, which in turn leads to the generalization in terms of rules of usage and finally finds unification in a theoretical statement" ["sic"]. That is, a text or collection of texts is examined with a bottom-up approach, so as to identify recurrent phraseological patterns in it/them. These patterns (i.e. recurrent word combinations) enable a researcher to formulate a theory about what the text(s) is/are about, what it/they is/are like, and about the cultural frames which they trigger (Stubbs 2002: 17).

For example, in his *corpus-driven* analysis of a large corpus, Stubbs (2002: 16) discovered that the words for different days of the week differed considerably in frequency: "the category days-of-the-week is culturally constructed, and there are cultural reasons why people talk most often about the weekend, less often about the beginning and end of the working week, and less often again about the days in the middle of the week" (Stubbs 2002: 17). He also noticed that the seven words tend to occur in different phrases, such as: *Friday night*; *Saturday night*; *Sunday afternoon*; *Monday morning*; *that Monday morning feeling*; *Monday morning blues* (Stubbs 2002: 17). He pointed out that

"it is formally possible (i.e. grammatical) to say *Sunday night*, but *Saturday night* is more frequent, and this is a fact with cultural significance (Stubbs 2002: 17).

The *corpus-driven* approach can be distinguished from the *corpus-based* approach, whose methodology "avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study" (Tognini-Bonelli 2001: 65). That is, starting with a set of preconceived ideas about the behaviour of linguistic categories, texts are analyzed in order to validate these assumptions. An example of the *corpus-based* approach can be found in Baker's (2006: 14-15) analysis of the meanings of the word *blind* in corpora of different periods (1960s and 1990s). The author found that in the 1960s corpus, the word *blind* almost always appeared in a literal sense, referring to people or animals who cannot see, while in the 1990s corpus, in about half its occurrences, *blind* was used in a range of metaphorical ways: *turn a blind eye*, *blind ambition*, *sheer blind anger*, *blind panic*, *blind patriotism*, *the blind lead the blind*, *blind to change* (Baker 2006: 15). The author concluded that the negative metaphorical meaning of *blind* appeared to have increased in written British English over time (Baker 2006: 15).

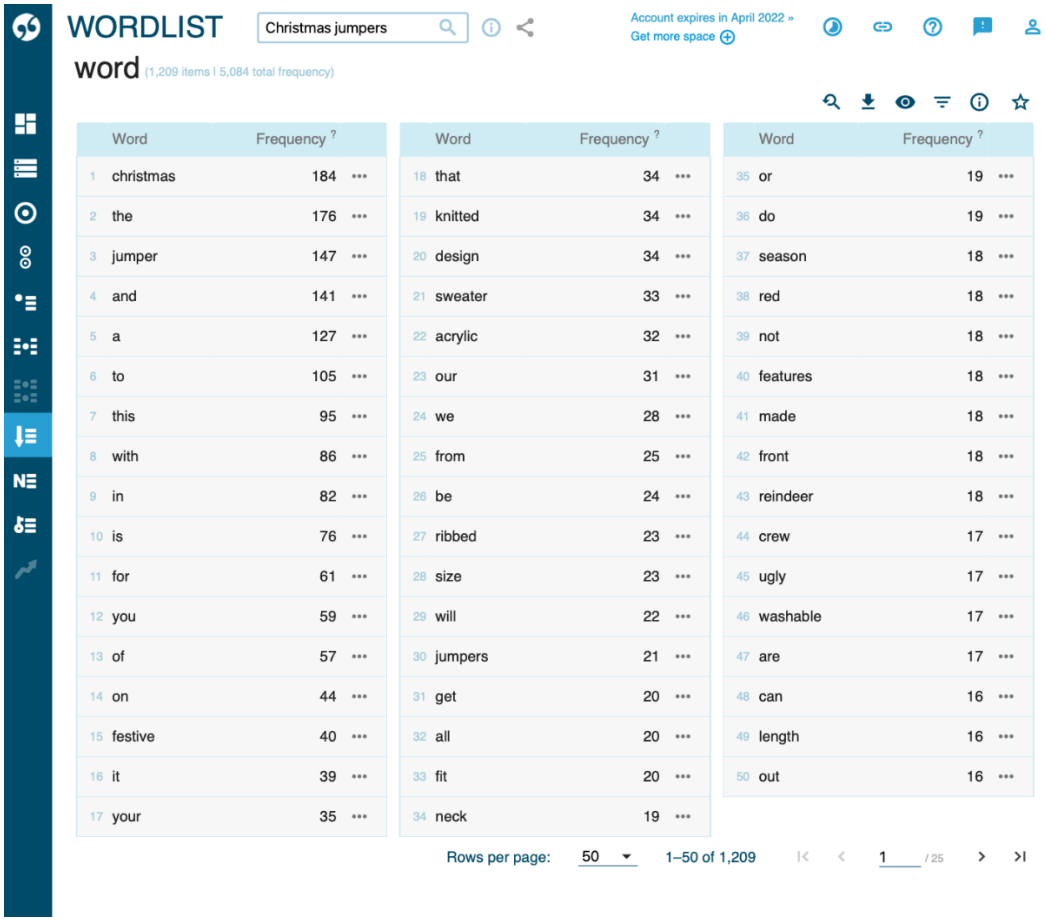
The main advantage of the *corpus-driven* approach is that it leads an analyst to uncover new grounds, posit new hypotheses and not always support old ones; in contrast, the *corpus-based* approach works always within accepted frameworks.

In terms of Saussure's famous distinction between *langue* and *parole*, in the *corpus-driven* approach, the starting point is *parole*, that is, the individual linguistic manifestation of the abstract system of grammar, while in the *corpus-based* approach, the starting point is *langue*, namely, the whole set of linguistic habits which enables the speaker to understand and to make him/herself understood (Tognini-Bonelli 2001: 169).

Both in the corpus-driven and in the corpus-based approach, data are typically stored in electronic format and specific text-analysis software programmes are used in order to search and process data. In particular, with corpus analysis software, it is possible to retrieve many kinds of information about the lexical make-up of texts. In the following illustrative descriptions, examples are drawn from the Sketch Engine platform (www.sketchengine.eu), which contains many ready-made corpora:

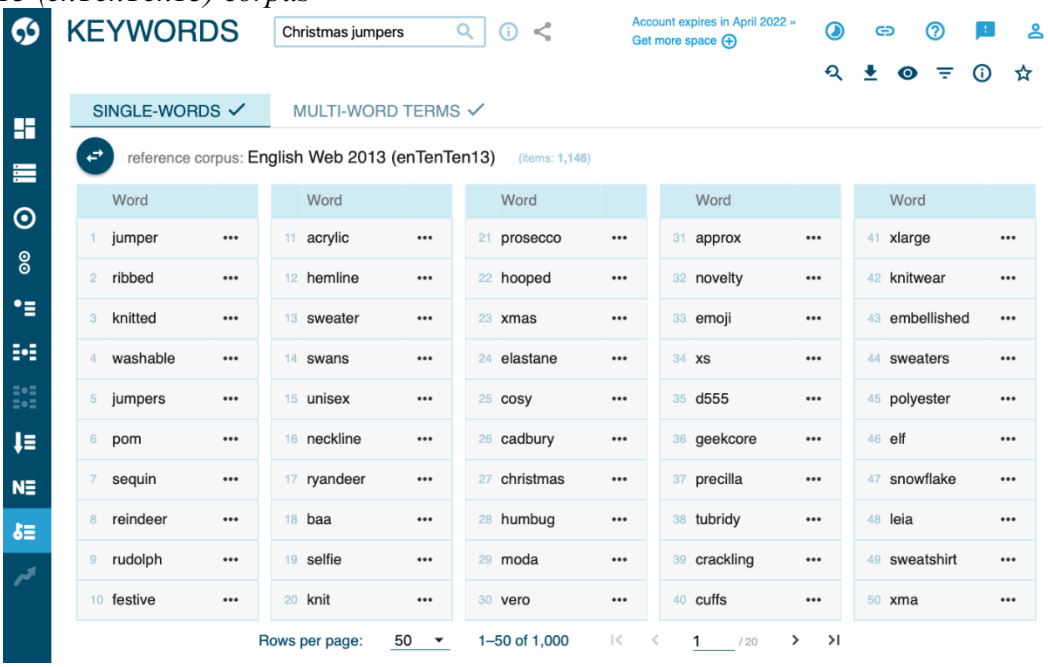
1) *word frequency lists* can be retrieved, that is, lists of the words that make up a corpus, in decreasing order of frequency. For example, Figure 3.1 shows a word frequency list of the '*Christmas jumpers*' corpus.

Figure 3.1: wordlist of the Christmas jumpers corpus



2) *keyword lists* can also be compiled, namely, lists of words which are unexpectedly frequent/infrequent in one, usually smaller, corpus when compared to another, usually much larger corpus, called a *reference corpus*. For example, Figure 3.2 shows a keyword list retrieved from a comparison between the *Christmas jumpers* corpus, a relatively small and specialized corpus, and the English *Web 2013 (enTenTen13)* corpus, a larger and more general corpus.

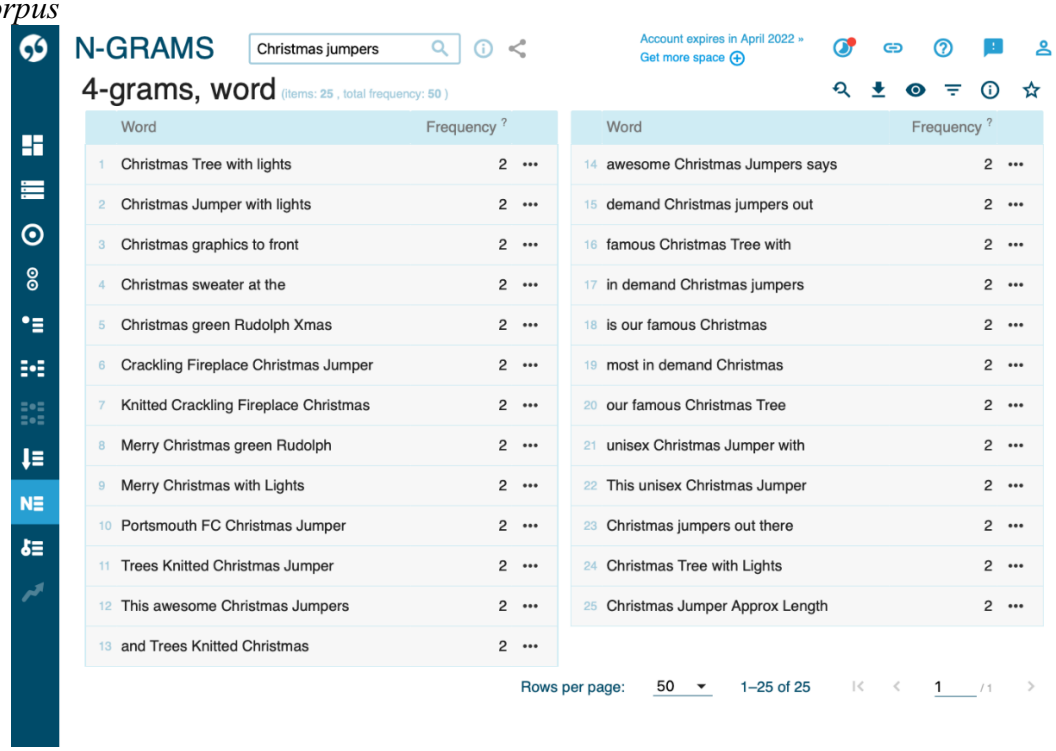
Figure 3.2: keyword list of the Christmas jumpers corpus compared to the English Web 2013 (enTenTen13) corpus



Both word frequency lists and keyword lists are useful for identifying the main topics/notions mentioned in a corpus. The former suggests what a given corpus is about, that is, its prominent topics; the latter show what is more distinctive about it, when compared to another, similar corpus relevant to the same subject matter.

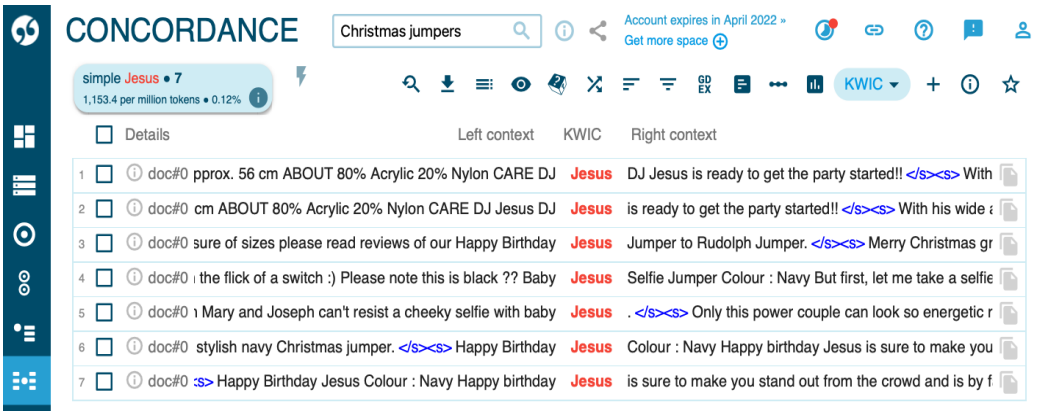
3) *Clusters* can also be identified with corpus software. They are recurrent multi-word units built around given node words. For example, Figure 3.3 shows a frequency list of 4-words units containing the word *Christmas* in the *Christmas jumper* corpus.

Figure 3.3: 4-words units containing the word 'Christmas' in the Christmas jumper corpus



4) *Concordances* are the output of corpus searches in which a chosen node word/phrase is presented in the centre of the computer screen, with the words that come before and after it to the left and to the right (Hunston 2002: 39). For example, Figure 3.4 shows the concordances of the word *Jesus* in the *Christmas jumper* corpus.

Figure 3.4: concordances of the word 'Jesus' in the Christmas jumpers corpus



Concordance lines allow a researcher to observe: 1) the 'central and typical'; 2) meaning distinctions; 3) meaning and pattern; and 4) detail (Hunston 2002: 42). I will illustrate these notions by drawing on Hunston's (2002) examples.

1) The 'central and typical' is the group of the most frequent meanings or patterns of an individual word or phrase (Hunston 2002: 42-43). For example, the phrase *recipe for* typically has a metaphoric rather than literal meaning, is more likely to be followed by negative terms (e.g. *damage*, *failure*, *chaos*, *slump*) and to be preceded by the verb *be,* thus being characteristic of the sequence "something *is a recipe for* something bad" (Hunston 2002: 43).

2) Meaning distinctions. As observed by Hunston (2002: 45), 'many words have meaning that are similar, and yet the words are not able to be substituted one for the other'. Differences in meaning can be retrieved from concordances: for example, the adjectives *sheer*, *pure*, *complete*, *utter* and *absolute* are regarded as synonyms by many dictionaries (Hunston 2002: 45). However, *sheer* is used with nouns of degree or magnitude (*sheer weight*, *sheer number*), while the other adjectives do not collocate with these nouns (Hunston 2002: 45).

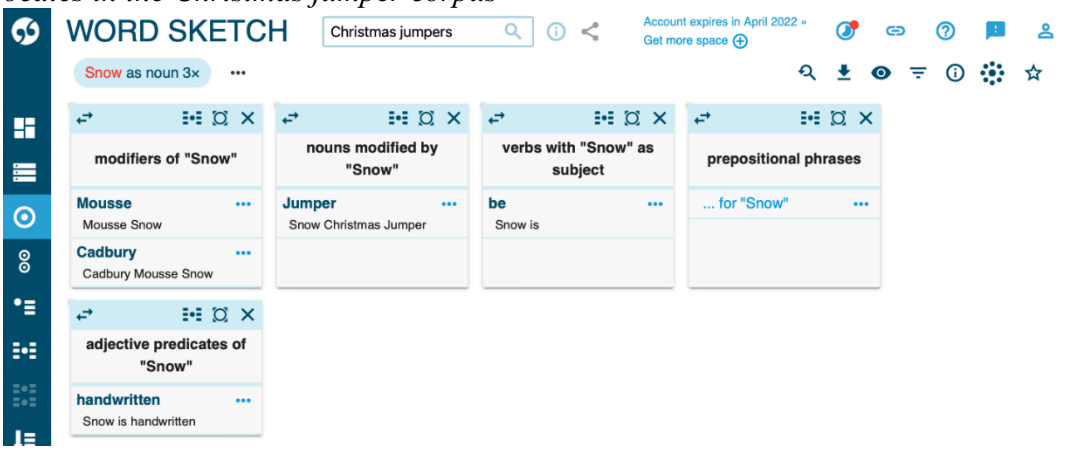
3) Meaning and pattern. According to Hunston (2002: 47-48), words which share patterns tend to have similar meanings, and vice versa. For example, she found that the verb *condemn* has several meanings which are associated with different patterns: the 'criticize' meaning is associated with the pattern 'condemn something as something', as in '*The Fable of the Bees was tried and condemned as a public nuisance by the Grand'* ; the 'pass sentence' meaning and the 'make something bad happen' meaning are associated with the pattern 'condemn something to something', as in *'people who a few days before had condemned him to death'*, and *'Four decades ago, Asia seemed condemned to poverty'* (Hunston 2002: 47).

4) Observing detail. For example, Hunston (2002: 51) found that the nouns *advice* and *answer* are often followed by *as to* and a clause beginning with a *wh*-word. However, more detail can be added to this patterning: *advice as to* often follows a verb indicating 'getting', 'giving', 'wanting', or 'offering', (as in '*I would need his advice as to how to cut around his ears*' (Hunston, 2002: 51)).

5) More generally, concordances show *collocations*, that is, the words with which a word most characteristically occurs. For example, Figure 3.4 shows the words

and classes of words with which the lemma[1] *snow* frequently collocates in the *Christmas jumper* corpus.

Figure 3.4: words and classes of words with which the lemma 'snow' frequently collocates in the Christmas jumper corpus



The lemma *snow* is the 'node', that is, the word-form being investigated; the lemmas *Mousse*, *Cadbury*, *Jumper*, *be*, *for*, *handwritten* are 'collocates' of the node, namely, the word forms with which the node co-occurs in the corpus (Stubbs 2002: 29). *Collocation* makes it possible to: 1) give a semantic profile of the word involved. For example, Stubbs (1995; 1996, quoted in Hunston 2002: 76) found that *cause* is typically used with nouns indicating 'something bad' (*anxiety*, *concern*, *AIDS*, *cancer*, etc.); and 2) determine if it belongs to a given semantic field(s). For example, the collocates of *bribe* and *bribery* show their relevance to the fields of wrong-doing (e.g. *allegations*, *scandal*, *corruption*), money (e.g. *dollar*, *money*, *tax*) and sport (e.g. *players*, *referee*) (Orpin 1997, quoted in Hunston 2002: 78).

A collocational analysis is important because:

> it gives the most salient and obvious lexical patterns surrounding a subject, from which a number of discourses can be obtained. When two words frequently collocate, there is evidence that the discourses surrounding them are particularly powerful. […] Collocates can therefore act as as triggers, suggesting unconscious associations through which discourses are maintained (Baker 2006: 114).

[1] A form of a word that appears as an entry in a dictionary and is used to represent all the other possible forms. For example, the lemma "build" represents "builds", "building", "built", etc.' (https://dictionary.cambridge.org/it/dizionario/inglese/lemma).

In other words, collocations can reveal: 1) if the searched word is used in specific contexts and 2) speakers and writers' unconscious attitudes and ideologies. For these two aspects collocation can be regarded as the basis for the identifying other patterns, that is, 1) *colligation*; 2) *semantic preference*; and 3) *semantic prosody*.

Colligation refers to "the syntactic constraints, or indeed preferences, that a specific word, seen as a unique lexical item rather than as member of its class, entertains with its environment" (Tognini-Bonelli 2001: 89), that is, the grammatical relations that a word establishes with other words in its surroundings. For example, Stubbs (2002: 65) noted that the word-form *cases* frequently occurs with the grammatical category of quantifiers, such as *in some cases*, *in many cases*, and Sinclair (quoted in Tognini-Bonelli 2001: 104) observed how *naked eye* is used preceded by the preposition *to* or *with* and the definite article *to/with the naked eye*.

Semantic preference refers to the relation between a lemma or word-form and a set of semantically related words. For example, Stubbs (2002: 64) found that the word-form *commit* is likely to co-occur with words like *suicide*, *crime*, *murder*, which share the same semantic field, namely violence. Investigating the semantic preference of a lemma in a corpus can reveal if the lemma is used in specific contexts.

The last pattern that can be explored with corpus software is *semantic prosody* (also called *discourse prosody* by Baker (2006: 87)), which is defined by Louw as "a consistent aura of meaning with which a form is imbued by its collocates" (Tognini-Bonelli, 2001: 111-12). In other words, if a word is regularly used in contexts of good news or bad news, it carries this kind of meaning around with it even outside its context. For example, Tognini-Bonelli (2001: 111) reports how Sinclair (1987f, 1991) investigated the pattern associated with the phrasal verb *set in*, noticing how its typical subject is something unpleasant, and listed *rot*, *decay*, *malaise*, *despair*, *ill-will*, *decadence*, *impoverishment*, *infection*, *prejudice*, *vicious (circle)*, as a result of which it has acquired an aura of negative meaning. Semantic prosodies reveal speakers and writers' unconscious attitudes and ideologies.

The difference between semantic preference and semantic prosody is not always clear cut: it depends on how open-ended the list of collocates is. As Baker (2006: 87) points out, a list of all of the words for 'drinks' indicates a semantic preference, but an

open-ended category such as 'unpleasant things' might be seen as a discourse prosody. In addition, semantic preference is independent of speakers, whereas discourse prosody focuses on the relationship of a word to speakers and hearers (Baker 2006: 87).

3.3 Data collection and description

The corpora considered for this study are made up of data downloaded from the online versions of the *New York Times* and *The Dawn* newspapers. These newspapers were chosen as data sources because they are amongst the leading newspapers in the USA and Pakistan, respectively. To have access to and download *New York Times* news articles, a €1 subscription fee is required. Instead, *The Dawn* gives its readers full access to its news articles for free.

The data were collected through the *Sketch Engine* platform (https://www.sketchengine.eu), which also serves as corpus-building software: they were downloaded directly from the newspapers' websites (www.nytimes.com and dawn.com) by using the following combination of search words: *Islam*, *Islamic*, *Islamically*, *Muslim* and *Muslims*. *Sketch Engine*'s default settings regarding the choice of search words and relevant sites – which can be specified under the *Web search settings* menu – were used for the search. Therefore, data were cleaned and non-text was eliminated automatically by *Sketch Engine*, as explained in the user guide. Since the data obtained sometimes contained pairs of texts with very similar or even identical content, I used the *Remove duplicated content* function so as to only keep one instance of each text.

As can be seen in Table 3.1, the final version of *The Dawn* corpus consisted of 138,542 tokens and 31 documents, while that of the *New York Times* consisted of 204,843 tokens and 106 documents. It thus appears that *The Dawn*'s articles are fewer than, but longer than, those of the *New York Times*. Moreover, it is interesting that there is an almost equal number of sentences and paragraphs in both newspapers (i.e. 9,509 sentences in *The Dawn* corpus and 8,787 sentences in the *New York Times* corpus; 5,917 paragraphs in *The Dawn* corpus and 3,918 paragraphs in the *New York Times* corpus). Given the different size of the corpora, this indicates that the sentences and paragraphs in *The Dawn* corpus are shorter than those in the *New York Times* corpus.

Table 3.1: Size of the corpora

	***The Dawn* corpus**	***New York Times* corpus**
Tokens	138,542	204,843
Words	114,977	175,756
Sentences	9,509	8,787
Paragraphs	5,917	3,918
Documents	31	106

3.4 Focus of the present study

The focus of this study is the representation of Islam and Muslims in the *New York Times* and *The Dawn* newspapers. The questions addressed are the following:

1) What are the most recurrent topics/notions mentioned?

2) How are Islam and/or Muslims talked about?

3) Do media mostly report good news or bad news about Islam and Muslims?

4) How similarly or differently are Islam and/or Muslims characterized in the two corpora?

In order to answer these questions, the following corpus-driven analyses will be carried out by using the corpus analysis tools available on the *Sketch Engine* platform:

1) The 'frequency *wordlist*' and the '*N-grams*' will show the most frequent words and multi-word units, respectively in each corpus; while the '*keyword*' lists obtained by comparing one corpus against the other and both against the *EnglishWeb enTenTen2020* corpus will identify the most unusually frequent words in each corpus, thus revealing the most recurrent topics/notions mentioned in the corpora;

2) The '*Concordances*' and the '*Word Sketch*' of the five most frequent words and the top five keywords will detect collocations and possible colligations and semantic preferences of those terms;

3) The results of the previous operation will be analyzed in depth to see if the terms which refer to Islam and Muslims carry particular discourse prosodies;

4) Finally, the '*Word Sketch Difference*' function will serve to highlight shared and unshared collocations of *Islam*, *Muslim* and *Islamic* in the two corpora.

The next chapter will report the results of the above analyses, which will then be discussed in Chapter 5

4. Findings

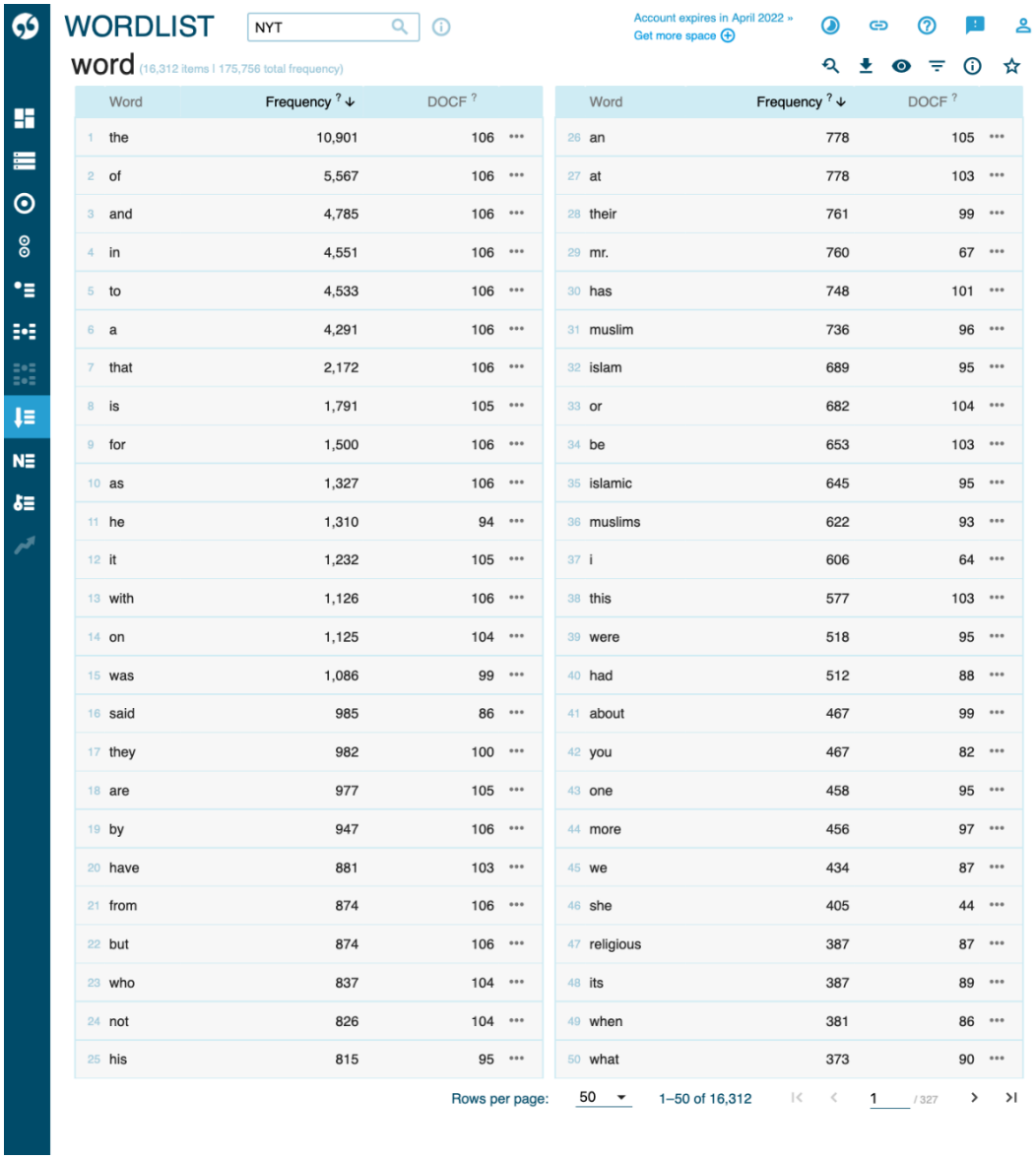
4.1 Introduction

In this chapter I am going to present the findings of my analysis of the *New York Times* and *The Dawn* corpora. Sections 4.2 and 4.3 will outline the most recurrent topics and notions mentioned in the *New York Times* and in *The Dawn* corpora respectively, identified by examining word frequency lists and keyword lists. It will also present the semantic preferences and semantic prosodies of the five most frequent words and the top five keywords, explored through an analysis of their concordances as well through the *Word Sketch* and the *Word Sketch Difference* functions available in Sketch Engine.

4.2 Findings about the *New York Times* Corpus

In order to find the most recurrent topics and notions mentioned in the *New York Times* corpus, first of all, it is useful to look at word frequency lists. Figures 4.1-A-B-C-D show the 200 most frequent words in the *New York Times* corpus listed in order of decreasing frequency.

Figure 4.1-A: word frequency list of the New York Times corpus, Part A



	Word	Frequency	DOCF
1	the	10,901	106
2	of	5,567	106
3	and	4,785	106
4	in	4,551	106
5	to	4,533	106
6	a	4,291	106
7	that	2,172	106
8	is	1,791	105
9	for	1,500	106
10	as	1,327	106
11	he	1,310	94
12	it	1,232	105
13	with	1,126	106
14	on	1,125	104
15	was	1,086	99
16	said	985	86
17	they	982	100
18	are	977	105
19	by	947	106
20	have	881	103
21	from	874	106
22	but	874	106
23	who	837	104
24	not	826	104
25	his	815	95

	Word	Frequency	DOCF
26	an	778	105
27	at	778	103
28	their	761	99
29	mr.	760	67
30	has	748	101
31	muslim	736	96
32	islam	689	95
33	or	682	104
34	be	653	103
35	islamic	645	95
36	muslims	622	93
37	i	606	64
38	this	577	103
39	were	518	95
40	had	512	88
41	about	467	99
42	you	467	82
43	one	458	95
44	more	456	97
45	we	434	87
46	she	405	44
47	religious	387	87
48	its	387	89
49	when	381	86
50	what	373	90

Figure 4.1-B: word frequency list of the New York Times corpus, Part B

	Word	Frequency	DOCF
51	been	371	89
52	there	354	94
53	them	346	86
54	all	344	88
55	some	337	93
56	her	336	34
57	which	330	94
58	people	324	84
59	also	324	92
60	like	323	91
61	n't	314	70
62	will	313	77
63	many	310	91
64	if	298	82
65	other	298	90
66	do	297	88
67	most	293	83
68	women	291	50
69	up	290	78
70	can	290	80
71	after	287	77
72	even	286	84
73	would	284	76
74	so	281	84
75	into	280	80

	Word	Frequency	DOCF
76	world	274	78
77	new	270	77
78	no	258	73
79	than	257	80
80	out	247	83
81	against	240	72
82	years	239	72
83	american	235	54
84	only	231	85
85	these	223	74
86	now	217	77
87	government	203	59
88	where	203	71
89	because	200	74
90	over	196	72
91	did	191	66
92	political	191	60
93	our	189	63
94	how	187	69
95	two	185	77
96	religion	183	60
97	could	183	64
98	last	180	75
99	first	174	65
100	mosque	171	40

Figure 4.1-C: word frequency list of the New York Times corpus, Part C

	Word	Frequency	DOCF
101	state	169	60
102	those	168	66
103	iran	165	23
104	him	164	57
105	just	161	59
106	say	161	60
107	group	160	59
108	such	158	67
109	time	157	68
110	states	155	59
111	al	155	34
112	while	154	73
113	my	152	36
114	before	152	66
115	among	150	67
116	should	149	56
117	united	149	52
118	then	148	64
119	any	148	70
120	saudi	145	40
121	middle	145	44
122	men	144	50
123	being	140	68
124	public	140	55
125	between	139	64

	Word	Frequency	DOCF
126	country	139	56
127	law	136	51
128	under	135	57
129	young	135	51
130	me	135	30
131	since	133	64
132	around	133	56
133	year	131	53
134	through	130	57
135	may	130	56
136	here	129	54
137	attacks	128	49
138	officials	128	44
139	says	126	38
140	see	126	65
141	power	125	40
142	imam	123	28
143	own	122	61
144	party	122	31
145	us	121	48
146	your	121	38
147	does	120	57
148	often	120	55
149	faith	119	45
150	told	118	45

Figure 4.1-D: word frequency list of the New York Times corpus, Part D

	Word	Frequency	DOCF
151	read	116	64
152	family	116	40
153	dr.	115	14
154	much	115	51
155	way	115	52
156	violence	114	46
157	long	114	54
158	life	113	52
159	home	113	49
160	east	113	42
161	xinjiang	111	6
162	society	111	48
163	leaders	111	46
164	black	111	26
165	community	111	46
166	part	109	61
167	god	109	43
168	must	108	48
169	president	108	51
170	school	108	42
171	including	107	53
172	children	107	38
173	called	107	54
174	iraq	107	27
175	percent	106	43

	Word	Frequency	DOCF
176	made	106	50
177	during	105	55
178	iranian	105	17
179	western	104	41
180	war	102	41
181	times	102	57
182	man	101	45
183	university	100	42
184	both	100	59
185	become	100	52
186	groups	100	41
187	few	99	51
188	city	99	38
189	french	97	21
190	center	97	47
191	three	97	50
192	recent	96	52
193	members	96	52
194	day	95	40
195	well	95	49
196	too	95	43
197	take	94	48
198	come	94	49
199	still	92	48
200	according	92	39

The most frequent words shown in Figures 4.1-A-B-C-D which are relevant to the Islamic world can be grouped into the following semantic categories: RELIGION (*faith*, *god*, *imam*, *islam*, *islamic*, *mosque*, *muslim*, *muslims*, *religion*, *religious*), POLITICS (*government*, *law*, *leaders*, *party*, *political*, *power*, *president*, *state*, *states*, *war*), GEOGRAPHICAL ORIGIN (*american*, *country*, *east*, *french*, *iran*, *iranian*, *iraq*, *middle*[2], *saudi*, *states*, *united*, *xinjang*, *western*) and VIOLENCE (*attacks*, *violence*, *war*).

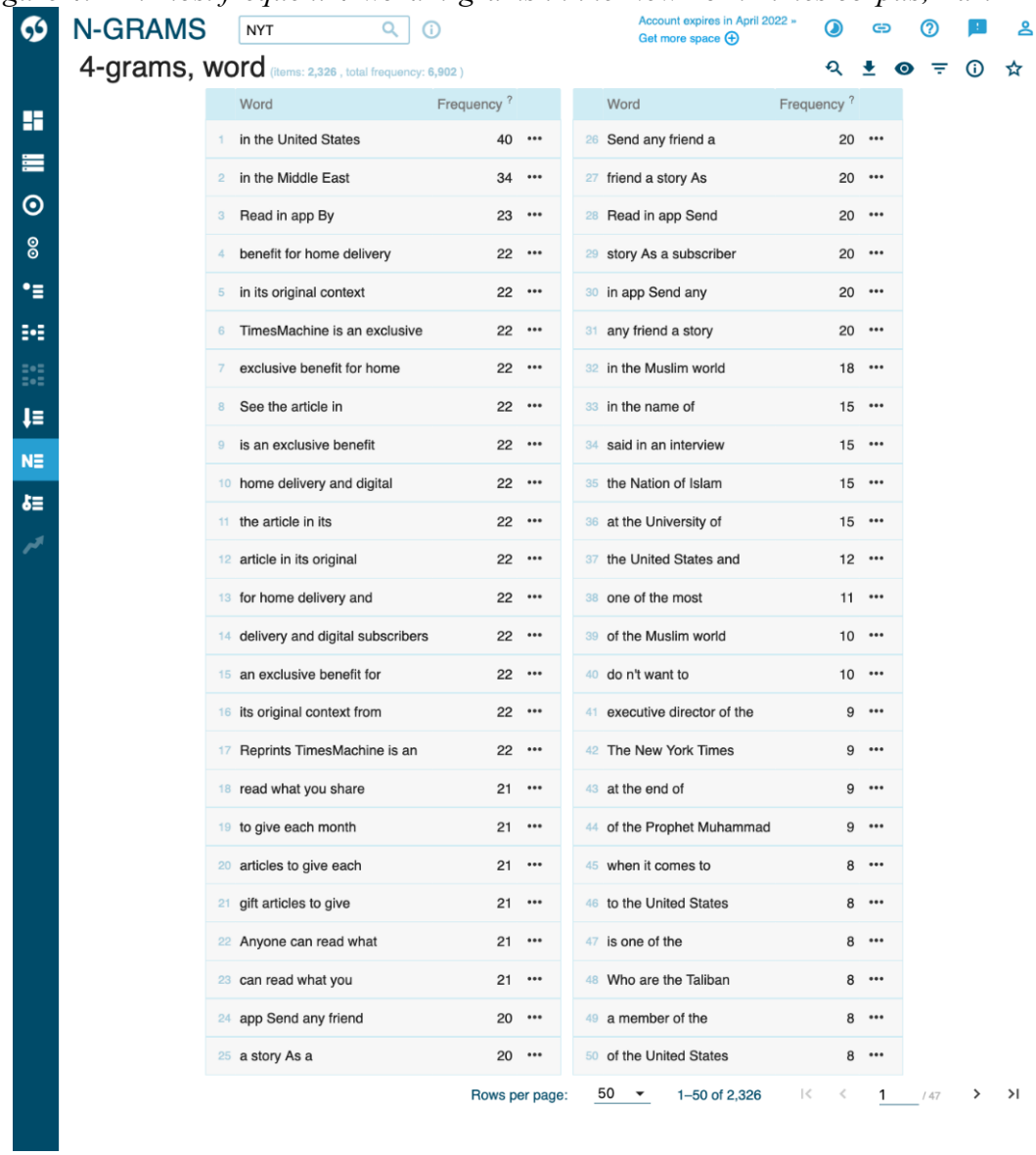
[2] This term is found in geographic expressions such as *Middle East*, *Middle Eastern*, *middle of a salt desert*, *middle of the floor*. However, the term is also found in a relatively small group of non-geographic expressions, such as *middle class*, *Middle Ages*, *middle finger*, *middle school*, *middle of the game*, *middle*

Other terms which appear in the frequency word list can be grouped into the following semantic categories: PEOPLE (*black*, *children*, *community*, *family*, *men*, *officials*, *people*, *public*, *young*, *members, women*), EDUCATION (*school, university*), SOCIETY (*group*, *home*, *society*, *world*) and OTHER (*center*, *life*, *time*). Although these words seem to cover other topics, they are actually relevant to the Islamic world too, because they collocate with words from that semantic field (e.g. *Muslim*, *Islamic*, *religious*, *observant*, etc.). The extent to which this occurs is specified next: *members*: 44.8% of the time; *world*: 43.4%; *community*: 36.1%; *society*: 32.6%; *group*: 31%; *university*: 30.6%; *young*: 26.7%; *women*: 24.8%; *center*: 23.9%; *black*: 23.3%; *school*: 22.6%; *home*: 22.4%; *officials*: 12.5%; *men*: 12.5%; *family*: 10.2%; *public*: 7.8%; *children*: 7.5%; *life*: 7.2%; *time*: 6%; *people*: 5.3%.

In order to have a more contextualized view of the discourse on Muslims and Islam, it is useful to identify recurrent phrases in the corpora, as can be retrieved through the *N-gram* function: Figures 4.2-A-B-C-D show the 200 most frequent sequences of four tokens with a minimum frequency of two occurrences in the *New York Times* corpus, listed in order of decreasing frequency.

of the traffic, *middle of the night*, *middle of the winter*, *critical middle*, *middle of a deal*, *middle school's traffic*, *The Middle East Journal*, *middle of the 13th century*.

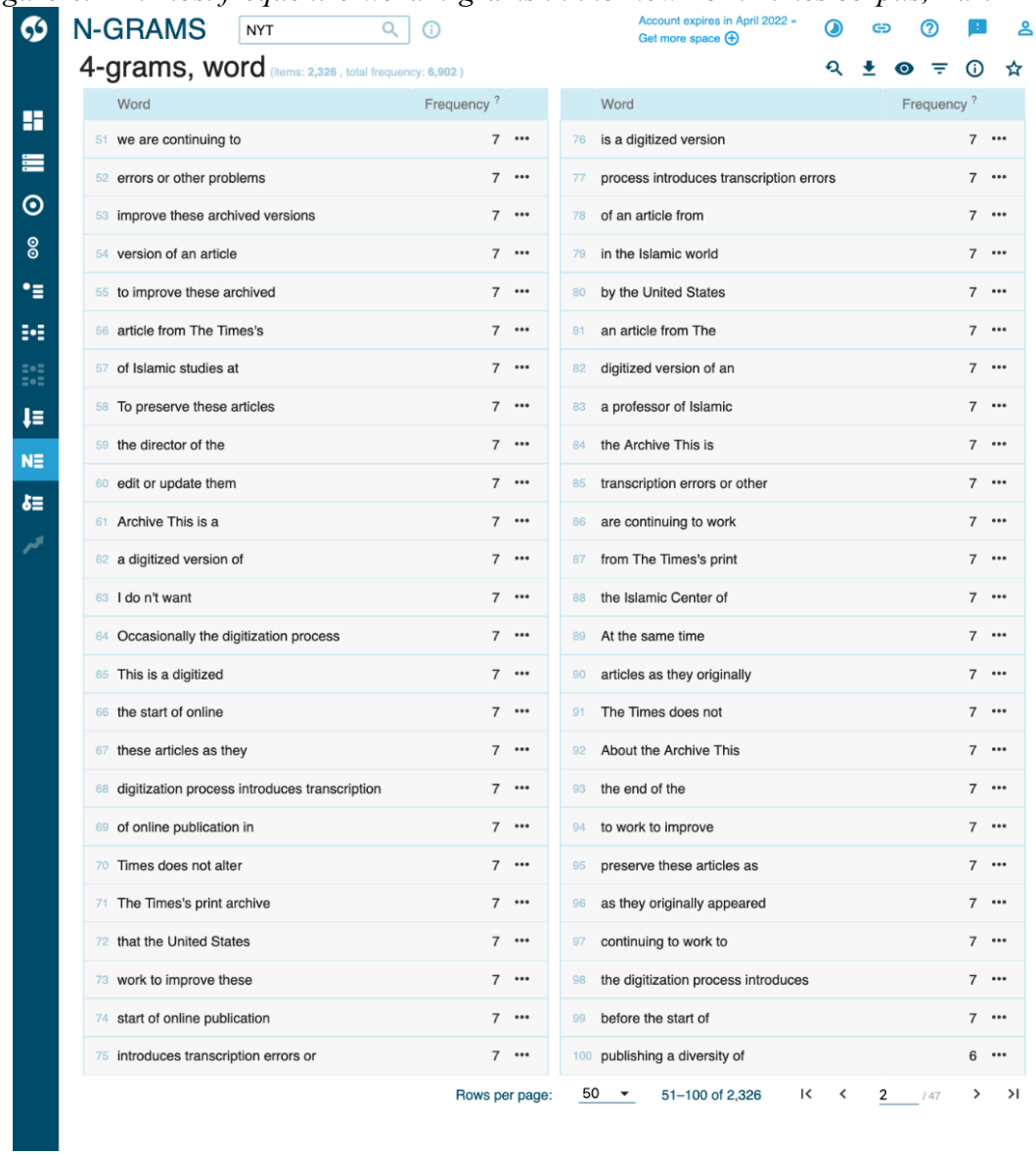
Figure 4.2-A: most frequent 4-word n-grams in the New York Times corpus, Part A



	Word	Frequency
1	in the United States	40
2	in the Middle East	34
3	Read in app By	23
4	benefit for home delivery	22
5	in its original context	22
6	TimesMachine is an exclusive	22
7	exclusive benefit for home	22
8	See the article in	22
9	is an exclusive benefit	22
10	home delivery and digital	22
11	the article in its	22
12	article in its original	22
13	for home delivery and	22
14	delivery and digital subscribers	22
15	an exclusive benefit for	22
16	its original context from	22
17	Reprints TimesMachine is an	22
18	read what you share	21
19	to give each month	21
20	articles to give each	21
21	gift articles to give	21
22	Anyone can read what	21
23	can read what you	21
24	app Send any friend	20
25	a story As a	20

	Word	Frequency
26	Send any friend a	20
27	friend a story As	20
28	Read in app Send	20
29	story As a subscriber	20
30	in app Send any	20
31	any friend a story	20
32	in the Muslim world	18
33	in the name of	15
34	said in an interview	15
35	the Nation of Islam	15
36	at the University of	15
37	the United States and	12
38	one of the most	11
39	of the Muslim world	10
40	do n't want to	10
41	executive director of the	9
42	The New York Times	9
43	at the end of	9
44	of the Prophet Muhammad	9
45	when it comes to	8
46	to the United States	8
47	is one of the	8
48	Who are the Taliban	8
49	a member of the	8
50	of the United States	8

Figure 4.2-B: most frequent 4-word n-grams in the New York Times corpus, Part B



N-GRAMS NYT

Account expires in April 2022 »
Get more space

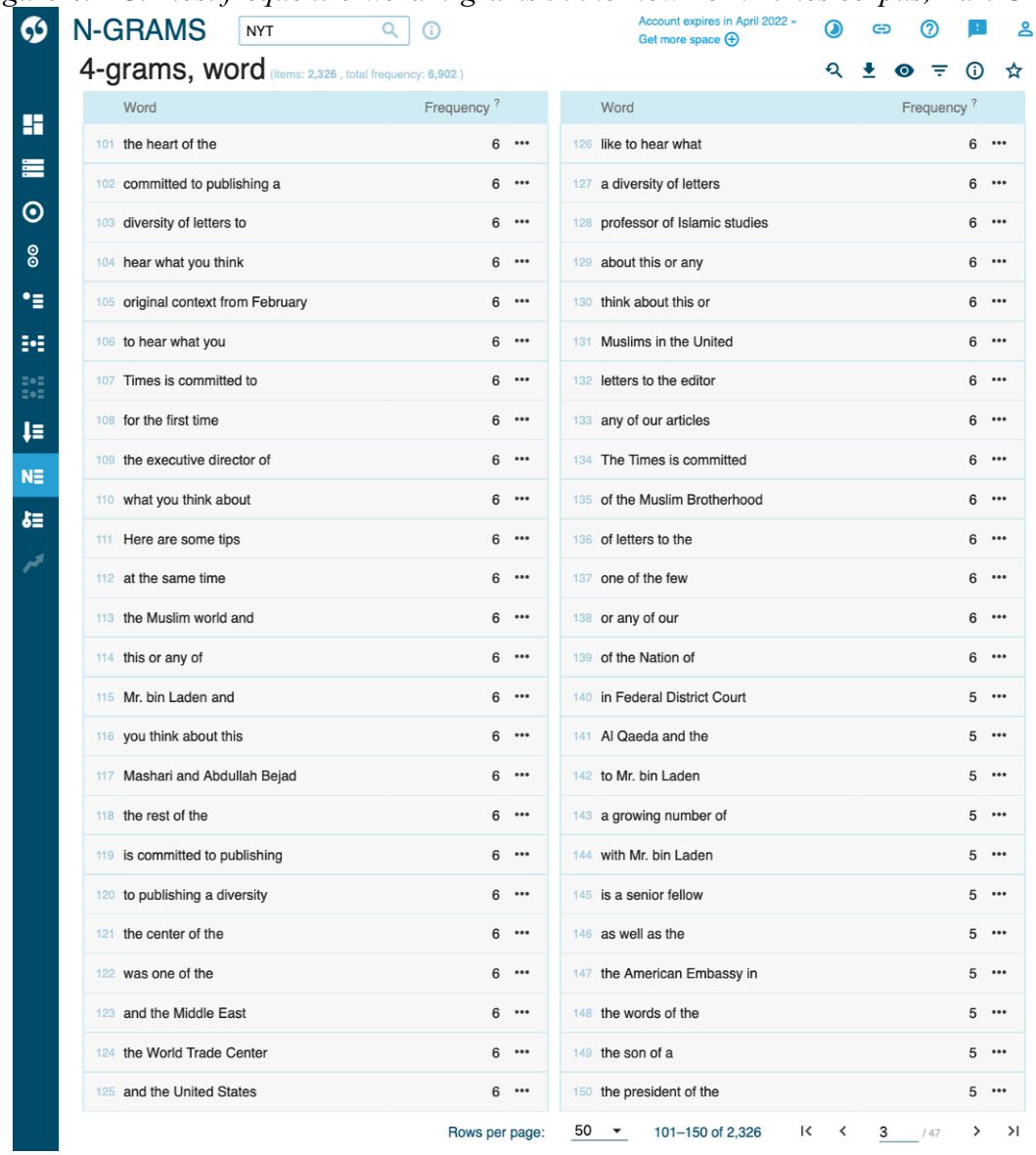
4-grams, word (items: 2,326 , total frequency: 6,902)

	Word	Frequency
51	we are continuing to	7
52	errors or other problems	7
53	improve these archived versions	7
54	version of an article	7
55	to improve these archived	7
56	article from The Times's	7
57	of Islamic studies at	7
58	To preserve these articles	7
59	the director of the	7
60	edit or update them	7
61	Archive This is a	7
62	a digitized version of	7
63	I do n't want	7
64	Occasionally the digitization process	7
65	This is a digitized	7
66	the start of online	7
67	these articles as they	7
68	digitization process introduces transcription	7
69	of online publication in	7
70	Times does not alter	7
71	The Times's print archive	7
72	that the United States	7
73	work to improve these	7
74	start of online publication	7
75	introduces transcription errors or	7

	Word	Frequency
76	is a digitized version	7
77	process introduces transcription errors	7
78	of an article from	7
79	in the Islamic world	7
80	by the United States	7
81	an article from The	7
82	digitized version of an	7
83	a professor of Islamic	7
84	the Archive This is	7
85	transcription errors or other	7
86	are continuing to work	7
87	from The Times's print	7
88	the Islamic Center of	7
89	At the same time	7
90	articles as they originally	7
91	The Times does not	7
92	About the Archive This	7
93	the end of the	7
94	to work to improve	7
95	preserve these articles as	7
96	as they originally appeared	7
97	continuing to work to	7
98	the digitization process introduces	7
99	before the start of	7
100	publishing a diversity of	6

Rows per page: 50 51–100 of 2,326 2 / 47

Figure 4.2-C: most frequent 4-word n-grams in the New York Times corpus, Part C



N-GRAMS NYT

Account expires in April 2022 »
Get more space

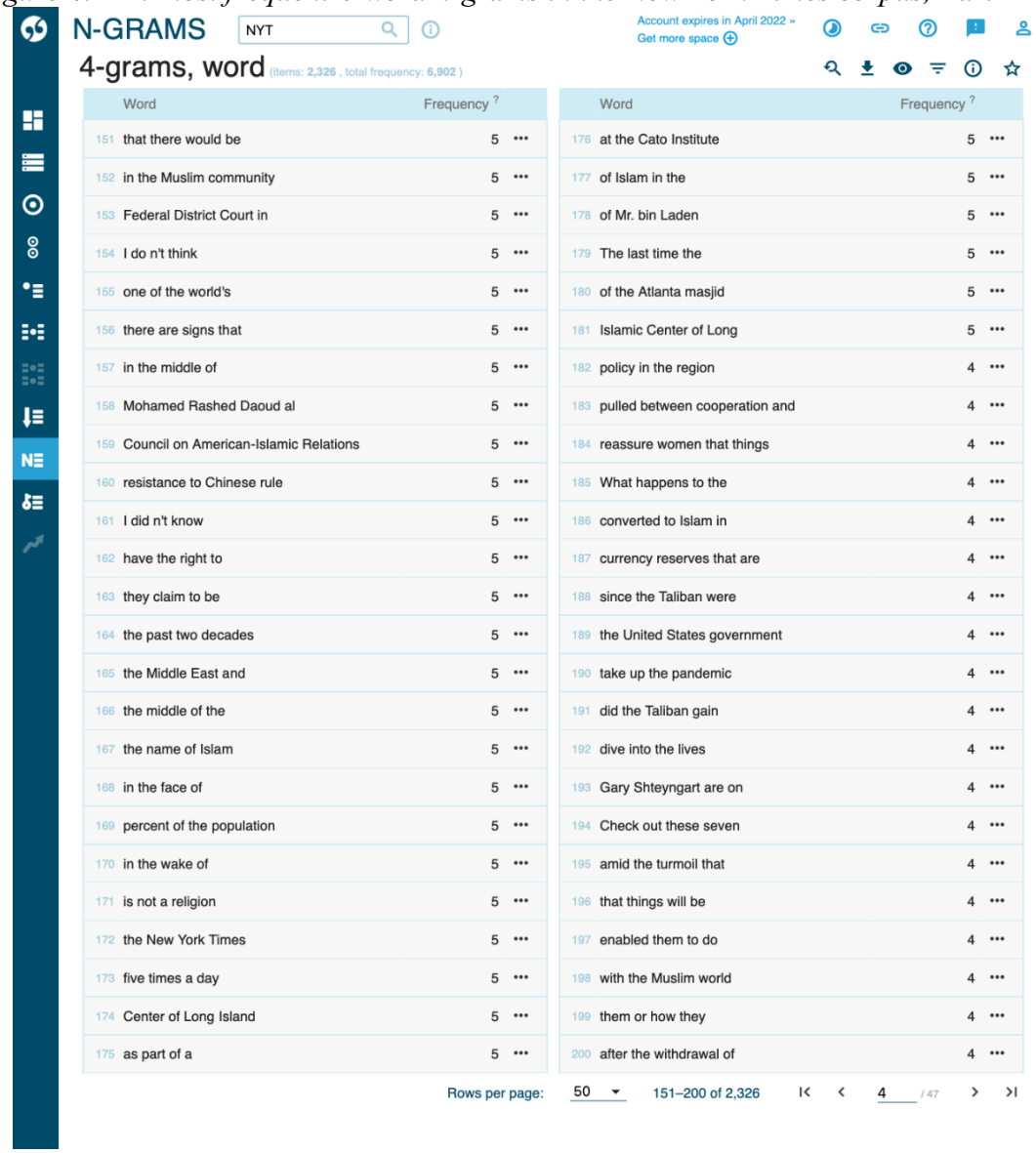
4-grams, word (items: 2,326 , total frequency: 6,902)

	Word	Frequency
101	the heart of the	6
102	committed to publishing a	6
103	diversity of letters to	6
104	hear what you think	6
105	original context from February	6
106	to hear what you	6
107	Times is committed to	6
108	for the first time	6
109	the executive director of	6
110	what you think about	6
111	Here are some tips	6
112	at the same time	6
113	the Muslim world and	6
114	this or any of	6
115	Mr. bin Laden and	6
116	you think about this	6
117	Mashari and Abdullah Bejad	6
118	the rest of the	6
119	is committed to publishing	6
120	to publishing a diversity	6
121	the center of the	6
122	was one of the	6
123	and the Middle East	6
124	the World Trade Center	6
125	and the United States	6

	Word	Frequency
126	like to hear what	6
127	a diversity of letters	6
128	professor of Islamic studies	6
129	about this or any	6
130	think about this or	6
131	Muslims in the United	6
132	letters to the editor	6
133	any of our articles	6
134	The Times is committed	6
135	of the Muslim Brotherhood	6
136	of letters to the	6
137	one of the few	6
138	or any of our	6
139	of the Nation of	6
140	in Federal District Court	5
141	Al Qaeda and the	5
142	to Mr. bin Laden	5
143	a growing number of	5
144	with Mr. bin Laden	5
145	is a senior fellow	5
146	as well as the	5
147	the American Embassy in	5
148	the words of the	5
149	the son of a	5
150	the president of the	5

Rows per page: 50 101–150 of 2,326 3 / 47

Figure 4.2-D: most frequent 4-word n-grams in the New York Times corpus, Part D



	Word	Frequency		Word	Frequency
151	that there would be	5	176	at the Cato Institute	5
152	in the Muslim community	5	177	of Islam in the	5
153	Federal District Court in	5	178	of Mr. bin Laden	5
154	I do n't think	5	179	The last time the	5
155	one of the world's	5	180	of the Atlanta masjid	5
156	there are signs that	5	181	Islamic Center of Long	5
157	in the middle of	5	182	policy in the region	4
158	Mohamed Rashed Daoud al	5	183	pulled between cooperation and	4
159	Council on American-Islamic Relations	5	184	reassure women that things	4
160	resistance to Chinese rule	5	185	What happens to the	4
161	I did n't know	5	186	converted to Islam in	4
162	have the right to	5	187	currency reserves that are	4
163	they claim to be	5	188	since the Taliban were	4
164	the past two decades	5	189	the United States government	4
165	the Middle East and	5	190	take up the pandemic	4
166	the middle of the	5	191	did the Taliban gain	4
167	the name of Islam	5	192	dive into the lives	4
168	in the face of	5	193	Gary Shteyngart are on	4
169	percent of the population	5	194	Check out these seven	4
170	in the wake of	5	195	amid the turmoil that	4
171	is not a religion	5	196	that things will be	4
172	the New York Times	5	197	enabled them to do	4
173	five times a day	5	198	with the Muslim world	4
174	Center of Long Island	5	199	them or how they	4
175	as part of a	5	200	after the withdrawal of	4

The output of the n-gram analysis comprises both sequences of words that form complete phrases (e.g. *of the Muslim*, *in the United States*, *mr. bin Laden*) and others that do not (e.g. *one of the*, *as well as*, *part of the*).

A great number of n-grams found in the corpus can be classified according to the following semantic categories: RELIGION (n= 19; *in the Muslim world, Muslims in the United, of the Muslim Brotherhood, in the Muslim community, with the Muslim world, the nation of Islam, of the Nation of, in the Islamic world, of the Prophet Muhammad, a professor of Islamic, the Islamic center of, the Muslim world and, the name of Islam, is not a religion, of Islam in the, Islamic Center of Long, converted to Islam in, of the Atlanta*

masjid[3], *five times a day*[4]), GEOGRAPHICAL LOCATIONS (n=18; *in the United States*, *the United States and*, *to the United States*, *of the United States*, *that the United States*, *by the United States*, *and the United States, the United States government, the American Embassy in*, *in the Middle East*, *and the Middle East*, *the Middle East and, at the University of*, *the World Trade Center*, *in Federal District Court*, *Federal District Court in*, *at the Cato Institute*, *Center of Long Island*), and NAMES OF INDIVIDUAL MUSLIMS AND/OR NAMES OF ISLAMIC GROUPS, often associated to discourse about terrorism (n=10; *Who are the Taliban*, *did the Taliban gain, since the Taliban were, Mr. bin Laden and*, *to Mr. bin Laden*, *with Mr. bin Laden*, *of Mr. bin Laden*, *Al Qaeda and the, Mohamed Rashed Daoud al*, *Mashari and Abdullah Bejad*).

Other n-grams found in the corpus refer to more varied subjects, each of which is, however, exemplified only once or a few times: POLITICS (*the president of the*, *resistance to Chinese rule, policy in the region*), WORK (*is a senior fellow*, *executive director of the*, *the director of the*), TIME (*the past two decades*), QUANTITY (*a growing number of*, *percent of the population*, *check out these seven*), PERSONAL RELATIONSHIPS (*the son of a*) and OTHER (*errors or other problems*, *have the right to*, *said in an interview*, *they claim to be*, *take up the pandemic*, *reassure women that things*, *dive into the lives*).

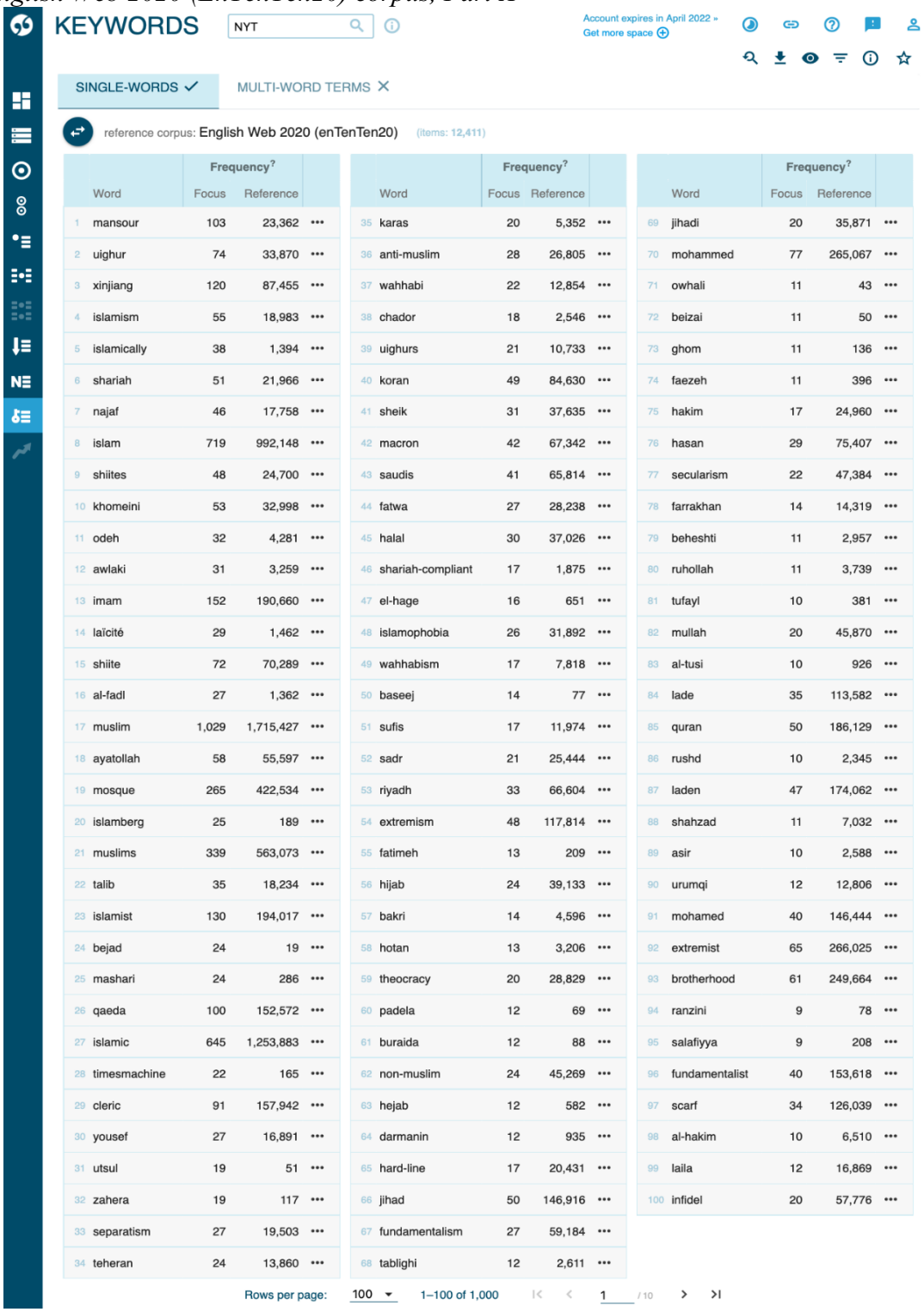
Finally, there are n-grams that exemplify the newspaper's metadiscourse, whose function is to guide the reader through the text, such as *Read in app by*, *read what you share*, *can read what you*, *Anyone can read what*, *for home delivery and*[5].

A complementary view on the lexical make-up of the *New York Times* corpus can be gained by identifying its most unusually frequent, and thus its most distinctive, words, as can be revealed through the *keyword* function. Figures 4.3-A-B show the top 200 keywords of the *New York Times* corpus (i.e. those that are much more frequent than might be expected considering the type/token ratio of the corpus), obtained by comparing it against the *English Web 2020 (enTenTen20)* corpus, used as a reference corpus[6].

---

[3] Majid: 'a mosque in an Arab country' (https://www.collinsdictionary.com/it/dizionario/inglese/masjid).
[4] This n-gram refers to the requirement of a practicing Muslim to pray five times per day.
[5] The presence of these n-grams shows that Sketch Engine only cleaned data from advertisement texts but not from words or phrases that exemplify the newspaper's metadiscourse.
[6] The keyness score of a word in Sketch Engine is calculated according to the following formula: $fpm_{focus} + n / fpm_{ref} + n$, where fpm_{focus} is the normalized (per million) frequency of the word in the focus corpus, fpm_{ref} is the normalized (per million) frequency of the word in the reference corpus, n is the simple Maths (smoothing) parameter (n = 1 is the default value).

Figure 4.3-A: top 200 keywords from the New York Times corpus compared against the English Web 2020 (EnTenTen20) corpus, Part A



	Word	Frequency Focus	Frequency Reference
1	mansour	103	23,362
2	uighur	74	33,870
3	xinjiang	120	87,455
4	islamism	55	18,983
5	islamically	38	1,394
6	shariah	51	21,966
7	najaf	46	17,758
8	islam	719	992,148
9	shiites	48	24,700
10	khomeini	53	32,998
11	odeh	32	4,281
12	awlaki	31	3,259
13	imam	152	190,660
14	laïcité	29	1,462
15	shiite	72	70,289
16	al-fadl	27	1,362
17	muslim	1,029	1,715,427
18	ayatollah	58	55,597
19	mosque	265	422,534
20	islamberg	25	189
21	muslims	339	563,073
22	talib	35	18,234
23	islamist	130	194,017
24	bejad	24	19
25	mashari	24	286
26	qaeda	100	152,572
27	islamic	645	1,253,883
28	timesmachine	22	165
29	cleric	91	157,942
30	yousef	27	16,891
31	utsul	19	51
32	zahera	19	117
33	separatism	27	19,503
34	teheran	24	13,860
35	karas	20	5,352
36	anti-muslim	28	26,805
37	wahhabi	22	12,854
38	chador	18	2,546
39	uighurs	21	10,733
40	koran	49	84,630
41	sheik	31	37,635
42	macron	42	67,342
43	saudis	41	65,814
44	fatwa	27	28,238
45	halal	30	37,026
46	shariah-compliant	17	1,875
47	el-hage	16	651
48	islamophobia	26	31,892
49	wahhabism	17	7,818
50	baseej	14	77
51	sufis	17	11,974
52	sadr	21	25,444
53	riyadh	33	66,604
54	extremism	48	117,814
55	fatimeh	13	209
56	hijab	24	39,133
57	bakri	14	4,596
58	hotan	13	3,206
59	theocracy	20	28,829
60	padela	12	69
61	buraida	12	88
62	non-muslim	24	45,269
63	hejab	12	582
64	darmanin	12	935
65	hard-line	17	20,431
66	jihad	50	146,916
67	fundamentalism	27	59,184
68	tablighi	12	2,611
69	jihadi	20	35,871
70	mohammed	77	265,067
71	owhali	11	43
72	beizai	11	50
73	ghom	11	136
74	faezeh	11	396
75	hakim	17	24,960
76	hasan	29	75,407
77	secularism	22	47,384
78	farrakhan	14	14,319
79	beheshti	11	2,957
80	ruhollah	11	3,739
81	tufayl	10	381
82	mullah	20	45,870
83	al-tusi	10	926
84	lade	35	113,582
85	quran	50	186,129
86	rushd	10	2,345
87	laden	47	174,062
88	shahzad	11	7,032
89	asir	10	2,588
90	urumqi	12	12,806
91	mohamed	40	146,444
92	extremist	65	266,025
93	brotherhood	61	249,664
94	ranzini	9	78
95	salafiyya	9	208
96	fundamentalist	40	153,618
97	scarf	34	126,039
98	al-hakim	10	6,510
99	laila	12	16,869
100	infidel	20	57,776

Figure 4.3-B: top 200 keywords from the New York Times corpus compared against the English Web 2020 (EnTenTen20) corpus, Part B

KEYWORDS

NYT

Account expires in April 2022 »
Get more space

SINGLE-WORDS ✓ MULTI-WORD TERMS ✕

reference corpus: English Web 2020 (enTenTen20) (items: 12,411)

	Word	Frequency? Focus	Reference
101	sanya	11	12,721
102	sunni	40	162,837
103	abdullah	38	152,825
104	crackdown	31	118,308
105	sufism	12	20,973
106	non-muslims	13	26,939
107	sufi	19	59,934
108	ramadan	23	82,063
109	anticep	8	0
110	suweidi	8	59
111	montaseri	8	75
112	juhayman	8	262
113	iranian	137	712,017
114	muslim-majority	10	11,402
115	khatami	10	11,641
116	salafi	10	12,246
117	jamaat	10	12,389
118	headscarf	10	12,883
119	nabila	8	2,057
120	blasphemy	21	77,574
121	taliban	64	328,160
122	uffizi	9	9,578
123	african-americans	19	69,077
124	saudi	146	829,571
125	saud	12	28,368
126	mecca	28	125,988
127	eamon	10	17,452
128	reformist	17	60,905
129	muhammad	78	442,558
130	maleky	7	3
131	esada	7	54
132	dastjerdi	7	204
133	khameni	7	219
134	counterterrorism	15	50,494
135	abdus-salaam	7	291
136	koranic	8	6,577
137	radicalism	13	38,987
138	thabet	7	1,135
139	akyol	7	1,170
140	khalfan	7	1,222
141	ningxia	8	8,150
142	yarkand	7	1,751
143	córdoba	10	21,384
144	salam	11	28,011
145	pipes	13	41,743
146	radicalization	10	22,513
147	hamad	10	22,746
148	andalusian	9	17,433
149	andalusia	10	24,238
150	umayyad	9	17,892
151	westernize	8	12,112
152	joko	8	12,562
153	belabbas	6	14
154	muppie	6	16
155	eldina	6	17
156	muktada	6	34
157	al-hesbah	6	44
158	kherchtou	6	77
159	hindu	83	560,092
160	kerbela	6	194
161	akrid	6	257
162	hebdo	8	14,896
163	ivry-sur-seine	6	335
164	mughniyah	6	527
165	ibn	41	257,194
166	militant	67	449,413
167	hayy	6	772
168	el-amin	6	1,076
169	xi	48	313,772
170	anti-western	7	8,697
171	polytheist	7	8,719
172	mila	9	23,652
173	ressam	6	1,429
174	bakr	10	31,492
175	osama	21	114,091
176	fahd	7	9,076
177	franzen	7	9,739
178	shah	46	307,145
179	hijacker	12	48,365
180	samir	9	25,442
181	al-sistani	6	2,669
182	baugh	7	10,364
183	baqir	6	2,863
184	worshiper	10	33,717
185	indoctrination	11	42,247
186	al-amin	6	3,466
187	paty	6	3,505
188	congregant	8	19,583
189	ahmed	36	240,722
190	arabia	78	573,968
191	scarves	13	59,655
192	far-right	14	68,051
193	puritanical	7	12,639
194	internment	12	52,633
195	iran	230	1,824,386
196	anwar	11	46,032
197	khalid	13	62,949
198	tyagi	6	5,989
199	fervor	11	47,990
200	sistani	6	7,380

Rows per page: 100 101–200 of 1,000 2 / 10

The top 200 keywords include PERSONAL NAMES TYPICAL OF MIDDLE-EASTERN AND NORTH AFRICAN COUNTRIES (e.g. *mansour*, *odeh*, *khomeini*,

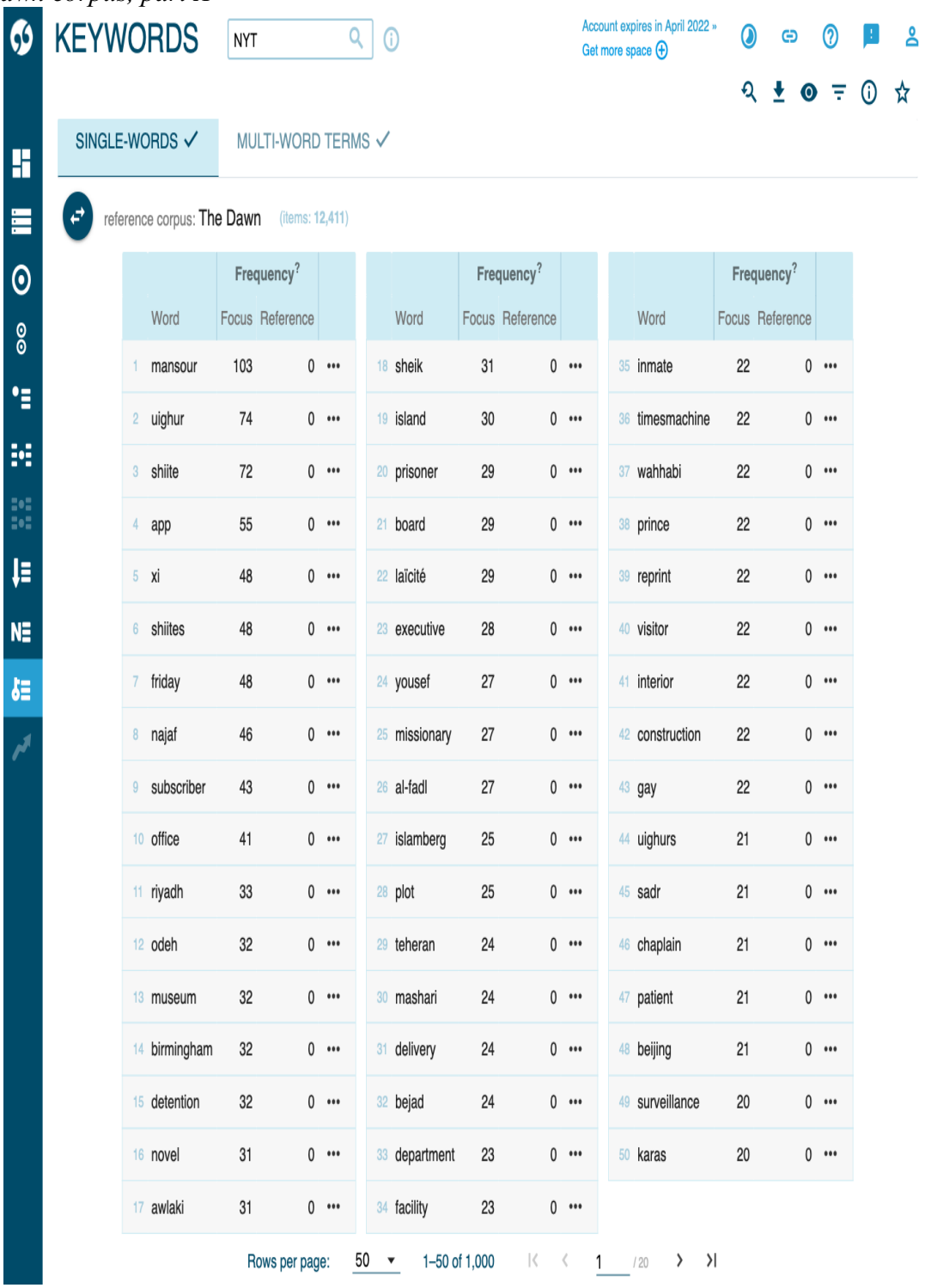
awlaki, *talib*, *bejad*, *mashari*, *qaeda*, *yousef*, *zahera*, *karas*, *el-hage*), terms which refer to ISLAMIC RELIGION AND THE ISLAMIC WORLD IN GENERAL (e.g. *islamism*, *islamically*, *shariah*, *islam*, *shiite(s)*, *imam*, *muslim*), POLITICS (e.g. *westernize*, *anti-western*, *macron*[7], *hard-line*, *reformist*, *militant*, *far-right*), LAW AND ORDER (e.g. *theocracy*, *counterterrorism*, *crackdown*, *hijacker*, *internment*), GEOGRAPHICAL LOCATIONS (e.g. *Córdoba*, *andalusia*, *ivry-sur-seine*, *arabia*, *iran*), ETHNIC GROUPS (e.g. *Uighur(s)*, *al-fadl*, *utsul*, *saudis*, *hindu*, *andalusian*, *Iranian*) and OTHER (e.g. *lacit*, *mashari*, *padela*, *sadr*, *riyadh*, *fatimeh*, *hijab*, *bakri*, *hotan*, *buraida*)[8].

Figure 4.27-A-B show the top 200 keywords of the *New York Times* corpus obtained by comparing it against *The Dawn* corpus, used as a reference corpus.

[7] This is the name of the current President of France, elected in 2017.

[8] For comparative purposes, I also had the top 200 keywords of the corpus automatically classified by the UCREL semantic tagger (http://ucrel-api.lancaster.ac.uk/usas/tagger.html). This identified the following semantic categories: RELIGION (*islam*, *imam*, *muslim*, *ayatollah*, *mosque*, *muslims*, *islamic*, *cleric*, *anti-muslim*, *fundamentalism*, *mullah*, *fundamentalist*, *sunni*, *ramadan*, *blasphemy*, *hindu*, *congregant*), POLITICS (*separatism*, *extremism*, *extremist*, *radicalism*, *far-right*), GEOGRAPHICAL NAMES (*teheran*, *Saudis*, *Iranian*, *saudi*, *blasphemy*, *hindu*, *congregant*), PEOPLE (*sheik*, *african-americans*, *shah*, *arabia*), GROUPS AND AFFILIATIONS (*brotherhood*, *Taliban*), CONSTRAINT (*hard-line*, *crackdown*, *internment*), PERSONAL NAMES (*mohammed*, *abdullah*, *salam*, *fahd*, *khalid*, *ahmed*), QUANTITIES (*laden*), CLOTHES AND PERSONAL BELONGINGS (*scarf*, *headscarf*, *scarves*), OBJECTS (*pipes*), CRIME, LAW AND ORDER (*militant*, *hijacker*, *puritanical*) and EDUCATION (*indoctrination*). A few words were incorrectly classified (i.e. *blasphemy*, *congregant* and *laden*, the last one being part of a name). In addition, the classification was not complete, in the sense that the software could not assign a semantic category to the following words, which were therefore tagged as UNMATCHED: *mansour*, *uighur*, *uighurs*, *xinjiang*, *andalusian*, *islamism*, *islamically*, *islamist*, *islamberg*, *islamophobia*, *koran*, *koranic*, *quran*, *halal*, *infidel*, *sufi*, *sufis*, *sufism*, *shariah*, *shariah-complaint*, *jihad*, *jihadi*, *non-muslim*, *non-muslims*, *muslim-majority*, *polytheist*, *worshiper*, *theocracy*, *reformist*, *secularism*, *radicalization*, *counterterrorism*, *westernize*, *anti-western*, *shiites*, *shiite*, *khomeini*, *odeh*, *awlaki*, *lacit*, *al-fadl*, *el-hage*, *osama*, *talib*, *mohamed*, *Muhammad*, *bejad*, *mashari*, *qaeda*, *yousef*, *padela*, *macron*, *timesmachine*, *utsul*, *zahera*, *karas*, *wahhabi*, *wahhabism*, *chador*, *fatwa*, *baseej*, *sadr*, *riyadh*, *fatimeh*, *hijab*, *bakri*, *hotan*, *buraida*, *hejab*, *darmain*, *tablighi*, *owhali*, *beizai*, *ghom*, *faezeh*, *hakim*, *hasan*, *farrakhan*, *beheshti*, *ruhollah*, *tufayl*, *al-tusi*, *lade*, *rushd*, *shahzad*, *asid*, *urumqi*, *ranzini*, *salafiyya*, *al-hakim*, *laila*, *sanya*, *anticep*, *suweidi*, *montaseri*, *juhayman*, *khatami*, *salafi*, *jamaat*, *nabila*, *uffizi*, *saud*, *eamon*, *maleky*, *esada*, *dastjerdi*, *khameni*, *abdus-salaam*, *thabet*, *akyol*, *khalfan*, *ningxia*, *yarkand*, *crdoba*, *hamad*, *umayyad*, *joko*, *belabbas*, *muppie*, *eldina*, *muktada*, *al-hesbah*, *kherchtou*, *kerbela*, *akrid*, *hebdo*, *ivry-sur-seine*, *mughniyah*, *ibn*, *hayy*, *el-amin*, *mila*, *ressam*, *bakr*, *franzen*, *samir*, *al-sistani*, *baugh*, *baqir*, *al-amin*, *paty*, *anwar*, *tyagi*, *fervor*, *najaf*, *sistani*.

Figure 4.4-A: Top 200 keywords from the New York Times corpus compared against The Dawn corpus, part A



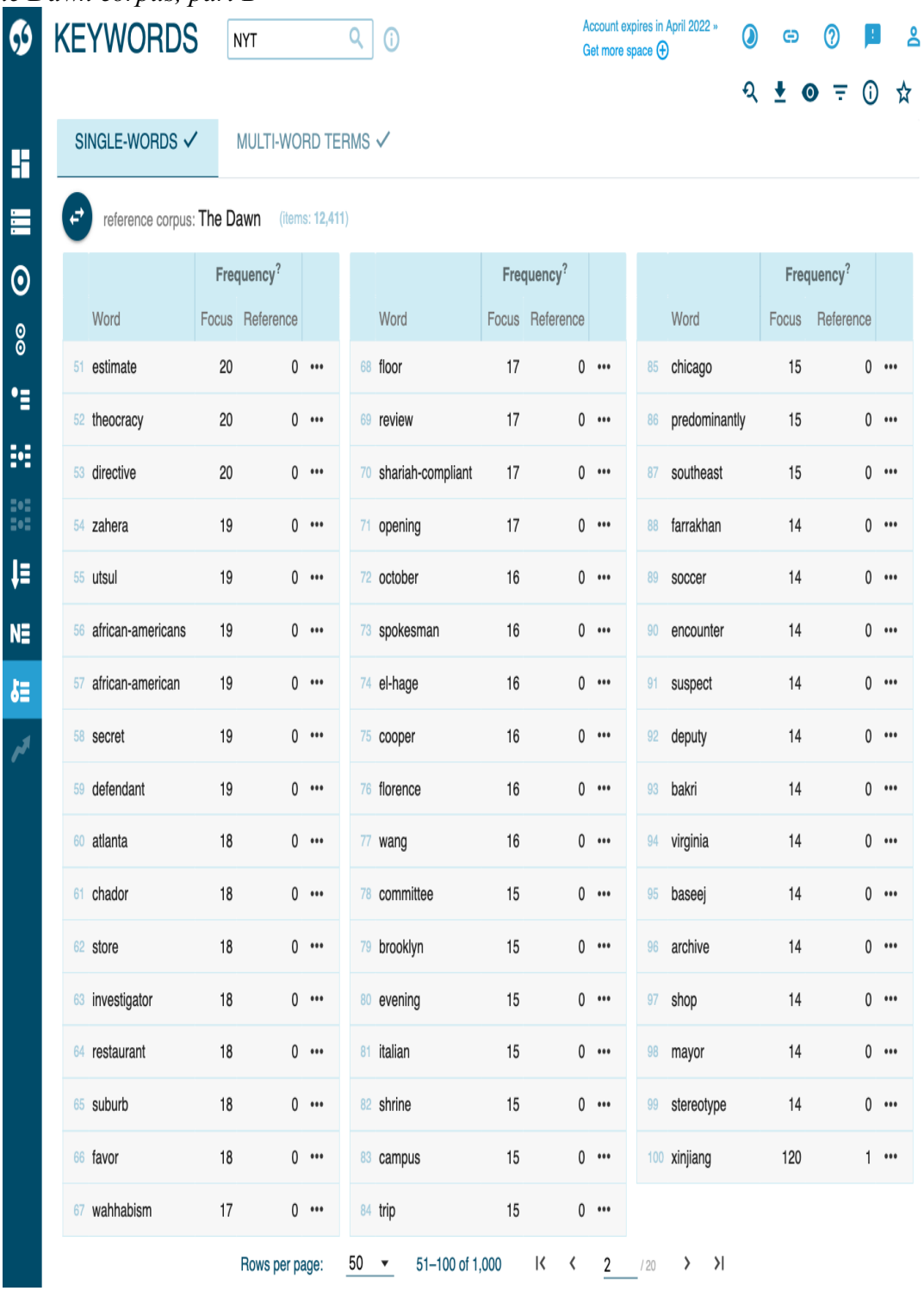
	Word	Frequency? Focus	Reference	
1	mansour	103	0	•••
2	uighur	74	0	•••
3	shiite	72	0	•••
4	app	55	0	•••
5	xi	48	0	•••
6	shiites	48	0	•••
7	friday	48	0	•••
8	najaf	46	0	•••
9	subscriber	43	0	•••
10	office	41	0	•••
11	riyadh	33	0	•••
12	odeh	32	0	•••
13	museum	32	0	•••
14	birmingham	32	0	•••
15	detention	32	0	•••
16	novel	31	0	•••
17	awlaki	31	0	•••

	Word	Frequency? Focus	Reference	
18	sheik	31	0	•••
19	island	30	0	•••
20	prisoner	29	0	•••
21	board	29	0	•••
22	laïcité	29	0	•••
23	executive	28	0	•••
24	yousef	27	0	•••
25	missionary	27	0	•••
26	al-fadl	27	0	•••
27	islamberg	25	0	•••
28	plot	25	0	•••
29	teheran	24	0	•••
30	mashari	24	0	•••
31	delivery	24	0	•••
32	bejad	24	0	•••
33	department	23	0	•••
34	facility	23	0	•••

	Word	Frequency? Focus	Reference	
35	inmate	22	0	•••
36	timesmachine	22	0	•••
37	wahhabi	22	0	•••
38	prince	22	0	•••
39	reprint	22	0	•••
40	visitor	22	0	•••
41	interior	22	0	•••
42	construction	22	0	•••
43	gay	22	0	•••
44	uighurs	21	0	•••
45	sadr	21	0	•••
46	chaplain	21	0	•••
47	patient	21	0	•••
48	beijing	21	0	•••
49	surveillance	20	0	•••
50	karas	20	0	•••

Rows per page: 50 1–50 of 1,000 1 / 20

Figure 4.4-B: Top 200 keywords from the New York Times corpus compared against The Dawn corpus, part B



	Word	Frequency Focus	Frequency Reference
51	estimate	20	0
52	theocracy	20	0
53	directive	20	0
54	zahera	19	0
55	utsul	19	0
56	african-americans	19	0
57	african-american	19	0
58	secret	19	0
59	defendant	19	0
60	atlanta	18	0
61	chador	18	0
62	store	18	0
63	investigator	18	0
64	restaurant	18	0
65	suburb	18	0
66	favor	18	0
67	wahhabism	17	0
68	floor	17	0
69	review	17	0
70	shariah-compliant	17	0
71	opening	17	0
72	october	16	0
73	spokesman	16	0
74	el-hage	16	0
75	cooper	16	0
76	florence	16	0
77	wang	16	0
78	committee	15	0
79	brooklyn	15	0
80	evening	15	0
81	italian	15	0
82	shrine	15	0
83	campus	15	0
84	trip	15	0
85	chicago	15	0
86	predominantly	15	0
87	southeast	15	0
88	farrakhan	14	0
89	soccer	14	0
90	encounter	14	0
91	suspect	14	0
92	deputy	14	0
93	bakri	14	0
94	virginia	14	0
95	baseej	14	0
96	archive	14	0
97	shop	14	0
98	mayor	14	0
99	stereotype	14	0
100	xinjiang	120	1

Most of the keywords shown in Figures 4.4-A-B can be assigned to the same semantic groups that were previously identified by comparing the *New York Times* corpus against *English Web 2020 (enTenTen20)* corpus, namely PERSONAL NAMES

TYPICAL OF MIDDLE-EASTERN AND NORTH AFRICAN COUNTRIES (*Mansour*, *Odeh*, *Awlaki*, *Yousef*, *Al-Fadl*, *mashari*, *bejad*, *sadr*, *Beijing*, *karas*, *zahera*, *el-hage*, *cooper*, *wang*, *Farrakhan*, *bakri*, *baseej*), terms which refer to the ISLAMIC RELIGION AND THE ISLAMIC WORLD IN GENERAL (*Shiite*, *Shiites*, *laïcité*, *missionary*, *islamberg*, *Wahhabi*, *Wahhabism*, *chaplain*, *shariah-complaint*, *shrine, chador*), POLITICS (*sheik*, *deputy*, *Xi*[9]), LAW AND ORDER (*theocracy*, *detention*, *surveillance*, *prisoner*, *inmate*, *defendant*, *investigator*), GEOGRAPHICAL LOCATIONS (*Najaf*, *Riyadh*, *Birmingham*, *island*, *Teheran*, *atlanta*, *suburb*, *Florence*, *Brooklyn*, *Chicago*, *southeast*, *virginia*, *xinjiang*), ETHNIC GROUPS (*Uighur*, *uighurs*, *utsul*, *african-americans*, *african-american*, *Italian*) and OTHER (*plot*, *interior*, *estimate*, *directive*, *secret*, *favor*, *review*, *opening*, *October*, *evening*, *trip*, *predominantly*, *soccer*, *encounter*, *stereotype*). But, in addition I assigned some words to new semantic fields (e.g. not previously identified), namely PEOPLE (*executive*, *prince*, *visitor*, *gay*, *patient*, *spokesman*, *committee*), PLACES (*department*, *facility*, *construction*, *office*, *store*, *floor*, *shop, campus*) and CULTURE (*museum*, *novel*, *board*[10]). Finally, as in the n-grams analysis, some words that exemplify the newspaper's metadiscourse, whose function is to guide the reader through the text, were identified (*app*, *Friday*, *subscriber*, *archive*, *timesmachine*, *reprint*, *delivery*).

Next, I analysed the collocates of the five most frequent lexical words (*muslim*, *islam*, *islamic*, *religious*, *american*) and the top five keywords obtained by comparing the *New York Times* corpus against the *English Web2020 enTenTen20* corpus (*uighur*, *xinjiang*, *islamism*, *islamically, shariah*) through the *Concordance* and *Word Sketch* functions.

In order to keep the analysis of concordances to a manageable size, I considered a random sample of only 200 concordance lines for each of the above-mentioned words. When I retrieved the concordances, I sorted them alphabetically by one, two and three tokens to the left and one, two and three tokens to the right of the searched word.

Figure 4.4-A shows a sample of 50 random concordance lines of the word *Muslim*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

[9] This is the name of the current President of China (Xi Jinping), elected in 2013.
[10] As in '*Boarding schools*'.

Figure 4.4-A: 50 sample concordance lines of the word "Muslim", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) *Muslim* is mostly used as a premodifier in noun phrases (e.g. *Black Muslims leaders denied any connection with the slayings*, *scholars have noted Muslim names*, *no Muslim societies offer their people rule of law*);

2) there is a frequent use of third person pronouns and possessives to the left of the word *Muslim* (*her position as a Muslim girl*, *they come from a Muslim country*, *herself a Muslim*, *their imagined Muslim takeover*, n=2: *many of them Muslims*, *they are neither Brothers nor are they Muslims*). On the other hand, the use of first person pronouns and/or other determiners is limited to those cases in which direct speech is reported (*I am a Muslim*, *Our Muslim minister*);

3) a distinction can be made between (groups of) words to the left of *Muslim* which carry NEGATIVE SEMANTIC PROSODY (*concern about*, *The mistrust between*, *despotic*, *starker problems in*, *the malaise of*, *oppression*, *Crackdown on*, *the problems of poor*, *ban several*, *silencing*, *disappointment with the*, *fear of us*, *we received verbal abuse from American Muslim leaders*) which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*engaging*, *noteworthy*, *a successful practicing*, *Respecting*, *enable the*, *catering to the*, *integrated*, *integrating*, *reassures*, *win for the*, *unite*, *a group organized by a Muslim civil rights organization*) which are less frequent;

4) the (groups of) words to the left of *Muslim* can be classified into the following semantic groups: QUANTITY/AMOUNT (n=2: *1.3 billion*, *200 billion*, *6 million*, *6.1 percent*, *one*, n=2: *two*, *three*, *the only*, n=2: *every*, n=4: *largest*, n=7: *many*, n=2: *most*, n=4: *predominantly*, n=2: *several*, n=3: *some*, *vast*, *in much of the*, *much of the*, *most parts of the*, *the rest of the*, *for the whole*, *the number of*), RELIGION (n=2: *conservative*, *faithful*, *fundamentalist*, *militant*, n=3: *observant*, *pious*, *practicing*, *Koran*, *secularism*), GEOGRAPHICAL ORIGIN (*Arab, Egyptian*, n=2: *French*, *Turks and other*, *local*, *American*, *India's 200 million*), PEOPLE (*young*, *no Muslim societies*, n=3: *immigrant*), POLITICS (*authoritarianism across the*, *minister to*, *banning forms of*), ETHNIC GROUPS (n=2: *black American*, n=2: *Black*), EVALUATION (*mainstream*, *good*, *true*), and OTHER (*rationalist*, *magazine*, *director*, *fixture*, *hallmark*, *life*, *handing out*, *designate*, *worldwide call to*, *thanks to*, *traditional, shot two*).

Figure 4.4-B shows a sample of 50 random concordance lines of the word *Muslim*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.4-B: 50 sample concordance lines of the word "Muslim", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) t
he word *Muslim* is occasionally (n=9) followed by a quotative expression like *he/she/someone said*, which serves to report direct speech;

2) some words and phrases to the right of *Muslim* carry NEGATIVE SEMANTIC PROSODY (*hostages*, *monolithic*, *the Muslim Brotherhood engages in terrorist activity*), while others carry POSITIVE SEMANTIC PROSODY (*being Muslim can be an asset*, *immigrant Muslims have been brotherly toward us*), the latter being less frequent;

3) the (groups of) words to the right of *Muslim* can be classified into the following semantic fields: GEOGRAPHICAL LOCATIONS (n=5: *in the United States*, *in North America*, *in Atlanta*, *areas*, *associations across Italy*, *in Egypt*, *across the world*, n=6: *country*, *countries of Southeast Asia (Indonesia and Malaysia)*, *country like Pakistan*, *country like Syria*, *Afghanistan or Yemen*, *in the Xinjiang region*, *Middle East*, *nations*, *to travel to Afghanistan*, n=17: *world*), RELIGION (*chaplain*, *and scholars of Islam*, *believes such apocalyptic prophecies*, *call to prayer*, *deviated from God's plan*, *from Jew or Christian*, *clerics*, *doctrine*, *faith*, *names and Islamic*, n=4: *leaders*, *missionaries*, *prayers*, *preachers*, *religious practice*, *to invoke Sharia*), SOCIAL GROUPS (n=2: *American groups*, n=2: *Americans*, *readers*, n=2: *patients*, *Users*, *gun owners*, n=6: *community*, n=2: *group*, *societies*, *opinion leaders*, n=3: *immigrants*, *migrants*, *astronomers*, *customers*, *on the job*, *professionals*, *traders*), PEOPLE (*residents*, n=8: *woman*, *teenagers*, *Children*, n=2: *girl*, *man, person*, n=2: *population*), INSTITUTIONS (*American Society for Muslim Advancement*, *Muslim Alliance*, *Muslim associations*, n=10: *Muslim Brotherhood*, n=2: *civil rights organization*, *center*), EDUCATION (*schools*, *students*, n=2: *academic*, *and a professor*, *at American colleges*, *friends at school*), CLOTHES AND PERSONAL BELONGINGS (*head covering*, *traditional religious garb*, *veil*, *carpets*, *veiling*, *scarf*), QUANTITY (*one of the largest*, *some of whom*), POLITICS (*congresswomen*, n=3: *minister*, *endorsed George W. Bush*[11]), POWER (*rulers*, *takeover of America*, *prison officials*), SCIENCE (*science*, *to science*, *Islamic science*), TERRORISM (*Terrorism*, *Muslim Brotherhood terrorist*), CULTURE

[11] He was President of the United States from 2001 to 2009.

(*culture*, *customs*, *literature*) and OTHER (*discourse*, n=2: *empires*, *fates*, n=2: *neighborhoods*, *traders are barred*, *call*, *endorsed*, *say*).

Figure 4.5-A shows a sample of 50 random concordance lines of the word *Islam*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.5-A: 50 sample concordance lines of the word "Islam", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islam* is a noun and is mostly used in prepositional phrases (e.g. *On Islam Trump takes a different approach*, *of Islam*, *between Islam and democracy*, *around Islam*);

2) a distinction can be made between (groups of) words to the left of *Islam* which carry NEGATIVE SEMANTIC PROSODY (*derogatory comments about*, *strong doubts about Islam's compatibility with the country's values*, *moral panic about*, *insults against*, *singled out Muslims and Islam as a civilizational threat*, *fear-based narrative around*, *conflict between Islam and democracy*, *criticized*, *criticizing*, *defaming*, *a vile vulgarity to describe*, *destroy*, *to eliminate*, *impose*, *the problem of Islamism*, *an irrational fear of, a general mistrust of,* n=2: *dark view of*, *hateful ideology of*, *hostile of*, *animosity toward*, *fanatic*, *ferocious*, *dangerous strain within*) which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*to nurture an "Islam of France"*, *"humanist"*, *serene practice of*, *in support of*, *business-friendly*, *a struggle to defend*, *Reinforcing and Improving*, *Liberal*, *modernized*, *"insults" of*, *the integration of*, *a peaceful version of*, *respect*) which are less frequent;

3) the (groups of) words to the right of *Islam* can be divided into the following semantic categories: RELIGION (n=3: *Muslims and*, n=2: *the Nation of Islam*, *conversion*, *converts*, n=3: *converted to*, *converting to*, *devotion to*, *blasphemers*, *The Apostates*, *Liberal*, n=2: *moderate*, *the sacred shrines*, *holy book of*, *Shiite branch of*, *mystical form of*, *scholars of*, *fundamentalist strain of*, *religious validity of*, n=4: *orthodox*, n=5: *radical*, *Christians and Jews*, *Sunni and Shia*, *Shiite, Sufi*, n=2: *Sunni*, *conservative Wahhabi*, *Muhammad*, *the prophet Muhammad*, *a sin against*), COMMUNICATION (*called*, *criticized*, *criticizing*, *declared*, n=2: *argued that*, *assertion that*, *contradict*, *confirm that*, *deny that*, *do not define*), POLITICS (n=4: *Communism and*, *figurehead*, n=2: *militant*, n=2: *political*, *public attitudes toward*), GEOGRAPHICAL ORIGIN (*Arabian Peninsula*, *American*, *Indonesian*, n=2: *Sarekat*), PERIOD (*in the seventh century*, *contemporary*, *medieval*), WAR (*an attack on*, *reform*, *at war with*, *soldier of*), PEOPLE (*women*, *population*, *senior fellow on*), TERRORISM (*ISIL*, n=2: *Fatah al Islam*, *soldier of*), ETHNIC GROUPS (*the Bolsheviks of*, *Italians*, *African-Americans*), SEPARATION (*abandoning*, *Leave/Leaving*, *to renounce*, *to separate*), REFLECTION (*Discussion of*, n=2: *interpretation of*, *the treatment of*, *understanding of*, *as proof that*,

misapprehensions connected to) and OTHER (*Traditionally*, *traditional*, *radiating*, n=2: *brand of*, *enemies of*, *the evolution of*, n=3: *in the name of*, *the picture of*, *resonant pieces of*, *the uncontested place of*, *the prescriptions of*, *the role of*, *"Sinicization of Islam"*, *source of*, *The story of*, *did not ban*).

Figure 4.5-B shows a sample of 50 random concordance lines of the word *Islam*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.5-B: 50 sample concordance lines of the word "Islam", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words which carry NEGATIVE SEMANTIC PROSODY (*was stabbed to death*, *and the terrorists*, *as a civilizational threat*, *can take revenge*, *conversion from Islam is punishable*, *ruined the lives*, *is anti-American*) which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*a great favor*, *has a peaceful effect*, *has encouraged science*, *in a brighter light*, *is promoted*) which are less frequent;

2) the (groups of) words to the right of *Islam* can be divided into the following semantic categories: RELIGION (*Muslim*, *Muslims*, *or Muslims*, *the black Muslims*, *the religion*, *as a religion*, *as the religion*, n=2: *is a religion*, n=2: *is not a religion*, n=2: *the Quran*, *holy book*, *apostates, Abrahamic past*), GEOGRAPHICAL LOCATION (*in New York City*, *and Harlem*, *India*, *Indonesia*, n=2: *in China*, *in America*, *in early America*, *Middle-East heartlands*, *in France*, n=3: *of France*, *in prison*), POLITICS (*government*, *Western governments*, *the Department of Justice*, *Trump*[12], *and democracy*, *taking over America*) and OTHER (*as armor or emancipator*, *ISIL*[13]).

Figure 4.6-A shows a sample of 50 random concordance lines of the word *Islamic*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

[12] He was President of the United States from 2017 to 2021.
[13] '*ISIL'* stands for 'Islamic State of Iraq and the Levant'.

Figure 4.6-A: 50 sample concordance lines of the word "Islamic", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islamic* is exclusively used as a premodifier and occurs in noun phrases (e.g. *Communists and Islamic leaders*, *Arab Islamic banks*, *Islamic Republic*, *Asian Islamic finance*) and prepositional phrases (e.g. *by Islamic financial reformists*, *in Islamic popular culture*, *for Islamic banks*, *of Islamic studies*, *within Islamic law*, *with Islamic scholarship*);

2) the word *Islamic* occurs with second and third person pronouns and possessives (n=2: *Their*, *them*, *your*), but not with first person pronouns and possessives. As in the case of the word *Muslim*, this could point to the phenomenon of "Othering", that is, "treating people from another group as essentially different from and generally inferior to the group you belong to" (MacMillan Dictionary on line);

3) a distinction can be made between the (groups of) words to the left of *Islamic* which carry NEGATIVE SEMANTIC PROSODY (*are blaming an*, *Despotic*, *immobile*, *obscure*, *assault of*, *destroy the*, *fight for the, victims of the, failure of the, gap between the Islamic world and the West*) and those which carry POSITIVE SEMANTIC PROSODY (*growing specialty of*, *perfect*, *support for the*), the latter being less frequent;

4) the word *Islamic* tends to occur in binomial expressions, underlined in the following excerpts, with terms belonging to either the semantic field of religion or that of politics (*It was undermined, she argues, by fundamentalism – Catholic and Islamic alike*; *In Indonesia, tensions between Communists and Islamic leaders had already begun to divide Sarekat Islam in the early 1920s*);

5) the (groups of) words to the left of the word *Islamic* can be divided into the following semantic categories: RELIGION (*Catholic and*, *conservatives*, *caliph of*, *Sharia*, *Umma*, *orthodox*, *radical*, *Ramadan*, *imam*), POLITICS (*Communists and*, *governments with*, *immigrants*, *proposed*), TIME (*After Iran's 1979*, *The 19th century*), EDUCATION (*Education in an*, n=2: *professor of*), GROUPS (*Indonesians*, *The Utsuls'*, *Arab*, *Asian*, *Al Qaeda*), COMMUNICATION (*Argues for*) and OTHER (*Fervent*, *fervor of the*, *immobile*, *players*, *interpreting*, *modern*, *most*, n=2: *new*, n=2: *head of*, *popular*, *private*, *proper*, *really*, *reinterpret*, *renowned*, *legitimacy of the*, *tenets of the*, n=2: *Traditional*, *Ptolemy's flaws*, n=3: *Classical*, *dons*, *militant*).

Figure 4.6-B shows a sample of 50 concordance lines of the word *Islamic*, in the *New York Times*, sorted alphabetically by one, two and three tokens to the right.

Figure 4.6-B: 50 sample concordance lines of the word "Islamic", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) the word *Islamic* is found in binomial expressions occurring with words referring to religion or culture (e.g. *Catholic and Islamic alike*, *Islamic and European civilization have often been not just compatible*);

2) a distinction can be made between the (groups of) words to the right of *Islamic* which carry NEGATIVE SEMANTIC PROSODY (*people killed*, *terrorists*), and those which carry POSITIVE SEMANTIC PROSODY (*awakening*), the latter being less frequent;

3) the (groups of) words to the right of *Islamic* can be divided into the following semantic categories: FINANCE (n=7: *bank*, *Banker*, n=4: *banking*, *economy*, *equity investments*, *investment banking unit*, *investment side*, n=10: *finance*, n=4: *financial*, *financiers*, *financing*, *fund*, *funds*, *market Empire*), GEOGRAPHICAL LOCATION (*center in Manhattan*, *Center of Long Island*, *Institute of Orange County*, *Republic of Iran*, *revolution in Iraq*, *school in Jamaica*, n=2: *Spain*, *State in Syria and Iraq*, *in Syria*), RELIGION (*blessing*, n=2: *extremists*, *reformists*, n=2: *fundamentalism*, *radicalism*, n=2: *guidance*, *Jesus*, n=12: *law*, n=2: *rule*, *mysticism*, *or Islamist*, *principles*, *religion*, *theological writing*, *theology*, n=2: *resurgence*), POLITICS (*political thought*, n=8: *republic*, *republics*, *insurgency*[14], n=2: *revolution*, *State militants*, *Nation*, n=15: *State*, *Activists*, *terrorists*), EDUCATION (*higher education*, n=3: *school*, *student group*, n=3: *studies*, n=3: *teachings*, *University*, n=4: *Scholar*, n=2: *scholarship*), PEOPLE (n=2: *Leaders*, *man*, *people killed*, *Consumers*, *Astronomers*), CULTURE (*cultural*, n=2: *culture*, *popular culture*) and OTHER (*court*, n=2: *jurisprudence*, n=2: *History*, n=2: *Science*, n=3: *Identity*, n=7: *World*, *Arm*, *Art collections*, n=2: *bonds*, n=5: *dress*, n=2: *awakening*).

Figure 4.7-A shows a sample of 50 concordance lines of the word *religious*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

[14] The complete sentence is '*In Libya, if Moammar Gadhafi falls it will probably be an obscure Islamic insurgency, in alliance with disaffected soldiers, that brings it about'.*

Figure 4.7-A: 50 sample concordance lines of the word "religious", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *religious* is exclusively used as a premodifier in noun phrases (e.g. *a religious appeal*, *some religious women*, *the religious debate*, *a religious dispute*);

2) the word *religious* mostly occurs with third person pronouns and possessives (*Their own*, n=2: *other*, *the party's*, *consider themselves*, *her*, *his*, *one*), while it occurs only once with a first person possessive (*"our"*). This could point to the phenomenon of "othering";

3) a distinction can be made between the (groups of) words to the left of *religious* which carry NEGATIVE SEMANTIC PROSODY (*Accused of*, *fears about*, *fight against*, *oppression against*, *struggle among*, *terrorism and*, *infected by*, *eliminate*, *concern for*, *murdered his*, *illegal*, *"symptoms" of*, *captured by*, *curtailing*, *murdered his*, *illegal*, *idiocy of*, *dangerous rhetoric of*, *toxicity of*, *victims of*, *an attack on*, *raided by the*, *erode the*, *problems related to*, *curbing violent*, *modernity collides with*) and those which carry POSITIVE SEMANTIC PROSODY (*preserve her*, *protect*, *defended*, *respect the*, *confident of their*, *freedoms of their*), the latter being less frequent;

4) the (groups of) words to the left of *religious* can be divided into the following semantic categories: RELIGION (*preachers and*, *extremist*, *Islamic*, n=4: *Muslim*, *Muslims the*, *fatwa*, *puritanical*, *Shiite*, *Islam is the*, *Shiites is their*), POLITICS (n=2: *political*, *autocrats*, *federal*, *terrorism and*), GEOGRAPHICAL ORIGIN or LOCATION (*Iranian-style*, n=2: *local*, *Saudi*, *Najaf's*, *Yemen's*), PEOPLE (*lower-class*, *Muslim soldiers'*) and OTHER (*captured by*, *confronted by*, *motivated by*, *Coronavirus Fans*, *new*, *individual*, *issuing*, n=2: *other*, n=2: *freedom of*, n=2: *prominent*, n=2: *Highest*, *highly*, *a professor of*).

Figure 4.7-B shows a sample of 50 concordance lines of the word *religious*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.7-B: sample of 50 concordance lines of the word "religious", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the words to the right of *religious* which carry NEGATIVE SEMANTIC PROSODY (n=3: *adversary*, *concerns*, *crisis*, *dictatorship*, n=2: *discrimination*, *dispute*, *hatred*, *obscurantism*) and those which carry POSITIVE SEMANTIC PROSODY (*values*, *support*), the latter being less frequent;

2) the (groups of) words to the right of *religious* can be grouped into the following semantic categories: RELIGION (n=5: *beliefs*, *devotion*, n=8: *extremism*, n=2: *faith*, *fundamentalism*, *radicalism*, *ritual*, *validity of Islam*), POLITICS (n=13: *leader*, *reformer*, *and political authorities*, n=4: *authority*, n=3: *extremist*, *dictatorship*, *hierarchies*, *nationalism*, *party*, *policies*, n=2: *politics*, *fighters*, *awakening and radicalization*, n=2: *establishment*), SOCIAL GROUPS (*fundamentalists*, *geeks*, *elite*, *merchant class*, n=4: *minority*, *vigilantes*, *officials*, *men*, n=2: *people*, n=3: *women*, *committees*, *community*, n=3: *group*), LAW and ORDER (*jurisprudence*, n=2: *power*, n=3: *ruling/s*, n=4: *police*), EDUCATION (*education*, *professors*, n=2: *scholars*, *school*, *teacher/s*), GEOGRAPHICAL LOCATION (*brethren in Iraq*, *in Najaf*), and OTHER (*affairs*, *appeal*, n=3: *buildings*, *culture,* n=2: *debate*, *diversity*, n=2: *establishment*, *experience*, n=2: *fervor*, n=3: *freedom*, *guidance*, n=2: *identity*, *ideology*, n=3: *issue*, n=3: *law*, n=6: *liberty*, n=2: *movement*, n=4: *practice*, *purity*, *reference to gays*, *revival*, n=3: *adversary*).

Figure 4.8-A shows a sample of 50 concordance lines of the word *American*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.8-A: sample of 50 concordance lines of the word "American", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *American* is mostly used as premodifier and occurs in noun phrases (e.g. *an American analyst*, *the American citizen*, *this American woman*, *the first American company*);

2) a distinction can be made between the (groups of) words which carry NEGATIVE SEMANTIC PROSODY (*condemning the*, *denounced*, *forgotten*, *verbal abuse from*, *hated*, *racism in*, *repression in*, *complicit*, *violent backlash against*, n=2: *hate crimes against*, n=2: *violence against*, *the attacks on*, *attacks against the*, *attack the*, *bomb the*, n=2: *bombings of the*, *destruction of the*, *subverting*, *opposition to*, *at odds with*, *condemning the*, *denounced*, *hated*) which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*proud*, *to shoulder with*), which are less frequent;

3) the (groups of) words to the left of *American* can be divided into the following semantic categories: RELIGION (*Muslim and*, *Muslims at*, *Muslims in*, *Islam is as*, *Hindu and Jewish*), POLITICS (*politically active*, *Liberalism in*, *liberalism within*), PEOPLE (n=4: *African*, n=9: *black*, *middle-class*, *Muslim*, n=3: *naturalized*), GEOGRAPHICAL ORIGIN/LOCATION (*Lebanese*, *European cities*), QUANTITY (*half of all*, n=2: *most*, *a small number of*, *thousands of*, *three*) and OTHER (*authentic*, *colloquial*, *complicit*, n=4: *dodging*, *seismic change*, *modern*, n=2: *new*, *drawdown*, *porn-grade*, *standard*, *professor of*, *terrorism among*).

Figure 4.8-B shows a sample of 50 concordance lines of the word *American*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.8-B: sample of 50 concordance lines of the word "American", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) the word *American* is virtually always used as an attribute, except once, when it is used as a noun in the following phrase: *whether the American is a diplomat*;

2) a distinction can be made between the (groups of) words to the right of *American* which carry NEGATIVE SEMANTIC PROSODY (*cultural invasion*, *enemy*, *hurt game*, *Muslims are anti-American*, n=4: *occupiers*) and those which carry POSITIVE SEMANTIC PROSODY (*support for Israel*), the latter being less frequent;

3) the (groups of) words to the right of *American* can be grouped into the following semantic categories: GEOGRAPHICAL ORIGIN/LOCATION (*natives*, *root*[15], *Embassies in Dar es Salaam*, *Embassies in Kenya*, *Embassies in Nairobi*, *Embassy in Kenya*, *Embassy in Riyadh*, *Embassy in Tanzania*, *suburbs*, *town*, *troops in Afghanistan*, *troops on Saudi soil*, *troops to Saudi Arabia*), RELIGION (n=3: *convert*, *Muslim women*, *Catholicism*, *Catholics*, *Christian*, *evangelicals*, *imam*, n=2: *Islam*, *Islamic Relations*, *Judaism*, *missionaries*, *mosque*, n=36: *Muslim*, *Protestant missionaries*, *Religion*, *Sufi cleric*, *ummah*, *and Muslim*, n=3: *Muslim community*), WAR (*allies*, *and Russian troops*, *attack*, *drone strike*, n=2: *drones*, *enemy*, *forces*, n=2: *intelligence*, *military presence*, n=4: *occupiers*, *onslaught*, n=3: *soldiers*, n=4: *troops*), PEOPLE (*adults*, *agents*, *analyst*, n=6: *citizen*, *woman*, *public*, *teen-agers*, *visitors*, *workers*, groups), POLITICS (n=10: *Embassy*, *democracy*, *government*, *imperialism*, *is a diplomat*, n=3: *leader*, *nationalists*, n=2: *policy*, *political establishment*, *Woman's Struggle*), EDUCATION (n=2: *colleges*, *curriculum*, *education*, *studies*), TIME (*before Sept. 11, 2001*, n=2: *and Muslim after 9/11*, *colleges in the 1950's*, *hands about 1996*), ETHNIC GROUPS (n=3: *blacks*, *and Russian*, n=3: *community*), INSTITUTIONS (*American Bankers Association*, *Center for American Progress*) and OTHER (*law*, *conception*, n=2: *culture*, *decision-making*, *event*, *eyes*, *films*, n=2: *flag*, *freedoms*, n=4: *history*, *identity*, *jeeps*, n=3: *life*, *teenage life*, n=3: *movie*, *pressure*, *Promise*, *sentiments*, n=3: *society*, *story*, n=3: *values*, *company*, n=2: *finance*, *journal*, *English*, *idioms*, *baseball caps*, *wears a uniform*).

I performed the same kind of analysis with the top five keywords, excluding proper names of individuals.

[15] The complete sentence is *'Imam Talib and other Black Muslims trace their American roots to the arrival of Muslims from West Africa as slaves in the South'.*

Figure 4.9-A shows a sample of 50 concordance lines (out of 74 total concordance lines) of the word *Uighur*, which refers to an ethnic group living in the Xinjiang, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.9-A: sample of 50 concordance lines of the word "Uighur", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Uighur* is mostly used as a premodifier in prepositional phrases (e.g. *of Uighur culture*, *on Uighur villages*, *of Uighurs*);

2) a distinction can be made between collocates which carry NEGATIVE SEMANTIC PROSODY, which are more frequent (*deadly attacks by*, *violence by*, *the backwardness of*, *trauma inflicted on*, *smothering*, *traumatized*, *detain*, *exile*, *the repression of*, n=3: *suppress*, *erase*), and those which carry POSITIVE SEMANTIC PROSODY, which are less frequent (*protecting fellow*, *watching over*, *shield*, *support*, *better understand*);

3) the (groups of) words to the left of *Uighur* can be classified into the following semantic groups: QUANTITY (*more than 100*, *900,000 residents are*, *more than a dozen*, n=2: *many*, *most*, *hundreds of*, *some*, n=2: *two*), ETHNICITY (*Hui and the*, *Xinjiang's brightest young*, *Ethnic identity*, n=5: *ethnic*), EDUCATION (*studies*, *textbooks on*, *teaching*, *taught in the*), TIME (*just weeks after*, *brief periods of*), RELIGION (n=2: *religion and*, *devout*), SOCIETY AND POLITICS (*numeric targets for*, *resistance in*, *indoctrinate*), and OTHER (*his parents and*, *break*, n=2: *control of the*, *the pressure on*, *a new generation of*, *the history*, *swaths of*, *ordinary*, *prominent*, *polarize*, *said*, *integrated than*, *the growth of*, *traditional*, *break up*, *interviews with*, *Mahmutjian Niyaz*).

Figure 4.9-B shows a sample of 50 concordance lines of the word *Uighur*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.9-B: sample of 50 concordance lines of the word "Uighur", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) one word and one phrase to the right of *Uighur* which carry NEGATIVE SEMANTIC PROSODY were identified (*detentions*, *have long bridled*); while no words or phrases characterized by a POSITIVE SEMANTIC PROSODY were found;

2) the (groups of) words to the right of *Uighur* can be grouped into the following semantic categories: PROFESSIONS (*A former official*, n=3: *officials*, *businessman*, *civil servants*, *educators*), ETHNIC GROUPS (*Kazakhs*, *and other Muslim minorities*, n=2: *Muslims*, *and the majority Han Chinese*), FAMILY (n=2: *families*, *parents*, n=3: *children*, *with relatives*), EDUCATION (*educators*, *students*, *to universities*), CULTURE (*history*, *language*, *literature*, n=4: *culture*), POLITICS (*activists*, n=5: *militants*, *independence*, n=4: *resistance to Chinese rule*), SOCIETY (n=3: *population*, *traditions*, *Chinese society*, n=5: *society*, *villages*, n=2: *identity*) and OTHER (*agree*, *and prevent*, *at a time*, *dress*, *groups*, *have adhered*, *self-rule*, n=2: *make up*, *primarily live*, *to the camps*, *who are secular*, *who have traveled*, n=2: *woman*, *moving about Hotan*, *in large numbers*, *and Gu Wensheng*).

Figure 4.10-A shows a sample of 50 concordance lines (out of 120 concordance lines) of the word *Xinjiang,* which refers to a geographical region in Northwest of China, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.10-A: sample of 50 concordance lines of the word "Xinjiang", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Xinjiang* is a noun mostly occurring in prepositional phrases (e.g. *in Xinjiang, for Xinjiang, across Xinjiang*);

2) a distinction can be made between the (groups of) words to the left of *Xinjiang* which carry NEGATIVE SEMANTIC PROSODY (*camp inmates*, *abuses in*, n=2: *attacks in*, *oppression of Muslims in*, *oppression in*, n=2: *repression in*, *stifle unrest in*, *violent unrest in*, *extremist violence in*, *the brutality of*), and those which carry POSITIVE SEMANTIC PROSODY (*reduce poverty in*, *ensuring stability in*, *a success of*, *praised*), the latter being less frequent;

3) the (groups of) words to the left of *Xinjiang* can be classified into the following semantic categories: POLITICS (*policy course for*, *leadership's strategy for*, *to govern*, *governing*, *party boss in*, *the party's top official in*, *party members in*, *policies in*, n=2: *leadership*, *conference on*, *the government sends*, n=2: *internment camps in*, *officials across*, n=3: *officials in*, n=4: *authorities in*, *authorities in the*, *"maintain stability" across*, *China's actions in*[16], *iron grip in*[17], *security lapses in*, *a harder line in*[18], n=2: *operating in*, *indoctrination programs in*, *protests in*, n=2: *Chinese rule in*, *Mr. Xi's goals for*), ETHNIC GROUPS (*to many Uighurs*, n=2: *Uighur population in*, *Uighurs in*, *minority groups in*, n=2: *ethnic minorities in the*, *Uighur society in*), GEOGRAPHICAL LOCATIONS (*eastern*, *northern*, *in parts of*, n=3: *western region of*, n=7: *southern*, n=2: *western*, *towns across*), RELIGION (*Islamic extremism in*, *radical Islam in*, *Muslims in the*), EDUCATION (*teachers for*, *elementary-school age children in*, n=2: *schools in*, *scholar at the*), QUANTITY (*just under half in*, *11 million in*, *many of them in*, *a fifth of*), and OTHER (*spilling from*, n=2: *children in*, *10 families in*, *big data in*, *security drive in*, *efforts in*, *live in*, *free-flowing monologues in*, *prisons in*, *residents in*, *security in*, *underway in*, n=2: *left*, *image of*, *an expert on*, *transferred to*, *trip to*, n=2: *visit to*, *in recent years*).

Figure 4.10-B shows a sample of 50 concordance lines of the word *Xinjiang*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

[16] The complete sentence is *'Why, in the U.S. view at least, do China's actions in Xinjiang amount to genocide?'*

[17] The complete sentence is '*China has sought for decades to restrict the practice of Islam and maintain an iron grip in Xinjiang'*.

[18] The complete sentence is *'Mr. Xi ordered the party to take a harder line in Xinjiang'*.

Figure 4.10-B: sample of 50 concordance lines of the word "Xinjiang", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Xinjiang* which carry NEGATIVE SEMANTIC PROSODY (*the current crackdown*, *the genocide declaration*, *amount to genocide*, *800-plus townships*, *for infractions*, *had committed crimes*, *have curbed*, *have detained*, *mass detentions*, *now loom*, *to detain Uighurs*, *were punished*), and those which carry POSITIVE SEMANTIC PROSODY (*watching over*, *with an array of technology*, *without putting them at risk*), the latter being less frequent;

2) the (groups of) words to the right of *Xinjiang* can be classified into the following semantic categories: GEOGRAPHY (*in the far northwestern region of China*, *In Hotan alone*, *capital Urumqi*, *from across China*, *from Tibet*, *into other parts of China*, *to the Middle East*, n=6: *region*, *A region*, *a resource-rich territory*), EDUCATION (*urged teachers*, *at Australian National University*, *boarding schools*, *Party Schools*, *teachers*), PEOPLE (*Mr. Chen*, *Mr. Xi*, *Mr. Xi's predecessor*, *Before Mr. Xi*, *Mr. Xi displayed*, *to the public*), TIME (*Since 2017*, *Within weeks*, *in April 2014*, *in the summer of 2017*, *over the past three years*), POLITICS (*as China's leader*, n=2: *government*, n=2: *policy*, *Including sloppy investigations*, *new security controls*, *have recruited*, *top security official*, *where the authorities*, *According to officials*, *according to internal documents*, *the internal report, drafted*[19]), QUANTITY (*Many of whom*, *or about 497,800 students*), ETHNIC GROUPS (*but the Hui*, *brightest young Uighurs*), METADISCOURSE (*the message*, *and 44 pages*) and OTHER (*nestled*, *and immigrated*, *as the work*, *expanded*, *has begun to bleed*, *has grown*, *has soared*, *have recognized*, *is known as*, *is spreading to*, *overall*, *played in*, *overcoming it*, *prisons*, *Production and Construction Corps*, *to give birth*, *who described*, *would require*, *He said*, *reporting on*).

Figure 4.11-A shows a sample of 50 concordance lines (out of 55 concordance lines) of the word *Islamism*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

[19] The complete sentence is '*Officials in Turpan, a city in eastern Xinjiang, drafted the question-and-answer script after the regional government warned local officials to prepare for the returning students*'.

Figure 4.11-A: sample of 50 concordance lines of the word "Islamism", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islamism* is always used as a head noun and mostly occurs in prepositional phrases (e.g. *of Islamism*, *for Islamism*, *toward Islamism*);

2) some (groups of) words to the left of *Islamism* which carry NEGATIVE SEMANTIC PROSODY were found (n=2: *law against*, *the fundamental contradiction that besets*, n=2: *to combat*, *would not destroy*, *complicit in*, *the problem of*, *terror on*, *is critical to*, *hostility toward*), while no (groups of) words which carry POSITIVE SEMANTIC PROSODY were found;

3) the (groups of) words to the left of *Islamism* can be classified into the following semantic categories: RELIGION (*militant Islam – or*, *Islam or*, *the Muslim world*, *Islam and of*, *the bible of*, *radical*), POLITICS (*Communism and*, *liberal strain of*, *Egypt was*, *the countries in which*) and OTHER (*Ideologically*, *crucial moment for*, *a drift away for*, *the target is*, *target*, *no "building of*, *the effects of*, *the evidence of*, *many forms of*, *influence of*, *the rise of*, *maintains that*, *today*, *the environment in which*, *explanation of why*, *Invoking, He says*).

Figure 4.11-B shows a sample of 50 concordance lines of the word *Islamism*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.11-B: sample of 50 concordance lines of the word "Islamism", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Islamism* which carry NEGATIVE SEMANTIC PROSODY (*a source of anger*, *could deepen the disparities*, *decline*, *defeat*, *hardened*, *is declining*, *is in decline*, *is not caused*, *must be eradicated*, *or deflect*, *remains an opposition*, *threatens*), and those which carry POSITIVE SEMANTIC PROSODY (*grew and flourished*, *has flourished*, *has triumphed*, *is capable*, *thrives on*), the latter being less frequent;

2) the (groups of) words to the right of Islamism can be classified into the following semantic categories: POLITICS (*A highly politicized*, *as a political instrument*, *in one country*, *in power*, *Turkey*, *in Turkey*), RELIGION (*and even Islam*, *drives Muslims to*) and OTHER (*From the Taliban*, *An ideology*, *in Decline*, *also rests*, *began by*, *by toughening rules*, n=2: *everywhere*, *first great fountainhead*, *has a tendency*, *is inherent*, *rise*).

Figure 4.12-A shows the concordance lines of the 38 occurrences of the word *Islamically*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.12-A: list of 38 concordance lines of the word "Islamically", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islamically* is always used as a premodifier of adjectives (e.g. *Islamically acceptable stocks*, *Islamically correct versions*, *what is Islamically correct*);

2) two words to the left of *Islamically* which carry NEGATIVE SEMANTIC PROSODY were found (*denied*, *incorrect*), while no words which carry POSITIVE SEMANTIC PROSODY were found;

3) the (groups of) words to the left of *Islamically* can be classified into the following semantic categories: POLITICS (*politically and*, *government by*, *politically incorrect or*, *political activism were*), RELIGION (n=2: *a Muslim could*, *religious scholars deem*, *more radicalized*), COOKING (*A halal*, *cooks*) CLOTHES (*dressed in an*, *the headscarf isn't*) and OTHER (*A sign advertising*, *over an*, *under an*, *needless and*, *market and are*, *to market*, *it is denied*, *$80 million in*, *endowed with*, *a list of*, *the so-called*, *caters to the*, *Al Quaeda was*, *acts that were*, *characters who were*, *They were*, *Made us all*, *I was not*).

Figure 4.12-B shows 38 concordance lines of the word *Islamically*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.12-B: list of 38 concordance lines of the word "Islamically", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the words to the right of *Islamically* which carry NEGATIVE SEMANTIC PROSODY (n=2: *incorrect*, *illegitimate*, *wrong*) and those which carry POSITIVE SEMANTIC PROSODY (*allowed*, *approvable*,

approved, n=4: *acceptable*, n=2: *pure*, n=8: *correct*, *sanctified*, *permissible*, *modest*, *legal*), the latter being more frequent;

2) the (groups of) words to the right of *Islamically* can be divided into the following semantic categories: EVALUATION (n=4: *acceptable*, *allowed*, n=2: *approvable*, n=8: *correct*, n=2: *incorrect*, *illegitimate*, *legal*, *mandated*, *permissible contract*, *pure*, *required*, *wrong*), RELIGION (*"Allah*, *sanctified loose tunic*) and OTHER (*He said*, *People have become*, *characterized state*, n=2: *conscious*, *excited*, *inflected writings*, n=2: *justify*, *modest bathing suits*, *pure kitchen*, *and politically*).

Figure 4.13-A shows a sample of 50 concordance lines (out of 51 total concordance lines) of the word *Shariah*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.13-A: sample of 50 concordance lines of the word "Shariah", in the New York Times corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Shariah*, which is a noun, occurs in noun phrases (e.g. *Shariah requires neither collecting nor paying*, *the Shariah*) and in prepositional phrases (e.g. *under Shariah*, *since Sharia*, *or Shariah*);

2) a distinction can be made between the (groups of) words to the left of *Shariah* which carry NEGATIVE SEMANTIC PROSODY (*abandon*, n=2: *creeping*, *imposed*, *toughening of*, *to prohibit*, *the danger that*, *violate the*, *no choice under*, *noncompliant with*), which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*offering*[20], *compatible with*), which are less frequent;

3) the (groups of) words to the left of *Shariah* can be divided into the following semantic categories: POLITICS (*states ruled by*, *implementing*, *implementation of*, *imposed*, *calling for*[21], *chief*, *abandon*, *applies*, *consulting*, *interpretation of*, *experts in*, *toughening of*, *to prohibit*, *violate the*, *compliance with the*, *by adhering to*, *noncompliant with*, *operating within*), RELIGION (*religious law* -, *regulations*, *jihadist and*, *hadith inform*[22], n=2: *religious law or*, n=2: *Islamic law or*, *Islamic law – the*) and OTHER (*overseen*, n=2: *creeping*, *indeed*, *Islamic finance*, *like*, *since*, *elements of*, *offering*, *only*, *far from spreading*, *the danger that*, *a reference to*, *under*, *while*, *compatible with*, *National*[23], *a board of*).

Figure 4.13-B shows a sample of 50 concordance lines (out of 51 total concordance lines) of the word *Shariah*, in the *New York Times* corpus, sorted alphabetically by one, two and three tokens to the right.

[20] The complete sentence is '*In the neighborhoods of Sparkbrook, Washwood Heath and Alum Rock, where many of Birmingham's Muslims live, Mosques dot the cityscape, some offering Shariah councils for family matters'.*

[21] The complete sentence is '*But he is an astute politician who has shouted for jihad and vowed to welcome home from self-imposed exile Rizieq Shihab, the head of the Islamic Defenders Front, which gained notoriety for attacking nightclubs in Jakarta, the capital, and calling for Shariah law*'.

[22] Hadith: 'the body of tradition and legend about Mohammed and his followers, uses as a basic of Islamic law' (https://www.collinsdictionary.com/it/dizionario/inglese/hadith).

[23] The complete phrase is '*National Shariah Board'.*

Figure 4.13-B: sample of 50 concordance lines of the word "Shariah", in the New York Times corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

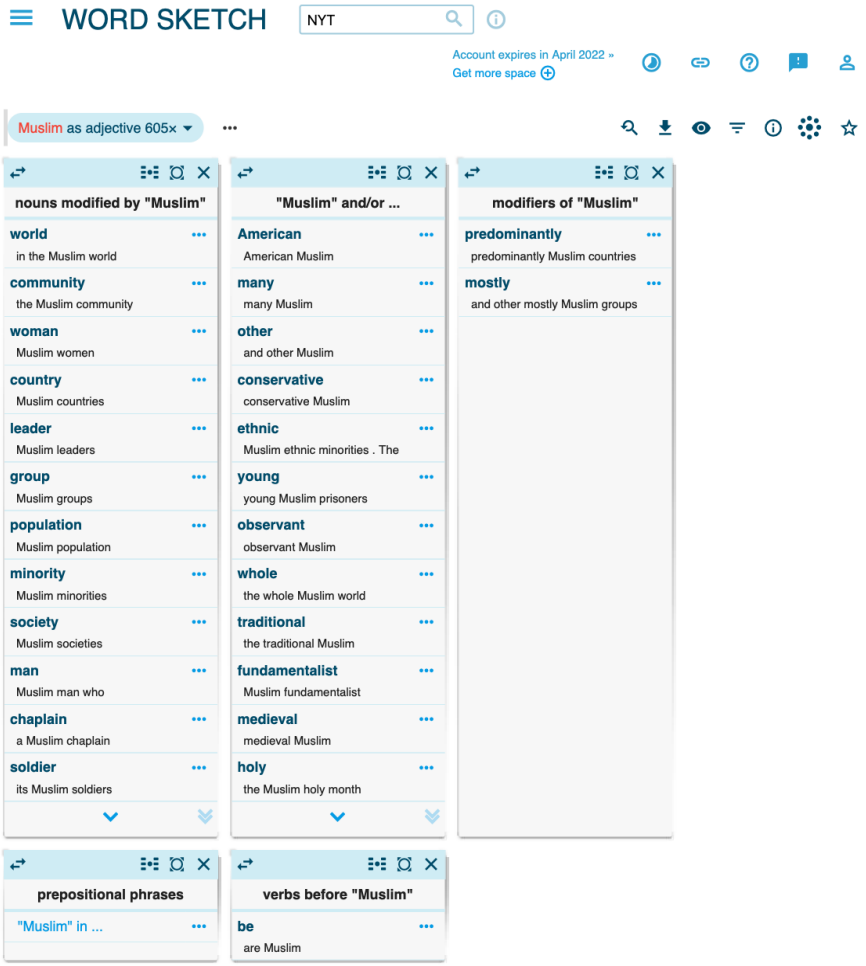
1) a distinction can be made between the (groups of) words to the right of *Shariah* which carry NEGATIVE SEMANTIC PROSODY (*as Islamophobes*, *they seem worried*, *with public flogging*, *rejected by*, *by throwing gays*[24], *forbidding of "riba"*) which are more frequent, and that which carries POSITIVE SEMANTIC PROSODY (*compliance*), the latter being less frequent;

2) the (groups of) words to the right of *Shariah* can be divided into the following semantic categories: RELIGION (*as Islamophobes*, *Muslims*, *which guides Muslims*, *or Islamic law*, *the legal code*, *the Quran*, *"The Islamic principles*, n=5: *law*), POLITICS (*and restoring a caliphate*, *committee*, *councils*, *compliance*, *officer*, *rules and regulations*), FINANCE (*cutoff*, *forbidding of "riba"*, *The largest Islamic financial institution*) and OTHER (*such as businesses*, *they seem worried*, *with public flogging*, *In 1986*, *It reflects*, *At the core*, *rejected by*, *that includes clauses*, *also speaks to*, *as interpreted by*, *Board*[25], *by throwing gays*, *Capital*, *is rooted*, *itself*, *requires*, *scholars*, *to everyday life*, *use*, *we are adhering*, *will serve*, *In Saudi Arabia*).

The collocates that I identified of the five most frequent words and the top five key keywords were based on an examination of a small set of concordances. To corroborate my analysis, I considered the lexical profile of the same words on the basis of the entire data. To this end, I used the *Word Sketch* function in Sketch Engine. Figure 4.14-A shows part of the *Word Sketch* of the word *Muslim* used as adjective in the *New York Times* corpus.

[24] The complete sentence is '*Meanwhile, Islamic State militants implement the most extreme interpretation of Shariah by throwing gays from rooftops*'.
[25] The complete phrase is *National Shariah Board*.

Figure 4.14-A: Partial Word Sketch of "Muslim" used as adjective in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

1) there is a tendency for the word *Muslim* to occur with words which refer to a particular ETHNIC GROUP (*American*, *ethnic minorities*, *immigrant*, *migrant*, *traditional*, *black*, *Indonesian*, *Hindu*, *Arab*, *mainstream*), thus indicating the existence of many different subgroups under the group 'Muslims';

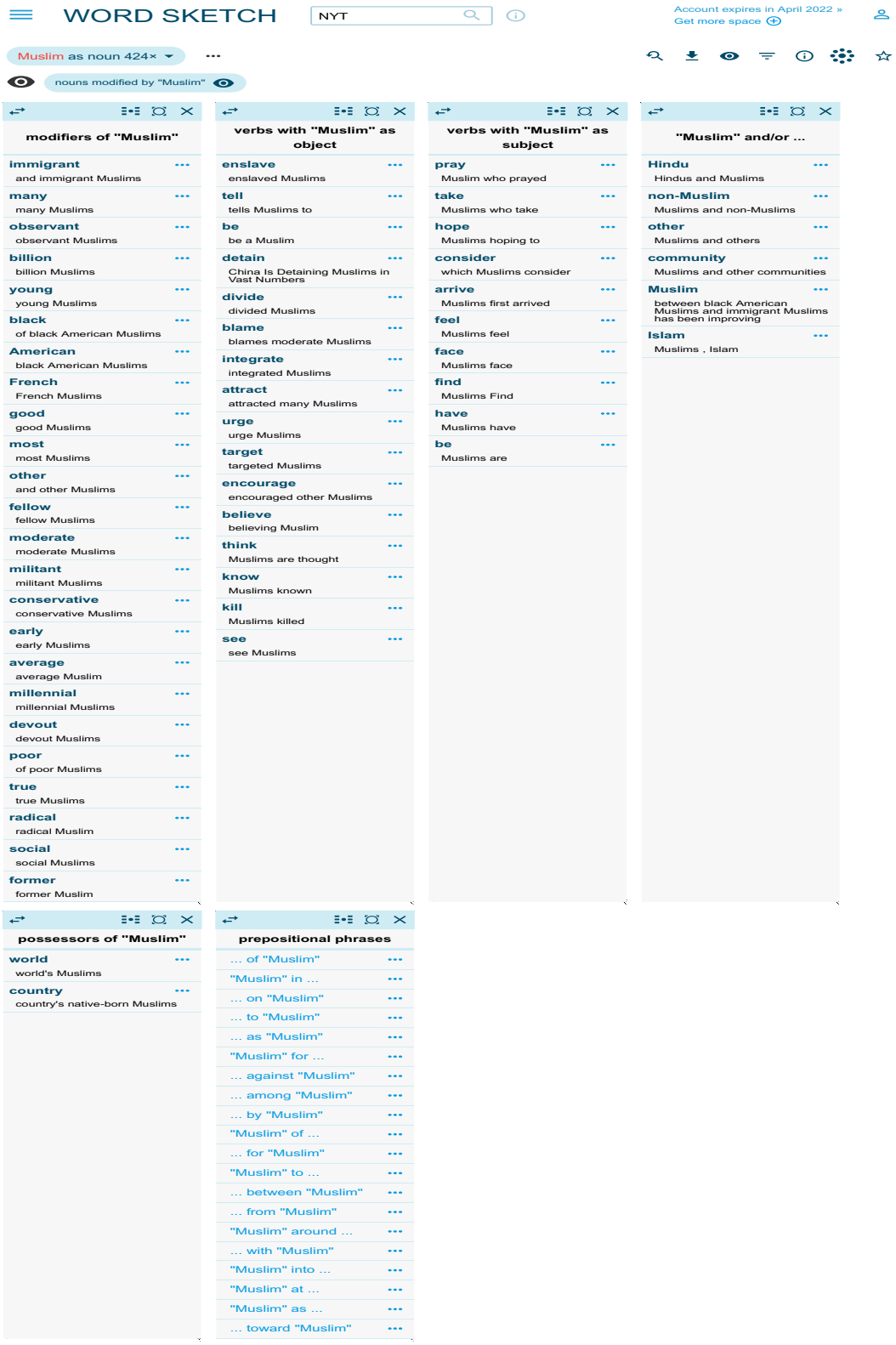
2) most of the collocates of the word *Muslim* used as an adjective can be divided into the same semantic fields identified through my concordance analysis, namely QUANTITY/AMOUNT (*many*, *whole*, *global*, *several*, *predominantly*, *mostly*), RELIGION (*world*, *chaplain*, *cleric*, *prayer*, *faith*, *radical*, *fundamentalist*, *fundamentalism*, *extremist*, *convert*, *religion*, *conservative*, *observant*, *holy*, *religious*, *pious*, *Christian*), GEOGRAPHICAL LOCATION (*enclave*, *land*, *area*, *neighborhood*), PEOPLE (*woman*, *man*, *girl*, *child*, *owner*, *young*, *respondent*, *family*, *people*, *adult*,

patient, *resident*, *youth*, *friend*, *public*), POLITICS (*country*, *leader*, *minister*, *nation*, *state*, *law*, *jurist*, *communism*, *advocate*, *movement*, *moderate*, *civil*), ETHNIC GROUPS (*American*, *ethnic*, *black*, *Indonesian*, *Hindu*, *Arab*), SOCIAL GROUPS (*group*, *population*, *minority*, *community*, *society*, *immigrant*, *migrant*, *organization*, *member*, *mainstream*), INSTITUTIONS (*association*, *center, Brotherhood*), POWER (*ruler*, *rule*), EDUCATION (*student*, *scholar*), PERSONAL BELONGINGS (*garb*, *veil*), TERRORISM (*soldier*, *terrorist*), CULTURE (*tradition*, *traditional*, *custom*, *identity*) and OTHER (*ban*, *greeting*, *holiday*, *name*, *case*, *month*, *life*, *other*, *medieval*, *lost*, *first*). However, I also assigned some words to two new semantic fields (i.e. different from those identified in my concordance analysis), namely WORK (*client*, *customer*, *merchant*, *employee*, *banker*, *worker*) and WAR (*prisoner*, *inmate*);

3) *Muslim* is often preceded by the verb *be*, as in *most of whom <u>are Muslim</u>*, *a man who wrongly assumed <u>he was Muslim</u>*, *I am Muslim*, *this is Muslim*.

Figure 4.14-B shows a part of the *Word Sketch* of *Muslim* used as noun in the *New York Times* corpus.

Figure 4.14-B: Partial Word Sketch of "Muslim" used as noun in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Muslim* used as noun can be assigned to the same semantic fields identified through collocations, namely POLITICS (*moderate*, *conservative*, *militant*, *radical*, *country*[26]), ETHNIC GROUPS (*Hindu*, *non-Muslim*, *Muslim*, *black*, *American*, *French*), EVALUATION (*average*, *true*, *good*), RELIGION (*world*, *observant*, *devout*, *pray*), SOCIAL GROUPS (*fellows*, *community*, *social*), QUANTITY/AMOUNT (*many*, *billion*, most), and OTHER (*other*, *early*, *millennial*, *poor*, *immigrant*, *young*, *former*);

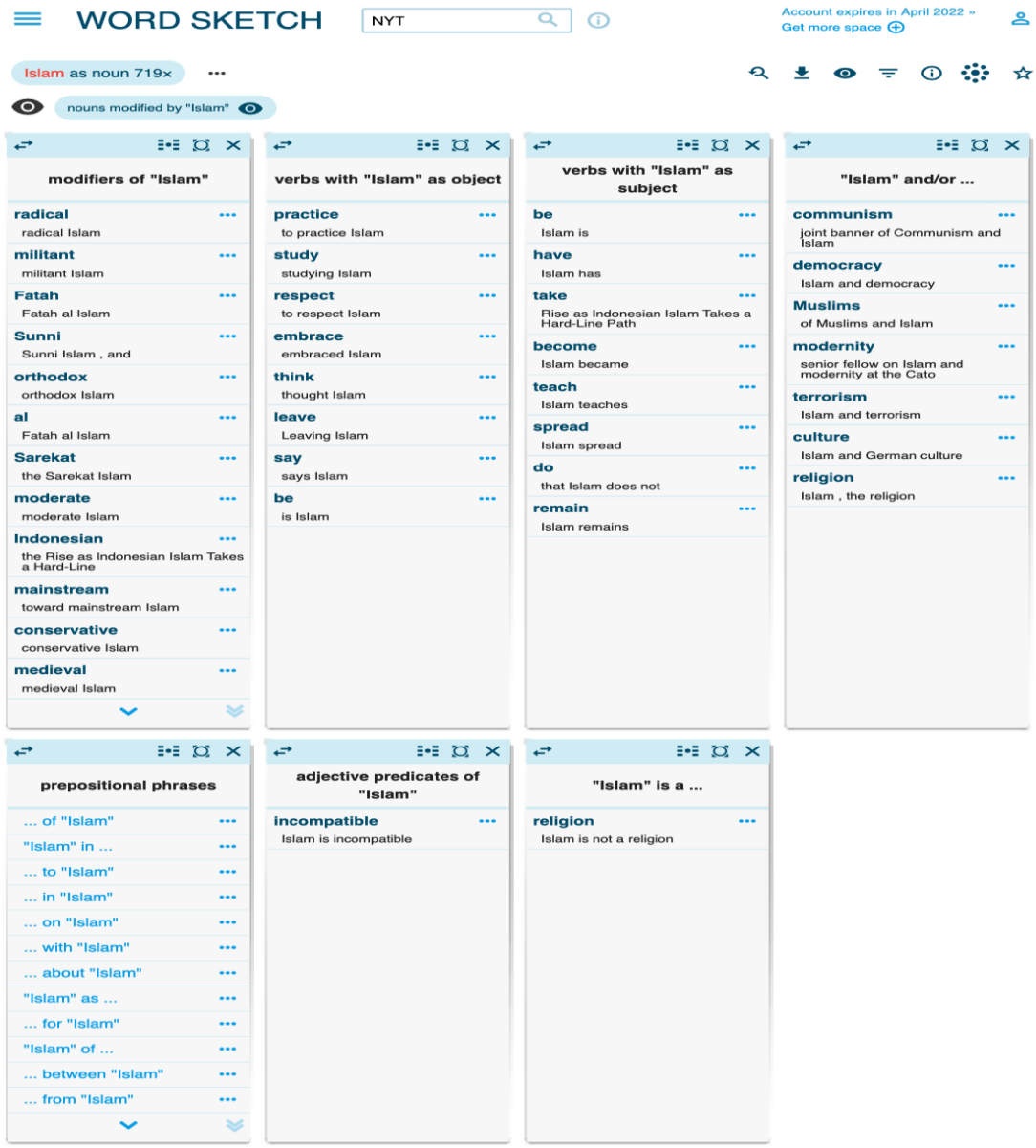
2) some collocates of the word *Muslim* not previously identified in my concordance analysis carry NEGATIVE SEMANTIC PROSODY (*enslave*, *detain*, *blame*, *kill*) and one collocate carries POSITIVE SEMANTIC PROSODY (*encourage*);

3) the verbs occurring with *Muslim* as a subject are fewer than those occurring with *Muslim* as an object, and the latter encode the notions of 'persuasion' (e.g. *integrate*, *attract*, *urge*, *encourage*) or 'conflict' (e.g. *detain*, *enslave*, *kill*, *blame*).

Figure 4.15 shows part of the *Word Sketch* of the word *Islam* used as noun in the *New York Times* corpus.

[26] Examples are: *For many African-American converts, Islam is an experience both spiritual and political, an expression of empowerment in a country they feel dominated by a white elite* and *I came here to build this country together*.

Figure 4.15: Partial Word Sketch of "Islam" used as noun in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

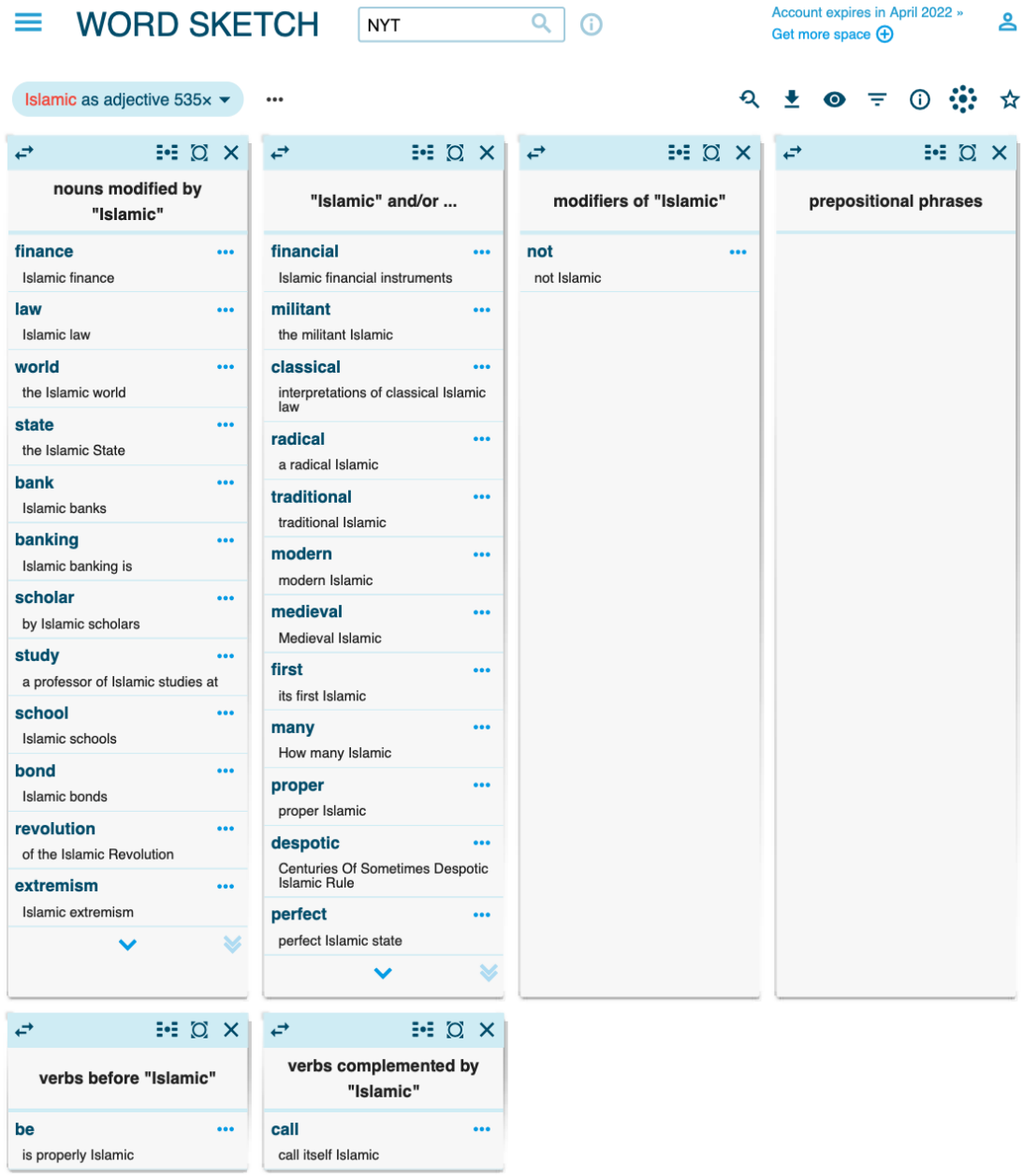
1) most of the collocates of the word *Islam* used as a noun can be assigned to the same semantic fields identified through concordance analysis, namely RELIGION (*radical*, *Fatah*, *Sunni*, *orthodox*, *al*, *conservative*, *Muslims*, *religion*), COMMUNICATION (*say*), POLITICS (*militant*, *Sarekat*, *moderate*, *political*, *communism*, *democracy*), PERIOD (*today*, *medieval*, *modernity*), PEOPLE (*mainstream*), TERRORISM (*terrorism*), ETHNIC GROUPS (*Indonesian*), SEPARATION (*incompatible*), OTHER (*practice*, *respect*, *embrace*, *think*, *leave*, *have*, *take*, *become*,

spread, *do*, *remain*, *culture*). Additionally, I identified some words which belong to a new semantic field, namely EDUCATION (*study*, *teach*);

2) the verbs occurring with *Islam* as the object mostly indicate people's attitude and/or adherence to religion (*practice*, *study*, *respect*, *embrace*, *leave*).

Figure 4.16 shows part of the *Word Sketch* of the word *Islamic* in the *New York Times* corpus.

Figure 4.16: Partial Word Sketch of "Islamic" in the New York Times corpus

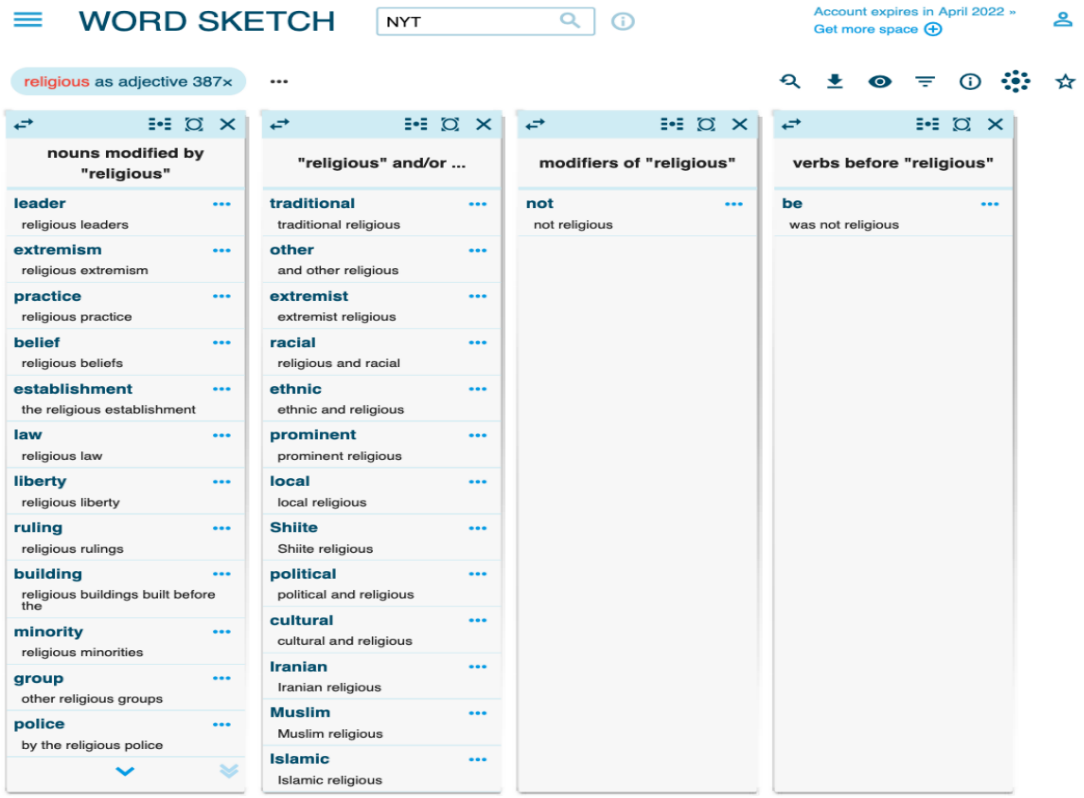


The analysis of the *Word Sketch* led to the following observations:

1) the collocates occurring after *Islamic* can be assigned to the same semantic fields identified through collocations, namely RELIGION (*world*, *mysticism*, *fundamentalism*, *religion*, *religious*, *theology*, *theological*, *piety*, *doctrine*, *faith*, *cleric*, *belief*, *guidance*, *ethic*), POLITICS (*law*, *state*, *revolution*, *extremism*, *jurisprudence*, *republic*, *rule*, *party*, *resurgence*, *leader*, *militant*, *despotic*, *political*, *radicalism*, *extremist*, *extremism*), TIME (*modern*, *medieval*), EDUCATION (*scholar*, *study*, *school*, *schooling*, *scholarship*, *teaching*, *text*, *history*), GROUPS (*group, European*), PEOPLE (*people*), CULTURE (*tradition*, *custom*, *identity*, *classical*), OTHER (*principle*, *science*, *organization*, *center*, *awakening*, *style*, *perspective*, *dress*, *view*, *apocalypticism*, *arm*, *astronomer*, *civilization*, *thinker*, *thought*, *influence*, *movement*, *first*, *many*, *proper*, *perfect*, *private*, *new*, *call*, *terrorist*), with a particular emphasis on the semantic field of FINANCE (*finance*, *financier*, *financing*, *bank*, *banking*, *banker*, *bond*, *fund*, *market*, *investment*, *instrument*, *product*, *system*, *consumer*, *sector*, *institutions*, *value*).

Figure 4.17 shows part of the *Word Sketch* of the word *religious* in the *New York Times* corpus.

Figure 4.17: Partial Word Sketch of "religious" in the New York Times corpus



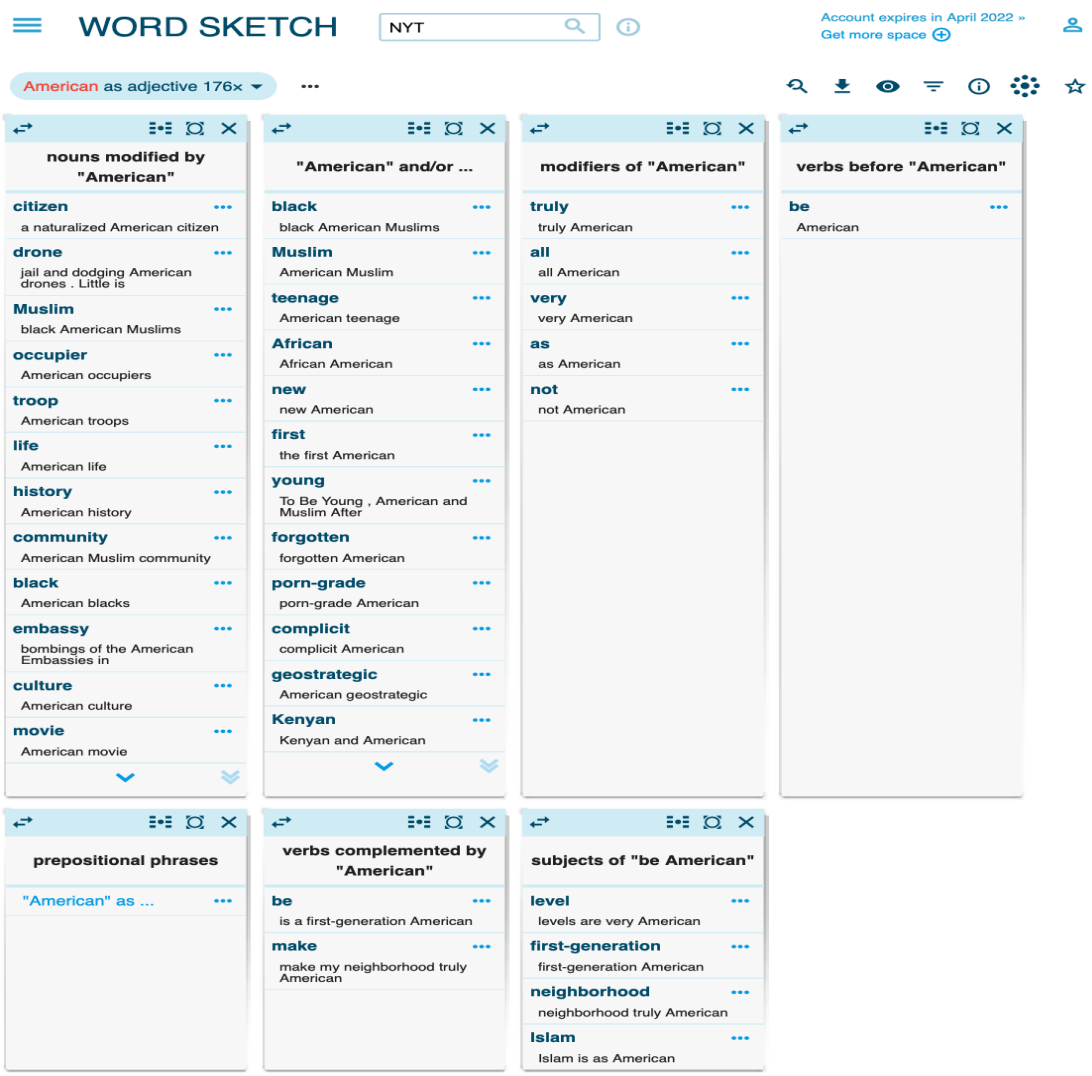
The analysis of the *Word Sketch* led to the following observations:

1) the word *religious* often occurs in binomial expressions with words referring to ETHNICITY, e.g. *religious and racial*, *ethnic and religious*, *cultural and religious*;

2) the collocates of *religious* belong to the same semantic field identified through concordance analysis, namely RELIGION (*extremism*, *belief*, *fervor*, *observance*, *extremist*, *commandment*, *devotion*, *tolerance*, *resurgence*, *fanatic*, *ritual*, *fundamentalism*, *fundamentalist*, *principle*, *faith*, *value*, *Shiite*, *Muslim*, *Islamic*, *practice*), POLITICS (*leader*, *establishment*, *ruling*, *police*, *authority*, *nationalism*, *hierarchy*, *politics*, *state*, *movement*, *political*, *war*), GEOGRAPHICAL ORIGIN or LOCATION (*local*, *Iranian*), PEOPLE (*people*, *man*, woman), SOCIAL GROUPS (*minority*, *group*, *organization*, *community*, *association*, family), LAW AND ORDER (*law*), EDUCATION (*education*, *study*, *student*, *teacher*, *text*, *scholar*, *school*) and OTHER (*liberty*, *building*, *adversary*, *freedom*, *identity, issue*, *revival*, *expression*, *activity*, *culture*, *need*, *discrimination*, *test*, *matter*, *affair*, *training*, *reason*, *idea*, *view*, *form*, *violence*, *debate*, *traditional*, *other*, *racial*, *ethnic*, *prominent*, *cultural*).

Figure 4.18 shows part of the *Word Sketch* of the word *American* in the *New York Times* corpus.

Figure 4.18: Partial Word Sketch of "American" in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

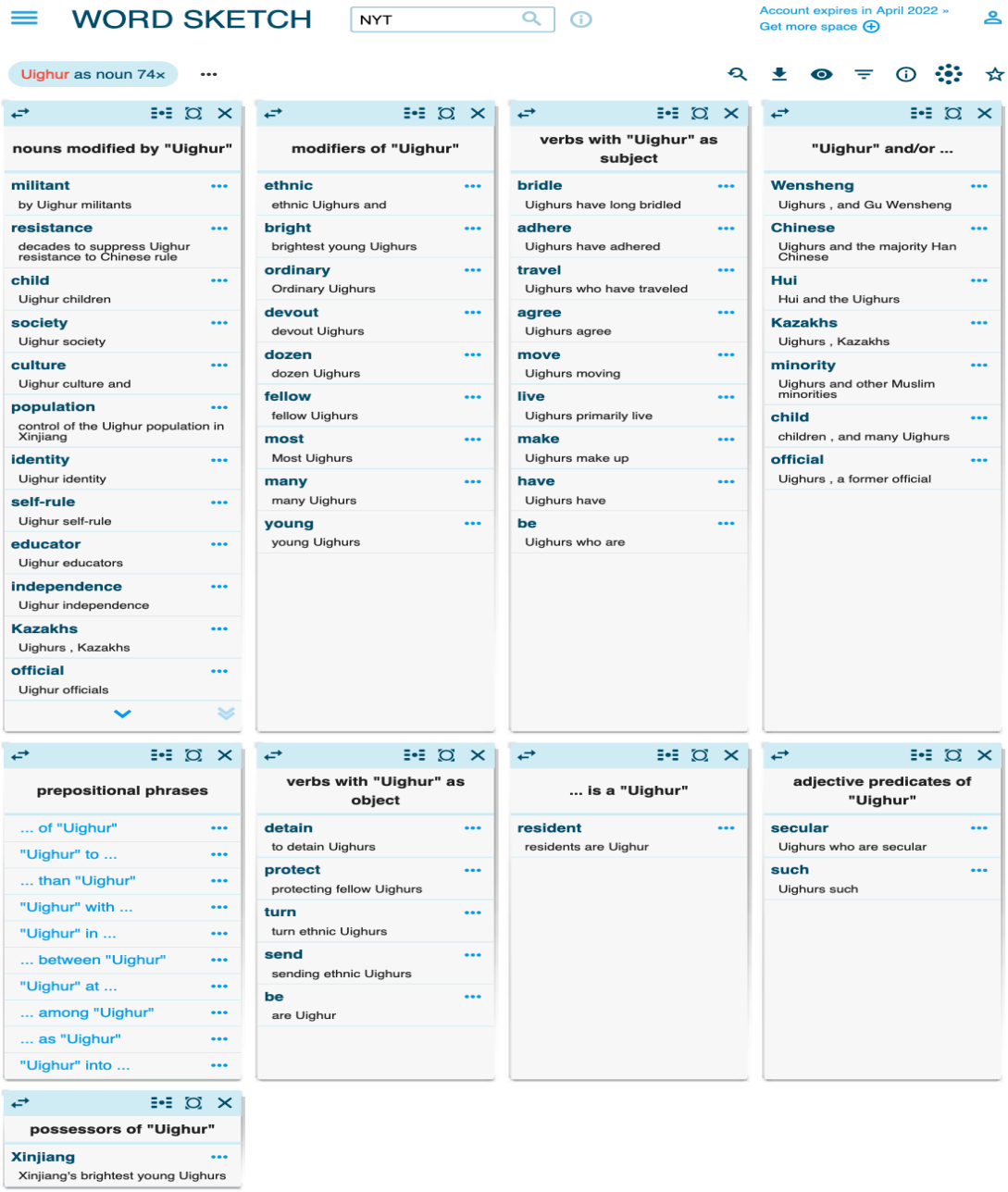
1) the word *American* tends to occur with words belonging to the semantic fields of WAR (*drone*, *occupier*, *soldier*, *force*, *ally*, *captor*, *onslaught*, *intelligence*, *strike*[27], *invasion*, *guard*, *enemy*, *attack*, *military*), RELIGION, especially Islam and Catholicism (*Muslim*, *ummah*, *imam*, *mosque*, *Jewish*, *convert*, *missionary*, *evangelicalism*, *evangelical*, *Catholicism*, *religion*, *Protestant*); and ETHNICITY (*Muslim*, *ummah black*, *African*, *Jewish*), thus indicating the existence of many different subgroups within the larger group 'Americans';

[27] A relevant example is: '*If something were to happen to me in Yemen by an American drone strike, it would be a day's news'.*

2) on three occasions, the word *American* is preceded by emphasizers (*all*, *very*, *truly*), two of which (*all*, *truly*) present *American* as a gradable adjective: *'He was met with gasps from the audience and social media complaints: "Outraged at a Muslim prayer at an all American event!"'*; '*Not that some parents don't choose their children's mates, sometimes even for daughters who on other levels are very American*'.

As with concordances, I performed the same kind of analysis with the top five keywords, excluding proper names of individuals. Figure 4.19 shows part of the *Word Sketch* of the word *Uighur* in the *New York Times* corpus.

Figure 4.19: Partial Word Sketch of "Uighur" in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

1) the *Word Sketch* shows that the word *Uighur* is used both as a head noun and as a premodifier (31% vs 69% of the time, respectively), while in the concordances I analysed I mostly identified cases in which the word *Uighur* is used as premodifier;

2) as found in the concordances, the words following *Uighur* tend to belong to the semantic field of POLITICS (*militant*, *resistance*, *official*, *detention*, *activist*, *self-rule*, *independence*);

3) as found in the concordances, most of the premodifiers of *Uighur* encode the notion of QUANTITY (*dozen*, *most*, *many*);

4) the verbs occurring with *Uighur* as the subject (*bridle*, *adhere*, *travel*, *agree*, *move*, *live*, *make*, *have*, *be*) are more varied and more numerous than the verbs with *Uighur* as object (*detain*, *protect*, *turn*, *send*, *be*);

5) the verbs with *Uighur* as object are action verbs presenting Uighurs as patients affected by others' paternalistic behavior toward them (*detain*, *protect*, *turn*[28], *send*[29]);

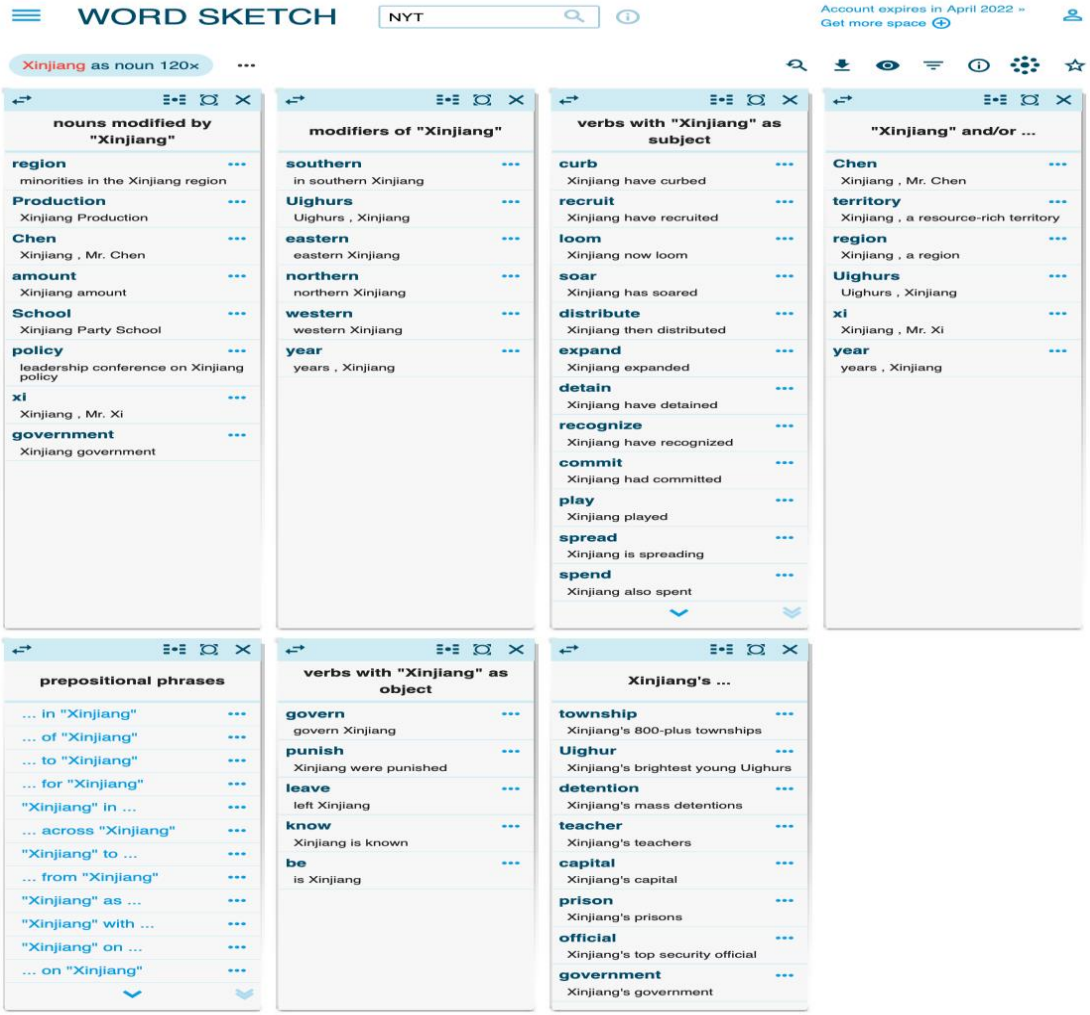
6) the word *Uighur* tends to occur in binomial expressions with words which refer to other ethnic groups (*Uighurs and the majority Han Chinese*, *Hui and the Uighurs*, *Uighurs, Kazakhs*, *Uighurs and other Muslim minorities*);

Figure 4.20 shows part of the *Word Sketch* of the word *Xinjiang*, which is a noun, in the *New York Times* corpus.

[28] The complete sentence is '*After a succession of violent antigovernment attacks reached a peak in 2014, the Communist Party chief, Xi Jinping, sharply escalated the crackdown, orchestrating an unforgiving drive to turn ethnic Uighurs and other Muslim minorities into loyal citizens and supporters of the party'*.

[29] The complete sentence is: '*The Times also discovered reports online written by teams of Chinese officials who were assigned to monitor families with detained relatives, and a study published last year that said officials in some places were indiscriminately sending ethnic Uighurs to the camps to meet numerical quotas*'.

Figure 4.20: Partial Word Sketch of the word "Xinjiang" in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

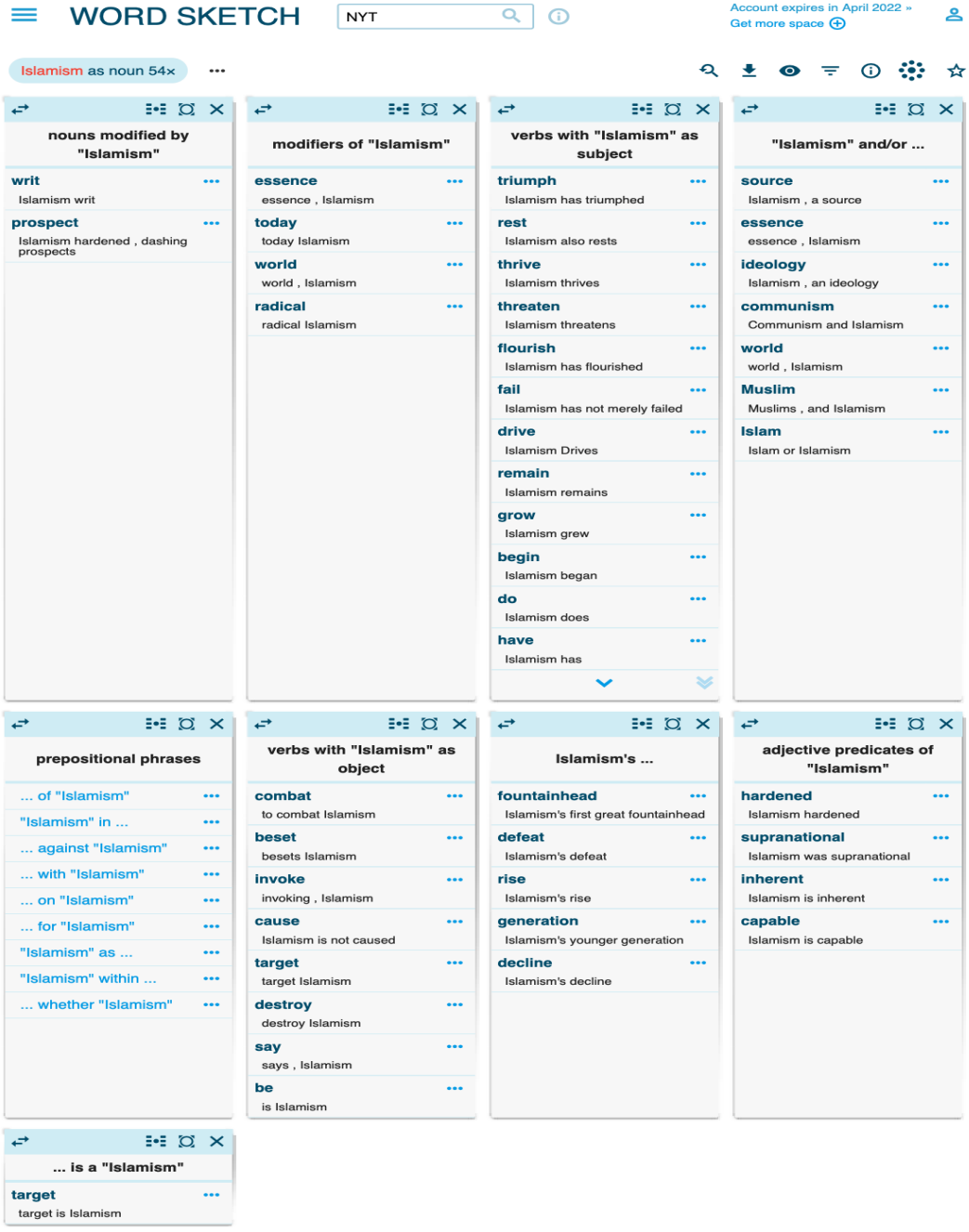
1) the only modifiers of *Xinjiang* encode GEOGRAPHICAL LOCATIONS (*eastern*, *western*, *southern*, *northern*)[30];

2) the verbs occurring with *Xinjiang* as the subject (*curb*, *recruit*, *loom*, *soar*, *distribute*, *expand*, *detain*, *recognize*, *commit*, *play*, *spread*, *spend*, *describe*, *grow*, *begin*, *have*, *say*, *be*) are more varied and more numerous than those occurring with *Xinjiang* as an object (*govern*, *punish*, *leave*, *know*, *be*). Moreover, among the verbs with *Xinjiang* as subject, there is a frequent presence of verbs which indicate expansion (*soar*, *expand*, *grow*).

[30] As you can see in the *Word Sketch*, the words *Uighurs* and *years* are not actually premodifiers of the word *Xinjiang*, since they occur before it, but are separated from it by a comma.

Figure 4.21 shows part of the *Word Sketch* of the word *Islamism* in the *New York Times* corpus.

Figure 4.21: Partial Word Sketch of the word "Islamism" in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

1) the only modifier of the word *Islamism* is *radical*[31];

[31] The other words (*essence, today, world*) are not modifiers of *Islamism* although *Sketch Engine* classifies them as such simply because they occur before it.

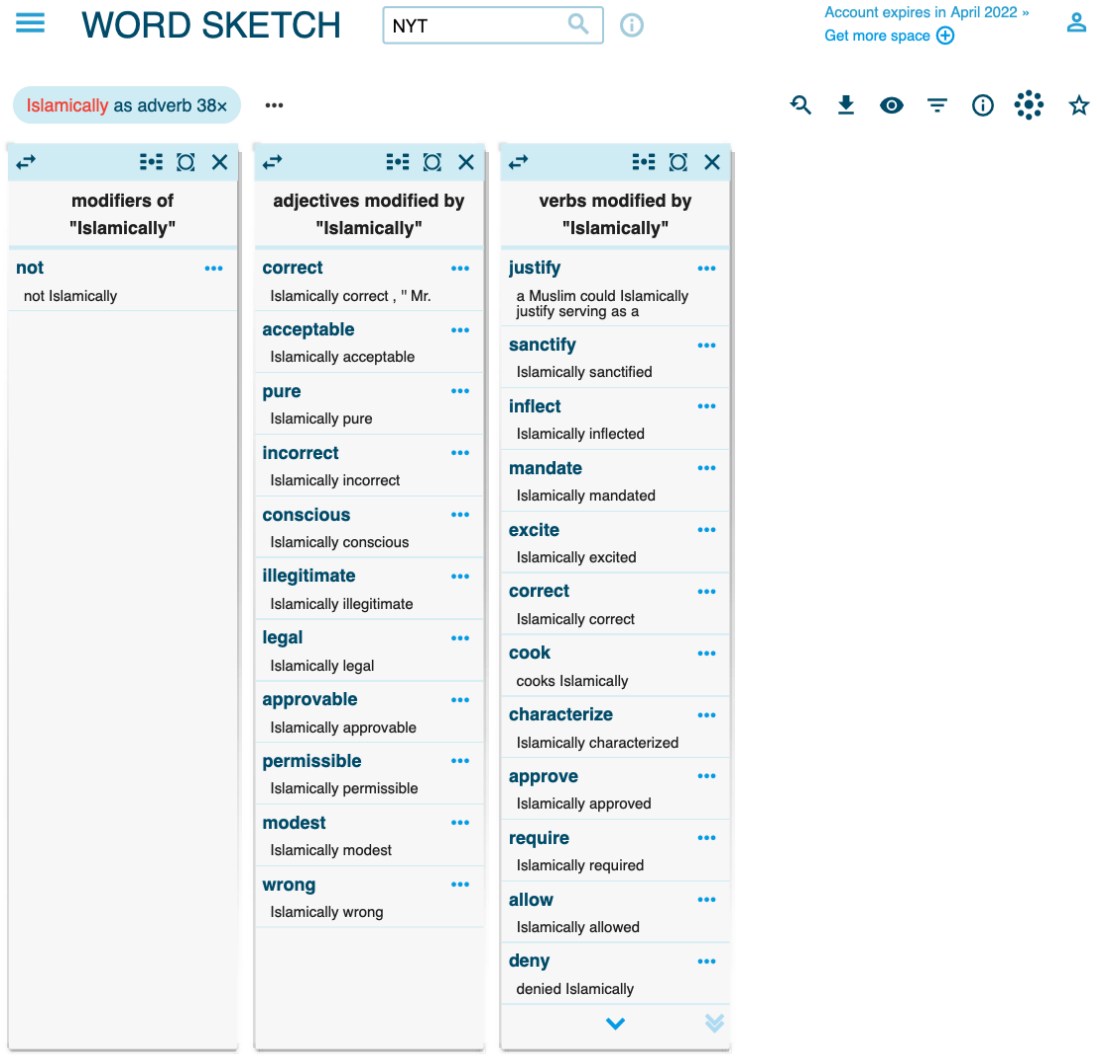
2) the verbs with *Islamism* as the subject (*triumph*, *rest*, *thrive*, *threaten*, *flourish*, *fail*, *drive*, *remain*, *grow*, *begin*, *do*, *have*, *be*) are more varied and more frequent than those with *Islamism* as an object (*combat*, *beset*, *invoke*, *cause*, *target*, *destroy*, *say*, *be*);

3) many verbs used with *Islamism* as subject carry POSITIVE SEMANTIC PROSODY (*triumph*, *thrive*, *flourish*, *grow*), while only two carry NEGATIVE SEMANTIC PROSODY (*threaten*, *fail*);

4) three verbs used with *Islamism* as object carry NEGATIVE SEMANTIC PROSODY (*combat*, *beset*, destroy), while none carry POSITIVE SEMANTIC PROSODY.

Figure 4.22 shows part of the *Word Sketch* of the word *Islamically* used as adverb in the *New York Times* corpus.

Figure 4.22: Partial Word Sketch of "Islamically" in the New York Times corpus



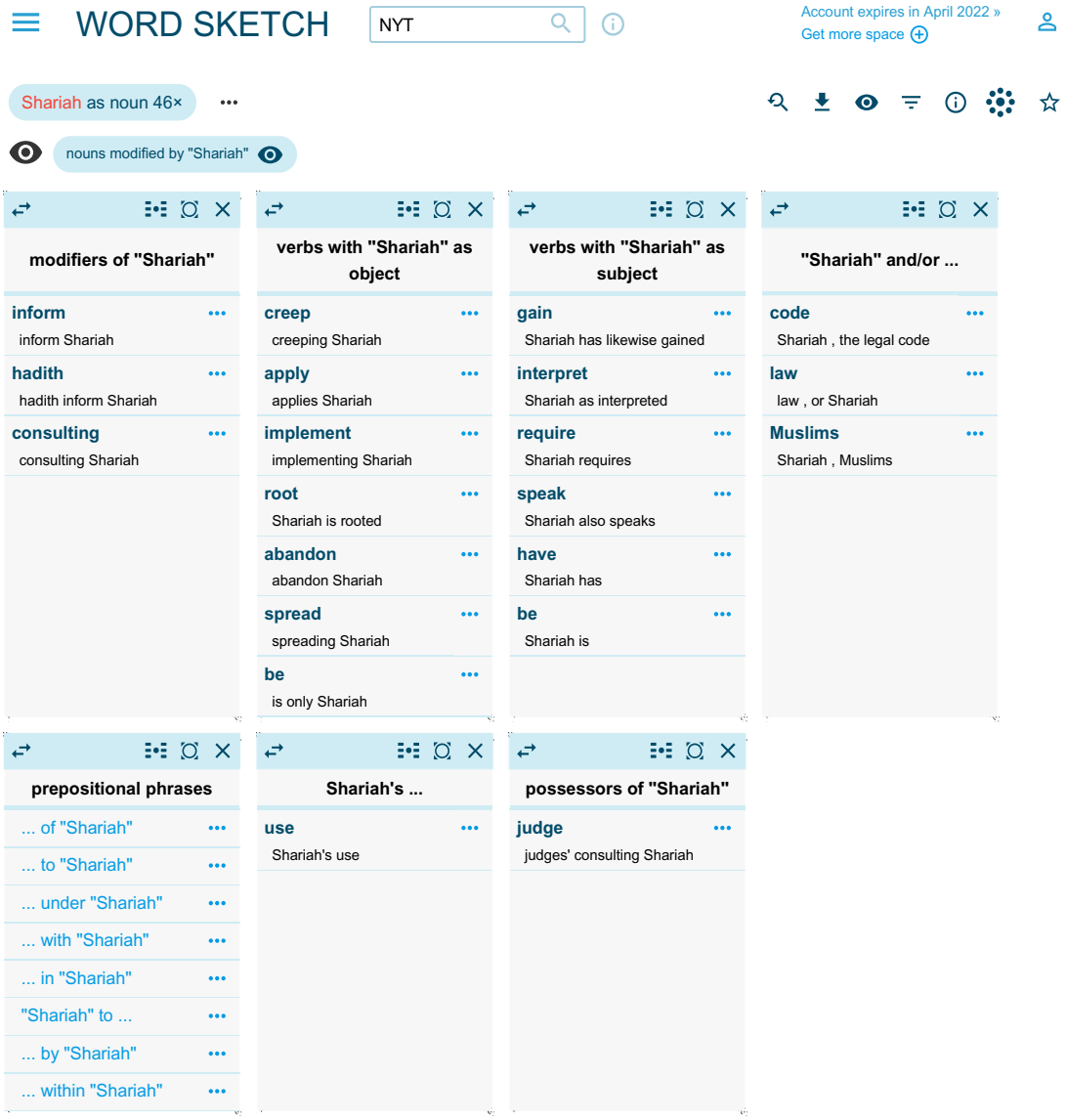
The analysis of the *Word Sketch* led to the following observations:

1) the word *Islamically* is used as a premodifier of adjectives (81.6% of the time; e.g. '*"There has been some discussion about what is Islamically correct," Mr. Karas told the jury*'; *'Malaysia's Securities Commission maintains a list of Islamically acceptable stocks'*; *'It has done nearly $80 million in Islamically approvable "mortgage-alternative" financing for residential and commercial real estate in 15 estate'*) and verbs (18.4% of the time; e.g. '*The only way a Muslim could Islamically justify serving as a soldier in the U.S. Army is if his intention is to follow the footsteps of men like Nidal"*'; '*Last month, after writing an essay arguing the headscarf isn't Islamically mandated, we received verbal abuse*'; '*it is denied Islamically*');

2) as found in the analysis of the concordances, a great number of collocates of *Islamically* belong to the semantic field of EVALUATION (*correct*, *incorrect*, *acceptable*, *illegitimate*, *approvable*, *modest*, *wrong*).

Figure 4.23 shows the complete *Word Sketch* of the word *Shariah* in the *New York Times* corpus.

Figure 4.23: complete Word Sketch of "Shariah" in the New York Times corpus



The analysis of the *Word Sketch* led to the following observations:

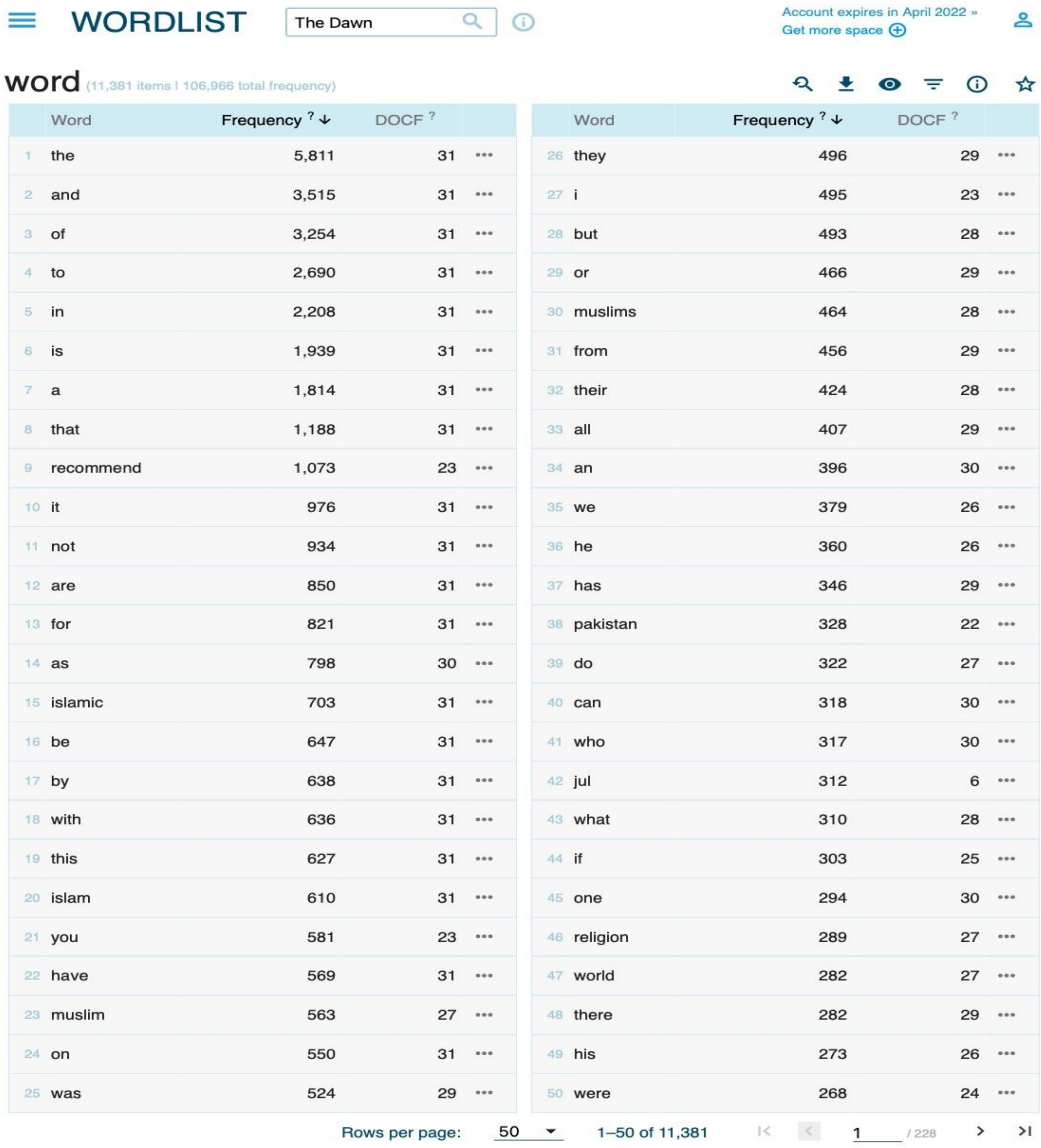
1) the verbs occurring with *Shariah* as object (*creep*, *apply*, *implement*, *root*, *abandon*, *spread*, *be*) are more varied and more numerous than the verbs with *Shariah* as subject (*gain*, *interpret*, *require*, *speak*, *have*, *be*);

2) many words which collocate with *Shariah* belong to the semantic field of LAW (*consulting*, *apply*, *implement*, *interpret*, *code*, *law*, *judge*).

4.3 Findings about *The Dawn* Corpus

In order to identify the most recurrent topics and notions mentioned in *The Dawn* corpus, first of all, it is useful to look at word frequency lists. Figures 4.24-A-B-C-D show the two hundred most frequent words in this corpus.

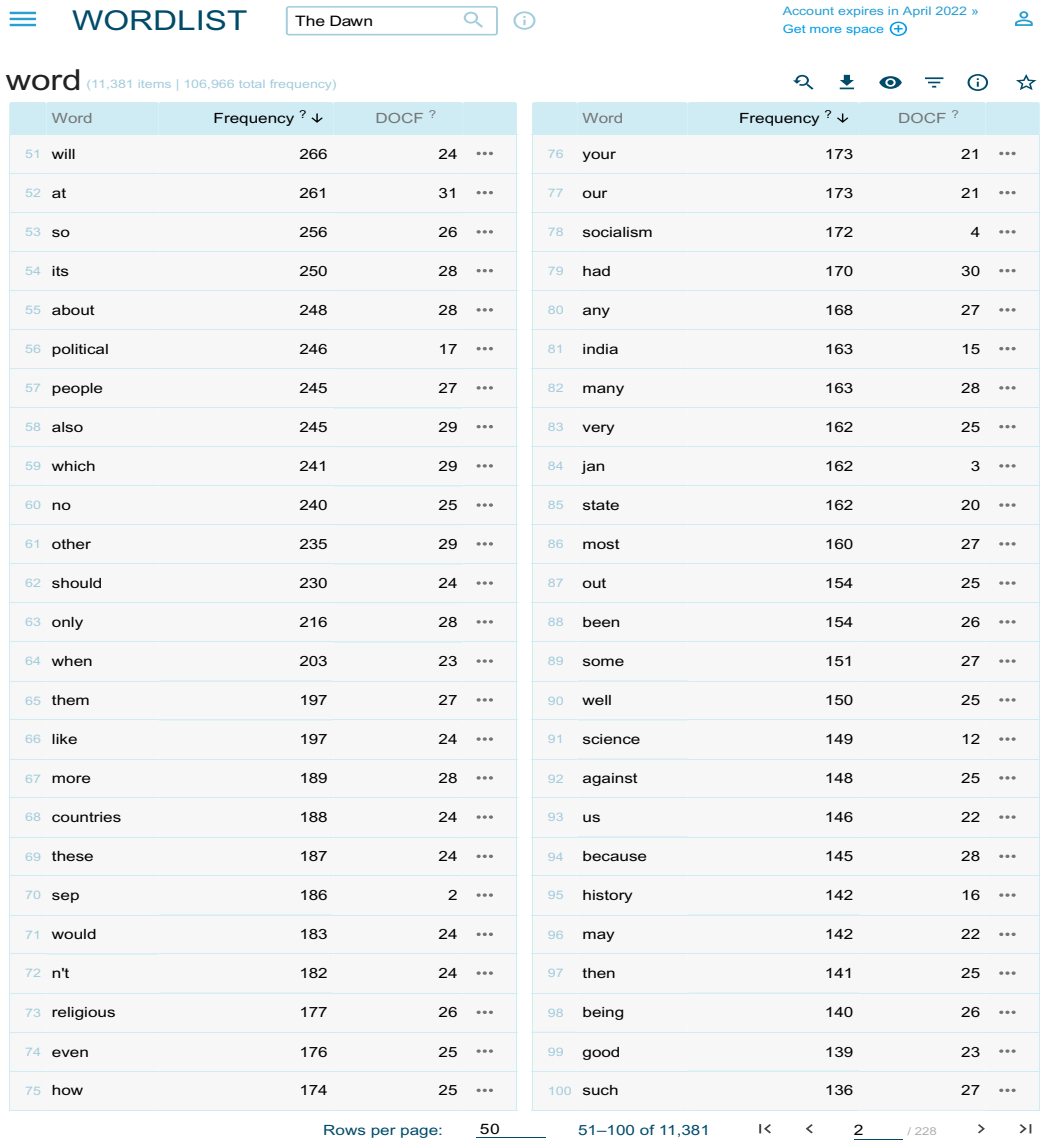
Figure 4.24-A: word frequency list of The Dawn corpus, Part A



	Word	Frequency	DOCF
1	the	5,811	31
2	and	3,515	31
3	of	3,254	31
4	to	2,690	31
5	in	2,208	31
6	is	1,939	31
7	a	1,814	31
8	that	1,188	31
9	recommend	1,073	23
10	it	976	31
11	not	934	31
12	are	850	31
13	for	821	31
14	as	798	30
15	islamic	703	31
16	be	647	31
17	by	638	31
18	with	636	31
19	this	627	31
20	islam	610	31
21	you	581	23
22	have	569	31
23	muslim	563	27
24	on	550	31
25	was	524	29

	Word	Frequency	DOCF
26	they	496	29
27	i	495	23
28	but	493	28
29	or	466	29
30	muslims	464	28
31	from	456	29
32	their	424	28
33	all	407	29
34	an	396	30
35	we	379	26
36	he	360	26
37	has	346	29
38	pakistan	328	22
39	do	322	27
40	can	318	30
41	who	317	30
42	jul	312	6
43	what	310	28
44	if	303	25
45	one	294	30
46	religion	289	27
47	world	282	27
48	there	282	29
49	his	273	26
50	were	268	24

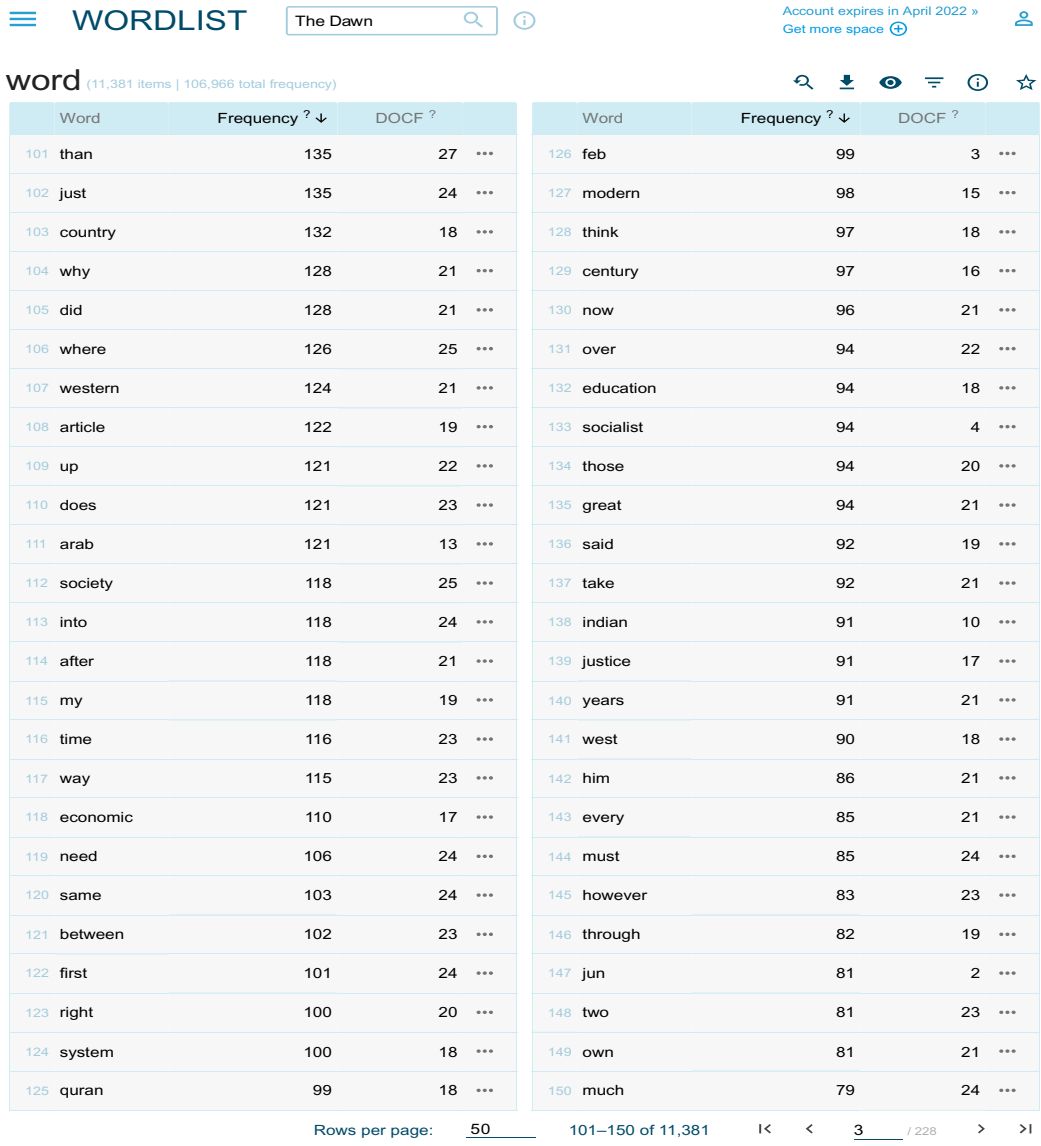
Figure 4.24-B: word frequency list of The Dawn corpus, Part B



	Word	Frequency	DOCF
51	will	266	24
52	at	261	31
53	so	256	26
54	its	250	28
55	about	248	28
56	political	246	17
57	people	245	27
58	also	245	29
59	which	241	29
60	no	240	25
61	other	235	29
62	should	230	24
63	only	216	28
64	when	203	23
65	them	197	27
66	like	197	24
67	more	189	28
68	countries	188	24
69	these	187	24
70	sep	186	2
71	would	183	24
72	n't	182	24
73	religious	177	26
74	even	176	25
75	how	174	25

	Word	Frequency	DOCF
76	your	173	21
77	our	173	21
78	socialism	172	4
79	had	170	30
80	any	168	27
81	india	163	15
82	many	163	28
83	very	162	25
84	jan	162	3
85	state	162	20
86	most	160	27
87	out	154	25
88	been	154	26
89	some	151	27
90	well	150	25
91	science	149	12
92	against	148	25
93	us	146	22
94	because	145	28
95	history	142	16
96	may	142	22
97	then	141	25
98	being	140	26
99	good	139	23
100	such	136	27

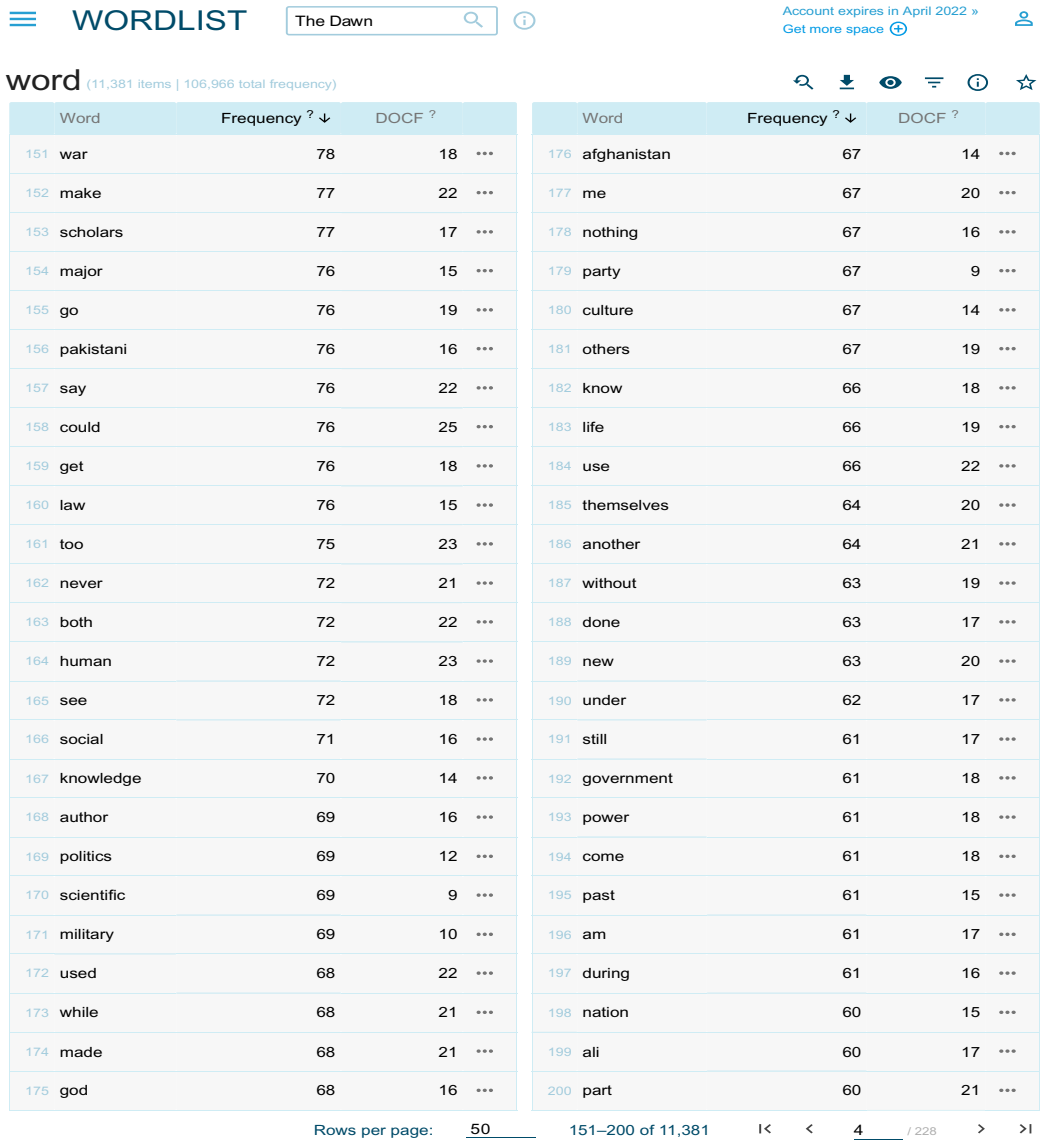
Figure 4.24-C: word frequency list of The Dawn corpus, Part C



	Word	Frequency	DOCF
101	than	135	27
102	just	135	24
103	country	132	18
104	why	128	21
105	did	128	21
106	where	126	25
107	western	124	21
108	article	122	19
109	up	121	22
110	does	121	23
111	arab	121	13
112	society	118	25
113	into	118	24
114	after	118	21
115	my	118	19
116	time	116	23
117	way	115	23
118	economic	110	17
119	need	106	24
120	same	103	24
121	between	102	23
122	first	101	24
123	right	100	20
124	system	100	18
125	quran	99	18

	Word	Frequency	DOCF
126	feb	99	3
127	modern	98	15
128	think	97	18
129	century	97	16
130	now	96	21
131	over	94	22
132	education	94	18
133	socialist	94	4
134	those	94	20
135	great	94	21
136	said	92	19
137	take	92	21
138	indian	91	10
139	justice	91	17
140	years	91	21
141	west	90	18
142	him	86	21
143	every	85	21
144	must	85	24
145	however	83	23
146	through	82	19
147	jun	81	2
148	two	81	23
149	own	81	21
150	much	79	24

Figure 4.24-D: word frequency list of The Dawn corpus, Part D



	Word	Frequency	DOCF
151	war	78	18
152	make	77	22
153	scholars	77	17
154	major	76	15
155	go	76	19
156	pakistani	76	16
157	say	76	22
158	could	76	25
159	get	76	18
160	law	76	15
161	too	75	23
162	never	72	21
163	both	72	22
164	human	72	23
165	see	72	18
166	social	71	16
167	knowledge	70	14
168	author	69	16
169	politics	69	12
170	scientific	69	9
171	military	69	10
172	used	68	22
173	while	68	21
174	made	68	21
175	god	68	16

	Word	Frequency	DOCF
176	afghanistan	67	14
177	me	67	20
178	nothing	67	16
179	party	67	9
180	culture	67	14
181	others	67	19
182	know	66	18
183	life	66	19
184	use	66	22
185	themselves	64	20
186	another	64	21
187	without	63	19
188	done	63	17
189	new	63	20
190	under	62	17
191	still	61	17
192	government	61	18
193	power	61	18
194	come	61	18
195	past	61	15
196	am	61	17
197	during	61	16
198	nation	60	15
199	ali	60	17
200	part	60	21

The most frequent words shown in Figures 4.24-A-B-C-D which are relevant to the Islamic world can be grouped into the following semantic categories: RELIGION (*god*, *islamic*, *islam*, *muslim*, *muslims*, *quran*, *religion*, *religious*), GEOGRAPHY (*afghanistan*, *arab*, *india*, *indian*, *pakistan*, *pakistani*, *saudi*, *west*, *western*) and POLITICS (*government*, *justice*, *law*, *military*, *nation*, *party*, *political*, *politics*, *power*, *socialism*, *socialist*, *state*, *system*, *war*).

Other terms which appear in the frequency word list can be grouped into the following semantic categories: KNOWLEDGE (*author*, *culture*, *education*, *knowledge*, *scholars*, *science*, *scientific*), SOCIETY (*country*, *countries*, *economic*, *history*, *human*, *modern*, *people*, *social*, *society*, *state*, *world*) and OTHER (*believe*, *good*, *major*, *media*).

As in the *New York Times* corpus, these "other" terms are actually relevant to the Islamic world too, because they collocate with words from that semantic field (e.g. *Pakistani*, *Islamic*, *religious*, *faith*, etc.), except for the word '*scientific*'. The extent to which this occurs is as follows: *countries*: 71.6% of the time; *scholars*: 59.8%; *world*: 41.2%; *state*: 40.5%; *history*: 36%; *society*: 27.9%; *modern*: 25%; *country*: 23.3%; *science*: 18.2%; *education*: 14.7%; *people*: 12%; *culture*: 11.5%; *knowledge*: 11.1%; *major*: 11.1%; *media*: 10.77%; *believe*: 9.6%; *good*: 5.7%; *social*: 4.9%; *economic*: 4.8%; *author*: 3,7%; *human*: 2.83%.

In order to have a more contextualized view of the discourse on Muslims and Islam, it is useful to identify recurrent phrases in the corpora, as can be retrieved through the *N-gram* function: Figures 4.25-A-B-C-D show the 200 most frequent sequences of four tokens with a minimum frequency of two occurrences in *The Dawn* corpus, listed in order of decreasing frequency.

Figure 4.25-A: most frequent 4-word n-grams in The Dawn corpus, Part A

N-GRAMS The Dawn

Account expires in April 2022 »
Get more space

4-grams, word (lowercase) (items: 1,727 , total frequency: 3,927)

	Word	Frequency	DOCF
1	in the muslim world	24	8
2	in the name of	24	12
3	on the other hand	13	9
4	the name of islam	12	8
5	is one of the	12	8
6	nothing to do with	11	6
7	the rest of the	10	8
8	all over the world	10	8
9	rest of the world	9	7
10	one of the most	9	6
11	with the help of	8	3
12	at the same time	8	7
13	has nothing to do	8	6
14	on the basis of	8	6
15	in the islamic world	7	4
16	as well as the	7	5
17	in the middle east	7	5
18	of the muslim world	7	4
19	of the qu'ran and	6	2
20	with the soviet union	6	1
21	the quran and the	6	5
22	of the quran and	6	6
23	the name of religion	6	6
24	islamic financial products and	6	2
25	expressions of political islam	6	2
26	during the cold war	5	1
27	the muslims of india	5	3
28	it is true that	5	3
29	was one of the	5	2
30	i remember being laughed	5	1
31	to do with religion	5	4
32	financial products and transactions	5	1
33	in the modern world	5	3
34	i do n't see	5	5
35	createdness of the quran	5	1
36	in the quran and	5	4
37	to the fact that	5	4
38	remember being laughed at	5	1
39	the islamic golden age	5	1
40	he is trying to	5	2
41	the createdness of the	5	1
42	a culture of science	5	2
43	the united states and	5	5
44	he should be tried	5	1
45	of the islamic state	5	4
46	and this is the	4	2
47	to be the most	4	3
48	producing four witnesses to	4	1
49	across the muslim world	4	4
50	the crime of rape	4	1

Rows per page: 50 1–50 of 1,727 1 / 35

Figure 4.25-B: most frequent 4-word n-grams in The Dawn corpus, Part B

N-GRAMS The Dawn

Account expires in April 2022 »
Get more space

4-grams, word (lowercase) (items: 1,727 , total frequency: 3,927)

	Word	Frequency	DOCF
51	the root cause of	4	4
52	in countries like pakistan	4	3
53	if you do n't	4	4
54	there is nothing wrong	4	3
55	in the hands of	4	2
56	i am not a	4	4
57	peace be upon him	4	3
58	but there is no	4	4
59	majority of the muslims	4	3
60	the soviet union and	4	2
61	after the creation of	4	2
62	in the shape of	4	3
63	the writer is a	4	4
64	islamic republic of pakistan	4	4
65	the ba'ath socialist party	4	1
66	interpretations of the qu'ran	4	2
67	of the createdness of	4	1
68	of political islam and	4	2
69	financial products and services	4	2
70	the quran and sunnah	4	2
71	as far as the	4	3
72	not be able to	4	4
73	witnesses to prove the	4	1
74	have they tried to	4	1
75	he should be punished	4	1
76	four witnesses to prove	4	1
77	one of the first	4	3
78	is supposed to be	4	4
79	religious beliefs strictly personal	4	1
80	you are trying to	4	3
81	when it comes to	4	3
82	at the time of	4	2
83	of the history of	3	3
84	the rightest expressions of	3	1
85	the rule of the	3	3
86	to the maximum extent	3	1
87	this is the reason	3	3
88	admissible in rape cases	3	1
89	one is not quite	3	3
90	religion and science are	3	2
91	be used as a	3	2
92	soviet invasion of afghanistan	3	2
93	but the question is	3	2
94	the hands of the	3	2
95	to accept the apology	3	1
96	of islam in the	3	3
97	in the context of	3	3
98	and the use of	3	2
99	became to be known	3	3
100	nation in the world	3	3

Rows per page: 50 51–100 of 1,727 2 / 35

Figure 4.25-C: most frequent 4-word n-grams in The Dawn corpus, Part C

N-GRAMS The Dawn

Account expires in April 2022 »
Get more space

4-grams, word (lowercase) (items: 1,727 , total frequency: 3,927)

	Word	Frequency	DOCF
101	quran and the sunnah	3	2
102	and its western and	3	2
103	is not quite sure	3	3
104	those who believed that	3	1
105	mainly due to the	3	2
106	extent of the law	3	1
107	not be used as	3	1
108	the emergence of a	3	3
109	all right for me	3	1
110	the state and the	3	3
111	v. c. bhutani jul	3	1
112	monarchy in a coup	3	2
113	this is due to	3	3
114	for the rest of	3	2
115	by the way i	3	3
116	in the form of	3	2
117	that some of the	3	3
118	the muslim world and	3	3
119	be tried and punished	3	1
120	history and islamic history	3	1
121	in the political context	3	2
122	the islamic republic of	3	3
123	the quran is not	3	3
124	is that it is	3	3
125	conventional financial products and	3	1
126	the muslims of south	3	1
127	the taliban and al	3	2
128	the founders of the	3	2
129	to the creation of	3	3
130	it is in the	3	3
131	why is it that	3	3
132	the doctrine of the	3	1
133	they need to be	3	3
134	and can not be	3	3
135	expressions of liberal islam	3	2
136	one of the best	3	3
137	after the cold war	3	2
138	in spite of the	3	2
139	abdul wahid mohammad jul	3	1
140	the fact that the	3	3
141	if there is no	3	3
142	the other side of	3	3
143	this day and age	3	3
144	for the betterment of	3	3
145	presence or absence of	3	1
146	i agree with you	3	3
147	a handful of people	3	3
148	rightest expressions of political	3	1
149	that one of the	3	2
150	the creation of pakistan	3	2

Rows per page: 50 101–150 of 1,727 3 / 35

Figure 4.25-D: most frequent 4-word n-grams in The Dawn corpus, part D

N-GRAMS The Dawn

Account expires in April 2022 »
Get more space

4-grams, word (lowercase) (items: 1,727 , total frequency: 3,927)

	Word	Frequency	DOCF
151	the maximum extent of	3	1
152	muslims of india as	3	1
153	swept under the carpet	3	3
154	right for me to	3	1
155	in the west and	3	2
156	religion out of politics	3	1
157	this guy should be	3	1
158	as the rest of	3	3
159	muslims of south asia	3	1
160	can be attributed to	3	2
161	is in the quran	3	2
162	is the key to	3	2
163	n't think for themselves	3	1
164	doctrine of the createdness	3	1
165	no matter what religion	3	2
166	for the most part	3	3
167	he can not do	3	2
168	between the eighth and	3	1
169	i do n't know	3	2
170	i do not agree	3	2
171	religion from the state	3	3
172	at the expense of	3	2
173	the creation of an	3	2
174	at the university of	3	2
175	and that is what	3	2

	Word	Frequency	DOCF
176	so its all right	3	1
177	in a state of	3	3
178	rest of the muslim	3	2
179	i fail to understand	3	3
180	in syria and iraq	3	3
181	sir syed ahmed khan	3	2
182	interesting to note that	3	3
183	is a matter of	3	3
184	from all over the	3	3
185	it should not be	3	3
186	off ties with the	3	1
187	using the name of	3	2
188	a spade a spade	3	3
189	in the middle of	3	3
190	in early islamic history	3	1
191	in the right direction	3	3
192	the author is making	3	2
193	from right to left	3	2
194	the help of the	3	3
195	the palestinians and kashmiris	3	1
196	the people of the	3	3
197	of the islamic world	3	3
198	there has been a	3	3
199	given in the quran	3	2
200	of the palestinians and	3	1

Rows per page: 50 151–200 of 1,727 4 / 35

As in the *n-grams* analysis conducted in the *New York Times* corpus, the n-grams retrieved comprise both sequences of words that are not complete phrases (e.g. *there has put the*, *over the age of*, *on the extent of*), but also some that are (e.g. *of the Quran and*, *in the Middle East*, *the United States and*).

A great number of n-grams found in the corpus can be assigned to the following semantic fields: RELIGION (n=37; *in the muslim world*, *of the muslim world*, *the muslims of india*, *muslims of india as*, *muslims of south asia*, *across the muslim world*, *majority of the muslims*, *the muslim world and*, *the muslims of south*, *rest of the muslim*, *the name of Islam*, *in the Islamic world*, *of the islamic state*, *of islam in the*, *expressions of liberal islam*, *of the islamic world*, *history and islamic history*, *of the qu'ran and*, *createdeness*

of the quran, *in the quran and*, *interpretations of the qu'ran*, *the quran and sunnah*, *quran and the sunnah*, *the quran is not*, *is in the quran*, *given in the quran*, *the name of religion*, *religion and science are*, *religious beliefs strictly personal*, *to do with religion*, *religion out of politics*, *religion from state*, *no matter what religion*, *the createdeness of the*, *of the createdness of*, *doctrine of the createdeness*, *the doctrine of the*), POLITICS (n=13; *with the soviet union*, *the soviet union and*, *soviet invasion of Afghanistan*, *expressions of political islam*, *of political islam and*, *the islamic republic of*, *the ba'ath socialist party*, *nation in the world*, *extent of the law*, *the state and the*, *monarchy in a coup*, *in the political context*, *rightest expressions of political*), GEOGRAPHICAL LOCATIONS (n=10; *all over the world*, *rest of the world*, *in the middle east*, *and its western and*, *in the west and*, *the united states and*, *in countries like Pakistan*, *islamic republic of pakistan*, *the creation of pakistan*, *in Syria and Iraq*), TIME (n=6; *during the cold war*, *in the modern world*, *the islamic golden age*, *after the cold war*, *this day and age*, *in early islamic history*) and FINANCE (n=3; *islamic financial products and*, *financial products and services*, *conventional financial products and*).

Other n-grams found in the corpus refer to more varied subjects, each of which is, however, exemplified only once or a few times: NAMES OF INDIVIDUAL MUSLIMS AND/OR NAMES OF ISLAMIC GROUPS (*v. c. bhutani jul*, *abdul wahid mohammad jul*, *sir syed ahmed khan*, *the Taliban and al*), KNOWLEDGE (*at the university of*, *A culture of science*), LEGAL MATTERS (*producing four witnesses to*, *witnesses to prove the*, *four witnesses to prove*, *the crime of rape*, *admissible in rape cases*, *he should be punished*, *be tried and punished*), PHENOMENA AND CIRMUMSTANCES (*the root cause of*, *after the creation of*, *the emergence of a*, *to the creation of*, *for the betterment of*, *presence or absence of*), ASSESSMENT (*the rightest expressions of*, *right for me to*, *in the right direction*) and OTHER (*peace be upon him*, *the writer is a*, *at the history of*, *the rule of*, *to the maximum extent*, *to accept the apology*, *the founders of the*, *a handful of people*, *swept under the carpet*, *between the eighth and*, *in a state of*, *a spade a spade, the author is making*, *from right to left*, *the help of the*, *of the Palestinians and*).

A complementary view on the lexical make-up of *The Dawn* corpus can be gained by identifying its most unusually frequent, and thus its most distinctive, words, as can be revealed through the *keyword* function. Figures 4.26-A-B show the top 200 keywords of

The Dawn corpus obtained by comparing it against the *English Web 2020 (enTenTen20)* corpus, used as a reference corpus.

Figure 4.26-A: top 200 keywords from The Dawn corpus compared against the English Web 2020 (enTenTen20) corpus, Part A

KEYWORDS — The Dawn

Account expires in April 2022 » Get more space

SINGLE-WORDS ✓ | MULTI-WORD TERMS ✓

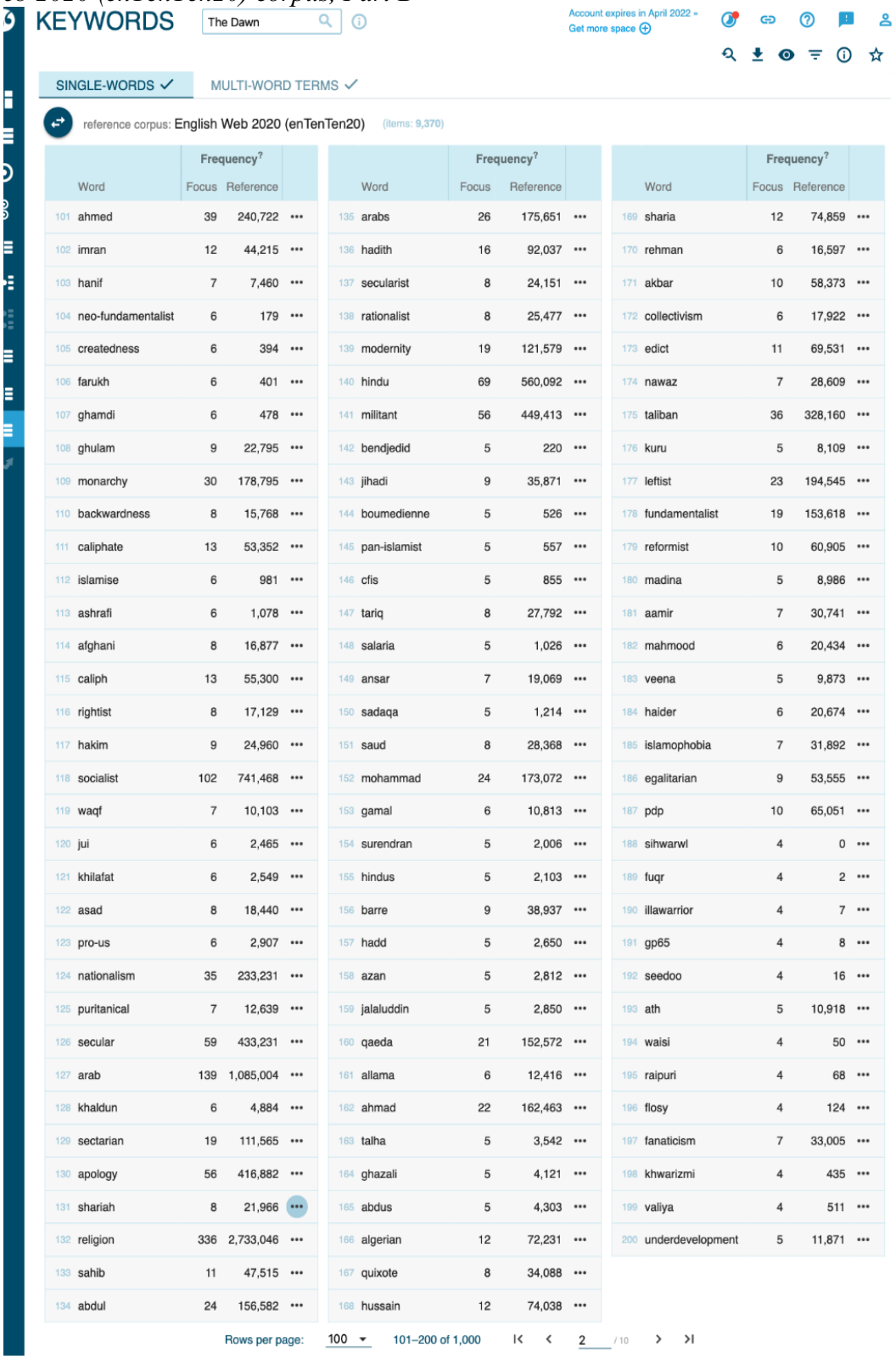
reference corpus: English Web 2020 (enTenTen20) (items: 9,370)

	Word	Frequency? Focus	Reference
1	akram	57	13,810
2	wasim	41	7,398
3	jul	312	364,202
4	islam	626	992,148
5	ulema	33	10,585
6	islamism	38	18,983
7	mullah	50	45,870
8	islamic	703	1,253,883
9	nfp	33	21,685
10	socialism	174	331,173
11	quran	102	186,129
12	muslim	775	1,715,427
13	sindhi	25	12,446
14	muslims	256	563,073
15	ummah	27	22,066
16	parvez	19	5,072
17	nasser	32	45,632
18	pakistani	114	296,383
19	sep	186	542,287
20	neo-fundamentalism	14	73
21	islamist	73	194,017
22	iqbal	25	38,815
23	nimeiry	13	576
24	saeed	24	39,153
25	non-muslims	20	26,939
26	madrassa	15	9,413
27	cii	17	17,020
28	pakistan	346	1,254,735
29	zakat	16	16,052
30	quranic	16	17,895
31	non-muslim	23	45,269
32	secularism	23	47,384
33	syed	27	63,623
34	pan-islamism	11	768
35	paracha	11	825
36	fundamentalism	25	59,184
37	abul	13	11,713
38	waseem	11	3,100
39	bhutto	17	29,351
40	recommend	1,078	4,688,314
41	abbastoronto	10	0
42	militancy	18	35,700
43	fln	11	5,706
44	nadeem	11	8,289
45	acculturate	10	4,047
46	ulema-state	9	1
47	sukarno	11	10,640
48	qutb	11	10,844
49	hoodbhoy	9	909
50	siddique	10	6,166

	Word	Frequency? Focus	Reference
51	extremism	32	117,814
52	shariati	9	1,443
53	jinnah	12	17,064
54	subcontinent	18	49,246
55	humayun	10	9,619
56	ttp	12	21,784
57	un-islamic	9	5,300
58	sadat	11	16,538
59	comparativereligionstudent	8	0
60	dăń	8	0
61	chris	8	1
62	mutazila	8	59
63	rayani	8	255
64	ubaidullah	8	629
65	aflaq	8	1,012
66	maududi	8	2,682
67	pbuh	12	26,468
68	munir	9	9,120
69	brr	8	3,551
70	fatwa	12	28,238
71	ijtihad	8	4,209
72	feb	99	555,577
73	jun	81	447,894
74	macron	18	67,342
75	shia	26	117,262
76	ramay	7	197
77	pia	16	57,527
78	jihad	30	146,916
79	jan	162	989,516
80	abbasid	9	14,614
81	malaka	7	1,751
82	karachi	21	93,841
83	adal	7	2,037
84	swt	10	22,155
85	mughal	17	68,988
86	anti-colonial	8	9,790
87	khalifa	13	44,833
88	abbasids	7	3,800
89	umayyad	9	17,892
90	nlf	7	4,136
91	ifis	7	4,436
92	muslim-majority	8	11,402
93	sunnah	10	25,781
94	maulana	10	26,198
95	crone	9	19,396
96	ideologue	11	34,052
97	algeria	27	148,497
98	salam	10	28,011
99	cleric	28	157,942
100	ifi	7	7,004

Rows per page: 100 | 1–100 of 1,000 | 1 / 10

Figure 4.26-B: top 200 keywords from The Dawn corpus compared against the English Web 2020 (enTenTen20) corpus, Part B



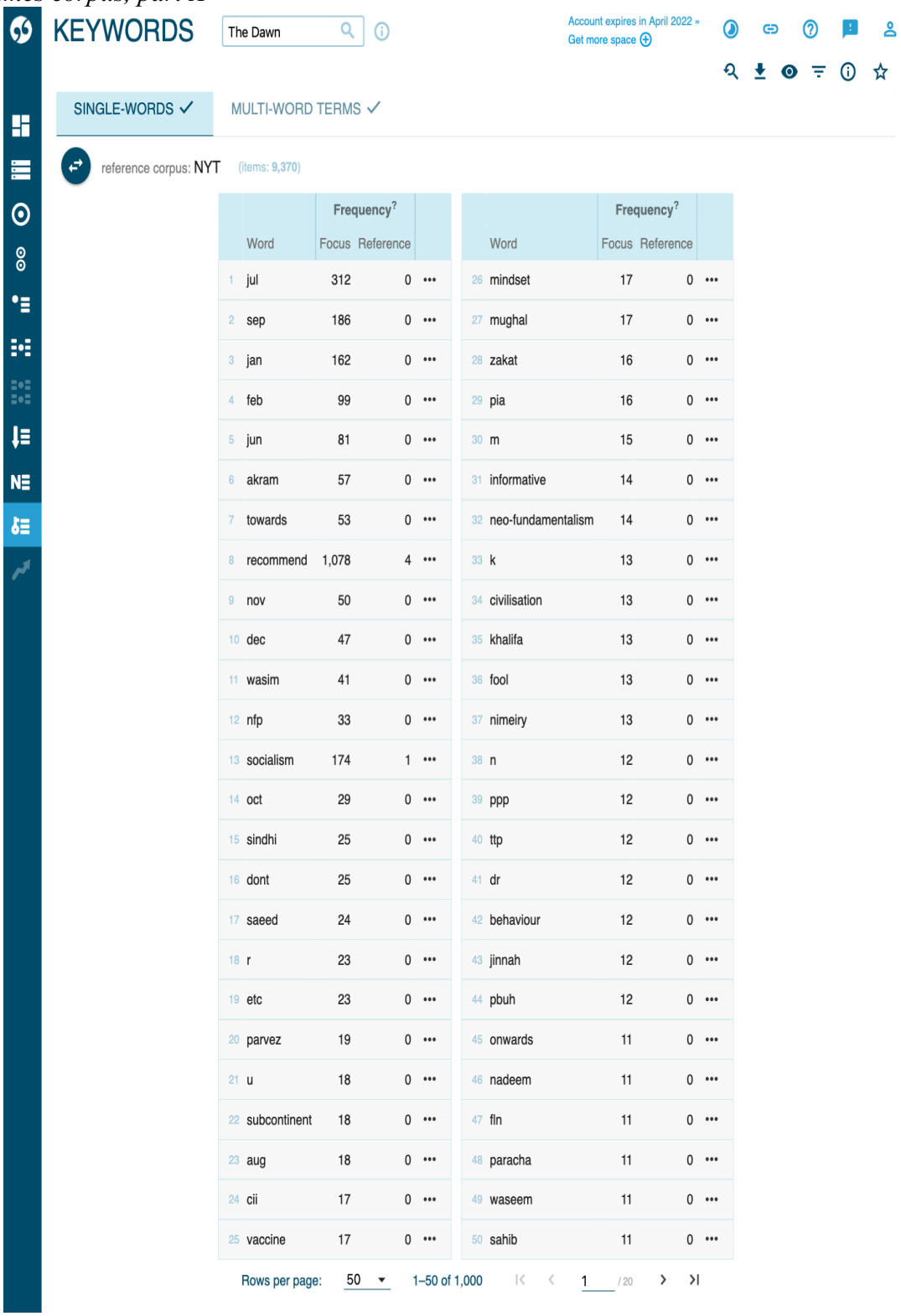
	Word	Frequency Focus	Frequency Reference
101	ahmed	39	240,722
102	imran	12	44,215
103	hanif	7	7,460
104	neo-fundamentalist	6	179
105	createdness	6	394
106	farukh	6	401
107	ghamdi	6	478
108	ghulam	9	22,795
109	monarchy	30	178,795
110	backwardness	8	15,768
111	caliphate	13	53,352
112	islamise	6	981
113	ashrafi	6	1,078
114	afghani	8	16,877
115	caliph	13	55,300
116	rightist	8	17,129
117	hakim	9	24,960
118	socialist	102	741,468
119	waqf	7	10,103
120	jui	6	2,465
121	khilafat	6	2,549
122	asad	8	18,440
123	pro-us	6	2,907
124	nationalism	35	233,231
125	puritanical	7	12,639
126	secular	59	433,231
127	arab	139	1,085,004
128	khaldun	6	4,884
129	sectarian	19	111,565
130	apology	56	416,882
131	shariah	8	21,966
132	religion	336	2,733,046
133	sahib	11	47,515
134	abdul	24	156,582
135	arabs	26	175,651
136	hadith	16	92,037
137	secularist	8	24,151
138	rationalist	8	25,477
139	modernity	19	121,579
140	hindu	69	560,092
141	militant	56	449,413
142	bendjedid	5	220
143	jihadi	9	35,871
144	boumedienne	5	526
145	pan-islamist	5	557
146	cfis	5	855
147	tariq	8	27,792
148	salaria	5	1,026
149	ansar	7	19,069
150	sadaqa	5	1,214
151	saud	8	28,368
152	mohammad	24	173,072
153	gamal	6	10,813
154	surendran	5	2,006
155	hindus	5	2,103
156	barre	9	38,937
157	hadd	5	2,650
158	azan	5	2,812
159	jalaluddin	5	2,850
160	qaeda	21	152,572
161	allama	6	12,416
162	ahmad	22	162,463
163	talha	5	3,542
164	ghazali	5	4,121
165	abdus	5	4,303
166	algerian	12	72,231
167	quixote	8	34,088
168	hussain	12	74,038
169	sharia	12	74,859
170	rehman	6	16,597
171	akbar	10	58,373
172	collectivism	6	17,922
173	edict	11	69,531
174	nawaz	7	28,609
175	taliban	36	328,160
176	kuru	5	8,109
177	leftist	23	194,545
178	fundamentalist	19	153,618
179	reformist	10	60,905
180	madina	5	8,986
181	aamir	7	30,741
182	mahmood	6	20,434
183	veena	5	9,873
184	haider	6	20,674
185	islamophobia	7	31,892
186	egalitarian	9	53,555
187	pdp	10	65,051
188	sihwarwl	4	0
189	fuqr	4	2
190	illawarrior	4	7
191	gp65	4	8
192	seedoo	4	16
193	ath	5	10,918
194	waisi	4	50
195	raipuri	4	68
196	flosy	4	124
197	fanaticism	7	33,005
198	khwarizmi	4	435
199	valiya	4	511
200	underdevelopment	5	11,871

The top 200 keywords (see Figure 4.26-A) include PERSONAL NAMES TYPICAL OF MIDDLE-EASTERN AND NORTH AFRICAN COUNTRIES (e.g. *Akram*, *Wasim*, *waseem*, *nfp*[32], *Paracha*, *Nadeem*, *Parvez*, *Nasser*, *Iqbal*, *Saeed*, *Nimeiry*, *Qutb*, *Syed*, *Abul*, *bhutto*, *sukarno*, *hoodboy*, *siddique*), POLITICS (e.g. *socialism*, *fln*[33], *militancy*), terms which refer to the ISLAMIC RELIGION AND ISLAMIC WORLD IN GENERAL (e.g. *islamism*, *islam*, *islamic*, *islamist*, *un-islamic*, *non-muslims*, *ulema*[34], *mullah*, *muslim*, *muslims*, *neo-fundamentalism*, *fundamentalism*, *quran*), GEOGRAPHICAL LOCATIONS (e.g. *subcontinent*, *algeria*, *afghani*, *arab*, *karachi*), ETHNIC GROUPS (e.g. *arabs*, *hindu*, *hindus*, *Algerian*, *Saudi*), and OTHER (e.g. *modernity*, *Recommend*, *acculturate*, *unwittingly*, *rationalist*)[35].

Figures 4.27-A-B show the top 200 keywords of *The Dawn* corpus obtained by comparing it against the *New York Times* corpus, used as a reference corpus.

[32] Nadeem Farooq Paracha is one of the most famous journalists in Pakistan, who currently writes for *The Dawn* (https://www.huffpost.com/entry/nadeem-farooq-paracha-abo_b_960013).
[33] It stands for Front de Libération Nationale (i.e. National Liberation Front, a political party in Algeria).
[34] This is the plural of *mullah*, that is, an Islamic religious teacher or leader (https://dictionary.cambridge.org/it/dizionario/inglese/mullah).
[35] For comparative purposes, I also had the top 200 keywords of the corpus automatically classified by the UCREL semantic tagger (http://ucrel-api.lancaster.ac.uk/usas/tagger.html). This identified the following semantic categories: PERSONAL NAMES (*Akram*, *Wasim*, *Quaeda*, *Waseem*, *Dan*, *salam*, *Ghazali*, *Khaldun*, *abdul*, *Talha*, *Abdus*, *Hussain*), OTHER PROPER NAMES (*NFP*, *Paracha*, *Nadeem*, *Parvez*, *Nasser*, *Iqbal*, *Saeed*, *Nimeiry*, *Qutb*, *Syed*, *Abul*, *Sindh*, *Pakistan*, *Bhutto*), GEOGRAPHICAL NAMES (*Pakistani*, *algeria*, *arab*, *Karachi*, *arabs*, *Algerian*, *Saudi*), MONEY (*ttp*), POLITICS (*socialism*, *militancy*, *extremism*, *rightist*, *socialist*, *nationalism*, *militant*, *leftist*), RELIGION (*islam*, *islamic*, *mullah*, *muslim*, *muslims*, *fundamentalism*, *fundamentalist*, *cleric*, *secular*, *religion*, *sectarian*, *hindu*, *hindus*), NUMBERS (*cii*), CRIME, LAW AND ORDER (*puritanical*, *edict*), GROUPS AND AFFILIATIONS (*Taliban*), DISCOURSE BIN (*ahmed*), POWER (*monarchy*), GOVERNMENT (*state-owned*), SPEECH ACTS (*Recommend*, *apology*), KNOWLEDGE (*unwittingly*), ABILITY (*backwardness*), SPORTS (*cricketing*), QUANTITIES (*abate*). In addition, the software could not assign a semantic category to the following words, which were therefore tagged as UNMATCHED: *Sindhi*, *Islamism*, *islamist*, *un-islamic*, *non-muslims*, *ulema*, *neo-fundamentalism*, *quran*, *ummah*, *madrassa*, *quranic*, *Secularism*, *zakat*, *non-Muslim*, *pan-islamism*, *comparativereligionstudent*, *shariah*, *neo-fundamentalist*, *Islamise*, *ideologue*, *muslim-majority*, *sunnah*, *secularist*, *jihadi*, *pan-islamist*, *sharia*, *swt*, *Pbuh*, *fatwa*, *ijtihad*, *umayyad*, *hadith*, *azan*, *Siddique*, *maududi*, *Humayun*, *sukarno*, *hoodbhoy*, *shariati*, *jinnah*, *Imran*, *sadat*, *chrs*, *mutazila*, *ramay*, *ubaidullah*, *aflaq*, *munir*, *hanif*, *malaka*, *Adal*, *saud*, *Ghulam*, *khalifa*, *Farukh*, *ghamdi*, *ashrafi*, *hakim*, *khilafat*, *asad*, *gamal*, *tariq*, *Sahib*, *Bendjedid*, *boumedienne*, *salaria*, *mohammad*, *surendran*, *barre*, *azan*, *jalaluddin*, *Allama*, *ahmad*, *Rehman*, *akbar*, *Nawaz*, *ulema-state*, *macron*, *reformist*, *anti-colonial*, *caliphate*, *caliph*, *left-leaning*, *collectivism*, *fln*, *NLF*, *abbasid*, *abbasids*, *shariat*, *Saur*, *subcontinent*, *afghani*, *shia*, *Mughal*, *modernity*, *Acculturate*, *Rationalist*, *maulana*, *crone*, *createdness*, *pro-us*, *oil-rich*, *Abbastoronto*, *rayani*, *pia*, *brr*, *IFIs*, *waqf*, *jui*, *CFIs*, *sadaqa*, *ath*, *had*.

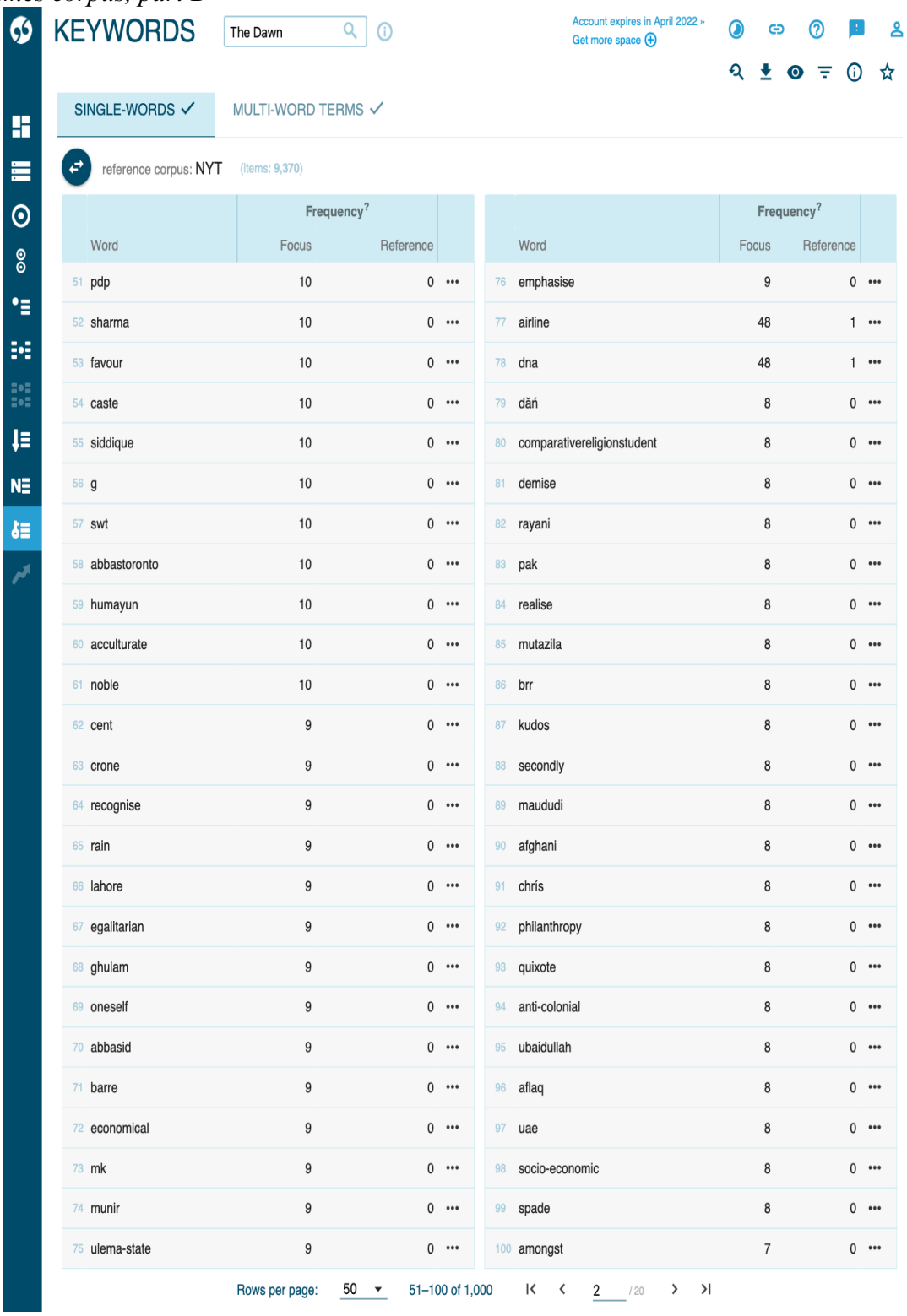
Figure 27-A: top 200 keywords from The Dawn corpus compared against the New York Times corpus, part A



	Word	Frequency? Focus	Reference
1	jul	312	0
2	sep	186	0
3	jan	162	0
4	feb	99	0
5	jun	81	0
6	akram	57	0
7	towards	53	0
8	recommend	1,078	4
9	nov	50	0
10	dec	47	0
11	wasim	41	0
12	nfp	33	0
13	socialism	174	1
14	oct	29	0
15	sindhi	25	0
16	dont	25	0
17	saeed	24	0
18	r	23	0
19	etc	23	0
20	parvez	19	0
21	u	18	0
22	subcontinent	18	0
23	aug	18	0
24	cii	17	0
25	vaccine	17	0

	Word	Frequency? Focus	Reference
26	mindset	17	0
27	mughal	17	0
28	zakat	16	0
29	pia	16	0
30	m	15	0
31	informative	14	0
32	neo-fundamentalism	14	0
33	k	13	0
34	civilisation	13	0
35	khalifa	13	0
36	fool	13	0
37	nimeiry	13	0
38	n	12	0
39	ppp	12	0
40	ttp	12	0
41	dr	12	0
42	behaviour	12	0
43	jinnah	12	0
44	pbuh	12	0
45	onwards	11	0
46	nadeem	11	0
47	fln	11	0
48	paracha	11	0
49	waseem	11	0
50	sahib	11	0

Figure 27-B: top 200 keywords from The Dawn corpus compared against the New York Times corpus, part B



	Word	Frequency? Focus	Reference
51	pdp	10	0
52	sharma	10	0
53	favour	10	0
54	caste	10	0
55	siddique	10	0
56	g	10	0
57	swt	10	0
58	abbastoronto	10	0
59	humayun	10	0
60	acculturate	10	0
61	noble	10	0
62	cent	9	0
63	crone	9	0
64	recognise	9	0
65	rain	9	0
66	lahore	9	0
67	egalitarian	9	0
68	ghulam	9	0
69	oneself	9	0
70	abbasid	9	0
71	barre	9	0
72	economical	9	0
73	mk	9	0
74	munir	9	0
75	ulema-state	9	0

	Word	Frequency? Focus	Reference
76	emphasise	9	0
77	airline	48	1
78	dna	48	1
79	dăń	8	0
80	comparativereligionstudent	8	0
81	demise	8	0
82	rayani	8	0
83	pak	8	0
84	realise	8	0
85	mutazila	8	0
86	brr	8	0
87	kudos	8	0
88	secondly	8	0
89	maududi	8	0
90	afghani	8	0
91	chrís	8	0
92	philanthropy	8	0
93	quixote	8	0
94	anti-colonial	8	0
95	ubaidullah	8	0
96	aflaq	8	0
97	uae	8	0
98	socio-economic	8	0
99	spade	8	0
100	amongst	7	0

Most of the keywords shown in Figures 27-A-B can be assigned to the same semantic fields that were previously identified by comparing *The Dawn* corpus against

the *English Web 2020 (enTenTen20)* corpus, namely PERSONAL NAMES TYPICAL OF MIDDLE-EASTERN AND NORTH AFRICAN COUNTRIES (e.g. *akram*, *wasim*, *saeed*, *parvez*, *khalifa*, *nimeiry*, *Jinnah*, *nadeem*, *paracha*, *waseem*, *sahib*, *sharma*, *crone*, *ghulam*, *Abbasid*, *barre*, *Mutazila*, *maududi*, *ubaidullah*, *aflaq*), POLITICS (*nfp*, *socialism*, *ppp*, *ttp*, *fln*, *pdp*, *anti-colonial*, *CII*[36]), terms which refer to the ISLAMIC RELIGION AND ISLAMIC WORLD IN GENERAL (*zakat*, *neo-fundamentalism*, *pbuh*), GEOGRAPHICAL LOCATIONS (*subcontinent*, *Lahore*, *UAE*[37]), ETHNIC GROUPS (*Sindhi*, *Mughal*, *afghani*) and OTHER (*towards*, *vaccine*, *mindset*, *informative*, *fool*, *dr*, *behaviour*, *onwards*, *favour*, *acculturate*, *noble*, *recognise*, *rain*, *egalitarian*, *oneself*, *mk*[38], *emphasise*, *dna*, *demise*, *realise*, *secondly*, *philantropy*, *amongst*, *cent, PIA*[39], *airline*, *rayani*). In addition, I assigned a few words to a new semantic field, namely ECONOMY and SOCIETY (*economical*, *socio-economic*, *civilisation*, *caste*). Finally, a great number of keywords belong to comments written by the readers of *The Dawn* online newspaper articles, which Sketch Engine included in the corpus (*recommend*, *Jul*, *sep*, *jan*, *feb*, *jun*, *nov*, *dec*, *oct*, *aug*, *dont [sic]*, *r*, *u*, *m*, *n*, *k*, *g*, *etc*, *siddique*, *swt*[40], *abbastoronto*, *humayun*, *munir*, *chrís*, *dan, comparativereligionstudent*, *pak*[41], *brr*, *kudos*, *quixote*, *spade*).

Then, by using the *Concordance* and *Word Sketch* functions, I analysed the collocates of the five most frequent lexical words (*islamic, islam, muslim, pakistan, religion*) and the top five keywords obtained by comparing *The Dawn* corpus against the *New York Times* corpus which did not coincide the five most frequent lexical words (*ulema, islamism, mullah, socialism, Quran*).

In order to keep the analysis of concordances to a manageable size, I considered a random sample of only 200 concordance lines for each of the above-mentioned words. When I retrieved the concordances, I sorted them alphabetically by one, two and three tokens to the left and one, two and three tokens to the right of the searched word.

Figure 4.28-A shows a sample of 50 concordance lines of the word *Islamic*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

[36] Council of Islamic Ideology.
[37] United Arab Emirates.
[38] Mujahideen-e-khalq.
[39] Pakistan International Airline.
[40] Allah.
[41] Island of Papua New Guinea.

Figure 4.28-A: 50 sample concordance lines of the word "Islamic", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islamic* is used as a premodifier of nouns (e.g. *Islamic scholars*, *Islamic finance*, *Islamic Socialism*);

2) a distinction can be made between the (groups of) words to the left of *Islamic* which carry NEGATIVE SEMANTIC PROSODY (*abolish*, *against*, *to bypass*, *denounce*, *imposition of harsh*, *negative aspects of*, *against basics of*, *go against the*, *the demise of*, *dark face of*, *had opposed*, *poor*, *vacuum in the*, *emergence of the*, *condemn un*), and which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*recognized by*[42], *very capable*, *genuine*, *glorious*, *the great*, *important in*, *a valuable part of*, *a true*), which are less frequent;

3) the (groups of) words to the right of *Islamic* can be divided into the following semantic categories: RELIGION (*Muslim and*, *Muslims as*, *many Muslims in*, *muslim nor*, *muslim History not*, *the Council of*, *theoreticians of*, *Quran and other*, *radical*, n=2: *reformist*, *traditional*), POLITICS (*republic with an*, *socialism and*, *FLN's, government on*, *politicised*, *right-wing*, *military and the*, *condemn un*, *ISIS is un*), EVALUATION (*genuine*, *glorious*, *the great*, *harsh*, *important in*, *negative aspects of*, *a valuable part of*, *poor*, *rational*, *true*, *very capable*), COMMUNICATION (*called as an*, *calling itself an*, *to be called*, n=2: *so called*, *was dubbed*, *explaining*, *not explained how*, *quote*), KNOWLEDGE (*scholar*, *basics of*, *scientific exploration in*, *Science in the*), GEOGRAPHICAL PLACES (*Arab/*, *Malaysia is an*, *Western and*, *Pakistan's*, *Middle East and the*, *all over the*), QUANTITY (*4,000*, *many*, *so much*, *a number of*, *the rest of*, *the whole*, *largest*), LAW (*abolish*, *advocating*[43]), TIME (*and 14th century*, *the 16th century*, *early 20th century*, *modern*, *modern-day*, *a half century of*, *present-day*, *whenever in the*, *today's*) and OTHER (e.g. *becoming an*, *music in an*, *suggests that an*, *themselves with an*, *views with an*, *airlines be*, *became*, *but*, *recognized by*, *to bypass*, *classical*, *country*, *country's first*, *groundwork for*, *other airlines from*, *different from*, *his power grab*, *start with implementing*, *includes*, *the name*, *noted*, *nothing*, *comprehensive analysis of*, *the demise of*).

[42] The complete sentence is '*Waqf is the permanent dedication, by a Muslim, of any property for any purpose recognized by Islamic law as religious, pious or charitable'*.
[43] The complete sentence is '*they were only interested in advocating Islamic laws, but never articulated a political plan that would carry these laws'*.

Figure 4.28-B shows a sample of 50 concordance lines of the word *Islamic*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.28-B: 50 sample concordance lines of the word "Islamic", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Islamic* which carry NEGATIVE SEMANTIC PROSODY (*and barbaric activities*, *Bomb*, *deaths*, n=8: *fundamentalism*, *is not good*, *opponents*, *separatism*, *terrorism*, *terrorist*, *warriors*, *way of punishment*) and those which carry POSITIVE SEMANTIC PROSODY (*credentials*, n=2: *Golden Age*, *great history*, *greatness*), the latter being less frequent;

2) the words to the right of *Islamic* can be grouped into the following semantic categories: RELIGION (*/Hindu/Christian/Jewish*, *and muslim*, *conservatism*, *ethics*, *evangelical*, n=8: *fundamentalism*, *hijab*, *Islam*, *Neo-Fundamentalism*, *preacher*, *theologian*, n=7: *world*), POLITICS (*about socialism*, *Council*, *Emirate*, *forces*, n=2: *ideology*, n=3: *injunctions*, *jurisprudence*, *jurists*, n=4: *law*, *leaders*, *militancy*, *militia*, n=2: *movements*, *nationalism*, *opponents*, n=2: *parties*, *political parties*, *political power*, n=6: *republic*, *Republican Party*, *restrictions*, *revolution*, *rules*, n=19: *socialism*, n=8: *socialist*, n=8: *state*, *terrorism*, *terrorist*, *way of governance*, *bomb*, *warriors*), FINANCE (*banking*, *and western business*, *business ethics*, n=2: *finance*, *financial products*), GEOGRAPHICAL LOCATIONS (*and Arabian country's*, *and western approaches*, n=7: *countries*, *Mountains*), ETHNIC GROUPS (*community*, n=3: *civilisation*, *groups*, *or Hindu*), EDUCATION (*Lectures*, *maths*, n=8: *scholar*, n=3: *teachings*, *way of teaching*), EVALUATION (*great*, *greatness*, *is not good*), TIME (n=2: *Golden Age*, n=2: *past*) and OTHER (n=2: *airline*, *and barbaric activities*, *anyway*, *calender*, *Calendar*, *credentials*, *deaths*, n=2: *heritage*, n=3: *history*, *in substance*, *justifies*, *only in form*, *option*, *or Un Islamic*, *perspective*, *society*, *texts*, n=3: *thought*, *values*, *way of life*, *way of punishment*, *windows*, *one*, n=3: *science*, *culture*, *descents*).

Figure 4.29-A shows a sample of 50 concordance lines of the word *Islam*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.29-A: 50 sample concordance lines of the word "Islam", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islam* mostly occurs in prepositional phrases (e.g. *about Islam*, *at Islam*, *between Islam*);

2) a distinction can be made between the (groups of) words to the left of *Islam* which carry NEGATIVE SEMANTIC PROSODY (n=3: *against*, *blaming*, *demonising*, *demonize*, *disrespecting*, *fighting*, *problem in*, *maligning*, *enemies of*, *negligence*, *dark ages of*), which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*integration*, *benefits that*), which are less frequent;

3) the (groups of) words to the left of *Islam* can be divided into the following semantic categories: RELIGION (n=2: *Allah and*, *Christianity and*, *conservative*, *follow*, *followers of*, *prophet of*, n=2: *religion of*, *religion*, n=2: *converted to*, *wahabi*), POLITICS (*politics around*, *justice in*, n=21: *political*), TIME (*ancient*, *early*, *golden age of*, *end of*[44], *dark ages of*), EDUCATION (*learn*, *Science and*, *education of*, *student of*, n=2: *studying*), LAW (*judge*, *advocated*, *forbidden in*, *principles of*, *tenets of*), BEHAVIOUR AND ATTITUDE (*Ignorance about*, *views about*, *behaviour as*, *laughs at*), REASONING (*distinction between*, *mocking*, *analyses of*, *distortion of*, *essence of*, *fundamental of*, *ideology of*, *knowledge of*, *perceptions of*, *questioning*) and OTHER (n=2: *given by*, *headed by*, *mandated by*, *global*, *hijacked*, *rapists in*, *liberation*, *arrival of*, *growth of*, n=3: *history of*, *manifestations of*, n=6: *in the name of*, *universality of*, *adhere to*, *indebted to*, *tribal*, *real*, *Pakistan and*).

Figure 4.29-B shows a sample of 50 concordance lines of the word *Islam*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

[44] The complete sentence is '*Centuries after the end of Islam's Golden Age (9th to 13th centuries), Arab culture is self-absorbed and centred on self-congratulation.*

Figure 4.29-B: 50 sample concordance lines of the word "Islam", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Islam* which carry NEGATIVE SEMANTIC PROSODY (*as a weapon*, *Authoritarianism and Underdevelopment*, *are illogical*, *largely failed*, *suggest severe punishment*), and those which carry POSITIVE SEMANTIC PROSODY (*as a unifying common denominator*, *compatible*, *is an attractive*, *is the best religion*, *is the most perfect religion*, *requires equity*), the latter being more frequent;

2) the (groups of) words to the right of *Islam* can be divided into the following semantic categories: RELIGION (*Hinduism*, *Messianic Judaism*, *and Islamism*, *and its followers*, *from Hinduism*, *from non-Muslims*, *in radicalizing*, *Prophet*, *scriptures*, *social and spiritual*, *through Shar'ia*, *was inherently secular*, *was preached*), POLITICS (*"Militant"*, *Democracy*, *a political ideology*, n=2: *and Communism*, *compatible with socialism*, *political and social*, *political parties*), TIME (*during his entire lifetime*, n=2: *Golden Age*, *in the 7th century*, *in the daily*, n=2: *rational past*, *till 1258*, *today*), GEOGRAPHICAL LOCATIONS (*western nations*, *while Europe*, *in the West*) and OTHER (*allows you*, *also developed*, *as a cultural heritage*, *came*, *can find*, *forbids*, *is an utopian*, *main mission*, *will not weaken*, *and its injunctions*, *at the University*).

Figure 4.30-A shows a sample of 50 concordance lines of the word *Muslim*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.30-A: 50 sample concordance lines of the word "Muslim", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Muslim* is mostly used as a premodifier of nouns (e.g. *Muslim country*, *Muslim name*, *Muslim minorities*);

2) as in the *New York Times* corpus, there is a frequent use of third person pronouns and possessives to the left of the word *Muslim* (e.g. *their own Muslim members*, *they are not the right kind of Muslims if they don't care for other Muslims first*, *people calling themselves Muslims*). On the other hand, contrarily to the *New York Times* corpus findings, no use of first person pronouns and/or other determiners referring to *Muslim* was found;

3) a distinction can be made between the (groups of) words to the left of *Muslim* which carry NEGATIVE SEMANTIC PROSODY (*against*, *against the*, *brutal*, *destroyed by*, *disarray of*, *oppressed*, *selfish*, *stagnation of the*), which are more frequent, and those which carry POSITIVE SEMANTIC PROSODY (*greeted by*, *creative*, *I love*, *successful*, *intellectually superior*, *Unity of the*), which are less frequent;

4) the (groups of) words to the left of *Muslim* can be classified into the following semantic categories: RELIGION (*Islamic or*, *Alawite*, *Hindus*, *hindu*, *Islamic scholars and*, *conservative*, *extremism in*, *liberalism in*, n=2: *liberal*, *radicals*, *radicalization as the*, *secular*), ETHNIC GROUPS (*Arab*, *non-Arab*, *Bangladeshi*, *Bharatiya*, *Danish*, *French*, n=5: *Indian*, *Pakistani*, *somali*, n=2: *Egypt's*, *Germany's*, *a group of*), QUANTITY (*30%*, *80%*, *one-fifth of 49*, *75*, *the entire*, n=5: *every*, *several major*, n=2: *many*, *200 million*, n=3: *most*, *mostly*, *70% of*, *lots of*, *small*, *some*, *most of the*, *various*), POLITICS (*nations are*, *anarchy in*, *left-leaning*, *militant*, *colonization of*), TIME (*last two decades*, *in 1947 as*, *early*), COMMUNICATION (*call*, *to call all*, *be called*, *to call on*) and OTHER (*become*, *inhabited by*, *invented by*, *famous*, *first*, *turf for*, *influences in*, *inspire*, *judge*, *leading*, n=3: *non*, *architects of*, *imagination of*, *movement of*, *reports of*, *ordinary*, *other*, *rest of the*, *their own*, *real*, *separate*, n=2: *process of the*, *traditional*, *treat*, *urged*, *boys*, n=3: *educated*).

Figure 4.30-B shows a sample of 50 concordance lines of the word *Muslim*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.30-B: 50 sample concordance lines of the word "Muslim", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Muslim* which carry NEGATIVE SEMANTIC PROSODY (*against anyone*, *against other non-Muslims*, *approving terrorism*, *are killing muslims*, *are not intelligent*, *have grievances*, *invasion*, *not killing*, *revolt*, *stop killing Muslim*, *terrorists*, n=2: *were killed*) and the only word which carries POSITIVE SEMANTIC PROSODY (*hero*), which is much less frequent;

2) the (groups of) words to the right of *Muslim* can be divided into the following semantic categories: RELIGION (*jew or Christian*, *the practitioner*, *and fundamental*, *and non-muslims*, *are converts*, *is the ex-Muslim*, *or Hindus*, *or Islam*, *ulema*, *ullema*, n=2: *ummah*, n=14: *world*), POLITICS (*liberal or hardliner*, n=2: *leader*, n=2: *leaders*, *military empires*, *military leaders*, *monarchies*, *movement*, n=6: *nationalism*, *nation's modernization movement*, *nations*, *polity's*, *regime*, *regimes*, *Revolt*, *rights*, *rulers*, *Socialist community*, *states*, *supporters of IS*, *terrorists*, *thinker/leader*), GEOGRAPHICAL LOCATIONS (*around the world*, n=16: *countries*, n=4: *country*, *from west*, *in Nigeria*, *in Pakistan*, *in the East*, *minorities in India*, *of subcontinent*, *outside Middle East*, *of our country*), PROFESSIONS (*farmers*, n=2: *female flight attendants*, *merchants*, *philosopher and poet*, *scientist*, *scientists*, *traders*), SOCIETY (*communities*, n=5: *community*, *members*, n=2: *population*, n=2: *populations*, *sects*, *women*, *societies*, *migrated out*, *refugees*), ATTITUDE (*do not care*, *don't need*, *obsessed*, *stand up against*, *will get respect*), INSTITUTIONS (n=3: *Brotherhood*, *Council*, *League*), FINANCE (*businessmen*, *cannot deal*), EDUCATION (n=4: *scholar*, *texts*), TIME (*of 7th century*, *tomorrow*) and OTHER (*For them*, *cool*, *histories*, n=2: *history*, *identity*, *in letter and spirit*, *Mind*, *Mindset*, *name*, *polymaths*, *to evolve*, *of my country*).

Figure 4.31-A shows a sample of 50 concordance lines of the word *Pakistan*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.31-A: 50 sample concordance lines of the word "Pakistan", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Pakistan* is mostly used as a noun (e.g. *Pakistan is 100 percent right*, *Pakistan and India comparison*, *Pakistan is more damaged*, *within Pakistan*), but in some cases it is also used as premodifier (e.g. *Pakistan peoples*, *Pakistan government*, *Pakistan Bengalis*, *Pakistan cricket captain*);

2) a distinction can be made between the (groups of) words to the left of *Pakistan* which carry NEGATIVE SEMANTIC PROSODY (*blame*, *defaming*, *Islamic indoctrination destroyed*, *Islamic nationalism dismembered*, *a problem for*, *arrested in*, *16 attacks in*, *danger in*, *discrimination in*, *havoc in*, *terrorism in*), and those which carry POSITIVE SEMANTIC PROSODY (*may God bless*, *bright*, *advancement in*, n=2: *Tolerant societies like*, *good citizen of*, *friends of*, *progress of*, *American support to*), the latter being less frequent;

3) the (groups of) words to the left of *Pakistan* can be classified into the following semantic categories: COUNTRIES (*Saudi*, *Saudi Arabia and*, *Arab countries and*, *Arab world and*, n=3: *India and*, *Iran and, Syria*, *US*, *United States/*, *Nigeria*, n=2: *Afghanistan*, *Muslim countries e.g.*, n=3: *countries like*, n=3: *country like*, *the Nation of*, n=2: *Islamic Republic of*, n=2: *East*, *everywhere in*, *Here in*, *West*), POLITICS (*governments in*, *political islam in*, n=2: *militancy in*, *Left parties in*, *Socialists in*, *populist*, *Pakistan's ruling*, *nationalists to serve*, *massive sanctions on*), RELIGION (*blasphemy law in*, *Deobandi movement in*, *Muslims in*, *the muslims of*, *their religion in*, *Islamic*), ETHNIC GROUPS (*Tehreek-e-Taliban*, *Hindus in*), TIME (n=2: *present day*, *today's*), SCHOLARSHIP (*Science*, *scientists from*, *university in*, *leading intellectual of*) and OTHER (*blinded*, *eluded*, *care for*, *a need for*, *the former*, *immigration from*, *pronouncement from*, *dormant in*, *gathering in*, *Dr Hoodboy in*, *saviors in*, *law*, *native*, *creation of*, n=2: *founder of*, *social history of*, n=3: *idea of*, *major issue of*, *nobody outside*, *stop*, *take*, *much better than*, *even bigger than*, *other than*, *behind the*, *under the*, *back to*, *belong to*, *came to*, *come to*, *migrated to*, *tour*, *wants*, *relations with*, *headlines within*, *no wonder*).

Figure 4.31-B shows a sample of 50 concordance lines of the word *Pakistan*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.31-B: 50 sample concordance lines of the word "Pakistan", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Pakistan* which carry NEGATIVE SEMANTIC PROSODY (*against minorities*, *everyone is poor*, *has been suffering*, *has no future*, *is more damaged*, *lawlessness is allowed*, *problems*, *who failed*) and those which carry POSITIVE SEMANTIC PROSODY (*is 100 percent right*, *will shine*), the latter being less frequent;

2) the (groups of) words to the right of *Pakistan* can be divided into the following semantic categories: COUNTRIES (*Algeria*, *Egypt*, *India*, *Saudi Arabia*, */Afghanistan*, */Bangladesh*, */Canada*, */Saudis*, *and Afghanistan*, n=3: *and India*, *and Iran*, *and Iraq*, *and other Muslim countries*, *and the mighty USSR*, *or any other muslim country*, *or any other muslim countries*, *or India or Afghanistan*, *or Pakistan*, *too was India*, *with UK*), POLITICS (*Awami National Party*, *Islamic Salvation Army*, *People's Democratic Party*, *the government*, *the great Islamic republic*, *agenda in Afghanistan*, *government*, *Islamic parties*, *Jamaat-e-Islami party*, *on a manifesto*, *Penal Code*, n=6: *Peoples Party*, *politics*, *ruling Pakistan Peoples Party*, *PLO*[45]), TIME (*in 1971*, *during the 90's*, *from about 1980 onwards*, *in 1947*, *in 1952*, *today*, *was before 1971*), RELIGION (*promoted Taliban*, *regarded jihadis*, *and Islam*) and OTHER (e.g. *End of the Past*, *saw the rise*, *shows knowledge*, *allows*, *along with other*, *attempt*, *call themselves*, *cannot co-exist*, *cannot grant them*, *cricket captain*, *decision*, *for support*, *founder*, *founding fathers*, *from progressing*, *functions as*, *had to counter*, *has asked for*, *has been for long*).

Figure 4.32-A shows a sample of 50 concordance lines of the word *religion*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

[45] It stands for *Palestine Liberation Organization*.

Figure 4.32-A: 50 sample concordance lines of the word "religion", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the noun *religion* occurs in noun phrases (e.g. *the religion*, *a religion*, *their religion*) or prepositional phrases (e.g. *with religion*, *of religion*, *if religion*);

2) the word *religion* is frequently found with words belonging to such semantic fields as politics and science, the concepts that they denote being presented in opposition to one another (e.g. '*science must be separated from Religion*'; '*politics separately from religion*'; '*keep religion out of politics*'; '*stop mixing religion to everything*');

3) a distinction can be made between the (groups of) words to the left of *religion* which carry NEGATIVE SEMANTIC PROSODY (*lame and*, *can't blame*, *media blamed*, *defaming*, *has ditched*, *critical issue like*, *misusing*, *hatred of*, *in the light of*, *I am tired of*, *poor*, *blames all ills on the*, *politics strangulates the*), and those which carry POSITIVE SEMANTIC PROSODY (*a good*, *all great*, *our great*, n=2: *natural*, *protect*, *all glory to the*, *promoting their*), the latter being less frequent;

4) the (groups of) words to the left of *religion* can be classified into the following semantic categories: RELIGION (*in Islam*, *Islam is not only a*, *God of a*, *monk*, *spiritual aspects of*, *last Prophet the*, *preaching their*), EVALUATION (*a good*, *all great*, *our great*, *true*), QUANTITY (n=3: *one*, *across all*, n=5: *all*, n=4: *any*, *the whole*, n=3: *every*), PERSONAL PRONOUNS AND POSSESSIVES (*they brought*, *their former*, n=10: *other*, n=2: *our*, *your own*, *our own*, *bringing their*, *to practice their*, *wear their*, n=2: *your*), EDUCATION (*teaching*, *interpreting the*, *culture and*, *Science without*), POLITICS (*state and*, *socio-economic ideologies are*) and OTHER (n=51; e.g. *commenting about*, *logics against*, *Ahmedi*, *allowing*, *apply any*, *conflating*, *intertwining*, *formula*, *divorced from*, *separated from*, n=2: *separately from*, *compulsion in*, *interest in*, *including*, *how important is*, *attempt to judge*, n=6: *keep*, *the main*, *mix*, *stop mixing*, *does not need*, *a new*, *basis of*, *branding of*, *contribution of*, *Freedom of*, *mosaics of*, n=3: *in the name of*).

Figure 4.32-B shows a sample of 50 concordance lines of the word *religion*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.32-B: 50 sample concordance lines of the word "religion", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *religion* which carry NEGATIVE SEMANTIC PROSODY (*as a weapon*, *beget violence*, *failed to become*, *for problems*, *is based on hocus pocus*, *is close to insulting*, *is divisive*, *is dragged into debate*, *is lame*, *is obscuring*, *with violence*), and those which carry POSITIVE SEMANTIC PROSODY (*has philanthropic components*, *innate ability to inspire*, *is fine*, *of peace*, *preach good things*, *should unite people*, *sound heavenly*, *too preaches respect for women*, *who raised awareness*), the latter being less frequent;

2) the (groups of) words to the right of *religion* can be classified into the following semantic categories: RELIGION (*not Islam only*, *and Islam*, *worship*, *and different sects*, *and the religious*, *belief*, *but muslim history*, *from all worldly life*, *from non-believers*, *if followers becomes*, *Islam*, *like other religions*, n=3: *of Islam*, *particularly Hinduism*, *preach good things*, *such as Judaism*, *through its followers*, *to be ethical*, *too preaches respect*, *color or creed*, *according to beliefs*), POLITICS (*not a political system*, *above nationalism*, *and politics*, *and state*, *are also socio-economic systems*, *as it's state policy*, *because politics*, *but as a political system*, *for political identity*, *from politics*, n=2: *from the state*, *into politics and governance*, *just as politics*, *or politics*, *out of politics*, *within state*, *are tied to economy*), SCIENCE (n=2: *and science*, *in science*, *restricting science*, *than his science*, *without* science), EVALUATION (*is great*, *lame*), GEOGRAPHY (*in Pakistan*, *of India*, *according to region*), QUANTITY (*most of them*), CULTURE (*culture*, *culture and ethnicities*), EDUCATION (*outside the classroom*, *should be taught*) and OTHER (n=53; e.g. *as a private affair*, *as a shield*, *but me myself*, *but open and social*, *can be strong*, *comes*, *did not invent*, *due to our obsession*, *enter into any*, *everywhere*, *for their benefits*, *got to do*, *has brought*, *has to be brought*, *has nothing to do*, *ideas of morality*, *in day*, *is a personal matter*, *is not a personal affair*, *is your private affair*, *is for the heart*, *is generally used*, n=2: *is not an easy task*, *is one's personal matter*, *is something*).

I performed the same kind of analysis with the top five keywords, excluding proper names of individuals. Figure 4.33-A shows the concordance lines of the 33 occurrences of the word *Ulema*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.33-A: list of 33 concordance lines of the word "Ulema", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Ulema* is mostly used as a head noun (e.g. *the Ulema*, *such Ulemas*, *the global Muslim Ulema*);

2) to the left of *Ulema* it was possible to identify one group of words carrying POSITIVE SEMANTIC PROSODY (*a hopeful future*), but none with NEGATIVE SEMANTIC PROSODY were identified;

3) the (groups of) words to the left of *Ulema* can be classified into the following semantic categories: RELIGION (*Muslim kings and*, n=7:

conservative, *different Islamic*, *the merchant-Shia*, *Shia*, *the global muslim*, *between muslim*, *clergy and the*), POLITICS (*Islamic parties and*, *alliances with the*) and OTHER (*many noted*, n=2: *our*, *our traditional*, *so-called*, *such*, *between the*, *under the*, *the real*, n=3: *scholars*, *Intellectual*, *ancient*).

Figure 4.33-B shows the concordance lines of the 33 occurrences of the word *Ulema*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.33-B: list of 33 concordance lines of the word "Ulema", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) one group of words to the right of *Ulema* with NEGATIVE SEMANTIC PROSODY was identified (*have hijacked Islam*), while no (groups of) words with POSITIVE SEMANTIC PROSODY were identified;

2) the (groups of) words to the right of *Ulema* can be divided into the following semantic categories: RELIGION (n=2: *and clergymen*, *clerics*, n=2: *the clergy and islamic evangelists*, */Pope/Pandits/Rabbis*, *and clergymen*, *and radical Islamists*, *developed ilm al-kalam*, *have hijacked Islam*), POLITICS (*instead of ulema-state*, *and Arab monarchies*, *and the military state*, *and the Mughal kings*, *had accepted socialism*) and OTHER (*Emerged*, *Rasikhona*, *accountable*, *and oil rents*, *covers all humans*, *influence*, *on these topics*, *should coordinate*, *should try*, *to say something*, *were used to be*, *who accused*, *who were keeping*, *were financed*).

Figure 4.34-A shows the concordance lines of the 38 occurrences of the word *Islamism*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.34-A: list of 38 concordance lines of the word "Islamism" in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Islamism* is always used as a head noun and mostly occurs in prepositional phrases (e.g. *with Islamism*, *to Islamism*, *of Islamism*, *in Islamism*);

2) no (groups of) words with clearly negative or positive semantic prosodies were identified;

3) the (groups of) words to the left of *Islamism* can be divided into the following semantic categories: POLITICS (*modern political context*, *political ambitions of*, *leading exponent of*, n=2: *forces of*, *the political goals of*), TIME (*20th century*, *modern-day*), RELIGION (*Islam and*) and OTHER (e.g. *the term*, *consequently*, *articulating as*, *remains to be*, *between*, *proclaimed by*, *defined*, *its fold*, *experiments in*, *noted modern*, *advocates of*, *'universal' ideals of*, *theory of*, *early*).

Figure 4.34-B shows the concordance lines of the 38 occurrences of the word *Islamism*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.34-B: list of 38 concordance lines of the word "Islamism", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Islamism* which carry NEGATIVE SEMANTIC PROSODY (*is also invalid*, *scorned at them*, *started to weaken and fragment*), and those which carry POSITIVE SEMANTIC PROSODY (*successfully steered the revolution*), the latter being less frequent;

2) the (groups of) words to the right of *Islamism* can be divided into the following semantic categories: RELIGION (n=2: *and 'Neo-Fundamentalism'*, *Islamic Neo-Fundamentalism*, *the politicised clergy*), POLITICS (*and Islamic Socialism*, *forces*, *leading organs*, *only tangible political enactment*), NAMES TYPICAL OF ISLAMIC COUTRIES (*such as Jamaat-i-Islami*, *Morsi*), ETHNIC GROUPS (*among the Shia Muslims*, *groups*) and OTHER (e.g. *settling instead*, *a must read*, *began almost immediately*, *came to explain*, *is the new opium*, *many notions*).

Figure 4.35-A shows the concordance lines of the 50 occurrences of the word *Mullah*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.35-A: list of 50 concordance lines of the word "Mullah", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Mullah* is used as head noun and frequently occurs in prepositional phrases (e.g. *of the Mullahs*, *by the Mullahs*, *to such Mullahs*, *of Mullahs*);

2) the following (groups of) words with NEGATIVE SEMANTIC PROSODY (*terrorism and*, *our main problem is*, *mad*, *closed minded*, *mindless*, *goaded on my*, *confronting the*, *crush these*) were identified, while no words with POSITIVE SEMANTIC PROSODY were identified;

3) the words to the left of *Mullah* can be divided into the following semantic fields: POLITICS (*the army and*, *is run by*, *militant*, *political*, *powers to such*), TIME (*the 18th century*, *since the*, *today's*), COMMUNICATION (*the so called*, *never questioned*, *presented by the*, *the narrative of the*) and OTHER (*Iqbal Aur*[46], *through bought*, *this is not only for*, *these are just*, *when these*, *I know*, *freak nature of*, *has lost and the*, *not because the*, *something the*, *gave it to the*, *to treat the*, *traditional*, n=2: *typical*, *clerics and*, *bin Laden or*, *in Pakistan but the*).

Figure 4.35-B shows a list of 50 concordance lines of the word *Mullah*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

[46] This is part of the book title '*Iqbal Aur Mullah*'.

Figure 4.35-B: list of 50 concordance lines of the word "Mullah", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) a distinction can be made between the (groups of) words to the right of *Mullah* which carry NEGATIVE SEMANTIC PROSODY (*to exploit the Muslims*, n=3: *manipulations*, *running havoc in the country*), which are more frequent, and the only one which carries POSITIVE SEMANTIC PROSODY (*have won*);

2) the (groups of) words to the right of *Mullah* can be classified into the following semantic categories: RELIGION (n=2: *and madrasas*, *nor a religious scholar*, *through bringing religion*, *will do to islam*), POLITICS (*and Military community*, *and politicians*, *to advance Pakistan's interests*, *took over*, *will take over*), THE ISLAMIC WORLD (*Khaled Husseni*, *Khalifa*, *Fazlullah*, n=3: *Omar*, *Omer*, *in case of Pakistan*, *in Pakistan*, *like the Taliban*) and OTHER (*the world nods in agreement*, *are only making*, *as an equal*, *do this to earn*, *for plundering*, *roaming*, *screams the loudest*, *should be looking*, *status*, *to me*, *who are put*, *who profusely quote*, *who want to take*, *will*, *wouldn't have it*, *in 1972-77*).

Figure 4.36-A shows a sample of 50 concordance lines (out of 174 total concordance lines) of the word *socialism*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

Figure 4.36-A: sample of 50 concordance lines of the word "socialism", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *socialism* is often a head noun in prepositional phrases (e.g. *by Islamic socialism*, *with socialism*, *higher than socialism*);

2) a distinction can be made between the (groups of) words to the left of *socialism* which carry NEGATIVE SEMANTIC PROSODY (*critique of*, *failure of*, *the unsuccessful rein of*), and those which carry POSITIVE SEMANTIC PROSODY (*fair distribution of wealth*, *ulema had accepted*, *Islam does supports*), both of them being equal in number;

3) the (groups of) words to the left of *socialism* can be divided into the following semantic categories: POLITICS (n=2: *capitalism*, n=2: *communism*, n=2: *Marxism*, *capitalism and*, n=16: *Ba'ath*[47], *corporate*, *national*, *PPP's*, *Marx predicted that*, n=2: *justice with*, *NFP claim that*), RELIGION (*religious colours*, *Islamic about*, *ulema had accepted*, *Islam and*, n=70: *Islamic*, n=2: *Islam compatible with*, *Islam has nothing to do with*, *equate Islam with*), GEOGRAPHICAL LOCATIONS (*Algerian*, n=16 : *Arab*, *Sudanese*, *Muslim countries where*), KNOWLEDGE (*science*, *scientific*, *studying*) and OTHER (*talking about*, *how can*, *is claiming*, *fusing*, *have believed in*, *and what is*, *to marry*, *to mix*, *the cause of*, *concepts of*, *this form of*, *second stage of*, *it is only*, *different than*, *far higher than*, *I deny that*, *Waisi defined their*, *Arab attached to*, *rich tribute to*, *many ways*, *whether*, *the words*).

Figure 4.36-B shows a sample of 50 concordance lines of the word *socialism*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

[47] 'Designating or of a political party formed in Syria in the 1940s: separate factions of the party have ruled Syria since 1970 and ruled Iraq from 1968 to 2003' (https://www.collinsdictionary.com/dictionary/english/baath).

Figure 4.36-B: sample of 50 concordance lines of the word "socialism", in the Dawn corpus, sorted by one, two and three tokens to the right

The analysis of the concordances led to the following observations:

1) three groups of words to the right of *socialism* with NEGATIVE SEMANTIC PROSODY (*aka Nazism*, *are not compatible with Islam*, *as harnessing the freedom of thought*) were identified, while no (groups of) words with POSITIVE SEMANTIC PROSODY were identified;

2) the (groups of) words to the right of *socialism* can be classified into the following semantic fields: POLITICS (*and 'Arab Nationalism'*, n=3: *and Ba'ath Socialism*, n=2: *'Ba'ath Socialism'*, */Ba'ath Socialism*, fascism, *Marxism*, *Egyptian leader*, *aka Nazism*, n=2: *and Arab nationalism*, *and Islamic Socialism*, *and Marxism*, *appealed to the unity*, *but justice*, *A History from left to right*, *became Iraq's central ideology*), RELIGION (*means an atheistic philosophy*, *whether Islamic or otherwise*, *and Islam*, *and Islamic liberals*, *and Liberal Islam*, *and secularism*, *and then between Islamism*, *as an atheistic concoction*, *best enforces Qur'anic dictums*), ECONOMY (*an economic concept*, n=3: *Capitalism*, *and Capitalism*, *which is an economic system*, *and Islamic economic order*), GEOGRAPHICAL LOCATIONS (*in Iraq*, *Syria and Egypt*, *and Western concepts*, *appeared in Russia*, *especially the Russian version*), NAMES COMMON AMONG PEOPLE FROM ISLAMIC COUNTRIES (*F. Rahman*, *NA Jawad*, *Nasser*[48]) and OTHER (*Should think again*, *(and its manifestations)*, *as an ideology*, *well worth reading*, *A documentary survey*, *A term first used by*, *Achievements*, *an ideology*, *are of the view that*, *and fiery polemics*, *and the egalitarian tenants*, *an essential part of*, *as meaning*, *began to wither*).

Figure 4.37-A shows a sample of 50 concordance lines (out of 102 total concordance lines) of the word *Quran*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the left.

[48] Gamal Abdel Nasser (1918-1970), 'Egyptian army officer, prime minister (1954-56), and then president (1956-70) of Egypt' (https://www.britannica.com/biography/Gamal-Abdel-Nasser).

Figure 4.37-A: sample of 50 concordance lines of the word "Quran", in the Dawn corpus, sorted by one, two and three tokens to the left

The analysis of the concordances led to the following observations:

1) the word *Quran* is often used as a head noun in prepositional phrases (e.g. *in the Quran*, *of the Quran*);

2) to the left of *Quran* only (groups of) words which carry POSITIVE SEMANTIC PROSODY were identified (n=3: *the Holy*, n=6: *holy*, *The Holy Book*, *noble*, *the miracle of the*);

3) the (groups of) words to the left of *Quran* can be divided into the following semantic categories: STUDY AND SCIENCE (*interpretation of the*, *when you read*, *understanding of the*, *understanding of the*, *understand the*, *read*, *message of the*, n=2: *answers in*, n=3: *mentioned in the*, *teachings of*, *reading the, science is in the*, *nitrogen in the*, *scientific facts in the*, *for science in the*), RELIGION (n=3: *the Holy*, n=6: *Holy*, *The Holy Book*, guidance *is in the*, *First commandment of Al*, *sins that the*, *the miracle of the*, n=5: *createdness of the*, *creator and the*, *quote from the*, *instructions given in the*, *quote the*, *memorizing*, *ignorance of the*), QUANTITY (*whole*, *part of the*, *essential part of*) and OTHER (*proof from the*, *directly from the*, *reflection of the Al*, *Hidayat in*, *provisions of the*, *provisions given in the*, *as an example the*, *there in the*, *regarding rain cycle the*, *in this world and*, *given in the*, *including the*, *contradict anything in*, *accepted what*, *according to the*, *to equate*, *was already in*, *predicted in*, *power of*).

Figure 4.37-B shows a sample of 50 concordance lines of the word *Quran*, in *The Dawn* corpus, sorted alphabetically by one, two and three tokens to the right.

Figure 4.37-B: sample of 50 concordance lines of the word "Quran", in the Dawn corpus, sorted by one, two and three tokens to the right

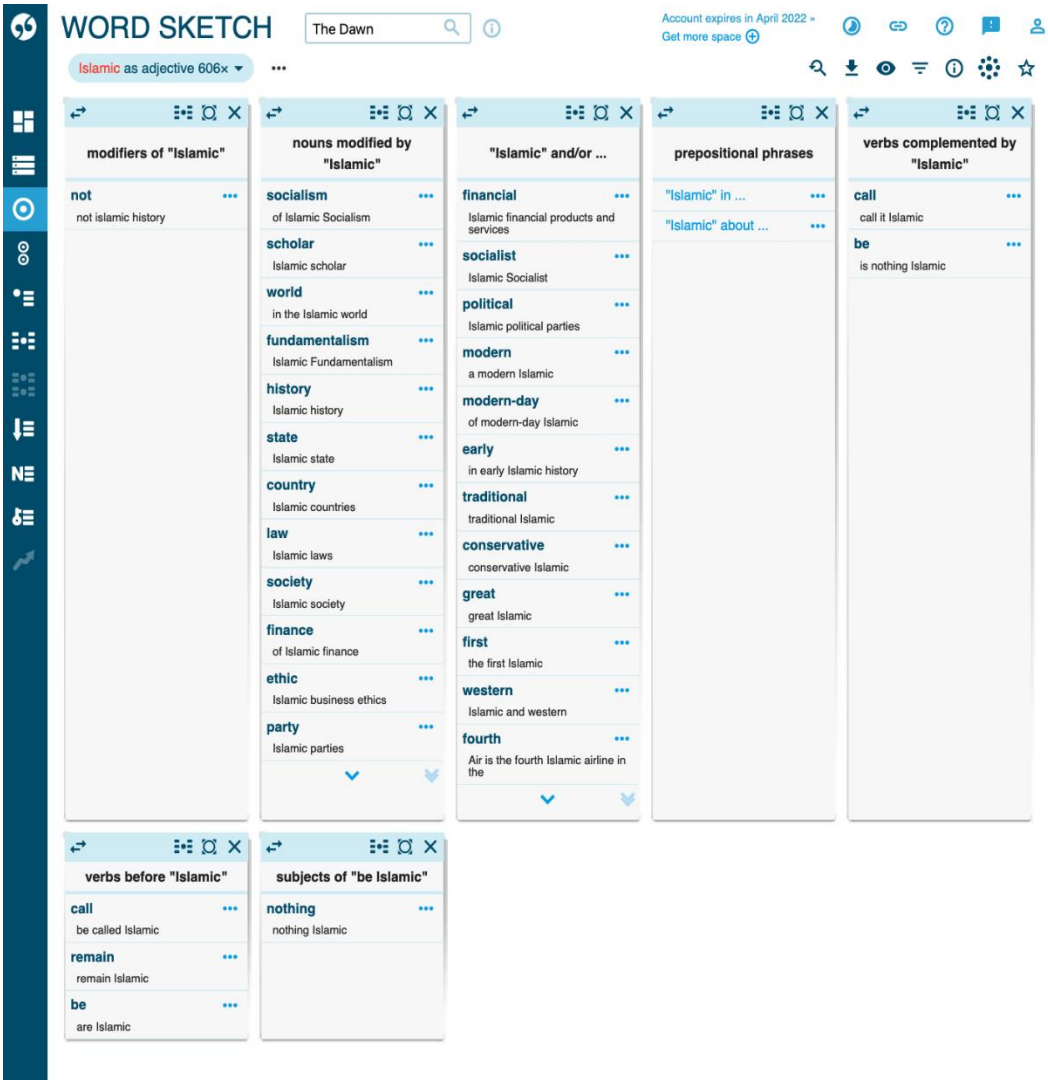
The analysis of the concordances led to the following observations:

1) to the right of *Quran* only one group of words which carry POSITIVE SEMANTIC PROSODY (*is a mercy on mankind*) was identified;

2) the words to the right of *Quran* can be assigned to the following semantic fields: RELIGION (*the hadith and Shariah*, *and Hadith*, *and the Hadiths*, *and Shariah*, n=4: *and Sunnah*, n=3: *and the Sunnah*, n=2: *or Sunnah and Hadees*, *(Zakat) charity*, *and all Muslims*, *has many verses*, *and abstain*, *and the traditions*), KNOWLEDGE (*is totally based on Education*, *The above verse*, *a book of science*, *with science*, *is to seek knowledge*, *also emphasizes*, *basic's emphasis*, *lays great emphasis*), JUSTICE (*concept of justice*, *distributive justice*, *justice*, *code of justice*), COMMUNICATION (*asks people*, *clearly mentioned*, *clearly talks*, *does not talk about*, *explains that its message*, *has warned against*, *not say that revelation*, n=4: *says*, *promised*, *pronounces*), QUANTITY (*is not enough*, *is not the only*, n=2: *is replete with*) and OTHER (e.g. *all the Ahadis*, *also uses*, *and become*, *and one can find*, *and the actions*, , *even goes to the extent*).

The *Word Sketch* function in Sketch Engine will be used to corroborate my analysis of the collocates of the five most frequent words and the top five keywords in *The Dawn* corpus. Figure 4.42 shows part of the *Word Sketch* of the word *Islamic* in *The Dawn* corpus.

Figure 4.42: partial Word Sketch of "Islamic" in The Dawn corpus



The analysis of the *Word Sketch* led to the following observations:

1) most of the collocates of the word *Islamic* can be assigned to the same semantic fields identified through my concordance analysis, namely RELIGION (*fundamentalism*, *fundamentalist*, *ethic*, *extremism*, *evangelist*, *conservative*, *radical, sect*), POLITICS (*socialism*, *socialist*, *democracy*, *state*, *party*, *nation*, *system*, *movement*, *republic*, *regime*, *nationalism*, *rule*, *emirate*, *revolution*, *leader*, *political*, *right-wing*), KNOWLEDGE (*scholar*, *teaching*, *text*, *thought*, *history*, *science*, *math*, *intellectualism*, *symbolism*, *study*, *notion*, *knowledge*, *education*, *culture*), TIME (*past*, *modern*, *modern-day*, *present-day*, *early*, *ancient*), FINANCE (*finance*, *product*, *institution*, *banking*, *financial*), SOCIETY (*group*, *society*, *civilization, world*), EVALUATION (*great*, *glorious*, *true*), COMMUNICATION (*so-called*, *call*), GEOGRAPHICAL PLACES

(*country*, *western*), LAW (*law*, *injunction*, *jurist*), CULTURE (*value*, *principle*, *ideology*, *heritage*, *traditional*,), and OTHER (*way*, *airline*, *window*, *branch*, *credential*, *outfit*, *expert*, *code*, *identity*, *thinker*, *power*, *first*, *fourth*, *other*, *remain*, *nothing*, *certain*, *many*). But in addition, I also assigned some words to a new semantic field (different from those identified in my concordance analysis), namely WAR (*warrior*, *bomb*, *force*, *militia*, *militancy*, *terrorist*, *terrorism*).

Figure 4.43 shows part of the *Word Sketch* of the word *Islam* in *The Dawn* corpus.

Figure 4.43: partial Word Sketch of "Islam" in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

1) most of the modifiers of *Islam* belong to the semantic field of POLITICS (*political*, *reform*, *right-wing*);

2) most of the verbs with *Islam* as object belong to the semantic field of EDUCATION (*study*, *learn*, *know*, *understand*);

3) the word *Islam* frequently occurs in binomial expression with words belonging to the semantic field of POLITICS (e.g. *Islam and Communism*, *Socialism and Islam*, *Islam and politics*, *Islam, Authoritarianism and Underdevelopment*, *Nationalism and Liberal Islam*);

4) the word *Islam* is often preceded by the preposition *against*, which indicates a conflict between *Islam* and something else (e.g. *a war against Islam*, *the intellectual onslaught against Islam*, *the UN is against Islam*);

5) one collocate of *Islam* not previously identified in my concordance analysis carries POSITIVE SEMANTIC PROSODY (*success*[49]);

6) most of the collocates of *Islam* can be assigned to the same semantic fields identified through my concordance analysis, namely RELIGION (*religion*, *world*, *faith*, *Muslim*, *Allah*, *Pan-Islamism*, *Christianity*, *follow*), POLITICS (*political*, *politics*, *reform*, *right-wing*, *authoritarianism*, *nationalism*, *communism*, *socialism*, *liberal*, *tribal*), TIME (*century*, *modern*, *today*, *past*), KNOWLEDGE (*book*, *study*, *learn*, *know*, *understand*, *science*, *history*, *concept*), MATERIAL ACTION (*use*, *make*, *do*, *give*, *bring*, *take*), BEHAVIOUR AND ATTITUDE (*disrespect*, *prohibit*) and OTHER (*context*, *instance*, *people*, *serve*, *base*, *characterize*, *see*, *have*, *lay*, *require*, *experience*, *seem*, *hold*, *believe*, *instance*, *compatible*, *more*, *true*, *real*, *tendency*, *word*, *only*). However, I noticed that there are no words belonging to some semantic fields identified through my concordance analysis, namely LAW, REASONING and GEOGRAPHICAL LOCATIONS. Lastly, I identified some words which belong to a new semantic field, namely COMMUNICATION (*call*, *say*, *tell*, *present*).

Figure 4.44 shows part of the *Word Sketch* of the word *Muslim* used as noun, in *The Dawn* corpus.

[49] The complete sentence is '*Although following path of Islam eventually people with get Adal(justice), But main purpose of Islam is the eternal success of mankind*'.

Figure 4.44: partial Word Sketch of "Muslim" used as a noun in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

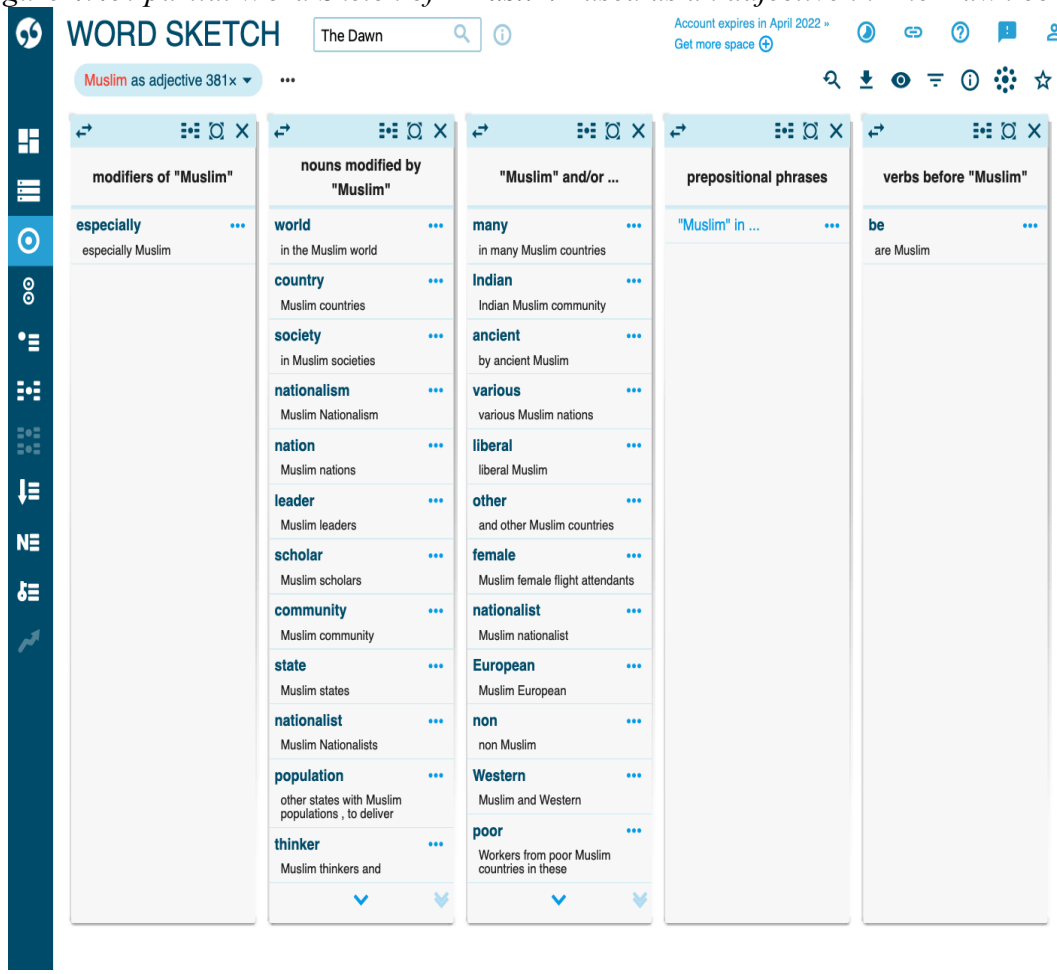
1) the collocates of *Muslim* used as noun can be assigned to the same semantic field identified through my concordance analysis, namely RELIGION (*liberal*,

redical [sic], *world*, *ummah*, *follow*, *Islamic*, *Islam*), ETHNIC GROUPS (*Indian*, *Hindu*, *Pakistani*, *French*, *Arab*), QUANTITY (*many*, *billion*, *most*), POLITICS (*militant*, *state*), GEOGRAPHICAL LOCATIONS (*country*, *Europe*), SOCIETY (*population*, *community*, *society*), ATTITUDE (*respect*, *reject*), INSTITUTIONS (*Brotherhood*, *League*, *Council*), KNOWLEDGE (*science*, *history*, *scholar*, *educate*, *know*), TIME (*early*), COMMUNICATION (*word*), and OTHER (*ordinary*, *non*, *other*, *real*, *first*, *kill*, *true*, *poor*, *fellow*, *good*, *see*, *do*, *have*, *live*, *choose*, *go*, *need*, *become*);

2) the most typical verb with *Muslim* as object is *kill*. Moreover, in three out of four occurrences this verb is associated with subjects/agents referring to Muslims who kill other Muslims (e.g. '*Why are Muslims killing Muslims?*', '*Were Arab Muslims not killing black Muslims in Darfur?*').

Figure 4.45 shows part of the *Word Sketch* of the word *Muslim* used as adjective in *The Dawn* corpus.

Figure 4.45: partial Word Sketch of "Muslim" used as an adjective in The Dawn corpus

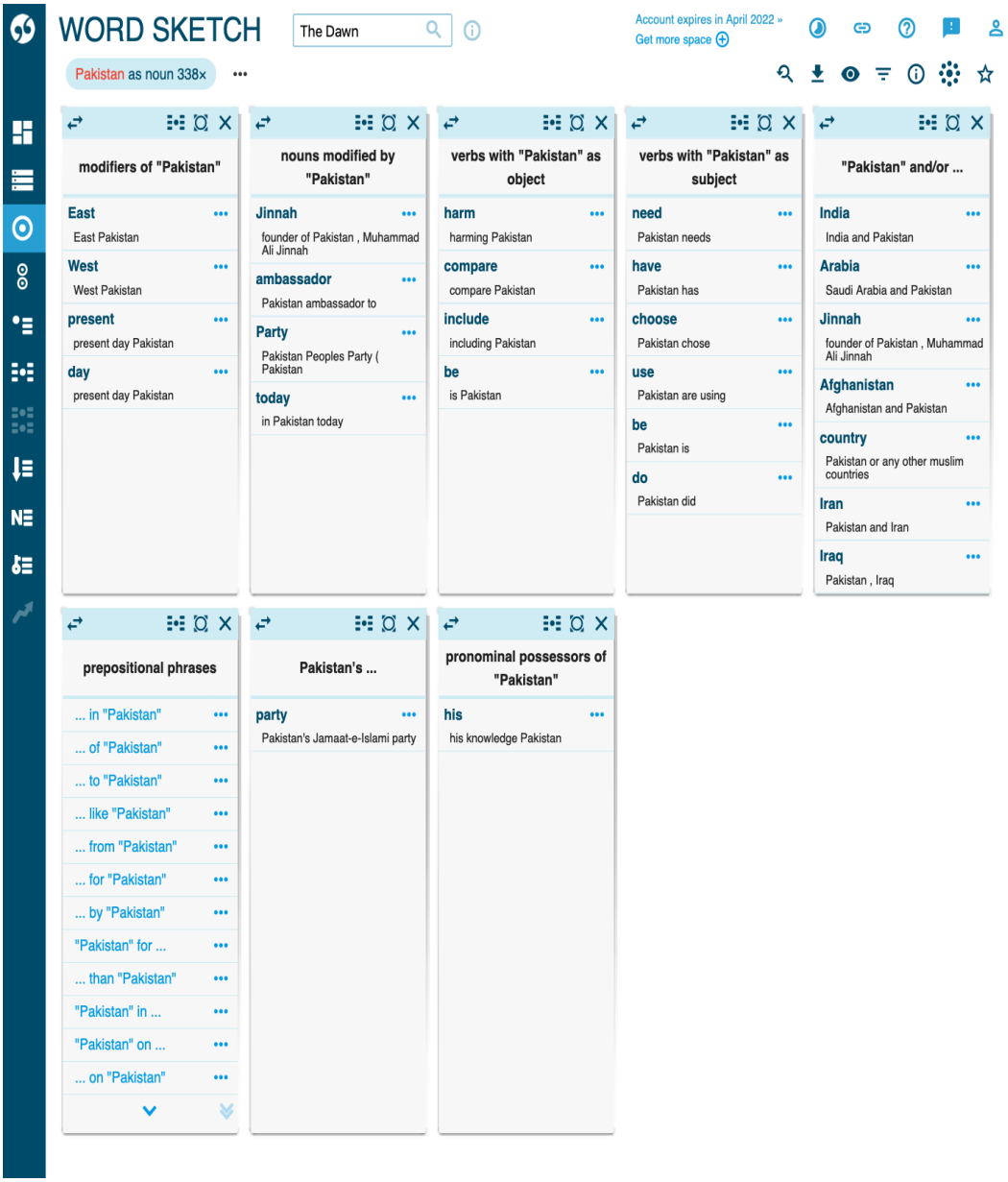


The analysis of the *Word Sketch* led to the following observations:

1) most of the collocates of *Muslim* used as adjective can be assigned to the same semantic fields identified through my collocations, namely POLITICS (*nationalism*, *nation*, *nationalist*, *leader*, *state*, *empire*, *right*, *regime*, *king*, *movement*, *military*), SOCIETY (*society*, *community*, *population*, *civilization*, *member*), SCHOLARSHIP (*scholar*, *history*, *scientist*, *polymath*, *intellectual*), RELIGION (*world*, *cleric*, *liberal*, *secular*), TIME (*today*, *modernist*, *modern*, *ancient*), GEOGRAPHICAL LOCATIONS (*country*, *European*, *Western*), ETHNIC GROUPS (*minority*, *Indian*), QUANTITY (*many*, *various*), OCCUPATIONS (*attendant*, *student*), and OTHER (*heritage*, *identity*, *other*, *non*, *poor*, *major*). However, I also assigned some words to a new semantic field namely PEOPLE (*female*, *woman*, *brother*, *youth*, *thinker*, *passenger*, *businessman*).

Figure 4.46 shows part of the *Word Sketch* of the word *Pakistan* in *The Dawn* corpus.

Figure 4.46: partial Word Sketch of "Pakistan" in The Dawn corpus



The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Pakistan* can be assigned to the same semantic field identified through my concordance analysis, namely COUNTRIES (*East*, *West*, *India*, *Arabia*, *Afghanistan*, *country*, *Iran*, *Iraq*), POLITICS (*ambassador*, *Party*), TIME (*present day*, *today*) and OTHER (*Jinnah*, *harm*, *compare*, *include*, *need*, *have*, *choose*, *use*, *do*);

2) the word *party* collocates with the word *Pakistan* to refer to a specific Pakistani political party, namely '*Pakistan Peoples Party*';

3) the verbs with *Pakistan* as subject (*need*, *have*, *choose*, *use*, *be*, *do*) are more varied and more numerous than those with *Pakistan* as object (*harm*, *compare*, *include*, *be*);

4) the prepositional phrase '*in Pakistan*' is often preceded (seven out of 54 times) by words with NEGATIVE SEMANTIC PROSODY (e.g. *attacks in Pakistan*, *violence in Pakistan*, *terrorism in Pakistan*, *discrimination in Pakistan*).

Figure 4.47 shows part of the *Word Sketch* of the word *religion* in *The Dawn* corpus.

Figure 4.47: partial Word Sketch of "religion" in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *religion* can be assigned to the same semantic fields identified through my concordance analysis, namely RELIGION (*Islam*, *preach*, *religion*), EVALUATION (*great*, *good*), PERSONAL PRONOUNS AND POSSESSIVES (*own*,

their, *our*, *your*), KNOWLEDGE (*interpret*, *understand*, *teach*, *science*, *culture*), POLITICS (*state*, *politics*), ARGUING (*say*, *blame*, *dispute*) and OTHER (*other*, *natural*, *separate*, *equate*, *keep*, *same*, *use*, *mix*, *practice*, *bring*, *ditch*, *promote*, *need*, *do*, *sound*, *accord*, *personal*, *matter*, *task*, *other*);

2) the evaluative words with POSITIVE SEMANTIC PROSODY (*great*, *good*) mostly refer to a specific religion, namely Islam (e.g. '*Islam is one among all great religions*', '*Islam is a great religion*', '*Islam is the best religion*');

3) the collocation '*keep religion*' frequently occurs with the word *politics*, when reference is made to a separation between the two areas (e.g. '*keep religion out of politics*', '*keep religion as a private affair away from politics*', '*keep away the religion from politics*');

4) the verbs with *religion* as object (*keep*, *use*, *mix*, *practice*, *bring*, *ditch*, *equate*, *interpret*, *preach*, *promote*, *blame*, *understand*, *need*, *do*, *say*, *have*, *be*) are more varied and more numerous than those with *religion* as subject (*preach*, *sound*, *accord*, *teach*, *have*, *be*);

5) the occurrences of the third person possessives '*their*' and '*your*' are much more frequent than the occurrences of the first person possessive '*our*' (20 vs 7 occurrences).

Figure 4.48 shows part of the *Word Sketch* of the word *Ulema* in *The Dawn* corpus.

Figure 4.48: partial Word Sketch of "Ulema" in The Dawn corpus

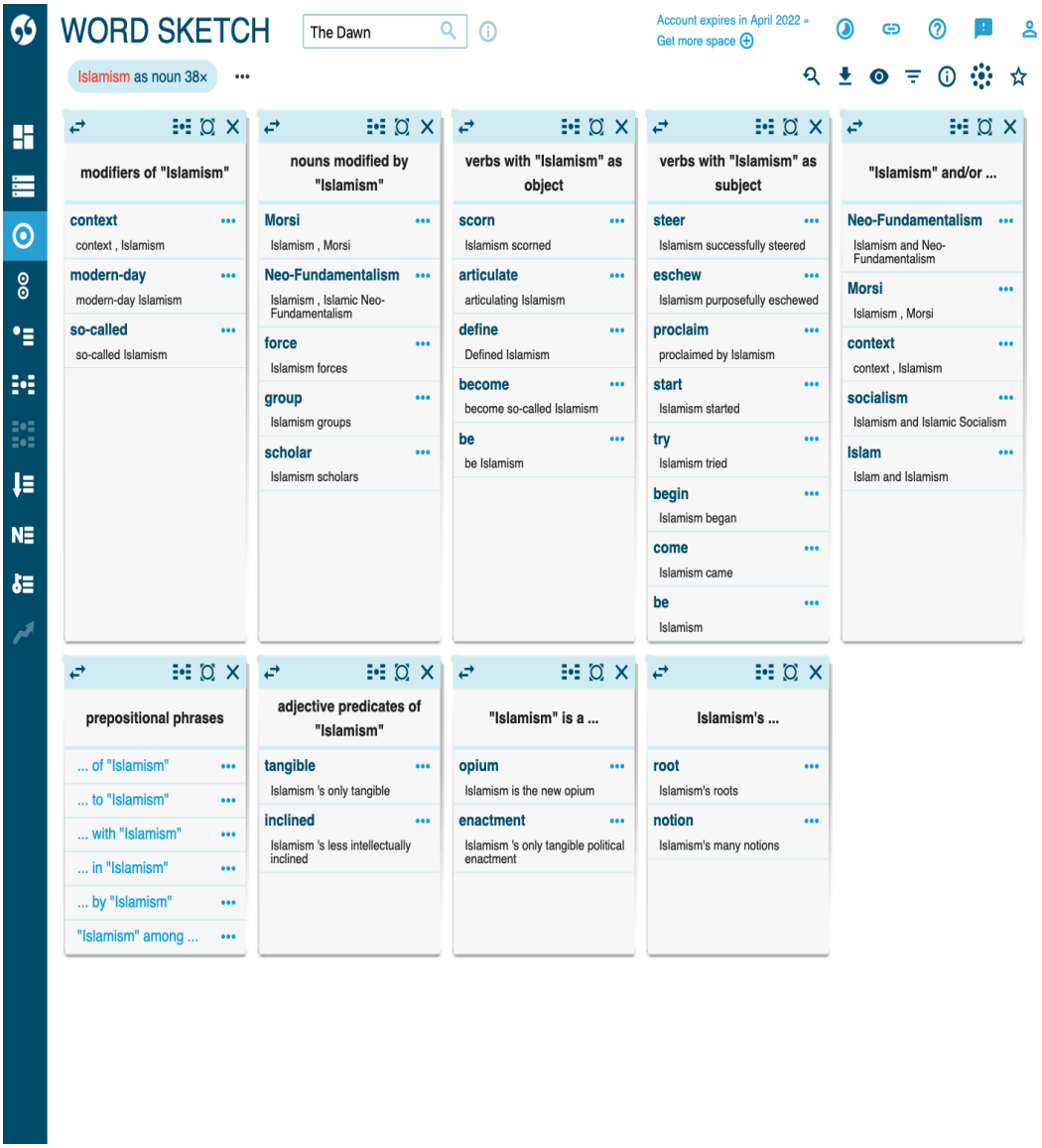
The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Ulema* can be assigned to the same semantic field identified through my concordance analysis, namely RELIGION (*conservative*, *merchant-Shia*, *Shia*, *clergy*, *clergyman*, *Muslim*, *Islamic*, *Islamist*, *cleric*, *clergy*), POLITICS (*apolitical*, *revolution*, *king*, *monarchy*, *party*, *state*), OTHER (*hopeful*, *individual*, *noted*, *future*, *ancient*, *so-called*, *global*, *traditional*, *different*, *real*, *such*,

many, *finance*, *develop*, *need*, *use*, *make*, *hijack*, *accuse*, *cover*, *keep*, *accept*, *have*, *be*, *decadence*, *rent*, *accountable*, *influence*, *our*).

Figure 4.49 shows part of the *Word Sketch* of the word *Islamism* in *The Dawn* corpus.

Figure 4.49: partial Word Sketch of "Islamism" in The Dawn corpus



The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Islamism* can be assigned to the same semantic fields identified through my concordance analysis, namely POLITICS (*socialism*, *enactment*), RELIGION (*Neo-Fundamentalism*, *Islam*), TIME (*modern-day*), NAMES TYPICAL OF ISLAMIC COUNTRIES (*Morsi*), and OTHER (*context*, *scholar*, *become*, *be*, *steer*,

eschew, *start*, *begin*, *try*, *come*, *force*, *tangible*, *inclined*, *opium*, *root*, *scorn*, *articulate*, *define*, *proclaim*, *notion*, *so-called*, *group*);

2) the verbs with *Islamism* as subject (*steer*, *eschew*, *proclaim*, *start*, *try*, *begin*, *come*, *be*) are more numerous and more various than those with *Islamism* as object (*articulate*, *define*, *become*, *be*)[50].

Figure 4.50 shows part of the *Word Sketch* of the word *Mullah* in *The Dawn* corpus.

[50] In the *scorn* case, the word *Islamism* is not used as subject of this verb ('*early thinkers of Islamism scorned at them*').

Figure 4.50: partial Word Sketch of "Mullah" in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Mullah* can be assigned to the same semantic fields identified through my concordance analysis, namely TIME (*century*, *[number]*, *today*), POLITICS (*militant*, *political*, *army*, *terrorism*, *politician*, *caliphate*), COMMUNICATION (*call*, *scream*, *quote*), RELIGION (*madrasa*, *cleric*, *world*) and OTHER (*Iqbar Aur*, *Fazlullah*, *Omar*, *traditional*, *such*, *manipulation*, *bashing*, *confront*,

crush, *buy*, *close*, *treat*, *put*, *see*, *be*, *run*, *roam*, *win*, *present*, *want*, *take*, *make*, *do*, *have*, *be*, *sanction*, *community*, *group*, *status*, *my*);

2) some words exemplify the semantic field of ATTITUDE (*mindless*, *minded*, *mad*) and CONFLICT (*manipulation*, *bashing*, *confront*, *problem*).

Figure 4.51 shows part of the *Word Sketch* of the word *socialism* in *The Dawn* corpus.

Figure 4.51: partial Word Sketch of "socialism" in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *socialism* can be assigned to the same semantic fields identified through concordance analysis, namely POLITICS (*Ba'ath*, *communism*, *fascism*, *commission*, *Marxism*, *Nazism*, *support*, *nationalism*, *polemic*, *order*, *PPP*, *FLN*), RELIGION (*Islamic*, *Islam*, *secular*), GEOGRAPHY (*Arab*, *Sudanese*, *Algerian*, *East*), KNOWLEDGE (*symbolism*, *reading*, *science*, *scientific*, *word*, *concept*, *ideology*, *thrust*), ECONOMY (*capitalism*, *wealth*, *corporate*), NAMES OF TYPICAL ISLAMIC COUNTRIES (*Paracha*, *Nasser*) and OTHER (e.g. *fall*, *concern*, *1970*, *modern*, *colour*, *start*, *work*, *evolve*, *formulate*, *originate*, *wither*, *profess*, *undertake*, *fuse*, *marry*, *accuse*, *mix*, *practice*, *attack*, *society*, *way*, *so-called*, *many*, *denounce*, *experience*). However, I also assigned some words to a new semantic field, namely ARGUING (*profess*, *accuse*, *attack*, *denounce*).

2) among the verbs with *socialism* as object, the verbs *mix*, *fuse* and *marry* refer to *socialism* as something to be combined with *Islam* (e.g. '*the economic maneuvers by regimes fusing socialism with certain aspect of Islam*', '*some people on the left attempted to marry Socialism and Marxism with Islam*', '*NFP and others have managed to mix socialism, an economic concept with Islam, a religious belief system*');

3) the words found in binomial expressions with the word *socialism* mostly belong to the semantic field of POLITICS (e.g. *capitalism, socialism*; *Marxism*, *socialism*; *socialism and Arab nationalism*; *Communism, socialism*; *Socialism and fiery polemics*; *socialism*, *fascism*);

4) the prepositional phrase *of Socialism* is occasionally (four out of 35 cases) preceded by a word which denotes demise, failure or collapse ('*The demise of Islamic Socialism*', '*the collapse of Algerian socialism*', '*the collapse of Islamic Socialism*', '*the failure of Socialism*').

Figure 4.52 shows part of the *Word Sketch* of the word *Quran* in *The Dawn* corpus.

Figure 4.52: partial Word Sketch of "Quran" in The Dawn corpus

The analysis of the *Word Sketch* led to the following observations:

1) the collocates of *Quran* can be assigned to the same semantic field identified through concordance analysis, namely KNOWLEDGE (*Read*, *memorize*, *read*, *understand*, *text*), RELIGION (*Holy*, *ordain*, *Sunnah*, *Hadith*, *creator*, *Shariah*), QUANTITY (*whole*, *thousand*, *replete*, *enough*), COMMUNICATION (*say*, *pronounce*,

promise, *warn*, *mention*, *talk*, *explain*, *ask*, *code*, *concept*) and OTHER (e.g. *sir*, *Al*, *noble*, *Sharief*, *justice*, *time*, *download*, *abandon*, *equate*, *consider*, *include*, *make*, *do*, *satates [sic]*, *use*, *lay*, *offer*, *go*, *be*, *do*, *have*);

2) the most frequent verbs with *Quran* as subject belong to the semantic field of COMMUNICATION (*say*, *pronounce*, *ordain*, *promise*, *warn*, *mention*, *talk*, *explain*);

3) the prepositional phrase *of Quran* is often (5 out of twenty cases) preceded by the word *createdness* (*createdness of the Quran*).

4.4 Findings about the *Word Sketch Difference* of key terms

Finally, the *Word Sketch Difference* function will show shared and unshared collocations of the words *Islam*, *Muslim* and *Islamic* in the *New York Times* and *The Dawn* corpora. Figure 4.38 shows part of the *Word Sketch Difference* of the word *Islam* in the *New York Times* and *The Dawn* corpora.

Figure 4.38: partial Word Sketch Difference of "Islam" in the New York Times and The Dawn corpora

The analysis of the *Word Sketch Difference* led to the following observations:

1) there are more concordances of the word *Islam* in the *New York Times* corpus than in *The Dawn* corpus (719 vs 671 concordances respectively);

2) all the collocates of *Islam* identified through the *Word Sketch Difference* function in the *New York Times* corpus can be assigned to the same semantic fields identified through concordances analysis (see Section 4.5-A-B), namely POLITICS (e.g. *democracy*, *communism*, *radical*, *militant*), KNOWLEDGE (*culture*, *teach*, *study*, *scholar*, *book, understanding*, *interpretation*), RELIGION (e.g. *Muslims*, *religion*, *orthodox*, *Sunni*), TIME (*modernity*, *medieval*, *today*), ATTITUDE (*practice*, *respect*, *embrace*, *practice*, *leave*, *incompatible*), GEOGRAPHY (*France*, *America*, *Europe*, *country*) and OTHER (*say*, *talk*, *terrorism*, *Indonesian*, *mainstream*, *think*, *spread*, *take*, *become*);

3) all the collocates of *Islam* identified through the *Word Sketch Difference* function in *The Dawn* corpus can be assigned to the same semantic fields identified through concordances analysis (see Section 4.29-A-B), namely POLITICS (e.g. *communism*, *politics*, *socialism*, *political*), RELIGION (*Muslims*, *religion*, *faith*, *convert*), KNOWLEDGE (*study*, *word*, *teaching*, *knowledge*, *history*), and OTHER (e.g. *context*, *be*, *see*, *base, present*). However, some words belonging to semantic fields identified through concordances analysis and *Word Skecth* function are absent in the *Word Sketch Difference* function, namely TIME (e.g. *ancient*, *early*, *end of*), LAW (e.g. *judge*, *advocated*, *tenets of*) BEHAVIOUR AND ATTITUDE (e.g. *ignorance about*, *views about*, *behaviour as*) REASONING (e.g. *distinction*, *mocking*, *distortion of*) and GEOGRAPHY (e.g. *western*, *while Europe*, *in the West*);

4) in the *New York Times* corpus one of the most typical collocations occurring in binomial expressions with the word *Islam* is the word *terrorism*, which carries NEGATIVE SEMANTIC PROSODY;

5) both in the *New York Times* and *The Dawn* corpora the most typical modifiers of *Islam* can be assigned to the semantic field of POLITICS. However, there is a different focus: the US paper mostly focus on radical Islam (e.g. *radical*, *Fatah*, *Sarekat*), while the Pakistani paper mostly focus on liberal Islam (e.g. *Liberal*);

6) in the *New York Times* corpus, the verbs with *Islam* as object mostly indicate people's attitude and/or adherence to religion (*practice*, *respect*, *embrace*, *leave*), while those in the *Dawn* corpus mostly indicate utility (*use*, *hijack*, *serve*);

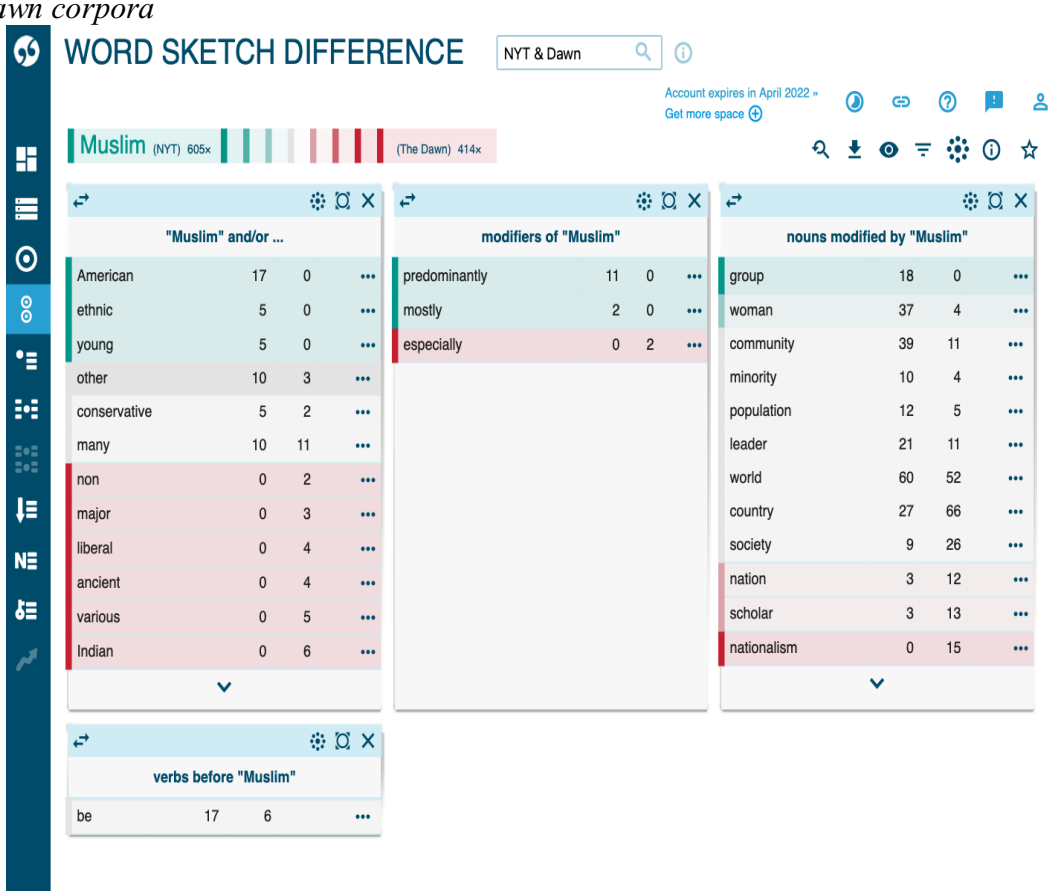
7) in the *New York Times* corpus, the most typical predicative adjective of *Islam* is *incompatible* (e.g. '*Islam is incompatible with democracy*', '*democracy and Islam were inherently incompatible*'), which contributes to negative semantic prosody; while in *The Dawn* corpus it is *compatible* (e.g. '*communism and Islam were compatible*', '*Is Islam compatible with socialism*?'), contributing to positive semantic prosody the latter;

8) in the *New York Times* corpus, most of the modifiers of the word *Islam* can be assigned to the semantic fields of RELIGION (*radical*, *orthodox*, *Sunni*, *conservative*) and POLITICS (*militant*, *Fatah*, *moderate*), while in *The Dawn* corpus they can be assigned only to the semantic field of POLITICS (*right-wing*, *Liberal*, *Political*);

9) in the *New York Times* corpus, most of the words occurring before the prepositional phrase '*to Islam*' indicate conversion (*turn*, *convert*).

Figure 4.39 shows part of the *Word Sketch Difference* of the word *Muslim* in the *New York Times* and *The Dawn* corpora.

Figure 4.39: partial Word Sketch Difference of "Muslim" in the New York Times and The Dawn corpora



The analysis of the *Word Sketch Difference* led to the following observations:

1) there are more concordances of the word *Muslim* in the *New York Times* corpus than in *The Dawn* corpus (605 vs 414 concordances respectively);

2) all the collocates of *Muslim* identified through the *Word Sketch Difference* function in the *New York Times* corpus can be assigned to the same semantic fields identified through concordances analysis (see Section 4.4-A-B), namely PEOPLE (e.g. *young*, *man*, *girl*, *people*), ETHNIC GROUPS (*American*, *ethnic*, *black*, *Indonesian*, *Arab*), RELIGION (e.g. *observant*, *fundamentalism*, *holy*, *religious*), POLITICS (e.g. *minister*, *communism*, *state*, *moderate, ruler*, *rule*, *leader*), SOCIAL GROUPS (e.g. *immigrant*, *migrant*, *group*, *minority*), QUANTITY (*whole*, *global*, *several*, *many*, *predominantly*), GEOGRAPHY (*neighborhood*, *enclave*, *land*, *area*, *country*), WORK (*employee*, *banker*, *merchant*, *worker*), CULTURE (*tradition*, *custom*, *garb*, *veil*), INSTITUTIONS (*Brotherhood*, *association*, *organization*, *center*), CRIME (*terrorist*, *soldier, inmate*, *prisoner*) and OTHER (e.g. *lost*, *civil*, *first*, *other*, *identity*, *ban*);

3) most of the collocates of *Muslim* identified through the *Word Sketch Difference* function in *The Dawn* corpus can be assigned to the same semantic fields identified through concordances and *Word Sketch* (see Sections 4.30-A-B, 4.44 and 4.45), namely RELIGION (e.g. *secular*, *convert*, *fundamentalism*, *religion*), POLITICS (e.g. *conservative*, *military*, *socialist*, *nationalist*), KNOWLEDGE (e.g. *scholar*, *intellectual*, *polymath*, *scientist*), SOCIETY (*civilization*, *community*, *minority*, *member*, *society*), ETHNIC GROUPS (*Indian*, *European*, *Western*), TIME (*modernist*, *today*, *modern*, *ancient*), PEOPLE (*brother*, *population*, *youth*), and OTHER (e.g. *other*, *traditional*, *poor*, *classical*, *businessman*, *many*, *various*). However, some words belonging to semantic fields identified through concordance analysis and the *Word Sketch* function are absent in the *Word Sketch Difference* function, namely COMMUNICATION (e.g. *call*, *to call all*, *be called*), GEOGRAPHY (e.g. *around the world*, *from west*, *in Nigeria*), PROFESSIONS (e.g. *farmers*, *attendants*, *merchants*), ATTITUDE (e.g. *do not care*, *don't need*, *obsessed*) and INSTITUTIONS (e.g. *Brotherhood*, *Council*, *League*);

4) in *The Dawn* corpus the most typical nouns modified by *Muslim* can be assigned to the semantic field of POLITICS (e.g. *nationalism*, *regime*, *empire*, *nationalist*); while in the *New York Times* corpus they can be assigned to the semantic field of PEOPLE (e.g. *group*, *man*, *immigrant*, *girl*);

5) among the nouns modified by the word *Muslim*, only one word which carries NEGATIVE SEMANTIC PROSODY was identified in *The Dawn* corpus (*revolt*), while four words (*ban*, *inmate*, *prisoner*, *terrorist*) were identified in the *New York Times* corpus;

6) among the nouns modified by the word *Muslim*, in *The Dawn* corpus the words *nationalist* and *nationalism* are very frequent (five and 15 occurrences, respectively), while in the *New York Times* corpus these words are not even mentioned;

7) in the *New York Times* corpus, the most typical collocation occurring before the word *Muslim* is *American*, while in *The Dawn* corpus it is *Indian*.

Figure 4.40 shows part of the *Word Sketch Difference* of the word *Islamic* in the *New York Times* and *The Dawn* corpora.

Figure 4.40: partial Word Sketch Difference of "Islamic" in the New York Times and The Dawn corpora

The analysis of the *Word Sketch Difference* led to the following observations:

1) there are many more concordances of the word *Islamic* in *The Dawn* corpus than in the *New York Times* corpus (663 vs 535 concordances respectively);

2) all the collocates of *Islamic* identified through the *Word Sketch Difference* function in the *New York Times* corpus can be assigned to the same semantic fields identified through concordance analysis and *Word Sketch* (see Sections 4.4), namely RELIGION (*world*, *fundamentalism*, *theological*, *religious*), POLITICS (e.g. *militant*, *radical*, *political*, *law*), TIME (*medieval*, *first*, *modern*), KNOWLEDGE (*study*, *scholar*, *history*), GEOGRAPHY (*European*), FINANCE (*financial*, *bank*, *banking*, *finance*), CULTURE (*classical*, *traditional*), COMMUNICATION (*call*), PEOPLE (*people*), GROUPS (*European*) and OTHER (e.g. *not*, *be*, *proper*, *perfect*);

3) all the collocates of *Islamic* identified through the *Word Sketch Difference* function in *The Dawn* corpus can be assigned to the same semantic fields identified through concordance analysis and *Word Sketch* (see Sections 4.28-A-B and 4.42), namely RELIGION (e.g. *puritanical*, *world*, *fundamentalism*, *fundamentalist*), POLITICS (e.g. *radical*, *conservative*, *political*, *right-wing*), EVALUATION (e.g. *true*, *glorious*, *great*), COMMUNICATION (e.g. *so-called*, *slogan*), KNOWLEDGE (e.g. *scholar*, *science*, *history*, *study*), GEOGRAPHY (e.g. *western*, *country*), QUANTITY (e.g. *many*, *certain*, *fourth*), LAW (e.g. *law*, *principle*, *code*), TIME (e.g. *first*, *new*, *old*, *modern*), FINANCE (e.g. *financial*, *banking*, *finance*, *product*), WAR (e.g. *terrorist*, *force*, *militia*) and OTHER (e.g. *classical*, *traditional*, *other*, *nothing*);

4) evaluative terms which carry POSITIVE SEMANTIC PROSODY (*glorious*, *great*) only emerge in *The Dawn* corpus, but not in the *New York Times* corpus;

5) the words belonging to the semantic field of TIME (*modern*, *present-day*, *ancient*, *modern-day*) are frequently used in *The Dawn* corpus compared against *The New York Times* corpus;

6) in the *New York Times* corpus, most of the nouns modified by *Islamic* can be assigned to the semantic field of FINANCE (e.g. *bank*, *bond*, *investment*, *instrument*), while in *The Dawn* corpus they can be assigned to the semantic field of POLITICS (e.g. *socialism*, *nation*, *regime*, *socialist*), with a particular focus on SOCIALISM;

7) in *The Dawn* corpus several nouns modified by *Islamic* can be assigned to the semantic field of WAR (e.g. *force*, *bomb*, *warrior*, *militia*), while no words modified by *Islamic* belonging to the semantic field of WAR were identified in the *New York Times* corpus.

5. Conclusion

5.1 Introduction

In this chapter, I summarise and discuss the results of my study. More specifically, in Section 5.2, I recapitulate the issue addressed and the method of investigation adopted in the present work. Then, in Sections 5.3, 5.4 and 5.5. I outline the major results obtained from the analysis of the *New York Times* and *The Dawn* corpora. In Section 5.6, I discuss the findings. In the last section, I present the limitations of my study and make suggestions for further research.

5.2 Research goal and approach

The topic of the present study was the representation of Islam and Muslims in the press. Previous studies on this topic had highlighted that western media tend to frame Islam within a discourse of conflict and violence, representing Muslims as a monolothic group of people whose values clash with those of Western civilization, with Muslim men being mostly portrayed as terrorists, Muslim women as victims of their own culture, and all Muslims as 'others', "them" (i.e. the topic of discussion) rather than "us" (i.e. the readership). Additionally, scholars who conducted comparative-contrastive studies on journalistic texts with a mixed-method approach showed that in countries with opposing ideologies (i.e. USA vs. Pakistan), the media tended to use different terms to represent the same entities and phenomena (e.g. *Pakistani Taliban* vs. *militants*).

Drawing on the findings and the methodological approaches adopted in previous studies, I carried out a comparative corpus-driven investigation into the representation of Islam and Muslims in an American and a Pakistani newspaper, namely the *New York Times* and *The Dawn*, respectively. The purpose was to identify similarities and differences in the characterization of Islam and Muslims in the two newspapers. This involved exploring the notions recurrently mentioned in association with the terms denoting Islam and Muslims, and exploring whether media mostly report good news or bad news about them.

5.3 Main findings about the *New York Times* corpus

The most recurrent topics and notions mentioned in the *New York Times* corpus were identified through word frequency lists, n-grams analysis and keyword lists. The findings showed that the most recurrent semantic fields identified through bottom-up lexical analysis were RELIGION (e.g. *faith*, *god*, *imam*), POLITICS (e.g. *government*, *law*, *leaders*), and GEOGRAPHY (e.g. *American*, *country*, *East*); however, also VIOLENCE (*attacks*, *violence*, *war*) and TERRORISM (*Who are the Taliban*, *Al Quaeda and*, *Mr. bin Laden*) emerged as relevant notions, thus showing that Islam and Muslims were thought of in terms of danger and negativity.

The analysis of concordances of the five most frequent words (*muslim*, *islam*, *islamic*, *religious*, *american*) and top five keywords (*uighur*, *xinjiang*, *islamism*, *islamically*, *shariah*) in the *New York Times* corpus, as well as an examination of their *Word Sketches*, led to the identification of their semantic preferences and semantic prosodies:

1) the pronouns *they*/*them*, and not *we*/*us*, were used to refer to Muslims and Islam, suggesting that they were subjected to "othering", that is, treated as different from and generally inferior to one's own group;
2) other frequent collocates of the word *Islamic* were assigned to the semantic field of FINANCE (21.5% of the time; e.g. *bank*, *banker*, *economy*), thus showing a focus on Islamic economy;
3) some of the most frequent collocates of the word *American* denoted words related to Islam and/or Muslims (20% of the time; e.g. *American Muslim*, *American Mosque*, *American ummah*), thus indicating a focus on national rather than international Islam-related phenomena;
4) the following frequent words and keywords – mostly related to Islam – were found to be frequent collocates of terms expressing notions of conflict and violence, as a result of which they appeared to be characterized by negative semantic prosody: *Muslim* (e.g. *concern about*, *mistrust between*, *oppression*); *Islam* (e.g. *dark view of*, *animosity toward*, *destroy*); *Islamic* (e.g. *are blaming an*, *obscure*, *destroy the*); *religious* (e.g. *adversary*, *concerns*, *crisis*); *Uighur* (e.g. *deadly attacks by*, *violence by*, *detain*); *Islamism* (e.g. *law against*, *to combat*, *terror on*); *Shariah*

(e.g. *they seem worried*, *with public fogging*, *rejected by*); *American* (e.g. *cultural invasion*, *enemy*, *occupiers*).

5.4 Main findings about *The Dawn* corpus

As in the *New York Times* corpus, the most recurrent topics/notions mentioned in *The Dawn* corpus were identified through word frequency lists, n-grams analysis and keyword lists. The findings showed that the most recurrent semantic fields identified through such a bottom-up lexical analysis were RELIGION (e.g. *faith*, *god*, *imam*), POLITICS (e.g. *government*, *law*, *leaders*), GEOGRAPHY (e.g. *American*, *country*, *East*), and FINANCE (*Islamic financial products and*, *conventional products and*, *financial products and services*). However, contrary to what emerged in the *New York Times* corpus, no semantic fields related to conflict and/or violence were identified.

The analysis of concordances of the five most frequent words (*islamic*, *islam*, *muslim*, *pakistan*, *religion*) and top five keywords (*ulema*, *islamism*, *mullah*, *socialism*, *Quran*) in *The Dawn* corpus, as well as an examination of their *Word Sketches*, led to the identification of their semantic preferences and semantic prosodies:

1) here too, the pronouns *they*/*them*, and not *we*/*us*, were used to refer to Muslims and Islam, which could point to the phenomenon of "othering";
2) frequent (12% of the time) collocates of the word *Muslim* were assigned to the semantic field of QUANTITY (e.g. *4000*, *many*, *the whole*), thus suggesting that Pakistani media tend to talk about Muslims as an undifferentiated group of people;
3) the following frequent words and keywords – mostly related to Islam – were found to be frequent collocates of terms expressing notions of conflict and violence, as a result of which they appeared to be characterized by negative semantic prosody: *Islamic* (e.g. *bomb*, *deaths*, *terrorist*); *Islam* (e.g. *against*, *blaming*, *demonizing*); *Muslim* (e.g. *against*, *brutal*, *selfish*); *Pakistan* (e.g. *defaming*, *against minorities*, *has no future*); *religion* (e.g. *lame and*, *as a weapon*, *is close to insulting*); *Islamism* (*is also invalid*, *scorned at them*, *started to weaken and fragment*); *Mullah* (e.g. *terrorism and*, *mindless*, *manipulations*); *socialism* (e.g. *aka Nazism*, *are not compatible with Islam*, *as harnessing the freedom of thought*).

5.5 Main findings from the *Word Sketch Difference* of key terms

Lastly, in order to better explore how similarly or differently Islam and Muslims were characterized in the two corpora, I used the *Word Sketch Difference* function to highlight shared and unshared collocations of the words *Islam*, *Muslim* and *Islamic*.

1) Some differences emerged regarding the word *Islam*: the US paper mostly focused on radical Islam, while the Pakistani paper mostly focused on liberal Islam, as evidenced by their most typical modifiers of *Islam* (e.g. *radical*, *Fatah*, *Sarekat* vs. *Liberal*);

2) the collocates of *Islam* carrying negative semantic prosody were more frequent in the *New York Times* corpus (*terrorism*, *incompatible*) than in *The Dawn* corpus (*war*); and

3) Islam was mostly represented within a discourse of religion in the *New York Times* corpus (e.g. *practice*, *embrace*, *Sunni*), but not in *The Dawn* corpus, more focused on politics (*militant*, *Fatah*, *moderate*);

The results of the *Word Sketch Difference* function of the word *Muslim* led to the following observations:

1) the semantic field of crime emerged in the *New York Times* corpus (e.g. *terrorist*, *soldier*, *inmate*, *prisoner*), while it was absent from *The Dawn* corpus; the nouns modified by the word *Muslim* carrying negative semantic prosody were more frequent in the *New York Times* corpus (*ban*, *inmate*, *prisoner*, *terrorist*) than those identified in *The Dawn* corpus (*revolt*); and

2) there was a focus on Indian Muslims in *The Dawn* corpus, while on American Muslims in the *New York Times* corpus.

The results of the *Word Sketch Difference* function of the word *Islamic* highlighted the following differences:

1) the semantic field of war (e.g. *terrorist*, *force*, *militia*) which showed a representation of Muslims within a discourse of violence, only emerged in *The Dawn* corpus, while it was absent in the *New York Times* corpus;

2) in the *New York Times* corpus, there was a focus on Islamic finance, as suggested by the most typical nouns modified by *Islamic* (e.g. *bank*, *bond*, *investment*);

3) instead, in *The Dawn* corpus, there was an emphasis on Islamic politics, particularly on socialism, as attested by the most typical nouns modified by *Islamic* (e.g. *socialism*, *nation*, *regime*);

4) evaluative words referring to *Islamic* which carry positive semantic prosody emerged (*glorious*, *great*) in *The Dawn* corpus, while they were absent in the *New York Times* corpus.

5.6 Conclusion

The results of the present work provide partial support for the findings of previous studies, that is, on the one hand, similarities in the representation of Islam and Muslims between the two corpora were detected; on the other some differences also emerged in line with the findings of previous studies.

Similarities show that both in the *New York Times* and *The Dawn* newspapers, Islam and Muslims were often represented within a discourse of violence and conflict (see the reference to terrorism), a distinction was often made between us (non-Muslims) and them (Muslims), and most of the collocates of the words referring to Muslims carried negative semantic prosody. Moreover, one unexpected pattern emerged, that is, although *The Dawn* newspaper is supposed to be the Muslim League mouthpiece, the phenomenon of 'othering' was detected there, too, with regard to the collocates of the word *Muslim*. This could be regarded as a tendency by Pakistani media to assume that most of the readers of *The Dawn* articles were non-Muslims. These findings differ from those of the study conducted by Mahmood et al. (2018) on the *New York Times* and *The Dawn* editorials about the portrayal of the Taliban terrorists vs. the victims killed and injured: in *New York Times* editorial, the Taliban were criticized, while in *The Dawn* editorial, the Taliban were not explicitly held responsible for the attack. In Mahmood et al. (2018), the different attitude upheld by Pakistani media could be regarded as a tendency to mitigate the actions of their compatriots, that is, the Taliban. In the present work, no mitigation strategies were found.

Nonetheless, the two newspapers presented some marked differences, that is, they represented the same truth through a different slant: in the American paper, the semantic fields of conflict and violence were more frequent than in the Pakistani paper; the Pakistani paper focused on politics more than the American paper; the American paper focused on radical Islam, while the Pakistani paper focused on liberal Islam.

This work suffered from a few limitations:

1) the *New York Times* and *The Dawn* corpora were not similar in size;
2) the temporal sampling of the texts was not precise: although in both corpora there are texts dating from 1990, 2000 and 2020, I do not know how many exemplify each decade;
3) only a limited amount of data was analysed, namely the 200 most frequent words, the 200 most frequent n-grams and the 200 most typical keywords;
4) the corpora were not perfectly "clean": the *New York Times* corpus contained phrases whose function is guiding the readers through the articles (e.g. *Reprints Timesmachine is an*, *benefit for home delivery*, *home delivery and digital*); and *The Dawn* corpus contained words belonging to comments written by readers of the newspaper (e.g. *recommend*, *abbastoronto*, *Jul*);
5) negative keywords, that is, those words which are much less frequent than what might be expected, were not explored, because they could not be retrieved in *Sketch Engine*.

In future research projects, it may be interesting to examine how Muslim women are represented in the press (i.e. the *New York Times* and *The Dawn*), for example whether they are described negatively, as victims of their own culture, or if a different representation prevails. In addition, I would suggest examining the ways an American newspaper represents American vs non-American Muslims: whether Muslim Americans are represented more positively, being regarded as compatriots, and if Muslim non-Americans are portrayed more negatively, being regarded as outsiders. Similarly, it would be interesting to investigate how differently a Western newspaper represents Christians and Muslims. In all these cases, a bottom-up lexical analysis of the data, similar to the analysis carried out here, would reveal patterns of language use and thus shed light on journalists' views of given social/ethnic groups.

Riassunto

In questa tesi affronto il tema di come si parla dell'Islam e dei musulmani nella stampa statunitense e pakistana. L'idea di svolgere questo studio mi è venuta immediatamente dopo il ritiro delle truppe americane dall'Afghanistan (30 agosto 2021) e la conseguente occupazione di questo stato da parte dai talebani. Essendo in Italia, ero esposto a notizie riguardanti questo tema solamente dai media occidentali, che si focalizzavano sulle conseguenze negative dell'occupazione dei talebani a causa del loro estremismo religioso. Ho cominciato a riflettere che nel mondo occidentale i media solitamente rappresentano la religione dell'Islam principalmente con un taglio negativo e mi domandavo se nel mondo medio-orientale i media ne parlassero in modo diverso. Inoltre, ho notato che associavo frequentemente il terrorismo ai musulmani; perciò, ho cominciato a chiedermi da dove provenisse questo nesso, e se potesse emergere o diventare addirittura preminente una rappresentazione diversa dei musulmani e/o di eventi connessi all'Islam in uno stato a maggioranza islamica.

Studi precedenti su questo tema mostrano che i media occidentali tendono a rappresentare l'Islam negativamente, in una narrazione di conflitto e violenza, conservatività o arretratezza, e in netto contrasto con i valori della società occidentale: gli uomini musulmani vengono presentati principalmente come terroristi, le donne musulmane come vittime della propria cultura e, più in generale, tutti i musulmani come un gruppo monolitico di 'altri' (cioè come dei *loro* contrapposti a *noi*). Questa rappresentazione negativa dell'Islam e dei musulmani risulta da ricorrenti scelte linguistiche (p. es. la frequente rappresentazione dei musulmani nel ruolo semantico di pazienti nelle frasi, l'uso di termini spregiativi per riferirsi ai musulmani quali *terrorists*, *extremists*, *fundamentalists*, *radicals*, *fanatics*). Tutto questo contribuisce a una visione stereotipata dei musulmani e dell'Islam, veicolando diffidenza verso di loro, e di conseguenza ostacolando, per quanto indirettamente, la loro integrazione in stati a maggioranza non islamica. Inoltre, studi che mettono a confronto i discorsi mediatici prodotti in stati caratterizzati da ideologie contrastanti (p. es. Pakistan e USA) illustrano come queste ideologie sembrino influenzare i discorsi su una stessa realtà sociale, presentandola, in un contesto, in modo neutro o positivo, ma nell'altro in modo negativo (p. es. *militants* rispetto a *Pakistani Taliban*).

La maggior parte dei suddetti studi, di tipo comparativo-contrastivo, e basati su un metodo misto (p. es. qualitativo e quantitativo), si è concentrata su testi giornalistici, ma occasionalmente anche riviste e notiziari radio o televisivi, ed ha esaminato le notizie prodotte dai media statunitensi, a confronto con quelle provenienti da varie fonti: Medio Oriente, Regno Unito, Italia e Spagna, Australia e Cina.

Questo studio si inserisce nell'alveo delle ricerche precedenti e prende in esame la rappresentazione dell'Islam e dei musulmani attraverso l'identificazione di ricorrenti pattern linguistici nel giornale americano *New York Times* e in quello Pakistano *The Dawn.* I testi considerati per questo studio sono composti da articoli giornalistici scaricati dalle versioni online dei suddetti giornali. Questi giornali sono stati scelti perché sono tra i giornali più importanti negli Stati Uniti e in Pakistan, rispettivamente. Per avere accesso agli articoli del *New York Times* e scaricarli, è stato necessario pagare 1€ per l'abbonamento. Invece, il *The Dawn* offre pieno accesso al pubblico per la lettura e il download degli articoli.

I dati sono stati raccolti mediante la piattaforma *Sketch Engine* (https://www.sketchengine.eu), una delle cui funzioni permette anche la compilazione di corpora. Gli articoli sono stati scaricati direttamente dai siti web dei giornali (www.nytimes.com and dawn.com) usando la seguente combinazione di parole di ricerca: *Islam*, *Islamic*, *Islamically*, *Muslim* and *Muslims*. Per la ricerca sono state utilizzate le impostazioni predefinite di *Sketch Engine* riguardo alla scelta delle parole di ricerca e dei siti rilevanti, che possono essere specificati sotto l'icona *Web search settings* menu. Successivamente, i testi sono stati puliti e gli elementi non testuali sono stati eliminati automaticamente da *Sketch Engine*, come spiegato nella guida per gli utenti. Dato che i testi ottenuti a volte presentavano coppie di testi con contenuto simile o identico, ho usato la funzione *Remove duplicated content* per tenere solamente una copia di ciascun testo. La versione finale del corpus *The Dawn* ammontava a 138,542 tokens e 31 documenti, mentre quella del *New York Times* ammontava a 204,843 tokens e 106 documenti. Risulta così che gli articoli del *The Dawn* sono di meno, ma più lunghi, rispetto a quelli del *New York Times*, anche se il numero di frasi e paragrafi è quasi uguale in entrambi i giornali (9,509 frasi nel corpus del *The Dawn* e 8,787 frasi nel corpus del *New York Times*; 5,917 paragrafi nel corpus del *The Dawn* e 3,918 paragrafi nel corpus del *New York Times*).

Date le dimensioni dei corpora, questo indica che le frasi e i paragrafi nel corpus del *The Dawn* sono più corti rispetto a quelli del corpus del *New York Times*.

I dati sono stati analizzati mediante un approccio *corpus-driven*, ossia attraverso un'analisi dal basso verso l'altro con lo scopo di identificare ricorrenti pattern fraseologici (Tognini-Bonelli, 2001: 17). Questi pattern (p. es. combinazioni ricorrenti di parole) permettono al ricercatore di formulare ipotesi sul contenuto del testo, e sui modelli culturali che essi implicano (Stubbs, 2002: 17). Il vantaggio principale dell'approccio *corpus-driven*, che, per dirla alla Saussure, esamina la *parole*, cioè la manifestazione linguistica individuale del sistema astratto della grammatica (Tognini-Bonelli, 2001: 169), consiste nel fatto che conduce alla scoperta di nuovi fenomeni, all'elaborazione di nuove ipotesi e la messa in dubbio di schemi interpretativi tradizionali.

L'analisi condotta si è avvalsa di software per l'analisi linguistica dei corpora. Con questi strumenti si sono individuate: A) le *word frequency lists*, ossia le liste delle parole che formano un corpus in ordine di frequenza decrescente; B) le *keyword lists*, cioè le liste di parole che sono inaspettatamente frequenti o infrequenti in un corpus, solitamente più piccolo, comparato ad un altro corpus, solitamente più grande, chiamato *reference corpus*; C) gli *N-grams,* cioè le unità ricorrenti composte da più parole; e D) le *concordances* (concordanze), che sono righe di testo in cui una certa parola o sintagma presa in esame (detta *parola/sintagma nodo*) sono presentati al centro dello schermo del computer con le parole che la precedono o la susseguono sulla sinistra e sulla destra (Hunston 2002: 39). Le *word frequency lists*, le *keyword lists* e gli *N-grams* sono utili per identificare le nozioni e i temi principali menzionati in un corpus; mentre le *concordances* permettono al ricercatore di osservare: ciò che è tipico e centrale, le distinzioni di significato, i pattern relativi a ogni significato e le sfumature di significato (Hunston 2002: 42).

Più in dettaglio, Hunston (2002) spiega che ciò che è tipico e centrale è l'insieme dei significati o pattern più frequenti di una parola o espressione (p. es. la sequenza *recipe for* è più probabile che sia seguìta da termini negativi piuttosto che positivi (p. es. *damage*, *failure*, *chaos*, *slump*) (p. 43)). Le distinzioni di significato, invece, sono quelle legate ai contesti d'uso specifici dei termini (p. es. gli aggettivi *sheer*, *pure*, *complete*, *utter* e *absolute*, benché siano considerati sinonimi da vari dizionari, sono usati in contesti differenti (p. 45)). I pattern dei significati, poi, sono le preferenze semantiche e le

colligazioni dei termini esaminati (p. es. il verbo *condemn* ha diversi significati che sono associati a pattern differenti: il significato 'critica' è associato al pattern 'condemn something as something', come in '*The Fable of the Bees was tried and condemned as a public nuisance by the Grand'*; il significato 'condanna' è associato al pattern 'condemn something to something', come in *'people who a few days before had condemned him to death'*, etc. (p. 47)). Infine, l'osservazione del dettaglio è l'individuazione di ciò che è più peculiare riguardo ad un certo termine o pattern (p. es. *advice as to* spesso è seguito da un verbo che indica 'wanting' (volizione), come in '*I would need his advice as to how to cut around his ears*' (p. 51)).

Più in generale, le *concordances* mostrano le collocazioni, cioè i termini con cui una certa parola ricorre più caratteristicamente. Le collocazioni permettono di ottenere un profilo semantico della parola esaminata e determinare l'appartenenza di una parola ad un certo campo semantico. Per queste ragioni, le collocazioni possono essere considerate le basi per identificare altri pattern, ossia *colligazioni*, *preferenze semantiche* e *prosodie semantiche*. Le colligazioni sono le relazioni grammaticali che una certa parola instaura con altre parole nelle sue vicinanze (p. es. la parola *cases* viene frequentemente usata con parole appartenenti alla categoria grammaticale dei quantificatori, come in *some cases*, *many cases* (Stubbs (2002: 65)); la preferenza semantica si riferisce alla probabilità che una certa parola si collochi con una serie di parole semanticamente correlate (p. es. la parola *commit* tende ad essere accompagnata da parole come *suicide*, *crime*, *murder*, che condividono il campo semantico della violenza (Stubbs (2002: 64)); e la prosodia semantica fa riferimento al fatto che, se una certa parola è regolarmente usata con termini positivi o negativi, diventa portatrice di un'accezione positiva o negativa anche al di fuori del contesto in cui essa è usata (p. es. il verbo frasale *set in* ha per soggetto più tipico qualcosa di spiacevole, come *rot*, *decay*, *malaise*, e di conseguenza ha acquisito un'accezione negativa (Sinclair (1987f, 1991), citato in Tognini-Bonelli (2001: 111)).

Le principali questioni affrontate nella presente ricerca sono le seguenti:

1) Quali sono i temi e le nozioni maggiormente menzionate?

2) Come sono presentati l'Islam e i musulmani?

3) I media riportano notizie più positive o più negative riguardo all'Islam e ai musulmani?

4) Quali sono le somiglianze e le differenze nella rappresentazione dell'Islam e dei musulmani nei due giornali?

I suddetti quesiti della ricerca sono stati operazionalizzati come segue:

1) Quali sono le parole, le combinazioni di parole più frequenti e le parole chiave nel *New York Times* e nel *The Dawn* identificabili attraverso le funzioni *frequency wordlist*, *N-grams* e *keyword* list in *Sketch Engine*?

2) Qual è il profilo lessicale, semantico e grammaticale del co-testo delle parole più frequenti e delle parole chiave più caratteristiche nel *New York Times* e nel *The Dawn* rilevabile con le funzioni *Concordances* e *Word Sketch* in *SketchEngine*?

3) Le cinque parole più frequenti e le cinque parole chiave più caratteristiche dei due corpora sono portatrici di una prosodia semantica positiva o negativa, considerando il significato positivo o negativo dei loro collocati?

4) Quali sono le somiglianze e le differenze che caratterizzano i collocati delle parole *Islam*, *Muslim* e *Islamic* nei due giornali rilevabili con la funzione *Word Sketch Difference*?

I risultati dell'analisi sul corpus del *New York Times* sono i seguenti:

1) Le parole, le combinazioni di parole più frequenti e le parole chiave sono state assegnate ai seguenti campi semantici: RELIGIONE (p. es. *faith*, *god*, *imam*), POLITICA (p. es. *government*, *law*, *leaders*), GEOGRAFIA (p. es. *American*, *country*, *East*), VIOLENZA (*attacks*, *violence*, *war*) e TERRORISMO (*who are the Taliban*, *Al Quaeda and*, *Mr. Bin Laden*), indicando così che i media si riferiscono all'Islam e ai musulmani in termini di pericolo e negatività.

2) L'analisi delle concordanze delle cinque parole più frequenti (*muslim*, *islam*, *islamic*, *religious*, *american*) e delle cinque parole chiave più caratteristiche (*uighur*, *xinjiang*, *islamism*, *islamically*, *shariah*) nel corpus del *New York Times* e l'analisi dei loro *Word Sketches* ha condotto alle seguenti osservazioni riguardo al profilo semantico, lessicale e grammaticale del loro co-testo: A) i pronomi *they*/*them*, e non *we/us*, sono usati per riferirsi all'Islam e ai musulmani, indicando così che questi ultimi sono soggetti al fenomeno dell''othering', ossia sono trattati in modo diverso, e generalmente ritenuti inferiori, rispetto al proprio gruppo sociale e/o etnico di appartenenza; B) i collocati frequenti della parola *Islamic* rientrano nel campo semantico della FINANZA (21.5% delle volte; p. es. *bank*, *banker*, *economy*), indicando così un focus sull'economia

islamica; C) alcuni dei più frequenti collocati della parola *American* denotano parole relative all'Islam o ai musulmani (il 20% delle volte; p. es. *American Muslim*, *American Mosque*, *ummah*), indicando così un'attenzione su fenomeni relativi all'Islam nazionali, piuttosto che quelli internazionali.

3) Le parole frequenti e le parole chiave, che per la maggior parte sono relative all'Islam, sono spesso usate vicino a termini che denotano nozioni di conflitto e di violenza; di conseguenza sono caratterizzate da una prosodia semantica negativa: *Muslim* (p. es. *cornern about*, *mistrust between*, *oppression*); *Islam* (p. es. *are blaming an*, *obscure*, *destroy the*); *religious* (p. es. *adversary*, *concern*, *crisis*); *Uighur* (p. es. *deadly attacks by*, *violence by*, *detain*); *Islamism* (p. es. *law against*, *to combat*, *terror on*); *Shariah* (p. es. *they seem worried*, *with public fogging*, *rejected by*); *American* (p. es. *cultural invasion*, *enemy*, *occupiers*).

I risultati dell'analisi sul corpus del *The Dawn* sono i seguenti:

1) Le parole, le combinazioni di parole più frequenti e le parole chiave sono state assegnate ai seguenti campi semantici: RELIGIONE (p. es. *faith*, *god*, *imam*), POLITICA (p. es. *government*, *law*, *leaders*), GEOGRAFIA (p. es. *American*, *country*, *East*), e FINANZA (*Islamic financial products and*, *conventional products and*, *financial products and services*). Contrariamente a ciò che è emerso nel corpus del *New York Times*, nessun campo semantico relativo al conflitto o alla violenza è stato identificato.

2) L'analisi delle concordanze delle cinque parole più frequenti (*islamic*, *islam*, *muslim*, *pakistan*, *religion*) e delle cinque parole chiave più caratteristiche (*ulema*, *islamism*, *mullah*, *socialism*, *Quran*) nel corpus del *The Dawn*, e l'analisi dei loro *Word Sketches*, ha condotto alle seguenti osservazioni riguardo al profilo semantico, lessicale e grammaticale del loro co-testo: A) anche in questo corpus, i pronomi *they*/*them*, e non *we/us*, sono stati usati per riferirsi ai musulmani e all'Islam, fatto che potrebbe indicare il fenomeno dell''othering'; B) collocati frequenti (12% delle volte) della parola *Muslim* sono stati assegnati al campo semantico della QUANTITÀ (p. es. *4000*, *many*, *the whole*), indicando così che i media Pakistani tendono a rappresentare i musulmani come un gruppo anziché come individui.

3) Le seguenti parole frequenti e parole chiave, la maggior parte relative all'Islam, sono state spesso usate vicino a termini che denotano nozioni di conflitto e di violenza; di conseguenza sono caratterizzate da una prosodia semantica negativa: *Islamic*

(p. es. *bomb*, *deaths*, *terrorist*); *Islam* (p. es. *against*, *blaming*, *demonizing*); *Muslim* (p. es. *against*, *brutal*, *selfish*); *Pakistan* (p. es. *defaming*, *against minorities*, *has no future*); *religion* (p. es. *lame and*, *as a weapon*, *is close to insulting*); *Islamism* (*is also invalid*, *scorned at them*, *started to weaken and fragment*); *Mullah* (p. es. *terrorism and*, *mindless*, *manipulations*); *socialism* (p. es. *aka Nazism*, *are not compatible with Islam*, *as harnessing the freedom of thought*).

4) Infine, per esplorare più approfonditamente le somiglianze e le differenze tra i due corpora nella rappresentazione dell'Islam e dei musulmani, ho usato la funzione *Word Sketch Difference* per identificare collocazioni delle parole *Islam*, *Muslim* e *Islamic*, sia quelle in comune sia quelle non-condivise. Quest'ultima analisi ha condotto alle seguenti osservazioni: A) il giornale americano si focalizza maggiormente sull'Islam radicale, mentre il giornale pakistano si focalizza maggiormente sull'Islam liberale, come rivelano i modificatori più tipici della parola *Islam* (p. es. *radical*, *Fatah*, *Sarekat* rispetto a *Liberal*); B) i collocati della parola *Islam* che hanno una prosodia semantica negativa sono più frequenti nel corpus del *New York Times* (*terrorism*, *incompatible*) rispetto a quelli del *The Dawn* (*war*); C) nel corpus del *New York Times*, l'Islam è stato maggiormente contestualizzato all'interno di un discorso sulla religione come rivelano i suoi collocati (p. es. *practice*, *embrace*, *Sunni*), ma non nel corpus del *The Dawn*, più focalizzato sulla politica (p. es. *militant*, *Fatah*, *moderate*); D) nel corpus del *New York Times*, alcuni collocati di *Muslim* sono stati assegnati al campo semantico del crimine (p. es. *terrorist*, *soldier*, *inmate*, *prisoner*), mentre quest'ultimo è assente nel corpus del *The Dawn*; E) i nomi modificati dalla parola *Muslim* che hanno una prosodia semantica negativa sono più frequenti nel corpus del *New York Times* (*ban*, *inmate*, *prisoner*, *terrorist*) rispetto a quelli identificati nel corpus del *The Dawn* (*revolt*); F) c'è un focus sui musulmani indiani nel corpus del *The Dawn*, e invece sui musulmani americani nel corpus del *New York Times*; G) nel corpus del *The Dawn*, alcuni collocati della parola *Islamic* sono stati assegnati al campo semantico della GUERRA (p. es. *terrorist*, *force*, *militia*), mentre questo campo semantico è assente nel corpus del *New York Times*; H) nel corpus del *New York Times*, c'è un focus sulla finanza islamica, come suggeriscono i termini più tipici modificati dalla parola *Islamic* (p. es. *bank*, *bond*, *investment*); I) d'altra parte, nel corpus del *The Dawn*, c'è un'enfasi sulla politica Islamica, particolarmente sul socialismo, come si può notare dai termini più tipici modificati dalla parola *Islamic* (e.g.

socialism, *nation*, *regime*); J) parole valutative che si riferiscono a *Islamic* che hanno una prosodia semantica positiva sono emerse (*glorious*, *great*) nel corpus del *The Dawn*, mentre sono assenti nel corpus del *New York Times*.

I risultati di questo lavoro confermano parzialmente le scoperte riportate da studi precedenti: da un lato si sono riscontrate somiglianze tra i due corpora nella rappresentazione dell'Islam; dall'altro sono emerse anche delle differenze in linea con studi precedenti.

Tra le somiglianze si è notato che, sia nel *New York Times* che nel *The Dawn*, l'Islam e i musulmani sono frequentemente rappresentati in termini di violenza e conflitto, viene fatta spesso una distinzione tra 'noi' (non musulmani) e loro (musulmani), e la maggior parte dei collocati delle parole che si riferiscono all'Islam e ai musulmani hanno una prosodia semantica negativa. Inoltre, inaspettatamente, è anche emerso che, sebbene il *The Dawn* sia considerato il giornale portavoce della Lega Islamica, il fenomeno dell''othering' è stato rilevato nei collocati della parola *Muslim*. Questa scoperta può essere considerata come una tendenza da parte dei media pakistani di presupporre che la maggior parte dei lettori degli articoli del *The Dawn* professino una religione diversa dall'Islam. Questo dato è in contrasto da quanto emerso dallo studio condotto da Mahmood e altri (2018) sugli editoriali del *New York Times* e del *The Dawn* sulla rappresentazione dei terroristi talebani rispetto alle vittime uccise e/o ferite: nell'editoriale del *New York Times*, i talebani sono stati criticati, mentre nell'editoriale del *The Dawn*, i talebani non sono stati esplicitamente accusati per l'attacco. Nello studio di Mahmood e altri (2018), la differente attitudine sostenuta dai media pakistani potrebbe essere considerata come una tendenza a mitigare le azioni dei loro compatrioti, ossia i talebani. Nel presente studio, nessun tipo di strategie mitigative è stato rilevato.

Tra le differenze, si è riscontrato che i due giornali rappresentano la stessa realtà sociale con un taglio diverso: nel giornale americano, i campi semantici del conflitto e della violenza sono più frequenti rispetto al giornale pakistano; il giornale pakistano si focalizza maggiormente sulla politica rispetto al giornale americano; il giornale americano si focalizza sull'Islam radicale, mentre il giornale pakistano si focalizza sull'Islam liberale.

Tornando ai quesiti della ricerca, i risultati si possono riassumere così:

1) I temi e le nozioni maggiormente menzionati nei due corpora sono RELIGIONE, POLITICA, GEOGRAFIA. Inoltre, altri frequenti temi solo nel corpus del *New York Times* sono VIOLENZA e TERRORISMO. Un ulteriore tema frequente solo nel corpus del *The Dawn* è FINANZA.

2) Nei due corpora l'Islam e i musulmani sono spesso presentati attraverso il fenomeno dell''othering', ossia in contrasto rispetto al gruppo etnico o sociale di appartenenza. Tuttavia, oltre a questo, nel corpus del *New York Times* l'Islam è rappresentato anche in termini economici e con un focus su fenomeni relativi all'Islam nazionali piuttosto che internazionali; mentre nel corpus del *The Dawn*, i musulmani sono spesso rappresentati come un gruppo anziché come individui.

3) I media riportano più notizie negative che positive riguardo all'Islam e ai musulmani. In particolare, questi ultimi vengono frequentemente menzionati all'interno di una narrazione di conflitto e violenza in entrambi i corpora.

4) Una somiglianza tra i due corpora consiste nella prosodia semantica negativa che caratterizza i termini che si riferiscono alle parole che denotano l'Islam e i musulmani. Tuttavia, nel corpus del *New York Times* c'è una maggiore frequenza di parole con prosodia semantica negativa rispetto al corpus del *The Dawn*. Inoltre, nel corpus del *New York Times* c'è un focus sull'Islam radicale, i musulmani americani e la finanza islamica. Invece, nel corpus del *The Dawn* c'è un focus sull'Islam liberale, i musulmani indiani e sulla politica, in particolare sul socialismo.

Questo studio presenta alcune limitazioni:

1) i corpora del *New York Times* e del *The Dawn* non hanno dimensioni simili;

2) la campionatura temporale dei testi non è stata precisa: sebbene in entrambi i corpora ci siano testi del 1990, 2000 e 2020, non ho potuto appurare quanti testi esemplifichino ciascun decennio;

3) solamente una quantità limitata di dati è stata analizzata, ossia le 200 parole più frequenti, le 200 combinazioni di parole più frequenti e le 200 parole chiave più tipiche;

4) i corpora non sono stati perfettamente "puliti": il corpus del *New York Times* conteneva anche stringhe di parole la cui funzione era quella di guidare il pubblico alla lettura degli articoli (per esempio, *Reprints Timesmachine is an*, *benefit for home delivery*, *home delivery and digital*), mentre il corpus del *The Dawn* conteneva parole

appartenenti a commenti scritti dai lettori del giornale (per esempio, *recommend*, *abbastoronto*, *Jul*);

5) le parole chiave negative, ossia quelle parole che sono meno frequenti rispetto a quanto ci si possa aspettare, non sono state esaminate poiché non potevano essere elaborate da *Sketch Engine*.

In futuri progetti di ricerca sarebbe interessante analizzare come le donne musulmane vengono rappresentate dalla stampa americana e medio-orientale, accertando se sono descritte negativamente, come vittime della propria cultura, o se prevale una rappresentazione differente nell'uno e nell'altro contesto mediatico. Inoltre, si potrebbero analizzare i modi in cui la stampa americana rappresenta i musulmani americani rispetto ai musulmani non-americani, verificando, cioè, se i musulmani americani vengono rappresentati più positivamente, essendo considerati compatrioti, e se i musulmani non-americani vengono presentati più negativamente, essendo considerati forestieri. Allo stesso modo, sarebbe interessante indagare sulle possibili differenze nella rappresentazione dei cristiani e dei musulmani da parte dei media occidentali. In tutti questi casi, un'analisi lessicale dei dati dal basso verso l'alto, simile all'analisi condotta in questo lavoro, rivelerebbe pattern linguistici e farebbe luce sui punti di vista adottati dai giornalisti su certi gruppi sociali e/o etnici.

References

Abdul-Rakkz, Noor Alaa; Khalil, H. Huda (2020) 'Linguistic evaluation of news reports on Islamophobic incidents', *International Linguistics Research*, 3(1): 46-56.

Acim, Rachid (2015) 'A critical linguistic study of the representation of Islam and Muslims in the discourse of the New York Times op-eds', *International Journal of Humanities and Cultural Studies*, 2(1): 29-41.

Ahmed, Saifuddin; Matthes, Jörg (2017) 'Media representation of Muslims and Islam from 2000 to 2015: a meta-analysis', *The International Communication Gazette*, 79(3): 219-244.

Baker, Paul (2006) *Using corpora in discourse analysis*, London & New York: *Continuum*.

Baker, Paul; Gabrielatos, Costas; Khosravinik, Majid; Kryzanowki, Michal; McEnery, Tony; Wodak, Ruth (2008) 'A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press', *Discourse and Society*, 19(3): 273-306.

Baker, Paul; Gabrielatos, Costas; Tony, McEnery (2013) 'Sketching Muslims: a corpus driven analysis of representations around the word 'Muslim' in the British press 1998-2009, *Applied Linguistics*, 34(3): 255-278.

Baker, Mona; Francis, Gill; Tognini-Bonelli, Elena (1993) *Text and technology*, Amsterdam & Philadelphia: John Benjamins Publishing Company.

Bowker, Lynne; Pearson Jennifer (2002) *Working with specialized language: a practical guide to using corpora*, London and New York: Routledge.

Britannica online (https://www.britannica.com/biography/Gamal-Abdel-Nasser). Last access: 15 January 2022.

Cambridge dictionary online (https://dictionary.cambridge.org/it/dizionario/inglese/mullah). Last access: 07 January 2022.

Cervi, Laura; Tejedor, Santiago; Gracia, Monica (2021) 'What kind of islamophobia? Representation of Muslims and Islam in Italian and Spanish media', *Religions*, 12(6): 427.

Collins Dictionary online (https://www.collinsdictionary.com/it/dizionario/inglese/azan). Last access: 08 January 2022.

Collins Dictionary online (https://www.collinsdictionary.com/dictionary/english/baath). Last access: 13 January 2022.

Collins Dictionary online (https://www.collinsdictionary.com/it/dizionario/inglese/hadith). Last access: 03 January 2022.

Collins Dictionary online (https://www.collinsdictionary.com/it/dizionario/inglese/masjid). Last access: 03 January 2022.

Collins Dictionary online (https://www.collinsdictionary.com/it/dizionario/inglese/ummah). Last access: 07 January 2022.

Dheskali, Vincenzo (2020) 'Attacks by the Hamas Resistance Movement or Terrorist Movement, Israeli Defence Forces or Occupation Forces? A corpus-based comparison of the framing of the Israeli-Palestine conflict in American, Arab-Speaking, Israeli and Palestinian online journals', *Conflicting Truths in Academic and Journalistic Writing*, 17: 51-74.

Fumo, Olimpia (2021) *Análisis lingüístico comparativo de la prosodia semántica en la prensa Italiana y Española: el discurso periodístico sobre la inmigración*. Tesi di laurea magistrale in Lingue Moderne per la Comunicazione e la Cooperazione Internazionale, Università di Padova.

Gandhara (https://gandhara.rferl.org/a/the-rise-of-the-new-pakistani-taliban/31261608.html). Last access: 07 January 2022.

Huffpost (https://www.huffpost.com/entry/nadeem-farooq-paracha-abo_b_960013). Last access: 07 January 2022.

Hunston, Susan (2002) *Corpora in applied linguistics*, Cambridge: *University Press*.

Learn Religions (https://www.learnreligions.com/islamic-abbreviation-swt-2004291). Last access: 07 January 2022.

Lemmouh, Zakaria (2008) 'A critical linguistic analysis of the representation of Muslims in The New York Times', *Hermes – Journal of Language and Communication Studies*, 40: 217-240.

Mahmood, Tariq; Kausar, Ghazara; Khan, Gul, Zamin (2018) 'A critical discourse analysis of the editorials of "Dawn" and "The New York Times" in the aftermath of

Army Public School attack. The "Us" versus "Them" ideology', *Journal of Research in Social Sciences*, 6(2): 1-18.

MacMillan Dictionary online (https://www.macmillandictionary.com/dictionary/british/othering). Last access: 31 December 2021.

Merriam-Webster Dictionary online (https://www.merriam-webster.com/dictionary/Abbasid). Last access: 07 January 2022.

Merriam-Webster Dictionary online (https://www.merriam-webster.com/dictionary/hadith). Last access: 08 January 2022.

Merriam-Webster Dictionary online (https://www.merriam-webster.com/dictionary/waqf). Last access: 07 January 2022.

Merriam-Webster Dictionary online (https://www.merriam-webster.com/dictionary/zakat). Last access: 07 January 2022.

Ming, Liu (2021) 'One city, two stories: competing discursive constructions of post-colonial Hong Kong in Chinese and American newspapers', *Critical Discourse and Corpus Approaches to Systemic Functional Grammar*, 16: 143-159.

Muhammad, Junaid Ghauri; Salma, Umber (2019) *Exploring the nature of representation of Islam and Muslims in the Australian press*, *SAGE Open*, 9(4): 1-9.

Noor Alaa, Abdul-Rakkz; Huda, H. Khalil (2020) 'Linguistic evaluation of news reports on islamophobic incidents', *International Linguistics Research*, 3(4): 46-56.

SketchEngine (https://app.sketchengine.eu/#dashboard?corpname=user%2Fkatherine.ackerley%2Fchristmas_jumpers). Last access: 31 December 2021.

Stubbs, Michael (2002) *Words and phrases, corpus studies of lexical semantics*, Oxford: Blackwell Publishing.

Tognini-Bonelli, Elena (2001) *Corpus Linguistics at Work*, Amsterdam & Philadelphia: John Benjamins Publishing Company.

Zein, Bani Younes; Isyaku, Hassan; Mohd Nazri, Latiff Azmi (2020) 'A pragmatic analysis of Islam-related terminologies in selected Eastern and Western mass media', *Arab World English Journal*, 11(2): 70-84.