# Volume 11 (2016)

## Contributors to this Issue:
Mona Abdel-Fadil
Susanne van der Beek
Camden Behrens
Frank G. Bosman
Claudia Carvalho
Gabriel Faimau
Polykarpos Karamoyzhs
Michalis Keffalas
Setareh Malekpour
Emad Mohamed
Tom van Nuenen
Stefan Piasecki
Moisés Sbardelotto
Sasha A.Q. Scott

© 2016

# Table of Contents

# *Jewish*, *Christian* and *Islamic* in the English Wikipedia

Emad Mohamed

## Abstract

In order to study how Judaism, Christianity, and Islam are represented in Wikipedia, I use corpus linguistics tools to extract the adjective noun collocates of the adjectives *Jewish*, *Christian*, and *Islamic* from the 2013 English Wikipedia in order find out their semantic prosody. I then rank the positive and negative noun collocates using the logdice scores in order to find whether there is a statistically significant difference between them. In the case of negative nouns, an ANOVA test found a statistically significant difference. Pair-wise comparisons suggest that Islamic is more negative than either Christian or Jewish while there is no statistically significant difference between Jewish and Christian. On the positive side, there is no statistically significant difference between the adjectives. Intra-adjectival comparisons suggest that there is no statistically significant difference between Islamic's positive and negative collocates while both Christian and Jewish are more positive than negative.

## Keywords

# 1    Introduction

Wikipedia is probably the largest book of facts available, and it may be the main source of information for millions of web users. It is one of the largest reference websites as it attracts 470 million unique visitors every month (https://en.wikipedia.org/wiki/Wikipedia:About). Wikipedia is continually growing and changing, and many articles get updated regularly, but "[o]lder articles tend to be more comprehensive and balanced; newer articles may contain misinformation and/or unencyclopedic content."

The English Wikipedia is the largest of Wikipedias with over 4 million pages of content. The English edition differs from many others in that it is global since English acts as the de facto lingua franca of the world and many non-native speakers of English use and edit the encyclopedia on a regular basis. Discovering how a certain thing is described in Wikipedia may give us an idea about

how that thing is perceived globally. In this paper, I attempt to discover how the three Abrahamic religions (Judaism, Christianity, and Islam ) are represented in the English Wikipedia through an examination of their attributive adjectives and the nouns they habitually modify.

This study aims to examine the positive and negative lexical associations of the three adjectives (*Jewish*, *Christian*, and *Islamic*) in the English Wikipedia. The paper seeks to answer two questions:

1.  Which negative/positive nouns do these adjectives tend to modify? and

2.  Is there a statistically significant difference between the semantic prosodies of these adjectives?

The answers to these questions may be useful, not only for linguistics purposes, but also for the study of religion and politics, but before providing answers, I will first briefly define *semantic prosody*.

## 1.1   What is Semantic Prosody?

Sinclair (1991:74) noticed that the verb SET IN has a tendency to have as its subject nouns that are not "conventionally desirable or attractive". The most common subjects of SET IN were *rot*, *decay*, *malaise*, *despair*, *ill-will*, *decadence*, and *impoverishment*. Louw, who first coined the term, defines **semantic prosody** as a "consistent aura of meaning with which a form is imbued by its collocates" (Louw, 1993: 157).

The main function of SP is to express the speaker/writer evaluations, which Stubbs (1996: 176) classifies into negative, positive and neutral semantic prosody. McEnery and Xiao (2006: 83) note that "[s]emantic prosodies are typically negative, with relatively few of them bearing an affectively positive meaning." Collocation and SP are very closely related. It may be, as Partington (2004) notes, that humans have more need for conveying bad news than good ones. Partington (2004) examined the verbs of the HAPPEN family and found that they mostly co-occur with bad news since it is usually bad news that is newsworthy. This human tendency may not be limited to news or to verbs.

Lou (ibid.:164) seems to see semantic prosody as a diachronic process as he states that "prosodies are undoubtedly the product of a long period of refinement through historical change". This focus on diachronicity is more expressed in Bublitz (1996:11): "we know from lexical semantics that constantly using a word in the same kind of context can eventually lead to a shift in its meaning: the word adopts semantic features from an adjacent item", or as Hoey (2005:8) states:

As a word is acquired through encounters with it in speech or writing, it becomes cumulatively loaded with contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. The same applies to word sequences built out of these words; these too become loaded with the contexts and co-texts in which they occur.

Semantic prosody and collocates thus go hand in hand. In the words of Xiao and McEnery (2006:84):

On the one hand, the item does not appear to have an affective meaning until it is in the context of its typical collocates. On the other hand, if a word has typical collocates with an affective meaning, it may take on that affective meaning even when used with atypical collocates. As the Chinese saying goes, 'he who stays near vermilion gets stained red, and he who stays near ink gets stained black'— one takes on the colour of one's company—the consequence of a word frequently keeping 'bad company' is that the use of the word alone may become enough to indicate something unfavourable.

Adopting the suggestion that a word may be stained by its company, and that this stain may stick even when the word is no longer accompanied, I investigate the collocational patterns of three adjectives. The adjectives *Jewish*, *Christian*, and *Islamic* may not be positive or negative in isolation, but they may combine with nouns that label them as such. They are adjectives that may have a neutral, negative or positive meaning potentials that are activated when they modify certain nouns. For example, ***Jewish philanthropist*** ascribes philanthropy to Jewishness while *Jewish criminal* does otherwise. This is in line with the suggestion that "*You shall know a word by the company it keeps*" (Firth, 1957:11). Things are not usually clearcut. While *Jewish criminal* is kind of obvious, what about *Jewish prisoner*? I do not consider this to be a negative collocate since this may be in the context of World War II, and prisoner by itself, unlike criminal, does not seem to pass a value judgment on the person it denotes. A less clear case is that of ***inmate***.

Semantic prosody is related to evaluation, or judgment by the speaker. Because the verb SET IN is usually used with negative subjects, a speaker who says *"the cold weather set in"* is more likely to be expressing a personal attitude than the one who utters *"the cold weather started"* (Stewart, 2010: 22). This may be true when you have alternatives, but sometimes, when there is no other way of saying it, the same verb that naturally conveys an attitude may just be used with negative subjects or objects with no attitudinal strings attached (ibid:22-3)

There have been studies that used collocates to explore how Muslims are represented in the media. Baker et al. (2013) examined a 143 million word corpus of British newspapers from 1998 to 2009 to find out how the word *Muslim* was represented. They found that the categories 'ethnic/national identity, characterizing/differentiating attributes, conflict, culture, religion, and

group/organizations' were referenced with the conflict category being especially lexically rich. Baker et al's paper did not focus on the negative or positive associations but it found that the nouns *extremist*, *fanatic*, and *terrorist* rank 10th, 18th, and 23rd on the collocate list of *Muslim*. Sadar (2014) used Critical Discourse Analysis (CDA) to study the image of veiled Muslim women in the British press between 2001 and 2014. Her study is, however, non-corpus-based. None of these two studies is especially concerned with semantic prosody. I am not aware of any corpus-based study of *Christian* or *Jewish*. No one seems to have compared Jewish, Christian and Islamic, and no one seems to have studied collocational patterns and their associated semantic prosodies in Wikipedia, which I try to do here.

The rest of this paper goes as follows: in section 2 I introduce the corpus and the methodology, in section 3 I present the results discuss some ramifications, and in section 4 I conclude the paper and suggest further research. At the end of the article is an appendix of the top collocates of *Jewish*, *Christian* and *Islamic*.

## 2    Data and Methods

Figure 1 summarizes the data and methods used in this paper.

| English Wikipedia 2013 Text Dump |
| Part of Speech Tagging (Stanford Tagger) |
| Dependency Parsing (MaltParser) |
| Lemmatization (NLTK Wordnet Lemmatizer) |
| Dependency Ngram Extraction |
| Ngram Scoring using Logdice |
| Extract Negative and Positive Collocates |
| Check for Differences Using ANOVA |

*Figure 1: A summary of the data and methods*

We use the Wikipedia 2013 text dump available from ([http://kopiwiki.dsd.sztaki.hu/](http://kopiwiki.dsd.sztaki.hu/)). To make things easier, we extract only those paragraphs that contain the words (*Jewish*, *Judaic*, *Christian*, *Islamic*, *Muslim*). This resulted in 759344 paragraphs and 53698038 words (not counting punctuation). The corpus was then processed through a pipeline of part of speech tagging, dependency parsing and lemmatization.

## 2.1   Part of speech tagging

For POS tagging, I used the Stanford tagger (Toutanova et al, 2003). I chose the Stanford tagger due to its very high accuracy (97.24% on the WSJ corpus). Given the sentence in (a), the Stanford tagger produces the sentence with grammatical tags assigned to the words (b). The tagger's role is then that of disambiguation since a word like **drive** could both be a noun and a verb. The tagger output is used for both lemmatization and dependency parsing.

   a. Melissa drives a white car.

   b. *Melissa_NNP drives_VBZ a_DET white_JJ car_NN ._.*

In the context of this study, and generally, both *Muslim* and *Christian* are ambiguous between nouns and adjectives, and I am interested only in their roles as attributive adjectives. Their adjectivehood is taken care of by the POS tagger while the dependency parser detects the attributive part.

## 2.2   Dependency parsing

In this study, I use dependency parsing to extract dependency bigrams. Dependency parsing models syntactic relations in sentences as binary relations between lexical items. For example, in the sentence *Melissa drives a white car.*, we can see a number of relationships: (a) drives *governs* Melissa and assigns it the role of *nominal subject (nsubj)*, (b) drives *governs* car and assigns it the role of *direct object (dobj)*, (c) *car* governs white and assigns it the role of adjectival modifier *(amod),* and governs *a* and assigns it the role of determiner *(det).*
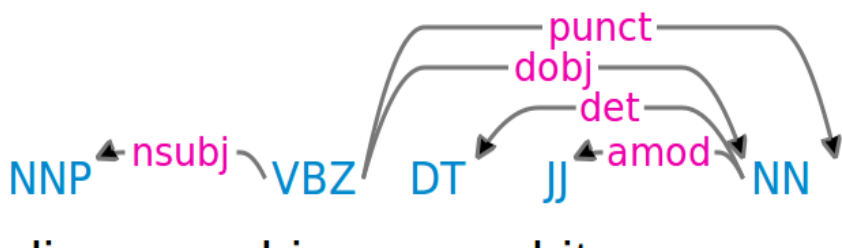


*Figure 2: A dependency graph of the sentence Melissa drives a white car.*

Each one of these relations is a dependency bigram, and in a large enough text, we can use these dependency bigrams to obtain a picture of the object of investigation. In this study, I use dependency bigrams to profile the three adjectives *Jewish*, *Christian*, and *Islamic* in the Wikipedia corpus. If "a word is known by the company it keeps", then heads and dependents are not merely neighbours, they are neighbors with close ties. I used the dependency parser MaltParser (Nivre, 2009) for extracting the dependency relations from the corpus. While the Stanford Parser can be used for the same purpose, MaltParser is much faster, which enabled me to finish parsing the relatively big corpus on an ordinary laptop in good time.

I focus on syntagmatic relations, rather than mere neighborhood, because these allow me to know what roles are played by our nouns of interest. Within the domain of lexicography, Hanks (2013: 93-5) states that in oder to know the meaning of a noun, "[o]ne must ask questions of two rank levels: (1) how the noun normally goes with modifiers within a phrase, and (2) how it normally functions as the subject, object, or prepositional object within a clause (what clause role does it normaly take?)." (Original emphasis). While my main focus here is on adjectives, adjectives modify nouns which can , in turn, be modified by other adjectives, establishing a network that may prove useful for studying words' semantic prosody.

The dependency parser output has 64509430 dependency bigrams including 3930762 relations in which an adjective modifies a noun. The pattern ***Christian_JJ + Noun*** occurs 134650 times, ***Jewish_JJ +Noun*** occurs 205969 times, ***Judaic + Noun*** occurs 1314 times, ***Islamic_JJ +*** Noun occurs 80663 times and ***Muslim_JJ*** + Noun occurs 95929 times.


## 2.3   Lemmatization

Lemmatization normalizes the word forms as it converts the verb *drives*, *drive*, *driving* and *driven* to the base form *drive*, thus enabling us to know that there is an object relation between *drive* and *car* regardless of the form of the verb or the form of the noun (car or cars). In this study, lemmatization is performed by the NLTK Wordnet-based lemmatizer (Bird et al, 2009). All the collocates used in this study are lemmatized collocates. The bigram ***Christian community*** covers both Christian ***community*** and Christian ***communities***.

## 2.4   Ranking the collocates

The collocates are ranked by the logdice score (Rychlý, 2008). The logdice score has the following features :

- Theoretical maximum is 14, in case when all occurrences of X co-occur with Y and all occurrences of Y co-occur with X. Usually the value is less then 10.

- Value 0 means there is less than 1 co-occurrence of XY per 16,000 X or 16,000 Y. We can say that negative values mean there is no statistical significance of XY collocation.

- Comparing two scores, plus 1 point means twice as often collocation, plus 7 points means roughly 100 times frequent collocation.

- The score does not depend on the total size of a corpus. The score combine relative frequencies of XY in relation to X and Y.

## 2.5   Extracting Negative and Positive Collocates

I limit my analysis to the top 200 collocates of each adjective. I adopt the following process:

1. **Treat *Islamic* and *Muslim* as the same and one adjective**. The same hold true for *Judaic* and *Jewish*. This does not mean that the adjectives mean exactly the same thing. In fact, *Islamic* is different from *Muslim* in some of its associations and so is the case for *Jewish* and *Judaic*. I do so for reasons of convenience.

2. **Extract the top 200 noun collocates of each adjective**. This results in the total of 600 nouns. The number is arbitrary and does not carry any significance. The nouns chosen are those that govern the adjective in a dependency graph.

3. **Extract the positive and negative nouns from the set of collocate nouns**. This is possibly the most subjective step in the process. I assign to the positive category those nouns that I feel are negative, and to the positive category those nouns that I feel are positive. One further complication is that I am not a native speaker of English, and my judgment may be blurred by my native tongue, or by a misunderstanding of cultural norms. I do, however, believe that my judgment will be generally correct, occasional errors and misjudgments notwithstanding.

4. **Compare the ranks in the three adjectives**. Using the logdice scores, compare the adjectives' scores on the nouns. If a certain noun is missing, assign it a logdice score of -3. This slightly less then the lowest recorded logdice in the 600 nouns. In order to determine

whether there is a statistically significant difference between the scores of the collocates among these adjectives, we use the One-way Analysis of Variance test.

# 3    Results and Discussion

## 3.1    Answer to Question 1:

### 3.1.1. Negative Nouns and Scores

community people family population law encyclopedia history descent cemetery state woman immigrant study tradition refugee society culture life museum settlement identity parent organization school origin quarter leader child treasure holiday ancestry question heritage group education faith immigration music editor student congregation literature scholar player philosophy mother settler religion resident writer background merchant synagogue view article movement week monotheism philosopher thought men prayer text review girl mysticism boy diaspora nation home ghetto practice actor source citizen scholarship news numismatics presence father historian revolt book root world inhabitant man victim newspaper calendar homeland prisoner lobby bibliography ethic center conspiracy name camp christian neighborhood artist category musician topic surname custom institution member survivor property physician character lore scientist experience soldier worker war activist person minority ethnicity service art youth belief ethnocentrism sect leadership acceptance dictionary scripture theology resistance value emigration author composer doctor mathematician authority influence banker publication subject lawyer friend organisation force year sage business perspective poet unit studies priest isra'iliyat christum athlete rabbi ritual needlework graeco-roman positivity liturgy tribe morbidity disquiet colony theme photoman denomination elective observance iconography pornographer section figure site nefesh language brigade specificity proselytization fog kashrut veneer village stricture control website tradition

*The top 200 collocates of Jewish in descending order according to their logdice score*

church denomination faith music band school missionary group community theology belief view format organization martyr tradition station art saint leader chart name democracy doctrine value life album education ministry literature college radio writer theologian artist teaching movement sect perspective family mission world population nation rock history woman scientist worship minister religion cross theme festival symbol writing ethic service hymn song party author culture scholar principle force site organisation institution country charity apologist university kingdom monk thought congregation state philosophy king musician burial text source era democrat influence society priest practice period apologetics holiday bishop singer program mysticism origin basilica people soldier article minority eschatology network child identity industry slave study message pastor metal spirituality programming convert book clergy

iconography label scripture evangelist man conference film right background living character monastery symbolism point work market fundamentalism post philosopher concept worldview thinker circle emperor prayer girl figure element mythology magazine virtue context interpretation subject meditation believer scene gospel camp idea activity pilgrim settlement politician ruler student love website cemetery environment morality village socialism preacher parent heresy classic ethos order aid root body home fundamentalist teacher response terrorism prince militia fellowship activist building center worker place monasticism household time unity show one demonology

### *The top 200 collocates of Christian in descending order according to their logdice score*

republic community law world league scholar revolution population study woman conquest country jurisprudence state calendar art culture tradition architecture society philosophy ruler army leader general terrorism fundamentalism prophet conference family civilization history minority faith student force school science center nation militant extremism dynasty view group cleric empire banking emirate rule education front extremist court teaching theology terrorist period movement organization caliphate association mosque thought troop council maghreb invasion men government soldier practice clan girl majority principle jurist influence historian literature saint text invader name belief refugee value regime youth scientist land fundamentalist party institution trader calligraphy cemetery era people term bank immigrant philosopher sect inhabitant dress economics learning guild month system mob source university concept convert medicine religion tribe man kingdom village quarter scholarship perspective control merchant holiday feminism doctrine finance area politician brother writer ideology centre editor knowledge astronomy prayer pilgrim identity code theologian charity prisoner resident rebel citizen ethic power household revival astronomer mysticism city rajput guidance presence festival site academy territory music insurgency style boy activist marriage authority edition mathematics shrine civilian scripture eschatology town governor book conqueror democracy website parent resistance thinker expansion monument fascism topic fighter context sultanate rebellion warrior intellectual attack member background heritage

### *The top 200 collocates of Islamic in descending order according to their logdice score*

*Table 1: the top 200 collocates of each adjective*

The top 200 collocate noun lemmas for each adjective are listed in Table 1. I have classified 27 of these collocated nouns as negative: *apologetics, apologist, attack, conqueror, conquest, conspiracy, ethnocentrism, extremism, extremist, fascism, fundamentalism, fundamentalist, heresy, insurgency, invader, invasion, lobby, militant, militia, mob, rebel, regime, stricture, terrorism, terrorist,* and *war.* The criterion for classifying a noun as negative is that if it associates with one of our religious adjectives, it could pass a negative value on the nature and/or adherents of the religion in question. For example, **Muslim prisoner** does not pass a negative value judgment on Muslims or Islam while *Muslim criminal* may.

|                | Jewish  | Christian | Islamic |
|----------------|---------|-----------|---------|
| attack         | 2.6837  | 2.0087    | 4.8622  |
| conqueror      | -2.5996 | 0.369     | 5.0115  |
| conquest       | -0.6958 | 2.4841    | 8.5427  |
| conspiracy     | 5.5848  | 0.8229    | 1.0361  |
| ethnocentrism  | 5.2361  | -0.3636   | -3.0    |
| extremism      | -1.0184 | 1.8484    | 7.1134  |
| extremist      | 2.5995  | 2.0308    | 6.9631  |
| fascism        | -1.0138 | 1.2182    | 4.9175  |
| fundamentalism | 3.713   | 5.0097    | 7.6058  |
| fundamentalist | 1.3983  | 4.6287    | 6.1949  |
| insurgency     | 0.7135  | -3.0      | 5.186   |
| invader        | -1.6038 | 0.3624    | 6.3211  |
| invasion       | 1.4335  | -0.4769   | 6.6416  |
| lobby          | 5.6875  | 3.276     | 1.0626  |
| militant       | 3.1413  | 2.4175    | 7.1135  |
| militia        | 3.3847  | 4.5616    | 4.6599  |
| mob            | 2.1464  | 3.0331    | 5.8065  |
| rebel          | 3.7948  | 1.724     | 5.3542  |
| regime         | -0.6593 | -1.0454   | 6.2386  |
| stricture      | 4.7796  | -1.9416   | -1.2156 |
| terrorism      | 3.6654  | 4.5698    | 7.7024  |
| terrorist      | 3.1072  | 2.7146    | 6.7726  |
| war            | 5.3437  | 1.4158    | 3.6251  |

*Table 2: Negative nouns and their logdice scores:. A value of -3 means this noun does not co-occur with the adjective in our corpus.*

I will discuss the top 2 negative nouns for each adjective:

**Jewish Lobby and Jewish Conspiracy**

*Jewish lobby* is the most salient negative collocation in which *Jewish* modifies a noun. While lobbying is a natural political activity, the association between *Jewish* and *lobby* indicates that Jews are a political entity rather than a religious group. The Wikipedia article on the term states that the term "[w]hile at times self-described, usage of the term is viewed as inaccurate, and, particularly when used to allege disproportionate Jewish influence, it can be perceived as pejorative or may constitute antisemitism". Both *Islamic* and *Christian* have *lobby* as their head noun, albeit with

ranks far down the list. *Islamic lobby* occupies the 1812th rank on the *Islmaic* collocate list while *Christian lobby* has a rank of 397 on the Christian Adj-Noun collocate list.

When we consider the word *lobby* itself, its top adjectival modifiers are: *apolitical*, **pro-israel**, *excessive*, *behind-the-curtain*, *Australian*, *middle-east*, *fruitless*, *intense*, *anti-Polish*, *intensive*, *makian*, *powerful*, *shameless*, **Z***ionist*, ***anti-Israel***, *all-powerful*, *pro-life*, *congressional*, *infamous*, and ***jewish*** . While *Jewish* is at rank 20, the second most salient adjective is *pro-Israel*. There are also the *Zionist lobby*, and the *anti-Israeli lobby*, which despite being in the opposite direction, is still strongly related. We can also notice that in the top twenty collocates of *lobby* there are negative adjectives: *excessive*, *behind-the-curtain*, *fruitless*, *shamemless*, and *infamous*, which is a good indication that the overall prosody of **lobby** is negative.
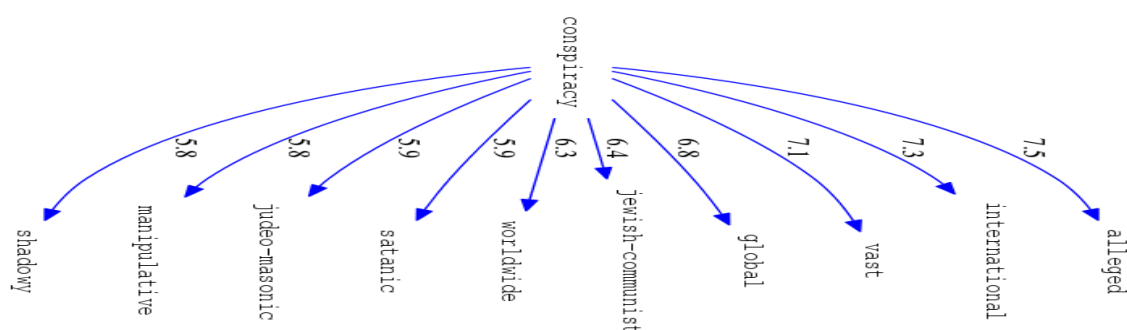


*Figure 3: The top 10 collocates of conspiracy*

*Jewish conspiracy* may refer to any conspiracy theory involving the Jews, for example the *blood libel* and *well poisoning*. In fact, in our corpus, many of the top collocates of conspiracy are Jewish: *alleged*, *international*, *vast*, *global*, ***jewish-communist***, *worldwide*, *satanic*, ***judeo-masonic***, *manipulative*, *shadowy*, *clerical*, *deliberate*, *seditious*, ***jewish***, *masonic*, ***jewish-masonic***, *hateful*, *anti-christian*, **zionist**, and *right-wing*. While the word *conspiracy* is in by itself a word of negative prosody, we can see that *Jewish conspiracy* is usually accompanied by another element: either *masonic* or *communist*. The Jewish-Communist connection denotes the anti-Jewish theory of a "secret collusion between the Jews and socialists to rule the world" (Mendes, 2014:250). de Poli (2014) explains that while the Judeo-Masonic conspiracy theory started in Europe, it is currently more common in Latin America, Asia (esp. Japan) and the Middle East. The following are example sentences in which Jewish conspiracy is used:

> In a May 2011 article Dankof protested the British government attempting to shut down Press TV, blaming it on "media outlets and correspondents with provable connections to the American **Jewish lobby**; Israeli intelligence; and Neo-Conservatives thirsting for a War of Civilizations with Iran specifically, and the Islamic world generally."

Although contemporary relations between Israel and Armenia are normally good, some anti-Jewish sentiments are still present th at may be due to several reasons such as: Israel's alliance with, and ongoing sale of weapons to, Azerbaijan; the fact that a number of the Ottoman empire's Young Turk instigators of the Armenian Genocide were Jewish or crypto-Jewish and the claim by some pseudo-historians that the genocide was actually part of a Zionist / Masonic plot; the continuing refusal of Israel's leaders to acknowledge the Armenian Genocide; and the prior active support given by **Jewish lobby** groups in America to Turkey's position of denying the Armenian Genocide.

Articles in many official Arab government newspapers claim that The Protocols of the Elders of Zion reflects facts, and thus points to an international **Jewish conspiracy** to take over the world.

In *Mein Kampf*, Hitler used the main thesis of "the Jewish peril", which speaks of an alleged **Jewish conspiracy** to gain world leadership.

Some white supremacist groups, such as the South African Boeremag, conflate elements of Christianity and Odinism. The World Church of the Creator (now called the Creativity Movement) is atheistic and denounces the Christian religion and other deistic religions. Aside from this, its ideology is similar to many Christian Identity groups, in their belief that there is a **Jewish conspiracy** in control of governments, the banking industry and the media. Matthew F. Hale, founder of the World Church of the Creator has published articles sta ting that all races other than white are "mud races", which the religion teaches.

### Islamic Terrorism and Islamic Fundamentalism

The top collocates for *Islamic* are *terrorism* and *fundamentalism*. They both portray Islam as a military religion whose followers do not accept others. Terrorism also occurs with Jewish (with a rank of 392), and Christian (rank = 185).
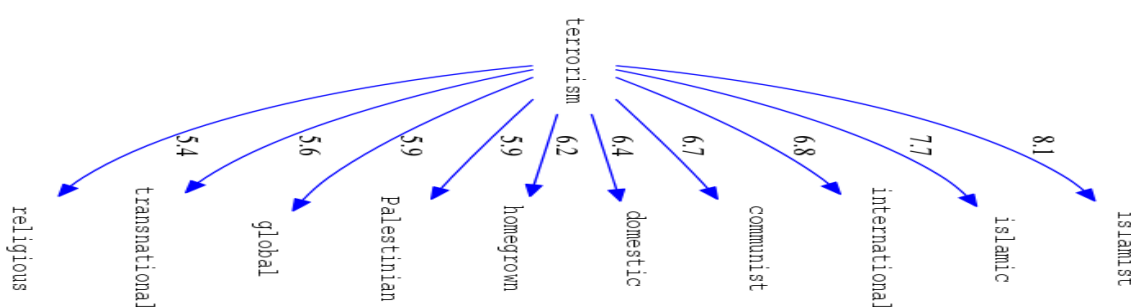


*Figure 4: the top 10 collocates of terrorism*

The picture of fundamentalism is similar to that of terrorism. The top collocates for fundamentalism are: *grass-root*, **islamic**, *militant*, **religious**, **christian**, *violent*, **mormon**, *radical*, *extreme*, *widespread*, *fundamentalist*, *seculaɩ*, **islamist**, **protestant**, *american*, *secret*, **Saudi-inspired**, **premillenarian**, *conservapedia-style*, *coming*, and **bible-belt**. The word clearly has a religious flavor, and is more related to Islam and Christianity than to Judaism, which has no related terms in the top 20 collocates of *fundamentalism*. Here are some sentences:

> The Shia and Sunni religious conflicts since the 7th century created an opening for radical ideologists, such as Ali Shariati (1933–77), to merge social revolution with **Islamic fundamentalism**, as exemplified by Ir an in the 1970s.

> Since 2001, the government of Indonesia h as co-operated with the U.S. in cracking down on **Islamic fundamentalism** and terrorist g roups.

> In January 2008, Liberal M P Sophie Mirabella launched an attack on Fraser, after a speech he gave at Melbourne University on "the Bush Administration (reversing) 60 years of progress in establishing a law-based international system", claiming errors and "either intellectual sloppiness or deliberate dishonesty", and that he tacitly supports **Islamic fundamentalism**, should have no influence on foreign policy, and that his stance on the war on terror has left him open to caricature as a "frothing-at-the-mouth leftie".

> Morris told The Sunday Times tha t the film, will seek to do for **Islamic terrorism** what Dad's Army, the classic BBC comedy, did for the Nazis by showing them as "scary but also ridiculous".

> However, allowing Coalition forces to be based in the country proved to be one of the issues that has led to an increase in **Islamic terrorism** in Saudi Arabia, as well as Islamic terrorist attacks in Western countries by Saudi nationals, the 9/11 attacks in New York being the most prominent example.

> On the issues pertaining t o the constant political turmoil in the Middle East, Hanson emphasises the lack of individual and political freedom in many Middle Eastern nations as a major factor retarding economic, technological and cultural progress. He further relates the root cause of radical **Islamic terrorism** to insecurities and a need to regain honour and reputation.

> **Islamic terrorism** became a problem in the Arab world in the 1970s to 1980s. While the Muslim Brotherhood had been active in Egypt since 1928, their militant actions were limited to assassination attempts on political leaders.

## Christian Fundamentalism and Christian Terrorism

The top 2 negative nouns associated with *Christian* are *fundamentalism* and *terrorism*, but they rank much lower for Christianity than they do for Islam. A comment on the discussion page may explain why terrorism is more likely to collocate with Islamic than Christian:

> Maybe so, but for better or worse Islamic terrorism is a topic that has been viewed by many as a phenomenon worth considering as part of a larger whole. I am not convinced that the concept of Christian terrorism is so widely recognized. (Indrian 21:46, Dec 18, 2004 (UTC) )

Maybe *Christian terrorism* is not recognized as a term in the context of English Wikipedia. Whether it is recognized by other Wikipedias is an empirical question worth investigating. A factor that may be at play is the (religious) beliefs commonly held by a Wikipedia contributors. One would expect the Arabic Wikipedia to generally hold different views from the Hebrew one, but with the lack of contributor background, the answers to these questions may be difficult to find.

This focus on Islam and terrorism shows when we consider the top adjectives modifying terrorism: ***Talibani***, ***Islamist***, ***Islamic***, *international*, *communist*, *domestic*, *homegrown*, *anti-abortion*, ***Palestinian***, *global*, *transnational*, ***religious***, *state-sponsored*, *nationalist*, ***zionist***, *so-called*, *radical*, ***religiously-motivated***, ***christian***, and *taboo*. Religion has a strong presence in describing terrorism, and Islam seems to have the strongest association as the top three adjective are directly associated with it. Another adjective (Palestinian) may also be related. While Judaism and Christianity are also on the list, they appear on the bottom, and in the case of *Judaism*, a related term, *Zionism*, is more often used than the literal Jewish adjective.

> The short novel, "If This Goes On—", describes a rebellion against an American theocracy and thus served as the vehicle for Heinlein to criticise the authoritarian potential of Protestant Christian fundamentalism.

> Liddle, a member of the Church of England, condemned the rise of evangelicalism and Christian fundamentalism in Britain, especially the anti-Darwinian influence of such beliefs in faith schools; and criticised the social te aching and cultural influence of this strand of Christianity.

> This push towards diversity has however thrown segments of Singapore's population into identity crises. It has resulted in the growth of a minority though vociferous Christian fundamentalism that adopts the discourses of the religious right wing in the United States.

> During the 1982 Lebanon War, while Sharon was Defense Minister, the Sabra and Shatila massacre occurred between 16 September and 18. Between 800 and 3,500 Palestinian civilians in the Sabra and Shatila refugee camps were killed by the Phalanges—Lebanese Maronite Christian militias.

In 1982, after an attack on a senior Israeli diplomat by Lebanese based Palestinian militants in Lebanon, Israel invaded Lebanon in a much larger scale in coordination with the Lebanese Christian militias, reaching Beirut and eventually resulting in ousting of the PLO headquarters in June that year.

Responding to the klephts' attacks, the Ottomans recruited the ablest amongst these groups, contracting Christian militias, known as "armatoloi" (αρματολοί), to secure endangered areas, especially mountain passes.

Analyses of his motivations have noted that he did not only display Christian terrorist inclinations, but also had non-religious, right-wing beliefs. Mark Juergensmeyer and John Mark Reynolds have stated that the events were Christian terrorism, whereas Brad Hirschfield has rejected the Christian terrorist label.

### 3.1.2. Positive Nouns Modified by the Adjectives

I have so far only discussed negative prosody, but there are positive words too. One reason positive prosody may not be as important as negative prosody is that positivity is only ephemeral while negativity is more permanent. I cannot think of a way terrorism can be turned into a positive word, no matter how you modify it. A collocate like *sweet terrorism* will not make terrorism sweet. Nor does something like *lovely invasion* make sense, although one wonders whether *benevolent dictators* makes sense. On the other hand, positive words can easily be transformed into negative ones. The words we use as positive below are no exception: *purposeful aid*, *ugly architecture*, *horrendous art*, *bad charity*, *bloody civilization*, *fake democracy*, and so on. I will try to examine these positive lemmas nonetheless.

By positive words I mean those nouns that, when modified by an adjective of the set *Jewish*, *Christian*, *Muslim*, give the adjective a positive flavor. Just like *Muslim criminal* stains *Muslim* with a negative impression, *Muslim art* can induce some admiration. The 200 top collocates of each adjective produced the following list of positive nouns, which table 3 presents along with their association scores:

|              | Jewish | Christian | Islamic |
|--------------|--------|-----------|---------|
| aid          | 1.1857 | 4.6565    | 4.238   |
| architecture | 1.7658 | 4.2685    | 7.918   |
| art          | 5.2797 | 7.3122    | 8.0439  |
| charity      | 4.2079 | 5.8403    | 5.3803  |
| civilization | 4.3982 | 3.773     | 7.4602  |
| culture      | 7.7277 | 5.9626    | 7.9451  |
| democracy    | 0.3691 | 7.1328    | 4.9902  |
| ethic        | 5.6705 | 6.1309    | 5.3356  |
| heritage     | 6.9398 | 4.0453    | 4.8455  |

| literature | 6.6418 | 6.8892 | 6.3711 |
| love | -0.3519 | 4.7904 | -1.6514 |
| martyr | 2.4432 | 7.3615 | 2.4094 |
| medicine | -0.0299 | -3.0 | 5.6926 |
| morality | -0.2831 | 4.739 | 3.7961 |
| music | 6.796 | 8.4306 | 5.2114 |
| philosopher | 6.1686 | 5.0035 | 5.9023 |
| philosophy | 6.4728 | 5.6663 | 7.8814 |
| positivity | 4.942 | -3.0 | -3.0 |
| revival | 2.2277 | 3.4587 | 5.2878 |
| sage | 5.0225 | -3.0 | 2.0731 |
| scholar | 6.6245 | 5.9611 | 8.9037 |
| scholarship | 5.8996 | 3.1332 | 5.5901 |
| science | 4.2471 | 4.441 | 7.1693 |
| scientist | 5.3622 | 6.2558 | 6.2003 |
| spirituality | 2.977 | 5.2531 | 3.2036 |
| thought | 6.1506 | 5.7316 | 6.7233 |
| value | 5.1811 | 7.0864 | 6.2593 |
| virtue | -0.2917 | 4.92 | 2.9778 |

*Table 3: Positive nouns and their logdice scores*

*Islamic* seems to score higher on positive nouns of knowledge as it scores higher than the other two (*Jewish* and *Christian*) on *architecture*, *art*, *civilization*, *culture*, *medicine*, *philosophy*, *revival*, *scholar*, and *science*. All these nouns indicate human effort and achievement. All these may be mainly related to the Islamic golden age. *Jewish* seems to score higher on *culture*, *heritage*, *philosopher*, *positivity*, *sage* and *thought,* which may all be subsumed under *Jewish Thought*. The similarity between how *Jewish* and *Islamic* are positively portrayed can be hardly overestimated as both can be seen in the light of thought and culture. *Christian* seems to score higher on *aid*, *charity*, *democracy*, *ethics*, *literature*, *love*, *martyr*, *morality*, *music*, *scientist*, *spirituality*, *value*, and *virtue*. *Christian* is thus portrayed more as a faith/spirituality adjective than a culture one, i.e. when *Christian* is used, it is used more in the religious sense, unlike *Jewish* and *Islamic* which are more used in the culture sense.

### 3.2    Answer to Question 2: *Is there a statistically significant difference between the semantic prosodies of these adjectives?*

#### 3.2.1. Negative Collocates

There are 23 negative collocates whose scores we measure in *Jewish*, *Christian*, and *Islamic*. The mean score for *Jewish* is 2.21 with a standard deviation of 2.5. The mean score for *Christian* is 1.64 with a standard deviation of 2.12. *Islamic* has a larger mean value than either *Christian* or *Jewish* with a mean score of 4.98 and a standard deviation of 2.9. A box-plot graph depicting the three adjectives is in Figure 5.
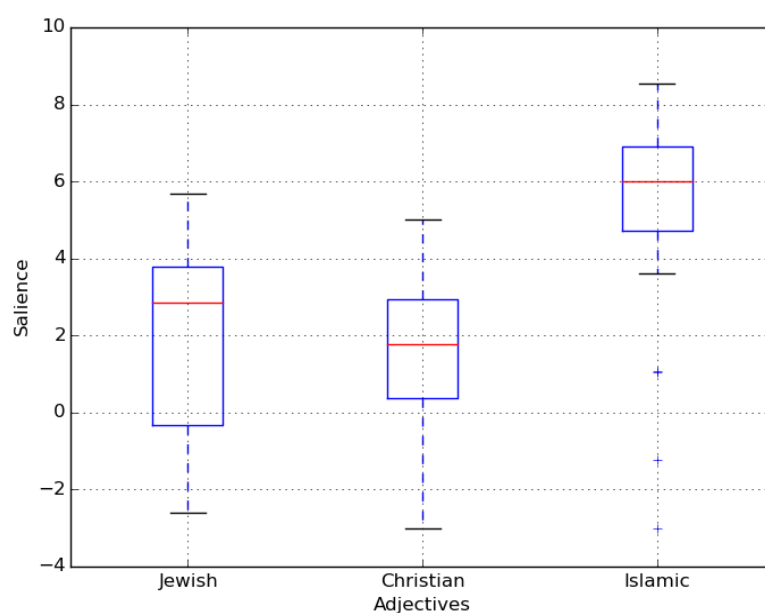


*Figure 5: Negative noun logdice scores*

There was a statistically significant difference between the three adjective groups as determined by one-way ANOVA ($F(2,66) = 11.439$, $p < 0.0001$). In light of this difference, I will use post-hoc tests to determine which pairs are significantly different from each other. For this purpose, I will run the independent t-test and will accept significance at $p \leq 0.05$. To adjust for the three groups, I will apply the Bonferroni correction and accept as significant $p$ values $< 0.01667$. Table 4 summarizes the results of the pair-wise comparisons:

| Pair | Means | t-value | p-value |
|---|---|---|---|
| Jewish vs. Christian | 2.21 vs. 1.64 | 0.84 | 0.408 |
| Jewish vs. Islamic | 2.21 vs. 4.98 | -3.46 | 0.0012 |
| Islamic vs. Christian | 4.98 vs. 1.64 | 4.45 | < 000001 |

*Table 4: independent t test results for the adjective pairs*

The t-test shows no significant difference between the scores for *Jewish* (M=2.21, SD=2.51) and *Christian* (M=1.64 , SD=2.12); t (44)=0.84, p = 0.41. These results suggest that *Jewish* and *Christian* have no difference in their negative semantic prosodies. When we compare *Jewish* and *Islamic* (M=4.98, SD=2.91), the t-test shows a significant difference; t(44) = -3.46, p = 0.0012. This suggests that *Islamic* has a more negative semantic prosody than *Jewish*. When we compare *Islamic* vs. *Christian*, the t-test shows a significant difference; t(44) = 4.45, p < 000001. This suggests that *Islamic* has a more negative prosody than *Christian*. The overall comparison suggests that *Islamic* is the adjective with the most negative semantic prosody while Christian and Jewish are similarly lower in their negative associations.

### 3.2.2. Positive Collocates

There are 28 positive noun collocates. The mean score on these nouns for *Jewish* is 4.1 with a standard deviation of 2.6 , *Christian* has a mean of 4.51 and a standard deviation of 3.26 while *Islamic* has a mean of 5.1 and a standard deviation of 2.9. There was no statistically significant difference between the three adjective groups as determined by one-way ANOVA ($F(2,81) = 0.856$, $p = 0.4286$). This suggests that the three adjectives have more or less the same level of positive semantic prosody. Figure 6 presents the mean scores.
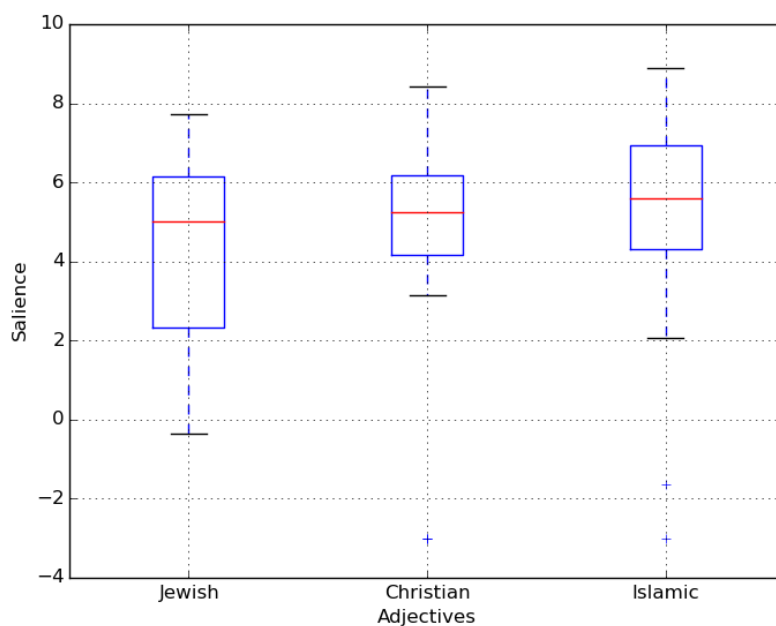


*Figure 6: Positive prosody*

## 3.3   Intra-adjectival Differences

The purpose of this section is to compare, not *Jewish* vs. *Christian* vs. *Islamic*, but the scores of each of these on the positive and negative scales. In other words, we ask the question whether the negative associations of *Islamic*, for example, have a higher mean than its positive associations. The rational behind this question is that we have examined positive and negative lexical associations independently of each other for each adjective, but for a reader, it may be the balance of these that ultimately form her perception of the adjective in question. We use the t-test for the comparison:

### Jewish Positive vs. Jewish Negative

The t-test shows a significant difference between *Jewish_positive* (M=4.1, SD=2.57) and *Jewish_negative* (M=2.21, SD= 2.5 ); t(49) = 2.599638, p = 0.0123967. This suggests that there is a significant difference between the means of the positive and negative values with positive values being higher than negative ones. We thus conclude that Jewish is more positive than negative. All things being equal, the reader of Wikipedia will probably conclude that Jewish is a positive adjective.

### Christian Positive vs. Christian Negative

The t-test shows a significant difference between *Christian_positive* (M=4.6, SD=2.98) and *Christian_negative* (M=1.64, SD= 2.12 ); t(49) = 4.1633630, p = 0.0001. This suggests that there is a significant difference between the means of the positive and negative values with positive values being higher than negative ones. We thus conclude that Christian is more positive than negative. All things being equal, the reader of Wikipedia will probably conclude that Jewish is a positive adjective.

### Islamic Positive vs. Islamic Negative

The t-test shows no significant difference between *Islamic_positive* (M=5.112786, SD=2.74) and *Islamic_negative* (M=4.978943, SD= 2.91 ); t(49) = 0.1676099, p = 0.868. This suggests that there is no significant difference between the means of the positive and negative values. We thus conclude that Islamic is ambivalent between positive and negative. All things being equal, the reader of Wikipedia will probably have difficulty determining whether Islamic is a positive or negative adjective.

## 4    Conclusion

I have so far presented my analysis of how the attributive adjectives *Jewish*, *Christian* and *Islamic* can be shaped by the negative and positive nouns they modify in the English Wikipedia. Religion is a sensitive issue, and I cannot claim to be neutral or impartial. I am a Muslim, and this may have affected my choices and analyses. Moreover, I am not a native speaker of English, and this may also have had some effect. These are not the only issues with these analysis thoughts

One issue is that comparing scores and ranks may be simplistic. All I have done is claim that X is more negative than Y because X associates with more negative words at higher ranks. This does not take into effect the degree of negativeness/positiveness of the modified nouns. While we may all agree that both *fundamentalism* and *extremism* are bad, we may not find it easy to agree which one is more negative. We may still agree that terrorism is worse than both, but how negative is it? If terrorism happened to be the only negative noun modified by an adjective, and it occupied a high rank, would it be more negative than extremism and fundamentalism combined? What if it had a lower rank against the two high ranking fundamentalism and extremism. This problem may not be easy to solve. Perhaps a corpus, or an ontology, of words and their degrees of negativeness may contribute to the solution.

Another problem is with the positive word list. It seems to me that it is easier to determine that a word is negative than to decide that it is positive. While intuitions may not be a good thing in corpus linguistics work, especially when they come from a nonnative speaker, but from a more universal perspective one can doubt that *art* is positive. It is true that *Islamic art* may be positive, but there is no rule against using it in a negative statement. One can say *ugly Islamic art*, or *feigned Christian love*.

These two issues aside, I have found that in Wikipedia *Islamic* is probably more negative than either *Christian* or *Jewish*, and if negativity is determined by the ranks and numbers of negative vs. positive nouns an adjective associates with, then it may not be too wrong if I claimed that both *Jewish* and *Christian* are positive words while *Islamic* has more negative semantic prosody.

Perhaps, as a sequel to this article, one should also examine the Arabic and Hebrew Wikipedias for the same issue, a project that may prove difficult since these two languages lack the computational tools readily available for English.

## Bibliography

Baker, P.; Gabrielatos, C. and McEnery, T. 2013. Sketching Muslims: A Corpus Driven Analysis of Representations Around the Word 'Muslim' in the British Press 1998–2009. Applied Linguistics 2013: 34/3: 255–278

Bird, S., Klein, E. and Loper, E. 2009. Natural Language Processing with Python, O'Reilly Media.

Bublitz, W. 1996. Semantic prosody and cohesive company: *somewhat predictable.*Leuvense Bijdragen. Tijdschrift voor Germanaase Filologie, 85/1-2:1-32

de Poli, B. 2014. The Judeo-Masonic Conspiracy: The Path from the Cemetery of Prague to Arab Anti_Zionist Propaganda. In Butter, M. & Reinkowsky, M. (Ed.). Conspiracy Theories in the United States and the Middle East (2014). Linguae & Litterae 29.

Toutanova, K.; Klein, D.; Manning, C. and Singer, Y. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.

Louw, M. 993. "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies". In Text and Technology: In Honour of John Sinclair . M. Baker et al. (Eds.) Amsterdam, Philadelphia, Benjamins. 240-251

McEnery, A. and Xiao, R. 2006. Corpus-Based Language Studies: An Advanced Resource Book. Routledge.

Mendes, P. 2014. Jews and the Left: The Rise and Fall of a Political Alliance. Palgrave Macmillan

Nivre, J. 2009. Non-Projective Dependency Parsing in Expected Linear Time. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 351-359

Partington, A. 2004. '"Utterly content in each other's company": Semantic prosody and semantic preference,' International Journal of Corpus Linguistics 9/1: 131–56.

Pataki, M.; Vajna, M. and Marosi, A. Wikipedia as Text. Ercim News – Special theme: Big Data. 2012, Vols. 89, pp. 48-49

Rychlý, P. 2008. A Lexicographer-Friendly Association Score in Sojka, P and Horák, A (Eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2008, pp. 6–9, 2008. Masaryk University, Brno 2008

Sabar. P. 2014. Exotic beauties, victims and terrorists: Representations of veiled women in the British press (2001–14). Journal of Arab & Muslim Media Research 7 (1) pp. 59–73 Intellect Limited 2014

Sinclair, J. 1991. Corpus, Concordance, Collocation. Oxford University Press.

Stewart, D. 2009. *Semantic Prosody: A Critical Evaluation*. Taylor & Francis.

Stubbs, M. 1996. Text and Corpus Linguistics. Oxford: Blackwell.

**Biography**

EMAD MOHAMED, PhD (Indiana University) is lecturer of Linguistics at Suez University, Egypt, and research associate of Computer Science at the Université du Québec à Montréal, where he currently lives. Emad's main interests are data science and digital humanities, especially as they relate to the study of religion.

Emad Mohamed
Suez University, Egypt
1112 rue Louis-Joliet.
Lasalle H8N2X6,
Quebec, Canada
emohamed@umail.iu.edu