

Ain Shams University
Faculty of Al-Alsun (Languages)
Department of English Language and Literature
Postgraduate Credit Hour System

**Perpetuation and Transmission of Bias: A Corpus-Driven Detection of
Gender-Oriented Linguistic Bias in Arabic and English Wikipedia Biography
Articles**

A Thesis Submitted in Partial Fulfillment of the Requirements for the Master's Degree of
Linguistics

Submitted to

**The Department of English Language and Literature
Faculty of Al-Alsun (Languages), Ain Shams University**

By

Reem Ali Mahmoud Alkashif

Teaching Assistant
Department of English
Faculty of Al-Alsun, Ain Shams University

Under the Supervision of

Prof. Khaled Mahmoud Tawfik

Professor of Translation and Linguistics
Faculty of Arts
Cairo University

Dr. Rania Mostafa Al-Sabbagh

Lecturer in Computational Linguistics
Faculty of Al-Alsun
Ain Shams University

2020

Acknowledgements

The present study would have not been possible without the help and support of my supervisors, Prof. Khaled Tawifk and Dr. Rania Al-Sabbagh. Their guidance in critical times has always been present. The time they took to review my work was key to its development. I am also grateful to both the current and former Chairs of the Department of English, Faculty of Alsun, Ain Shams University; Prof. Samar Abdel Salam and Prof. Fadwa Abdel Rahman for their encouragement and help in creating an atmosphere conducive to productivity.

I am immensely grateful to the Wikidata and Wikipedia community. They gave me insights that pushed forward the thesis. Gratitude also goes to Tilman Bayer, Dr. Ghada Attia, and Amine Tifratine for the fruitful discussions.

Finally, I am thankful to my parents and brother, my fiancé Matej Grochal, and Radwa Ali for their unfailing support and a strong emotional push.

Table of Contents

Acknowledgements.....	I
Abstract	V
List of Tables	VI
List of Figures	VIII
List of Abbreviations	IX
Chapter One.....	1
Introduction	1
1.1. Objectives of the Study.....	6
1.2. Problem of the Study	7
1.3. Research Questions.....	8
1.4. Scope of the Study	8
1.5. Chapterization.....	9
Chapter Two	10
Review of the Literature.....	10
2.1. Manifestations of Gender Bias in Wikipedia	10
2.2. Bias Detection Using LCM	14
2.3. Significance of the current study	17
Chapter Three.....	20
Methodology	20
3.1. Theoretical Background	20
3.1.1. LCM Framework: An Overview.....	20

3.1.1.1. The Psychological Implications of LCM Categories	24
3.2. Source of the Data	27
3.3. Tools	27
3.3.1. Tokenizers.....	28
3.3.2. Part-of-speech Taggers	28
3.3.3. Lemmatizers.....	33
3.3.4. MPQA Subjectivity Lexicons	33
3.3.5. One-way ANOVA and Post-hoc Tukey HSD	37
3.4. Procedures and Techniques of Data Analysis	38
Chapter Four.....	41
Analysis and Findings	41
4.1. Corpus Description	41
4.2. Corpus Collection	43
4.3. Corpus Preprocessing	53
4.4. Corpus Analysis.....	54
4.4.1. Tokenization, Lemmatization, and Part-of-speech Tagging.....	54
4.4.2. Finding Subjective Words Using MPQA Lexicon	56
4.4.3. Measuring Statistical Significance.....	57
4.5. Findings	64
Chapter Five	72
Conclusion.....	72
5.1. Summing-up of the Findings and Contribution of the Study	72
5.2. Answering Research Questions	74
5.3. Limitations of the Study	75

5.4. Recommendations and Suggestions for Further Research	76
References	78
Books and Papers	78
Websites	85
Appendices: MPQA English and Arabic Lexicons	90
Appendix A: MPQA English Lexicon.....	90
Appendix B: MPQA Arabic Lexicon	127

Abstract

It is a non-trivial task to maintain language neutrality, especially in Wikipedia biographical articles about men and women given the open, crowdsourced nature of Wikipedia. For that purpose, an automatic system is developed to detect gender-based bias and its intensity in Wikipedia biographical articles about men and women. Based on the tenets of the Linguistic Category Model (LCM) about the abstract biased language and concrete neutral language, the study uses the English Multi-perspective Question Answering (MPQA) Subjectivity lexicon and its Arabic translation to detect verbs, adjectives, nouns, and adverbs in a corpus of 1600 Wikipedia articles covering both men and women from different professions in both English and Arabic. The current study tries to answer the following questions: 1) What is the evidence of gender bias in each of the Arabic and English male and female articles? 2) Which of the Arabic and English Wikipedia male and female datasets show more intense gender bias? and 3) How does the intensity of gender bias change when comparing specific professions between the Arabic and English male and female datasets? The results show evidence of intense gender bias particularly in the analyzed Arabic male Wikipedia articles compared with Arabic female and English male and female Wikipedia articles. It has also been found that Arabic Wikipedia articles about male politicians are more intensely biased compared to sportsmen and male writers.

Keywords: Wikipedia, Linguistic Category Model, Lexicon, Gender Bias, Biographies

List of Tables

Table 1 Criteria of abstractness and concreteness.	22
Table 2 Overview of LCM.....	26
Table 3 Summary of POS taggers accuracy test results	32
Table 4 Total number of entries in the English and Arabic MPQA as used in the present study.....	37
Table 5 Word count of each corpus by gender.	42
Table 6 Word count of each profession data subset.	42
Table 7 Basic corpus data matrix.....	43
Table 8 The Wikipedia profession categories used in the present study.	46
Table 9 Number of duplicated articles in profession categories of females in the Arabic data.....	50
Table 10 Number of duplicated articles in profession categories of males in the Arabic data.....	51
Table 11 Number of duplicated articles in profession categories of females in the English data.....	52

Table 12 Number of duplicated articles in profession categories of males in the English data.....	52
Table 13 Number of strong subjective words (nouns, adjectives, adverbs, verbs) in the Arabic data.....	58
Table 14 Number of strong subjective words (nouns, adjectives, adverbs, verbs) in the English data.....	58
Table 15 ANOVA Test for using formulaic content words in the Arabic and English datasets.....	61
Table 16 Post-hoc Tukey HSD Test among the different groups.....	62
Table 17 ANOVA Test for using formulaic content words among Arabic users at the article level.....	63
Table 18 Post-hoc Tukey HSD Test among the different groups.....	64

List of Figures

Figure 1 Meta-data in Wikipedia.	12
Figure 2 Example of an English POS tagger interface.	30
Figure 3 Example of an Arabic POS tagger interface.....	30
Figure 4 Example of an English POS tagger output.	30
Figure 5 Example of an Arabic POS tagger simplified output.	31
Figure 6 Entries in MPQA subjectivity lexicon.....	34
Figure 7 Entries in the edited and lemmatized translation of MPQA lexicon.	36
Figure 8 Corpus building process.	53
Figure 9 One-way ANOVA algorithm.	60
Figure 10 Post-hoc Tukey HSD algorithm.	62

List of Abbreviations

Abbreviation	Explanation
ADJ/Adj	Adjective
ADV/Adv	Adverb
ANOVA	Analysis of Variance
DAV	Descriptive Action Verb
HSD	Honestly Significant Difference
IAV	Interpretive Action Verb
LCM	Linguistic Category Model
MPQA	Multi-perspective Question Answering
N/n	Noun
NLP	Natural Language Processing
NPOV	Neutral Point of View
POS	Part of speech
SV	State Verb
V/v	Verb

Chapter One

Introduction

Wikipedia is a free, online encyclopedia available in 302 languages (*List of Wikipedias*, n.d.). People around the world make Wikipedia content to produce free knowledge to everyone everywhere. Wikipedia is the fifth most visited website in the world (*Wikipedia.org Traffic, Demographics and Competitors*, n.d.) hosting more than 47 million articles in different languages and receiving more than 15 billion views and more than 10 million edits per month (*Wikimedia Report Card*, 2016).

To ensure content quality in all languages, Wikipedia has five nonnegotiable rules called the ‘five pillars’. The rules dictate that Wikipedia accepts only 1) unbiased 2) encyclopedic entries 3) added in an environment of mutual respect 4) that does not have firm rules and 5) under a free license (*Wikipedia: Five Pillars*, n.d.). Keeping articles unbiased is essential in guaranteeing quality and avoiding the spread of false information. Wikipedia, thus, upholds a ‘Neutral Point of View Policy (NPOV)’ which is defined as “representing fairly, proportionately, and, as far as possible, without editorial bias, all of the significant views that have been published by reliable sources on a topic” (*Wikipedia: Neutral Point of View*, n.d.).

This definition of the Wikipedia Neutral Point of View Policy has triggered research in elements jeopardizing articles’ neutrality. Among the examined elements are race, geography, and gender. Perhaps the importance of gender equality in the Wikimedia Movement (*Gender Equity Report*, n.d.) and the world is what makes the study of

gender-oriented bias a topic worthy of research. In their study of gender-oriented bias, Glott et al., (2010) have linked unequal participation rates from different groups and communities to reducing articles' neutrality. This has been particularly stark in the ratio of male to female contributors as the researchers have found that the percentage of female contributors was less than 13% in 2010 (Glott et al.). In 2011, the percentage of female contributors has declined to 8.5% (AyushKhanna, 2011). Glott et al. have suggested that this difference in participation affects the type of content on Wikipedia in different ways making it more male biased.

This gender bias against women is also found in meta-data, network and structural properties, coverage, notability, and language. First, meta-data bias refers to the disproportionate number and length of articles about women in comparison to men. It also includes an analysis of the time period in which the largest amount of biographies of male and female figures has been created. In addition, some meta-data research focuses on the frequency and type of information found in the infoboxes, which are panes added to articles to give quick information about the people, in male and female biographies. Meta-data bias is usually measured using DBpedia; a structured version of the said meta-data elements as found in Wikipedia articles. Wagner et al. (2016) have showed that there are differences in the frequency of information found in the infoboxes; for example, women had information about their spouses more often than men. Wagner et al. (2016) have also found that the hyperlinks network in Wikipedia puts biography articles about men at an advantage in page ranking algorithms.

Closely related to meta-data are networks or structural properties. Wagner et al. (2016) have found gender-oriented bias in the network or structural properties of Wikipedia. Wikipedia network refers to the way Wikipedia connects articles to other related articles inside Wikipedia using internal hyperlinks. Network bias happens when the density of internal links connecting female biographies to male biographies is higher than the density of female biographies that are related to female biographies. More specifically, it is a measure of how much Wikipedia biography articles on males and females are mutually exclusive. Such analyses also aim to explore the effects of network bias, together with other elements like the discrepancy between the number of male and female biographies, on the visibility and reachability of female biographies for users. Wagner et al. have suggested that this bias in network connections added to the lower number of articles on female figures on Wikipedia make the existing articles less visible.

In terms of coverage of female figures on Wikipedia, there are two areas of research. On one hand, Reagle & Rhue (2011) and Lam et al. (2011) have focused on the coverage of notable females in Wikipedia in terms of the number and length of articles about women in Wikipedia in comparison to males. Reagle & Rhue have assigned a reference point, the Encyclopedia Britannica, to which Wikipedia has been compared regarding gender bias. They have found that the number of articles about women in Wikipedia is smaller than that in Britannica. In a similar vein, Lam et al. have used articles' length as a measure of quality and have found that articles which tend to be particularly interesting to males are higher in quality than those of interest to females. On the other hand, Wagner et al. (2016) have focused on analyzing the conditions and factors that determine which figures can have articles on Wikipedia. Those conditions are

referred to in the Wikipedia editing environment as the notability criteria. More specifically, they have compared whether notability differs between males and females. Wagner et al. found that women are subjected to more strict and complex notability conditions, such as having to be recently in the news and becoming search trends on Google, before getting articles on Wikipedia (Wagner et al., 2016).

Topical bias is another element that introduces and perpetuates gender-oriented bias on Wikipedia. Wagner et al. (2015) and Graells-Garrido et al. (2015) have focused on gender bias in Wikipedia through exploring topical bias which involves exploring the topics usually included in the articles on male figures in comparison to the ones on females. Finkbeiner's test (Brainard, 2013), a list of phrases compiled by the journalist Christie Aschwanden to highlight lexical bias against women, has inspired analysis of the topics prevalent in articles about women. Using Finkbeiner-based lexicons and computer assisted methods, family, romantic relationship status, gender, personal life, arts have been the topics found to be discussed more frequently and in much detail in articles about women in comparison to the ones about men. For example, Wagner et al. (2015) have found that divorce as a life event tends to be mentioned 3.66 times more in articles about women more than in articles about men.

Linguistic analysis is another main area of Wikipedia is the lexical exploration of linguistic manifestations of bias. Wagner et al. (2015) employed the Linguistic Category Model (LCM) (Semin & Fiedler, 1991; Coenen et al., 2006) to automatically identify bias in language editions of Wikipedia. LCM shows the underlying psychological explanation of how abstractness and concreteness in language contribute to creating and perpetuating

bias. The model makes a distinction between grammatical categories in interpersonal use of language and splits them into adjectives, nouns, and adverbs (the abstract category) and verbs (the concrete category). On one hand, adjectives, nouns, and adverbs are the most abstract lexical items; they give less verifiable, disputable information about a person and make the information seem like a stable quality that is unlikely to change in the future. For example, in the sentence (the Prime Minister is *aggressive*), the use of the adjective (*aggressive*) is highly abstract, according to LCM, hence more biased. Similarly, in the following example: (the Prime Minister's *aggression*), the use of the noun (*aggression*) introduces a highly abstract idea about the Prime Minister. The use of adverbs in (the Prime Minister spoke *aggressively*) also introduces similar abstract ideas about the traits of the subject as do adjectives and nouns. Verbs, on the other hand, are more concrete as they tend to give a description of an action, limited in time and space, done by someone rather than conveying a judgment of that person. For example, a sentence like (the Prime Minister *attacks* the opposition) is a concrete verb which gives a less biased idea about the Prime Minister as it refers to a verb that could be interpreted in variant ways. Using LCM, Wagner et al. (2016) have found that Wikipedia articles on women show a subtle level of bias encoded in words. The researchers have found that negative descriptions of women were frequently referred to using more abstract terms. Abstract terms convey a negative attitude towards women since they show that this negativity is permanent, not context-bound, general information about the person or thing more than the particular situation. However, when it comes to positive descriptions of women, more concrete language is used which is cognitively understood as a reference to temporary, disputable traits.