

Gender Construction in User Comments on Selected Facebook Pages: Integrating Topic Modeling with Feminist Critical Discourse Analysis^(*)

**Under the Supervision of
Norice W. Methias
Faculty of Arts, Cairo University**

**Amira Nasr Ekara
Language Instructor, Faculty of Arts, Damanhour University**

Abstract

Interactive spaces on social networking sites (SNSs) provide a significant site for the (re)production, negotiation and/or contestation of gender identity. The comment sections on social media platforms, in particular, allow users to engage in the construction of their gender identities, notably via the digitally-mediated textual interactions hosted by these spaces. Against this backdrop, this study aimed to examine gender construction in a sample of user-generated comments drawn from a number of English and Arabic Facebook Pages pertaining to the themes of feminism and women's rights and violence against women. The study combined topic modeling, a computational method for classifying texts into topics by detecting the clusters and frequencies of co-occurring words, and feminist critical discourse analysis (FCDA) with a view to gaining a better understanding of the research problem by integrating insights from both quantitative and qualitative methods. To investigate the workings of gender construction in the data under investigation using the FCDA paradigm, this study specifically drew on the analytical toolkit proposed by Bucholtz and Hall (2004, 2005, 2010) for analyzing identity in interaction. The results showed that users utilized a number of indexical processes, such as references to identity categories and labels, pronoun use, metaphors, and stance-taking, to

(*) Gender Construction in User Comments on Selected Facebook Pages: Integrating Topic Modeling with Feminist Critical Discourse Analysis, Vol.14, Issue No.3, July 2025, pp.95-132.

construct themselves and others as particular types of gendered beings as well as to align themselves with or against particular gender discourses and ideologies. The findings also revealed the multidimensionality of the construction of gender identity in the studied comments, particularly insofar as gender intersects with other identity-based factors such as region, culture, and religion. Thus, the gendered representations that emerge from the data exhibit a range of complex and multifaceted gender-based positions and identifications.

Keywords: gender, user-generated comments, Facebook Pages, feminist critical discourse analysis (FCDA), topic modeling,

الملخص

تمثل المساحات التفاعلية على مواقع التواصل الاجتماعي فضاءً مهمًا لإعادة إنتاج الهويات الجندرية والتفاوض حولها وخلقتها كذلك. وبشكل خاص، يُمكن القسم الخاص بالتعليقات على الشبكات الاجتماعية المستخدمين والمستخدمات من بناء هوياتهم/نّ الجندرية من خلال التفاعلات النصية التي تقام عبر هذه المساحات. في هذا الإطار، تهدف هذه الدراسة إلى تقصي بناء الجندر في عينة من تعليقات المستخدمين/ين مستقاة من عدد من الصفحات العامة على موقع فيسبوك باللغتين الإنجليزية والعربية، تحديدًا تلك التي تتناول موضوعات النسوية وحقوق النساء والعنف ضد النساء. تعتمد الدراسة بصورة رئيسة على الدمج بين تقنية "نمذجة الموضوعات"، وهي طريقة حاسوبية لتصنيف النصوص إلى موضوعات اعتمادًا على استكشاف الكلمات المتلازمة في النصوص وتواتر استخدامها، وتحليل الخطاب النقدي النسوي. ويهدف الدمج بين هذين المنهجين إلى الجمع بين مزايا نمذجة الموضوعات بوصفه منهجًا كمياً وتحليل الخطاب النقدي النسوي بوصفه منهجاً كيفياً ومن ثم السعي إلى فهم أكثر تعمقاً للمشكلة البحثية محل الدراسة. فيما يتعلّق بتحليل الخطاب النقدي النسوي، توظف الدراسة بشكل خاص الإطار الذي طرحته باكلتزر وهول (Bucholtz & Hall, 2004, 2005, 2010) لدراسة الهوية من خلال التفاعلات اللغوية. وقد أظهرت نتائج الدراسة أنّ المستخدمين والمستخدمات اعتمدوا/اعتمدن على عدّة أدوات لغوية، مثل الإشارة إلى الفئات والتصنيفات الهويةيّة واستخدام الضمائر والاستعارات واتخاذ المواقف (stance-taking)، لبناء الهويات الجندرية ولإظهار الاتفاق أو الاختلاف مع الخطابات والأيديولوجيات المتعلقة بالجندر. أظهرت النتائج كذلك الأبعاد المتعددة لبناء الهوية الجندرية في التعليقات محل الدراسة، لا سيّما في تقاطعها مع عوامل أخرى مرتبطة بالهوية مثل الإقليم والثقافة والدين. من ثم، تُظهر تمثيلات الجندر في التعليقات محل الدراسة عدّة تعاريف ومواقع جندرية تتسم بالتعقيد وتعدد الأوجه.

الكلمات المفتاحية: الجندر، تعليقات المستخدمين/ين، صفحات فيسبوك، تحليل الخطاب النقدي النسوي، نمذجة الموضوعات

1. Introduction

In a world marked by ubiquitous digitality, online meaning-making spaces have become a significant site for the public discussion and negotiation of a host of different issues. The development of Web 2.0 technologies in particular has fostered a new type of sociality that is fundamentally premised on facilitating interaction and user participation (Jenkins, 2006), rendering user-generated content one of its most distinctive features. Thus, the modern digital landscape is principally marked by what has come to be known as "participatory culture" (Jenkins, 2006) which has transformed the practices and roles of digital media users. Notably, this resulted in blurring the lines between the roles of consumers and producers of media content, thereby giving rise to "the hybrid terms *prosumers* and *produsers*" (Gnach et al., 2023, p. 29), the coinage of which was necessitated by the need to capture the essence of the participatory practices of the social web.

This shift in the dynamics of digital communication in Web 2.0 is perhaps best epitomized by social networking sites (SNSs) which, in Barton and Lee's (2013) terms, have been "domesticated," that is, have become an integral part of users' everyday lives. Given the inherently interactive nature of digital communication on SNSs, the issue of identity has come to the fore, not least because we construct our identities within and in accordance with an increasingly digitized world (Ching & Foley, 2012). For instance, the different communicative spaces and modes provided by digital media allow for the display and documentation of various aspects of users' everyday lives (Barton & Lee, 2013), thus enabling them to engage in the constant performance and (re)shaping of their identities, particularly through the set of linguistic and other semiotic resources at their disposal (Georgalou, 2017). In addition, people's pervasive online presence in modern digital spaces has an impact on how they perform their identities, especially inasmuch as they are enacted in relation to others and with regard to the relations of sameness and difference that

are brought out in the process (Cover, 2016, p. xxi).

Among the interactive spaces on SNSs, the comment section in particular is a rich source of public discussions pertaining to a range of different issues. Social media comments can, therefore, provide significant insights into the "mundane forms of public discourse" constituting the mass communication of millions of users (Lovink, 2011, p. 51). As an interactive meaning-making practice, commenting on social media platforms allows users to "inhabit different sorts of roles and responsibilities vis-à-vis their interlocutors" (Jones et al., 2015, p. 12). Further, social media users utilize these interactive spaces to perform their identities by announcing alignments with "different groups, opinions and cultural issues" (Seargeant & Tagg, 2014, p. 9). As such, commenting on social media can itself be viewed as an act of stance-taking by virtue of allowing users to position themselves and others while engaging in the digitally-mediated discussions in the comment sections (Barton & Lee, 2013).

1.1 Research Problem

As an interactive digital practice, social media comments can yield significant insights into how gender is discursively constructed in the digital domain. Currently standing as "the world's most popular and widely used social media platform" (Howarth, 2025), with around 3.07 billion monthly active users globally accessing the site as of 2025 (Kumar, 2025), Facebook, in particular, offers a rich source of data for such an investigation, specifically via the textually-mediated online interactions taking place within the site's comment sections.

While the issues of self-presentation and identity construction on Facebook have been the subject of extensive research (Aboh et al., 2023; Bouvier, 2012; Emara, 2017; Georgalou, 2015; Gil-Lopez et al., 2018; Grasmuck et al., 2009; Wittkower, 2014; Zhao et al., 2008), this body of research has mainly focused on the investigation of user profiles and personal usages rather than on the public interactive spaces on Facebook Pages (Rieder et al., 2015). Besides, in their review of studies on user-generated content, Naab and Sehl (2017)

pointed out that internationally-oriented comparative research into user-generated content is scarce despite the accessibility of transnational digital data. Similarly, Ehrlich and Meyerhoff (2014, p. 14) pointed out that language and gender research has been dominated by "the hegemony of research conducted on English and/or in accordance with the scholarly paradigms adopted in anglophone or Western European locations." In a similar vein, other scholars have emphasized the importance of expanding the field of language and gender by embracing more contributions, perspectives, and epistemologies from non-Western contexts (Hall et al., 2020; Milani & Lazar, 2017; Singh, 2021).

1.2 Aim of the Study

The current study aimed to fill a gap in the literature by focusing on the analysis of gender construction in data drawn from the public interactive spaces on Facebook Pages rather than from user profiles, which have been widely researched. More specifically, this study sought to explore how female and male commenters on Facebook Pages in both English and Arabic (1) construct themselves and others as particular kinds of gendered beings, (2) align themselves with or against particular gender discourses and ideologies, and (3) draw on various indexical processes to enact their gender identities and/or their positions on gender.

1.3 Research Questions

The overarching question that guides this study is:

How is gender constructed in the discussions included in user comments on Facebook Pages?

This broad research question was broken down into the following sub-questions:

RQ1: What are the topics that prevail in the corpus of English and Arabic user comments under investigation?

RQ2: How do female and male commenters on Facebook Pages (re)produce, negotiate, and/or challenge normative gendered discourse

in English?

RQ3: How do female and male commenters on Facebook Pages (re)produce, negotiate, and/or challenge normative gendered discourse in Arabic?

2. Literature Review

2.1 Language and Gender

Language and gender is an interdisciplinary field that draws on a range of diverse approaches and methods to theorize the relationship between gender and language use. Robin Lakoff's article entitled "Language and Woman's Place" (1973), which was published as a book with the same title two years later, is often credited as launching the field. Since its inception, language and gender research has witnessed significant paradigm shifts. Lakoff's work, for example, focused on examining the ways in which male dominance played out in women's language use. This line of research has been commonly referred to as the "dominance approach." In the 1980s and 1990s, the (cultural) difference approach to the study of language and gender gained currency (Gormley, 2015). This approach traces the differences between women's and men's conversational styles back to early gendered socialization. Consequently, within this view, mixed-sex interaction is understood as being similar to cross-cultural communication, thereby typically featuring frequent instances of miscommunication (Gormley, 2015).

The "dominance" and "difference" research traditions shared a number of shortcomings. Work within both traditions, for example, was preoccupied with the study of the differences, rather than the similarities, between women's and men's speech styles and typically focused on examining the speech of white, straight, middle-class women and men (Cameron, 2005). Such limitations ushered in a major paradigm shift in language and gender studies. More specifically, in the 1990s, essentialist understandings of gender that are premised on conceptualizing gender identity as "something stable

and enduring and that it underpins and reveals itself in speech" gave way to "ideas associated with constructionism and poststructural theories of meaning" (Weatherall, 2016, p. 1).

2.2 Computer-Mediated Communication (CMC)

Beißwenger and Storrer (2008) define computer-mediated communication (CMC) as "a research field that explores the social, communicative and linguistic impact of communication technologies, which have continually evolved in connection with the use of computer networks (esp. the Internet)" (p. 292). While CMC was initially mainly text-based (Herring, 2019), Web 2.0 technologies have introduced diverse modes of multimodal communication (Herring & Androutsopoulos, 2015). Linguistic research into CMC is deeply tied to the evolution and advancement of communication technology. Early language-based research into CMC, for instance, focused on pre-Web CMC modes that were "largely restricted to interpersonal exchanges carried out on language-heavy modes such as mailing lists, newsgroups, and Internet Relay Chat" (Androutsopoulos, 2014, p. 76). Current CMC scholarship, on the other hand, typically engages with a wide range of communicative modes as well as with the digital practices of content consumption and production enacted in SNSs, blogs, wikis, and other types of participatory media (Androutsopoulos, 2014).

2.3 Feminist Critical Discourse Analysis (FCDA)

Developed by Michelle Lazar (2005, 2007, 2014, 2018), feminist critical discourse analysis (FCDA) is a research paradigm that brings together CDA and feminist scholarship with the aim of examining "the complex and diverse ways by which gender ideologies that entrench power asymmetries become 'common sense' in particular communities and discourse contexts, and how they may be challenged" (Lazar, 2018, p. 372). As Lazar (2005) argues, FCDA is motivated by the need for an explicit feminist perspective in CDA studies given the significance of analyzing "from a critical feminist

perspective the particularly insidious and oppressive nature of gender as an omni-relevant category in most social practices" (p. 3). What CDA can offer to the study of gender is a range of analytical tools in addition to the theoretical insights it offers into how social practices and discourse structures are interconnected (Lazar, 2005). Hence, the analytical methods and approaches that FCDA research incorporates are wide-ranging. They encompass the tools of analysis situated within semantics, systemic-functional linguistics, corpus linguistics, conversation analysis, social semiotics, in addition to the many frameworks within CDA studies (Lazar, 2007).

2.4 Previous Studies

2.4.1 Studies on Gender in CMC

Investigating gender-related issues within CMC has received considerable attention (Barton & Lee, 2013). Previous research into the subject explored gender in different CMC modes, addressing the topic from different perspectives and employing different research methods. Rellstab (2007) examined plays with gender in Swiss internet relay chats (IRCs). The study showed that while gender in IRCs is played with, explored, and performed, these gender plays, which are particularly facilitated by anonymous communication in IRCs and the need to break free of gender-based social constraints, are no more than brief attempts to performatively explore and transgress traditional gender limits. Thelwall (2008) studied gendered swearing in a corpus of MySpace member home pages in the U.S. and the U.K. Whilst the U.S. MySpace data showed significantly more strong swearing for males than for females, the U.K. data showed no gender difference in strong swearing, particularly for younger users. Thelwall points out that this finding can be attributed to changes in gender roles and expectations in the U.K.

Wang, Burke, and Kraut (2013) examined a large corpus of Facebook status updates to identify the topics that sparked more audience feedback, that is, likes and comments. The study also

investigated gender differences in the topics shared by women and men, showing that women preferred to share about personal issues while men opted for sharing content relating to general, public events such as politics or sports. The study also revealed that men generally received fewer comments than women. However, "male" topics were found to generate more comments than did "female" ones.

Zhukova and Herring (2024) examined the differences in emoji perception among different demographic groups, especially insofar as toxic emoji use is concerned. Using an online survey to probe this issue, the study found that emoji interpretation differed by gender, generation, and emoji type. With respect to gender, the findings revealed that, generally speaking, there were no differences between female and male respondents in their perception of (potentially toxic) emoji use. However, female and male users' understanding of emoji use differed from non-binary respondents and those who did not disclose their gender, thus supporting earlier research showing that "marginalized gender groups face increased online toxicity, priming them to perceive potential threats" (Zhukova & Herring, 2024, p. 95).

2.4.2 Studies Employing FCDA

Several studies drew on FCDA for understanding the workings of gender and its intersection with power and ideology in a host of different texts. Lazar (2009) examined the construction of a postfeminist identity, called "entitled femininity," as it emerged in a set of print beauty advertisements in Singapore. Using FCDA, the study identified three major discursive themes constituting postfeminist feminine identity in the data. Despite the existence of the traditional, stereotypical media images of women as sexual objects and in terms of their domestic roles in Singapore, another discourse that emphasizes women's public presence, assertiveness, and autonomy is becoming increasingly common, "indexing a global neo-liberal discourse of postfeminism" (Lazar, 2009, p. 371). A postfeminist discourse rests on the assumption that "feminist struggles have ended, viz., that women today enjoy full equality and can 'have it

all' if they put their minds to it; in fact, that it is becoming a women's world, with a celebration of all things feminine" (Lazar, 2009, pp. 371-372). Lazar critiques the discourse of "entitled femininity" on the grounds that it tends to both celebrate and repudiate feminism, and it also contributes to reproducing established gendered stereotypes.

Makoni (2013) studied the maintenance of gendered ideologies in the context of migration. More specifically, the study examined responses from Zimbabwean migrants based in the UK to photographs of men in non-traditional gender roles (e.g., caring for babies and washing dishes). Using FCDA, Makoni was concerned with investigating how the female and male respondents linguistically constructed traditional notions about feminine and masculine identities. The findings revealed that the responses basically reflected the impact of patriarchal gender norms as well as African cultural values.

Peng et al. (2024) examined how Chinese sportswomen are viewed from the perspective of male Chinese sports fans. The study more specifically analyzed the postings of male users on Hupu, China's leading sports fandom platform, about two elite sportswomen at the Tokyo 2020 Olympics. Drawing on FCDA, the authors showed how Hupu male users engaged in sexualizing Chinese sportswomen and trivializing their achievements by using referential and predication strategies which either focused on assessing the physical appearance and sexual attractiveness of the sportswomen or comparing them with their male counterparts to undermine their athletic accomplishments.

3. Data and Methodology

3.1 Data Collection and Sampling Procedures

The data examined in the present study comprise a sample of user comments ($N = 12,021$) drawn from a number of Facebook Pages that discuss the issues of (1) feminism and women's rights and (2) violence against women in English and Arabic. Data collection for the present study involved the following steps. First, using the Facebook search

function, a preliminary search was conducted to identify the relevant Pages. To this end, a number of keywords, such as "feminism," "women's rights," "women's empowerment," "equality," "sexual harassment," "sexual assault," "rape," and "violence against women," were used to capture the relevant Pages for the two themes in English and Arabic. Second, each of these Pages was checked to ensure the availability of relevant content and frequent posting and commenting activity. For each theme, the top 10 Pages with the most likes and the highest number of followers were selected for investigation, comprising a total of 40 Pages from which the entire sample of user comments was extracted. Table 1 and Table 2 show the Pages selected for collecting the English and Arabic data for the present study.

Subsequently, 20 posts with the associated comments were collected from each Page if they met the following criteria: (1) the content of the post was relevant to the theme in question, and (2) the post had at least five comments on it (including replies). The data collection process described above resulted in retrieving a total of 800 posts for the two themes in both English and Arabic. Lastly, using systematic sampling, a sample that comprises 25% of the population of collected posts was drawn, the comments on which provided the data for the present study. Table 3 provides a numerical breakdown of the data by language, theme, and gender. The posts and the associated comments were collected using Export Comments (<https://exportcomments.com/>), an online tool that retrieves and exports social media comments to Excel files.

3.2 Research Design

To examine the construction of gender in the sample of user comments under analysis, this study drew on the methodological synergy first suggested by Törnberg and Törnberg (2016) by combining a quantitative research method, more specifically topic modeling, and a qualitative method, namely, FCDA. Topic modeling refers to "a collection of methods and algorithms that uncover the hidden thematic structure in document collections by revealing

recurring clusters of co-occurring words" (Törnberg & Törnberg, 2016, p. 405). As an essentially inductive method, topic modeling contributes to reducing researcher bias (Törnberg & Törnberg, 2016) by guarding against a common criticism of critical discourse studies, namely, the "cherry-picking" of data, or the tendency to select the texts and linguistic features that suit the researcher's preconceived assumptions (Widdowson, 2004). As such, topic modeling "constitutes an enriching complement to CDA, aiding discovery and adding analytical rigor" (Törnberg & Törnberg, 2016, p. 402). Jaworska and Nanda (2016) similarly argue for integrating topic modeling and discourse analysis, highlighting the usefulness of the former in detecting the prevalent themes in a given corpus as well as guiding the researcher as to which data points to particularly focus on in the qualitative analysis.

3.3 Data Preparation

To prepare the data for analysis, the data collection and sampling for the present study were followed by three more phases. First, data filtering was performed to remove any irrelevant data from the sample (e.g., non-textual comments, duplicate comments). Next, to disaggregate the data by gender, gender identification was automatically performed using Gender API (<https://gender-api.com/>) and NamSor (<https://namesorts.com/api/>), both of which are gender prediction tools that infer a person's gender from their name. In keeping with the ethical practices of digital research (Ditchfield, 2021; Sula, 2016; Townsend & Wallace, 2016), the final data preparation phase involved data anonymization which included removing personally identifiable information such as commenters' names or pseudonyms, mentions, tags, and comment and profile IDs.

Table 1

Pages Selected for Collecting Data in English

Theme: Feminism and Women's Rights		
Page name and URL	Number of likes	Number of followers
Women's Rights News https://www.facebook.com/WOMENSRIGHTSNEWS	1,752,881	1,714,612
Feminist News https://www.facebook.com/feministnews.us	Not listed	1,551,828
Women's March https://www.facebook.com/womensmarchonwash	Not listed	911,000
Feminist Info https://www.facebook.com/feministinfo	610,762	633,957
My Favorite F Word Is Feminism https://www.facebook.com/feminismonbustle	591,818	585,450
Everyday Feminism https://www.facebook.com/everydayfeminism	556,911	544,680
HeForShe https://www.facebook.com/HeForShe	537,791	541,125
UniteWomen.org https://www.facebook.com/UniteWomen	198,403	196,419
Rabid Feminist https://www.facebook.com/RabidFeminist	108,786	142,995
Feminist Current https://www.facebook.com/FeministCurrent	32,213	33,880
Theme: Violence Against Women		
Page name and URL	Number of likes	Number of followers
SayNOtoEndViolenceAgainstWomen https://www.facebook.com/SayNO.UNiTE	206,201	216,619
World Wide Women https://www.facebook.com/WorldWideWomenAll	155,910	158,422
Me Too. Move#metoo https://www.facebook.com/MeTooMVMt	98,142	109,002
NSVRC https://www.facebook.com/nsvrc	Not listed	72,000
Texas Association Against Sexual Assault (TAASA) https://www.facebook.com/taasa	44,731	44,307
Stop Street Harassment https://www.facebook.com/StopStreetHarassment	37,818	38,000
Women's Liberation Front https://www.facebook.com/womensliberationfront	Not listed	27,000
REAL for women https://www.facebook.com/getrealforwomen	18,877	19,792
Australia Against Male Violence Towards Women https://www.facebook.com/endmaleviolence	16,490	17,207
No Shame: Rape/Abuse Survivors Speak Out https://www.facebook.com/noshamespeakout	8,384	8,792

Table 2

Pages Selected for Collecting Data in Arabic

Theme: Feminism and Women's Rights		
Page name and URL	Number of likes	Number of followers
Nasawyia https://www.facebook.com/nasawyia/?ref=br_rs	408,730	417,739
ثورة البنات - Girls Revolution https://www.facebook.com/EgyGirlsRev	151,103	151,285
Egypt girls' diaries https://www.facebook.com/egygirlsdiaary	150,617	172,171
المرأة ثورة و ليست عورة https://www.facebook.com/Revolution.Women	Not listed	60,000
ثورة على المجتمع الذكوري https://www.facebook.com/Revolution.on.the.males.society	54,150	54,410
ثورة نسوية https://www.facebook.com/feminist.revolution	37,541	52,300
حرية بنت https://www.facebook.com/HMHS12	33,811	34,003
Feminists of Jordan https://www.facebook.com/FeministsOfJordan/?ref=page_internal	33,016	33,758
ثوري على العالم تمردي على المجتمع https://www.facebook.com/Thouri.fal.thawra.ontha	20,437	20,404
حقوق المرأة العربية https://www.facebook.com/WomenRightsar	19,856	20,268
Theme: Violence Against Women		
Page name and URL	Number of likes	Number of followers
ضد التحرش - Anti Harassment Movement https://www.facebook.com/AntiHarassmentMovement/	213,071	213,802
Abaad https://www.facebook.com/abaadmena	179,000	197,000
KAFA (enough) Violence & Exploitation https://www.facebook.com/KAFALebanon	148,901	151,638
انتفاضة المرأة في العالم العربي https://www.facebook.com/intifadat.almar2a	110,726	107,367
خريطة التحرش الجنسي HarassMap https://www.facebook.com/HarassMapEgypt	70,679	71,002
شفت تحرش Shoft Ta7rosh https://www.facebook.com/Shoft.Ta7rosh	53,467	53,630
لا لكل أشكال العنف ضد الفتيات والنساء https://www.facebook.com/114575301907797/	43,924	44,028
توليب لدعم المرأة و الطفل - Tswk https://www.facebook.com/tulip.wk/?ref=br_rs	28,556	29,304
تدوين / Tadwein https://www.facebook.com/Tadwein	27,920	29,552
انا مش هاسكت على التحرش https://www.facebook.com/AnaMeshHaskot	17,000	17,000

Table 3

Breakdown of Data by Language, Theme, and Gender

Language	Theme	Gender	Number of Comments	Word Count	Mean Words per Comment
English	Feminism and women's rights	Female	4,253	156,498	37
English	Feminism and women's rights	Male	1,770	68,758	39
English	Violence against women	Female	1,675	67,825	40
English	Violence against women	Male	515	26,139	51
Arabic	Feminism and women's rights	Female	1,288	34,922	27
Arabic	Feminism and women's rights	Male	921	25,365	28
Arabic	Violence against women	Female	843	23,122	27
Arabic	Violence against women	Male	756	21,797	29
Total			12,021	424,426	35 (overall mean comment length)

3.4 Analytical Framework

The present study addressed the research questions it posed by undertaking the analysis of the data in two consecutive stages. First, topic modeling, more particularly Latent Dirichlet Allocation (LDA), which is the most popular method for performing topic modeling, was applied to the corpus of comments under study so as to get an overview of the recurrent themes in the corpus. The second stage of analysis drew on FCDA for an in-depth qualitative analysis of the data. For the purposes of the present study, data analysis at this stage utilized the analytical toolkit proposed by Bucholtz and Hall (2004, 2005, 2010) for analyzing identity as produced in linguistic interaction.

Drawing on insights from a host of different disciplines, Bucholtz and Hall's (2004, 2005, 2010) framework for the linguistic analysis of

identity in interaction rests on an understanding of identity as a sociocultural phenomenon broadly defined as "*the social positioning of self and other*" (Bucholtz & Hall, 2005, p. 586). Bucholtz and Hall's framework is based on five principles: emergence, positionality, indexicality, relationality, and partialness. The emergence principle posits that identity is an interactionally emergent sociocultural phenomenon and that, as such, it does not exist prior to its linguistic reflexes (Bucholtz & Hall, 2005). The positionality principle is premised on a broad perspective on identity that is not limited to macro-level social categories such as gender, class, and race, but also encompasses local ethnographic categories and cultural positions as well as the temporary stances and participant roles that unfold in interaction (Bucholtz & Hall, 2010). The indexicality principle pertains to the mechanism of identity formation. More specifically, it is concerned with how linguistic forms are used to produce identities in discourse. In this regard, indexical processes are taken to refer to the direct and indirect relations created between given linguistic forms and social meanings. They involve all levels of linguistic structure and use including direct references to identity categories and labels, implicatures, presuppositions, styles, and stance-taking (Bucholtz & Hall, 2010).

The relationality principle suggests that identity is not an independent or autonomous phenomenon but a "relational" one, not least because identities "acquire social meaning in relation to other available identity positions and other social actors" (Bucholtz & Hall, 2005, p. 598). Within this view, the intersubjective construction of identities involves a range of complementary relations. The first pair of identity relations is called *adequation* and *distinction*. While adequation concerns the ways in which sameness between two or more entities is constructed, distinction typically works by suppressing similarities and highlighting differences (Bucholtz & Hall, 2004). The second pair of identity relations, namely, *authentication* and *denaturalization*, focuses on the discursive

construction of realness and artifice, respectively. Authentication emphasizes the ways in which identities are verified and produced as genuine; conversely, denaturalization concerns "the ways in which identity is crafted, fragmented, problematic or false" (Bucholtz & Hall, 2010, p. 24). Denaturalization also focuses on how the supposed "naturalness" of identities is disrupted and destabilized (Bucholtz & Hall, 2004).

The third pair of identity relations addresses "the structural and institutional aspects of identity formation" (Bucholtz & Hall, 2010, p. 24). While authorization pertains to the impact of ideology and power relations on affirming or imposing particular identities, illegitimation "addresses the ways in which identities are dismissed, censored, or simply ignored by these same structures" (Bucholtz & Hall, 2005, p. 603). Given that identity formation is both relational and governed by interactional and ideological constraints, Bucholtz and Hall's final principle, that of partialness, stresses the partial nature of any account of identity (Bucholtz & Hall, 2005).

4. Analysis and Results

4.1 LDA Topic Modeling Results

For convenience, the English and Arabic data samples were each broken down into four datasets on the basis of theme and gender. Table 4 shows the eight datasets resulting from this division.

Table 4

Description and Titles of Datasets

English Dataset (Dataset 1)			Arabic Dataset (Dataset 2)		
Dataset Title	Theme	Gender	Dataset Title	Theme	Gender
Dataset 1a	Feminism and women's rights	Female	Dataset 2a	Feminism and women's rights	Female
Dataset 1b	Feminism and women's rights	Male	Dataset 2b	Feminism and women's rights	Male
Dataset 1c	Violence against women	Female	Dataset 2c	Violence against women	Female
Dataset 1d	Violence against women	Male	Dataset 2d	Violence against women	Male

LDA was implemented using the Gensim library in Python (<https://radimrehurek.com/gensim/>). To help determine the optimal number of topics, the coherence score was calculated for each dataset. Ultimately, it was decided to set the datasets under analysis to five topics each.

The output of the LDA topic modeling typically includes (1) the topics for each dataset and their distribution across the entire dataset, and (2) a list of the most frequent words for each topic and their weights. The topics that are generated are not labelled, but are rather given a numerical ID from 0 to 4. Figure 1 shows the top 20 frequently occurring words representing Topic 3 in Dataset 1a. Besides, as the figure shows, the weight that each word contributes to the topic in question is also provided, showing how important each keyword is for the topic in question.

Figure 1

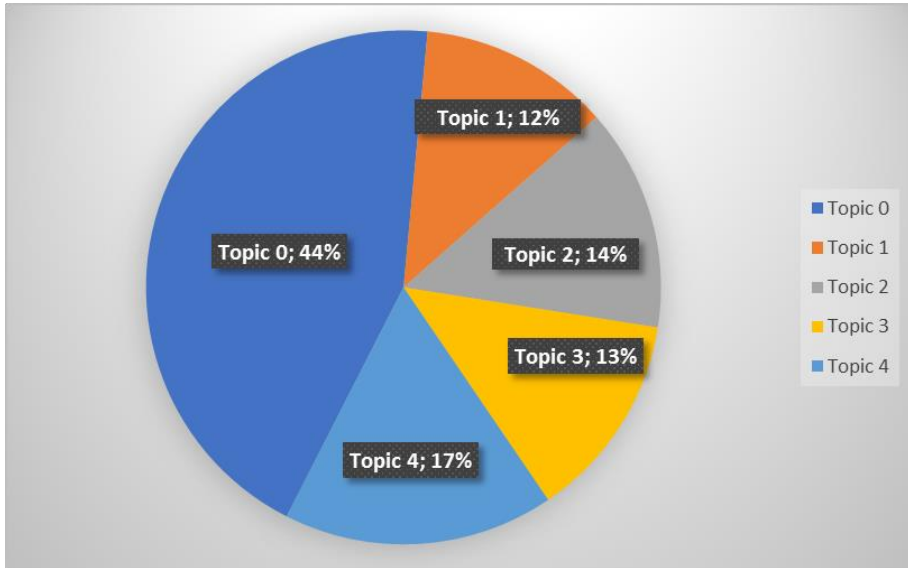
Top 20 Words for Topic 3 in Dataset 1a

```
Topic 3: (3, '0.008*gender" + 0.006*"people" + 0.006*"sex" + 0.006*"trans" + 0.005*"men"
+ 0.004*"women" + 0.004*"equal" + 0.004*"treated" + 0.004*"biological" + 0.003*"female"
+ 0.003*"masculinity" + 0.003*"male" + 0.003*"systematic" + 0.003*"hate" + 0.003*"rights"
+ 0.003*"experience" + 0.002*"oppression" + 0.002*"violence" + 0.002*"decision" +
0.002*"patriarchal"')
```

As mentioned earlier, the distribution of the extracted topics across the entire dataset is also provided. Figure 2 shows the topic weights for the five topics generated for Dataset 1a. As the figure indicates, Topic 0 is the most prominent topic, comprising 44% of the entire dataset and Topic 1, which accounts for only 12% of the dataset in question, is the least prominent topic.

Figure 2

Topic Weights for Dataset 1a



Because topic models only identify the underlying topics in a collection of texts without labeling them, the task of assigning labels to the generated topics is typically left to the researcher. Following the approach adopted in previous research (DiMaggio et al., 2013; Gillings & Hardie, 2023; Nanda et al., 2023; Törnberg & Törnberg, 2016), topic labeling for the present study was conducted by closely examining the top 20 words as well as the top 20 documents representing each topic. To further validate the results of the topic label assignment process, subsamples of the data were reviewed by an independent rater. In case there were any discrepancies in labeling the data in question, these were resolved through discussion.

4.2 Analysis of Gender Construction in a Sample of Facebook User Comments

By identifying the major topics in each dataset as well as the top 20 texts associated with them, the topic modeling of the data under study guided this stage of data analysis by providing insights into which texts to particularly focus on. Thus, informed by the major tenets of FCDA, this section aimed to provide a textual analysis of the top 20 texts representing the major topics in the data under examination. As noted earlier, the analysis particularly drew on Bucholtz and Hall's

(2004, 2005, 2010) framework for examining identity in interaction (see Section 3.4) with a view to exploring how gender is constructed in the collection of comments under study. More specifically, the analysis focused on the indexical processes and identity relations that showcase how gender is (re)produced, negotiated, and/or challenged in the comments in question.

It is worth noting that due to space limitations, the comments discussed in this section are representative of only some of the recurring patterns of gender construction in the datasets under study. All the instances presented below are reproduced verbatim, including any misspellings, typographic, grammatical, and punctuation mistakes as well as any repetitions and unconventional usages included in the original comments.

4.2.1 English Comments by Female Authors (Datasets 1a and 1c)

According to Bucholtz and Hall (2005, p. 594), introducing identity categories and labels constitutes "the most obvious and direct" resource for the construction of identities in discourse. In this regard, the data under study in this section, as well as across all the other datasets, abound with references to the social categories of "women" and "men," often in the generic sense. At the most basic level, these references constitute an act of gender construction in the sense that they "cumulatively produce not only each individual's gender identity, but gender itself as a socially meaningful system" (Bucholtz & Hall, 2005, p. 590).

Example (1), (Dataset 1c, Topic 1)

"Too many people, perhaps most of us, grew up thinking that part of being **a man** involves his ability to dominate & control **women**. Thus, **women** are often put in a position of 'letting **a man** be **a man**' so that he doesn't suffer even a small humiliation; because as I have been lectured '**a real woman** judges herself by the happiness she brings others.' The truth is, authentic happiness comes from within. It is time for **men** to deal with the real sources of their anger, pain & fear. **Man-**

made systems have them down on their knees 'taking it.' *Real men* need to work with **women** so we can create lives of abundance where boundaries matter."

According to Bucholtz and Hall (2010), modifiers, predicates, and any other linguistic forms that are used to qualify or elaborate identity categories also contribute to the constitution of identities in discourse. In the last sentence in Example (1), the author uses the modifier "real" to contest hegemonic masculinity and mainstream gender politics by establishing a relation of distinction between the norm-abiding category of men on the one hand and another category of supposedly "real" men on the other. Additionally, attaching the modifier "real" to "men" in this context also serves to denaturalize the dominant strand of masculinity by dismissing it as false, or rather, problematic. This is particularly reinforced by the use of the modifier "real" with "men" but not with "women" in the same sentence, suggesting that it is this particular category of men and the associated perspective on masculinity it underlies that are being contested by the commenter.

In Example (2), reference to the category of "men" is metaphorically enacted as indexed in the use of the lexical item "penis." This synecdochical reference invokes a sex-based perspective on gender which frames the gender inequalities discussed in the example as primarily based on biological differences.

Example (2), (Dataset 1a, Topic 0)

"[. . .] Well, every retired **woman** who earned a salary deserves reparations. I struggle to get by and an additional 25% in my SS retirement and pension fund would make it right. I have been cheated out of 25% of my income because **I don't have a penis**. Money talks! When **a man** sees **a woman** as economically equal, then **a woman** can have control of her own body."

4.2.2 English Comments by Male Authors (Datasets 1b and 1d)

In the following example, the author invokes a postfeminist

discourse to frame gender inequality and women's oppression as a thing of the past:

Example (3), (Dataset 1b, Topic 3):

"[...] *feminism isn't needed anymore*, feminism has gone from actually helping women get past *real oppression* to now instilling the idea in women's heads that they are in a society that still see's them as second class citizens like in the 50's and 60's etc, *you* want a man that cooks and cleans and knows how to fix things? *Us men* want women who know who do things too, *we* don't call it oppression when *you* want *us* to fix *your* car *we* just do it cause it needs to be done, *you* had a female running for President and *you* still think *your* oppressed? Go to Africa and the Middle East where woman arnt allowed to drive or vote or even be allowed to walk in the street without a man walking with them THATS oppression but *you* never seem to address that do *you*?"

Example (3) illustrates how the inclusive pronouns "we" and "us" are used to typically establish men as an in-group positioned vis-à-vis women, addressed in this example by the second-person pronoun "you." By directly attaching the inclusive pronoun "us" to "men" in the "*Us men*" phrase, the author indexes a relation of distinction from women, thereby signifying a men-only group, typically defined by mutual gender-based interests and concerns. Additionally, the author uses sweeping generalizations to set up a distinction in status between women in the West and women in Africa and the Middle East. By so doing, he presupposes that all women in the West and in regions as vast and diverse as Africa and the Middle East constitute homogeneous groups. More importantly, this example illustrates a case where the yardstick for defining women's oppression is set by a man who defines women's rights narrowly and declares that "*feminism isn't needed anymore*" regardless of the said women's gendered lived experiences. Thus, bringing up postfeminism in this example signals an attempt to undermine the struggles of the women in question to advance gender equality.

A different approach to women's issues and experiences is offered in the following example:

Example (4), (Dataset 1d, Topic 1)

"[Male name omitted] . . . Now as to harassment, *we men just have no idea how serious it is. You can intellectually understand pregnancy or rape, but you will never understand how a woman understands it. Most women do spend their whole lives being harassed almost all the time. They* have no rest. That's why *they're*, as you say, so 'mad.'"

Unlike the previous example where men are conceptualized as an in-group sharing similar interests and positions in opposition to women, and where women's issues are the subject of conceptualization from an androcentric perspective, a distinction is made in the example above between the "*we men*" group and women to point out men's inability to fully understand what women go through. Additionally, the author uses the qualifiers "most" in "*Most women*," "whole" in "*their whole lives*," and "all" in "*almost all the time*" to reinforce his point about the ubiquity of sexual harassment and its consequences as an experience shared by large numbers of women. In this fashion, the author constructs the sexual harassment of women as a gendered experience that only women can *fully* understand. Gender here is thus constructed as a prism through which experiences can take on different meanings, and gender differences are conceptualized as a way for leveraging understanding and solidarity rather than hostility and antagonism.

4.2.3 Arabic Comments by Female Authors (Datasets 2a and 2c)

Example (5), (Dataset 2a, Topic 1):

النساء شقائق الرجال ما اكرمهن الا كريم "معظم الناس فاهمه الشريعة غلط! الرسول قال عشان يقولوا انهم احسن الرجال قوامون على النساء' بيستخدموا آية 'وما اهانهن الا لنيم مننا مثلاً بس هما قوامون بالنفقة اللي لو سقطت سقطت القوامة عنهم وفي مفسرين قالوا انه اللي بينفق هو اللي ليه القوامة الدين حمى حقوق المرأة بس الناس مش فاهمين وعشان

أنا مجتمع نكوري وفيه ثغرات كتير ف بيستخدموا الدين حجة ليهم مع انه في الحقيقة حجة عليهم عشان كذا لازم المرأة تفهم اللي ليها واللي عليها عشان تعرف تاخذ حقوقها في مجتمع زي دا ومعظم النساء ميعرفوش حاجة عن ابسط حقوقهم ك بشر حتى..."

(Most people misinterpret the Sharia law. The Prophet said, "*Women are the counterparts of men*. Only an honorable **man** honors **them** and only a vile **man** degrades **them**." **They just** use the verse "*men are the maintainers of women*" to say, for example, that **they** are better than **us**, **but they** have the right to guardianship over **women** by virtue of maintenance, which if abrogated, male guardianship is revoked. Some interpreters say that guardianship should be granted to whoever maintains the family. *Religion protected women's rights, but people don't understand this*. And because *ours is a patriarchal society with lots of deficiencies*, they use religion to their advantage, although it isn't actually in their favor. For this reason, it's necessary that women know their rights and duties in order to be able to get their rights in such a society. Most women don't have the slightest idea even about their basic rights as humans.)

In the example above, the author uses the category labels "النساء/المرأة" (women) and "الرجال" (men) alongside the associated pronouns "انـ(هم)" (they) and "منـ(نا)" (us) to discuss women's situation vis-à-vis men in the context in question. By making direct references to the hadith and Quranic verse she cites in the comment, the author draws on religious discourse to construct women's oppression as an outcome of people's misinterpretation of religion, rather than of religion per se. This can be seen in the relation of distinction the author establishes between the two religiously-based gendered discourses she invokes: "women as the counterparts of men" versus "men as the guardians of women." By establishing these discourses as antithetical, she exposes how people, particularly men, misinterpret and selectively use religious texts to claim their authority over women, and how they prioritize particular aspects of religious discourse over others. Additionally, the author sets up another relation of distinction between religion as a protector of women's rights and

society's patriarchal norms, ascribing women's situation to the latter. Using the adversative conjunction "ليس" (but) in the statement "الدين " (Religion protected women's rights, but people don't understand this) reinforces this distinction and serves to represent gender-based inequality as socially constructed.

Example (6), (Dataset 2c, Topic 0):

"إلى متى وتموت النساء على يد الرجال بحجة الشرف وغير الشرف لئلا يفسد القانون داعم للرجل حتى في حالات القتل... متى يتم تعديل القانون الذي لا يعاقب الجاني حتى لا تتكرر الجريمة"

(*How long* will **women** die at the hands of **men** under the pretext of honor and dishonor? *Unfortunately, the law supports men even in cases of murder. When is the law that does not punish the perpetrator amended so that this crime is not repeated?*)

In Example (6), the law is established as a supporter of men by enabling them to commit honor crimes. Using the adverb "حتى" (even) in the statement "القانون داعم للرجل حتى في حالات القتل" (the law supports men *even* in cases of murder) implies that the law supports men all along, so much so that it does not stop when violence is as extreme as taking women's lives. The use of a rhetorical question starting with the interrogative "إلى متى" (how long) serves to index the recurring pattern of such acts of violence against women as well as the authors' impatience with them. Thus, in this example, the legal system is constructed as a domain of gender differentiation and discrimination against women, thus exposing the unequal gender relations that mark the sociocultural milieu in question as well as the differential gender-based identity positions and roles women and men typically take up within this domain.

4.2.4 Arabic Comments by Male Authors (Datasets 2b and 2d)

Example (7), (Dataset 2b, Topic 2)

"الديك ديك .. والفرخة فرخة .. بس كده"

(A rooster is a rooster, and a hen is a hen. That's it!)

In Example (7), the author's use of the roosters and hens' metaphor positions him in line with a gender essentialist ideology which suggests that gender differences are rooted in biology and that there are certain essential qualities that make women and men who they are. Notably, by referring to roosters and hens instead of men and women and by suggesting that a given category is its own definition ("a rooster is a rooster"), the author seems to point to a universal, unmistakable essence that marks women and men as purely fixed, unitary entities. The order of the clauses also suggests the primacy of men over women, further reasserting the "male as norm" principle. Finally, the use of the idiomatic expression "ليس كذا" (That's it) at the end of the comment suggests that the author's statement is self-evident, and that there is no more to the whole conversation about gender than this.

Example (8) below draws on a trope that is familiar in the context of sexual harassment, one that both objectifies women and dehumanizes men. Here, the metaphor of women as uncovered meat and men as animals going after them is used to establish women as "equally" responsible for sexual harassment. In addition, by prefacing the statement with "لا يمكن لعاقل" (No sane person can), the author constructs this position as rational and self-evident. Further, this framing serves to reorient the discussion to be about women's clothing and, consequently, to enjoin women to abide by strict gendered codes of conduct. As such, the author's position aligns with gender-based victim blaming narratives.

Example (8), (Dataset 2d, Topic 1)

نكرر انه ليس لاني مهمما كانت ديانتها ان تخرج من بيتها عاريه متبرجه مخالفه لكل [...] "
معايير الاديان ثم نطالب بحق هي اول من تخلي عنه يا فندم لا يمكن لعاقل ان يرمي لحم
امام حيوان ثم يطالبه بعدم محاولته المساس بها "

(We repeat that it is not permissible for any female, regardless of her religion, to go out of her home undressed and unveiled, violating all religious standards, and then we call for a right she was the first to abandon. Sir, *no sane person can throw meat in front of an animal and then ask the animal not to attempt to touch it.*)

5. Discussion and Conclusion

This study sought to investigate the construction of gender in a sample of user comments drawn from the public discussions on Facebook Pages in English and Arabic. The first phase of data analysis involved an LDA topic modeling which identified the major topics in the collection of studied comments. By so doing, the findings of the topic modeling analysis offered an insight into the discursive frameworks within which gender is constructed in the comments in question. In other words, the topics detected in the data under analysis shed light on the discursive strategies that recurred in the comments. For example, the identified topics across all the datasets revealed that society strongly featured in the discussions around gender in the comments, thereby showing how gender issues are inextricably tied to the social realm. The prevalence of religious references in the Arabic datasets attests to the particular importance of this framework for the construction of gender in the Arabic data. Reference to issues relating the gender pay gap, which featured exclusively in the English data, or law enforcement, which was frequent across the English and Arabic datasets, testify to the impact of the institutionalized aspects of gender construction on framing users' gendered positions.

The second phase of data analysis involved a textual analysis of the data from a feminist critical discourse analytic perspective. Overall, the results revealed that as female and male commenters engaged in discussing gender-related issues on Facebook Pages, they drew on a range of indexical processes and identity relations, in Bucholtz and Hall's (2004, 2005, 2010) terms, to construct themselves and others as particular kinds of women and men as well as to index their affiliation, or disaffiliation, with particular gender discourses and ideologies. Commonly used indexical markers included references to identity categories and labels as well as the associated modifiers and/or qualifiers, pronoun use, metaphors, and stance-taking. Besides, all three sets of complementary intersubjective relations identified by Bucholtz and Hall (2004, 2005, 2010) were used in constructing

gender in the studied comments. Commenters, for example, frequently employed the relations of adequation and distinction to show their identification and/or disidentification with particular individuals, groups, or ideologies, all centrally defined on the basis of gender. Relations of sameness and difference were likewise used to establish in-group versus out-group divisions based on the presence (or lack thereof) of mutual gendered interests and perspectives.

Notably, gender was frequently established across all datasets in terms of difference, whether seen as biologically-based or socially-conditioned. Such oppositions include women versus men, a particular type of masculinity or femininity versus another, individual women versus women as a group, and women in the West versus women in non-Western countries. Nonetheless, gender differences sometimes took on new meanings. For instance, instead of using difference to establish opposition or stir hostility between men and women, some male commenters capitalized on gender-based differences to promote allegiance and solidity with women, thus also constructing themselves as pro-women. On the other hand, some female commenters marked their distinction and dissonance from other women who prioritize their personal interests over the public good for women as a class, or those who perpetuate traditional gender roles and thus contribute to reproducing the existing gender order. In this sense, such intra-gender distinctions are instigated by differences in how gender issues are conceptualized.

The intersubjective tactics of authentication and denaturalization were also commonly used to verify particular gender identities and ideologies or to dismiss others as crafted or problematic. Questioning hegemonic masculinity, for instance, frequently involved destabilizing its unfairness, fragility, and unreasonableness. In addition, when gender dynamics and power structures were in question, the relations of authorization and illegitimation were drawn upon to reassert or challenge the institutional and structural constraints governing gender relations. Use of these various indexical processes was also associated

with the invocation of various gender discourses and ideologies such as gender essentialism, gender constructionism, patriarchy, heteronormativity, gender parity, anti-feminist backlash, and postfeminism.

Although some instances in the studied comments feature patterns that have been reported in previous studies to be typical of women's or men's online speech styles (e.g., Herring, 1993), the ways gender construction plays out in user comments reveals a more complex and dynamic picture wherein users draw on multiple resources to construct various gender-based positions and enact a range of identity relations. In line with Lazar's work (2006, 2009), users were also found to draw on the discourse of postfeminism to establish feminist endeavors in particular sociocultural contexts as unwarranted. Such positions are oftentimes based on uninformed conceptualizations of women's rights and gender oppression and are, more importantly, premised on presupposing the homogeneity of diverse groups of women. Furthermore, several of the tactics and discursive strategies used in conceptualizing gender-based violence across the datasets also echo the findings of prior research on the subject. For instance, when discussing various forms of violence against women in the comments, (mostly male) users drew on various linguistic and discursive devices to blame the victims (typically women) and mitigate the responsibility of the perpetrators (typically men), which is consistent with the findings reported in previous research (Clark, 1992/2017; Ehrlich, 2001). This is also particularly consistent with Lazar's (2018) findings which showed that "'victim blaming' and 'perpetrator mitigation' function as wider *translocal/transcultural* discursive logics at work" (p. 378). Crucially, these findings also support a body of research that pointed out that violence against women is gender-based and rooted in structural inequalities (Hourani et al., 2021; Montesanti, 2015; World Health Organization, 2020).

The findings of the present study contribute to the existing literature on the subject by providing insights into how commenting

on SNSs constitutes a significant site for gender construction. The current study also extends scholarship on gender and language and CMC by examining Arabic data. However, these findings should be considered in the light of some limitations. First, the analysis was limited to data from users belonging to the gender binary. Relatedly, it is worth noting that existing gender prediction tools are likewise typically based on a binary gender classification. An additional limitation of name-based gender identification tools concerns the question of how far the classification of users on the basis of their female and male names corresponds to their actual gender identifications. Besides, owing to the need to disaggregate the studied comments by gender, the current analysis could not examine comment threads in their entirety, with the whole comments and replies structure typically marking them undisrupted.

Future research could extend the current inquiry by addressing these limitations. For example, investigating how gender non-binary and genderqueer individuals navigate and construct their gender identities in the commenting spaces on social media could offer valuable insights into the subject, particularly if further research could also examine how gender intersects with other identities such as age, race, or sexuality. Future studies might as well incorporate larger sample sizes as well as data from other languages to further expand the global scope of the literature on the subject. Analysis of full sequences of comment threads is also needed to fully capture the dialogic aspects of gender construction in social media comment sections.

References

- Aboh, R., Mensah, E., Inyabri, I., & Ushuple, L. (2023). My stunning angel: endearment terms as strategies of gender identity construction in Facebook picture uploads in Nigeria. *New Review of Hypermedia and Multimedia*, 29(3-4), 206- 224. <https://doi.org/10.1080/13614568.2024.2328837>
- Androutsopoulos, J. (2014). Computer-mediated communication and linguistic landscapes. In J. Holmes & K. Hazen (Eds.), *Research methods in sociolinguistics: A practical guide* (pp. 74- 90). Wiley Blackwell.
- Barton, D., & Lee, C. (2013). *Language online: Investigating digital texts and practices*. Routledge.
- Beißwenger, M., & Storrer, A. (2008). Corpora of computer-mediated communication. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 292–309). De Gruyter.
- Bouvier, G. (2012). How Facebook users select identity categories for self- presentation. *Journal of Multicultural Discourse*, 7(1), 37- 57. DOI:[10.1080/17447143.2011.652781](https://doi.org/10.1080/17447143.2011.652781)
- Bucholtz, M., & Hall, K. (2004). Language and identity. In A. Duranti (Ed.), *A companion to linguistic anthropology* (pp. 369-94). Blackwell.
- _____. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585-614. 10.1177/1461445605054407
- _____. (2010). Locating identity in language. In C. Llamas, & D.Watt (Eds.), *Language and identities* (pp. 18-28). Edinburgh University Press. <https://doi.org/10.1515/9780748635788-006>
- Cameron, D. (2005). Language, gender, and sexuality: Current issues and new directions. *Applied Linguistics*, 26(4), 482-502. <http://doi.org/10.1093/applin/ami027>

- Ching, C. C., & Foley, B. J. (2012). Introduction: Connecting conversations about technology, learning, and identity. In C. C. Ching & B. J. Foley (Eds.), *Constructing the self in a digital world* (pp. 1-13). Cambridge University Press.
- Clark, K. (2017). The linguistics of blame: Representations of women in *The Sun's* reporting of crimes of sexual violence. In M. Toolan (Ed.), *Language, text and context: Essays in stylistics* (Vol. 7, pp. 208–224). Routledge. (Original work published 1992).
- Cover, R. (2016). *Digital identities: Creating and communicating the online self*. Academic Press.
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* 41(6), 570–606. doi:10.1016/j.poetic.2013.08.004
- Ditchfield, H. (2021). Ethical challenges in collecting and analysing online interactions. In J. Meredith, D. Giles & W. Stommel (Eds.), *Analysing digital interaction* (pp. 23-40). Palgrave Macmillan.
- Ehrlich, S. (2001). *Representing rape: Language and sexual consent*. Routledge.
- Ehrlich, S., & Meyerhoff, M. (2014). Introduction: Language, gender, and sexuality. In S. Ehrlich, M. Meyerhoff & J. Holmes (Eds.), *The handbook of language, gender, and sexuality* (2nd ed., pp. 1-20). John Wiley.
- Emara, I. (2017). Gender identity construction in Facebook statuses of Egyptian young adults. *Cairo Studies in English*, 2017(1), 86-111. [https://doi.org/ 10.21608/cse.2017.22698](https://doi.org/10.21608/cse.2017.22698)
- Georgalou, M. (2015). Beyond the timeline: Constructing time and age identities on Facebook. *Discourse, Context and Media*, 9, 24-33. <https://doi.org/10.1016/j.dcm.2015.07.001>

- _____. (2017). *Discourse and identity on Facebook*. Bloomsbury.
- Gillings, M., & Hardie, A. (2023). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities* 38(2), 530–543. <https://doi.org/10.1093/dlc/fqac075>
- Gil-Lopez, T., Shen, C., Benefield, G. A., Palomares, N. A., Kosinski, M., & Stillwell, D. (2018). One size fits all: Context collapse, self-presentation strategies and language styles on Facebook. *Journal of Computer-Mediated Communication*, 23(3), 127–145. <https://doi.org/10.1093/jcmc/zmy006>
- Gnach, A., Weber, W., Engebretsen, M., & Perrin, D. (2023). *Digital communication and media linguistics: With case studies in journalism, PR, and community communication*. Cambridge University Press.
- Gormley, S. (2015). Language and Gender. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed.).doi.org%2F10.1016%2FB978-0-08-097086-8.53055-4
- Grasmuck, S., Martin, J., & Zhao, S. (2009). Ethno-racial identity displays on Facebook. *Journal of Computer-Mediated Communication*, 15(1), 158-188. <https://doi.org/10.1111/j.1083-6101.2009.01498.x>
- Hall, K., Borba, R., & Hiramoto, M. (2020). Language and gender. *The International Encyclopedia of Linguistic Anthropology*, 1-22. doi: 10.1002/9781118786093. iela0143
- Herring, S.C. (1993). Gender and democracy in computer-mediated communication. *Electronic Journal of Communication*, 3(2). <https://www.cios.org/EJCPUBLIC/003/2/00328.HTML>
- _____. (2019). The co-evolution of computer-mediated communication and computer-mediated discourse analysis. In P. Bou-Franch & P. Garcés-Conejos Blitvich (Eds.), *Analysing digital discourse: New insights and future directions* (pp. 25-

- 67). Palgrave Macmillan.
- Herring, S. C., & Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. In D. Tannen, H. E. Hamilton & D. Schiffrin (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 127-151). John Wiley & Sons.
- Hourani, J., Block, K., Phillimore, J., Bradby, H., Ozcurumez, S., Goodson, L., & Vaughan, C. (2021). Structural and symbolic violence exacerbates the risks and consequences of sexual and gender-Based violence for forced migrant women. *Frontiers in Human Dynamics*, (3)769611, 1-13. <https://doi.org/10.3389/fhumd.2021>.
- Howarth, J. (2025, January 21). Top 35 social media platforms (September2024). *Exploding Topics*. <https://explodingtopics.com/blog/top-social-media-Platforms>
- Jaworska, S., & Nanda, A. (2016). Doing well by talking good: A topic modelling-assisted discourse study of corporate social responsibility. *Applied Linguistics*, 39(3), 1-28. <https://doi.org/10.1093/applin/amw014>
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. New York University Press.
- Jones, R. H., Chik, A., & Hafner, C. A. (2015). Introduction: Discourse analysis and digital practices. In R. H. Jones, A. Chik & C. A. Hafner (Eds.), *Discourse analysis and digital practices: Doing discourse analysis in the digital age* (pp. 1- 17). Routledge.
- Kumar, N. (2025, January 1). Facebook users statistics (2025) — Worldwide data. *Demandsage*. <https://www.demandsage.com/facebook-statistics/>
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45-80. <http://www.jstor.org/stable/4166707?origin=JSTOR-pdf>

- Lakoff, R. (1975). *Language and woman's place*. Harper & Row.
- Lazar, M. M. (2005). Politicizing gender in discourse: Feminist critical discourse analysis as political perspective and praxis. In M. M. Lazar (Ed.), *Feminist critical discourse analysis: Gender power and ideology in discourse* (pp. 1-28). Palgrave Macmillan.
- _____. (2006). "Discover the power of femininity!" Analyzing global 'power femininity' in local advertising. *Feminist Media Studies*, 6(4), 505- 517. DOI:10.1080/14680770600990002
- _____. (2007). Feminist critical discourse analysis: Articulating a feminist discourse praxis. *Critical Discourse Studies*, 4(2), 141-164. <http://dx.doi.org/10.1080/17405900701464816>
- _____. (2009). Entitled to consume: Postfeminist femininity and a culture of post-critique. *Discourse & Communication*, 3(4), 371-400. <http://dx.doi.org/10.1177/1750481309343872>
- _____. (2014). Feminist critical discourse analysis: Relevance for current gender and language research. In S. Ehrlich, M. Meyerhoff & J. Holmes (Eds.), *The handbook of language, gender, and sexuality* (2nd ed., pp. 180-199). Wiley & Sons.
- _____. (2018). Feminist critical discourse analysis. In J. Flowerdew & J. E. Richardson (Eds.), *The Routledge handbook of critical discourse studies* (pp. 372- 387). Routledge.
- Lovink, G. (2011). *Networks without a cause: A critique of social media*. Polity Press.
- Makoni, B. (2013). "Women of the diaspora": A feminist critical discourse analysis of migration narratives of dual career zimbabwean migrants. *Gender and Language*, 7(2), 201–229. <http://dx.doi.org/10.1558/genl.v7i2.203>
- Milani, T. M., & Lazar, M. M. (2017). Seeing from the south: Discourse, gender, and sexuality from southern perspectives. *Journal of Sociolinguistics*, 21(3), 307–319. <https://doi.org/10.1111/josl.12241>

- Montesanti, S. R. (2015). The role of structural and interpersonal violence in the lives of women: A conceptual shift in prevention of gender-based violence. *BMC Women's Health*, 15(93), 1-3. <https://doi.org/10.1186/s12905-015-0247-5>
- Naab, T. K., & Sehl, A. (2017). Studies of user-generated content: A systematic review. *Journalism*, 18(10), 1256–1273. DOI:10.1177/1464884916673557
- Nanda, G., Jaiswal, A., Castellanos, H., Zhou, Y., Choi, A., & Magana, A.J. (2023). Evaluating the coverage and depth of latent dirichlet allocation topic model in comparison with human coding of qualitative data: The case of education research. *Machine Learning & Knowledge Extraction*, 5(2), 473–490. <https://doi.org/10.3390/make5020029>
- Peng, A. Y., Wu, C., & Chen, M. (2024). Sportswomen under the Chinese male gaze: A feminist critical discourse analysis, *Critical Discourse Studies*, 21(1), 34-51. DOI: 10.1080/17405904.2022.2098150
- Rellstab, D. (2007). Staging gender online: Gender plays in Swiss internet relay chats. *Discourse & Society*, 18(6), 765-787. <http://dx.doi.org/10.1177/0957926507082195>
- Rieder, B., Abdulla, R., Poell, T., Woltering, R., & Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said". *Big Data & Society*, 2(2), 1-22. <http://dx.doi.org/10.1177/2053951715614980>
- Sergeant, P., & Tagg, C. (2014). Introduction: The language of social media. In P. Sergeant & C. Tagg (Eds.), *The Language of social media: Identity and community on the internet* (pp. 1-20). Palgrave Macmillan.
- Singh, J. N. (2021). Language, gender and sexuality in 2020: Forward Global South. *Gender and Language*, 15(2), 207-230. <https://doi.org/10.1558/genl.20311>

- Sula, C. (2016). Research ethics in an age of big data. *Bulletin of the Association for Information Science and Technology*, 42(2), 17-21. doi:10.1002/bul2. 2016.1720420207
- Thelwall, M. (2008). Fk yea I swear: Cursing and gender in MySpace. *Corpora*, 3(1), 83-107. <http://dx.doi.org/10.3366/e1749503208000087>
- Törnberg, A., & Törnberg, P. (2016). Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, 27(4), 401-422. <http://dx.doi.org/10.1177/0957926516634546>
- Townsend, L., & Wallace, C. (2016). Social media research: A guide to ethics. University of Aberdeen. Retrieved from <https://www.bolton.ac.uk/assets/Uploads/Social-media-ethics-study-Aberdeen-2018.pdf>
- Wang Y., Burke, M., & Kraut, R. E. (2013). Gender, topic, and audience response: An analysis of user-generated content on Facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 31–34). ACM. <http://dx.doi.org/10.1145/2470654.2470659>
- Weatherall, A. (2016). Discourse and gender. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*, 1-5. doi: 10.1002/9781118663219.wbegss150
- Widdowson, H. G. (2004). *Text, context, pretext: Critical issues in discourse analysis*. Blackwell.
- Wittkower, D. E. (2014). Facebook and dramauthentic identity: A post-Goffmanian theory of identity performance on SNS. *First Monday*, 19(4). DOI: <https://doi.org/10.5210/fm.v19i4.4858>
- World Health Organization. (2020, July 8). *Violence against women*. https://www.who.int/health-topics/violence-against-women#tab=tab_1

- Zhao, S., Grasmuck, S., & Martin, J. (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior*, 24(5), 1816–1836. DOI: [10.1016/j.chb.2008.02.012](https://doi.org/10.1016/j.chb.2008.02.012)
- Zhukova, M., & Herring, S. (2024). Benign or toxic? Differences in emoji interpretation by gender, generation, and emoji type. *Language@internet*, 22(Special Issue), 74-108. DOI: 10.14434/li.v22.40800