

# A Computerized Analysis of Gender Linguistic Patterns as Reflected in the Jordanians' Facebook Statuses: Lexical Items, Affect, Theme, Identity and More

Awni Shati Etaywe

Jordanian Association of Translators and Applied Linguists  
Amman, Jordan

## ABSTRACT

Facebook has revolutionized the way people share information and use language which has the potential to mark individual and collective identity. This makes exploring patterns in language use of a paramount significance. One conventional marker of what is being expressed and how is the word used in writing. Focusing on the used words, this study aims at exploring the potential distinguishing patterns in language used by Jordanian male and female authors of self-generated Facebook-status texts. Data were collected from a purposive sample by means of the 'social network' model, and then categorized and analyzed statistically using eclectic computerized (stylometric and online automated) analysis tools, including LIWC. Results demonstrated the impact of gender on discriminating language patterns and, thus, might facilitate gender specific status-predictability. Men and women wrote their statuses in statistically different ways. Women exhibited higher frequency of use of (social) lexical items, first person and second person pronouns, and optimistic/'upbeat' tone. Men's statuses revealed more third person pronouns, positive affect and clout in their writing. Geopolitical and sport theme-related words characterized male-authored themes and male semantic orientation. This study emphasizes the value of linguistic work with focus on resources used in social media to mirror themes, affect, tone, clout and more. It also provides implications for (electronic) discourse analysis, text/corpus analysis, sociolinguistics and forensic linguistics.

## General Terms

Pattern, Facebook status, Linguistic Inquiry, Word counting.

## Keywords

Linguistic pattern, gender, lexical item, theme, affect, tone

## 1. INTRODUCTION

*"[W]e write [...] with the whole person. The nerve which controls the pen [or the keyboard] winds itself about every fib[er] of our being, threads the heart, [and] pierces the liver." Virginia Woolf, Orlando*

*"High frequency lexicon could discriminate the works of authors from each other." Olsson*

*"There are predictable differences in language used between sexes [making it] possible to predict the sex of the user without knowledge of the true sex." Tausczik and Pennebaker*

Departing from Virginia Woolf's words, words of a Facebook status-author could be looked at as a mirror of an author's personality, daily topics, concerns, emotions, thoughts, relationships and above all "how we are organizing and analyzing our worlds" [1]. The style of the individual authors can be discriminated by investigating the frequency of lexical items (LIs) use [2]. Such investigation could be applied to

discriminate the works of authors of different genders, as gender may influence the language style and, hence, emphasize the need to investigate the patterns and frequency of word use by men and women. Research on the influence of gender has drawn the interest of several researchers and for various purposes [3].

This study has come, however, with an interest in social media authorship to help in showing how very common computerized tools may help differentiate between different genders of authors. Besides the qualitative methods, researchers have recently tended to apply automated/computerized text analysis applications and tools such as Coh-Metrix, word-it-out and Linguistic Inquiry and Word Count (LIWC) to text analysis tasks. The capabilities of such tools seem to be the reason behind the use of them. Applying the computerized analysis approach, this study aims at exploring the frequency of patterns in language use of Jordanian men and women, as users of Facebook, as reflected in their writing of Facebook statuses. The study explores and marks statistically the frequent LIs choice (i.e. the content words), in addition to the person pronouns, and the characteristic affect, emotional tone, clout and themes in Jordanians' Facebook-statuses.

Like many people around the world, affected by the rapid development of computer mediated communication and mobile phone technology, Jordanians have entered an unprecedented digital-like society that uses many advances in communication. Among these advances Facebook stands out as an inseparable part of people lives; hence, Facebook is influencing the Jordanian users' writing style which has received little attention by researchers despite the abundance of data for such studies. Data in statuses are getting more and more important with the rise in the daily active status users of Facebook. The data can give information about users' attitudes, identities, mood and so forth, making it a good reference for analysts working in linguistics, security and law enforcement field, intelligence, psychology, sociology, and so forth. The data and information can be considered countless, particularly when we consider that 'Statista Website' has revealed that "as of the third quarter of 2017, Facebook had 2.07 billion monthly active users".

The significance of this study, therefore, lies in the target population of the study, the Jordanian users of Facebook. In addition, the study comes to add to the existing literature on Facebook statuses as a source of linguistic data, considering that Facebook statuses have received little attention so far. Reviewing the literature on gender specific-authorship verification and gender linguistic style of Facebook status-writing in general, and the writing of Jordanian users in particular has revealed an absence of research dealing with the characteristic style of Jordanian male and female users. This gap was the impetus of this study. It is believed that this is one

of the first studies, if not the first, to investigate the Jordanian male versus female users' frequent lexical choice, themes, affect, tone and clout in Facebook statuses as informed by computerized tools.

This study contributes to the existing literature and seeks to establish a practice for concerned agencies in Jordan as well as other countries in the area of social networking language and individual and collective identity finding, taking into account the urgent need for such a practice to be introduced to such agencies in Jordan as security agencies (and, mainly, among intelligence and cyber crime practitioners). Therefore, a line of enquiry linked to the gender (of one's specific identity) marking is badly needed to narrow the field of investigation. Identifying the gender of an author could be a steering point, especially to those involved in cyber crimes and narrowing the pool of search. Therefore, this significance of gender specific words- exploration could be of value to whatever agency or practitioner with interest in language user profiling, identification and comparison. The study claims that it establishes a practice ground to distinguish male users from female users, and thus may help concerned agencies identify gendered authorship in civil cases. It facilitates decision making in regards to the gender of author in disputes. It would encourage other researchers and practitioners to apply similar techniques to find answers for similar questions related to real world cases based on larger corpora of target 'individual or collective' statuses.

The study is, however, limited to a sample of a network of friends' (400) statuses; this would impact the generalization of findings. Moreover, a challenge facing differentiation between the gender of authors of Facebook statuses consists in that statuses are generally "short texts [whose] analysis [is likely to reflect] low inter-author variation [given that] the texts are of the same type", and thus imposing the "short text stability problem" [4]. This has led the researcher to create a corpus of (200) statuses authored by several users of the same gender in pursuit of more and more words and thus patterns of language use of the entire group of that specific gender. The study is also limited to the use of statistical analysis. A combination of quantitative and statistical methods might be perceived better in such type of studies. This study seeks, though, to answer the following focus questions:

1. What are the distinctive lexical-items and person pronouns used by male and female Facebook status text authors?
2. What are the gender-reference affect, tone and clout that are conveyed in Jordanians' Facebook statuses?
3. What are the typical themes (underpinning identities) of Jordanian users that are reflected in Facebook statuses?

To elucidate confusion related to the study terms, it is of a benefit to mention that 'Facebook Website' defines a Facebook user's status as an "update feature [posted on user's wall and shown in the news feeds] which allows users to [give and] discuss their thoughts, whereabouts, or important information with their friends" (Facebook.com). Also, the term affect is used to refer to 'the author's evaluations, opinions, emotions and speculations' [5]. Clout is used to refer to "the relative social status, confidence or leadership that people display through their writing [, and] it taps writing that is authoritative, confident and exhibiting leadership" (LIWC.net).

## **2. THEORETICAL BACKGROUND AND LITERATURE REVIEW**

### **2.1 Theoretical background**

With the rise of numbers of users of social media networks such as Twitter, WhatsApp and Facebook, electronic discourse analysts tend to continue to check the influence of gender on language and they do that by various techniques and methodologies, including the computerized analyses. This could be ascribed to the importance of gender as an influencing variable when studying writing style and "for lack of agreement over the best way to analyze language" [6]. With that in mind, this study positions itself within the context of 'identification' area of linguistic evidence [7] and 'authorship comparison' [8] that applies stylometric automated approaches to "distinguishing between and categorizing texts and authors on the basis of the relative frequencies of pre-defined linguistic features and various statistical techniques" [9] such as the computerized analysis tradition followed by pioneering researchers such as [10, 11]. It is also an extension to Olsson's (2008) notion 'on LIs as being markers to discriminate the works of different authors' and [12] emphasis on the value of focusing on linguistic work in the areas of discourse analysis with the intention to highlight discourse resources used in social media to render attitude and values.

In this study, it is hypothesized that the socio-cultural context in which male and female Facebook users live and interact has an impact on the LIs employed, the themes discussed and the affect, tone and clout conveyed, but at varying degrees and frequencies due to differences in gender. It is also hypothesized that frequencies might be so distinctive that they signal different gender-identities. The study hypothesizes the validity of the argument made, first, by [13] about LIs as being discriminating markers and, second, by [14] about people's language use as a behavioral manifestation that can reflect predictable differences between sexes, in such terms as marking negative and positive attitude [15] and highlighting theme (and more) codes which are most of the time stimulated by ideology [16] and triggering underpinned identities and ideologies.

### **2.2 Literature Review**

Computer mediated communication and social media applications have been the center of interest of recent studies in so many disciplines because of their multi-dimensional effects and relationships with key issues in our life. Many studies in linguistics, pedagogy, politics, and sociology etcetera have been carried out in relation to computer mediated communication and social media due to the influential role they have played over the past few years. Socially and politically, for example, social media played a pivotal role in shaping the reality of revolution in Egypt [17], when the Egyptian activists used it as a platform of communication for coordinating and planning for the historic event of January 25, 2011 which forced President Mubarak to step down [18]. Pedagogically, studies have found an influence of social media on tertiary students' learning performance [19]. Using such applications has been found with negative consequences on users; this includes lack of concentration in lectures, problems of language spellings and grammars, distraction from studies, and submitting incomplete assignments due to constant use of such applications [20].

Linguistically speaking, others see that social media are rich source for linguistic studies. Computer and mobile phone mediated communication has involved the exchange of

messages synchronously or asynchronously using applications such as emails, mobile text messages and social networking websites, and interest in language choice among language users such as the Jordanian university students [21, 22]. Despite the previous focus on emails and mobile text messages, social networks have become the trendy focus with the rise in numbers of users of such networks as Facebook, WhatsApp and Twitter. Today, they constitute a platform to investigate certain linguistic patterns, and thus could be considered a ground for evidence and indicators of people's characteristics [23, 24]. Even affiliations of people could be traced linguistically by investigating social media. In her study on how people use language to create affiliation on the Web, Zappavigna in [25] studied the discourse of Twitter and social media with a focus on linguistic work in the areas of pragmatics and discourse analysis. She explored the prevailing communicative conventions arising via Twitter (e.g. retweeting and hashtagging). Highlighting discourse resources used to render attitude and values, Zappavigna introduced the term 'searchable talk' to refer to the discourse rendering opinion and sentiment findable through resources such as social tagging which help uncover affiliations.

Shedding light on Facebook and its statuses, Facebook is looked at as an appealing experience for billions of people around the world as can be realized through the eight users' online experience measures developed by [26]. These measures are as follows:

1. Stimulation and Inspiration (e.g. Facebook makes one a more interesting person).
2. Social Facilitation (e.g. Facebook often gives one something to talk about).
3. Temporal (e.g. logging on Facebook is part of one's routine).
4. Self-Esteem and Civic Mindedness (e.g. one may use Facebook to reflect his/her values as a better part of his/her community).
5. Intrinsic Enjoyment (e.g. Facebook is a treat for oneself).
6. Utilitarian (e.g. one gives advice and tips to people he/she knows).
7. Participation and Socializing (e.g. one contributes to the conversation on this site).
8. Community (e.g. a big reason one likes Facebook is what he/she gets from other users; Facebook does a good job of getting its visitors to contribute or provide feedback).

These measures inform us about the major reasons why people refer to use Facebook as a medium of communication, presenting it as an area of research. Few researchers such as [27, 28, 29] have investigated the semantic patterns and the speech acts in users' statuses. Some others have focused on the frequency of positive and negative emotional words in participants' status updates, and the relationship between gender and emotional expressiveness, see [30] for example, highlighting a line of enquiry that focuses on affect and attitude reflected in statuses. Detecting LIs loaded with affect could trigger how LIs may help trace the feared situations or people, and capture information about people's stance and evaluation of an external or internal event and beliefs about something or someone [31]. This shows the relatedness of the present study to the wider research circles.

Other studies have sought to identify the themes shared in users' Facebook statuses [32, 33]. This tradition of thematic

analysis has started ages ago. Thematic apperception test (TAT) researchers [34] analyzed the stories told in response to drawings of people and could identify key clues that revealed the writers' power and needs for affiliation and achievement by analyzing tagged words or phrases related to pre-identified dimensions. In the second millennium, researchers like [35], however, investigated the themes among other issues in their largest Facebook study in Sweden by using a self-report questionnaire; however, this data collection technique might not reflect the actual generated themes as would be realized in real sample of statuses instead of the questionnaire. This drawback is avoided in the present study by analyzing a sample of statuses instead of using a questionnaire; yet, it shares with [35] the interest in themes.

Unlike [36], [37] used automated analysis. They used statistical generative method, Latent Dirichlet Allocation, looking for clusters of co-occurring words to identify hidden topics and then classify them as grasped in about half a million Facebook statuses. They could identify 25 themes such as family, politics, work, religion, etc. Although automated analysis may not provide an in-depth user-generated content analysis of the meaning of the collected data (as the inductive bottom-up thematic analysis otherwise would), automated analysis is faster and could be looked at as replicable, for excluding the personal analysis and relying solely on the computerized analysis.

Affect, themes and more dimensions can be investigated by shedding light on the content words (as well as the function words). This study, however, will be mainly focusing on the content words, given their rich properties, in addition to the person pronouns from the list of function words for its significance in pointing to author's style and other cognitive and emotional characteristics. Focus on content words is ascribed to the large number of them (i.e. nouns, verbs, adjectives and adverbs) that represent 99.5% of the total English words in our vocabulary and are used to convey the 'content of a communication' which is the focus of the study, while there are only about 0.05% function/style words which, nevertheless, make up about 55% of all the words we use and are mentally processed differently from the content words [38, 39].

Out of the function words, person pronouns are dealt with in here due to their reference to the power, status and relationship that the author demonstrates in writing. Higher-status individuals 'more often make statements that involve others whereas lower-status individuals are more self-focused'; the use of first-person plural pronoun may indicate a higher rank or a higher status of the author/speaker while the first-person singular could be a good indicator of lower status [40]. Interestingly, even an attention to the 'personalization' loaded in person pronouns used in statuses can easily be achieved by means of the online applications. Shedding light on person pronouns may help recognize how much attention in statuses is drawn to the SNW-user or the reader of the statuses. This focus on pronouns is sought in this study to explore how distinctively a gender specific writer uses person pronouns. Personalization generally helps position the readers/audience as agreeing to what is stated in the status. It also creates a feeling of solidarity and intimacy, and/or an impression of interactivity. Impersonalization, in contrast, creates a feeling of objectivity and freedom of personal bias by using clauses like 'it is disappointing/necessary that...'. Personalization is generally managed by pronoun (which is the area of interest in this study), directives, rhetorical questions and questions coming from imagined participant.

Regarding the emotional tone inherent in one's linguistic characteristics, [41] state that tone can "help identify individual differences" for it is influenced by the author's "age, sex and personality". In LIWC, for example, emotional tone is scored such that higher numbers are more positive and upbeat tone, and lower numbers are more negative tone. That said, when it comes to the connection between 'gender' and 'the used content words (and pronouns) and their themes and loaded tone and affect', it is worth considering Tausczik and Pennebaker's argument about content words and pronouns as one's language features that can be used to make predictions about individuals of different sexes. Tausczik and Pennebaker mention that women use more social words, pronouns and references to others, and men use more complex language and more words.

Based on the above, a language user's choices of language resources help uncover characteristics such as themes, tone, relationships, affect and 'ideological stance and perception on the world' [42]. The same choice may be interpreted meaningfully based on the socio-cultural context in which resources are exchanged [43]. This underlines the value of considering the context of language use. In other words, paying attention to what resources are used by a language user and in which context or society carries a great value to social issues and understanding society and culture-specific aspects such as attitudes, common themes and interpersonal relations; it also reveals semantic orientation and one's identity and ideology [44].

Correlated to the author comparison and identification area of research is the study of and defining one's identity which can be mapped onto the used LIs which, according to [45], usually express the author's identity that informs social relationships and communicative exchanges within established contexts (yet, defined by social variables such as gender). According to [46], an individual can have an identity as a woman, a black, a Muslim and so forth, and that identity may serve as "a frame of reference within which our recognition of an entity takes place". People may possess "multiple identities on the basis of the multiple roles they are capable of representing in the socio-cultural relationships in which they participate" [47]. This means that language presents self-presentation [48]. Exploring the characteristic style of language use by a gender could therefore be used for "labeling [...such] normative characteristics or practices as constituting the core of an individual or group which are then used to define them" since language is considered "an acceptable identity marker" [49]. According to [50], there are two broad identity types, individual and collective, to which the following belong:

1. Ethnic identity which is associated with a cultural group;
2. Racial identity which is biological genetic make-up related;
3. National identity which is associated with a nation state;
4. Gendered identity which refers to socially constructed notions of femininities and masculinities;
5. Social class identity; and
6. Language and dialect.

Highlighting identities based on the emerging themes gains importance when we consider the reality that identities in the Jordanian society (as in other societies) emerge "under security stress, conflict and lack of security" Suleiman, cited in [51] that I, as an insider, see them evident in the Jordanian

society. This reflects the potential of Facebook statuses as an area for identity exploration by means of thematic analysis. Not only identities but also social norms and cultural aspects may rise from statuses. [52] states that because people of a certain society and culture would value specific things and perform them in a certain way, their culture and social norms are likely to be reflected in the language they employ- presenting a potential to reflect a particular speech community's shared set of norms, local knowledge, experiences, beliefs and values besides identities. Matching themes with underpinned identities, as described by [53], would be a secondary task achieved in this study.

Given all that, it is assumed that a computerized analysis of SM real statuses can turn them into self-reports that would mirror their authors socio-culturally. Tracing all elements raised in the study questions is believed to be feasibly fast and replicable when considering using a computerized tool, and this is why computerized tools are used in this study. Researchers today can use, for example, Coh-Metrix, LIWC, etc. Coh-Metrix provides 53 language and discourse measures at levels related to conceptual knowledge, cohesion, lexical difficulty, syntactic complexity and simple incidence scores (<http://cohmetrix.memphis.edu>).

According to [54], LIWC was introduced to provide an "efficient and effective method for studying the various emotional, cognitive, and structural components present in individuals' verbal and written speech samples". With its text processing module and an internal default dictionary, LIWC can classify words into around 80 linguistic and psychological categories, mainly: 22 standard linguistic dimensions (e.g., percentage of words in the text that are pronouns, articles, auxiliary verbs, etc.) and 32 word categories tapping psychological constructs (e.g., affect including the positive and negative emotions, cognition, and biological processes) [55]. This means that LIWC is text analysis software that enables the extraction of features from natural language text, which makes it suitable for the present study, especially after developers upgraded it from LIWC 2007 into LIWC 2015. Identifying patterns by means of LIWC is likely to "provide windows into [Facebook users'] emotional and cognitive worlds" [56], including tone, clout and affects which are part of this study's interest.

Researchers, today, may also use accessible corpus analysis systems for larger corpus and data for various functions. Researchers may use CLAWS (the Constituent Likelihood Automatic Word-tagging System) Web Tagger, software for English texts developed by Paul Rayson and Roger Garside at the University of Lancaster. CLAWS part-of-speech (POS) tagger is the most common form of corpus annotation, and was the first form of annotation developed by University Centre for Computer Corpus Research on Language, at Lancaster. It provides the user a very accurate result in a corpus analysis task; the latest version of the tagger, CLAWS4, uses 100 million words of the British National Corpus (BNC) (see [ucrel.lancs.ac.uk/claws/](http://ucrel.lancs.ac.uk/claws/)).

To conclude the above, Facebook users' statuses are receiving more attention as a rich resource of data on users' emotions, thoughts, plans, ideologies, themes and so forth. Status texts can turn the very statuses into 'self-reports' of users who might be unaware of unveiling their personal, socio-cultural and ideological views when authoring their statuses. Statuses may also report on gender, society and culture. This potential of statuses should be a call for researchers of authorship and (forensic) stylistics to bend over status analysis, whether via a

qualitative or/and statistical analysis. This study solely depends on statistical analysis, using computerized tools.

### 3. METHODOLOGY

The Jordanian society represents the population of the study. The sample of the study is made up of 400 Facebook users, who represents the purposive sample of the study and their statuses were the source of data of the study. The data was collected by means of the 'social network' model, suggested by [57], whereby varied sample statuses are approached with the help of networks of the researcher's friends who could provide him with the statuses appearing on their Facebook statuses. The same friends were requested to randomly collect and send their Facebook statuses as well as their friends' statuses in a document file for male Facebook users and another for female Facebook users. Friends' personal accounts were the source of data, and they had no restricted privacy on their statuses, which made it possible to collect their statuses for the purpose of completing the study.

400 Facebook statuses presented the material of this study. The randomly chosen statuses were then equally divided into a male group and a female group, according to the gender of the account user, since the gender variable is the one of interest for this study. The collected statuses were copied and pasted in two word files: one for male statuses and another for the female's. Male users' (200) statuses made up the 4,246 word male corpus, while the female users' (200) statuses built the 3,558 word female corpus. The statuses were found written in Arabic (the native language of the Jordanians) and English (as a language occasionally used in Jordan, formally taught as a foreign language in schools, and used as a medium of teaching in higher education institutions). The language used is excluded from this study. Statuses in Arabic were translated into English to fit the entry language of computerized tools. The following are some examples of the statuses (along with transcription and translation of those found written in Arabic):

- 1/ ?attojur ʕala ʔaʃkæliha taqaʕ/ (الطيور على أشكالها تقع)  
[Birds of a feather flock together]
2. Happy 23rd Birthday to my best friend, my everything!  
I love you always Rasha.
- 3/badaʔtu ʕamali ʔaldʒadid ka ʔustæð musæfed fi  
aldʒamiʕah ʔalordunjah fi qism dʒirahat ʔaddimæy  
walʔaʕsab  
(بدأت عملي الجديد كأستاذ مساعد - الجامعة الأردنية - قسم جراحة  
الدماع والأعصاب)  
[I started my new job as Assistant Professor of  
Neurosurgery at the university of Jordan]
4. /surja ʔalʔæn ʔaʃbah bi leʕbbt ʔetrandʒ saqat ʔlwazir  
walqilæʕ walfijalah tabaqa faqat keʕ malek/  
(سوريا الآن أشبه بلعبة شطرنج: سقط الوزير والقلاع والبيادق  
والفيلة، تبقى فقط كش ملك)  
[Syria now is like a chess game: bishop, rooks, pawns  
and knights are down. The only thing remaining is  
King in check]
5. /dʒaw ʔalxalidʒ wa ʔasʕar oroba waʕifat ʔasumæl:  
alordun juraheb bekom  
(جو الخليج ..أسعار أوروبا ..عيشة الصومال ... الأردن ترحب  
بكم)  
[Here you have the Gulf weather, European prices and  
Somali life: Welcome to Jordan]

6. Ramadan Kareem to everyone.
7. Merry Christmas dear friends, enjoy the holiday!
8. Ladies you are all welcome to join me in our new  
"Group Therapy Session" on Saturday June 14th  
from 1-2 pm at Fitness One.
9. /Fi ʔasjæsaʕ lajsa hunæka ʕado dæʔem ʔaw sadiq  
dæʔem hunæk masaleh dæʔimah/  
(في السياسة ليس هناك عدو دائم أو صديق دائم هناك مصالح  
دائمة)

[A nation has no permanent enemies nor permanent friends;  
only permanent interests it has]

Data analysis was conducted under the hypothesis that the gender of Facebook users and the socio-cultural context to which the Facebook status users belong and in which they interact has a fundamental impact on the language resources used. It is also hypothesized that frequent LIs, pronouns, themes, affect, tone and clout in statuses would be distinctively different in male users' statuses from those of female users. Following the tradition of such researchers as [58] and others who tend to conduct computerized analysis by such programs and applications as 'Language Inquiry and Word Count (LIWC) and the online word cloud-creators and word-it-out tools', this study assumes the validity of that statistics-based analysis is considered able to provide 'objective, replicable and reliable' [59] results if repeated by different users at any time using the same corpus.

For frequent LIs and person pronouns used by Jordanian users in Facebook statuses, the online word counting tool 'countwordsworth' was used for counting the number of words and person pronouns- a thing the tool can perform along with many other services (like calculating words per sentence and prepositions, etc). Using this tool was a matter of 'copy and paste' of each gender group's corpus in the 'paste' field of the tool (<http://countwordsworth.com/>), which made it easy to obtain the most frequently used LIs and person pronouns under appropriate headings that showed the frequency of use of each person pronouns type and number of use of each word in corpus. Every LI with a frequency of 0.05% of repetition of use (in comparison to the total word count of each corpus) was considered significant/ distinctive for the purpose of this study, given the conventional significance level of  $p < 0.05$  in statistics.

For the emerging gender-reference affect, tone and clout in status writing, LIWC 2015 (LIWC.wpengine.com) along with its relative reference frequencies of pre-defined linguistic features (of affect, tone and clout) was used. LIWC 2015 was used instead of LIWC 2007 due to the better characteristics of LIWC 2015 pertinent to accuracy and the slight variations in "word count, words per sentence, and selected punctuations [of] more precise counting metrics. [...] LIWC 2015 is the gold standard in computerized text analysis [and through which one learns how words] reveal out thoughts, feelings, personality, and motivations" (LIWC.wpengine.com). LIWC 2015 classifies the analyzed texts into personal writing, personal email correspondence, professional correspondence, social media (Facebook, Twitter, blog), commercial writing, professional or scientific writing, and other. Among which, social media (Facebook) option fits the present study's focus question two. Therefore, it seemed ideal to explore affect, whether negative or positive, emotional tone and clout as loaded in words used by each gender group by means of LIWC 2015.

Furthermore, all it took was only to copy each gender group's corpus at a time and then to paste it in the 'paste field' after

choosing the classification of the sample text (i.e. 'social media' text). The results were shown simultaneously and distributed into LIWC dimensions (which include among others: positive emotions, negative emotions, emotional tone and clout) together with the percentage within sample data, and LIWC average frequencies for social media/Facebook affect, tone and clout. Bearing in mind the LIWC reference average for social media, gender groups' variation in percentages would characterize the Jordanian male and female statuses.

VocabGrabber, the online word-it-out tool and sentence counting tool of visual thesaurus, was used to provide us with the characteristic themes (which would underpin traced identities) by means of the VocabGrabber-generated 'tag cloud' of list of words and phrases used in corpus. All it required was to access VocabGrabber, copy texts of each gender group's statuses from the word file and paste it in the 'paste' field on VocabGrabber, and then click on the 'Grab Vocabulary'

button (<https://www.visualthesaurus.com/vocabgrabber/>). The computerized online tool automatically generated a list of the used vocabulary in a tag cloud view, showing how words were used in context and providing options to sort the vocabulary the way that serves study along with tool-suggested theme categories colored differently. In this first step, the tag/word cloud help highlight the potential themes of words used in statuses of each gender group. However, the VocabGrabber word related themes were extended and modified to better represent the corpus, leading to the second step.

In the second step, themes are analyzed through 'field analysis' of LIs. In other words, the subjects or domains of experience to which the LIs refer were identified based on knowledge of dictionary meaning of LIs and the researcher's intuition as an insider of his society. LIs were linked up with their interrelated sub-themes/sub fields and then categorized under a theme category such as religious and socio-economical themes. The result of step one and step two provided us with the sought after themes. Making use of field-related words and the rising labeling of theme categories, matching is done, in the third step, between the potential Jordanian Facebook users' identities and the general categorization of identities, set by [60], (ethnic identity, racial identity, national identity, gendered identity, social class identity, and language and dialect), taking into account that themes are expressive of social norms and identities and 'mostly stimulated by ideology' [61].

## 4. RESULTS AND DISCUSSION

### 4.1 Distinctive lexical-items and person pronouns

Using the online word counting tool 'count words worth', female group used 3,558 words with an average of 8.64 words per sentence, while male group used 4,246 words in the same number of statuses with an average of 9.25 words per sentence. The maximum sentence length of female users was 80-word long, whereas the male longest sentence was 151-word long. This indicates that female Jordanian Facebook status-authors tend to write slightly shorter statuses than men's. Nevertheless, the difference seems to be clearer when it comes to the most frequently used LIs and person pronouns. See Table 1 below for distinctive LIs use by each gender group, and Table 2 for variation in person pronouns use by the two gender groups.

**Table 1. Distinctive lexical items in Facebook statuses**

Groups	Most frequent LI	Frequency
Female status authors	Love	0.155
	Friends	0.11
	Make	0.085
	Happy	0.075
	Day life; thank	0.07 (each)
	People; life; family	0.06 (each)
Male status authors	Good	0.095
	success; know; nothing	0.09
	Day	0.085
	Man	0.08
	New	0.055

Table 1 shows the distinctive LIs in Facebook statuses of both gender groups. The LIs that were found to be used more than 10 times in every 200 statuses were tabulated along with their frequency of use. In other words, every LI repeated in 0.05 of the statuses or more was considered significant/ distinctive. On the first hand, the words love, friend, make, happy, day life, thank, people, life and family were distinctive in female statuses, with love (0.155%) which appeared 31 times and friends (0.11%) which appeared 22 times as the two most frequently used LIs. On the other hand, the words good, 'success, know and nothing', day, man and new were the distinctively present in male statuses, with good 'success, know and nothing' appearing almost equally in statuses (0.095% and 0.09% respectively), about 19 times each, followed by day (0.085%), man (0.08%) and new (0.055% each). In addition, female status-authors tend to repeat the use of certain LIs such as love and friends more frequently than male status-authors do. These identified content words may help make predictions about individuals' gender, as raised by [62]; women showed preference of use of more social words. Concerning person pronouns, now look at Table 2 below.

**Table 2. Person pronouns in Facebook statuses**

Groups	1st person pronoun	2nd person pronoun	3rd person pronoun
Female status authors	50%	31%	17%
Male status authors	43%	24%	31%

Table 2 presents how person pronouns are used by female and male authors of Facebook statuses. Both female and male authors showed preference of use of first person pronouns. Female authors used first person pronouns 50% of the times of person pronoun use, while male authors used them in 43% of the times. This preference is likely driven by Jordanian

status-authors' concern with creating a feeling of solidarity, intimacy, agreeability and interactivity which Facebook facilitates. It is a way to achieve personalization.

The second person pronoun was the second in frequency of appearance in female authored statuses (31%) and the third person pronouns were the least used (17%). Conversely, male authored statuses showed third person pronouns as the second in frequency of appearance (31%) while the second person pronouns were the least used with a frequency of 24%. The results may help indicate that the statuses with least third person pronouns are likely to be authored by female users. Drawing such results on pronouns may help make predictions about the gender of authors, as the pronoun-use is defended by [63] as individual's style marker. The use of far more third person pronouns by male authors (31%) than the use of it by female authors (17%) could be considered as a clear marker of male's statuses which tend to position the readers/audience as imagined ones.

#### 4.2 The gender-reference emerging affect, tone and clout

LIWC analysis results distributed into their respective dimensions against LIWC reference average for social (Facebook, in particular) are presented in Table 3 below.

**Table 3. Gender-reference emerging affect, tone and clout**

LIWC dimension	Percentage in female statuses	Percentage in male statuses	LIWC reference frequency
Affect: positive emotions	5.2	6.5	4.57
Affect: negative emotions	1.7	1.3	2.10
Emotional tone	86.1	83.7	63.35
Clout	79.7	87.7	55.45

Table 3 illustrates the affect, tone and clout in Facebook statuses written by female and male authors. It shows that positive emotions in both gender groups' statuses are higher than LIWC reference frequency while the negative emotions are less, which may highlight the positive stance of authors towards events, things and people, and their evident positive emotions when communicating with others through statuses. That said, it is important to notice that men were found using more words of positive emotions and less words of negative emotions than those of women. Women used words loaded with positive emotions with a percentage of 5.2% in comparison of the reference 4.57%, whereas men used them with a frequency of 6.5% which adds up almost 50% of the reference percentage. Likewise, negative emotions in men's statuses are about the half of the reference percentage (i.e. 1.3% versus 2.10%), but it was 1.7% in women's statuses. Analysis has emphasized what [64, 65] previously concluded about the value of tracing emotion-loaded words: it could capture information about authors' stance and evaluation.

Emotional tone in male and female authored statuses was higher than the reference frequency, with 86.1% for women's statuses and 83.7% for men's statuses with reference to LIWC's 63.35%. Frequencies show that female authors tone is more positive and 'upbeat' than male authors, a thing that might be related to Jordanian women's general nature as being described more emotional than men. Along the line of thought of [66], this female versus male authors' tone difference can help identify individual distinctiveness in authorship which is impacted by the author's gender.

Concerning clout, men's statuses seemed to reflect more clout than women's statuses did. This could be ascribed to the social status and leadership role that men enjoy in the Jordanian society where male language users would have recourse to writing style that is authoritative and confident. Nevertheless, both men and women tended to reflect clout far more than the reference percentage, with 87.7%, 79.7% and 55.45% for men, women and reference percentage respectively.

#### 4.3 The characteristic themes (and identities)

Field analysis of LIs resulted in the (sub) themes categorized as tabulated in Table 4 below which demonstrates the gender characteristic themes in Facebook statuses.

**Table 4. Gender characteristic themes in Facebook status**

Groups	Themes	Sub-themes	% of theme per 200 statuses
Female corpus	Socio-economical	Family/friendship occasions; Events in society; Science and arts training; Satisfaction with life standard ; Spending holidays and weekends	76.5%
	Geopolitical	Actions in areas of interest; Consequences of actions	6%
	Religious	Religious occasions; Supplications	17.5%
Male corpus	Socio-economical	Sport; Events in society;	63%

		Family/ friendship occasions; Science and arts training; Satisfaction with life standard	
	Geopolitical	Actions in areas of interest; Consequences of actions	24.5%
	Religious	Religious occasions; Supplications	12.5%

Table 4 demonstrates the major gender characteristic themes (and sub-themes) in Facebook statuses. The themes and sub-themes are as follows:

1. Socio-economical theme which comprises the following sub-themes of social and economical field, with examples of words related to the field(s):
  - a. Family/friendship relations, by using such words used by female writers as hug, friend, sister, love, cake, love, kid, marry, thank and family and such words used by men as cousin, journey and good day.
  - b. Events in society, using words (by female writers) for example, accident, life, flood, electricity, lighting and graduates and other words by male authors including independence day, trial, weather, crime, drug, suicide, orphan and love story.
  - c. Science and arts training, using female author words like multimedia, techno and university, and male author words such as processor, physics, medicine, drug, surgical, Folic acid, Laptop, Windows, Zinc and gastroenterology.
  - d. Satisfaction with life standards, for which women used words such as tired, grateful, laughter and happy, and men used words like lovely, happy, wonderful, amazing, approve, appreciate, suggest, blame and challenge.
  - e. Spending holidays and weekends, for which women mentioned related words like musical, trip, holiday, Aqaba, Dead Sea, painting, rainbow, cinema, concert and memory.
  - f. Sport, using related words to the field of sports such as Barcelona, game, football, final and fan.
2. Geo-political theme comprising the following sub-themes:
  - a. Actions in areas of interest, which included the following words among female authored statuses (Egypt, Germany, Canada, Gaza, Libya, Istanbul and Kuwait) and far more words among male authored statuses (e.g. Syria, USA, Somalia, Palestine, Nazareth, Pittsburg, Iraq, Haiti, Germany, Jordan, Dubai, West Bank, Algeria, Saudi, Egypt, Jerusalem, Cuba, Emirate and Arabs).
  - b. Consequences of actions, as in female words (e.g. struggle, make and frustrate) and male broader spectrum of words (e.g. terrorist, catastrophic, attack, boycott, risks, confederation, council, counterbalance, destroy, election and occupation).

3. Religious theme, and its sub-themes:

- a. Religious occasions, as referred to in words like Ramadan, Islam, Easter and prayer.
- b. Supplications as in LIs quoted from Quran and the Hadith of Prophet Mohammad. That includes such words as merry and Kareem.

These themes and sub-themes were evident in the two corpuses except for the sub-theme of 'sport' which appeared only in men's statuses, and 'spending holidays and weekends' topic which appeared among women authors. However, the frequencies of writing about the major themes were not the same in the two corpuses. Female writers tended to write more about socio-economical topics and less about geopolitical topics. Similarly to women, men tended to write more about socio-economical topics, but, unlike women, men were found to write far more than women about geopolitical topics (24.5% versus 6%), implicating that statuses with more words about geo-political topics are most likely male-authored.

Matching the identified themes and their filed-related words with the general categorization of [67] identities introduced the common gender characteristic identities in Facebook statuses, as follows: ethnic identity, national identity, gendered identity, and language and dialect. Ethnic identity is Muslim and Christian Arab identity. National identity is associated with Jordan (as a nation state) of status authors who are linked with and interested in their national occasions, social relations and the wider regional events. Gendered identity appeared associated with the socially constructed notions of women (i.e. feminities) as mostly interested in social, emotional and spending holiday-topics and (unlike men) least interested in geopolitical topics. Women also showed less clout in their writing in comparison with men as discussed earlier (in 4.2) and less power and status as demonstrated in use of person pronouns (in 4.1), a thing that represents a kind of stereotypical image of women as less authoritative in a society like Jordan. Language and dialect were also obvious in that authors were Arabic speakers who used (Jordanian/Levant) Arabic, and English users. Identities rose from the employed resources (namely, LIs and pronouns) which are expressive of identities and reflective of the socio-cultural context and indicative of semantic orientation and author's background, a notion stressed by [68, 69].

## 5. SUMMARY, CONCLUSION AND RECOMMENDATIONS

Taking into account the significance of LIs in discriminating the texts of authors and interpreting socio-culturally influenced aspects of language use, this study has emphasized the value of linguistic work with focus on resources used in social media to render affect, tone, clout, themes and more, a thing accentuated by Zappavigna (2012). This study attempted



to investigate the potential distinguishing patterns between female authors and male authors of Facebook statuses based on word/LI analysis, using computerized tools. As a reported (by Facebook.com) favorite and most popular social site in Jordan, Facebook communicated texts have become a good database for linguists who can support different agencies through text analyses that are of such benefits as predicting the gender of author founded on the differences in languages.

This study has provided a number of distinctive differences that facilitate gender-predictability, which would support the study hypothesis that identified resources may help make predictions about individuals' gender based on the frequencies that are so distinctive that they could signal different gender-identities. Female status-authors, on the one hand, had a tendency to repeat the use of certain (social) LIs such as love and friends more frequently than other words and far more than any other word used by male authors. Female authors also conveyed more optimistic tone. On the other hand, male authored statuses could characterize male authors by inclination to use more third person pronouns. Male authors were also found to convey more positive emotions (affect), and to demonstrate more clout (and authoritative style). Male authors, also, dominated the use of sport (sub-theme) and geo-political theme-related words. Differences could be ascribed to gender difference.

Having said that, however, male and female authors showed writing about common major themes (and sub-themes) utilizing field related words. Themes were 'socio-economical' (pertaining to family/friendship relations, events in society, science and arts training, satisfaction with life standards, spending holidays and weekends, and sport), geo-political (associated with actions in areas of interest, and consequences of actions), and religious (linked to religious occasions, and supplications). The discussed themes and employed resources (namely, LIs and pronouns) could inform us about the identities of authors, a thing that underlines the notion that linguistic resources are expressive of identities and reflective of the socio-cultural context and indicative of semantic preference and ideological background, as stressed by Eggins (2005) and Seals (2012). The identified lexical choice and patterns could be ascribed to the author's socio-cultural context in which Jordanians Facebook users interact as well as the author's gender.

Used resources could generally mirror issues of social interest and help understand society and culture specific authors' attitudes, identities, themes, interpersonal relations and more. This validates the hypothesis that the socio-cultural context in which Facebook users of various genders live and interact has an impact on the LIs and person pronouns employed, the themes discussed, emerging identities, and affect, tone and clout conveyed by Jordanian Facebook status-authors. Differences in frequencies of results between male and female authors who belong to the same socio-cultural context stresses on that gender of Facebook users has a fundamental impact on the resources used which could be predictable, particularly in relation to distinctive differences that have appeared in this study between male status-authors and female status-authors. The computerized analysis of real statuses has proved the potential of turning them into self gender-report.

In light of the results of this study, it is recommended that more studies be conducted on Facebook statuses for specific purposes, and/or for comparison between two users of two different Facebook accounts' statuses. A corpus of particular author's statuses is also likely to reflect his/her style which might be compared with another's. A similar study with larger

corpus might produce more generalizable findings. A computerized analysis with a concordancer or collocater may help achieve a better computerized thematic analysis based on clusters of co-occurring words. Further studies may investigate distinctive parts of speech of status authors by using online taggers. A similar study with computerized analysis is also advised to focus on WhatsApp statuses and tweets on Twitter. A qualitative investigation of gender differences explored in this study would be more useful avenue for further research, because such a qualitative investigation would help gain a better understanding of the ways author's gender contribute to differences in language use, particularly when it come to thematic analysis.

## 6. REFERENCES

- [1] Tausczik, Y. and Pennebaker, J. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29 (1): 24–54.
- [2] Olsson, J. 2008. *Forensic Linguistics*, 2<sup>nd</sup> edn. London: Continuum International Publishing Group.
- [3] Argamon, S., Koppel, M., Fine, J., and Shimoni, A. 2003. Gender, genre, and writing style in formal written texts. *Text*, 23(3).
- [4] Wiebe, J., Bruce, R. and O'Hara, T. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *ACL-99*: 246–253.
- [5] Newman, M., Groom, C., Handelman, L. and Pennebaker, J. 2008. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45:211–236. DOI: 10.1080/01638530802073712
- [6] Gibbons, J. 2009. Language and law. In Wei, L., and V. Cook, (Eds.), *Contemporary Applied Linguistics: Language for the Real World*, vol 2. London and New York: Continuum International Publishing Group.
- [7] Wright, D. 2014. Stylistics versus statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails. PhD thesis. University of Leeds.
- [8] Pennebaker, J., Chung, C., Ireland, M., Gonzales, A. and Booth, R. 2007. The Development and Psychometric Properties of LIWC 2007, in *LIWC Manual 2007*. LIWC.net.
- [9] Wang, Y., Burke, M. and Kraut, R. 2013. Gender, topic, and audience response: An analysis of user-generated content on Facebook. In Mackay, W. (Ed.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* held 27 April-2 May at Paris, France. New York, NY: ACM, 31-34. <https://doi.org/10.1145/2470654.2470659>
- [10] Zappavigna, M. 2012. Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web. London: Bloomsbury.
- [11] Thetela, P. 2001. Critique discourses and ideology in newspaper reports: a discourse analysis of the South African press reports on the 1998 SADC's military intervention in Lesotho. *Discourse and Society*, 12 (3): 347-370.
- [12] Van Dijk, T. 1995. Discourse semantics and ideology. *Discourse and Society*, 6 (2): 243-298. London: Sage.

- [13] Samuel, M. 2012. The Egyptian revolution, Al-Jazeera, Twitter and Facebook- The interaction effect of new media on the Egyptian revolution. MA Thesis. Leiden University.
- [14] Preston, J., Kirkpatrick, D., Fahim, K. and Shadid, A. 2011. Movement began with outrage and a Facebook page that gave it an outlet. *New York Times*, 6: 10.
- [15] Bere, A. 2012. A comparative study of student experiences of unquitting learning via mobile devices and learner management systems at a South African university. *Proceedings of the 14<sup>th</sup> Annual Conference on World Wide Web Applications*. Cape Peninsula University of Technology, <http://www.zaw3.co.za/index.php/ZA-WWW/2012/paper/viewFile/537/160>.
- [16] Yeboah, J. and Ewur, G. 2014. The impact of WhatsApp messenger usage on students performance in tertiary institutions in Ghana. *Journal of Education and Practice*, 5(6): 157-164.
- [17] Al-Khatib, M. and Enaq, H. 2008. Language choice in mobile text messages among Jordanian university students. *SKY Journal of Linguistics*, 21:37-65.
- [18] Baron, N. 2008. *Always On: Language in An Online And Mobile World*. Oxford University Press.
- [19] Pennebaker, J. and Graybeal, A. 2001. Patterns of natural language use: Disclosure, personality, and social integration. *Current Directions in Psychological Science* 10, 90–93.
- [20] Pennebaker, J., Mehl, M. and Niederhoffer, K. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54: 547–577.
- [21] Calder, B., Malthouse, E. and Schaedel, U. 2009. An experimental study of Relationship between online engagement and advertising effectiveness. *Journal of Interactive Marketing*, 23: 321-331
- [22] Carr, C., Schrock, D. and Dauterman, P. 2012. Speech acts within Facebook status messages. *Journal of Language and Social Psychology*, 31(2): 176-196. <https://doi.org/10.1177/0261927X12438535>.
- [23] Ilyas, S., and Khushi, Q. 2012. Facebook status updates: A speech act analysis. *Academic Research International*, 3(2): 500-507.
- [24] Lin, H. and Qiu, L. 2013. Two sites, two voices: Linguistic differences between Facebook status updates and tweets. In Rau, P. (Eds.), *Proceedings of 5<sup>th</sup> International Conference, CCD 2013 Cross-Cultural Design: Cultural Differences in Everyday Life*, held 21-26 July at Las Vegas, NV, USA.
- [25] Parkins, R. 2012. Gender and emotional expressiveness: An analysis of prosodic features in emotional expression. *Pragmatics and Intercultural Communication*, 5(1), 46-54.
- [26] Jurafsky, D. and Martin, J. 2018. *Speech and Language Processing (3<sup>rd</sup> edn)*. Stanford University.
- [27] Denti, L., Barbopoulos, I., Nilsson, I., Holmberg, L., Thulin, M., Wendeblad, M., Andén, L. and Davidsson, E. 2012. *Sweden's Largest Facebook Study*. Göteborg: Gothenburg Research Institute.
- [28] Winter, D. and McClelland, D. 1978. Thematic analysis: An empirically derived measure of the effects of liberal arts education. *Journal of Educational Psychology*, 70: 8-16.
- [29] Miller, G. 1995. *The Science of Words*. New York: Scientific American Library.
- [30] Kacewicz, E., Pennebaker, J., Davis, M., Jeon, M. and Graesser, A. 2013. Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33, 125143.
- [31] Fowler, R. 1986. *Linguistic Criticism*. Oxford: Oxford University Press.
- [32] Eggins, S. 2005. *An Introduction to Systemic Functional Linguistics (2<sup>nd</sup> Ed)*. New York and London: Continuum.
- [33] Seals, C. 2012. When a 'non-issue' becomes an issue in discourse surrounding LGBT communities. *Journal of Language and Sexuality*, 1 (2): 230-255.
- [34] Omoniyi, T. and White, G. 2006. *Sociolinguistics of Identity*. London and New York: Continuum.
- [35] Omoniyi, T. 2006. Hierarchy of identities. In Omoniyi, T. and G. White (Eds), *Sociolinguistics of Identity*. London and New York: Continuum.
- [36] Tabouret-keller, A. 1997. Language and identity. In F. Coulmas (Eds). *The handbook of Sociolinguistics*. Oxford: Blackwell, 315-26.
- [37] Block, D. 2006. Identity in applied linguistics. In Omoniyi, T. and G. White, (Eds), *Sociolinguistics of Identity*. London and New York: Continuum.
- [38] Wardhaugh, R. 2002. *An Introduction to Sociolinguistics (4<sup>th</sup> ed.)*. Oxford: Blackwell Publishers.
- [39] Milroy, J. and Milroy, L. 1978. Belfast: Change and variation in an urban vernacular. In P. Trudgill (Eds). *Sociolinguistic Patterns in British English*, London: Arnold.
- [40] Scherer, K. 2000. Psychological models of emotion. In Borod, J. (Eds.), *The Neuropsychology of Emotion*, 137–162. Oxford.