A Pilot Study on the Collection and Computational Analysis of Linguistic Differences Amongst Men and Women in a Kuwaiti Arabic WhatsApp Dataset

Hesah Aldihan[1,2] and **Robert Gaizauskas**[1] and **Susan Fitzmaurice**[1]
[1] University of Sheffield, UK
[2]University of Kuwait, Kuwait
{haldihan1,r.gaizauskas,s.fitzmaurice}@sheffield.ac.uk

Abstract

This study focuses on the collection and computational analysis of Kuwaiti Arabic, which is considered a low resource dialect, to test different sociolinguistic hypotheses related to gendered language use. In this paper, we describe the collection and analysis of a corpus of WhatsApp Group chats with mixed gender Kuwaiti participants. This corpus, which we are making publicly available, is the first corpus of Kuwaiti Arabic conversational data. We analyse different interactional and linguistic features to get insights about features that may be indicative of gender to inform the development of a gender classification system for Kuwaiti Arabic in an upcoming study. Statistical analysis of our data shows that there is insufficient evidence to claim that there are significant differences amongst men and women with respect to number of turns, length of turns and number of emojis. However, qualitative analysis shows that men and women differ substantially in the types of emojis they use and in their use of lengthened words.

1 Introduction

A wide range of sociolinguistic gender studies have been carried out in English speaking cultures and in the Arab world too. However, there is a lack of research on Gulf Arabic (GA) dialects, and especially the Kuwaiti dialect, from a sociolinguistic perspective. The GA dialects vary tremendously with regards to morpho-phonological features, lexical structures and the effect of language borrowing from different languages (Khalifa et al., 2016). There are some interesting linguistic phenomena in the Kuwaiti dialect. The way men and women speak is different and this can be noticed in their choice of words when communicating or expressing feelings or reacting to situations. It can be noticed that there are some words which men would refrain from using because they represent femininity. For example, the word اينن "eyanen", which means "amazing" is a word used to convey a positive sentiment towards an entity and is usually only used by women. This word can for example be used to describe a movie by Kuwaiti women, whereas men might use the word جبار "jbar" which is a polysemous adjective that in this context means "amazing", to describe the movie. Moreover, يا حافظ "ya hafeth" is a phrase that is only used by women. It can be translated into "Oh saviour (God)" to convey dissatisfaction or disappointment. If a man uses this expression, he would be described as someone who is feminine in the way he speaks.

Advances in the field of Arabic Natural Language Processing (ANLP) have made it possible to study such variation in lexical usage between genders as well to explore other features that are indicative of gender. However, the lack of KA textual resources and preprocessing tools make it a challenging task.

This study contributes to the field of ANLP in two ways. First, we have compiled and made publicly available a new, gender-labelled KA dataset, which can be used by researchers interested in the Kuwaiti dialect or gender studies. This dataset consists of textual book club conversations conducted on the WhatsApp online instant messaging mobile application. To the best of our knowledge this is the first published dataset of mixed gender KA conversational data. Second, we have carried out an analysis of interactional and linguistic features that may inform the development of a gender classification system for KA.

This paper is structured as follows. In the next section we review related work. In section 3 we first discuss how we have collected the raw data, then describe how this raw data has been preprocessed to prepare the dataset for analysis and finally discuss the features that will be explored and analysed. In section 4 we present our results and analysis. Finally, we conclude and discuss future work, as well as pointing out some of the limitations of

Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP), pages 372 - 380
December 8, 2022 ©2022 Association for Computational Linguistics

our work.

2 Related Work

Language is a rich source for analysis and many studies have been conducted to infer the relationship between different social variables and the language they construct (Holmes and Meyerhoff, 2008; Eckert and McConnell-Ginet, 2013). One of the social variables that is studied in relation to language is gender. Traditional studies of language and gender that have been conducted in the humanities and social sciences have had inconsistent findings and have received some criticism. For example, Wareing (1996) criticised conclusions drawn about the relationship between language and gender that are dependent on small samples of data. The implication of this criticism is that gender and language studies should be improved by using larger samples of data and different contexts (Litosseliti and Sunderland, 2002). However, now that we are in the era of 'big data', extracting large amounts of data for gender analysis has become possible. Moreover, sociolinguistic studies of gender have mostly been explored using qualitative methods such as interviews, surveys, recordings and manual observations. Bamman et al. (2014) argue that qualitative and quantitative analysis of sociolinguistic gender studies are complementary as qualitative analysis may shed light on phenomena and quantitative analysis provides the opportunity to explore phenomena through large scale studies and also identify cases that can be analysed qualitatively. Litosseliti and Sunderland (2002) explain:

> Language and gender may, then, legitimately be viewed from different perspectives: a pragmatic combination of methods and approaches, along with an acknowledgment of their possibilities and limitations, might allow us to focus on different aspects of the relationship between language and gender, or have a wider range of things to say about this.

In the context of studies that have explored gender differences in language use, Rosenfeld et al. (2016) looked into gender differences in language usage of WhatsApp groups. They analysed over 4 million WhatsApp messages from more than 100 users to find and understand differences between different age and gender demographic groups. In analysing the data, they relied on metadata only such as message lengths, size of the WhatsApp groups, time, average number of sentences sent per day, time between messages. In relation to gender, analysing the length of messages sent by both genders showed that women send and receive more messages than men. They also concluded that women are more active in small WhatsApp groups, whereas men are more active in larger WhatsApp groups. These differences were then employed in building age and gender prediction models. They performed a 10-fold cross validation for these tasks using decision trees and a Bayesian network. For the gender prediction task, using users' metadata with decision trees achieved 70.27% accuracy and 73.87% accuracy when used with a Bayesian network.

Other studies have looked into differences amongst genders in the use of emojis. Chen et al. (2018) compiled a large dataset of 401 million smartphone messages in 58 different languages and labelled them according to the gender of users. They used emojis from the dataset to study how they are used by males and females in terms of emoji frequency, emoji preference and sentiment conveyed by the emojis. They also studied the extent in which emojis are indicative of gender when used in a gender classification system. The results obtained from this study showed that not only are there considerable differences in the use of emojis between males and females, but also that a gender classification system that uses emojis alone as features can achieve an accuracy of 81%.

Shared NLP tasks that are organized for the research community have started off by tackling problems with the English language and in recent years have added Arabic datasets, reflecting the increasing interest in Arabic NLP. For example, the PAN 2017 Author Profiling Shared Task included two tasks: gender identification and language variety identification of Twitter users. Arabic, English, Portuguese, and Spanish datasets consisting of tweets were provided for training and testing. The system that achieved the highest accuracy result on gender identification in the Arabic dataset was the system developed by Basile et al. (2017). They used an SVM classifier in combination with word unigrams and character 3- to 5-grams and achieved an accuracy of 0.80.

As for studies that have targeted the Arabic language, Alsmearat et al. (2014) studied gender text classification of Arabic articles using the Bag-of-

Words (BoW) approach. They collected and manually labelled 500 Arabic articles from different Arabic news websites. The number of articles was distributed equally across both genders. They wanted to explore the result of performing feature reduction techniques such as PCA and correlation analysis on the high-dimensional data in combination with different machine learning algorithms for the gender classification task. Results showed that Stochastic Gradient Descent (SGD), Naive Bayes Multinomial (NBM) and Support Vector Machines (SVM) were the classifiers that performed best on the original dataset where the accuracy results surpassed 90%.

Furthermore, Mubarak et al. (2022) compiled a dataset of 166K Arabic tweets and labelled them with gender and geo location labels. They used this dataset for gender analysis and to build a gender classification system using SVMs that was tested on different features such as usernames of the twitter users, the profile pictures of the users, tweets and gender distribution of users' friends. Their study showed that using usernames alone as features for gender prediction achieved the highest F1 score of 82.1 %. In addition, Hussein et al. (2019) attempted to build a gender classification system for Egyptian Arabic. They created a dataset of 140K tweets that were retrieved from famous Egyptian influencers and active Egyptian users of Twitter. They labelled the dataset according to the gender of the Twitter users by referring to the users' profile image and names. They experimented with different features such as gender discriminative emojis, female suffixes, manually created dictionaries of swear words, emotion words, political words, flirting words, technological words and word embeddings. They used ensemble weighted average on a mixed feature vector fed into a Random Forest classifier and an N-gram feature vector fed into a Logistic Regression classifier. They achieved an accuracy score of 87.6%.

Not many gender studies in NLP have provided much insight into linguistic characteristics of gendered language, especially those related to dialectal Arabic. Furthermore, the field of ANLP still lacks enough dialectal arabic datasets to help inform the development of Arabic natural language processing tools. Khalifa et al. (2016) compiled Gumar corpus which consists of 100 million GA words from 1200 forum novels annotated according to the dialect, novel name and writer name. The corpus was also used to develop dialectal Arabic orthography. However, although Gumar corpus contains some KA text, the text is not naturally occuring conversational KA. Therefore, there is still a need to compile conversational KA resources. We aim to address this gap by contributing towards providing resources for the KA dialect and analysing sociolinguistic features of that dialect that can be used to inform NLP applications, such as gender classification systems.

3 Methodology

3.1 Data Collection

Since we are interested in studying the features of conversational data of Kuwaiti men and women, we chose to collect textual data from WhatsApp reading club groups.

As part of the data collection process, we applied for ethical approval before conducting the study. This involved ensuring that all participants were aware of the nature and purpose of the study and their role in it. We obtained informed consent from all participants.

The dataset was collected from three Kuwaiti reading club WhatsApp groups. These were already existing WhatsApp reading club groups that have been running for years and are managed by Kuwaiti admins. All participants were native Kuwaiti speakers whose first language is KA. The researcher was added to the groups to be able to export the chat after 9 months of being added. The chats were then exported from the mobile phone and saved in the researcher's computer for processing.

The dataset consists of 4479 turns (2623 turns by females and 1856 turns by males). The dataset will be made publicly available for researchers in the research field.[1]

3.2 Preprocessing

A number of steps were taken prior to exporting the chats from the researcher's mobile. This involved anonymising the names of the WhatsApp members. The usernames were replaced with the word "USER" concatenated with a number and a letter to represent the gender of the user (e.g, USER1F). The chats were then exported to the researcher's computer to prepare the data for computational pro-

[1] Interested parties can contact the first author for dataset access.

Gender		Emoji Count	Word Count	Num of Turns
Women (28 participants)	Total Number	2144	17388	2623
	Mean	76	621	94
	Median	23	163	29
	Std. Deviation	123	1132	144
	Minimum	2	6	2
	Maximum	506	5611	655
Men (14 participants)	Total Number	801	14005	1856
	Mean	57	1000	133
	Median	36	432	102
	Std. Deviation	68	1197	134
	Minimum	1	5	3
	Maximum	249	3941	444

Table 1: Descriptive Statistics of the Features Analysed

cessing. The following preprocessing steps were performed:

1. All sensitive and personal information was removed.
2. Real names that were mentioned in the chat were replaced with fictitious names.
3. URL links were removed.
4. Two versions of the dataset were created using the CAMel tools, built by Obeid et al. (2020), for preprocessing: one that involves tokenisation, removal of digits, diacritics and punctuation and changing alef variants to ا and alef maksura to ي and teh marbuta to ه; and another version that involves tokenisation and punctuatation removal. Depending on the type of textual analysis required, the dataset version was chosen.

3.3 Feature Analysis

We were interested in exploring interactional features and lexical features pertaining to the KA dialect. We chose to study how the following features were used amongst men and women participating in the study:

- Number of turns per gender.
- Length of turns per gender (word count).
- Use of emojis amongst females and males, especially in the context of the view that certain emjois are considered too feminine and others too masculine in the Kuwaiti society.
- Whether there are KA words or expressions that are exclusive to each gender.
- Most frequently used words.
- Lengthened or elongated words.

Table 1. presents the descriptive statistics of the first three features.

4 Results and Analysis

To analyse the results of this study, two approaches were taken: a quantitative statistical approach and a qualitative linguistic approach. As for the statistical approach, the Mann Whitney U test was used for analysis due to it being suitable for data, like ours, which is not normally distributed. It was done using SPSS [2]. One limitation of using a statistical approach in analysing the data is that it does not take into account the contextual information and meanings embedded within the text. Therefore, it was important to perform an in-depth manual analysis of the data to be able to describe the patterns found and provide interpretations for points that the statistical analysis could not capture.

4.1 Quantitative Analysis

We tested the distribution of each feature using normality tests, namely Shapiro-Wilk test (sample size less than 50) which indicated that the features were not normally distributed P values: (< 0.01). The Mann Whitney U test was used to test if there are significant differences between men and women

[2] Statistical Package for the Social Sciences: a statistical analysis software package. https://www.ibm.com/products/spss-statistics

with regards to three features: number of emojis used in the chat, number of turns taken, and total number of words (word count) for each user. This test is based on two hypotheses; a null hypothesis (H_0):

> H_0: states that there is no significant difference between men and women with regards to the features mentioned above.

and an alternative hypothesis (H_1):

> H_1 : states that there is a significant difference between men and women with regards to the features tested.

The hypotheses are accepted or rejected after comparing the P values to the threshold (0.05).

As can be seen in Table.2, all the P - values for all the features are larger than 0.05. This means that we lack enough evidence to suggest that there are significant differences between men and women in terms of number of emojis used, number of turns taken and word count.

In the following subsections, we look into the analysis of each feature in detail.

4.1.1 Number of Turns

We were interested in analysing the number of turns used by each user and gender. We were also interested in computing the percentage of turns for men and women from the total number of turns. We noticed that 59% of the total number of turns were by women, and the remaining 41% of turns were by men. However, the ratio of women to men in the corpus is 2:1 and based on the results we obtained from Mann Whitney U test: (women: median= 29, IQR = 105), (men: median= 102, IQR = 198), P - value > 0.05 as shown in Table 1, we lack enough evidence to suggest that there is a significant difference amongst men and women in terms of number of turns.

4.1.2 Length of Turns/ Word Count

The length of turns was computed to test the hypothesis that women speak more than men. This was done by counting the total number of words used in the chats for each user and the total word counts for each gender. Details are shown in Table 1.

On average, men speak more than woman (1000 words per male participant vs 621 words per female participant). However, Mann Whitney U test results for word counts (women: median= 163, IQR = 582), (men: median= 432, IQR = 1778), P - value > 0.05 as shown in Table 1, suggest that we lack enough evidence to claim that there is a significant difference amongst men and women in word usage.

We were also interested in comparing the average number of words per turn for women as compared with men. Referring to Table 1 we can see that for women the average number of words per turn is 17388/2623 = 6.62 while for men the average words per turn is 14005/1856 = 7.55. The difference here does not appear to be that great, but we have not carried out statistical analysis to see if that difference is significant.

4.1.3 Emoji Usage

We were interested in analysing how likely it is for men and women to use emojis when interacting in the chat groups. We noticed that on average women used .82 emojis per turn, while men used on average .43 emojis per turn. Therefore, the odds of using emojis amongst women compared to men is 1.9:1, indicating that women were almost 2 times more likely to use emojis than men. However, based on the results we retrieved from Mann Whitney U test: (women: median= 23, IQR = 84), (men: median= 36, IQR = 95), P - value > 0.05 as shown in Table 1, we lack enough evidence to suggest that there is a significant difference amongst men and women in emoji usage.

Nonetheless, it was important to explore the types of emojis, exclusivity of emojis and patterns of emojis used by men and women to achieve a better understanding of emoji usage amongst genders. This is discussed in the following section.

4.2 Qualitative Analysis

4.2.1 Frequency and Types of Emojis

Emojis were significant features observed in the group chats and were commonly used by both men and women. Women used a total of 2144 emojis, while men used a total of 801 emojis. As for the types of emojis used, various differences were observed. Emojis used by women are from a wide range of emoji categories and are colorful, whereas men used a limited set of emojis from certain categories. 68% of women used heart emojis, whereas only 29% of men used heart emojis. It was also noticed that women used different types and colors of heart emojis. However, men used limited heart emojis 💙, 💔, 💕. Further more, women used a large variety of flowers and plants 🌷, 🌺, 🌻, 💐,

Features	P Value	U Value	Median of Females	Median of Males
Num of Emojis	0.779	185.500	23.00	35.50
Num of Turns	0.298	157.00	29.00	101.50
Word Count	0.350	161.00	163.00	431.50

Table 2: Mann-Whitney Test Results for Emojis, Number of Turns and Word Count Features

	Women		Men	
Rank	Emoji	Count	Emoji	Count
1	👏🏻	218	👍🏻	156
2	🙏🏻	211	😂	143
3	😍	193	🌹	43
4	👍🏻	140	🤣	42
5	🌸	116	😅	28
6	💜	95	😄	26
7	😂	91	💕	20
8	🌹	85	😜	18
9	🌺	75	😊	17
10	💐	75	😍	15

Table 3: Top Ten Emojis Used by Kuwaiti Men and Women

🌸, 🌹, 🌱, 🌿, whereas men used only two types of flowers 💐 and 🌹.

The analysis also involved computing the 10 most frequently used emojis by men and women as shown in Table 3. As it can be seen, the top used emojis for both men and women are (👏🏻 and 👍🏻) which shows that both men and women are encouraging and applauding each other. It was observed that men used (😂 and 👍🏻) significantly more than all the other emojis extracted, which were mainly smileys. In comparing the top 10 lists of emojis by men and women, it was noticed that women used 😍 (193 times) notably higher than men (15 times) and used flowers more than smileys as opposed to men.

4.2.2 Exclusivity of Emojis

There are some stereotypes regarding emoji usage such as that there are certain emojis that are not used by men due to them implying a feminine sense and other emojis not used by women because they are masculine. This study examined this stereotype to explore if this can be considered a feature indicative of gender. The emojis that were exclusively used by each gender were extracted and compared. It was noticed that men refrained from using certain emojis that are stereo-typically considered feminine and were used by women in the group chats such as 💋, 😚, 😻, 🤗, 💖, 💗, 💞, 💕, 💓, 🦋. This observation also supports the hypothesis that women are more emotionally expressive than men (Goldshmidt and Weller, 2000). The emojis that were exclusively used by men mainly consisted of male character emojis such as 🤦🏾‍♂️, 🤷🏻‍♂️, 🙇🏻‍♂️, 🕺🏽, 🏃🏻.

4.2.3 Patterns of Emoji Usage

A number of observations were made related to patterns of emoji usage. Women used a larger variety of emojis across different categories (smileys and people, activity, travel and places, food and drink , nature .. etc) than men to express themselves. Men used limited types of emojis from certain categories (smileys and people, nature) and very limited use of hearts or emojis that express emotions.

A pattern was also noticed regarding the number of emojis used per turn. Most users used one or two emojis in a turn and this lead to interest in analysing bigrams of emojis used by men and women to explore if there are any patterns of use or certain emoji combinations used. The most frequently used bigrams consisted of the same emoji repeated rather than a combination of two different emojis. It was observed that certain combinations were used significantly more by each gender. For example, 😂 😂 was used 70 times by men and 38 times by women, 😍 😍 was used 3 times by men and 64 times by women, and 👏🏻 👏🏻 was used 4 times by men and 80 times by women. This showed certain emoji combinations may be used with different frequencies amongst men and women.

4.2.4 KA Lexical choices and Features

Other exploratory data analysis was conducted to analyse the lexical choices amongst men and women in the WhatsApp groups. Features such as the most frequently used words, the exclusively used words and other lexical features were analysed.

Analysis regarding the most frequently used words showed that the word "Allah", "الله" was one of the highly repeated words amongst both men (262 times) and women (325 times). "Allah" means "God" and could appear in a sentence as a separate word or part of a phrase such as "masha'Allah",ماشاءالله which is an expression used to express appreciation when someone hears good news, and "inshaAllah", ان شاءالله which is an expression used to convey willingness to do something. The high repetition of these phrases could indicate cooperativeness and politeness in the conversations. The word "alketab" الكتاب which means "book" was also amongst the highest repeated words amongst men (32 times) and women (99 times). This is due to the conversations mainly revolving around reading books. Figure 1 and Figure 2 show the most frequent words in both the women's and men's chats.

Analysis was also done on the exclusively used words amongst both men and women. One aim of extracting the gender exclusive words was to find KA gendered words that denote femininity or masculinity to inform the development of a gender classification system. However, due to the formal nature of the reading club WhatsApp groups, only a few examples of this phenomenon were captured and they were mostly in women's messages. Some of the examples of female exclusive words found are: "shatoora" شطورة , meaning "smart girl", "b'khatri" بخاطري meaning "I really want ..", "habeebty" حبيبتي , meaning "my dear", "s'ghairoona" صغيروونه, meaning "very small", "katkoota" كتكووته , meaning "so cute" and "please" بليز.

Analysis of the chat also showed high occurrence of lengthened or elongated words which are words that include repeated letters to emphasise different meanings such as هههههههه "hhhhhhhhh" expressing laughter and واااااو "wooooow" expressing amazement. Lengthened words can be indicators of expressing feelings which is stereotypically attached to women's speech, and therefore we wanted to test this hypothesis by determining the number of lengthened words used by men and women per turn on average. There were some interesting observations. Women used 0.057 lengthened words per turn on average (so about once per 18 turns), whereas men used 0.037 (about once in 28 turns). This indicates that women tend to lengthen words roughly 1.5 times as often as men. After performing further inspection to the lengthened words, it was observed that women tend to perform this with a large variety of words when laughing هههههه "hhhhhh" , complimenting جمممميله"beautifuuuul", congratulating مبروووك "congraaatulations" , encouraging برااڤووووو "bravooooo" , agreeing ايي "yeees" , greeting صباح النوور "good mooorning" and expressing feelings such as missing the members مشتاقيين "miiis you". However, men's use of lengthened words were less diverse. They mostly used lengthening when laughing ههههههههههه "hhhhhhhhhhh" and greeting هلاا "hiii".

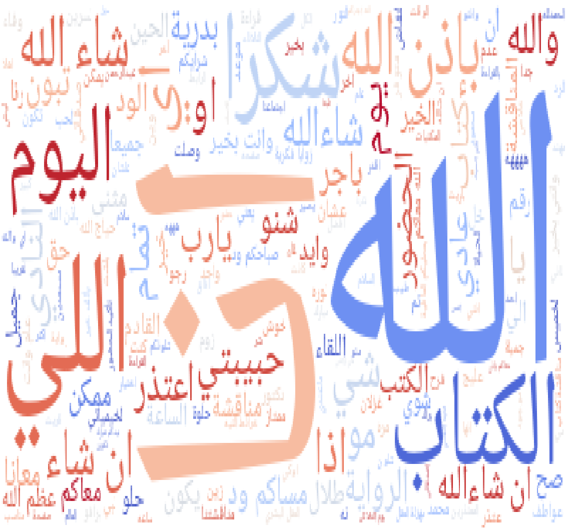

Figure 1: Most Frequent Words Used by Women

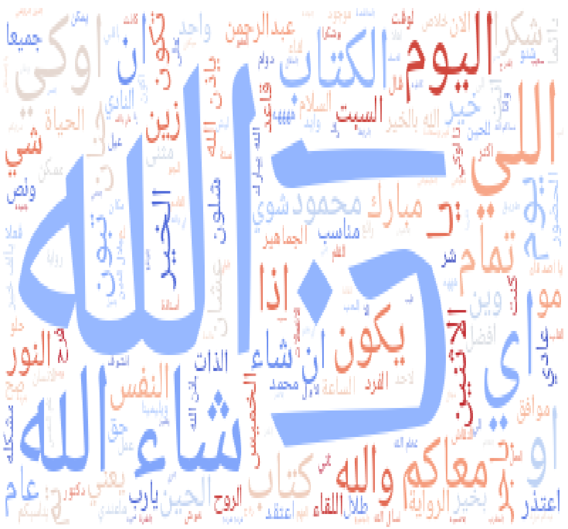

Figure 2: Most Frequent Words Used by Men

5 Conclusion

We have described the first publicly available dataset of conversational Kuwaiti Arabic that is labelled by gender. We analysed the dataset by looking into interactional and linguistic features that are performed in mixed gender WhatsApp groups. We described the WhatsApp data collection process and analysed features such as number of turns, length of turns, emoji counts and Kuwaiti Arabic lexical features. Statistical analysis shows that our dataset does not allow us to conclude that significant differences between men's and women's language exist with respect to the features number of turns, length of turns, and number of emojis used. However, substantial differences in these features are observed. Furthermore, qualitative analysis of other features such as the range and specific types of emojis used, certain lexical choices and the phenomenon of word lengthening revealed considerable differences between women and men's language use.

Going forward we intend to build a gender classification system for Kuwaiti Arabic trained and tested on the dataset reported here. We intend to use insights gained in the study reported here to inform our feature selection, with the longer term aim of better understanding differences in men and women's language use in Kuwaiti Arabic.

Limitations

Our study is limited in several ways. The first relates to the dataset as a basis for studying differences in men and women's language differences in conversational KA. The compiled dataset is of limited size and unbalanced in gender labels. Since we wanted to study KA conversational data, it was only possible to get ethical approval for formal WhatsApp groups. This had an impact on both size and type of data collected. The size of data was subject to participants' level of interaction in the WhatsApp groups. Furthermore, the type of conversational data collected tends to have a formal tone due to the groups conversation revolving around discussing books. This means there may be a lack of certain sociolinguistic phenomena being present in the conversations. Moreover, the language usage of participants who are book club readers may not be representative of the KA dialect more generally. The second sort of limitations pertain to the restricted amount of analysis carried out as yet on our dataset. To date we have not built a gender classification system using this dataset to see, for example, how well word or emoji unigrams or bigrams might serve as a basis for predicting gender. As noted above in section 5, this is next on our agenda.

Ethics Statement

To gather the data we submitted an application to the University of Sheffield Ethics Review process and had this application approved. Participants were provided with an information sheet describing the aims and objectives of our research, what they would be expected to do, what data we would collect, how that data would be used and how it would be stored. We then obtained informed consent from each participant for our proposed work. Regarding potential use of our work we see both potential benefits and potential harms. On the benefits side, better understanding of the differences in language use between genders may help us identify and better understand the causes of these differences. Insights from this could lead to change in perception of gender roles and positive change in gender equality. On the negative side, ability to predict gender from language use could lead to targeting of individuals in various ways including advertising, political messaging or even persecution for expressing certain beliefs.

Acknowledgements

The authors would like to thank the reading club group members that participated in the study. We would also like to thank the statisticians who helped out with the analysis and the University of Kuwait for funding this work.

References

Kholoud Alsmearat, Mahmoud Al-Ayyoub, and Riyad Al-Shalabi. 2014. An extensive study of the bag-of-words approach for gender identification of arabic articles. In *2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA)*, pages 601–608. IEEE.

David Bamman, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160.

Angelo Basile, Gareth Dwyer, Maria Medvedeva, Josine Rawee, Hessel Haagsma, and Malvina Nissim. 2017. N-gram: New groningen author-profiling model. *arXiv preprint arXiv:1707.03764*.

Zhenpeng Chen, Xuan Lu, Wei Ai, Huoran Li, Qiaozhu Mei, and Xuanzhe Liu. 2018. Through a gender lens: Learning usage patterns of emojis from large-scale android users. In *Proceedings of the 2018 World Wide Web Conference*, pages 763–772.

Penelope Eckert and Sally McConnell-Ginet. 2013. *Language and gender*. Cambridge University Press.

Orly Turgeman Goldshmidt and Leonard Weller. 2000. talking emotions: Gender differences in a variety of conversational contexts. *Symbolic Interaction*, 23(2):117–134.

Janet Holmes and Miriam Meyerhoff. 2008. *The handbook of language and gender*, volume 25. John Wiley & Sons.

Shereen Hussein, Mona Farouk, and ElSayed Hemayed. 2019. Gender identification of egyptian dialect in twitter. *Egyptian Informatics Journal*, 20(2):109–116.

Salam Khalifa, Nizar Habash, Dana Abdulrahim, and Sara Hassan. 2016. A large scale corpus of gulf arabic. *arXiv preprint arXiv:1609.02960*.

Lia Litosseliti and Jane Sunderland. 2002. *Gender identity and discourse analysis*, volume 2. John Benjamins Publishing.

Hamdy Mubarak, Shammur Absar Chowdhury, and Firoj Alam. 2022. Arabgend: Gender analysis and inference on arabic twitter. *arXiv preprint arXiv:2203.00271*.

Ossama Obeid, Nasser Zalmout, Salam Khalifa, Dima Taji, Mai Oudah, Bashar Alhafni, Go Inoue, Fadhl Eryani, Alexander Erdmann, and Nizar Habash. 2020. CAMeL tools: An open source python toolkit for Arabic natural language processing. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 7022–7032, Marseille, France. European Language Resources Association.

Avi Rosenfeld, Sigal Sina, David Sarne, Or Avidov, and Sarit Kraus. 2016. Whatsapp usage patterns and prediction models. In *ICWSM/IUSSP Workshop on Social Media and Demographic Research*.

Shan Wareing. 1996. What do we know about language and gender. In *eleventh sociolinguistic symposium, Cardiff, September*, pages 5–7.