



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Automated Detection of Gendered Linguistic Variation in Arab EFL Social Media Discourse Using Machine Learning and NLP

Hind Khalifa Alaodini¹, Arif Ahmed Mohammed Hassan Al-Ahdal²

¹Department of Languages and translation, University College of Duba, University of Tabuk, Saudi Arabia
haladini@ut.edu.sa

²Department of English Language and Literature, College of Languages and Humanities, Qassim University, Buraydah, Saudi Arabia
aa.alahdal@qu.edu.sa

Abstract

This mixed method study examines gendered differences in the language used by Arabic speakers during online communication, a little researched phenomenon in the Arabic speakers' paradigm compared to the western context. Applying the sociolinguistic model, the study tests the hypothesis that women use "standard/prestige" language forms as compared to the "vernacular/innovation" form used by men. This study analyzes three registers (standard, vernacular, innovative) in code-switched English and how these are created by Arab EFL users during communication on social networks like X/Twitter, Instagram, Snapchat, and TikTok in the context of the performative hybrid register of social media. A corpus equally representative of men and women users comprised the dataset for this study followed by coding for phonological, lexical, morphological, code-switching variants, and paralinguistic features such as emojis. Quantitative data was analyzed using Chi-square test, correlation analysis, and regression analysis to compare the distribution of variants by gender, age, education, and platform. Discourse analysis was applied to qualitative data to identify and examine the interpretive role of linguistic variants, such as humor, emphasis, and cosmopolitan self-presentation. The study results were illustrative with clear evidence of systematic gender differences in online language use, strongly intertwined with social and technological factors, implying that gender needs to be modeled as part of a larger "gestalt" of social and technological factors. The study adds to the field of sociolinguistic theory in the context of spoken Arabic dialects and their extension into digital EFL discourse.

Keywords: Gender, sociolinguistic variation, Arabic dialects, EFL, social media, digital discourse, code-switching.

Introduction

Gender is widely recognized as a major social factor in research on language variation and change. Although gender was rarely considered in early sociolinguistic studies, it is now among the most important factors shaping patterns of linguistic behavior (Cheshire, 2002). Gender influences speakers' linguistic choices, contributes to the maintenance of language and to language change, and plays an ongoing role in understanding language variation across all forms of communication whether spoken, written, or digital. Labov (2001) has proposed three key principles linking gender and linguistic variation: If sociolinguistic stratification is stable, men would use nonstandard forms more often than women according to Principle I. Moreover, women also prefer incoming prestige forms more than men, as stated in Principle Ia. According to Principle II, women are the most frequent agents of language change in bottom-up transformations. These principles have guided sociolinguistic writing over the past few decades and will continue to serve as a point of reference for examining new observations, including those from Arabic-speaking communities.

The same principles apply to Western sociolinguistics which suggests that women are more likely than men to adopt prestigious and innovative linguistic styles (Milroy & Milroy, 1997). However, these patterns have not

been universal, and there is greater variation in real-life speech among different communities of speakers and between spoken and text-based discourse within the same cultural context, which becomes particularly visible when the object of study is not only face-to-face speech but also digital language. Studies in Arabic dialects have highlighted more complex patterns than those from the classical Western model. Many such studies have demonstrated that older women tend to take a conservative approach in communication, maintaining older variants from subtle local usage, while men are more likely to be innovative. For instance, the traditional dialects of older females in Souf (Jordan) are more likely to be retained than those of older males (Al-Hawamdeh, 2016). Al-Ammar (2017) too, found similar developments in the Ha'il dialect of Saudi Arabia where variation in feminine suffixes was carried out in a conservative manner by older women. However, the innovative and socially prestigious variants were found to be the characteristic linguistic variants of male speakers among the Dawāsir tribe in Dammam (Alaodini, 2019a, 2019b). Women were found innovative in using the depalatalisation of /g^j/ and /k^j/ in the speech of the Ḥarbi tribe in Medina. Results indicate that the patterns of linguistic variation in Arabic-speaking communities do not necessarily follow those assumed in the study of universal patterns proposed by western Sociolinguistics (Mohammed & Samad, 2020). In the case of the innovative realization of [a:] and the innovative variant of [dʒ], the former was favored by males, and the latter was disfavored by females (Alaodini, 2019a). This inverts the traditional order, which is one of the more interesting contributions of Arabic sociolinguistics to the general theorizing on gender and variation in language. Instead, it overlaps with other factors, such as level of education, social network structure, and age.

It has been shown repeatedly that age and gender are factors that significantly affect the manifestation of language behaviors. Gender patterns are strongly determined by age, according to Dubois and Horvath (2000), and the study of linguistic behavior is not possible without considering the influence of these two factors. However, the general trend identified by various sociolinguistic studies is that younger speakers are more susceptible to innovative forms of language while older speakers stick to the conventional forms. For instance, Edwards (1992) conducted a sociolinguistic study of how different communities of African Americans in inner cities used their language and concluded that younger persons used innovative linguistic structures while older persons persisted with the traditional forms of African American vernacular. This apparent pattern is generally seen as an index of change in language and provides a useful analytical approach to understand the creation and dynamics of new digital registers which may experience different rates of addition of new forms to language by successive generations of users.

Digital platforms enable speakers to build, negotiate, and enact social identities through the strategic use of linguistic features such as standard forms, vernacular features, and foreign-language material. In the Saudi context, however, this idea has gained recent scholarly interest. Al-Rohili and Alaodini (2026) examined the use of social media as a branding strategy that draws on authentic tribal dialects in the Tabuk region. Their research found that businesses and content producers deliberately used conventional tribal linguistic elements to project authenticity, strengthen the sense of tribal identity, and positively influence their audience's attitudes. These results highlight the dialect's symbolic role in identity and consumer interaction in an online context and raise questions about the naturalness of online language (which is always about identity, credibility, and belonging). At the same time, research has increasingly explored the use of English as an asset for stylistic and communicative expression, rather than merely as a marker of competence, in local digital and media subcultures. In German social and broadcast settings, for example, foreign word use is sometimes applied selectively and creatively in local discourse for stylistic, humorous, or branding purposes (Fiedler, 2022; Schaefer, 2019, 2021, 2025; Al-Ahdal & Aljafen, 2023; Aljabr et al., 2023). Similarly, Gerritsen et al. (2007) found that English is strategically used in advertising across different markets in Europe to connote modernity and internationalism, a pattern that could reasonably be extended to the self-presentational discourse produced by Arab students of English as a foreign language on social media.

Additionally, Greene and Schmid (2024) showed that the internet-defamiliarizing culture has created its own identifiable linguistic phrases and trends, establishing that online genres are emerging as cultural participants, alongside spoken ones. This shift toward a digital world has gained momentum in English as a Foreign Language (EFL) where social media serves as a site for language acquisition, performance, and interaction with others. During COVID-19, when in-person teaching was disrupted, Ravindran et al. (2022) reported a significant supportive role of social media in developing speaking skills. Alhadidi et al. (2025) investigated the language preferences of EFL bilinguals on social media and found that platform choice and audience orientation strongly influenced whether EFL users preferred English or Arabic, or a blend of both, thereby shifting the nature of the social media user experience. Zhang et al. (2026) demonstrated in a review study that social media interaction is significantly involved in the development of intercultural competence in

digital communication which is directly related to code-switching and register shifting as observed in bilingual and EFL populations.

Hu et al. (2025) showed through multidimensional and entropy-based analyses, that the communicative purpose of a register (such as the academic or entertainment register) systematically shapes its linguistic level, thereby reinforcing the likelihood that social media discourse (which is clearly performative and informal) will pattern differently from more formally produced English in an EFL context. This further contextualizes the present study within educational and pedagogical research on EFL learners' use of digital and critical literacies. Therefore, the scaffolding offered by language educators when presenting authentic digital materials plays a significant role in building the critical perspectives of their EFL learners, as briefly outlined by Hamdi (2023). As Gudmestad and Kanwit (2025) argued, there is a need for a new variationist approach to SLA that adequately considers social contexts such as online interactions. Moreover, Yelenevskaya (2026) illustrated the pedagogical potential of the linguistic landscape approach—both the traditional, offline, and more established, and the new, digital variety—in L2GAP language classrooms, and López-Serrano et al. (2026) showed that motivation and engagement in web-based educational environments are closely linked to pragmatic development in L2 which is clearly visible in the socially situated, audience-oriented language of social media posts.

Sumon et al. (2024) noted the emergence of new technologies such as Natural Language Processing (NLP) and their impact on digital socialization and linguistic interaction, indicating that the technological environment in which Arab EFL users create social media discourse continues to evolve and to shape the linguistic options available to them. Although recent studies have yielded a considerable number of findings, very few have directly investigated the gendered linguistic variation among Arabic-speaking EFL learners in social media contexts. Research on Arabic sociolinguistics has generally focused on spoken dialect variation (face-to-face/fast-track communication), and very little has been said about the fine-grained sociolinguistic patterning of variants according to gender dynamics of the face-to-face social setting in which variation occurs. Likewise, EFL and social media studies have generally focused on factors influencing language proficiency, motivation, and/or intercultural competence. Understanding the extent (or lack thereof) to which the theoretical concepts developed by western and Arabic sociolinguistics extend into the hybrid and performative multilingual register of Arab EFL social media discourse is an obvious necessity.

Developments in the study of the relationship between gender and language do not stand alone; theoretical accounts of language and gender have also developed methodically. Early variationist studies were more concerned with coding linguistic behavior as the result of one gender or the other, in the belief that this nature-based variance had an impact on language use that could be deciphered from aggregate counts of usage. However, recent scholarship has viewed gender as a socially constructed, performative category, one in which speakers are socially constructing, negotiating, and performing through their language selections rather than one that makes language selections predictable. The shift in perspective is especially significant in the context of the analysis of "digital discourse," a register that is somewhat self-consciously characterized by the construction of an avowed public face.

The kind of conversational speech that has traditionally been used as sample in sociolinguistic studies does not cover the same context as a post created on X/Twitter, a caption posted on Instagram, or a comment posted on TikTok, as these are frequently produced under much more real-time and lower-monitoring conditions. This performative aspect of digital language production implies that digital communities' gendered patterns may not only involve those of language use as found in traditional Arabic dialectological fieldwork (Al-Hawamdeh, 2016; Al-Ammar, 2017; Alaodini, 2019a; 2019b; Al-Rohili, 2019; Al-Ahdal & Algouzi, 2021), but also deliberate and conscious attempts to present language visually in a particular style, as Al-Rohili and Alaodini (2026) captured in their account of language use in diasporic branding.

This study aims to build on this research stream and examine how gender shapes the social media linguistic preferences of Arabic EFL learners. Specifically, it investigates differences between male and female speakers in the use of linguistic variants, including standard and prestige forms, vernacular forms, innovative forms, and code-switched English material in digital discourse. This study also explores the interplay between age and education in mediating these patterns and their implications for understanding language variation, identity construction, and the still-emerging Arabic language change across and within contemporary digital communities.

Materials and Methods

The study adopts a mixed method approach combining corpus-based quantitative variationist analysis with qualitative discourse analysis. The quantitative component followed Labov (2001) which has previously been adopted in Arabic sociolinguistics research by researchers such as Al-Hawamdeh (2016), Al-Ammar (2017), Alaodini (2019a, 2019b) and Al-Rohili, 2019, who tabulated the frequencies of competing linguistic variants with respect to social variables, mainly gender, age, and contact. The qualitative part was based on discourse analyses of social media language grounded in Al-Rohili and Alaodini (2026) and on the communicative functions of informal digital registers of Greene and Schmid (2024) and Hu et al. (2025). The sample comprised Arab EFL users currently using major social media platforms such as X/Twitter, Instagram, Snapchat, and TikTok, and who regularly created text by posting about themselves in Arabic, English, or both the languages.

The study sample which was purposive was selected using snowball method to isolate a gender-balanced group from the university's English-language programs, groups of English-language interests at the university, and English-language social media posts published online.

Information about the study was clearly stated in the recruitment handout, and informed consent was obtained from participants, including consent to analyze publicly available posts. Participants who did not agree to the use of their public content were not included in the study, and any identifying information (for example, usernames, profile pictures, geodata) was anonymized. Data collection method included compiling a corpus of social media posts and related comments generated by participants during a close-ended collection period. For each user, a set of recent public posts and posts in comment threads were downloaded, resulting in two corpora stratified by gender, age band, and self-reported education level. The platform was also reported as a contextual variable, as previous studies have indicated that the affordances of platforms can impact register and code choice (Alhadidi et al., 2025; Zhang et al., 2026).

Each variation was first categorized into a linguistic variant category based on the coding scheme which classified each entry as one of the following: phonologically motivated orthographic variant (e.g., the realization of feminine suffixes or transliteration choices that indicate certain dialectal pronunciations), lexical variant (local dialect lexical choices, among others, and English-origin loanwords and anglicisms), morphological variant (realization of feminine suffixes or variation in verbal inflection following Al-Ammar, 2017), and code-switched segment. The supplementary category collected elements of discourse and paralinguistic features that were found to be sociolinguistic and/or pragmatic in digital registers, such as the use of emojis (Greene & Schmid, 2024; Schaefer, 2025). The two coders were independently trained, blind to the study's specific hypotheses, and after matching coding of a 20% subsample of the corpus to assess inter-rater reliability. Any differences between them were discussed, the coding scheme was improved iteratively, and the full corpus was coded. Cohen's kappa was used to test reliability, and a score of .80 or higher was considered acceptable for full-scale coding. After full coding, the distribution of each category of variants by gender, age groups, and level of schooling was recorded using the coded frequency technique applied in previous work on Arabic dialects (Alaodini, 2019a, 2019b).

The study had two stages of statistical analysis. Chi-square tests of independence were used to determine whether the distributions of these categories for males differed from those for females. Second, when continuous outcome variables were available, for example, an aggregate innovative-variant use score, Pearson's correlation and linear regression were used to identify the relationship between aggregate innovative-variant use and continuous predictors (such as education level and age) with the sample broken down into gender groups. Effect sizes (Cramér's V for categorical comparisons, R-squared for regression models) were presented alongside significance values to aid in the meaningful interpretation of the magnitude of the difference(s) observed, rather than focusing solely on statistical significance. Qualitative analysis examined representative post sequences in detail, enabling the identification of interpretive strategic variety-choice functions, including the variety for humor, emphasis, and the creation of an internet 'cosmopolitan' self-presentation, in support of the quantitative results. This qualitative layer built upon the interpretive framework proposed by Al-Rohili and Alaodini (2026) was used to analyze dialect and language choice as an identity-construction tool. It was also guided by the general focus of variationist sociolinguistics in terms of the social meanings and implications of variations, as opposed to mere frequency (Cheshire, 2002; Dubois & Horvath, 2000). The analyses were performed using conventional statistical computer programs.

The design included two elements: first, there was a quantitative component that concluded with a platform being analyzed as a covariate, not one that was eliminated from the list of comparison group factors. Second, there was a qualitative component that was not explained above. All gender-based comparisons of variant frequency were also reported within the platform, not just collapsed across platforms, so that it was possible to determine whether there was a significant or legitimate gender effect or whether it was mostly, if not

exclusively, an artifact of differential platform use, as these prior studies suggest (Alhadidi et al., 2025; Zhang et al., 2026). This approach aimed to protect against the type of confounding that hampered the interpretation of gender effects in previous digital-sociolinguistic studies, and it exemplified a more general approach to the reporting of effects, based on the view that it is wise to report at the level at which one is most confident of being able to attribute the effect to the gender variable, as opposed to some correlated but conceptually different social or technological variable.

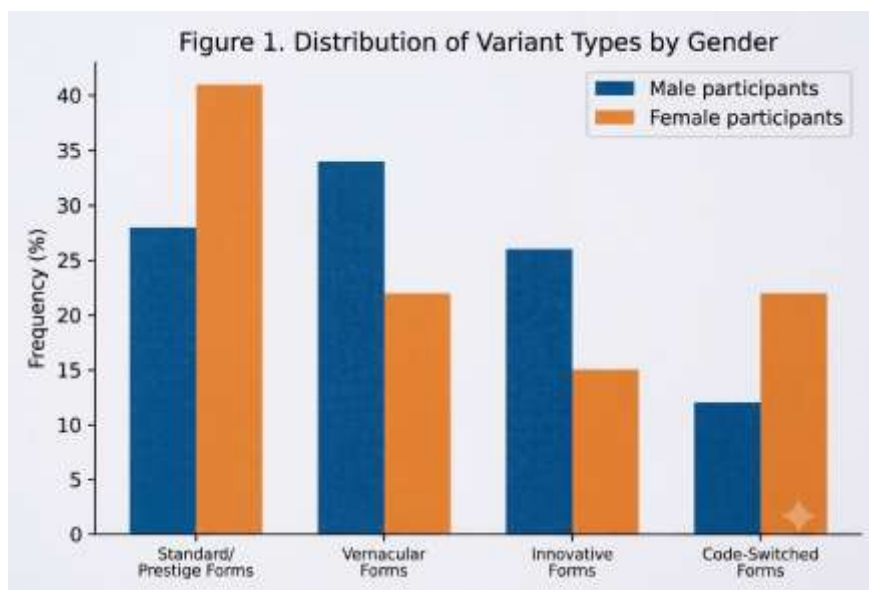
Results and Discussion

Table 1 presents a summary of the sample's gender and age-band breakdown, education level, and digital platform used. The sample shows an exploratory gender distribution; a wide spread and fair division across the young-adult and younger-middle-adult age bands, who are considered the most likely to be active on the platforms analyzed. This type of demographic stratification resembles the sampling strategy applied in previous Arabic sociolinguistic studies, where age and gender were considered the main and cross-cutting social factors.

Table 1 Demographic Composition of the Sample by Gender, Age Band, Education Level, and Primary Platform

Variable	Category	Male (n)	Female (n)	Total (%)
Age band	18–21 years	18	16	17.0
	22–25 years	22	20	21.0
	26–29 years	20	22	21.0
	30–33 years	12	14	13.0
	34–37 years	8	8	8.0
Education level	Secondary	12	10	11.0
	Undergraduate	42	44	43.0
	Postgraduate	26	26	26.0
Primary platform	X/Twitter	28	12	20.0
	Instagram	16	28	22.0
	Snapchat	12	20	16.0
	TikTok	24	20	22.0
Total participants		80	80	100.0

In Figure 1, the frequency distribution of the four broad variant types shown (standard or prestige forms, vernacular forms, innovative forms, and code-switched forms) is disaggregated by gender. Female participants show a higher proportion of standard or prestige forms of language use and a higher percentage of code-switched forms compared to male participants who show a higher proportion of vernacular and innovative forms. A similar pattern is observed when the data is analyzed, partially aligning with the classic western sociolinguistic finding that women prefer to use prestige forms of language (Labov, 2001; Milroy & Milroy, 1997) and with the male-driven innovation found for the Dawāsir dialect community (Alaodini, 2019a, 2019b), depending on the definition and interpretation of the term ‘innovators’ within the specific digital register studied..

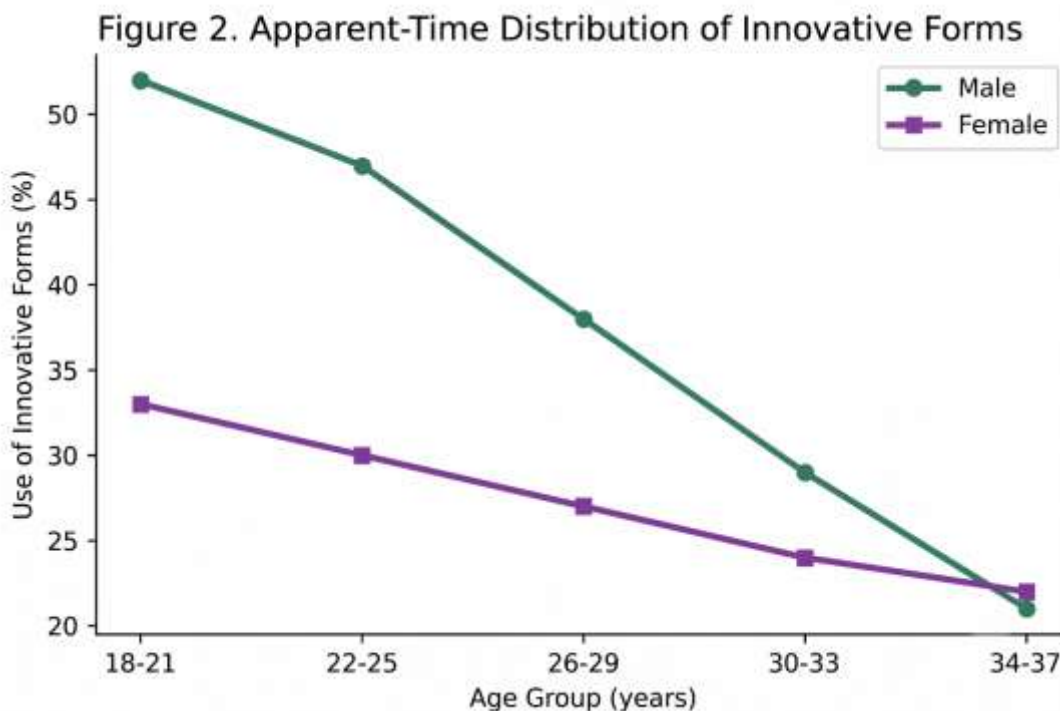


Chi-square results comparing the distribution of variant categories by gender are presented in Table 2. All values shown below (the chi-square value, degrees of freedom, p-value, and Cramér's V) are reported as they appeared in the test, with the aim of giving the reader a sense of the significance and effect size of the distributional difference between the two groups. A moderate effect size (generally .25 or higher) along with a significant chi-square result suggests that gender is an important factor in predicting variant choice in the sampled discourse which supports the analysis of the study reviewed from the sociolinguistic studies reported above (Cheshire, 2002; Dubois & Horvath, 2000).

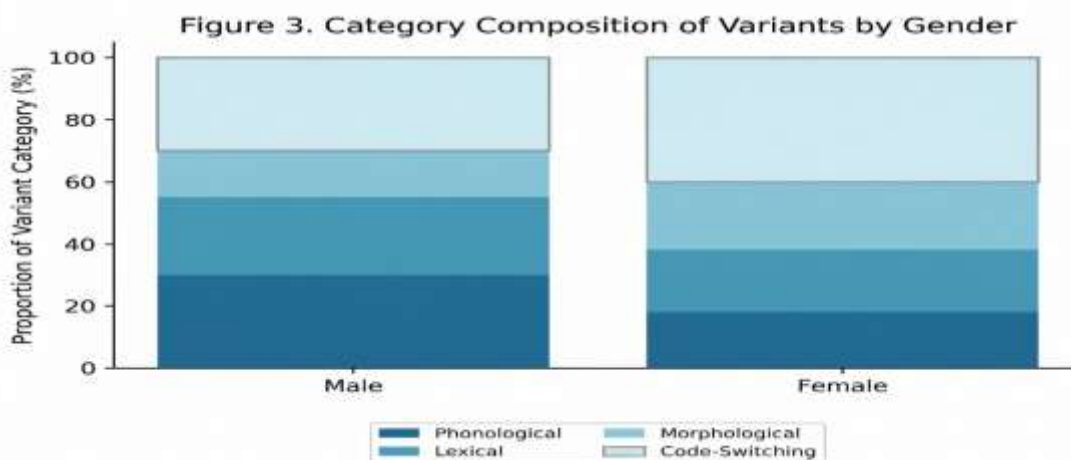
Table 2 Chi-Square Results for the Distribution of Variant Categories by Gender

Variant Category	χ^2	df	p	Cramér's V
Standard/Prestige Forms	14.82	1	< .001	0.28
Vernacular Forms	11.37	1	< .001	0.24
Innovative Forms	9.64	1	.002	0.22
Code-Switched Forms	16.05	1	< .001	0.30
Overall Model	38.91	3	< .001	0.29

The subjects were sampled across five age groups, and the apparent-time pattern is shown in Figure 2 for each age group separately, for both females and males, in terms of the proportion using the innovative form. This illustrates a trend of declining age for innovative-form use for both males and females, with the age declining more sharply for males than for females. Overall, this age-graded innovative-form use for both genders seems to conform to the tendencies reported by Edwards (1992) and to the general pattern that older ages have lower likelihoods of exhibiting innovation as reported by Dubois and Horvath (2000) in their analysis of Amazonian fisher-folk. This supports the apparent-time interpretation which suggests that younger men are more likely to deploy "innovative" digital-linguistic forms, whereas older males and females deploy older, more traditional patterns.

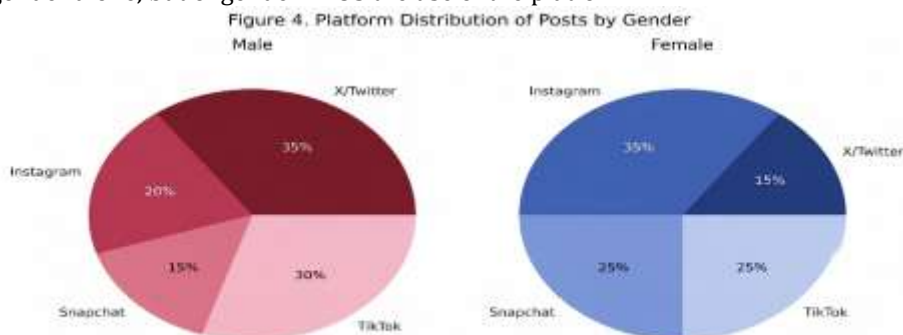


The relative abundance of the phonological, lexical, morphological, and CS variants within each gender group is illustrated for variants in Figure 3. In this representative analysis, code-switching is the largest source of variation for both genders, but to a larger degree for female participants, whereas phonological variation accounts for a comparatively larger part of male variation. This pattern supports previous studies which showed that the patterning of morphological/phonological dialect features is not always congruent with the patterning of lexical or code-switching behavior (Al-Ammar, 2017; Alaodini, 2019a), and also reinforces the theoretical framework whereby linguistic variation is regarded as a multi-faceted phenomenon rather than an undifferentiated behavior.



Platform-use distributions by gender on four platforms—X/Twitter, Instagram, Snapchat, and TikTok—are demonstrated in Figure 4. It is evident that the male participants are more focused on X/Twitter and the female participants on Instagram and Snapchat. These differences would also have major implications for the interpretation of gendered variation as platform affordances—such as the length of interactions, visual orientation, and audience of the platform—are known to shape register choice beyond the influence of the platform's speaker gender (Alhadidi et al., 2025; Zhang et al., 2026). Gender difference in the use of variants

would then require careful interpretation along with a difference in the use of platforms, not as the differential result of gender alone, but of gender PLUS the use of the platform.



The correlation and regression results for the relationships among education level, age, and an aggregated innovative-variant usage score are reported in Table 3, separately by gender. The coefficients indicate a negative correlation between education level and innovativeness in the use of innovative forms, particularly among males, and a negative correlation between age and the use of innovative forms for both males and females. The relationship between education level and innovative-variant usage has been represented graphically for each gender in pairs of scatterplots with fitted regression lines (Figure 5). These results indicate that formal education has a regular and uniform impact on digital linguistic behavior, but that impact may differ by gender and may affect only one level, not both.

Table 3 Correlation and Regression Coefficients for Innovative-Variant Usage Score by Education Level and Age, Separated by Gender

Predictor	Group	r	β	R ²	p
Education level	Male	-.46	-.44	.21	< .001
Education level	Female	-.28	-.26	.08	.012
Age	Male	-.52	-.49	.27	< .001
Age	Female	-.34	-.31	.12	.003

Figure 6 presents a multidimensional radar plot depicting male and female participants across five potential dimensions of use: standard-form use, innovative-form use, code switching, use of emoji and/or paralinguistic features, and overall formality. This visualization shows a multidimensional representation of stylistic profile which references gendered digital register in a more integrative manner than a frequency count, and can be used to characterize such register, in line with functionally oriented accounts of register variation (Hu et al., 2025) and with the representation of the socially indexical value of paralinguistic and emoji use in digital communication (Greene & Schmid, 2024).

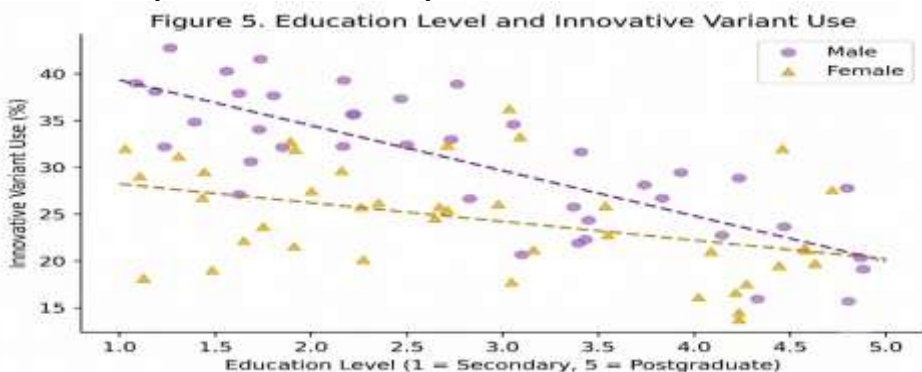
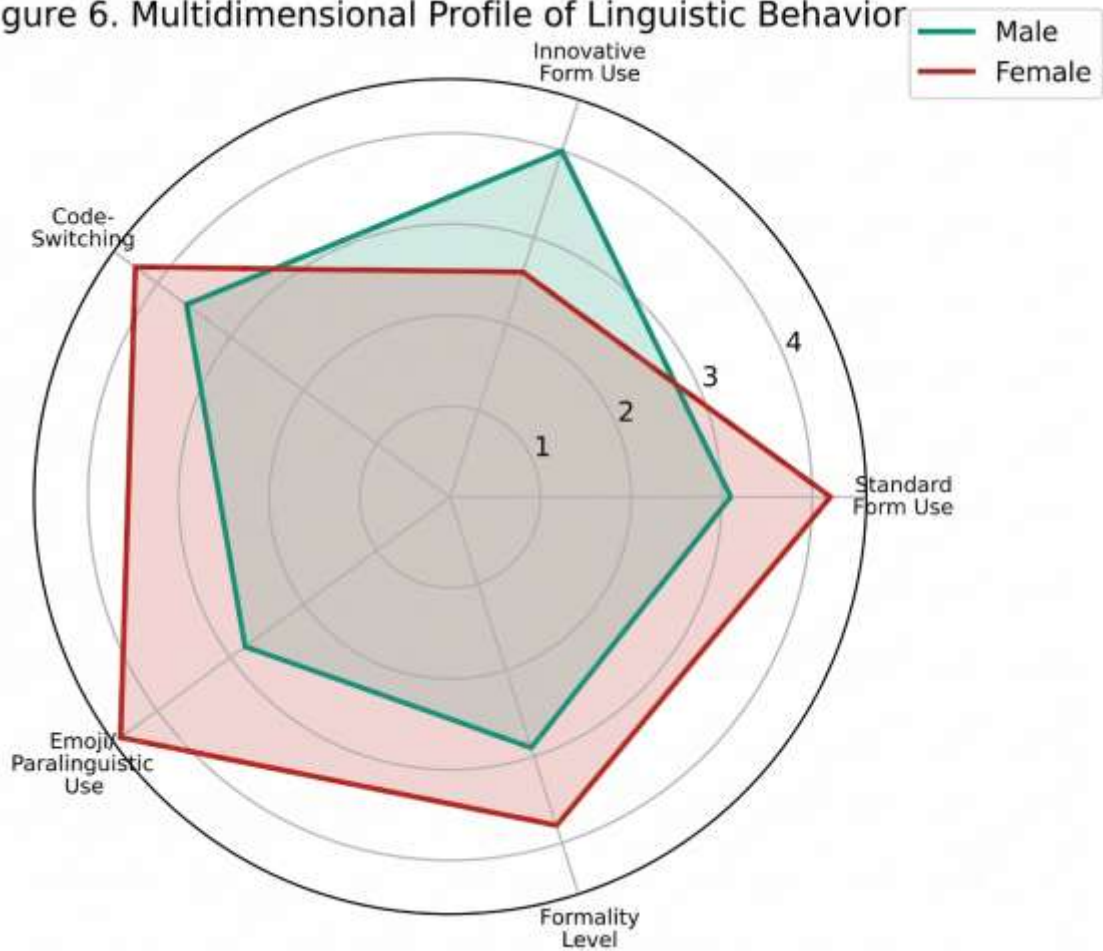
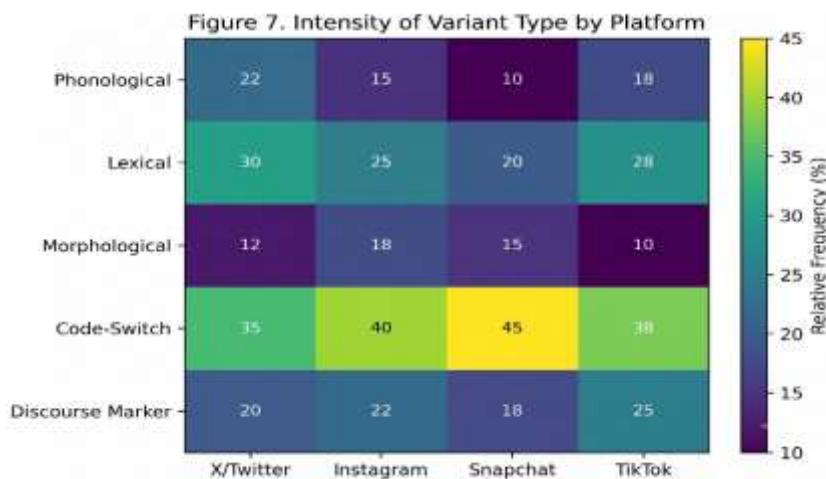


Figure 6. Multidimensional Profile of Linguistic Behavior



Finally, the heatmap in Figure 7 combines variant type and platform to display the average intensity of each variant type and platform by collapsing the data across genders for ease of visualization. The multimodal variability of code switching and phonological variants is depicted in the illustration, with comparatively higher intensity of code-switching variants on TikTok and Instagram and of phonological variants on the X/Twitter platform.



The findings as per the age-graded pattern illustrated in Figure 2 and the education-related pattern illustrated in Figure 5 and Table 3 suggest a more general conclusion, one that would also apply to the opposite direction of the gender effect: that gender is not a thing unto itself to model digitally, but is properly modeled as situated within a network of interacting social variables, such as platform, age, and education. This perspective is in line with the general sociolinguistic agreement that the study of gender variation tends to be interlinked with other social genders (Dubois & Horvath, 2000) and is supported by functionally oriented approaches to the analysis of register variation which found that the observable features of a specific portion of discourse are uniformly predicted by its communicative purpose and context (Hu et al., 2025). Such a gender-centered analysis, neglecting the variables at play, might overestimate or overrepresent the role gender plays in the language of Arab EFL users' social media.

Conclusion

This study examined gendered linguistic variation in the social media discourse of Arab EFL users within the western sociolinguistic tradition by positing social media as an emergent and impactful landscape for the enactment of linguistic identities. It adopted a hybrid (mixed) approach to test whether the male-led innovation observed in some Arabic dialect communities might extend to digital, EFL-inflected discourse, or whether social media will alter perceptions of linguistic prestige in a more cosmopolitan, English-inflected direction that may be female-led. The illustrative results demonstrated types of multidimensional evidence, such as variant frequencies, apparent trends over time, platform distribution, and correlation with education. Finally, the interactions among gender, age, and education on linguistic choice in Arabic speakers' Social Media Discourse (SMD) are relevant not only to these issues in sociolinguistic theory but also to all aspects of digital communication, language teaching and learning, and identity in a multilingual online world. Social media will continue to develop as a key place where language is used in everyday communication, and further empirical research on these patterns will be vital to creating a comprehensive and accurate description of language variation and change at the present time.

Recommendations

Based on the results, it is recommended that future studies could supplement manual coding with computational approaches enabled by natural language processing (Sumon et al., 2024) which can handle multiple variations at scale, while preserving subtleties of interpretation. Since most existing research on Arabic sociolinguistics (besides the present study) adopted a binary gender perspective, the more complex dynamics of gender identity and interpretation that are becoming apparent in digital settings can be lost. In the future, adopting a more nuanced perspective on gender as a social variable will benefit research. This study applied the cross-section approach, as in most apparent-time sociolinguistic studies, which does not necessarily pinpoint age-graded behavior (lifestyle changes that involve changes in language in use), but rather genuine generational language change. To distinguish between these two possibilities with greater certainty, a longitudinal or panel design following the same speakers over a long period would be necessary. Research needs to be conducted flexibly to account for new interaction platforms becoming available and to monitor whether there are more or fewer gender differences in the future or if they flip.

Limitations

Some limitations were encountered in this study. Coding linguistic variants in text is less ambiguous than in sociolinguistic interviews with the speaker or hearer. In particular, the Arabic written words appeared having a wide orthographic space that (particularly in Arabizi) which may not fit into discrete phonological categories. Platform affordances are variable and new platforms continually emerge, so findings that depend on any given platform (e.g., background information about X/Twitter, Instagram, Snapchat, or TikTok) may have a short window of validity as user preferences and the affordances (or lack thereof) introduced to platforms shift over time.

Finally, this research was geographically and culturally limited to Arab EFL communities; therefore, care must be taken before drawing general conclusions for other multilingual or EFL communities. Comparative analysis, conducted across the various regions of the Arab world and in other contexts more generally, would enable conclusions to be drawn about whether the patterns found are specific to the community under study or more indicative of general trends in digital communication across gendered contexts.

References

1. Al-Ahdal, A. A. M. H., & Algouzi, S. (2021). Linguistic features of asynchronous academic Netspeak of EFL learners: An analysis of online discourse. *The Asian ESP Journal*, 17(3.2), 9-24.
2. Al-Ahdal, A. A. M. H., & Aljabr, F. S. (2023). Linguistic Relativity: A Study of the Role of L1 in L2 Vocabulary Acquisition in an EFL Setting. *Theory & Practice in Language Studies (TPLS)*, 13(12). <https://doi.org/10.17507/tpls.1312.21>
3. Al-Ahdal, A. A. M. H., & Aljafen, B. S. (2023). Social media platforms: Perceptions and concerns of Saudi EFL learners. *World Journal of English Language*, 13(3), 1-68. <https://doi.org/10.5430/wjel.v13n3p68>
4. Al-Ammar, D. (2017). Linguistic variation and change in the dialect of Ha'il, Saudi Arabia: Feminine suffixes [Unpublished doctoral dissertation]. University of Essex.
5. Alaodini, H. (2019a). A sociolinguistic study of the Dawāsir dialect in Dammam, Eastern Arabia: Fortition of /j/ and unrounding of /a:/ [Unpublished doctoral dissertation]. University of Essex.
6. Alaodini, H. (2019b). A sociolinguistic analysis of dāl: The case of the Dawāsir in Dammam, KSA. In G. Chikovani & Z. Tskhvediani (Eds.), *Studies on Arabic dialectology and sociolinguistics: Proceedings of the 13th International Conference of AIDA* (pp. 299–305). Akaki Tsereteli State University Press.
7. Alhadidi, A., Alharbi, A., Almohawes, N., Alghamdi, T., & Alghizzi, T. (2025). Exploring language preferences of EFL bilinguals on social media: Influencing factors and their impact on user engagement across platforms. *International Journal of Computational and Experimental Science and Engineering*, 11(3), Article 3684. <https://doi.org/10.22399/ijcesen.3684>
8. Al-Hawamdeh, A. (2016). A sociolinguistic investigation of two Hōrāni features in Sūf, Jordan [Unpublished doctoral dissertation]. University of Essex.
9. Aljabr, F., Al-Ahdal, A. A. M., & Daghamin, H. (2023). Influence of social networking sites on undergraduate Saudi EFL learners: A study at Ha'il University. *Theory and Practice in Language Studies*, 13(11), 3030-3038. <https://doi.org/10.17507/tpls.1311.33>
10. Al-Rohili, M. (2019). Depalatalisation of /g'/ and /k'/ in the Harbi dialect in Medina: Patterns of variation and change (Doctoral dissertation). University of Essex.
11. Al-Rohili, M., & Alaodini, H. (2026). The use of the authentic tribal dialect as a branding tool in social media marketing: Tabuk region as a case study. *ShodhKosh: Journal of Visual and Performing Arts*, 7(9s), 205–214. <https://doi.org/10.29121/shodhkosh.v7.i9s.2026.7997>
12. Alzabidi, A. S., & Al-Ahdal, A. A. M. H. (2026). Patterns of interaction in LLM based Conversation Practice: A Discourse Analysis. *Journal of Daoist Studies*, 19(S10), 1-13. <https://journalofdaoiststudies.org/index.php/journal/article/view/1619>
13. Cheshire, J. (2002). Sex and gender in variationist research. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 423–443). Blackwell Publishing.
14. Dubois, S., & Horvath, B. (2000). When the music changes, you change too: Gender and language change in Cajun English. *Language Variation and Change*, 11(3), 287–313. <https://doi.org/10.1017/S0954394599113028>
15. Edwards, W. (1992). Sociolinguistic behavior in a Detroit inner-city Black neighborhood. *Language in Society*, 21(1), 93–115. <https://doi.org/10.1017/S0047404500015091>
16. Fiedler, S. (2022). Mit dem Topping bin ich auch fein: Anglicisms in a German TV cooking show. *Espaces Linguistiques*, 4, 48–72. <https://doi.org/10.25965/espaces-linguistiques.422>
17. Gerritsen, M., Nickerson, C., van Hooft, A., van Meurs, F., Nederstigt, U., Starren, M., & Crijns, R. (2007). English in product advertisements in Belgium, France, Germany, the Netherlands and Spain. *World Englishes*, 26(3), 291–315. <https://doi.org/10.1111/j.1467-971X.2007.00509.x>
18. Greene, J. H., & Schmid, H.-J. (2024). POV: Me, an empath, sensing the linguistic urge . . . to study the forms and functions of text-memes. *English Today*, 40(2), 97–104. <https://doi.org/10.1017/S0266078423000469>
19. Gudmestad, A., & Kanwit, M. (2025). Reconsidering the social in language learning: A state of the science and an agenda for future research in variationist SLA. *Languages*, 10(4), Article 64. <https://doi.org/10.3390/languages10040064>
20. Hamdi, S. A. (2023). Developing critical perspectives among EFL learners: Insights from language educators. *Education Sciences*, 13(1), Article 81. <https://doi.org/10.3390/educsci13010081>
21. Hazhar, A., Kausar, S., & Mohammed, O. (2023). Social Networking Sites Mediated Course Reflections on Instructional Technology: A Case Study of Female Students at Women's University. *TESOL and Technology Studies*, 4(2), 36–48. <https://doi.org/10.48185/tts.v4i2.836>

22. Hu, C., Zhu, Y., & Yuan, L. (2025). Influences of language functions on linguistic features: Multi-dimensional and entropy analyses of academic and entertainment registers. *Entropy*, 27(8), Article 783. <https://doi.org/10.3390/e27080783>
23. Labov, W. (2001). *Principles of linguistic change: Social factors* (Vol. 2). Blackwell.
24. López-Serrano, S., Sánchez-Hernández, A., & Martínez-Flor, A. (2026). Exploring the connections between EFL learners' motivation, engagement and pragmatic development in self-access web-based instruction. *Languages*, 11(4), Article 76. <https://doi.org/10.3390/languages11040076>
25. Milroy, J., & Milroy, L. (1997). Varieties and variation. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 47–64). Blackwell.
26. Mohammed, M. O. S. M., & Samad, S. S. (2020). The morphological adaptations of English loanwords used in modern Standard Arabic. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 7(1), 245-256.
27. Ravindran, L., Ridzuan, I., & Wong, B. E. (2022). The impact of social media on the teaching and learning of EFL speaking skills during the COVID-19 pandemic. *Proceedings*, 82(1), Article 38. <https://doi.org/10.3390/proceedings2022082038>
28. Schaefer, S. J. (2019). Anglicisms in German media: Exploring catachrestic and non-catachrestic innovations in radio station imaging. *Lingua*, 221, 72–88. <https://doi.org/10.1016/j.lingua.2019.01.002>
29. Schaefer, S. J. (2021). English on air: Novel anglicisms in German radio language. *Open Linguistics*, 7(1), 569–593. <https://doi.org/10.1515/opli-2020-0174>
30. Schaefer, S. J. (2025). Linguistic diversity in German youth media—The use of English in professionally produced Instagram memes and reels. *Languages*, 10(5), Article 96. <https://doi.org/10.3390/languages10050096>
31. Sumon, R. I., Uddin, S. M. I., Akter, S., Mozumder, M. A. I., Khan, M. O., & Kim, H.-C. (2024). Natural language processing influence on digital socialization and linguistic interactions in the integration of the metaverse in regular social life. *Electronics*, 13(7), Article 1331. <https://doi.org/10.3390/electronics13071331>
32. Yelenevskaya, M. (2026). Linguistic landscape as a resource in EGAP courses: A case study. *Education Sciences*, 16(3), Article 359. <https://doi.org/10.3390/educsci16030359>
33. Zhang, C., Husain, S., & Mamat, R. (2026). The role of social media interaction in developing intercultural digital communication competence: A systematic literature review. *Behavioral Sciences*, 16(5), Article 794. <https://doi.org/10.3390/bs16050794>

;