

A Modernized Discourse among Algerian Youth on Social Networking: Case Study of Romanized Arabic

Nawel Bengrait *¹

¹ University 8 Mai 1945-Guelma (Algeria),

Linguisticstoday@gmail.com

Received: 20/10/2021

Accepted: 04/12/2021

Published: 31/12/2021

Abstract

The accessibility to new technological devices and social media platforms led to the emergence of a new form of written Arabic script among Algerian young users named Arabizi, composed of Roman alphabet and which adheres the rule of writing simultaneously as speaking. The present paper aims at (i) synthesizing the samples of Arabizi in order to determine its characteristic features, and (ii) to understand and explain the reasons behind these tendencies. The Findings were collected by adopting a qualitative research design with the use of a Text Corpus analysis approach applied on written discourse uttered by Algerian EFL students on a Facebook page, who study at the Department of English, University of Guelma, Algeria. The compiled results revealed that the users produce an electronic discourse, which varies in terms of transliteration from one interlocutor to another, and in which features of code-switching into French and English are spotted. The non-standard application of the Latin orthography caused character ambiguity in relation to the corresponding Algerian Arabic alphabet. Moreover, a distinction has been made between the transliterated structures and electronic texting as many of the extracted forms comprise chat abbreviations in English and French; which are not linked to Arabic conventions.

Keywords:

Algerian Arabic, Arabizi, Electronic Algerian Arabic Dialect, Romanized Arabic,

1. INTRODUCTION

Algeria, besides Berbers, its original inhabitants, conquerors including Phoenicians, Romans, Vandals, Byzantines, Arabs, Spaniards, Turks, and French, colonized the country for centuries due to its richness of resources and strategic position in the North African coast (Arezki, 2008; Benrabah, 1999a; Ibrahimi, 1997; Morsly, 1986). This situation resulted in

¹ Corresponding author : Nawel Bengrait

cultural and pluralism diversity, and brought new modifications into the native language of the indigenous society to become what is known as ‘*Algerian Arabic*’ or ‘*Darja*’. Arabization was imposed since independence 1962, and Amazigh Language or Tamazight was first granted a national language status in 2002, and then became an official language in 2005. Classical Arabic is the language of the Quran and it is not frequently used among Algerian speakers, unlike Modern Standard Arabic (MSA) that is taught at school and employed in administrative paper work and the news broadcasts, etc. To this end, young Algerians surfing online developed a new Algerian Arabic written script or what is termed as EAAD; that serves their communicative necessities and copes with the social media interactive needs. In light of this, it became a necessity to understand the functionality of this language resulted from online social exchange, which differs remarkably from Modern Standard Arabic (MSA).

2. Literature Review

During the 1990s, the American Standard Code for Information Interchange (ASCII), which is a system for encoding and it designates a particular number to a given character in text representation, was manipulated restrictedly in online communication. Arabic script was developed to comprise Roman characters when typing in an online exchange environment, and which was adopted by users all over the world. The main reasons for this applicability is the users’ familiarity of the French/English keyboard in early years of the web when Arabic keyboard was not available, situation of multilingualism, and the accessibility to technological devices (Chalabi & Hany Gerges, 2012).

Arabizi is characterized with the use of *sound-to-letter* principle which is associated with the spoken dialect, that may differ in terms of pronunciation depending on the geographical area. The spontaneous orthographic conventions of Arabizi make use of Latin alphabet, numerals, symbols, and punctuation to convert the spoken language into text when conversing on social networking, chat, and Short Message Service (SMS). The Arabizi writing turns a phoneme into grapheme with correspondence to Arabic pronunciation into Roman script.

In light of this, transliteration is the mapping or conversion of one text script to another. This system focuses on the representation of characters or letters not sounds, unlike transcription. Transliteration involves the Romanization system that aims to construct a spoken dialect into a written script with the employment of Latin alphabet, and displaying pronunciation of the language understood by its readers/users. However, the transliteration of Arabic script is not a mere substitution of Arabic letter with a Roman letter. In this regard, Arabizi shows some particular elements. Sometimes vowels in Arabizi are represented with its Roman counterparts but usually are deleted in words compositions. The consonants representation in Arabizi with Latin letters are in ambiguous relationship due to the fact that a Latin grapheme may have more than one phoneme matches (Al-Badrashiny et al., 2014).

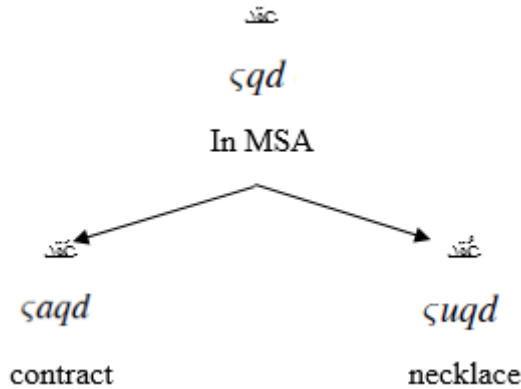
A Conventionalized Orthography for Dialectal Arabic (CODA) is a conjoined framework used to write Arabic dialects, in which every word is represented with a unique orthography (Habash et al., 2012a; Eskander et al., 2013). CODA has some characteristics as it works in a coherent and consistent way, it is developed for computational goals, and it is easy to manipulate by Arabic users. Furthermore, when transforming Arabizi into CODA the Arabic script is employed similarly as in MSA and no adoption of symbols is considered and it is unique to any given Arabic dialect.

Varieties of written Arabic scripts are examined in terms of Natural Language Processing (NLP), which demonstrate challenges at morphological level for possessing a large amount of word forms, dialectal differences, and inconsistent and ambiguous orthography (Habash et al., 2019). Arabic words are subjected to inflectional morphology that depends on patterns as number, gender, person, mood, and clitics including negative particles, conjunctions, etc., which results in massive amount of words that exhibits the morphological richness of Arabic.

With reference to dialectal Arabic, variations can be classified as regional involving Levantine, Egyptian, Gulf, Maghreb, etc. These dialects

are distinct from MSA and from one another at morphological, lexicon, and phonological levels (Watson, 2007 as cited in Habash et al., 2019).

Arabic words are associated with diacritical marks in consonants gemination and short vowels, which may cause ambiguity at orthographic and semantic levels in case these cues are missing. In this regard, Habash (2010) stated that the absence of diacritics can be challenging to non-native speakers that may lead to difficulties in NLP. For instance,



It is observed that online written texts comprise spelling errors which is considered problematic in NLP. Zaghouani et al. (2014) pointed out that 32% of composed words on social networking are misspelled. Similarly, Eskander et al. (2013) noted that about 24% of Egyptian online users' productions do not conform to Egyptian CODA due to spelling errors and as result of Arabizi (Darwish, 2013); consequently, this demonstrates orthographic inconsistency of Arabic script.

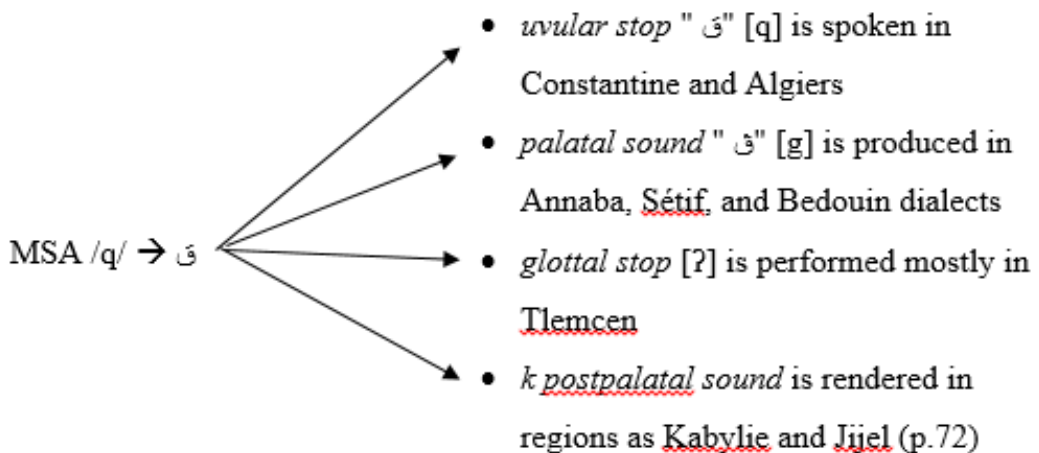
There is a growing interest in processing the Arabic dialectal varieties due to the progressive development in communication via Internet. In this respect, research focuses recently on elaborating new instruments to investigate the nature of the modernized Arabic that displays a non-normalized and non-standard orthography. In this regard, attempts were made to develop guidelines for dialectal Arabic transcription and conventions for the annotation of orthography as Maamouri et al. (2004) rules of Levantine and Habash et al. (2012a) of Egyptian CODA.

Similarly, Harrat et al. (2014) came up with the *Grapheme-to-Phoneme* transcription in which a written word is converted to its corresponding pronunciation. Consequently, the researchers created Algerian Arabic corpora as a way to realize Machine Translation (MT) that treats issues of Algerian dialect and MSA. For this reason, they built models for morphological synthesis and suggested a system for diacritization, which permitted the generalization of findings in relation to MT. Furthermore, Harrat et al. (2015) presented a research that comprises a number of Arabic dialects including the Middle Eastern and Maghreb for the purpose of comparison with MSA. This allowed to collect resources Algerian dialects of Algiers and Annaba, Tunisian of Sfax and South of Tunisia, and Middle East varieties in Syria and Palestine.

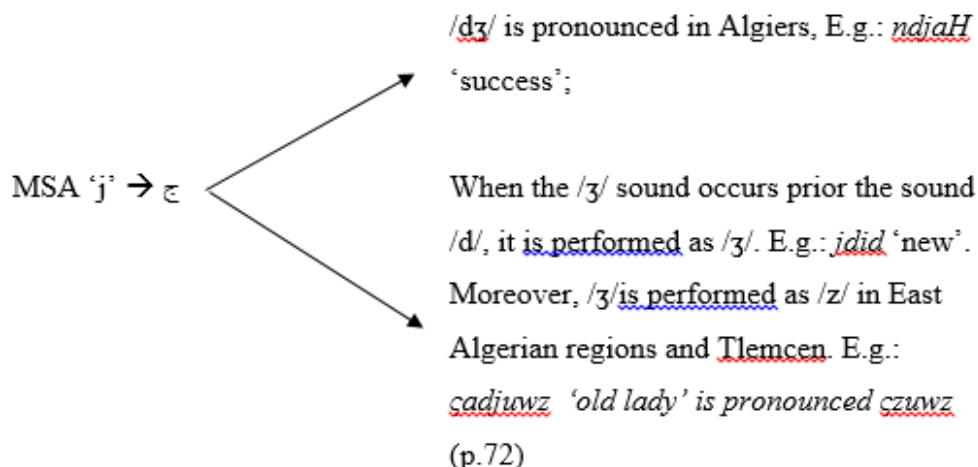
It is found that Algerian dialects differ from MSA at phonological, morphological, orthographic, and lexical levels. These distinctions are explained and illustrated as follows (Saadane & Habash, 2015):

2.1. Phonological Differences

Some sounds in MSA are rendered sometimes in Algerian Arabic differently, for example, the /q/ sound has several realizations. These include the following compositions:

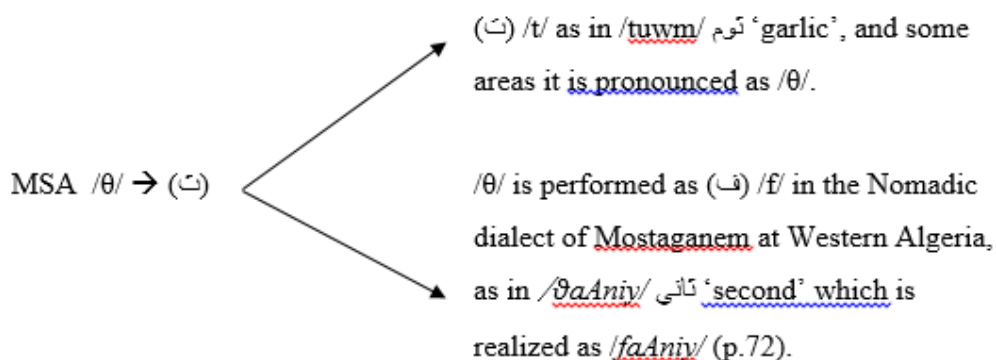


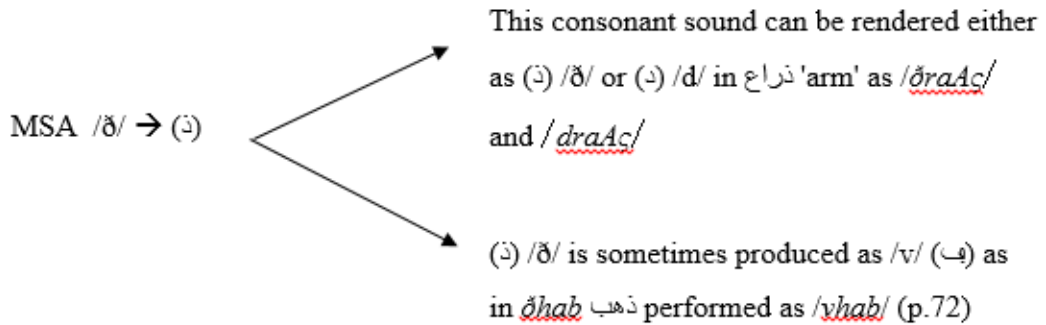
Additionally, the production of the consonant sound /ʒ/ → ڄ is distinct from one region to another in Algeria. For instance:



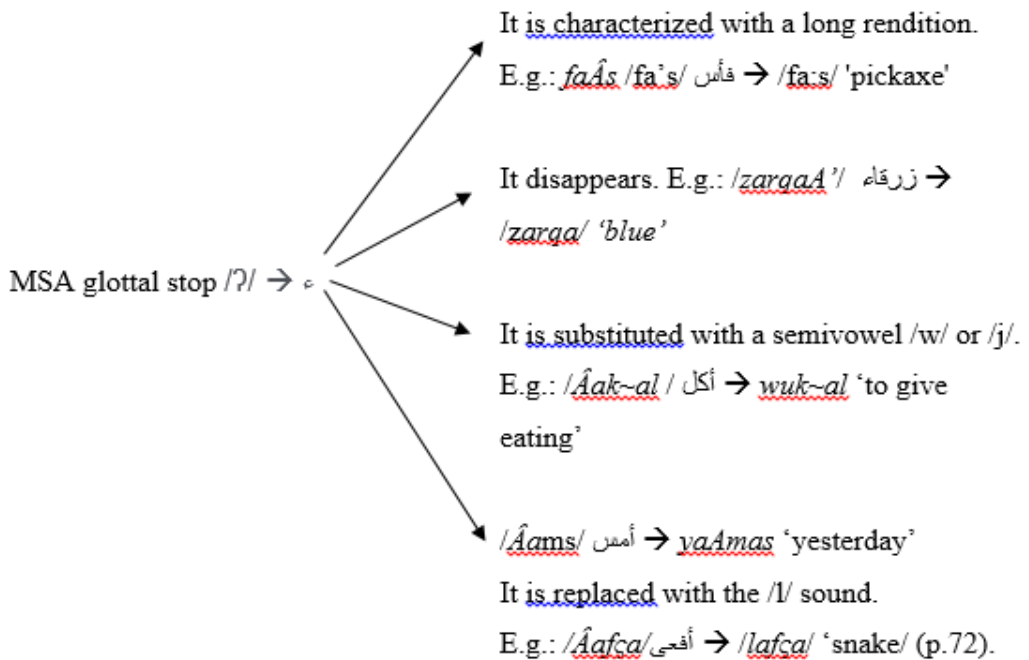
Moreover, in MSA, the assimilation of the consonant sound (غ) /ɣ/ results in the realization of (ق) /q/ and (خ) /x/ in Eastern Algerian regions and Sahara. For instance, /sɣayraħ/ صغيرة 'small' is uttered as /sqayraħ/, and /ɣssel/ غسل 'washed' is produced as /xssel/, respectively.

Another case in MSA of interdental consonants (ث) /θ/ and (ڌ) /ð/ are tend to be pronounced distinctly by Algerian speakers. These are illustrated in the following:





Furthermore, the MSA glottal stop has diverse pronunciations in Algerian Arabic. These are summarized as follows (p.72):



Algerian Arabic exhibits the tendency of deleting short vowels in unstressed words; this feature is shown in MSA as in *jamal* جَمَل 'camel', which is performed as *jmal* جُمَل. Therefore, this results in the use of successive consonants at word initial position, particularly in verb form, for instance, */qatal/* قَتَلَ 'killed' in MSA becomes */qtal/* قُتِل in Darja (Saadane & Habash, 2015, p. 72).

2.2. Morphological Differences

According to Saadane and Habash (2015, pp.73-74), Algerian Arabic possesses some morphological elements that are distinct from MSA. Some of these aspects are summed up in terms of inflexion and addition of certain clitics, as follows:

- Feminine plural behaves in the same way as masculine plural. E.g. /šakartun~a/ شَكَرْتُنْ 'they thank' becomes /škar-tuwa/ شَكَرْتُوا
- The conjugation in the first and second person are quite the same in Algerian Arabic as in /škart/ شَكَرْتُ 'I thank' and 'you thank', unlike in MSA /šakartu/ شَكَرْتُ and /šakarta/ شَكَرْتَ, respectively.
- Insertion of long vowel /i:/ between the root word and the consonant suffixes. E.g.: šad~a/šadadtu شَدَّ شَدَدْتُ 'he/I pulled' is produced as šad~/šad~iyt/ شَدَّ شَدِيْثٌ, respectively.
- The passive voice form is built by prefixing the verb unlike in MSA which uses vowel change. E.g. Adding the prefix *tn-* as in /tnaqtal/ تَنْقُتْلُ 'to be killed'.
- The utilization of particle 'n' especially in verb formation. E.g.: /naktab/ نَكْتَبُ 'I write'.
- The employment of clitics which are basically a reduced MSA structure.

Examples:

- The demonstrative proclitic + هـ Ha+ precedes the article *Al* as in /hađihi AldunyaA/ هذه الدنيا 'this life' in MSA becomes /haAldinyaA/ in Algerian Arabic.
- The preposition proclitic + عـ ɟa+ is reduced as in /ɟalay AlTaAwilaḥ/ على الطاولة 'on the table' is pronounced as /ɟaAlmaAydaḥ/ عالمائدة.

2.3. Orthographic Differences

Algerian Arabic is found to vary in orthography from MSA due to the fact that there is no normalized or codified form of the dialect, and the phonological variations result in a non-standard orthography. The dialect spelling depends on the etymological and phonological forms of MSA; which caused inconsistency in the Algerian written compositions. For instance, some words are assimilated as in /*Tuwmuwbiyl*/ طوموبيل and *Tuwnuwbiyl* طونوبيل 'car'. In other situations, this can lead to semantic ambiguity, for example, in the third person singular and the plural: /*šrbw*/ شربو and /*šarbuwA*/ شربوا 'they drank' is confused with /*šarbuh*/ شربه 'he drank it' (Saadane & Habash, 2015, p.74).

2.4. Lexical Differences

Colonization of Algeria resulted in language mixing that led to the transfer of linguistic and cultural features. At syntactic level, many loan words are used in Algerian Arabic dialect. Table 1 shows some of these words (Benrabah 2002, pp.235-242):

Table 1. Borrowed Words in Algerian Arabic

Loan Word in Algerian Arabic	Originated From:
<i>rfissa</i> and <i>barkukas</i> (traditional dishes)	Latin
<i>boukraj</i> (kettle) and <i>braniya</i> (aubergine)	Turkish
<i>kanasta</i> (basket) and <i>essekouila</i> (school)	Spanish
<i>miziriya</i> (poverty) and <i>kamyoun</i> (lorry)	French

3. Research Methodology

The current study adopts a *qualitative Corpus Analysis* methodology that enables to conduct an in-depth examination of a linguistic situation related to a communicative act, which displays authentic language corpora. This approach is exploratory in nature centered on investigating the interactive connections established between language meanings and functions, and the features associated with the utilized language for communicative purposes. These characteristics comprise interlocutors' gender, age, socioeconomic class, educational level, etc. Furthermore, the approach focusses on processing real-time data produced in written and/or oral discourse setting manipulated with computer-aided device (Hasko, 2012, p. 4758).

The text corpus analysis to discourse will enable to reflect on the tendencies applied in Algerian discourse exchanged on social media context. According to Schiffrin, et al. (2001), discourse analysis can be categorized into three groups: (i) the language use study; (ii) the linguistic structure study that is referred to as 'beyond sentence'; and (iii) the social activities related to communicative and/or language usage. First, language use addresses difficulties of language forms entities as diverse structures expressing quite identical meanings such as adverbial particle shift from one position to another. Second, beyond sentence structure extends the synthesis of utterances and focus on systematic texts constructions. Third, the social-cultural concern of discourse is to explore the interlocutors' actions in a given communicative situation that allows portraying discoursal cues with regard to the speech community (Upton & Cohen, 2009). Thus, the paper will investigate the gathered texts from different perspectives, which will permit to describe and annotate the application of EAAD scripts produced by young Algerians.

3.1. Purpose of the Study

The study aims are threefold:

1. To access dialectal content of EAAD utilized on social networking for the collection of corpora.

2. To examine the EAAD written script and extract its characteristics.
3. To contribute in the body of literature and resources regarding the Romanized Arabic application on the web.

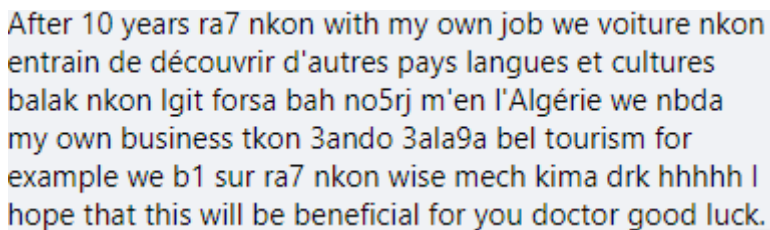
3.2. Materials and Procedure

The data was generated via raising a discussion on the social media platform Facebook, among English as a Foreign Language (EFL) Master Two students at the Department of English, University of Guelma, Algeria. The data was in the shape of comments corpus motivated by the question ‘*Where do you see yourself in ten years?*’ that was randomly selected, and information was stored on computer to facilitate the process of retrieval and analysis.

4. Findings

The collected results are described and analyzed in terms of code switching/multilingualism and Romanized characters’ opacity.

The EAAD produced by Algerian EFL students on social media comprises informal language, in which the users are not constrained to particular orthographic conventions or grammatical rules, but rather, they are free to express themselves in writing as they speak. Additionally, the utilized language style exhibits the utilization of borrowings from French and English in EAAD discourse as demonstrated in Figure 1.



After 10 years ra7 nkon with my own job we voiture nkon
entraîné de découvrir d'autres pays langues et cultures
balak nkon lgit forsa bah no5rj m'en l'Algérie we nbda
my own business tkon 3ando 3ala9a bel tourism for
example we b1 sur ra7 nkon wise mech kima drk hhhhh I
hope that this will be beneficial for you doctor good luck.

Figure 1. EAAD Script Sample

The retrieved text is translated in English as:

After 10 years I will be with my own job and car [,] I will be in the process of discovering other countries' languages and cultures [,] may be I will find the opportunity to leave Algeria and start my own business that will have a relation with tourism for example [,] of course I will be wise not as in this moment [laughter] [,] I hope that this will be beneficial for you doctor [,] good luck.

The given structures are grouped in the three languages to illustrate code switching in Table 2.

Table 2. EAAD Code-Switching

	Algerian Arabic	French	English
Code Switching	• ra7 <u>nkon</u> (I will be) • we (and) • <u>nkon</u> (I will be) • <u>balak nkon lgit forsa</u> <u>bah no5rj</u> (may be I will have the opportunity to leave the country = to travel) • <u>m'en</u> (from) • we <u>nhda</u> (I start) • <u>tkon</u> 3ando 3ala9a <u>bel</u> (to have a relationship with) • <u>mech kima drk</u> (not as in the moment)	• <u>voiture</u> (car) • <u>entraîn de découvrir</u> <u>d'autres pays langues</u> et cultures (...in the process of discovering other countries languages, and cultures) • <u>l'Algérie</u> (Algeria) • b1 <u>sur</u> (of course)	• After 10 years • with my own job • my own business • tourism for example • I hope that this will be beneficial for <u>you</u> doctor good luck. • wise

The fact that EAAD is not confined to an explicit vocabulary system makes it difficult to predict and transliterate the form of the structure. Some of these forms are summarized in Table 3.

Table 3. EAAD Orthographic Ambiguity

EAAD	Transliteration	English
nkoun and nkon	nkun	I will be
we, wa and w	wa	And
wl youm and youm	elyum/jum	Today
wel and b	bil/bi	With
nchAllah and inshAllah	inshAllah	With God's will
fi and F	Fi	In
min and mn	min	Since
hilimt/hlimt/hlmt	hlimt	[I] dreamed

The students seem not to have control over vowel choice, as they tend sometimes to delete the vowel or replace it with the apostrophe (') in order to cope with the quick and simple requirements of communicating content in an online environment. The users are found to employ numerals to substitute the consonants that do not have an approximate phonetic construction in the Roman script, as in “no5rj” – نخرج , “ra7” – راح , “3ando” – عندو , “3ala9a” – علاقة, etc. These numbers correspond to the following orthography:

2 – ء → '

 3 – ع → '

 5 – خ → kh

 6 – ط → t

 7 – ظ → h

 8 & 9 – ق → q

The annotation of the Algerian youth written discourse displays that at word level there is a confusion related to the used Roman character and its EAAD counterpart. It is found that a given character can have more than two Arabic matches in relation to the application setting and what the user prefers (Table 4).

Table 4. EAAD Morphological Variations

Roman Character	Arabic Match	EAAD Transliteration
Kh	خ ق + ه ك + ح	khali – خالي (uncle) Khar – قهر (defated) khal – كحل (black)
Sh	ش س + ه ص + ه س + ح	inshAllah – انشالله (with God's will) shalt - سهالت (easier) 5sha – خصها (she/it needs) Shar – سحر (charmed)
dh	ض ذ ده	dhrob – ضرب (hit) dhork – نرك (now) 9dhom – قدمهم (they/it suit/s them/it)
th	ت + ه ث ت + ح ط + ه ط + ح	miditha – مدتھا (she gave) th9ila – ثقيلة (heavy) tithal – تتحل (it opens) 7athom – طھم (s/he put them) than – طحن (grinding)

Additionally, it is observed that the EAAD script does not differentiate between lowercase and upper case letters, thus, it does not respect the Latin conventions. Moreover, the users seem to have the tendency of utilizing the French chat abbreviations as “b1 sûr” (of course), and “qlq” (some) for ‘quelque’ in French. The prepositions are embedded in the word and are not written separately, this confirms the rule of ‘writing as they speak’ and little attention is paid to the grammatical structure such as: ‘flkhadma’ which is transliterated as ‘fil (at) khadma (work)’.

5. Discussion of Results

The reported results are found to be in accordance with the reviewed literature as the Algerian written script makes use of some MSA conventions but it may differ at some phonological, lexical, morphological, and orthographic aspects.

The manual transliteration of Algerian EFL students EAAD corpus revealed the richness of the young users’ repertoire who show flexibility in shifting from Algerian Arabic, into English and French. During this process, the learners do not give much attention to respecting the grammar rules including sentence structure or punctuation, which results in a non-standard form of EAAD. Moreover, it is noted that the participants transferred English language structure into French. For instance, in the case of possessive, the user behaved in the same way as in English when referring to “*d’autres pays langues*” which means “*other countries’ languages*” instead of “*langues d’autres pays*”.

Ambiguity at word level regarding the selection of the Latin character and its diverse correspondences to the Arabic letter may cause confusion from the part of the reader if s/he is not familiar with these characteristics. Nonetheless, it is observed that the Algerian speakers use the preposition proclitic + ف *fa*, and + م *m*+, that are reduced forms of the prepositions /fi/ في ‘in’ and /min/ من ‘from’, and which occur by deleting the vowel. Additionally, the users employ the particle ‘n’ in verb formation as in /nkun/ نكون instead of /akun/ اكون ‘I will be’.

Finally, the non-standard orthographic system of Algerian Arabic dialects resulted in typographical errors, which is a challenge to build a unified Algerian CODA.

6. Conclusion

The present paper reported on the non-standard conventional orthography used by the Algerian young speakers to express themselves in written discourse. This research serves as the first step to understand the functionality of the Algerian CODA. The growing interest in EAAD calls for the generation of larger corpus to allow the generalization of results. Its analysis would achieve more outcomes with the manipulation of machine transliteration, which may enable the construction of a well-defined system that governs EAAD use.

7. Bibliography List:

1. Al-Badrashiny, M., Eskander, R., Habash, N., & Rambow, O. (2014). Proceedings of the Eighteenth Conference on Computational Language Learning, pages 30–38, Baltimore, Maryland USA, June 26-27 2014. © 2014 The Association for Computational Linguistics
2. Arezki, A. (2008). *Le rôle et la place du français dans le système éducatif algérien. Revue du Réseau des Observatoires du Français Contemporain en Afrique*, 23, 21-31.
3. Benrabah, M. (1999a). *Langue et pouvoir en Algérie : Histoire d'un traumatisme linguistique*. Paris: éditions Séguier.
4. Benrabah, M. (2002). Language and modernity in Algeria. In Eliezer Ben- Raphael & Yitzhak Sternberg (eds.) *Identity, Culture and Globalization*, (pp.235-242). Leiden: Brill.
5. Chalabi, A., & Gerges, H. (2012). Romanized Arabic transliteration. In Proceedings of the Second Workshop on Advances in Text Input Methods, pages 89–96, Mumbai, India, December. The COLING 2012 Organizing Committee.
6. Darwish, K. (2013). *Arabizi Detection and Conversion to Arabic*. CoRR.

7. Eskander, R., Habash, N., Rambow, O., & Tomeh, N. (2013). Processing Spontaneous Orthography. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Atlanta, GA
8. Habash, N. (2010). *Introduction to Arabic Natural Language Processing*. Synthesis Lectures
9. Habash, N., Diab, M., & Rambow, O. (2012a). Conventional Orthography for Dialectal Arabic: Principles and Guidelines – Egyptian Arabic. Technical Report CCLS-12-02, Columbia University Center for Computational Learning Systems.
10. Habash, N., Bouamor, H., & Chung, C. (2019). Automatic Gender Identification and Reinflection in Arabic. In Proceedings of the First Workshop on Gender Bias in Natural Language Processing, Florence, Italy.
11. Harrat, S., Meftouh, K., Abbas, M., & Smaïli, K. (2014). *Grapheme to Phoneme Conversion- An Arabic Dialect Case*. In Spoken Language Technologies for Under-Resourced Languages.
12. Harrat, S., Meftouh, K., Abbas, M., Jamoussi, S., Saad, M., & Smaili, K. (2015). *Cross-Dialectal Arabic Processing*. In Computational Linguistics and Intelligent Text Processing (pp.620-632). Springer International Publishing.
13. Hasko, V. (2012). Qualitative corpus analysis. In: *The Encyclopedia of Applied Linguistics* (pp.4758-4164). Hoboken, NJ: John Wiley Online Library.
14. Maamouri, M., Bies, A., Kulick, S., Ciul, M., Habash, N., and Eskander, R. (2014). Developing an Egyptian Arabic treebank: Impact of dialectal morphology on annotation and tool development. In Proceedings of the Ninth International

- Conference on Language Resources and Evaluation (LREC-2014). European Language Resources Association (ELRA).
15. Morsly, D. (1986 a). Multilingualism in Algeria, in Fishman, J., Tabouret-Keller, A. et alii, *The Fergusonian impact. In Honor of Charles A. Ferguson on the occasion of his 65th Birthday* (pp. 253-263), Berlin: Mouton de Gruyter.
 16. Saadane, H. and Habash, N. (2015). A Conventional Orthography for Algerian Arabic. In ANLP Workshop 2015, page 69.
 17. Schiffrin, D., Tannen, D., & Hamilton, H. (2001). *The Handbook of Discourse Analysis*. Oxford : Blackwell.
 18. Taleb Ibrahim, K. (1997). Les Algériens et leurs langues. Éléments pour une approche sociolinguistique de la société algérienne. Alger: Editions El Hikma.
 19. Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *SAGE Publications*, 5, 585–605.
 20. Zaghouani, W., Mohit, B., Habash, N., Obeid, O., Tomeh, N., Rozovskaya, A., Farra, N., Alkuhlani, S., & Oflazer, K. (2014). Large scale Arabic error annotation: Guidelines and framework. In International Conference on Language Resources and Evaluation (LREC 2014).