

Cet article est une publication en ligne avancée. Il faut le citer comme :

Halawi, A., Messarra, N. & HArane, J. (2022). La fréquence de l'alternance codique dans les groupes WhatsApp des étudiants libanais. *Revue canadienne de linguistique appliquée*. Publication en ligne avancée.

La fréquence de l'alternance codique dans les groupes WhatsApp des étudiants libanais

Ayman Halawi

Université Saint-Joseph de Beyrouth

Nasri Messarra

Université Saint-Joseph de Beyrouth

Jarjoura Hardane

Université Saint-Joseph de Beyrouth

Résumé

Les moyens de communication médiée par ordinateur (CMO) et spécifiquement l'application WhatsApp, ont mené à des pratiques langagières innovantes au niveau de la communication écrite. Parmi ces pratiques, le recours à l'alternance codique (AC), qui est définie dans cette étude, comme un passage d'un code écrit à un autre au sein du même message. Cette étude quantitative visait à identifier automatiquement les occurrences de l'alternance codique dans les discussions de groupes WhatsApp durant 14 mois. Nous avons collecté 168 219 messages dans 30 groupes WhatsApp. L'échantillon de l'étude comprend 1 482 étudiants bilingues issus de 7 établissements universitaires libanais. Un outil informatique 'DACA' (détection automatique de l'alternance codique et l'arabizi) a été développé pour détecter la fréquence de ce phénomène résultant du contact des langues. Les résultats montrent que dans le corpus il y a 15 342 occurrences de l'AC soit 9,1% du total des messages. 70,5% de ces ACs sont détectés dans les messages en arabizi et 17,9% dans les messages en anglais, 10,6% dans les messages en arabe et 1% dans les messages en français. Les résultats ont montré aussi que les ACs dans les messages composés en arabizi sont assez souvent vers l'anglais (91,3% du total de ces ACs) et vers l'arabizi dans les messages composés en anglais avec le même pourcentage.

Abstract

The means of computer-mediated communication (CMC) and specifically the WhatsApp application, have led to innovative language practices in written communication. Among these practices is the high frequency of Code-Switching (CS), which is defined in this study as a switch from one written code to another within the same message. This quantitative study aims to automatically identify occurrences of Code-Switching in WhatsApp group chats. All through 14 months, we collected 168 219 messages from 30 WhatsApp groups. The study sample encompasses 1 482 bilingual students from 7 Lebanese universities. A computer tool "DACA" (automatic detection of Code-Switching and arabizi) has been developed to detect the frequency of this phenomenon resulting from languages contact. The results show that in the corpus, there are 15 342 occurrences of CS or 9,1% of the total number of messages. 70,5% of these CS occurrences are detected in messages in Arabizi, 17,9% in messages in English, 10,6% in messages in Arabic and 1% in messages in French.

The results also reveal that CS in messages composed in Arabizi are quite often towards English (91,3% of the total number of these CS occurrences) and towards Arabizi in messages composed in English with the same percentage.

La fréquence de l'alternance codique dans les groupes WhatsApp des étudiants libanais

Le comportement langagier sur le plan individuel ou collectif a été un sujet abordé par de différentes disciplines et selon plusieurs perspectives. Lüdi et Py (2003) cités dans Ali-bencherif (2009) considèrent que « *dans le monde d'aujourd'hui, le plurilinguisme est le plus souvent la règle que l'exception.*[...] ». (Ali-bencherif, 2009, p. 43) Beaucoup de chercheurs ont essayé d'étudier les motifs qui poussent un locuteur à utiliser deux langues ou alterner leurs usages à un moment donné de la conversation (Auer, 1998; Bassam, 2017; Heller et Pfaff, 1996; Ndlovu, 2015; Rodrigues, 2012, et Woolard, 2004). Le phénomène de l'alternance codique, qui signifie le passage d'une langue à une autre au cours de la même conversation (voir Poplack, 1980), a été perçu comme un résultat du bilinguisme voire du plurilinguisme selon Mati (2013). Elle parle d'une compétence plurilingue qui permet « *de passer d'une langue à l'autre dans de nombreuses situations si cela est possible ou nécessaire, même avec une compétence considérablement asymétrique* ». (p. 19)

Avec l'usage accru des moyens de CMO à partir des années 1990s, l'AC au sein d'une conversation est devenue une source de créativité qui aboutit à de nouvelles formes du langage. Selon Melouah et Maiche (2017), cette pratique multilingue contribue à la création d'un nouveau langage : « *Les conversations multilingues procurent de la liberté productive où l'interaction n'est pas réduite à une réponse, mais à un acte de création fructueux qui s'ouvre sur un néo-langage multilingue et universel* ». (p. 260)

Au Liban, le terrain de la présente étude, le biplurilinguisme¹ dans l'enseignement, rend le parler libanais riche en emprunts et en AC. Le français et l'anglais, en tant que langues de l'enseignement scolaire et universitaire, s'intègrent de plus en plus dans le dialecte parlé par les Libanais. Des mots tels que : bonjour, bonsoir, hi, salut, sorry, pardon, thank you, take care, entre autres sont présents dans la communication quotidienne même entre des monolingues. De plus, les jeunes étudiants biplurilingues libanais alternent les codes (arabe, français et/ou anglais) dans leurs conversations d'une façon perçue comme curieuse : « *L'observation d'alternances codiques dans leurs discours est curieuse car ils partagent, parlent et comprennent une même langue maternelle* » (Kanaan 2011, p. 43).

Également, les emprunts et les ACs sont fréquents au niveau de l'écrit dans le contexte de la CMO : « *Concrètement, écrire un message par exemple reviendrait pour nos compatriotes à choisir dans trois lexiques et dans trois syntaxes, les mots et les constructions selon des combinaisons répondant aux visées communicationnelles* » (Kanaan 2011, p. 44). Le parler plurilingue est perçu par un grand pourcentage de la

¹ D'après le dictionnaire de linguistique, le bilinguisme est défini comme une situation linguistique dans laquelle les sujets parlants sont conduits à utiliser alternativement, selon les milieux où les situations, deux langues différentes. Cependant, un sujet parlant est plurilingue quand il utilise à l'intérieur d'une même communauté plusieurs langues selon le type de communication.

population libanaise comme un signe de culture, d'éducation, de modernisation et de l'appartenance à une classe sociale aisée. Pourtant, un autre point de vue considère les mots empruntés et les ACs comme une menace à l'identité culturelle arabe (Hoteit, 2011). Selon Yaghan (2017), les tenants de l'arabisation considèrent que les mots étrangers sont devenus un substitut linguistique national de leurs synonymes existants en arabe.

Recension des écrits

Dans cette section, il est question de la définition de la notion de l'alternance codique, l'application WhatsApp qui a été l'outil de collection du corpus et l'arabizi qui se présente comme une pratique d'écriture en ligne bien présente dans le contexte libanais.

L'alternance codique

Depuis des décennies, l'AC a été le sujet de nombreuses études spécialement dans les recherches sur le bilinguisme, l'apprentissage et le contact des langues. Ce terme est utilisé pour décrire le passage d'une langue à une autre au cours de la même conversation (Auer, 1988; Causa, 2007; Green et Wei, 2016; Grosjean, 2010; Myers-Scotton, 2003; Poplack, 1980).

Cougnon et Thomas (2010), de leur côté, considèrent que dans l'AC, il s'agit de séquences de mots de longueur variable provenant de deux langues distinctes, juxtaposées au sein du même échange verbal tout en montrant tous les signes d'un échange monolingue et en respectant les règles grammaticales des deux langues.

D'autres chercheurs lient entre ce phénomène et le contact des langues et considèrent que l'AC apparaît tout naturellement lors de la rencontre de plusieurs langues. Dans ce contexte, Grosjean (2010) avance que l'AC résulte d'une interaction entre des bilingues qui partagent les mêmes langues. Sa définition se base sur une approche linguistique syntaxique comme celle de Poplack (1980) : « L'alternance codique est l'usage alternatif de deux langues, c'est-à-dire que le locuteur passe complètement à une autre langue pour un mot, une expression ou une phrase, puis revient à la langue de base » (Grosjean, 2010, p. 51).

Bien que la majorité des recherches se soit concentrée sur l'aspect oral de l'AC, il semble que les pratiques d'écriture en ligne dans le contexte de la CMO rapprochent l'écrit de l'oral (Lanchantin et coll., 2014). Les nouveaux moyens de la CMO sont à l'origine de nouvelles pratiques avec lesquelles « *la frontière existante entre l'oral et l'écrit s'estompe progressivement et, de ce fait, parler et écrire deviennent deux catégories continues* » (Fiore, 2016, cité dans Montesino, 2017). De même, Grebeshkova (2017) constate que le développement des moyens de communication est à la base d'un intérêt croissant à étudier ce phénomène au niveau de l'écrit:

« L'alternance codique n'est pas qu'un phénomène oral. Elle existe bel et bien dans des formes écrites. Avec le progrès technologique, beaucoup de données multilingues écrites sont devenues accessibles pour les chercheurs sur Internet, ce qui a donné lieu à des études sur l'alternance codique dans la correspondance par courriel, les blogs, les journaux en ligne, etc. » (Grebeshkova, 2017, p. 1)

Sans doute, le phénomène de l'AC a été étudié par des chercheurs provenant de différentes disciplines et selon des approches divergentes et parfois convergentes, d'où une littérature abondante, très riche en terminologie pour décrire, définir, modéliser et comprendre ce phénomène résultant du contact des langues. Il importe de présenter un tableau récapitulatif de plusieurs typologies de l'AC selon quelques différentes approches avec les objectifs de chacune d'entre elles.

Tableau 1*Les typologies de l'AC selon les différentes approches*

Approche	Auteur(es)	Typologie	Objectif(s)
Linguistique	Shana Poplack	- Les alternances intraphrastiques - Les alternances interphrastiques - Les interjections (Tag-switchings)	- dégage les règles formelles dans les segments mixtes - détecte les contraintes qui gouvernent l'alternance codique
	John Gumperz	- We code - They code	- Analyse les effets de contact de langues - étudie les fonctions conversationnelles et pragmatiques des alternances codiques
Sociolinguistique	Peter Auer	- L'alternance codique relative au discours (discourse-related CS). - L'alternance codique relative-au-participant (participant-related CS)	- Étudie les interactions bilingues dans le cadre de la théorie de l'analyse conversationnelle
	Carol Myers-Scotton	- Les alternances marquées - Les alternances non marquées	- (La théorie de la marque) étudie les motivations de l'alternance codique en fonction du locuteur et de l'activité langagière

Dans le contexte de cette étude, nous proposons la définition suivante de l'AC : c'est un passage d'un code écrit à un autre, par un destinataire au sein du même message dans une discussion menée en ligne. La détection automatique des ACs se limite à l'intérieur des messages, ce qui ressemble aux alternances intraphrastiques de Poplack (1980). Nous négligeons les alternances interphrastiques (entre les messages, c'est-à-dire quand un membre d'un groupe WhatsApp compose un message en langue A, un autre membre lui répond en ayant recourt à une langue B) parce que leur détection présuppose une analyse de contenu pour toutes les discussions ce qui est coûteux en matière de temps dans un grand corpus.

Beaucoup de chercheurs ont constaté que l'AC remplit plusieurs fonctions à savoir : combler un manque de vocabulaire dans une langue, utiliser le mot approprié, montrer l'implication de l'auteur dans une conversation, marquer l'identité, exclure quelqu'un, montrer l'expertise linguistique, souligner l'importance de certains mots ou expression (Akeel, 2016; Altarriba et Basnight-Brown 2009; Halim et Maros, 2014; Heredia et Altarriba, 2001; Kim, 2006; Myers-Scotton, 2003; Urback, 2007; Woolard, 2004).

Concernant la fréquence de l'alternance codique dans les messages écrits, il semble qu'elle ait commencé à accroître avec la communication par SMS au début des années 1990s (Pekarek-Doehler, 2011). Androutsopoulos (2010) considère que le phénomène de l'AC est bien présent dans le contexte des publicités et de l'écriture journalistique. Mais pour elle, le contexte de la CMO permet de détecter une fréquence élevée de cette pratique linguistique. De plus, elle précise que les modes synchrones de CMO contiennent plus d'AC conversationnelle que dans son mode asynchrone.

WhatsApp

WhatsApp Inc. a été fondée en 2009 par Jan Koum et Brian Acton à Santa Clara en Californie pour le système opératoire IOS de iPhone. Au mois d'août 2009, une nouvelle version de WhatsApp a été lancée contenant, pour la première fois la messagerie instantanée. Très rapidement, elle devient disponible pour le système Android et pour Windows mobile, BlackBerry et Nokia. En février 2014, Facebook inc. a acheté WhatsApp pour la somme de 19 milliards de dollars (Seufert et coll., 2016).

Montag et coll. (2015) soulignent le rôle de WhatsApp dans la dépendance aux téléphones intelligents. Ils la définissent comme « l'application de messagerie instantanée sur téléphone mobile la plus utilisée dans le monde ». Parler sur WhatsApp est devenu une façon de communiquer, et non pas simplement un moyen de communication entre deux ou plusieurs personnes (Montesino, 2017). Cette application offre la possibilité de créer un groupe, le nommer et sélectionner ses membres (Ghliss et Jahjah, 2019). Ces discussions de groupe constituent le point fort de WhatsApp (Seufert et coll., 2016). Les records des conversations sur WhatsApp ont constitué les corpus de nombreux projets de recherche spécifiquement pour analyser le phénomène de l'AC (Ndlovu, 2015; Nurhamidah, 2017; Pérez-Sabater, 2015). Au Liban, terrain de notre étude, WhatsApp est en tête du classement des applications de messagerie en ligne les plus utilisées selon le rapport des réseaux sociaux arabes (TNS, 2015).

Arabizi

Dans les années 90 et sous l'effet de la mondialisation, le Liban comme la plupart des pays arabes a été sous l'effet d'un "boom technologique" dans plusieurs domaines. L'anglais a acquis plus d'importance en dominant les nouveaux moyens de communication. L'Internet Relay Chat (IRC), la messagerie via téléphone mobile (SMS) et les courriels ne supportaient pas à l'époque le script arabe ce qui a conduit à la romanisation de l'arabe en utilisant l'alphabet latin (Allehaiby, 2013). Dans ce contexte, Bortzmeyer (2012) note qu'au début de l'année 2011, le système opérateur d'Android 9 ne gérait pas le script arabe ni son système d'écriture de droite à gauche.

L'arabizi est un terme qui décrit un système d'écriture de l'arabe en utilisant des caractères latins et des chiffres pour désigner les consonnes de l'arabe qui ne figurent pas dans l'alphabet latin (Bianchi, 2012; Yaghan, 2008). L'intégration des chiffres dans l'écriture pour désigner des lettres ou des syllabes a commencé avec la communication par SMS comme dans "2mr" pour "demain" dans le but d'abrégier la longueur du message qui a un nombre limité de caractères (Bernicot et coll., 2015). Il se peut que cette pratique ait inspiré les internautes arabophones pour compenser l'absence de quelques consonnes arabes de l'alphabet latin.

L'écriture en arabizi se caractérise par son aspect non standard guidé par la prononciation et très sensible aux variations dialectales (van der Wees et coll., 2016). À leur tour, Bies et coll. (2014) soulignent l'impact des langues étrangères sur les déviations par rapport à l'orthographe standard. Par exemple les anglophones écrivent le phonème /ʃ/ en arabizi avec "s" et "h" alors que les francophones le font avec "c" et "h". Ainsi Darwish (2013) souligne l'importance du mot en contexte pour distinguer entre certains mots arabes et anglais partageant une orthographe commune comme la préposition من (de) qui s'écrit en arabizi (men). Il affirme aussi que l'arabizi utilisé souvent pour écrire l'arabe dialectal sur les réseaux sociaux est mêlé à l'anglais. Eskander et coll. (2014) expliquent ce phénomène par l'écriture des deux codes en caractères latins, le bilinguisme caractéristique d'un haut pourcentage des locuteurs arabes et le statut diglossique de cette langue (arabe standard et dialectal). La présente étude constitue une occasion pour savoir si cette pratique d'écriture en arabizi dans les autres pays arabes est semblable à son usage dans le contexte libanais.

Méthodologie

La nature informelle et non standard de l'arabizi, qui est souvent mêlé à l'anglais et au français et même au script arabe dans les discussions de groupes WhatsApp collectées pour cette étude, constitue un vrai défi aux chercheurs travaillant sur le traitement automatique de la langue et plus spécifiquement sur le développement des systèmes de reconnaissance de la parole, de translittération de l'arabizi, d'intelligence artificielle, voire de traduction automatique.

La question principale à résoudre dans cet article est la suivante : sera-t-il possible de détecter automatiquement les occurrences de l'AC dans un corpus mixte composé de messages échangés via WhatsApp et contenant de l'arabizi avec son aspect informel et non standard ?

- Dès lors, à partir de cette question nous pouvons poser d'autres questions de recherche :
- Quelle est la fréquence de l'AC dans le corpus ?
- La fréquence de l'AC dans les messages en arabizi sera-t-elle plus élevée que dans les messages composés en autres codes d'écriture ?
- Les ACs sont-elles le plus souvent vers l'anglais ?

Pour répondre à ces questions, nous allons détecter à l'aide du logiciel DACA les occurrences de l'AC dans chacun des 4 codes d'écriture (arabizi, anglais, français et arabe) dans tout le corpus. Les réponses à ces questions pourraient aider à mieux comprendre les pratiques linguistiques des étudiants, guider vers les meilleures stratégies de communication avec eux et comprendre si un 'nouveau langage' est en train de se développer dans les différentes applications de la communication médiée par ordinateur. Une approche quantitative statistique est adoptée, ce qui offre la possibilité d'obtenir des résultats représentatifs basés sur un grand échantillon.

Pérez-Sabater (2015) explique que l'utilisation des corpus et des logiciels rend l'analyse du langage indépendante et offre des données empiriques. D'après lui, les études quantitatives statistiques des corpus permettent d'identifier les principaux modèles (patterns) du langage. Donc, l'analyse quantitative statistique nous permettra de donner une vue d'ensemble de notre corpus constitué de conversations de groupes WhatsApp entre étudiants universitaires. L'analyse quantitative statistique présente ici le comportement linguistique (l'AC) en chiffres (les occurrences).

Le corpus

Des discussions dans 30 groupes WhatsApp ont été collectées durant une période de 14 mois (entre octobre 2017 et décembre 2018) auprès de 7 établissements universitaires, dont 1 seul public (l'université libanaise) et six autres privés. Les groupes sont constitués par les étudiants pour des fins de divertissement ou de communication entre eux et parfois avec leurs enseignants. Un numéro de téléphone et une tablette ont été utilisés pour recueillir les données de l'étude. Pour des questions d'ordre éthique, les participants ont consenti à participer à l'étude après leur avoir expliqué le but du projet et la façon selon laquelle les données seraient stockées et utilisées, sans oublier d'affirmer l'engagement à crypter les données identitaires (nom, numéro de téléphone), de sorte que toutes les données personnelles soient anonymisées et toute indication permettant de reconnaître un utilisateur sera obfusquée. Les discussions ont été exportées à l'aide d'un courriel dans un fichier en format texte (TXT). Les messages automatiquement générés par le système signalant la création de groupes, l'ajout d'un nouveau membre, les membres quittant le groupe, etc.) ont été exclus de l'analyse de l'alternance codique. De même, les émojis, GIF, memes, etc. ont été filtrés parce qu'ils ne font pas l'objet de la présente étude. Le nombre total de messages après le filtrage était 168 219 messages collectés de 30 groupes incluant 1 482 étudiants dans 7 universités :

Tableau 2

La répartition des messages du corpus et le compte des membres des groupes par université

Université	Nombre de groupes	Nombre de messages	Nombre d'étudiants	Moyenne : message par étudiant
LAU	3	14 189	81	174,17
LIU	13	17 220	312	55,19
LU	6	86 932	773	112,46
MU	4	26 091	155	168,32
PU	1	20 924	17	1230,82
USJ	2	872	27	32,29
AUB	1	1 991	117	17
Grand total	30	168 219	1482	113,5

Analyse du corpus

Le lexique utilisé dans toutes les conversations constituant le corpus a été automatiquement détecté par TXM “ une plateforme logicielle code ouvert pour la textométrie ”. Le lexique constitue principalement un mélange de l’arabe, du français et de l’anglais, sans oublier la forte présence de l’arabizi comme une pratique quotidienne de l’écriture en ligne au Liban. Ensuite, le lexique a été trié manuellement pour obtenir 4 listes séparées en arabizi, anglais, français et arabe. Avec l’aide d’un programmeur et grâce à une série d’expressions régulières, un outil informatique ‘DACA’ (détection automatique de l’alternance codique et l’arabizi) a été développé pour identifier automatiquement les ACs en suivant les étapes suivantes :

- Diviser le contenu des messages en mots dans des colonnes sur Excel.
- Introduire les 4 listes des lexiques en tant que pages Excel et identifier les mots composant les messages selon ces listes.
- Utiliser une formule pour convertir les mots en code².
- Effacer les colonnes correspondantes à la formule (Wrd11, Wrd12.Wrd13, Wrd14) et copier les données dans un nouveau fichier pour améliorer la performance du logiciel.
- Identifier la langue d’écriture par mot (word level) dans un nouveau fichier.
- Calculer le total des mots selon la langue d’écriture dans chaque message.
- Calculer le pourcentage des langues d’écriture dans chaque message.
- Déterminer la langue de base dans chaque message ‘basic language’³.
- Ajouter une colonne ‘langue d’écriture’ (typing language) pour distinguer entre les messages ‘mono’ et ‘mixte’⁴

² Un code (informatique) est défini selon le dictionnaire Larousse en ligne comme un ensemble d’instructions en langage machine ou symbolique constituant un programme.

³ La détermination de la langue de base ‘ basic language’ dans chaque message permettra de quantifier et déterminer la fréquence de l’AC vers chacun des 4 codes d’écriture.

⁴ Nous avons désigné par mixte ‘mixed’ les messages tapés avec 2 codes et par ‘mono’ les messages écrits avec un seul code.

- Détecter les ACs sur le plan intraphrastique (à l'intérieur de chaque message)⁵.
- Trier et regrouper les occurrences de l'AC par université.

Résultats

Le tableau récapitulatif ci-dessous représente les résultats obtenus (la fréquence des ACs) après l'analyse du corpus par DACA au niveau des universités et montre également la somme des occurrences de l'AC de la langue de base du message vers les trois autres codes d'écriture. Dans le tableau 3, l'arabizi est désigné par (A), l'anglais par (E), le français par (F) et l'arabe par (Ar).

Dans le corpus composé de 168 219 messages, 15 342 occurrences de l'AC ont été détectées dont 10 820 dans les messages en arabizi, 2 743 dans les messages en anglais, 154 dans les messages français et 1 625 dans les messages en arabe.

La distribution des pourcentages dans les messages où il y a eu des ACs est comme suit : 70,5% dans les messages tapés en arabizi, 17,9% dans les messages en anglais, 1% dans les messages en français et 10,6% dans les messages en arabe. Le pourcentage des messages dans lesquels il ya eu des ACs est 9,1% du total des messages envoyés dans tous les groupes qui composent le corpus.

Dans les messages en arabizi

9 847 occurrences de l'AC vers l'anglais (91,3%), 754 vers le français (7%) et 192 vers l'arabe (1,7%).

Dans les messages en anglais

2 504 occurrences vers l'arabizi (91,3% des ACs), 169 vers le français (6,2%) et 70 vers l'arabe (2,5%).

Dans les messages en français

78 occurrences vers l'arabizi (50,5%), 75 vers l'anglais (48,7%) et 1 vers l'arabe (0,6%).

Dans les messages en arabe

369 vers l'arabizi (22,7%), 1,041 vers l'anglais (64,1%) et 215 vers le français (13,2%).

⁵ Selon l'approche qui se base sur *l'analyse conversationnelle*, l'AC est étudiée en fonction du choix de la langue à chaque tour de parole. Nous avons exclu cette approche parce qu'elle présuppose la détermination des frontières de chaque conversation dans les groupes WhatsApp, ce qui est couteux en termes du temps dans un grand corpus pareil à celui de la présente étude.

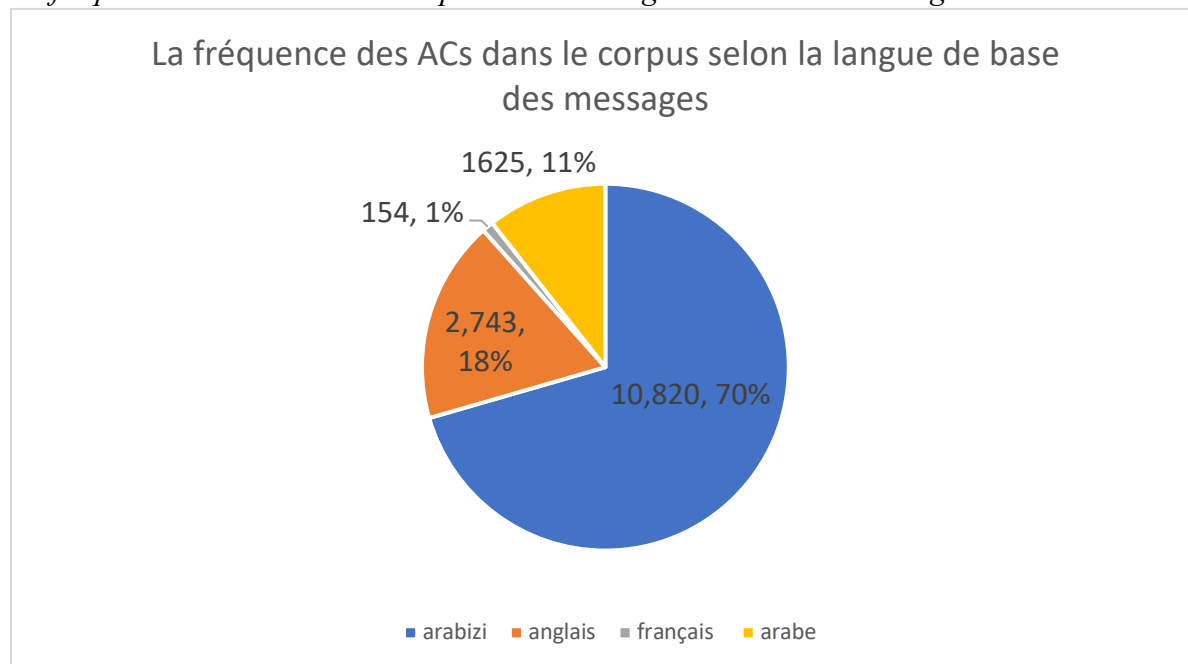
Tableau 3*La fréquence des alternances codiques dans les universités selon la langue de base des messages*

Université	Messages / Université	AC / messages en arabizi				AC / messages en anglais				AC / messages en français				AC / messages en arabe				AC par université
		A / E	A / F	A / Ar	Total	E / A	E / F	E / Ar	Total	F / A	F / E	F / Ar	Total	Ar / A	Ar / E	Ar / F	Total	
AUB	1991	246	14	1	261	101	2	0	103	0	0	0	0	0	0	0	0	364
LAU	14189	414	2	4	420	1387	117	36	1540	3	4	0	7	0	0	0	0	1967
LIU	17220	1948	112	16	2076	365	13	3	381	1	0	0	1	9	14	4	27	2485
LU	86932	3649	498	114	4261	313	14	17	344	9	4	0	13	263	657	190	1110	5728
MU	26091	475	13	9	497	50	2	9	61	0	0	0	0	69	264	18	351	909
PU	20924	3133	108	48	3289	263	2	4	269	1	0	0	1	28	106	3	137	3696
USJ	872	9	7	0	16	25	19	1	45	64	67	1	132	0	0	0	0	193
Total	168219	9874	754	192	10820	2504	169	70	2743	78	75	1	154	369	1041	215	1625	15342

La figure 1 montre que la fréquence la plus élevée de l'AC a été détectée dans les messages en arabizi (70% des occurrences). L'arabizi écrit à l'aide du clavier latin facilite le passage vers d'autres langues ou codes d'écriture.

Figure 1

La fréquence des ACs dans le corpus selon la langue de base des messages



Discussion et conclusion

Les résultats ont montré qu'il y a eu des ACs dans 9,1% du total des messages composant le corpus de l'étude. La fréquence de l'AC dans les messages en arabizi est plus élevée que dans les messages composés en d'autres codes d'écriture (70,5% du total des ACs). Les ACs dans les messages en arabizi sont le plus souvent vers l'anglais (91.3%).

Les résultats de la présente étude sont, en effet, harmonieux avec ceux d'autres études menées dans des pays arabophones ou bien parlant d'autres langues du monde. Ces études ont montré que les mots anglais envahissent la CMO résultant du contact entre les langues locales et l'anglais global (Abdulbari et coll., 2018; Bianchi, 2012; Eskander et coll., 2014; Ndlovu, 2015). David et coll. (2009) considèrent que l'anglais gagne de plus en plus du terrain sous l'effet de la mondialisation et l'utilisation accrue des moyens de la CMO. Pour Bianchi (2012), le code mixte composé de 'l'arabe vernaculaire' (l'arabizi) et l'anglais pourrait produire de nouvelles formes du langage. De même, Melouah et Maiche (2017) soutiennent que l'alternance de codes au sein d'une conversation dans les moyens de CMO, est devenue une source de créativité procurant de la 'liberté productive'. Les moyens de CMO constituent pour Martin (2017) l'un des supports facilitant l'usage de mots anglais surtout par les jeunes dans toutes les langues du monde. De même, Darwish (2013) affirme que la fréquence des ACs dans les échanges en ligne écrits en arabizi est très élevée, ce qui rejoint les résultats de la présente étude (70,5% des ACs ont eu lieu dans les messages en arabizi). Eskander et coll. (2014) le justifient par l'écriture des deux codes

avec le clavier latin. Dans sa thèse dans laquelle il a analysé un corpus composé de gazouillis, Sullivan (2017) considère qu'une partie importante des gazouillis écrits en arabizi impliquait également des ACs dans lesquels les twitteurs ont utilisé plus d'un code. Dans 37% des gazouillis de son corpus, Sullivan a détecté des alternances entre l'arabizi et une autre langue. Identiquement à nos résultats, il avance que la fréquence des ACs la plus élevée était entre l'arabizi et l'anglais (95% du total des gazouillis en arabizi dans lesquels il y a eu des ACs). Ainsi les résultats de Bies et coll. (2014) se rapprochent des nôtres. Dans son corpus de plus de cent mille messages SMS de 26 participants arabes égyptiens, 19% des SMS étaient composés de 2 codes d'écriture différents ce qui pouvait constituer des occurrences de l'AC qu'ils ont nommées 'un mélange' de plusieurs codes. Une autre étude en Égypte par Eskander et coll. (2014) a constaté que les échanges en ligne sur les réseaux sociaux écrits en arabizi étaient riches en emprunts et en alternance codique. La plupart de ces occurrences dans leur corpus composé de messages en arabizi collecté en Égypte étaient en anglais.

Il nous semble que l'alternance de codes vers l'anglais est appréciée dans les pays arabophones et spécifiquement au Liban par les nouvelles générations⁶. C'est ce qu'a montré l'étude d' Abdulbari et coll. (2018) sur l'AC conversationnelle par les étudiants yéménites. De même, dans une autre étude à Oman sur l'attitude des enseignants et des étudiants vis-à-vis de l'AC sur les réseaux sociaux, Al-Qaysi et Al-Emran (2017) ont conclu que la fréquence des ACs dans les discussions WhatsApp était la plus élevée par rapport aux autres réseaux sociaux et que ce comportement linguistique avait une connotation positive par les étudiants.

Pour conclure, l'arabizi permet aux étudiants de composer leurs messages en se servant de leur dialecte parlé (l'arabe libanais) tout en alternant avec l'anglais, la langue de la mondialisation, de modernisation et de l'éducation, sans le besoin de changer de clavier puisque les 2 codes s'écrivent à l'aide du clavier latin. Alors dans les messages tapés en anglais, les ACs sont assez souvent vers l'arabizi. Donc, quand les étudiants composent leurs messages en choisissant (l'arabizi ou l'anglais), ils alternent vers l'autre code. L'écriture en arabizi en alternant assez souvent avec l'anglais pourrait constituer avec les émojis un miroir de la langue parlée par les nouvelles générations, surtout les étudiants. De même, pour les messages tapés en français, les ACs sont soit vers l'arabizi soit vers l'anglais. Tandis que dans les messages composés en arabe, les étudiants changent de clavier pour alterner le plus souvent avec l'anglais puis avec l'arabizi et enfin avec le français. Mais cette pratique ralentit la communication puisqu'elle présuppose le changement du clavier en passant du clavier latin vers le clavier arabe. Ne pas pouvoir trouver un équivalent d'un mot en arabe pourrait, dans certaines occurrences, justifier cette alternance de code. La CMO est en quelque sorte une simulation pour reproduire au niveau de l'écrit la communication de face-à-face.

Dans le futur, il sera utile de tester notre outil informatique 'DACA' sur d'autres corpus pour améliorer sa performance, et comparer les résultats de cette étude avec d'autres réseaux sociaux (Twitter, Facebook, Instagram...).

⁶ Les systèmes scolaire et universitaire au Liban sont parfaitement bilingues. Le français et l'anglais, qui viennent en seconde place après l'arabe, sont compris à des degrés différents par un grand pourcentage de la population et surtout par les jeunes. Ce sont les langues d'enseignement dans tous les établissements scolaires à côté de l'arabe.

Références

- Abdulbari, A., Abdulmalik, A., Abdulraheem, M., & Alsabri, S. (2018). Types of code-switching between Yemeni dialect and English language among Yemeni undergraduate students at university of Sheba region, *Language in India* 18(49042), 1–10.
- Akeel, E. S. (2016). Investigating code switching between Arabic/English bilingual speakers. *English Linguistics Research* 5 (2). <https://doi.org/10.5430/elr.v5n2p57>
- Ali-Bencherif, M. Z. (2009). *L'alternance codique arabe dialectal / français dans des conversations bilingues de locuteurs algériens immigrés / non-immigrés*. Thèse de doctorat, Université Abou Bakr BELKAÏD de Tlemcen (Algérie).
- Allehaiby, W. H. (2013). Arabizi: An analysis of the romanization of the Arabic script from a sociolinguistic perspective. *Arab World English Journal*, 4(3).
- Al-Qaysi, N., & Al-Emran, M. (2017). Code-switching usage in social media: A case study from Oman. *International Journal of Information Technology and Language Studies*, 1(1), 25–38.
- Altarriba, J., & Basnight-Brown, D. M. (2009). Empirical approaches to the study of code-switching in sentential contexts. *Multidisciplinary approaches to code switching*, 3–25.
- Androutsopoulos, J. (2013). Code-switching in computer-mediated communication. In *Pragmatics of computer-mediated communication* (pp. 667–694). de Gruyter Mouton.
- Auer, P. (1988). A conversation analytic approach to codeswitching and transfer. In *Codeswitching: Anthropological and Sociolinguistic Perspectives*, 187–213. Berlin: Mouton de Gruyter.
- Auer, P. (1998). Introduction: Bilingual conversation revisited. In *Code-Switching in conversation, Language interaction and identity* (pp. 1–24). Routledge, Taylor & Francis Group. London and New York.
- Bassam, L. (2017). *Gender Differences in SMS Code-Switching by Lebanese Undergraduates* (Doctoral dissertation, Universitat Rovira i Virgili).
- Bernicot, J., Bert-erboul, A., Goumi, A., & Volckaert-legrier, O. (2015). Analyse d'un corpus longitudinal de SMS produits par de jeunes adolescents. *Revue Tranel (Travaux Neuchâtelois de Linguistique)* 63, 1, 15–29.
- Bianchi, R. M. (2012). 3arabizi-When local Arabic meets global English on the Internet. *Acta Linguistica Asiatica*, 2(1), 89–100.
- Bies, A., Song, Z., Maamouri, M., Grimes, S., Lee, H., Wright, J., Strassel, S., Habash, N., Eskandar, R. & Rambow, O. (2014). Transliteration of arabizi into arabic orthography: Developing a parallel annotated arabizi-arabic script sms/chat corpus. In *Proceedings of the EMNLP 2014 workshop on Arabic natural language processing (ANLP)* (pp. 93–103).
- Bortzmeyer, S. (2012). Multilingualism and the internet's standardisation. *trans. John Rossbottom. In NET. LANG: Towards the Multilingual Cyberspace*, ed. Laurent Vannini and Hervé Le Crosnier, 104–17.
- Causa, M. (2007). L'indispensable alternance codique. *Français dans le monde*, 351, 18.
- Cougnon, L. A., & François, T. (2010). Quelques contributions des statistiques à l'analyse

- sociolinguistique d'un corpus de SMS. In *Actes du colloque JADT 2010* (Vol. 1, pp. 619-630).
- Darwish, K. (2013). Arabizi detection and conversion to Arabic. *arXiv preprint arXiv:1306.6755*.
- David, M. K., Cavallaro, F., & Coluzzi, P. (2009). Language policies-impact on language maintenance and teaching: Focus on Malaysia, Singapore, Brunei and the Philippines. *Linguistics Journal*, 4.
- Eskander, R., Al-Badrashiny, M., Habash, N., & Rambow, O. (2014). Foreign words and the automatic processing of arabic social media text written in roman script. In *Proceedings of The First Workshop on Computational Approaches to Code Switching* (pp. 1-12).
- Montesino, J. M. C. (2017). *L'application WhatsApp dans la négociation franco-espagnole: un catalyseur de la confiance interculturelle*. Thèse de doctorat, Université de Nanterre - Paris X.
- Ghliiss, Y. & Jahjah, M. (2019). Habiter WhatsApp? Éléments d'analyse postdualiste des interactions en espace numérique. *Langage et société*, (2), 29-50.
- Grebeshkova, O. (2017). *Written code-switching in the notes of second-language learners in bilingual classroom environments*. Doctoral dissertation, Lyon.
- Green, D. W., & Wei, L. (2016). Code-switching and language control. *Bilingualism: Language and Cognition*, 19(5), 883-884.
- Grosjean, F. (2010). *Bilingual*. Harvard university press.
- Halim, N. S., & Maros, M. (2014). The functions of code-switching in Facebook interactions. *Procedia-Social and Behavioral Sciences*, 118, 126-133.
- Heller, M., & Pfaff, C. W. (1996). Code-switching In H. Goebel et al.(Eds.) *Kontaktlinguistik/Contact linguistics* 1.
- Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch?. *Current Directions in Psychological Science*, 10(5), 164-168.
- Hoteit, S. (2011). *Enseignement-apprentissage du français au sud du Liban: didactique contextualisée et intégration dans une dynamique culturelle francophone?* (Thèse de doctorat, Université Rennes 2; Université Européenne de Bretagne).
- Kanaan, L. (2011). *Reformulations, contacts de langues et compétence de communication: analyse linguistique et interactionnelle dans des discussions entre jeunes Libanais francophones* (Thèse de doctorat, Université d'Orléans).
- Kim, E. (2006). Reasons and motivations for code-mixing and code-switching. *Issues in EFL*, 4(1), 43-61.
- Lanchantin, T., Simoës-Perlant, A., & Largy, P. (2014). Good spellers write more textism than bad spellers in Instant Messaging: The case of French. *PsychNology Journal*, 12.
- Martin, E. (2017). *A sociolinguistic comparison of the French and Anglo-Saxon cultures from codeswitched substantives to borrowings: the issue of grammatical gender* (Doctoral dissertation, Lyon).
- Mati, N. (2013). *Alterner les codes en contexte didactique algérien: aide ou écueil pour l'appropriation de la langue de l'autre? Analyse d'échanges alternés entre enseignant de FLE et sujet* (Thèse de doctorat, Université de Franche-Comté).
- Melouah, S., & Maiche, H. (2017). Analyse des pratiques langagières de jeunes utilisateurs algériens de la messagerie instantanée Facebook. *El-Tawassol: Langues et Littératures* 23 (52): 256–67.

- Montag, C., Błaszkiwicz, K., Sariyska, R., Lachmann, B., Andone, I., & Trendafilov, B. & Markowetz, A. (2015). Smartphone usage in the 21st century: who is active on WhatsApp. *BMC research notes*, 8(1), 331.
- Montesino, J. M. (2017). *L'application WhatsApp dans la négociation franco-espagnole: un catalyseur de la confiance interculturelle* (Thèse de doctorat, Paris 10).
- Myers-Scotton, C. (2003). *Social motivations for codeswitching: Evidence from Africa*. Oxford University Press.
- Ndlovu, A. (2015). Code-switching in WhatsApp chat messages in Botswana. *NAWA Journal of Language & Communication*, 10(1).
- Nurhamidah, I. (2017). Code-Switching in WhatsApp exchanges: Cultural or language barriers?. In *Proceedings Education and Language International Conference* (Vol. 1, No. 1).
- Pekarek-Doehler, S. (2011). Dans la communication par SMS. *Linguistic Online* 48 (4/11): 49–70.
- Pérez-Sabater, C. (2015). Discovering language variation in WhatsApp text interactions. *Onomázein: Revista semestral de lingüística, filología y traducción*, 31(1), 113-126.
- Poplack, S. (1980). Sometimes i'll start a sentence in Spanish y termino en Espanol: toward a typology of code-switching. *Linguistics* 18, pp. 581- 618.
<https://doi.org/10.1515/ling.1980.18.7-8.581>
- Rodrigues, P. (2012). *Processing highly variant language using incremental model selection* (Doctoral dissertation, Indiana University).
- Seufert, M., Hoßfeld, T., Schwind, A., Burger, V., & Tran-Gia, P. (2016). Group-based communication in WhatsApp. In *2016 IFIP networking conference (IFIP networking) and workshops* (pp. 536-541). IEEE.
- Sullivan, N. (2017). *Writing Arabizi: Orthographic variation in romanized Lebanese Arabic on Twitter* (Doctoral dissertation, The University of Texas at Austin).
- T.N.S. (2015). Arab Social Media Report. *Arab Social Media Influencers Summit*, <http://dmc.ae/img/pdf/white-papers/ArabSocialMediaReport-2015.pdf>.
- Urbäck, K. (2007). *Code-switching in computer-mediated communication: The use of Swedish and English in an Internet discussion forum*. Undergraduate Thesis, Växjö University
- van der Wees, M., Bisazza, A., & Monz, C. (2016). A simple but effective approach to improve Arabizi-to-English statistical machine translation. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)* (pp. 43-50).
- Wehbe, O. (2017). *Questions que pose une didactique plurilingue au Liban, pratiques et représentations*. (Doctoral dissertation, Université Sorbonne Nouvelle – Paris3)
- Woolard, K. A. (2004). Codeswitching. *A companion to linguistic anthropology*, 73-94. Blackwell Publishing.
- Yaghan, M. A. (2008). “Arabizi”: A contemporary style of Arabic Slang. *Design Issues*, 24(2), 39-52.