ORIGINAL RESEARCH

A Semi-supervised Approach for Sentiment Analysis of Arab(ic+izi) Messages: Application to the Algerian Dialect

Imane Guellil[1,2] · Ahsan Adeel[3] · Faical Azouaou[1,2] · Fodil Benali[4] · Ala-Eddine Hachani[5] · Kia Dashtipour[6] · Mandar Gogate[6] · Cosimo Ieracitano[7] · Reza Kashani[6] · Amir Hussain[8]

Abstract

In this paper, we propose a semi-supervised approach for sentiment analysis of Arabic and its dialects. This approach is based on a sentiment corpus, constructed automatically and reviewed manually by Algerian dialect native speakers. This approach consists of constructing and applying a set of deep learning algorithms to classify the sentiment of Arabic messages as positive or negative. It was applied on Facebook messages written in Modern Standard Arabic (MSA) as well as in Algerian dialect (DALG, which is a low resourced-dialect, spoken by more than 40 million people) with both scripts Arabic and Arabizi. To handle Arabizi, we consider both options: transliteration (largely used in the research literature for handling Arabizi) and translation (never used in the research literature for handling Arabizi). For highlighting the effectiveness of a semi-supervised approach, we carried out different experiments using both corpora for the training (i.e. the corpus constructed automatically and the one that was reviewed manually). The experiments were done on many test corpora dedicated to MSA/DALG, which were proposed and evaluated in the research literature. Both classifiers are used, shallow and deep learning classifiers such as Random Forest (RF), Logistic Regression(LR) Convolutional Neural Network (CNN) and Long short-term memory (LSTM). These classifiers are combined with word embedding models such as Word2vec and fastText that were used for sentiment classification. Experimental results (F1 score up to 95% for intrinsic experiments and up to 89% for extrinsic experiments) showed that the proposed system outperforms the existing state-of-the-art methodologies (the best improvement is up to 25%).

Keywords Arabizi · Sentiment analysis · Arabic · Arabic dialect · Translation · Transliteration

---

✉ Imane Guellil
   i.guellil@aston.ac.uk

1   Aston University and Folding Space, Birmingham, UK

2   Ecole Nationale Supérieure d'Informatique (ESI), Algiers, Algeria

3   School of Mathematics and Computer Science, University of Wolverhampton, Wolverhampton, UK

4   Capgemini, Paris, France

5   Legal Doctrine, Algiers, Algeria

6   University of Stirling, Stirling, UK

7   University "Mediterranea" of Reggio Calabria, Reggio Calabria, Italy

8   Napier University, Edinburgh, UK

Introduction

Sentiment analysis (SA) helps to analyse people's opinions, sentiments, appraisals, attitudes, and emotions towards entities such as products, services, organisations, individuals, issues, events, topics [1]. Two main approaches are commonly used to determine the valence of documents (i.e. positive or negative): lexicon-based approach [2] and machine learning-based approach (ML) [3]. English has the greatest number of sentiment analysis studies, while research is more limited for other languages, including Arabic [4–6]. There are three different variants of Arabic: Classical Arabic (CA, for Quran), Modern Standard Arabic (MSA, used in formal exchange) and Arabic dialects (AD, used in informal exchange).

Moreover, Arabic can be written in both scripts, Arabic and Arabizi (corresponds to Arabic written with Latin letters, numerals and punctuation [7, 8]). However, one of the

main issues related to the treatment of Arabic and its dialects is the lack of resources. Also, other dominant problems include the non-standard romanisation (called Arabizi) that Arabic speakers often use in social media. Arabizi uses the Latin alphabet, numbers, punctuation for writing an Arabic word. For example, the word "mli7", combined with Latin letters and numbers, becomes the romanised form of the Arabic word "مليح" meaning "good". Due to the challenging problems related to transliteration, most of the ongoing research works are focused on Arabic sentiment analysis written in Arabic script. To the best of our knowledge, only a few works have been presented in the literature on Arabizi sentiment analysis [9, 10] or on Arabic and Arabizi sentiment analysis ([11, 12]). Furthermore, some dialects (such as Egyptian, Gulf or Iraqi (belonging to Mashreq dialects)) are more studied than others; indeed, a few works have been conducted on Maghrebi dialects, such as Morocco or Algerian dialects.

To address the challenges mentioned earlier, this paper proposes a semi-supervised sentiment analysis approach of Arabic messages extracted from social media (i.e. Facebook). The main idea behind this approach is to construct the sentiment corpus automatically and review it manually. The interesting aspect of this approach is that it considers both Arabic and Arabizi. For transforming Arabizi into Arabic, we consider both options, transliteration and translation. The impact of both techniques on sentiment analysis is shown to highlight the most suitable approach to adopt for Arabizi handling. The proposed approach consists of four main steps which are (1) Corpus extraction. (2) Arabizi transliteration. (3) Arabizi translation. (4) Arabic sentiment analysis. At the end of this paper, we aim to answer a set of research questions where each answer opens the door to a research perspective. The research questions addressed in this paper are the following:

1. What is the best option for handling Arabizi (i.e. transliteration or translation)?
2. How could we improve the used transliteration approach?
3. How could we improve the proposed translation approach?
4. What is the best technique for Arabic sentiment analysis (i.e. supervised or semi-supervised approach)?
5. How could we improve the automatic annotation approach?

This paper is organised as follows. The next section presents the different challenges related to Arabic sentiment analysis followed by which the related work done on sentiment corpus construction and the new trends related to Arabic SA are presented. The subsequent section presents the methodology that we follow. Then the different experimentation that we carry out and the different results that we obtained by comparing our results to those obtained in the research literature are presented. Before the concluding section, a discussion and some errors analysis are presented. We conclude by presenting a synthesis and some opening for futures works.

Arab(ic+izi) Sentiment Analysis: Challenges

Most of the works on short text sentiment classification concentrate on Twitter [13–17]. Facebook has more than one billion clients. Facebook users spend approximately 120 minutes, consistently communicating with family and companions [18]. Although Facebook is the biggest social network, only a few approaches targeting Facebook posts and comments have been proposed. This is mainly due to the lack of labelled dataset for such a purpose. Facebook is also a popular social media platform in Arabic countries, where users typically write in Arabic and its dialects. Table 1 illustrates few messages, as examples, extracted from Facebook, highlighting the following characteristics:

– Different Arabic variants are present in Social media (Facebook in particular) including (1) Classical Arabic (CA), message 1; (2) Modern Standard Arabic (MSA), message 2; (3) Arabic Dialects (AD), messages 3 and 4; (4) Arabizi, messages 5 and 6.
– Some messages are written using Arabic script (messages 1, 2, 3, and 4) and others using Arabizi (messages 5 and 6).
– Inappropriate use of punctuation, space, exaggeration and links, as the text in social media, is recognised to be unstructured (messages 3, 4, 6, and 10).
– Code-switching between languages. The combined use of Arabic and English can be seen in Mechrek countries such as Egypt and the Gulf. The combined use of Arabic and French can be seen in Maghreb countries such as Tunisia and Algeria (message 10).
– Code-switching between scripts, where some messages are written using Arabizi and Arabic (message 9).
– A massive use of emoticons (a convenient way to express opinions, sentiments and emotions (message 4)).

In the context of this paper, we are focusing on four of the presented challenges, which are Arabizi, code-switching between Arabic and Arabizi, inappropriate use of punctuation and the extensive use of emoticons.

As presented above, one of the most important challenges behind Arabic sentiment analysis is the use of Arabizi. The challenge behind Arabizi is the presence of many forms of the same word. For example, Cottrell et al. [19] argued that the word ان شَاء الَله (meaning if God is willing) could be written in 69 different manners. Another challenge is related

Table 1 Example messages: Arabic corpus extracted from Facebook

Number	Message	Language and dialects	Specificities	English translation
1	بسم اللَه الرحمَان الرحيم	CA	Arabic script	In the name of Allah(God), The most Gracious, The most Merciful
2	اللَه يرحمهم وجعلهم من اهل الجنة	MSA or Dialects	Arabic script	God bring them peace and make them from the people of paradise
3	حشم رَاك كبيِّر	Algerian dialect	Arabic script = exaggeration	shame on you you are old
4	#زَاني زعفَان :(#	Algerian dialect	Arabic script Emoticon	I am angry # :(
5	n7bha ou t3rf tmathl	Arabizi Maghrebi	Latin letters + numbers	I love her and she knows acting
6	Hahah yea bt7bii tsma3ii la firouz	Arabizi Egyptian	Latin letters + numbers	Hahah yea you like listening to fairou
7	Very beautiful Nancy Ajram	English	Transliterated name "Nancy Ajram"	The same
8	Vous êtes superbe	French	Latin	You are superb
9	mashaAllahجزَاك الله خيرًا	Arabizi+Arabic	Mix script	Mashallah God give you good things
10	That so funny walh hhhhh	English+Arabizi	Mix language + repetition	That so funny I swear hhhhh

to the annotation process. Almost all the works presented in the research literature rely on a manual annotation for the sentiment corpus (used in the training phase) [11, 20–22]. However, manual annotation is time and effort consuming. Some works dedicated to English and Dutch [15, 23, 24] present approaches using emoticons for automatically tag a large corpus. However, relying on emoticons only leads to many errors where some users express a contradictory sentiment between the text and the emotions that they used. More recently, Gamal et al. [25] presented a large sentiment corpus dedicated to MSA and Egyptian dialect. They also relied on a sentiment lexicon for the automatic annotation. However, they used only the sentiment score for the annotation. Also, they carry out only intrinsic experiments (i.e. the constructed corpus was split into train and test corpora). The challenge behind the automatic annotation is to propose an approach combining between the emoticons and text and other features for increasing the annotation precision. For validating the constructed corpus, it would be better to choose an external test corpus for showing the efficiency of the training corpus with real-world examples.

Algerian dialect (DALG) is a Maghrebi Dialect, primarily used in informal communication including social media [26, 27]. DALG is not used in school education or within television news. It is used more in everyday life, music and series broadcast. This dialect is considered as a language of low variety, meaning that DALG is lowly standardised and normalised. DALG has been enriched with the influence of the language of countries colonising the Algerian population. Among these languages: Turc, Italian and more recently French. Hence, DALG resulted from different languages,

including MSA (representing the major part of this dialect). The challenge behind DALG is the lack of works and resources. To the best of our knowledge, in addition to the corpora that we presented for DALG (and that we present in more details in the experimentation part), only three corpora are publicly available for DALG. The first one is Cottrel's corpus [19], which is an Algerian Arabizi corpus extracted from Facebook. The second one is PADIC corpus [28], which is a parallel corpus between MSA and many dialects, including DALG. The last one is SANA_Alg [29] that is a recent, an annotated sentiment corpus (which we are using for our experiments, to evaluate the proposed approach on a test corpus presented in the literature).

Related Works

Arabic Sentiment Analysis

The classification of Arabic messages into two/three main classes (i.e. positive/negative or positive/negatives/neural) is done using two main approaches: lexicon-based approach and corpus-based approach. Both approaches require annotated data. The lexicon-based approach requires an annotated lexicon where each word is annotated as positive/negative/neutral. Some lexicons also contain a sentiment orientation score (generally a number from 0 to 5) estimating the strength of the sentiment. Corpus-based approaches require an annotated corpus were each sentence contains a label (defining if it is positive/negative or neutral). For constructing both lexicons and corpora, three trends are emerging:

(1) manual construction, (2) automatic construction and (3) semi-automatic construction.

Manual Construction

Only few lexicons are constructed manually [30, 31]. The work in [30] described the process of creating SIFAAT, a manually created lexicon of 3325 Arabic adjectives labelled with one of the following tags, positive (Pos), negative (Neg), neutral (Obj). The adjectives in SIFAAT pertained to the newswire domain and were extracted from the first four parts of the Penn Arabic Treebank [32]. In [31], the authors focused on Algerian dialect by constructing three lexicons: (1) keyword lexicon; (2) negation word lexicon; and (3) intensification word lexicon. All these lexicons were constructed manually using existing MSA and Egyptian lexicons. The translation from MSA and Egyptian to Algerian dialect was done manually. The resulted lexicon contains 3093 words where 2380 are positives, and 713 words are negatives.

However, Almost all the corpora were constructed manually [11, 20–22, 33, 34]. In the majority of cases, the annotation is done by natives annotators. In [20], the authors presented OCA, which contains 500 movie reviews collected from different Arabic web pages and blogs in Arabic (250 positive and 250 negatives). The reviews were also manually pre-processed, segmented, and roots were extracted. In [21] the authors presented AWATIF, a multi-genre corpus containing 10,723 Arabic sentences from three sources, namely the Penn Arabic Treebank (ATB) [32], Wikipedia talk pages, and web forums. The sentences are manually annotated as objective or subjective, and subjective sentences are annotated as positive or negative. Authors of [11] presented the TSAC (Tunisian Sentiment Analysis Corpus) corpus. It contains 17,060 Tunisian Facebook comments. These comments were manually annotated, and they include 8215 positive and 8845 negative statements. This corpus was collected from comments written on official pages of Tunisian radios and TV channels. In [22], the authors constructed ASTD, an Arabic Sentiment Tweets Dataset. This corpus contains 10,000 Arabic tweets that are annotated using Amazon Mechanical Turk as objective, subjective positive, subjective negative, or subjective mixed. The corpus presented in [33] is composed of DARDASHA (2798 chat messages from Maktoob[1]), TAGREED (3,015 Arabic tweets), TAHRIR (3,008 sentences from Wikipedia Talk Page), and MONTADA (3,097 Web forum sentences). Two natives speakers manually annotated these corpora. The corpus used [34] contains 2300 tweets that are manually annotated.

Automatic Construction

Almost all the lexicons presented in literature were constructed automatically. To automatically construct an Arabic sentiment lexicon, three tendencies have emerged: (1) construction based on automatic translation [30, 35–39]. (2) Construction based on resources linking [40–43]. (3) Construction based on both translation and resources linking [44, 45]. The main idea behind automatic translation construction is to start with an English sentiment lexicon (i.e. Bing Liu lexicon [46], SentiWordnet [47], SentiStrength [48], etc.) and translate them using Google translate. Some translations are done using an Arabic/English dictionary [45]. In resources linking different existing English/Arabic resources such as Sentiwordnet, Arabic WordNet [49], Arabic Morphological Analyzer [50, 51] are combined. The main idea behind the construction combining automatic translation and the resources linking is to use a reduce seed of English sentiment words, translate them to Arabic and expand them using Arabic Wordnet or Arabic synonyms dictionaries.

Only a few works have been conducted on automatic construction, and two techniques have been used: (1) using rating reviews [52, 53] and (2) using sentiment lexicons [12]. In the context of using rating reviews, [52] presents LABR containing 63,257 book reviews, each rated on a scale from 1 to 5 stars. The authors considered reviews with 4 or 5 stars as positive, those with 1 or 2 stars as negative, and ones with 3 stars as neutral. In [53] the authors fellow the same annotation principle used in [52] for constructing 7 data sets (ATT, HTL, MOV, PROD, RES1, RES2, RES). ATT is a dataset of Attraction Reviews scrapped from TripAdvisor.com, containing 2154 reviews. HTL is a dataset of Hotel Reviews scrapped from TripAdvisor.com too and containing 15,572 reviews. MOV is a dataset of Movie Reviews scrapped from elcinema.com, containing 1524 reviews. PROD is a dataset of product reviews scrapped from souq.com, containing 4272 reviews. RES1 is a dataset of restaurant reviews scrapped from qaym.com containing 8364 reviews. RES2 is a dataset of restaurant reviews scrapped from tripadvisor.com containing 2642 reviews and RES is a combination between RES1 and RES2; hence it contains 10,970 reviews. In the context of using lexicon, the work in [12] create and use an Algerian sentiment lexicon for tagging a large set of MSA and Algerian messages. However, these authors concentrate on a reduced annotated corpus containing only 8000 messages (where 4000 are for Arabic and 4000 for Arabizi).

Semi-automatic Construction

Few works only have been done on semi-automatic construction for both resources (lexicon and corpora) [54–56]. [54] presents NileULex, an Arabic sentiment lexicon containing 45% of Egyptian (EGY) and 55% of MSA. This

---

[1] chat.mymaktoob.com.

lexicon contains 5953 unique terms. [55] presents SANA, a large-scale multi-genre, a multi-dialectal multi-lingual lexical resource for subjectivity and sentiment analysis of the Arabic and dialects. In addition to MSA, SANA also covers both EGY and LEV, along with providing English glosses. A significant portion of SANA entries is also augmented with POS, diacritics, gender and number. SANA is developed both manually and automatically, and it contains 224 564 entries. Finally, in [56] the authors present a Saudi corpus. This corpus contains 17,573 Saudi tweets that were manually reviewed into four classes: positive, negative, neutral and mixed. To construct this corpus, the authors target a set of sentiment words and use them to extract tweets containing these words. After the phase of cleaning and processing, they charge native speakers of Arabic/Saudi to review the constructed corpus manually.

After analysing the presented works using the constructed resources, we conclude that

1. The corpus-based approaches gives better results than the lexicon-based approaches. Also, almost all the recent works are relying on a corpus-based approach.
2. The resources constructed manually give the best results for both lexicon-based and corpus-based approaches. However, the size of the resource is a crucial factor in the quality of the results.
3. The voluminous resources give the best results (mainly where the resources were constructed manually). However, manual construction represents time and effort consuming.
4. Semi-automatic construction seems to be the solution resolving both problems: precision and time/effort consuming. However, only a few approaches were proposed in this category.
5. Almost all the recent work in the research literature rely on word-embedding and deep learning approaches (detailed in the following parts).

Word Embedding and Deep Learning Approaches

In the supervised approach (corpus-based approach), the text is represented as a feature vector. A bag of words (BOW) representation is commonly used, mainly due to its simplicity as well as its efficiency[57]. Despite its popularity, this approach has two significant weaknesses: (1) loss of word order in the sentence, and (2) semantic ignorance of words [58]. Moreover, the application of this approach may require additional pretreatment of data and an appropriate word feature extraction technique [58, 59]. More recently, word and document embedding have emerged as an alternative representation [58–61]. Among the most used word/document embedding methods, those presented in [58–61]. Al-Azani and El-Alfy [59] and Altowayan et al. [61] relied on large

Arabic corpora to train word2vec models [62] to improve sentiment analysis. They generated features and used these features for training different classifiers. Barhoumi in [58] applied doc2vec model [63] for the sentiment classification of the corpus LABR [52]). El Mahdaouy et [60] affirm that using document embeddings improve text classification. All these works are based on Word2vec and Doc2vec. More recently, another algorithm is appearing, which is fastText [64]. As for Word2vec, fastText models are also based on either the skip-gram (SG) or the continuous bag-of-words (CBOW) architectures. fastText is often compared to Word-2vec for the classification task [65, 66]. However, to the best of our knowledge, fastText has not been used for Arabic classification or sentiment analysis.

Recently deep learning algorithms such as convolutional neural network (CNN), long short-term memory (LSTM), bidirectional LSTM (Bi-LSTM), etc. take an essential place for classifying sentiments. In this context, [67] presents a scheme of Arabic sentiment classification, which evaluates and detects the sentiment polarity from Arabic reviews. The authors used Word2vec for features extraction (with Both CBOW and Skip-gram architecture). A convolutional neural network (CNN) was trained on top of pre-trained Arabic word embeddings for sentiment classification. For CNN, the authors used the same architecture defined in [68] relying on one channel that allows the adaptation of pre-trained vectors for each task. They apply their approach to different corpus presented in the literature such as LABR, ASTD, ATT, HTL, and MOV. More recently, [69] present a model (language-independent) for multi-class sentiment analysis using a simple neural network architecture of different layers. The advantage of the proposed model is that it does not rely on language-specific features such as ontologies, dictionaries, morphological or syntactic pre-processing. The authors applied their model for three languages which are: English, German and Arabic. For Arabic, they relied on ASTD corpus constructed in [22].

Arabizi Sentiment Analysis

Limited works have been conducted on Arabizi sentiment analysis [9, 10, 12]. In [9], the authors present a transliteration step before proceeding to the sentiment classification. However, their approach presents two majors drawbacks: (1) they relied on a fundamental table for the passage from Arabizi to Arabic, which cannot handle Arabizi ambiguities. (2) They constructed a small annotated corpus manually (containing 3026 messages). This corpus contains Arabizi messages which therefore transliterated into Arabic. In [12], the authors automatically construct an annotated sentiment Arabizi corpus and directly applied sentiment classification without calling the transliteration/translation process. However, the authors confronted several ambiguity

problems which resulted in low F1 score of 66%. The same test corpus used in [12] was also used in [10], where the authors improved the results by calling a transliteration step. The authors used a large sentiment corpus constructed automatically by relying on a sentiment lexicon (also constructed automatically [39]). The results were up to 76% for automatic transliteration and up to 78% for manual transliteration.

Hence, it can be seen that for handling Arabizi, two trends are emerging: 1) considering the Arabizi as a proper language and rely on an annotated Arabizi corpus. 2) Transliterating Arabizi into Arabic and rely on the transliterated annotated corpus. Many works have been proposed to transliterate Arabizi to Arabic. Some of them consider a set of rules [10, 70, 71]. Others rely on a parallel corpus (Arabizi/Arabic) and consider the transliteration task as a translation task at a character level [7, 72, 73]. The usefulness of transliteration was shown and illustrated in different researches. Almost all the annotate sentiment corpora are in Arabic (not in Arabizi). Then, behind transliteration, we are aiming to transform Arabizi into Arabic. However, another way could lead to Arabic, the translation. Although the translation allows us to transform Arabizi into Arabic, no research work considers this way. In this paper, we consider this new perspective for handling Arabizi, which involves machine translation. Although no work was proposed for Arabizi sentiment analysis after translation, many works were proposed for Arabic machine translation. Some works also consider the effect of translation of the sentiment analysis results. The following part briefly describes some of these works.

Arabic Translation and Sentiment Analysis

During the last decades, several approaches have been proposed for translating Arabic to and from other spoken languages [74–76]. Arabic is also considered as a pivot for many works concentrating on Dialectal Arabic [77]. The proximity of dialectal Arabic to MSA makes the mapping easier than direct MT, and several researchers have explored this direction [77, 78]. The main challenge in developing any MT system is the lack of data. This challenge is accentuated in the case of Arabic and its dialects where parallel corpora are rarely publicly available. Some dialects are more suffering from this lack than others. For example, for Algerian dialect, only one parallel corpus is publicly available (PADIC) [28] which contains 6,412 sentences translated from Algerian Dialect to MSA. Some work have been done on Arabizi translation [70, 73, 79, 80]. However, these work consider transliteration before the translation step.

The idea of analysing sentiments after the automatic translation of messages was explored in many works [81, 82]. However, to the best of our knowledge, two works only have been done on Arabic [36, 83]. Rafaee et al. [83]

presented a sentiment analysis approach using freely available MT systems to translate Arabic tweets to English, which the authors then label for sentiment using a state-of-the-art English SA system. The authors of the cited work affirm that MT-based SA is a cheap and effective alternative to building a complete SA system when dealing with under-resourced languages. Salameh et al. [36] achieved competitive results even with automatic translation. Both papers present the same idea: The translation of Arabic messages into English and then use the English resources for determining the sentiment.

However, both papers concentrate on Arabic only (omitting its dialects and specially Arabizi).

Table 2 summarises and classifies the main works and resources presented in this section.

Methodology

Figure 1 summarizes the main steps of the proposed approach, including

– Corpus extraction
– Arabizi transliteration
– Arabizi translation
– Arabic sentiment analysis

Corpus Extraction

Text messages written in MSA/DALG from Facebook are extracted using two methods. In the first method, the comments from 226 popular Algerian pages such as Ooredoo[2], HamoudBoualem[3], and Ruiba[4] (which belongs to commercial companies, press, and public personalities) are extracted. The most popular Facebook pages using the statistics offered by the SocialBakers website[5] are identified. For the second method, Facebook content is searched using Facebook Rest API[6] with MSA/DALG words. The DALG terms are obtained using two sources. The first source is PADIC, which is a parallel multi-dialectal corpus containing parallel DALG–MSA pairs [28]. The second one is our translated lexicon that is described above. Using both methods, a corpus containing **15,407,910** messages is collected. After filtering out non-Arabic messages, **7,926,504** messages are retained. To extract the Arabic message, the

---

[2] https://fr-fr.facebook.com/OoredooDZ/.

[3] https://www.facebook.com/HamoudBoualemOfficielle/.

[4] https://fr-fr.facebook.com/Rouiba.News/.

[5] https://www.socialbakers.com/statistics/facebook/pages/total/algeria/.

[6] https://developers.facebook.com/docs/graph-api/.

Table 2 Works on Arabic/Arabizi sentiment analysis

| Area	Sub-area	Categories	Sub-category	Works and resource
Arabic sentiment analysis	Manualconstruction	Lexicon		[30, 31]
		Corpora		[11, 20–22, 33, 34]
	Automatic construction	Lexicon	Based on automatic translation	[30, 35–39]
			Based on resources linking	[40–43]
			Based on both translation and resources linking	[44, 45]
		Corpora	Using rating reviews	[52, 53]
			Using sentiment lexicon	[12]
	Semi-automatic Corpora	Lexicon		[54, 55]
		Corpora		[56]
	Feature extraction and classification	Word embedding		[58–61]
		Deep learning classification		[67, 69]
	Sentiment analysis after translation			[36, 83]
Arabizi processing	Transliteration	Rules-based approach		[10, 70, 71]
		Corpus-based approach		[7, 72, 73]
	Sentiment analysis	Arabizi as a proper language		[12]
		Bason on Arbizi Transliteration		[9, 10]
	Translation after transliteration			[70, 73, 79, 80]
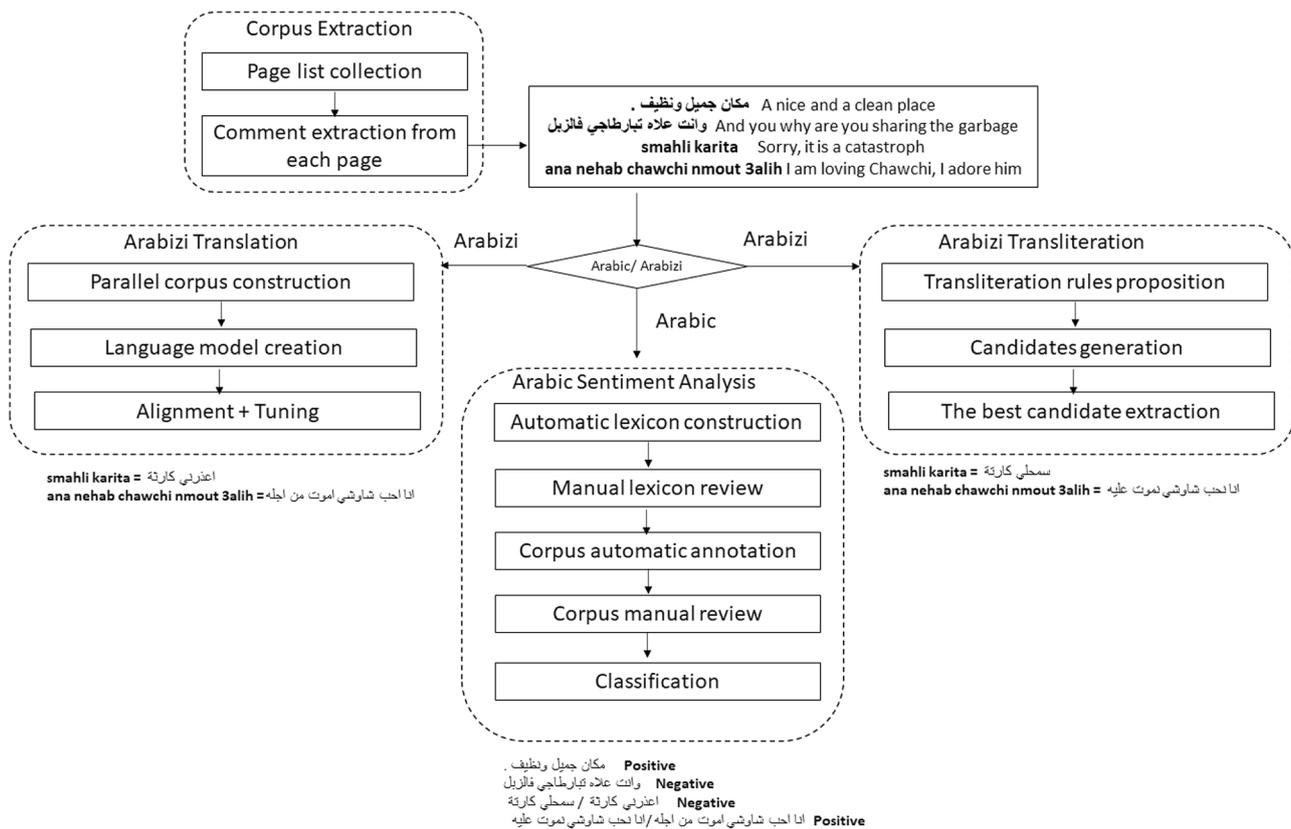
Fig. 1 A semi-automatic approach for Arabic/Arabizi sentiment analysis

Fig. 2 Sample of DALG Arabic corpus

انضمو لصفحتنا لتعيشو المتعة الكروية ﻻﻟ

و اله اعلم

من الاحسن يعطي فرصة لبنجاح

اسلام لانه يلعب في مستوي اعلى

بونجاح احسن

دايرين اشهار فالتلفاز و الراديو و الانترنت و مكتوقعوش 160 الف واحد ؟؟ خاصكم ديرو بحساب الملاين . استحمار الشعب و الله و لكن الغبي

إهو لي غادي يصدق فعلا بانه كاينة تلفاز او هاتف ذكي من النوع الممتاز فقط ب 100درهم

لقد نجحنا 😂😂😂

ريم غزالي نجمة ستار اكاديمي مع شمسو نجم الكوميديا الجزائريا   ربي يبارك نشاله تكون مليحة

حفظكم اله وذرياتكم من كل بلاء ورحم الغفار موتاكم وثبتكم على نشر الخير كي يملا الانسانية العالم كله.

صبر روحك.😊😊😊😊😊

كرهتونا فل بالون بسب المدعو سفينة الصحراء اسلام

بنات صعيبة بزاف اله يرحمهم يارب😊😊😊😊😊😊

صحيح هو الاول

بلاصتو في البنك

سليماني المعوج

messages were classified into two categories: (1) messages not containing Latine letters and (2) messages containing Latin letters. The messages were classified based on their script. The messages, including only Arabic letters, are used in the sentiment analysis step. The other messages are used in both steps related to transliteration/translation of Arabizi. This corpus was extracted in November 2017. As the interest of this study is dedicated to text analysis, only the textual messages were extracted. Figure 2 illustrates some samples from the resulting Arabic corpus.

To handle the extracted corpus, a set of pre-processing methods are used: (1) delete repeated messages; (2) delete exaggerations (for example the word نحّحّحّحب is transformed into: نحب and nhhhhhab is transformed into nhab. Different repetitions of the different letter 'ح' and 'h' are removed to keep a single occurrence); (3) delete the character '#' and different punctuation '.,!,?'; (4) delete consecutive whites spaces as well as Tatweel ('–').

Arabizi Transliteration

For Arabizi transliteration, we rely on the approach proposed by Guellil et al. in [71] and used for sentiment analysis purpose by Guellil et al. in [10]. This approach includes four main steps: (1) pretreatment of the Arabic corpus and the Arabizi message. (2) Proposal and application of the rules for the Algerian Arabizi. (3) Generating different candidates. (4) Extraction of the best candidate. It receives input, a set of messages written in Arabizi and a voluminous corpus written in MSA/DALG extracted from Facebook.

All these messages are pretreated. Afterwards, a set of passages rules are proposed (i.e. the letter 'a' could be replaced by 'ا، ة، ى، ع، ع'', etc. It could also be replaced by '', none letters when it represents a diacritic). By applying different replacements, as well as different rules

developed, each Arabizi word is corresponding to several words in Arabic. For example the word "kraht" (meaning, *I hate*) generates 32 possible candidates, such as 'كراهت'، 'قرهت'، 'كرهت'. The correctly transliterated word is 'كرهت'. To extract the best candidate for the transliteration of a given Arabizi word into Arabic, a language model is constructed and applied.

Arabizi Translation

From the corpus automatically extracted from Facebook, 2,924 were randomly selected. These comments were manually translated into Arabic (MSA). Table 3 presents the set of samples included in our parallel corpora. Our parallel corpus in between the pair Arabizi/MSA. The English translation is only added on the table for clarity. Inspired by the work presented in [28, 77, 78] on statistical machine translation of Arabic and its dialect, we propose three main steps: (1) language model training, (2) alignment, and (3) tuning. For training the language model, the large Arabic corpus in Arabic from Facebook is used. The parallel corpus was divided into two parts. The first one contains 90% of the whole corpus (representing 2,632 comments) is used for the training. The second one, containing 10% of the corpus (representing 292 comments, is used for the validation). Subsequently, alignment model and tuning methods are used to select the best translation. Inspired by [26], we used the open-source Moses toolkit [84] to build a phrase-based MT system with default settings: bidirectional phrase and lexical translation probabilities, distortion model, a word and a phrase penalty and a trigram language model. We used GIZA++ [85] for alignment and KenLM [86] to compute trigram language models.

Table 3 Some samples of the constructed parallel corpus Arabizi/MSA

Arabizi	MSA translation	English translation
tahadrou	كلكمْ تحيِّدون الكَلَام	All of you master the speaking
salam fi lhakika mabrouk 3la l'akhe chahadat al bakaloria	سلَام فَى الحقيقة مبروك علَى الأخ شهَادت البّكَالوريَا	Hi, actually congratulation to the brother for obtaining the baccalaureate certificate.
rebi m3ahom nchalah	ربي معهم ان شَاع الَه	God is with them if god willing
rana m3ak ya sa3dan	انَّنَا معك سعدَانِ	We are with you Saadane
sa7	حقيقة	True
had lblad li matou 3liha 1,5 milions d'algeriens.	من اجلهَا هده البلَاد التي مَات 1,5 مليون جزَاعيري	This is the country who 1.5 millions Algerians died for it.

Arabic Sentiment Analysis

Lexicon Construction and Review

For lexicon construction, we rely on the same approach proposed by Guellil et al. [39]. The main idea behind this construction is to automatically translate an existing English lexicon to DALG and MSA using Glosbe API[7]. In this work, we automatically translate SOCAL lexicon (containing 6769 terms among the Adjectives, verbs, nouns, and adverbs) [2]. The same score is assigned to all the translated words. This score corresponds to the score of the English word from which they are translated. For example, all the translations of the English word 'excellent' with a score of +5, such as بَاهي (bAhy) meaning *brilliant*, لطيف (lTyf) meaning *nice*, and 'مليح' (mlyH), meaning *good*, are assigned a score of +5. Since some Arabic words result from different English words having different sentiment scores, an average score is assigned to such Arabic words. For example, the word 'مليح' (mlyH), meaning *good* can be the translation of the English term 'excellent' (with an associated score of +5); however, it can also be translated from the English term "good" (with an associated score of +3). Hence, the Arabic term مليح is associated with the average of all sentiment scores of the English terms it is translated from.

The resulted lexicon after applying this approach contains 2,384 entries. Afterwards, we manually review this lexicon, to delete ambiguous words, to increase the annotation precision. Finally, we obtain a sentiment lexicon containing 1745 terms, of which 968 are negative, 771 are positive, and 6 are neutral, in both MSA and DALG.

Corpus Construction and Review

The constructed lexicon is used to provide a sentiment score for DALG utterances automatically. This process provides a baseline for different experiments. The lexicon is then used to build a large sentiment corpus. To calculate the score, we considered (1) opposition which is generally expressed in DALG with the keyword ''<b.s.h>' (bSH – but); (2) multi-word expressions because the constructed lexicon contains multi-word entries; (3) handling DALG morphology by employing a simple rule-based light stemmer that handles DALG prefixes and suffixes; (4) negation which can reverse polarity. Negation in DALG is usually expressed as an attached prefix, suffix, or a combination of both. To score a message, the sentiment scores of all the words in the message are averaged. Finally, balanced dataset (by keeping the same number of messages in a positive and negative dataset) is constructed. The resulted corpus contains 255,008 messages (where both positive and negative corpus contains 127, 504 messages).

By analysing the corpus annotated automatically, we observe that some messages were wrongly annotated. For example, the message: جَابو فخَامة الأسم تكفي meaning *Djabou the excellency of the name is sufficient* was annotated negative (where it is positive). Another example with the message: بهدي العب اله جيب الخير meaning *guide the play, we hope God brings the good things* (we hope God bring the good things is an expression used to speak about bad things). This message is wrongly annotated as positive. To construct the corpus, the messages that are correctly annotated were kept, and those that are wrongly annotated were corrected. Also, some objectives messages (not holding a sentiment) were deleted. The resulted corpus contains 3048 messages (where 1488 are positives, and 1560 are negatives). This corpus is considered, to the best of our knowledge, as the first annotated sentiment corpus (manually checked) which

---

[7] https://glosbe.com/en/arq/excellent.

handles DALG as well as MSA. We also use it for evaluating our automatic annotation. Among the 3048 messages that are manually reviewed, 2596 messages representing 85.17% were correctly annotated.

Sentiment Classification

For classification, we use two kinds of Algorithms, shallow and deep. For both classifications, we extract features with word embedding techniques. With shallow classification, we use Word2vec algorithm, while we use both word2Vec and fastText for deep classification.

Word2Vec + Classical Machine Learning Algorithms

For Word2vec, we used a context of 10 words to produce representations for both CBOW and SG of length 300. We trained the Word2vec models on the messages that appear in the training sets. In this work, we used the model presented by Altowayan et al. [61]. However, this work relies only on CBOW representation, where we rely on both representations CBOW and SG. For classification, we use five Algorithms such as GaussianNB (GNB), LogisticRegression (LR), RandomForset (RF), SGDClassifier (SGD, with loss='log' and penalty='l1') and LinearSVC (LSVC with C='1e1').

Word2/fastText + Deep Learning Algorithms

Three deep learning classifiers were used: CNN, LSTM and Bi-LSTM. For each model, six layers were used. The first layer is a randomly-initialised word embedding layer that turns words in sentences into a feature map. The weights of embedding_matrix are calculated using word2vec and fastText (with both SG and CBOW implementation). This layer is followed by a CNN/ LSTM/BiLSTM layer that scans the feature map (depending on the model that we defined). These layers are used with 300 filters and a width of 7, which means that each filter is trained to detect a particular pattern in a 7-gram window of words. Global max-pooling is applied to the output generated by CNN/LSTM/BiLSTM layer to take the maximum score of each pattern. The main function of the pooling layer is to reduce the dimensionality of the CNN/LSTM/BiLSTM representations by downsampling the output and to keep the maximum value. For reducing over-fitting by preventing complex co-adaptations on training data, a Dropout layer with a probability equal to 0.5 is added. The obtained scores are then fed to a single feed-forward (fully connected) layer with Relu activation. Finally, the output of that layer goes through a sigmoid layer that predicts the output classes. For all the models, we used Adam optimisers with epoch 100 and an early_stopping parameter for stopping the iteration in the absence of improvements.

Evaluation and Results

Dataset

For evaluating the proposed approach, different corpora were constructed and used:

– *Ar_corpus1*, automatically extracted from Facebook. This corpus was extracted by targeting the 226, most famous Algerian pages. It was extracted in November 2017 that contains 15,407,910 messages with 7,926,504 written in Arabic letters. This corpus is rich in term of opinions, sentiments and emotions.
– *ALG_Senti_auto*, the Algerian annotated (automatically) sentiment corpus. ALG_Senti is an annotated sentiment corpus which was automatically constructed based on the sentiment lexicon. In the context of this paper and by following the majority of work that constructed a balanced training corpus[8]. Hence, after deleting the repeated messages, we obtained a corpus containing 255,008 messages where both positives and negatives classes respectively contain 127,504 messages.
– *ALG_Senti_manu*, representing the manually reviewed corpus and containing 3048 messages (where 1488 are positives and 1560 are negatives).
– *Test_Ar_Tr_auto*, which is an Arabizi sentiment corpus, firstly used in [12] and transliterated automatically in [10] with an accuracy of 72.05% and containing 500 Facebook comments (250 are positives and 250 are negatives).
– *Test_Ar_Tr_manu*, which is the same Arabizi sentiment corpus [12] and transliterated manually in [10],
– *Test_Ar_Translation_auto*, which is an Arabizi sentiment corpus, firstly used in [12] and translated automatically (the BLEU score of the automatic translation is up to 8.13).
– *Test_Ar_Translation_manu* which is the same Arabizi sentiment corpus [12] and translated manually.
– *SANA_Alg*[9], an Algerian sentiment corpus containing 513 messages (236 positives; 194 negatives; 83 neutral) extracted from news, political, religion, sports, and society articles selected at the following Algerian Arabic newspaper web sites.

---

[8] A corpus is balanced if the number of positives messages equals the number of negative ones. In the other case, it is unbalanced

[9] http://rahab.e-monsite.com/medias/files/corpus.rar.

– *ASTD/QCRI/ARTwitter*[10], a corpus containing (4349 messages), including both MSA and Egyptian dialect.

Metrics

In total, five metrics are used for evaluating the proposed system. To evaluate the transliteration module, the Accuracy (A) is used. Accuracy, as shown in Eq. 1, represents the number of words correctly transliterated divided by the total number of words. In order to evaluate the translation module, the BLEU score is used [87]. BLEU score, as shown in Eq. 2 represents the geometric mean of the test corpus using modified precision scores and multiplied by an exponential brevity penalty factor. In order to evaluate the sentiment analysis module, three metrics are used (Precision(P), Recall(R) and F1 score (F1)). Precision, as shown in Eq. 3, represents the number of sentiments correctly labelled as belonging to the positive class divided by the total number of sentiments labelled as belonging to the positive class. Recall, as shown in Eq. 4, represents the number of true positives divided by the total number of opinions that belongs to the positive class. Finally, F score, as shown in Eq. 5, represents the harmonic mean of precision and recall [88].

$$A = \frac{NB\_Correct}{NB\_Total} \tag{1}$$

$$BLEU = BP \cdot e^{\sum_{n=1}^{N} w_n \log p_n} \tag{2}$$

$$P = \frac{TP}{TP + FP} \tag{3}$$

$$R = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2 * P * R}{P + R} = \frac{2 * TP}{2 * TP + FP + FN}, \tag{5}$$

where NB_Correct represents the number of words correctly transliterated. NB_Total represents the total number of words. BP, as shown in Eq. 6, represents the brevity penalty comparing the length of the candidate translation c and the effective reference corpus length r. TP represents true positive (i.e. manually annotated as positive and predicted by the model as positive). TN represents true negative (i.e. manually annotated as negative and predicted by the model as negative). FP represents false positive (i.e. manually annotated as negative and predicted by the model as positive).

[10] https://github.com/iamaziz/ar-embeddings/tree/master/datasets/tweets.

And FN represents false negative (i.e. manually annotated as positive and predicted by the model by negative.

$$BP = \begin{cases} 1, & \text{if } c > r \\ e^{1-\frac{r}{c}}, & \text{if } c <= r. \end{cases} \tag{6}$$

Experimental Results

Our aim behind this experiments is to first synthesise and compare the results obtained using both training corpora (i.e. constructed automatically and that reviewed manually). Second, the sentiments analysis results using both techniques (i.e. transliteration and translation) also need to be compared. Third, the best model for extracting features (i.e. Word2vec and fastText) needs to be extracted. Fourth, the most suitable classification algorithms (classical ones and deep learning ones). Finally, the most suitable deep learning algorithm for Arabic sentiment analysis (i.e. CNN, LSTM, Bi-LSTM) needs to be highlighted.

Results Using Word2Vec + Classical Machine Learning Algorithms

Both SG and CBOW models where used. However, the CBOW model gives the best results where it is associated to the classical algorithms. Then, table 4 presents results, on both training corpora (constructed automatically and reviewed manually) for all the used test corpora. It can be from this table that the manual review of the corpus consequently improves the results. For Arabizi, we observe that automatic transliteration (F1=0.74/0.78) gives better results than the automatic translation (F1= 0.71/0.73). However, the manual translation (F1 = 0.78/0.82) outperforms the manual transliteration (F1= 0.76/0.80). The best results obtained are on Test_SentiAlg (F1= 0.89) which we constructed in the context of previous researches. The results on both SANA_Alg (0.80) and ASTD/QCRI/ARTwitter (0.80) are very encouraging.

Results Using Word2Vec/fastText + Deep Learning Algorithms

Same to the previous experiments, both models CBOW and SG were used. However, now we present the results obtained using the SG model because this model outperforms the CBOW model. It can be seen from Table 5 that manual reviewing on the automatic annotation improves the performances. The obtained results with *Test_SentiAlg* are up to 0.82 where they are up to 0.89 on *Corpus_manu*.

Concerning the models extracting features such as Word2vec and FastText, it can be seen that the best results were obtained using both models with *Corpus_auto*. However,

Table 4 CBOW Word2Vec + classical machine learning algorithms results

Models	Test_corpus	Classifier	Corpus_auto			Corpus_manu		
			P	R	F1	P	R	F1
Word2vec	Test_SentiAlg	GNB	0.93	0.75	0.83	0.94	0.71	0.81
		LR	0.82	0.88	0.85	0.94	0.82	0.88
		SGD	0.85	0.90	**0.88**	0.87	0.91	**0.89**
		LSVC	0.83	0.88	0.85	0.90	0.86	0.88
		RF	0.85	0.75	0.80	0.86	0.70	0.77
	Test_Ar_Tr_auto	GNB	0.82	0.57	0.67	0.85	0.51	0.63
		LR	0.69	0.74	0.72	0.87	0.71	**0.78**
		SGD	0.69	0.80	**0.74**	0.84	0.66	0.74
		LSVC	0.70	0.72	0.71	0.81	0.74	0.77
		RF	0.73	0.64	0.68	0.85	0.57	0.68
	Test_Ar_Tr_manu	GNB	0.85	0.59	0.70	0.87	0.52	0.69
		LR	0.75	0.77	**0.76**	0.89	0.72	**0.80**
		SGD	0.74	0.74	0.74	0.87	0.75	**0.80**
		LSVC	0.76	0.76	**0.76**	0.85	0.75	0.79
		RF	0.72	0.64	0.68	0.87	0.57	0.69
	Test_Ar_Translat_auto	GNB	0.63	0.79	0.70	0.61	0.83	0.71
		LR	0.58	0.88	0.70	0.68	0.79	0.73
		SGD	0.60	0.87	**0.71**	0.65	0.82	**0.73**
		LSVC	0.58	0.87	0.69	0.65	0.80	0.72
		RF	0.62	0.66	0.64	0.68	0.66	0.67
	Test_Ar_Translat_manu	GNB	0.69	0.88	**0.78**	0.71	0.88	0.78
		LR	0.65	0.88	0.75	0.78	0.87	**0.82**
		SGD	0.68	0.86	0.76	0.72	0.92	0.80
		LSVC	0.66	0.87	0.75	0.76	0.84	0.80
		RF	0.70	0.79	0.74	0.74	0.74	0.74
	SANA_Alg	GNB	0.81	0.81	**0.81**	0.78	0.68	0.73
		LR	0.76	0.71	0.73	0.88	0.73	**0.80**
		SGD	0.80	0.73	0.76	0.91	0.62	0.74
		LSVC	0.74	0.69	0.72	0.90	0.64	0.75
		RF	0.72	0.77	0.74	0.85	0.55	0.67
	ASTD/QCRI/ARTwitter	GNB	0.76	0.86	**0.81**	0.78	0.81	0.79
		LR	0.72	0.75	0.73	0.82	0.79	**0.80**
		SGD	0.78	0.73	0.75	0.73	0.88	**0.80**
		LSVC	0.71	0.75	0.73	0.84	0.69	0.76
		RF	0.70	0.68	0.69	0.80	0.67	0.73

fastText literally outperforms word2vec with *Corpus_manu*. It can also be seen from Table 5 that CNN outperforms all the others classifiers with *Corpus_manu*. However, both CNN and Bi-LSTM give remarkable results on *Corpus_auto*.

Finally, from Table 5, it can be seen that the results of Arabizi after transliteration process (up to 0.71/0.80 for the automatic transliteration and up to 0.74/0.80 for the manual transliteration). These results are more promising than those obtained after the translation (up to 0.61/0.69 for the automatic translation and up to 0.66/0.79 for the manual translation). However, it can also be seen that the difference of results between automatic/manual transliteration is less significant than the difference of the results between automatic/manual

translation. The results related to manual transliteration/translation are almost the same for *corpus_manu*, the corpus constructed in a semi-supervised way. This highlights the effectiveness of both techniques for handling Arabizi. However, the translation approaches require many improvements, starting by enriching the parallel corpus.

Table 5 Deep learning classification results

Models	Test_corpus	Classifier	Corpus_auto			Corpus_manu		
			P	R	F1	P	R	F1
Word2Vec	Test_SentiAlg	CNN	0.81	0.81	**0.82**	0.88	0.88	**0.88**
		LSTM	0.80	0.80	0.80	0.87	0.87	0.87
		Bi-LSTM	0.80	0.80	0.80	0.87	0.87	0.87
	Test_Ar_Tr_auto	CNN	0.72	0.72	**0.71**	0.80	0.80	**0.80**
		LSTM	0.67	0.67	0.67	0.78	0.78	0.78
		Bi-LSTM	0.69	0.69	0.69	0.76	0.76	0.76
	Test_Ar_Tr_manu	CNN	0.73	0.73	**0.74**	0.84	0.84	**0.84**
		LSTM	0.72	0.72	0.72	0.78	0.78	0.78
		Bi-LSTM	0.70	0.70	0.69	0.80	0.80	0.80
	Test_Ar_Translat_auto	CNN	0.63	0.63	0.63	0.67	0.67	0.67
		LSTM	0.63	0.58	0.54	0.68	0.68	**0.68**
		Bi-LSTM	0.66	0.66	**0.66**	0.67	0.66	0.66
	Test_Ar_Translat_manu	CNN	0.68	0.67	**0.67**	0.78	0.77	**0.76**
		LSTM	0.68	0.66	0.65	0.77	0.76	**0.76**
		Bi-LSTM	0.68	0.67	**0.67**	0.76	0.75	0.75
	SANA_Alg	CNN	0.59	0.60	0.59	0.74	0.73	0.73
		LSTM	0.63	0.63	0.62	0.75	0.73	0.73
		Bi-LSTM	0.60	0.60	0.60	0.74	0.74	0.74
	ASTD/QCRI/ARTwitter	CNN	0.64	0.64	0.64	0.74	0.74	0.74
		LSTM	0.63	0.62	0.61	0.73	0.73	0.73
		Bi-LSTM	0.62	0.62	0.62	0.72	0.72	0.72
fastText	Test_SentiAlg	CNN	0.80	0.80	**0.80**	0.89	0.89	**0.89**
		LSTM	0.79	0.79	0.79	0.87	0.87	0.87
		Bi-LSTM	0.79	0.79	0.79	0.88	0.88	0.88
	Test_Ar_Tr_auto	CNN	0.69	0.69	**0.69**	0.83	0.83	**0.83**
		LSTM	0.68	0.68	0.68	0.80	0.80	0.80
		Bi-LSTM	0.68	0.68	0.68	0.79	0.79	0.79
	Test_Ar_Tr_manu	CNN	0.73	0.73	**0.73**	0.82	0.82	**0.82**
		LSTM	0.71	0.71	0.71	0.79	0.78	0.78
		Bi-LSTM	0.70	0.70	0.70	0.81	0.81	0.81
	Test_Ar_Translat_auto	CNN	0.61	0.61	**0.61**	0.68	0.68	0.68
		LSTM	0.64	0.62	**0.61**	0.66	0.66	0.65
		Bi-LSTM	0.61	0.61	**0.61**	0.70	0.69	**0.69**
	Test_Ar_Translat_manu	CNN	0.67	0.66	**0.66**	0.80	0.79	**0.79**
		LSTM	0.80	0.63	0.63	0.80	0.78	0.78
		Bi-LSTM	0.65	0.64	0.64	0.81	0.80	**0.79**
	SANA_Alg	CNN	0.64	0.64	**0.64**	0.71	0.70	0.70
		LSTM	0.60	0.59	0.59	0.71	0.70	0.70
		Bi-LSTM	0.57	0.58	0.58	0.72	0.71	**0.72**
	ASTD/QCRI/ARTwitter	CNN	063	0.63	**0.63**	0.75	0.75	**0.75**
		LSTM	0.61	0.60	0.60	0.73	0.73	0.73
		Bi-LSTM	0.60	0.60	0.60	0.73	0.73	0.73

Discussion and Analysis

Discussion

For showing the efficiency of our approach and corpus, we carried out many experiments on several test corpora (previously used in the research literature). The corpora *Senti_Alg* (i.e. *Senti_Alg_test_Arabic*, *Senti_Alg_test_trauto* and *Senti_Alg_test_trmanu* were presented and used in many research papers [10, 12, 39]. The results related to *Test_SentiAlg_Arabic* (up to 87.77%) are very encouraging. These results were obtained using the CBOW model associated

to the SGD classifier. The research literature presents an F1 score of up to 68%.

The best results obtained on SANA_Alg are up to 81.00% (for F1 score). This result outperforms the results presented in the research literature, where the F1 score presented by Rahab et al. [29] was up to 75%. Hence, our approach and corpus lead to an improvement of 6% on this corpus.

Finally, our corpus and approach were also evaluated on MSA and another dialect (Egyptian dialect) using corpus ASTD/QCRI/ArTwitter, which was used by Altowayen et al. [61]. In [61], the corpus was classified in two classes only (i.e. positive and negative classes). As we also focus on binary classification, it was more practical to compare our results to the results obtained by these authors ([61]) rather than comparing them to the results obtained for each corpus separately. The best results obtained by Altowayen et al. [61] are up to 79.62% (for F1 score). The best results that we obtained are up to 80.58% (for F1 score). Moreover, This corpus is dedicated to MSA with a focus on Egyptian dialect (for ASTD). Hence, our approach and corpus which are dedicated to Algerian dialect outperform the results presented for corpora dedicated to MSA and Egyptian dialect.

Analysis

After presenting and comparing all the results related to the presented approach, we can answer different research questions presented in the *Introduction* part. We present the different answers in the following part.

1. **What is the best option for handling Arabizi (i.e. transliteration or translation)?** From the presented results, it can be seen that the transliteration is more suitable for Arabizi sentiment analysis. However, it was also highlighted that bad results associated with the translation are not related to the technique itself but the proposed approach. An approach is principally relying on a small parallel corpus including only 2924 parallel sentences.
2. **How could we improve the used transliteration approach?** The principal error appears in transliteration process is related to the technique of choosing the best candidate. The idea of a language model is to extract the candidate having the most significant number of occurrence. However, in some cases, these techniques return an incorrect candidate. For example, the word "rakom" meaning "you are" is transliterated as "رقم" meaning "a number" rather than "راكم" (which is the correct transliteration). The solution to this problem is to integrate other parameters for determining the best candidate, such as distance.
3. **How could we improve the proposed translation approach?** To improve the translation results, particular

attention should be first given to the parallel corpus construction and enrichment. 2,924 parallel sentences are not enough for training a statistical machine translation system. Relying on neural machine translation will also certainly improve the results. However, neural networks models require large corpora for the training phase.

4. **What is the best technique for Arabic sentiment analysis (i.e. supervised or semi-supervised approach)?** The presented results highlight the fact that the corpus constructed semi-automatically outperform the corpus constructed purely automatically. Hence, a semi-supervised approach is less effort and time consuming than a manual one, and its results are better than an automatic one.
5. **How could we improve the automatic annotation approach?** Some sentiment classification errors are due to transliteration errors for Arabizi. For example, "khlwiya" meaning excellent and quiet is wrongly transliterated to "خليَا" (meaning empty) rather than خلوية.. Improving transliteration will improve sentiment classification. Also, the automatic annotation is based on a reduced lexicon counting only 1,745 terms. Then, the vocabulary on which we based our annotation is relatively small. The manually reviewed corpus also has a reduced size (only 3,048 messages) where most corpora on the literature contain more than 10,000 messages. Hence, enriching both the sentiment lexicon and the sentiment annotated corpus will undoubtedly improve the results.

Open Issues

From this study and results, we have identified several research directions that deserve a more in-depth study.

Proposing a Statistical Machine Transliteration System

To improve our transliteration approach, we plan to use the presented system for automatically transliterate an Arabizi corpus. Afterwards, we manually review the transliteration pairs. Our systems gave us a precision more than (70%). Hence correcting (30%) of wrongly transliterated messages is better than constructing (100%) (from constructing a parallel corpus from scratch). Then, we could consider the transliteration as a translation task.

Proposing an Arabizi Identification Module

Among the issues related to Arabizi treatment, the confusion among Arabizi, English and French. In the context of this paper, we assume that our input is messages written in Arabic, Arabizi or both scripts. However, in real life, it is

not the unique case. The problem that we could face with this system is to transliterate a message written in French or English. To resolve this problem, we plan to work on an identification system. We previously proposed a bilingual lexicon that we constructed [5]. Using this lexicon, we proposed a rules-based identification system [89]. However, we plan to improve our identification system by considering the identification task as any classification problem containing three classes: (1) messages which are written in Arabizi. (2) messages which are written in French. (3) messages which are written in English. Hence, we could use machine learning algorithms to detect Arabizi messages.

Enriching the Proposed Lexicon Automatically

Our proposed lexicon was constructed automatically by translating an existing English lexicon. The constructed lexicon was then manually reviewed. The resulted lexicon contains only 1745 entries. The problem with a reduced lexicon is that it is not covering all the vocabularies, and then it could not analyse the sentiment of all messages. The proposed lexicon could be enriched using Word2vec. The idea of Word2vec is to return the most semantically close words to a given word (i.e. the words which have similar vectors). However, the problem with this technique is that the two words "good" and "bad" are returned simultaneously as they are very close. It is perfectly understandable, where these two words frequently appear in the same context. Hence, our major problematic by handling these issues is to resolve the "good/bad" situation.

The Application of This Approach to More Dialects and Languages

It can be seen from the obtained results that our approach outperforms the results presented in the research literature, even with an Algerian corpus. To have a multi-dialect sentiment analysis, we need a training corpus for each dialect. For obtaining these training corpora, we propose to extend this approach to other dialects. This approach could also be applied to other languages. Moreover, it could be employed with other NLP problems requiring training corpus (especially in the training corpus is used in the context of classification).

Using This Approach in a Real-Life Application Case

Lots of recent application and problematic need sentiment analysis—for example, Hate-speech detection. According to Nockleby, *Hate speech is commonly defined as any communication that disparages or defames a person or a group based on some characteristic such as race, colour, ethnicity, gender, sexual orientation, nationality, religion, or other characteristics*. Many approaches are proposed for hate-speech detection. Some researched consider hate-speech as a strong negative sentiment. Hence, We could use a part of our corpus in the context of Hate-speech detection. Also, we could propose the same approach to construct a corpus dedicated to hate-speech detection automatically.

To sum up, this paper handles sentiment analysis of Arabic and its dialects by focusing on both scripts: Arabic and Arabizi. It proposes new techniques and approaches for handling Arabizi and for constructing resources with a minimum of efforts. It also showed that reviewing a resource constructed automatically is better than constructing it from scratch in term of effort, time and results. However, this approach, as all the approaches presented in the research literature is not perfect. Some issues were observed. To handle these issues, we need to develop other approaches related to other NLP fields. Then, we join Erik Cambria qualifying sentiment analysis as *a big suitcase of natural language processing (NLP) problems. Sentiment analysis has long been mistaken for the task of polarity detection. However, it is just one of the many NLP problems that need to be solved to achieve human-like performance in sentiment analysis* [90].

Conclusion and Perspectives

In this paper, we proposed a sentiment analysis approach dedicated to Arabic and its dialects, and we applied it on DALG/MSA Facebook messages. The principal strengths of this approach are that we automatically constructed a sentiment corpus that we reviewed manually (for increasing the classification precision) and we handle both scripts Arabic and Arabizi. Another important aspect is that we relied on different word embedding models and different deep learning classifiers (for comparing the results). The obtained results are very encouraging (F1 up to 89% for extrinsic experiments using CNN), and they outperform the results obtained in the research literature (with a difference up to 25%). Also, for handling Arabizi, both techniques, transliteration and translation were used.

After analysing the different classifications errors, we highlighted different issues that we plan to address in our future works by integrating the following points:

- Proposing a transliteration system based on a corpus-based approach.
- Enriching the parallel corpora and proposed a neural machine translation system.
- Extending the constructed lexicon using Word2vec.
- Extending the constructed annotated corpus.
- Proposing classifiers which combine between different models.

– Extending the approach to other dialects by starting with Maghrebi dialect which shares many characteristics with DALG.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Liu B. Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol. 2012;5(1):1–167.
2. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput linguist. 2011;37(2):267–307.
3. Maas AL, Daly RE, Pham PT. Huang D, Ng AY, Potts C (2011) Learning word vectors for sentiment analysis. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies-volume 1, pp. 142–150. Association for computational linguistics
4. Guellil I, Boukhalfa K (2015) Social big data mining: a survey focused on opinion mining and sentiments analysis. In: Programming and systems (ISPS), 2015 12th international symposium on, pp. 1–10. IEEE
5. Guellil I, Faical A (2017) Bilingual lexicon for algerian arabic dialect treatment in social media. In: WiNLP: Women & Underrepresented Minorities in Natural Language Processing (co-located with ACL 2017). http://www.winlp.org/wp-content/uploads/2017/final_papers_2017/92_Paper.pdf
6. Guellil I, Azouaou F, Valitutti A (2019) English vs arabic sentiment analysis: A survey presenting 100 work studies, resources and tools. In: 2019 IEEE/ACS 16th international conference on computer systems and applications (AICCSA), pp. 1–8. IEEE
7. Darwish K ((2013)) Arabizi detection and conversion to arabic. arXiv preprint arXiv:1306.6755
8. Guellil I, Saâdane H, Azouaou F, Gueni B, Nouvel D (2019) Arabic natural language processing: an overview. J King Saud Univ Comput Inf Sci (2019)
9. Duwairi RM, Alfaqeh M, Wardat M, Alrabadi A (2016) Sentiment analysis for arabizi text. In: information and communication systems (ICICS), 2016 7th international conference on, pp. 127–132. IEEE
10. Guellil I, Adeel A, Azouaou F, Benali F, Hachani AE, Hussain A (2018) Arabizi sentiment analysis based on transliteration and automatic corpus annotation. In: Proceedings of the 9th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 335–341
11. Medhaffar S, Bougares F, Esteve Y, Hadrich-Belguith L (2017) Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In: Proceedings of the third arabic natural language processing workshop, pp. 55–61
12. Guellil I, Adeel A, Azouaou F, Hussain A (2018) Sentialg: automated corpus annotation for Algerian sentiment analysis. In: 9th international conference on brain inspired cognitive systems (BICS 2018)
13. Pfitzner R, Garas A, Schweitzer F. Emotional divergence influences information spreading in twitter. ICWSM. 2012;12:2–5.
14. Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B (2014) Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd Annual meeting of the association for computational linguistics (Volume 1: Long Papers), vol. 1, pp. 1555–1565
15. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining
16. Refaee, E., Rieser, V (2014) An arabic twitter corpus for subjectivity and sentiment analysis. In: LREC
17. Al-Twairesh N, Al-Khalifa H, AlSalman A (2016) Arasenti: large-scale twitter-specific arabic sentiment lexicons. In: Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 1: Long Papers). 1:697–705.
18. Salloum SA, Mhamdi C, Al-Emran M, Shaalan K (2017) Analysis and classification of arabic newspapers' facebook pages using text mining techniques
19. Ryan C, Renduchintala A, Saphra N, Callison-Burch C (2014) An Algerian Arabic-French code-switched corpus. In: Workshop on free/open-source arabic corpora and corpora processing tools workshop programme, p. 34
20. Rushdi-Saleh M, Martín-Valdivia MT, Ure na-López LA, Perea-Ortega JM, Oca: opinion corpus for Arabic. J Assoc Inf Sci Technol. 2011;62(10):2045–54.
21. Abdul-Mageed M, Diab MT ()2012 Awatif: A multi-genre corpus for modern standard arabic subjectivity and sentiment analysis. In: LREC, pp. 3907–3914
22. Nabil M, Aly M, Atiya A (2015) Astd: Arabic sentiment tweets dataset. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 2515–2519
23. Hogenboom A, Bal D, Frasincar F, Bal M, de Jong F, Kaymak U (2013) Exploiting emoticons in sentiment analysis. In: Proceedings of the 28th annual ACM symposium on applied computing, pp. 703–710. ACM
24. Yadav P, Pandya D (2017) Sentireview: sentiment analysis based on text and emoticons. In: Innovative mechanisms for industry applications (ICIMIA), 2017 International conference on, pp. 467–472. IEEE
25. Gamal D, Alfonse M, El-Horbaty ESM, Salem ABM. Twitter benchmark dataset for Arabic sentiment analysis. Int J Mod Educ Comput Sci. 2019;11(1):33.
26. Meftouh K, Bouchemal N, Smaïli K (2012) A study of a non-resourced language: the case of one of the algerian dialects. In: The third international workshop on spoken languages technologies for under-resourced languages-SLTU'12
27. Harrat S, Meftouh K, Smaïli K (2017) Machine translation for Arabic dialects (survey). Inf Process & Manag
28. Meftouh K, Harrat S, Jamoussi S, Abbas M, Smaili K (2015) Machine translation experiments on padic: a parallel Arabic dialect corpus. In: The 29th Pacific Asia conference on language, information and computation

29. Rahab H, Zitouni A, Djoudi M (2019) Sana: sentiment analysis on newspapers comments in algeria. J King Saud Univ Comput Inf Sci

30. Abdul-Mageed M, Diab M (2012) Toward building a large-scale arabic sentiment lexicon. In: Proceedings of the 6th international global WordNet conference 18–22.

31. Mataoui M, Zelmati O, Boumechache M. A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. Res Comput Sci. 2016;110:55–70.

32. Maamouri M, Bies A, Buckwalter T, Mekki W (2004) The penn arabic treebank: Building a large-scale annotated arabic corpus. In: NEMLAR conference on Arabic language resources and tools, vol. 27, pp. 466–467. Cairo

33. Abdul-Mageed M, Diab M, Kübler S. Samar: Subjectivity and sentiment analysis for Arabic social media. Comput Speech Lang. 2014;28(1):20–37.

34. Mourad A, Darwish K (2013) Subjectivity and sentiment analysis of modern standard Arabic and Arabic microblogs. In: Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis, pp. 55–64

35. Mohammad S, Salameh M, Kiritchenko S (2016) Sentiment lexicons for Arabic social media. In: LREC

36. Salameh M, Mohammad S, Kiritchenko S (2015) Sentiment after translation: a case-study on arabic social media posts. In: Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies, pp. 767–777

37. Mohammad SM, Salameh M, Kiritchenko S. How translation alters sentiment. J Artif Intell Res. 2016;55:95–130.

38. Abdulla N, Mohammed S, Al-Ayyoub M, Al-Kabi M, et al. (2014) Automatic lexicon construction for arabic sentiment analysis. In: Future internet of things and cloud (FiCloud), 2014 international conference on, pp. 547–552. IEEE

39. Guellil I, Azouaou F, Saâdane H, Semmar N (2017) Une approche fondée sur les lexiques d'analyse de sentiments du dialecte algérien

40. Badaro G, Baly R, Hajj H, Habash N, El-Hajj W (2014) A large scale arabic sentiment lexicon for arabic opinion mining. In: Proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing (ANLP), pp. 165–173

41. Gilbert B, Hussein J, Hazem H, Wassim EH, Nizar H (2018) Arsel: a large scale arabic sentiment and emotion lexicon

42. Eskander R, Rambow O (2015) Slsa: a sentiment lexicon for standard Arabic. In: Proceedings of the 2015 conference on empirical methods in natural language processing, pp. 2545–2550

43. Altrabsheh, N., El-Masri, M., Mansour, H (2017) Combining sentiment lexicons of Arabic terms

44. Mahyoub FH, Siddiqui MA, Dahab MY. Building an Arabic sentiment lexicon using semi-supervised learning. J King Saud Univ Comput Inf Sci. 2014;26(4):417–24.

45. Abdulla NA, Ahmed NA, Shehab MA, Al-Ayyoub M, Al-Kabi MN, Al-rifai S. Towards improving the lexicon-based approach for arabic sentiment analysis. Int J Inf Technol Web Eng (IJITWE). 2014;9(3):55–71.

46. Ding X. Liu B, Yu PS (2008) A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 international conference on web search and data mining, pp. 231–240. ACM

47. Esuli A, Sebastiani F. Sentiwordnet: a high-coverage lexical resource for opinion mining. Evaluation. 2007;17:1–26.

48. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A. Sentiment strength detection in short informal text. J Am Soc Inf Sci Technol. 2010;61(12):2544–58.

49. Fellbaum C, Alkhalifa M, Black W, Elkateb S, Pease A, Rodriguez H, Vossen P (2006) Introducing the Arabic wordnet project

50. Graff D, Buckwalter T, Jin H, Maamouri M (2006) Lexicon development for varieties of spoken colloquial arabic. In: LREC

51. Buckwalter T (2004) Buckwalter arabic morphological analyzer version 2.0. linguistic data consortium, university of pennsylvania, 2002. ldc cat alog no.: Ldc2004l02. Tech. rep., ISBN 1-58563-324-0

52. Aly, M., Atiya, A (2013) Labr: a large scale arabic book reviews dataset. In: Proceedings of the 51st annual meeting of the association for computational linguistics (Volume 2: Short Papers), vol. 2, pp. 494–498

53. El Sahar H, El-Beltagy SR (2015) Building large arabic multi-domain resources for sentiment analysis. In: International conference on intelligent text processing and computational linguistics, pp. 23–34. Springer

54. El-Beltagy SR (2016) Nileulex: A phrase and word level sentiment lexicon for egyptian and modern standard arabic. In: LREC

55. Abdul-Mageed M, Diab MT (2016) Sana: A large scale multi-genre, multi-dialect lexicon for arabic subjectivity and sentiment analysis. In: LREC

56. Al-Twairesh N, Al-Khalifa H, Al-Salman A, Al-Ohali Y. Arasenti-tweet: a corpus for Arabic sentiment analysis of Saudi tweets. Procedia Comput Sci. 2017;117:63–72.

57. Alowaidi S, Saleh M, Abulnaja O. Semantic sentiment analysis of Arabic texts. Int J Adv Comput Sci Appl. 2017;8(2):256–62.

58. Barhoumi A, Aloulou YEC, Belguith LH (2017) Document embeddings for Arabic sentiment analysis

59. Al-Azani S, El-Alfy ESM. Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text. Procedia Comput Sci. 2017;109:359–66.

60. El Mahdaouy A, Gaussier E, El Alaoui SO (2016) Arabic text classification based on word and document embeddings. In: International conference on advanced intelligent systems and informatics, pp. 32–41. Springer

61. Altowayan AA, Tao L (2016) Word embeddings for arabic sentiment analysis. In: Big Data (Big Data), 2016 IEEE international conference on, pp. 3820–3825. IEEE.

62. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. Adv Neural Inf Process Syst. 2013;1:3111–9.

63. Le Q, Mikolov T (2014) Distributed representations of sentences and documents. In: International conference on machine learning, pp. 1188–1196

64. Joulin A, Grave E, Bojanowski P, Mikolov T (2016) Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759

65. Tafreshi S, Diab M (2018) Emotion detection and classification in a multigenre corpus with joint multi-task deep learning. In: Proceedings of the 27th international conference on computational linguistics, pp. 2905–2913

66. Schmitt M, Steinheber S, Schreiber K, Roth B (2018) Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks. arXiv preprint arXiv:1808.09238

67. Dahou A, Xiong S, Zhou J, Haddoud MH, Duan P (2016) Word embeddings and convolutional neural network for Arabic sentiment classification. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers, pp. 2418–2427

68. Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882

69. Attia M, Samih Y, El-Kahky A, Kallmeyer L (2018) Multilingual multi-class sentiment classification using convolutional neural networks. In: LREC

70. May J, Benjira Y, Echihabi A (2014) An arabizi-english social media statistical machine translation system. In: Proceedings of the 11th conference of the association for machine translation in the Americas, pp. 329–341

71. Guellil I. Azouaou F, Benali F, Hachani AE, Saadane H (2018) Approche hybride pour la translitération de l'arabizi algérien : une étude préliminaire. In: Conference: 25e conférence sur le Traitement Automatique des Langues Naturelles (TALN), May 2018, Rennes, FranceAt: Rennes, France. https://www.researchgate.net/publication

72. Guellil I, Azouaou F, Abbas M, Fatiha S (2017) Arabizi transliteration of algerian arabic dialect into modern standard Arabic. In: Social MT 2017/First workshop on social media and user generated content machine translation

73. Guellil I, Azouaou F, Abbas M (2017) Comparison between neural and sta-tistical translation after translitera-tion of algerian arabic dialect. In: WiNLP: women & underrepresented minorities in natural language processing (co-located withACL 2017)

74. Shirko O, Omar N, Arshad H, Albared M. Machine translation of noun phrases from Arabic to English using transfer-based approach. J Comput Sci. 2010;6(3):350.

75. Samy D, Sandoval AM, Guirao JM, Alfonseca E (2006) Building a parallel multilingual corpus (arabic-spanish-english). In: Proceedings of the 5th Intl. Conf. on Language Resources and Evaluations, LREC

76. Shquier MMA, Atoum MS, Shqeer OMA. Arabic to English machine translation. New Trends Inf Technol. 2017;1:118.

77. Zbib R, Malchiodi E, Devlin J, Stallard D, Matsoukas S, Schwartz R, Makhoul J, Zaidan OF, Callison-Burch C (2012) Machine translation of arabic dialects. In: Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies, pp. 49–59. Association for computational linguistics

78. Salloum W, Habash N (2011) Dialectal to standard arabic paraphrasing to improve arabic-english statistical machine translation. In: Proceedings of the first workshop on algorithms and resources for modelling of dialects and language varieties, pp. 10–21. Association for computational linguistics

79. 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

80. Guellil I, Azouaou F (2017) Neural vs statistical translation of Algerian Arabic dialect written with Arabizi and Arabic letter

81. Banea C, Mihalcea R, Wiebe J, Hassan S (2008) Multilingual subjectivity analysis using machine translation. In: Proceedings of the conference on empirical methods in natural language processing, pp. 127–135. Association for computational linguistics

82. Balahur, A., Turchi, M (2012) Multilingual sentiment analysis using machine translation? In: Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis, pp. 52–60. Association for computational linguistics

83. Refaee E, Rieser V (2015) Benchmarking machine translated sentiment analysis for arabic tweets. In: Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: student research workshop, pp. 71–78

84. Koehn P, Hoang H, Birch A, Callison-Burch C, Federico M, Bertoldi N, Cowan B, Shen W, Moran C, Zens R, et al. (2007) Moses: Open source toolkit for statistical machine translation. In: Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions, pp. 177–180. Association for computational linguistics

85. Och FJ, Ney H. A systematic comparison of various statistical alignment models. Comput Linguist. 2003;29(1):19–511.

86. Heafield K (2011) Kenlm: faster and smaller language model queries. In: Proceedings of the sixth workshop on statistical machine translation, pp. 187–197. Association for computational linguistics

87. Papineni K, Roukos S, Ward T, Zhu WJ (2002) Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting on association for computational linguistics, pp. 311–318. Association for computational linguistics

88. Keyvanpour M, Zandian ZK, Heidarypanah M. Omlml: a helpful opinion mining method based on lexicon and machine learning in social networks. Soc Netw Anal Min. 2020;10(1):1–17.

89. Guellil I, Azouaou F (2016) Arabic dialect identification with an unsupervised learning (based on a lexicon). application case: Algerian dialect. In: Computational science and engineering (CSE) and IEEE intl conference on embedded and ubiquitous computing (EUC) and 15th intl symposium on distributed computing and applications for business engineering (DCABES), 2016 IEEE intl conference on, pp. 724–731. IEEE

90. Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment analysis is a big suitcase. IEEE Intell Syst. 2017;32(6):74–80.