

# Opinion mining and sentiment analysis for Arabic on-line texts: application on the political domain

Dhekra Najar<sup>1</sup> · Slim Mesfar<sup>1</sup>

Received: 10 March 2015 / Accepted: 22 April 2017 / Published online: 13 June 2017  
© Springer Science+Business Media New York 2017

**Abstract** Since the continuous proliferation of the journalistic content online and the changing political landscape in many Arabic countries, we started our current research in order to implement a media monitoring system about the opinion mining in political field. This system allows political actors, despite of the large volume of online data, to be constantly informed about opinions expressed on the web in order to properly monitor their actual standing, orient their communication strategy and prepare the election campaigns. The developed system is based on a linguistic approach using NooJ's linguistic engine to formalize the automatic recognition rules and apply them to a dynamic corpus composed of journalistic articles. The first implemented rules allow identifying and annotating the different political entities (political actors and organizations). Then these annotations are used in our system of media monitoring in order to identify the opinions associated with the extracted named entities. The system is mainly based on a set of local grammars developed for the identification of different structures of the political opinion phrases. These grammars are using the entries of the opinion lexicon that contain the different opinion words (verbs, adjectives, nouns) where each entry is associated with the corresponding semantic marker (polarity and intensity). Our developed system is able to identify and properly annotate the opinion holder, the opinion target and the polarity (positive or negative) of the phraseological expression (nominal or verbal) expressing the opinion. Our experiments showed

that the adopted method of extraction is consistent with 0.83 F-measure.

**Keywords** Opinion mining · Sentiment analysis · Arabic language · TAL · Politic · Web media

## 1 Introduction

Thanks to the democratic transitional phase that the Arab spring's countries are passing through, Arabic citizens are becoming more actively engaged in political issues and more and more relying on online newspapers (than the traditional media) to get informed of the latest news and events. It's a normal change that is due to the easy access and use of on-line contents. In fact, online medias publish constantly a large variety of political information and opinions. As a result, they began to have the power to influence the people's political decisions (elections, voting...) in a much stronger manner. That's why political actors consider that they must have knowledge about their actual standing on the web media (Online newspapers...). In this paper, we develop an opinion mining system capable of measuring, sentiments vis-à-vis political actors, in the content of web media using NooJ<sup>1</sup>). NooJ is a linguistic engine that is based on large coverage dictionaries and grammars (Silberztein 2015).

Opinion mining is a new challenging application in natural language processing domain that is born from the proliferation of the internet and online medias. It's the process that allows a user to analyze a huge amount of unstructured text data transmitted over the web in order to extract information related to the opinions and evaluations of the author

---

✉ Dhekra Najar  
dhekra.najar@gmail.com  
Slim Mesfar  
mesfarslim@yahoo.fr

<sup>1</sup> RIADI, ENSI, University de la manouba, Manouba, Tunisia

<sup>1</sup> <http://www.nooj4nlp.net/>.

as they are expressed vis-à-vis objects (a product, organization...) or a concept (a service, an individual, a decision...) (Bing 2010). This paper is divided into four parts. The first part deals with a discussion about the related works in opinion mining field. The second part of this paper begins by laying out the theoretical dimensions of the research in order to specify the adopted approach. The third part presents a description of the preconized approach based on the linguistic platform NooJ. The fourth part describes the evaluation and synthesis of our opinion mining system and some of challenges faced. Finally, we discuss the results and perspectives of our research.

## 2 Related work

There is a large volume of published studies describing various methods of opinion mining especially, in Latin languages. Several studies have led to the creation of lexical resources, in particular for English such as the dictionary WordNet-Affect (Strapparava and Valitutti 2004) which contains 1903 concepts related to mental or emotional and SentiWordNet (Esuli and Sebastiani 2006). For the French language, there is the lexicon of feelings developed by (Yannick 2005) which has a 1000 simple words expressing feelings, emotions and psychological states. It is an ontology containing words which are divided into 38 different classes. Previous cited studies have based their methods on using large coverage linguistic resources. In contrast, several studies have been introduced in the literature based only on machine learning techniques. Pang et al. (2002) tested different learning techniques (Naive Bayes, Maximum Entropy and SVM) in the field of ‘movie reviews’ in order to classify comments and advices into two classes (positive, negative). They showed that the use of maximum entropy classifier gave a higher value of F-measure. As for Wiebe et al. (1999), they used a naive Bayes classifier to determine whether a sentence is objective or negative. This system has obtained an F-measure equal to 81.5%. Thus, we mention the study of (Farra et al. 2010) that used a SVM classifier to automatically classify sentiments in the political field. The classification is based in two levels: at the document level or at the sentence level. This system has achieved 89.3% F-measure for classification at the sentence level and 87% for the document level. Generally, manually developed dictionaries and resources (SentiWordNet, WordNet-Affect...) are often combined with automatic techniques (SVM) in order to take advantage of machine learning approaches. In this context that the work of (Vernier and Monceaux 2009) attempt to construct automatically a lexicon of subjective terms from 5000 blog posts and comments. This method relies on the indexation of the

language constituents (adjectives, adverbs, nominal and verbal expressions) by using a web search engine and a large number of queries.

On the other hand, few amount of literature has been published on sentiment analysis in Arab world. Most of the studies are mainly based on machine learning techniques. In a study of (Abdul-Mageed and Diab 2012b) which is an extension of their work (Abdul-Mageed and Diab 2012a, b), the authors present the system AWATIF which is an Arabic multi-genre annotated corpus for sentiment and subjectivity analysis. The corpus is tagged using two methods of sourcing. It can be used to assist in the construction of subjectivity and sentiment analysis systems, such as the attempt of (Abdul-Mageed and Diab 2012b). We note that this corpus is not available to the public. In another major and recent study, (Abdul-Mageed et al. 2014) have developed the system SAMAR which is a supervised machine learning system for subjectivity analysis in Arabic language. The advantage of this study is that the authors have created a multi-genre corpus (from four types of social media) of a text written in Standard Arabic (MSA) and Dialectal Arabic (DA). In fact, this is the first research that addresses the dialectal Arabic language (Egyptian dialect). The authors have created manually a lexicon of 3982 adjectives. Each of the adjectives is marked as positive, negative or neutral adjective. For classification, they have adopted a two-phased approach: (objective/subjective classification and positive/negative classification). Indeed, most of the proposed automatic approaches in Arabic language are based on the semantic similarities between words in order to classify the words with unknown polarity. For example, (Amira et al. 2013) propose an automatic approach for emotion detection in Arabic texts. This approach relies on the construction of a moderate size lexicon that contains emotions and feelings terms to annotate stories for children. SVM model aims to classify these stories into six basic emotions (joy, fear, sadness, anger, disgust and surprise) by calculating the degree of similarity between words and emotions already annotated. This approach has achieved 64.5% F-measure for the six emotions. According to the recent study of (Alaa El-Dine and Fatma El-zahraa 2013), the authors have developed an annotated corpus in Arabic language for sentiment analysis. This corpus is used to classify new comments. Then, they used a various learning algorithms (decision tree, SVM...) to implement the sentiment analyzer. The best result was obtained by the SVM classifier with an F-measure reached 73.4%. We also include the study of (Abdul-Mageed et al. 2011) who annotated a corpus of 200 documents from the Penn Treebank Arabic (Maamouri et al. 2004), which is composed of subjective journalistic texts. They applied three

**Table 1** Summary table of Arabic opinion mining studies

Studies	Meth.	Corpus	Advantages	Disadvantages
Abdul-Mageed et al. (2011)	ML	Journalistic	200 annotated documents corpus Many classifiers F-measure: 99.48% (SVM)	Only classification on the phrase level MSA only
Farra et al. (2010)	ML (SVM)	Journalistic	Bilevel classification (phrase and document) Generalization of the opinion phrase structure F-measure: 89.3% on the phrase level et 87% on the document level	A small amount of test data MSA only
AWATIF (Abdul-Mageed and Diab 2012a)	ML	Annotated multi-type	Classification of the subjectivity et the sentiments F-measure: 71.44	Only classification on the phrase level
SAMAR (Abdul-Mageed et al. 2014)	ML	Social media	Building a lexicon of 3982 adjectives, manually Classification of the subjectivity et the sentiments	The system does not take into account neutral and mixed polarity DA and MSA
Amira et al. (2013)	ML (SVM)	Stories for kids	Construction of sentiment lexicon Considering the similarity between words Classification according to six emotions F-measure: 64.5%	Moderate volume lexicon MSA only
Alaa El-Dine and Fatma El-zahraa (2013)	ML	Annotated corpus	More than a method of automatic classification The SVM classifier has obtained the best result: F-measure: 73.4%	MSA only

ML machine learning, SVM support vector machine, MSA modern standard Arabic, DA dialectal Arabic

different learning algorithms on the corpus to automatically classify sentences. This approach has achieved a very high F-measure which is equal to 99.48% with the SVM classifier (Table 1).

We notice that the most of the works in Arabic opinion mining field has mainly used machine learning methods in order to classify sentiments in Arabic texts. We also notice that the opinion analysis is confined to classify the extracted opinion sentence into two polarities (positive or negative) without taking into account the other opinion variables (opinion holder, opinion target, intensity...).

### 3 Implementation

The approach we follow for developing our Arabic opinion mining system is mainly based on a linguistic approach. The named entities recognition is a potentially important pre-processing for the opinion mining field. However, this task represents a serious challenge, given the specificities of the Arabic language. For this purpose, we have adapted a rules-based approach to recognize the Arabic named entity and political organization, using different grammars and gazetteer.

NooJ linguistic engine is based on large coverage dictionaries and grammars. It uses finite state transducers (FSTs) to parse text corpora made up of hundreds of text files in real time and associate each recognized entry with its related information, such as morpho-syntactic information (POS—part of speech, gender, number, etc.), syntactic and semantic information (e.g., transitive, human, etc.). NooJ is a well known linguistic environment (see <http://www.nooj4nlp.net>) that is already used to formalize more than 20 languages. Nevertheless, we used to make some specific tools to be able to deal with the specificities of the Arabic language (Mesfar 2010). For the following parts of this work, we will use the Electronic Dictionary for Arabic “*El-DicAr*” resources (Mesfar 2008) as the basis for our sentiment analysis system.

The recognition process consists on identification of lexical entities using dictionaries and grammars, and the transformation of grammars into transducers. For the realization of our process, we will adopt the spiral life cycle model witch manage risks and evaluate all alternatives in each step. Changes can be introduced later in the life cycle as well.

**Table 2** Number of the opinion dictionary entries

Category	Number of entries	Polarity		Total
		Positive	Negative	
Verbs	127	46	71	117
Nouns	480	225	335	560
Adjectives	256	132	124	256
Total		403	530	933

### 3.1 Linguistic resources

We begin with building a new lexicon representing the different opinion and political vocabularies (verbs opinions, adjectives, nouns). Each opinion word is integrated in the dictionary associated with a set of linguistic and semantic information: grammatical category, gender and number, syntactic information. We have extracted 933 subjective terms from our training corpus, which are commonly used in the journalistic media. Since we have limited our research in one field (politics), we consider these rates high. We note that the nouns represent more than half of all entries in the opinion lexicon (60%), followed by adjectives (27.5%), verbs (12.5%). All the recognized subjective words in our dictionary are associated with the corresponding semantic markup: **+Polarity=pos** for positive terms and **+Polarity=neg** for negative terms (Table 2).

We show in the Table 3 below some examples of entries in our opinion mining lexicon and grammars:

We have also extracted some neutral verbs that express opinions depending on the constituents of the phrase in which they appear. To determine the polarity of the phrases containing these verbs, we use syntactic rules describing their structure. We give some examples of neutral verbs such as (تَظَاهَرَ, to manifest), (أَعْلَنَ, to announce), (أَعْرَبَ, to express)... These verbs can express positive or negative expressions all depends on the phrase's constituents.

### 3.2 Named entities recognition: ENAMEX

The political opinion mining application first requires that the corpus annotation using political entities and parties. To do that, we developed series of grammars in order

to annotate, in a first stage, the political organizations and actors.

The approach we take for Named Entity recognition is a rule based one which is quite similar to that used by (Mesfar 2008). The named entity recognition module is able to find mentions of persons, locations and organizations, as potential opinion targets. For this purpose, The NER system is based on the use of lists containing gazetteers and lexical markers that are known beforehand and have been classified into named entity types. We also use lists of trigger words which indicate that the surrounding tokens are probably named entity constituents and may reliably permit the type of the political named entity to be determined (minister, president...). Lists of triggers, such as (رئيس, president) and (سيد, sir) for political actors recognition and (وزارة, Ministry) (حركة, Movement) for political organizations recognition, were produced manually from the training corpus. The triggers (43 political actors and 54 political organizations) are tagged as result of the morphological analysis, and are used in named entity grammar rules. A syntactic grammar represents word sequences described by manually created rules, and then produces some kind of linguistic information such as function of the recognized political actor. We note that all NooJ's syntactic codes are described in Annex (Fig. 1).

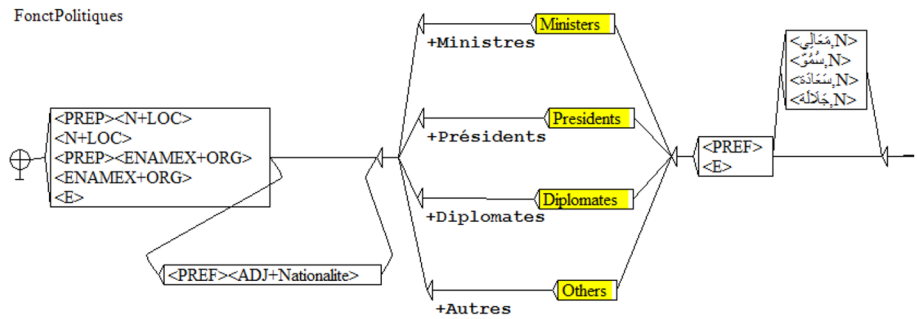
Beyond the recognition of the political actors name's and their functions, we will also save the recognized sequence in a variable to be used later. We will use this NooJ's functionality to store the name of the named entity in the variable "NOM" and its political function (president, minister...) in the variable "FONCTION". We will use these variables later in our opinion mining system in order to identify the different variables of an extracted opinion (Fig. 2).

The NER grammar "ENAMEX PERS" will be launched in the linguistic analysis of the corpus. As shown in the example below, its role is to enhance the corpus by associating annotations to its various textual segments: named entities, political organizations. The generated annotations are used in the description of syntactic rules for the identification of opinions in the journalistic texts.

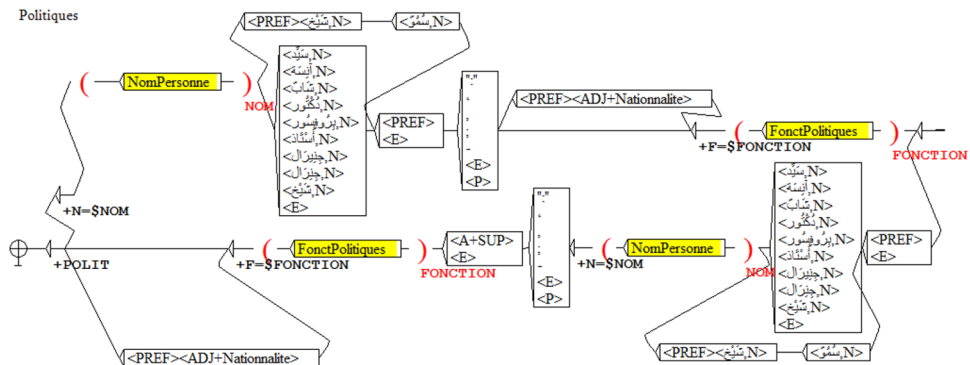
**Table 3** Examples of the lexicon's entries

Category	Polarity: positive	Polarity: negative
Subjective Verbs	(مَدَحَ, to praise), (طَوَّرَ, to develop), (أَيَّدَ, to support)...	(اِسْتَشْكَى, to complain), (اِسْتَنْكَرَ, to denounce), (فَشَلَ, to fail)...
Subjective Nouns	(رِبْحَ, profit), (اِنْتِصَارَ, victory), (تَنْمِيَّةَ, development)...	(اِنْتِهَاكَ, violation), (هُجُومَ, attack) (اِهْمَالَ, negligence)...
Subjective Adjectives	(مُحَادِدَ, neutral), (مُنْصِيفَ, fair), (عَاقِلَ, sane)...	(اِسْتَبْدَادِيَّ, tyrannical), (دِيكْتَاتُورِيَّ, dictatorial), (خَطِيرَ, dangerous)...

**Fig. 1** ENAMEX NooJ syntactic grammar (28 subgraph)



**Fig. 2** ENAMEX NooJ syntactic grammar



The subgraph “NomPersonne” contains a combination of context for the categorization of a person’s name in a text. In order to formalize the rules, we will use the list of names and family names that exist in “Al-DicAr”. They are recognizable using the semantic markup (N + Prenom) or (N + Prenom + Compound).

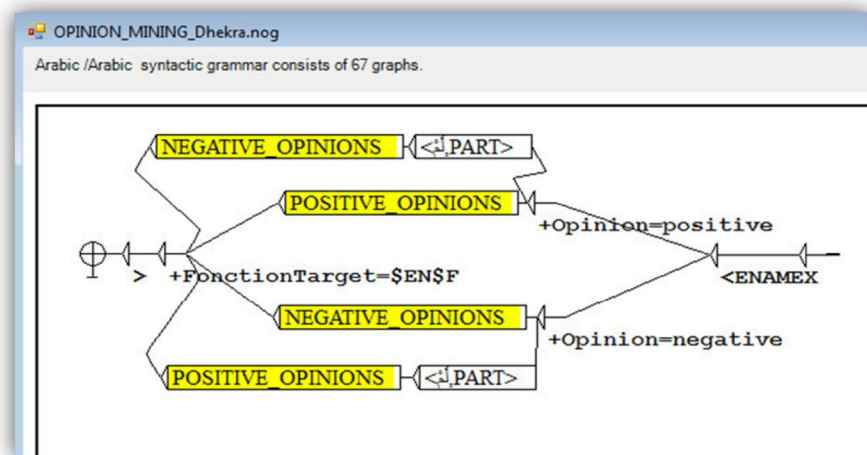
**3.3 Opinion mining grammar**

The approach we take for sentiment analysis is mainly based on the entries of our opinion lexicon (933 opinion entries) and the generated annotations of the named entities grammar. Initially, we try to collect the maximum of morphological and structural information for contextual forms of the opinion sentences in Arabic language. In

this purpose, we need to deeply understand the morphological nature of Arabic political texts. Indeed, we need to study the syntactic–semantic organization of political Arabic text-structure. There are several challenges that complicate this task in Arabic language. Those difficulties are essentially due to the complexity and the specific characteristics of Arabic such as the vocalization and the agglutinated words that complicate the syntactic analysis. Texts in web contents (especially in social media) make use of noisy language with capitalization, unusual punctuation, use of hashtags and derivations. This lack of precision in texts can give rise to ambiguities. A second challenge is that people do not always express opinions the same way (each author expresses the information in its own way). However, based on the training corpus, we have identified three general structures of opinion sentences (either verbal sentences or nominal or sentences containing a relative pronoun). To better explain this, we take the example below that express the same opinion phrase and can be written in three different grammatical structures.



**Fig. 4** Opinion mining syntactic grammar (67 subgraph)



Our system is able to extract and classify the expressions of opinion in journalistic texts and identify their different variables.

## 4 Evaluation

Data were collected using a program of online extracting regular journalistic texts. Then, the texts are analyzed using NooJ's linguistic engine. The studied corpus is composed of a set of journalistic articles published during the period between 05/12/2013 and 12/08/2013: At this time we loaded 1000 articles from different Arabic web media.

Traditionally, the scoring report compares the answer file with a manually annotated file. The system was evaluated in terms of the complementary precision and recall metrics. Briefly, precision evaluates the noise of a system while recall evaluates its coverage. These metrics are often combined using a weighted harmonic called the F-measure.

### 4.1 Challenges of Arabic language processing using NooJ

Compared to the big amount of available resources and lexicons in Latin languages, particularly English, the Arabic sentiment analysis domain is still immature. Unfortunately, the lack of Arabic opinion mining resources that are publicly available makes the progress of sentiment analysis in this language very difficult.

Several obstacles make the analysis of Arabic text-structures so complicated: the high inflectional nature (uses internal patterns for its grammatical processes), the agglutinated form of pronouns and prepositions, the variant

sources of ambiguity (unvowled journalistic texts, rich metaphoric script), the dual forms for pronouns and verbs, and the absence of the upper case in the beginning of named entities. These specificities of Arabic language represent the most challenging problems for Arabic NLP researchers.

Finally, another important limitation regarding Arabic sentiment analysis field may result from the fact that the absence of a robust general grammatical structures of Arabic opinion phrases makes it difficult to extract opinion sentences hence harden the extraction of its variables. Arabic expressions in news articles are generally so long.

### 4.2 Evaluation of electronic dictionaries

To test the lexical coverage of our dictionary, we launch the linguistic analysis of our corpora. The linguistic analysis shows that the corpus contains about 57,957 different forms in which we have recognized about 57,303 forms (654 unknown forms). In other words, the result of lexical analysis shows that the vocabulary of the corpora is recognized to 98.3% by our lexical and morphological resources. In fact, the non-recognition is due to two main reasons or the absence of certain words in our dictionaries or the frequent mistakes in journalistic texts. For example:

- False vocalization of words such as (misplaced vowels);
- Common typographical errors such as confusion between Alif and Hamza or the substitution of (ة, ة) and (ي, ي) at the end of the word;
- The false writing of Hamza (جاءوا, جاؤوا, they arrived);
- The addition or omission of a character in a word;
- The lack of white between two terms as (وهذا, and this);
- The transcription of foreign names. (ديفيد David);
- Neologisms such as (ديپلوم, Diploma) (twitter, تويتر).

**Table 4** ENAMEX grammar experiments on our corpus

Precision	Recall	F-measure
0.86	0.77	0.81

4.2.1 Discussion

The recognition rate in the corpus is high. This shows that the developed resources *El-DicAr* in the thesis of (Mesfar 2008) as well as our specialized political lexicons are rich and wide. They can probably cover the journalistic discourse used in articles. This will ensure a maximum of lexical recognition for our opinion mining system.

4.3 Evaluation of the NER grammar

To evaluate our NER local grammars, we analyze our corpus to extract manually all the named entities (the whole named entity’s sentence is considered: name of the person and his function). Then, we compare the results of our system with those obtained by our manual extraction. The application of our local grammar gives the following result (Table 4):

The table above presents the recall and precision and F-measure obtained by the application of the NER grammar on our corpus. It is clear that we have reached a

reasonable result of recognition with 0.81 F-measures. This result is encouraging given the rate achieved by the systems participating in MUC. In fact, in our system, we consider only the information completely extracted (named entities and their functions).

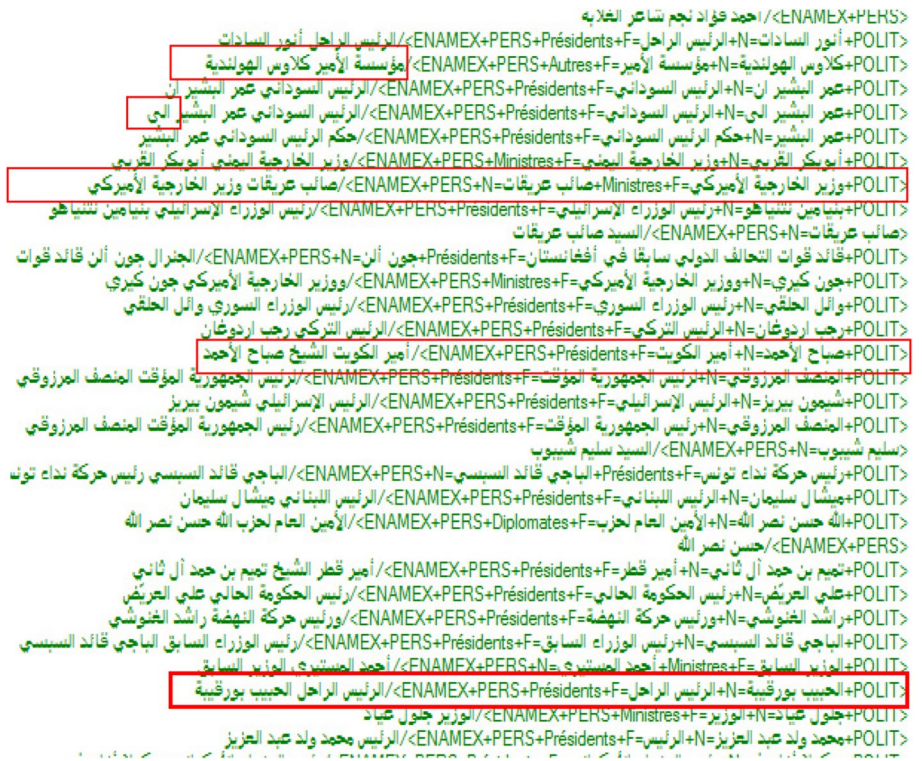
In the figure below there are several sources for error. The main error is the bad formalization of the recognition rules. Concerning the silence in our NER module, it is often due to the absence of some recognition rules; naive indices (upper case) and triggers. This explains why the rate of recall is low 0.77%. Other reasons, we can mention the obstacle of the transcription into Arabic language (different variants of a single word). The application of the NER grammar gives the results below.

As we see in the correlation table, the grammar is able to extract the name of the political entity as well as his function (job) (Fig. 5).

This is an example of an annotation generated by our system as shown in the figure below (the example circled in bold red):

- الرئيس الراحل الحبيب بورقيبة.

**Fig. 5** ENAMEX + PERS’s correlation table







#### 4.4.1 Automatic annotation

/<ENAMEX+Opinion=positive+Structure=Verbal+OpinionHolder=جاكوب زوما+Target=نيلسون مانديلا+FonctionTarget=رئيس جنوب أفريقيا الأسبق >

#### 4.4.2 Translation of the matching sequence

The Current President Jacob Zuma has praised the previous President of South Africa Nelson Mandela.

#### 4.4.3 Translated annotation

/<ENAMEX + Opinion = positive + Structure = Verbal + OpinionHolder = Jacob Zuma + Target = Nelson Mandela + FonctionTarget = previous President of South Africa >.

As shown in the correlation table and the example below, our developed syntactic resources are able to distinguish and identify all the opinion elements:

- The opinion target: can only be a political actor (recognizable through grammar ENAMEX + PERS).
- The opinion holder: can be a political actor (ENAMEX + PERS), an organization (ENAMEX + ORG), a country (N + LOC) or other sources of opinion holders such as (الشعب, the people) (المتظاهرين, protesters) (المواطنين, citizens).
- Features: we identified a few explicit features associated with political features such as (جيش بشار الأسد, the army of Bashar al-Asad), (نظام بشار الأسد, the system of Bashar Al-Asad) (أنصار بشار الأسد, supporters of Bashar al-Asad).
- The polarity of opinion: we classified the extracted opinion sentences using annotations: = Positive Opinion for positive sentences and = negative for negative sentences.
- Intensity of opinion: we distinguished between adverbs of intensity and frequency adverbs that come to reinforce or reduce the degree of polarity. Then, we annotated and classified adverbs of intensity and frequency according to their intensity (+Intensity = HIGH, +Intensity = LOW...)
- The political function of the opinion target.

In fact, the target of the opinion and polarity are two obligatory variables for the categorization and classification of an opinion sentence. The other variables are complementary and optional depending on the structure of the opinion sentence. We note that in our system, it is useful to implement the task of identifying the source of the opinion (which can be a person or organization).

#### By visualizing the extracted terms in concordance tables

after the application of our syntactic grammars, we see the emergence of some false sequences. This is due to:

- Problems related to the procedure for recognition of named entities;
- Problems related to ambiguity of Arabic language (journalistic texts are usually unvowled);
- A word can be both a personal name and an entry in our lexicon of opinion. This can lead the system to extract noisy information (noise).

#### 4.4.4 Discussion

Errors are often due to the complexity of opinion sentences and the presence of the agglutinated form of pronouns and prepositions in Arabic language that can restrain recognizing opinion expression's structures. The errors are due also to the absence of some structures in our system (and in the learning corpus). In fact, the Arabic sentences in news articles are usually very long, which set up obstacles for opinion mining analysis. Despite the problems described above, the developed method seems to be adequate and shows very encouraging extraction rates. However, other rules must be added to improve the result.

## 5 Conclusion and perspectives

Information, thoughts and opinions are shared prolifically on the social web these days. Tracking these opinions in web contents using NLP tools has become a crucial step in order to make sense of all the information that could influence the public's choices.

Through this paper, we have described a new method of political opinion mining system.

With this system, we have successfully analyzed sentiment in Arabic language and achieved its main tasks; the development of Arabic subjectivity linguistic resources and lexicons, the opinion polarity identification, Opinion elements identification task. Our experiments showed that the adopted method of extraction is consistent. We have also developed a system allowing the identification of NEs (organizations and political actors) in Arabic texts.

More broadly, research is also needed to compress the long sentences in news articles, which can set up obstacles

for further opinion mining steps. Sentence compression is the task of producing a brief summary at the sentence level.

## 6 Appendix

See Table 6.

**Table 6** NooJ's syntactic categories

Syntactic codes	
<ADJ>	Adjective
<V>	Verb
<N>	Noun
<ADV>	Adverb
<CONJ>	Conjunction
<PREP>	Preposition
<PREF>	Prefix
<PRON>	Pronoun
<REL>	Relative pronoun
<PART>	Particle
<E>	Empty character
<P>	Punctuation sign
Inflectional codes	
<s>	Singular
<p>	Plurial
<m>	Male
<f>	Female
Semantic codes	
<LOC>	Location
<Nationalite>	Nationality
<ENAMEX+ORG>	Organisation named entity
<ENAMEX+PERS>	Person named entity

### Compliance with ethical standards

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

## References

Abdul-Mageed, M. & Diab M. (2012a). AWATIF: A multi-genre corpora for modern standard arabic subjectivity and sentiment analysis. *LREC*, Istanbul, Turkey

- Abdul-Mageed, M. & Diab, M. (2012b). Toward building a large-scale arabic sentiment lexicon. In *Proceedings of the 6th International Global WordNet Conference*.
- Abdul-Mageed, M., Diab, M. T., & Korayem, M. (2011). Subjectivity and sentiment analysis of modern standard Arabic. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers* (vol. 2, pp. 587–591). Association for Computational Linguistics.
- Abdul-Mageed, M., Diab, M., & Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social media. *Computer Speech & Language*, 28(1), 20–37.
- Alaa El-Dine, A. H., & Fatma El-zahraa, E.-t. (2013). Sentiment analyzer for Arabic comments system. *International Journal of Advanced Computer Science and Applications*.
- Amira, F. E. G., Torky, I. S., Maha, A. H., Mohamed, M. E. D. (2013). A computational approach for analyzing and detecting emotions in Arabic text. *International Journal of Engineering Research and Applications (IJERA)*, 3, 100–107.
- Bing, L. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, (pp. 627–666). Boca Raton, FL: CRC Press.
- Esuli, A. & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *LREC06*.
- Farra, N., Challita, E., Assi, R. A., & Hajj, H. (2010). Sentence-level and document-level sentiment mining for Arabic texts. In *IEEE International Conference*, (pp. 1114–1119).
- 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).
- Mathieu, Y. Y. (2005). Annotation of emotions and feelings in texts. In *International Conference on Affective Computing and Intelligent Interaction* (pp. 350–357). Berlin: Springer.
- Mesfar, S. (2008). *Analyse morpho-syntaxique et reconnaissance des entités nommées en arabe standard*. Thèse, Université de franche-comté, France.
- Mesfar, S. (2010). Towards a cascade of morpho-syntactic tools for Arabic natural language processing. In *Proceedings of CICLing 2010, LNCS 6008*, (pp. 150–162).
- Pang, B., Lee, L. & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Association for Computational Linguistics, (pp. 79–86).
- Strapparava, C. & Valitutti, A. (2004). WordNet-Affect: An affective extension of WordNet. In *Proceedings of LREC* (vol. 4, pp. 1083–1086).
- Silberztein, M. (2015). *La formalisation des langues : l'approche de NooJ*. ISTE Ed.: Londres (p. 426).
- Vernier, M. & Monceaux, L. (2009). Enrichissement d'un lexique de termes subjectifs à partir de tests sémantiques.
- Wiebe, J. M., Bruce, R. F. & O'hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics (pp. 246–253).