



HAL
open science

Opinion Mining from Arabic Quotations

Motasem Alrahabi

► **To cite this version:**

Motasem Alrahabi. Opinion Mining from Arabic Quotations. 6th International ACM Conference on Management of Emergent Digital Ecosystems, 2014, Buraidah Al Qassim, Saudi Arabia. 10.1145/2668260.2668291 . hal-03199841

HAL Id: hal-03199841

<https://hal.science/hal-03199841>

Submitted on 16 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Opinion Mining from Arabic Quotations

Motasem Alrahabi

► **To cite this version:**

Motasem Alrahabi. Opinion Mining from Arabic Quotations. 6th International ACM Conference on Management of Emergent Digital Ecosystems, 2014, Buraidah Al Qassim, Saudi Arabia. hal-03199841

HAL Id: hal-03199841

<https://hal.archives-ouvertes.fr/hal-03199841>

Submitted on 16 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Opinion Mining from Arabic Quotations

Motasem Alrahabi

Paris-Sorbonne University in Abu Dhabi

Abu Dhabi - UAE

almoatasem.alrahabi@psuad.ac.ae

ABSTRACT

We present a contribution to the Arabic Opinion Mining area with the development of a new system for quotations detection and categorization. Our approach differs from the existing work by proposing an operational tool that enables search for keywords in the categorized quotations. The categorization is performed with a rule-based system and is based on the reporting markers' meaning. The system end-user interface enables two query types within the categorized quotations: searching for keywords in the part containing eventually the quotation source (*reporting clause*) or in the part concerning the topic (*reported clause*). Our system was applied to process a corpus of Arabic newspaper articles. Both evaluations were carried out with the following results: 88 of recall and 93 of precision for the identification task; 85 of recall and a 93 of precision for the categorization task. The system is publicly accessible on the web site (<http://e-quotes.net>).

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications: *Data mining*; H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing: *Linguistic processing*; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval: *Information filtering*; I.2.7 [Artificial Intelligence]: Natural Language Processing: *Text analysis*

General Terms

Algorithms, Experimentation, Languages

Keywords

Direct reported speech, opinion mining and sentiment analysis, Arabic language, semantic categorization, rule-based system.

1. INTRODUCTION

The Reported Speech (RS) is an important linguistic phenomenon which is characterized by its syntactic structure: a matrix clause, containing usually a reporting marker, and a subordinate clause that embeds the conveyed information [1]. Among the various forms of RS (*direct speech, indirect paraphrases, direct speech introduced by "that"...*), we are particularly interested in the direct RS or quotations. Many text mining applications use quotations to analyze, organize and summarize information because they are a major vehicle of communication in the news genre. We believe

that a system that automatically identifies and categorizes quotations would enable readers, journalists and researchers to place news in the context of all comments made on a given topic, and especially to know how these comments were interpreted and reported in the media.

Our work is a contribution in this direction since we aim to automatically detect and categorize Arabic quotations according to semantic and enunciative criteria. For this purpose, we rely on the *reporting markers* that specifically denote a verbal communication (*reporting verbs and their derived gerunds or nouns...*). These markers are usually used by the enunciator in order to report the quotation, to exhibit the speaker's attitude and eventually to indicate his own position towards the reported content. Let us consider this sentence:

ولحسن الحظ اعترف أوباما 'بعدم قدرة أميركا لوجدها على ضمان الأمن والسلام العالميين'.

Fortunately, Obama acknowledged "the inability of the United States to ensure international peace and security alone."

the different elements are analysed as follows:

Element	Label
أوباما <i>Obama</i>	<i>Source (speaker)</i>
اعترف <i>Admit</i>	<i>Reporting marker + Speaker modality marker</i>
ولحسن الحظ <i>Fortunately</i>	<i>Enunciator modality marker</i>

Our primary contributions making our current research significant are: i) Developing linguistic resources (markers and rules) to identify and categorize quotations from texts in Arabic; ii) creating a simple and operational working implementation which allows the user to directly query the annotated corpus by both classical and semantic criteria.

The remainder of this paper is organized as follows: we first show how difficult Opinion analysis from quotations can be (§2). We then describe our proposed method (§3) and give an overview of the system (§4). In (§5), we present evaluation results and discuss them. We present in (§6) the related work and finally, in (§7) we draw our conclusions and future work.

2. OPINION MINING CHALLENGES FROM DIRECT REPORTED SPEECH

The RS is a standardized way to relate opinion, sentiment or attitude expression of a certain *Source* regarding a certain *Target*. Inspired by Banfield, Uspensky and Quirk, [2] considers that Subjectivity refers to aspects of language used to express opinions, feelings, evaluations, and speculations, including sentiments. Therefore, from a computational point of view, current researches distinguish between subjectivity and objectivity in opinions along with determining these elements: Opinion *polarities* which tell us whether the opinion's orientation or valence is positive, neutral, negative or, sometimes, mixed; the

opinion *strength* (attitude's degree, i.e., low, medium, high); opinion *holder* (the people who express opinion), and opinion *target* (the object of this opinion).

Nevertheless, characterizing the opinions and sentiments analysis from quotations remains challenging for at least these three reasons: the target, the source and the expressed opinion:

- **Target:** in Opinion Mining (OM) over movie, product or book reviews, the target or topic is clearly identified. On the contrary, the target in news articles is not a concrete object [3] because when a text is argumentative and when it opposes different points of view on a topic, journalists may span larger subject domains, more complex event descriptions and a whole range of targets [4]. Thus, identifying a concrete target that can be resolved back to named entities does not work for quotations [5], because quotes may not necessarily mention the debate topic (*implicit targets*), and there may be many relevant targets for a single topic (*mixed speeches, selective and partial targets*).
- **Source:** the opinion source (holder) identification aims to extract entities that express opinions in texts [6]. There are many major challenges in the task of automatically attributing each quote to its correct speaker [7]. Sometimes, the source may not be located near the quotation, so syntactic parsing and named entity recognition may be necessary. The use of pronouns is also common, such that anaphora and co-reference resolution are needed to determine the name of the source. In the case of quotations, a source can be a title or role (*Prime Minister*); personal name (*Vladimir Poutine*); pronominal reference (*she said...*); anonymous (*a passenger, a witness...*), etc.. The source of quotation is not always explicitly mentioned, as in the following cases in Arabic:

إليك الخبر الآتي: ...
 أتانا ما يلي: ...
 جاء في البيان أن...
 نص القرار على...
 وقال أن... (*implicit pronoun*)

For Arabic language, several challenges complicate the opinion source identification [8]: the lack of resources, the high inflectual nature of Arabic language, the variant sources of ambiguity, the rich metaphoric script usage and the absence of robust Arabic parser that understand the sentence structure [9].

- **Expressed opinion:** most work on OM has been carried out on subjective text types such as blogs and product or movie reviews where individuals express their opinions quite freely. On the contrary, the position of the journalist in relation to what s/he reports in newspaper articles is often more subtle [10], because the authors of newspaper articles try to make their articles to look objective concerning the topics they are covering. In these cases opinion or sentiment is most of the time not expressed explicitly in the text. This issue makes it often hard to directly identify sentiment in the news. However, journalists continue to keep a large flexibility to exhibit their attitude in relation to what they are reporting: it goes for instance by highlighting some facts while omitting others, but especially by the choice of words to introduce the RS, to show his own point of view and to describe the position of the different actors of the original utterance situation.

For all these reasons, OM from quotations may not guarantee perfect results. Our current aim in this study is not trying to tackle all of these complex issues, but to focus our efforts on the last point (*expressed opinion*) for Arabic language, i.e. how a quotation is reported and interpreted by the enunciator?

3. THE PROPOSED APPROACH

We describe here the different aspects of our approach for quotations identification and categorization.

First, we distinguish between the opinions expressed within the reported content (*inside quote, the speaker's opinion*) and those attributed directly to the enunciator. For this purpose, we refer in our analysis to the observable reporting markers (*which are lexically expressed*) in order to categorize quotations and distinguish between the speaker opinion and the enunciator opinion. In the enunciative approach [11], [12], the enunciator (*utterer*) is the entity that reports the whole speech (*generally the author*), whereas the speaker (or *locutor*) is the last source or holder of the speech. For instance, when enunciator uses the verb *to claim*, he reports the purpose of the speaker, in addition, he shows his own position by doubting the speaker's credibility.

3.1 Semantic categorization

Moreover, our added value is that the analysis covers a larger scale of phenomena which are not easily classifiable when using only the categories of positive and negative.

The enunciator can use the mechanism of RS not only to reproduce the original utterance, but also to interpret it and to give other information using at least three types of markers:

- Reporting markers such as *according to, to confirm, declaration, etc.*
- Modality markers (modalizers) that indicate:
 - the position of the enunciator towards the speech act (*unfortunately she admitted..*) or towards its specifications (*by adding this short / unacceptable comment...*);
 - the intersubjective relation: *X criticizes Y, X encourages Y, X informs Y, etc.*
 - the attitude of the speaker: *He said, fawning / with skepticism... / He cries... / She whispers...*
- circumstantial information that clarify the original utterance speech act: spatio-temporal and audience settings [1], theme or topic (*concerning /about...*), communication medium (*He said in a letter...*).

The corpus examination¹ allowed us to identify more than 150 reporting markers [13], which have been manually listed and stored with their derived forms. We organized these markers in several dozens of categories and sub-categories into a semantic map (*linguistic ontology*) where we have taken into consideration intersubjective, semantic and discursive parameters:

- the axis "Good – Bad": *criticize, encourage, threat...*
- the axis "True – False": *confirm, deny, believe...*
- the axis "Information": *declare, explain, diffuse...*
- the axis "Will": *order, decide, solicit...*
- the axis "Speech organization": *add, answer, summarize..*
- ...

¹ Our corpus is a collection from internet-based Arabic media (Al-Jazeera, BBC Arabic, CNN Arabic, Al-Nahar, Arabic LMD...).

For our current experiment, we only covered from this semantic map the “Good – Bad” axis. This part contains two categories and it is organized as follows:

Table 1. Selected semantic categories

Enunciator’s opinion			Speaker’s opinion		
Neg.	Pos.	Neutr.	Neg.	Pos.	Neutr.
Claim	...	Disengagement	Humiliation	Apology	Observation
...		...	Complaint	Praise	Definition
			Criticism	Encourag ^{nt}	Declaration
			Mockery	...	Transmission
			Accusation		...
			Contestation		
			...		

3.2 Markers and structures of Direct RS

Direct RS can be expressed in various ways. We consider, on the formal level, that a quotation is any kind of speech delimited by meta-characters (*the typographical signs of quotation*) and introduced by, at least, one reporting marker referring to an act of communication, whether the speaker is explicitly defined or not. By convention, we consider the quotation span as a verbatim transcription of the source utterance, despite the existence of rare cases where the quoted words are not certainly attributed to the speaker, as in: *Has he really said “...” ? / He did not say “...” / He probably said “...”*, or in the following case:

ولسان حال الشاب الذي يتقدم لزواجها يقول: "لا يصلح العطار ما أفسد الدهر".
It's as if the young man said: "the perfumer cannot fix what time has spoiled"

Here are some examples of the several different possible constructions in Arabic with verbal, nominal or adverbial reporting markers:

يشير فلان الى أن « ... »
 و يؤكد فلان: « ... »
 « ... » يضيف فلان.
 ويحتج فلان قائلاً: « ... »
 ...
 في ما يلي تصريح فلان: « ... »
 أرسل فلان التعليق الآتي: « ... »
 ...
 بحسب فلان: « ... »
 « ... » بحسب ما يدعي / قول فلان
 ...

Sometimes, one or more intermediate entities can be part of the transmission chain between utterer and speaker. We call this entity the “transmitter” (T):

I heard from T that X said... / According to T, X said...

Thus, we can formalize a quotation (Q) by this general formula:

Q ≡ [Time] [Place] [audience] [Topic] [Medium]
 [Transmitter(s)] [Speaker] [Hearer]
 [Reporting marker(s)] [modalizer(s)]
 (Reported word(s))

The position of each element within a text can vary and, except the quoted words, all other parameters are optional.

4. SYSTEM OVERVIEW

In this section, we describe the system pipeline configuration and how it is deployed in practice. For processing resources, we use EXCOM-2 [14], a rule-based system that performs annotations by using surface markers and heuristic rules. Annotated texts are then indexed for rapid retrieval at query time with Solr search engine platform.

4.1 Corpus preparation

Technically, to annotate a corpus, EXCOM-2 needs one pre-treatment phase of segmentation (*splitting*). It helps in determining the search fields for linguistic markers, and the textual snippets which are to be annotated. This consists in defining by heuristic rules the boundaries of sections, titles, paragraphs and sentences. For this, all corpus documents have to be normalized and converted to raw texts files in UTF-8 encoding002E

4.2 Annotation: Quotations recognition and categorization

EXCOM-2 algorithm is based on the Contextual Exploration (CE) method [15]; the tool does not deal with any preliminary morpho-syntactic analysis or named entities recognition. The starting point of the system is the corpus preparation and its core is the semantic annotation task. In our perspective, we consider two types of surface markers: indicators and clues. The presence of a potential marker [16] in the search space triggers the associated CE rules, and then, additional markers (*clues*) are searched in a specified context. If all the rule conditions are satisfied, the segment specified by the rule is annotated. For our processing, we consider quotations marks as indicators and reporting markers as clues.

Different types of rules can be implemented in EXCOM-2, depending on the research space or the type of markers (*linguistic units, regular expressions, text structure tags...*). The tool enables to use the already annotated segments, to use the text structure (*titles, paragraphs...*), to sort the rules according to their importance and to use *negative clues* that cancel certain rules.

For our task, the annotation of each semantic category requires the creation of three rules on average.

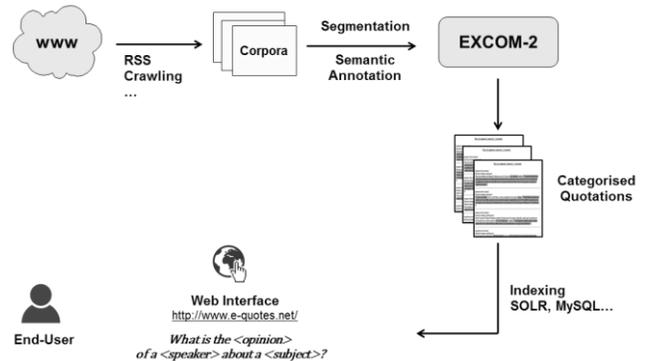


Figure 1. General architecture of E-Quotes.

A basic version of EXCOM-2 is available online at this address: <http://www.excom.fr/>.

4.3 Indexing and Web Interface

For the information retrieval module, we use a typical Vector Space Model based system (*Apache Solr*) and we store all annotated XML documents in an inverted index that enables flexible search for keywords in all quotations of the corpus.

E-Quotes end-user web interface supports search in many different ways:

- Search for keywords in all identified quotations;
- Search for keywords in specific identified and categorized quotations. For example, search for a word in a *negative* quotation and more especially in a *negative* quotation that has the value « *accusation* »;
- Search for keywords in the reporting clause or in the reported clause of a specific category of quotations.

Since we do not proceed to the recognition of speakers or targets, the last feature allows user to find answers to such a question:

*What is the <opinion> of a <speaker>
about a <subject>?*

where the parameters <speaker> and <subject> are specified by users as combination of keywords or entities. The list of <opinion> categories is extracted from the semantic map.

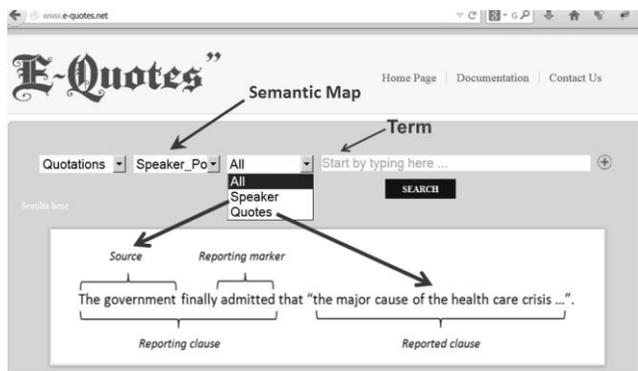


Figure 2. E-Quotes homepage.

The system supports also boolean combinations of multiple fields, i.e. AND, OR, NOT.

5. EVALUATION

We conducted two evaluations by computing traditional measures in order to test the capacity of our system to *identify* and to *categorize* quotations. We also performed a detailed analysis of error cases introduced by our system and their root causes.

5.1 Quotations Recognition Evaluation

We randomly selected out 21 new documents from online newspaper articles which include more than 1000 sentences and 25000 words. Topics covered in these articles are mostly political, economic, social news and events. We then annotated these texts with EXCOM-2 that identified exactly 271 quotations. In parallel, we asked an expert (*journalist*) to read the selected articles and to highlight manually only the snippets she judges as quotations. After the comparison, we obtained the following results: 88 for Recall and 93 for Precision.

5.1.1 Identification task results analysis

We conducted a manual inspection over all evaluation documents to identify detection errors. Here are the Recall result analysis:

- The value of silence is due to the fact that some markers are not yet covered by our resource base:

- Nominal markers and gerunds derived from reporting verbs: *declaration, by saying...*).
- Passive construction markers like: *قيل, سمي / It was said ...*
- Self-quotation markers: *انتهى الى مسامعنا... / We have heard...*

- The quotation is introduced by some markers indicating the speaker's attitude but that are not reporting markers:

أما المعنيون فيهزون أكتافهم غير مبالين: "أنظروا ليس هنا سوى أشخاص مسالمين ينشطون من أجل مصلحة المجتمع".

*Concerned people shake their shoulders, without caring:
"look, there is no one but peaceful persons who work for the sake
of their society".*

Concerning the Precision rate, we can mention the cases of misleading quotation marks and the polysomic reporting markers.

- A large number of noise is usually caused by the presence of misleading quotation marks (*quotation marks that do not surround quotes*). Example:

لكن المنظمة الاسرائيلية لحقوق الانسان "بيتسالم" نشرت تقريراً يؤكد رواية شهود العيان الفلسطينيين.

But the Israeli human rights organization "B'Tselem" published a report confirming the Palestinian eyewitnesses' version.

- In Arabic, the surface forms are generally polysemics [17], especially the forms that have a three-letter root (عبر, بين, علق, ...). This difficulty is due to the morphological ambiguity in Arabic, caused, above all, by the absence of vocalization, the agglutination and the relatively free word order in a sentence. Here is an example for wrongly assigned quotation, caused by the presence of a polysemic reporting marker (نقلت) in the context of misleading quotation marks:

ونقلت سيارات الاغاثة التابعة للمنظمة "مواد إغاثة وطبية" إلى داخل المدينة.
The cars of the Relief Organization have transported « relief and medical materials » to the center of the city.

We also mention the difficulty of nested quotation marks, where a quotation can contain another one. This case can produce errors in the annotation.

5.2 Quotations Categorization Evaluation

To obtain a preliminary assessment of the categorization task, we carried out a limited evaluation, mostly to guide our future efforts. Thus, we simplified the annotators' task by only testing if a given quotation belongs or not to the main categories: Positive, Negative or Neutral. A deeper evaluation layer could be certainly necessary to determine the sub-category of a given one.

For this second evaluation, we annotated with EXCOM-2 a new set of data and selected randomly 20 annotated quotations from each category of the speaker's opinion. We then obtained 60 mixed annotated quotations.

Three Arabic native speakers with language-related academic background tagged each quotation and decide whether the text snippet (*reporting clause*) is being talked about in a positive or in a negative light, or if the statement is rather neutral. For this, we

decided to hide the contents of the quotes (*reported clause*) to ensure that annotators will judge based only the words of the enunciator, without mixing with the words of the speaker. The conflicts of tagging were resolved using majority voting principle and the average final agreement is 87% between annotators. The system achieved a Recall value of 85 and a Precision value of 93.

5.2.1 Categorization task results analysis

Taking into account the complexity of the analysis, we consider the overall results to be rather good.

The major difficulty encountered in the categorization task is the mixed and nested opinions [16]. In fact, different cases are envisaged:

- Same source (speaker), several opinion markers. In the following example, the quotation should be annotated as *definition* and *denunciation*:

وقد استنكر المسؤولون الفلسطينيون إطلاق النار هذا وسّمّوه "رصاص صوب الشمس".

Palestinian reponsible have condemned the shooting and called it as "bullets directed towards the sun".

- Different sources, different opinion markers. Here's an example:

وصف قائد المجلس العسكري الأعلى تصريح "أوباما" بإرسال الأسلحة للمعارضة بأنه "تصريح شجاع جدا".

The commander of the Supreme Military Council described the declaration of "Obama" to send weapons to the opposition as "a very brave declaration".

It's difficult here to decide automatically to which speaker attribute this or that opinion.

We need to refine our linguistic analysis in order to improve the attribution rules and to have better results.

6. RELATED WORKS

Quotation extraction has been previously approached using different techniques and for several languages. But, to our knowledge, there are only few operational systems that detect quotations from Arabic texts, and even less for the opinion mining task from quotations [18].

NewsExplorer [19] is developed in the European Commission's Joint Research Centre². This tool detects quotations from multilingual live news feeds, including Arabic. The system is able to extract quotes, the name of the entity making the quote and also entities mentioned in the quote. According to the authors, the system recognizes quotations only if it successfully detects three parts: the speaker name, the reporting verb and the quotation. For each person in the system's database, the most recent quotations from and about the person are listed on this person's dedicated information page. For the English language evaluation, the system aimed for high precision (87.5%) at the expense of low recall, as their data contained many redundant quotes.

[20] propose a quotation extraction and attribution tool from English newspapers. The system is implemented in GATE and combines a lexicon of 53 common reporting verbs and a hand-built grammar to detect constructions that match 6 general lexical patterns. The authors evaluate their work on 7 newspaper articles from the WSJ, which contain 133 quotations. For the detection of reporting verb and source, the system achieved a recall value of

0.79 and a precision value of 1.00, thus an F-measure of 0.88. For quotation span detection the results are: 99% of precision and 74% of recall.

Google's InQuotes application³ allows users to search for quotes made - in English - by a small selected set of politicians. The web-based interface is structured in topics and displays side-by-side quotes from two actors. Users can search for any keywords in the search area and quotes containing the keywords would be returned. They do not enable search on the speaker itself other than from the selected set, and no implementation details is published about this system.

[21] describe SAPIENS, a system that relies on a deep linguistic processing chain (NE extraction, anaphora resolution, deep parsing...) in order to extract quotations from French news with their author and context. The evaluation was carried out both for the span of the quotation and for the correctness of the author. The evaluation found that 19 out of 40 quotes had a correct span and author, while a further 19 had an incorrect author, and 4 had an incorrect span.

We can observe that most of the prior approaches deliberately choose to focus on the more frequent syntactic structures and on limited lists of reporting markers. In the other hand, all these works carry out a pre-recognition of quotations' sources (*holders or speakers*) and retain only the quotations where the speaker is identified unambiguously. Finally, none of these systems applies linguistic and semantic analysis like we do in order to categorize automatically quotations according only to the reporting markers.

7. CONCLUSION AND FUTUR WORKS

Our system makes the semantic information explicit and accessible for end-users. We demonstrated it by adapting the standard IR technologies (i.e. keywords queries matched against bag-of-words document representation) to semantically tagged natural texts. By indexing semantic annotations using such a keyword search engine, we provide a highly scalable and fast semantic search capability by enabling users to search for quotes made by a particular person or about an entity. Each quotation is categorized according to the opinion of the source (speaker) and eventually to that of the enunciator.

The used method is simple and does not require morpho-syntactic pre-processing or NER. For the categorization task we achieve a recall rate of 83% and a precision of 98%.

As future work, we envision to do the following:

- Extend the lexical resources with new markers like adjectives (*doubtful, boring...*); adverbs (*finally, unfortunately...*) and gerunds (*laughing, shouting...*). This allows to have more fine-grained categorization and to analyze the intensity of opinions (*strong, medium or weak...*).
- Evaluate the impact of using markers that modify the polarity of an expressed opinion such as valence shifters (*negations, intensifiers...* [22]), connectives or even modals.
- Classify the content of quotes (*the reported clause*). This feature will give us a complete vision of the polarity of each quotation. Using a classifier could also help us to assign topic tags to each quote.

² <http://press.jrc.it/NewsExplorer>

³ <http://labs.google.com/inquotes/> (*deprecated*).

- Extend the analysis of RS in Arabic and cover indirect, mixed and unmarked RS forms [23].
- Last, integrate the annotation module as a webservice that can be automatically queried by the user interface in order to directly process new submitted documents in different formats.

The application is publicly available at the address: <http://e-quotes.net>.

8. REFERENCES

- [1] Bergler, S., 1993. Semantic Dimensions in the Field of Reporting Verbs, in *Proceedings of the Ninth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary and Text Research*, Oxford, England, September 1993.
- [2] Wiebe, J. 1994. *Tracking point of view in narrative*. Computational Linguistics, 20.
- [3] Pang, B. and Lee, L. 2008. "Opinion mining and sentiment analysis". In *Foundations and Trends in Information Retrieval*, Vol. 2, Nos. 1-2, pp. 1-135, 2008.
- [4] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., and Belyaeva, J. 2010. *Sentiment analysis in the news*. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 2216–2220.
- [5] O’Keefe, T., Curran, J. R., Ashwell, P., Koprinska, I. 2013. An Annotated Corpus of Quoted Opinions in News Articles. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*
- [6] Lun-Wei, K., Chia-Ying, L. and Hsin-Hsi, C. 2009. Identification of Opinion Holders. *International Journal of Computational Linguistics and Chinese Language Processing*, volume 14, number 4, pages 383-402, December 2009.
- [7] Paret, S. 2012. The independent encoding of Attribution Relations. In *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-8)*, Pisa, October 3-5.
- [8] Elarnaoty, M., AbdelRahman, S., and Fahmy, A. 2012. A Machine Learning Approach For Opinion Holder Extraction Arabic Language. *CoRR*, abs/1206.1011, 2012.
- [9] Habash, N., 2010. *Introduction to Arabic Natural Language Processing*, Synthesis Lectures on Human Language Technologies, Graeme Hirst, editor. Morgan & Claypool Publishers. 187 pages.
- [10] Balahur, A., Steinberger, R., Van der Goot, E., Pouliquen, B., and Kabadjov, M. 2009. *Opinion mining on newspaper quotations*. In 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pages 523–526. IEEE.
- [11] Johansson, M., Suomela-Salmi, E. 2011. Enonciation: French pragmatic approach(es) in Zienkowski, J., J-O Östman & J. Verschueren (eds) *Discursive Pragmatics*. Amsterdam / Philadelphia : John Benjamins: 71-101.
- [12] Desclés, J.-P., Guentchéva, Z. 2000. "Enonciateur, locuteur, médiateur" In *Les Rituels du dialogue : Promenades ethnolinguistiques en terres amérindiennes* / A. Monod Becquelin, P. Erikson (eds.) -- Nanterre : Société d’ethnologie (Recherches thématiques 6), 2000, p.79-112
- [13] Alrahabi M., Desclés, J.-P., Suh J.-Y. 2010, Direct Reported Speech in multilingual texts: Automatic annotation and semantic categorization, in *Proceeding of FLAIRS 2010*, Florida, USA
- [14] Alrahabi M., Desclés, J.-P. 2009, EXCOM : Plate-forme d’annotation sémantique de textes multilingues, in *TALN 2009*, Senlis, 24-26 June 2009
- [15] Desclés J.-P. 2006. Contextual Exploration Processing for Discourse Automatic Annotations of Texts. Actes de FLAIRS 2006, 281-284.
- [16] Wiebe, J. Wilson, T., Bruce, R., Bell, M., and Martin, M. 2004. Learning subjective language. *Computational Linguistics* 30.
- [17] Dichy, J., 2000. Morphosyntactic Specifiers to be associated to arabic lexical entries-Methodological and theoretical aspects, in *Proceedings of the ACIDCA’2000, Volume Corpora and Natural Language Processing, Monastir-Tunisie*.
- [18] Korayem, M., Crandall, D. J., and Abdul-Mageed, M. 2012. Subjectivity and Sentiment Analysis of Arabic: A Survey, in *Advanced Machine Learning Technologies and Applications*, AMLTA 2012: 128-139.
- [19] Pouliquen B., Steinberger R. and Best C. 2008. Automatic Detection of Quotations in Multilingual News. In *Proceedings of the 6th International Conference on Natural Language Processing, GoTAL 2008*. August 25-27, 2008, Gothenburg, Sweden.
- [20] Krestel R., Bergler S., and Witte R. 2008. Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles. *Proceedings of the Sixth International Language Resources and Evaluation (LREC 2008)*, May 28-30, 2008, Marrakech, Morocco.
- [21] De la Clergerie, E., Sagot, B., Stern, R., Denis, P., Recource, G., and Mignot, V. 2009. Extracting and visualizing quotations from news wires. In *Proceedings of L&TC 2009*, Poznan, Poland.
- [22] Polanyi, L. and Zaenen, A. 2004. "Contextual valence shifters". *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, 2004.
- [23] Paret, S., O’Keefe, T., Konstas, I., Curran, J. R., and Koprinska, I. 2013. Automatically Detecting and Attributing Indirect Quotations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, U.S.