

# E-Quotes: Enunciative Modalities Analysis Tool for Direct Reported Speech in Arabic

Motasem Alrahabi

Paris-Sorbonne University in Abou Dhabi,  
Abou Dhabi - UAE  
motasem.alrahabi@gmail.com

**Abstract.** With rapidly growing Arabic online sources aimed to encourage people's discussions concerning personal, public or social issues (*news, blogs, forums...*), there is a critical need in development of computational tools for the Enunciative Modalities analysis (*attitude, opinion, commitment...*). We present a new system that identifies and categorizes quotations in Arabic texts and proposes a strategy to determine whether a given speaker's quotation conveys some enunciative modalities and potentially its evaluation by the enunciator. Our system enables two query types search for keywords within the "categorized" quotations: searching for keywords in the part potentially containing the reported speech source (*the reporting clause*) or searching for keywords in the part concerning the topic (*the reported clause*). The annotation is performed with a rule-based system using the reporting markers' meaning. We applied our system to process a corpus of Arabic newspaper articles and we obtained promising results for the evaluation.

**Keywords:** Direct reported speech, Enunciative Modalities, Opinion Mining, Sentiment Analysis, categorization, Arabic language, rule-based system.

## 1 Introduction

The Reported Speech (RS) is an important linguistic phenomenon characterized by its syntactic structure: a matrix clause usually containing a reporting marker, and a subordinate clause embedding 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 (*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 tool which identifies and semantically categorizes quotations would enable readers, journalists or researchers to place news in the context of all comments made on a given topic, and specifically to know how these comments were interpreted in the media.

Our work heads in this direction since we aim to automatically detect and categorize Arabic quotations according to the enunciative modalities. The automatic identification and interpretation of modalized statements in texts is a major concern in a large number of applications, especially with the recent attention to Opinion Mining,

Semantic Analysis and Appraisal Theory [2]. In our approach, enunciative modalities concern the manner in which the enunciator reports, interprets and evaluates the words of the speaker (*disengagement, commitment, attitude...*). It also concerns the manner by which the speaker expresses an attitude towards his interlocutor and towards the contents of his utterance (*commitment, control, opinion, judgment...*). For this purpose, we rely on the semantic *reporting markers* that introduce and modalize the reported words. Let us consider this sentence:

ولحسن الحظ اعترف أوباما "بعدم قدرة أميركا لوحدها على ضمان الأمن والسلام العالميين".  
*Fortunately, Obama acknowledged "the inability of the United States alone to ensure international peace and security."*

The different elements are analyzed as follows:

Element		Label
أوباما	<i>Obama</i>	<i>Speaker (Source)</i>
اعترف	<i>acknowledged</i>	<i>Speaker commitment</i>
ولحسن الحظ	<i>Fortunately</i>	<i>Enunciator (author) positive attitude</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 an operational application which allows users to directly query an annotated corpus by both classical and semantic criteria.

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

## 2 Automatic Analysis Challenges from Direct Reported Speech

We situate our current work on Direct RS in the field of Opinion Mining and Sentiment Analysis which we consider as a part of modalities studies. In fact, 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, the author of [3] 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 research distinguishes 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 or person expressing the 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 [4] because when a text is argumentative and when it opposes different points of view, journalists may span larger subject domains, more complex event descriptions and a whole range of targets [5]. Thus, identifying a concrete target that can be resolved back to named entities recognition (NER) does not work for quotations [6], because quotes may not necessarily mention the debate topic (*implicit targets*), and there may be multiple 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 [7]. There are many challenges in the task of automatically attributing each quote to its correct speaker [8]. Sometimes, the source may not be located near the quotation, so syntactic parsing and NER may be necessary. The use of pronouns is also common, so that anaphora and co-reference resolution are needed to determine the name of the source. In the case of quotations, a source can be title or role (*Prime Minister*); proper name (*Vladimir Poutine*); pronominal reference (*she said...*); anonymous (*a passenger, a witness...*). In the following cases in Arabic, the source of quotation is not explicitly mentioned:

إلېكم الخبر الآتي: ...	* Take the following news: ...
أنا ما يلي: ...	* We got the following: ...
جاء في البيان أن...	* It came in the report that...
نص القرار على...	* The decision consisted of ...

For Arabic language, several challenges complicate the opinion source identification [9]: 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 [10].

- **Expressed Opinion:** Most work on OM has been carried out on subjective text types such as product reviews, blogs or even social media [11], 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 [12], because the authors of newspaper articles try to make their articles look objective concerning the topics they are covering. In these cases, opinion or sentiment is not always expressed explicitly in the text. However, journalists try to remain flexible in exhibiting their attitude towards what they report : it goes for instance by highlighting some facts while omitting others, but especially by the choice of words to introduce the RS and to describe the position of the different actors of the original utterance situation.

For all these reasons, the OM and more generally the semantic analysis for 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 Our Proposed Approach

Here, we will describe the different aspects of our approach for quotations identification and categorization.

#### 3.1 Markers and Structures of Direct RS

We consider, on the formal level, that a Direct RS is any kind of speech delimited by meta-characters (*the typographical signs of quotations*) 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 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 the following case:

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

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

يشير فلان الى أن "...".	<i>X points that "..."</i>
يوكد فلان: "...".	<i>X affirms: "..."</i>
"...", "اضاف فلان.	<i>* "... X added.</i>
يحتج فلان قائلاً: "...".	<i>X objects saying: "..."</i>
في ما يلي تصريح فلان: "...".	<i>Following is the declaration of X: "..."</i>
أرسل فلان التعليق الآتي: "...".	<i>X sent the following comment: "..."</i>
بحسب فلان: "...".	<i>According to X: "..."</i>
"..." بحسب ما يقول فلان.	<i>"..." according to what X is saying.</i>
...	...

Sometimes, one or more intermediate entities can be part of the transmission chain between the utterer and the speaker. We call this entity the "transmitter". In the following sentence, (*Anadolu news agency* / وكالة أنباء الأناضول) is the speaker, while (*the local authorities* / السلطات المحلية) is the transmitter.

في غضون ذلك ذكرت وكالة أنباء الأناضول نقلاً عن السلطات المحلية "تدفق مئات اللاجئين السوريين في الساعات الأخيرة إلى تركيا".

Meanwhile, *Anadolu news agency* mentioned, according to the local authorities :  
*"the flow of hundreds of Syrian refugees in the last hours to Turkey."*

By ignoring the aspecto-temporal parameters in this analysis, the standard meta-linguistic formula of a *reporting act* can be expressed by *operators* acting onto *operands*:

**I-SAY (T<sub>n</sub>-SAYS (X-SAYS (λ) to X') to T<sub>n</sub>') to YOU**

where the operator "I-SAY" represents the original utterer's act of speaking. "T<sub>n</sub>" represent the transmitter(s) and "X" the speaker (*the source*). The symbols "YOU", "T<sub>n</sub>'" and "X'" represent respectively the interlocutors of the Enunciator (I), the Transmitter (T) and the Speaker (X). Finally, the "λ" represents the reported speech.

### 3.2 Semantic Categorization

We distinguish between the enunciator modalities and those attributed by the enunciator to the speaker (*in this current work we don't analyze the content of quotes*). In the enunciative approach [13], the enunciator (*utterer*) is the entity that reports the whole speech (*generally the author*), whereas the speaker (*or locutor*) is the last source or speech holder. For instance, when an enunciator uses the verb *to claim*, he attributes a modality to the speaker (*commitment*), in addition, he exhibits his own position towards the speaker's credibility.

Moreover, our added value is that the analysis covers a larger scale of semantic phenomena which are not easily classifiable when using only the categories of positive and negative opinions. Actually, the enunciator can use the mechanism of RS not only to reproduce the original utterance, but also to interpret it and to provide other information using several types of markers:

**i) Reporting markers** that introduce the quotes (*according to, to inform...*);

**ii) Modality markers** or modalizers [14] that indicate:

- the position of the enunciator towards the speech act (*unfortunately she admitted..*) or towards its characteristics (*by adding this short comment...* );
- the attitude of the speaker (*he said with skepticism that ...*);
- circumstantial information which clarifies the speech act of the original utterance: spatio-temporal and audience settings [1], theme or topic (*concerning / about...*), communication medium (*He said in a letter...*).

In our current analysis, we only refer to the observable *reporting markers* which are lexically expressed in texts. An empirical examination of the corpora<sup>1</sup> allowed us to identify more than 150 reporting markers. The latter have been manually listed and organized with their derived forms (*gerunds or nouns*) into a semantic map (*linguistic ontology*) which includes, for instance, the following categories:

- Neutral: *say, observe, define...*
- Positive opinion: *encourage, praise...*
- Negative opinion: *criticize, denounce...*
- Commitment: *confirm, affirm, believe...*
- Disengagement: *deny, refute...*
- Control: *order, decide, refuse...*
- Speech organization: *add, ask, answer, conclude...*

Due to the polysemy phenomenon, a given marker can naturally belong to one or more categories. The general enunciative formula will become more complex to receive the operators **OP** which represent the different values of the semantic map:

**I-SAY(OP<sub>I</sub>(T<sub>n</sub>-SAYS(OP<sub>T</sub>(X-SAYS(OP<sub>X</sub>(λ))to X'))to T<sub>n</sub>'))to YOU**

The enunciative dimension of our analysis permits us to associate to each set of markers whether the speaker is “sending” information (*X informs X'...*), “receiving” information (*X reads in the journal that...*), “transmitting” information (*X reports that...*) or even he is totally absent as a source (*I heard that...*).

---

<sup>1</sup> Our corpus is a collection from internet-based Arabic media (Al-Jazeera, BBC Arabic, CNN Arabic, Al-Nahar, Le Monde Diplomatique in Arabic...).

## 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 [15], a rule-based system that performs annotations by using surface markers and heuristic rules. Annotated texts are indexed for rapid retrieval at query time with Solr search engine platform.

### 4.1 Corpus Preparation

The starting point of the system is the corpus preparation. Technically, to annotate a corpus, EXCOM-2 needs a 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 encoding.

### 4.2 Annotation: Quotations Recognition and Categorization

The core of the system is the semantic annotation task. In our perspective, we consider two types of surface markers: “indicators” and “clues”. The presence of a potential indicator in the search space triggers the associated CE rules, and then, additional clues are searched in a specified context. If all the rule conditions are satisfied, the segment specified by the rule will be annotated and assigned one or more semantic values. 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 re-search 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 can inhibit certain rules. For our current task, the annotation of each semantic category requires the creation of three rules on average.

EXCOM-2 is based on the Contextual Exploration (CE) method [16]. The tool does not deal with any preliminary morpho-syntactic analysis or NER. We believe that this system can be an addition to these approaches, not a substitution (*for example, identifying the quotations sources by NER*). A basic version of the tool is available online at this address: <http://www.excom.fr/>.

### 4.3 Indexing and Web Interface

The output of the annotation processing pipeline is indexed using the Apache Solr framework, which is based on the Lucene engine. We store all annotated XML documents in an inverted index that enables flexible search for keywords in all annotated quotations.

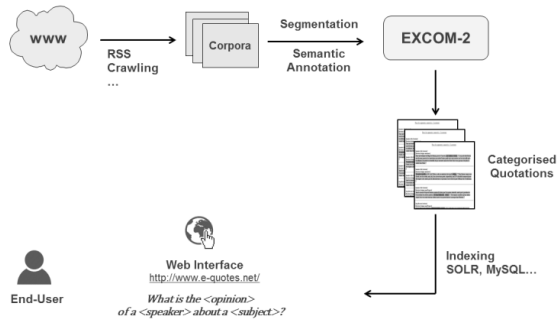


Fig. 1. General architecture of E-Quotes

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

- Search for keywords in all identified quotations;
- Filter the quotations according to a specific category and then search for keywords in one of its sub-categories. For example, the user can search for a word in all the quotations that are annotated as “Negative Opinion” and more especially in those that hold the value “Accusation” ;
- Filter the quotations according to a specific category or sub-category and then search for keywords in one of the two options:
  - the space containing the quotation’s topic (*the reported clause*) ;
  - the space containing potentially the quotation’s source (*the reporting clause*).

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 <attitude> of a <speaker> towards a <subject>?**

where the parameters <speaker> and <subject> are specified by the user as combination of keywords or entities. The list of <attitude> (*or enunciative modalities*) categories is automatically extracted from the semantic map and proposed to the user.

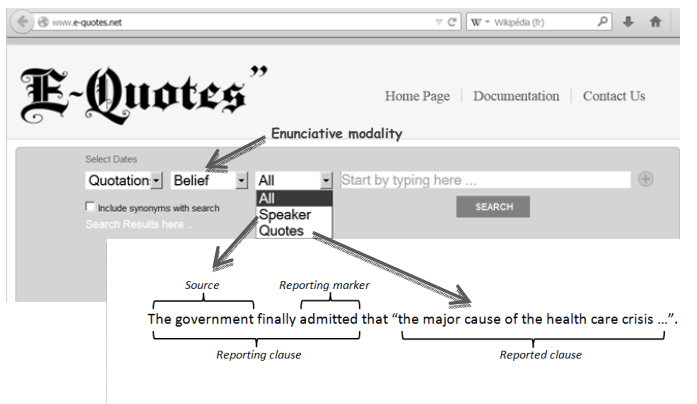


Fig. 2. E-Quotes homepage

The system supports 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 21 new documents from online newspaper articles which included 1049 sentences and over 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 269 quotations. In parallel, we asked three Arabic native speakers with language-related academic background to read the selected articles and to highlight manually only the snippets they judge as quotations. After the comparison, we obtained the following results: 89 for recall and 97 for precision.

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

i - The value of silence is due to the fact that some markers are not yet covered by our resource collection, such as nominal markers and gerunds derived from reporting verbs: *declaration, by adding...*

ii - Some quotations are introduced by markers indicating the speaker's attitude but are not considered as reporting markers:

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

*The people concerned shake their shoulders, indifferently: "Look, there is no one but peaceful persons working for the sake of their society".*

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

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

لكن المنظمة الاسرائيلية لحقوق الانسان "بينتسالم" نشرت تقريراً يؤكد رواية شهود العيان الفلسطينيين.  
*But the Israeli human rights organization "B'Tselem" published a report  
confirming the Palestinian eyewitnesses' version.*

ii - In Arabic, the surface forms are generally polysemics [17], especially the forms that have a three-letter root (عبر, بين, علق, شرح). 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 of wrongly assigned quotation, caused by the presence of a polysemic reporting marker (نقلت) in the context of misguiding quotation marks:

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

We also mention a difficulty caused by the nested quotation marks, where a quotation contains 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 only tested the following categories of the semantic map: *positive opinion*; *negative opinion*; *commitment* and *disengagement*. For this evaluation, we used the same corpus of the first evaluation, and we selected only the well annotated quotations that belong to the aforementioned four categories. We then obtained 57 quotations.

The same three evaluators were asked to tag each quotation and decide whether the text snippet (*the reporting clause*) is being talked about one or more of our categories. Thus, we decided to hide the contents of the quotes (*the reported clause*) to ensure that the annotators will judge based only on the words of the enunciator, without mixing them with the words of the speaker. The conflicts of tagging were resolved using majority voting principle and the average final agreement was 84% between annotators. The system achieved a recall value of 87 and a precision value of 94.

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. In fact, different cases can be found:

- One source (*speaker*), several opinion markers. In the following example, the quotation should be annotated as a Definition and a Denunciation:

وقد استنكر المسؤولون الفلسطينيون إطلاق النار هذا وسّمّوه "رصاص صوب الشمس".  
*The Palestinian officials have condemned this shooting and called it "bullets targeting 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".*

This issue makes it often hard to automatically decide to which speaker the system attributes this or that opinion. At that point, we need to refine the linguistic analysis in order to improve the attribution rules.

## 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 OM or enunciative modalities task from quotations [18].

NewsExplorer [19] is developed in the European Commission's JRC<sup>2</sup>. 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

<sup>2</sup> <http://press.jrc.it/NewsExplorer>

quotation. 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, 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 application<sup>3</sup> 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 are 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. On the other hand, all these works carry out a pre-recognition of quotations' sources (*holders*) and retain only the quotations where the speaker is identified unambiguously. Finally, none of these systems applies enunciative and semantic analysis like we do in order to automatically categorize quotations according to the reporting markers.

## 7 Conclusion and Future 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 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 potentially 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 87% and a precision of 94%. As future work, we envision to do the following:

- Extend the lexical resources with new markers such as adjectives (*doubtful, boring...*); adverbs (*finally, unfortunately...*) and gerunds (*laughing, shouting...*).

---

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

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 and analyze 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.
- 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>.

## References

1. Bergler, S.: Semantic Dimensions in the Field of Reporting Verbs. In: The 9th Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary and Text Research, Oxford (1993)
2. Martin, J.R., White, P.R.R.: The Language of Evaluation: Appraisal in English. Palgrave Macmillan, London & New York (2005)
3. Wiebe, J.: Tracking point of view in narrative. *Computational Linguistics* 20 (1994)
4. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1-2), 1–135 (2008)
5. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news. In: 7th International Conference on Language Resources and Evaluation, pp. 2216–2220 (2010)
6. O’Keefe, T., Curran, J.R., Ashwell, P., Koprinska, I.: An Annotated Corpus of Quoted Opinions in News Articles. In: 51st Annual Meeting of the Association for Computational Linguistics (2013)
7. Lun-Wei, K., Chia-Ying, L., Hsin-Hsi, C.: Identification of Opinion Holders. *International Journal of Computational Linguistics and Chinese Language Processing* 14(4), 383–402 (2009)
8. Pareti, S.: The independent encoding of Attribution Relations. In: Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-8), Pisa (2012)
9. Elarnaoty, M., AbdelRahman, S., Fahmy, A.: A Machine Learning Approach For Opinion Holder Extraction Arabic Language. In: CoRR, abs/1206.1011 (2012)
10. Habash, N.: Introduction to Arabic Natural Language Processing. In: Hirst, G. (ed.) *Synthesis Lectures on Human Language Technologies*, 187p. Morgan Kaufmann (2010)
11. Abdul-Mageed, M., Kuebler, S., Diab, M.: SAMAR: A system for subjectivity and sentiment analysis of social media Arabic. In: 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), ICC Jeju, Republic of Korea (2012)
12. Balahur, A., Steinberger, R., Van der Goot, E., Pouliquen, B., Kabadjov, M.: Opinion mining on newspaper quotations. In: 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 523–526. IEEE (2009)

13. Desclés, J.-P., Guentchéva, Z.: Enonciateur, locuteur, médiateur. In: Monod Becquelin, A., Erikson, P. (eds.) *Les Rituels du dialogue: Promenades ethnolinguistiques en terres amérindiennes*, pp. 79–112. Société d'ethnologie (Recherches thématiques 6), Nanterre (2000)
14. Alrahabi M., Desclés, J.-P., Suh J.-Y.: Direct Reported Speech in multilingual texts: Automatic annotation and semantic categorization. In: *FLAIRS 2010, Florida* (2010)
15. Alrahabi, M., Desclés, J.-P.: EXCOM: Plate-forme d'annotation sémantique de textes multilingues. In: *TALN 2009, Senlis* (2009)
16. Desclés, J.-P.: Contextual Exploration Processing for Discourse Automatic Annotations of Texts. In: *FLAIRS 2006, Florida* (2006)
17. Dichy, J.: Morphosyntactic Specifiers to be associated to arabic lexical entries-Methodological and theoretical aspects. In: *ACIDCA 2000, Volume Corpora and Natural Language Processing, Monastir* (2000)
18. Korayem, M., Crandall, D.J., Abdul-Mageed, M.: Subjectivity and Sentiment Analysis of Arabic: A Survey. In: *Advanced Machine Learning Technologies and Applications, AMLTA* (2012)
19. Pouliquen, B., Steinberger, R., Best, C.: Automatic Detection of Quotations in Multilingual News. In: *The 6th International Conference on Natural Language Processing, GoTAL 2008, Gothenburg* (2008)
20. Krestel, R., Bergler, S., Witte, R.: Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles. In: *6th International Language Resources and Evaluation (LREC 2008), Marrakech* (2008)
21. De la Clergerie, E., Sagot, B., Stern, R., Denis, P., Recource, G., Mignot, V.: Extracting and visualizing quotations from news wires. In: *L&TC 2009, Poznan* (2008)
22. Polanyi, L., Zaenen, A.: Contextual valence shifters. In: *AAAI Spring Symposium on Exploring Attitude and Affect inText: Theories and Applications* (2004)
23. Pareti, S., O'Keefe, T., Konstas, I., Curran, J.R., Koprinska, I.: Automatically Detecting and Attributing Indirect Quotations. In: *Empirical Methods in Natural Language Processing (EMNLP), Seattle* (2013)