Automatic Annotation of Direct Reported Speech in Arabic and French, According to a Semantic Map of Enunciative Modalities

Motasem Alrahabi and Jean-Pierre Desclés

Maison de la Recherche – Université de Paris-Sorbonne
28, Rue Serpente, 75006 Paris, France
{motasem.alrahabi,jean-pierre.descles}@paris4.sorbonne.fr

Abstract. We present an analysis of the linguistic markers of the enunciative modalities in direct reported speech, in a multilingual framework concerning Arabic and French. Furthermore, we present a platform for automatic annotation of semantic relations, based on the Contextual Exploration method. This platform allows the automatic annotation and categorisation of quotational segments in both languages, exploiting a semantic map based on the notion of speaker commitment in enunciation.

Keywords: Automatic annotation, multilingual approach, commitment, semantic map, Contextual Exploration.

1 Introduction

Reported speech, both in the form of direct quotation and indirect paraphrases, is the most frequent expression found in newspapers, where it can occur in up to 90% of the sentences of the latter[1]. Nevertheless, the existing Natural Language Processing systems do not usually target on reported speech itself, since they mainly focus on an automated retrieval of quoted segments, without going further into neither the linguistic analysis of the introduction markers of the reported speech, nor the enunciative modality. The latter concerns, for instance, the position of the enunciator towards what he reports, or the manner in which he describes the reported enunciation or the attitude of the others speakers, etc.

Some research on the *Opinion Mining* uses automatic procedures, generally based on statistical methods, in order to assign a subjective or objective character to a word, a sentence or a text, and to determine the attitude of the speaker (orientation: positive, negative or neutral) or the degree (strength) of this attitude [3] ; [4] ; [5] ; [6]. This lexical approach seems to be limited because the terms in a certain context may have an emotional value which is exactly opposed to their values if considered individually, for example:

"This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up", cited in [7].

A. Ranta (Ed.): GoTAL 2008, LNAI 5221, pp. 40–51, 2008.

Many researchers in this field have observed the same phenomenon and speak in favour of combining lexical information with more complex linguistic analysis: Polanyi and Zaenen [8] (cited in [9]), state the necessity of taking into consideration the negations, some connectors (*Although Boris is brilliant in math, he is a horrific teacher*) and the modal operators (*If Mary were a terrible person, she would be mean to her dogs*).

We can also mention other existing works [2] that develop a linguistic analysis of the modalities concerning events (in the aspectual sense). The term modality for the author refers to certain degrees of possibility, beliefs, opinions, evidentiality, etc., however the polysemy of modal auxiliaries (must, may…) is not taken into consideration.

In our work, we propose a linguistic analysis of enunicative modalities in direct reported speech (D-RS) in Arabic and French. This analysis takes into account the marks of the enunciator in the discourse (his attitude towards what is reported), and allows the organisation of modality values in a semantic map. This map is exploited by an automatic system of grammatical and discursive annotations, based on the Contextual Exploration. This method, unlike the above-stated approach [2], requires only the analysis of surface linguistic forms.

Our presentation is organized around two main lines: a linguistic one, which exposes the theoretical principles of the analysis and categorization of linguistic data; and an other, computational, which explains the architecture of the implemented system on the different processing levels.

2 Contextual Exploration Methodology

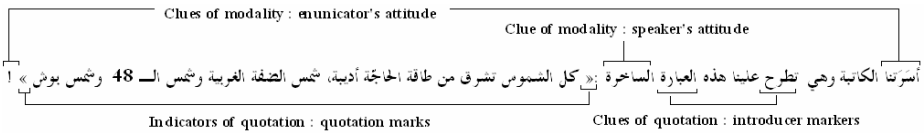
In the linguistic study of enunciation, the construction of an utterance (or a text) has to take into account some language operations such as predication, discourse operations and operations of commitment, the expression of which leaves a certain amount of surface linguistic traces. By analysing these linguistic indicators, the linguist is able to reconstruct, according to the process of abduction [10], the underlying operations of language production. Our methodology, the Contextual Exploration (CE) [11], is based on the analysis of these surface linguistic indicators, caracterising the textual representations used by the enunciator and which correspond to a given *point of view*[1] frame, such as citations, definitions or causal relations, etc. And because in natural languages the relationship between operations and linguistic indicators is rarely an one-to-one function, we need to explore the context in order to identify complementary *clues* that confirm or falsify the pertinence of the hypothesis first motivated by the indicators. The following example is analyzed by the CE' strategy:

أسَرَتنا الكاتبة وهي تطرح علينا هذه العبارة الساحرة : « كل الشموس تشرق من طاقة الحاجّة أدبية، شمس الضفة الغربية وشمس الـ 48 وشمس بوش » !

The writer fascinated us by giving to us this mocking sentence: "all suns rise from the dormer window of Hajja Adiba, the sun of the West Bank, the sun of 48th and the sun of Bush"!

[1] The notion of "point of view" in our approach corresponds to the analysis of a concrete task defined by the user.

In this example[2], the indicator of quotation is the quotation marks, and the complementary clues are combination of the verb « تطـرح » / « by giving » and the declarative noun « العبـــارة » / « sentence ». As far as the modal clues are concerned, they are: « الســاخرة » / « mocking » which marks the attitude of the speaker ; « أسَرَتنا » / « fascinated us » and « ! », that denote the attitude of the enunciator.



The linguistic analysis performed by the CE method does not rely on any preliminary morpho-syntactic analysis or statistical method, and is composed mainly of two parts of procedure made by the linguist: one that consists in studying various types of texts in different languages and distinguishes between two types of linguistic markers: indicators and clues ; and another that concerns CE's rules and exploits these markers (by using a CE engine), in order to find the surface level markers of the textual representations corresponding to the given *point of view*.

This method has been used in different computer applications, such as automatic summarization [12] [13] [14], extraction of causal relationships [15] and relationships between concepts [16].

3 The Linguistic Markers of Enunciative Modalities in D-RS

The notion of modality has been studied from many different perspectives: Logic, Philosophy and Linguistics [28] ; [29] ; [30]. In the field of Linguistics, modality can be considered from a syntactic, semantic or enunciative perspective.

We shall consider modality in an enunciative approach, according to Ch. Bally [31], E. Benveniste [32] and A. Culioli [33], so we distinguish between the enunciator and the speaker [34]: in reported speech, the enunciator makes a commitment to the utterance in its totality (*the author*), and the speaker is the third person quoted by the author, the "*last enunciator who directly makes commitment to the predicative relation*" [35].

In the theory of Enunciation, the commitment of an enunciator of an utterance introduces aspect and tense variations or enunciative modalities, marked in the utterance by traces that the enunciator leaves in his speech. In our case, these traces can manifest themselves in the introductory portion of the direct speech in different forms: they may indicate the enunciator's position towards what is reported, describe the speaker's attitude towards what is being said (in general) or towards what the speaker himself is saying; or refer to the relationship between the speaker and the enunciator, etc. In each of the processed languages, the enunciative modality markers in D-RS are either those that introduce the citation, or other markers that are otherwise appointed.

[2] *http://www.arabicstory.net*

3.1 Citational Linguistic Markers

The term *reported speech* covers a number of forms [17]: direct and indirect speech, free indirect speech, direct speech introduced by "*that"* [18], etc. We are particularly interested in direct reported speech (quotations). This linguistic act permits the enunciator to make a commitment for what is said or written by the speaker, without modifying it.

As we have already mentioned above, we can distinguish two types of linguistic markers: indicators and clues. For us, the indicators of D-RS are typographical signs (in French, Arabic and also in a number of many other languages) that define the scope of the D-RS. These signs are the quotation marks surrounding the clause[3], sometimes preceded by a colon (therefore the clause constitutes a syntactically independent sentence) or by the conjunction '*that*'.

As to the contextual clues for D-RS, they are the declarative linguistic markers that introduce or succeed the citation:

- verbs (*X denied the facts: "…"*): Arabic examples: ...أشار إلى, عبّر عن, أعلن, زعم ;
  French examples: *écrire, souligner, avouer, affirmer, critiquer, …*
- nouns (*This is the declaration of X: "…"*): Arabic examples: بيان, تصريح, اعلان ;
  French examples: *déclaration, annonce, slogan, appel…*
- gerunds (*X affirmed this by adding: "…"*): Arabic ex.: ...مؤكداً, مضيفاً, قائلاً ;
  French examples: *en soulignant, en affirmant, en ajoutant…*
- adverbials (*According to X: "…"*): Arabic ex.: ...وفقاً لـ, على ذمة, نقلاً عن, بحسب
  French examples: *d'après, selon…*

We note that in French, unlike Arabic, verbs can be positioned in the middle of the citation ("…, *affirme-t-il*, …").

3.2 Linguistic Modality Markers

In a D-RS, the enunciator can take into account the oral or written speech of the speaker, or describe the speaker's attitude towards his own speech or that of his interlocutor. On the other hand, the enunciator can show his position towards what he reports. We shall see in more detail some of these enunciative relationships:

- The enunciator reports the speaker's declaration (*he says, he declares, he adds, he repeats,* etc.). The latter makes a commitment to the predicative relation without getting involved.
- The enunciator reports the speaker's commitment (*he confirms, he asserts, he certifies*, etc.). The latter takes the responsibility for the matter of the clause. This language act is a commitment concerning the validity of a predicative relation.
- The enunciator describes the relationship between the speaker and the interlocutor: this relationship can be related to the volutive modalities (*to encourage, to forbid, to command*) or to evaluative modalities (*to make fun, to denounce, to apologize*). In a question, for example, the speaker demands the

[3] We do not take into consideration the '*textual islands*' (*She criticised the president's "machiavelism"*).

commitment of the interlocutor in regard to the content suggested by the speaker (*he asked his son: "…"*).

- The enunciator relates the act of locution (*I say that X said: « … »*)
- The enunciator makes a commitment to the truth of the locution act: *I affirme / it's sur that X said: « … » ;*
- The enunciator indicates a judgement on the spatio-temporal realisation of the predicative relation: *X said in Goteborg yesterday that: « … ».*
- The enunciator makes commitment on the evaluative modalities that pertain to the veracity of the speaker's utterance (an *untrue* declaration, *a credible* explanation…) or on positive or negative values (*good / bad* explanation…).
- The enunciator makes commitment on the evaluative modalities pertaining on his own attitude (*happily he confessed*…) or to the attitude of other speakers who are implied in the reported enunciation (sincerity: *to pretend* ; agreement/disagreement: *wrongly* ; pronunciation: *to babble*).

These modalities can be marked by the choice of the declarative expression introducing the citation seen above (*to say, to whisper*…). But they can also be tagged by polysemic declarative markers (*to make fun of, to humiliate*…) or by non declarative markers that denote the speaker's attitude for example (*to interrupt, to blush*…). Other grammatical categories are also to be observed, such as adverbs (*alas, finally*) and adjectives (*untrue, credible*).

Among the markers of modality, some expressions introducing citations, especially the verbs, have been subject to diverse syntactic, semantic and pragmatic analysis [19] ; [20] ; [21] ; [22] ; [23] ; [24] ; [25]; [26] ; [27]. The analysis that we have adopted here differs on several points: the linguistic markers of the reported speech and modality concern all lexical categories (verbs, nouns, adverbs…) ; these markers have been studied cross-linguistically (Arabic and French); the framework is that of the Enunciative Linguistics where the mark of the enunciator is analysed in the discourse ; finally, this work is carried out in the perspective of automatic language processing and our final goal is to provide automatic applications that respond to concrete needs.

4 Categorisation of the Enunciative Modalities: Semantic Map

We have analysed, in a contrastive manner, the markers of enuncicative modalities in Arabic and French and we then organized them according to a semantic map, based on the principle of commitment [35], [36].

This semantic map (SM) is a "linguistic ontology" of grammatical or discursive categories, interlinked by the elements of specification, opposition, application, value attribution, etc. It corresponds to one or more *points of view*. The values of the SM (the nodes of the graph) are represented in texts by different indicators and clues (node instances) in one language or another[4], and by the CE rules that are associated to these instances. Thus, we have organised the enunciative relations [37] in the following figure:

[4] Some values of the SM can be attested in one language and not in another.

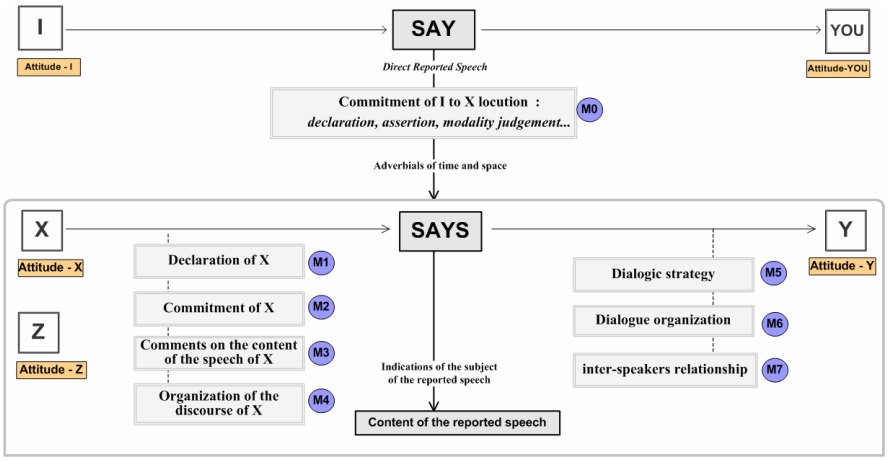


Fig. 1. Semantic Map of enunciative modalities in D-RS

The fundamental dialogical dipole is between the enunciator (I) and the co-enunciator (YOU). The (I) reports to (YOU) the utterance of a speaker (X), absent from the dialogue. In the same way, the D-RS is centered on the dialogical relation speaker-interlocutor. All the enunciative relations in this SM are under the commitment of the enunciator. Some of them concern the simple I-YOU speech relation (M0), and others concern the D-RS (M1 to M7).

The categories M1 to M4 denote the relation between the speaker and his utterance ; the categories M5 to M7 indicate the relation between the speaker and the interlocutor. We have also added other categories to the SM, in order to describe the attitudes of the different speakers implicated in the D-RS: Attitude-I, Attitude-YOU, Attitude-X (speaker), Attitude-Y (interlocutor) and Attitude-Z (speaker absent).

The categories of the SM can be related by application, incompatibility, specification, etc. In this way, the spatio-temporal category that depends on the enunciator can be applied to all the sub-categories from M0 to M7. Similarly, all the categories contain other sub-classes, such as assertion of the speaker, that can take several values (individual assertion, universal or collective assertion, etc. [35]).

5 Program Implementation of the Automatic Annotation

The applicative part of our work [38] consists in the implementation of an automatic annotation tool for grammatical or discursive categories. This tool, EXCOM[5], is composed by a CE engine and the supplementary modules[6] connected to it.

The automatic annotation requires pre-processing of the linguistic resources, which means: corpora segmentation, markers organization and CE rules construction.

[5] For “EXploration COntextuelle Multilingue”. Our system in its second edition is freely available online for the use of researchers on the following address: `https://www.excom.fr`

[6] For the implementation we have used Java, XML, JDOM, XLINK, JNLP, etc.

5.1 Automatic Segmentation

In our work, the segmentation of a text into smaller parts helps in determining the search fields for linguistic markers, and the textual segments which are to be annotated. This consists in defining the boundaries of sections, titles, paragraphs and sentences.

The sections are determined by the presence of titles in the text, the titles are defined by several heuristics[7], and the paragraphs are delimited by the sign of carriage return. In order to split the paragraphs into sentences, we used a set of rules, which can be modified by the user, and based on disambiguation of typographical signs[8] and linguistic terms[9]. This method [39] takes into account the difficulties encountered in Arabic (lack of capitalization and of vocalization) and in French (many abbreviations).

The input files for the segmentation module are raw text files in UTF-8 encoding, in different languages, and the output files are in the XML DocBook format for articles. The results of segmentation are satisfactory, however they must be evaluated in a large scale, and improved, for example, by the identification of item lists and the hierarchy between titles and sub-titles.

5.2 Automatic Annotation

The core of the EXCOM architecture (Fig. 2) consists of a CE engine that manipulates the CE rules and linguistic markers associated with the annotations. The annotation process consists in the research of the indicators in the search fields defined by the segmentation process. The presence of indicators calls the application of CE rules and then the conditions of these rules are examined (research of contextual clues). If all the conditions are satisfied, the CE engine either attributes the corresponding annotation to the segment, or calls (recursively) another CE rule.

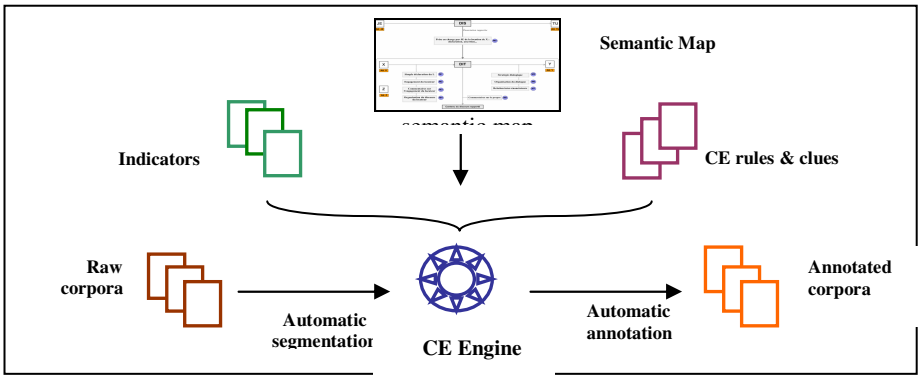


Fig. 2. Simplified architecture of the platform EXCOM

[7] For example, if the line is not longer than *n* words and if it is terminated by colon.
[8] For example: period, semicolon, question mark, etc.
[9] Connectors like *but*, *however*, *nevertheless*, *so that*, etc.

As to the CE rules, they are automatically generated (with XML format) by using a graphical user interface that offers the possibility of defining and modifing all of the functional parameters of the CE engine, in order to separate maximally the data from the implementation. The efficiency of the CE rules corresponds to the functionalities of localization and disambiguation of markers, offered by the CE engine, such as for example:

- The hierarchy between the indicators that carry the semantic meaning of the category in question and the contextual clues. From a theoretical perspective, this principle is fundamental for us and concerns one of the differences between the CE method and other context analysis methods.
- The contextual clues can be of two types: positive (their presence is obligatory) and negative (their presence cancels the action of annotation).
- The markers (indicators or clues) can be either linguistic units (words or regular expressions), or already annotated segments (in an already annonated text).
- The targeting of a part of a text can be specified and used as search field by using the XML structure of the file and the identified titles in the segmented document (for ex.: find indicators in all the titles, or find indicators in the last sentence of the second paragraph of the first section).
- The research of the contextual clues is carried out in the context of the indicators: before, after or inside the indicator (for ex.: a morphem in a word).
- Different types of clues can be combined by logical operators: between the positive clues, before and after the indicator, between the negative clues, before and after the indicator. These operators can be: AND, OR, XOR. (for ex. looking at clues before OR exclusively (XOR) after indicators).
- An order can be defined between the positive and the negative clues in the context before or after each indicator (for ex.: in the context before the indicator, a negative clue cannot occur before a positive clue).
- Meta-rules can be set by the user in order to define priority between rules or the navigation mode in the semantic map (this second section is under construction).

The annotation of the segments contains the following meta-data: the semantic category of the annotation, the class of the indicator that has triggered the annotation, the identifier of the CE rule that has carried out the annotation, etc. This information allows the linguist to improve the rules, as well as the relevance of the linguistic markers.

This is an example of a simple EC rule for the citation, in a declarative form:

CE rule # 5:
Given P the following research space: all sentences of the first paragraph of the last section
If (indicator from the class “2-quotes” exists in P)
If (in the before-indicator-context does not exists a negative clue from the class “references”)
If (in the middle-indicator context exists a positive clue from “declaration-verb-reversed”)
Then: Give the semantic annotation "quotation-middle-conclusive" to P

This rule can annotate some sentences found in the first paragraph of the last section of an article which carry a conclusive value in addition to their enunciative modality. For example:

"*Nevertheless, he concludes, this rapprochement between the European business companies is one of the challenges of tomorrow***"**.

Once the texts are annotated, the user can then proceed to the post-processing treatment by using a module which compiles all annotated segments of the corpus in a database, with an interface allowing the navigation between these segments and their original contexts.

6 Evaluation

We set up a first evaluation test consisting in the judgment of the capacity of the system to categorize[10] the marked segments of quotations, according to the SM. The corpus on which we worked was composed of 250 texts of journalistic articles per language handling various subjects[11]. We made preliminary tests on 80 % of the corpus aiming the EC rules adjustment. Then we performed the evaluation test on the other part of the texts (20 %).

Based on the results on each corpus, we extracted randomly 39 segments annotated according to the following three semantic categories[12]:

1) the declaration of the speaker without commitment to the content (category M1), for instance, *X says: "…"* ;
2) the commitment of the speaker in regard to the content (category M2): for instance, *X affirms: "…"* ;
3) the enunciator's comments about the speech of the speaker, concerning the degree of the sincerity of the speaker or the truth of his speech (category M3): for instance, *X claims: "…"*.

The test consisted then in asking the subjects (15 French-speaking persons[13], and 9 Arabic-speaking persons[14]) to annotate manually the 39 extracted segments, according to the same semantic categories. For each segment the subjects had to choose one of the proposed categories and assign it to the sentence. For the calculation of the evaluation measures, we have used the evaluation interface EVA-2[15].

In order to calculate the precision and recall measures, the "correct" annotations were determined on the basis of the set of human annotations. These correct annotations are defined as the most frequent annotations attributed by the subjects. The results for the Arabic and French corpus are the following:

---

[10] As for quotations localization by means of typographical indicators, no evaluation has been made due to the relative simplicity of the task.

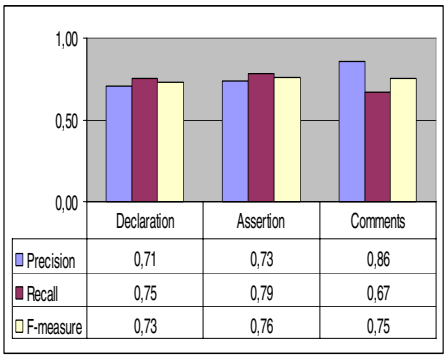
[11] The French corpus was taken from the following newspapers: *Le Monde Diplomatique, le Figaro, l'Humanité* and *Libération* ; and the Arabic one from: *Al-Nahar, AL-Ahram, Tishreen, Al-Jazeera, Al-Sabah, Al-Alam* and *Al-Quds*.

[12] The only restriction was that the segment numbers by category had to be the same for both languages.

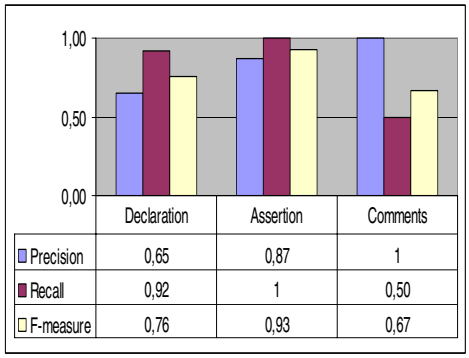
[13] PhD and Master students of Human Sciences Department of Paris-Sorbonne University and Paris 7 - Denis Diderot University.

[14] Arabic-speaker students of Sorbonne Univ., Paris-Jussieu Univ., Lyon and Damascus Univ.

[15] *Evaluation d'Annotation Automatique*, developped by I. Atanassova and M. Bertin (LaLIC).



	Declaration	Assertion	Comments
Precision	0,71	0,73	0,86
Recall	0,75	0,79	0,67
F-measure	0,73	0,76	0,75



	Declaration	Assertion	Comments
Precision	0,65	0,87	1
Recall	0,92	1	0,50
F-measure	0,76	0,93	0,67

Fig.3. The results of evaluation on the Arabic (left) and French (right) corpora

8 Perspectives

Our first evaluation results are quite encouraging. We are currently drawing up a deeper protocol for the evaluation of all categories in the SM, in order to validate this part of the work.

These tests have allowed us to draw comparisons between French and Arabic on several levels. First, we have noticed that in Arabic the surface forms are generally more polysemous than in French, especially the forms that have a three-letter root. This difficulty, already well known [40] ; [41] ; [42] ; [43], is due to the morphological ambiguity in Arabic caused, above all, by the absence of vocalisation, the agglutination and the relatively free word order in a sentence. To resolve this problem, we have used clues for the disambiguation of certain markers, in order to validate or not their correspondence to the researched forms. Secondly, we remark that the occurrences of direct speech in French texts and the use of enunciative modalities are much more frequent than in texts in Arabic. Finally, we should elaborate more our reflexion on our categorization of the assertion class and its sub-categories. Some of these categories were manually annotated as a declaration.

As the annotation procedure is independent of a given *point of view* (the D-RS in our case) and of the processed languages (Arabic and French), it is absolutely possible to annotate other types of texts with different linguistic resources, in different languages and according to other *points of view*. The annotation platform has also been then tested in the following works: identification of the hypotheses in biological articles in English[16], annotation of D-RS in Korean[17] and of events in articles in French and Polish[18]. These works are in progress and will be published shortly. It becomes also conceivable, in futur work, to intersect the annotations according to different semantic maps, such as, for example, that of D-RS and the SM of bibliosemantics or contact between people (meetings).

[16] Desclés J., Ecole Normal Supérieure de Paris.
[17] Suh J., Seoul Women's Univ.
[18] Gwiazdecka E., Univ. of Warsaw.

Acknowledgments. The authors would like to thank all readers of this paper, and everyone who participated in the evaluation tests.

References

1. Bergler, S.: Conveying Attitude with Reported Speech. In: Shanahan, J.C., Qu, Y., Wiebe, J. (eds.) Computing Attitude and Affect in Text: Theory and Applications. Springer, Heidelberg (2005)
2. Sauri, R., Verhagen, M., Pustejovsky, J.: Annotating and recognizing event modality in text. In: Proceedings of the 19th international Flairs conference, Florida (2006)
3. Riloff, E., Wiebe, J., Wilson, T.: Learning subjective nouns using extraction pattern bootstrapping. In: Proceedings of the 7th Conference on Natural Language Learning, pp. 25–32 (2003)
4. Wiebe, J., Wilson, T., Bruce, R., Bell, M., Martin, M.: Learning Subjective Language. Computational Linguistics 30, 277–308 (2004)
5. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 129–136 (2003)
6. Turney, P.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, pp. 417–424 (2002)
7. Pang, B., Lee, L., Vaithyanathan, V.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the 2002 Conference on Empirical Methods in natural language processing, pp. 79–86 (2002)
8. Polanyi, L., Zaenen, A.: Shifting attitudes. In: Lagerwerf, L., Spooren, W. (eds.) Determination of information and tenor in texts: Multidisciplinary approaches to discourse, Nodus Publikationen, pp. 61–69 (2003)
9. Bestgen, Y., Fairon, C., Kevers, L.: Un baromètre affectif effectif. In: Purnelle, G., Fairon, C., Dister, A. (eds.) Actes des septième Journées internationales d'Analyse statistique des Données Textuelles, Louvain-la-Neuve, pp. 182–191 (2004)
10. Peirce, Ch. The Collected Papers vol. V: Pragmatism and Pragmaticism. Harvard University Press, Cambridge, MA (1934)
11. Desclés, J.-P., Jouis, C., Oh, H.-G., Reppert, D.: Exploration Contextuelle et sémantique: un système expert qui trouve les valeurs sémantiques des temps de l'indicatif dans un texte. In: Herin-Aime, R., Dieng, J.-P., Regourd, J.P. (eds.) Knowledge modeling andexpertise transfer, pp. 371–400. Amsterdam (1991)
12. Berri, J.: Contribution à la méthode d'Exploration Contextuelle: Applications au résumé automatique et aux représentations temporelles. Réalisation informatique du système SERAPHIN, Thèse de doctorat, Université de ParisIV-Sorbonne, Paris (1996)
13. Minel, J.-L.: Filtrage sémantique, du résumé automatique à la fouille de textes. Hermès, Paris (2002)
14. Blais, A., Desclés, J.-P., Djioua, B.: Le résumé automatique dans la plate-forme EXCOM, Digital Humanities, Paris (2006)
15. Jackiewicz, A.: La notion de cause pour le filtrage de phrases importantes d'un texte. In: International Conference on Natural Language Processing and Industrial Applications, Moncton, Canada (1996)
16. Le Priol, F.: A data processing sequence to extract terms and semantics relations between terms. In: 10th Human Centered Processes Conference, pp. 241–248 (1999)

17. Rosier, L.: Le discours rapporté. Histoire, théories, pratiques, Duculot, Paris et Bruxelles (1999)
18. Bruna-Cuevas, M.: le discours indirect introduit par que. Le français moderne 54(1), 28–50 (1996)
19. Gross, M.: Méthodes en syntaxe, Hermann, Paris (1975)
20. Charolles, M.: Exercices sur les verbes de communication. Pratiques 9, 83–107 (1976)
21. Lamiroy, B., Charolles, M.: Les verbes de parole et la question de l'(in)transitivité. In: colloque lexique et grammaire, Thessaloniki (2003)
22. Mourad, G.: Analyse informatique des signes typographiques pour la segmentation de textes et l'extraction automatique de citations: Réalisation des Applications informatiques, SegATex et CitaRE, thèse de doctorat, Université de Paris-Sorbonne (2001)
23. Austin, J.: How To Do Things With Words. Oxford University Press, Oxford (1962)
24. Searle, J.-R.: les actes du langage: essai philosophique du langage, Hermann, Paris (1972)
25. Vanderveken, D.: Les actes de discours, Mardaga, Liège (1988)
26. Ross, J.-R.: On declarative sentences. In: Jacobs, R.-A., Rosen-baum, P.-S. (eds.) Readings in english transformational grammar, Massachussets (1970)
27. Banfield, A.: Le style narratif et la grammaire des discours direct et indirect. Change 16(17), 190–226 (1973)
28. Bybee, J., Perkins, R., Pagliuca, W.: The Evolution of Grammar: tense, aspect and modality in the languages of the world. The University of Chicago Press, Chicago (1994)
29. Palmer, F.R. (ed.): Mood and Modality. Cambridge University Press, Cambridge (1986)
30. Nuyts, J.: Modality: overview and linguistic issues, in The expression of modality, pp. 1–26. Mouton de Gruyter, Berlin (2006)
31. Bally, Ch.: Linguistique générale et linguistique française, Francke, Berne (1932)
32. Benveniste, E.: Problèmes de linguistique générale 1, Gallimard, Paris (1966)
33. Culioli, A.: sur quelques contradictions en linguistiques. Communications 20, 83–91 (1973)
34. Desclés, J.-P.: Quelques opérations énonciatives. In: David, J. (ed.) Logique et niveaux d'analyse linguistique, Klinckseieck, Paris, pp. 213–242 (1976)
35. Desclés, J.-P., Guentcheva, Z.: Enonciateur, locuteur, médiateur. Les Rituels du dialogue, Nanterre: Société d'ethnologie, Recherches thématiques 6, 79–112 (2000)
36. Desclés, J.-P.: Prise en charge, engagement et désengagement, Anvers (2007)
37. Alrahabi, M., Desclés, J.-P.: Vers une carte sémantique du discours rapporté. In: 75th congrès de l'ACFAS, Trois-Rivières (2007)
38. Alrahabi, M., Ibrahim, A.H., Desclés, J.-P.: Semantic Annotation of Reported Information in Arabic. In: Flairs 2006, Floride (2006)
39. Alrahabi, M., Djioua, B., Desclés, J.-P.: Annotation Sémantique des Énonciations en Arabe. In: Inforsid 2006, Hammamet (2006)
40. Alrahabi, M., Mourad, G., Djioua, B.: Filtrage sémantique de textes en arabe. In: JEP-TALN 2004, Fès (2004)
41. Alrahabi, M.: Désambiguiser par le contexte, le cas de la graphie alef-noun en arabe. In: vers de nouveaux accès aux savoirs, Paris (2006)
42. Diab, M., Kadri, H., Daniel, J.: Automatic Tagging of Arabic Text: From raw text to Base Phrase Chunks. In: Proceedings of HLT-NAACL (2004)
43. Dichy, J.: On lemmatization in Arabic. In: Proceedings of the Workshop on Arabic Language Processing, Toulouse (2001)