# An Approach for Detecting Modality and Negation in Texts by using Rule-based Techniques.

Sara Lana-Serrano[1,4], Daniel Sánchez-Cisneros[2], Paloma Martinez-Fernandez[2], Antonio Moreno-Sandoval[3], Leonardo Campillo-Llanos[3]

[1] Universidad Politécnica de Madrid
[2] Universidad Carlos III de Madrid
[3] Universidad Autónoma de Madrid
[4] DAEDALUS - Data, Decisions and Language, S.A.
slana@diatel.upm.es, dscisner@inf.uc3m.es, pmf@inf.uc3m.es, antonio.msandoval@uam.es, leonardo.campillos@uam.es

**Abstract.** The automatically processing of texts has become a very important aim in the recent years because of the huge amount of time that could be saved in sectors such as education, jurisprudence or medicine. Thus, a very important task in the automatically processing of texts is the detection of language mechanism, such as modality and negation. In this paper we present our first approach in the field of detection of modality and negation by using a rule-based system. The system uses lexical and syntactic information found in the text to determine where exist evidentially of modality or negation. The best results obtained by our approach in detecting modality and negation achieves a macroavaraged F1 measurement of 0.5339, a Microavaraged F1 of 0.6395, and an overall accuracy of 0.6551.

**Keywords:** Natural Language processing, modality, negativity, verbal group.

## 1 Introduction.

The automatically processing of texts has become a very important aim in the recent years because of the huge amount of time that could be saved in sectors such as education, jurisprudence or medicine. The language mechanisms allow to the human being express his knowledge and ideas with plenty freedom. This makes that a writer can express things with different types of certainty and security. Thus, a very important task in the automatically processing of texts is the detection of language mechanism, such as modality and negation.

*Morante* [1] defines modality as "a grammatical category that allows the expression of aspects related to the attitude of the speaker towards her statements in terms of degree of certainty, reliability, subjectivity, sources of information, and perspective", but it can be found more specific description of modality in *Salkie, Busuttil, and van der Auwera* [2] or *Jespersen* [3]. *Saurí, Verhagen, and Pustejovsky* [4] exposes that modality can be expressed by a variety of different strategies and constructions. In English, these include both lexical items (verbal predicates, nominal predicates, etc.) and syntactic constructions (relative clauses, subordinated temporal clauses, etc.).

On the other hand, a negation clause is defined by *Payne* [5] as "the one that asserts that some event, situation, or state of affairs does not hold. Negative clauses usually occur in the context of some presupposition, functioning to negate or counter-assert that presupposition".

Detecting negation and modality in texts is the main topic of the pilot task called "Processing Modality and Negation for Machine Reading"[1] from the QA4MRE lab[2], held in the main conference "CLEF 2012: Information Access Evaluation meets Multilinguality, Multimodality, and Visual Analytics"[3]. The aim of this task is the automatic unsupervised understanding of texts. Thus, the task is an annotation task where systems have to determine whether an event mentioned in a text is presented as negated, modalised (i.e. affected by an expression of modality), or both.

Our team is composed by four different entities: LaBDA group[4] (Universidad Carlos III de Madrid), GSI group[5] (Universidad Politécnica de Madrid), LLI group[6] (Universidad Autónoma de Madrid) and Daedalus[7]. Each entity has held a research line in different fields of natural language processing that have been fused in this task.

The four groups are integrated in Multimedica[8] project. The aim of this project is to define and develop information extraction and retrieval techniques from biomedical texts. This is being carried out following two basic tasks: firstly, processing scientific documents in English about pharmacology, and secondly, processing informative texts about health topics in other languages such as Spanish and Arabic. The multidisciplinary character of each entity allows the chance of combining different types of knowledge in order to detect modality and negation in texts.

In this paper we present our first approach in the field of detection of modality and negation by using a rule-based system. The system uses lexical and syntactic information found in the text to determine where exist evidentially of modality or negation. In this paper we will introduce this approach following this structure: in section 2 we expose a study of the systems that have treated the modality and negation in the recent years. Section 3 describes the architecture of our system. Later, we present the results obtained in the task in section 4. Finally, in section 5 we present our conclusions and error analysis.

---

1 http://celct.fbk.eu/QA4MRE/index.php?page=Pages/modalityTask.html
2 http://celct.fbk.eu/QA4MRE/
3 http://clef2012.org/
4 http://labda.inf.uc3m.es
5 http://www.gsi.dit.upm.es
6 http://www.lllf.uam.es/ESP/
7 http://www.daedalus.es/
8 http://labda.inf.uc3m.es/multimedica/

## 2 Related Work.

In recent years, there have been achieved different research lines that share a main goal: the processing of modality and negation in texts. To do so, there exist different approaches depending on the type of information the systems uses to detect this modality and negation.

Modality has been treated in different ways because it can be understood as an event, or as a predicate that shows modality in word or phrases.

EvITA [4] is a system based on patter-matching techniques that is oriented to improve the identifying of the scope of modality in natural language and proposes a solution for its automatic identification. In one hand, this system is able to identify events that denote modality, and on the other hand, analyse the event-based grammatical features that are relevant for temporal reasoning. The pattern-matching techniques are applied over the chunked text, enriched with part-of-speech tagging. This system achieves a precision of 74.55%, a recall of 78.61% and an F1-measure of 76.53%. This performance has been evaluated against TimeBank corpus[9].

SlinkET [6] is a parser for identifying context of event modality in text, following the research line of EvITA system above described. This system has been developed under the TARSQI framework [10] for identifying, annotating and reasoning about temporal information in texts. SlinkET also treat modality based on events, introducing modality at the syntactic level, involving subordination relations between two clauses. The system has been evaluated over the 10% of the TimeBank corpus containing 681 events, and has achieved a precision of 92%, a recall of 56% and an F1-measure o 70%.

*Baker et al* [7] head modality without a processing based on events. This system detects modality based on a modality annotation scheme, a modality lexicon, and two automated modality taggers. Entries in the modality lexicon consist of a string of one or more words, a part of speech for each word, a modality type from its scheme, a head word, and subcategorization codes. On the other hand, the modality taggers are based on two different approaches: a string-based tagger that operates on text that has been tagged with parts of speech by a Collins-style statistical parser; and a structure-based tagger that (1) processes flattened trees and (2) finds modality trigger words, its target and the action insert tags. The evaluation was done over 249 modality-tagged sentences from the English side of the NIST 09 MTEval[11] training sentences. There has been performed just the structure-based tagger, that achieves a precision of 86.

Detecting negation on texts have been studied on may approaches that goes from rule-based system that codifies grammar rules to machine learning systems that has been trained with lexical, semantic and syntax features.

*Huang and Lowe* [8] have elaborated a system that detects negation in texts based on syntactical categories of negation signals and patterns. To do so they created a grammar based on 6 different types of negation depending of the syntactical

[9] http://www.timeml.org/site/index.html
[10] http://www.timeml.org/site/tarsqi/index.html
[11] http://www.ldc.upenn.edu/Catalog/project_index.jsp

information found in the text. In a first step, they manually annotated the negation found in the texts of a corpus with clinical radiology reports and then they created the grammar, which later was improved using a PERL script. This system was evaluated through manual inspection of validated radiology reports, obtaining identified negated phrases with sensitivity of 92.6%, positive predictive value of 98.6% and specificity of 99.87%.

Another approach that uses regular expression-based algorithm is Negex [9] and is oriented to medical domain. The text is processed in a first step to index UMLS terms. Later, NegEx expands on the baseline algorithm with more regular expression syntax and identify two groups of negation: the “pseudo-negation” phrases that consist of phrases that appear to indicate negation but instead identify double negatives, modified meanings, and ambiguous phrasing; and the phrases believed to be used to deny findings and diseases when used followed or preceded by UMLS terms. This approach was evaluated on the 28 UMLS terms that occurred in the test set at least 10 times. This dataset were manually tagged by physicians. NegEx had a specificity of 94.5%, a positive predictive value of 84.5%, and a sensitivity of 77.8%.

*Rokach, Romano, and Maimon* [10] proposes a pattern-based solution that uses two regular expressions representation: one for the string that precedes the targeted medical term, and one for the string that follows it. This system is also based in two algorithms to learn regular expressions: the longest common subsequence algorithm (LCS) [11] and Teiresias algorithm [12]. This approach was evaluated over a set of 1,766 instances parsed from de-identified discharge summaries that were obtained from Mount Sinai Hospital in New York This system obtains the best results when using cascade decision trees with LCS algorithm, achieving an F-measure of 95.9%.

In most above studies exposed in this section, the detection of modality and negation has been faced with rule-based approaches that use regular expressions. Thus we purpose a new approach based on a regular expression recognition system that uses English grammar rules for detecting negation and modality.

## 3 Description of the System.

We approach this task with the construction of a standalone application using the GATE framework [13]. The architecture of our solution (**Fig. 1**) presents a modular structure where each of the modules carries out one step of the text processing. For the three first steps (tokenization, sentence splitting and POS tagging), native GATE modules (ANNIE) have been used, so their description is out of the scope of this paper.

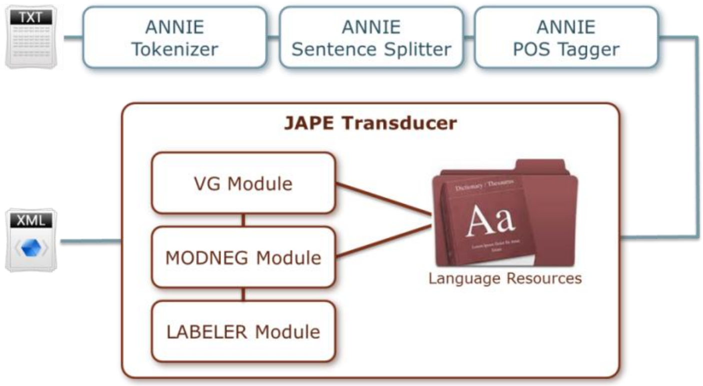


**Fig. 1.** System Arquitecture

The tagging process has been implemented with a specification of a rule system defined in the JAPE (Java Annotation Patterns Engine). Rules have been grouped in different JAPE modules, according to their functionality.

For us, one of the most important indicators to consider in the analysis of modality and negation of a verb event within a given text is the identification and characterization of the verbal group in which it is contained, as morphosyntactic features such as verb mode or tense can be decisive. For this objective, a general-purpose rule module (VG Module) has been developed to tag verbal groups in a document, annotating them with the following features: lexical category (cat), mode, tense, aspect, voice and modality. **Table 1** shows the different values that can be assigned to each feature.

The VG Module performs the following steps:

1. Tagging of non-finite verbs and verbal groups and finite verbs/verbal groups in indicative
2. Tagging of verbs and verbal groups in subjunctive mode and the context that determines them. Rules defined in this step are supported by a knowledge base (KB) including verbs, phrasal verbs, expressions and grammatical structures.
3. Tagging of the *modality* feature attending to the negative and/or modal aspects of the adverbs that modify the main verb of the verbal group. Rules defined in this

step are supported by a Language Resource set that includes a list of modal and negation adverbs.

**Table 1.** Verbal group tagset

| Feature | Values | Feature | Values |
|---|---|---|---|
| **cat** | V (verbal group) | | |
| **Mode** | - (none)<br>I (indicative)<br>S (subjunctive)<br>M (imperative)<br>F (infinitive)<br>P (past participle)<br>G (present participle)<br>W (modal verb) | **aspect** | - (none)<br>S (simple)<br>P (perfect)<br>C (continuous)<br>B (perfect and continuous) |
| | | **voice** | - (none)<br>A (active)<br>P (passive) |
| **tense** | - (none)<br>P (present)<br>A (past)<br>F (future)<br>C (conditional) | **modality** | - (none)<br>N (negative)<br>M (modal)<br>B (negative and modal) |

4. Tagging of the modal value of the *modality* feature, attending to the modal aspect derived from the grammatical structure of the verbal group. For this purpose, the definition that the Longman grammar [14] proposes about the following concepts is used: modal forms, modal verbs, semi-modal verbs and lexical modal verbs.

This module does not consider the (semantic) factuality of the main verb to determine the value associated to the *modality* feature of the verbal group, and only the form is considered.

**Fig. 2.** and **Fig. 3** show several examples of the verb tagging carried out by this module over different fragments of the test documents.

**Input text**

```
When economists run into the limitations of their models, they tend to
heed the Wittgensteinian injunction: Whereof one can not speak, thereof
one must be silent.
```

**Tagged text**

```
When economists
<VG type="VIPSA-" modality="none">run</VG> into the limitations of their
models, they <VG type="VIPSA-" modality="none">tend</VG><VG type="VF-SA-"
modality="none">to heed</VG> the Wittgensteinian injunction: Whereof one
<VG type="VW-SAB" modality="modneg">can not speak</VG>, thereof one <VG
type="VW-SAM" modality="mod">must be</VG> silent.
```

**Source document**

```
climate-are-economists-erring-on-climate-change.txt
```

**Fig. 2.** Example of tagging of verbal groups (I).

**Input text**

```
But crime and starvation are hardly the only options. Even without a
high-school diploma, a black man can probably find a job if he looks.
```

**Tagged text**

```
But crime and starvation
<VG type="VIPSA-" modality="none">are</VG> hardly the only options. Even
without a high-school diploma, a black man <VG type="VW-SAM"
modality="mod">can probably find</VG> a job <SUBJUNCTIVE_CONTEXT
type="conditional">if he</SUBJUNCTIVE_CONTEXT>
<VG type="VSPSA-" modality="none">looks</VG>.
```

**Source document**

```
music-can-hiphop-change-the-world.txt
```

**Fig. 3.** Example of tagging of verbal groups (II).

The MODNEG Module is in charge of tagging all those particles that may be associated to the grammatical categories of modality and/or negation. It is structured in the following steps:

1. Tagging of negation and modal particles. Defined rules are supported by a KB including mainly adverbs, prepositions, conjunctions, pronouns, nouns, prefixes, conditional structures, consecutive expressions, nominal modifiers and subordinate clauses.
2. Tagging of the modality/negation of verbs attending to their semantics. In this case, the KB is formed by lexical entries tagged according to their lexical category and factual type [15] and grammatical structures. **Table 2** shows the Factual types that have been considered.

**Table 2.** Factual types

| Category | Factual Type |
|---|---|
| **Modality** | uncertainty<br>certainty<br>want<br>conjecture<br>imagine<br>expected<br>lookLike<br>pretend<br>suggest |
| **Negation** | refuse |

The last module (LABELER Module) is build up with a set of rules that determine the tagging of the event under analysis, according to the modality/negation of its context.

Globally, the standalone application generates, for each of the input documents, an output XML document annotated with information about the modality and negation of each of the events under analysis. Each event is tagged with two labels:

- ***MODNEG_CONTEXT:*** determines the context that has been used as reference to infer the modality associated to that event.
- ***EVENT:*** identifies the event under analysis.

The attributes that are associated to the context (MODNEG_CONTEXT label) try to sum up the decisions made by the rule system during the analysis, without considering the inherent factuality of the verb.

The attributes associated to the event (EVENT label) are inferred from the contexts in which the event is contained and the factual class associated to the event itself.

Both labels have the following attributes:

- ***type***: information about the rules that have been applied to determine the modality/negation.
- ***modality***: modality associated to the context or the event. Possible values are the following: *mod*, *neg*, *modneg* or *none*.
- ***confidenceMod***: confidence in the annotated modal character. Possible values are:
  - **100**: default value. This value indicates that there is no evidence to determine the modal character. It corresponds to a *neg* or *none* value in the *modality attribute*. It is only applicable to the MODNEG_CONTEXT label.
  - **0**: the best confidence level. The verbal group to which the event belongs is modal or there is a modal *particle* (word, multiword unit or grammatical expression) in the verbal group or adjacent (0 distance) to it.
  - **1**: there is a modal particle in a distance from the verbal group ranging from 1 to 5 *tokens* belonging to the same sentence and not including punctuation marks.
  - **2**: lowest confidence level. There is a modal particle in a distance from 1 to 25 *tokens* belonging to the same sentence, including punctuation marks.
- ***confidenceNeg:*** confidence in the annotated negative character. Possible values:
  - **100**: default value. It shows that there is no evidence that allows to determine the negative character. It corresponds to a *mod* or *none* value in the *modality* attribute.
  - **0**: highest confidence level. The verbal group is negative or there is a negative *particle* in the verbal group or adjacent to it.
  - **1**: there is a negative particle in a distance from the verbal group up to 5 *tokens* in the same sentence and different from punctuation marks.
  - **2**: lowest confidence level. There is a negative particle in a distance from the verbal group up to 25 *tokens* belonging to the same sentence, including punctuation marks.

In the case that the event corresponds to a verb of a *modal* and/or *negative factual class*, the EVENT label will include, in addition, the following attributes:

- ***kind***: concatenation of the different factual classes to which the event belongs.
- ***type:*** concatenation of the modal or negative character associated to the different classes to which the event belongs.

- ***modClass***: modality associated to the event, attending to the combination of *modality* and *type*.
- ***negClass***: negation value associated to the event, attending to the combination of *confidenceNeg* and *type*. This attribute enables a correct tagging in the case of negative contexts and events with negative polarity.

**Fig. 4** and **Fig. 5** show some examples of fragments of the output XML document for different test documents.

```
Input text
When economists <event id=49>run</event> into the limitations of their
models, they <event id=50>tend</event> to <event id=51>heed</event> the
Wittgensteinian injunction: Whereof one can not <event
id=52>speak</event>, thereof one must <event id=53>be</event> silent.
Output XML
When economists
<MODNEG_CONTEXT confidenceMod="100" confidenceNeg="100" modality="none"
type="withingVG">
   <EVENT confidenceNeg="0" confidenceMod="0" modality="none">run</EVENT>
</MODNEG_CONTEXT>

 into the limitations of their models, they
<MODNEG_CONTEXT confidenceMod="100" confidenceNeg="100" modality="none"
type="withingVG">
   <EVENT confidenceNeg="0" confidenceMod="0" modality="none">tend</EVENT>
</MODNEG_CONTEXT>

<MODNEG_CONTEXT confidenceMod="100" confidenceNeg="100" modality="none"
type="withingVG">
   to
   <EVENT confidenceNeg="0" confidenceMod="0" modality="none">heed</EVENT>
</MODNEG_CONTEXT>

the Wittgensteinian injunction: Whereof one
<MODNEG_CONTEXT confidenceMod="0" confidenceNeg="0" modality="modneg"
 type="withingVG followingNEG">
   can not
   <EVENT confidenceNeg="0" confidenceMod="0" modality="modneg">
    speak</EVENT>
</MODNEG_CONTEXT>

, thereof one
<MODNEG_CONTEXT confidenceMod="0" confidenceNeg="100" modality="mod"
 type="withingVG">
   must
   <EVENT confidenceNeg="0" confidenceMod="0" modality="mod">be</EVENT>
</MODNEG_CONTEXT>
silent.
Source document
climate-are-economists-erring-on-climate-change-out.xml
```

**Fig. 4.** Example of an XML output document (I).

**Input text**

```
Even without a high-school diploma, a black man can probably <event
id=79>find</event> a job if he <event id=80>looks</event>.
And some manual jobs, such as plumber or cable technician, pay quite
well.
```

**Output XML**

```
Even without a high-school diploma, a black man
<MODNEG_CONTEXT confidenceMod="0" confidenceNeg="100" modality="mod"
 type="followingVG followingMOD withingVG_subjunctive">
   can probably
   <EVENT confidenceNeg="0" confidenceMod="0" modality="mod">find</EVENT>
   a job
   if he
   <EVENT confidenceMod="0" confidenceNeg="0" negClass="none"
    modClass="mod" class="lookLike" modality="mod"
    kind="mod">looks</EVENT>
</MODNEG_CONTEXT>.
And some manual jobs, such as plumber or cable technician, pay quite
well.
```

**Source document**

```
music-can-hiphop-change-the-world-out.xml
```

**Fig. 5.** Example of an XML output document (II).

## 4 Results.

Attending to the confidence of the modal (*confidenceMod*) and negative (*confidenceNeg*) characters of the event that are annotated by the application, three runs have been generated for the task:

- ***Run 1*** (R1): only annotations (*modality* attribute) with the highest confidence levels (confidenceMod=0, confidenceNeg=0) are considered as positive.
- ***Run 2*** (R2): annotations with the highest and medium confidence levels (confidenceMod=[0|1], confidenceNeg=[0|1]) are considered as positive.
- ***Run 3*** (R3): the tagging is inferred from the combination of the annotations with the highest confidence and the evaluation derived from the factuality of the event (modClass and negClass attributes).

The results of the evaluation of the different runs are shown in the following tables. The first of them (**Table 3**) displays the confusion matrix, the second table (**Table 4**) shows the precision, recall and F-measure values for each run and class, and finally, **Table 5** shows the computing average scores.

**Table 3.** Evaluation results in terms of confusion matrix.

| | TP | | | FP | | | TN | | | FN | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **R1** | **R2** | **R3** | **R1** | **R2** | **R3** | **R1** | **R2** | **R3** | **R1** | **R2** | **R3** |
| **MOD** | 203 | **229** | 228 | **88** | 128 | 163 | **682** | 642 | 607 | 271 | **245** | 246 |
| **NEG** | **32** | 28 | 28 | **32** | 39 | 35 | **1148** | 1141 | 1145 | **32** | 36 | 36 |
| **NEGMOD** | 15 | **16** | **16** | **28** | 45 | 36 | **1175** | 1158 | 1167 | 26 | **25** | **25** |
| **NONE** | **565** | 516 | 490 | 281 | **243** | 248 | 298 | **336** | 331 | **100** | 149 | 175 |

**Table 4.** Evaluation results in terms of precision, recall, and F1-measure averaged.

| | Precision | | | Recall | | | F-measure (beta=1.0) | | |
|---|---|---|---|---|---|---|---|---|---|
| | R1 | R2 | R3 | R1 | R2 | R3 | R1 | R2 | R3 |
| **MOD** | **0.6976** | 0.6415 | 0.5831 | 0.4283 | **0.4831** | 0.4810 | 0.5307 | **0.5511** | 0.5272 |
| **NEG** | **0.5000** | 0.4179 | 0.4444 | **0.5000** | 0.4375 | 0.4375 | **0.5000** | 0.4275 | 0.4409 |
| **NEGMOD** | **0.3488** | 0.2623 | 0.3077 | 0.3659 | **0.3902** | **0.3902** | **0.3571** | 0.3137 | 0.3441 |
| **NONE** | 0.6678 | **0.6798** | 0.6640 | **0.8496** | 0.7759 | 0.7368 | **0.7478** | 0.7247 | 0.6985 |

**Table 5.** Evaluation results in terms of computing averaged.

| | R1 | R2 | R3 |
|---|---|---|---|
| **Macroaveraged F1 (beta=1.0)** | **0.5339** | 0.5043 | 0.5027 |
| **Microaveraged F1 (beta=1.0)** | **0.6395** | 0.6297 | 0.6083 |
| **Overall accuracy** | **0.6551 (815 out of 1244)** | 0.6342 (789 out of 1244) | 0.6125 (762 out of 1244) |

It can be observed that, given the same context for analysis, results improve when considering the inherent factuality of the event (R1 vs. R3) for the MOD and NEGMOD classes, but are worse for the NEG class. This is because of the fact that many events tagged as NEG have been retagged to NEGMOD when considering the factuality, which raises some confusion about how the factuality of the event should have been treated in the task.

The increase in the considered context (R2) has not a deterministic behaviour, contributing sometimes to increment the TP and the FP in other occasions. This is probably because the strategy of the definition of the context based only in the proximity of tokens is quite simple and the syntactic tree structure of the sentence where the event is located should have been considered.

## 5 Conclusions and Future work.

After a preliminary evaluation of the results, we have observed that, in general, the GOLD standard considers as modals those cases that include speculation, conjecture or raise some hypothesis in sentences with abstract or generic meaning, which our solution has not considered:

`(text 6, event 130): When you` **`look`** `at historical`
`(text 6, event 100): Populations` **`are`** `highly mobile`
`(text 6, event 43): You start` **`seeing`** `wild, arbitrary`

However, in the GOLD standard itself, some counter-examples that have not been annotated as MOD can easily be found.

In other cases, the GOLD standard has wrongly annotated NONE, especially in hypothetical contexts with the *one* pronoun (first example):

```
one is <event id=184>forced</event> to <event
id=185>confront</event> the problem that

A complex society <event id=134>develops</event> within
a local environment
```

Furthermore, our solution has not considered cases of subordinated verbs in expressions of evaluation or judgment as modal, as we think that they are more speculative than factual:

```
(text 6, event 125) It’s important to understand
(text 6, event 127) but it’s strange to think
```

We have observed many disagreements and ambiguities in the annotations: equivalent contexts have been differently annotated in different cases, in cases that include subordination sometimes the main verb is annotated and the subordinated verb in others, experts clearly disagree with the annotation of the GOLD standard, and even the expert themselves disagree with their own annotations.

These evidences show that there is still much to do in this field. Probably the first task should be to disambiguate, as far as possible, the cases to consider, and to elaborate clear guidelines with the annotation criteria in function of the context.

## Acknowledgements.

This work has been partially supported by several Spanish research projects: MULTIMEDICA: Multilingual Information Extraction in Health domain and application to scientific and informative documents (TIN2010-20644-C03-01), and MA2VICMR: Improving the access, analysis and visibility of the multilingual and multimedia information in web for the Region of Madrid (S2009/TIC-1542).

Authors would like to thank all partners for their knowledge and support.

## References.

1. Morante, R.: Modality and negation: *An introduction to the special issue. Computational Linguistics*. 2012. Vol 38, number 2.
2. Salkie R., Busuttil, B., and van der Auwera, J.: *Modality in English: Theory and Description*. 2009. Mouton de Gruyter, Berlin, Pages 1-8.
3. Jespersen, O.: *The philosophy of grammar*. 1924. Allen and Unwin, London.
4. Saurí, R., Verhagen, M., and Pustejovsky, J.: *Annotating and recognizing event modality in text*. 2006. Proceedings of FLAIRS 2006. Pages 333-339.

5. Payne, T. E.: *Describing Morphosyntax*. 1997. Cambridge University Press. Cabridge, UK. Page 282.
6. Saurí, R., Verhagen, M., and Pustejovsky, J.: *SlinkET: A partial modality parser for events*. 2006. Proceedings of LREC 2006, pages 1332-1337, Genoa.
7. Baker, K., Bloodgood, M., Dorr, B.J., Filardo, N.W., Levin, L. Piatko, C.: *A modality lexicon and its use in automatic tagging*. 2010. Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10).
8. Huang, Y., Lowe, H.J.: *A novel hybrid approach to automated negation detection in clinical radiology reports*. 2007. Journal of the American medical informatics association, 14(3):304-311.
9. Chapman, W.W., Bridewell, W., Hanbury, P., Cooper, G.F., and Buchanan, B.G.: *A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries*. 2002. Journal of Biomedical Informatics 34, 301–310.
10. Rokach, L., Romano, R., and Maimon, O.: *Negation recognition in medical narrative reports*. 2008. Information retrieval online, 11(6):446-538.
11. Myers, E.W., An, O.: *Difference algorithm and its variations*. 1986. Algorithmica, l (1), 251-266. Doi: 10.1007/BF01840446.
12. Rigoutsos, I., Floratos, A.: *Combinatorial pattern discovery in biological sequences: The TEIRESIAS algorithm*. 1998. Bioinformatics (Oxford, England), 14(2), 229.
13. GATE: General Arquitecture for Text Engineering. http://gate.ac.uk.
14. Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E.: *Longman Grammar Spoken and Written English (Grammar Reference)*. Pearson Longman, 2009.
15. Saurí, R.: *A Factuality Profiler for Eventualities in Text*. 2008.