Annotating Modality and Negation for a Machine Reading Evaluation

Roser Morante and Walter Daelemans

CLiPS - University of Antwerp
Prinsstraat 13, B-2000 Antwerpen, Belgium
{roser.morante,walter.daelemans}@ua.ac.be

Abstract. This paper describes the pilot task *Processing modality and negation*, which was organized in the framework of the Question Answering for Machine Reading Evaluation Lab at CLEF 2012. This task was defined as an annotation exercise consisting on determining whether an event mentioned in a text is presented as negated, modalised (i.e. affected by an expression of modality), or both. Three teams participated in the task submitting a total of 6 runs. The highest score obtained by a system was 0.6368 macroaveraged F_1 measure.

1 Introduction

This report describes the task *Processing modality and negation*, which is a pilot task of the *Question Answering for Machine Reading Evaluation* (QA4MRE)[1] Lab (14; 15) at CLEF 2012. Machine reading (4; 21) is a task that aims at automatic unsupervised understanding of texts. The goal of the QA4MRE evaluation is to develop a methodology for evaluating machine reading systems through question answering and reading comprehension tests. Participating systems are provided with a background collection on a certain topic, which has to be "understood" in order to answer multiple choice questions about documents of the same topic. Finding the correct answer might require performing some kind of inference and processing previously acquired background knowledge from the background collection. The main characteristics of the reading comprehension tests is that they assume a cognitive process that involves deriving implications and presuppositions, retrieving the stored information, and performing inferences to make implicit information explicit. Since modality and negation are very relevant phenomena for understanding texts, we proposed this pilot task in the framework of the QA4MRE Lab.

Modality and negation are two main grammatical categories used to express extra-propositional aspects of meaning, i.e. aspects of meaning, such as factuality, uncertainty, or subjectivity, which are used to convey information beyond the propositional content of the type "who does what when and where". Generally speaking, modality is a grammatical category that allows the expression of aspects related to the attitude of the speaker towards her statements. Negation

[1] http://celct.fbk.eu/QA4MRE/

is a grammatical category that allows to change the truth value of a proposition. Research on modality and negation in Natural Language Processing (NLP) has progressed thanks to the availability of data sets where extra-propositional aspects of meaning are annotated, such as the certainty corpus (17), the ACE 2008 corpus (9), the BioScope corpus (26), the FactBank corpus (20), and the annotation undertaken as part of the SIMT SCALE project (2). The CoNLL-2010 shared task (5) on hedge detection and the *SEM shared task on resolving the scope and focus of negation (11) have also boosted research on these topics. More information about modality and negation in NLP can be found in the special issue of the journal Computational Linguistics on Modality and Negation (12).

The paper is organized as follows. Section 2 defines the phenomena that are the focus of the task. Section 3 describes the task. Section 4 presents the participating teams and their results, and Section 5 puts forward some conclusions.

2 Modality and negation

2.1 Negation

In the context of this task, negation is understood as a grammatical phenomenon used to state that some event, situation, or state of affairs does not hold. Negation can be expressed by a variety of grammatical categories, as shown in the examples below.[2]

- Nouns:
  In the face of an international **inability** *to put the sort of price on carbon use that would drive its emission down, an increasing number of policy wonks, and the politicians they advise, are taking a more serious look at these other factors as possible ways of controlling climate change.*
- Verbs:
  A large gate at the front **prevents** *people from sleeping rough in the disused courtyard.*
- Prepositions:
  They simply asserted claims about Mr Obama **without** *providing the court (or anyone else) with any convincing reason to believe those claims.*
- Adverbs:
  The witnesses whom Ms Taitz called to testify (you can read them here, in the transcript) were **never** *tendered as experts.*
- Determiners:
  They usually have **no** *experience of the company's products or markets.*
- Pronouns:
  None *of these measures has come close to solving the problem.*
- Prefixes:
  The new pact has left some important problems **un***solved.*

[2] In the examples below cues are marked in bold and the negated and/or modalised event is underlined.

– Conjunctions:
Neither *the decision* **nor** *the changes themselves were based on anyones political beliefs or ideology.*

In the examples above negation is expressed explicitly, but it can also be expressed implicitly, for example, by means of certain linguistic contexts:

– Conditional constructions:
If matter and antimatter were truly symmetrical, then they would have come into existence in equal amounts during the Big Bang.
In this sentence the conditional construction determines that the events expressed by were and come into existence are implicitly negated.
– Combination of certain types of verbs and verb tenses:
The process to determine the Democratic nominee was supposed to have ended four years ago.
In this case, the use of the verb suppose and the past perfect tense indicates that the event 'end' has not happened.

A description of negative contexts is presented in (25), and a description of negation in English in (22). (6) is an exhaustive study about how negation has been treated throughout history. A list of negation cues in biomedical language can be extracted from the BioScope corpus (26).

2.2 Modality

From a theoretical perspective, modality can be defined as a philosophical concept, as a subject of the study of logic, or as a grammatical category. There are many definitions and classifications of modal phenomena. For this task we understand modality in a broad sense. Modality will not only refer to epistemic modality (typically expressed by modal verbs), but also to concepts such as hedging, uncertainty, factuality, evidentiality, and subjectivity. These concepts are related to the expression of the attitude of the speaker towards her statements in terms of degree of certainty, reliability, subjectivity, sources of information, and perspective.

Epistemic modality, as described by Lyons (10, p.793), is concerned with matters of knowledge and belief, "the speakers opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes". Palmer defines two types of propositional modality: epistemic, used by speakers "to express their judgement about the factual status of the proposition, and evidential, used to indicate the evidence that they have for its factual status (13, p.8–9). The term *hedging* is originally due to Lakoff, who describes hedges as "words whose job is to make things more or less fuzzy" (8, p. 195). Evidentiality is related to the expression of the information source of a statement (1, p.1). Certainty is a type of subjective information that can be conceived of as a variety of epistemic modality (18). Factuality involves polarity, epistemic modality, evidentiality and mood. It is defined by Saurí (19, p.1) as: "the level

of information expressing the commitment of relevant sources towards the factual nature of eventualities in text. That is, it is in charge of conveying whether eventualities are characterized as corresponding to a fact, to a possibility, or to a situation that does not hold in the world." The term subjectivity is introduced by Banfield (3). According to Wiebe et al. (27, p.279), "subjective language is language used to express private states in the context of a text or conversation. Private state is a general covering term for opinions, evaluations, emotions, and speculations."

In the context of this task, we will consider that an event is modalised when it is not presented as certain or factual. Many linguistic devices can be used to present an event as modalised.

- Modal verbs:
  Mr Sakurai fears many of Minamisomas evacuees **may** *never come back.*
  These alternatives **could** *also improve the content and prospects of other climate action.*
  Global greenhouse-gas emissions **must** *fall by half to limit climate change.*
- Epistemic adjectives:
  Providing most of that energy from wind, sunshine, plants and rivers, along with a bit of nuclear, is **possible**.
- Epistemic adverbs:
  It will **probably** *never again generate the majority of America's energy.*
- Epistemic nouns:
  Insiders reckon the **possibility** *of being let off the hook by a new Administration.*
- Propositional attitude verbs and adjectives:
  We do not **believe** *these attacks breached the servers that support our Domain Name System network.*
  We **hope** *to unveil it before the month is up. The ECB was* **considering** *writing down the value of its Greek bonds to the price it paid for them*
- Generics and habituals:
  American universities are **usually** *happy to accept such good students.*
  There **is** *a big difference between drawing a map and following it.*
- Future tense:
  They **will** *start to decide that its not worth the money.*
- Conditional constructions:
  If *you are highly motivated to minimise your taxes, you can hunt for every possible deduction for which youre eligible.*
  The investment required to decarbonise power **would** *average about 30 billion ($ 42 billion) a year over 40 years.*
- Expression of purpose/goal:
  Europe **has set a goal** *of reducing emissions by 80-95% by 2050.*
  The investment required **to** *decarbonise power would average about 30 billion ($ 42 billion) a year over 40 years.*
- Expression of need:
  China has less urgent **need** to bolster growth.

- Expression of obligation:
  All that gassy baggage **must** *go.*
  Rich countries **should** *cut the most.*
- Expression of desire: They **want** it raised to 30%.
- Epistemic judgment verbs:
  Suggesting *that such a large number of Americans are doing a job that is no longer necessary was perhaps not the wisest move politically.*
  We can **assume** *that this has probably been known about since the beginning of this century.*
- Epistemic evidential verbs:
  Turkey **seems** *to favour a rival Russian-backed project.*
- Epistemic deductive verbs:
  Some **deduce** *from the overall picture that as China and other authoritarian states get more educated and richer, their people will agitate for greater political freedom, culminating in a shift to a more democratic form of government.*

More information about modality can be found in the study by Portner (16). A description of modality types is presented in (2), and an exhaustive description about how to annotate information related to factuality can be found in (19). A description of hedging in scientific text is presented in (7), and from the BioScope corpus (26) a list can be extracted of hedge cues in biomedical texts.

3 Task description

The exercise *Processing modality and negation* is defined as an annotation task in which 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. This information can be relevant for machine reading systems, since negated and modalised events should be treated differently than factual events in the inference making process. The term *event* is understood in a broad sense to refer to events and states.

The input for a system is a text where all events expressed by verbs are identified automatically. The identification of verbs is done automatically with the Stanford POS Tagger (v. 3.0, 2010-05-10) (24; 23). Verbs that are not identified by the tagger are not marked. Only the main verb of a verbal form is marked. The output should be a label per event. The possible values of the label are four: MOD, NEG, NEGMOD, NONE.

- An event is assigned the tag NONE when it is presented as certain and it happened (e.g., Half of Europe's electricity comes from fossil fuels).
- An event is assigned the tag NEG when it is presented as certain and did not happen (e.g., Half of Europe's electricity does not come from fossil fuels).
- An event is assigned the tag MOD when it is not presented as certain and is not negated (e.g., Half of Europe's electricity might come from fossil fuels).

– An event is assigned the tag NEGMOD when it is not presented as certain and is negated (e.g., Half of Europe's electricity might not come from fossil fuels).

Figure 1 shows an example of an input sentence and the output expected from a system.

INPUT
Some <event id=1>deduce</event> from the overall picture that as China and other authoritarian states <event id=2>get</event> more educated and richer, their people will <event id=3>agitate<event> for greater political freedom, <event id=4>culminating</event> in a shift to a more democratic form of government.

OUTPUT
e1=NONE e2=MOD e3=MOD e4=MOD

Fig. 1. Example of imput-ouput.

The organization provided one example document with gold annotations and 8 test documents in English for evaluation, two for each of the topics of the QA4MRE task: AIDS, climate change, music and society, and Alzheimer's disease. The test documents are articles from the journal *The Economist*.[3] The gold annotations of the test documents were released when the evaluation period finished. No training documents were provided for this edition. The total number of events and the class distribution is presented in Table 1. The event labels are distributed similarly across topics, with the exception of MOD, which is more frequent in the climate documents.

Topic	**# Events**	**NONE**	%	**MOD**	%	**NEG**	%	**NEGMOD**	%
Aids	247	139	56,27	95	38,46	5	2,02	8	3,23
Alzheimer	180	99	55,00	65	36,11	9	5,00	7	3,88
Climate	590	264	44,74	269	45,59	33	5,60	24	4,06
Music	247	139	56,27	95	38,46	5	2,02	8	3,24
All	1244	655	52,65	474	38,10	64	5,14	41	3,29

Table 1. Number of events and class distribution.

In order to solve the task, participants could use any existing resources such as corpora, lexicons or NLP tools such as factuality profilers, scope resolvers, hedge and negation detectors, etc. The only requirement was that the task be solved automatically.

[3] *The Economist* kindly made available the texts for non-commercial research purposes.

The output of systems was evaluated against a manually annotated gold standard. The test documents were annotated by two annotators. As indicated above, the labels to be annotated were four: NONE, MOD, NEG, NEGMOD. The annotators used the decision tree presented in Figure 2 in order to decide which label was to be assigned to an event.

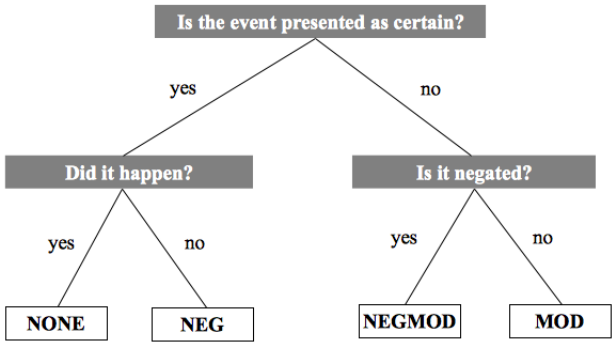

Fig. 2. Decision tree.

4 Participation and results

3 groups participated submitting 6 runs. Table 2 shows the list of participating teams and the reference to their reports.

Team	Institutions	Reference
CLaC	CLaC Lab Concordia University, Canada	Rosenberg et al.
desancis	Universidad Politécnica de Madrid, Spain	Lana-Serrano et al.
	Universidad Carlos III de Madrid, Spain	
	Universidad Autónoma de Madrid, Spain	
	DAEDALUS, Madrid	
JUCSENLP	Jadavpur University, India	Pakray et al.
	National Polytechnic Institute, Mexico	

Table 2. Participating teams with reference to their reports.

The task was evaluated at the event level in terms of F_1. The highest score obtained by a system was 0.6368 macroaveraged F_1measure. Table 3 indicates the number of runs submitted per team and the highest macroaveraged F_1obtained. A baseline was calculated by assigning the majority class (MOD) to all events, which scored 0.1380. All submitted runs score above baseline.

Team	# of runs	Highest F_1
CLaC	2	0.6368
desancis	3	0.5339
JUCSENLP	1	0.3378
baseline (all MOD)	-	0.1380

Table 3. Number of runs and highest scores per team.

The CLaC team extended NEGATOR, a heuristics-based system initially developed to process negation. The system is composed of three modules: the first component detects and annotates negation and modality cues; the second component detects and annotates the syntactic scope of the detected negation and modality cues based on information from the dependency graph. The third component determines which of the annotated events are within the detected scopes of negation, modality or within the intersection of both in order to assign the event labels.

The desancis team developed also a rule-based system defined in JAPE (Java Annotation Patterns Engine). The system is composed of three modules: the VG module tags verbal groups with lexical category, mode, tense, aspect, voice, and modality; the MODNEG module tags particles that may be related to modality and/or negation; the LABELER module tags the event under analysis by means of rules that use contextual information about modality and negation.

The JUCSENLP team developed a system that relies on a list of modality and negation cues. The label assigned to an event depends on whether the event is preceded in the sentence by the modality and negation cues contained in the list. More information about the systems can be found in the individual description papers referenced in Table 2.

The scores per run are provided in Table 4 in terms of overall macroaveraged F_1, overall accuracy and F_1 per label. Considering that modality and negation tagging is a recently emerged task and that the corpus provided by the organization is small, we consider that the results obtained by systems are encouraging.

	Overall results		F_1 per label			
Run	Macro. F_1	Accuracy	NONE	MOD	NEG	NEGMOD
CLaC 1	0.6368	0.7130	0.7657	0.6653	0.6545	0.4615
CLaC 2	0.6196	0.6688	0.7008	0.6593	0.6667	0.4516
desancis 1	0.5339	0.6551	0.7478	0.5307	0.5000	0.3571
desancis 2	0.5043	0.6342	0.7247	0.5511	0.4275	0.3137
desancis 3	0.5027	0.6125	0.6985	0.5272	0.4409	0.3441
JUCSENLP 1	0.3378	0.6262	0.7219	0.5933	0.0000	0.0360

Table 4. Overall results per run.

5 Conclusions

In this paper we described the task *Processing modality and negation* organized in the framework of the QA4MRE Lab at CLEF. The task was defined as an annotation exercise in which systems had to determine whether an event mentioned in a text is negated and/or modalised. Participants were provided with 8 test documents, 2 per each of the following topics: AIDS, music, climate change, Alzheimer's disease. No training documents were provided, since the task was not intended as a machine learning task. The three teams that participated submitted heuristics based systems that outperformed a majority class baseline. Taking into consideration that this is a new task and that the task is difficult, we consider that the results obtained by participants with heuristics based systems are encouraging.

Acknowledgments

This study was made possible through financial support from the University of Antwerp (GOA project BIOGRAPH). We are grateful to the organizers of the QA4MRE lab at CLEF 2012 for their support and for hosting the pilot task. We also thank Janneke van de Loo for helping with the data annotation.

Bibliography

[1] Aikhenvald, A.: Evidentiality. Oxford University Press, New York, USA (2004)

[2] Baker, K., Bloodgood, M., Dorr, B., Filardo, N.W., Levin, L., Piatko, C.: A modality lexicon and its use in automatic tagging. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10). pp. 1402–1407. European Language Resources Association (ELRA), Valetta, Malta (2010)

[3] Banfield, A.: Unspeakable sentences. Routledge and Kegan Paul, Boston (1982)

[4] Etzioni, O., Banko, M., Cafarella, M.J.: Machine reading. In: Proceedings of the 21st National Conference on Artificial Intelligence. vol. 2, pp. 1517–1519. Boston, Massachusetts (2006)

[5] Farkas, R., Vincze, V., Móra, G., Csirik, J., Szarvas, G.: The CoNLL 2010 shared task: Learning to detect hedges and their scope in natural language text. In: Proceedings of the CoNLL2010 Shared Task. Association for Computational Linguistics, Uppsala, Sweden (2010)

[6] Horn, L.: A natural history of negation. Chicago University Press, Chicago (1989)

[7] Hyland, K.: Hedging in scientific research articles. John Benjamins B.V, Amsterdam (1998)

[8] Lakoff, G.: Hedges: a study in meaning criteria and the logic of fuzzy concepts. Chicago Linguistics Society Papers 8, 183–228 (1972)

[9] Linguistic Data Consortium: ACE (Automatic Content Extraction) English annotation guidelines for relations. Tech. Rep. Version 6.2 2008.04.28, LDC (2008)

[10] Lyons, J.: Semantics. CUP, Cambridge (1977)

[11] Morante, R., Blanco, E.: *SEM 2012 shared task: Resolving the scope and focus of negation. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics. Montreal (2012)

[12] Morante, R., Sporleder, C.: Special issue on modality and negation: An introduction. Computational Linguistics (2012)

[13] Palmer, F.: Mood and modality. CUP, Cambridge, UK (1986)

[14] Peñas, A., Hovy, E.H., Forner, P., Rodrigo, Á., Sutcliffe, R.F.E., Forascu, C., Sporleder, C.: Overview of QA4MRE at CLEF 2011: Question answering for machine reading evaluation. In: CLEF (Notebook Papers/Labs/Workshop) (2011)

[15] Peñas, A., Hovy, E.H., Forner, P., Rodrigo, Á., Sutcliffe, R.F.E., Sporleder, C., Forascu, C., Benajiba, Y., Osenova, P.: Overview of QA4MRE at CLEF 2012: Question answering for machine reading evaluation. In: CLEF 2012 Evaluation Labs and Workshop - Working Notes Papers (2012)

[16] Portner, P.: Modality. Oxford University Press, Oxford, UK (2009)

[17] Rubin, V.L.: Stating with certainty or stating with doubt: intercoder reliability results for manual annotation of epistemically modalized statements. In: NAACL '07: Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers on XX. pp. 141–144. Association for Computational Linguistics, Morristown, NJ, USA (2007)

[18] Rubin, V., Liddy, E., Kando, N.: Certainty identification in texts: Categorization model and manual tagging results. In: Computing Attitude and Affect in Text: Theory and Applications, Information Retrieval Series, vol. 20, pp. 61–76. Springer-Verlag, New York (2005)

[19] Saurí, R.: A factuality profiler for eventualities in text. Ph.D. thesis, Brandeis University, Waltham, MA, USA (2008)
[20] Saurí, R., Pustejovsky, J.: FactBank: A corpus annotated with event factuality. Language Resources and Evaluation 43(3), 227–268 (2009)
[21] Strassel, S., Adams, D., Goldberg, H., Herr, J., Keesing, R., Oblinger, D., Simpson, H., Schrag, R., Wright, J.: The DARPA machine reading program - encouraging linguistic and reasoning research with a series of reading tasks. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valletta, Malta (2010)
[22] Tottie, G.: Negation in English speech and writing: a study in variation. Academic Press, New York (1991)
[23] Toutanova, K., Klein, D., Manning, C., Singer, Y.: Feature-rich part-of-speech tagging with a cyclic dependency network. In: Proceedings of HLT-NAACL 2003. pp. 252–259. Hong Kong (2003)
[24] Toutanova, K., Manning, C.: Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000). Hong Kong (2000)
[25] van der Wouden, T.: Negative contexts: collocation, polarity, and multiple negation. Routledge, London (1997)
[26] Vincze, V., Szarvas, G., Farkas, R., Móra, G., Csirik, J.: The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics 9((Suppl 11)), S9 (2008)
[27] Wiebe, J., Wilson, T., Bruce, R., Bell, M., Martin, M.: Learning subjective language. Computational Linguistics 30(3), 277–308 (2004)