

Track Guidelines

TASK DESCRIPTION

Until recently, research on Natural Language Processing (NLP) has focused on propositional aspects of meaning. For example, semantic role labeling, question answering or text mining tasks aim at extracting information of the type "who does what when and where". However, understanding language involves also processing extra-propositional aspects of meaning, such as factuality, uncertainty, or subjectivity. Modality and negation are two main grammatical categories used to express extra-propositional aspects of meaning. Generally speaking, *modality* is a grammatical category that allows the expression aspects related to the attitude of the speaker towards her statements. *Negation* is a grammatical category that allows to change the truth value of a proposition.

Research on modality and negation in NLP has progressed thanks to the availability of data sets annotated with various aspects of modality and negation, such as the certainty corpus (Rubin 2007), the ACE 2008 corpus (Linguistic Data Consortium, 2008), the BioScope corpus (Vincze et al., 2008), the FactBank corpus (Saurí and Pustejovsky, 2009), and the annotation undertaken as part of the SIMT SCALE project (Baker et al., 2010), as well as the organization of the CoNLL-2010 Shared Task (Farkas et al., 2010). More information about modality and negation in NLP can be found in the forthcoming special issue of the journal *Computational Linguistics* on Modality and Negation (Morante and Sporleder 2012).

Machine Reading (Etzioni et al., 2006, Strassel et al., 2010) 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 propose this pilot task on 'Processing modality and negation' in the framework of the Question Answering for Machine Reading Evaluation (QA4MRE) at CLEF 2012.

This task is defined as 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. 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 will be a text where all events expressed by verbs are identified, and the output should be a label per event. The possible values of the label are: MOD, NEG, NEGMOD, NONE.

In order to solve the task, participants can use any existing resources such as corpora, lexicons or NLP tools such as factuality profilers, scope labelers, hedge taggers, etc. The only requirement is that the task has to be solved automatically.

MODALITY AND NEGATION

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.¹

- **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>.*
- **Conjunctions:** *[Neither] the decision [nor] the changes themselves were <based> on anyone's 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 van der Wouden (1997), and a description of negation in English in Tottie (1991). Horn (1989) 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 (Vincze et al., 2008).

¹ In the examples, negation and modality cues, i.e., words that express negation or modality, are marked between square brackets and the negated or modalised event is marked between angle brackets. Examples are extracted from articles in the journal *The Economist*.

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 (1977:793), is concerned with matters of knowledge and belief, "the speaker's 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" (Palmer 1986: 8–9). The term 'hedging' is originally due to Lakoff (1972), who describes hedges as "words whose job is to make things more or less fuzzy" (Lakoff 1972: 195). Evidentiality is related to the expression of the information source of a statement (Aikhenvald 2004:1). Certainty is a type of subjective information that can be conceived of as a variety of epistemic modality (Rubin et al., 2005). Factuality involves polarity, epistemic modality, evidentiality and mood. It is defined by Sauri (2008, 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 (1982). According to Wiebe et al., (2004: 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 Minamisoma's 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 it'[s] not worth the money.*
- **Conditional constructions:** *[If] you <are> highly motivated to <minimise> your taxes, you can <hunt> for every possible deduction for which you're <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 (Portner 2009). A description of modality types is presented in Baker et al., (2009), and an exhaustive description about how to annotate information related to factuality can be found in Saurí (2008). A description of hedging in scientific text is presented in Hyland (1998), and from the BioScope corpus (Vincze et al., 2008) a list can be extracted of hedge cues in biomedical texts.

ANNOTATION

The labels to be annotated are four: NONE, MOD, NEG, NEGMOD.

For the annotation of the test documents, the decision tree presented in Figure 1 is used to decide which label has to be assigned to an event.

- 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*).

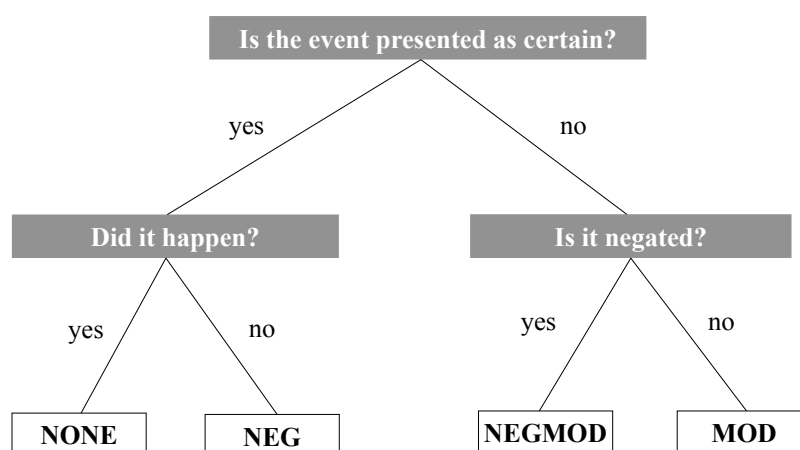


Figure 1. Decision tree.

TEST DOCUMENTS

The test set will be composed of 8 documents in English, two for each of the topics of the main task: AIDS, climate change, music and society, and Alzheimer's disease. The test documents are articles from the journal *The Economist*².

An example test document is provided in the Appendix I. The test documents are provided in text format. All verbs in the document are marked in the text with xml tags of the type `<event id=[id number]>[verb]</event>`. The identification of verbs is done automatically with the Stanford POS Tagger (v. 3.0, 2010-05-10) (Toutanova and Manning 2000, Toutanova et al. 2003). Verbs that are not identified by the tagger are not marked. Only the main verb of a verbal form is marked. There are around 100 verbs per document.

Given the text and the identified verbs, systems have to produce the annotation, as in the example test in Appendix I.

EVALUATION

Systems will be evaluated in terms of F measures: precision, recall, and F1. The output of the systems will be evaluated against gold-standard data. The test documents will be released on June 5. Submissions should be due by June 15.

RUNS

Participants are allowed to submit up to three runs.

² *The Economist* kindly made available the texts for non-commercial research purposes.

SUBMISSION PROCEDURE

The output of systems should be submitted in a compressed file by e-mail to the task organisers (see below). The subject of the e-mail should be "[qa4mre-pilot modality] submission". The output for all the tests should be included in the same file. The output for a test starts with a line "OUTPUT TEST[test number]" and is followed by lines containing the label per event following the format "e[event number]=LABEL". The information for each event is written in a line, as in:

```
OUTPUT TEST[test number]
e[event number]=LABEL
e[event number]=LABEL
e[event number]=LABEL
OUTPUT TEST[test number]
e[event number]=LABEL
e[event number]=LABEL
e[event number]=LABEL
```

With the system output, participants should submit a brief description of the system and of the characteristics of each run, if they submit more than one.

IMPORTANT DATES

Test set release	June 5
Run submissions	June 15
Individual results to participants	June 25
Submissions of Working Notes Papers	August 17
CLEF Workshop	September 17-20, Rome, Italy

ORGANISERS

Roser Morante
Walter Daelemans
CLiPS, University of Antwerp, Belgium
{Roser.Morante,Walter.Daelemans}@ua.ac.be

References

- Aikhenvald, A. Y. (2004) *Evidentiality*. Oxford University Press, New York, USA.
- Baker, K., M. Bloodgood, M. Diab, B. J. Dorr, E. Hovy, L. Levin, M. McShane, T. Mitamura, S. Nirenburg, C. Piatko, O. Rambow, and G. Richardson (2010). SIMT SCALE 2009 - modality annotation guidelines. Technical Report 004, Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore.
- Banfield, A. (1982) *Unspeakable sentences*. Routledge and Kegan Paul, Boston.
- Etzioni, O., M. Banko, and M. J. Cafarella (2006) Machine reading. *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI 2006)*.
- Etzioni, O., A. Fader, J. Christensen, and S. Soderland (2011) Mausam: Open information extraction: The second generation. In *Proceedings IJCAI 2011*, pages 3-10).
- Farkas, R., V. Vincze, G. Móra, J. Csirik, and G. Szarvas (2010). The CoNLL 2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the CoNLL2010 Shared Task*, Uppsala, Sweden. Association for Computational Linguistics.
- Hyland, K. (1998) *Hedging in scientific research articles*. John Benjamins B.V, Amsterdam.
- Lakoff, G. (1972) Hedges: a study in meaning criteria and the logic of fuzzy concepts. *Chicago Linguistics Society Papers*, 8:183–228.
- Linguistic Data Consortium (2008). ACE (Automatic Content Extraction) English annotation guidelines for relations. Technical Report Version 6.2 2008.04.28, LDC.
- Lyons, J. (1977) *Semantics*. CUP, Cambridge.
- Morante, R. and C. Sporleder (2012)(Eds.) Special Issue on Modality and Negation. *Computational Linguistics*. To appear.
- Palmer, F. R. (1986) *Mood and modality*. CUP, Cambridge, UK.
- Portner, P. (2009) *Modality*. Oxford University Press, Oxford, UK.
- Rubin, V. L., E. Liddy, and N. Kando (2005) Certainty identification in texts: Categorization model and manual tagging results. In *Computing Attitude and Affect in Text: Theory and Applications*, volume 20 of Information Retrieval Series. Springer-Verlag, New York, pages 61–76.
- Rubin, V. L. (2006). Identifying certainty in texts. Ph. D. thesis, Siracuse University, Syracuse, NY, USA.
- Rubin, V. L. (2007). 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, Morristown, NJ, USA, pp. 141–144. Association for Computational Linguistics.
- Saurí, R. (2008) A factuality profiler for eventualities in text. Ph.D. thesis, Brandeis University, Waltham, MA, USA.
- Saurí, R. and J. Pustejovsky (2009). FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation* 43(3), 227–268.
- Strassel, S., D. Adams, H. Goldberg, J. Herr, R. Keesing, D. Oblinger, H. Simpson, R. Schrag, and J. Wright (2010) The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Tottie, G. (1991) *Negation in English speech and writing: a study in variation*. Academic Press, New York.

- Toutanova, K. and Ch. D. Manning (2000) Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, Hong Kong.
- Toutanova, K., D. Klein, Ch. Manning, and Y. Singer (2003) Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003* pages 252-259.
- van der Wouden, T. (1997) *Negative contexts: collocation, polarity, and multiple negation*. Routledge, London.
- Vincze, V., G. Szarvas, R. Farkas, G. Móra, and J. Csirik (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* 9((Suppl 11)), S9.
- Wiebe, J., Th. Wilson, R. Bruce, M. Bell, and M. Martin. 2004. Learning subjective language. *Computational Linguistics*, 30(3):277–308.

APPENDIX I: Example test document

Europe's climate policy

<event id=1>Being</event> ambitious

The European Commission maps a path to a low-carbon future. Now to <event id=2>walk</event> it

Mar 10th 2011 from the print edition

About half Europe's electricity <event id=3>comes</event> from fossil fuels, with CO2 emissions as an unwanted by-product. By 2050, <event id=4>proposes</event> a ``road map'' <event id=5>released</event> by the European Commission this week, all that gassy baggage must <event id=6>go</event>. <event id=7>Believing</event> that global greenhouse-gas emissions must <event id=8>fall</event> by half to <event id=9>limit</event> climate change, and that rich countries should <event id=10>cut</event> the most, Europe has <event id=11>set</event> a goal of <event id=12>reducing</event> emissions by 80-95% by 2050. The road map <event id=13>is</event> its first stab at <event id=14>sharing</event> out the cuts (<event id=15>see</event> chart). Since some greenhouse-gas emissions, such as those from farming, <event id=16>are</event> hard to <event id=17>curb</event>, this target <event id=18>means</event> that CO2 from power generation must <event id=19>shrink</event> almost to nothing -- even as electricity production ramps up by 50% to <event id=20>provide</event> more juice for electric cars and a lot more houses. <event id=21>Providing</event> most of that energy from wind, sunshine, plants and rivers, along with a bit of nuclear, <event id=22>is</event> possible. The biggest technical caveat <event id=23>is</event> that most studies <event id=24>include</event> a lot of fossil-fuel plants in which CO2 is <event id=25>whisked</event> out of chimneys and <event id=26>tucked</event> underground without <event id=27>troubling</event> the atmosphere, and this capture and storage technology has yet to be <event id=28>proved</event> on the scale <event id=29>needed</event>. The plan <event id=30>requires</event> a lot of investment in power generation and smarter grids, best <event id=31>done</event> in the context of -- at long last -- a <event id=32>reformed</event> and competitive energy market. The commission <event id=33>says</event> the investment <event id=34>required</event> to decarbonise power would <event id=35>average</event> about 30 billion (\$ 42 billion) a year over 40 years. This <event id=36>is</event> one of the cheaper parts of the plan; the total cost <event id=37>is</event> about 270 billion a year, with 80 billion <event id=38>going</event> on buildings and appliances and 150 billion on transport. But the commission's modelling also <event id=39>points</event> to savings on fuel costs, which <event id=40>are</event> low for nuclear and zero for most renewables, of between 175 billion and 320 billion. Other benefits <event id=41>include</event> more energy security and cleaner air. Some governments, <event id=42>including</event> France's and Britain's, <event id=43>are</event> so convinced of the benefits of green investment that they <event id=44>think</event> the commission's near-term goal of a 20% emission reduction by 2020 <event id=45>is</event> too low. They <event id=46>want</event> it <event id=47>raised</event> to 30%. Others <event id=48>argue</event> strenuously for the limit to be <event id=49>kept</event> as it <event id=50>is</event>, as <event id=51>do</event> some industry lobbies. The road map <event id=52>says</event> that an optimal path would <event id=53>require</event> a larger cut by 2020 -- but that, as luck would <event id=54>have</event> it, the optimal reduction, <event id=55>set</event> at 25%, <event id=56>is</event> what will be <event id=57>realized</event> by <event id=58>existing</event> plans for more renewable energy and <event id=59>improved</event> energy efficiency if they <event id=60>work</event> out. One

problem with this <event id=61>is</event> that the efficiency goal of a 20% improvement by 2020 <event id=62>seems</event> high. The commission <event id=63>says</event> the EU <event id=64>is</event> on course to <event id=65>do</event> only half as well. In an energy-efficiency plan <event id=66>released</event> with its road map, the commission talks of <event id=67>making</event> the goals binding in 2013. And if the targets are <event id=68>met</event>, the carbon market may <event id=69>suffer</event>. The number of permits to be <event id=70>issued</event> to carbon emitters from 2013 to 2020 has already been <event id=71>fixed</event>. <event id=72>Hitting</event> the energy-efficiency targets would <event id=73>reduce</event> demand for those permits, in which case their price might <event id=74>fall</event> too low to <event id=75>encourage</event> further investment into low-carbon technologies. In a victory over some of her colleagues, Connie Hedegaard, the climate commissioner, has <event id=76>got</event> the commission to <event id=77>agree</event> to <event id=78>propose</event> <event id=79>tightening</event> up on emissions allowances in 2013-20. A political decision to <event id=80>approve</event> this, which will <event id=81>involve</event> <event id=82>haggling</event> over where revenues from the trading scheme <event id=83>go</event>, <event id=84>is</event> probably not within reach until the Danes <event id=85>take</event> the EU's <event id=86>rotating</event> presidency in 2012. That <event id=87>offers</event> a chance for Denmark to <event id=88>redeem</event> itself, <event id=89>quips</event> Abyd Karmali, head of carbon markets for Bank of America Merrill Lynch, <event id=90>referring</event> to the dismal 2009 Copenhagen climate conference that Ms Hedegaard <event id=91>ran</event>. The road map <event id=92>shows</event> that some of that conference's ambitions still <event id=93>resonate</event> in Brussels. But there <event id=94>is</event> a big difference between <event id=95>drawing</event> a map and <event id=96>following</event> it.

OUTPUT TEST1

e1=NONE
e2=MOD
e3=NONE
e4=NONE
e5=NONE
e6=MOD
e7=NONE
e8=MOD
e9=MOD
e10=MOD
e11=NONE
e12=MOD
e13=NONE
e14=MOD
e15=MOD
e16=NONE
e17=MOD
e18=NONE
e19=NONE
e20=MOD
e21=MOD
e22=NONE
e23=NONE
e24=NONE

e25=NONE
e26=NONE
e27=NEG
e28=MOD
e29=NONE
e30=NONE
e31=MOD
e32=MOD
e33=NONE
e34=NONE
e35=MOD
e36=NONE
e37=NONE
e38=MOD
e39=NONE
e40=MOD
e41=MOD
e42=NONE
e43=NONE
e44=NONE
e45=MOD
e46=NONE
e47=MOD
e48=NONE
e49=MOD
e50=NONE
e51=NONE
e52=NONE
e53=MOD
e54=MOD
e55=NONE
e56=NONE
e57=MOD
e58=NONE
e59=MOD
e60=MOD
e61=NONE
e62=NONE
e63=NONE
e64=MOD
e65=MOD
e66=NONE
e67=MOD
e68=MOD
e69=MOD
e70=MOD
e71=NONE
e72=MOD
e73=MOD
e74=MOD
e75=MOD

e76=NONE
e77=NONE
e78=NONE
e79=MOD
e80=MOD
e81=MOD
e82=MOD
e83=MOD
e84=MOD
e85=MOD
e86=NONE
e87=NONE
e88=MOD
e89=NONE
e90=NONE
e91=NONE
e92=NONE
e93=NONE
e94=NONE
e95=MOD
e96=MOD