Modality and Negation in Natural Language Processing

Roser Morante

CLiPS - University of Antwerp

November 8, 2011

IJCNLP 2011 Tutorial, Chiang Mai, Thailand
Somalia has **not** had an effective central government since 1991, when the former government was toppled by clan militias that later turned on each other. For decades, generals, warlords and warrior types have reduced this once languid coastal country in Eastern Africa to rubble. Somalia remains a raging battle zone today, with jihadists pouring in from overseas, intent on toppling the transitional government.

**No amount of** outside firepower has brought the country to heel. **Not** thousands of American Marines in the early 1990s. **Not** the enormous United Nations mission that followed. **Not** the Ethiopian Army storming into Somalia in 2006. **Not** the current African Union peacekeepers, who are steadily wearing out their welcome.

Modality

Environmentalists may regard such schemes with mixed feelings. Carbon-neutral extraction would do nothing to cut the bulk of oil-related emissions that come from combustion. Eco-friendlier tar sands could also encourage unconventional development elsewhere: Jordan, Madagascar, Congo and Venezuela, where the government claims a reserve of bitumen even greater than Alberta’s, may be less open to environmental scrutiny. Kill Alberta’s tar sands, say some, and rising crude prices would choke oil consumption and force an era of clean energy into being.

Source: http://www.economist.com/node/17959688?story_id=17959688
Uncertainty about exoplanets

Habitable Super-Earth?

The planet HD 85512 b orbits within its star’s habitable zone. Liquid water, a vital requirement for life as we know it, could exist on its surface.

Distance from Earth: 35 light-years
Mass: 3.6 times that of Earth
Surface temperature: 77 degrees F (25 degrees C)

ARTIST’S CONCEPTION OF HD 85512 b (CREDIT: M. KORNMESSER, EUROPEAN SOUTHERN OBSERVATORY).

HD 85512
Spectral type K5V

RELATIVE DISTANCE OF PLANETS IN OUR SOLAR SYSTEM:

SOURCE: EUROPEAN SOUTHERN OBSERVATORY
http://www.exoplanet.hanno-rein.de/
Statements about exoplanets

**Same proposition, different meanings**

- Other types of life have taken root in planet HD85512b
- Other types of life *could conceivably* take root in planet HD85512b
- Have other types of life taken root in planet HD85512b?
- Other types of life *will never* take root in planet HD85512b
- Other types of life *might* have taken root in planet HD85512b
- If 60% of the planet is covered in cloud, other types of life *will probably* take root in planet HD85512b
- *It is expected that* other types of life have taken root in planet HD85512b
- *It has been denied* that other types of life have taken root in planet HD85512b
Opinions about exoplanets

just think we could have come from those planets in the first place. Maybe we screwed up those first and had to get away from them. And possibly we lost the technology due to war of our people from along time ago. maybe the (otherside) won to the point of almost human annihilation here on earth and we are just survivors from the war. once you colonize a planet or lands you kill the competition. you win. them stupidity follows suit again history repeats it self. we need to realize we are from the universe understandin ourselves.

Source:
Last consulted 14 October 2011
Outline

Part 1: Introduction: Modality and Negation
Part 2: Categorising and Annotating Modality and Negation
Part 3: Tasks Related to Processing Modality and Negation
Part 4: Modality and Negation in Applications
Part I

Introduction: Modality and Negation
Outline

1. Defining modality
   - Related concepts

2. Defining negation

3. Why is it interesting to process modality and negation?

4. References
Outline

1. Defining modality
   - Related concepts

2. Defining negation

3. Why is it interesting to process modality and negation?

4. References
Modality (von Fintel 2006)

“Modality is a category of linguistic meaning having to do with the expression of possibility and necessity. A modalized sentence locates an underlying or prejacent proposition in the space of possibilities.

*Sandy might be home*
says that there is a possibility that Sandy is home.

*Sandy must be home*
says that in all possibilities, Sandy is home.”
Modality as displacement (von Fintel 2006)

“The counterpart of modality in the temporal domain should be called “temporality”, but it is more common to talk of tense and aspect, the prototypical verbal expressions of temporality.

Together, modality and temporality are at the heart of the property of “displacement” (...) that enables natural language to talk about affairs beyond the actual here and now.”
Defining modality

**Modality, tense, aspect** (Palmer 2001)

- **Tense**: is concerned with the time of the event
- **Aspect**: is concerned with the nature of the event, in terms of the internal temporal constituency
- **Modality**: is concerned with the status of the proposition that describes the event

- All three are categories of the clause
- All three are concerned with the event or situation that is reported by the utterance
Defining modality

Expressions with modal meanings (von Fintel 2006)

1. Modal auxiliaries
   Sandy must/should/might/may/could be home

2. Semimodal Verbs
   Sandy has to/ought to/needs to be home

3. Adverbs
   Perhaps, Sandy is home

4. Nouns
   There is a slight possibility that Sandy is home

5. Adjectives
   It is far from necessary that Sandy is home

6. Conditionals
   If the light is on, Sandy is home
Defining modality

Distribution of modal cues in different text types

<table>
<thead>
<tr>
<th></th>
<th>Hyland (Biology)</th>
<th>Holmes (gen. academic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical verbs</td>
<td>27.4%</td>
<td>35.9%</td>
</tr>
<tr>
<td>Adverbials</td>
<td>24.7%</td>
<td>12.8%</td>
</tr>
<tr>
<td>Adjectives</td>
<td>22.1%</td>
<td>6.6%</td>
</tr>
<tr>
<td>Modal verbs</td>
<td>19.4%</td>
<td>36.8%</td>
</tr>
<tr>
<td>Nouns</td>
<td>6.4%</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

(Table from Thompson et al. (2008) Categorising modality in biomedical texts. Proceedings of LREC 2008, page 27.)
Modality categories (Palmer 2001)

**Propositional modality**: speaker’s judgement of the truth value or factual status of the proposition

- **Epistemic**: speakers express judgement about the factual status of the proposition
  - Speculative: express uncertainty
    - John may be in his office
  - Deductive: indicate an inference from observable evidence
    - John must be in his office, the lights are on
  - Assumptive: indicate inference from what is generally known
    - John’ll be in his office, he is always there at this time

- **Evidential**: speakers indicate the evidence they have about the factual status of the proposition
  - Reported
  - Sensory
Defining modality

**Modality categories** (Palmer 2001)

**Event modality:** speaker’s attitude towards a potential future event that has not taken place

- **Deontic:** relates to obligation or permission
  - Permissive: *John can come in now*
  - Obligative: *John must come in now*
  - Commissive: *John promises to come back*

- **Dynamic:** relates to ability or willingness
  - Abilitive: *John can speak French*
  - Volitive: *John will do it for you*
Defining modality

Types of modal meaning (von Fintel 2006)

- **Epistemic modality** concerns what is possible or necessary given what is known and what the available evidence is.

- **Deontic modality** concerns what is possible, necessary, permissible, or obligatory, given a body of law or a set of moral principles or the like.

- **Bouletic modality** concerns what is possible or necessary, given a person’s desires.

- **Circumstantial modality** concerns what is possible or necessary, given a particular set of circumstances.

- **Teleological modality** concerns what means are possible or necessary for achieving a particular goal.
'Have' - ambiguity of modality triggers (von Fintel 2006)

1. It has to be raining.
   [after observing people coming inside with wet umbrellas; epistemic modality]

2. Visitors have to leave by six pm.
   [hospital regulations; deontic]

3. You have to go to bed in ten minutes.
   [stern father; bouletic]

4. I have to sneeze.
   [given the current state of one's nose; circumstantial]

5. To get home in time, you have to take a taxi.
   [teleleological]
Another classification (Portner 2009) depending on at what level the modal meaning is expressed

- **Sentential modality**: at the level of the sentence
  Modal auxiliaries, sentential adverbs
- **Sub-sentential modality**: at the level of constituents smaller than a full clause
  Within the predicate, modifying a noun phrase, verbal mood
- **Discourse modality**: any contribution of modality to meaning in discourse
  Any modal meaning that it is not part of sentential truth conditions
Defining modality

**Sentential modality** (Portner 2009)

- **Modal auxiliaries and modal verbs**: must, can, might, should, ...
- **Modal adverbs**: ought, need (to)
- **Generics, habituals and individual level predicates**:
  - G: A dog is a wonderful animal
  - H: Ben drinks chocolate milk
  - ILP: Noah is smart

- **Tense and aspect**: future, use of past to express “unreality”, progressive, perfect
  
  Even if Mary stayed until tomorrow, I’d be sad

- **Conditionals**: if ... then constructions

- **Covert modality**: it seems that no overt material in the sentence expresses modal meaning
  
  Ben knows how to solve the problem = ‘Tim knows how he can solve the problem’
Defining modality

Sub-sentential modality (Portner 2009)

- Modal adjectives and nouns: possible, necessary, certain, possibility, ...
- Propositional attitude verbs and adjectives: believe, hope, know, remember, certain, pleased, ...
- Verbal mood: indicative, subjunctive
- Infinitives
- Dependent modals
  I’d be surprised if David should win
- Negative polarity items: words and phrases that must be licensed by another element
  David will *(not) ever leave
Defining modality

**Discourse modality** (Portner 2009)

- **Evidentiality**: a speaker’s assessment of her grounds for saying something
- **Clause types**: declarative, interrogative, and imperative sentences
- **Performativity of sentential modals**
- **Modality in discourse semantics, modal subordination**: pragmatic phenomenon in which one sentence involving (sentential) modality affects the interpretations of subsequent modal sentences
  - John might go to the store. He should buy some fruit
  - Meaning of second sentence: ‘If he goes to the store, he should buy some fruit’.

R. Morante (CLiPS - University of Antwerp )

Modality and Negation in NLP

November 8, 2011 23 / 448
Defining modality

Defining modality

(Baker et al. 2010)

- “Modality might be construed broadly to include several types of attitudes that a speaker might have toward an event or state”.
- Modality might indicate:
  - **Factivity** is related to whether an event, state, or proposition happened or didn’t happen.
    It distinguishes things that happened from things that are desired, planned, or probable.
  - **Evidentiality** deals with the scope of information and may provide clues to the reliability of the information.
    Did the speaker have first hand knowledge of what he or she is reporting or was it inferred from indirect evidence?
  - **Sentiment** deals with a speaker’s positive or negative feelings toward an event, state, or proposition’.
Evidentiality (Aikhenvald 2003)

“In a number of languages, the nature of the evidence on which a statement is based must be specified for every statement - whether the speaker saw it, or heard it, or inferred from indirect evidence, or learnt it from someone else. This grammatical category, referring to an information source, is called ‘evidentiality’.”
Evidentiality (von Fintel 2006)

“Various languages regularly add markers, inflectional or otherwise, to sentences that indicate the nature of the evidence that the speaker has for the prejacent proposition.

A typical evidential system might centrally distinguish between direct evidence and indirect evidence.”

“The standard European languages do not have elaborate evidential systems but find other ways of expressing evidentiality when needed”

Kim has apparently been offered a new job
Related concepts: Hedging

Lakoff’s (1972) hedges
“Words whose job it is to make things more or less fuzzy”.

Hyland’s (1998) hedging
- “Linguistic devices used to qualify a speaker’s confidence in the truth of a proposition, the kind of caveats like I think, perhaps, might and maybe which we routinely add to out statements to avoid commitment to categorial assertions.”
- Any linguistics means used to indicate either
  - a) a lack of complete commitment to the truth value of an accompanying proposition, or
  - b) a desire not to express that commitment categorically.”
- “Hedging is one part of epistemic modality; it indicates an unwillingness to make an explicit and complete commitment to the truth of propositions”.

R. Morante (CLiPS - University of Antwerp)
Related concepts: Hedging

Hyland categories of surface realizations of hedging in scientific articles

Lexical
- Modal auxiliaries: may, might, could, would, should
- Epistemic judgment verbs: suggest, indicate, speculate, believe, assume
- Epistemic evidential verbs: appear, seem
- Epistemic deductive verbs: conclude, infer, deduce
- Epistemic adjectives: likely, probable, possible
- Epistemic adverbs: probably, possibly, perhaps, generally
- Epistemic nouns: possibility, suggestion
Hyland categories of surface realizations of hedging in scientific articles

Non-lexical features

- Reference to limiting experimental conditions, reference to a model or theory or admission to a lack of knowledge.
- Their surface realizations typically go beyond words and even phrases.
  - Whereas much attention has focused on elucidating basic mechanisms governing axon development, relatively little is known about the genetic programs required for the establishment of dendrite arborization patterns that are hallmarks of distinct neuronal types.
Hedge examples from Medlock and Briscoe (2007)

- DI and Ser have been proposed to act redundantly in the sensory bristle lineage
- How endocytosis of DI leads to the activation of N remains to be elucidated
- A second important question is whether the roX genes have the same, overlapping or complementing functions
- To test whether the reported sea urchin sequences represent a true RAG1-like match, we repeated the BLASTP search against all GenBank proteins
- This hypothesis is supported by our finding that both pupariation rate and survival are affected by EL9
Hedge instances as defined in Medlock and Briscoe (2007)

- Speculative question
  A second important question is whether the roX genes have the same, overlapping or complementing functions

- Statement of speculative hypothesis
  To test whether the reported sea urchin sequences represent a true RAG1-like match, we repeated the BLASTP search against all GenBank proteins

- Anaphoric hedge reference
  This hypothesis is supported by our finding that both pupariation rate and survival are affected by EL9
Related concepts: Hedging

What is not hedging? (Medlock and Briscoe 2007)

- Indication of experimentally observed nonuniversal behaviour
  Proteins with single BIR domains can also have functions in cell cycle regulation and cytokinesis

- Confident assertion based on external work
  Two distinct E3 ubiquitin ligases have been shown to regulate DI signaling in Drosophila melanogaster

- Statement of existence of proposed alternatives
  Different models have been proposed to explain how endocytosis of the ligand, which removes the ligand from the cell surface, results in N receptor activation
Experimentally-supported confirmation of previous speculation
Here we show that the hemocytes are the main regulator of adenosine in the Drosophila larva, as was speculated previously for mammals

Negation of previous hedge
Although the adgf-a mutation leads to larval or pupal death, we have shown that this is not due to the adenosine or deoxyadenosine simply blocking cellular proliferation or survival, as the experiments in vitro would suggest
Related concepts: Factuality

Factuality (Saurí and Pustejovsky 2009)

“Information conveying whether events mentioned in text correspond to real situations in the world or, instead, to situations of uncertain status.”

“The level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse”
Related concepts: Factuality

(Saurí and Pustejovsky 2009)

- Events in language are couched in terms of a continuum that ranges from truly factual to counterfactual.

- Depending on the polarity, events are then depicted as either facts or counterfacts.

  Five U.N. inspection teams visited a total of nine other sites.
  The size of the contingent was not disclosed.

- Depending on the level of uncertainty combined with polarity, events will be presented as possibly factual or possibly counterfactual.

  United States may extend its naval quarantine to Jordan’s Red Sea port of Aqaba.
  They may not have enthused him for their particular brand of political idealism.
Linguistic means of expressing factuality (Saurí and Pustejovsky 2009)

- **Polarity particles** express the positive or negative factuality of events mentioned in text (*no, not*)
- **Modality particles** contribute different degrees of certainty to a given event
- **Event-selecting predicates**: predicates that select for an argument denoting an event of some sort
  - They project factuality information on the event denoted by its argument through syntactic means: *claim, suggest, promise, offer, avoid, try, delay, think*

The Human Rights Committee regretted that discrimination against women persisted in practice
Linguistic means of expressing factuality (Saurí and Pustejovsky 2009)

- **Syntactic constructions**: some syntactic constructions involving subordination introduce factuality information of some sort.
  - The embedded event is presupposed as holding as fact
    Rice, who became secretary of state two months ago today, took stock of a period of tumultuous change
  - The embedded event is presented as underspecified with respect to its factuality status
    The environmental commission has adopted regulations to ensure that people are not exposed to radioactive waste
Related concepts: Factuality

Linguistic means of expressing factuality (Saurí and Pustejovsky 2009)

- Discourse structure: Some events may first have their factual status characterized in one way, but then be presented differently in a subsequent sentence.

Yesterday, the police denied that [drug dealers were tipped off before the operation]. However, it emerged last night that [a reporter from London Weekend Television unwittingly tipped off residents about the raid] when he phoned contacts on the estate to ask if there had been a raid—before it had actually happened.
Defining modality: Subjectivity

- Term introduced by Banfield (1982)

Wiebe et al 2004

“Subjectivity 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.”
Defining modality: Subjectivity


- Main types of subjectivity
  - Evaluation: emotions, evaluations, judgements, opinions
  - Speculation: “anything that removes the presuppositions of events occurring or states holding, such as speculation and uncertainty”.

- Many expressions are not subjective in all contexts. A subjective element is an instance of a potential subjective element, in a particular context, that is subjective in that context.

- A subjective element expresses the subjectivity of a source.

- There can be multiple subjective elements in a sentence, of different types and attributed to different sources and targets.

- Subjective elements might be complex expressions.

- Syntactic or morphological devices may also be subjective elements.
Rubin et al. 2005

“Certainty is viewed as a type of subjective information available in texts and a form of epistemic modality expressed through explicitly-coded linguistic means.

- Such devices as subjectivity expressions, epistemic comments, evidentials, reporting verbs, attitudinal adverbials, hedges, shields, approximators, understatements, tentatives, intensifiers, emphatics, boosters, and assertives, often overlap in their definitions, classifications, and lexical representations in English.

- They explicitly signal presence of certainty information that covers a full continuum of writer’s confidence, ranging from uncertain possibility and withholding full commitment to statements.”
1 Defining modality
   ● Related concepts

2 Defining negation

3 Why is it interesting to process modality and negation?

4 References
Defining negation

Lawler (2007)

“Negation is a linguistic, cognitive, and intellectual phenomenon. Ubiquitous and richly diverse in its manifestations, it is fundamentally important to all human thought. As Horn and Kato 2000 put it:

“Negative utterances are a core feature of every system of human communication and of no system of animal communication. Negation and its correlates – truth-values, false messages, contradiction, and irony – can thus be seen as defining characteristics of the human species.” (p.1)"
In natural language, negation functions as an operator along with quantifiers and modals.

Operators have a scope: elements to which negative, modals and quantifiers refer are in the scope of the negative operator.

Negation interacts with other operators (modals, quantifiers) in complex ways.

- Ambiguous cases:
  - Every boy didn’t leave

- Idiosyncratic combination with modals:
  - Deontic *may not*: You may not go (‘not possible’)
  - Epistemic *may not*: This may not be the place (‘possibly not’)

R. Morante (CLiPS - University of Antwerp)
Modality and Negation in NLP
November 8, 2011
Negation has been studied from a philosophical perspective since Aristotle (2500 years ago!)

It has been studied in terms of truth values: how does the truth value of a sentence change if we add a negative element?

Aristotle distinctions:
Defining negation: Philosophical tradition

Contradictory negation
Found in pairs such as:
   Socrates is sitting
   Socrates is not sitting
If one is true the other is necessarily false.
   *Socrates is neither sitting nor not sitting
The following truth values apply:

<table>
<thead>
<tr>
<th>p</th>
<th>( \neg p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
</tr>
</tbody>
</table>

Contrary negation
Found in pairs such as:
   Socrates is a good man
   Socrates is a bad man
Only one sentence can be true at any point in time and both sentences can be false at the same time.
   Socrates is neither a good man nor a bad man
Defining negation: Philosophical tradition

Law of Contradiction
A statement cannot be true and false at the same time
\[ \neg \exists x \ (Px \land \neg Px) \]
- Applies to contrary and contradictory negation

Law of the Excluded Middle
A statement must be either true or false
\[ \forall x \ (Px \lor \neg Px) \]
- Applies to contradictory negation
Defining negation: Philosophical tradition

Aristotle studies also negation combined with quantifiers

1. Every man is white
2. Some men are white
3. No man is white
4. Not every man is white

The pairs (2, 3) and (1, 4) are contradictory. LC and LEM apply
The pair (1, 3) is a contrary pair to which the LC applies
These oppositions are represented in the Square of Oppositions
Defining negation: Philosophical tradition

**Square of Oppositions** by Apuleius and Boethius

![Square of Oppositions Diagram]

- **AI vs. EO:** affirmative vs. negative opposition
- **AE vs. IO:** universal vs. particular

Partee (2007)

- Negation ¬ is a unary (or monadic) sentential operator.
- Monadic means it has just one argument, unlike ∨ and ∧, which are binary.
- Sentential means that its argument must be a formula, an expression of type t.
- Semantically it is a function of type t → t: It maps 1 onto 0 and 0 onto 1.
- In predicate logic and propositional logic, it is assumed that every formula gets a truth value (relative to a model and an assignment).
Defining negation: Grammar


Negation

The presence of a negative in a sentence or constituent, or the addition of such an element, or the effect of such an element when present.

Negative

1. A grammatical element which, when added to a sentence expressing a proposition, reverses the truth value of that proposition. [...] A negative element is an operator which takes some part of its sentence as its scope; that scope may be the entire proposition [...] or only some part of it [...]

R. Morante (CLiPS - University of Antwerp )
Modality and Negation in NLP
November 8, 2011 52 / 448

**Negative concord**

The phenomenon by which the presence of an overt negative requires other elements in the sentence to be marked as negative.

Sp. No he visto nada
Eng. I didn’t see anything
Defining negation: Grammar

**Negative polarity item**

Any of various items which can only occur within the scope of a negative and possibly also in certain other specified grammatical circumstances, notably in questions.

We don’t have **any** wine

Do we have **any** wine?

**any, anyone, anything, anywhere, ever, at all; give a damn, lift a finger, move a muscle, pay the slightest attention**
Defining negation: Traditional negation problems

Truth value and presuppositions

1. The King of France is not bald

   - If France is a monarchy, the proposition is either true or false
     Presupposition: there is a King of France
     Internal negation

   - If France is a republic, does the proposition have a truth value? It can continue as:
     1. The King of France is not bald, because there is no King of France

     The presupposition that there is a King of France is cancelled

     External negation

According to Horn (1985) this is case of pragmatic ambiguity and metalinguistic negation
Defining negation: Traditional negation problems

Scope of negation and quantifiers

1. All the boys did not leave
   - Interpretation 1: not boys at all left, but some did
   - Interpretation 2: all the boys stayed

Use of negative polarity items: use of *some* versus *any*

Neg-raising

1. I don’t think he is here
2. I think that he is not here
Defining negation: Types

Clausal negation (Tottie 1991)

- **Denials**: The audio system on this television is not very good, but the picture is amazing.
- **Rejections**: one participant rejects an offer or suggestion of another. Appear in expository text where a writer explicitly rejects a previous supposition or expectation. Given the poor reputation of the manufacturer, I expected to be disappointed with the device. This was not the case.
- **Imperatives**: directing an audience away from a particular action. Do not neglect to order their delicious garlic bread.
- **Questions**: Why couldn’t they include a decent speaker in this phone?
- **Supports** and **repetitions**: express agreement and add emphasis or clarity.
Defining negation: Types

Sentential versus inter-sentential negation

Intersentential negation

The language used in one sentence may explicitly negate a proposition or implication found in another sentence:

Rejections and supports

Sentential negation

Negations within the scope of a single sentence:

Sentential denials, imperatives, and questions
Clausal versus constituent negation (Payne 1997)

- Clausal negations negate an entire proposition
  - *I don’t have books*

- Constituent negation is associated with particular constituents or clauses
  - *I have no books*

- The effect of clausal and constituent negation can be very similar or identical, but constituent negation is less common as a grammatical device

- Most languages possess more than one type of clausal negation. The functional difference has to do with:
  - Negation of existence
  - Negation of fact
  - Negation of different aspects, modes or speech acts
Defining negation: Types of clausal negation (Payne 1997)

**Lexical negation**

“Describes a situation in which the concept of negation is part and parcel of the lexical semantics of a particular verb.”

- **Lack** as the lexical negative of *have*

**Morphological negation**

Morphemes that express clausal negation are associated with the verb

**Analytic negation**

Negative particles are normally associated with the main verb of the clause

- Negative particles: *n’t, not, never*
- Finite negative verbs (not in English)
Defining negation: Types of clausal negation (Payne 1997)

**Derivational negation**

“Languages will allow a stem to change into its “opposite” by use of some derivational morphology.

- Prefixes: unhappy, non-smoker
- Suffixes: motionless

**Negative quantifiers**

“Many languages employ quantifiers that are either inherently negative (none, nothing) or are negated independently of clausal negation (not many).”
Defining negation: Negation atlas

http://wals.info

Chapter 112: negative morphemes by Matthew S. Dryer

- Shows the nature of morphemes signalling clausal negation in declarative sentences
- All of the ways of indicating negation involve negative morphemes
- There are no known instances of languages in which negation is realized by a change in word order or by intonation
- All languages have negative morphemes
- Both negative particles and negative affixes are widely distributed throughout the world
Defining negation: Negation atlas

neg. affix, neg. particle, neg. aux. verb, neg. word, neg. wordaffix, double neg.

http://wals.info/feature/112
Defining negation: Negation vs negative polarity

- **Negation**: grammatical phenomenon used to state that some event, situation, or state of affairs does not hold.

- **Polarity**: a relation between semantic opposites.
  
  > “polarity encompasses not just the logical relation between negative and affirmative propositions, but also the conceptual relations defining contrary pairs like *hot–cold*, *long–short*, and *good–bad*” (Israel 2004).

The relation between negation and polarity lies in the fact that negation can reverse the polarity of an expression.

In the context of sentiment analysis positive and negative polarity is used in the sense of positive and negative opinions, emotions, and evaluations.
Outline

1 Defining modality
   - Related concepts

2 Defining negation

3 Why is it interesting to process modality and negation?

4 References
Why is it interesting to process modality and negation?

- Some NLP applications aim at extracting factual information from texts.
- As Prabhakaran et al. (2010) put it: “There is more to “meaning than” just propositional content”:
  1. GM will lay off workers
  2. A spokesman for GM said GM will lay off workers
  3. GM may lay off workers
  4. The politician claimed that GM will lay off workers
  5. Some wish GM would lay off workers
  6. Will GM lay off workers?
  7. Many wonder if GM will lay off workers

Examples from Prabhakaran et al. (2010)
Why is it interesting to process modality?

(Saurí and Pustejovsky 2009)

- **Opinion Mining**: the same situation can be presented as a fact in the world, a mere possibility, or a counterfact according to different sources.

- **Textual Entailment**:
  - Factuality-related information has been taken as a basic feature in some systems using the data from PASCAL RTE challenges (Tatu and Moldovan 2005, de Marneffe et al. 2006, and Snow and Vanderwende 2006).
  - The system that obtained the best absolute result in the three RTE challenges, scoring an 80% accuracy (Hickl and Bensley 2007), is based on identifying the set of publicly-expressed beliefs of the author.
Why is it interesting to process modality and negation?

**Textual Entailment** Dagan et al. include negation and modality as an aspect of the logical structure:

- **Factivity**: Uncovering the context in which a verb phrase is embedded
  - The terrorists tried to enter the building.
  - The terrorists entered the building.
- **Polarity**: negative markers or a negation-denoting verb (e.g. *deny*, *refuse*, *fail*)
  - The terrorists failed to enter the building.
  - The terrorists entered the building.
- **Modality/Negation**: Dealing with modal auxiliary verbs (can, must, should), that modify verbs’ meanings and with the identification of the scope of negation.
- **Superlatives/Comparatives/Monotonicity**: inflecting adjectives or adverbs.
- **Quantifiers, determiners and articles**
Why is it interesting to process modality and negation?

Summarization

- Fiszman et al. (2006) report that the majority of the system errors were due to two phenomena: missed negation and complicated sentence structure.

  ▶ Example of missed negation:
  
  Selegiline was found **unable to** inhibit deamination of beta-PEA.

  ▶ System output:
  
  **Selegiline INTERACTS**_WITH_ Phenethylamine

Why is it interesting to process modality and negation?

Information Extraction

The atovaquone/proguanil combination has not been widely used yet in West Africa so it is unlikely that the patient was initially infected with an atovaquone-resistant strain.

- Extracted information that falls under the scope of a negation signal cannot be presented as factual information (Vincze et al. 2008)
- More than 13% of the sentences in the BioScope corpus contain negation signals (Szarvas et al. 2008)
Why is it interesting to process modality and negation?

**Information Extraction**

- Not being able to recognize negation can hinder automated indexing systems (Mutalik et al. 1991)
- Approximately half of the conditions indexed in dictated reports are negated (Chapman et al. 2001)
- Negation status was the most important feature for classifying patients based on whether they had an acute lower respiratory syndrome; including negation status contributed significantly to classification accuracy (Chu et al. 2006)
Why is it interesting to process modality and negation?

Medlock 2008

“it is clear that interactive bioinformation systems that take account of hedging can render a significantly more effective service to curators and researchers alike”

- 30% of sentences in the results and discussion sections of biomedical papers contain speculative assertions, and this figure increases to around 40% for the conclusions section (Mercer and Marco 2004) (Medlock 2008)

- a significant part of the gene names mentioned (638 occurrences out of a total of 1968) appears in a speculative sentence. This means that approximately 1 in every 3 genes should be excluded from the interaction detection process (Szarvas 2008)
Why is it interesting to process modality and negation?

Biomedical information extraction

Light 2004

“The scientific process involves making hypotheses, gathering evidence, using inductive reasoning to reach a conclusion based on the data, and then making new hypotheses. Scientists are often not completely certain of a conclusion. This lack of definite belief is often reflected in the way scientists discuss their work.”

(Light 2004)

- 11% of sentences in MEDLINE contain speculative language.
- Extracting tables of protein-protein interactions would benefit from knowing which interactions were speculative and which were definite.
- In the context of knowledge discovery (KR), current speculative statements about a topic of interest can be used as a seed for the automated knowledge discovery process.
Why is it interesting to process modality and negation?

Examples from Light (2004)

1. Pdcd4 may thus constitute a useful molecular target for cancer prevention. (1131400)

2. On the basis of these complementary results, it has been concluded that curcumin shows very high binding to BSA, probably at the hydrophobic cavities inside the protein. (12870844)

3. Removal of the carboxy terminus enables ERP to interact with a variety of ets-binding sites including the E74 site, the IgH enhancer pisite, and the lck promoter ets site, suggesting a carboxy-terminal negative regulatory domain. (7909357)

4. Results suggest that one of the mechanisms of curcumin inhibition of prostate cancer may be via inhibition of Akt. (12682902)

5. To date, we find that the signaling pathway triggered by each type of insult is distinct. (10556169)
Why is it interesting to process modality and negation?

Biomedical information extraction

- Biomedical information extraction focuses on identifying biomedical entities and their relations
- Biomedical information retrieval focuses on finding documents that are relevant for specific database curation tasks

“However, the fact that a gene is mentioned, and even information about it is provided, does not necessarily imply that the information is reliable or useful in satisfying the scientist’s information need (Shatkay et al. 2008).”

“We believe that an important first step towards more accurate text-mining lies in the ability to identify and characterize text that satisfies various types of information needs.” (Wilbur et al. 2006)
1 Defining modality
   - Related concepts

2 Defining negation

3 Why is it interesting to process modality and negation?

4 References
References: Modality


References: Evidentiality


References: Hedging


References: Factuality


References: Subjectivity


References: Negation


Part II

Categorising and Annotating Modality and Negation
Outline

5 Annotation schemes

6 Existing resources

7 Future directions

8 References
Outline

5 Annotation schemes

6 Existing resources

7 Future directions

8 References
Annotation schemes: Modality in OntoSem


- **Framework:** **OntoSem project**
- **Text processing environment** that takes as input unrestricted raw text and carries out several levels of linguistic analysis, including modality at the semantic level
- **The output of the semantic analysis is represented as formal text-meaning representations (TMRs)**
Overall architecture of the OntoSem semantic analyzer

Modality information is encoded as part of the semantic module in the lexical entries of the modality cues.

Four modality attributes are encoded:

- **Modality type**
- **Scalar value** ranges from zero to one
- **Scope** attribute: the predicate that is affected by the modality
- **Attributed-to** attribute: indicates to whom is the modality assigned (default value = speaker)
Modality type

- **Polarity**, whether a proposition is positive or negated;
- **Volition**, the extent to which someone wants or does not want the event/state to occur;
- **Obligation**, the extent to which someone considers the event/state to be necessary;
- **Belief**, the extent to which someone believes the content of the proposition;
- **Potential**, the extent to which someone believes that the event/state is possible;
- **Permission**, the extent to which someone believes that the event/state is permitted;
- **Evaluative**, the extent to which someone believes the event/state is a good thing.
Examples

- **Polarity**: Reed refused to back down demanding the Republican led intelligence committee finish a long awaited report on whether the Bush administration twisted intelligence.

- **Volition**: He's trying to get Hamas to co-exist with Israel.

- **Obligation**: For payment, we have to forecast the money two days out.

- **Belief**: This week, the government arrested Jose Abello Silva, said to be the fourth-ranking cartel leader.

- **Potential**: If I can get all of the information today, I can tell you this afternoon.

- **Permission**: Flights are not permitted into Iraq.

- **Evaluative**: Ditches: They are better than road bumps because they are harder to see.
Values

- **Volition**: do not want 0 – really want 1
- **Obligation**: need not 0 – must 1
- **Belief**: do not believe it 0 – strongly believe it 1
- **Potential**: can’t achieve/be achieved 0 – can 1
- **Permission**: may not 0 – may 1
- **Evaluative**: evaluation really poor 0 – evaluation really highly 1

“Assigning values of modality to lexical items is a judgment call, not a science. No value that anyone assigns is set in stone. While some might argue that something so inherently inexact will not be helpful in text processing, we fervently disagree. The reason for using scalar values for modalities lies not in their absolute values but in their relative values. Whether disfavor is given the value (.3) or (<> .2 .3) or (<.4) on the scale of evaluative modality is less important than the fact that it has a much lower value than adore.” (Nirenburg and McShane 2008)
In the sentence

Entrance to the tower *should* be totally camouflage

*should* is identified as a modality cue and characterized with:

- Type obligative
- Value 0.8
- Scope *camouflage*
- and is attributed to the speaker
Modality entry in OntoSem

(certain-adj1
(cat adj)
(anno
(def "sure, convinced")
(ex "I am certain this is correct")
(comments "a certain winner is a bit different; I'm leaving it for now")))

(syn-struc
((np ((root $var1) (cat n)))
(root $var2) (cat v) (root be)
(adj ((root $var0) (cat adj)))
(xcomp ((root $var3) (cat v))))))

(sem-struc
(modality
(type belief)
(value 1)
(scope (value ^$var3))
(attributed-to (value ^$var1))
(^$var2 (null-sem +))))

From Nirenburg and McShane (2008)
Annotation schemes: Private States


- Context: sentiment analysis, opinion mining
- Annotation scheme that identifies key components and properties of opinions, emotions, sentiments, speculations, evaluations, and other private states
  - **Private states**: internal states that cannot be directly observed by others
- Goal: identifying private state expressions in context
Two types of frames to distinguish between opinion-oriented material and factual material

- Objective speech event frames that represent “material that is attributed to some source, but is presented as an objective fact”.
  - The source is the speaker or writer;
  - The target, what the private state is about;

- Private state frame for every expression of private state
  - The source of the private state, whose private state is being expressed;
  - The target, what the private state is about;
  - Properties like intensity, significance, and type of attitude (positive, negative, other, none).
Private States

Three **types of private state expressions** are considered for the annotation:

- **Explicit mentions**
  The U.S. fears a spill-over,” said Xirao-Nima

- **Speech events**
  Sargeant O’Leary said the incident took place at 2:00pm

- **Expressive subjective elements**,
  The report is full of absurdities,” Xirao-Nima said

Private states are expressed by the words and the style of language that is used
Nested sources: attribution of private states

“private states are often filtered through the “eyes” of another source, and private states are often directed toward the private states of others ” (Wiebe et. al 2005)

“The U.S. fears a spill-over,” said Xirao-Nima.

- According to the writer, according to Xirao-Nima, the U.S. fears a spill-over.
- Nested source of private state fears: [writer, Xirao-Nima, U.S.]
- The concept of source is very relevant for the annotation of modalities
Annotation at word/phrase level

Annotators were not limited to marking a type or list of words

Large variety of words appearing in subjective expressions

Many sentences are mixtures of subjectivity and objectivity:
44% of the sentences analysed are mixtures of two or more subjectivity intensity ratings or mixtures of subjectivity and objectivity
Private States

MPQA Opinion Corpus
http://www.cs.pitt.edu/mpqa/
- 10,657 sentences
- 535 documents of English newswire
- Annotated with information about private states at the word and phrase level.

- Discourse connectives and their arguments are assigned attribution-related features
- Goal: to capture the source and degrees of factuality of abstract objects
  - **SOURCE**: writer, other, arbitrary
  - **TYPE**: reflects the nature of the relation between the agent and the abstract object
  - **SCOPAL POLARITY** of attribution: identifies cases when verbs of attribution (say, think, ...) are negated syntactically (didn’t say) or lexically (denied).
  - **DETERMINACY**: indicates the presence of contexts canceling the entailment of attribution
Attribution in PDTB

**TYPE**: nature of the relation between the agent and the abstract object

- **Propositions**: “attribution to an agent of his/her (varying degrees of) commitment towards the truth of a proposition”
  - **Assertions**: identified by assertive predicates or verbs of communication
  - **Beliefs**: identified by “propositional attitude verbs” (believe, think, expect, suppose, imagine, etc.)
- **Facts**: “attribution to an agent of an evaluation towards or knowledge of a proposition whose truth is taken for granted (i.e., a presupposed proposition)”
- **Eventualities**: “attribution to an agent of an intention/attitude towards an eventuality”
SCOPAL POLARITY: the negation reverses the polarity of the attributed relation or argument content

- ***Null***: the neg-lowered interpretations are not present
- ***Neg***: the interpretation of the connective requires the surface negation to take semantic scope over the lower argument.

Example (Prasad et al. 2006): “Having the dividend increases is a supportive element in the market outlook, **but I don’t think it’s a main consideration,**” he says.

The polarity of Arg2 **I don’t think** is **Neg** because it scopes over the embedded clause **it’s a main consideration.**
Annotation schemes: ACE 2008


Automatic Content Extraction (ACE) 2008 corpus

- Goal: relation detection and recognition
- English and Arabic texts from several sources
- Relations are ordered pairs of entities annotated with modality and tense attributes
- Modality attributes
  - **Asserted**: relations pertain to situations in the real world
  - **Other**: relations pertain to situations in “some other world defined by counterfactual constraints elsewhere in the context”
- Example: We are afraid Al-Qaeda terrorists will be in Baghdad
  - ORG-Aff.Membership relation between terrorists and Al-Qaeda: asserted
  - Physical.Located relation between terrorists and Baghdad: other


- Uncertainty is understood as the speculative type of subjectivity
- Subjectivity: aspects of language used to express opinions and evaluations (Wiebe 1994)
- Certainty can also be seen as a variety of epistemic modality expressed through epistemic comments (*probably, perhaps*)
- Certainty is a pragmatic position rather than a grammatical feature
Certainty

View on certainty (Rubin et al 2005)
“Certainty is viewed as a type of subjective information available in texts and a form of epistemic modality expressed through explicitly-coded linguistic means”

Explicit markers of certainty
“explicitly signal presence of certainty information that covers a full continuum of writer’s confidence”

Devices
Subjectivity expressions, epistemic comments, evidentials, reporting verbs, attitudinal adverbials, hedges, shields, approximators, understatements, tentatives, intensifiers, emphatics, boosters, and assertives
Certainty identification (Rubin et al 2005)

“Certainty identification is defined as an automated process of extracting information from certainty-qualified texts or individual statements along four hypothesized dimensions of certainty”

- **Level**: degree of certainty
- **Perspective**: whose certainty is involved
- **Focus**: what the object of certainty is
- **Time**: what time the certainty is expressed
Certainty

Uncertainty model (Rubin 2006)

Explicit Certainty Categorization Model

D1: LEVEL
- Absolute Certainty (i.e., unambiguous or undisputable conviction, reassurance)
- High Certainty (i.e., high probability or firm knowledge)
- Moderate Certainty (i.e., average likelihood or reasonable chances)
- Low Certainty (i.e., distant possibility)
- Uncertainty (i.e., hesitancy, lack of knowledge or lack of clarity)

Primary Dimension

D2: PERSPECTIVE
- Writer’s Point of View
  - Reported
    - Participant’s Account (e.g., witnesses, victims, negotiating parties)
    - Expert’s View (e.g., analysts, authorities)

Contextual Dimensions

D3: FOCUS
- Opinions, Emotions, or Judgments (i.e., beliefs, attitudes, assessments, predictions)
- Facts and Events (or concrete states of affairs)

D4: TIME
- Past Time (i.e., completed, recent in the past)
- Present Time (i.e., immediate, incomplete, habitual)
- Future Time (i.e., predicted, scheduled)
- Irrelevant (or ambiguous)

(Images from Rubin 2010)
Certainty

Examples: certainty level

1. **Certain** An enduring lesson of the Reagan years, of course, is that it really does take smoke and mirrors to produce tax cuts, spending initiatives and a balanced budget at the same time.

2. **Less certain** So far the presidential candidates are more interested in talking about what a surplus might buy than in the painful choices that lie ahead.
Certainty

Examples: perspective

1 Writer More evenhanded coverage of the presidential race would help enhance the legitimacy of the eventual winner, which now appears likely to be Putin.

2 Reported The Dutch recruited settlers with an advertisement that promised to provide them with slaves who “would accomplish more work for their masters, ...”
Examples: focus

1 **Abstract information**: statements that reflect an idea that does not represent an external reality, but rather a hypothesized world
   In Iraq, the first steps must be taken to put a hard-won new security council resolution on arms inspections into effect.

2 **Factual information**: based on facts that have an actual existence in the world of events
   The settlement may not fully compensate survivors for the delay in justice, ...
Certainty

Data
- 32 articles from The New York Times
- 685 sentences, excluding headlines
- Sentence-level
- Average of 0.53 explicit certainty markers per sentence
- The distinction of focus into factual and abstract information presented the most difficulties for annotation
Factuality (Saurí and Pustejovský 2009)

“Information conveying whether events mentioned in text correspond to real situations in the world or, instead, to situations of uncertain status.”

“The level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse”
- **FactBank** A corpus of events annotated with factuality information
- Define a discrete set of factuality values and a battery of criteria that allow annotators to differentiate among these values
Difficulties in annotating factuality:

- “To find an expressive enough set of discrete factuality values that is grounded on linguistic intuitions but also supported by commonsense reasoning”
- “Factuality is expressed through a complex interaction of many different aspects of the overall linguistic expression”:
  - Polarity, epistemic modality, evidentiality, mood.
  - Component in the semantics of specific syntactic structures with presuppositional effects
  - Component in certain types of predicates (e.g. factive and implicative predicates)
Factuality

**Challenges** (Saurí and Pustejovsky 2009)

- Distinguishing among factuality degrees
- Interaction between factuality markers

1. The Royal Family will continue to allow detailed fire brigade inspections of their private quarters
2. The Royal Family will continue to refuse to allow detailed fire brigade inspections of their private quarters
3. The Royal Family may refuse to allow detailed fire brigade inspections of their private quarters

- Relevant sources: different discourse participants may present divergent views about the factuality nature of the very same event

4. Slobodan Milosevic’s son said Tuesday that the former Yugoslav president had been murdered at the detention center of the UN war crimes tribunal in The Hague
### Factuality values in Saurí and Pustejovsky (2009)

<table>
<thead>
<tr>
<th></th>
<th>Positive (+)</th>
<th>Negative (−)</th>
<th>Underspecified (u)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certain (ct)</td>
<td>Fact: (&lt;\text{ct},+&gt;)</td>
<td>Counterfact: (&lt;\text{ct},-&gt;)</td>
<td>Certain but unknown output: (&lt;\text{ct},u&gt;)</td>
</tr>
<tr>
<td>Probable (pr)</td>
<td>Probable: (&lt;\text{pr},+&gt;)</td>
<td>Not probable: (&lt;\text{pr},-&gt;)</td>
<td>(NA)</td>
</tr>
<tr>
<td>Possible (ps)</td>
<td>Possible: (&lt;\text{ps},+&gt;)</td>
<td>Not certain: (&lt;\text{ps},-&gt;)</td>
<td>(NA)</td>
</tr>
<tr>
<td>Underspecified (u)</td>
<td>(NA)</td>
<td>(NA)</td>
<td>Unknown or uncommitted: (&lt;u,u&gt;)</td>
</tr>
</tbody>
</table>
Specified Values:

- CT1 According to the source, it is certainly the case that X.
- PR1 According to the source, it is probably the case that X.
- PS1 According to the source, it is possibly the case that X.
- CT2 According to the source, it is certainly not the case that X.
- PR2 According to the source it is probably not the case that X.
- PS2 According to the source it is possibly not the case that X.

Underspecified Values:

- CTu The source knows whether it is the case that X or that not X.
- Uu The source does not know what is the factual status of the event, or does not commit to it.

Discriminatory tests are applied to discriminate between values
FactBank Data:

- 208 documents
- 9,488 manually annotated events
- 0.81 $K_{cohen}$ agreement
- FactBank as a second layer on top of TimeBank

Example

Newspaper reports have said Amir was infatuated with Har-Shefi and may have been trying to impress her by killing the prime minister.
Factuality

TimeBank annotation
Newspaper reports have

\[
\begin{align*}
&<\text{EVENT eid="e22" class="REPORTING" tense="PRESENT" aspect="PERFECTIVE"> said \</EVENT> \\
&\text{Amir was} \\
&<\text{EVENT eid="e23" class="STATE" tense="PAST"> infatuated \</EVENT> \\
&\text{with Har-Shefi and may have been} \\
&<\text{EVENT eid="e24" class="I\_ACTION" modality="may" tense="NONE" aspect="PERF\_PROG"> trying \</EVENT> \\
&\text{to} \\
&<\text{EVENT eid="e25" class="OCCURRENCE" tense="INFINITIVE"> impress \</EVENT> \\
&\text{her by} \\
&<\text{EVENT eid="e26" class="OCCURRENCE" tense="PRES\_PART" aspect="NONE"> killing \</EVENT> the prime minister. \\
&<\text{SLINK lid="l50" eventId="e22" subordinatedEventId="e23" relType="EVIDENTIAL"> \\
&<\text{SLINK lid="l51" eventId="e22" subordinatedEventId="e24" relType="EVIDENTIAL"> \\
&<\text{SLINK lid="l52" eventId="e24" subordinatedEventId="e25" relType="MODAL">}
\end{align*}
\]
FactBank annotation

<table>
<thead>
<tr>
<th>Event (ID)</th>
<th>Source (ID)</th>
<th>Fact. value</th>
</tr>
</thead>
<tbody>
<tr>
<td>said (e22)</td>
<td>author ($s_0$)</td>
<td>CT+</td>
</tr>
<tr>
<td>infatuated (e23)</td>
<td>reports_author ($s_2$,$s_0$)</td>
<td>CT+</td>
</tr>
<tr>
<td>trying (e24)</td>
<td>author ($s_0$)</td>
<td>Uu</td>
</tr>
<tr>
<td>impress (e25)</td>
<td>reports_author ($s_2$,$s_0$)</td>
<td>PS+</td>
</tr>
<tr>
<td>killing (e26)</td>
<td>author ($s_0$)</td>
<td>Uu</td>
</tr>
</tbody>
</table>
Annotation tasks by 2 annotators

- Identifying source-introducing predicates (SIP) (reporting, knowledge and opinion) \([0.88 \ k_{cohen}]\)
  - SIP contribute new source to the discourse

- Identifying sources \([0.95 \ k_{cohen}]\)
  - In mid-2001, Colin Powell_{source} and Condoleezza Rice_{source} both publically denied_{SIP} that Iraq had_{event} weapons of mass destruction

- Assigning factuality values \([0.81 \ k_{cohen}]\)

Goal

- Recognize what the writer of the text intends the reader to believe about various people’s beliefs about the world (including the writer’s own)

Assumption

- Discourse participants model each other’s cognitive state during discourse
  - They model cognitive states as beliefs, desires, and intentions
- Language provides cues for the discourse participants to do the modeling
Committed belief

Annotated categories (Diab et al. 2009)

- Each verbal proposition is annotated with the tags:
  - **Committed belief (CB)**: the writer indicates in this utterance that he or she believes the proposition
    - *We know that GM has laid off workers*
  - **Non-committed belief (NCB)**: the writer identifies the proposition as something which he or she could believe, but he or she happens not to have a strong belief in
    - *GM may lay off workers*
  - **Not applicable (NA)**: for the writer, the proposition is not of the type in which he or she is expressing a belief, or could express a belief
    - ★ Expressions of desire: *Some wish GM would lay off workers*
    - ★ Questions: *Will GM lay off workers?*
    - ★ Expressions of requirements: *GM is required to lay off workers*
Committed belief

Corpus (Diab et al. 2009)

- 10,000 words annotated for speaker belief of stated propositions.
- They annotate the writer’s beliefs
- Nested beliefs are excluded
- Annotation at proposition level
- Different domains and genres
Annotation schemes: Categorising modality


- **Focus**: epistemic modality in biomedical text
  - The expression of the author’s level of confidence towards a proposition
  - The type of knowledge, assumptions or evidence on which the proposition is based

- **Corpus**: 113 abstracts from MEDLINE, E. Coli, annotated with gene regulation events

- **Goal**: annotate a corpus with modality categories if the modality information is under the scope of a gene regulation event

- **Results**: 202 MEDLINE abstracts annotated, 1469 gene regulation events
Categorising modality

Dimensions of the categorisation scheme

Knowledge Type

“The type of “knowledge” that underlies a statement, encapsulating both whether the statement is a speculation or based on evidence and how the evidence is to be interpreted”

- Speculative: predict, hypothesis, view, in theory
- Deductive: interpret, indication, infer, imply
- Sensory: observation, see, appear
- Demonstrative: show, confirm, demonstrate

(Speculative, deductive and demonstrative based on Palmer’s model)
Categorising modality

Dimensions of the categorisation scheme

Level of certainty

“Indicating how certain the author (or cited author) is about the statement”

- Absolute: certainly, known
- High: likely, probably, generally
- Moderate: possibly, perhaps, may, could
- Low: unlikely, unknown
Categorising modality

Dimensions of the categorisation scheme

Point of View

“Indicating whether the statement is based on the author’s own or a cited point of view or experimental findings.”

- Writer: *we, our results*
- Other: citations
Categorising modality

We suggest that overproduction of SlyA in E. coli derepresses clyA transcription by counteracting H-NS.

(Figure from Thompson et al. (2008)
www.nactem.ac.uk/workshops/lrec08_ws/slides/Thompson_et_al.pdf)
Categorising modality

Distribution per category in the annotated corpus

- Knowledge Type and Certainty Level
  - Knowledge Type only
  - Certainty Level only
  - Knowledge Type + Certainty Level

- If values were assigned to Knowledge Type and/or Certainty Level dimensions, Point of View dimension also instantiated

(Figure from Thompson et al. (2008)
www.nactem.ac.uk/workshops/lrec08_ws/slides/Thompson_et_al.pdf)
Categorising modality

Distribution per subcategory in the annotated corpus

(Based on Thompson et al. (2008)
www.nactem.ac.uk/workshops/lrec08_ws/slides/Thompson_et_al.pdf)
Annotation schemes: A modality lexicon


- Goal: Exploring whether structured annotations of entities and modalities can improve translation output in the face of sparse training data
- Focus on modal words that are related (H = holder; P = proposition) to factivity
  - Requirement: does H require P?
  - Permissive: does H allow P?
  - Success: does H succeed in P?
  - Effort: does H try to do P?
  - Intention: does H intend P?
  - Ability: can H do P?
  - Want: does H want P?
  - Belief: with what strength does H believe P?
A modality lexicon

Annotation scheme

- Three components for sentences that express modality
  - Trigger: word or string that expresses modality
  - Target: event, state or relation that the modality scopes over
  - Holder: experiencer or cognizer of the modality

- Modality can be expressed without a lexical trigger
Simplifications

- Scope of modality and negation. Same annotation for:
  - I do not believe that he left
  - I believe he didn’t leave

- Duality of meaning require and permit
  - not require P to be true = Permit P to be false
  - not permit P to be true = Require P to be false.

- Entailment between modalities. Annotators were provided a specificity-ordered modality list
  - requires → permits
  - succeeds → tries → intends → is able → wants

- Sentences without an overt trigger word are tagged as Firmly Believes

- Nested modalities are not marked, only one modality is marked

- The holder is not marked
A modality lexicon

Entry definition

1. A string of one or more words: for example, should or have need of
2. A part of speech for each word
3. A modality: one of the thirteen modalities
4. A head word (or trigger): the primary phrasal constituent to cover cases where an entry is a multiword unit, e.g., the word hope in hope for
5. One or more subcategorization codes
Example entries

### Able
- capable "JJ of "IN$Able & capable, JJ-of-basic, JJ-of-VBG
- able "JJ to "TO$Able & able, JJ-infinitive
- can "MD$Able & can, modal-auxiliary-basic
- could "MD$Able & could, modal-auxiliary-basic
- ready "JJ$Able & ready, JJ-infinitive

### NotAble
- powerless "JJ$NotAble & powerless, JJ-infinitive
- unable "JJ$NotAble & unable, JJ-infinitive
A modality lexicon

Modality tagging example

**Input:** He managed to hold general elections in the year 2002, but he can not be ignorant of the fact that the world at large did not accept these elections.

**Output:** He <TrigSucceed managed> to <TrigSucceed hold> general elections in the year 2002, but he <TrigAble can> <TrigNegation not> <TrigNOTAble be> ignorant of the fact that the world at large did <TrigNegation not> <TrigBelief accept> these <TrigBelief elections>.
Modality tagger

- A modality tagger produces text or structured text in which modality triggers and/or targets are identified

  **Tagger 1: string-based**
  - Input: text with PoS
  - Marks spans of words/phrases that exactly match modality trigger words in the modality lexicon
  - It identifies the target by tagging the next non-auxiliary verb to the right of the trigger

  **Tagger 2: structure-based**
  - Input: parsed text
  - TSurgeon patterns are automatically generated from the verb class codes in the modality lexicon along with a set of templates
  - The patterns are matched with part of a parse tree
A modality lexicon

Output of modality tagger

(TOP
  (S
    (NP
      (NNP Pakistan)
      (SBAR (WDT which)
        (S (MD TrigAble could)
          (RB TrigNegation not)
          (VB B TargAble TrigSucceed
            TargNegation reach)
          (ADJP
            (JJ TargSucceed semi-final))
          , ,
          (PP (IN in) (DT a)
            (NN match) (PP (IN against)
              (ADJP (JJ South) (JJ African))
              (NN team))
            (PP (IN for) (DT the)
              (JJ fifth) (NN position))
            (NP (NNP Pakistan))))))
  (VB D defeated))

- Sentence level annotation of speculation
  - 6 papers from the functional genomics literature

- They define what is and what is not hedging
  - Guidelines to be found in
    http://www.cl.cam.ac.uk/research/nl/nl-download/hedging.html

- Sentences are classified into speculative or non-speculative
  - **Spec**: This unusual substrate specificity may explain why Dronc is resistant to inhibition by the pan-caspase inhibitor p.

- 380 out of 1,157 sentences are speculative
Annotation schemes: Focus of Negation


- Based on distinction between scope and focus of negation (Huddleston and Pullum 2002)
  - Scope is the part of the meaning that is negated
  - Focus is that part of the scope that is most prominently or explicitly negated
- Focus of negation annotated on 3,993 verbal negations signaled as MNEG in PropBank
- For each instance, annotators decide the focus given the full syntactic tree, as well as the previous and next sentence
- Inter-annotator agreement was 0.72
Annotation examples (Table from Blanco and Moldovan 2011)

<table>
<thead>
<tr>
<th>Statement</th>
<th>V</th>
<th>A0</th>
<th>A1</th>
<th>A2</th>
<th>A4</th>
<th>TMP</th>
<th>MNR</th>
<th>ADV</th>
<th>LOC</th>
<th>PNC</th>
<th>EXT</th>
<th>DIS</th>
<th>MOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Even if [that deal]$<em>{A1}$ isn’t [revived]$</em>{V}$, NBC hopes to find another.</td>
<td>*</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>– Even if that deal is suppressed, NBC hopes to find another one.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2  [He]$<em>{A0}$ [simply]$</em>{MDIS}$ [ca]$<em>{MMOD}$n’t [stomach]$</em>{V}$ [the taste of Heinz]$_{A1}$, she says.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– He simply can stomach any ketchup but Heinz’s.</td>
<td>+</td>
<td>+</td>
<td>*</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3  [A decision]$<em>{A1}$ isn’t [expected]$</em>{V}$ [until some time next year]$_{MTMP}$.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– A decision is expected at some time next year.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4  […] it told the SEC [it]$<em>{A0}$ [could]$</em>{MMOD}$n’t [provide]$<em>{V}$ [financial statements]$</em>{A1}$ [by the end of its first extension]$<em>{MTMP}$ “[without unreasonable burden or expense]$</em>{MMNR}$”.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>– It could provide them by that time with a huge overhead.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
http://www.inf.u-szeged.hu/rgai/bioscope

Corpus annotated with negation and speculation cues and their scopes in English biomedical texts
### Table 1: Statistics of the three subcorpora

<table>
<thead>
<tr>
<th></th>
<th>Clinical</th>
<th>Full Paper</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Documents</td>
<td>1954</td>
<td>9</td>
<td>1273</td>
</tr>
<tr>
<td>#Sentences</td>
<td>6383</td>
<td>2670</td>
<td>11871</td>
</tr>
<tr>
<td>Negation sentences</td>
<td>13.55%</td>
<td>12.70%</td>
<td>13.45%</td>
</tr>
<tr>
<td>#Negation cues</td>
<td>877</td>
<td>389</td>
<td>1848</td>
</tr>
<tr>
<td>Hedge sentences</td>
<td>13.39%</td>
<td>19.44%</td>
<td>17.70%</td>
</tr>
<tr>
<td>#Hedge cues</td>
<td>1189</td>
<td>714</td>
<td>2769</td>
</tr>
</tbody>
</table>
Example

When U937 cells were infected with HIV-1, <xcope id="X1.6.3"><cure type="negation" ref="X1.6.3">no</cure> induction of NF-KB factor was detected</xcope>, whereas high level of progeny virions was produced, <xcope id="X1.6.2"><cure type="speculation" ref="X1.6.2">suggesting</cure> that this factor was <xcope id="X1.6.1"><cure type="negation" ref="X1.6.1">not</cure> required for viral replication</xcope>.</xcope>
Hedges modify the factuality of a statement or reflect the author’s attitude towards the content of the text.

Categories:

- **Auxiliaries**: may, might, can, would, should, could, etc.
- **Verbs of hedging or verbs with speculative content**: suggest, question, presume, suspect, indicate, suppose, seem, appear, favor, etc.
- **Adjectives or adverbs**: probable, likely, possible, unsure, etc.
- **Conjunctions**: or, andor, either ... or, etc.
- **Complex keywords**: Mild bladder wall thickening raises the question of cystitis.
Annotation strategy

- Marking the keywords: the minimal unit that expresses hedging and determines the actual strength of hedging was marked as a keyword.
- Marking scope: all constituents that fell within the uncertain interpretation were included in the scope
  - Motivation: disregarding the marked text span, the rest of the sentence can be used for extracting factual information. In:

    Mild bladder wall thickening raises the question of cystitis.

    Mild bladder wall thickening is a fact
    Cystitis is an uncertain fact
**Scope and syntax:** the scope of a speculative element can be determined on the basis of syntax.

- Verbs, auxiliaries, adjectives and adverbs usually start right with the keyword.
  - Verbal elements: verbs and auxiliaries, it ends at the end of the clause or sentence, all complements and adjuncts are included.
  - Attributive adjectives: scope generally extends to the following noun phrase
  - Predicative adjectives: scope includes the whole sentence.

- Sentential adverbs have a scope over the entire sentence.
- The scope of other adverbs usually ends at the end of the clause or sentence.

- Conjunctions generally have a scope over the syntactic unit whose members they coordinate.
Annotation schemes: Scopes in ConanDoyle-neg

Same corpus as SemEval Task on Linking Events and Their Participants in Discourse (Ruppenhofer et al. 2010)

- Already annotated with semantic roles, coreference and null instantiations of semantic roles
- Different domain than BioScope
- Not subject to copyright
- Linear narrative
- **But** older variety of English

- HB: 14 chapters (2700 sentences)
- WL: 2 chapters (600 sentences)

- **Negation cues**: words that express negation
- **Scope of negation cues**: tokens in the sentence that are affected by the negation
- **Negated event or property**
Scopes in ConanDoyle-neg

The most of them would by no means advance, but three of them, the boldest, or it may be the most drunken, rode forward down the goyal. [HB 2-59]
The annotation format is based on the BioScope format, but there are differences:

- Different scope model
  - All participants of the event that is negated fall under the scope of the negation cue
  - Discontinuous scope is allowed
- Affixal negation is annotated
- Negated events are annotated
Not all negations cues negate a fact
” Do you not find it interesting? ” [HB 2.74]
Not all negations cues negate a fact

Had the prosaic finding of the coroner not finally put an end to the romantic stories which have been whispered in connection with the affair, it might have been difficult to find a tenant for Baskerville Hall. [HB 2.113]
Not all negations cues negate a fact

For both these reasons I thought that I was justified in telling rather less than I knew, since no practical good could result from it, but with you there is no reason why I should not be perfectly frank. [HB 2.127]
The annotation guidelines are published soon as a CLiPS Technical Report: 
http://www.clips.ua.ac.be/
annotation-of-negation-cues-and-their-scope-guidelines-v10
Annotation schemes: High utility text


The contents of scientific statements can be characterized along certain general dimensions.

In turn, the characteristics of each phrase, sentence or paragraph along these dimensions can help to determine whether the text is useful to a particular user with specific information needs.

Different users have different information needs.
A **database curator** who is looking for experimental evidence that the gene was expressed under certain conditions, would only be satisfied with sentences discussing experimental evidence and stating with high confidence that the gene was indeed expressed under the reported conditions.

A **scientist** looking for all the information published about a certain gene, may be satisfied by obtaining all the papers or all the sentences mentioning this gene.
Goal

- Enabling the creation of well-focused subsets of biomedical text that have certain properties
- Identifying information-bearing fragments within scientific text
  - Reducing the document search space for a specific domain in order to improve retrieval and extraction
    * Identifying regions that are rich in experimental evidence and methodological details
    * Focusing extraction efforts on these regions
  - Providing users with candidate sentences that:
    * Describe the desired phenomenon
    * Bear the evidence for the phenomenon or describe the methods by which the phenomenon was identified
- Differentiate informative fragments from non-informative ones automatically
Annotation scheme

- 10,000 sentences selected at random from both full-text articles and abstracts
- Each statement in the corpus is characterized and marked-up along 5 dimensions
- A statement may be a sentence or just a fragment of a sentence
Dimensions

Focus

The type of the information conveyed by the statement

- **Scientific (S):** discussing findings and discovery
- **Generic (G):** general state of knowledge and science outside the scope of the paper, the structure of the paper itself or the state of the world
- **Methodology (M):** describing a procedure or a method

annotate methodology when the sentence under annotation contains an indication that methodology is being discussed

- Not every sentence appearing in a Methodology section discusses methodology, and not every sentence discussing methodology appears in the Methodology section
A fragment with any focus can be stated either positively (P) or negatively (N).

Each fragment conveys a degree of certainty about the validity of the assertion it makes.

- (0) represents complete uncertainty, that is, the fragment explicitly states that there is an uncertainty or lack of knowledge about a particular phenomenon ("it is unknown..." or "it is unclear whether..." etc.).
- (1) represents a low certainty
- (2) is assigned to high-likelihood expressions that are still short of complete certainty.
- (3) represents complete certainty, reflecting an accepted, known and/or proven fact.
Dimensions

Evidence

This dimension indicates for any fragment if its assertion is supported by evidence.

- E0: No indication of evidence in the fragment whatsoever, or an explicit statement in the text indicates lack of evidence.
- E1: A claim of evidence, but no verifying information is explicitly given. "Previous experiments show that...," followed by the fragment, "therefore, it is likely that ...".
- E2: Evidence is not given within the sentence/fragment, but explicit reference is made to other papers (citations) to support the assertion.
- E3: Evidence is provided, within the fragment, in one of the following forms:
  - A reference to experiments previously reported within the body of the paper (Our results show that ...)
  - A verb within the statement indicates an observation or an experimental finding (We found that ...)
  - A reference to an experimental figure or a table of data given within the paper.
Dimensions

Direction-trend

The signs + or - indicate respectively whether the assertion reports a qualitatively high or low level or an increase/ decrease in a specific phenomenon, finding or activity.

"In fact, as demonstrated using several SOD assays including pulse radiolysis, 2-ME does not inhibit SOD"

Negative polarity, negative trend (inhibit)
Examples

The binding of both forms of $\beta$-catenin to CBP is completely inhibited by ICG-001 (Fig. 3B Top, lane 4).

We demonstrate that ICG-001 binds specifically to CBP but not the related transcriptional coactivator p300.

A statement may be supported by several types of evidence:

...the overexpression of phospho-H2Av did not induce G2/M arrest or affect DSB-dependent G2/M arrest (fig. S10) (14,21),
Data

- 10000 sentences annotated by 3 annotators
- For experiments, sentences in which the three annotators agree
- Each dimension is examined separately

**Table 1.** The number of sentences and of fragments, for which there is complete agreement in annotation, along each dimension

<table>
<thead>
<tr>
<th>Dataset name</th>
<th>Foc. &amp; Ev.</th>
<th>Focus</th>
<th>Evidence</th>
<th>Certainty</th>
<th>Polarity</th>
<th>Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences</td>
<td>Frag_FE</td>
<td>1977</td>
<td>4068</td>
<td>2964</td>
<td>5644</td>
<td>6430</td>
</tr>
<tr>
<td>Fragments</td>
<td>Frag_F</td>
<td>4068</td>
<td>2964</td>
<td>5644</td>
<td>6430</td>
<td>5907</td>
</tr>
<tr>
<td>No. of terms</td>
<td>Frag_E</td>
<td>2109</td>
<td>4447</td>
<td>3133</td>
<td>5992</td>
<td>6945</td>
</tr>
<tr>
<td>selected</td>
<td>Frag_C</td>
<td>2109</td>
<td>4447</td>
<td>3133</td>
<td>5992</td>
<td>6330</td>
</tr>
<tr>
<td></td>
<td>Frag_P</td>
<td>2109</td>
<td>4447</td>
<td>3133</td>
<td>5992</td>
<td>6945</td>
</tr>
<tr>
<td></td>
<td>Frag_T</td>
<td>2109</td>
<td>4447</td>
<td>3133</td>
<td>5992</td>
<td>6330</td>
</tr>
</tbody>
</table>

(From Shatkay et al. 2008)
BiographTA
Text Analytics in the Biograph Project

Processing modality and negation for machine reading

QA4MRE
CLEF 2011
Amsterdam

Brief description
This is a pilot task of the Machine Reading Evaluation QA4MRE at CLEF 2011.
**Goal**: evaluating whether machine reading systems understand extra-propositional aspects of meaning beyond propositional content, focusing mostly on phenomena related to modality and negation.

**Background collections**: same as main QA4MRE task

**Test sets**: 12 texts from The Economist, 4 per topic (climate change, AIDS, music and society)

- Two pilot test documents were released first

**Questions** for each document there are ten multiple choice questions

- 5 candidate answers
- 1 clearly correct answer

**Evaluation**: same as main task
## Modality and negation for MR

### Table: Test documents from The Economist

<table>
<thead>
<tr>
<th>Topic</th>
<th>Number</th>
<th>Title</th>
<th># of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aids</td>
<td>1</td>
<td>All colours of the brainbow</td>
<td>915</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>DARC continent</td>
<td>817</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Double, not quits</td>
<td>779</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Win some, lose some</td>
<td>1919</td>
</tr>
<tr>
<td>Climate change</td>
<td>1</td>
<td>A record-making effort</td>
<td>2841</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Are economists erring on climate change?</td>
<td>1412</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Climate change and evolution</td>
<td>1256</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Climate change in black and white</td>
<td>2850</td>
</tr>
<tr>
<td>Music and society</td>
<td>1</td>
<td>The politics of hip-hop</td>
<td>1004</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>How to sink pirates</td>
<td>773</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Singing a different tune</td>
<td>1042</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Turn that noise off</td>
<td>677</td>
</tr>
</tbody>
</table>
Event Description
Given a multiple choice question, systems have to choose the answer that best characterises an event along five aspects of meaning:

- Negation
- Perspective
- Certainty
- Modality
- Condition for another event or conditioned by another event
Negative

1. A grammatical element which, when added to a sentence expressing a proposition, reverses the truth value of that proposition. [...] A negative element is an operator which takes some part of its sentence as its scope; (R.L. Trask (1993) A dictionary of grammatical terms in linguistics. Routledge.)

1. But these new types of climate action do not REPLACE the need to reduce carbon emissions.

2. 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.
A statement is presented from the point of view of someone. By default the statement is presented from the perspective of the author of the text, but the author might be mentioning the view from someone else.

The European Union has named a dozen prefectures that need radiation tests, yet traders in these places report a LACK of testing equipment.
Certainty

Events can be presented with a range of certainty values, including underspecified certainty. Here we include all not certain events under the category of *uncertain* events, without distinguishing degrees.

1. Even though external radiation has since returned to near-harmless levels, Mr Sakurai *fears many of Minamisoma’s evacuees may never COME BACK.*

2. As well as having charms that efforts to reduce carbon-dioxide emissions lack, *these alternatives could also IMPROVE the content and prospects of other climate action.*
Modality

Five options:

- **Non-modal event.** This is the default category for events that do not fall under the modal categories below and do not have other modal meanings. In the questions we refer to it as event.

- **Purpose event.** Purpose, aim or goal.
  
  Neighbouring South Korea expressed concern that it was not warned about TEPCO’s decision to dump low-level radioactive waste into the sea to **MAKE** room to store more toxic stuff on land.

- **Need event.** Need or requirement.

  The plan **requires** a lot of **INVESTMENT** in power generation and **smarter grids**, best done in the context of –at long last– reformed and competitive energy market.
**Obligation.**

Believing that global greenhouse-gas emissions must **fall** by half to limit climate change, and that rich countries should **cut** the most, Europe has set a goal of reducing emissions by 80-95% by 2050.

**Desire.** Desires, intentions and plans.

Neighbouring South Korea expressed concern that it was not warned about TEPCO’s **decision** to **dump** low-level radioactive waste into the sea to make room to store more toxic stuff on land.
Condition-conditioned

An event can be presented as a condition for another event or as conditioned by another event.

1. If you are highly motivated to minimise your taxes, you can HUNT for every possible deduction for which you’re eligible.
Event description

- An event description consists at most of one value per aspect of meaning.
- An event description consists at least of one modality value:
  - Event, purpose event, need event, obligation event, desire event
- If applicable, events can additionally be described with the following aspects of meaning that systems have to identify:
  - Negated
  - Perspective of someone other than the author
  - Uncertain
  - Condition for another event, conditioned by another event
Controlling black carbon by giving poor people cleaner ways to burn various fuels could not only forestall a decade or two of global warming, it would also save hundreds of thousands of lives currently blighted by smoke and disease.

Event - controlling black carbon by giving poor people cleaner ways to burn various fuels forestall a decade or two of global warming - is presented in the text as:

- MOD-NEED
- MOD-WANT
- COND-BY MOD-NON
- UNCERT MOD-NON
- NEG UNCERT MOD-NON
The combinations of codes that conform the answers to the questions can be summarized with the following regular expression:

$$[\text{COND}|\text{COND} - \text{BY}]? \text{NEG}? \text{PERS}? \text{UNCERT}? \text{MOD} [-\text{NEED}| - \text{NON}| - \text{PURP}| - \text{MUST}| - \text{WANT}]$$

In total there are 120 combinations, although not all of them will be represented in the test set of 12 documents because not all of them are equally frequent.
Outline

5 Annotation schemes

6 Existing resources

7 Future directions

8 References
Why do we need to have annotated resources?

- To have a better insight into the surface realization of negation and modality and their role in NLP
- To train systems that:
  - Detect non-factual information
  - Detect statements with negative polarity
  - Detect contrastive information
- This can be useful for several NLP applications:
  - Information extraction
  - Opinion mining, sentiment analysis
  - Paraphrasing
  - Recognizing textual entailment
  - Machine translation
Existing resources

Negated biomedical events

- **Genia Event** [http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/home/wiki.cgi?page=Event+Annotation](http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/home/wiki.cgi?page=Event+Annotation)
- **BioNLP Shared Task 2010** data [http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/](http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/)
- **GREC** [http://www.nactem.ac.uk/GREC/](http://www.nactem.ac.uk/GREC/)

Scopes

- **BioScope** [http://www.inf.u-szeged.hu/rgai/bioscope](http://www.inf.u-szeged.hu/rgai/bioscope)
- **CoNLL Shared Task 2010** data [http://www.inf.u-szeged.hu/rgai/conll2010st/](http://www.inf.u-szeged.hu/rgai/conll2010st/)
Existing resources

Meta-knowledge annotation including modality and negation

- **Meta-Knowledge Genia Corpus**
  http://www.nactem.ac.uk/meta-knowledge/

- **Statement Map corpus of Japanese**
  http://www.cl.ecei.tohoku.ac.jp/stmap/sem_corpus.html

Lexicons

- Modality lexicon described in Baker et al. (2010)
  http://www.umiacs.umd.edu/~bonnie/ModalityLexicon.txt
Existing resources

Negation and modality for machine reading

- Test set QA4MRE pilot task 2011

Factuality

- FactBank [http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009T23](http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009T23)

Subjectivity

- MPQA Opinion Corpus

Discourse

- PDTB [http://www.seas.upenn.edu/~pdtb/](http://www.seas.upenn.edu/~pdtb/)
Outline

5 Annotation schemes

6 Existing resources

7 Future directions

8 References
Future directions

- Creating a unified annotation scheme for modality?
- Merging existing annotations?
- Defining guidelines?
- Annotating fine-grained modality types?
- Annotating larger corpora, different genres
Outline

5 Annotation schemes

6 Existing resources

7 Future directions

8 References
References


Part III

Tasks Related to Processing Modality and Negation
Detecting speculated sentences

Processing negation in biomedical texts

Scope resolution

Finding negated and speculated events

Modality tagging

Belief categorisation

Processing contradiction and contrast

Visualising negation features

References
Detecting speculated sentences

Defining the task

Medlock and Briscoe (2007)

Given a collection of sentences, $S$, the task is to label each sentence as either speculative or nonspeculative. Specifically, $S$ is to be partitioned into two disjoint sets, one representing sentences that contain some form of hedging, and the other representing those that do not.
Detecting speculated sentences

- Light et al. (2004) used a handcrafted list of hedge cues to identify speculative sentences in MEDLINE abstracts.
- Medlock and Briscoe (2007) used single words as input features in order to classify sentences from biological articles (FlyBase) as speculative or non-speculative based on semi-automatically collected training examples.
- Szarvas (2008) extended the methodology of Medlock and Briscoe (2007) to use n-gram features and a semi-supervised selection of the keyword features.
- Kilicoglu and Bergler (2008) proposed a linguistically motivated approach based on syntactic information to semi-automatically refine a list of hedge cues.
- Ganter and Strube (2009) proposed an approach for the automatic detection of sentences containing uncertainty based on Wikipedia weasel tags and syntactic patterns.
Detecting speculated sentences


- **Corpus**: sentences marked as highly speculative, low speculative, or definite.
  - 1456 sentences
  - 173 speculative

- **Bag-of-words** representation of text sentences occurring in MEDLINE abstracts

- **Algorithm**: SVM\textsubscript{light}

- **Baseline**: checking whether any cue is present in the sentence
  - suggest, potential, likely, may, at least, in part, possible, potential, further investigation, unlikely, putative, insights, point toward, promise, propose

- **Accuracy results** for SVM = 92% vs. 89% baseline
Detecting speculated sentences


System

- Bag-of-words approach
- Semi-supervised learning: of labelled training data, from which a supervised classifier can subsequently be learned
- Test corpus: manually annotated 380 spec sentences and 1157 nspec sentences
  - Step 1: a weakly supervised Bayesian learning model in order to derive the probability of each word to represent a hedge cue
  - Step 2: feature selection based on these probabilities, only the most indicative features of the spec class are retained
  - Step 3: classifier trained on a given number of selected features
Detecting speculated sentences

Seed generation

- Seeds for the spec class: all sentences from U containing either (or both) of the terms suggest or likely. 6423 spec seeds
- Nspec seeds: 7541 sentences
  1. Create initial $S_{nspec}$ by sampling randomly from $U$.
  2. Manually remove more ‘obvious’ speculative sentences using pattern matching
  3. Iterate:
     - Order $S_{nspec}$ by $P(spec|x_j)$ using estimates from $S_{spec}$ and current $S_{nspec}$
     - Examine most probable sentences and remove speculative instances

Results

- Baseline: 0.60 recall/precision break-even point
- System results: 0.76 recall/precision break-even point
Detecting speculated sentences

Error analysis

- The model is unsuccessful in identifying assertive statements of knowledge paucity which are generally marked rather syntactically than lexically.

  There is no clear evidence for cytochrome c release during apoptosis in C elegans or Drosophila.

- Distinguishing between a speculative assertion and one relating to a pattern of observed non-universal behaviour is often difficult.

Sentence chosen as spec:

  Each component consists of a set of subcomponents that can be localized within a larger distributed neural system.

The sentence does not, in fact, contain a hedge but rather a statement of observed non-universal behaviour.
Detecting speculated sentences


- Same dataset as Medlock and Briscoe (2007)
- Experiments with additional features:
  - Part-of-speech tags
  - Stems
  - Bigrams: in some instances combinations of terms represent more reliable hedge cues than just single terms

SPEC: In addition several studies indicate that in mammals the Rel proteins could probably be involved in CNS processes such as neuronal development and synaptic plasticity

NSPEC: In the row marked dgqa the stippled exons indicate regions that are not found in the dgqa cDNAs identified by us

- Conclusions:
  - Adding PoS features and stems to a bag-of-words input representation can slightly improve the accuracy
  - Adding bigrams brings a statistically significant improvement over a simple bag-of-words representation
Detecting speculated sentences

Learning curves stemming

Figure from Medlock (2008)
Detecting speculated sentences

Learning curves bigrams

Figure from Medlock (2008)
### Detecting speculated sentences

#### Informative features

A table showing single term + bigram features ranked by $P(spec|x_k)$ with $\alpha = 5$:

<table>
<thead>
<tr>
<th>Rank</th>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>suggest</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>suggest that</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>might</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>may_be</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>possibl_that</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>might_be</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>appear_to</td>
<td>37</td>
</tr>
<tr>
<td>8</td>
<td>result_suggest</td>
<td>38</td>
</tr>
<tr>
<td>9</td>
<td>propos_that</td>
<td>39</td>
</tr>
<tr>
<td>10</td>
<td>is_like</td>
<td>40</td>
</tr>
<tr>
<td>11</td>
<td>thought_to</td>
<td>41</td>
</tr>
<tr>
<td>12</td>
<td>suggest_a</td>
<td>42</td>
</tr>
<tr>
<td>13</td>
<td>thi_suggest</td>
<td>43</td>
</tr>
<tr>
<td>14</td>
<td>seem_to</td>
<td>44</td>
</tr>
<tr>
<td>15</td>
<td>whether_the</td>
<td>45</td>
</tr>
<tr>
<td>16</td>
<td>whether</td>
<td>46</td>
</tr>
<tr>
<td>17</td>
<td>data_suggest</td>
<td>47</td>
</tr>
<tr>
<td>18</td>
<td>like_to</td>
<td>48</td>
</tr>
<tr>
<td>19</td>
<td>like_that</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>may_have</td>
<td>50</td>
</tr>
<tr>
<td>21</td>
<td>may_also</td>
<td>51</td>
</tr>
<tr>
<td>22</td>
<td>seem</td>
<td>52</td>
</tr>
<tr>
<td>23</td>
<td>may</td>
<td>53</td>
</tr>
<tr>
<td>24</td>
<td>the_posibl</td>
<td>54</td>
</tr>
<tr>
<td>25</td>
<td>thought</td>
<td>55</td>
</tr>
<tr>
<td>26</td>
<td>determin_whether</td>
<td>56</td>
</tr>
<tr>
<td>27</td>
<td>ar_like</td>
<td>57</td>
</tr>
<tr>
<td>28</td>
<td>is_posibl</td>
<td>58</td>
</tr>
<tr>
<td>29</td>
<td>is_thought</td>
<td>59</td>
</tr>
<tr>
<td>30</td>
<td>the_idea</td>
<td>60</td>
</tr>
</tbody>
</table>

*Figure from Medlock (2008)*
Error analysis

- 20% Statements of knowledge paucity
  This brings us to the largest of all mysteries, namely how the DCC is spread along the X chromosome
- Cases where speculativity is indicated by a particular term, while the general construction of the sentence does not fit the usual spec mold
  We then tested the putative RNA-binding property of MOF directly using electromobility shift assays
- Genuine hedge cues were not induced with enough certainty
  Invertebrates in vivo RAG-mediated transpositions are strongly suppressed, probably to minimize potential harm to genome function
Detecting speculated sentences


- Hedge detection in radiology records (newly annotated) and biomedical texts (dataset from Medlock anb Briscoe 2007)
- Complex feature selection
- Maximum Entropy Model
- Weakly supervised machine learning
Detecting speculated sentences

Feature selection

- Ranking the features $x$ by frequency and their class conditional probability $P(\text{spec}|x)$.
  - Select features with $P(\text{spec}|x) > 0.94$ and appeared in the training dataset with reasonable frequency
  - Result: 2407 candidates

- For trigrams, bigrams and unigrams, calculate a new class-conditional probability for each feature $x$, discarding those observations of $x$ in speculative instances where $x$ was not among the two highest ranked candidate.
  - Separately for the uni-, bi- and trigrams
  - Results: filtered out 85% of all the keyword candidates and kept 362 uni-, bi-, and trigrams altogether

- Re-evaluate all 362 candidates together and filtered out all phrases that had a shorter substring of themselves among the features, with a similar class-conditional probability on the speculative class
  - Results: discard 30% of the candidates and kept 253 uni-, bi-, and trigrams altogether
Detecting speculated sentences

Evaluation settings

- Automatic feature selection
- Manual feature selection
  - A phrase was irrelevant if they could consider no situation in which the phrase could be used to express hedging
  - 63 out of the 253 keywords found to be potentially relevant in hedge classification
- Adding external dictionaries
  - Keywords used in (Light et al., 2004) and those gathered for the author’s ICD-9-CM hedge detection module
  - Added only keywords found to be reliable enough by the maxent model trained on the training dataset
  - From 63 to 71 features
Detecting speculated sentences

Results

<table>
<thead>
<tr>
<th></th>
<th>Biomedical papers</th>
<th>Medical reports</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$BEP(spec)$</td>
<td>$F_{\beta=1}(spec)$</td>
</tr>
<tr>
<td>Baseline 1</td>
<td>60.00</td>
<td>–</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>76.30</td>
<td>–</td>
</tr>
<tr>
<td>All features</td>
<td>76.05</td>
<td>73.61</td>
</tr>
<tr>
<td>Feature selection</td>
<td>78.68</td>
<td>78.09</td>
</tr>
<tr>
<td>Manual feat. sel.</td>
<td>82.02</td>
<td>80.88</td>
</tr>
<tr>
<td>Outer dictionary</td>
<td>85.29</td>
<td>85.08</td>
</tr>
</tbody>
</table>

Baseline 1: substring matching of Light et al. (2004)
Baseline 2: Medlock and Briscoe (2007) system
Detecting speculated sentences

Conclusions

- The radiology reports had mainly unambiguous single-term hedge cues
- It proved to be useful to consider bi- and trigrams as hedge cues in scientific texts
- The hedge classification task reduces to a lookup for informative single keywords or phrases. Removing uninformative features did not produce any difference in the scores
- The analysis of errors indicate that more complex features like dependency structure and clausal phrase information could only help in allocating the scope of hedge cues detected in a sentence, not the detection of any itself
- Worse results on biomedical scientific papers from a different source showed that the portability of hedge classifiers is limited

  - The keywords *possible* and *likely* are apparently always used as speculative terms in the FlyBase articles, while the articles from BMC Bioinformatics frequently used such cliche phrases as *all possible combinations* or *less likely / more likely* ...
Detecting speculated sentences


- Linguistically motivated approach
- Lexical resources and syntactic patterns
Detecting speculated sentences

Data

- Fruit fly dataset by Medlock and Briscoe (2007)
  - Semiautomatically annotated
  - Noisy and biased towards the hedging cues used as seed terms (suggest, likely).
- Manually annotated data from the fruit fly dataset
  - 523 sentences training, 213 speculative
  - Balanced distribution of surface realization features: epistemic verbs (30%), adverbs (20%), adjectives (16%), modal verbs (23%)
- Manually annotated data from Szarvas (2008)
Detecting speculated sentences

Methodology

- Expansion of lexical hedging cues (190 entries)
  - Hyland cues
  - Synonyms from WordNet
  - Nominalizations from UMLS

- Quantification of hedging strength
  - Semi-automatic weighting depending on the type of cue and how it was obtained (SA)
  - Information gain weighting schemes: hedging cues that occur frequently in the speculative sentences but never in non-speculative sentences will have a higher IG weight
  - Accumulate the weights of the hedging cues found in a sentence to assign an overall hedging score to each sentence
Detecting speculated sentences

Methodology: syntax

- Identification of the most salient syntactic patterns in the train corpus that play a role in hedging and their contribution to hedging strength

<table>
<thead>
<tr>
<th>Syntactic Pattern</th>
<th>Effect on strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;EPISTEMIC VERB&gt; to(inf) VB</td>
<td>+1</td>
</tr>
<tr>
<td>&lt;EPISTEMIC VERB&gt; that(comp) VB</td>
<td>+2</td>
</tr>
<tr>
<td>Otherwise</td>
<td>-1</td>
</tr>
<tr>
<td>&lt;EPISTEMIC NOUN&gt; followed by that(comp)</td>
<td>+2</td>
</tr>
<tr>
<td>Otherwise</td>
<td>-1</td>
</tr>
<tr>
<td>not &lt;UNHEDGING VERB&gt;</td>
<td>+1</td>
</tr>
<tr>
<td>no</td>
<td>not &lt;UNHEDGING NOUN&gt;</td>
</tr>
<tr>
<td>no</td>
<td>not immediately followed by &lt;UNHEDGING ADVERB&gt;</td>
</tr>
<tr>
<td>no</td>
<td>not immediately followed by &lt;UNHEDGING ADJECTIVE&gt;</td>
</tr>
<tr>
<td>whether</td>
<td>if in a clausal complement context</td>
</tr>
<tr>
<td></td>
<td>1.58(IG)</td>
</tr>
</tbody>
</table>

(Table from Kiligoglu and Bergler 2008)
Detecting speculated sentences

Results

Table 9: Recall/precision break-even point (BEP) results

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall/Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline1</td>
<td>0.60</td>
</tr>
<tr>
<td>baseline2</td>
<td>0.76</td>
</tr>
<tr>
<td>Our system on the fruit-fly dataset with SA weighting</td>
<td>0.85</td>
</tr>
<tr>
<td>Our system on the fruit-fly dataset with IG weighting</td>
<td>0.80</td>
</tr>
<tr>
<td>Our system on the BMC dataset with SA weighting</td>
<td>0.82</td>
</tr>
<tr>
<td>Our system on the BMC dataset with IG weighting</td>
<td>0.70</td>
</tr>
</tbody>
</table>

(Table from Kiligoglu and Bergler 2008)

- Baseline 2: substring matching, with the top 15 ranked term features reported in reported in Medlock and Briscoe (2007)
Conclusions

- The SA weighting scheme gives better results: “a weighting scheme relying on the particular semantic properties of the indicators is likely to capture the hedging strengths more accurately”
- SA weighting provides relatively stable results across datasets
- A larger training set will yield a more accurate weighting scheme based on IG measure
- The IG weighting scheme is less portable
Detecting speculated sentences

**Error analysis: false negatives**

- Syntactic patterns not addressed by the method
  - Negation of “unhedgers” was used as a syntactic pattern; while this pattern correctly identified *know* as an “unhedger”, it did not recognize *little* as a negative quantifier

  Little was known however about the specific role of the roX RNAs during the formation of the DCC

- Certain derivational forms of epistemic words
  - The adjective *suggestive* is not recognized as a hedging cue, even though its base form *suggest* is an epistemic verb

  Phenotypic differences are suggestive of distinct functions for some of these genes in regulating dendrite arborization

- Incorrect dependency relations
Detecting speculated sentences

Error analysis: false positives

• Word sense ambiguity of hedging cues
  Also we *could* not find any RAG-like sequences in the recently sequenced sea urchin lancelet hydra and sea anemone genomes, which encode RAG-like sequences

• “Weak” hedging cues, such as epistemic deductive verbs (*conclude, estimate*) as well as some adverbs (*essentially, usually*) and nominalizations (*implication, assumption*)
Detecting speculated sentences


Detecting speculative language in Wikipedia

- Wikipedia as a source of training data for hedge classification
- Adopt Wikipedia’s notion of weasel words: “Some people say”, “I think”, “Clearly“, “is widely regarded as”, “it has been said/suggested/noticed”, “It may be that”
Detecting speculated sentences

http://en.wikipedia.org/wiki/Weasel_word
Detecting speculated sentences

Data

- Several Wikipedia dumps from the years 2006 to 2008
- Only those articles that contained the string "{{weasel".
- 168,923 unique sentences containing 437 weasel tags

Datasets

- Development: dump completed on July 14, 2008
- Test: dump completed on March 6, 2009
  - Created a balanced test set choosing one random, non-tagged sentence per tagged sentence
- 246 manually annotated sentences for evaluation
Detecting speculated sentences

**Features**

- Words preceding the weasel tags. Each word within these 5-grams receives an individual score, based on:
  - The relative frequency of this word in weasel contexts and the corpus in general
  - The average distance the word has to a weasel tag, if found in a weasel context

- Shallow linguistic features: three types of syntactic patterns:
  - Numerically underspecified subjects ("Some people", "Experts", "Many")
  - Passive constructions ("It is believed", "It is considered")
  - Adverbs ("Often", "Probably")
Results

<table>
<thead>
<tr>
<th></th>
<th>( \sigma )</th>
<th>.60</th>
<th>.70</th>
<th>.76</th>
<th>.80</th>
<th>.90</th>
<th>.98</th>
</tr>
</thead>
<tbody>
<tr>
<td>balanced set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( wpw )</td>
<td></td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.69</td>
<td>.69</td>
<td>.70</td>
</tr>
<tr>
<td>( asp )</td>
<td></td>
<td>.67</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.61</td>
<td>.59</td>
</tr>
<tr>
<td>manual annot.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( wpw )</td>
<td></td>
<td>.68</td>
<td>.69</td>
<td>.69</td>
<td>.69</td>
<td>.70</td>
<td>.65</td>
</tr>
<tr>
<td>( asp )</td>
<td></td>
<td>.67</td>
<td>.69</td>
<td>.68</td>
<td>.68</td>
<td>.61</td>
<td>.59</td>
</tr>
</tbody>
</table>

- The syntactic patterns do not contribute to the regeneration of weasel tags.
- The decreasing precision of both approaches when trained on more tagged sentences might be caused by the great number of unannotated weasel words.
- The difference between \( wpw \) and \( asp \) becomes more distinct when the manually annotated data form the test set.
  - The added syntactic patterns indeed manage to detect weasels that have not yet been tagged.
CoNLL-2010 Shared Task
Learning to detect hedges and their scope in natural language

Task 1  Learning to detect sentences containing uncertainty: identify sentences in texts which contain unreliable or uncertain information

- Task1B: Biological abstracts and full articles
- Task1W: Wikipedia paragraphs

Task 2  Learning to resolve the in-sentence scope of hedge cues: in-sentence scope resolvers have to be developed

- Biological abstracts and full articles

### Results Task 1

<table>
<thead>
<tr>
<th>Name</th>
<th>P / R / F</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Georgescul</td>
<td>72.0 / 51.7 / 60.2</td>
<td>C</td>
</tr>
<tr>
<td>Ji</td>
<td>62.7 / 55.3 / 58.7</td>
<td>X</td>
</tr>
<tr>
<td>Chen</td>
<td>68.0 / 49.7 / 57.4</td>
<td>C</td>
</tr>
<tr>
<td>Morante</td>
<td>80.6 / 44.5 / 57.3</td>
<td>C</td>
</tr>
<tr>
<td>Zhang</td>
<td>76.6 / 44.4 / 56.2</td>
<td>C</td>
</tr>
<tr>
<td>Zheng</td>
<td>76.3 / 43.6 / 55.5</td>
<td>C</td>
</tr>
<tr>
<td>Täckström</td>
<td>78.3 / 42.8 / 55.4</td>
<td>C</td>
</tr>
<tr>
<td>Mamani Sánchez</td>
<td>68.3 / 46.2 / 55.1</td>
<td>C</td>
</tr>
<tr>
<td>Tang</td>
<td>82.3 / 41.4 / 55.0</td>
<td>C</td>
</tr>
<tr>
<td>Kilicoglu</td>
<td>67.9 / 46.0 / 54.9</td>
<td>O</td>
</tr>
<tr>
<td>Tjong Kim Sang</td>
<td>74.0 / 43.0 / 54.4</td>
<td>C</td>
</tr>
<tr>
<td>Clausen</td>
<td>75.1 / 42.0 / 53.9</td>
<td>C</td>
</tr>
<tr>
<td>Özgür</td>
<td>59.4 / 47.9 / 53.1</td>
<td>C</td>
</tr>
<tr>
<td>Zhou</td>
<td>85.3 / 36.5 / 51.1</td>
<td>C</td>
</tr>
<tr>
<td>Li</td>
<td>88.4 / 31.9 / 46.9</td>
<td>C</td>
</tr>
<tr>
<td>Prabhakaran</td>
<td>88.0 / 28.4 / 43.0</td>
<td>C</td>
</tr>
<tr>
<td>Ji</td>
<td>94.2 / 6.6 / 12.3</td>
<td>C</td>
</tr>
</tbody>
</table>

**Table 1**: Task1 Wikipedia results (type ∈ {Closed(C), Cross(X), Open(O)}).
Classifying Wikipedia sentences as uncertain - best system


- Motivation: test whether a list of cues suffices for automatic hedge detection
- System based on SVM parameter tuning
- Features: lexical information, i.e. features extracted from the list of hedge cues provided with the training corpus.
Detecting speculated sentences. CoNLL ST’10

Baseline classifying as “uncertain” any sentence that contains any of the multi-word expressions labeled as hedge cues in the training corpus.

- Small percentage of false negatives on the BioScope test data: only a small percentage of “uncertain” sentences in the reference test dataset do not contain a hedge cue that occurs in the training dataset.
- Precision of baseline algorithm has values under 0.5 on all four datasets: ambiguous hedge cues are frequently used in “certain” sentences.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#sentences</th>
<th>%uncertain sentences</th>
<th>#distinct cues</th>
<th>#ambiguous cues</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia training</td>
<td>11111</td>
<td>22%</td>
<td>1912</td>
<td></td>
<td>0.32</td>
<td>0.96</td>
<td>0.48</td>
</tr>
<tr>
<td>Wikipedia test</td>
<td>9634</td>
<td>23%</td>
<td>-</td>
<td>188</td>
<td>0.45</td>
<td>0.86</td>
<td>0.59</td>
</tr>
<tr>
<td>BioScope training</td>
<td>14541</td>
<td>18%</td>
<td>168</td>
<td></td>
<td>0.46</td>
<td>0.99</td>
<td>0.63</td>
</tr>
<tr>
<td>BioScope test</td>
<td>5003</td>
<td>16%</td>
<td>-</td>
<td>96</td>
<td>0.42</td>
<td>0.98</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Table 1: The percentage of “uncertain” sentences (% uncertain sentences) given the total number of available sentences (#sentences) together with the number of distinct cues in the training corpus and the performance of the baseline algorithm based on the list of cues extracted from the training corpus.
System characteristics (Wikipedia)

- **Features**
  - Frequency of each hedge cue provided with the training corpus in each sentence
  - 2-grams and 3-grams extracted from the list of hedge cues provided with the training corpus

- **SVM**
  - Gaussian Radial Basis kernel function
  - Width of the RBF kernel = $\gamma = 0.0625$
  - Regularization parameter $C = 10$
Results with best parameters

<table>
<thead>
<tr>
<th>Dataset</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wikipedia training</td>
<td>1899</td>
<td>1586</td>
<td>585</td>
<td>0.5449</td>
<td>0.7644</td>
<td>0.6362</td>
<td>49.1 seconds</td>
</tr>
<tr>
<td>Wikipedia test</td>
<td>1213</td>
<td>471</td>
<td>1021</td>
<td>0.7203</td>
<td>0.5429</td>
<td>0.6191</td>
<td>21.5 seconds</td>
</tr>
<tr>
<td>BioScope training</td>
<td>2508</td>
<td>515</td>
<td>112</td>
<td>0.8296</td>
<td>0.9572</td>
<td>0.8888</td>
<td>19.5 seconds</td>
</tr>
<tr>
<td>BioScope test</td>
<td>719</td>
<td>322</td>
<td>71</td>
<td>0.6907</td>
<td>0.9101</td>
<td>0.7854</td>
<td>2.6 seconds</td>
</tr>
</tbody>
</table>

Table 2: The performance of our system corresponding to the best parameter values. The performance is denoted in terms of true positives (TP), false positives (FP), false negatives (FN), precision (P), recall (R) and F-score (F).
Optimized results for biomedical train corpus

<table>
<thead>
<tr>
<th>Dataset content</th>
<th>#sentences used for training</th>
<th>#sentences used for test</th>
<th>SVM</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Abstracts only</td>
<td>9871</td>
<td>2000</td>
<td>0.85</td>
<td>0.94</td>
</tr>
<tr>
<td>Full articles only</td>
<td>2170</td>
<td>500</td>
<td>0.72</td>
<td>0.87</td>
</tr>
<tr>
<td>Abstracts and full articles</td>
<td>11541</td>
<td>3000</td>
<td>0.81</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 4: Performances when considering separately the dataset containing abstracts only and the dataset containing articles from BioScope corpus. The SVM classifier was trained with gamma = 1 and c=10. Approximately 80% of the CoNLL train corpus was used for training and 20% of the train corpus was held out for testing.

- Learning curves show that the system is more efficient on abstracts than on full articles
- On test data the results are lower
Outline

9 Detecting speculated sentences
10 Processing negation in biomedical texts
11 Scope resolution
12 Finding negated and speculated events
13 Modality tagging
14 Belief categorisation
15 Processing contradiction and contrast
16 Visualising negation features
17 References
Related work

- NegEx (Chapman et al. 2001) uses a regular expression algorithm that implements phrases indicating negation in discharge summaries.
- NegFinder (Mutalik et al. 2001) uses rules to recognise negated patterns occurring in medical narrative.
- Elkin et al. (2005) apply a grammar that assigns to each concept an attribute (positive/negative/uncertain assertion).
- Boytcheva et al. (2005) use negation rules based on regular expressions to mark negated phrases (Bulgarian).
Determining whether a finding, disease or concept is negated

  207 sentences from hospital reports
  Naïve Bayes, Decision Trees 90 F1

- Averbuch et al. (2004)
  Algorithm that uses information gain to learn negative context patterns
  7 medical terms
  97.47 F1

- Huang and Lowe (2007) develop a hybrid system that combines regular expression matching with parsing in order to locate negated concepts
Negfinder is a rule-based system that recognizes a large set of negated patterns occurring in medical narrative.

Described in:

Research Paper

Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents:

A Quantitative Study Using the UMLS

Pradeep G. Mutalik, MD, Aniruddha Deshpande, MD, Prakash M. Nadkarni, MD
**Motivation**

To increase the utility of concept indexing of medical documents, it is necessary to record whether the concepts have been negated or not.

Medical personnel are trained to include pertinent negatives in their reports.

- Databases need to be searched to find relevant information for clinical and research purposes.
- Documents pertaining to a specific domain may also be concept indexed.
- Phrases in the document are identified and matched to concepts in a domain-specific thesaurus.
- For a medical document, however, the presence of a concept does not necessarily make the document relevant for that concept.
  - The concept may refer to a finding that was looked for but found to be absent or that occurred in the remote past.
Motivation

- In medical narrative, negations are direct and straightforward, since clinicians are trained to convey the salient features of a case concisely and unambiguously.
- Hypothesis: negations in dictated medical narrative are unlikely to cross sentence boundaries and are also likely to be simple in structure.
  - Simple syntactic methods to identify negations might therefore be reasonably successful.
Components

1. **Concept-finding**: identifies UMLS concepts
2. **Input transformation**: replace every instance of a concept or compound concept in the original document with the UMLS ID
3. **Lexing/parsing step**:
   - **Lexer**: identifies a very large number of negation signals and classifies them on the basis of properties such as whether they generally precede or succeed the concept they negate and whether they can negate multiple concepts
   - **Parser**: applies its grammar rules to associate the negation signal with a single concept or with multiple concepts preceding or succeeding it
4. **Verification step**: marks up the original document by color-coding the text to assist human validation of the program’s output
Processing negation in biomedical texts: Negfinder

*DIAGNOSES: Pneumonia. The patient has a history of peptic ulcer disease with a Billroth I operation as well as vomiting... He denies shortness of breath, chest pain, fever, chills, nausea, vomiting, diarrhea, or abdominal pain. The patient had no other complaints... *PHYSICAL EXAMINATION... JVP estimated at 6cm. 2+ carotid pulsations bilaterally. No bruits... *CARDIAC: Irregularly irregular, with normal S1, split S2, with no murmurs, rubs, or gallops... He was without complaint and had no evidence of cough or sputum production. ...Sputum and blood cultures were negative for evidence of infection.


(From Mutalik et al. (2001) Negation Detection to Augment Concept Indexing, JAMIA 2001 8: 598-609.)
*DIAGNOSES: **Pneumonia.** The patient has a history of **peptic ulcer disease** with a Billroth I operation as well as **vomiting**... He **denies shortness of breath, chest pain, fever, chills, nausea, vomiting, diarrhea, or abdominal pain.** The patient had **no other complaints.** *PHYSICAL EXAMINATION... JVP estimated at 6cm. 2+ carotid pulsations** bilaterally. **No bruits.** *CARDIAC:* Irregularly **irregular,** with normal S1, split S2, with **no murmurs, rubs, or gallops...** He was **without complaint** and had **no evidence of cough or sputum production...** Sputum and **blood cultures** were **negative for evidence of infection.**

(From Mutalik et al. (2001) Negation Detection to Augment Concept Indexing, JAMIA 2001 8: 598-609.)
Negation complexities

- The negation signals were quite heterogeneous, from single words (“no”, “without”, “negative”) to simple phrases (“no evidence of”) and complex verb phrases (“could not be currently identified”).

- There is a large set of verbs that, when preceded by the word “not”, negate their subject or object concept (“X is not seen”, “does not show X”); but there are also a large number of verbs that do not do so (“X did not decrease”, “does not affect X”). These need to be correctly distinguished.

- The negation signals may precede or succeed the concepts they have scope over, and there may be several words between the two (“there was absence of this type of X”, “X, in this instance, is absent”).

- A single negation signal may serve to negate a whole list of concepts either preceding or following it (“A, B, C, and D are absent” “without evidence of A, B, C, or D”); or it may scope over some but not all of them (“there is no A, B and C, and D seemed normal”).
Negation recognition

The Lexer

- It recognizes 60 distinct words or patterns that express negation.
- It passes a specific token to the parser to represent the exact way in which the NegP is used for negation.
- A token is a combination of characteristics:
  - Does the NegP precede or follow the concepts it negates? (“No” – “not present”)
  - Can the NegP negate multiple concepts? (“No” – “non”)
  - Is the terminal conjunction an “or” or an “and”?
    - “no murmurs, rubs or gallops”
    - “murmurs, rubs, and gallops are absent”
- It outputs a “negation-termination” token
The **Parser**

- It assembles contiguous concepts into a list,
- It associates a concept or a list of concepts with a negative phrase that either precedes or follows it to form a negation
- It accurately determines where the negation starts and ends.
Evaluation

<table>
<thead>
<tr>
<th>Negation found by program</th>
<th>Negation not found by observer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True positives:</strong> 135</td>
<td><strong>False positives:</strong> 12</td>
</tr>
<tr>
<td><strong>False negatives:</strong> 6</td>
<td><strong>True negatives:</strong> 1,716</td>
</tr>
</tbody>
</table>

**Specificity:** 91.8%  **Sensitivity:** 95.7%
Error analysis

- no seizure activity throughout his detoxification: also marked “detoxification” as being negated, because the word “throughout” was not on its list of negation terminators
- several blood cultures, six in all, had been negative: could not identify the “blood cultures” as the concept that was being negated by the word “negative” because it was too far away
- Correctly parses some double negatives, such as X-rays were negative except for... but fails on others such as The patient was unable to walk for long periods without dyspnea, where it identified dyspnea as being negated
- non-distended: does not recognize single words with contained negatives
ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports

Henk Harkema a,*, John N. Dowling a, Tyler Thornblade b, Wendy W. Chapman a

a Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA 15260, USA
b Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA

**ConText** determines whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient

**ConText** can be integrated with any application that indexes clinical conditions from text
Motivation

- Clinical documents: source of information for detection and characterization of outbreaks, decision support, recruiting patients for clinical trials, and translational research
- Improving precision of information retrieval and extraction from clinical records by reducing false positives:
  - ruled out pneumonia
  - family history of pneumonia
  - past history of pneumonia
- Most medical language processing applications index or extract individual clinical conditions but do not model much information found in the context of the condition
Algorithm

Assumption

A clinical condition in text is affirmed by default and that a departure from the default value, i.e., the condition is absent, can be inferred from simple lexical clues occurring in the context of the condition.

- ConText is a regular-expression based algorithm that searches for trigger terms preceding or following the indexed clinical conditions.
- If a condition falls within the scope of the trigger term, ConText changes the default value to the value indicated by that trigger term.
"No history of chest tightness but family history of CHF."

- chest tightness
  - Negation: affirmed
  - Experiencer: patient
  - Temporality: recent

- CHF
  - Negation: affirmed
  - Experiencer: patient
  - Temporality: recent

- chest tightness
  - Negation: negated
  - Experiencer: patient
  - Temporality: historical

- CHF
  - Negation: affirmed
  - Experiencer: other
  - Temporality: historical
Trigger term

Trigger terms prompt ConText to change the default value of a contextual property for a condition, provided the condition falls within the scope of the trigger term.

- 143 for negated, 10 for historical, 11 for hypothetical, and 26 for other

Pseudo-trigger terms for terms that contain trigger terms but do not act as contextual property triggers

- 17 pseudo-triggers for negated (e.g., "no increase", "not cause")
Pseudo trigger terms

http://code.google.com/p/negex/wiki/NegExTerms

- no increase
- no suspicious change
- no significant change
- no change
- no interval change
- no definite change
- no significant interval change
- not extend

- not cause
- not drain
- not certain if
- not certain whether
- gram negative
- without difficulty
- not necessarily
- not only
Processing negation in biomedical texts: Context

Trigger terms [http://code.google.com/p/negex/wiki/NegExTerms](http://code.google.com/p/negex/wiki/NegExTerms)

- absence of
- cannot
- cannot see
- checked for
- declined
- declines
- denied
- denies
- denying
- evaluate for
- fails to reveal
- free of
- negative for
- never developed
- never had
- no
- no abnormal
- no cause of
- no complaints of
- no evidence
- no new evidence
- no other evidence
- no evidence to suggest
- no findings of
- no findings to indicate
- no sign of
- no significant
- no signs of
- no suggestion of
- no suspicious
- not
- not appear
- not appreciate
- not associated with
- not complain of
- not demonstrate
- not exhibit
- not know of
- not known to have
- not reveal
- not see
- not to be
- patient was not
- rather than
- resolved
- test for
- to exclude
- unremarkable for
- with no
- without
- without any
- evidence of
- without evidence
- without indication of
- without sign of
- ...
Processing negation in biomedical texts: Context

Termination terms [http://code.google.com/p/negex/wiki/NegExTerms](http://code.google.com/p/negex/wiki/NegExTerms)

- but
- however
- nevertheless
- yet
- though
- although
- still
- aside from
- except
- apart from
- secondary to
- as the cause of
- as the source of
- as the reason of
- as the etiology of
- as the origin of
- as the cause for
- as the source for
- as the reason for
- as the etiology for
- as the origin for
- as the secondary cause of
- as the secondary source of
- as the secondary reason of
- as the secondary etiology of
- as the secondary origin of
- as the secondary cause for
- as the secondary source for
- as the secondary reason for
- as the secondary etiology for
- as the secondary origin for
- as a cause of
- as a source of
- as a reason of
- as a cause for
- as a source for
- as a reason for
- as an etiology for
- as an origin for
- as a trigger event for
- as an etiology of
- as an origin of
- as an etiology for
- as an origin for
- as an etiology for
- as an origin for
- as an etiology for
- as an origin for
- as an etiology for
- as an origin for
The default scope of a trigger term includes all clinical conditions following the trigger term until the end of the sentence or a termination term, but this scope can be overridden

**History of COPD**, presenting with shortness of breath
Definition of scope for negation

There is a set of 14 “left-looking” trigger terms or post-triggers. The scope of these trigger terms runs from the trigger term leftward to the beginning of the sentence, and can be terminated by any regular, intervening termination term.

Eg.: “is ruled out”, “are not seen”, “negative”
Algorithm

1. Mark up all trigger terms, pseudo-trigger terms, and termination terms in the sentence.
2. Iterate through the trigger terms in the sentence from left to right:
   - If the trigger term is a pseudo-trigger term, skip to the next trigger term.
   - Otherwise, determine the scope of the trigger term and assign the appropriate contextual property value to all indexed clinical conditions within the scope of the trigger term.
Evaluation

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>P</th>
<th>F</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Negation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical pathology</td>
<td>.75 .30–.95</td>
<td>.75 .30–.95</td>
<td>.75</td>
<td>4.6%</td>
</tr>
<tr>
<td>Operative procedure</td>
<td>.94 .73–.99</td>
<td>.84 .62–.94</td>
<td>.89</td>
<td>17.10%</td>
</tr>
<tr>
<td>Radiology</td>
<td>.86 .71–.94</td>
<td>1.0 .89–1.0</td>
<td>.93</td>
<td>35.23%</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>.91 .78–.97</td>
<td>.97 .85–.97</td>
<td>.94</td>
<td>35.6%</td>
</tr>
<tr>
<td>Discharge summary</td>
<td>.89 .79–.94</td>
<td><strong>.84 .74–.90</strong></td>
<td>.86</td>
<td>74.18%</td>
</tr>
<tr>
<td>Emergency department</td>
<td>.93 .90–.95</td>
<td>.96 .93–.98</td>
<td>.95</td>
<td>325.36%</td>
</tr>
<tr>
<td>All</td>
<td>.92 .89–.94</td>
<td>.94 .91–.96</td>
<td>.93</td>
<td>490.22%</td>
</tr>
</tbody>
</table>

- Performs comparably well on all report types, apart from discharge summaries
- FP in discharge summaries are due to missing terms, the pseudo-trigger “with/without”
- Access to linguistic knowledge will improve performance by making the determination of the scope of a trigger term more precise
- Lexical clues or trigger words for negation, when they occur in multiple report types, have the same interpretation across report types
Outline

9 Detecting speculated sentences
10 Processing negation in biomedical texts
11 Scope resolution
12 Finding negated and speculated events
13 Modality tagging
14 Belief categorisation
15 Processing contradiction and contrast
16 Visualising negation features
17 References
Scope resolution: Negation

Task definition

Finding the scope of a negation signal means determining at a sentence level which words in the sentence are affected by the negation(s).

Analysis at the phenotype and genetic level showed that lack of CD5 expression was due neither to segregation of human autosome 11, on which the CD5 gene has been mapped, nor to deletion of the CD5 structural gene.
Modelling the task

- We model the scope finding task as two consecutive classification tasks:
  1. Finding negation signals: a token is classified as being at the beginning of a negation signal, inside or outside
  2. Finding the scope: a token is classified as being the first element or the last of a scope sequence

- Supervised machine learning approach
Corpora: BioScope

- Abstracts corpus:
  10 fold cross-validation experiments
- Clinical and full papers corpora: robustness test
  - Training on abstracts
  - Testing on clinical and full papers
Scope resolution: Negation

System architecture
## Scope resolution: Negation

### Preprocessing

<table>
<thead>
<tr>
<th>N.</th>
<th>TOKEN</th>
<th>LEMMA</th>
<th>POS</th>
<th>CHUNK</th>
<th>NE</th>
<th>NEG SIGNAL</th>
<th>SCOPE</th>
<th>SCOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>These</td>
<td>These</td>
<td>DT</td>
<td>B-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>syncytia</td>
<td>syncytia</td>
<td>NN</td>
<td>I-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>lack</td>
<td>lack</td>
<td>NN</td>
<td>I-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>activated</td>
<td>activate</td>
<td>VBN</td>
<td>B-VP</td>
<td>O</td>
<td>0</td>
<td>B-NEG</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>cells</td>
<td>cell</td>
<td>NNS</td>
<td>B-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>as</td>
<td>as</td>
<td>IN</td>
<td>B-SBAR</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>determined</td>
<td>determine</td>
<td>VBN</td>
<td>B-VP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>by</td>
<td>by</td>
<td>IN</td>
<td>B-PP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>an</td>
<td>an</td>
<td>DT</td>
<td>B-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>absence</td>
<td>absence</td>
<td>NN</td>
<td>I-NP</td>
<td>O</td>
<td>0</td>
<td>B-NEG</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>of</td>
<td>of</td>
<td>IN</td>
<td>B-PP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>staining</td>
<td>staining</td>
<td>NN</td>
<td>B-NP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>for</td>
<td>for</td>
<td>IN</td>
<td>B-PP</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>Ki-67</td>
<td>Ki-67</td>
<td>NN</td>
<td>B-NP</td>
<td>B-protein</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>cell</td>
<td>cell</td>
<td>NN</td>
<td>I-NP</td>
<td>I-protein</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>cycle</td>
<td>cycle</td>
<td>NN</td>
<td>I-NP</td>
<td>I-protein</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>antigen</td>
<td>antigen</td>
<td>NN</td>
<td>I-NP</td>
<td>I-protein</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>O</td>
<td>O</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Finding negation cues

- We filter out negation signals that are unambiguous in the training corpus (17 out of 30).
- For the rest a classifier predicts whether a token is the first token of a negation signal, inside or outside of it:
  - Algorithm: IGTREE as implemented in TiMBL (Daelemans et al. 2007).
  - Instances represent all tokens in a sentence.
  - Features about the token:
    - Lemma, word, POS and IOB chunk tag.
  - Features about the token context:
    - Word, POS and IOB chunk tag of 3 tokens to the right and 3 to the left.

<table>
<thead>
<tr>
<th>N.</th>
<th>TOKEN</th>
<th>NEG SIGNAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>These</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>syncytia</td>
<td>O</td>
</tr>
<tr>
<td>3</td>
<td>lack</td>
<td>O</td>
</tr>
<tr>
<td>4</td>
<td>activated</td>
<td>O</td>
</tr>
<tr>
<td>5</td>
<td>cells</td>
<td>O</td>
</tr>
<tr>
<td>6</td>
<td>as</td>
<td>O</td>
</tr>
<tr>
<td>7</td>
<td>determined</td>
<td>O</td>
</tr>
<tr>
<td>8</td>
<td>by</td>
<td>O</td>
</tr>
<tr>
<td>9</td>
<td>an</td>
<td>O</td>
</tr>
<tr>
<td>10</td>
<td>absence</td>
<td>O</td>
</tr>
<tr>
<td>11</td>
<td>of</td>
<td>O</td>
</tr>
<tr>
<td>12</td>
<td>staining</td>
<td>O</td>
</tr>
<tr>
<td>13</td>
<td>for</td>
<td>O</td>
</tr>
<tr>
<td>14</td>
<td>Ki-67</td>
<td>O</td>
</tr>
<tr>
<td>15</td>
<td>cell</td>
<td>O</td>
</tr>
<tr>
<td>16</td>
<td>cycle</td>
<td>O</td>
</tr>
<tr>
<td>17</td>
<td>antigen</td>
<td>O</td>
</tr>
<tr>
<td>18</td>
<td>.</td>
<td>O</td>
</tr>
</tbody>
</table>
Finding negation cues: results

- Baseline: tagging as negation signals tokens that are negation signals at least in 50% of the occurrences in the training corpus

<table>
<thead>
<tr>
<th>BASELINE</th>
<th>PREC</th>
<th>RECALL</th>
<th>F1</th>
<th>IAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>82.00</td>
<td>95.17</td>
<td>88.09</td>
<td>94.46</td>
</tr>
<tr>
<td>Papers</td>
<td>84.01</td>
<td>92.46</td>
<td>88.03</td>
<td>79.42</td>
</tr>
<tr>
<td>Clinical</td>
<td>97.31</td>
<td>97.53</td>
<td>97.42</td>
<td>90.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>PREC</th>
<th>RECALL</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>84.72</td>
<td>98.75</td>
<td>91.20(+3.11)</td>
</tr>
<tr>
<td>Papers</td>
<td>87.18</td>
<td>95.72</td>
<td>91.25(+3.22)</td>
</tr>
<tr>
<td>Clinical</td>
<td>97.33</td>
<td>98.09</td>
<td>97.71(+0.29)</td>
</tr>
</tbody>
</table>
Finding negation cues: system versus baseline
Finding negation cues: results in 3 corpora

- Prec
- Recall
- F1

- Abstracts
- Papers
- Clinical
Discussion

- Cause of lower recall on papers corpus:

<table>
<thead>
<tr>
<th>NOT</th>
<th>% negation signals</th>
<th>% classified correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>58.89</td>
<td>98.25</td>
</tr>
<tr>
<td>Papers</td>
<td>53.22</td>
<td>93.68</td>
</tr>
<tr>
<td>Clinical</td>
<td>6.72</td>
<td>91.22</td>
</tr>
</tbody>
</table>

- Errors: *not* is classified as negation signal
  However, programs for tRNA identification [...] do not necessarily perform well on unknown ones
  The evaluation of this ratio is difficult because not all true interactions are known
The features used by the object classifiers and the metalearner are different.
## Scope finding

<table>
<thead>
<tr>
<th>N.</th>
<th>TOKEN</th>
<th>NEG</th>
<th>SCOPE</th>
<th>N.</th>
<th>TOKEN</th>
<th>NEG</th>
<th>SCOPE</th>
<th>N.</th>
<th>TOKEN</th>
<th>NEG</th>
<th>SCOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>These</td>
<td>lack</td>
<td>0</td>
<td>1</td>
<td>These</td>
<td>absence</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>syncytia</td>
<td>lack</td>
<td>0</td>
<td>2</td>
<td>syncytia</td>
<td>absence</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>lack</td>
<td>lack</td>
<td>0</td>
<td>3</td>
<td>lack</td>
<td>absence</td>
<td>FIRST 0</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>activated</td>
<td>lack</td>
<td>LAST 0</td>
<td>4</td>
<td>activated</td>
<td>absence</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>cells</td>
<td>lack</td>
<td>0</td>
<td>5</td>
<td>cells</td>
<td>absence</td>
<td>0</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>as</td>
<td>lack</td>
<td>0</td>
<td>6</td>
<td>as</td>
<td>absence</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>determined</td>
<td>lack</td>
<td>0</td>
<td>7</td>
<td>determined</td>
<td>absence</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>by</td>
<td>lack</td>
<td>0</td>
<td>8</td>
<td>by</td>
<td>absence</td>
<td>0</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>an</td>
<td>lack</td>
<td>0</td>
<td>9</td>
<td>an</td>
<td>absence</td>
<td>0</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>absence</td>
<td>lack</td>
<td>0</td>
<td>10</td>
<td>absence</td>
<td>absence</td>
<td>0</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>of</td>
<td>lack</td>
<td>0</td>
<td>11</td>
<td>of</td>
<td>absence</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>staining</td>
<td>lack</td>
<td>0</td>
<td>12</td>
<td>staining</td>
<td>absence</td>
<td>0</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>for</td>
<td>lack</td>
<td>0</td>
<td>13</td>
<td>for</td>
<td>absence</td>
<td>0</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Ki-67</td>
<td>lack</td>
<td>0</td>
<td>14</td>
<td>Ki-67</td>
<td>absence</td>
<td>FIRST 0</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>cell</td>
<td>lack</td>
<td>0</td>
<td>15</td>
<td>cell</td>
<td>absence</td>
<td>0</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>cycle</td>
<td>lack</td>
<td>0</td>
<td>16</td>
<td>cycle</td>
<td>absence</td>
<td>0</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>antigen</td>
<td>lack</td>
<td>0</td>
<td>17</td>
<td>antigen</td>
<td>absence</td>
<td>LAST 0</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>lack</td>
<td>0</td>
<td>18</td>
<td></td>
<td>absence</td>
<td>0</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Scope finding: features classifiers

- **Of the negation signal**: Chain of words
- **Of the paired token**: Lemma, POS, chunk IOB tag, type of chunk; lemma of the second and third tokens to the left; lemma, POS, chunk IOB tag, and type of chunk of the first token to the left and three tokens to the right; first word, last word, chain of words, and chain of POSs of the chunk of the paired token and of two chunks to the left and two chunks to the right.
- **Of the tokens between the negation signal and the token in focus**: Chain of POS types, distance in number of tokens, and chain of chunk IOB tags.
- **Others**: A feature indicating the location of the token relative to the negation signal (pre, post, same).
Scope resolution: Negation

Scope finding: features metalearner

- **Of the negation signal**: Chain of words, chain of POS, word of the two tokens to the right and two tokens to the left, token number divided by the total number of tokens in the sentence.
- **Of the paired token**: Lemma, POS, word of two tokens to the right and two tokens to the left, token number divided by the total number of tokens in the sentence.
- **Of the tokens between the negation signal and the token in focus**: Binary features indicating if there are commas, colons, semicolons, verbal phrases or one of the following words between the negation signal and the token in focus: Whereas, but, although, nevertheless, notwithstanding, however, consequently, hence, therefore, thus, instead, otherwise, alternatively, furthermore, moreover.
- **About the predictions of the three classifiers**: prediction, previous and next predictions of each of the classifiers, full sequence of previous and full sequence of next predictions of each of the classifiers.
- **Others**: A feature indicating the location of the token relative to the negation signal (pre, post, same).
Scope resolution: Negation

Scope finding: postprocessing

- Scope is always a consecutive block of scope tokens, including the negation signal
- The classifiers predict the first and last token of the scope sequence: None or more than one FIRST and one LAST elements are predicted
- In the post-processing we apply some rules to select one FIRST and one LAST token
  - Example: If more than one token has been predicted as FIRST, take as FIRST the first token of the negation signal
Scope finding: baseline

- Baseline: calculating the average length of the scope to the right of the negation signal and tagging that number of tokens as scope tokens
- Motivation: 85.70% of scopes to the right

<table>
<thead>
<tr>
<th>BASELINE</th>
<th>PCS</th>
<th>PCS-2</th>
<th>IAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>7.11</td>
<td>37.45</td>
<td>92.46</td>
</tr>
<tr>
<td>Papers</td>
<td>4.76</td>
<td>24.86</td>
<td>70.86</td>
</tr>
<tr>
<td>Clinical</td>
<td>12.95</td>
<td>62.27</td>
<td>76.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>PCS</th>
<th>PCS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>66.07</td>
<td>66.93</td>
</tr>
<tr>
<td>Papers</td>
<td>41.00</td>
<td>44.44</td>
</tr>
<tr>
<td>Clinical</td>
<td>70.75</td>
<td>71.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM gold negs</th>
<th>PCS</th>
<th>PCS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>+7.29</td>
<td>+7.17</td>
</tr>
<tr>
<td>Papers</td>
<td>+9.26</td>
<td>+9.79</td>
</tr>
<tr>
<td>Clinical</td>
<td>+16.52</td>
<td>+16.74</td>
</tr>
</tbody>
</table>
The system performs clearly better than baseline
There is a higher upperbound calculated with gold standard negation signals
The system is portable
Lower results in the papers corpus
Scope finding: discussion

- Clinical reports are easier to process than abstracts and papers
- Negation signal *no* is very frequent (76.65 %) and has a high PCS (73.10 %)
  - No findings to account for symptoms
  - No signs of tuberculosis
- Sentences are shorter in clinical reports than in abstracts and papers:
  - Average length in clinical reports is 7.8 tokens vs. 26.43 in abstracts and 26.24 in full papers
  - 75.85 % of the sentences have 10 or less tokens
Scope finding: discussion

- Papers are more difficult to process than abstracts
  - Negation signal *not* is frequent (53.22%) and has a low PCS (39.50) in papers. Why?

<table>
<thead>
<tr>
<th></th>
<th>Papers</th>
<th>Abstracts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguity (%)&lt;sup&gt;-neg&lt;/sup&gt;</td>
<td>25.56</td>
<td>14.29</td>
</tr>
<tr>
<td>Av. scope length</td>
<td>6.45</td>
<td>8.85</td>
</tr>
<tr>
<td>% Scopes left</td>
<td>23.28</td>
<td>16.41</td>
</tr>
<tr>
<td>Av. scope left</td>
<td>5.60</td>
<td>8.82</td>
</tr>
</tbody>
</table>
The metalearner performs better than the three object classifiers (except SVMs on the clinical corpus).
Resolving the scope of negation for sentiment analysis

Scope resolution: Negation

Goal

To construct a negation system that can correctly identify the presence or absence of negation in spans of text that are expressions of sentiment

- Focus on explicit negation mentions
- Conditional Random Fields (Lafferty, McCallum and Pereira 2001)
  - Structured prediction learning framework
- Features from dependency syntax
- Evaluation on corpus of product reviews and BioScope corpus
Datasets

- BioScope corpus
- Product Reviews corpus (by Google, not publicly available)
  - 268 product reviews sampled from Google Product Search
  - 2111 sentences, 679 sentences with negation
  - 91% inter-annotator agreement, strict exact span
  - Robustness: ungrammatical sentences, misspelling
### Lexicon of negation cues

<table>
<thead>
<tr>
<th>hardly</th>
<th>lack</th>
<th>lacking</th>
<th>lacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>neither</td>
<td>nor</td>
<td>never</td>
<td>no</td>
</tr>
<tr>
<td>nobody</td>
<td>none</td>
<td>nothing</td>
<td>nowhere</td>
</tr>
<tr>
<td>not</td>
<td>n’t</td>
<td>aint</td>
<td>cant</td>
</tr>
<tr>
<td>cannot</td>
<td>darent</td>
<td>dont</td>
<td>doesn’t</td>
</tr>
<tr>
<td>didnt</td>
<td>hadnt</td>
<td>hasnt</td>
<td>havnt</td>
</tr>
<tr>
<td>havent</td>
<td>isnt</td>
<td>mightnt</td>
<td>mustnt</td>
</tr>
<tr>
<td>neednt</td>
<td>oughtnt</td>
<td>shant</td>
<td>shouldn’t</td>
</tr>
<tr>
<td>wasn’t</td>
<td>wouldn’t</td>
<td>without</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Lexicon of explicit negation cues.
Scope resolution: Negation

System description

1. Negation cues are detected using a lexicon
2. Scopes are processed by the negation annotator:
   - Input: sentence boundary + dependency (MaltParser) annotations
   - Algorithm: CRF++
     - Label set of size two indicating whether a token is within or outside of a negation span
   - Features: (next slide)
     - Only unigram features are employed, but each unigram feature vector is expanded to include bigram and trigram representations derived from the current token in conjunction with the prior and subsequent tokens
Scope resolution: Negation

Features

- Lowercased token string
- POS of a token
- Linear token-wise distance to the nearest explicit negation cue to the right of a token
- Linear token-wise distance to the nearest explicit negation cue to the left of a token
- POS of the first order dependency of a token
- Minimum number of dependency relations that must be traversed to from the first order dependency head of a token to an explicit negation cue
- POS of the second order dependency of a token
- The minimum number of dependency relations that must be traversed to from the second order dependency head of a token to an explicit negation cue
Scope resolution: Negation

Evaluation

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
<th>PCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviews</td>
<td>81.9</td>
<td>78.2</td>
<td>80.0</td>
<td>39.8</td>
</tr>
<tr>
<td>BioScope</td>
<td>80.8</td>
<td>70.8</td>
<td>75.5</td>
<td>53.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
<th>PCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioScope, trained on Reviews</td>
<td>72.2</td>
<td>42.1</td>
<td>53.5</td>
<td>52.2</td>
</tr>
<tr>
<td>Reviews, trained on BioScope</td>
<td>58.8</td>
<td>68.8</td>
<td>63.4</td>
<td>45.7</td>
</tr>
</tbody>
</table>

- Punctuation tokens are not counted
- BioScope: 5-f cv; Reviews: 7-f cv
Scope resolution: Negation

Negation system built into a sentiment analysis pipeline

1. Sentence boundary detection: finds and scores mentions of n-grams found in a large lexicon of sentiment terms and phrases
2. Sentiment detection:
3. Negation scope detection
4. Sentence sentiment scoring:
   ▶ Determines whether any scored sentiment terms fall within the scope of a negation, and flips the sign of the sentiment score for all negated sentiment terms
   ▶ Sums all sentiment scores within each sentence and computes overall sentence sentiment scores
Effect on sentiment classification

- 1135 sentences
- Human raters were asked to classify each sentence as expressing one of the following types of sentiment:
  - positive
  - negative
  - neutral
  - mixed positive and negative
- 216 sentences (19% contained negations)
  - positive: 73
  - negative: 114
  - neutral: 12
  - mixed positive and negative: 17
- The effect of the negation system on sentiment classification was evaluated on the smaller subset of 216 sentences
**Scope resolution: Negation**

**Effect on sentiment classification**

Figure 1: Precision-recall curve showing the effect of negation detection on positive sentiment prediction.

Figure 2: Precision-recall curve showing the effect of negation detection on negative sentiment prediction.

- A significant improvement is apparent at all recall levels
Effect on sentiment classification

<table>
<thead>
<tr>
<th>Metric</th>
<th>w/o Neg.</th>
<th>w/ Neg.</th>
<th>% Improv.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive Sentiment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prec.</td>
<td>44.0</td>
<td>64.1</td>
<td>35.9</td>
</tr>
<tr>
<td>Recall</td>
<td>54.8</td>
<td>63.7</td>
<td>20.0</td>
</tr>
<tr>
<td>F1</td>
<td>48.8</td>
<td>63.9</td>
<td>29.5</td>
</tr>
<tr>
<td><strong>Negative Sentiment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prec.</td>
<td>68.6</td>
<td>83.3</td>
<td>46.8</td>
</tr>
<tr>
<td>Recall</td>
<td>21.1</td>
<td>26.3</td>
<td>6.6</td>
</tr>
<tr>
<td>F1</td>
<td>32.3</td>
<td>40.0</td>
<td>11.4</td>
</tr>
</tbody>
</table>

- Performance is improved by introducing negation scope detection
- The precision of positive sentiment predictions sees the largest improvement, largely due to the inherent bias in the sentiment scoring algorithm

Table 5: Sentiment classification results, showing the percentage improvement obtained from including negation scope detection (w/ Neg.) over results obtained without including negation scope detection (w/o Neg.).
Task definition

Finding the scope of a hedge cue means determining at a sentence level which words in the sentence are affected by the hedge(s)

These results [suggest that expression of c-jun, jun B and jun D genes [might be involved in terminal granulocyte differentiation [or in regulating granulocyte functionality]]].
Related work

- Machine learning systems
  - Morante and Daelemans (2009a,b)
  - Agarwal and Yu (2010)
  - Zhu et al. (2010)

- Rule-based systems: use syntactic information
  - Jia et al. (2009)
  - Özgür and Radev (2009)
  - Øvrelid et al. (2010)

- CoNLL Shared Task 2010
Scope resolution: Hedges


- System based on the system that processed negation cues
- Goal: investigate whether the same system can process hedge cues
- We model the scope finding task as two consecutive classification tasks:
  1. Finding negation signals
  2. Finding the scope

- BioScope corpus
  - Abstracts corpus:
    10 fold cross-validation experiments
  - Clinical and papers corpora: robustness test
    Training on abstracts - Testing on clinical and papers
## Data preprocessing

<table>
<thead>
<tr>
<th>N.</th>
<th>TOKEN</th>
<th>LEMMA</th>
<th>POS</th>
<th>CHUNK</th>
<th>NE</th>
<th>HEDGE CUE</th>
<th>SCOPE</th>
<th>SCOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Together</td>
<td>Together</td>
<td>RB</td>
<td>B-ADVP</td>
<td>0</td>
<td>0</td>
<td>O-SPEC</td>
<td>O-SPEC</td>
</tr>
<tr>
<td>2</td>
<td>these</td>
<td>these</td>
<td>DT</td>
<td>B-NP</td>
<td>0</td>
<td>0</td>
<td>O-SPEC</td>
<td>O-SPEC</td>
</tr>
<tr>
<td>3</td>
<td>data</td>
<td>datum</td>
<td>NNS</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>O-SPEC</td>
<td>O-SPEC</td>
</tr>
<tr>
<td>4</td>
<td>suggest</td>
<td>suggest</td>
<td>VBP</td>
<td>B-VP</td>
<td>0</td>
<td>B-speculation</td>
<td>0</td>
<td>B-SPEC</td>
</tr>
<tr>
<td>5</td>
<td>that</td>
<td>that</td>
<td>IN</td>
<td>B-SBAR</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>O-SPEC</td>
</tr>
<tr>
<td>6</td>
<td>ETS1</td>
<td>ETS1</td>
<td>NN</td>
<td>B-NP</td>
<td>0</td>
<td>B-speculation</td>
<td>0</td>
<td>B-SPEC</td>
</tr>
<tr>
<td>7</td>
<td>may</td>
<td>may</td>
<td>MD</td>
<td>B-VP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>8</td>
<td>be</td>
<td>be</td>
<td>VB</td>
<td>I-VP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>9</td>
<td>involved</td>
<td>involve</td>
<td>VBN</td>
<td>I-VP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>10</td>
<td>in</td>
<td>in</td>
<td>IN</td>
<td>B-PP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>11</td>
<td>mediating</td>
<td>mediate</td>
<td>VBG</td>
<td>B-VP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>12</td>
<td>the</td>
<td>the</td>
<td>DT</td>
<td>B-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>13</td>
<td>increased</td>
<td>increase</td>
<td>VBN</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>14</td>
<td>GM-CSF</td>
<td>GM-CSF</td>
<td>NN</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>15</td>
<td>production</td>
<td>production</td>
<td>NN</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>16</td>
<td>associated</td>
<td>associate</td>
<td>VBN</td>
<td>B-VP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>17</td>
<td>with</td>
<td>with</td>
<td>IN</td>
<td>B-PP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>18</td>
<td>T</td>
<td>T</td>
<td>NN</td>
<td>B-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>19</td>
<td>cell</td>
<td>cell</td>
<td>NN</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>20</td>
<td>activation</td>
<td>activation</td>
<td>NN</td>
<td>I-NP</td>
<td>0</td>
<td>0</td>
<td>I-SPEC</td>
<td>I-SPEC</td>
</tr>
<tr>
<td>21</td>
<td>,</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>O</td>
<td>0</td>
<td>O-SPEC</td>
<td>O-SPEC</td>
</tr>
</tbody>
</table>
Scope resolution: Hedges

System architecture

PREPROCESSING
GENIA Tagger
POS, Chunks, NE

FINDING HEDGE CUES
C0 MBL
Predictions C0

POSTPROCESSING
Selecting consecutive block of scope tokens

SCOPE FINDING
C1 MBL
Predictions C1
C2 SVM
Predictions C2
C3 CRF
Predictions C3
C4 CRF
Predictions C4
## Scope resolution: Hedges

### Results cue finding

- Baseline: tagging as hedge cues a list of words extracted from the abstracts corpus

<table>
<thead>
<tr>
<th>BASELINE</th>
<th>PREC</th>
<th>RECALL</th>
<th>F1</th>
<th>IAA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>55.62</td>
<td>71.77</td>
<td>62.67</td>
<td>79.12</td>
</tr>
<tr>
<td>Papers</td>
<td>54.39</td>
<td>61.21</td>
<td>57.60</td>
<td>77.60</td>
</tr>
<tr>
<td>Clinical</td>
<td>66.55</td>
<td>40.78</td>
<td>50.57</td>
<td>84.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>PREC</th>
<th>RECALL</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>90.81</td>
<td>79.84</td>
<td>84.77</td>
</tr>
<tr>
<td>Papers</td>
<td>75.35</td>
<td>68.18</td>
<td>71.59</td>
</tr>
<tr>
<td>Clinical</td>
<td>88.10</td>
<td>27.51</td>
<td>41.92</td>
</tr>
</tbody>
</table>
Results cue finding across corpora

Scope resolution: Hedges
Discussion

- Cause of lower recall on clinical corpus:

<table>
<thead>
<tr>
<th>OR</th>
<th>total #</th>
<th>% as hedge</th>
<th># as hedge</th>
<th>% of hedges</th>
<th>recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>1062</td>
<td>11.29</td>
<td>118</td>
<td><strong>4.42</strong></td>
<td>0.129</td>
</tr>
<tr>
<td>Papers</td>
<td>153</td>
<td>16.99</td>
<td>27</td>
<td><strong>4.04</strong></td>
<td>0.137</td>
</tr>
<tr>
<td>Clinical</td>
<td>281</td>
<td>98.22</td>
<td>276</td>
<td><strong>24.62</strong></td>
<td>0.007</td>
</tr>
</tbody>
</table>

- The use of OR as hedge cue is difficult to interpret

+CUE: Nucleotide sequence and PCR analyses demonstrated the presence of novel duplications or deletions involving the NF-kappa B motif.

-CUE: In nuclear extracts from monocytes or macrophages, induction of NF-KB occurred only if the cells were previously infected with HIV-1. (= AND)
Scope resolution: Hedges

Results scope resolution across corpora

- Baseline: calculating the average length of the scope to the right of the hedge cue and tagging that number of tokens as scope tokens
  - Motivation: 82.45 % of scopes to the right

<table>
<thead>
<tr>
<th>BASELINE</th>
<th>PCS</th>
<th>PCS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>3.15</td>
<td>3.17</td>
</tr>
<tr>
<td>Papers</td>
<td>2.19</td>
<td>2.26</td>
</tr>
<tr>
<td>Clinical</td>
<td>2.72</td>
<td>3.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>PCS</th>
<th>PCS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts</td>
<td>65.55</td>
<td>66.10</td>
</tr>
<tr>
<td>Papers</td>
<td>35.92</td>
<td>42.37</td>
</tr>
<tr>
<td>Clinical</td>
<td>26.21</td>
<td>27.44</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SYSTEM</th>
<th>PCS</th>
<th>PCS-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstracts gold cues</td>
<td>+11.58</td>
<td>+12.11</td>
</tr>
<tr>
<td>Papers</td>
<td>+12.02</td>
<td>+15.84</td>
</tr>
<tr>
<td>Clinical</td>
<td>+34.38</td>
<td>+36.50</td>
</tr>
</tbody>
</table>
Results scope resolution

- Prec
- Rec
- F1
- PCS
- PCS2

- Abstracts
- Papers
- Clinical
Discussion

- Why are the results in papers lower?
  - 41 cues (47.00%) in papers are not in abstracts
  - Some cues that occur in abstracts and are frequent in papers get low scores. They are used differently. (Ex. suggest: 92.33 PCS in abstracts vs. 62.85 PCS in papers)

- Why are the results in clinical lower?
  - 68 cues (35.45%) in clinical are not in abstracts
  - Frequent hedge cues in clinical are not represented in abstracts
Comparison negation - hedge processing systems

- Gold hedge cues = no error propagation from the first phase
- The abstracts results show that the same system can be applied to finding the scope of negation and hedge processing
- The systems are equally portable to the papers corpus
- The negation system is better portable to the clinical corpus
Comparison negation - hedge processing systems

PCS - Predicted Cues Systems

- Error propagation from the first phase:
  - The hedge system is much less portable to the clinical corpus than the negation system

- Supervised hedge detection
  - Algorithm: SVM
- Scope finding based on syntactic information
  - Data parsed with Stanford Dependency Parser (de Marneffe et al. 2006)
- Data: BioScope corpus
Features for hedge detection

**Dependency syntax**

- **Clausal Complement** set to 1 if the keyword has a child which is connected to it with a clausal complement or infinitival clause dependency type
- **Negation**: set to 1 if the keyword (1) has a child which is connected to it with a negation dependency type or the determiner “no” is a child of the keyword
- **Auxiliary**: set to 1 if the keyword has a child which is connected to it with an auxiliary dependency type

**Positional features for abstracts**

Motivation: different parts of a text might have different characteristics in terms of the usage of speculative language

- For abstracts: title, first sentence, last sentence
- For full papers: title, first sentence, last sentence, background, results, methods, conclusion, legend

**Co-occurring keywords**
Results hedge detection in abstracts

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1</td>
<td>52.84</td>
<td>92.71</td>
<td>67.25</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>97.54</td>
<td>43.66</td>
<td>60.30</td>
</tr>
<tr>
<td>BOW 3 - stemmed</td>
<td>81.47</td>
<td>92.36</td>
<td>86.51</td>
</tr>
<tr>
<td>BOW 2 - stemmed</td>
<td>81.56</td>
<td>93.29</td>
<td>86.97</td>
</tr>
<tr>
<td>BOW 1 - stemmed</td>
<td>83.08</td>
<td>93.83</td>
<td>88.05</td>
</tr>
<tr>
<td>BOW 3</td>
<td>82.58</td>
<td>92.04</td>
<td>86.98</td>
</tr>
<tr>
<td>BOW 2</td>
<td>82.77</td>
<td>92.74</td>
<td>87.41</td>
</tr>
<tr>
<td>BOW 1</td>
<td>83.27</td>
<td>93.67</td>
<td>88.10</td>
</tr>
<tr>
<td>KW: kw, kw-stem, kw-pos</td>
<td>88.62</td>
<td>92.77</td>
<td>90.61</td>
</tr>
<tr>
<td>KW, DEP</td>
<td>88.77</td>
<td>92.67</td>
<td>90.64</td>
</tr>
<tr>
<td>KW, DEP, BOW 1</td>
<td>88.46</td>
<td>94.71</td>
<td>91.43</td>
</tr>
<tr>
<td>KW, DEP, BOW 1, POS</td>
<td>88.16</td>
<td>95.21</td>
<td>91.50</td>
</tr>
<tr>
<td>KW, DEP, BOW 1, POS, CO-KW</td>
<td>88.22</td>
<td>95.56</td>
<td>91.69</td>
</tr>
</tbody>
</table>

Baseline 1: 14 keywords Light et al. (2004)
Baseline 2: keywords from train corpus
## Results hedge detection in papers

<table>
<thead>
<tr>
<th>Method</th>
<th>Recall</th>
<th>Precision</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1</td>
<td>33.77</td>
<td>86.75</td>
<td>47.13</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>88.22</td>
<td>52.57</td>
<td>64.70</td>
</tr>
<tr>
<td>BOW 3 - stemmed</td>
<td>70.79</td>
<td>83.88</td>
<td>76.58</td>
</tr>
<tr>
<td>BOW 2 - stemmed</td>
<td>72.31</td>
<td>85.49</td>
<td>78.11</td>
</tr>
<tr>
<td>BOW 1 - stemmed</td>
<td>73.49</td>
<td>84.35</td>
<td>78.41</td>
</tr>
<tr>
<td>BOW 3</td>
<td>70.54</td>
<td>82.56</td>
<td>75.88</td>
</tr>
<tr>
<td>BOW 2</td>
<td>71.52</td>
<td>85.93</td>
<td>77.94</td>
</tr>
<tr>
<td>BOW 1</td>
<td>73.72</td>
<td>86.27</td>
<td>79.43</td>
</tr>
<tr>
<td>KW: kw, kw-stem, kw-pos</td>
<td>75.21</td>
<td>87.08</td>
<td>80.57</td>
</tr>
<tr>
<td>KW, DEP</td>
<td>75.02</td>
<td>89.49</td>
<td>81.53</td>
</tr>
<tr>
<td>KW, DEP, BOW 1</td>
<td>76.15</td>
<td>89.54</td>
<td>82.27</td>
</tr>
<tr>
<td><strong>KW, DEP, BOW 1, POS</strong></td>
<td><strong>76.17</strong></td>
<td><strong>90.81</strong></td>
<td><strong>82.82</strong></td>
</tr>
<tr>
<td>KW, DEP, BOW 1, POS, CO-KW</td>
<td>75.76</td>
<td>90.82</td>
<td>82.58</td>
</tr>
</tbody>
</table>

Baseline 1: 14 keywords Light et al. (2004)
Baseline 2: keywords from train corpus
Resolving the scopes

- Assumption: the scope of a keyword can be characterized by its part-of-speech and the syntactic structure of the sentence in which it occurs
- Rule-based approach
  - The scope of a conjunction or a determiner is the syntactic phrase to which it is attached
  - The scope of a modal verb is the “VP” to which it is attached
  - The scope of an adjective or an adverb starts with the keyword and ends with the last token of the highest level “NP” which dominates the adjective or the adverb
  - The scope of a verb followed by an infinitival clause extends to the whole sentence
  - The scope of a verb in passive voice extends to the whole sentence
  - If none of the above rules apply, the scope of a keyword starts with the keyword and ends at the end of the sentence
## Results scope resolution

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy-Abstracts</th>
<th>Accuracy-Full text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline 1</td>
<td>4.82</td>
<td>4.29</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>67.60</td>
<td>42.82</td>
</tr>
<tr>
<td>Rule-based method</td>
<td>79.89</td>
<td>61.13</td>
</tr>
</tbody>
</table>

Baseline 1: assign scope to the whole sentence
Baseline 2: assign scope from keyword to the end of the sentence
Scope resolution: Hedges


- Supervised system using CRF as implemented in the ABNER library
- Pipeline system: cue identification + scope resolution
- Task modelled as in Morante and Daelemans (2009) and Özgür and Radev (2009)
- The corpus partitions and the evaluation measures are different. Systems are not comparable
## Scope resolution: Hedges

### Systems (From Agarwal et al. 2010)

<table>
<thead>
<tr>
<th>System name</th>
<th>Detects</th>
<th>Features used</th>
</tr>
</thead>
<tbody>
<tr>
<td>HedgeCue</td>
<td>Hedge cues</td>
<td>Words</td>
</tr>
<tr>
<td>BaselineCue</td>
<td>Hedge cues</td>
<td>Words</td>
</tr>
<tr>
<td>HedgeScope</td>
<td>Scope of a hedge cue</td>
<td>Words, POS tags, Cue phrase words not replaced with POS tags, POS tags</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cue phrase words replaced with custom tag ‘CUE’, POS tags</td>
</tr>
<tr>
<td>BaselineScope</td>
<td>Scope of a hedge cue</td>
<td>Words</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Words</td>
</tr>
</tbody>
</table>
### Results (From Agarwal et al. 2010)

<table>
<thead>
<tr>
<th>Features used</th>
<th>HedgeScope</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Part of speech</td>
<td>Part of speech</td>
<td>Part of speech</td>
<td>Part of speech</td>
</tr>
<tr>
<td>Cue phrase identified using</td>
<td>–</td>
<td>HedgeCue</td>
<td>HedgeCue</td>
<td>BaselineCue</td>
<td>BaselineCue</td>
</tr>
<tr>
<td>Cue phrase replaced</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Scope limited by</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Recall</td>
<td>78.81 ± 0.02</td>
<td>82.47 ± 0.02</td>
<td>83.91 ± 0.02</td>
<td>90.78 ± 0.01</td>
<td>91.59 ± 0.01</td>
</tr>
<tr>
<td>Precision</td>
<td>84.82 ± 0.01</td>
<td>88.98 ± 0.01</td>
<td>88.54 ± 0.01</td>
<td>74.46 ± 0.04</td>
<td>74.6 ± 0.06</td>
</tr>
<tr>
<td>F1-score</td>
<td>81.7 ± 0.02</td>
<td>85.6 ± 0.01</td>
<td><strong>86.16 ± 0.01</strong></td>
<td>81.81 ± 0.02</td>
<td>82.23 ± 0.03</td>
</tr>
<tr>
<td>Accuracy</td>
<td>88.92 ± 0.01</td>
<td>91.29 ± 0.01</td>
<td><strong>91.54 ± 0.01</strong></td>
<td>87.34 ± 0.01</td>
<td>87.58 ± 0.02</td>
</tr>
<tr>
<td>PCS</td>
<td>76.79 ± 3.32</td>
<td><strong>80.0 ± 2.27</strong></td>
<td>79.73 ± 2.02</td>
<td>70.57 ± 3.55</td>
<td>70.23 ± 3.04</td>
</tr>
</tbody>
</table>
HedgeScope: Automatic biomedical hedge scope detection algorithm

HedgeScope is also available as a Java API. Click here to go to the Java API download page.
Enter sentence to tag here:
For example motifs which occur in an incorrect cellular compartment, or outside the known taxonomic range, are unlikely to be functional as are those which are not conserved in closely related proteins or buried in a globular domain inaccessible for interaction.

Show feature selection panel

Sentence type:
- Biological
- Clinical
- Unknown

Submit

Result (scope shown in bold; if none of the words are bold, then there was no hedging detected by the algorithm):
For example motifs which occur in an incorrect cellular compartment, or outside the known taxonomic range, are unlikely to be functional as are those which are not conserved in closely related proteins or buried in a globular domain inaccessible for interaction.

http://snake.ims.uwm.edu/hedgescope/index.php
Scope resolution: Hedges

CoNLL-2010 Shared Task
Learning to detect hedges and their scope in natural language

| Introduction | FAQ | Task definitions | Download | Results | Program | Organise |

**Task 1** Learning to detect sentences containing uncertainty: identify sentences in texts which contain unreliable or uncertain information

- Task1B: Biological abstracts and full articles
- Task1W: Wikipedia paragraphs

**Task 2** Learning to resolve the in-sentence scope of hedge cues: in–sentence scope resolvers have to be developed

- Biological abstracts and full articles

Scope resolution: Hedges

Approaches (Table from Farkas et al. 2010)

<table>
<thead>
<tr>
<th>NAME</th>
<th>approach</th>
<th>scope</th>
<th>ML</th>
<th>postproc</th>
<th>tree</th>
<th>dep</th>
<th>multihedge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fernandes</td>
<td>TC</td>
<td>FL</td>
<td>ETL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ji</td>
<td>TC</td>
<td>I</td>
<td>AP</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Kilicoglu</td>
<td>HC</td>
<td>man</td>
<td>manual</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Li</td>
<td>SL</td>
<td>FL</td>
<td>CRF, SVMHMM</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morante</td>
<td>TC</td>
<td>FL</td>
<td>KNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rei</td>
<td>SL</td>
<td>FIL</td>
<td>manual+CRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Täckström</td>
<td>TC</td>
<td>FI</td>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velldal</td>
<td>HC</td>
<td>man</td>
<td>manual</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Vlachos</td>
<td>TC</td>
<td>I</td>
<td>Bayesian MaxEnt</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>SL</td>
<td>FIL</td>
<td>CRF</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: System architectures overview for Task2. Approaches: sequence labeling (SL), token classification (TC), hand-crafted rules (HC); Machine learners: Entropy Guided Transformation Learning (ETL), Averaged Perceptron (AP), k-nearest neighbour (KNN); The way of identifying scopes: predicting first/last tokens (FL), first/inside/last tokens (FIL), just inside tokens (I); Multiple Hedges: the system applied a mechanism for handling multiple hedges inside a sentence.
Scope resolution: Hedges

Evaluation

**Task 1**
- Sentence level
- $F_1$ of the uncertain class

**Task 2**
- A scope-level $F_1$ measure
- True positives were scopes which exactly matched the gold standard cue phrases and gold standard scope boundaries assigned to the cue word.
- Extract match: including or excluding punctuations, citations or some bracketed expressions
Datasets

**Biological data**
- **Train:** BioScope corpus (abstracts from Genia corpus, 5 full articles from functional genomics literature, 4 articles from BMC Bioinformatics.
- **Test:** 15 biomedical articles from PubMedCentral

**Wikipedia data**
- **Train:** 2186 paragraphs (11111 sentences)
- **Test:** 2346 paragraphs (9634 sentences total, of which 2234 uncertain)
Scope resolution: Hedges

Datasets

<table>
<thead>
<tr>
<th></th>
<th>Abstracts</th>
<th>Papers</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Documents</td>
<td>1273</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>#Sentences</td>
<td>11871</td>
<td>2670</td>
<td>5003</td>
</tr>
<tr>
<td>%Hedge sent.</td>
<td>17.70</td>
<td>19.44</td>
<td>15.75</td>
</tr>
<tr>
<td>#Hedges</td>
<td>2694</td>
<td>682</td>
<td>1043</td>
</tr>
<tr>
<td>#AvL. of sent.</td>
<td>30.43</td>
<td>27.95</td>
<td>31.30</td>
</tr>
<tr>
<td>#AvL. of scopes</td>
<td>17.27</td>
<td>14.17</td>
<td>17.51</td>
</tr>
</tbody>
</table>

Table 1: The detailed information of BioScope corpus. "AvL." stands for average length.

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Documents</td>
<td>2186</td>
<td>2737</td>
</tr>
<tr>
<td>#Sentences</td>
<td>11111</td>
<td>9634</td>
</tr>
<tr>
<td>%Hedge sentences</td>
<td>22.36</td>
<td>23.19</td>
</tr>
<tr>
<td>#Hedges</td>
<td>3133</td>
<td>3143</td>
</tr>
<tr>
<td>#AvL. of sentences</td>
<td>23.07</td>
<td>20.82</td>
</tr>
</tbody>
</table>

Table 2: The detail information of Wikipedia corpus. "AvL." stands for average length.
## Scope resolution: Hedges

### Results Task 2 cues (Table from Farkas et al. 2010)

<table>
<thead>
<tr>
<th>Name</th>
<th>P / R / F</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tang</td>
<td>85.0 / 87.7 / 86.4</td>
<td>C</td>
</tr>
<tr>
<td>Zhou</td>
<td>86.5 / 85.1 / 85.8</td>
<td>C</td>
</tr>
<tr>
<td>Li</td>
<td>90.4 / 81.0 / 85.4</td>
<td>C</td>
</tr>
<tr>
<td>Velldal</td>
<td>85.5 / 84.9 / 85.2</td>
<td>C</td>
</tr>
<tr>
<td>Vlachos</td>
<td>85.5 / 84.9 / 85.2</td>
<td>C</td>
</tr>
<tr>
<td>Täckström</td>
<td>87.1 / 83.4 / 85.2</td>
<td>C</td>
</tr>
<tr>
<td>Shimizu</td>
<td>88.1 / 82.3 / 85.1</td>
<td>C</td>
</tr>
<tr>
<td>Zhao</td>
<td>83.4 / 84.8 / 84.1</td>
<td>X</td>
</tr>
<tr>
<td>Ö zgür</td>
<td>77.8 / 91.3 / 84.0</td>
<td>C</td>
</tr>
<tr>
<td>Rei</td>
<td>83.8 / 84.2 / 84.0</td>
<td>C</td>
</tr>
<tr>
<td>Zhang</td>
<td>82.6 / 84.7 / 83.6</td>
<td>C</td>
</tr>
<tr>
<td>Kilicoglu</td>
<td>92.1 / 74.9 / 82.6</td>
<td>O</td>
</tr>
<tr>
<td>Morante</td>
<td>80.5 / 83.3 / 81.9</td>
<td>X</td>
</tr>
<tr>
<td>Morante</td>
<td>81.1 / 82.3 / 81.7</td>
<td>C</td>
</tr>
<tr>
<td>Zheng</td>
<td>73.3 / 90.8 / 81.1</td>
<td>C</td>
</tr>
<tr>
<td>Tjong Kim Sang</td>
<td>74.3 / 87.1 / 80.2</td>
<td>C</td>
</tr>
<tr>
<td>Clausen</td>
<td>79.3 / 80.6 / 80.0</td>
<td>C</td>
</tr>
<tr>
<td>Szidarovszky</td>
<td>70.3 / 91.0 / 79.3</td>
<td>C</td>
</tr>
<tr>
<td>Georgescul</td>
<td>69.1 / 91.0 / 78.5</td>
<td>C</td>
</tr>
<tr>
<td>Zhao</td>
<td>71.0 / 86.6 / 78.0</td>
<td>C</td>
</tr>
<tr>
<td>Ji</td>
<td>79.4 / 76.3 / 77.9</td>
<td>C</td>
</tr>
<tr>
<td>Chen</td>
<td>74.9 / 79.1 / 76.9</td>
<td>C</td>
</tr>
<tr>
<td>Fernandes</td>
<td>70.1 / 71.1 / 70.6</td>
<td>C</td>
</tr>
<tr>
<td>Prabhakaran</td>
<td>67.5 / 19.5 / 30.3</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 3: Task1 biological results (type ∈ \{Closed(C), Cross(X), Open(O)\}).
Scope resolution: Hedges

Results Task 2 scopes (Table from Farkas et al. 2010)

<table>
<thead>
<tr>
<th>Name</th>
<th>P</th>
<th>R</th>
<th>F</th>
<th>type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morante</td>
<td>59.6</td>
<td>55.2</td>
<td>57.3</td>
<td>C</td>
</tr>
<tr>
<td>Rei</td>
<td>56.7</td>
<td>54.6</td>
<td>55.6</td>
<td>C</td>
</tr>
<tr>
<td>Velldal</td>
<td>56.7</td>
<td>54.0</td>
<td>55.3</td>
<td>C</td>
</tr>
<tr>
<td>Kilicoglu</td>
<td>62.5</td>
<td>49.5</td>
<td>55.2</td>
<td>O</td>
</tr>
<tr>
<td>Li</td>
<td>57.4</td>
<td>47.9</td>
<td>52.2</td>
<td>C</td>
</tr>
<tr>
<td>Zhou</td>
<td>45.6</td>
<td>43.9</td>
<td>44.7</td>
<td>O</td>
</tr>
<tr>
<td>Zhou</td>
<td>45.3</td>
<td>43.6</td>
<td>44.4</td>
<td>C</td>
</tr>
<tr>
<td>Zhang</td>
<td>46.0</td>
<td>42.9</td>
<td>44.4</td>
<td>C</td>
</tr>
<tr>
<td>Fernandes</td>
<td>46.0</td>
<td>38.0</td>
<td>41.6</td>
<td>C</td>
</tr>
<tr>
<td>Vlachos</td>
<td>41.2</td>
<td>35.9</td>
<td>38.4</td>
<td>C</td>
</tr>
<tr>
<td>Zhao</td>
<td>34.8</td>
<td>41.0</td>
<td>37.7</td>
<td>C</td>
</tr>
<tr>
<td>Tang</td>
<td>34.5</td>
<td>31.8</td>
<td>33.1</td>
<td>C</td>
</tr>
<tr>
<td>Ji</td>
<td>21.9</td>
<td>17.2</td>
<td>19.3</td>
<td>C</td>
</tr>
<tr>
<td>Täckström</td>
<td>2.3</td>
<td>2.0</td>
<td>2.1</td>
<td>C</td>
</tr>
</tbody>
</table>

Table 2: Task2 results (type ∈ {Closed(C), Open(O)}).
## Scope resolution: Hedges

### Approaches (Table from Farkas et al. 2010)

<table>
<thead>
<tr>
<th>NAME</th>
<th>approach</th>
<th>scope</th>
<th>ML</th>
<th>postproc</th>
<th>tree</th>
<th>dep</th>
<th>multihedge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fernandes</td>
<td>TC</td>
<td>FL</td>
<td>ETL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ji</td>
<td>TC</td>
<td>I</td>
<td>AP</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Kilicoglu</td>
<td>HC</td>
<td>FL</td>
<td>manual</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Li</td>
<td>SL</td>
<td>FL</td>
<td>CRF, SVMHMM</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>Morante</td>
<td>TC</td>
<td>FL</td>
<td>KNN</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rei</td>
<td>SL</td>
<td>FIL</td>
<td>manual+CRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Täckström</td>
<td>TC</td>
<td>FI</td>
<td>SVM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velldal</td>
<td>HC</td>
<td>I</td>
<td>manual</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vlachos</td>
<td>TC</td>
<td>I</td>
<td>Bayesian MaxEnt</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang</td>
<td>SL</td>
<td>FIL</td>
<td>CRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhao</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou</td>
<td>SL</td>
<td>FL</td>
<td>CRF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7: System architectures overview for Task2. Approaches: sequence labeling (SL), token classification (TC), hand-crafted rules (HC); Machine learners: Entropy Guided Transformation Learning (ETL), Averaged Perceptron (AP), k-nearest neighbour (KNN); The way of identifying scopes: predicting first/last tokens (FL), first/inside/last tokens (FIL), just inside tokens (I); Multiple Hedges: the system applied a mechanism for handling multiple hedges inside a sentence.
Finding hedge cues in biomedical texts - best system


- CRF-based system
- Cascaded system: hedge detection $\rightarrow$ scope detection
- First-Last classification for scope
Scope resolution: Hedges

Figure 1: System architecture

(Figure from Tang et al. 2010)
Scope resolution: Hedges

Features hedge detection - first layer

- Word and Word Shape of the lemma
- Prefix and Suffix with length 3-5.
- Context of the lemma, POS and the chunk in the window [-2,2].
- Combined features lemma-chunk, lemma-POS of focus token and previous and next token
- The type of a chunk; the lemma and POS sequences of it
- Whether a token is a part of the pairs "neither ... nor" and "either ... or"
- From dictionary (training corpus): whether a token can possibly be classified into B cue, I cue or O cue; its lemma, POS and chunk tag for each possible case:
Features hedge detection - first layer

- Same as first layer
- The lemma and POS sequences of the hedge predicted by each classifier.
- The times of a token classified into B cue, I cue and O cue by the first two classifiers.
- Whether a token is the last token of the hedge predicted by each classifier
Features scope detection

- Same as first layer
- Word
- Context of the lemma, POS, the chunk, the hedge and the dependency relation in the window [-2,2].
- Combined features including \(L_0C_0, L_0H_0, L_0D_0, L_iP_0, P_iC_0, P_iH_0, C_iH_0, P_iD_0, C_iD_0\), where \(-1 \leq i \leq 1\). \(L\) denotes the lemma of a word, \(P\) denotes a POS, \(C\) denotes a chunk tag, \(H\) denotes a hedge tag and \(D\) denotes a dependency relation tag.
- The type of a chunk; the lemma and POS sequences of it
- The type of a hedge; the lemma, POS and chunk sequences of it
Features scope detection

- The lemma, POS, chunk, hedge and dependency relation sequences of 1st and 2nd dependency relation edges; the lemma, POS, chunk, hedge and dependency relation sequences of the path from a token to the root
- Whether there are hedges in the 1st, 2nd dependency relation edges or path from a token to the root
- The location of a token relative to the negation signal: previous the first hedge, in the first hedge, between two hedge cues, in the last hedge, post the last hedge
Scope resolution: Hedges

Postprocessing

- If a hedge is bracketed by a F scope and a L scope, its scope is formed by the tokens between them.
- If a hedge is only bracketed by a F scope, and there is no L scope in the sentence, search for the first possible word from the end of the sentence according to a dictionary, which extracted from the training corpus, and assign it as L scope.
The scope of the hedge is formed by the tokens between them.
- If a hedge is only bracketed by a F scope, and there are at least one L scope in the sentence, the last L scope is the L scope of the hedge, and its scope is formed by the tokens between them.
Scope resolution: Hedges

Postprocessing

- If a hedge is only bracketed by a L scope, and there is no F scope in the sentence, search for the first possible word from the beginning of the sentence to the hedge according to the dictionary, and assign it as F scope. The scope of the hedge is formed by the tokens between them.

- If a hedge is only bracketed by a L scope, and there are at least one F scope in the sentence, search for the first possible word from the hedge to the beginning of the sentence according to the dictionary, and think it as the F scope of the hedge. The scope of the hedge is formed by the tokens between them.

- If a hedge is bracketed by neither of them, remove it.
Scope resolution: Hedges

**Results hedge detection** (Table from Tang et al. 2010)

Algorithms: CRR++ and SVMLight

<table>
<thead>
<tr>
<th>Corpus</th>
<th>System</th>
<th>Prec.</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioScope</td>
<td>CRF</td>
<td>87.12</td>
<td>86.46</td>
<td>86.79</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>85.24</td>
<td>87.72</td>
<td>86.46</td>
</tr>
<tr>
<td></td>
<td>CAS</td>
<td>85.03</td>
<td>87.72</td>
<td>86.36</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>CRF</td>
<td>86.10</td>
<td>35.77</td>
<td>50.54</td>
</tr>
<tr>
<td></td>
<td>LM</td>
<td>82.28</td>
<td>41.36</td>
<td>55.05</td>
</tr>
<tr>
<td></td>
<td>CAS</td>
<td>82.28</td>
<td>41.36</td>
<td>55.05</td>
</tr>
</tbody>
</table>

Table 3: In-sentence performance of the hedge detection subsystem for in-domain test. ”Prec.” stands for precision, ”LM” stands for large margin, and ”CAS” stands for cascaded system.
Results hedge detection (Table from Tang et al. 2010)
Algorithm: CRF++

<table>
<thead>
<tr>
<th>HD subsystem</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>gold</td>
<td>57.92</td>
<td>55.95</td>
<td>56.92</td>
</tr>
<tr>
<td>CRF</td>
<td>52.36</td>
<td>48.40</td>
<td>50.30</td>
</tr>
<tr>
<td>LM</td>
<td>51.06</td>
<td>48.89</td>
<td>49.95</td>
</tr>
<tr>
<td>CAS</td>
<td>50.96</td>
<td>48.98</td>
<td>49.95</td>
</tr>
</tbody>
</table>

Table 6: Results of the hedge scope in-sentence. ”HD” stands for hedge detection subsystem we used, ”LM” stands for large margin, and ”CAS” stands for cascaded system.
Finding the scopes of hedge cues in biomedical texts - best system


- Memory-based learning
- Features from dependency trees
A sentence where scopes should be found

The conservation from Drosophila to mammals of these two structurally distinct but functionally similar E3 ubiquitin ligases is likely to reflect a combination of evolutionary advantages associated with: (i) specialized expression pattern, as evidenced by the cell-specific expression of the neur gene in sensory organ precursor cells [52]; (ii) specialized function, as suggested by the role of murine MIB in TNFα signaling [32]; (iii) regulation of protein stability, localization, and/or activity.
A sentence where scopes should be not found

For example, the word *may* in sentence 1 *indicates that* there is some uncertainty about the truth of the event, whilst the phrase *Our results show that* in 2) *indicates that* there is experimental evidence to back up the event described by encodes...
Scope resolution: Hedges

PREPROCESSING
GENIA Tagger
POS, Chunks, NE
CoNLL format

CLASSIFICATION

MBL CUE FINDER

MBL SCOPE FINDER

POSTPROCESSING
P-SCOPE: consecutive block of scope tokens
P-REF: end sentence references out of scope
Different version of system in Morante and Daelemans (2009)

- One classifier per task, instead of a metalearner combining three classifiers
- Features from the dependency tree instead of shallow features only
- Better treatment of multiword cues
- Postprocessing of references
Scope resolution: Hedges

Data are converted into the CoNLL format

<table>
<thead>
<tr>
<th>WORD</th>
<th>LEMMA</th>
<th>PoS</th>
<th>CHUNK</th>
<th>D</th>
<th>LABEL</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>The</td>
<td>The</td>
<td>DT</td>
<td>B-NP</td>
<td>3</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>structural</td>
<td>structural</td>
<td>JJ</td>
<td>I-NP</td>
<td>3</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>evidence</td>
<td>evidence</td>
<td>NN</td>
<td>I-NP</td>
<td>4</td>
<td>SUB</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>lends</td>
<td>lend</td>
<td>VBZ</td>
<td>B-VP</td>
<td>0</td>
<td>ROOT</td>
<td>B</td>
<td>F</td>
</tr>
<tr>
<td>strong</td>
<td>strong</td>
<td>JJ</td>
<td>B-NP</td>
<td>6</td>
<td>NMOD</td>
<td>I</td>
<td>O</td>
</tr>
<tr>
<td>support</td>
<td>support</td>
<td>NN</td>
<td>I-NP</td>
<td>4</td>
<td>OBJ</td>
<td>I</td>
<td>O</td>
</tr>
<tr>
<td>to</td>
<td>to</td>
<td>TO</td>
<td>B-PP</td>
<td>6</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>the</td>
<td>the</td>
<td>DT</td>
<td>B-NP</td>
<td>11</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>inferred</td>
<td>inferred</td>
<td>JJ</td>
<td>I-NP</td>
<td>11</td>
<td>NMOD</td>
<td>B</td>
<td>O</td>
</tr>
<tr>
<td>domain</td>
<td>domain</td>
<td>NN</td>
<td>I-NP</td>
<td>11</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>pair</td>
<td>pair</td>
<td>NN</td>
<td>I-NP</td>
<td>7</td>
<td>PMOD</td>
<td>O</td>
<td>L</td>
</tr>
<tr>
<td>,</td>
<td>,</td>
<td>O</td>
<td></td>
<td>4</td>
<td>P</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>resulting</td>
<td>result</td>
<td>VBG</td>
<td>B-VP</td>
<td>4</td>
<td>VMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>in</td>
<td>in</td>
<td>IN</td>
<td>B-PP</td>
<td>13</td>
<td>VMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>a</td>
<td>a</td>
<td>DT</td>
<td>B-NP</td>
<td>18</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>high</td>
<td>high</td>
<td>JJ</td>
<td>I-NP</td>
<td>18</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>confidence</td>
<td>confidence</td>
<td>NN</td>
<td>I-NP</td>
<td>18</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>set</td>
<td>set</td>
<td>NN</td>
<td>I-NP</td>
<td>14</td>
<td>PMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>of</td>
<td>of</td>
<td>IN</td>
<td>B-PP</td>
<td>18</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>domain</td>
<td>domain</td>
<td>NN</td>
<td>B-NP</td>
<td>21</td>
<td>NMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>pairs</td>
<td>pair</td>
<td>NNS</td>
<td>I-NP</td>
<td>19</td>
<td>PMOD</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td></td>
<td>O</td>
<td></td>
<td>4</td>
<td>P</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
**Evaluation of the conversion of the corpus into CoNLL format**

<table>
<thead>
<tr>
<th>TASK 1 (cues)</th>
<th>TASK 2 (scope)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIKI</td>
<td>BIO-ART</td>
</tr>
<tr>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>BIO-ART</td>
</tr>
<tr>
<td></td>
<td>99.10</td>
</tr>
</tbody>
</table>
Classification 1: cues

- Instances represent tokens
- BIO classification of tokens
- IGTree as implemented in TiMBL
- Features
  - Token
  - Token context in string of words and dependency tree
  - Lexicon of cues from training data
Scope resolution: Hedges

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO</td>
<td>81.15</td>
<td>82.28</td>
<td>81.71</td>
</tr>
<tr>
<td>BIO-cd</td>
<td>80.54</td>
<td>83.29</td>
<td>81.89</td>
</tr>
<tr>
<td>WIKI</td>
<td>80.55</td>
<td>44.49</td>
<td>57.32</td>
</tr>
<tr>
<td>WIKI-cd</td>
<td>80.64</td>
<td>44.94</td>
<td>57.71</td>
</tr>
</tbody>
</table>
Scope resolution: Hedges

Classification 2: scope

- An instance represents a pair of a predicted cue and a token
- Tokens are classified as being FIRST, LAST or none in scope sequence for as many cues as there are in the sentence
- IB1 as implemented in TiMBL
Features classification scope

- Features about cue, token, and their context in the string of words and in the dependency tree
- Features indicating whether token is candidate to be the FIRST and to be LAST
  - Values are assigned by a heuristics that takes into account detailed information from the dependency tree (voice of clause, PoS of cue, lemma of cue, etc.)
Postprocessing steps

- P-SCOPE builds a sequence of scope tokens based on 7 rules
  - Classifier predicts only FIRST and LAST element in the scope
- P-REF eliminates references from the scope at the end of clause and sentence
Scope resolution: Hedges

Results Task 2

<table>
<thead>
<tr>
<th></th>
<th>P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIO</td>
<td>59.62</td>
<td>55.18</td>
<td>57.32</td>
</tr>
<tr>
<td>BIO before P-REF</td>
<td>51.98</td>
<td>48.2</td>
<td>50.02</td>
</tr>
<tr>
<td>BIO before P-SCOPE</td>
<td>48.82</td>
<td>44.43</td>
<td>46.52</td>
</tr>
</tbody>
</table>
Scope resolution: Hedges

Error analysis

- **Task 1**: system fails to treat Or
  - BIO papers: 3 TP, 8 FP, 49 FN
- **Task 2**:
  - error propagation from Task 1
  - errors derived from incorrect dependency trees
  - errors derived from wrong encoding of features with dependency information
  - subordinate clauses are kept within the scope of cues in the main clause

- The test corpus contained a full paper with metalanguage
Scope resolution: Hedges

BiographTA: NeSp scope labeler

BiographTA
Text Analytics in the Biograph Project

NeSp demo
Finding negation and speculation cues and their scopes in biomedical texts.

The system accepts as input a text and it returns the text splitted into sentences, where negation and speculation cues and their scope are marked. It has been trained to process English biomedical texts.
BiographTA: NeSp scope labeler

\[ \text{neg0} \] The question \([\text{spec1} \ \text{whether}]\) Vif alters transcription controlled by the A3G promoter \([\text{spec1}]\) has \text{not} been analyzed so far \([\text{neg0}]\).

Our analysis \([\text{spec2} \ \text{indicates that}]\) transcription from the A3G promoter is unaffected by Vif or other HIV-1 proteins \([\text{spec2}]\).

Taken together, in T cell lines, the A3G promoter appears constitutively active.
Scope resolution: Hedges

Hybrid, two-level approach for hedge resolution,

- A statistical classifier (MaxEnt) detects cue words
- A small set of manually crafted rules operating over syntactic structures resolve scope

Syntactic information contributes to the resolution of in-sentence scope of hedge cues
Scope resolution: Hedges

Rules for scope resolution

- Input for rules: a parsed sentence which has been further tagged with hedge cues.
- Rules operate over the dependency structures and additional features provided by the parser (MaltParser)

Evaluation (Table from Ovrelid et al. 2010)

<table>
<thead>
<tr>
<th>Configuration</th>
<th>F₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default, Gold Cues</td>
<td>45.21</td>
</tr>
<tr>
<td>Rules, Gold Cues</td>
<td>72.31</td>
</tr>
<tr>
<td>Rules, System Cues</td>
<td>64.77</td>
</tr>
<tr>
<td>BSP</td>
<td></td>
</tr>
<tr>
<td>Rules, Gold Cues</td>
<td>66.73</td>
</tr>
<tr>
<td>Rules, System Cues</td>
<td>55.75</td>
</tr>
</tbody>
</table>

Default: scope from cue to end of sentence. BSE: evaluation on CoNLL test set
Outline

9 Detecting speculated sentences
10 Processing negation in biomedical texts
11 Scope resolution
12 Finding negated and speculated events
13 Modality tagging
14 Belief categorisation
15 Processing contradiction and contrast
16 Visualising negation features
17 References
Finding negated and speculated events: BioNLP ST 2009

BioNLP'09 Shared Task on Event Extraction
in conjunction with BioNLP, a NAACL-HLT 2009 workshop, June 4-5 2009, Boulder, Colorado

Task 3. Negation and speculation recognition (optional)

Participants are required to find negations and speculations regarding events extracted by Task 1.

- e.g.) TRADD did not interact with TES2
  -> (Negation (Type: Binding, Theme: TRADD, Theme: TES2))

http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA/SharedTask/

Results (From Kim et al. 2009)

<table>
<thead>
<tr>
<th>Team</th>
<th>Negation</th>
<th>Speculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ConcordU</td>
<td>14.98 / 50.75 / 23.13</td>
<td>16.83 / 50.72 / 25.27</td>
</tr>
<tr>
<td>VIBGhent</td>
<td>10.57 / 45.10 / 17.13</td>
<td>08.65 / 15.79 / 11.18</td>
</tr>
<tr>
<td>ASU+HU+BU</td>
<td>03.96 / 27.27 / 06.92</td>
<td>06.25 / 28.26 / 10.24</td>
</tr>
<tr>
<td>NICTA</td>
<td>05.29 / 34.48 / 09.17</td>
<td>04.81 / 30.30 / 08.30</td>
</tr>
<tr>
<td>USzeged</td>
<td>05.29 / 01.94 / 02.84</td>
<td>12.02 / 03.88 / 05.87</td>
</tr>
<tr>
<td>CCP-BTMG</td>
<td>01.76 / 05.26 / 02.64</td>
<td>06.73 / 13.33 / 08.95</td>
</tr>
</tbody>
</table>
Outline

9 Detecting speculated sentences
10 Processing negation in biomedical texts
11 Scope resolution
12 Finding negated and speculated events
13 Modality tagging
14 Belief categorisation
15 Processing contradiction and contrast
16 Visualising negation features
17 References
Modality tagging


- **Modality tagger**: produces text or structured text in which modality triggers and/or targets are identified
- **Two modality taggers**:
  - String-based English tagger
  - Structure-based English tagger
Modality tagging

String-based modality tagger

- Input: text with POS tags from a Collins-style statistical parser
- Marks spans of words/phrases that exactly match modality trigger words in the modality lexicon
- Identifies the target of each modality using the heuristic of tagging the next non-auxiliary verb to the right of the trigger
Structure-based modality tagger

- Input: text that has been parsed
- The parsed sentences are processed by TSurgeon rules
- TSurgeon rules:
  - Pattern: matches part of a parse tree
    Finds a modality trigger word and its target
  - Action: alters the parse tree
    Inserts tags such as TrigRequire and TargRequire for triggers and targets
    for the modality Require
Modality tagging

Output from structure-based modality tagger (Figure from Baker et al. 2010)

(TOP
  (S
    (NP
      (NNP Pakistan)
      (SBAR (WDT which)
        (S (MD TrigAble could)
          (RB TrigNegation not)
          (VB B TargAble TrigSucceed
            TargNegation reach)
          (ADJP
            (JJ TargSucceed semi-final))
          (, ,)
          (PP (IN in) (DT a)
            (NN match) (PP (IN against)
              (ADJP (JJ South) (JJ African))
              (NN team))
            (PP (IN for) (DT the)
              (JJ fifth) (NN position))
            (NP (NNP Pakistan))))))
    (VB D defeated)
    (NP (NNP South) (NNP Africa))
    (PP (IN by) (CD 41) (NNS runs)) (.) )
  )
)
Evaluation

- Agreement between taggers (Kappa)
  - 0.82 for triggers
  - 0.76 for targets

- Precision of structure-based tagger on 249 sentences: 86.3 %
Errors

- Light verbs tagged as semantic target
  The decision **should** be **taken** on delayed cases on the basis of merit
  “Decision” should have been marked

- Wrong word sense
  Sikhs attacked a train
  Attack is not used in the sense of ‘try’ (e.g. attack the problem)

- Coordinate structures

- Non-heads of compound nouns tagged as target, instead of head
Detecting speculated sentences
Processing negation in biomedical texts
Scope resolution
Finding negated and speculated events
Modality tagging
Belief categorisation
Processing contradiction and contrast
Visualising negation features
References
Belief categorisation: Committed belief tagging


- “We need to abandon a simple view of text as a repository of propositions about the world”
- “the result of text processing is not a list of facts about the world, but a list of facts about different people’s cognitive states”
- **Goal**: to recognize what the writer of the text intends the reader to believe about various people’s beliefs about the world (including the writer’s own)
- To determine which propositions he or she intends us to believe he or she holds as beliefs, and with what strength
Belief categorisation: Committed belief tagging

Corpus (Diab et al. 2009)

- 10,000 words annotated for speaker belief of stated propositions. Each verbal proposition is annotated with the tags:
  - **Committed belief** (CB): the writer indicates in this utterance that he or she believes the proposition
    *We know that GM has laid off workers*
  - **Non-committed belief** (NCB): the writer identifies the proposition as something which he or she could believe, but he or she happens not to have a strong belief in
    *GM may lay off workers*
  - **Not applicable** (NA): for the writer, the proposition is not of the type in which he or she is expressing a belief, or could express a belief
    - Expressions of desire: *Some wish GM would lay off workers*
    - Questions: *Will GM lay off workers?*
    - Expressions of requirements: *GM is required to lay off workers*
Belief categorisation: Committed belief tagging

Experiments

- Algorithms:
  - SVM, YAMCHA (Kudo and Matsumoto, 2000) sequence labeling system
  - CRF implementation of the MALLET toolkit (McCallum, 2002)
- Features: syntactic and lexical
- Models
  - Joint: a four-way classification task where each token is tagged as one of four classes – CB, NCB, NA, or O
  - Pipeline:
    1. Identifying the propositions
    2. Classifying each proposition as CB, NCB, or NA
- Evaluation: 4-fold cv
Belief categorisation: Committed belief tagging

Results (Table from Prabhakaran et al. 2010)

<table>
<thead>
<tr>
<th>Class</th>
<th>Feature Set</th>
<th>Parm</th>
<th>P</th>
<th>R</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>YAMCHA - Joint Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_C$</td>
<td>POS, whichModalAmI, verbType, isNumeric</td>
<td>CW=3</td>
<td>61.9</td>
<td>52.7</td>
<td>56.9</td>
</tr>
<tr>
<td>$L_N S_N$</td>
<td>POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModal-</td>
<td>CW=0</td>
<td>62.5</td>
<td>57.5</td>
<td>59.9</td>
</tr>
<tr>
<td></td>
<td>MyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, amVBwith-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DaughterTo, haveDaughterWh, haveDaughterShould</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_C S_N$</td>
<td>POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModal-</td>
<td>CW=2</td>
<td>67.4</td>
<td>58.1</td>
<td>62.4</td>
</tr>
<tr>
<td></td>
<td>MyDaughter, whichAuxIsMyDaughter, haveDaughterShould</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_C S_C$</td>
<td>POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModal-</td>
<td>CW=2</td>
<td>68.5</td>
<td>60.0</td>
<td>64.0</td>
</tr>
<tr>
<td></td>
<td>MyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughter-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wh, haveDaughterShould</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>MALLETT - Joint Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L$</td>
<td>POS, whichModalAmI, verbType</td>
<td>GV=1</td>
<td>55.1</td>
<td>45.0</td>
<td>49.6</td>
</tr>
<tr>
<td>$L S$</td>
<td>POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModal-</td>
<td>GV=1</td>
<td>64.5</td>
<td>54.4</td>
<td>59.0</td>
</tr>
<tr>
<td></td>
<td>MyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughter-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wh, haveDaughterShould</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Pipeline Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_C S_C$</td>
<td>POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModal-</td>
<td>CW=2</td>
<td>49.8</td>
<td>42.9</td>
<td>46.1</td>
</tr>
<tr>
<td></td>
<td>MyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughter-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wh, haveDaughterShould</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Overall Results. CW = Context Width, GV = Gaussian Variance, P = Precision, R = Recall, F = F-Measure
Belief categorisation: Committed belief tagging

Features that were useful (Table from Prabhakaran et al. 2010)

<table>
<thead>
<tr>
<th>Features that performed well</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. isNumeric</td>
<td>L</td>
</tr>
<tr>
<td>2. POS</td>
<td>L</td>
</tr>
<tr>
<td>3. verbType</td>
<td>L</td>
</tr>
<tr>
<td>4. whichModalAmI</td>
<td>L</td>
</tr>
<tr>
<td>3. amVBwithDaughterTo</td>
<td>S</td>
</tr>
<tr>
<td>4. haveDaughterPerfect</td>
<td>S</td>
</tr>
<tr>
<td>5. haveDaughterShould</td>
<td>S</td>
</tr>
<tr>
<td>6. haveDaughterWh</td>
<td>S</td>
</tr>
<tr>
<td>7. haveReportingAncestor</td>
<td>S</td>
</tr>
<tr>
<td>8. parentPOS</td>
<td>S</td>
</tr>
<tr>
<td>9. whichAuxIsMyDaughter</td>
<td>S</td>
</tr>
<tr>
<td>10. whichModalIsMyDaughter</td>
<td>S</td>
</tr>
</tbody>
</table>

Word is Alphabet or Numeric?
Word’s POS tag
Modal/Aux/Reg ( = ’nil’ if the word is not a verb)
If I am a modal, what am I? ( = ’nil’ if I am not a modal)
Am I a VB with a daughter to?
Do I have a daughter which is one of has, have, had?
Do I have a daughter should?
Do I have a daughter who is one of where, when, while, who, why?
Am I a verb/predicate with an ancestor whose lemma is one of tell, accuse, insist, seem, believe, say, find, conclude, claim, trust, think, suspect, doubt, suppose?
What is my parent’s POS tag?
If I have a daughter which is an auxiliary, what is it? ( = ’nil’ if I do not have an auxiliary daughter)
If I have a daughter which is a modal, what is it? ( = ’nil’ if I do not have a modal daughter)
### Features that were not useful (Table from Prabhakaran et al. 2010)

<table>
<thead>
<tr>
<th></th>
<th>Feature</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lemma</td>
<td>L</td>
<td>Word's Lemma</td>
</tr>
<tr>
<td>2</td>
<td>Stem</td>
<td>L</td>
<td>Word stem (Using Porter Stemmer)</td>
</tr>
<tr>
<td>3</td>
<td>Drole</td>
<td>S</td>
<td>Deep role (drole in MICA features)</td>
</tr>
<tr>
<td>4</td>
<td>isRoot</td>
<td>S</td>
<td>Is the word the root of the MICA Parse tree?</td>
</tr>
<tr>
<td>5</td>
<td>parentLemma</td>
<td>S</td>
<td>Parent word’s Lemma</td>
</tr>
<tr>
<td>6</td>
<td>parentStem</td>
<td>S</td>
<td>Parent word stem (Using Porter Stemmer)</td>
</tr>
<tr>
<td>7</td>
<td>parentSupertag</td>
<td>S</td>
<td>Parent word’s super tag (from Penn Treebank)</td>
</tr>
<tr>
<td>8</td>
<td>Pred</td>
<td>S</td>
<td>Is the word a predicate? (pred in MICA features)</td>
</tr>
<tr>
<td>9</td>
<td>wordSupertag</td>
<td>S</td>
<td>Word’s Super Tag (from Penn Treebank)</td>
</tr>
</tbody>
</table>
Some conclusions YAMCHA

- Syntactic features improve the classifier performance
- Syntactic features with no context improve Recall by 4.8 % over only lexical features with context
- Adding back context to lexical features further improves Precision by 4.9 %
- Adding context of syntactic features improves both Precision and Recall
- NCB performs much worse than the other two categories
Detecting speculated sentences

Processing negation in biomedical texts

Scope resolution

Finding negated and speculated events

Modality tagging

Belief categorisation

Processing contradiction and contrast

Visualising negation features

References
Contradictions occur whenever information that is communicated in two different texts is incompatible.

- Framework for recognizing contradictions between multiple text sources by relying on three forms of linguistic information:
  - negation
  - antonymy
  - semantic and pragmatic information associated with the discourse relations

- Contradictions need to be recognized by QA systems or by Multi-Document Summarization (MDS) systems.
## Processing contradiction and contrast

<table>
<thead>
<tr>
<th>Contradiction</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
</tr>
<tr>
<td>T1: Joachim Johansson held off a dramatic fightback from defending champion Andy Roddick, to reach the semi-finals of the US Open on Thursday night.</td>
</tr>
<tr>
<td>T2: Defending champion Andy Roddick <em>never</em> took on Joachim Johansson.</td>
</tr>
<tr>
<td>(b)</td>
</tr>
<tr>
<td>T3: In California, one hundred twenty Central Americans, due to be deported, began a hunger strike when their deportation was delayed.</td>
</tr>
<tr>
<td>T4: A hunger strike was <em>called off</em>.</td>
</tr>
<tr>
<td>(c)</td>
</tr>
<tr>
<td>T5: The explosion wounded the arm of Beatriz Iero, damaged the doors and walls of the offices, and broke the windows of neighboring buildings.</td>
</tr>
<tr>
<td>T6: Beatriz Iero <em>emerged unscathed</em> from an explosion.</td>
</tr>
</tbody>
</table>

(From Harabagiu et al. 2006)
The recognition of contradictions is useful to fusion operators, that consider information originating in different texts.

- When contradictory information is discovered, the answer selects information from only one of the texts, discarding its contradiction.

**Question:** When did Pakistan test its Shaheen-2 ballistic missile?

**Answer**₁: The source noted that the Shaheen-2, with a range of 2400 km, has never been tested by Pakistan.

**Answer**₂: Pakistan has said that it performed several tests of its 2300 km-range Shaheen-2 missile in September 2004.
Two views for contradiction detection:

**View 1** Contradictions are recognized by identifying and removing negations of propositions and then testing for textual entailment

**View 2** Contradictions are recognized by deriving linguistic information from the text inputs, including information that identifies negations, contrasts, or oppositions and by training a classifier based on examples
System architecture (From Harabagiu et al. 2006)

Figure 3: The Architecture Used for Recognizing Contradictions with the Help of Textual Entailment.
Processing contradiction and contrast

- **Types of negation detected**
  - Overt negation
    - the morpheme n’t and not
    - negative quantifiers like no (also “no one” and “nothing”)
    - strong negative adverbs like “never”
  - Indirectly licensed negation.
    - verbs ( “deny”, “fail”, “refuse”, “keep from” )
    - prepositions ( “without”, “except” )
    - weak quantifiers ( “few”, “any”, “some” )
    - traditional negative polarity items such as “a red cent” or “any more”

- **Types of negated constituents**: events, states and entities
Negation detection steps

1. Preprocessing: negation markers are flagged
2. Detect negated events: filter out events without predicates marked as negated
   An predicate is negated if it falls within the scope of a negative marker
3. Detect negated entities: any noun phrase that falls within the scope of an overt negative quantifier (“no”) or a non-veridical quantifier (“few, some, many”)
4. Detect negated states:
   ▶ Detect states based on WordNet
   ▶ A state is negated if it falls within the scope of a negative marker

The system eliminates negations and reverts the polarity of negated events, entities and states by using antonyms and paraphrases
Protein and interaction databases have been compiled from experimental data and published literature. However, the information captured in these resources are typically individual positive facts of the kind such as ‘protein A binds to protein B’. Kim et al. (2006) extract contrastive information between proteins from the biomedical literature to augment the information in current protein databases.
Contrasts are effective conceptual vehicles for learning processes such as correcting, highlighting, contrasting, and grouping central concepts. Thus, they are useful for exploring the unknown. They can provide much invaluable insights and explanations about the observed phenomena. For example, contrasts between proteins in terms of their biological interactions can reveal what similarities, divergences, and relations there are of the proteins, leading to additional useful insights about the underlying functional nature of the proteins.
With the **BioContrast database** users can

- Search for contrasts of proteins of interest with their Swiss-Prot IDs or names
- Browse and navigate networks of protein–protein contrasts graphically
- Search for contrasts that are associated with KEGG pathways, InterPro domain entries, and Gene Ontology concepts, which may be useful for enhancement of KEGG pathway, inference over contrasts between protein domains, and subcategorization of Gene Ontology concepts.
Processing contradiction and contrast

NAT1 binds eIF4A but not eIF4E and inhibits both cap-dependent and cap-independent translation (PMID: 90306851).

Truncated N-terminal mutant huntingtin repressed transcription, whereas the corresponding wild-type fragment did not repress transcription (PMID: 11739372).

Parts:

1. **Focused objects**: a contrastive pair of two or more objects that are so contrasted (e.g. eIF4A, eIF4E, wild-type huntingtin, mutant huntingtin)

2. **Presupposed property**: a biological property or process that the contrast is based on (e.g. binding to NAT1, transcription repression).
A protein-protein contrast is a contrast between two proteins A and B, called as “focused proteins”, which indicates that A but not B is involved in a biological property C, called as “presupposed property”, or vice versa.

- Contrast information is often encoded by contrastive negation patterns such as “A but not B” in the biomedical literature.
- Such contrast:
  - explicitly describes a difference between focused proteins in terms of its presupposed property
  - implicitly indicates that the focused proteins are semantically similar

This combination of difference and similarity between proteins is useful for augmenting proteomics databases and also for discovering novel knowledge.
Extracting contrastive relations

Given a MEDLINE abstract:

1. The system first locates sentences that contain the negative ‘not’.
2. It then identifies contrastive expressions from these sentences using either subclausal coordination or clause level parallelism.
3. If the contrastive expressions are, or can be reduced to, protein names, the system produces a contrast between the two proteins.
4. It then cross-links (i.e. grounds) the contrastive protein names with entries of a standard protein database (namely, Swiss-Prot).

The net result is a database of useful biological contrastive relations between actual Swiss-Prot entries.
Processing contradiction and contrast

Extracting contrastive relations

1. Input Sentence: NAT1 binds eIF4A but not eIF4E and inhibits both cap-dependent and cap-independent translation.
   - Find ‘not’ and a coordinating conjunction

2. Input Sentence: NAT1 binds eIF4A but not eIF4E and inhibits both cap-dependent and cap-independent translation.
   - Find candidates for the conjuncts

3. Input Sentence: NAT1 binds eIF4A but not eIF4E and inhibits both cap-dependent and cap-independent translation.
   - Extract contrastive information

4. Presupposed Property: NAT1 binds X
   - Focused Protein Names: X = eIF4A, X ≠ eIF4E

5. Ground protein names

6. Presupposed Property: NAT1 binds X
   - Grounded Protein Entry: X = IF4A_HUMAN, X ≠ IF4E_HUMAN
Immunohistochemical studies revealed that HAI-1 but not HAI-2 was detected more strongly in regenerative epithelium than in normal epithelium, although both proteins were detected throughout the human gastrointestinal tract.

Positive: HAI-1
PosSprot: SPIT1_HUMAN: “HAI-1”
Negative: HAI-2
NegSprot: SPIT2_HUMAN: “HAI-2”
Property: CONTRAST_OBJ was detected more strongly in regenerative epithelium than in normal epithelium
Identifying subclausal coordinations

1. Given a sentence that contains a ‘not’, the system first tries to identify contrastive expressions using subclausal coordination patterns.

<table>
<thead>
<tr>
<th>General patterns</th>
<th>Specific patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>A but not B</td>
<td>NP but not NP</td>
</tr>
<tr>
<td>V NP but not V NP</td>
<td>V PREP NP but not V PREP NP</td>
</tr>
<tr>
<td>not A but B</td>
<td>not NP but NP</td>
</tr>
<tr>
<td>not V NP but V NP</td>
<td>not V PREP NP but V PREP NP</td>
</tr>
<tr>
<td>A, not B</td>
<td>NP, not NP</td>
</tr>
<tr>
<td></td>
<td>PREP NP, not PREP NP</td>
</tr>
</tbody>
</table>

A and B denote the pair of focused objects in a general subclausal coordination pattern. NP indicates a noun phrase, PREP a preposition, V a verb, and ADJ an adjective.

2. The system analyzes the word-level similarity by checking whether the variable-matching phrases are semantically identical or at least in a subsumption relation.
In contrast, IFN-gamma priming did not affect the expression of p105 transcripts but enhanced the expression of p65 mRNA.

Matching pattern: ‘not V NP but V NP’

1. Match the V variables to the verbs ‘affect’ and ‘enhanced’
2. Match the NP variables to the noun phrases ‘the expression of p105 transcripts’ and ‘the expression of p65 mRNA’
3. Analyze the similarity between the verbs and the similarity between the noun phrases
   - Synonymy and hypernymy relations in WordNet for verbs and adjectives
   - Biomedical databases, WordNet and own resource
4. Determine the presupposed property for the focused proteins by extracting the subject phrase and the verb whose object phrases correspond to the focused proteins
Processing contradiction and contrast

Identifying clause-level parallelisms

The system checks whether

1. the linguistic expressions that match the variables with the same subscript (e.g. \{V_1, V'_1\}) are either semantically identical (e.g. \{‘repress’, ‘repressed’\}) or are in a subsumption relation (e.g. \{‘affect’, ‘activate’\})

2. the variables with the subscript ‘C’ (e.g. \{Subj_C, Subj'_C\}), which indicate focused objects of the pattern, are matched to semantically similar expressions (e.g. \{‘eIF4A’, ‘eIF4E’\}).

<table>
<thead>
<tr>
<th>Negative patterns</th>
<th>Positive patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj_C not V_1 Obj_2</td>
<td>Subj'_C V_1 Obj'_2</td>
</tr>
<tr>
<td>Subj_2 not V_1 Obj_C</td>
<td>Subj'_2 V'_1 Obj'_C</td>
</tr>
<tr>
<td>Subj_C not V_1 PREP_3 Obj_2</td>
<td>Subj'_C V_1 PREP'_3 Obj'_2</td>
</tr>
<tr>
<td>Subj_2 not V_1 PREP_3 Obj_C</td>
<td>Subj'_2 V'_1 PREP'_3 Obj'_C</td>
</tr>
<tr>
<td>Subj_C BeV not ADJ_1 PREP_3 NP_2</td>
<td>Subj'_C BeV ADJ'_1 PREP'_3 NP'_2</td>
</tr>
<tr>
<td>Subj_2 BeV not ADJ_1 PREP_3 NP_C</td>
<td>Subj'_2 BeV ADJ'_1 PREP'_3 NP'_C</td>
</tr>
<tr>
<td>NN_1 of NP_C with NP_2 not V</td>
<td>N'_1 of NP'_C with NP'_2 V</td>
</tr>
<tr>
<td>NN_1 between NP_C and NP_2 not V</td>
<td>NN'_1 between NP'_C and NP'_2 V</td>
</tr>
</tbody>
</table>
Truncated N-terminal mutant huntingtin repressed transcription, whereas the corresponding wild-type fragment did **not** repress transcription.

1. Locate the verb ‘repress’ in the subordinate clause which is negated by ‘not’.
2. Locate the positive verb ‘repressed’ of the main clause.
3. Identify the corresponding subject phrases and the object phrases in the two clauses.
   - Subject phrase in the main clause = ‘Truncated N-terminal mutant huntingtin’
   - Object phrase = ‘transcription’
   - Subject phrase of the subordinate clause = ‘the corresponding wild-type fragment’,
   - Object phrase = ‘transcription’.
4. Check that the two verb phrases and the two object phrases are all semantically identical.

Contrastive relation extracted here is one between the two protein names at the corresponding subject positions with respect to the presupposed biological property of ‘CONTRAST OBJ repressed transcription’. 
Evaluation

- Processed data:
  - 2.5 million corpus from MEDLINE abstracts processed
  - 799169 pairs of contrastive expressions
  - 11284 pairs contrastive protein names
  - 41471 contrasts between Swiss-Prot entries (a protein maybe grounded with multiple Swiss-Prot entries)

- Test data:
  - 100 pairs of constrastive proteins examined
  - 97 % precision
  - 61.5 % recall from previous system
  - 91 contrastive patterns ‘A but not B’
  - 5 parallelism patterns (40% precision)
In the pathway for well-studied Huntington’s disease (HD) a key node in the pathway was labeled generically as ‘caspase’.

‘Caspase’ can be resolved as caspase-3 andor caspase-6.
Processing contradiction and contrast

KEGG Huntington’s disease pathway
Refining pathway roles of similar proteins

- A contrast between caspase-3 and caspase-6 is extracted by BioContrasts:

  Importantly, Mch2, but not Yama or LAP3, is capable of cleaving lamin A to its signature apoptotic fragment, indicating that Mch2 is an apoptotic laminase (PMID:8663580).

- It suggests that the two proteins may not function identically

- An article from MEDLINE explains the difference between the two proteins in terms of the cleavage sites at Htt:

  We have previously shown that Htt is cleaved in vitro by caspase-3 at amino acids 513 and 552, and by caspase-6 at amino-acid position 586 (PMID:10770929).
Outline

9 Detecting speculated sentences
10 Processing negation in biomedical texts
11 Scope resolution
12 Finding negated and speculated events
13 Modality tagging
14 Belief categorisation
15 Processing contradiction and contrast
16 Visualising negation features
17 References
Visualising negation features


“The major challenge in computational text analysis is the gap between automatically computable text features and the users’ ability to control and evaluate these features.”

- Application of document fingerprinting for visualizing text features as part of an interactive feedback loop between evaluation and feature engineering
- Based on Literature Fingerprint (Keim and Oelke 2007)
  - Documents are represented by a pixel-based visualization in which each pixel represents one unit of text
  - The color of each pixel is mapped to its feature value
  - The visualization takes the document structure into account
Pipeline for visual evaluation of text features applied for document summarization and analysis
Opinion mining experiments

- Classifying reviews of digital cameras as positive or negative
- Lexical approach: dictionary of negative and positive polarity words
- To get values on sentence level, for each sentence the number of negative words is subtracted from the number of positive words
Visualising negation features

The visualization has been annotated with comments on some of the wrongly classified statements. Figure from Oelke et al. (2008:78)
Error analysis

- Errors: negation is not taken into account, and nouns are not included in the list of opinion words
- Improvements:
  - Negation is taken into account by inverting the value of a word if one of the three preceding words is a negation signal word
  - Nouns with negative positive connotations are added to the list of opinion words
- Evaluate whether the extensions result in improvement
Visualising negation features

Visualising the effect of the extensions. Figure from Oelke et al. (2008:78)
Detecting Speculative Language


Processing Negation in Biomedical Texts


Resolving the Scope of Negation


Resolving the Scope of Hedges


References


All papers from the Proceedings of the CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text

Committed Belief Tagging


Contradiction


Part IV

Modality and Negation in Applications
Sentiment analysis

Recognizing textual entailment

Machine translation

Text mining

Identifying the structure of scientific articles

Trustworthiness detection

References
Outline

18 Sentiment analysis

19 Recognizing textual entailment

20 Machine translation

21 Text mining

22 Identifying the structure of scientific articles

23 Trustworthiness detection

24 References
Information sources


Sentiment analysis

- Types of textual information
  - Facts
  - Opinions

- Most current information processing systems work with factual information

- In 2001 a new research area emerged: sentiment analysis

- Why then?
  - Word-of-mouth on the web: the web contains huge amounts of opinionated text
  - User-generated media: one can express opinions on anything in forums, discussion groups, blogs, social networks, ...

(Slide adapted from B. Liu, Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences. Workshop on Social Theory and Social Computing 2010)
“A sizeable number of papers mentioning “sentiment analysis” focus on the specific application of classifying reviews as to their polarity (either positive or negative), a fact that appears to have caused some authors to suggest that the phrase refers specifically to this narrowly defined task. However, nowadays many construe the term more broadly to mean the computational treatment of opinion, sentiment, and subjectivity in text.”
Sentiment classification

- **Document level (Pang et al. 2002, Turney 2002):** classify a document as positive or negative based on the overall sentiment expressed by opinion holder.

- **Sentence level (Wiebe et al. 2004):** classify a sentence as
  - Objective or subjective
  - Having positive or negative polarity

- **Feature level (Hu and Liu 2004):** finding opinions related to features of objects.
“Sentiment analysis is not simply the problem of determining whether a document, a paragraph or even a sentence expresses a positive or negative sentiment or opinion. It is also about entities. Without such information, any sentiment is of little practical use. So one should not only talk about sentiment analysis of documents, paragraphs or sentences, but also about the entities that sentiments have been expressed upon. Here an entity can be a product, service, person, organisation, event or topic”

(Liu 2009, An Interview on Sentiment Analysis and Opinion Mining by textAnalyticsNews.com, April 20, 2009)
Negation in sentiment analysis

(Wiegand et al. 2010)

- Negation words can change the polarity of an expression:
  
  I like\(^+\) this new Nokia model – I do [not like\(^+\)]\(^-\) this new Nokia model

- Not all negation words change the polarity
  
  Not only is this phone expensive but it is also heavy and difficult to use

- The presence of an actual negation word in a sentence does not mean that all its polar opinions are inverted
  
  [I do [not like\(^+\)]\(^-\) the design of new Nokia model] but [it contains some intriguing\(^+\) new functions]

- Surface realization of negation is variable
  
  ▶ Diminishers/valence shifters:
    
    I find the functionality of the new phone less practical
  
  ▶ Connectives:
    
    Perhaps it is a great phone, but I fail to see why
  
  ▶ Modals:
    
    In theory, the phone should have worked even under water
Negation in sentiment analysis: computational models

Model: contextual valence shifting (Polanyi and Zaenen, 2004)

- The model assigns scores to polar expressions
- If a polar expression is negated, its polarity score is simply inverted
  clever (+2) ← not clever (-2)
- For diminishers, the score is only reduced rather than shifted to the other polarity type
  efficient (+2) ← rather efficient (+1)
Bag of words approach (Pang et al., 2002)

- Fairly effective
- The supervised classifier has to figure out by itself which words in the dataset are polar and which are not
- It does not contain any explicit knowledge of polar expressions
- Negation modeling: adding artificial words
  I do not NOT like NOT this NOT new NOT_Nokia NOT model
  increases the feature space with more sparse features
- The scope of negation cannot be properly modeled with this representation
- The impact of negation modeling on this level of representation is limited
Movie Review Data

This page is a distribution site for movie-review data for use in sentiment-analysis experiments. Available are collections of movie-review documents labeled with respect to their overall sentiment polarity (positive or negative) or subjective rating (e.g., "two and a half stars") and sentences labeled with respect to their subjectivity status (subjective or objective) or polarity. These data sets were introduced in the following papers:


We also have available an additional sentiment-analysis dataset, Congressional floor-debate transcripts, with support/oppose labels.

If you have results to report on these corpora, please send email to Bo Pang and/or Lillian Lee so we can add you to our list of other papers using this data. Thanks!

Please cite the version number of the dataset you used in any publications, in order to facilitate comparison of results. Thank you.

Sentiment polarity datasets

- polarity dataset v2.0 (3.0Mb) (includes README v2.0): 1000 positive and 1000 negative processed reviews. Introduced in Pang/Lee ACL 2004. Released June 2004.
- Pool of 27886 unprocessed html files (81.1Mb) from which the polarity dataset v2.0 was derived. (This file is identical to movie.zip from data release v1.0.)

archive:
- polarity dataset v1.0 (2.8Mb) (includes README): 700 positive and 700 negative processed reviews. Released July 2002.
- polarity dataset v1.1 (2.2Mb) (includes README,1,1): approximately 700 positive and 700 negative processed reviews. Released November 2002. This alternative version was created by Nathan Trefler, who removed a few non-English/incomplete reviews and changing some of the labels (judging some polarities to be different from the original author's rating). The complete list of changes made to v1.1 can be found in diff.txt.
- polarity dataset v0.9 (2.8Mb) (includes a README): 700 positive and 700 negative processed reviews. Introduced in Pang/Lee/Vaithyanathan EMNLP 2002. Released July 2002. Please read the "Rating Information - WARNING" section of the README.
- movie.zip (81.1Mb): all html files we collected from the IMDb archive.
Expression-level polarity classification (Wilson et al. 2005, 2009)

- Supervised machine learning where negation modeling is mostly encoded as features using polar expressions
- Three feature types (next slide)
- Adding these three feature groups to a feature set comprising bag of words and features counting polar expressions results in a significant improvement
Features

- **Negation features**
  - Check whether a negation expression occurs in a fixed window of four words preceding the polar expression
  - Does the polar predicate have a negated subject?
    
    \[ \text{[No politically prudent Israeli]}_{\text{subject}} \text{ could support}_{\text{polarpred}} \text{ either of them} \]
  - Negation expressions are additionally disambiguated

- **Shifter features**: binary features checking the presence of different types of polarity shifters (e.g. *little*)

- **Polarity modification features**: describe polar expressions of a particular type modifying or being modified by other polar expressions
Negation in sentiment analysis: benchmark

http://www.cs.pitt.edu/mpqa/

MPQA Releases - Corpus and Opinion Recognition System

MPQA Opinion Corpus
annotated for opinions and sentiments

The MPQA Opinion Corpus contains news articles from a wide variety of news sources manually annotated for opinions and other private states (i.e., beliefs, emotions, sentiments, speculations, etc.). The corpus was initially collected and annotated as part of the summer 2002 NRRC Workshop on Multi-Perspective Question Answering (MPQA) sponsored by ARDA. To learn more about the subjectivity and sentiment research that produced MFQA, please visit Dr. Janye Webe's page of related publications and the CERATOPS site.

To download the MPQA Opinion Corpus click here.

OpinionFinder

OpinionFinder is a system that processes documents and automatically identifies subjective sentences as well as various aspects of subjectivity within sentences, including agents who are sources of opinion, direct subjective expressions and speech events, and sentiment expressions. OpinionFinder was developed by researchers at the University of Pittsburgh, Cornell University, and the University of Utah.

In addition to OpinionFinder, we are also releasing the automatic annotations produced by running OpinionFinder on a subset of the Penn Treebank.

To go to the OpinionFinder download page click here.

Please note that OpinionFinder only runs on Linux.
Negation in sentiment analysis: benchmark

http://www.cs.pitt.edu/opinionfinderrelease/

OpinionFinder Release Page

OpinionFinder Available versions

LICENSE AGREEMENT
FAQ

Version 1.5

- README - OpinionFinder 1.5
- Request OpinionFinder 1.5

Version 1.4

- README - OpinionFinder 1.4
- Request OpinionFinder 1.4

OpinionFinder Sample Automatic Annotations

Penn Treebank

- README - Penn Treebank Automatic Opinion Annotations
- Request Penn Treebank Automatic Opinion Annotations

This research was supported in part by NSF Grants IIS-0208798 and IIS-0208985

References

Negation in sentiment analysis: shallow semantic composition

**Compositional semantics** (Choi and Cardie 2008)

- The polarity of a phrase can be computed in two steps:
  - The assessment of polarity of the constituents
  - The subsequent application of a set of previously defined inference rules.

Example of a rule:

\[
Polarity([NP1\_\neg \ [IN] \ [NP2\_\neg ] \ ) = +
\]

\[
[lack\_\neg _{NP1} \ [of]_{IN} \ [crime\_\neg _{NP2} \text{ in rural areas}]
\]

- They define syntactic contexts of the polar expressions
- From each context a direct polarity for the entire expression can be derived
- Advantage: they restrict the scope of negation to specific constituents rather than using the scope of the entire target expression
Negation in sentiment analysis: bad vs. not good

Polarity as a continuum (Liu and Seneff 2009)

- Not bad and good may have the same polarity but they differ in their respective polar strength, i.e. not bad is less positive than good
- Unifying account for intensifiers (e.g. very), diminishers, polarity shifters and negation words
  - Compositional rules for polar phrases, such as adverb-adjective or negation-adverb-adjective are defined exclusively using the scores of the individual words
  - Adverbs function like universal quantifiers scaling either up or down the polar strength of the specific polar adjectives they modify
- Polarity is treated compositionally and is interpreted as a continuum rather than a binary classification
Lexicon induction

The process of acquiring lexical resources that compile knowledge of which natural language expressions are polar

- The observation that negations co-occur with polar expressions has been used for inducing polarity lexicons on Chinese in an unsupervised manner (Zagibalov and Carroll, 2008)
- The model relies on the observation that a polar expression can be negated but it occurs more frequently without the negation.
  - The distributional behaviour of an expression, i.e. significantly often co-occurring with a negation word but significantly more often occurring without a negation word makes up a property of a polar expression.
Many polar expressions, such as *disease* are ambiguous

*He is a disease to every team he has gone to*

*Early symptoms of the disease are headaches, fevers, cold chills and body pain*

Some polar opinions are not lexicalized. World knowledge is needed

*The next time I hear this song on the radio, I’ll throw my radio out of the window*

The use of irony can reflect an implicit negation of what is conveyed through the literal use of the words (Carvalho et al. 2009)

A polarity classifier should also be able to decompose words and carry out negation modeling within words

*not-so-nice, anti-war or offensiveless*
Outline

18 Sentiment analysis

19 Recognizing textual entailment

20 Machine translation

21 Text mining

22 Identifying the structure of scientific articles

23 Trustworthiness detection

24 References
Recognizing textual entailment


- Machine learning system. Alignment is followed by a classification step.

- The system uses features from polarity and modality.
  - **Polarity features** “capture the presence (or absence) of linguistic markers of negative polarity contexts in both the text and the hypothesis, such as simple negation (not), downward-monotone quantifiers (no, few), restricting prepositions (without, except) and superlatives (tallest)”.
  - **Modality features** “capture simple patterns of modal reasoning”. The text and the hypothesis is mapped to one of six modalities: possible, not possible, actual, not actual, necessary, and not necessary.
  - **Factuality features**: a list of factive, implicative and non-factive verbs, clustered according to the kinds of entailments they create.

Snow et al. (2006) present a RTE system that incorporates negation and modality in order to recognize false entailment.

- The system checks whether nodes that are aligned in the hypothesis and text sentence have a negation or modality mismatch.
- If the mismatch exists, it is predicted that the entailment is false.
Recognizing textual entailment


Hickl and Bensley 2007: system that obtained the best absolute result in the RTE-3 challenge (80% accuracy)

- Based on identifying the set of publicly-expressed beliefs of the author (discourse commitments)
- A set of commitments are extracted from a text-hypothesis pair, so that the RTE task can be reduced to the identification of the commitments from a text that support the inference of the hypothesis.
- A discourse commitment represents any of the set of propositions that can be inferred to be true, given a conventional reading of the passage.
18 Sentiment analysis

19 Recognizing textual entailment

20 Machine translation

21 Text mining

22 Identifying the structure of scientific articles

23 Trustworthiness detection

24 References


- Measure the effect of modality tagging on the quality of machine translation output in Urdu-English MT.
  - Modality annotation: Bleu measure from 26.4 to 26.7
  - Modality + NE: from 26.4 to 26.9
Outline

18 Sentiment analysis

19 Recognizing textual entailment

20 Machine translation

21 Text mining

22 Identifying the structure of scientific articles

23 Trustworthiness detection

24 References
Negfinder is a rule-based system that recognizes a large set of negated patterns occurring in medical narrative.

Described in:

Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents:

A Quantitative Study Using the UMLS

PRADEEP G. MUTALIK, MD, ANIRUDDHA DESHPANDE, MD, PRakash M. NADKARNI, MD
ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports

Henk Harkema a,*, John N. Dowling a, Tyler Thornblade b, Wendy W. Chapman a

a Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA 15260, USA
b Department of Computer Science, University of Pittsburgh, Pittsburgh, PA 15260, USA

ConText determines whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient

ConText can be integrated with any application that indexes clinical conditions from text
Text mining: BioCaster


A disease outbreak report text mining system

- The system scans online news reports for stories about infectious disease outbreaks and sends e-mail alerts to registered users.
- Additionally, a topic classifier filters data which are used to populate the Global Health Monitor.

www.biocaster.org
The BioCaster corpus consists of 1,000 news articles classified as being a disease outbreak report or not.

Conway et al. 2009 find that the frequency of hedge cues differs in the two categories of the BioCaster corpus, being more frequent in the documents classified as reports.

The classifier is augmented with a binary hedge feature that is true if one of the 105 hedge cues occurs in the text within 5 words of a disease named entity.

The accuracy of this classifier is 0.8% better than the accuracy of a classifier that uses only unigrams, but it does not outperform the best classifier that incorporates feature selection.

Hedge information is also used to assign a speculative metric to the input documents of the BioCaster system, based on the frequency of hedge cues in 10,000 Reuters documents.
18. Sentiment analysis

19. Recognizing textual entailment

20. Machine translation

21. Text mining

22. Identifying the structure of scientific articles

23. Trustworthiness detection

24. References
Identifying the structure of scientific articles


- Automatically categorize article sections (abstract, introduction, material and methods, results, discussion) based on the speculation cues that the sections contain
- The features are 363 speculation cues collected from biomedical articles, which are classified into groups according to their strength.
- When using all features, the sections abstract, results and materials and methods can be classified with high accuracy.
- Strong cues are specific of results, discussion and abstract, and non strong cues of materials and methods.
18. Sentiment analysis

19. Recognizing textual entailment

20. Machine translation

21. Text mining

22. Identifying the structure of scientific articles

23. Trustworthiness detection

24. References
Incorporate evidentiality information to predict trustworthiness of text information in the context of collaborative question answering.

In this context, trustworthiness is useful to find the best answers of the system.

Hypothesis: evidentials will be used in less reliable answers.

Evidentiality is incorporated in the form of lexical features of a classifier that detects best answers and non-best answers.

Results show a 14.85% increase in performance of the classifier with evidentiality information over the baseline classifier (bag-of-words).
Outline

18 Sentiment analysis
19 Recognizing textual entailment
20 Machine translation
21 Text mining
22 Identifying the structure of scientific articles
23 Trustworthiness detection
24 References


References: sentiment analysis


References

**Machine Translation**


**Textual Entailment**


**Classifying citations**