Modality and Negation in Natural Language Processing

Roser Morante

CLiPS - University of Antwerp

November 8, 2011

IJCNLP 2011 Tutorial, Chiang Mai, Thailand

CLiPS

WORLD > COUNTRIES AND TERRITORIES > S

Somalia

Updated: Oct. 15, 2010

Somalia has not had an effective central
former government was toppled by clan
other. For decades, generals, warlords an
this once languid coastal country in East
remains a raging battle zone today, with
overseas, intent on toppling the transitio

No amount of outside firepower has brou
thousands of American Marines in the ea
United Nations mission that followed. N
into Somalia in 2006. Not the current Af
are steadily wearing out their welcome.

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.

Source: `http://topics.nytimes.com/top/news/international/countriesandterritories/somalia/index.html`

Canada's tar sands

Muck and brass

Rising oil prices and falling production co
from Alberta's tar sands. But environmen

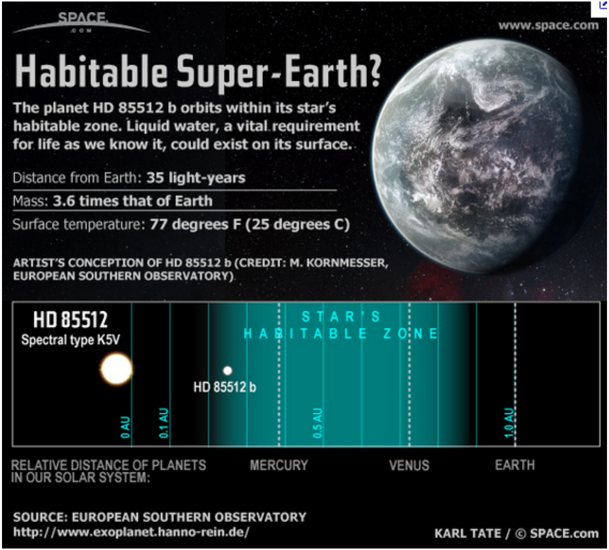
Jan 20th 2011 | CALGARY AND OTTAWA | from PRINT EDI

SMOKESTACKS dot the horizon; a whiff of oil han
the highway. There is a din of heavy machinery, p
scaring birds away from toxic lakes. But golf cour
place liveable, and some locals have grown attach
McMurray, the town at the centre of them. "I'd lik
says a worker at one of Shell's mines.

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

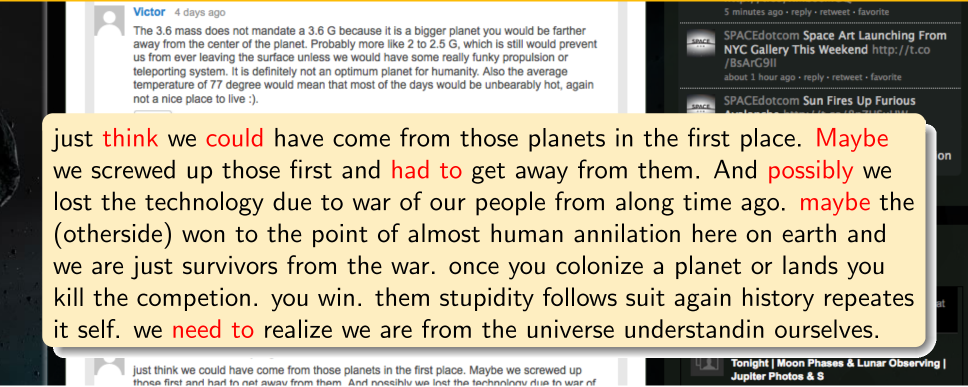


Statements about exoplanets

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 annilation here on earth and we are just survivors from the war. once you colonize a planet or lands you kill the competion. you win. them stupidity follows suit again history repeates it self. we need to realize we are from the universe understandin ourselves.

Source:
http://www.space.com/12918-habitable-alien-planet-hd-85512b-super-earth-infographic.html

Last consulted 14 October 2011

Outline

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

Defining modality

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."

Defining modality

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

	Hyland (Biology)	Holmes (gen. academic)
Lexical verbs	27.4%	35.9 %
Adverbials	24.7%	12.8 %
Adjectives	22.1 %	6.6 %
Modal verbs	19.4 %	36.8 %
Nouns	6.4 %	7.7 %

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

Defining modality

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.

Defining modality

'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.
   [telelological]

Defining modality

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)

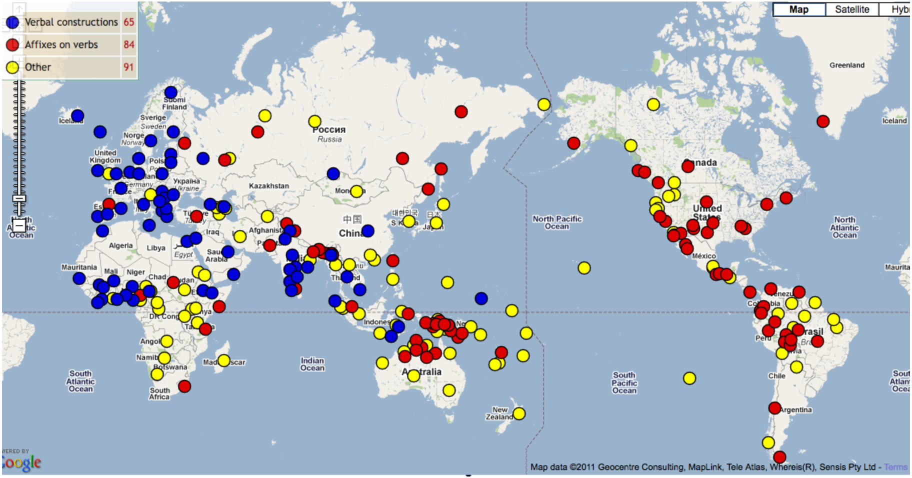
- **Modal adjectives and nouns**: possible, necessary, certain, possibility, ...
- **Propositional attitude verbs and adjectives**: believe, hope, know, remember, certain, pleased, ...
- **Verbal mood**: indicative, subjunctive
- **Infinitives**
- **Dependent modals**
  I'd be surprised if David should 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'.

Defining modality



Source: J. van der Auwera and V. Plungian (2008) Epistemic possibility. In: Haspelmath, M., Dryer, M. S., Gil, D. and Comrie, B. (eds.) The World Atlas of Language Structures Online. Munich: Max Planck Digital Library, chapter 39. Available online at `http://wals.info/refdb/record/van-der-Auwera-and-Plungian-1998`. Accessed on 21 Jan 2011

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'.

Related concepts: Evidentiality

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'."

Related concepts: 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"

1. 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".

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

Related concepts: Hedging

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.

Related concepts: Hedging

Hedge examples from Medlock and Briscoe (2007)

- Dl and Ser have been proposed to act redundantly in the sensory bristle lineage
- How endocytosis of Dl 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

Related concepts: Hedging

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 Dl 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

Related concepts: Hedging

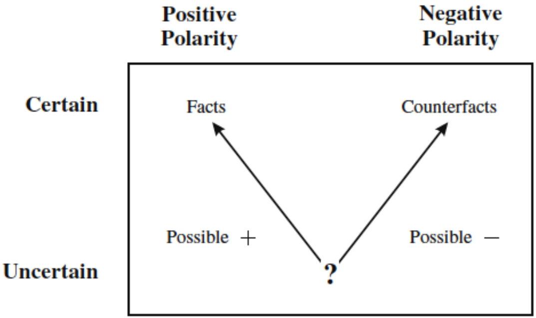
- Experimentally-supported confirmation of previous speculation
  Here we show that the hemocytes are the main regulator of adenosine in the Drosophila larva, as was speculated previously for mammals
- Negation of previous hedge
  Although the adgf-a mutation leads to larval or pupal death, we have shown 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

Related concepts: Factuality

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

Related concepts: Factuality

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)
- Work on subjectivity in computational linguistics is initially due to Wiebe, Wilson, and collaborators (Wiebe 1994, Wiebe et al 2001, 2005, Wilson et al 2006, Wilson 2008, ...) and focuses on learning subjectivity from corpora

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

(Wiebe et al. 2001, 2004)

- 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

Defining modality: 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

- 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"

Outline

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)"

Defining negation

- 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')

Defining negation: Philosophical tradition

- 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:

p	¬p
T	F
F	T

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

$\neg \exists x (Px \wedge \neg Px)$

- Applies to contrary and contradictory negation

Law of the Excluded Middle

A statement must be either true or false

$\forall x (Px \vee \neg Px)$

- Applies to contradictory negation

Defining negation: Philosophical tradition

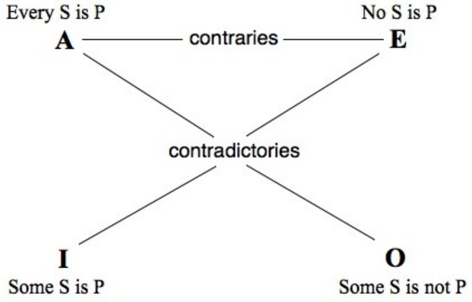
Aristotle studies also negation combined with quantifiers

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



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

More information: Stanford Encyclopedia of Philosophy; `http://plato.stanford.edu/entries/square/`

Defining negation: Logics

Partee (2007)

- Negation $\neg$ is a unary (or monadic) sentential operator.
- Monadic means it has just one argument, unlike $\vee$ and $\wedge$, which are binary
- Sentential means that its argument must be a formula, an expression of type t.
- Semantically it is a function of type $t \rightarrow t$: 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

R.L. Trask (1993) A dictionary of grammatical terms in linguistics. Routledge.

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 [...]

Defining negation: Grammar

R.L. Trask (1993) A dictionary of grammatical terms in linguistics. Routledge.

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 mononarchy, 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

Defining negation: Types

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 consituent negation can be very similar or identical, but constituent negation is less common as a grammatical device
- Most languages posses 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

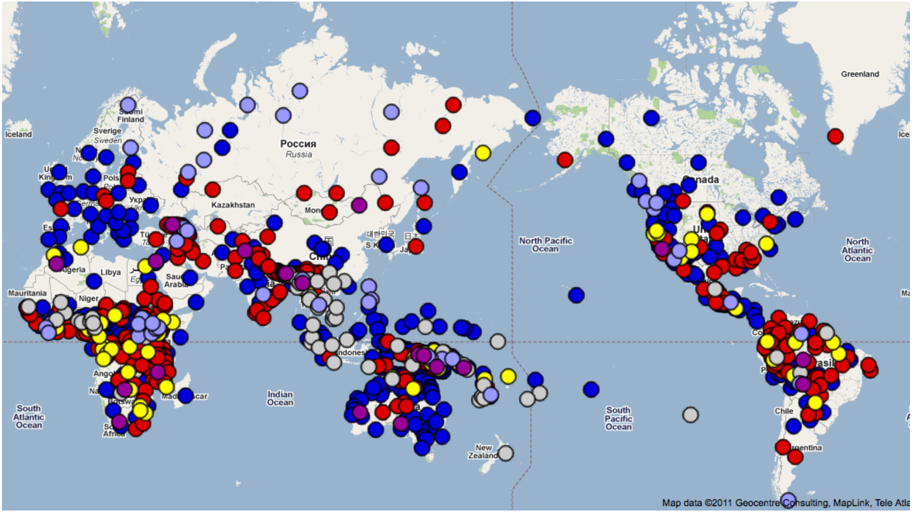
http://wals.info

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 encompases 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

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 of workers
6. Will GM lay off workers?
7. Many wonder if GM will lay off workers

Examples from Prabhakaran et al. (2010)

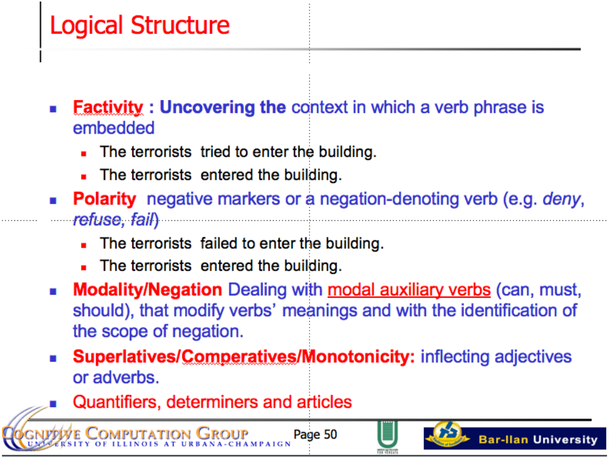
Why is it interesting to process modality?

(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:



Slide borrowed from the Tutorial on Textual Entailment - ACL 2007, by Ido Dagan, DanRoth and Fabio Massimo Zanzotto. `www.cs.biu.ac.il/~dagan/TE-Tutorial-ACL07.ppt`

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

M. Fiszman, Th. C. Rindflesch, and H. Kilicoglu (2006) Summarizing Drug Information in Medline Citations. AMIA Annu Symp Proc. 2006; 2006: 254–258.

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 occurences 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 resoning to reach a conclusion based on the data, and then making new hypotehses. Scientist 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% sentence 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 th curcumin shows very high binding to BSA, probably at the hydrophobi cavities inside the protein. (12870844)
3. Removal of the carboxy terminus enables ERP to interact with a varie of ets-binding sites including the E74 site, the IgH enhancer pisite, anc 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)

Outline

References: Modality

- Bybee, J. L., R. Perkins, and W. Pagliuca (1994) The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World. Chicago: University of Chicago Press
- de Haan, F. (1997) The Interaction of Modality and Negation. A typological study. New York: Garland Publishing, Inc.
- Frawley, W. (2005) (ed.) The expression of modality. Berlin: Mouton de Gruyter
- Palmer, F. R. (1979) Modality and the English modals. London: Longman
- Palmer, F. R. (2001) Mood and Modality. Cambridge: CUP
- Portner, P. (2009) Modality. Oxford University Press, Oxford, UK
- Rizomilioti, V. (2006). Exploring Epistemic Modality in Academic Discourse Using Corpora. Information Technology in Languages for Specific Purposes (7), pp. 53-71
- Salkie, R., P. Busuttil and J. van der Auwera (2009) (eds.). Modality in English. Theory and Description. Berlin: Mouton de Gruyter
- von Fintel, K. (2006) "Modality and Language". In Encyclopedia of Philosophy - Second Edition, edited by D. M. Borchert. Detroit: MacMillan Reference USA. Most recent version online at `http://mit.edu/fintel/www/modality.pdf`

References: Evidentiality

Aikenvald, A. (2004) Evidentiality. Oxford: Oxford University Press

Cornillie, B. (2009) Evidentiality and epistemic modality. Functions of Language 16:1, 44-62

Wallace, Ch. and J. (1986) (eds.) Evidentiality: The Linguistic Coding of Epistemology. Norwood, New Jersey: Ablex

References: Certainty

- Rubin, V.L., E. D. Liddy, and N. Kando (2005) Computing Attitude and Affect in Text: Theory and Applications, chapter Certainty identification in texts: Categorization model and manual tagging results. Springer-Verlag, New York
- Rubin, V.L. (2006) Identifying certainty in texts. Ph.D. thesis, Syracuse, NY, USA
- Rubin, V.L. (2007) Stating with certainty or stating with doubt: intercoder reliability results for manual annotation of epistemically modalized statements. In Proceedings of NAACL'07: Human Language Technologies 2007. Companion Volume, pages 141–144, Morristown, NJ, USA
- Rubin, V.L. (2010) Epistemic modality: from uncertainty to certainty in the context of information seeking as interactions with texts. Information processing and management 46:533–540

References: Hedging

- Hyland, K. (1994) Hedging in academic writing and eaptextbooks. English for Specific Purposes, 13:239-256
- Hyland, K. (1998) Hedging in scientific research articles. Amsterdam, The Netherlands: John Benjamins
- Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. Journal of Philosophical Logic, 2:458-508
- Medlock, B. and T. Briscoe (2007) Weakly supervised learning for hedge classfication in scientifi literature. In Proceedings of ACL 2007, pages 992–999
- Vincze, V., G. Szarvas, R. Farkas, G. Móra, and J. Csirik (2008) The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics, 9((Suppl 11)):S9.

References: Factuality

- Pustejovsky J., P. Hanks, R. Saurí, A. See , R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, and M. Lazo (2003) The TimeBank corpus. Proceedings of Corpus Linguistics 2003:647-656
- Saurí R., M. Verhagen and J. Pustejovsky (2006) SlinkET: a partial modal parser for events. Proceedings of 5th International Conference on Language Resources and Evaluation 2006
- Saurí, R. and J. Pustejovsky (2009). FactBank: a corpus annotated with event factuality. Language Resources and Evaluation 43: 227–268

References: Subjectivity

- Banfield, A. (1982) Unspeakable sentences. Routledge and Kegan Paul, Boston
- Prabhakaran, V., O. Rambow, and M. Diab (2010) Automatic committed belief tagging. In Proceedings of COLING 2010, pages 1014–1022
- Wiebe, J., R. Bruce, M. Bell, M. Martin and Th. Wilson (2001) A corpus study of evaluative and speculative language. Proceeding SIGDIAL '01 Proceedings of the Second SIGdial Workshop on Discourse and Dialogue
- Wiebe, J., Th. Wilson, R. Bruce, M. Bell, and M. Martin (2004) Learning subjective language. Computational Linguistics, 30(3):277–308
- Wilson, Th, J. Wiebe, and R. Hwa (2006) Recognizing strong and weak opinion clauses. Computational Intelligence, 22(2):73–99
- Wilson, Th. (2008) Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states. Ph.D. thesis, University of Pittsburgh, Pittsburgh, PA, USA

References: Negation

- Dryer, M. S. (2008) Negative Morphemes. In: M. Haspelmath, M. S. Dryer, D. Gil, and B. Comrie (eds.) The World Atlas of Language Structures Online. Munich: Max Planck Digital Library, chapter 39. Available online at `http://wals.info/feature/39`. Accessed on 20 Jan 2011
- de Haan, F. (1997) The Interaction of Modality and Negation. A typological study. New York: Garland Publishing, Inc.
- Hoeksema, F., H. Rullmann, V. Sánchez-Valencia and T. van der Wouden (2001) (eds.) Perspectives on Negation and Polarity Items. Amsterdam: John Benjamins
- Horn, L.R., and Y. Kato (2000) (eds) Studies in Negation and Polarity. Oxford University Press
- Jespersen, O. (1924) The Philosophy of Grammar. Chicago: The University of Chicago Press, edition 1992
- Klima, E. S. (1964) Negation in English. Readings in the Philosophy of Language. Ed. J. A. Fodor and J. J. Katz. Prentice Hall, Englewood Cliffs, NJ: 246-323
- Lawler, J. (2007) Negation and negative polarity. Cambridge Encyclopedia of the Language Sciences `http://www.umich.edu/~jlawler/CELS-Negation.pdf`
- Tottie, G. (1991) Negation in English Speech and Writing: A Study in Variation Academic, San Diego, CA
- van der Wouden, T. (1997) Negative Contexts. Collocation, polarity, and multiple negation. London and New York: Routledge

Additional References Part I

- Chapman, W. W., W. Bridewell, P. Hanbury, G. F. Cooper, and B. G. Buchanan (2001) A simple algorithm for identifying negatedfindings and diseases in discharge summaries. J Biomed Inform, 34:301–310
- Chu D, J. N. Dowling, and W. W. Chapman (2006) Evaluating the effectiveness of four contextual features in classifying annotated clinical conditions in emergency department reports. AMIA Annu Symp Proc 2006:141–5
- Light, M., X. Y. Qiu, and P. Srinivasan (2004) The language of bioscience: facts, speculations, and statements in between. In Proceedings of BioLINK 2004, pages 17–24
- Medlock, B. (2008) Exploring hedge identfication in biomedical literature. JBI, 41:636–654
- Mercer, R. E., Ch. DiMarco, and F. W. Kroon (2004) The frequency of hedging cues in citation contexts in scientific writing. In Proceedings of the Canadian Conference on AI, pages 75–88
- Mutalik, P. G., A. Deshpande, and P. M. Nadkarni (2001) Use of general-purpose negation detection to augment concept indexing of medquantitative study using the UMLS. J Am Med Inform Assoc, 8(6):598–609.
- Shatkay, H., F. Pan, A. Rzhetsky, and W. J. Wilbur (2008) Multi-dimensional classfication of biomedical texts: toward automated practical provision of high-utility text to diverse users. Bioinformatics, 24(18):2086–2093
- Wilbur, W. J., A. Rzhetsky, and H. Shatkay (2006) New directions in biomedical textannotations: dfinitions, guidelines and corpus construction. BMC Bioinformatics, 7:356

Part II

Categorising and Annotating Modality and Negation

Outline

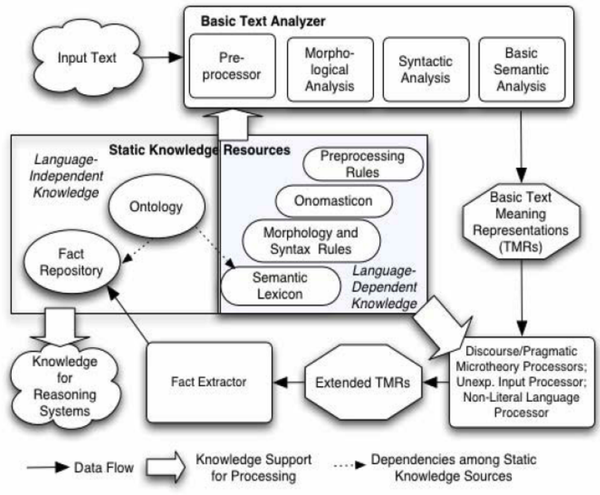
Outline

Annotation schemes: Modality in OntoSem

Nirenburg, S. and M. McShane (2008) Annotating modality. OntoSem final project report. March.

- 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



From Nirenburg et al (2004) Evaluating the performance of the OntoSem semantic analyzer. Proc. of the ACL Workshop on Text Meaning Representation.

OntoSem

- 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)

OntoSem

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.

OntoSem

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

OntoSem

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)

OntoSem

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

OntoSem

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

Wiebe, J., Th. Wilson, and C. Cardie (2005) Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 38:165–210.

- 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: identiying private state expressions in context

Private States

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

Private States

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)

1. "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

Private States

- 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.

Annotation schemes: Attribution in PDTB

Prasad, R. , N. Dinesh, A. Lee, A. Joshi, and B. Webber. 2006. Annotating attribution in the Penn Discourse TreeBank. In SST '06: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pages 31-38, Morristown, NJ, USA. ACL.

- 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"

Attribution in PDTB

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

Linguistic Data Consortium. 2008. ACE (Automatic Content Extraction) English annotation guidelines for relations. Technical Report Version 6.2 2008.04.28, LDC.

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

Annotation schemes: Certainty

Rubin, V.L., E. D. Liddy, and N. Kando (2005) Computing Attitude and Affect in Text: Theory and Applications, chapter Certainty identification in texts: Categorization model and manual tagging results. Springer-Verlag, New York.

Rubin, V.L. (2010) Epistemic modality: from uncertainty to certainty in the context of information seeking as interactions with texts. Information processing and management 46:533-540.

- 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

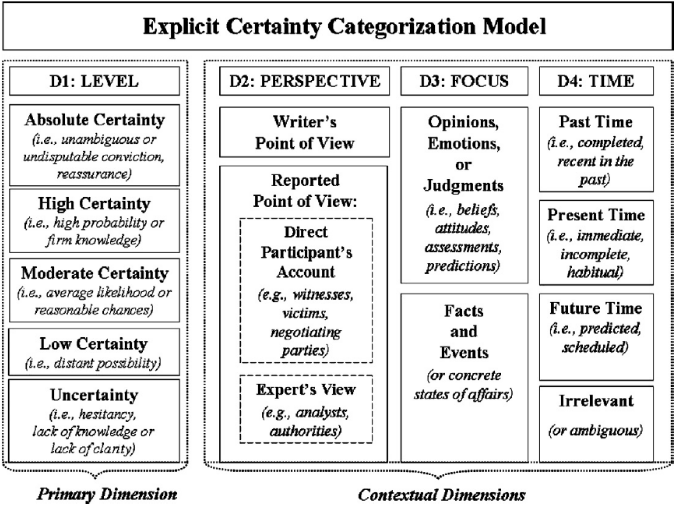
Certainty identificaiton (Rubin et al 2005)

"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)



(Image 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 e 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, ..."

Certainty

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 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

Annotation schemes: Factuality

Saurí, R. and J. Pustejovsky (2009). FactBank: a corpus annotated with event factuality. Language Resources and Evaluation 43: 227–268.

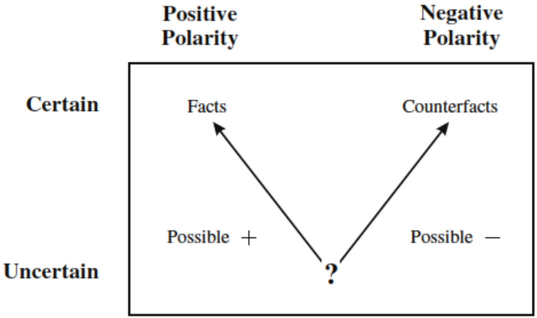
Factuality (Saurí and Pustejovsky 2009)

"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"

Factuality

- **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
  1. 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

	Positive (+)	Negative (−)	Underspecified (u)
Certain (CT)	Fact: <CT,+>	Counterfact: <CT,−>	Certain but unknown output: <CT,u>
Probable (PR)	Probable: <PR,+>	Not probable: <PR,−>	(NA)
Possible (PS)	Possible: <PS,+>	Not certain: <PS,−>	(NA)
Underspecified (U)	(NA)	(NA)	Unknown or uncommitted: <U,u>

Factuality values in Saurí and Pustejovsky (2009)

Factuality

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

Factuality

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

<EVENT eid="e22" class="REPORTING" tense="PRESENT" aspect="PERFECTIVE">

said </EVENT>

Amir was

<EVENT eid="e23" class="STATE" tense="PAST">

infatuated </EVENT>

with Har-Shefi and may have been

<EVENT eid="e24" class="I_ACTION" modality="may" tense="NONE" aspect="PERF_PROG">

trying </EVENT>

to

<EVENT eid="e25" class="OCCURRENCE" tense="INFINITIVE">

impress </EVENT>

her by

<EVENT eid="e26" class="OCCURRENCE" tense="PRESPART" aspect="NONE">

killing </EVENT> the prime minister.

<SLINK lid="l50" eventId="e22" subordinatedEventId="e23" relType="EVIDENTIAL"/>

<SLINK lid="l51" eventId="e22" subordinatedEventId="e24" relType="EVIDENTIAL"/>

<SLINK lid="l52" eventId="e24" subordinatedEventId="e25" relType="MODAL"/>

FactBank annotation

Event (ID):	Source (ID):	Fact. value:
said (e22)	*author* (s_0)	CT+
infatuated (e23)	*reports_author* ($s_2\_s_0$)	CT+
	author (s_0)	Uu
trying (e24)	*reports_author* ($s_2\_s_0$)	PS+
	author (s_0)	Uu
impress (e25)	*reports_author* ($s_2\_s_0$)	Uu
	author (s_0)	Uu
killing (e26)	*reports_author* ($s_2\_s_0$)	Uu
	author (s_0)	Uu

Factuality

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}]

Annotation schemes: Committed belief

Diab, M. T., L. Levin, T. Mitamura, O. Rambow, V. Prabhakaran, and W. Guo (2009) Committed belief annotation and tagging. In ACL-IJCNLP '09: Proceedings of the Third Linguistic Annotation Workshop, pages 68-73.

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 of 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

Thompson P, Venturi G, McNaught J, Montemagni S, Ananiadou S: Categorising modality in biomedical texts. Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining 2008.

- **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

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, probaly, generally
- Moderate: possibly, perhaps, may, could
- Low: unlikely, unknown

Categorising modality

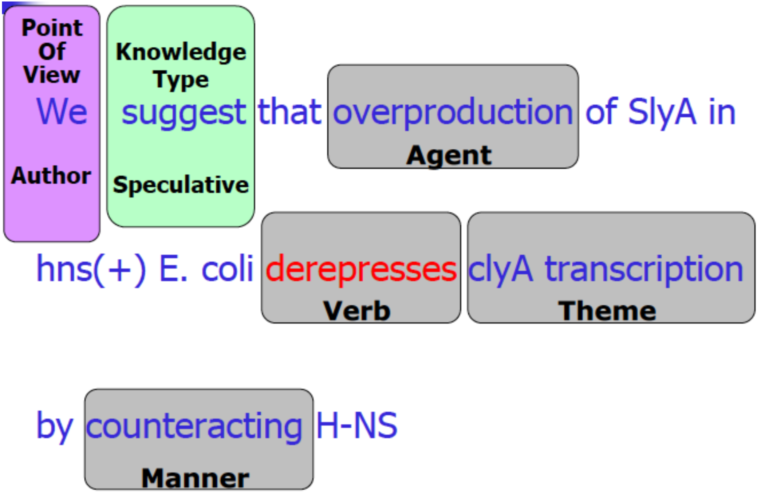
Dimensions of the categorisation scheme

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

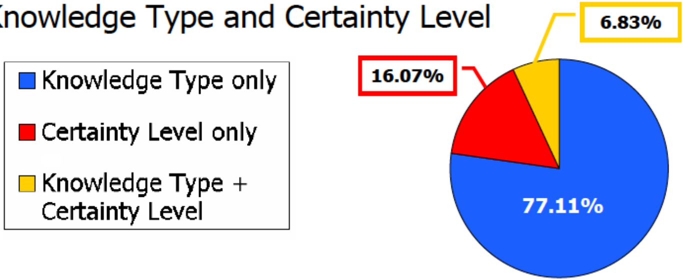


(Figure from Thompson et al. (2008)
www.nactem.ac.uk/workshops/lrec08_ws/slides/Thompson_et_al.pdf)

Categorising modality

Distribution per category in the annotated corpus

- Knowledge Type and Certainty Level

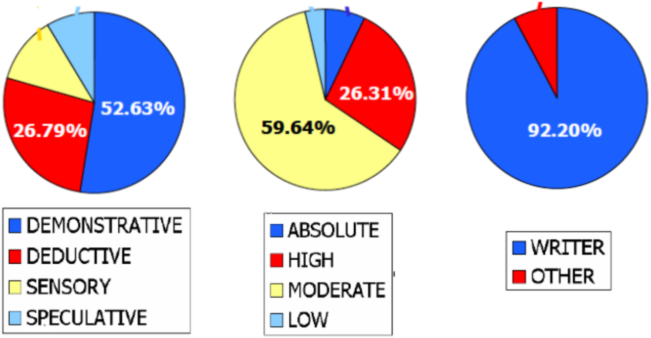


- If values were assigned to Knowledge Type and/or Certainty Level dimensions, Point of View dimension also instantiated

(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

K. Baker, M. Bloodgood, B. J. Dorr , N. W. Filardo , L. Levin, and Christine Piatko. A modality lexicon and its use in automatic tagging. Proceedings of LREC 2010, pages 1402–1407.

Baker, K. et al. (2009) SIMT SCALE 2009 - Modality Annotation Guidelines. Technical Report 4, Human Language Technology Center of Excellence, Johns Hopkins University, 2010.

- 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

A modality lexicon

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 ˆJJ of ˆIN$Able & capable, JJ-of-basic, JJ-of-VBG

able ˆJJ to ˆ"TO$Able & able, JJ-infinitive

can ˆMD$Able & can, modal-auxiliary-basic

could ˆMD$Able & could, modal-auxiliary-basic

ready ˆJJ$Able & ready, JJ-infinitive

 # NotAble

powerless ˆJJ$NotAble & powerless, JJ-infinitive

unable ˆJJ$NotAble & unable, JJ-infinitive

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 <TargSucceed hold> general elections in the year 2002, but he <TrigAble can> <TrigNegation not> <TargNOTAble be> ignorant of the fact that the world at large did <TrigNegation not> <TrigBelief accept> these <TargBelief elections>.

A modality lexicon

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 wordsphrases 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)
```

Annotation schemes: Speculated sentences

Medlock, B. and T. Briscoe (2007) Weakly Supervised Learning for Hedge Classification in Scientific Literature. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics.

- 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

Blanco, E. and D. Moldovan (2011) Semantic representation of negation using focus detection. Proceedings of ACL 2011, pges 581-589. ACL.

- 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 schemes: Focus of Negation

Annotation examples (Table from Blanco and Moldovan 2011)

	Statement	V	A0	A1	A2	A4	TMP	MNR	ADV	LOC	PNC	EXT	DIS	MOD
1	Even if [that deal]$_{A1}$ isn't [revived]$_{V}$, NBC hopes to find another.													
	– Even if that deal is suppressed, NBC hopes to find another one.	★	-	+	-	-	-	-	-	-	-	-	-	-
2	[He]$_{A0}$ [simply]$_{MDIS}$ [ca]$_{MMOD}$n't [stomach]$_{V}$ [the taste of Heinz]$_{A1}$, she says.													
	– He simply can stomach any ketchup but Heinz's.	+	+	★	-	-	-	-	-	-	-	-	+	+
3	[A decision]$_{A1}$ isn't [expected]$_{V}$ [until some time next year]$_{MTMP}$.													
	– A decision is expected at some time next year.	+	-	+	-	-	★	-	-	-	-	-	-	-
4	[...] it told the SEC [it]$_{A0}$ [could]$_{MMOD}$n't [provide]$_{V}$ [financial statements]$_{A1}$ [by the end of its first extension]$_{MTMP}$ "[without unreasonable burden or expense]$_{MMNR}$".													
	– It could provide them by that time with a huge overhead.	+	+	+	-	-	+	★	-	-	-	-	-	+

http://www.inf.u-szeged.hu/rgai/bioscope

The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes

Veronika Vincze[1], **György Szarvas**[1], **Richárd Farkas**[2*], **György Móra**[1*] and **János Csirik**[2*]

1 University of Szeged, Department of Informatics, Human Language Technology Group, Árpád tér 2., Szeged, Hungary

2 Research Group on Artificial Intelligence of the Hungarian Academy of Sciences and University of Szeged, Aradi Vértanúk tere 1., Szeged, Hungary

author email corresponding author email * Contributed equally

BMC Bioinformatics 2008, **9**(Suppl 11):S9 doi:10.1186/1471-2105-9-S11-S9

Corpus annotated with negation and speculation cues and their scopes in English biomedical texts

Scopes in BioScope

Table 1: Statistics of the three subcorpora

	Clinical	Full Paper	Abstract
#Documents	1954	9	1273
#Sentences	6383	2670	11871
Negation sentences	13.55%	12.70%	13.45%
#Negation cues	877	389	1848
Hedge sentences	13.39%	19.44%	17.70%
#Hedge cues	1189	714	2769

Scopes in BioScope

Example

When U937 cells were infected with HIV-1, <xcope id="X1.6.3"><cue type="negation" ref="X1.6.3">**no**</cue> induction of NF-KB factor was detected</xcope>, whereas high level of progeny virions was produced, <xcope id="X1.6.2"><cue type="speculation" ref="X1.6.2">suggesting</cue> that this factor was <xcope id="X1.6.1"><cue type="negation" ref="X1.6.1">**not**</cue> required for viral replication</xcope></xcope>.

Scopes in BioScope

- Hedges modify the factuality of an 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.

Scopes in BioScope

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.

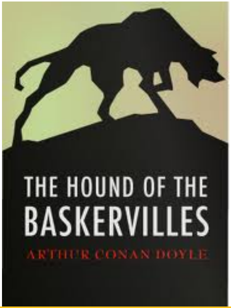
    Mild bladder wall thickening is a fact
    Cystitis is an uncertain fact

Scopes in BioScope

Scope and syntax: the scope of a speculative element can be determined on the basis of syntax.

- Verbs, auxiliaries, adjectives and adverbs usually starts 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: scopegenerally 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)

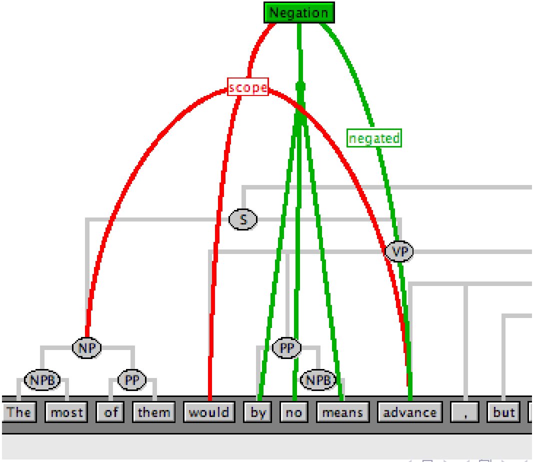
Scopes in ConanDoyle-neg

R. Morante, S. Schrauwen and W. Daelemans. Annotation of negation cues and their scope. CLiPS Technical Report 3. University of Antwerp.

- **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]



Scopes in ConanDoyle-neg

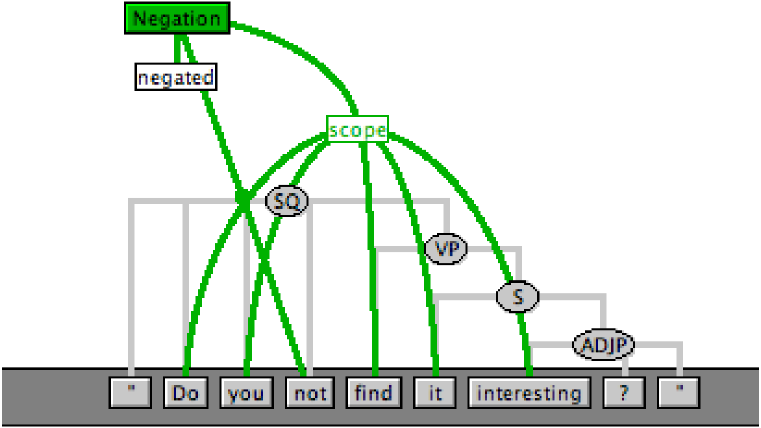
The annotation format is based on the BioScope format, but there are differences:

- 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

Scopes in ConanDoyle-neg

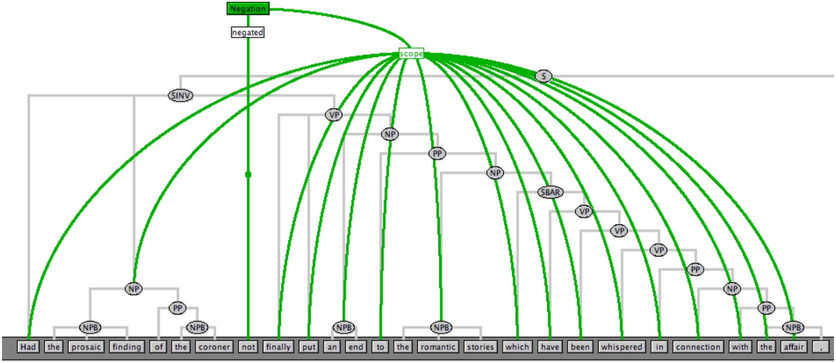
Not all negations cues negate a fact

" Do you not find it interesting ? " [HB 2.74]



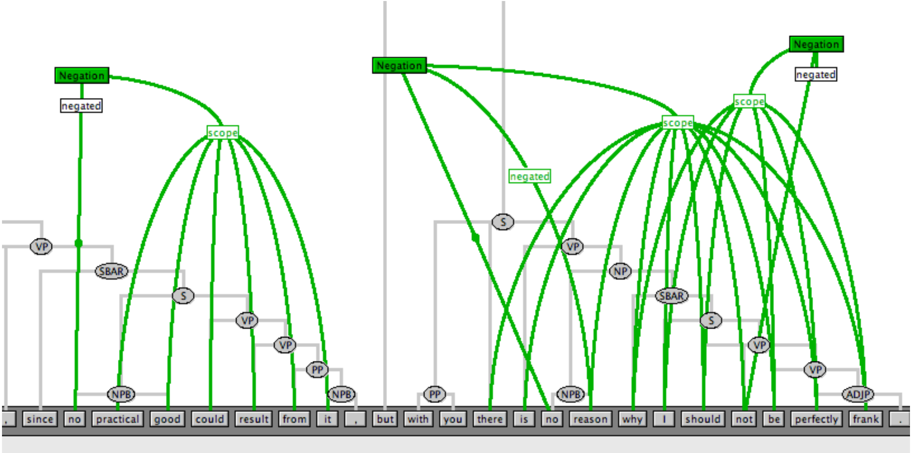
Not all negations cues negate a fact

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]



Scopes in ConanDoyle-neg

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`

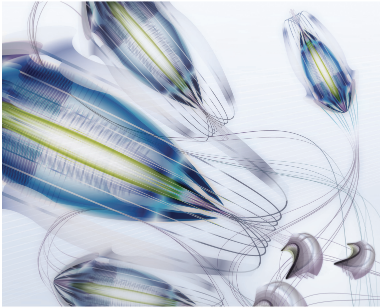
ISSN 2033-3544

CLiPS

COMPUTATIONAL LINGUISTICS & PSYCHOLINGUISTICS
TECHNICAL REPORT SERIES, CTRS-003, JANUARY 2010
WWW.CLIPS.UA.AC.BE/CTRS

Annotation of Negation Cues and their Scope Guidelines v1.0

Roser Morante, Sara Schrauwen, Walter Daelemans



Universiteit Antwerpen

COMPUTATIONAL LINGUISTICS PSYCHOLINGUISTICS RESEARCH CENTER CLiPS

Annotation schemes: High utility text

W John Wilbur, Andrey Rzhetsky, and Hagit Shatkay (2006) New directions in biomedical text annotation: definitions, guidelines and corpus construction. BMC Bioinformatics 2006, 7:356

Hagit Shatkay, Fengxia Pan, Andrey Rzhetsky, and W. John Wilbur (2008) Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users. Bioinformatics 24(18), pages 2086–2093

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

High utility text

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

High utility text

Goal

- Enabling the creation of well-focused subsets of biomedical text that have certain properties
- Identifing 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

High utility text

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

High utility text

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

High utility text

Dimensions

Polarity

A fragment with any focus can be stated either positively (P) or negatively (N)

Certainty

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.

High utility text

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

High utility text

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)

High utility text

Examples

The binding of both forms of β-catenin to CBP is completely inhibited by ICG-001 (Fig. 3B Top, lane 4). **1SP3E3-

We demonstrate that ICG-001 binds specifically to CBP **1SP2E3 but not the related transcriptional coactivator p300, **2SN2E3

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), **1SN3E23+

High utility text

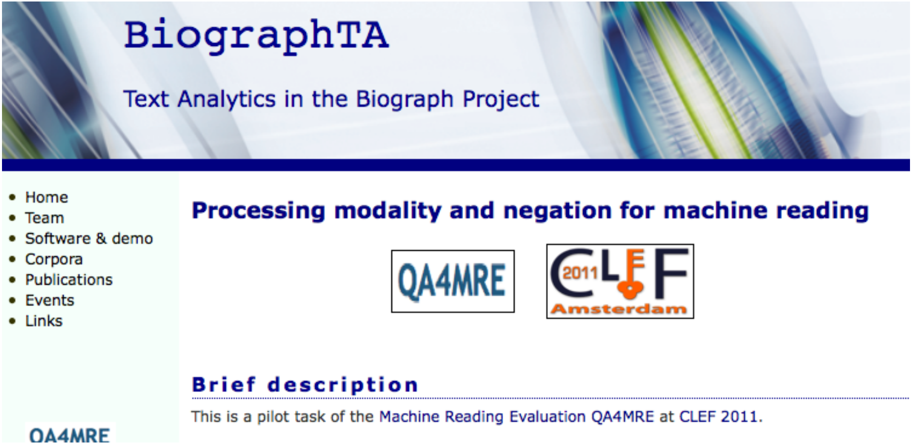
Data

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

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

	Foc. & Ev.	Focus	Evidence	Certainty	Polarity	Trend
Dataset name	*Frag_FE*	*Frag_F*	*Frag_E*	*Frag_C*	*Frag_P*	*Frag_T*
Sentences	1977	4068	2964	5644	6430	5907
Fragments	2109	4447	3133	5992	6945	6330
No. of terms selected	F: 600 + E: 200	1500	1500	100	600	100

(From Shatkay et al. 2008)


BiographTA
Text Analytics in the Biograph Project
Home
Team
Software & demo
Corpora
Publications
Events
Links
Processing modality and negation for machine reading
QA4MRE
2011 CLEF
Amsterdam
Brief description
This is a pilot task of the Machine Reading Evaluation QA4MRE at CLEF 2011.
QA4MRE

Annotation schemes: Modality and negation for MR

- **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 document 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

Topic	Number	Title	# of words
Aids	1	All colours of the brainbow	915
	2	DARC continent	817
	3	Double, not quits	779
	4	Win some, lose some	1919
Climate change	1	A record-making effort	2841
	2	Are economists erring on climate change?	1412
	3	Climate change and evolution	1256
	4	Climate change in black and white	2850
Music and society	1	The politics of hip-hop	1004
	2	How to sink pirates	773
	3	Singing a different tune	1042
	4	Turn that noise off	677

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

Modality and negation for MR

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.

Perspective

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.

1. 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 and negation for MR

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.

Modality and negation for MR

- **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

Modality and negation for MR

Example question

Text

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.

Question

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

<answer a_id="1">MOD-NEED</answer>
<answer a_id="2">MOD-WANT</answer>
<answer a_id="3">COND-BY MOD-NON</answer>
<answer a_id="4" correct="Yes">UNCERT MOD-NON</answer>
<answer a_id="5">NEG UNCERT MOD-NON</answer>

Modality and negation for MR

- The combinations of codes that conform the answers to the questions can be summarized with the following regular expression:

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

- 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

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

- **BioInfer** http://mars.cs.utu.fi/BioInfer/
- **Genia Event** 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/
- **GREC** http://www.nactem.ac.uk/GREC/

Scopes

- **BioScope** http://www.inf.u-szeged.hu/rgai/bioscope
- **CoNLL Shared Task 2010 data** http://www.inf.u-szeged.hu/rgai/conll2010st/
- **ConanDoyle-neg** http://www.clips.ua.ac.be/BiographTA/corpora-files/ConanDoyle-neg-v1.zip

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**
  http://www.clips.ua.ac.be/BiographTA/qa4mre.html

Factuality

- **FactBank** http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2009T23

Subjectivity

- **MPQA Opinion Corpus**
  http://www.cs.pitt.edu/mpqa/

Discourse

- **PDTB** http://www.seas.upenn.edu/~pdtb/

Outline

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

Baker K., M. Bloodgood, B. J. Dorr , N. W. Filardo , L. Levin, and Ch. Piatko (2010) A modality lexicon and its use in automatic tagging. Proceedings of LREC 2010, pages 1402–1407.

Baker, K. et al. (2009) SIMT SCALE 2009 - Modality Annotation Guidelines. Technical Report 4, Human Language Technology Center of Excellence, Johns Hopkins University, 2010.

Dalianis, H. and S. Velupillai. 2010. How certain are clinical assessments? Annotating swedish clinical text for (un)certainties, speculations and negations. In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta.

Diab, M. T., L. Levin, T. Mitamura, O. Rambow, V. Prabhakaran, and W. Guo (2009) Committed belief annotation and tagging. In ACL-IJCNLP '09: Proceedings of the Third Linguistic Annotation Workshop, pages 68-73.

Linguistic Data Consortium. 2008. ACE (Automatic Content Extraction) English annotation guidelines for relations. Technical Report Version 6.2 2008.04.28, LDC.

Medlock, B. and T. Briscoe (2007) Weakly Supervised Learning for Hedge Classification in Scientific Literature. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics.

Medlock, B. (2006) Guidelines for Speculative Sentence Annotation. http://www.benmedlock.co.uk/annotation.pdf

References

Morante, R., S. Schrauwen and W. Daelemans. Annotation of negation cues and their scope. CLiPS Technical Report 3. University of Antwerp.

Nirenburg, S. and M. McShane (2008) Annotating modality. OntoSem final project report. March.

Prasad, R. , N. Dinesh, A. Lee, A. Joshi, and B. Webber. 2006. Annotating attribution in the Penn Discourse TreeBank. In SST '06: Proceedings of the Workshop on Sentiment and Subjectivity in Text, pages 31-38, Morristown, NJ, USA. Association for Computational Linguistics.

Rubin, V.L., E. D. Liddy, and N. Kando (2005) Computing Attitude and Affect in Text: Theory and Applications, chapter Certainty identification in texts: Categorization model and manual tagging results. Springer-Verlag, New York.

Rubin, V.L. (2010) Epistemic modality: from uncertainty to certainty in the context of information seeking as interactions with texts. Information processing and management 46:533-540.

References

Saurí, R. and J. Pustejovsky (2009). FactBank: a corpus annotated with event factuality. Language Resources and Evaluation 43: 227–268.

Thompson P., Venturi G., McNaught J., Montemagni S., Ananiadou S. (2008) Categorising modality in biomedical texts. Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining 2008.

Thompson P, Nawaz, R, McNaught J, Ananiadou SEnriching a biomedical event corpus with meta-knowledge annotation. BMC Bioinformatics 2011, 12:393

Vincze V., G. Szarvas, R. Farkas, G. Móra, and J. Csirik (2008) The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC Bioinformatics, 9(Suppl 11):S9, 2008.

Vincze, V., G. Szarvas, G. Móra, T. Ohta, and R. Farkas (2011) Linguistic scope-based and biological event based speculation and negation annotations in the BioScope and Genia Event corpora. Journal of Biomedical Semantics 2011, 2(Suppl 5):S8

Wiebe, J., Th. Wilson, and C. Cardie (2005) Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 38:165–210.

Wilbur W.J., Rzhetsky A., Shatkay H. (2006) New directions in biomedical text annotations: definitions, guidelines and corpus construction. BMC Bioinformatics 2006, 7:356.

Part III

Tasks Related to Processing Modality and Negation

Outline

Outline

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

M. Light, X. Ying Qiu, and P. Srinivasan. 2004. The Language of Bioscience: Facts, Speculations, and Statements in between. In Proceedings of the HLT BioLINK.

- **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**: $\text{SVM}_{\mathit{light}}$
- **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

B. Medlock, and T. Briscoe. 2007. Weakly Supervised Learning for Hedge Classification in Scientific Literature. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics.

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 $\mathcal{S}_{nspec}$ by sampling randomly from $\mathcal{U}$.
2. Manually remove more 'obvious' speculative sentences using pattern matching
3. Iterate:
   - Order $\mathcal{S}_{nspec}$ by $P(spec|\mathrm{x}_j)$ using estimates from $\mathcal{S}_{spec}$ and current $\mathcal{S}_{nspec}$
   - 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

Ben Medlock. 2008. Exploring hedge identification in biomedical literature. Journal of Biomedical Informatics, 41:636-54

- 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

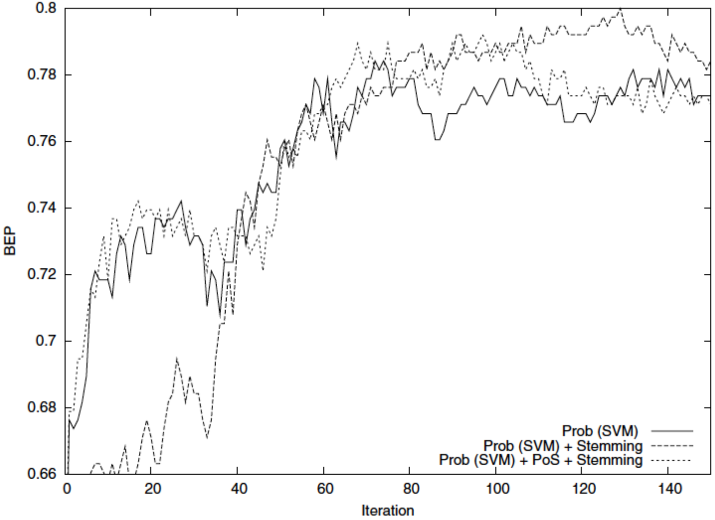


Figure from Medlock (2008)

Learning curves bigrams

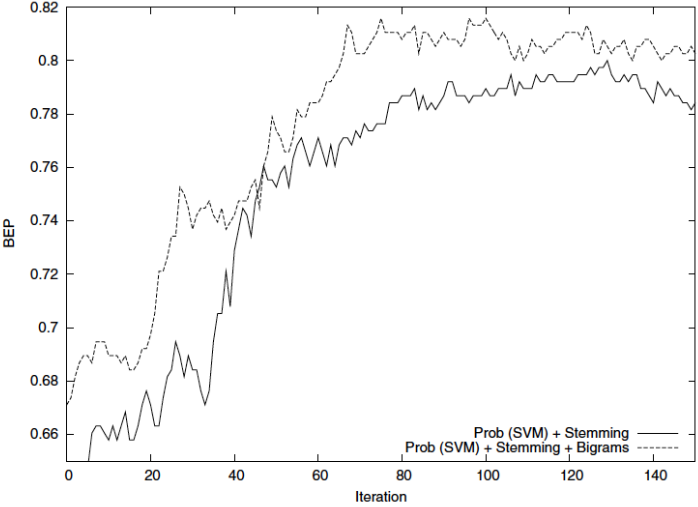


Figure from Medlock (2008)

Informative features

Single term + bigram features ranked by $P(spec|x_k)$ with $\alpha = 5$

1	suggest	31	may_not
2	suggest_that	32	idea_that
3	might	33	be_due
4	may_be	34	it_may
5	possibl_that	35	most_like
6	might_be	36	result_indic
7	appear_to	37	and_may
8	result_suggest	38	it_seem
9	propos_that	39	hypothesi_that
10	is_like	40	suggest_the
11	thought_to	41	been_suggest
12	suggest_a	42	the_hypothesi
13	thi_suggest	43	we_propos
14	seem_to	44	test_whether
15	whether_the	45	possibl
16	whether	46	specul
17	data_suggest	47	that_may
18	like_to	48	observ_suggest
19	like_that	49	strongli_suggest
20	may_have	50	possibl_is
21	may_also	51	rais_the
22	seem	52	appear_that
23	may	53	also_be
24	the_possibl	54	ar_thought
25	thought	55	and_suggest
26	determin_whether	56	be_involv
27	ar_like	57	thi_may
28	is_possibl	58	propos
29	is_thought	59	specul_that
30	the_idea	60	a_role

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

G. Szarvas. 2008. Hedge classification in biomedical texts with a weakly supervised selection of keywords. In Proceedings of the ACL-08: HLT

- 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(spec|x).
  - Select features with P(spec|x) $>$ 0.94 and appeared in the training dataset with reasonable frequency
  - Result: 2407 candidates
- For trigrams, bigrams and unigrams, calculate a new classconditional 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

	Biomedical papers		Medical reports
	$BEP(spec)$	$F_{\beta=1}(spec)$	$F_{\beta=1}(spec)$
Baseline 1	60.00	–	48.99
Baseline 2	76.30	–	–
All features	76.05	73.61	64.04
Feature selection	78.68	78.09	79.73
Manual feat. sel.	82.02	80.88	81.96
Outer dictionary	85.29	85.08	82.07

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

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 uninformaive 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

H. Kilicoglu and S. Bergler. 2008. Recognizing Speculative Language in Biomedical Research Articles: A Linguistically Motivated Perspective. In Proceedings of Current Trends in Biomedical Natural Language Processing (BioNLP), Columbus, Ohio, USA

- 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 datase
  - 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 2: Syntactic patterns and their effect on hedging strength

Syntactic Pattern	Effect on strength
<EPISTEMIC VERB> *to(inf)* VB	+1
<EPISTEMIC VERB> *that(comp)* VB	+2
Otherwise	-1
<EPISTEMIC NOUN> followed by *that(comp)*	+2
Otherwise	-1
not <UNHEDGING VERB>	+1
no|not <UNHEDGING NOUN>	+2
no|not immediately followed by <UNHEDGING ADVERB>	+1
no|not immediately followed by <UNHEDGING ADJECTIVE>	+1
whether|if in a clausal complement context	3(SA) 1.58(IG)

(Table from Kiligoglu and Bergler 2008)

Detecting speculated sentences

Results

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

Method	Recall/Precision
baseline1	0.60
baseline2	0.76
Our system on the fruit-fly dataset with SA weighting	0.85
Our system on the fruit-fly dataset with IG weighting	0.80
Our system on the BMC dataset with SA weighting	0.82
Our system on the BMC dataset with IG weighting	0.70

(Table from Kiligoglu and Bergler 2008)

- Baseline 1: substring matching from Light et al (2004)
- 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

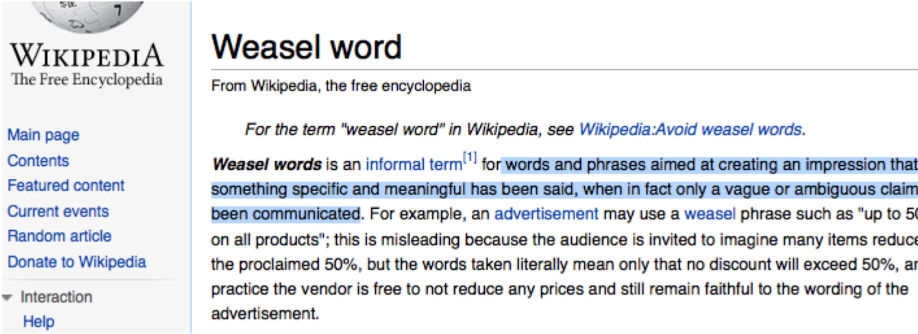
- "Weak" hedging cues, such as epistemic deductive verbs (*conclude, estimate*) as well as some adverbs (*essentially, usually*) and nominalizations (*implication, assumption*)

V. Ganter and M. Strube: Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 173-176, Suntec, Singapore, August 2009. Association for Computational Linguistics

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



WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia

Interaction
Help

Weasel word

From Wikipedia, the free encyclopedia

For the term "weasel word" in Wikipedia, see Wikipedia:Avoid weasel words.

Weasel words is an informal term[1] for words and phrases aimed at creating an impression that something specific and meaningful has been said, when in fact only a vague or ambiguous claim been communicated. For example, an advertisement may use a weasel phrase such as "up to 5 on all products"; this is misleading because the audience is invited to imagine many items reduce the proclaimed 50%, but the words taken literally mean only that no discount will exceed 50%, ar practice the vendor is free to not reduce any prices and still remain faithful to the wording of the advertisement.

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 chosing 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:
  - Numericallly underspecified subjects ("Some people", "Experts", "Many")
  - Passive constructions ("It is believed", "It is considered")
  - Adverbs ("Often", "Probably")

Detecting speculated sentences

Results

σ	.60	.70	**.76**	.80	.90	**.98**
balanced set						
wpw	.68	.68	.68	.69	.69	**.70**
asp	.67	.68	**.68**	.68	.61	.59
manual annot.						
wpw	-	.59	-	-	-	**.59**
asp	.68	.69	**.69**	.69	.70	.65

- 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

Detecting speculated sentences. CoNLL ST'10

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

Information source: R. Farkas, V. Vincze, G. Móra, J. Csirik, and G. Szarvas. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. Proceedings of the Fourteenth Conference on Computational Natural Language Learning: Shared Task, pages 1-12

Detecting speculated sentences. CoNLL ST'10

Results Task 1

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

Table 1: Task1 Wikipedia results (type $\in$ {Closed(C), Cross(X), Open(O)}).

Detecting speculated sentences. CoNLL ST'10

Classifying Wikipedia sentences as uncertain - best system

M. Georgescul. A Hedgehop over a Max-Margin Framework Using Hedge Cues. Proceedings of the Fourteenth Conference on Computational Natural Language Learning: Shared Task, pages 26-31

- 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.

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

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.

Detecting speculated sentences. CoNLL ST'10

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 = γ 0.0625
  - Regularization parameter C = 10

Results with best parameters

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

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)

Detecting speculated sentences. CoNLL ST'10

Optimized results for biomedical train corpus

Dataset content	#sentences used for training	#sentences used for test	SVM			Baseline		
			P	R	F	P	R	F
Abstracts only	9871	2000	0.85	0.94	0.90	0.49	0.97	0.65
Full articles only	2170	500	0.72	0.87	0.79	0.46	0.91	0.61
Abstracts and full articles	11541	3000	0.81	0.92	0.86	0.47	0.98	0.64

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

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)
- Sanchez-Graillet and Poesio (2007) develop heuristics for extracting negative protein interactions

Processing negation in biomedical texts

Determining whether a finding, disease or concept is negated

- Golding and Chapman (2003)
  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

Processing negation in biomedical texts: Negfinder

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

Described in:

Original Investigations JAMIA

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

Processing negation in biomedical texts: Negfinder

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

Processing negation in biomedical texts: Negfinder

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

Processing negation in biomedical texts: Negfinder

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.

*DIAGNOSES: ~32285:17:9. The patient has a ~48604:340:7 of ~30920:351:20 with a ~192440:379:20 as well as ~42963:411:19... He denies shortness of ~225386:703:6, ~8031:711:10, ~15967:723:5, ~85593:730:6, ~27497:738:6, ~42963:746:8, ~11991:756:8, or ~737:769:14. The patient had no other ~231216:846:10... *PHYSICAL EXAMINATION... ~332254:1683:6. ~750572:1692:13 at 6cm. 2+ ~741982:1718:18 bilaterally. No ~6318:1651:6. ...*~205041:1812:7 Irregularly ~205271:1833:9, with normal S1, split S2, with no murmurs, rubs, or gallops... He was without ~231216:2489:9 and had no evidence of ~10200:2522:5 or ~242104:2531:17... ~38056:2869:6 and ~T213890:2880:14 were negative for evidence of ~21311:2925:9.

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

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.)

Processing negation in biomedical texts: Negfinder

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").

Processing negation in biomedical texts: Negfinder

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

Processing negation in biomedical texts: Negfinder

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

	Negation found by observer	Negation not found by observer
Negation found by program	True positives: 135	False positives: 12
Negation not found by program	False negatives: 6	True negatives: 1,716

Figure 3 Results of Evaluation 2, showing performance of Negfinder on a test set of 10 documents (5 discharge summaries, 5 surgical notes), using an unbiased design of independent evaluation by a human observer and Negfinder.

Specificity: 91.8% Sensitivity: 95.7%

Processing negation in biomedical texts: Negfinder

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

Processing negation in biomedical texts: Context

Journal of Biomedical Informatics 42 (2009) 839–851

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

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

Processing negation in biomedical texts: Context

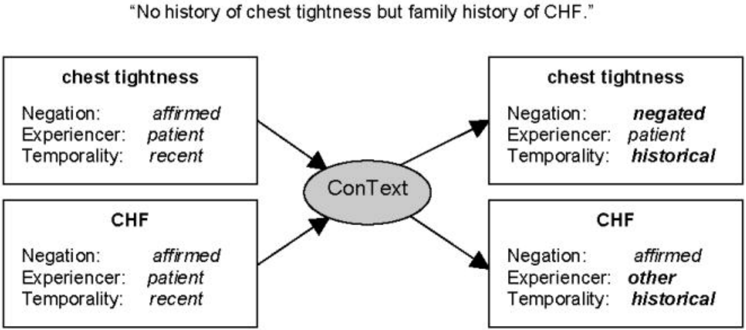
Algorithm

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

Processing negation in biomedical texts: Context



Processing negation in biomedical texts: Context

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"')

Processing negation in biomedical texts: Context

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`

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`

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 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

cause of
cause for
causes of
causes for
source of
source for
sources of
sources for
reason of
reason for
reasons of
reasons for
etiology of
etiology for
trigger event for
origin of
origin for
origins of
origins for
...

Processing negation in biomedical texts: Context

Scope of trigger terms

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$_{termination\ term}$ with shortness of breath

Processing negation in biomedical texts: Context

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"

Processing negation in biomedical texts: Context

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.

Processing negation in biomedical texts: Context

Evaluation

	R	P	F	N
(a) Negation				
Surgical pathology	.75 .30–.95	.75 .30–.95	.75	4 6%
Operative procedure	.94 .73–.99	.84 .62–.94	.89	17 10%
Radiology	.86 .71–.94	1.0 .89–1.0	.93	35 23%
Echocardiogram	.91 .78–.97	.97 .85–.97	.94	35 6%
Discharge summary	.89 .79–.94	.84 .74–.90	.86	74 18%
Emergency department	.93 .90–.95	.96 .93–.98	.95	325 36%
All	.92 .89–.94	.94 .91–.96	.93	490 22%

- 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

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.

Scope resolution: Negation

R. Morante and W. Daelemans (2009) A metalearning approach to processing the scope of negation. Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL), pages 21-29, Boulder, Colorado, June 2009. Association for Computational Linguistics.

Modelling the task

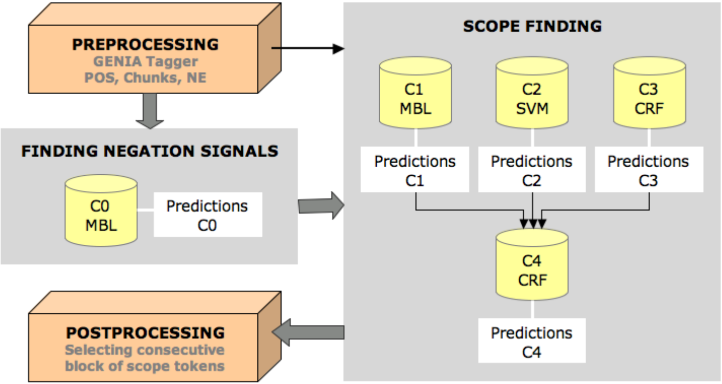
- We model the scope finding task as two consecutive classification tasks:
  1. Finding negation signals: a token is classified as being at the beginning of a negation signal, inside or outside
  2. Finding the scope: a token is classified as being the first element or the last of a scope sequence
-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

Scope resolution: Negation

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

System architecture



Scope resolution: Negation

Preprocessing

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

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

N.	TOKEN	NEG SIGNAL
1	These	O
2	syncytia	O
3	lack	B-NEG
4	activated	O
5	cells	O
6	as	O
7	determined	O
8	by	O
9	an	O
10	absence	B-NEG
11	of	O
12	staining	O
13	for	O
14	Ki-67	O
15	cell	O
16	cycle	O
17	antigen	O
18	.	O

Scope resolution: Negation

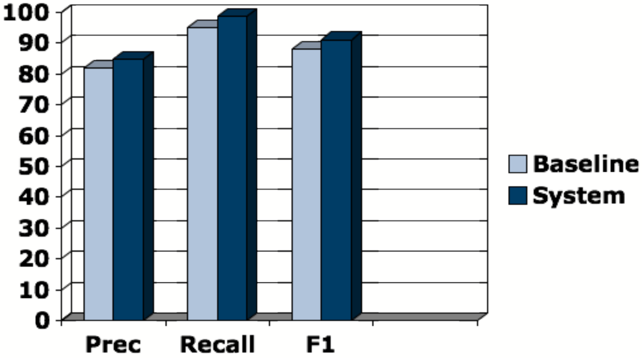
Finding negation cues: results

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

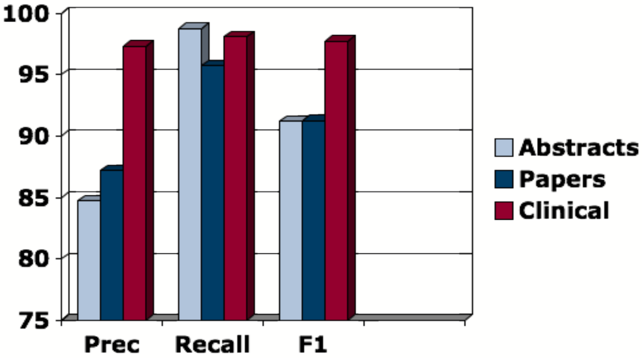
BASELINE	PREC	RECALL	F1	IAA
Abstracts	82.00	95.17	88.09	94.46
Papers	84.01	92.46	88.03	79.42
Clinical	97.31	97.53	97.42	90.70

SYSTEM	PREC	RECALL	F1
Abstracts	84.72	98.75	91.20 (+3.11)
Papers	87.18	95.72	91.25 (+3.22)
Clinical	97.33	98.09	97.71 (+0.29)

Finding negation cues: system versus baseline



Finding negation cues: results in 3 corpora



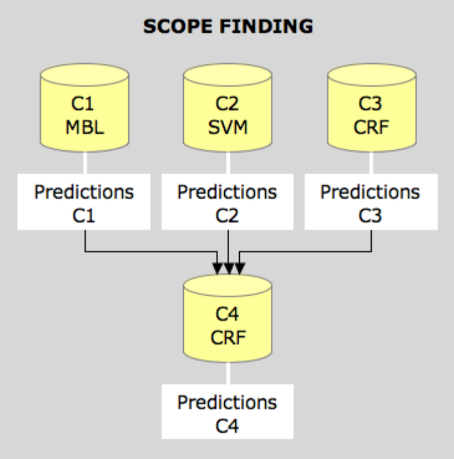
Discussion

- Cause of lower recall on papers corpus:

NOT	% negation signals	% classified correctly
Abstracts	58.89	98.25
Papers	53.22	93.68
Clinical	6.72	91.22

- 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

Scope finding



- The features used by the object classifiers and the metalearner are different

Scope resolution: Negation

Scope finding

N.	TOKEN	NEG	SCOPE
1	These	lack	O
2	syncytia	lack	O
3	lack	lack	FIRST
4	activated	lack	O
5	cells	lack	LAST
6	as	lack	O
7	determined	lack	O
8	by	lack	O
9	an	lack	O
10	absence	lack	O
11	of	lack	O
12	staining	lack	O
13	for	lack	O
14	Ki-67	lack	O
15	cell	lack	O
16	cycle	lack	O
17	antigen	lack	O
18	.	lack	O

N.	TOKEN	NEG	SCOPE
1	These	absence	O
2	syncytia	absence	O
3	lack	absence	O
4	activated	absence	O
5	cells	absence	O
6	as	absence	O
7	determined	absence	O
8	by	absence	O
9	an	absence	O
10	absence	absence	FIRST
11	of	absence	O
12	staining	absence	O
13	for	absence	O
14	Ki-67	absence	O
15	cell	absence	O
16	cycle	absence	O
17	antigen	absence	LAST
18	.	absence	O

Scope resolution: Negation

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 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 finding: postprocesing

- 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

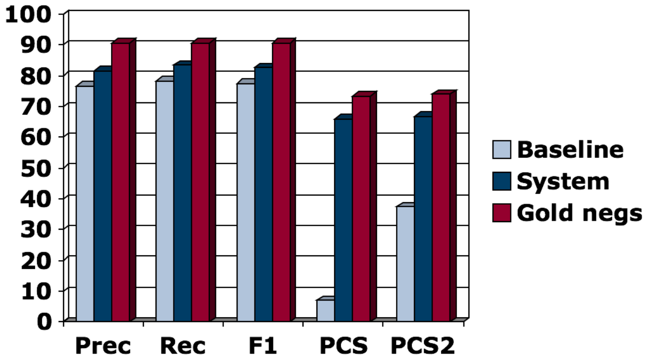
- Baseline: calculating the average length of the scope to the right of the negation signal and tagging that number of tokens as scope tokens
  Motivation: 85.70 % of scopes to the right

BASELINE	PCS	PCS-2	IAA
Abstracts	7.11	37.45	92.46
Papers	4.76	24.86	70.86
Clinical	12.95	62.27	76.29

SYSTEM	PCS	PCS-2
Abstracts	**66.07**	**66.93**
Papers	**41.00**	**44.44**
Clinical	**70.75**	**71.21**

SYSTEM gold negs	PCS	PCS-2
Abstracts	+7.29	+7.17
Papers	+9.26	+9.79
Clinical	+16.52	+16.74

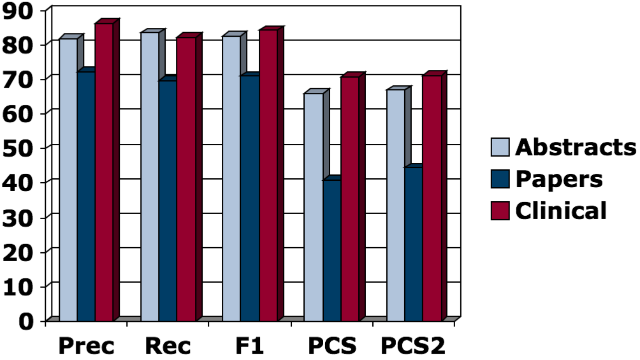
Scope resolution: Negation



- The system performs clearly better than baseline
- There is a higher upperbound calculated with gold standard negation signals

Scope resolution: Negation

Scope finding: results



- 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

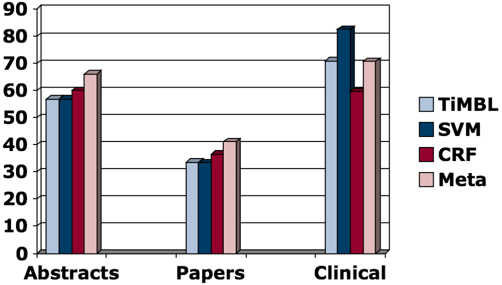
Scope finding: discussion

- 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?

NOT	Papers	Abstracts
Ambiguity (%¬neg)	25.56	14.29
Av. scope length	6.45	8.85
% Scopes left	23.28	16.41
Av. scope left	5.60	8.82

Scope finding: discussion

- The metalearner performs better than the three object classifiers (except SVMs on the clinical corpus)



Scope resolution: Negation

Resolving the scope of negation for sentiment analysis

I. Councill, R. McDonald, and L. Velikovich (2010) What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. Proceedings of the Workshop on Negation and Speculation in NLP. Uppsala.

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 depedency 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

hardly	lack	lacking	lacks
neither	nor	never	no
nobody	none	nothing	nowhere
not	n't	aint	cant
cannot	darent	dont	doesnt
didnt	hadnt	hasnt	havnt
havent	isnt	mightnt	mustnt
neednt	oughtnt	shant	shouldnt
wasnt	wouldnt	without	

Table 1: Lexicon of explicit negation cues.

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 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 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

Corpus	Prec.	Recall	F1	PCS
Reviews	81.9	78.2	80.0	39.8
BioScope	80.8	70.8	75.5	53.7

Condition	Prec.	Recall	F1	PCS
BioScope, trained on Reviews	72.2	42.1	53.5	52.2
Reviews, trained on Bioscope	58.8	68.8	63.4	45.7

- 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

Scope resolution: Negation

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

Effect on sentiment classification

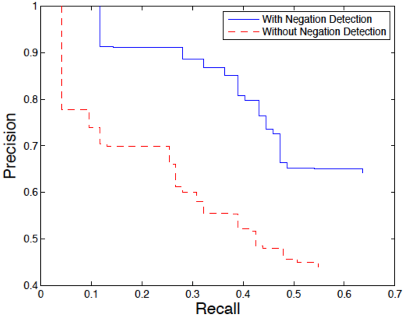


Figure 1: Precision-recall curve showing the effect of negation detection on positive sentiment prediction.

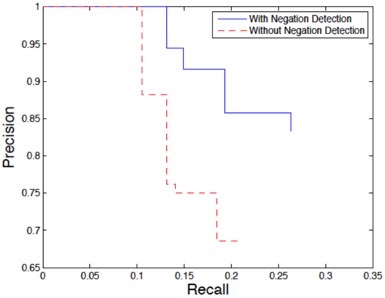


Figure 2: Precision-recall curve showing the effect of negation detection on negative sentiment prediction.

- A significant improvement is apparent at all recall levels

Scope resolution: Negation

Effect on sentiment classification

Metric	w/o Neg.	w/ Neg.	% Improv.
Positive Sentiment			
Prec.	44.0	64.1	35.9
Recall	54.8	63.7	20.0
F1	48.8	63.9	29.5
Negative Sentiment			
Prec.	68.6	83.3	46.8
Recall	21.1	26.3	6.6
F1	32.3	40.0	11.4

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.).

- 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

Scope resolution: Hedges

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]]].

Scope resolution: Hedges

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

R. Morante and W. Daelemans: Learning the scope of hedge cues in biomedical texts. In Proceedings of the BioNLP 2009 Workshop, pages 28-36, Boulder, Colorado, June 2009. Association for Computational Linguistics.

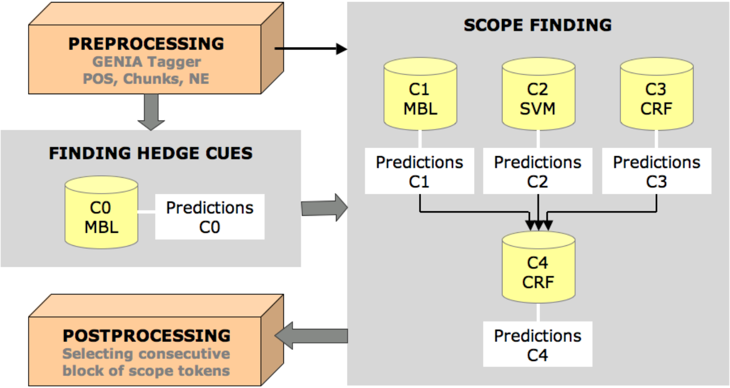
- System based on the system that processed negation cues
- Goal: investigate whether the same system can process hedge cues
- We model the scope finding task as two consecutive classification tasks:
  1. Finding negation signals
  2. Finding the scope
- BioScope corpus
  - Abstracts corpus:
    10 fold cross-validation experiments
  - Clinical and papers corpora: robustness test
    Training on abstracts - Testing on clinical and papers

Data preprocessing

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

Scope resolution: Hedges

System architecture



Results cue finding

BASELINE TOKENS

appear, apparent, apparently, believe, estimate, hypothesis, hypothesize, if, imply, likely, may, might, or, perhaps, possible, possibly, postulate, potentially, presumably, probably, propose, putatitve, should, seem, speculate, suggest, support, suppose, suspect, think, uncertain, unclear, unknown, unlikely, whether, would

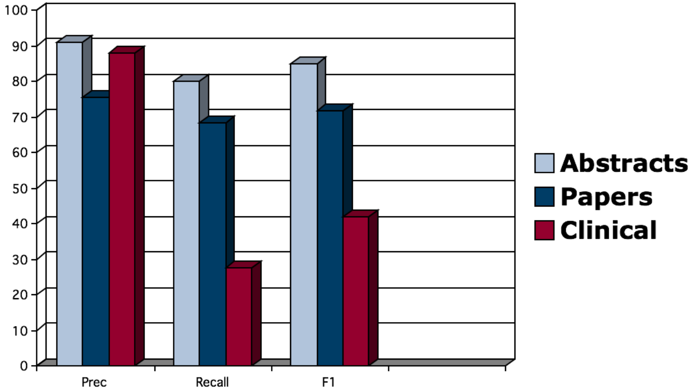
- Baseline: tagging as hedge cues a list of words extracted from the abstracts corpus

BASELINE	PREC	RECALL	F1	IAA
Abstracts	55.62	71.77	62.67	79.12
Papers	54.39	61.21	57.60	77.60
Clinical	66.55	40.78	50.57	84.01

SYSTEM	PREC	RECALL	F1
Abstracts	90.81	79.84	84.77
Papers	75.35	68.18	71.59
Clinical	88.10	27.51	41.92

Scope resolution: Hedges

Results cue finding across corpora



Discussion

- Cause of lower recall on clinical corpus:

OR	total #	% as hedge	# as hedge	% of hedges	recall
Abstracts	1062	11.29	118	**4.42**	0.129
Papers	153	16.99	27	**4.04**	0.137
Clinical	281	98.22	276	**24.62**	0.007

- 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

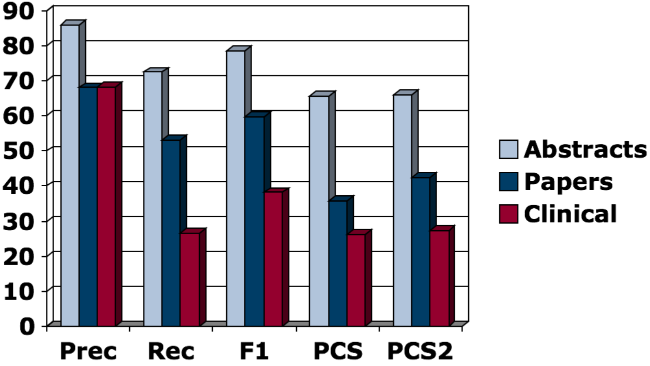
BASELINE	PCS	PCS-2
Abstracts	3.15	3.17
Papers	2.19	2.26
Clinical	2.72	3.53

SYSTEM	PCS	PCS-2
Abstracts	65.55	66.10
Papers	35.92	42.37
Clinical	26.21	27.44

SYSTEM gold cues	PCS	PCS-2
Abstracts	+11.58	+12.11
Papers	+12.02	+15.84
Clinical	+34.38	+36.50

Scope resolution: Hedges

Results scope resolution

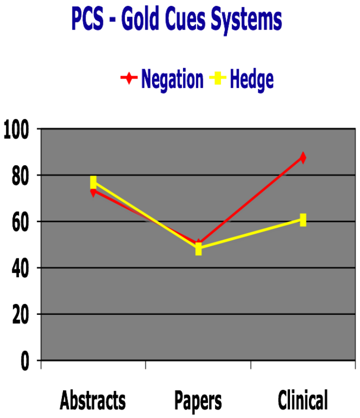


Scope resolution: Hedges

Discussion

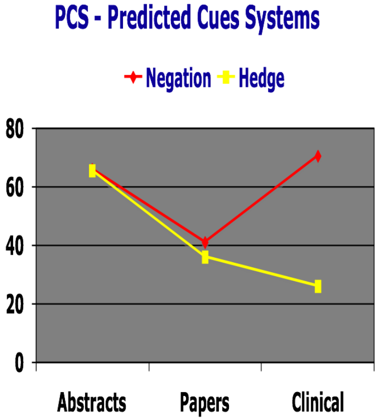
- Why are the results in papers lower?
  - 41 cues (47.00%) in papers are not in abstracts
  - Some cues that occur in abstracts and are frequent in papers get low scores. They are used differently.
    (Ex. *suggest*: 92.33 PCS in abstracts vs. 62.85 PCS in papers)
- Why are the results in clinical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



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

Scope resolution: Hedges

A. Özgür and D.R. Radev. 2009. Detecting speculations and their scopes in scientific text. In Proc. of EMNLP 2009, pages 1398-1407, Singapore.

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

Scope resolution: Hedges

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

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

Baseline 1: 14 keywords Light et al. (2004)

Baseline 2: keywords from train corpus

Results hedge detection in papers

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

Baseline 1: 14 keywords Light et al. (2004)

Baseline 2: keywords from train corpus

Scope resolution: Hedges

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

Scope resolution: Hedges

Results scope resolution

Method	Accuracy-Abstracts	Accuracy-Full text
Baseline 1	4.82	4.29
Baseline 2	67.60	42.82
Rule-based method	79.89	61.13

Baseline 1: assign scope to the whole sentence

Baseline 2: assign scope from keyword to the end of the sentence

Scope resolution: Hedges

Agarwal, Sh. and H. Yu (2010) Detecting hedge cues and their scope in biomedical text with conditional random fields. J Biomed Inform. 2010 Dec;43(6):953-61

- 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

Systems (From Agarwal et al. 2010)

System name	Detects	Features used
HedgeCue	Hedge cues	Words
BaselineCue	Hedge cues	Words
HedgeScope	Scope of a hedge cue	Words POS tags Cue phrase words not replaced with POS tags POS tags Cue phrase words not replaced with POS tags POS tags Cue phrase words replaced with custom tag 'CUE' POS tags Cue phrase words replaced with custom tag 'CUE'
BaselineScope	Scope of a hedge cue	Words
		Words

Scope resolution: Hedges

Results (From Agarwal et al. 2010)

	HedgeScope				
Features used	Words	Part of speech	Part of speech	Part of speech	Part of speech
Cue phrase identified using	–	HedgeCue	HedgeCue	BaselineCue	BaselineCue
Cue phrase replaced	–	No	Yes	No	Yes
Scope limited by	–	–	–	–	–
Recall	78.81 ± 0.02	82.47 ± 0.02	83.91 ± 0.02	90.78 ± 0.01	91.59 ± 0.01
Precision	84.82 ± 0.01	88.98 ± 0.01	88.54 ± 0.01	74.46 ± 0.04	74.6 ± 0.06
F1-score	81.7 ± 0.02	85.6 ± 0.01	**86.16 ± 0.01**	81.81 ± 0.02	82.23 ± 0.03
Accuracy	88.92 ± 0.01	91.29 ± 0.01	**91.54 ± 0.01**	87.34 ± 0.01	87.58 ± 0.02
PCS	76.79 ± 3.32	**80.0 ± 2.27**	79.73 ± 2.02	70.57 ± 3.55	70.23 ± 3.04

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 | Organiser

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

Information source: R. Farkas, V. Vincze, G. Móra, J. Csirik, and G. Szarvas. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. Proceedings of the Fourteenth Conference on Computational Natural Language Learning: Shared Task, pages 1-12

Scope resolution: Hedges

Approaches (Table from Farkas et al. 2010)

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

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.
- Extrict match: including or excluding punctuations, citations or some bracketed expressions

Scope resolution: Hedges

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

	Abstracts	Papers	Test
#Documents	1273	9	15
#Sentences	11871	2670	5003
%Hedge sent.	17.70	19.44	15.75
#Hedges	2694	682	1043
#AvL. of sent.	30.43	27.95	31.30
#AvL. of scopes	17.27	14.17	17.51

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

	Train	Test
#Documents	2186	2737
#Sentences	11111	9634
%Hedge sentences	22.36	23.19
#Hedges	3133	3143
#AvL. of sentences	23.07	20.82

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

(Tables from B. Tang, X. Wang, X. Wang, B. Yuan, and Sh. Fan. 2010. A Cascade Method for Detecting Hedges and their Scope in Natural Language Text. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task, pages

Scope resolution: Hedges

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

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

Table 3: Task1 biological results (type $\in$ {Closed(C), Cross(X), Open(O)}).

Scope resolution: Hedges

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

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

Table 2: Task2 results (type $\in$ {Closed(C), Open(O)}).

Scope resolution: Hedges

Approaches (Table from Farkas et al. 2010)

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

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

Finding hedge cues in biomedical texts - best system

B. Tang, X. Wang, X. Wang, B. Yuan, and Sh. Fan. 2010. A Cascade Method for Detecting Hedges and their Scope in Natural Language Text. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task, pages 25-29.

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

Scope resolution: Hedges

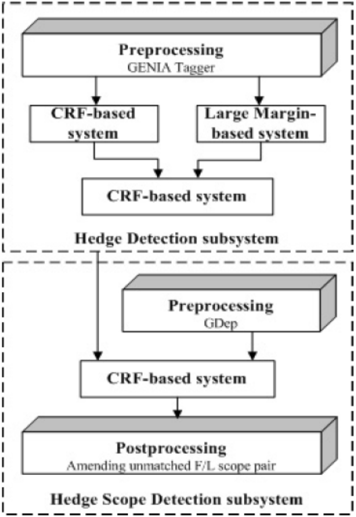


Figure 1: System architecture

(Figure from Tang et al. 2010)

Scope resolution: Hedges

Features hedge detection - first layer

- Word andWord 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

Scope resolution: Hedges

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 $\geq$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

Scope resolution: Hedges

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

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.

Results hedge detection (Table from Tang et al. 2010)
Algorithms: CRR++ and SVMLight

Corpus	System	Prec.	Recall	F1
BioScope	CRF	87.12	86.46	86.79
	LM	85.24	87.72	86.46
	CAS	85.03	87.72	86.36
Wikipedia	CRF	86.10	35.77	50.54
	LM	82.28	41.36	55.05
	CAS	82.28	41.36	55.05

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++

HD subsystem	Precision	Recall	F1
gold	57.92	55.95	56.92
CRF	52.36	48.40	50.30
LM	51.06	48.89	49.95
CAS	50.96	48.98	49.95

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.

Scope resolution: Hedges

Finding the scopes of hedge cues in biomedical texts - best system

Roser Morante, Vincent Van Asch, and Walter Daelemans. 2010. Memory-based Resolution of Insentence Scopes of Hedge Cues. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task, pages 48–55.

- Memory-based learning
- Features from dependency trees

Scope resolution: Hedges

A sentence where scopes should be found

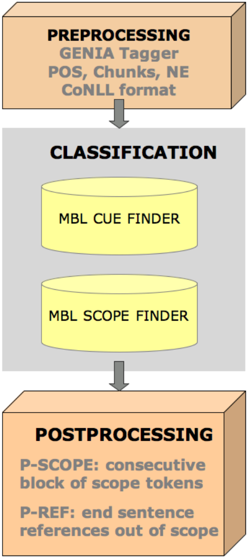
<The conservation from Drosophila to mammals of these two structurally distinct but functionally similar E3 ubiquitin ligases is likely to reflect a combination of evolutionary advantages associated with: **(i)** specialized expression pattern, as evidenced by the cell-specific expression of the neur gene in sensory organ precursor cells **[52]**; **(ii)** specialized function, as <suggested by the role of murine MIB in TNFa signaling> **[32]**; **(iii)** regulation of protein stability, localization, and/or activity>.

Scope resolution: Hedges

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



Scope resolution: Hedges

- 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

	WORD	LEMMA	PoS	CHUNK	D	LABEL	C	S
1	The	The	DT	B-NP	3	NMOD	O	O O
2	structural	structural	JJ	I-NP	3	NMOD	O	O O
3	evidence	evidence	NN	I-NP	4	SUB	O	O O
4	lends	lend	VBZ	B-VP	0	ROOT	B	F O
5	strong	strong	JJ	B-NP	6	NMOD	I	O O
6	support	support	NN	I-NP	4	OBJ	I	O O
7	to	to	TO	B-PP	6	NMOD	O	O O
8	the	the	DT	B-NP	11	NMOD	O	O O
9	inferred	inferred	JJ	I-NP	11	NMOD	B	O F
10	domain	domain	NN	I-NP	11	NMOD	O	O O
11	pair	pair	NN	I-NP	7	PMOD	O	L L
12	,	,	,	O	4	P	O	O O
13	resulting	result	VBG	B-VP	4	VMOD	O	O O
14	in	in	IN	B-PP	13	VMOD	O	O O
15	a	a	DT	B-NP	18	NMOD	O	O O
16	high	high	JJ	I-NP	18	NMOD	O	O O
17	confidence	confidence	NN	I-NP	18	NMOD	O	O O
18	set	set	NN	I-NP	14	PMOD	O	O O
19	of	of	IN	B-PP	18	NMOD	O	O O
20	domain	domain	NN	B-NP	21	NMOD	O	O O
21	pairs	pair	NNS	I-NP	19	PMOD	O	O O
22	.	.	.	O	4	P	O	O O

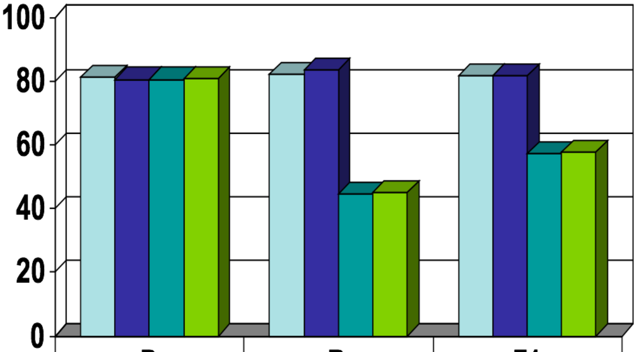
Evaluation of the conversion of the corpus into CoNLL format

TASK 1 (cues)			TASK 2 (scope)	
WIKI	BIO-ART	BIO-ABS	BIO-ART	BIO-ABS
100.00	100.00	100.00	99.10	99.66

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



	P	R	F1
BIO	81,15	82,28	81,71
BIO-cd	80,54	83,29	81,89
WIKI	80,55	44,49	57,32
WIKI-cd	80,64	44,94	57,71

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

Scope resolution: Hedges

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.)

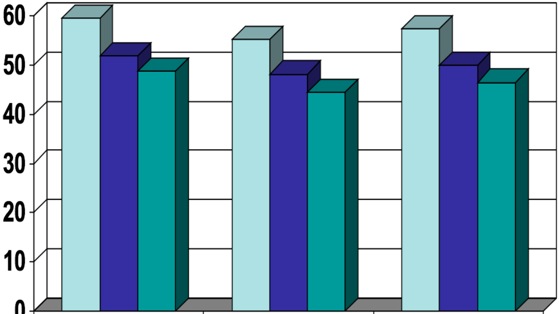
Scope resolution: Hedges

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



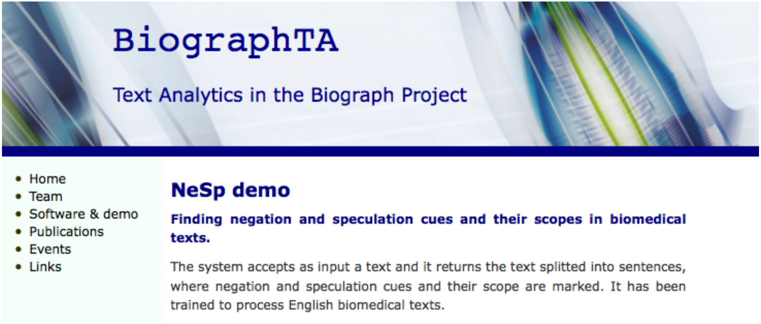
	P	R	F1
BIO	59,62	55,18	57,32
BIO before P-REF	51,98	48,2	50,02
BIO before P-SCOPE	48,82	44,43	46,52

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

BiographTA: NeSp scope labeler



BiographTA: NeSp scope labeler

[NEG0 The question [SPEC1 **whether** Vif alters transcription controlled by the A3G promoter SPEC1] has **not** been analyzed so far NEG0] .

Our analysis [SPEC2 **indicates that** transcription from the A3G promoter is unaffected by Vif or other HIV-1 proteins SPEC2] .

Taken together , in T cell lines , the A3G promoter appears constitutively active .

Scope resolution: Hedges

Øvrelid,L., E. Velldal, and S. Oepen (2010) Syntactic Scope Resolution in Uncertainty Analysis Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010) Beijing, China, 2010

- 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

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)

	Configuration	F_1
BSP	Default, Gold Cues	45.21
	Rules, Gold Cues	72.31
	Rules, System Cues	64.77
BSE	Rules, Gold Cues	66.73
	Rules, System Cues	55.75

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

Outline

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/`

J-D. Kim, T. Ohta, S. Pyysalo, Y. Kano, and J. Tsujii (2009) Overview of BioNLP'09 Shared Task on Event Extraction Proceedings of the Workshop on BioNLP: Shared Task, pages 1-9, Boulder, Colorado, ACL.

Results (From Kim et al. 2009)

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

Outline

Modality tagging

K. Baker, M. Bloodgood, B. J. Dorr , N. W. Filardo , L. Levin, and Christine Piatko. A modality lexicon and its use in automatic tagging. Proceedings of LREC 2010, pages 1402–1407.

- **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

Modality tagging

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)) (. .)))
```

Modality tagging

Evaluation

- Agreement between taggers (Kappa)
  - 0.82 for triggers
  - 0.76 for targets
- Precision of structure-based tagger on 249 sentences: 86.3 %

Modality tagging

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

Outline

Belief categorisation: Committed belief tagging

Prabhakaran, V., O. Rambow and M. Diab (2010) Automatic committed belief tagging. Proceedings of COLING 2010, pages 1014–1022.

- "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 of 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 classifiation 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)

Class	Feature Set	Parm	P	R	F
YAMCHA - Joint Model					
L_C	POS, whichModalAmI, verbType, isNumeric	CW=3	61.9	52.7	56.9
$L_N S_N$	POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModalIsMyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, amVBwithDaughterTo, haveDaughterWh, haveDaughterShould	CW=0	62.5	57.5	59.9
$L_C S_N$	POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModalIsMyDaughter, whichAuxIsMyDaughter, haveDaughterShould	CW=2	67.4	58.1	62.4
$L_C S_C$	POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModalIsMyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughterWh, haveDaughterShould	CW=2	68.5	60.0	64.0
MALLET - Joint Model					
L	POS, whichModalAmI, verbType	GV=1	55.1	45.0	49.6
LS	POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModalIsMyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughterWh, haveDaughterShould	GV=1	64.5	54.4	59.0
Pipeline Model					
$L_C S_C$	POS, whichModalAmI, parentPOS, haveReportingAncestor, whichModalIsMyDaughter, haveDaughterPerfect, whichAuxIsMyDaughter, haveDaughterWh, haveDaughterShould	CW=2	49.8	42.9	46.1

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)

Features that performed well			
1	isNumeric	L	Word is Alphabet or Numeric?
2	POS	L	Word's POS tag
3	verbType	L	Modal/Aux/Reg (= 'nil' if the word is not a verb)
4	whichModalAmI	L	If I am a modal, what am I? (= 'nil' if I am not a modal)
3	amVBwithDaughterTo	S	Am I a VB with a daughter *to*?
4	haveDaughterPerfect	S	Do I have a daughter which is one of *has, have, had*?
5	haveDaughterShould	S	Do I have a daughter *should*?
6	haveDaughterWh	S	Do I have a daughter who is one of *where, when, while, who, why*?
7	haveReportingAncestor	S	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*?
8	parentPOS	S	What is my parent's POS tag?
9	whichAuxIsMyDaughter	S	If I have a daughter which is an auxiliary, what is it? (= 'nil' if I do not have an auxiliary daughter)
10	whichModalIsMyDaughter	S	If I have a daughter which is a modal, what is it? (= 'nil' if I do not have a modal daughter)

Belief categorisation: Committed belief tagging

Features that were not useful (Table from Prabhakaran et al. 2010)

1	Lemma	L	Word's Lemma
2	Stem	L	Word stem (Using Porter Stemmer)
3	Drole	S	Deep role (drole in MICA features)
4	isRoot	S	Is the word the root of the MICA Parse tree?
5	parentLemma	S	Parent word's Lemma
6	parentStem	S	Parent word stem (Using Porter Stemmer)
7	parentSupertag	S	Parent word's super tag (from Penn Treebank)
8	Pred	S	Is the word a predicate? (pred in MICA features)
9	wordSupertag	S	Word's Super Tag (from Penn Treebank)

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

Outline

Processing contradiction and contrast

S. Harabagiu, A. Hickl and F. Lacatusu (2006) Negation, contrast and contradiction in text processing. Proceedings of the 21st national conference on Artificial intelligence - Volume 1, pages 755-762

Processing contradiction and contrast

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

(a)	Contradiction	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.
		T2:	Defending champion Andy Roddick **never** took on Joachim Johansson.
(b)	Contradiction	T3:	In California, one hundred twenty Central Americans, due to be deported, began a hunger strike when their deportation was delayed.
		T4:	A hunger strike was **called off.**
(c)	Contradiction	T5:	The explosion wounded the arm of Beatriz Iero, damaged the doors and walls of the offices, and broke the windows of neighboring buildings.
		T6:	Beatriz Iero **emerged unscathed** from an explosion.

(From Harabagiu et al. 2006)

Processing contradiction and contrast

- 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**.

Processing contradiction and contrast

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)

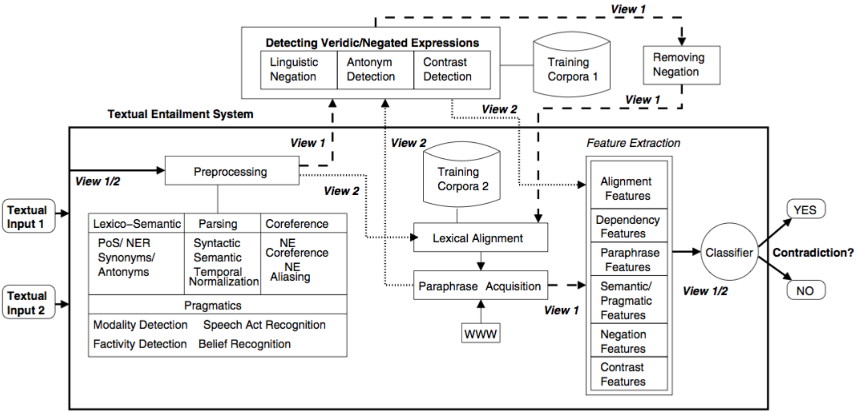


Figure 3: The Architecture Used for Recognizing Contradictions with the Help of Textual Entailment.

- **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. Detec negated entities: any noun phrase that falls within the scope of an overt negative quantifier ("no") or a non-veridical quntifier ("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

Processing contradiction and contrast

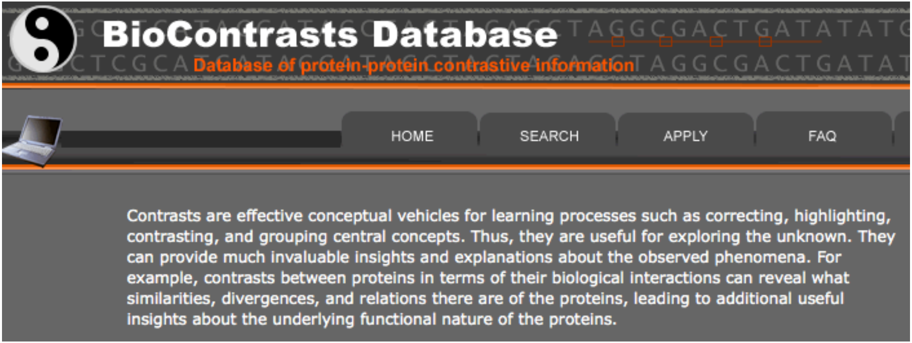
Jung-jae Kim, Zhuo Zhang, Jong C. Park and See-Kiong Ng (2006) BioContrasts: extracting and exploiting protein-protein contrastive relations from biomedical literature. Bioinformatics 22 (5): 597-605.

Processing contradiction and contrast

- 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.

Processing contradiction and contrast

http://biocontrasts.biopathway.org/index.php



Processing contradiction and contrast

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).

Processing contradiction and contrast

PPI Contrast

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.

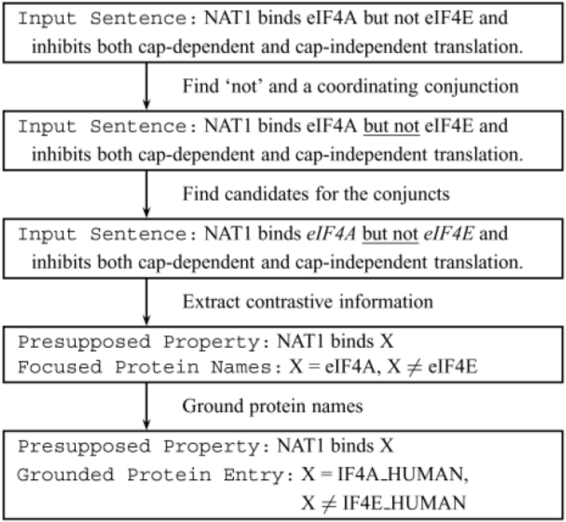
Extracting contrastive relations

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.

Extracting contrastive relations



Processing contradiction and contrast

PMID: 10762618
Year: 2000
Sent: **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

Processing contradiction and contrast

Identifying subclausal coordinations

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

General patterns	Specific patterns	
A but not B	NP but not NP	PREP NP but not PREP NP
	V NP but not V NP	V PREP NP but not V PREP NP
not A but B	not NP but NP	not PREP NP but PREP NP
	not V NP but V NP	not V PREP NP but V PREP NP
A, not B	NP, not NP	PREP NP, not PREP NP

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

Processing contradiction and contrast

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'}).

Negative patterns	Positive patterns
$Subj_C$ not V_1 Obj_2	$Subj'_C$ V'_1 Obj'_2
$Subj_2$ not V_1 Obj_C	$Subj'_2$ V'_1 Obj'_C
$Subj_C$ not V_1 $PREP_3$ Obj_2	$Subj'_C$ V'_1 $PREP'_3$ Obj'_2
$Subj_2$ not V_1 $PREP_3$ Obj_C	$Subj'_2$ V'_1 $PREP'_3$ Obj'_C
$Subj_C$ BeV not ADJ_1 $PREP_3$ NP_2	$Subj'_C$ BeV ADJ'_1 $PREP'_3$ NP'_2
$Subj_2$ BeV not ADJ_1 $PREP_3$ NP_C	$Subj'_2$ BeV ADJ'_1 $PREP'_3$ NP'_C
NN_1 of NP_C with NP_2 not V	N'_1 of NP'_C with NP'_2 V
NN_1 between NP_C and NP_2 not V	NN'_1 between NP'_C and NP'_2 V

Processing contradiction and contrast

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’.

Processing contradiction and contrast

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)

Processing contradiction and contrast

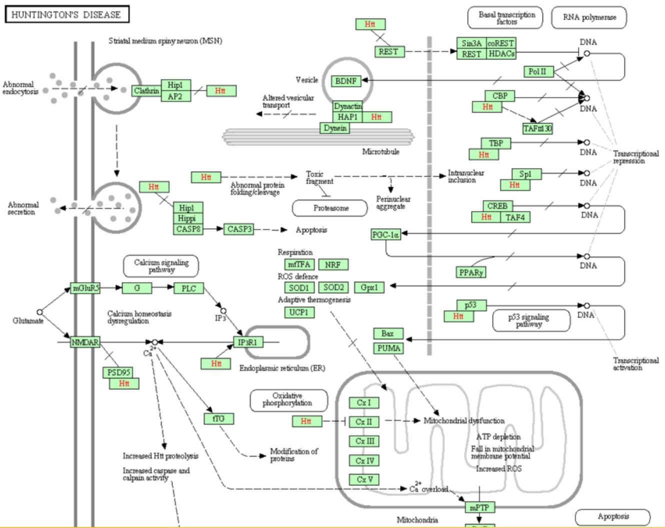
Refining pathway roles of similar proteins

- 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

http://www.genome.jp/kegg/pathway/hsa/hsa05016.html



Processing contradiction and contrast

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

Visualising negation features

D. Oelke, P. Bak, D. A. Keim, M. Last, G. Danon, Visual Evaluation of Text Features for Document Summarization and Analysis, Proceedings of the IEEE Symposium on Visual Analytics Science and Technology 2008 (IEEE VAST 2008), Columbus, OH, USA, 2008.

"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 documentfingerprinting 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

Visualising negation features

Pipeline for visual evaluation of text features applied for document summarization and analysis

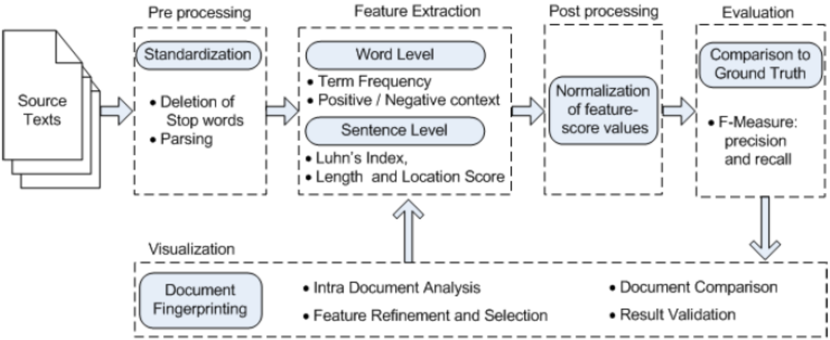


Figure from Oelke et al. 2008:76

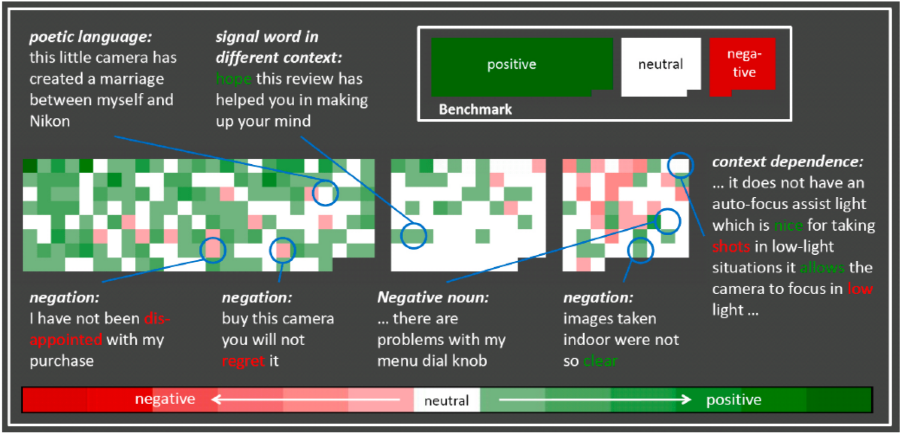
Visualising negation features

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 substracted 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)



Visualising negation features

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)

Outline

References

Detecting Speculative Language

- Di Marco, Ch. and R. E. Mercer (2004) Hedging in Scientific Articles as a Means of Classifying Citations. In Proceedings of Working Notes of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, Stanford University.
- Ganter, V. and M. Strube (2009) Finding hedges by chasing weasels: Hedge detection using wikipedia tags and shallow linguistic features. In Proceedings of the ACL-IJCNLP 2009 Conference Short Papers, pages 173-176.
- Kilicoglu, H. and S. Bergler (2008) Recognizing Speculative Language in Biomedical Research Articles: A Linguistically Motivated Perspective. In Proceedings of Current Trends in Biomedical Natural Language Processing (BioNLP), Columbus, Ohio, USA.
- Light, M., X. Ying Qiu, and P. Srinivasan (2004) The Language of Bioscience: Facts, Speculations, and Statements in between. In Proceedings of the HLT BioLINK.
- Medlock, B. (2008) Exploring hedge identification in biomedical literature. Journal of Biomedical Informatics, 41:636-54.
- Medlock, B. and T. Briscoe (2007) Weakly Supervised Learning for Hedge Classification in Scientific Literature. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics.
- Szarvas, G. (2008) Hedge classification in biomedical texts with a weakly supervised selection of keywords. In Proceedings of the ACL-08: HLT.

References

Velldal, E. (2011) Predicting Speculation: A Simple Disambiguation Approach to Hedge Detection in Biomedical Literature Journal of Biomedical Semantics, Vol. 2, Suppl. 5, BioMed Central, October 2011.

Verbeke, M., P. Frasconi, V. Van Asch, R. Morante.3, W. Daelemans, and L. De Raedt. Kernel-based Logical and Relational Learning with kLog for Hedge Cue Detection. ILP 2011.

Processing Negation in Biomedical Texts

S. Boytcheva, A. Strupchanska, E. Paskaleva, and D. Tcharaktchiev (2005) Some aspects of negation processing in electronic health records. In Proc. of International Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries, pages 1-8, Borovets, Bulgaria.

W.W. Chapman,W. Bridewell, P. Hanbury, G. F. Cooper, and B.G. Buchanan (2001) A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform, 34:301-310.

P. L. Elkin, S. H. Brown, B. A. Bauer, C.S. Husser, W. Carruth, L.R. Bergstrom, and D. L. Wahner-Roedler (2005) A controlled trial of automated classification of negation from clinical notes. BMC Medical Informatics and Decision Making, 5(13).

I. M. Goldin and W.W. Chapman (2003) Learning to detect negation with 'Not' in medical texts. In Proceedings of ACM-SIGIR 2003.

References

S. Goryachev, M. Sordo, Q.T. Zeng, and L. Ngo (2006) Implementation and evaluation of four different methods of negation detection. Technical report, DSG.

Y. Huang and H.J. Lowe (2007) A novel hybrid approach to automated negation detection in clinical radiology reports. J Am Med Inform Assoc, 14(3):304-311

A.G. Mutalik, A. Deshpande, and P.M. Nadkarni (2001) Use of general-purpose negation detection to augment concept indexing of medical documents. a quantitative study using the UMLS. J Am Med Inform Assoc, 8(6):598-609.

L. Rokach, R.Romano, and O. Maimon (2008) Negation recognition in medical narrative reports. Information Retrieval Online.

O. Sanchez-Graillet and M. Poesio (2007) Negation of protein-protein interactions: analysis and extraction. Bioinformatics, 23(13):424-432

References

Resolving the Scope of Negation

Agarwal, Sh. and H. Yu (2010) Biomedical negation scope detection with conditional random fields. J Am Med Inform Assoc. 2010 Nov 1;17(6):696-701

Blanco, E. and D. Moldovan (2011) Semantic Representation of Negation Using Focus Detection. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 581-589, Portland, Oregon, ACL.

Councill, I., R. McDonald, and L. Velikovich (2010) What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. Proceedings of the Workshop on Negation and Speculation in NLP. Uppsala.

Morante, R., A. Liekens, and W. Daelemans (2008) Learning the Scope of Negation in Biomedical Texts. In Proceedings of EMNLP

Morante, R. and W. Daelemans (2009) A Metalearning Approach to Processing the Scope of Negation. In Proceedings of CoNLL

Resolving the Scope of Hedges

Agarwal, Sh. and H. Yu (2010) Detecting hedge cues and their scope in biomedical text with conditional random fields. J Biomed Inform. 2010 Dec;43(6):953-61

Farkas, R., V. Vincze, G. Móra, J. Csirik, and G. Szarvas (2010) The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. Proceedings of the Fourteenth Conference on Computational Natural Language Learning: Shared Task, pages 1-12

Morante, R. and W. Daelemans (2009) Learning the scope of hedge cues in biomedical texts. In Proceedings of the BioNLP 2009 Workshop, pages 28-36.

Morante, R., V. Van Asch, and W. Daelemans (2010) Memory-based Resolution of Insentence Scopes of Hedge Cues. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task, pages 48–55.

Özgür, A. and D.R. Radev (2009) Detecting speculations and their scopes in scientific text. In Proceedings of EMNLP 2009, pages 1398-1407.

Øvrelid,L., E. Velldal, and S. Oepen (2010) Syntactic Scope Resolution in Uncertainty Analysis Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010) Beijing, China, 2010

References

Tang, B., X. Wang, X. Wang, B. Yuan, and Sh. Fan. 2010. A Cascade Method for Detecting Hedges and their Scope in Natural Language Text. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task, pages 25-29.

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

Prabhakaran, V., O. Rambow and M. Diab (2010) Automatic committed belief tagging. Proceedings of COLING 2010, pages 1014–1022.

Diab, M. T., L. Levin, T. Mitamura, O. Rambow, V. Prabhakaran, and W. Guo (2009) Committed belief annotation and tagging. In ACL-IJCNLP '09: Proceedings of the Third Linguistic Annotation Workshop, pages 68-73.

Contradiction

Harabagiu, S., A. Hickl and F. Lacatusu (2006) Negation, contrast and contradiction in text processing. Proceedings of the 21st national conference on Artificial intelligence - Volume 1, pages 755-762

Kim, J., Zh. Zhang, J. C. Park and See-Kiong Ng (2006) BioContrasts: extracting and exploiting protein-protein contrastive relations from biomedical literature. Bioinformatics 22 (5): 597-605.

Part IV

Modality and Negation in Applications

Outline

Outline

Information sources

B. Pang and L. Lee (2008) Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol. 2, Nos 1-2, 1-135.

M. Wiegand, A. Balahur, B. Roth, D. Klakow and A. Montoyo (2010) A survey of the role of negation in sentiment analysis. Proc. of the Workshop on Negation and Speculation in Natural Language Processing, Uppsala, pages 60-68.

B. Liu. Opinion Mining and Sentiment Analysis: NLP Meets Social Sciences. `http://www.cs.uic.edu/~liub/FBS/Liu-Opinion-Mining-STSC.ppt`

Sentiment analysis

- Types of textual information
  - Facts
  - Opinions
- Most current information processing sytems 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)

Sentiment analysis

Sentiment analysis (Pang and Lee 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 classification feature based

"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)

Negation in sentiment analysis: ML approaches

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

Negation in sentiment analysis: benchmark

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:

- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, *Proceedings of EMNLP 2002*.
- Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, *Proceedings of ACL 2004*.
- Bo Pang and Lillian Lee, Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales, *Proceedings of ACL 2005*.

We also have available an addtional 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.)
- sentence polarity dataset v1.0 (includes sentence polarity dataset README v1.0: 5331 positive and 5331 negative processed sentences / snippets. Introduced in Pang/Lee ACL 2005. Released July 2005.

- 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 Treloar, 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.

Negation in sentiment analysis: ML approaches

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

Negation in sentiment analysis: ML approaches

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?

    [No politically prudent Israeli]$_{subject}$ could support$_{polarpred}$ 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 Recogntion 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 MPQA, please visit Dr. Janyce Wiebe'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.

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

Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan (2005). Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. *Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005).*

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:

    $Polarity([\text{NP1}]^{-}\ [\text{IN}]\ [\text{NP2}]^{-}) = +$

    $[\text{lack}]^{-}{}_{NP1}\ [\text{of}]_{IN}\ [\text{crime}]^{-}{}_{NP2}$ 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

Negation in sentiment analysis: using negation in lexicon induction

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.

Negation in sentiment analysis: limits of negation modeling

- 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

Recognizing textual entailment

de Marneffe, M. C., B. MacCartney, T. Grenager, D. Cer, A. Rafferty, and Ch. D. Manning (2006) Learning to distinguish valid textual entailments. In Proceedings of the Second PASCAL Challenges Workshop on Recognising 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 quantfiers (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.

Recognizing textual entailment

Snow, R., L. Vanderwende, and A. Menezes (2006) Effectively using syntax for recognizing false entailment. In Proceedings of HLT-NAACL, pages 33-40, Morristown, NJ, USA. ACL.

- 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 mistmatch.
  - If the mismatch exists, it is predicted that the entailment is false.

Recognizing textual entailment

Hickl, A. and J. Bensley (2007) A discourse commitment-based framework for recognizing textual entailment. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pages 171-176, Stroudsburg, PA, USA. ACL

- 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.

Outline

Machine translation

Baker K., M. Bloodgood, B. J. Dorr , N. W. Filardo , L. Levin, and Ch. Piatko (2010) A modality lexicon and its use in automatic tagging. Proceedings of LREC 2010, pages 1402–1407.

Baker, K., M. Bloodgood, Ch. Vallison-Burch, B. J. Dorr, N. W. Filardo, L. Levin, S. Miller, Ch. Piatko (2010) Semantically-Informed Syntactic Machine Translation: A Tree-Grafting Approach. Proceedings of AMTA 2010.

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

Outline

Negfinder

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

Described in:

Original Investigations JAMIA

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

ConText

Journal of Biomedical Informatics 42 (2009) 839–851

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

ELSEVIER

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

Conway, M., S. Doan, and N. Collier (2009) Using hedges to enhance a disease outbreak report text mining system. In Proceedings of the BioNLP 2009 workshop 2009, pages 142-143, Boulder, Colorado. ACL.

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`

Text mining: BioCaster

- 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 classifer 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 frequecy of hedge cues in 10,000 Reuters documents.

Outline

Identifying the structure of scientific articles

Grabar, N. and Th. Hamon (2009) Exploitation of speculation markers to identify the structure of biomedical scientifc writing. In AMIA 2009 Symposium Proceedings.

- 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 acoording 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*.

Outline

Trustworthiness detection

Su, Q., Ch-R. Huang, and H. Kai-yun Chen (2010) Evidentiality for text trustworthiness detection. In Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, pages 10-17, Uppsala, Sweden. ACL.

- 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

References: sentiment analysis

Carvalho, P., L. Sarmento, M. J. Silva, and E. de Oliveira (2009) Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). In Proceedings of CIKM-Workshop TSA

Choi, Y. and C. Cardie (2008) Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In Proceedings of EMNLP

Minqing, H.and B. Liu (2004) Mining and summarizing customer reviews. In KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168-177, New York, NY, USA. ACM.

Liu, J. and S. Seneff (2009) Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In Proceedings of EMNLP

Pang B., L. Lee, and S. Vaithyanathan (2002) Thumbs up? Sentiment Classification Using Machine Learning Techniques. In Proceedings of EMNLP

Pang B. and L. Lee (2008) Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, Vol. 2, Nos 1-2, 1-135

Polanyi L. and A. Zaenen (2004) Context Valence Shifters. In Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text

References: sentiment analysis

Turney, P. D. (2002) Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. ACL 2002: 417-424

Wiebe J., Th. Wilson, R. Bruce, M. Bell, and M. Martin (2004) Learning subjective language. Comput. Linguist., 30(3):277-308

Wiegand, M., A. Balahur, B. Roth, D. Klakow and A. Montoyo (2010) A survey of the role of negation in sentiment analysis. Proc. of the Workshop on Negation and Speculation in Natural Language Processing, Uppsala, pages 60-68

Wilson, T., J. Wiebe, and P. Hoffmann (2005) Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In Proceedings of HLT/EMNLP

Wilson, T., J. Wiebe, and P. Hoffmann (2009) Recognizing Contextual Polarity: An Exploration for Phrase-level Analysis. Computational Linguistics, 35:3

Zagibalov, T. and J. Carroll (2008) Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text. In Proceedings of COLING

References

Machine Translation

Baker K., M. Bloodgood, B. J. Dorr , N. W. Filardo , L. Levin, and Ch. Piatko (2010) A modality lexicon and its use in automatic tagging. Proceedings of LREC 2010, pages 1402–1407.

Baker, K., M. Bloodgood, Ch. Vallison-Burch, B. J. Dorr, N. W. Filardo, L. Levin, S. Miller, Ch. Piatko (2010) Semantically-Informed Syntactic Machine Translation: A Tree-Grafting Approach. Proceedings of AMTA 2010.

Textual Entailment

de Marneffe, M. C., B. Maccartney, T. Grenager, D. Cer, A. Rafferty, and Ch. D. Manning (2006) Learning to distinguish valid textual entailments. In Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment.

Hickl, A. and J. Bensley (2007) A discourse commitment-based framework for recognizing textual entailment. In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pages 171-176, Stroudsburg, PA, USA. ACL

Snow, R., L. Vanderwende, and A. Menezes (2006) Effectively using syntax for recognizing false entailment. In Proceedings of HLT-NAACL, pages 33-40, Morristown, NJ, USA. ACL.

Classifying citations

Marco, Ch, F. Kroon, and R. Mercer (2006) Using Hedges to Classify Citations in Scientific Articles. Computing Attitude and Affect in Text Theory and Applications, pages: 247-263