Computational Linguistics Lecture: Textual Entailment

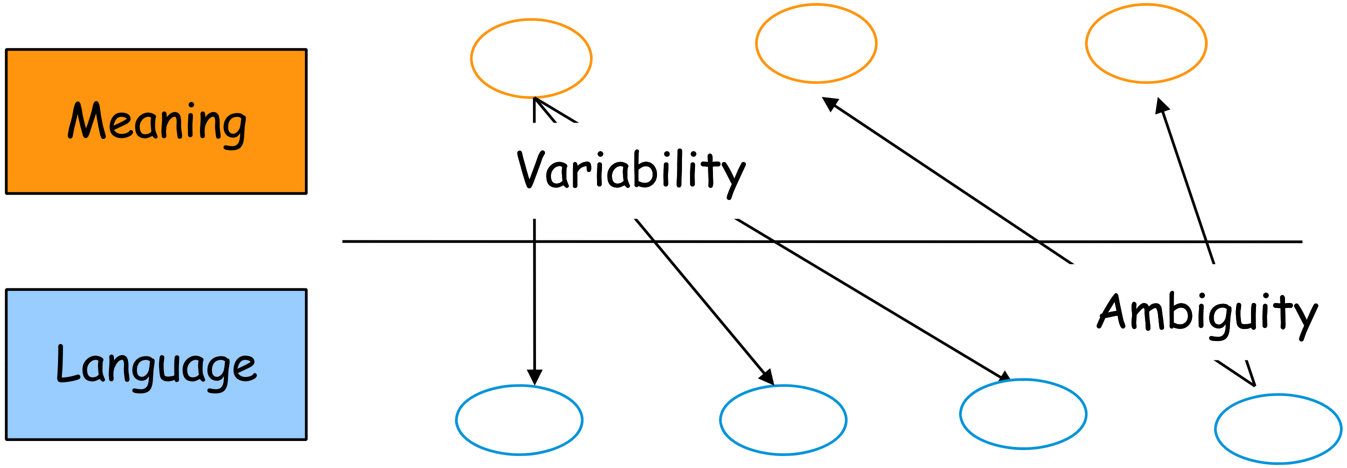
Elena Cabrio, FBK-Irst
mail: cabrio@fbk.eu

Outline:

- Introduction
  - Language variability
  - Meaning representation
  - Models of interpretation

- Textual Entailment
  - Motivation
  - Textual Entailment as a task: RTE
  - Textual Entailment as a new framework for applied semantics
  - Approaches

Natural Language and Meaning:

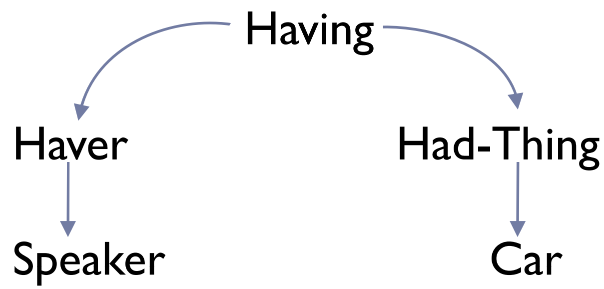


Meaning Representation Languages:

"I have a car".

$\exists x,y\ Having(x) \wedge Haver(Speaker,x) \wedge HadThing(y,x) \wedge Car(y)$

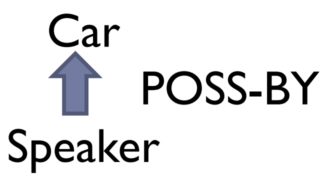
First Order Logic



Semantic Network graph

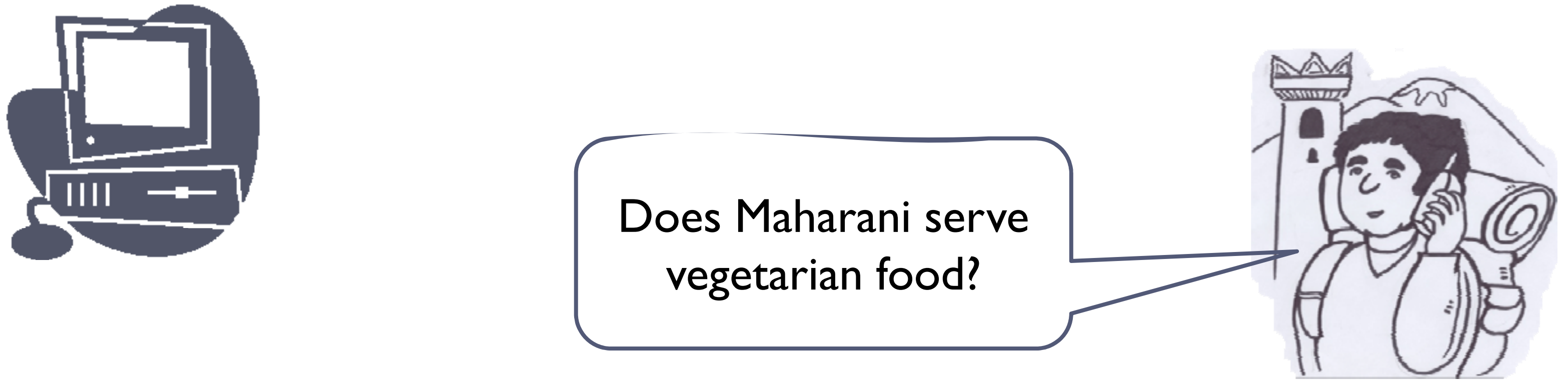
```
Having
Haver:     Speaker
HadThing:  Car
```

Frame-based representation



Conceptual Dependency diagram

Why do we need meaning representation?



- ***Basic requirement:*** to determine the relationship between the meaning of a sentence and the world as we know it.
- Most straightforward implementation → a system that compares the representation of the meaning of an input againsta the representation in its ***knowledge base***.

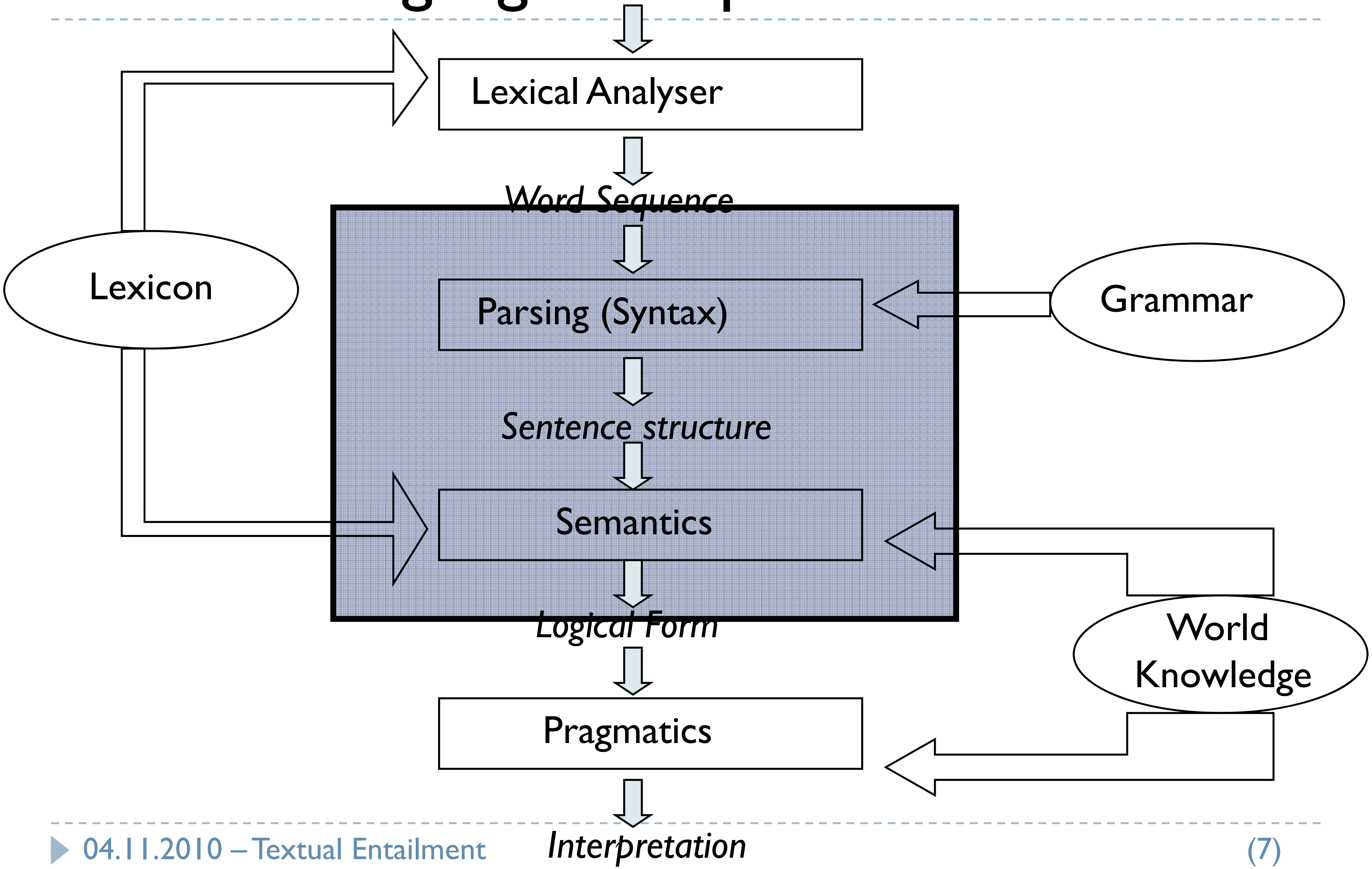
Why do we need meaning representation?

Does Maharani serve vegetarian food?

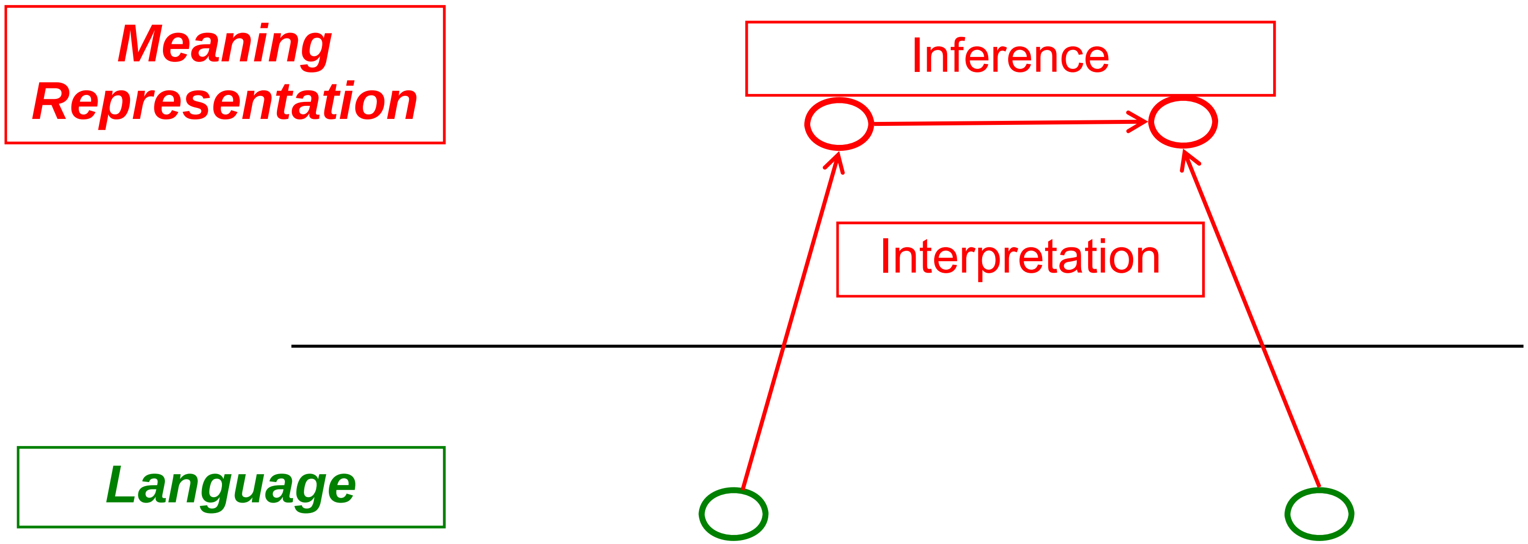
Serves(Maharani,VegetarianFood)

- ***Verifiability:*** a system's ability to compare the state of affairs described by a representation, to the state of affairs in some world as modeled by the knowledge base
- ***Interpretation****:* transforming a natural language sentence into a logical formula (First order predicate logic (FOL) is typically used)

Natural Language Interpretation:

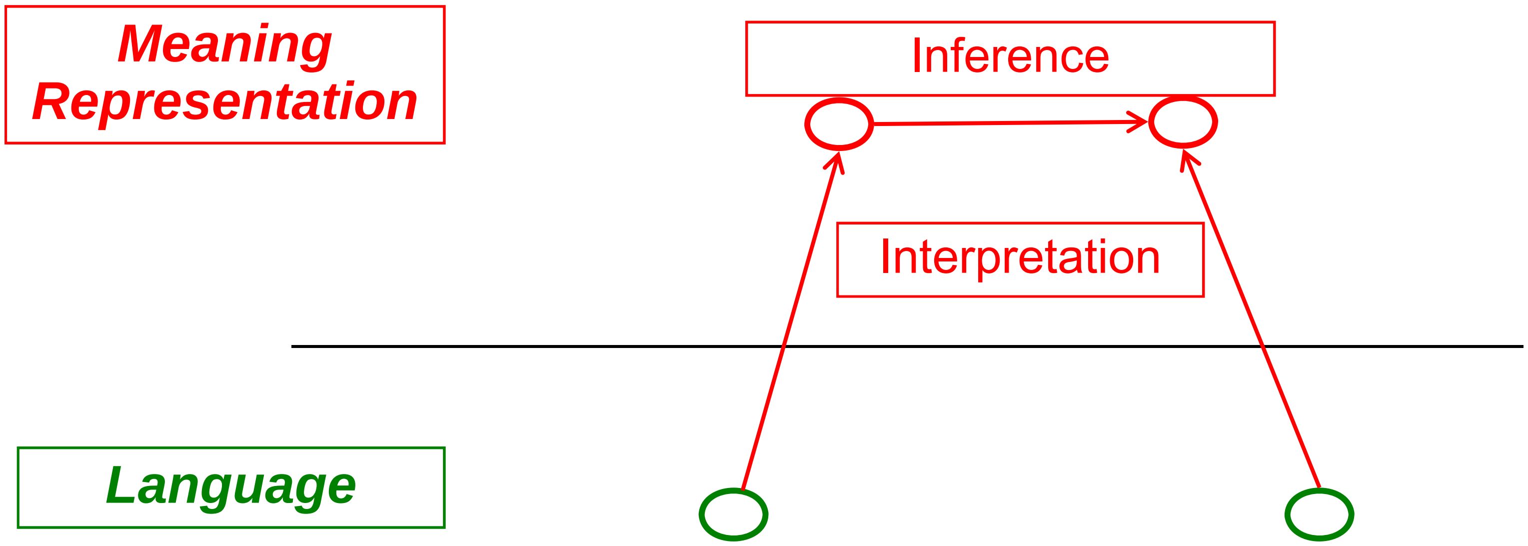


Interpretation model:

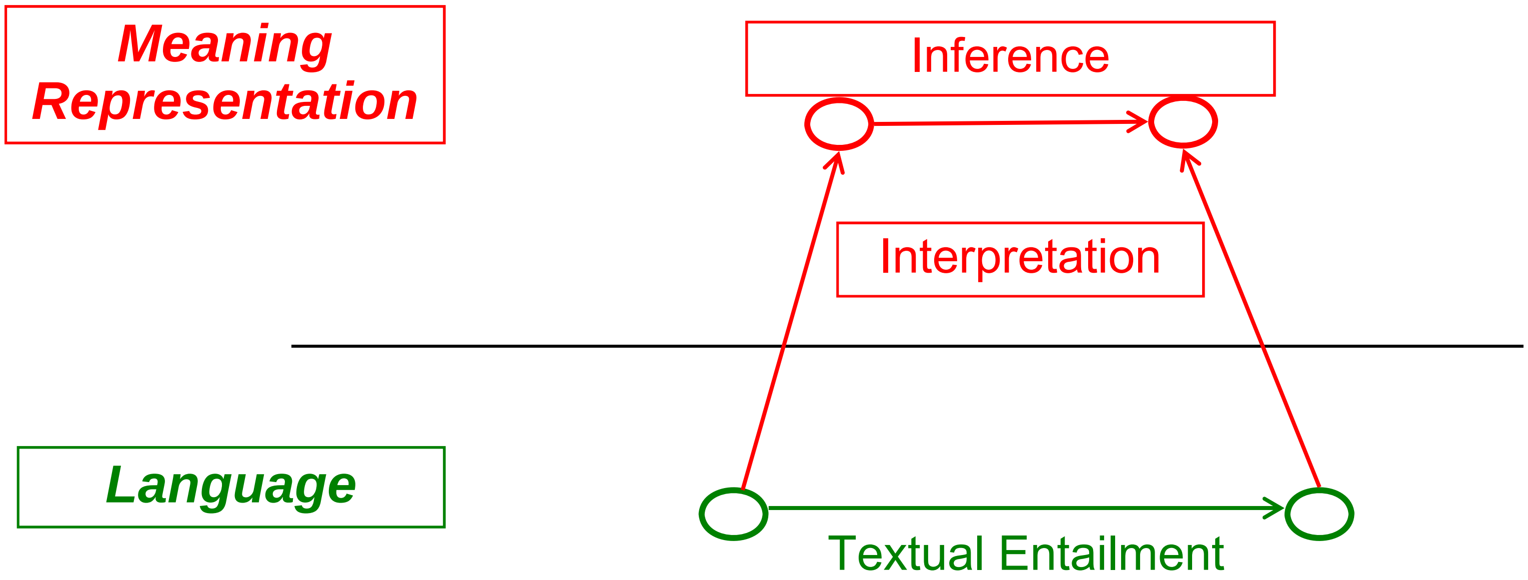


- Linguistic expressions are mapped into conceptual structures
- **Open issues** for large-scale applications:
  - Which predicates? Can we agree on them? (see issues in WSD)
  - Very few training data (meaning representations are not natural)

Interpretation model:



Textual Entailment:



- Linguistic Interpretation is a mean
- ***Benefits:***
  - No need to agree on shared predicates
  - Potentially lot of training data (texts are naturally produced by humans)

Motivation:

- Text applications require *semantic* inference

  What is the highest mountain in Europe?*

  … the village lies at the foot of the towering Mont Blanc, with the highest summit in Europe.

- Inference is done today in an application-dependent manner
- A common framework for applied semantics is needed, but still missing
- Textual entailment may provide such framework

Variability of semantic expressions:

The Dow Jones Industrial Average closed up 255

Dow ends up

Dow gains 255 points

Dow climbs 255

Stock market hits a record high

Model variability as relations between text expressions:

- Equivalence: *text1* ⇔ *text2* (paraphrasing)
- **Entailment:** *text1* ⇒ *text2* – the general case

Classical Entailment Definition:

- Chierchia & McConnell-Ginet (1990):
  A text t entails a hypothesis h if h is true in every circumstance (possible world) in which t is true

- Strict entailment - doesn't account for some uncertainty allowed in applications

"Almost certain" entailments:

T: *The technological triumph known as GPS was incubated in the mind of Ivan Getting.*

H: *Ivan Getting invented the GPS.*

Applied Textual Entailment:

- Directional relation between two text fragments: *Text (t)* and *Hypothesis (h)*:

> *t* entails *h* ($t \Rightarrow h$) if **humans** reading *t* will infer that *h* is **most likely** true

- Operational (applied) definition:
  - Human gold standard - as in NLP applications
  - Assuming common background knowledge – which is indeed expected from applications

Probabilistic Interpretation:

- *t probabilistically entails h* if:
  - P(*h is true* | *t*) > P(*h is true*)
    - *t* increases the likelihood of *h* being true
    - ≡ Positive PMI – *t* provides information on *h*'s truth
- P(*h is true* | *t*): *entailment confidence*
  - The relevant entailment score for applications
  - In practice: "most likely" entailment expected

The role of knowledge:

- For textual entailment to hold we require:
  - *text* AND *knowledge* $\Rightarrow$ *h*

    but
  - *knowledge* should not entail *h* alone

- Systems are not supposed to validate *h*'s truth regardless of *t* (e.g. by searching *h* on the web)

T: *The technological triumph known as GPS was incubated in the mind of Ivan Getting.*

H: *Ivan Getting invented the GPS.*

Typical Application Inference:

Question
Who bought Overture? >> *Expected answer form* X bought Overture

Overture's acquisition by Yahoo ⇒ Yahoo bought Overture
entails

text — *hypothesized answer*

- Similar for IE: → *X buy Y*
- IR
- Summarization (multi-document)
- Educational applications

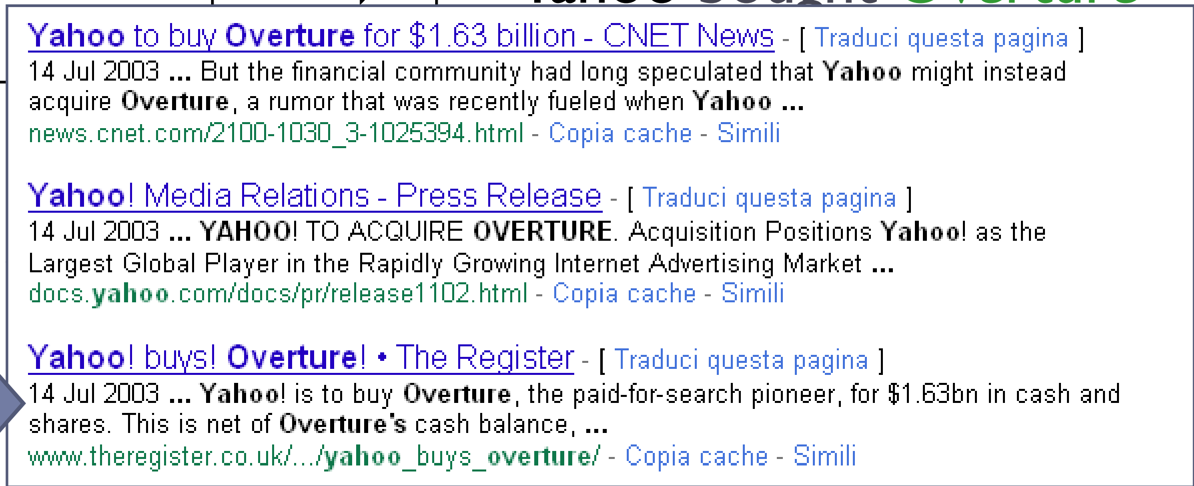
Typical Application Inference:

Question

Who bought Overture? >> *Expected answer form* X bought Overture

Overture's acquisition by Yahoo ⇒ Yahoo bought Overture

text



- Similar for IE:
- IR (q: *Overture acquisition*)
- Summarization (multi-document)
- Educational applications

Typical Application Inference:

Question
Who bought Overture? >> *Expected answer form*
X bought Overture

Overture's acquisition by Yahoo ⇒ (entails) Yahoo bought Overture

text — *hypothesized answer*

- Similar for IE:
- IR
- Summarization (multi-document) – identify redundant info
- Educational applications

Typical Application Inference:

Question
Who bought Overture?

\>>

Expected answer form
X bought Overture

Overture's acquisition by Yahoo

entails

Yahoo bought Overture

text

hypothesized answer

- Similar for IE:
- IR
- Summarization (multi-document)
- Educational applications – Reading comprehensions, ….

RTE evaluation campaigns:

PASCAL Recognizing Textual Entailment (RTE) Challenges

EU FP-6 Funded PASCAL Network of Excellence
2004-7

- Bar-Ilan University (Israel)
- CELCT, Trento (Italy)
- FBK-Irst, Trento (Italy)
- MITRE (USA)
- Microsoft Research (USA)
- NIST (USA)

Recognizing Textual Entailment challenge:

- ***TASK:*** given a T-H pair, automatically determine whether an entailment relation holds between T and H or not.

- ***DATASET:*** typical T-H pairs corresponding to success and failure cases of actual text processing applications (IR, IE, QA, SUM)
- ***SYSTEMS OUTPUT:***
  - Two-way judgment (entailment yes/no)
  - Three way judgment (entailment, contradiction, unknown)
- ***SYSTEMS EVALUATION:***
  - ***Accuracy:*** percentage of correct judgments against the Gold Standard
  - ***Average precision (optionally)***: measures the ability of the system to rank all the T-H pairs in the test set according to their entailment confidence *(for systems which returned a confidence score)*

RTE examples:

	TEXT	HYPOTHESIS	TASK	ENTAILMENT
1	*Regan attended a ceremony in Washington to commemorate the landings in Normandy.*	*Washington is located in Normandy.*	IE	False
2	*Google files for its long awaited IPO.*	*Google goes public.*	IR	True
3	*…: a shootout at the Guadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.*	*Cardinal Juan Jesus Posadas Ocampo died in 1993.*	QA	True
4	*The SPD got just 21.5% of the vote in the European Parliament elections, while the conservative opposition parties polled 44.5%.*	*The SPD is defeated by the opposition parties.*	IE	True

RTE dataset:

- Development Set and Test Set
- T-H pairs: 1,200 (600 Dev Set + 600 Test Set)
- Application settings
  - IE (200+200), IR (200+200), QA (200+200)
- NO SUM (in RTE-5)
- Distribution wrt the entailment judgment:
  - –50% *YES, 35% UNKNOWN, 15% CONTRADICTION*
- Longer *T's (100 words vs. 40 words in RTE-4)*
- T's not edited from their source documents (in RTE-5)

RTE datasets:

- T-H pairs involve various levels of entailment reasoning (lexical, syntactic, morphological and logical).

<pair id=”10” entailment=**”ENTAILMENT”** task=”IR”>
<t>In the end, defeated, Antony committed suicide and so did Cleopatra, according to legend, by putting an asp to her breast.</t>
<h>Cleopatra committed suicide.</h> </pair>

<pair id=”794” entailment=**”CONTRADICTION”** task=”IE”>
<t>Charged with murdering his wife, the jury acquitted Blake due to lack of evidence.</t>
<h>Blake was condemned for the murder of his wife.</h> </pair>

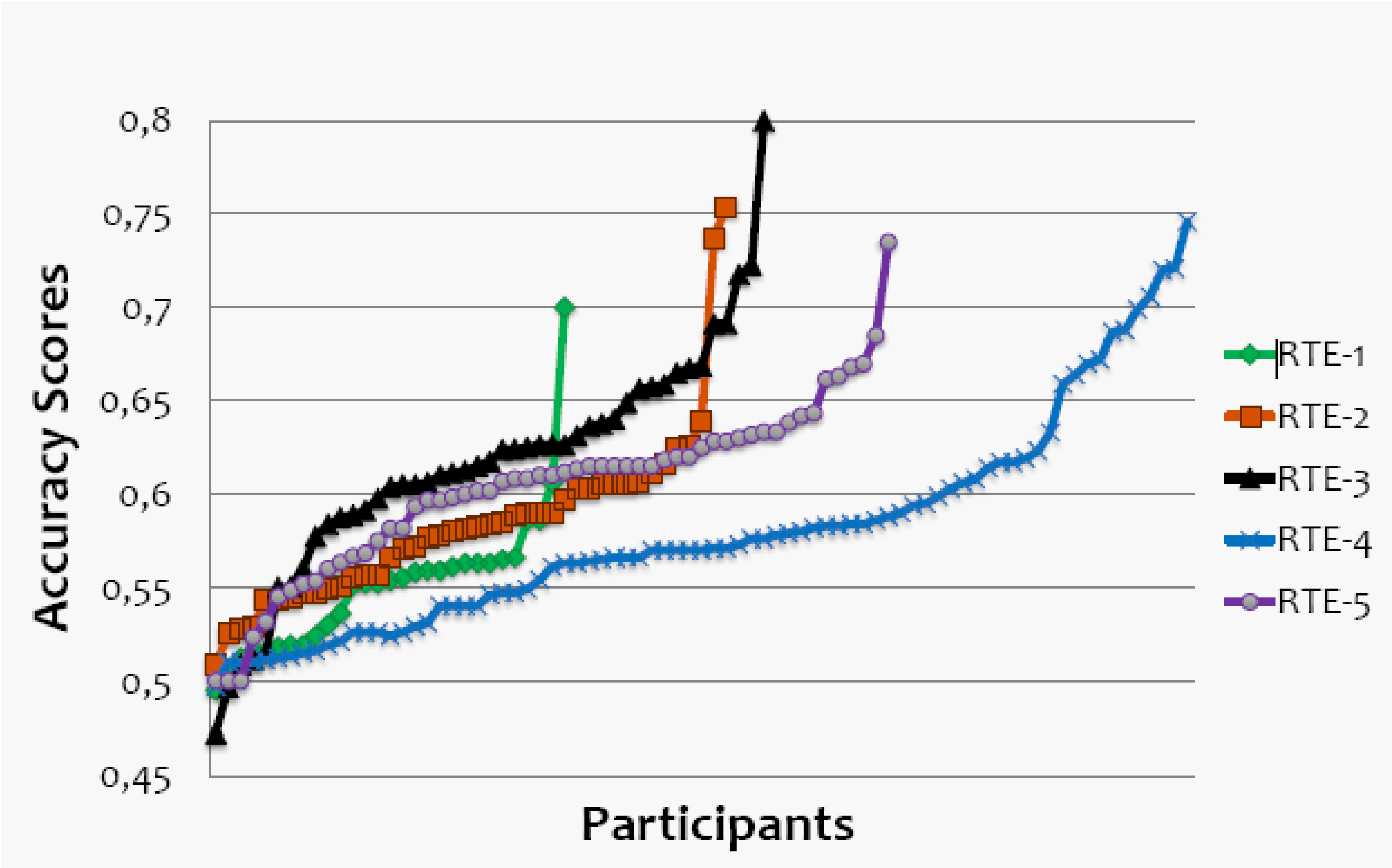
<pair id=”577” entailment=**”UNKNOWN”** task=”SUM”>
<t>Honda has released the first official images of the new Jazz. Honda claims its new Jazz will build on the qualities that made the old model so popular in the UK.</t>
<h>The new-generation Honda Jazz goes on sale in the UK.</h> </pair>

Participation and impact:

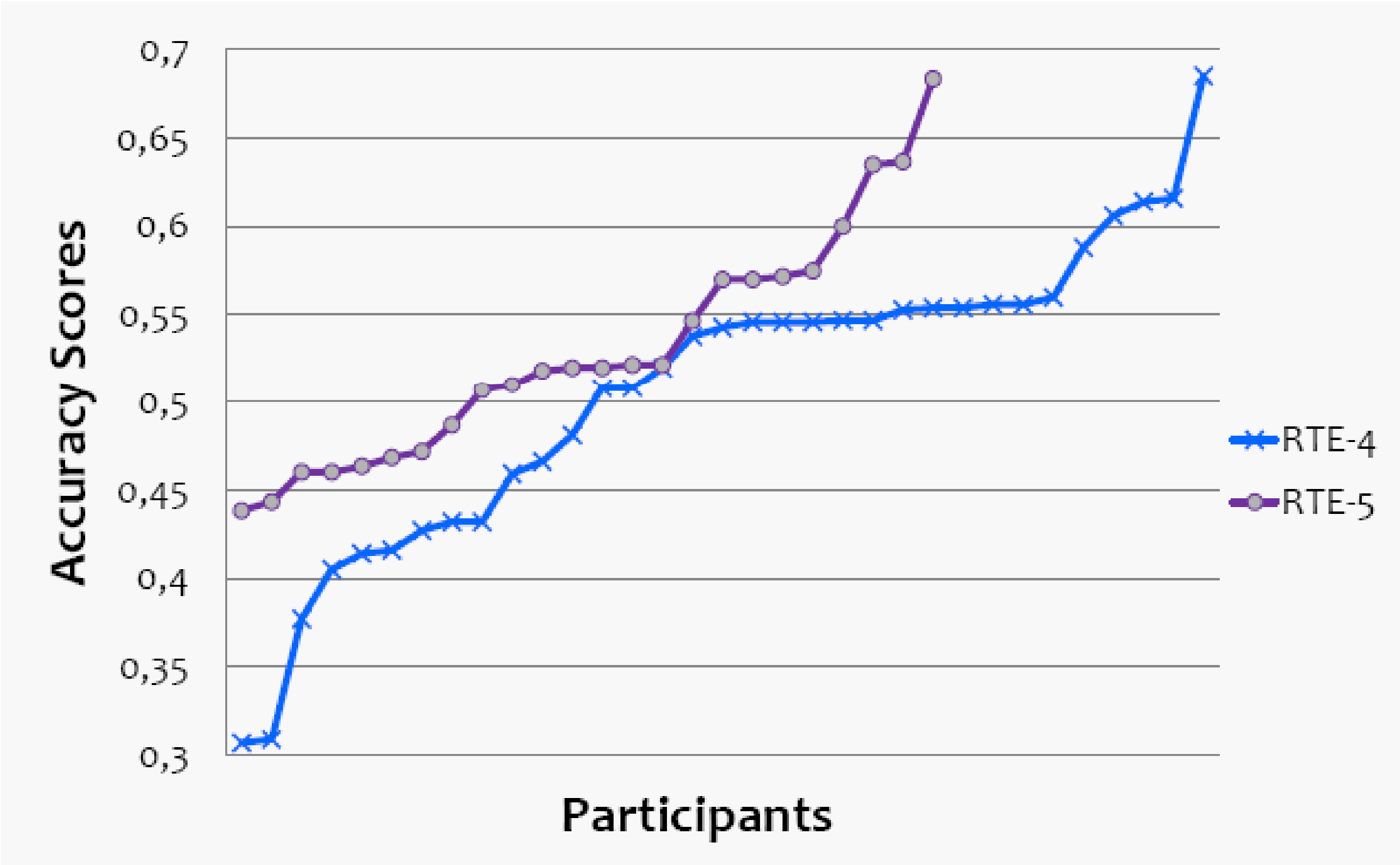
- Very successful challenges, world wide:
  - RTE-1 - 17 groups
  - RTE-2 - 23 groups
  - RTE-3 - 25 groups
  - RTE-4, under NIST coordination
  - RTE-5 - 21 groups
  - RTE-6 (Workshop on 17 November)

- High interest in the research community

RTE-1 to RTE-5 results for 2 way task



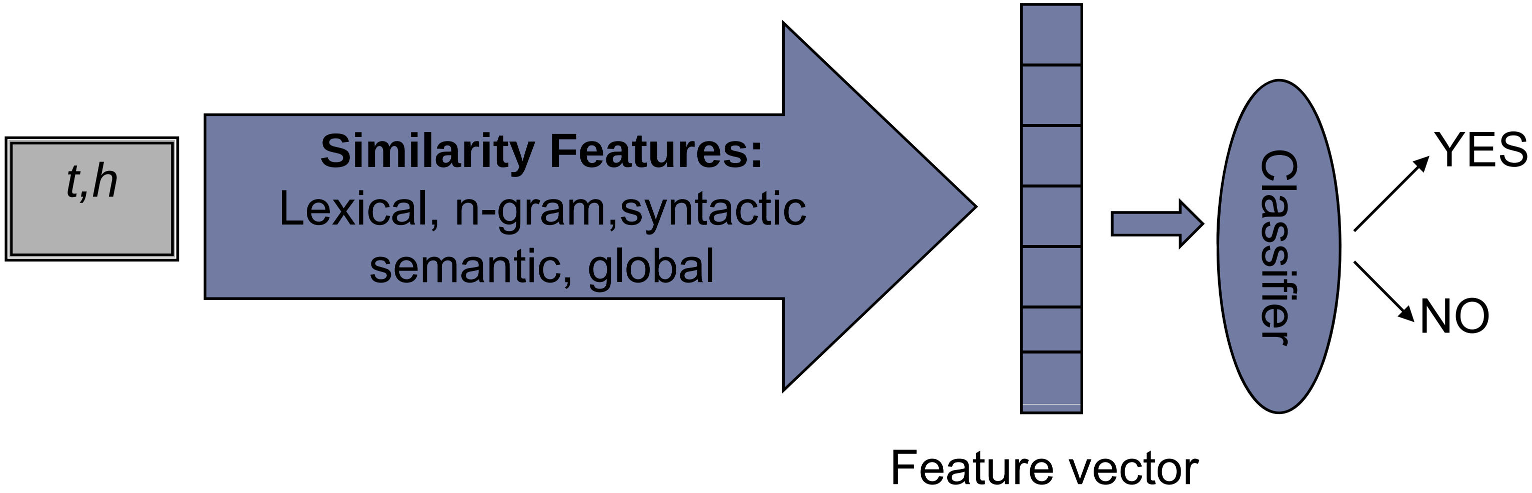
RTE-1 to RTE-5 results for 3 way task



Methods and approaches:

- Measure similarity match between *t* and *h* (*coverage* of *h* by *t*):
  - Lexical overlap (unigram, N-gram, subsequence)
  - Lexical substitution (WordNet, statistical)
  - Syntactic matching / transformations
  - Lexical-syntactic variations ("paraphrases")
  - Semantic role labeling and matching
  - Global similarity parameters (e.g. negation, modality)
- Detect mismatch (for non-entailment)
- Interpretation to logic representation + logic inference
- A dominant approach: supervised learning

Dominant approach: supervised learning



- Features model similarity and mismatch
- Classifier determines relative weights of information sources
- Train on development set and auxiliary *t-h* corpora

Research directions:

- Inference
  - Principled framework for inference
  - *Are we happy with bags of features?*

- Knowledge acquisition
  - Unsupervised acquisition of linguistic and world knowledge from general corpora and web
  - Acquiring larger entailment corpora

Textual Entailment:

Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc. last year

Entails

Subsumed by

$\subseteq$

Yahoo acquired Overture

Google owns Overture

Entity matching

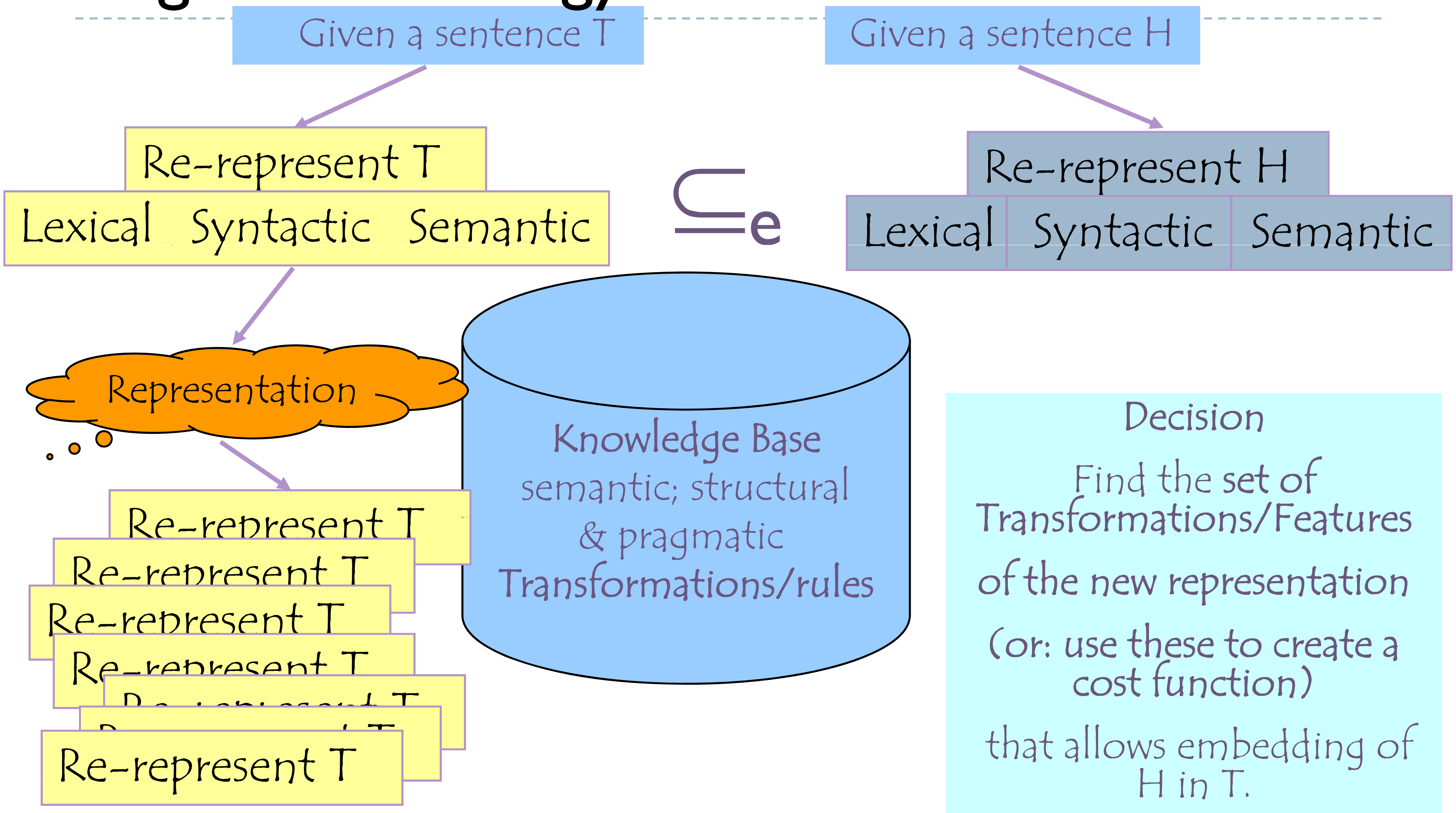
Alignment

How?

Semantic Role Labeling

Integration

A general strategy for TE:



Details of the entailment strategy:

- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

The case of shallow lexical approaches:

- Preprocessing
  - Identify stop words
- Representation
  - Bag of words
- Knowledge source
  - Shallow lexical resources (typically WordNet)
- Control Strategy and Decision Making
  - Single pass
  - Compute similarity; use threshold tuned on a development set
- Justification
  - It works

Shallow Lexical Approaches (example):

- Lexical/word-based semantic overlap: score based on matching each word in H with some word in T
  - Word similarity measu
  - May take account of su
  - 'Learn' threshold on m

Clearly, this may not appeal to what we think as *understanding*, and it is easy to generate cases for which this does not work well.
However, it works (surprisingly) well with respect to current evaluation metrics (data sets?)

Text: The Cassini spacec

Text: NASA's Cassini-Huy
2006.

Text: The Cassini spacecraft has taken images that show rivers on Saturn's moon Titan.

Hyp: The Cassini spacecraft has reached Titan.

An algorithm: LocalLexicalMatching

- For each word in Hypothesis, Text
  - if word matches stopword – remove word
  - if no words left in Hypothesis or Text   return 0
- numberMatched = 0;
  - for each word W_H in Hypothesis
  - for each word W_T in Text
  - HYP_LEMMAS = Lemmatize(W_H);
  - TEXT_LEMMAS = Lemmatize(W_T);
    - Use Wordnet's
- if any term in HYP_LEMMAS matches any term in TEXT_LEMMAS
  - using LexicalCompare()
- numberMatched++;
- Return: numberMatched/|HYP_Lemmas|

An algorithm: LocalLexicalMatching

LLM Performance:

RTE2:	Dev: 63.00	Test: 60.50
RTE 3:	Dev: 67.50	Test: 65.63

- LexicalCompare()

```
if(LEMMA_H == LEMMA_T)
    return TRUE;
if(HypernymDistanceFromTo(textWord, hypothesisWord) <= 3)
    return TRUE;
if(MeronymyDistanceFromTo(textWord, hypothesisWord) <= 3)
    returnTRUE;
if(MemberOfDistanceFromTo(textWord, hypothesisWord) <= 3)
    return TRUE:
if(SynonymOf(textWord, hypothesisWord)
     return TRUE;
```

- Notes:
  - LexicalCompare is Asymmetric & makes use of single relation type
  - Additional differences could be attributed to stop word list (e.g, including aux verbs)
  - Straightforward improvements such as bi-grams do not help.
  - More sophisticated lexical knowledge (entities; time) should help.

Details of the entailment strategy:

- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

Preprocessing:

- Syntactic Processing:
  - Syntactic Parsing (Collins; Charniak; CCG)
  - Dependency Parsing (+types)
- Lexical Processing
  - Tokenization; lemmatization
  - For each word in Hypothesis, Text
  - Phrasal verbs
  - Idiom processing
  - Named Entities + Normalization
  - Date/Time arguments + Normalization
- Semantic Processing
  - Semantic Role Labeling
  - Nominalization
  - Modality/Polarity/Factive
  - Co-reference

Only a few systems

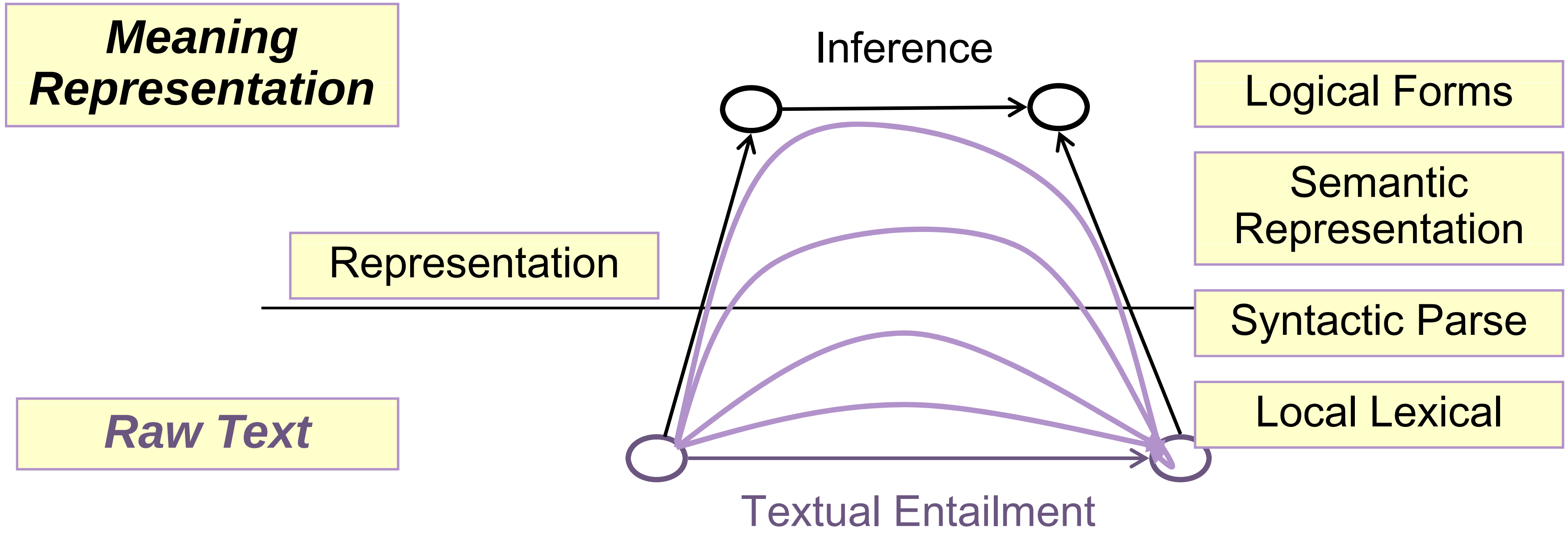
often used only during decision making

often used only during decision making

Details of the entailment strategy:

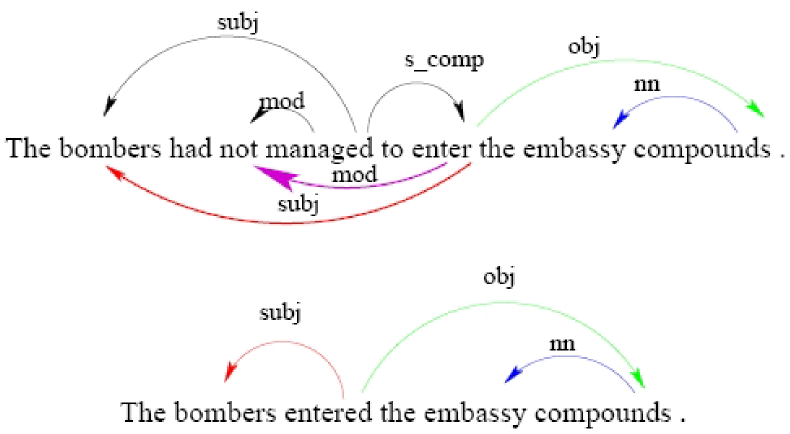
- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

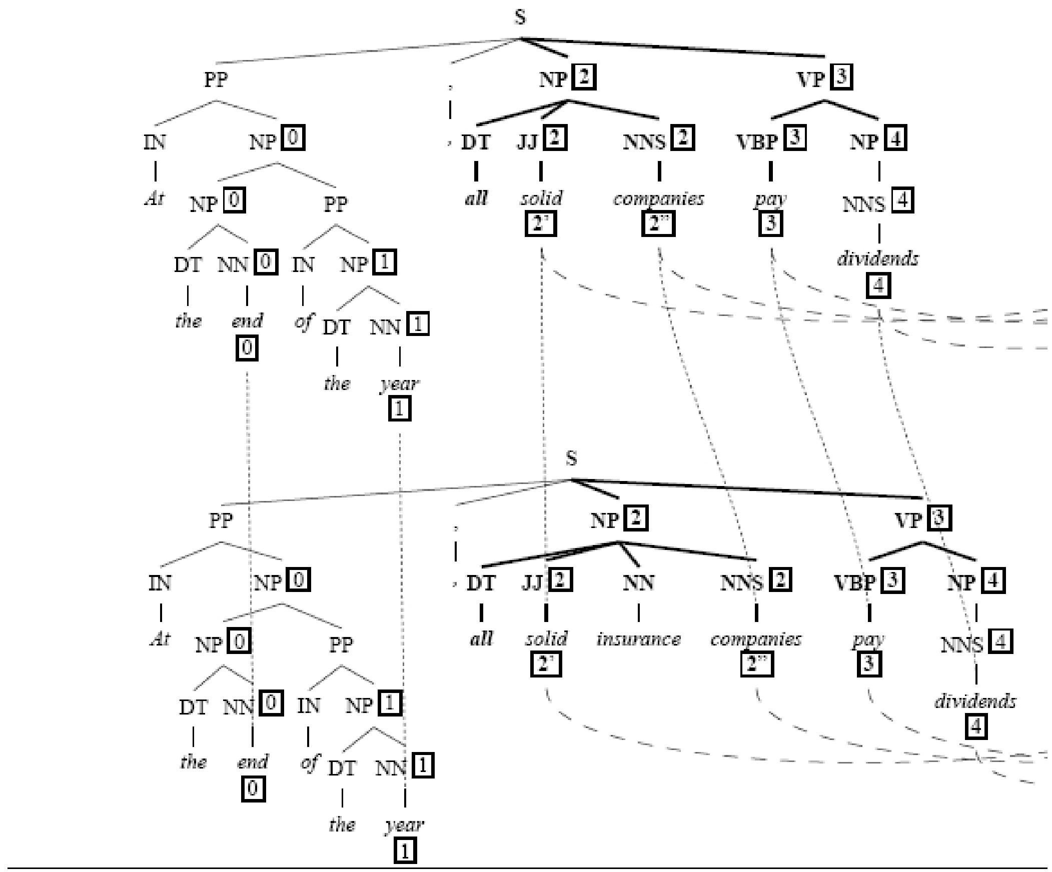
Basic representations:



- Most approaches augment the basic structure defined by the processing level with additional annotation and make use of a tree/graph/frame-based system.

Basic representations: syntax





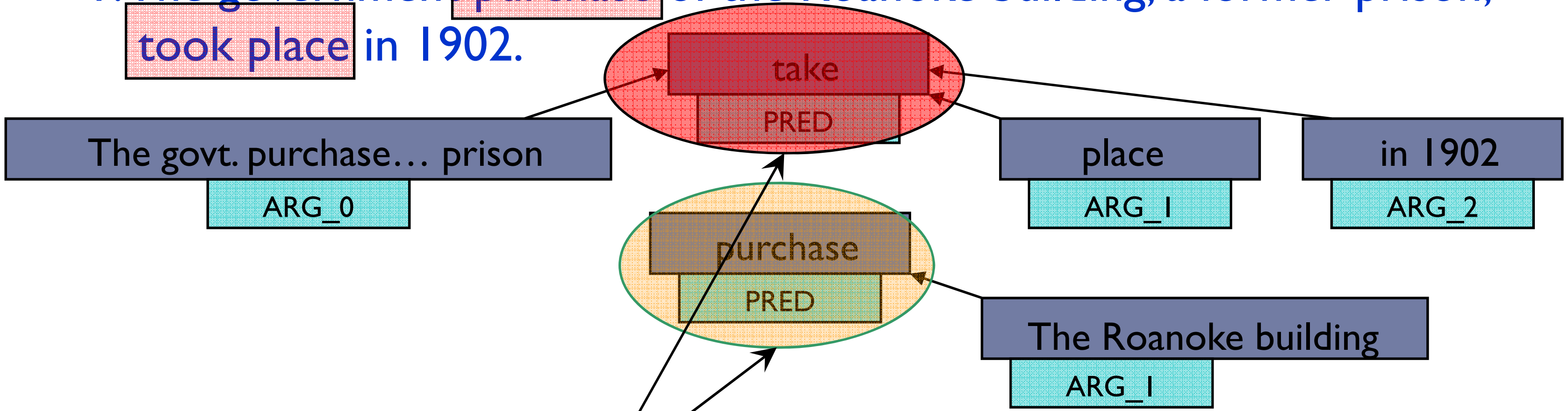
Syntactic Parse

Hyp: The Cassini spacecraft has reached Titan.

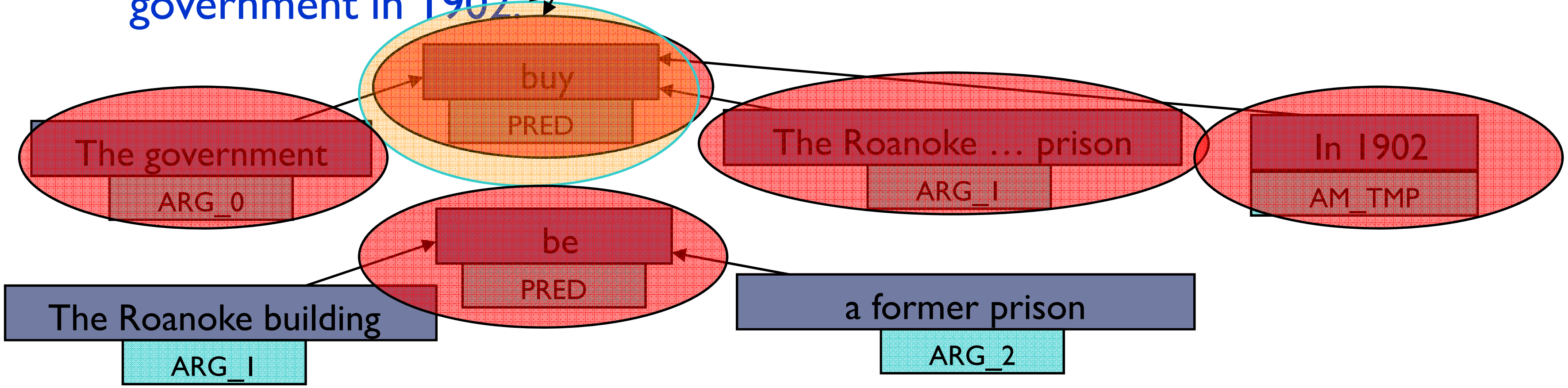
Local Lexical

Basic representation (shallow semantics: pred-arg)

T: The government purchase of the Roanoke building, a former prison, took place in 1902.



H: The Roanoke building, which was a former prison, was bought by the government in 1902.



(Roth & Sammons 2007)

Representing knowledge sources

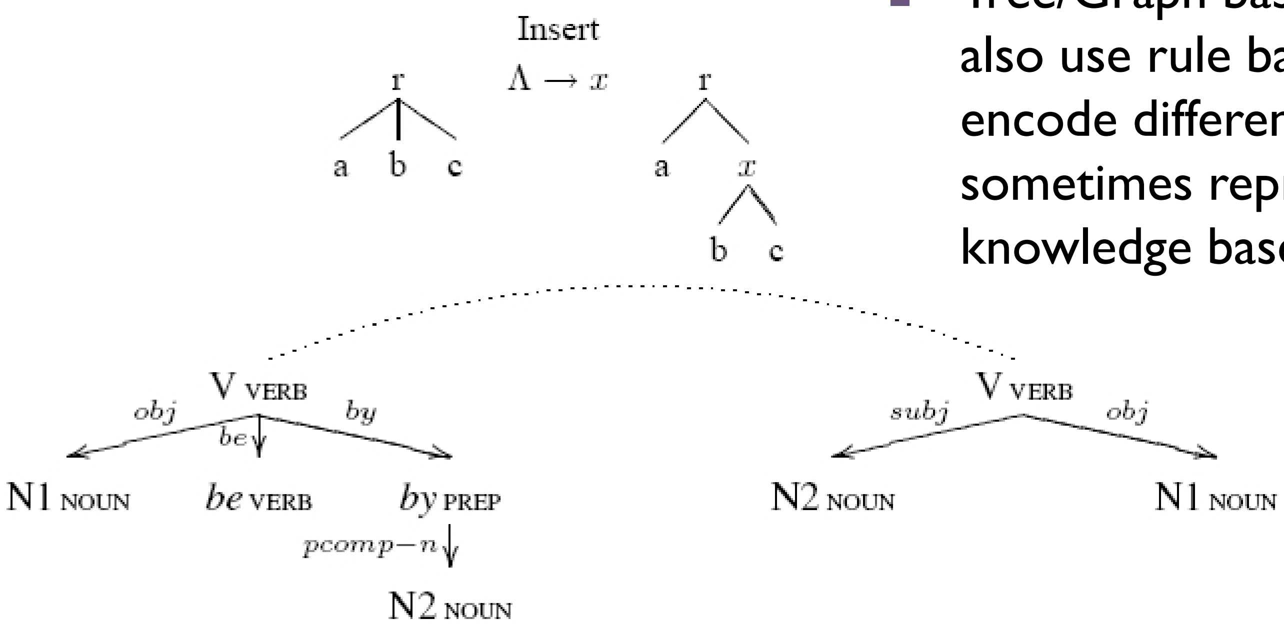
- Rather straightforward in the Logical Framework:

$$\forall x \forall y (paris(x) \wedge france(y) \rightarrow in(x,y))$$

$$\forall x (clinton(x) \rightarrow person(x))$$
$$\forall x (artifact(x) \rightarrow \neg person(x))$$

$$\forall x (book(x) \rightarrow artifact(x))$$
$$\forall x (soar(x) \rightarrow rise(x))$$

- Tree/Graph base representation may also use rule based transformations to encode different kinds of knowledge, sometimes represented as generic or knowledge based tree transformations.



Representing knowledge sources

- In general, there is a mix of procedural and rule based encodings of knowledge sources
  - done by hanging more information on parse tree or predicate argumerent presentation.

> **T:** The Bills [now]$_{2006-01-01}$ appear ready to hand the reins over to one of their two-top picks from [a year ago]$_{2005-01-01}$ in quarterback J.P. Losman, who missed most of last season with a broken leg.

> **T:** [The Bills]$_{team}$ now appear ready to hand the reins over to one of their two-top picks from a year ago in quarterback [J.P. Losman]$_{person}$, who missed most of last season with a broken leg.

  - A different frame-based annotation systems for encoding information, that are processed procedurally.

Details of the entailment strategy:

- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

Knowledge sources

- The knowledge sources available to the system are the most significant component of supporting TE.
- Different systems draw differently the line between ***preprocessing capabilities*** and ***knowledge resources.***
- The way resources are handled is also different across different approaches.

Enriching Preprocessing

- In addition to syntactic parsing several approaches enrich the representation with various linguistics resources:
  - Pos tagging
  - Stemming
  - Predicate argument representation: verb predicates and nominalization
  - Entity Annotation: Stand alone NERs with a variable number of classes
  - Acronym handling and Entity Normalization: mapping mentions of the same entity mentioned in different ways to a single ID.
  - Co-reference resolution
  - Dates, times and numeric values; identification and normalization.
  - Identification of semantic relations: complex nominals, genitives, adjectival phrases, and adjectival clauses.
  - Event identification and frame construction.

Lexical Resources

- Recognizing that a ***word*** or a ***phrase*** in ***T*** entails a word or a phrase in ***H*** is essential in determining Textual Entailment.

- WordNet is the most commonly used resoruce
  - In most cases, a Wordnet based **similarity measure** between words is used. This is typically a symmetric relation.
  - **Lexical chains** over Wordnet are used; in some cases, care is taken to disallow some chains of specific relations.
  - Extended Wordnet is being used to make use of **Entities**
  - **Derivation** relation which links verbs with their corresponding **nominalized nouns**.

Lexical Resources (cont.)

- Lexical Paraphrasing Rules
  - A number of efforts to acquire relational paraphrase rules are under way, and several systems are making use of resources such as DIRT and TEASE.
  - Some systems seems to have acquired paraphrase rules that are in the RTE corpus
    - person killed --> claimed one life
    - hand reins over to --> give starting job to
    - same-sex marriage --> gay nuptials
    - cast ballots in the election -> vote
    - dominant firm --> monopoly power
    - death toll --> kill
    - try to kill --> attack
    - lost their lives --> were killed
    - left people dead --> people were killed

Representing knowledge sources

- A large number of semantic phenomena have been identified as significant to Textual Entailment.
- A large number of them are being handled (in a restricted way) by some of the systems. Very little quantification per-phenomena has been done, if at all.
- Semantic implications of interpreting syntactic structures (Braz et. al. 2005; Bar-Haim et. al. 2007)

- Conjunctions
  - Jake and Jill ran up the hill — Jake ran up the hill
  - Jake and Jill met on the hill — *Jake met on the hill

- Clausal modifiers
  - But celebrations were muted as many Iranians observed a Shi'ite mourning month.
  - Many Iranians observed a Shi'ite mourning month.
  - Semantic Role Labeling handles this phenomena automatically

Representing knowledge sources

- **Relative clauses**
  - The assailants fired six bullets at the car, which carried Vladimir Skobtsov.
  - The car carried Vladimir Skobtsov.
  - Semantic Role Labeling handles this phenomena automatically

- **Appositives**
  - Frank Robinson, a one-time manager of the Indians, has the distinction for the NL.
  - Frank Robinson is a one-time manager of the Indians.

- **Passive**
  - We have been approached by the investment banker.
  - The investment banker approached us.
  - Semantic Role Labeling handles this phenomena automatically

- **Genitive modifier**
  - Malaysia's crude palm oil output is estimated to have risen..
  - The crude palm oil output of Malasia is estimated to have risen .

Logical Structure

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

Some examples

- T: Legally, John could drive.
- H: John drove.

.

- T: Bush said that Khan sold centrifuges to North Korea.
- H: Centrifuges were sold to North Korea.

.

- T: No US congressman visited Iraq until the war.
- H: Some US congressmen visited Iraq before the war.

- T: The room was full of women.
- H: The room was full of intelligent women.

- T: The New York Times reported that Hanssen sold FBI secrets to the Russians and could face the death penalty.
- H: Hanssen sold FBI secrets to the Russians.

- T: All soldiers were killed in the ambush.
- H: Many soldiers were killed in the ambush.

Details of the entailment strategy:

- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

Control strategy and decision making

- Single Iteration
  - Strict Logical approaches are, in principle, a single stage computation.
  - The pair is processed and transform into the logic form.
  - Existing Theorem Provers act on the pair along with the KB.
- Multiple iterations
  - Graph based algorithms are typically iterative.
  - Following (Punyakanok et. al '04) transformations are applied and entailment test is done after each transformation is applied.
  - Transformation can be chained, but sometimes the order makes a difference. The algorithm can be a greedy algorithm or can be more exhaustive, and search for the best path found (Braz et. al'05;Bar-Haim et.al 07)

Details of the entailment strategy:

- Preprocessing
  - Multiple levels of lexical pre-processing
  - Syntactic Parsing
  - Shallow semantic parsing
  - Annotating semantic phenomena
- Representation
  - Bag of words, n-grams through tree/graphs based representation
  - Logical representations
- Knowledge source
  - Syntactic mapping rules
  - Lexical Resources
  - Semantic phenomena specific modules
  - RTE specific knowledge source
  - Additional Corpora/Web resources
- Control Strategy and Decision Making
  - Pass/iterative processing
  - Strict vs. parameter based
- Justification
  - What can be said about the decision?

Justification

- For most approaches justification is given only by the data Preprocessed
  - Empirical Evaluation
- Logical Approaches
  - There is a proof theoretic justification
  - Depending on the power of the resources and the ability to map a sentence to a logical form.
- Graph/tree based approaches
  - There is a model theoretic justification
  - The approach is sound, but not complete, depending on the availably of resources.

What is TE missing?

- It is completely clear that the key resource missing is knowledge.
  - Better resources translate immediately to better results.
  - At this point existing resources seem to be lacking in coverage and accuracy.
  - Not enough high quality public resources; no quantification.
- Some Examples
  - Lexical Knowledge: Some cases are difficult to acquire systematically.
    - A bought Y → A has/owns Y
    - Many of the current lexical resources are very noisy.
  - Numbers, quantitative reasoning
  - Time and Date; Temporal Reasoning.
  - Robust event based reasoning and information integration

BIBLIOGRAPHY:

- Bar-Haim, R. Dagan I., Greental I., Szpektor I. and Fridman M. (2007). **Semantic Inference at the Lexical-Syntactic Level for Textual Entailment Recognition**. In *Proceedings of WTEP 2007, Prague.*
- Braz R., Girju R., Punyakanok V., Roth D., and Sammons M. (2005). An Inference Model for Semantic Entailment in Natural Language. Twentieth National Conference on Artificial Intelligence (AAAI-05).
- Chierchia G. and McConnell-Ginet, S.(1990). *Meaning and Grammar: An Introduction to Meaning*. The MIT Press, Cambridge, Mass,
- Jurafsky D., Martin J.H., *Speech and Language Processing: An introduction to natural language processing, computational lingustics and speech recognition. Chapter19.*
- Punyakanok V., Roth D.and Yih W. (2004) Natural Language Inference via Dependency Tree Mapping: An Application to Question Answering, Computational Linguistics.
- Roth D. and Sammons M., (2007) Semantic and Logical Inference Model for Textual Entailment Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing – 2007

- CREDITS: thanks to Bernardo Magnini (Fbk-Irst), Ido Dagan and Shahar Mirkin (Bar-Ilan University), Fabio Massimo Zanzotto (Università Tor Vergata), Luisa Bentivogli (FBK-Irst)