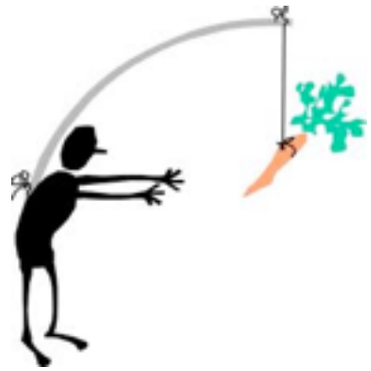
# Textual Inference - Methods and Applications

Günter Neumann, LT Lab, DFKI, December 2013

Some slides are from Ido Dagan (BIU, Israel), Bill Dolan (Microsoft Research, USA), and Arindam Bhattacharya (Indian Institute of Technology, Indian).

# Motivation

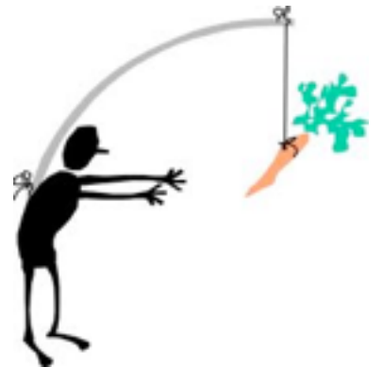


- Text-based applications need robust semantic inference engines
- Example: Open domain question answering

*Q: Who is John Lennon's widow?*

*A: Yoko Ono unveiled a bronze statue of her late husband, John Lennon, to complete the official renaming of England's Liverpool Airport as Liverpool John Lennon Airport.*

# Motivation



- Text-based applications need robust semantic inference engines
- Example: Open domain question answering

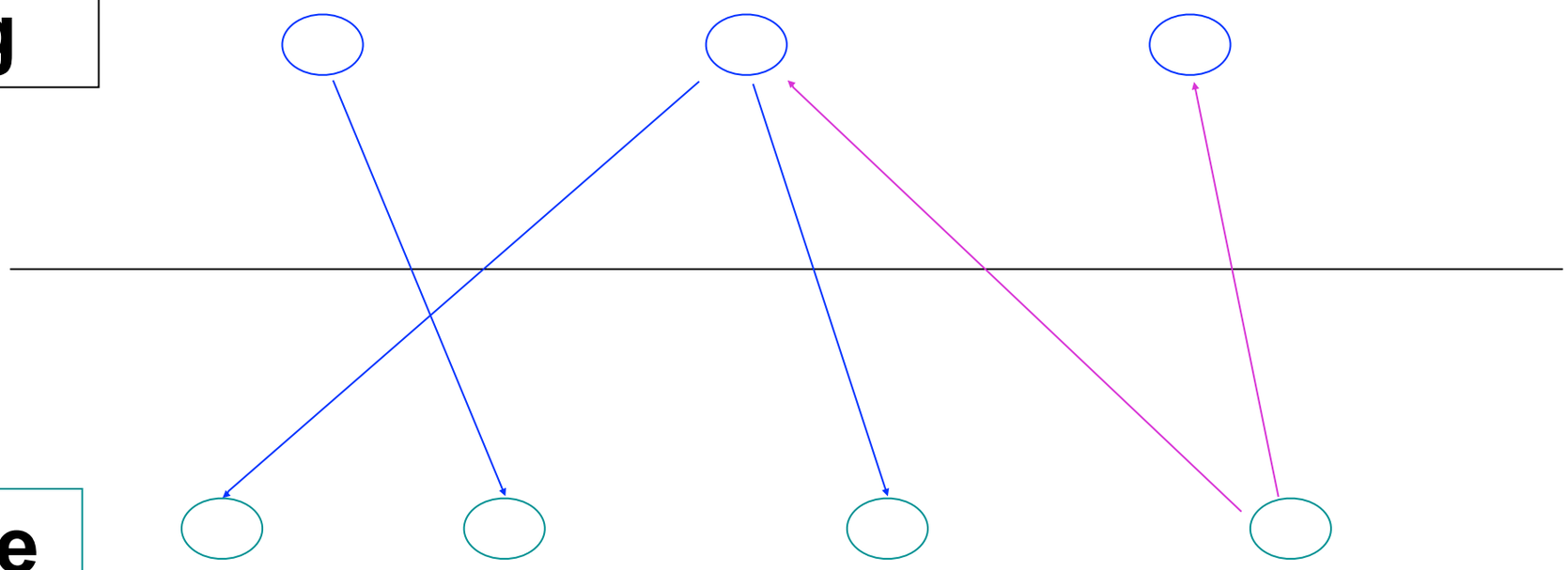
*Q: Who is John Lennon's widow?*

*A: Yoko Ono unveiled a bronze statue of her late husband, John Lennon, to complete the official renaming of England's Liverpool Airport as Liverpool John Lennon Airport.*

# Natural Language and Meaning

**Meaning**

Variability



**Language**

Ambiguity

# Variability of Semantic Expression

**All major stock markets surged**

**Dow gains 255 points**

**Dow ends up**


**Stock market hits a record high**

**Dow climbs 255**

**The Dow Jones Industrial Average closed up 255**

# Text-based Applications

- Question answering:
  „Who acquired Overture?“ vs. „Yahoos‘ buyout of Overture was approved ...“

- Open information extraction:
  Clustering of extracted semantically similar relations, e.g., all instances of the business acquisition relation found in a set of online newspapers

- Web query understanding:
  „johny depp movies 2010“ vs. „what are the movies of 2010 in which johny depp stars ?“

# Text-based Applications

- E-learning:
  Automatically score students‘ free-text answers to open questions relative to the „expected answers“.

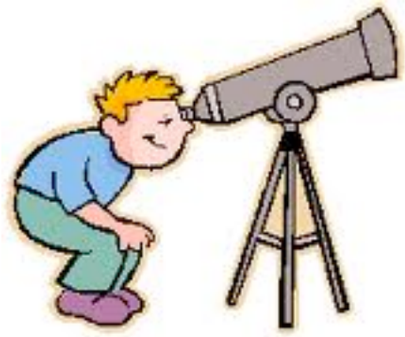
- Text summarization:
  Identify redundant information from multiple documents.

- Machine Reading:
  Text extraction and automatic linkage to knowledge bases.

# Text-based Applications


- Common challenges
  - textual variability of semantic expressions
  - un-precise language usage of semantic relationships
  - noisy language use and text data
- Still dominating approach: Individual solutions
  - task specific solutions, e.g, answer extraction, empirical co-occurrence, narrow „procedural“ lexical semantics
  - no generic approach (no „parsing“ equivalence)

# Scientific Perspective



- The usage of discrete NLP components alone are not sufficient, e.g., POS tagging, dependency parsing, word sense disambiguation, reference resolution.

- Because: text understanding applications need to be able to

    - determine whether two strings „mean the same“ in a certain context independently of their surface realizations.

    - determine whether one string semantically entails another string.

    - reformulate strings in a meaning preserving manner.

- Hence: empirical models of semantic overlap are needed

    - a common framework for applied semantics which renders possible scalable, robust, efficient semantic inference.

# Definition

## Classical Definition

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 real world uncertainties.
- Example:
  - **T:** *Ram was born and brought up in Maharashtra.*
  - **H:** *Ram can speak Marathi.*

## Applied Definition

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

# Probabilistic Interpretation

- Applied definition sounds good.
- But doesn't sound concrete of mathematical.

## Probabilistic interpretation

$t$ probabilistically entails $h$ if:

$$P(h \text{ is true} \mid t) > P(h \text{ is true})$$

- **P($h$ is true $|t$)** is called Entailment Confidence.

# Goal

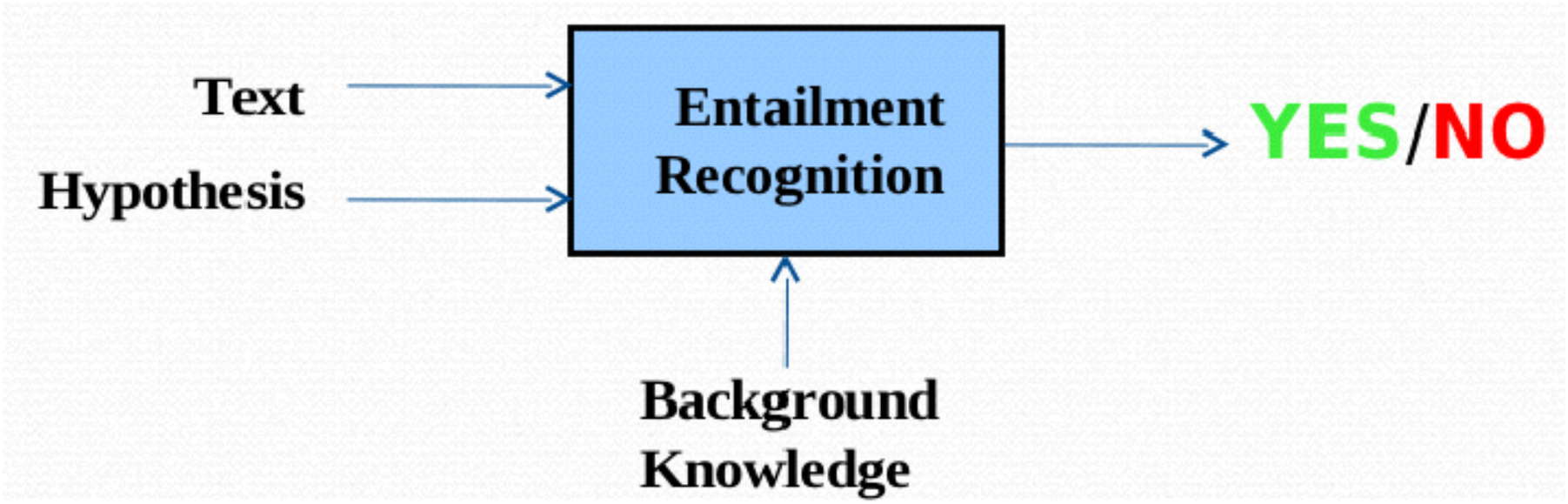



Figure: Textual Entailment

# Entailment Triggers

- Triggers are linguistic features that affect entailment [**?**].
- Here are some examples to show how these various factors affect entailment.

Synonymy: Very common form of entailment trigger, where a word is replaced by its synonym.

T: *World War I* **began** *in 1914.*

H: *World War I* **started** *in 1914.*

# Entailment Triggers

Hypernymy/Hyponymy: Certain concept can be either *generalized* or *specialized*, leading to entailment.

T: **Reptiles** *have scale.*

H: **Snakes** *have scale. (Specialization or Hyponymy)*

T: *Beckham plays* **football**.

H: *Beckham plays* **a game**. *(Generalization or Hypernymy)*

# Entailment Triggers

Co-reference: One of the main sources for text entailment. Especially with long text containing paragraphs!

T: **Barrack Obama** *came to India.* **The American President** *had a meeting with Manmahon Singh.*

H: **Barrack Obama** *had a meeting with Manmahon Singh.*

# Entailment Triggers

**Modality/Polarity/Factive:** Plays critical role in entailment as they affect the degree of reliability on the remaining sentence. Especially troublesome for *lexical* approaches.

**Modality** denotes possibility or necessity and sometimes may lead to wrong entailment. e.g. may, can, shall, must etc. are modality triggers.

T: *The government* **may** *approve the anti-corruption bill.*

H: *The government approved the anti-corruption bill.*

# Entailment Triggers

**Polarity** determines whether the fact asserted or its negation is going to occur. e.g. not, never, deny etc. are polarity triggers.

T: *The watchman* **denied** *that he was sleeping.*

H: *The watchman was sleeping.*

**Factivity** deals with presupposition. It states a fact assuming another has occurred. e.g. realize, regret etc. are factivity triggers.

T: *Martha* **regrets** *eating John's homemade cake.*

H: *Martha ate John's homemade cake.*

# Entailment Triggers

Dropping or Inserting Adjunct: Adding or dropping adjuncts affect entailment based on which of $T$ or $H$ is modified, and the polarity.

T: *Bob was running quickly.*
H: *Bob was running.*
T: *Carl was eating.*
H: *Carl was eating slowly. [Incorrect entailment]*
T: *Alice was not driving.*
H: *Alice was not driving fast.*
T: *Derek was not writing properly.*
H: *Derek was not writing. [Incorrect entailment]*

# Role of Knowledge

- Background knowledge is *crucial* in entailment as in any AI application!

**Example**

T: *President of Russia visited Paris.*

H: *President of Russia visited France.*

B: *Paris is situated in France.*

- Background knowledge $B$ alone should not entail the hypothesis $H$ and text $T$ must contain necessary information (may not be sufficient).

$$(T \wedge B) \models H$$

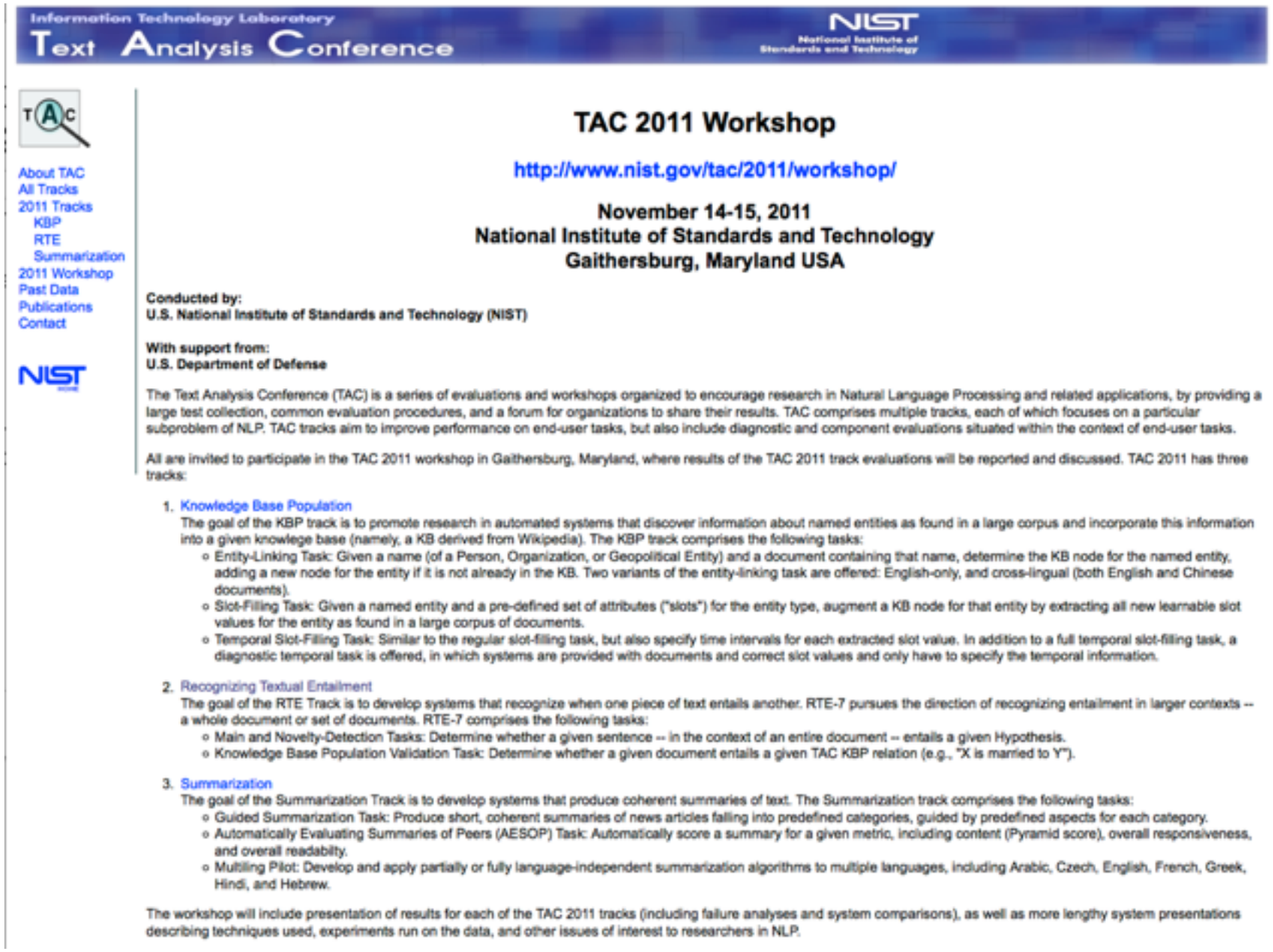
but

$$B \not\models H$$

# Recognizing Textual Entailment (RTE) Challenge – A Scientific Competition

- Since 2005 until today - RTE-1 to RTE-7

- Main motivation: Bring together scientists from all over the world, in order to commonly push forward the scientific field of „applied semantics“ („open collaboration“).



Information Technology Laboratory
Text Analysis Conference
NIST National Institute of Standards and Technology

About TAC
All Tracks
2011 Tracks
KBP
RTE
Summarization
2011 Workshop
Past Data
Publications
Contact

**TAC 2011 Workshop**

http://www.nist.gov/tac/2011/workshop/

**November 14-15, 2011**
**National Institute of Standards and Technology**
**Gaithersburg, Maryland USA**

**Conducted by:**
**U.S. National Institute of Standards and Technology (NIST)**

**With support from:**
**U.S. Department of Defense**

The Text Analysis Conference (TAC) is a series of evaluations and workshops organized to encourage research in Natural Language Processing and related applications, by providing a large test collection, common evaluation procedures, and a forum for organizations to share their results. TAC comprises multiple tracks, each of which focuses on a particular subproblem of NLP. TAC tracks aim to improve performance on end-user tasks, but also include diagnostic and component evaluations situated within the context of end-user tasks.

All are invited to participate in the TAC 2011 workshop in Gaithersburg, Maryland, where results of the TAC 2011 track evaluations will be reported and discussed. TAC 2011 has three tracks:

1. Knowledge Base Population
   The goal of the KBP track is to promote research in automated systems that discover information about named entities as found in a large corpus and incorporate this information into a given knowlege base (namely, a KB derived from Wikipedia). The KBP track comprises the following tasks:
   - Entity-Linking Task: Given a name (of a Person, Organization, or Geopolitical Entity) and a document containing that name, determine the KB node for the named entity, adding a new node for the entity if it is not already in the KB. Two variants of the entity-linking task are offered: English-only, and cross-lingual (both English and Chinese documents).
   - Slot-Filling Task: Given a named entity and a pre-defined set of attributes ("slots") for the entity type, augment a KB node for that entity by extracting all new learnable slot values for the entity as found in a large corpus of documents.
   - Temporal Slot-Filling Task: Similar to the regular slot-filling task, but also specify time intervals for each extracted slot value. In addition to a full temporal slot-filling task, a diagnostic temporal task is offered, in which systems are provided with documents and correct slot values and only have to specify the temporal information.

2. Recognizing Textual Entailment
   The goal of the RTE Track is to develop systems that recognize when one piece of text entails another. RTE-7 pursues the direction of recognizing entailment in larger contexts -- a whole document or set of documents. RTE-7 comprises the following tasks:
   - Main and Novelty-Detection Tasks: Determine whether a given sentence -- in the context of an entire document -- entails a given Hypothesis.
   - Knowledge Base Population Validation Task: Determine whether a given document entails a given TAC KBP relation (e.g., "X is married to Y").

3. Summarization
   The goal of the Summarization Track is to develop systems that produce coherent summaries of text. The Summarization track comprises the following tasks:
   - Guided Summarization Task: Produce short, coherent summaries of news articles falling into predefined categories, guided by predefined aspects for each category.
   - Automatically Evaluating Summaries of Peers (AESOP) Task: Automatically score a summary for a given metric, including content (Pyramid score), overall responsiveness, and overall readability.
   - Multiling Pilot: Develop and apply partially or fully language-independent summarization algorithms to multiple languages, including Arabic, Czech, English, French, Greek, Hindi, and Hebrew.

The workshop will include presentation of results for each of the TAC 2011 tracks (including failure analyses and system comparisons), as well as more lengthy system presentations describing techniques used, experiments run on the data, and other issues of interest to researchers in NLP.

# Differences between RTE-1-5 and RTE-6-7

## RTE1-5 vs. RTE6 Main Task

| RTE1-5 | RTE6 |
| --- | --- |
| • RTE on isolated T-H pairs | • RTE within a corpus |
| • T-H pairs drawn from multiple applications | • Summarization application setting |
| • T and H do not contain references to information outside the pair itself | • Both T and H are to be interpreted within the context of the corpus |
| • The distribution of entailment is determined a priori | • Reflects the natural distribution of entailment in a corpus |

NIST - November 16, 2010 RTE-6@TAC2010

# Data format for RTE-1-5

```
<pair id="1" entailment="YES" task="IE" length="short" >
<t>The sale was made to pay Yukos' US$ 27.5 billion tax bill, Yuganskneftegaz was originally
sold for US$ 9.4 billion to a little known company Baikalfinansgroup which was later bought
by the Russian state-owned oil company Rosneft .</t>
<h>Baikalfinansgroup was sold to Rosneft.</h> </pair>

<pair id="2" entailment="NO" task="IE" length="short" >
<t>The sale was made to pay Yukos' US$ 27.5 billion tax bill, Yuganskneftegaz was originally
sold for US$9.4 billion to a little known company Baikalfinansgroup which was later bought by
the Russian state-owned oil company Rosneft .</t>
<h>Yuganskneftegaz cost US$ 27.5 billion.</h> </pair>

<pair id="3" entailment="NO" task="IE" length="long" >
<t>Loraine besides participating in Broadway's Dreamgirls, also participated in the Off-
Broadway production of "Does A Tiger Have A Necktie". In 1999, Loraine went to London, United
Kingdom. There she participated in the production of "RENT" where she was cast as "Mimi" the
understudy.</t>
<h>"Does A Tiger Have A Necktie" was produced in London.</h> </pair>

<pair id="4" entailment="YES" task="IE" length="long" >
<t>"The Extra Girl" (1923) is a story of a small-town girl, Sue Graham (played by Mabel
Normand) who comes to Hollywood to be in the pictures. This Mabel Normand vehicle, produced
by Mack Sennett, followed earlier films about the film industry and also paved the way for
later films about Hollywood, such as King Vidor's "Show People" (1928).</t>
<h>"The Extra Girl" was produced by Sennett.</h> </pair>
```

# RTE-6 Example

## RTE-6 Main Task Example

Topic 918: Betty Friedan

**Hs SET**

**H380** :Betty Friedan is the author of "The Feminine Mystique."
**H391 :** "The Feminine Mystique" was published in 1963.
**H401 :** In 1962, Judy Mott was laid off from her job with Sears.

### Document 1

**S1:** Betty Friedan, a founder of the modern feminist movement in the United States, died here Saturday of congestive heart failure, feminist leaders announced.

**S2**: She was 85.

**S3**: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women who were expected to find fulfillment through the achievements of their husbands and children.

**S4:** The book sparked a movement for a re-evaluation of women's role in American society and is credited with laying the foundation of modern feminism.

**S5:** She was a founder of the National Organization for Women and a leading advocate of the Equal Rights Amendment, a proposed amendment to the US constitution banning sex-based discrimination, women's rights activists said.

**S6:** "The movement that Friedan's energy sparked continues to grow, and is bigger today than she could ever have dreamed …

...

### Document 2

**S1**: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died Saturday, which was her 85th birthday.

**S2:** Friedan died of congestive heart failure at her home in Washington, D.C., according to Emily Bazelon, a cousin who was speaking for the family.

**S3:** She said Friedan had been in failing health for some time.

**S4:** Her best-selling book identified "the problem that has no name," the unhappiness of post-World War II American women unfulfilled by traditional notions of female domesticity.

**S5:**. Melding sociology and humanistic psychology, the book became the cornerstone of one of the last century's most profound movements, unleashing the first full flowering of American feminism since the 1800s.

**S6:** It gave Friedan, an obscure suburban New York housewife and freelance writer, the mantle to...

...

### Document 3

**S26:** What is perhaps most surprising, though, is not that feminists like Hirshman believe homemaking is second-class drudgery, but that so many people still get worked up over the issue.

**S27:** After all, feminist thinkers have been proclaiming the need to free women from the bondage of housework for a long time..

**S28:** It is, as Hirshman freely acknowledges, precisely what Friedan argued in "The Feminine Mystique," first published more than 40 years ago.

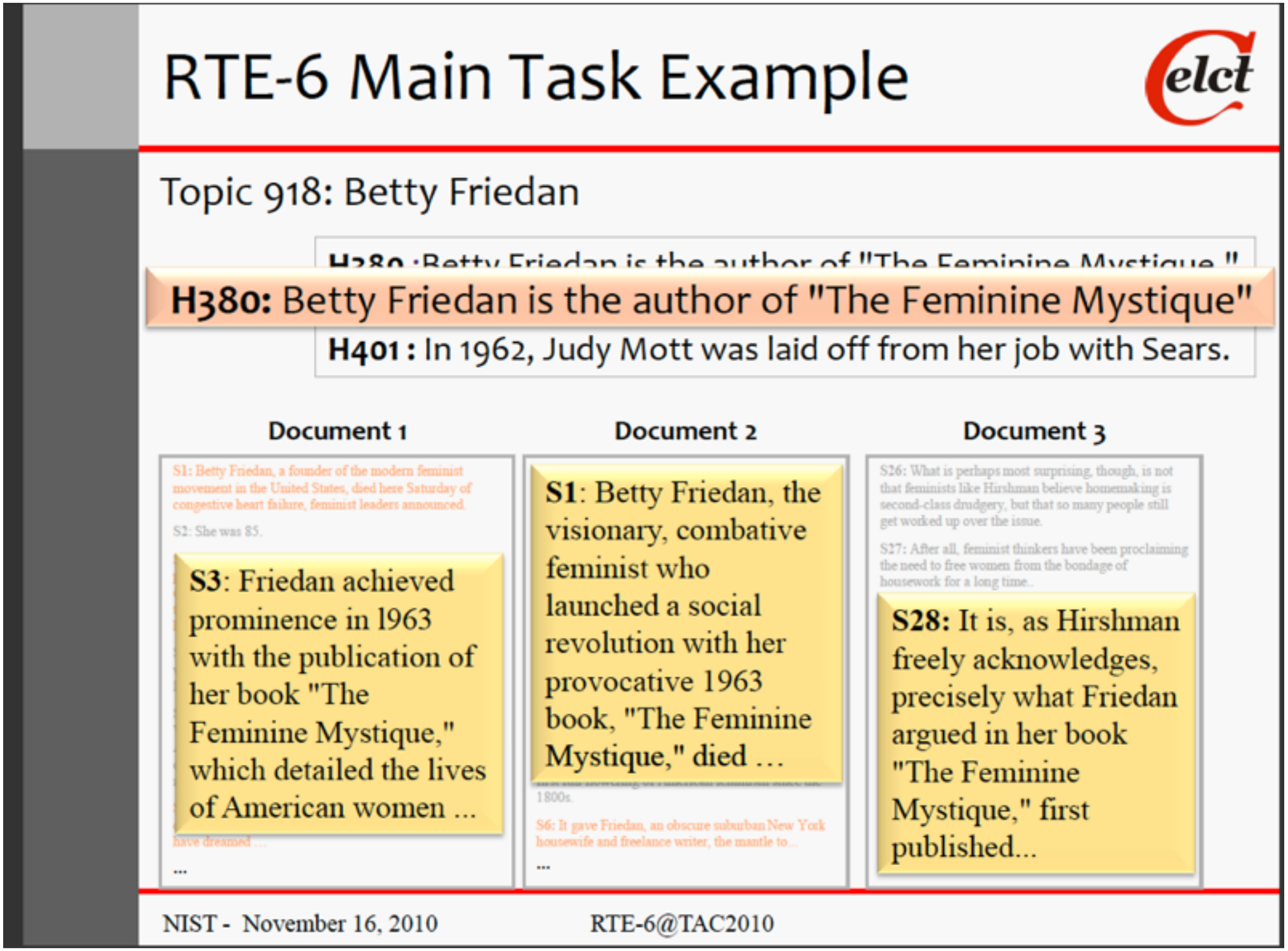
**S29** "The only kind of work which permits an able woman to realize her abilities fully," Friedan wrote, "is the kind that was forbidden by the feminine mystique, the lifelong commitment to an art or science, to politics or profession.".

**S30:** Not homemaking, not motherhood.

**S31:** In an interview, Hirshman said that in the course of researching a book, she began to wonder when feminism switched from offering a clear blueprint for liberation to choosing from Column A and Column B.

...

# RTE-6 Example



# Example Application: Multi-document Summarization


## Hypothesis

*Obama gave a speech last night in the Israeli lobby conference*

*Barack Obama's AIPAC address yesterday ...*

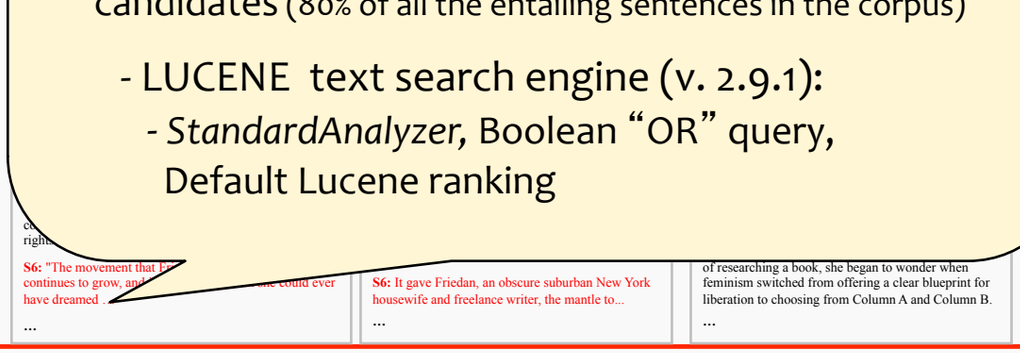
*In his speech at the American Israel Public Affairs Committee yesterday, the president challenged …*

## Texts

# RTE-7 Data Set Similar to RTE-6

## RTE-7 Main Data Set (2/2)

Up to 100 "candidate" entailing sentences
- Information Retrieval filtering phase:
  - The H is the query
  - The corpus sentences are "the documents" to be retrieved for the query
  - the 100 top-ranked sentences are selected as candidates (80% of all the entailing sentences in the corpus)
- LUCENE text search engine (v. 2.9.1):
  - *StandardAnalyzer*, Boolean "OR" query, Default Lucene ranking






## Data Set Composition

| DEVELOPMENT SET | | TEST SET | |
|---|---|---|---|
| Topics | 10 | Topics | 10 |
| **Hypotheses**<br>Entailment: yes \|no<br>Summaries: yes \|no | **284**<br>174 \| 110<br>193 \| 91 | **Hypotheses**<br>Entailment: yes \| no<br>Summaries: yes \| no | **269**<br>186 \| 83<br>192 \|77 |
| **Annotations** | **21,420** | **Annotations** | **22,426** |
| **"entailment" judg.** | **1,136** | **"entailment" judg.** | **1,308** |

- 3 annotations for the whole data set
- IAA (*Kappa*): 98.35% (Dev), 98.51% (Test)

# RTE-6 Main Task Description

- Given
  - a corpus
  - a hypothesis H
    - a set of "candidate" entailing sentences for that H retrieved by Lucene from the corpus
- RTE systems are required
  - to identify all the sentences among the candidate sentences that entail a given Hypothesis
  - „find all mentions  (Ts) of a sentence (H) in a corpus“

# Current Approaches and Methods

- Conventional methods
  - Assumption of independencies between words (Bag of Words) (*Corley and Mihalcea, 2005*)

- Measuring the distances between syntactic trees (*Kouylekov and Magnini, 2006*)

# Current Approaches and Methods

- Logical based rules
  - Logic rules (*Bos and Markert, 2005*)
  - Sequences of allowed transformations (*de Salvo Braz et al., 2005*)
  - Models of Knowledge Representation which is based on logical prove systems (*Tatu et al., 2006*)

# Current Approaches and Methods

- Machine Learning based approaches
  - Automatic determination of additional training material (*Hickl et al., 2006*) (**1st** in RTE-2)
  - Machine Learning methods based on tree kernels (*Zanzotto and Moschitti, 2006*) (3rd in RTE-2)

# Details for Two Approaches

- Transformation-based Approach
- Classification-based Approach

# Matching vs. Transformations

- Matching

The boy was located by the police.

Eventually, the police found the child.

- Sequence of transformations (Ako proof)

$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$

  - Tree-Edits
    - Complete proofs
    - Estimate confidence
  - Knowledge based Entailment Rules
    - Linguistically motivated
    - Formalize many types of knowledge

Next 7 slides from Stern et al. (2011), „ BIUTEE - Knowledge and Tree-Edits in Learnable Entailment Proofs“, RTE-7 workshop

# Transformation based RTE - Example

$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$

**Text:** The boy was located by the police.

**Hypothesis:** Eventually, the police found the child.

# Transformation based RTE - Example

$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$

**Text:** The boy was located by the police.

⇩

The police located the boy.

⇩

The police found the boy.

⇩

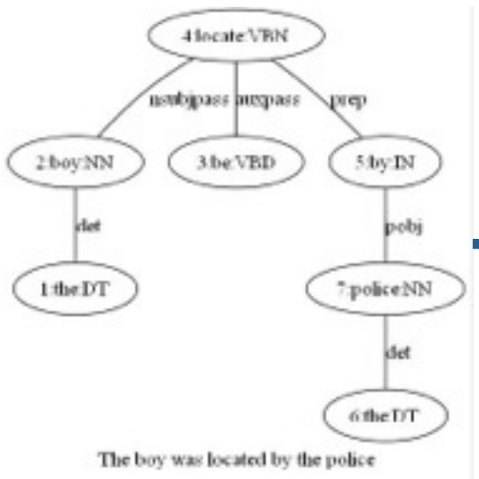
The police found the child.

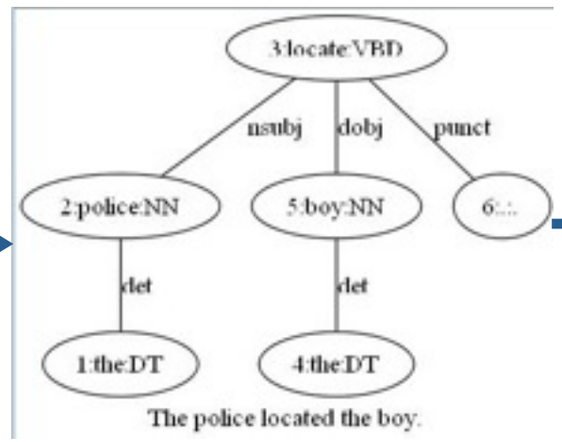
⇩

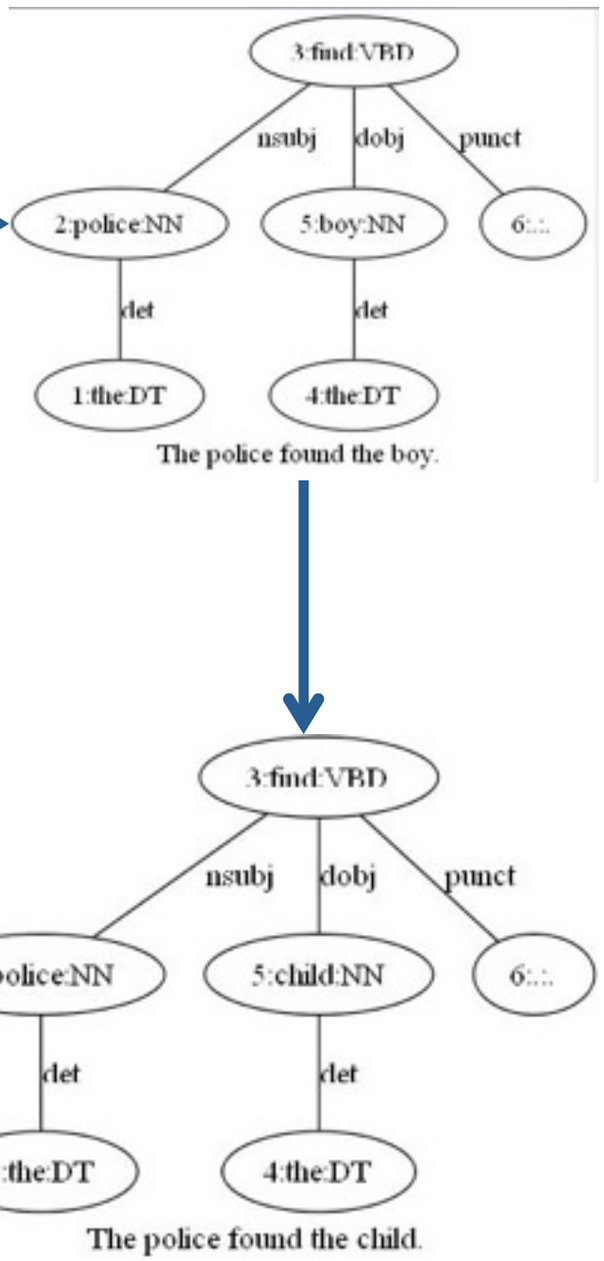
**Hypothesis:** Eventually, the police found the child.

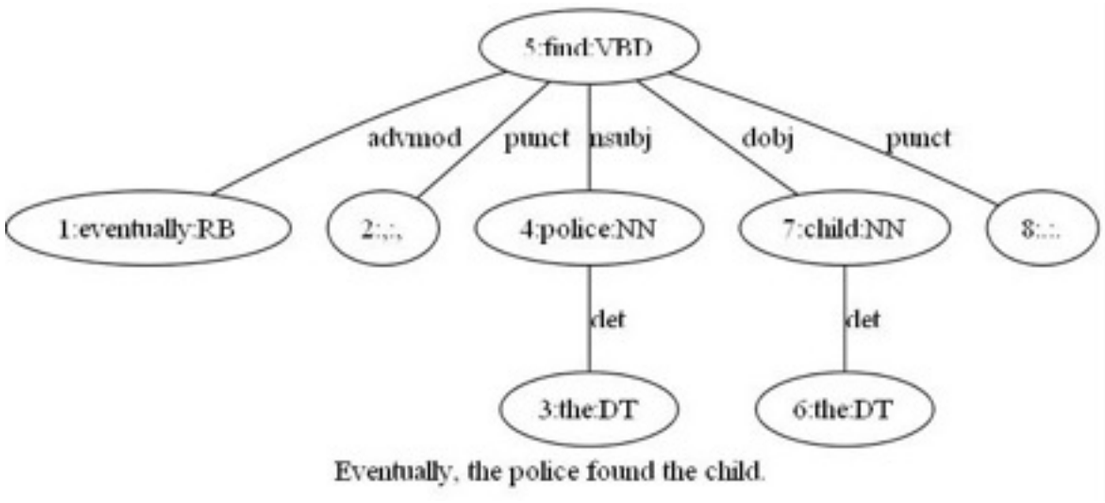
# Transformation based RTE - Example

$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$



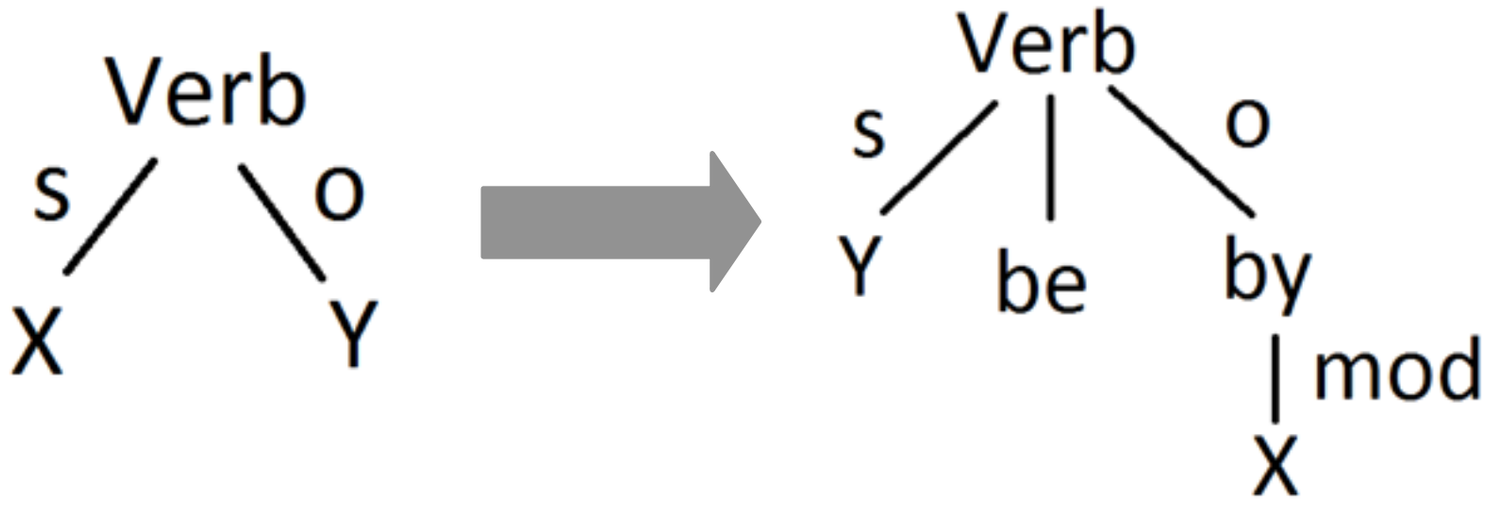



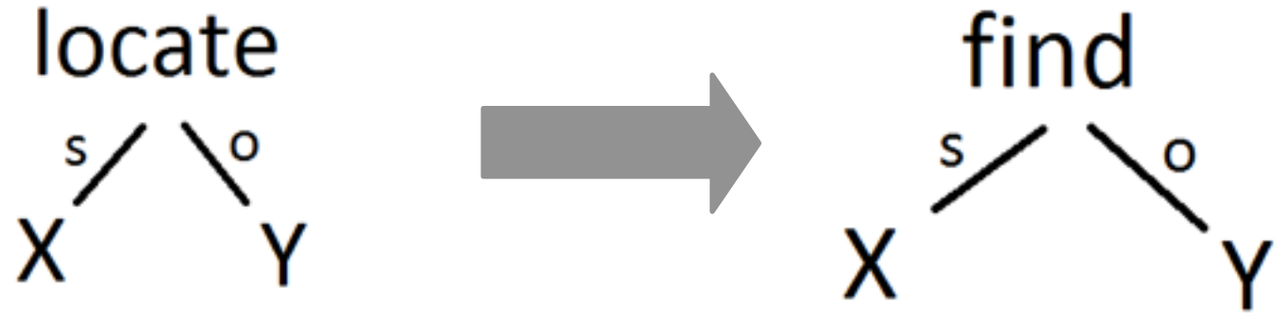



# Entailment Rules

Generic Syntactic




Lexical Syntactic




Lexical boy → child

Bar-Haim et al. 2007. ***Semantic inference at the lexical-syntactic level***.

# Proof over Parse Trees - Example

$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$

**Text:** The boy was located by the police.

Passive to active

The police located the boy.

X locate Y → X find Y

The police found the boy.

Boy → child

The police found the child.

Insertion on the fly

**Hypothesis:** Eventually, the police found the child.

# Results RTE7

| ID | Knowledge Resources | Precision % | Recall % | F1 % |
|---|---|---|---|---|
| BIU1 | WordNet, Directional Similarity | 38.97 | **47.40** | 42.77 |
| BIU2 | WordNet, Directional Similarity, Wikipedia | **41.81** | 44.11 | **42.93** |
| BIU3 | WordNet, Directional Similarity, Wikipedia, FrameNet, Geographical database | 39.26 | 45.95 | 42.34 |

DFKI-RTE7 result:

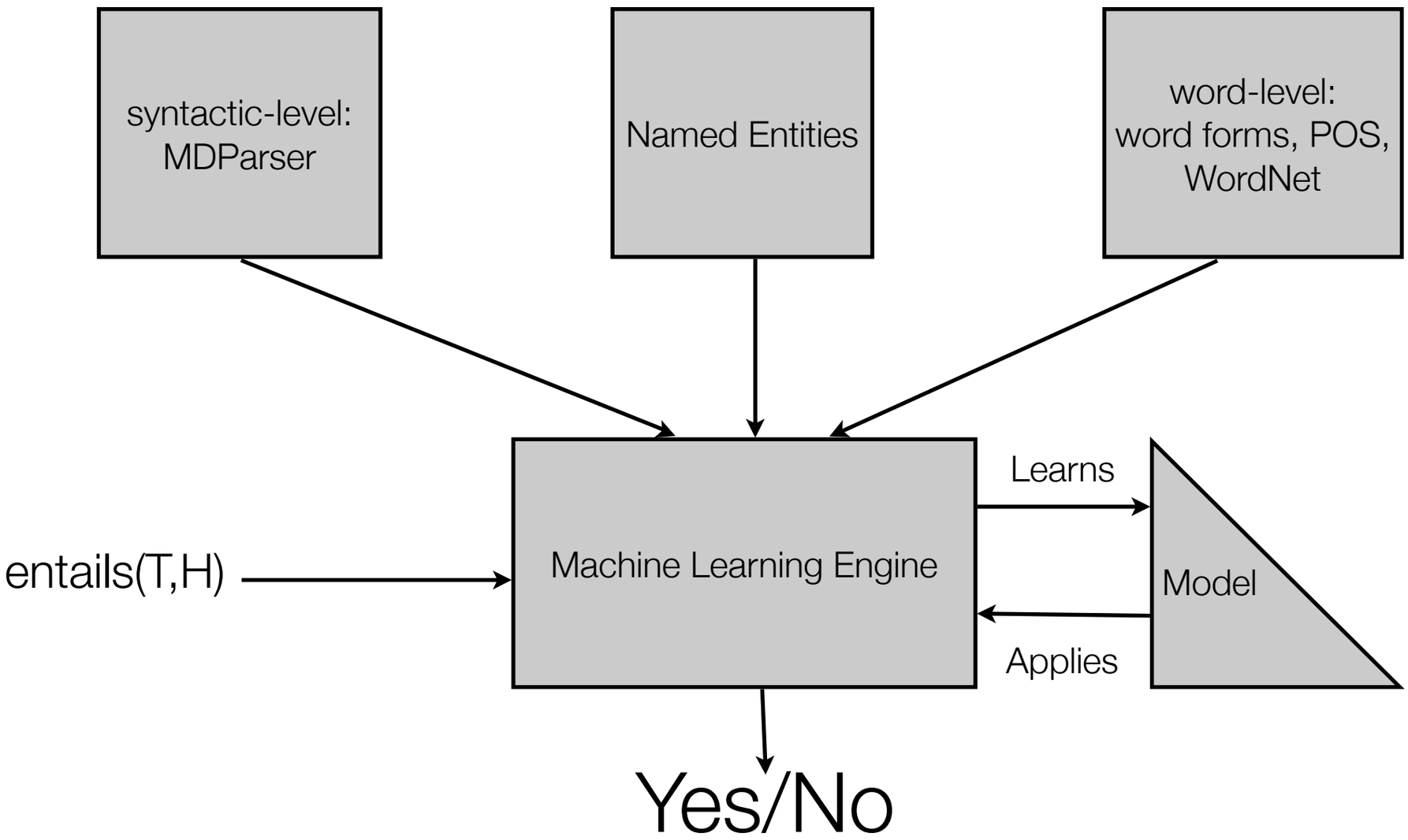
43.41 %

# DFKI System LITE

- LITE (Linear Classification for Textual Entailment)
- A single uniform machine learning classifier
- Focus on robustness, efficiency
- Features from
  - Linguistics tools (e.g., POS tagging, dependency parsing)
  - Knowledge bases (e.g., NER)
  - Text Alignment Tools

# DFKI-LITE for RTE-6

- A single machine learning engine (a linear SVM) is fed with features extracted from many different sources and learns to select the best



# RTE-6 - Results

## Best Results

| Team | Precision | Recall | F-measure |
|---|---|---|---|
| PKUTM2 | 68.57 | 36.93 | 48.01 |
| deb_iitb2 | 53.43 | 42.86 | 47.56 |
| IKOMA1 | 39.71 | 51.43 | 44.81 |
| FBK_irst3 | 43.46 | 46.03 | 44.71 |
| Boeing1 | 55.1 | 36.61 | 43.99 |
| DirRelCond21 | 38.99 | 41.8 | 40.35 |
| DFKI2 | 55.94 | 30.9 | 39.81 |
| SJTU_CIT3 | [illegible] | 46.67 | 39.57 |
| BIU1 | 37.54 | 37.46 | 37.5 |
| JU_CSE_TAC1 | 38.63 | 31.64 | 34.79 |
| *Baseline_Lucene5* | *30.78* | *39.58* | *34.63* |
| *Baseline_LuceneAll* | *4.73* | *100.00* | *9.03* |

NIST - November 16, 2010 RTE-6@TAC2010

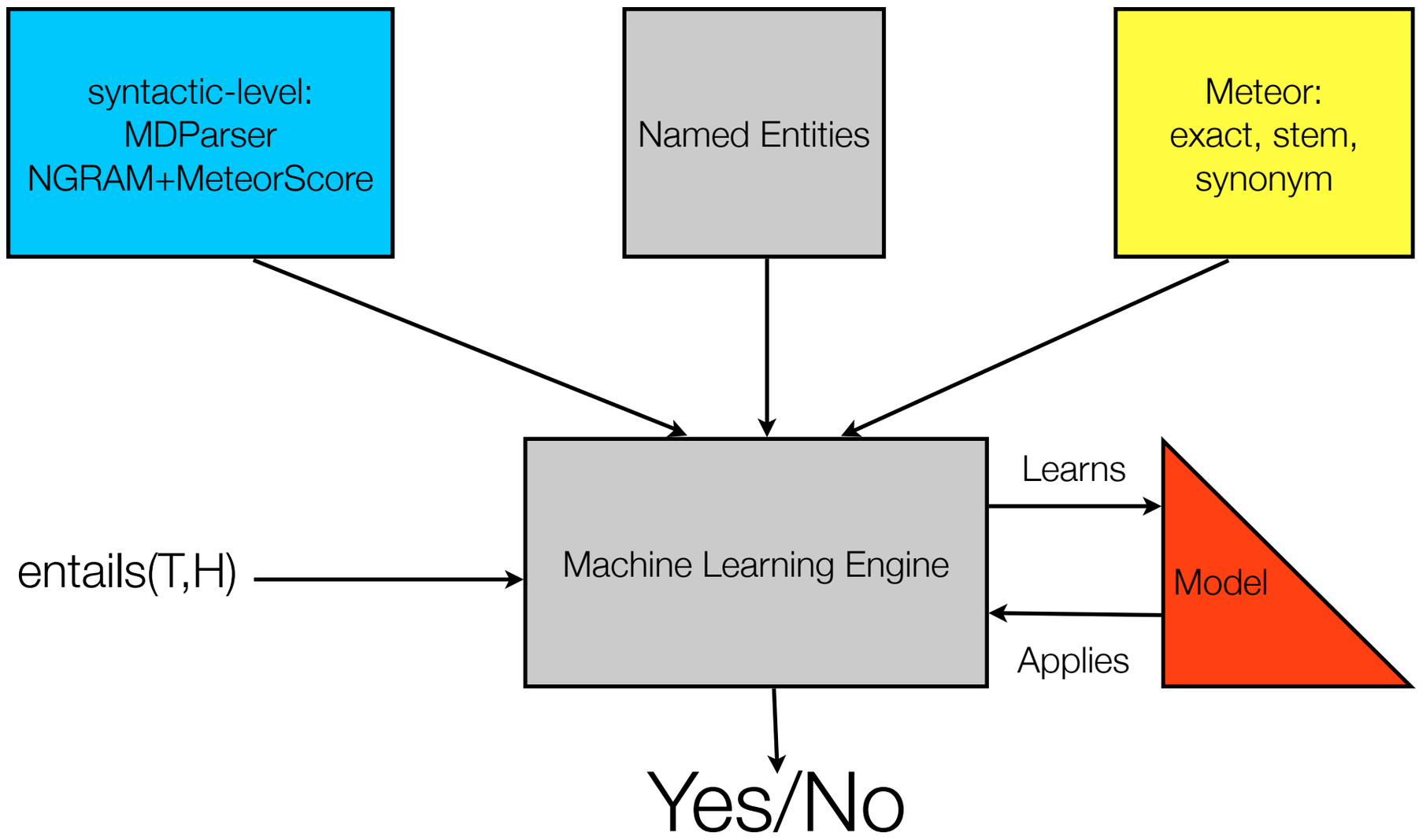
(Note: numbers of previous RTE-1-5 cannot be used for comparison; accuracy vs. F-Measure)

## Ablation test

| Test # | F Measure | Impact | Left out Features |
|---|---|---|---|
| 1 | 35.31 | 2.95 | 1 (root features) |
| 2 | 39.11 | -0.85 | 2 (depth 1 features) |
| 3 | 38.54 | -0.28 | 3 (depth 2 features) |
| 4 | 33.27 | 4.99 | 4 (word form and pos features) |
| 5 | 19.22 | 19.04 | 6 (content word features) |
| 6 | 36.04 | 2.22 | 7 (named entities features) |
| 7 | 38.49 | -0.23 | 5 (WordNet similarity features) |
| 8 | 36.72 | 1.54 | Coreference resolution features. No additional features were introduced or left out, but the content of all T-H-pairs was first processed with the LingPipe coreference resolution tool. |
| 9 | 39.10 | -0.84 | Threshold 0.15 |
| 10 | 39.11 | -0.85 | Threshold 0.13 |

# DFKI LITE - RTE-7

- A single machine learning engine (a linear SVM) is fed with features extracted from many different sources and learns to select the best (Volokh & Neumann, 2011)



# Main Task Evaluation

## 13 participants (33 runs)

- Evaluation measures:
  - Precision, Recall, F-measure (micro-averaged)
- IR Baselines:

| | Precision | Recall | F1 |
|---|---|---|---|
| Lucene_5 | 37.00 | 37.84 | 37.41 |
| Lucene_10 | 27.07 | 55.20 | 36.33 |
| Lucene_15 | 21.15 | 64.65 | 31.85 |
| Lucene_20 | 17.71 | 71.64 | 28.40 |
| Lucene_100 | 5.83 | 100 | 11.02 |



- 43.41 micro-average F1-score
- 46.34 macro-average F1-score

  → Above median, big improvement over the last year
- Very robust solution to an exremely large amount of data

  - >50% can be solved this way if account for weaknesses
- Problem-specific alternatives can still be included for the rest of the data

# Best Results

| Team | Precision | Recall | F-measure |
|---|---|---|---|
| IKOMA1 | 46.96 | 49.08 | 48.00 |
| u_tokyo3 | 46.84 | 43.58 | 45.15 |
| BUPTTeam1 | 45.02 | 44.95 | 44.99 |
| CELI1 | 41.88 | 46.56 | 44.10 |
| DFKI2 | 50.77 | 37.92 | 43.41 |
| BIU2 | 41.81 | 44.11 | 42.93 |
| FBK_irst3 | 46.59 | 38.07 | 41.90 |
| *Baseline_Lucene5* | *30.78* | *39.58* | *34.63* |
| te_iitb1 | 20.67 | 60.24 | 30.78 |
| JU_CSE_TAC2 | 26.66 | 35.55 | 30.47 |
| ICL1 | 47.88 | 21.56 | 29.73 |
| UAIC20112 | 30.21 | 25.84 | 27.85 |
| SJTU_CIT3 | 17.92 | 33.33 | 23.31 |
| SINAI3 | 47.3 | 8.72 | 14.72 |
| *Baseline_LuceneAll* | *4.73* | *100.00* | *9.03* |

# Ablation Tests - Results

The following table summarises the results:

| Test # | F Measure | Impact | Left out Features |
|---|---|---|---|
| 1 | 42.80 | -0.14 | WordNet features |
| 2 | 40.58 | 2.08 | LingPipe features |
| 3 | 42.71 | -0.05 | Time regular expressions |
| 4 | 40.51 | 2.15 | Meteor score feature |
| 5 | 40.42 | 2.24 | Meteor n-gram feature |
| 6 | 42.80 | -0.14 | Length feature |

The tests show that Meteor features are the most useful ones in the system. Named entity features have proven beneficial as well. However, other features seem to be not robust enough and were useful only for the development set and did not work for the test data.

# Summary

- Text inference is a hot topic
- EU project Excitement will further boost text inference for real-world research and applications:
  - We are providing an open-source platform for Textual Entailment
  - http://hltfbk.github.io/Excitement-Open-Platform/
- Web-scale RTE required
- New applications have to be considered ? -> what is the the RTE killer app?