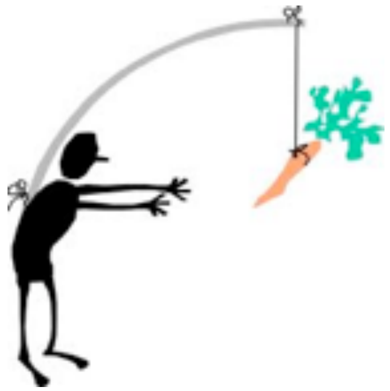
Textual Inference - Methods and Applications

Günter Neumann, LT Lab, DFKI, December 2012

Some slides are from Ido Dagan (BIU, Israel) and Bill Dolan (Microsoft Research, Seattle)

Motivation



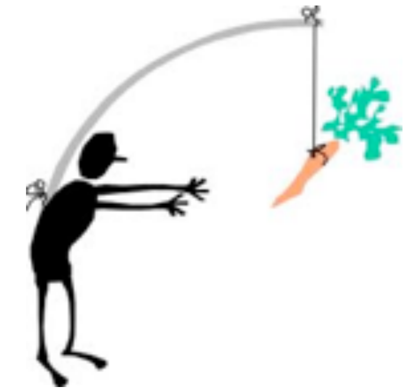
- Text-based applications need robust semantic inference engines

- 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



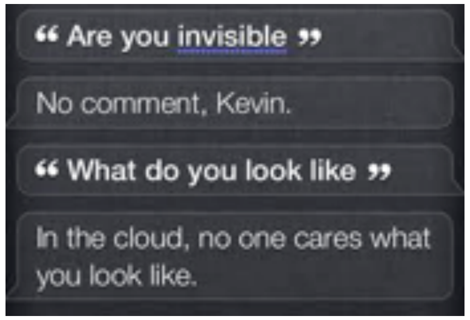
- Text-based applications need robust semantic inference engines

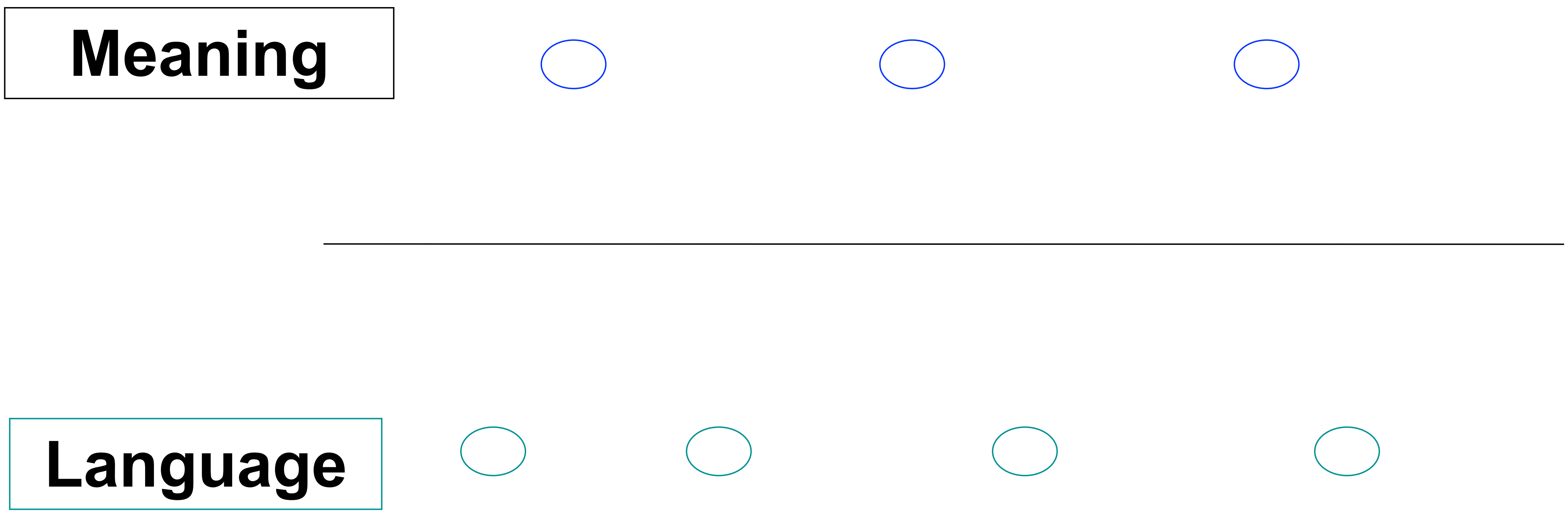
- Example: Open domain question answering

Q: Who is John Lennon's widow?

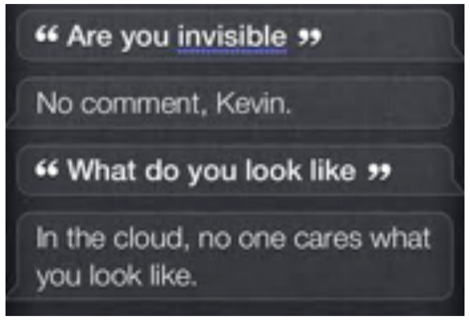
A: <u>Yoko Ono</u> unveiled a bronze statue of <u>her late husband</u>, <u>John Lennon</u>, to complete the official renaming of England's Liverpool Airport as Liverpool John Lennon Airport.

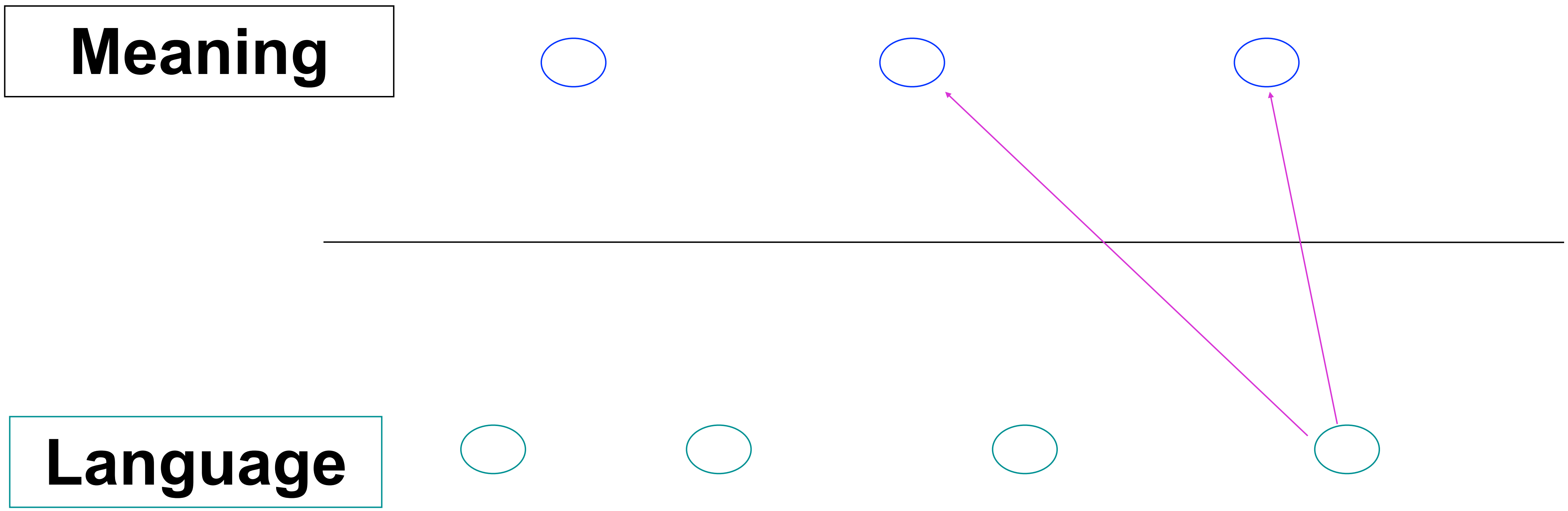
Natural Language and Meaning





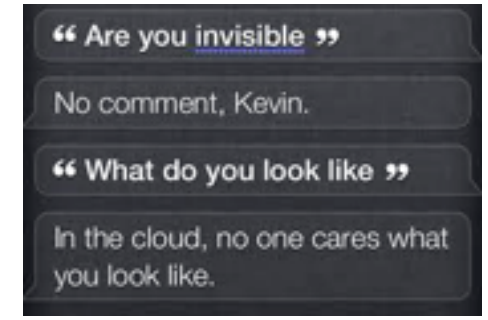
Natural Language and Meaning





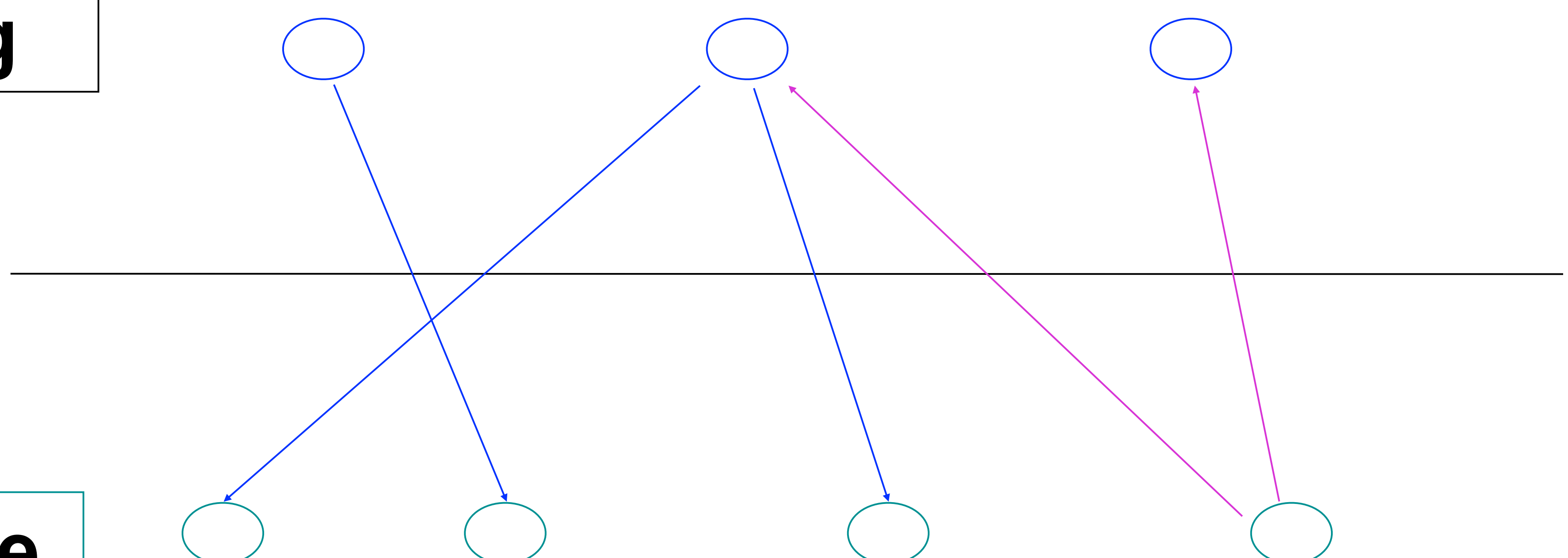
Ambiguity

Natural Language and Meaning



Variability

Meaning



Language

Ambiguity

Variability of Semantic Expression

Variability of Semantic Expression

All major stock markets surged

Dow gains 255 points

Dow ends up

Dow climbs 255

Stock market hits a record high

The Dow Jones Industrial Average closed up 255

Text-based Applications

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

- Unsupervised relation 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“.

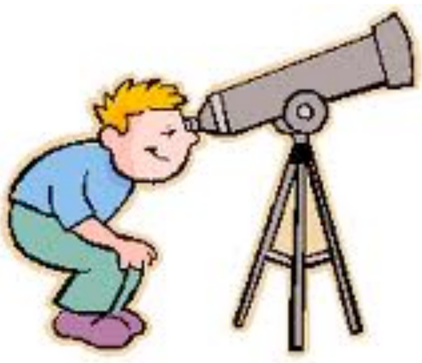
- Text summarization:
  Identify redundant information from multiple documents.

- 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.

Applied Textual Entailment: Relations between texts wrt. semantic entailment

Hypothesis (h): John Wayne was born in Iowa

Question: “Where was John Wayne Born ?“

Answer: Iowa

inference

Text (t): The birthplace of John Wayne is in Iowa

Generic Entailment as a Task

Hypothesis (h): John Wayne was born in Iowa

Given text t, is it possible to infer that h (quite likely) is true ?

inference

Text (t): The birthplace of John Wayne is in Iowa

Classical Entailment

- Chierchia & McConnell-Ginet (2001):
  A text t entails a hypothesis h, if h is true in all circumstances (possible worlds) where t is true.

- Very strict - does not consider uncertainties which are common in real-world applications.

“Nearly exact” Entailment

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

h: Ivan Getting invented the GPS.

t: According to the Encyclopedia Britannica, Indonesia is the largest archipelagic nation in the world, consisting of 13,670 islands.

h: 13,670 islands make up Indonesia.

Textual Entailment ≈ Human Reading Comprehension

- From a school book (Sela and Greenberg):

- Reference test: *"…The Bermuda Triangle lies in the Atlantic Ocean, <u>off the coast</u> of <u>Florida</u>. …"*

- Hypotheses (True/False?): *The Bermuda Triangle is <u>near</u> the <u>United States</u>*

???

Machine Reading

By Canadian Broadcasting Corporation

T: The school has turned its one-time metal shop – lost to budget cuts almost two years ago - into a money-making professional fitness club.

Q: When did the metal shop close?

A: Almost two years ago

Machine Reading

By Canadian Broadcasting Corporation

T: The school has turned its one-time metal shop – <u>lost to budget cuts</u> almost two years ago - into a money-making professional fitness club.

Q: When did the metal shop <u>close</u>?

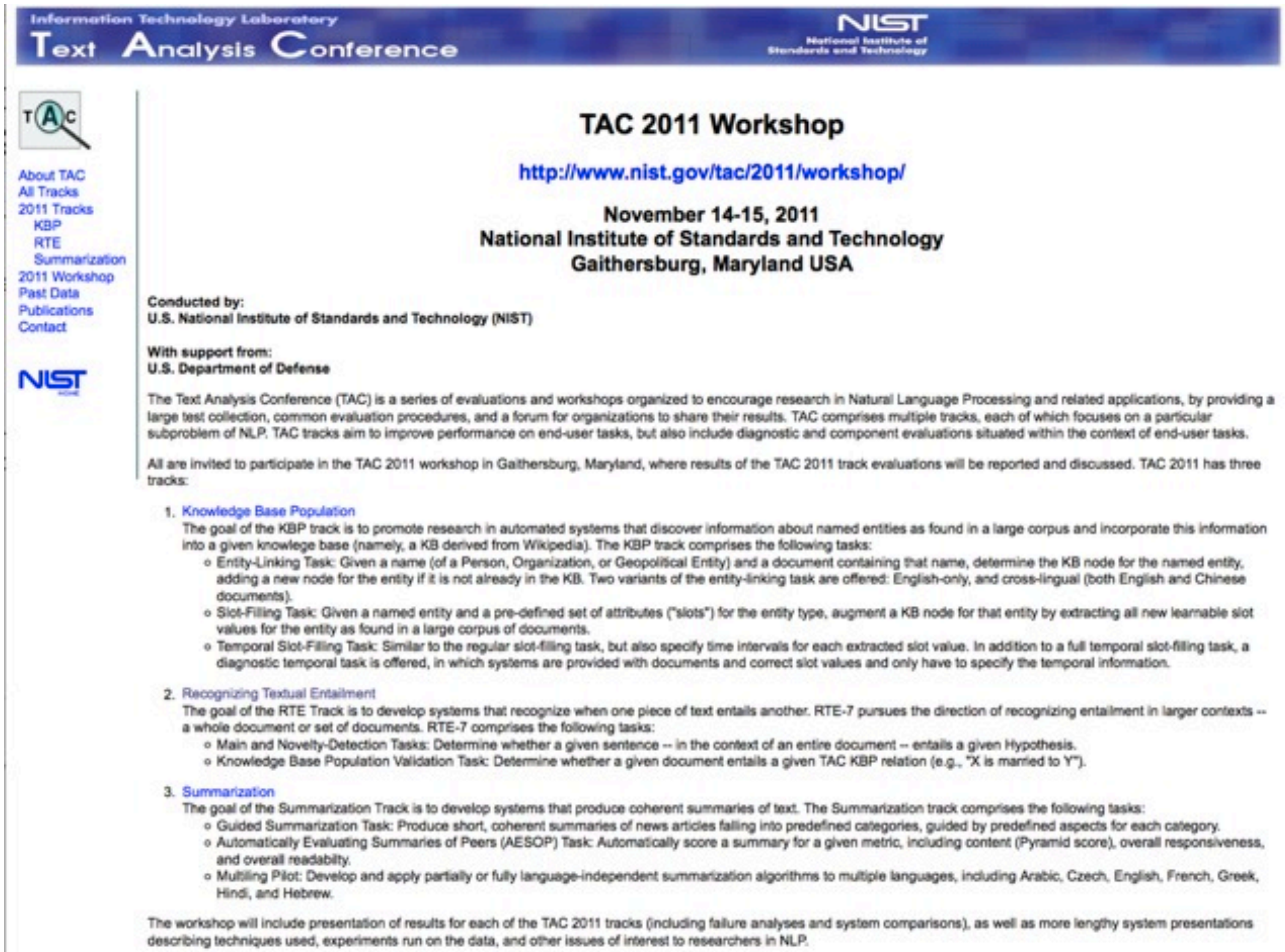
A: Almost two years ago

Two possible approaches:

a) System answers questions, which come from outside (QA)
b) System generate its own question, which are answered from outside (E-Learning)

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:

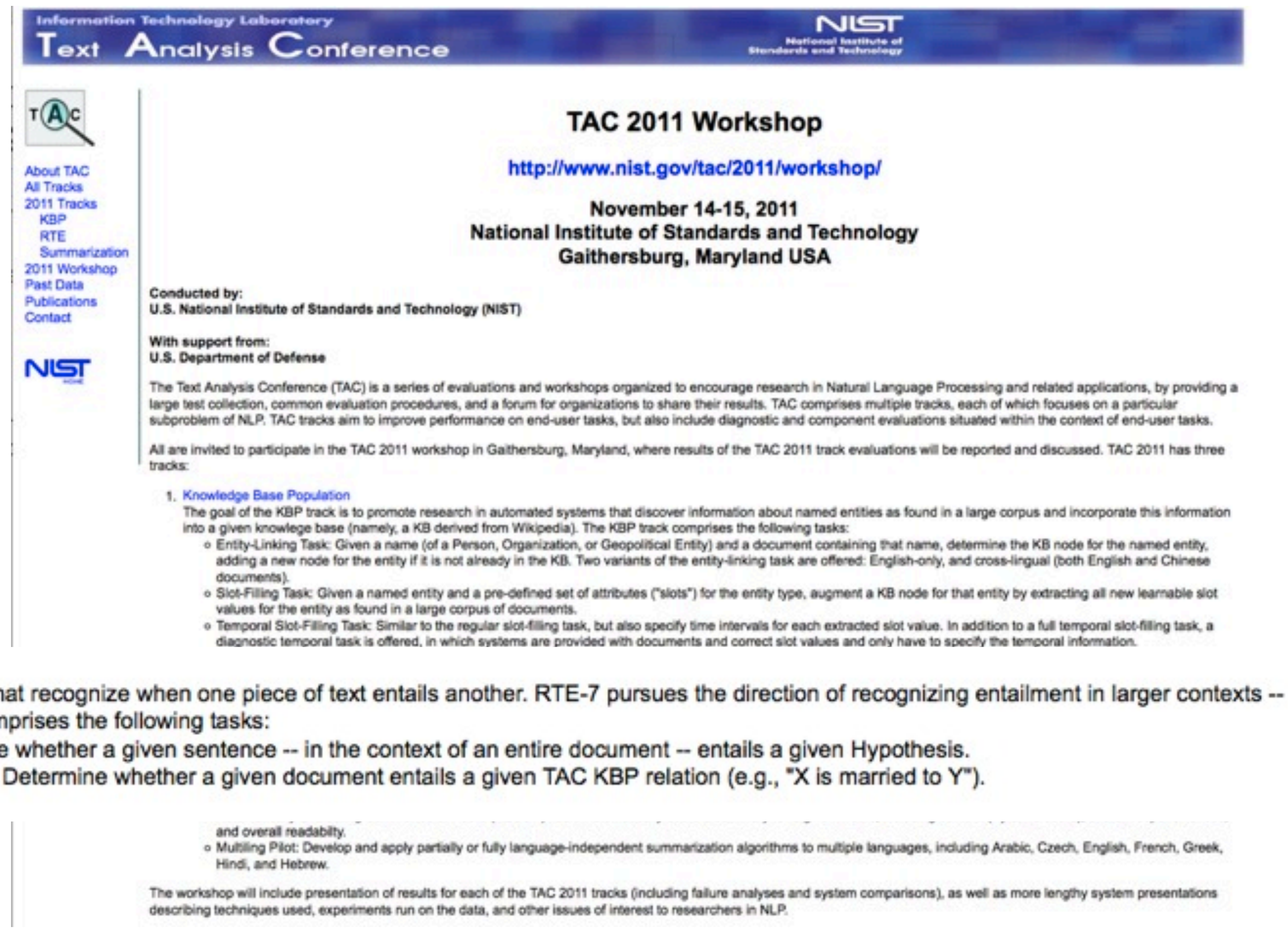
1. Knowledge Base Population
   The goal of the KBP track is to promote research in automated systems that discover information about named entities as found in a large corpus and incorporate this information into a given knowlege base (namely, a KB derived from Wikipedia). The KBP track comprises the 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.

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 knowledge 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").

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

- RTE on isolated T-H pairs
- T-H pairs drawn from multiple applications
- T and H do not contain references to information outside the pair itself
- The distribution of entailment is determined a priori

RTE6

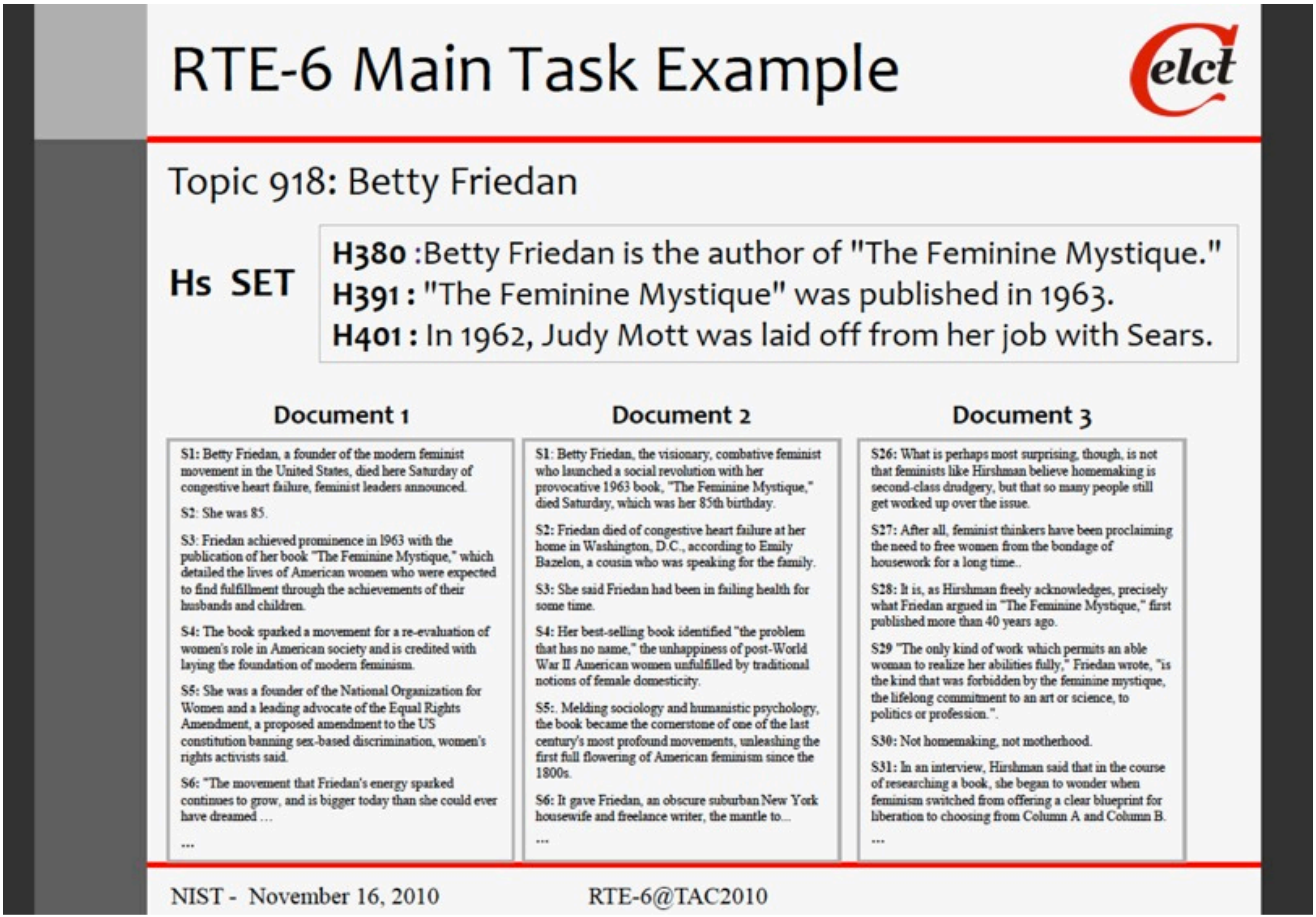
- RTE within a corpus
- Summarization application setting
- Both T and H are to be interpreted within the context of the corpus
- Reflects the natural distribution of entailment in a corpus

NIST - November 16, 2010 RTE-6@TAC2010

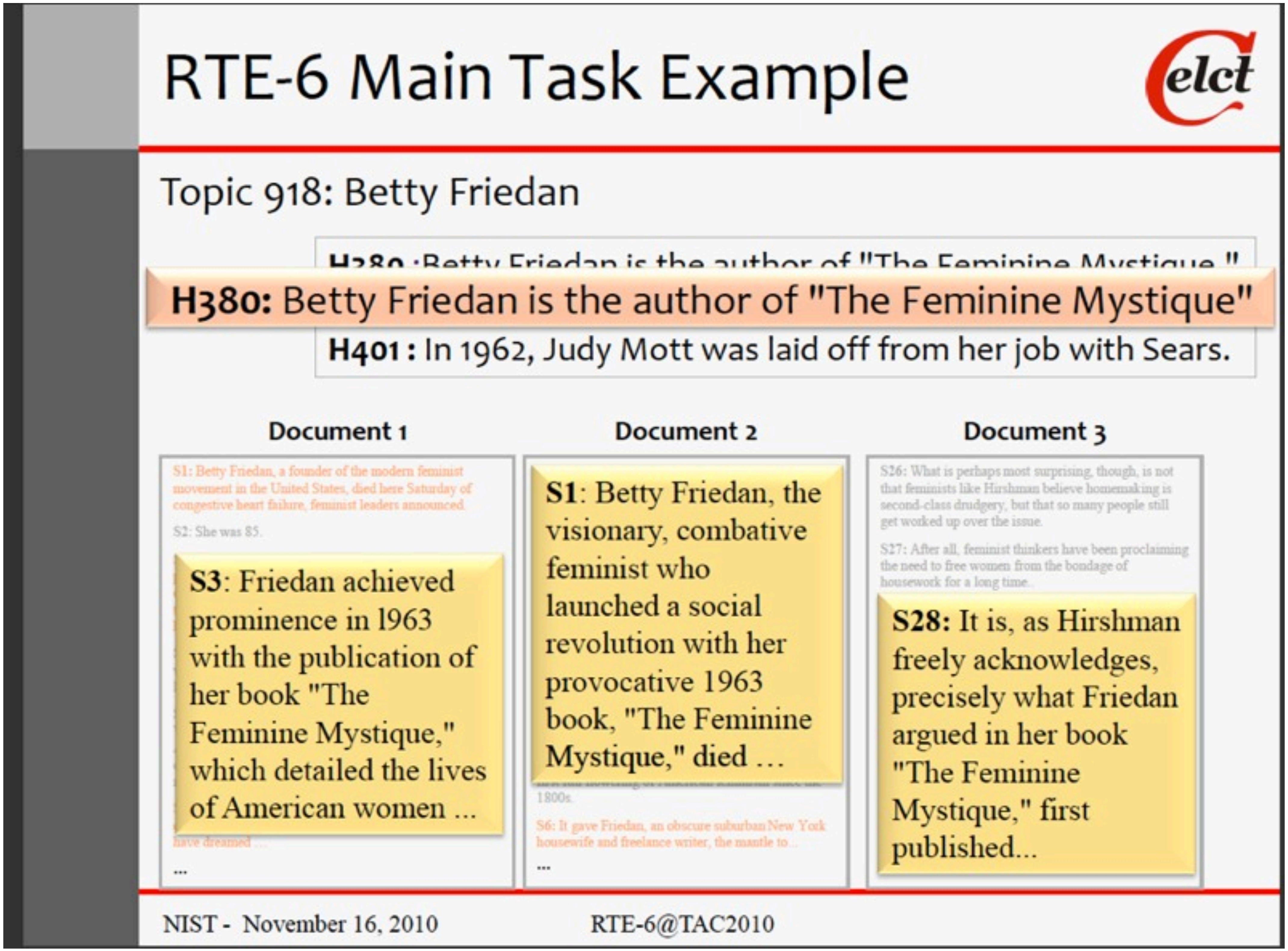
Data format for RTE-1-5

```
<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 Example



Another Example in XML Style

RTE6_Main_DEVSET_GS.xml


```xml
<entailment_corpus>
    <TOPIC t_id="D0929">
        <H h_id="2">
            <H_sentence>Rita picked up strength.</H_sentence>
            <text doc_id="AFP_ENG_20050920.0413" s_id="2" evaluation="YES">Hurricane Rita was upgraded from a tropical storm as it threatened the southeastern United States, forcing an alert in southern Florida and scuttling plans to repopulate New Orleans after Hurricane Katrina turned it into a ghost city three weeks earlier.</text>
            <text doc_id="AFP_ENG_20050920.0534" s_id="0" evaluation="YES">Hurricane Rita strengthens to category two</text>
            <text doc_id="AFP_ENG_20050920.0534" s_id="1" evaluation="YES">Hurricane Rita strengthened to category two, packing winds of 160 kilometers per hour (100 mph) early Tuesday afternoon, the Miami-based National Weather Center announced.</text>
            <text doc_id="AFP_ENG_20050920.0545" s_id="3" evaluation="YES">The Miami-based National Weather Center said the hurricane strengthened Tuesday, packing winds of 160 kilometers per hour (100 mph).</text>
            <text doc_id="AFP_ENG_20050920.0561" s_id="0" evaluation="YES">Hurricane Rita strengthens to category two</text>
            <text doc_id="AFP_ENG_20050920.0561" s_id="1" evaluation="YES">Hurricane Rita strengthened to category two, packing winds of 160 kilometers per hour (100 mph) early Tuesday afternoon, the Miami-based National Weather Center announced.</text>
            <text doc_id="AFP_ENG_20050920.0633" s_id="1" evaluation="YES">Louisiana Governor Kathleen Blanco declared a state of emergency in western Louisiana on Tuesday and urged people to evacuate as strengthening Hurricane Rita headed toward the Gulf of Mexico.</text>
            <text doc_id="NYT_ENG_20050920.0246" s_id="0" evaluation="YES">STATES ON GULF TAKE PRECAUTION AS STORM GROWS</text>
        </H>
        <H h_id="3">
            <H_sentence>Hurricane Rita could reach top intensity before it hits land.</H_sentence>
        </H>
        <H h_id="4">
            <H_sentence>Rita could slam Texas.</H_sentence>
            <text doc_id="AFP_ENG_20050920.0614" s_id="0" evaluation="YES">Katrina refugees forced to flee Texas shelters amid Rita threat</text>
            <text doc_id="AFP_ENG_20050920.0614" s_id="1" evaluation="YES">More than 7,000 Hurricane Katrina survivors taking refuge in Texas shelters were being uprooted again Tuesday as Hurricane Rita threatened the state neighboring Louisiana, officials said.</text>
```


Montag, 17. Dezember 12

Another Example in XML Style

RTE6_Main_DEVSET_GS.xml


```
<entailment_corpus>
    <TOPIC t_id="D0929">
        <H h_id="2">
            <H_sentence>Rita picked up strength.</H_sentence>
            <text doc_id="AFP_ENG_20050920.0413" s_id="2" evaluation="YES">Hurricane Rita was upgraded from a tropical storm as it threatened the
southeastern Unit
ghost city three
            <text
            <text
per hour (100 mph)
            <text
Tuesday, packing
            <text
            <text
per hour (100 mph)
            <text
Louisiana on Tues
            <text
        </H>
        <H h_id="
            <H_se
        </H>
        <H h_id="
            <H_se
            <text
            <text
were being uproot
```


AFP_ENG_20050920.0413.xml


```
<DOC doc_id="AFP_ENG_20050920.0413" type="story">
<HEADLINE>
Oil prices drop amid OPEC output pledge, Hurricane Rita
</HEADLINE>
<DATELINE>
LONDON, Sept 20
</DATELINE>
<TEXT>
<S s_id="0">
Oil prices drop amid OPEC output pledge, Hurricane Rita
</S>
<S s_id="1">
World oil prices fell further on Tuesday, despite a new hurricane powering towards oil facilities in the Gulf of Mexico, and as OPEC pledged to supply more crude from the start of October if required.
</S>
<S s_id="2">
Hurricane Rita was upgraded from a tropical storm as it threatened the southeastern United States, forcing an alert in southern Florida and scuttling plans to repopulate New Orleans after Hurricane Katrina turned it into a ghost city three weeks earlier.
</S>
<S s_id="3">
The Organization of Petroleum Exporting Countries, meanwhile, pledged to make an extra two million more barrels per day available to oil markets from October 1 if there were adequate demand, while holding its official production ceiling at 28 million barrels per day.
</S>
<S s_id="4">
New York's main contract, light sweet crude for delivery in October, which lapses later Tuesday, sank 1.09 dollars to 66.30 dollars per barrel during early trading.
</S>
<S s_id="5">
In London, the price of Brent North Sea crude for November delivery lost 1.29 dollars to 64.32 dollars per barrel.
</S>
<S s_id="6">
Crude futures extended earlier losses, made on profit taking, after surging more than four dollars on Monday as Rita looked set to reach hurricane strength over the oil-rich Gulf Coast.
</S>
</TEXT>
</DOC>
```

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

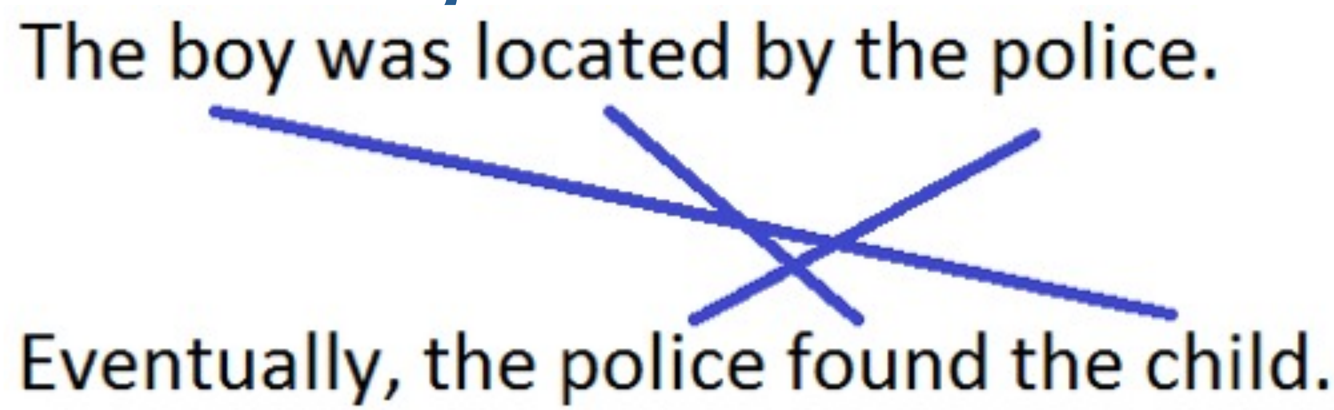
- Machine Learning based approaches
  - Automatic determination of additional training material (*Hickl et al., 2006*) (**1st** in RTE-2)
  - Machine Learning methods based on tree kernels (*Zanzotto and Moschitti, 2006*) (3rd in RTE-2)

Matching vs. Transformations

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

Matching vs. Transformations

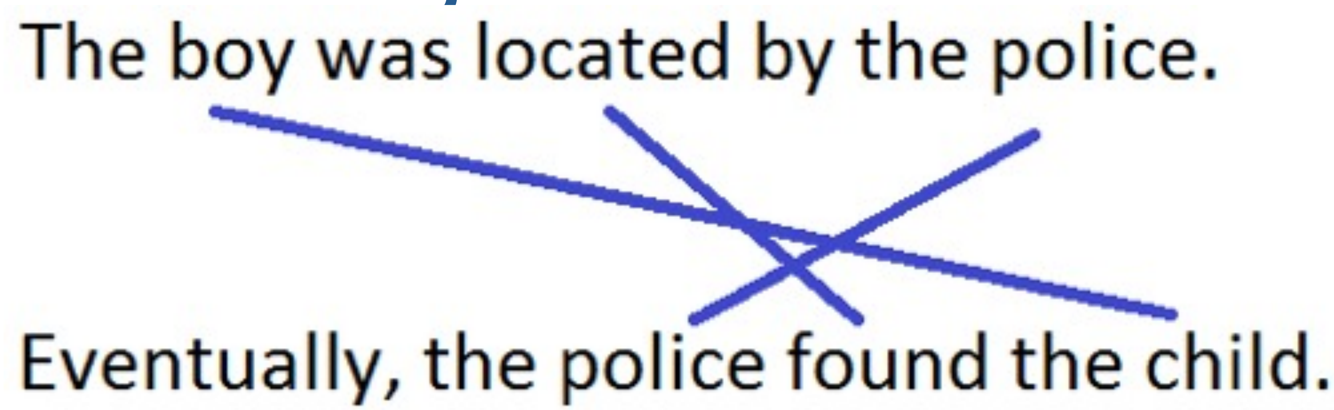
- Matching



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

Matching vs. Transformations

- Matching

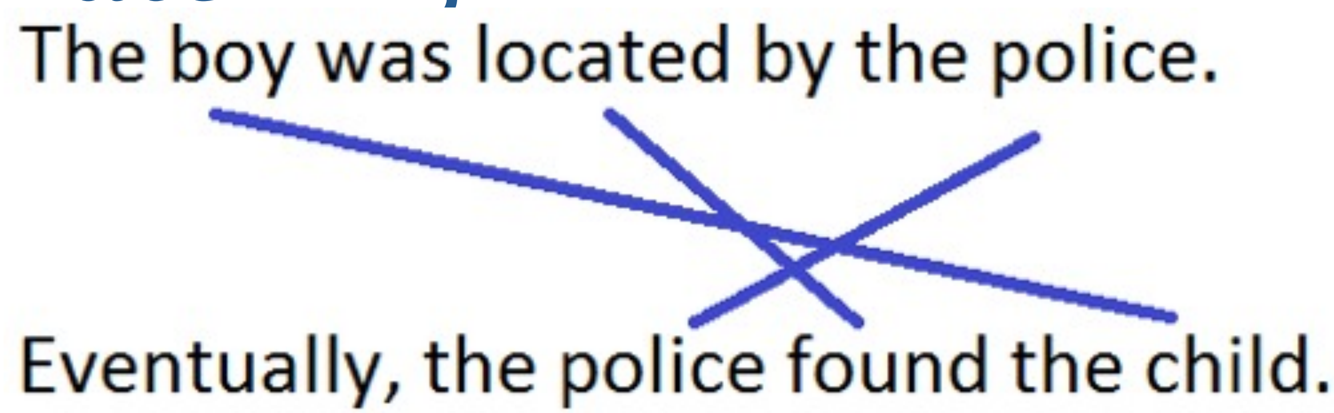


$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \ldots \rightarrow T_n = H$$

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

Matching vs. Transformations

- Matching

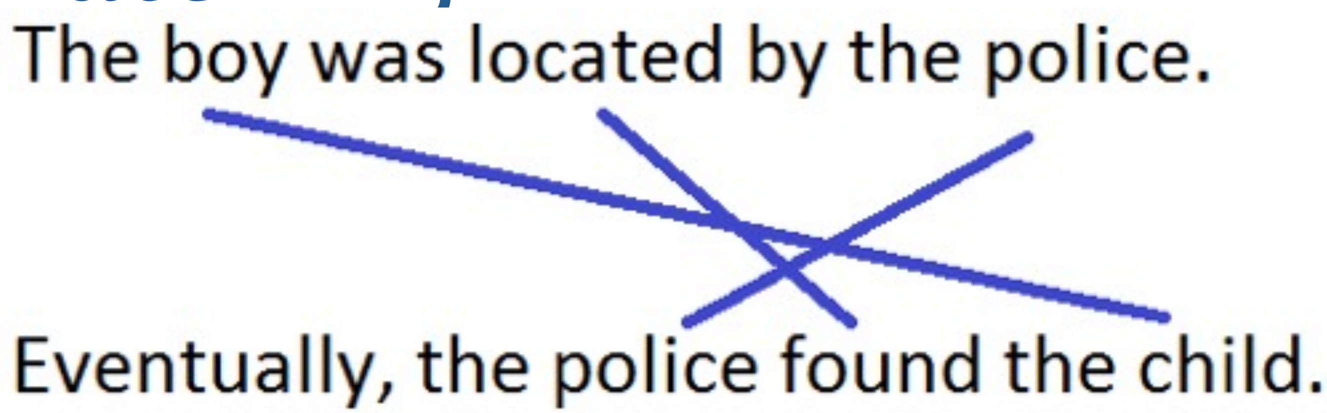


$$T = T_0 \rightarrow T_1 \rightarrow T_2 \rightarrow \dots \rightarrow T_n = H$$

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

Matching vs. Transformations

- Matching



- 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

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.

Hypothesis: Eventually, the police found the child.

Transformation based RTE – Example

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

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

Text: The boy was located by the police.

⇩

<u>The police located the boy.</u>

⇩

The police <u>found</u> the boy.

⇩

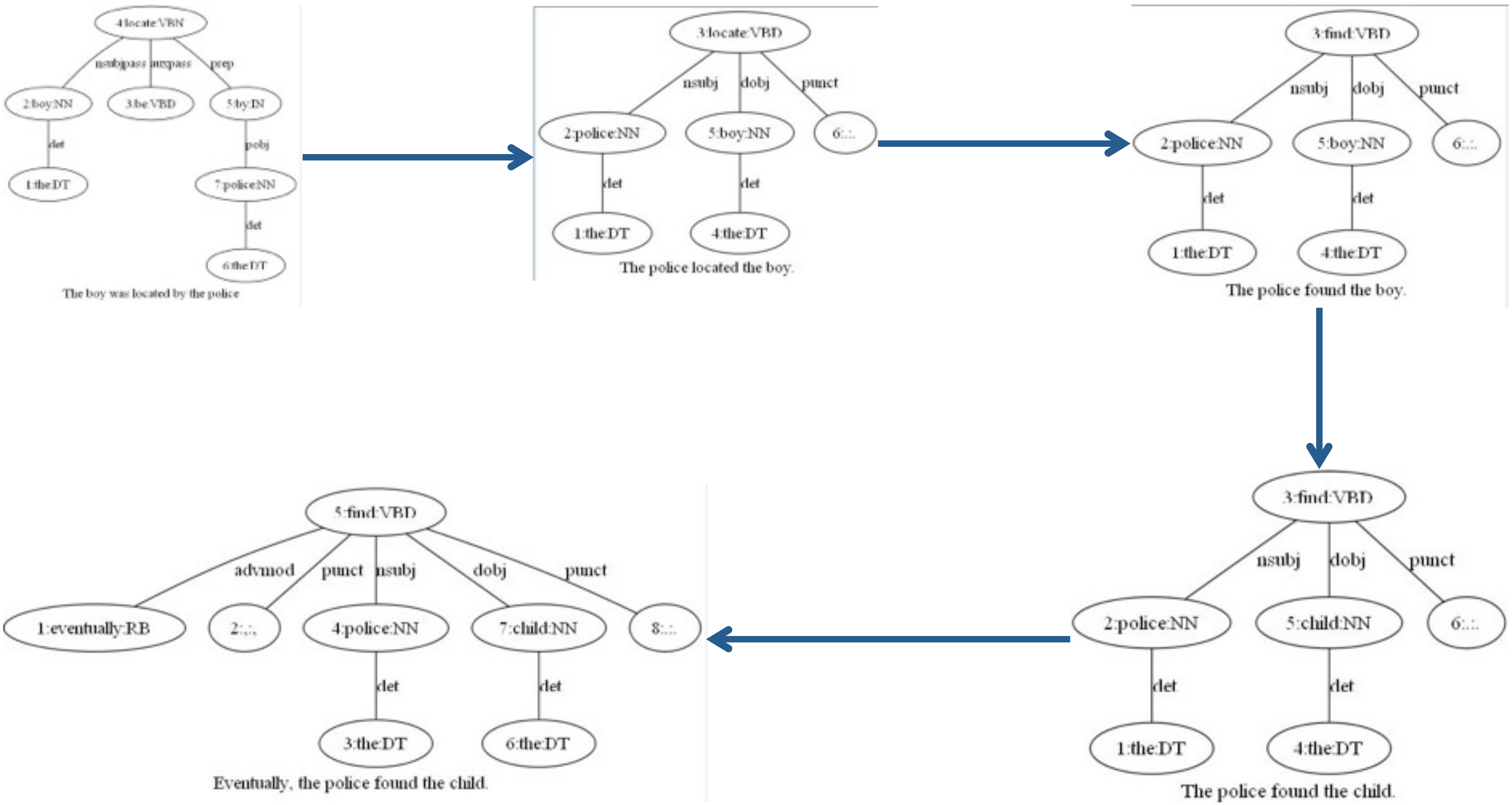
The police found the <u>child</u>.

⇩

Hypothesis: <u>Eventually</u>, the police found the child.

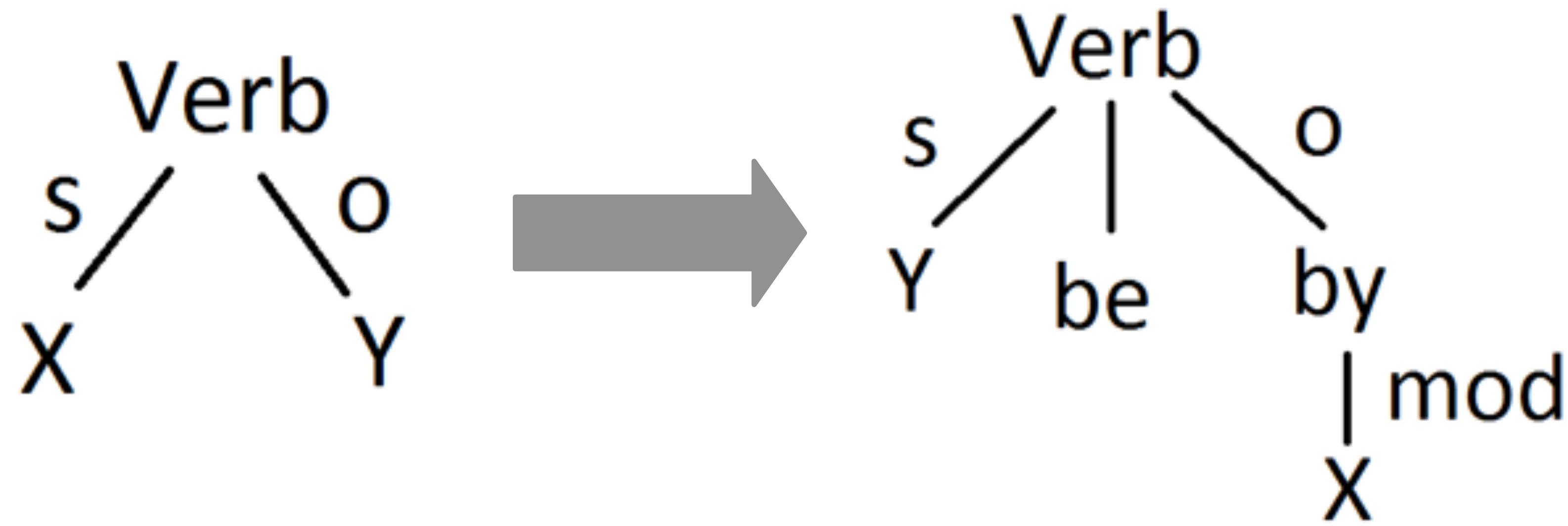
Transformation based RTE – Example

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

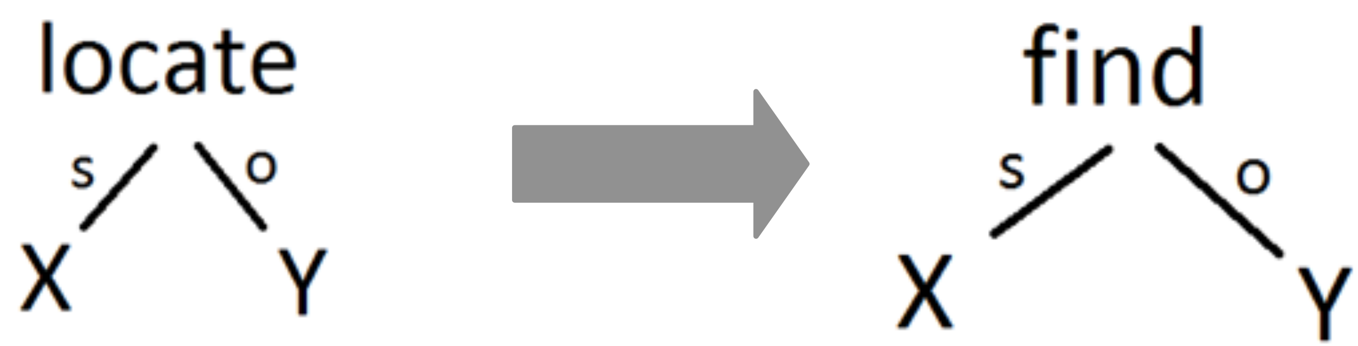


Entailment Rules

Generic Syntactic



Lexical Syntactic



Lexical

boy → child

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

Proof over Parse Trees – Example

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.

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.

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

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.

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

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.

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

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.

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

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 %

Results RTE7

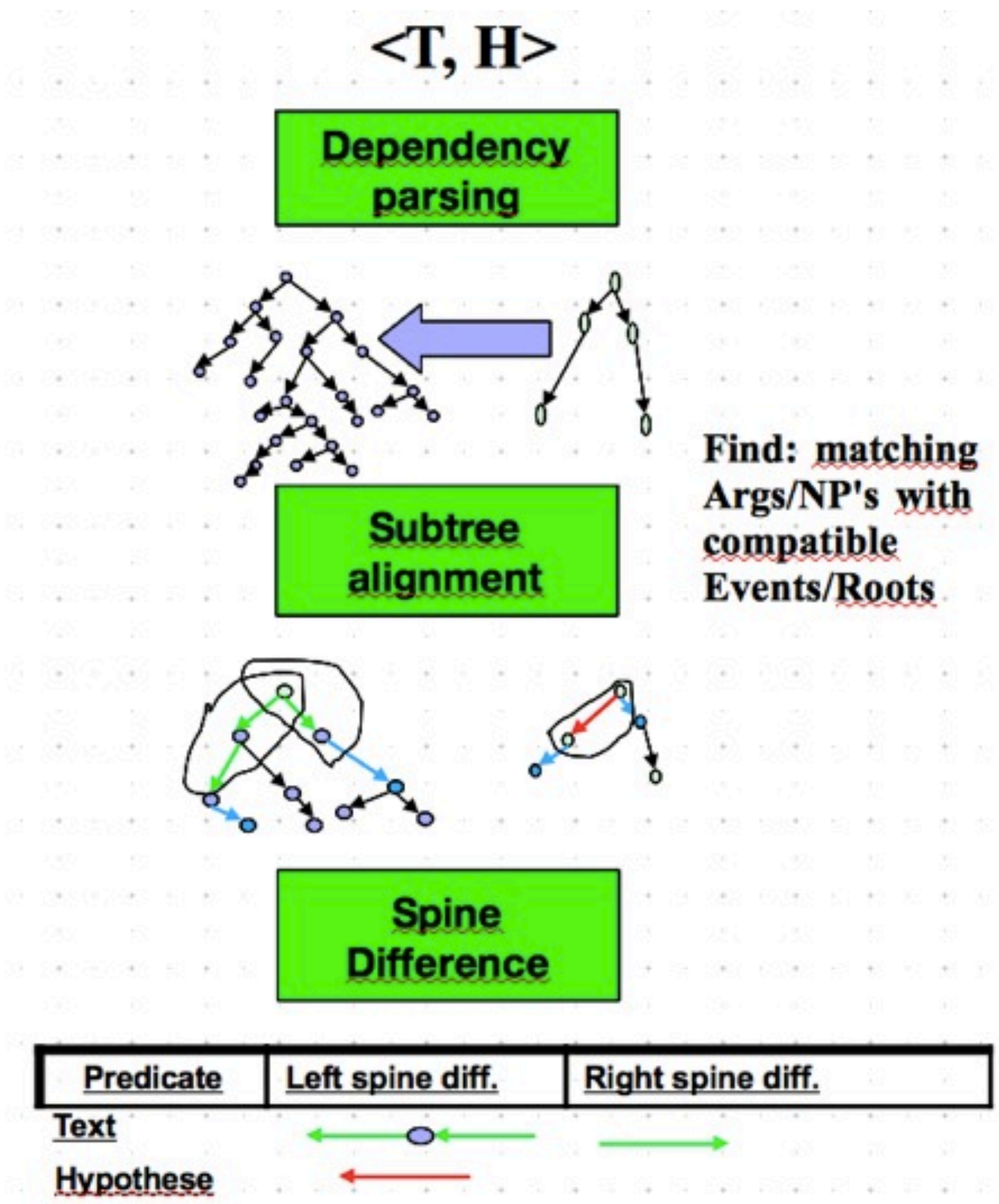
ID	Knowledge Resources	Precision %	Recall %	F1 %
BIU1	WordNet, Directional Similarity	38.97	**47.40**	42.77
BIU2	WordNet, Directional Similarity, Wikipedia	**41.81**	44.11	**42.93**
BIU3	WordNet, Directional Similarity, Wikipedia, FrameNet, Geographical database	39.26	45.95	42.34

BIUTEE 2011 on RTE 6 (F1 %)	
Base line (Use IR top-5 relevance)	34.63
Median (September 2010)	36.14
Best (September 2010)	48.01
Our system	**49.54**

DFKI-RTE7 result:
43.41 %

DFKI - How far can we go with syntax only ? cf. Wang & Neumann, AAAI, 2007.

- Goal: Achieve a possible maximal syntactic baseline
- Method:
- Compare similarity of dependency trees of H and T
- Tree compression: only consider relevant parts of the dependency trees
    - avoid noise generated by the parsers
    - can be used to construct compressed syntactic path information
- Feature extraction on basis of partial sequences
    - Consider all possible sequences of path differences
    - Linear SMV for learning classification (binary threshold)



Performance of the Puristic Syntax Approach using RTE-3 results

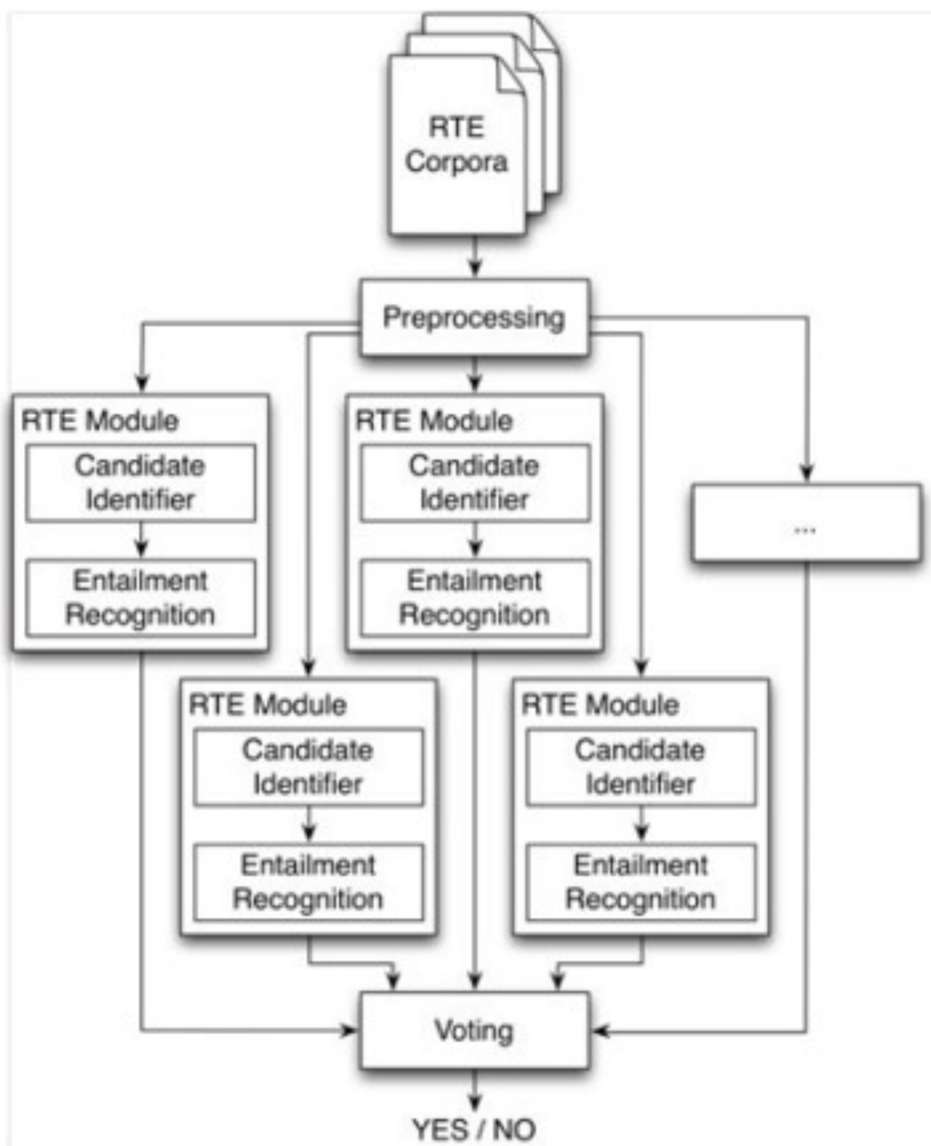
Systems	Acc. %	Lx*	Ng	Sy	Se	LI	C	ML	B
Hickl et al.	80,0	X	X	X	X		X	X	X
Tatu et al.	72,3	X				X			X
Iftene	69,1	X		X					X
Adams	67,0	X	X				X	X	
DFKI	**66,9**			**X**				**X**	

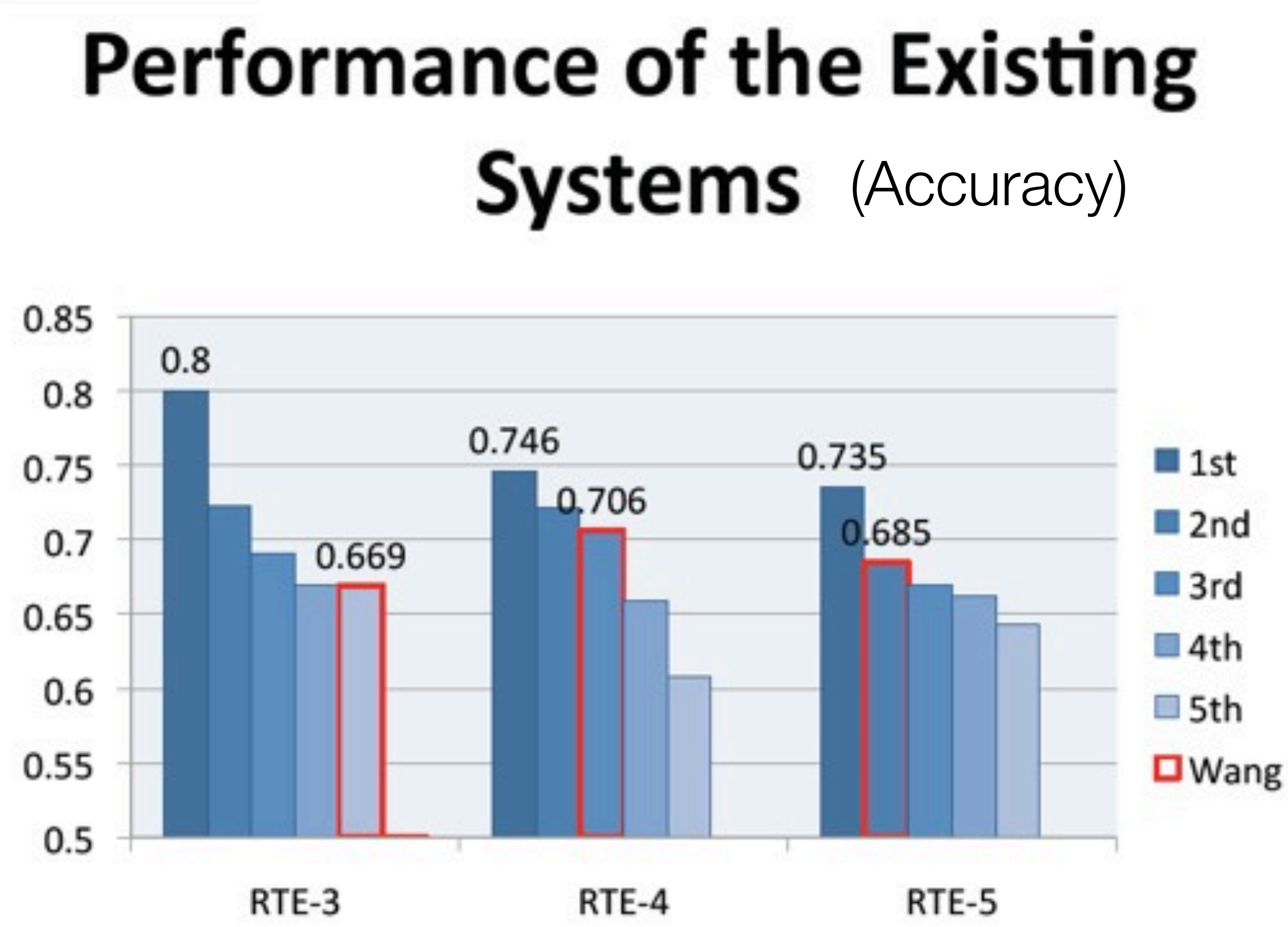
* Notation von (*Giampiccolo et al., 2007*):

Lx: Lexical Relation DB; Ng: N-Gram / Subsequence overlap; Sy: Syntactic Matching / Alignment; Se: Semantic Role Labeling; LI: Logical Inference; C: Corpus/Web; ML: ML Classification; B: Entailment corpora/Background Knowledge;

RTE-3-5 DFKI Voting-based Approach

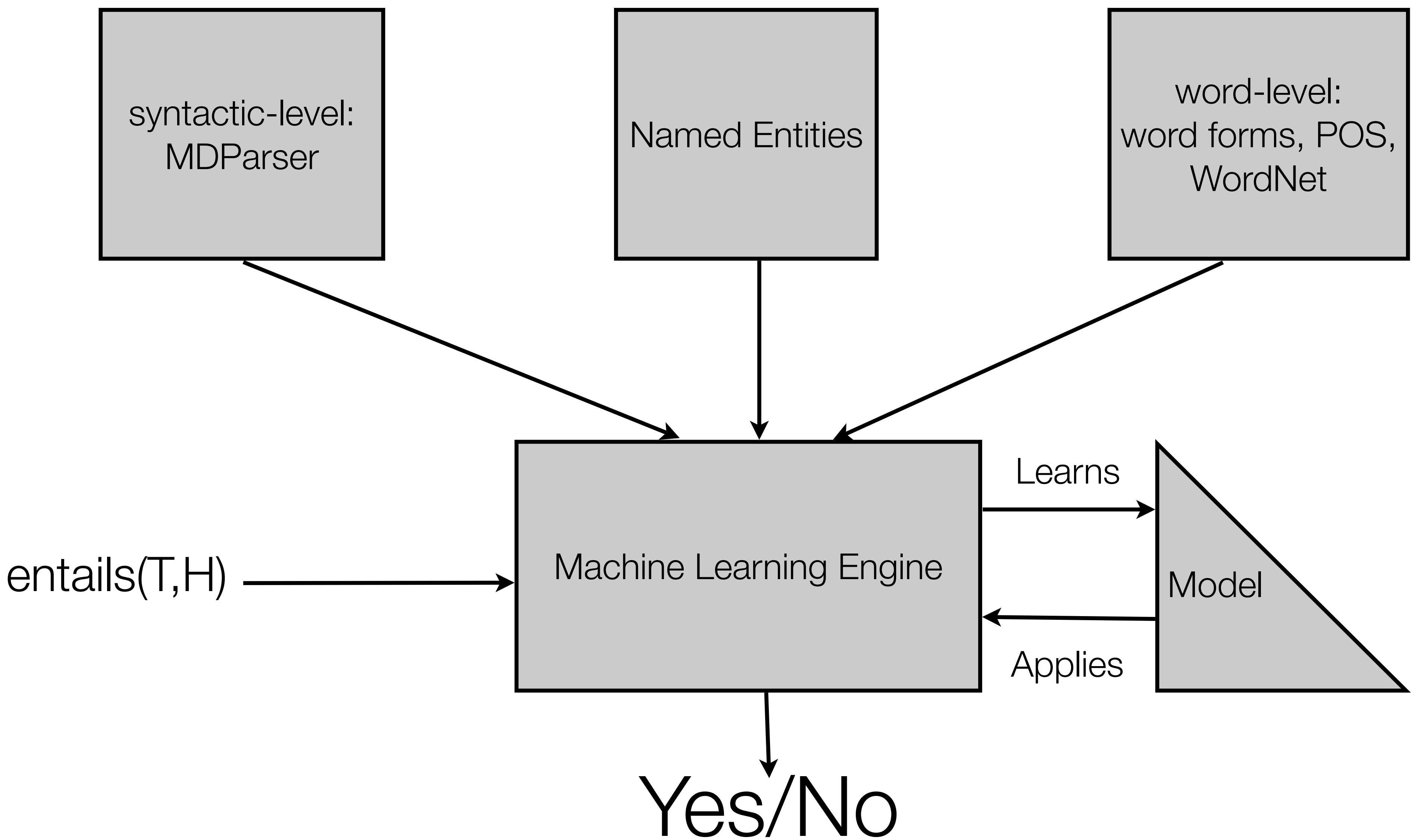
- Specialized RTE-engines which are integrated via a voting mechanism, cf. Wang & Neumann, AAAI, 2007; PhD Rui Wang, 2011





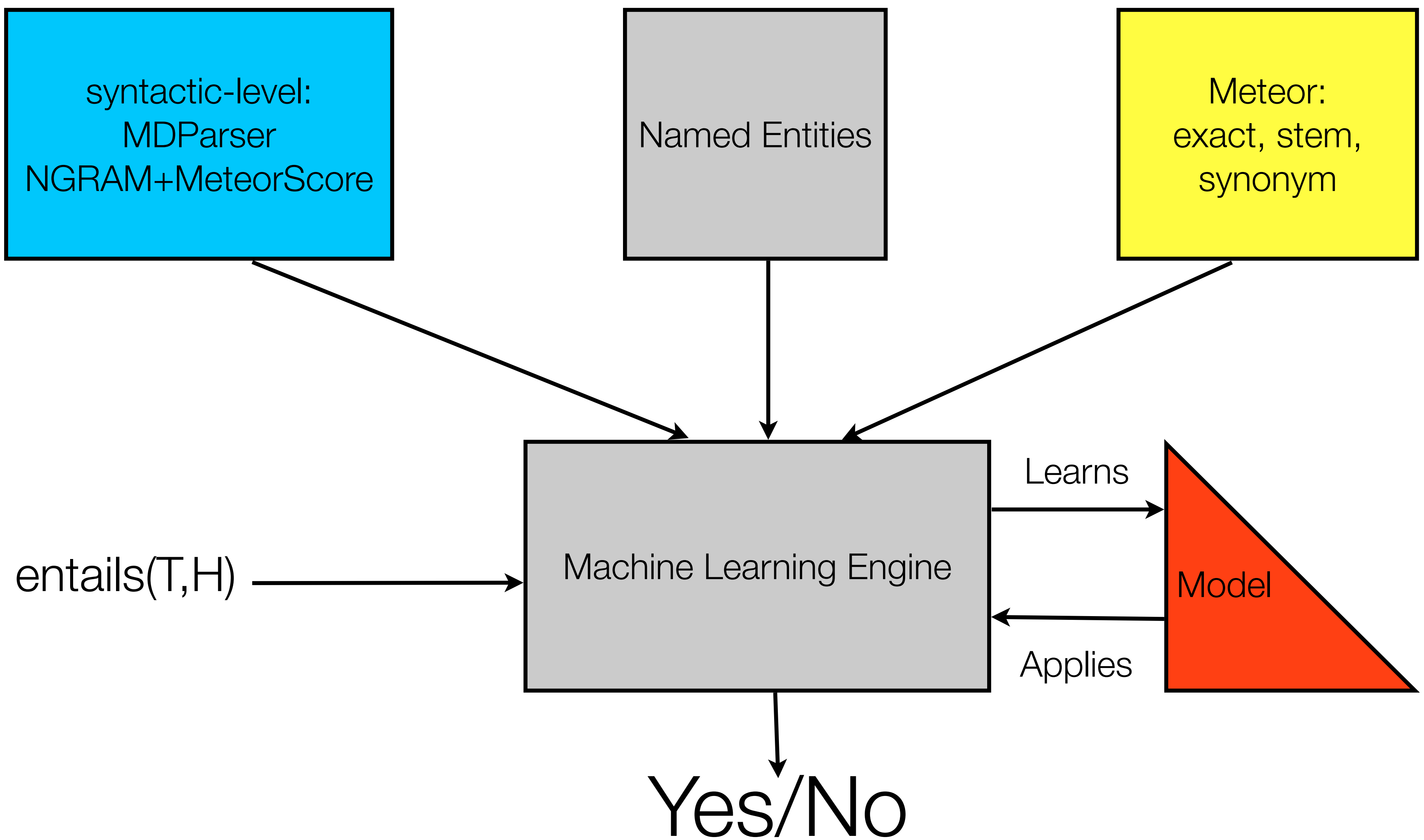
RTE-6: DFKI Machine Learning based Approach

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



RTE-7: DFKI LITE - Linear Machine Learning for Textual Entailment

- 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)



RTE-7 results

RUN	Micro-Average			Macro-Average (by TOPIC)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
BIU1	38.97	47.4	42.77	41.3	48.2	44.48
BIU2	41.81	44.11	**42.93**	43.16	45.12	44.12
BIU3	39.26	45.95	42.34	41.00	47.07	43.83
BUPTTeam1	45.02	44.95	**44.99**	47.53	46.41	46.96
BUPTTeam2	48.93	40.37	44.24	52.22	41.88	46.48
BUPTTeam3	51.99	36.93	43.18	56.21	38.63	45.79
CELI1	41.88	46.56	**44.10**	46.63	47.65	47.14
DFKI1	49.4	37.54	42.66	53.98	38.85	45.19
DFKI2	50.77	37.92	**43.41**	56.03	39.5	46.34
DFKI3	53.07	36.31	43.12	58.85	37.63	45.9
FBK_irst1	52.43	32.19	39.89	56.42	33.55	42.08
FBK_irst2	52.33	31.73	39.50	55.46	32.96	41.35
FBK_irst3	46.59	38.07	**41.90**	51.07	39.86	44.78
ICL1	47.88	21.56	**29.73**	49.23	24.59	32.8
IKOMA1	46.96	49.08	**48.00**	48.94	50.22	49.58
IKOMA2	58.48	30.05	39.70	58.87	31.95	41.42
IKOMA3	46.51	49.46	47.94	48.37	50.53	49.43
JU_CSE_TAC1	58.92	19.95	29.81	66.59	20.74	31.63
JU_CSE_TAC2	26.66	35.55	**30.47**	40.63	35.65	37.98
JU_CSE_TAC3	25.16	36.85	29.90	38.99	36.95	37.94
SINAI1	47.08	8.64	14.60	50.15	9.21	15.56
SINAI2	42.99	3.52	6.50	42.95	3.75	6.89
SINAI3	47.3	8.72	**14.72**	50.6	9.27	15.68
SJTU_CIT1	18.52	27.6	22.17	18.35	27.03	21.86
SJTU_CIT2	16.5	38.3	23.07	16.1	37.24	22.48
SJTU_CIT3	17.92	33.33	**23.31**	17.49	32.49	22.74
te_iitb1	20.67	60.24	**30.78**	25.06	63.11	35.87
u_tokyo1	46.49	43.58	44.99	48.44	45.24	46.78
u_tokyo2	47.55	42.35	44.80	48.75	43.61	46.04
u_tokyo3	46.84	43.58	**45.15**	48.63	45.24	46.87
UAIC20111	45.4	18.12	25.90	54.17	19.03	28.17
UAIC20112	30.21	25.84	**27.85**	35.18	27.51	30.88
UAIC20113	18.04	29.66	22.43	23.78	32.29	27.39

Table 1. Main Task results (in bold Best run of each system)

- 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

Summary

- Text inference is a hot topic
- New EU project Excitement will further boost text inference for real-world research and applications:
  - We will provide a open-source platform for RTE
- Web-scale RTE required
- New applications have to be considered ? -> what is the the RTE killer app?