Linguistic Inferencs

Outline

- Introduction
- RTE in PASCAL and TAC
- Techniques
- Systems
- Paraphrasing

Introduction

- **Textual Entailment Community**:
  - The *RTE Resource Pool* can now be accessed from: http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Resource_Pool
  - The *Textual Entailment Subzone* can now be accessed from: http://aclweb.org/aclwiki/index.php?title=Textual_Entailment_Portal
- **PASCAL Challenges**
  - **RTE-1 2005**
  - **RTE-2 2006**
  - **RTE-3 2007**
- **TAC** has been proposed as a generic task that captures major semantic inference needs across many natural language processing applications.
- **TAC challenges**
  - **RTE-4 TAC 2008**
  - **RTE-5 TAC 2009**
  - **RTE-6 TAC 2010**
  - **RTE-7 TAC 2011**

Readings

- **Workshops**
  - ACL 2005 Workshop on Empirical Modeling of Semantic Equivalence and Entailment, 2005
  - Pascal workshops 2005, 2006, 2007
  - TAC workshops since 2008
  - Answer Validation Exercise CLEF 2006, 2007
- **Surveys**
  - [Ghuge, Bhattacharya, 2013]

Readings

- **Thesis**
  - Oren Glickman (PHD, 2006)
  - Idan Szpecktor (MSC, 2005, PHD, 2009)
  - Milen Kouylekov (PHD, 2006)
  - Regina Barzilay (PHD, 2004)
  - Elena Cabrio (PHD, 2011)
  - Óscar Ferrández (PHD, 2009)
  - Prodromos Malakasiotis (PHD, 2011)
  - Annisa Ihsani (MSC, 2012)
  - Roy Bar Haim (PHD, 2010)
  - Shachar Mirkin (PHD, 2011)
  - Marta Vila (PHD, 2015)

- RTE is the task of deciding, given two text fragments, whether the meaning of one text is entailed (can be inferred) from another text. This task captures generically a broad range of inferences that are relevant for multiple applications.
- For example, a QA system has to identify texts that entail the expected answer. Given the question "Who killed Kennedy?", the text "the assassination of Kennedy by Oswald" entails the expected answer form "Oswald killed Kennedy".

.

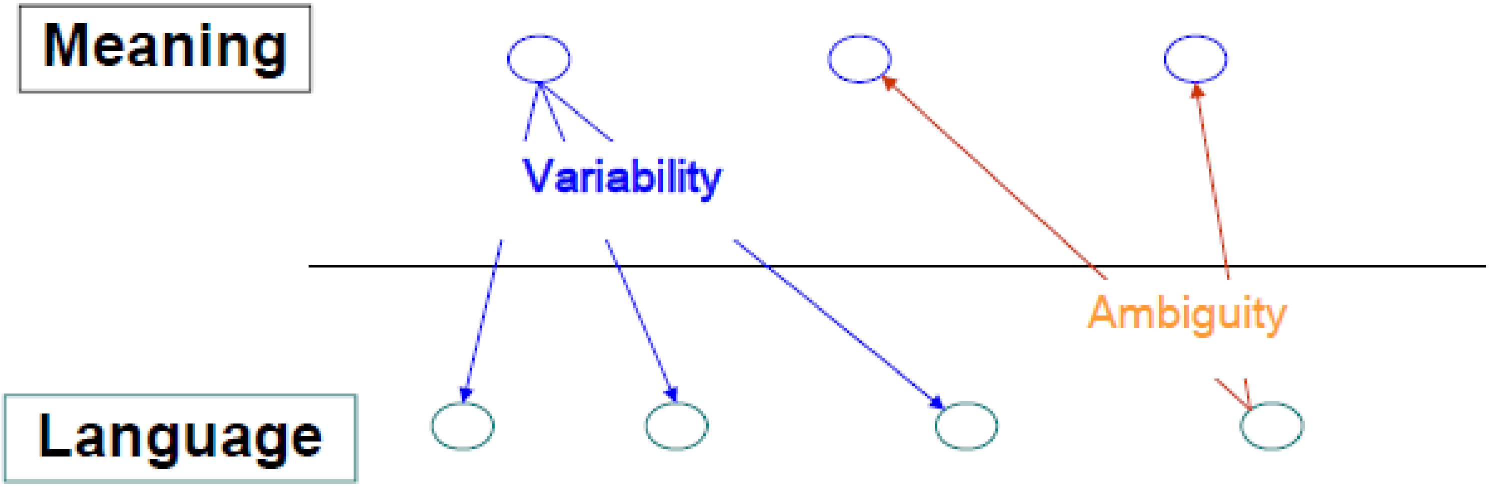
Why

- Limitations of NLP systems based only on shallow processing.
- Need of semantic processing for some tasks
- Need of World Knowledge, Common Sense Knowledge
- Acquisition of this knowledge.

Why

Motivation:

- Text applications require *semantic* inference
- A common framework for applied semantics is needed, but still missing
- Textual entailment may provide such framework



Linguistic Inference Applications

- Question-Answering
- Information Extraction
- Information Retrieval
- Multi-Document Summarization
- Named Entity Recognition
- Temporal and Spatial Normalization
- Semantic Parsing
- Natural Language Generation

Linguistic Inference

- Equivalence (**Paraphrase**): *expr1* $\Leftrightarrow$ *expr2*
- **Entailment:** *expr1* $\Rightarrow$ *expr2* – more general

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

t **entails** *h* ***(t***$\Rightarrow$***h)*** if, typically, a **human** reading *t* would infer that *h* is **most likely** true"

Linguistic Inference

The Dow Jones Industrial Average closed up 255

Dow ends up

Dow climbs 255

Dow gains 255 points

Stock market hits a record high

Linguistic Inference examples

TEXT	HYPOTHESIS	ENTAILMENT
Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc last year.	*Yahoo bought Overture.*	**TRUE**
Microsoft's rival Sun Microsystems Inc. bought Star Office last month and plans to boost its development as a Web-based device running over the Net on personal computers and Internet appliances.	*Microsoft bought Star Office.*	**FALSE**
The National Institute for Psychobiology in Israel was established in May 1971 as the Israel Center for Psychobiology by Prof. Joel.	*Israel was established in May 1971.*	**FALSE**

Methods and Approaches

- Word overlap
  - lexical, syntactic, and semantic
- Logical approaches
  - Raina et al, 2005
  - Bos et al, 2005, 2006
  - Moldovan et al, 2003
- Graph matching approaches
  - Haghighi et al, 2005
  - de Salvo et al, 2005
  - de Marneffe et al, 2005, 2006
- Paraphrases and Entailment Rules
  - Moldovan and Rus, 2001
  - Lin and Pantel, 2001 QA
  - Shinyama et al, 2002 IE

Methods and Approaches

- Probabilistic interpretation:

> *t* **probabilistically entails** *h* ($\boldsymbol{t} \Rightarrow \boldsymbol{h}$) 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

Methods and Approaches

- The role of knowledge:
  - For textual entailment to hold we require:

    $\boxed{\textit{text} \text{ AND } \textit{knowledge} \Rightarrow h}$

  - 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)
- The knowledge sources available to the system are the most significant component of supporting TE

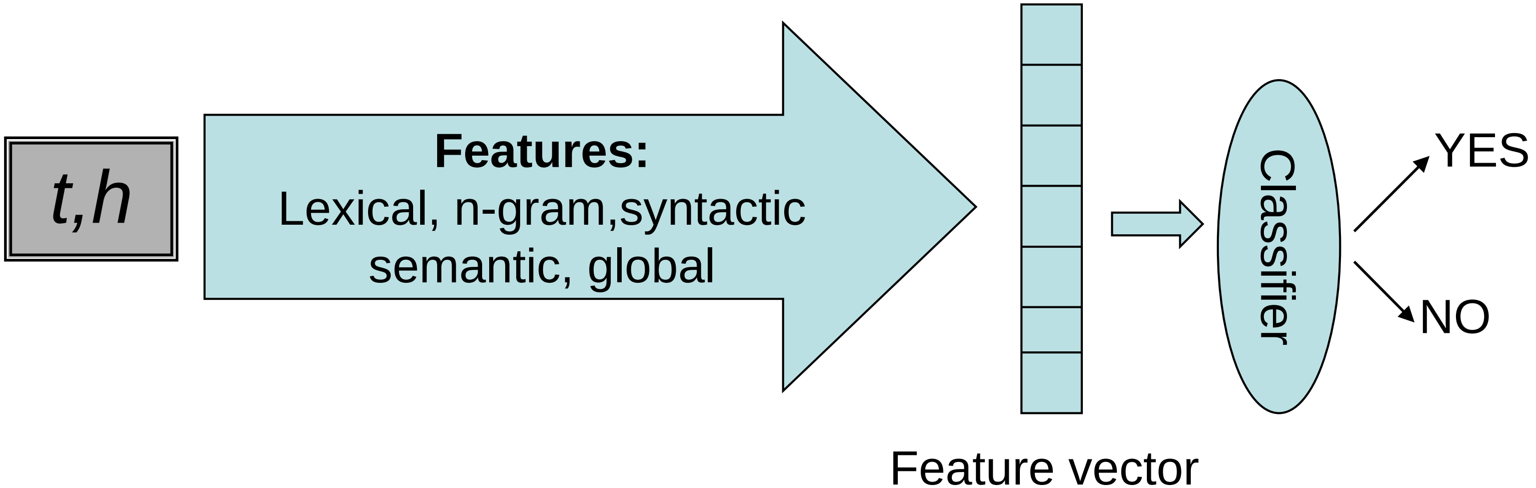
Methods and Approaches

- Measure similarity between *t* and *h* (*coverage* of *h* by *t)*:
  - Lexical overlap (unigram, N-gram, subsequence)
  - Average Matched Word Displacement
  - Lexical substitution (WordNet, statistical)
  - Syntactic matching/transformations
  - Lexical-syntactic variations (“paraphrases”)
  - Semantic role labeling and matching
  - Global similarity parameters (e.g. negation, modality)
- Sentence Alignment
  - Exhaustive Sentence Alignment
    - *parallel corpora*
    - *comparable corpora*
  - Web-based Sentence Alignment
  - Bigrams
    - Syncronous grammars
    - Inversion Transduction grammars
- Cross-pair similarity
- Detect mismatch (for non-entailment)
- Logical inference

Methods and Approaches

- Thesaurus-based Term Expansion
  - WN
- Distributional Similarity
- BLEU (BiLingual Evaluation Understudy)
- ROUGE (Recall-Oriented Understudy for Gisting Evaluation)
- classical statistical machine translation model
  - giza++ software (Och and Ney, 2003)

Dominant approach: Supervised Learning



- Features model both similarity and mismatch
- Train on development set and auxiliary *t-h* corpora

General View

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

Overture is a search company

Google is a search company

……….

Phrasal verb paraphrasing

Entity matching

Alignment

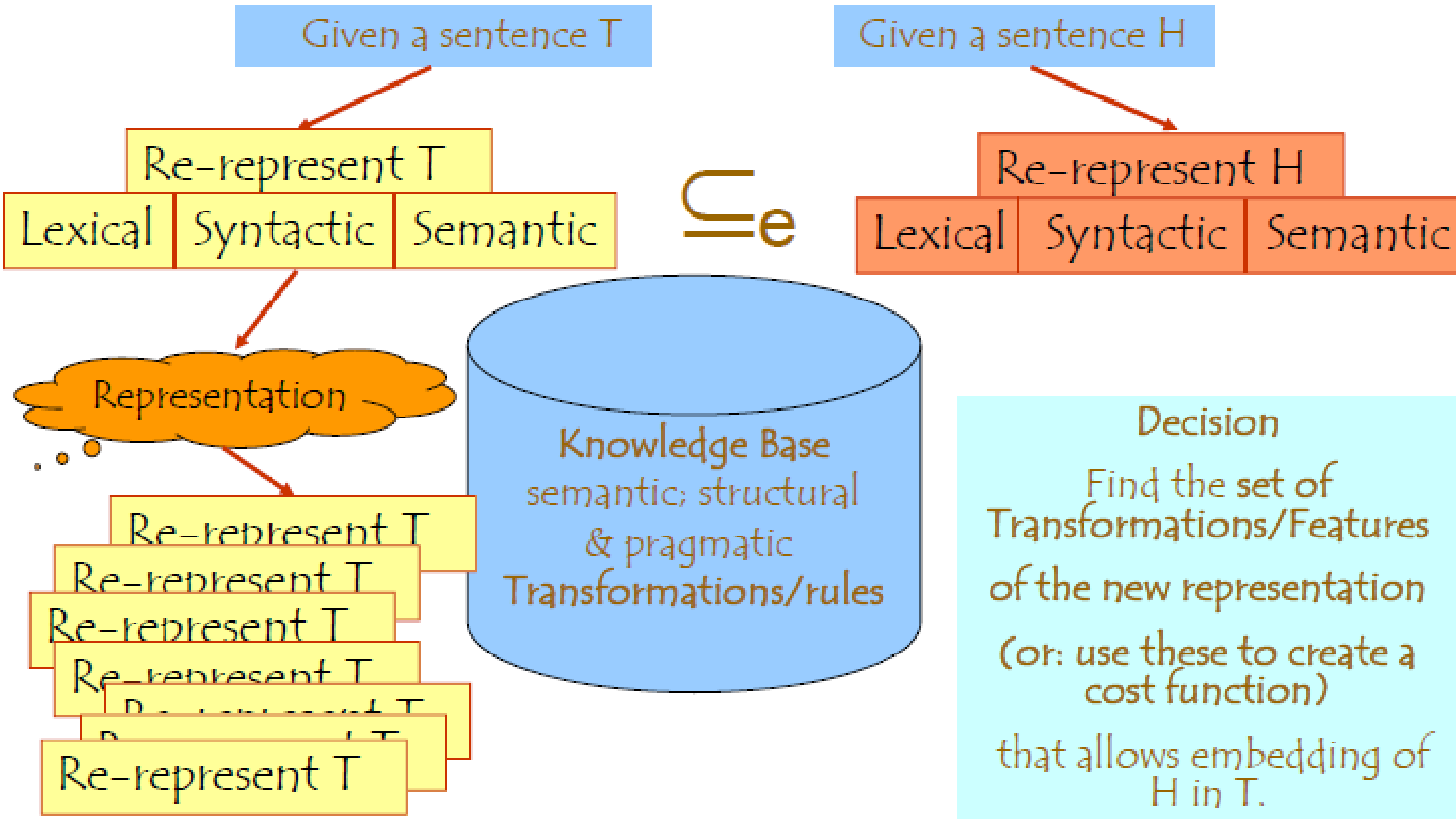
How?

Semantic Role Labeling

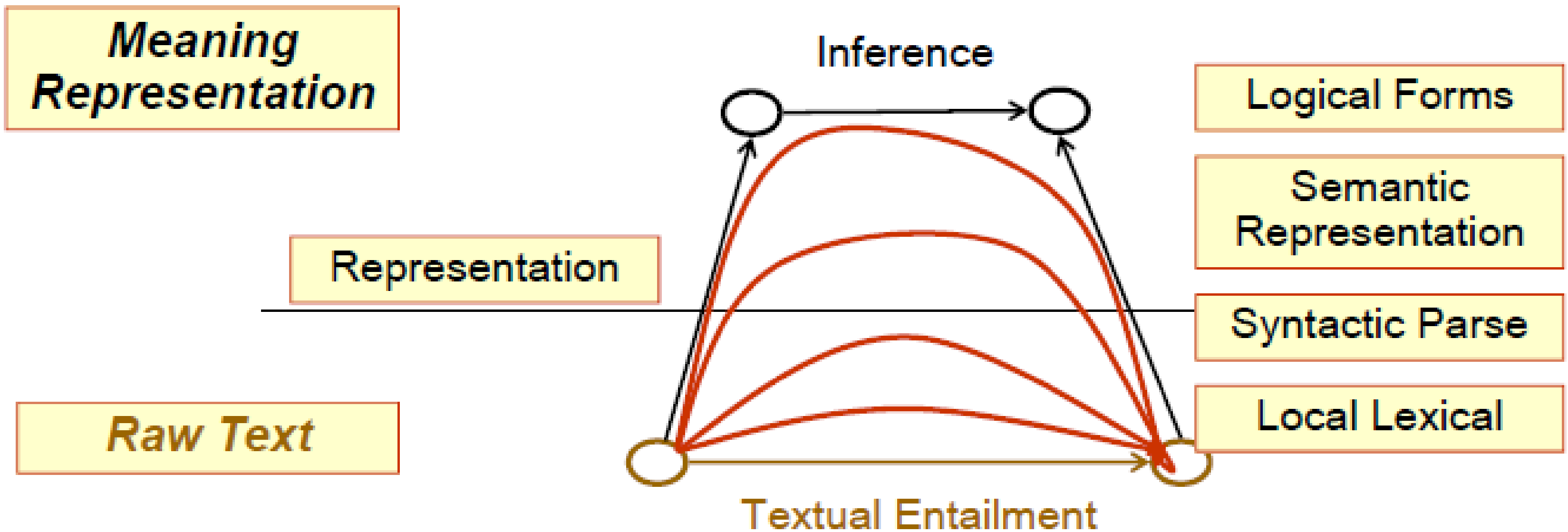
Integration

Textual Entailment 30

A general Strategy for Textual Entailment



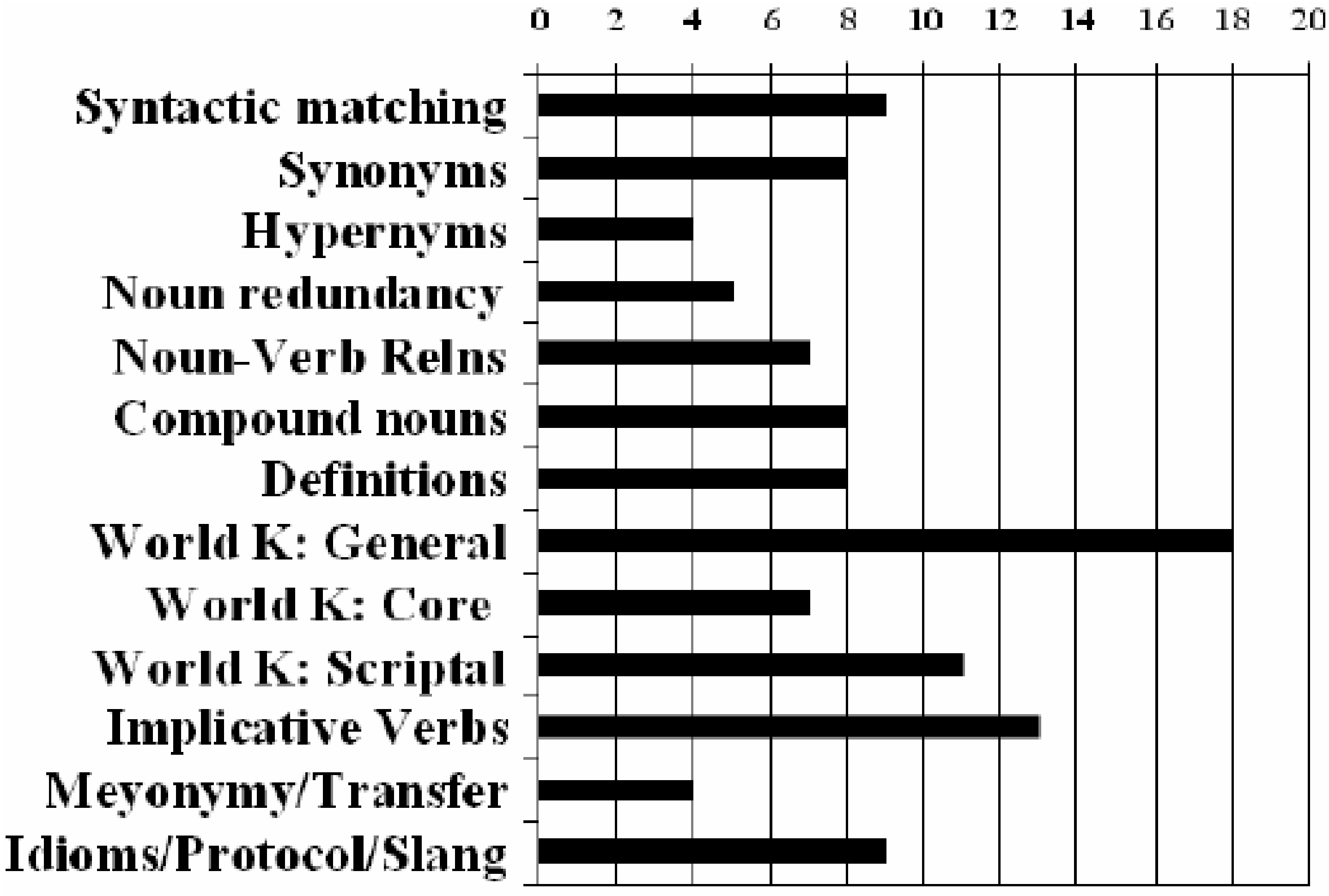
Representation



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

PASCAL RTE-3

Frequency



PASCAL RTE-3

- Resources
  - WordNet
  - Extended WordNet
  - WordNet3.0
  - TimeML
  - IKRIS
  - DIRT paraphrase database
  - FrameNet
  - VerbNet
  - VerbOcean
  - Component Library (U. Texas)
  - OpenCyC
  - SUMO
  - Tuple Database (Boeing)
  - Stanford's additions to Wn

Notable Systems

- TEASE & improvements
- Glickman
- DFKI
- COGEX, Groundhog, Hickl at LCC
- Stanford
- Tor Vergata
- TALP UPC
- Nutcracker

Notable Systems

TEASE and improvements

- Group of Ido Dagan at Bar Ilan University (Israel)
- Idan Szpektor (2005) Scaling Web-based Acquisition of Entailment Relations (Ms. thesis)
- Idan Szpektor et al (2004)
- Idan Szpektor and Ido Dagan (2007)
- Lorenza Romano et al (2007)
- Ido Dagan et al (2008)

TEASE

Input Verb	Learned Templates	
X approach *Y*	*X* go to *Y* *X* step to *Y* *X* walk to *Y*	*X* near *Y* *X* stride to *Y* *X* pass a note to *Y*
X defeat *Y*	*Y* lose to *X* *X* beat *Y* *X* victory over *Y*	*X* destroy *Y* *X* win *Y* *X* conquer *Y*
X host *Y*	bring *Y* to *X* *Y* is held in *X* *Y* come to *X*	*Y* is played in *X* *X* venue of *Y* *X* play host to *Y*
X preclude *Y*	*X* prevent *Y* *X* bar *Y* *X* prohibit *Y*	*X* exclude *Y* *X* deny *Y* *X* forbid *Y*
X plant *Y*	*X* grow *Y* *X* produce *Y* *X* cultivate *Y*	*X* raise *Y* *X* sow *Y* *X* farm *Y*

TEASE

- A *template*, T_i, is a connected parse graph fragment (or dependency parse-tree fragment) with optional variables at some nodes.
  - example

$$X \overset{subj}{\leftarrow} prevent \overset{obj}{\rightarrow} Y$$

- A *pair of templates* T_1 and T_2 is denoted as $<T_1, T_2>$.
- A pair of templates is called an *entailment relation* if T_1 and T_2 contain the same variables and the meaning of T_2 can be inferred from the meaning of T_1, or vice versa, in some contexts, under the same variable instantiation.

TEASE

- example $X \xleftarrow{subj} prevent \xrightarrow{obj} Y$ entails $X \xleftarrow{subj} reduce \xrightarrow{obj} Y\ risk$

  "*aspirin reduces heart attacks risk*" can be inferred from "*aspirin prevents heart attacks*"

- An entailment relation does not need to hold under all possible variable instantiations, i.e. the correctness/validity of an entailment relation depends on specific variable instantiations
- A *pivot* P is a lexical phrase, such as a verb, a phrasal verb or a noun phrase (typically a nominalization) that expresses a semantic relation.

  Ex: aquire, fall to, prevent, victory over, near, ...
- A *pivot template*, denoted $\{P, T_P\}$, is a pivot P with its syntactic template T_P in the form of a parse graph fragment including at least two variable slots (nodes).

TEASE

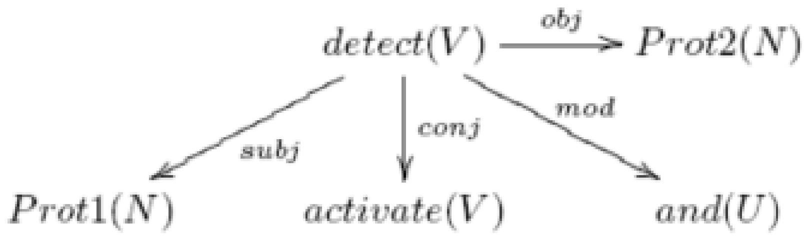


Figure 1: The dependency parse graph of the sentence "*Prot1 detected and activated Prot2*".

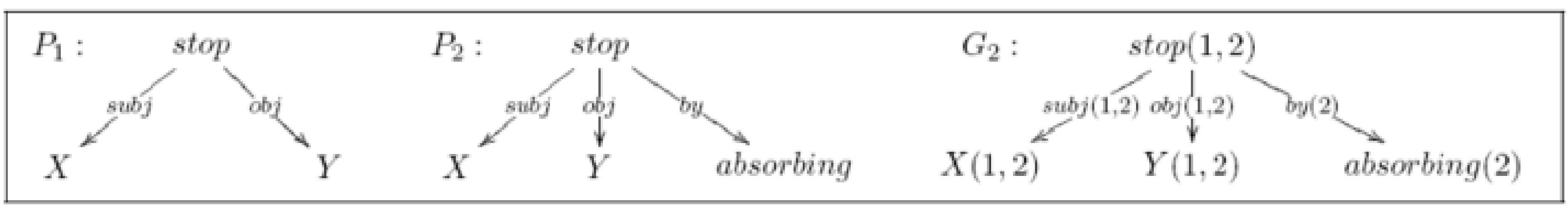


Figure 2: Two parse trees and their compact representation (sentence sets are shown in parentheses).

TEASE

First step: create a complete template T_P for the input pivot P

- Variable slots are added for the major types of syntactic relations that interact with P, based on its syntactic type

- input template for a transitive verb V $$X \xleftarrow{subj} V \xrightarrow{obj} Y$$

- phrasal verb consisting of a verb V and a preposition

$$X \xleftarrow{subj} V \xrightarrow{prep} Preposition \xrightarrow{\text{mod}} Y$$

- The output of the TEASE algorithm is a ranked list of templates $\{T_j\}$ $<T_P, T_j>$ is an entailment relation candidate

TEASE

- An *anchor-set* is a set of words (or terms). Each member of an anchor-set is termed an *anchor*
- An anchor in an anchor-set that is designated to be an instantiation of a template variable in a sentence is termed a *slot anchor*
- An anchor in an anchor-set that is not designated to be an instantiation of a template variable in a sentence is termed a *context anchor*.
- A *matching* of a template *T* in a sentence *s* is the embedding of the parse-graph of *T* as a sub-graph in the parse-graph of *s*
- An *instantiation* of a template *T* by an anchor-set *AS* in a sentence *s* is a matching of *T* in *s* where each variable of *T* is instantiated with the corresponding value of a slot anchor in *AS*, and all the values of the context anchors in *AS* appear elsewhere in *s*

"Antibiotics in pregnancy prevent miscarriage"

Slot anchors: $\{antibiotics \overset{subj}{\longleftarrow}, miscarriage \overset{obj}{\longleftarrow}\}$

Context anchor: *pregnancy*

$$X \overset{subj}{\leftarrow} prevent \overset{obj}{\rightarrow} Y$$

TEASE

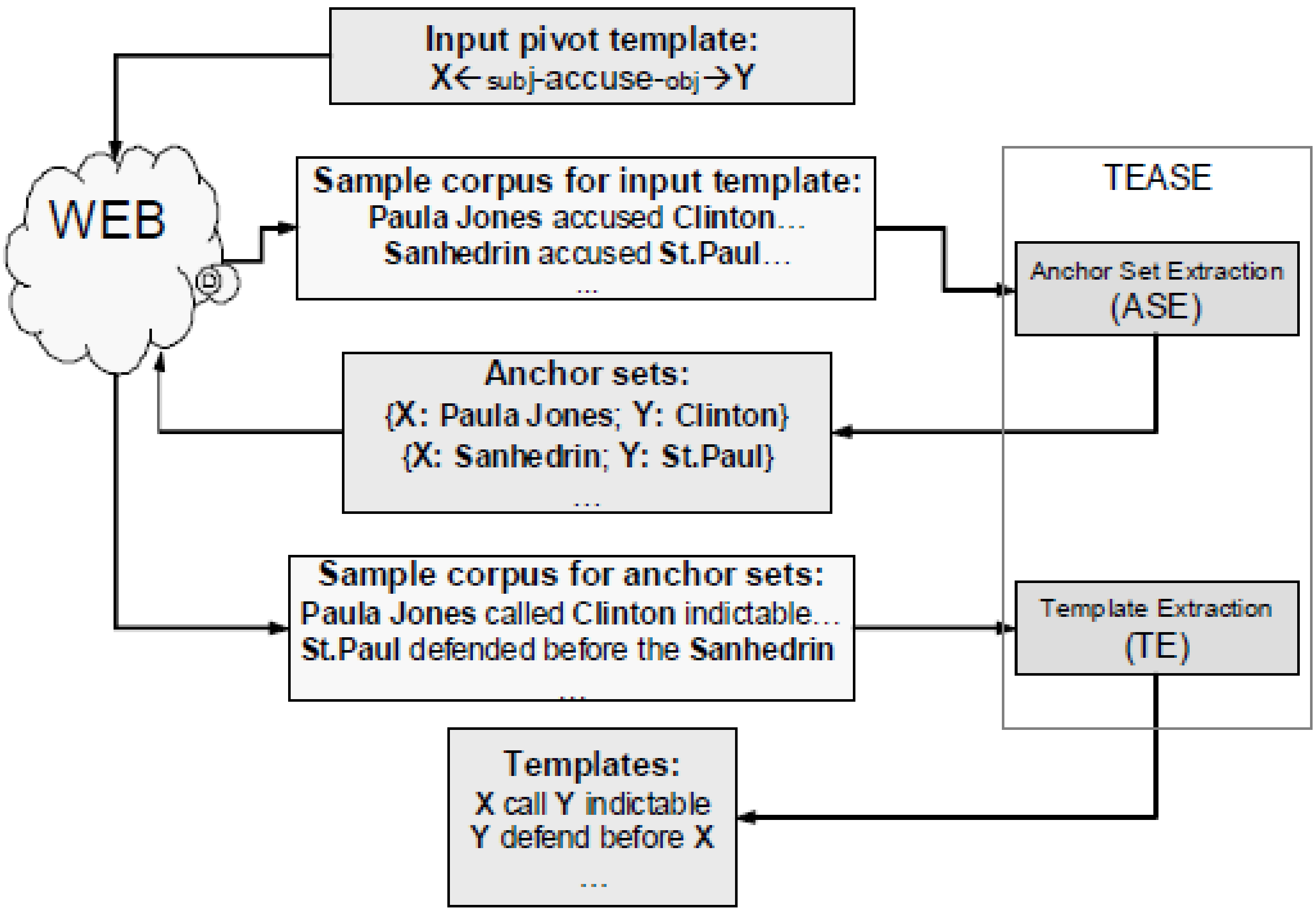
A good *AS* should satisfy a proper balance between specificity and generality

- A *minimal anchor-set* is an anchor-set that contains only slot anchors
- *AS* is termed a *characteristic anchor-set* of a template *T* if for every template T_i that is instantiated by *AS* in some sentence, the entailment relation $<T, T_i>$ holds between *T* and T_i
- A *diverse anchor-set* is an anchor-set that instantiates more than one template in sentences
- An anchor-set that is both characteristic of a template *T* and diverse is termed a *productive anchor-set* for a template *T*
- Contex anchors provide for specificity

TEASE

- Algorithm
  - For each input pivot template T_P:
    - Extract productive anchor-sets for the pivot template (ASE phase)
      - Construct a sample corpus for the pivot template by retrieving sentences containing the pivot template from the Web.
      - Extract candidate anchor-sets from the sentences in the sample corpus.
      - Filter out candidate anchor-sets that fail certain criteria
    - Extract templates (TE phase)
      - Construct a sample corpus by retrieving sentences containing the anchor-sets extracted in phase 1 from the Web.
      - Extract repeated sub-structures in the sample corpus to be template candidates T_i.
    - Rank each extracted template T_i according to the confidence level in the correctness of the entailment relation $<T_P, T_i>$.

TEASE


Input pivot template:
X← subj-accuse-obj →Y
WEB
Sample corpus for input template:
Paula Jones accused Clinton…
Sanhedrin accused St.Paul…
…
TEASE
Anchor Set Extraction
(ASE)
Anchor sets:
{X: Paula Jones; Y: Clinton}
{X: Sanhedrin; Y: St.Paul}
…
Sample corpus for anchor sets:
Paula Jones called Clinton indictable…
St.Paul defended before the Sanhedrin
…
Template Extraction
(TE)
Templates:
X call Y indictable
Y defend before X
…

TEASE

- ASE Algorithm
  - For each input pivot template T_P:
    - Construct a sample corpus that consists of sentences containing T_P.
      - Retrieve sentences the Web using a query containing the template's words
      - Retrieve more sentences from the Web using refined queries, based on the sentences retrieved at step
    - Extract productive anchor-set candidates from the constructed corpus.
      - Extract one minimal anchor-set, containing only the slot anchors, from each sentence in the sample corpus.
      - Extract one more anchor-set from each sentence, containing one context anchor in addition to the slot anchors, if possible.
    - Filter out candidates that fail certain filtering criteria:
      - Applying thresholds over individual anchor-set statistics.
      - Filtering anchor-sets that are redundant or inconsistent relative to other anchor-sets.

TEASE – ASE phase

Pivot Template	Learned Anchor Sets
X establish Y	X = epa, Y = national emission standard, C_1 = asbestos X = canada agricultural products act, Y = review tribunal X = school district, Y = breakfast program X = federal government, Y = conservation corps X = erisa, Y = minimum standards X = constantine, Y = new rome
X write Y	X = laurie, Y = numerous songs X = lewis carrol, Y = alice's adventures X = plato, Y = detailed account, C_1 = atlantis X = mendelssohn, Y = incidental music X = shakespeare, Y = great tragedies X = thomas malthus, Y = essay
X calculate Y	X = katz equation, Y = membrane potential X = eratosthenes, Y = circumference X = nernst equation, Y = equilibrium potential X = language model, Y = probabilities X = following table, Y = annual cost X = acos, Y = arc cosine

TEASE – TE phase

Pivot Template	Learned Templates	
X establish *Y*	*X* set *Y* *X* develop *Y* *X* create *Y* *X* found *Y* *X* enforce *Y* *X* form *Y* *X* offer *Y* *X* release *Y*	*X* promulgate *Y* *X* issue *Y* *X* implement *Y* *X* provide *Y* *X* make *Y* *X* launch *Y* *X* institute *Y* *X* for the establishment of *Y*
X write *Y*	*X* who write *Y* *X* publish *Y* *X* compose *Y* read *Y* by *X* *Y* attributed to *X* perform *Y* by *X* *X* writer of *Y* selected *Y* of *X*	*X* produce *Y* *X* pen *Y* *X* create *Y* *X*'s *Y* *X* complete *Y* *X* book of *Y* *X* say in *Y* *X* work include *Y*
X calculate *Y*	*X* determine *Y* *X* compute *Y* *X* give estimate of *Y* *X* return *Y* *X* assess *Y* *X* generate *Y* *X* recalculate *Y* *X* work out *Y*	*X* measure *Y* *X* calculation of *Y* *X* yield *Y* *X* get *Y* *X* produce *Y* *Y* according to *X* *Y* obtained from *X* *X* evaluate *Y*

TEASE

Pivot template	Learned template	Reason
$X \xleftarrow{subj} \text{write} \xrightarrow{obj} Y$	$X \xleftarrow{subj} \text{present} \xrightarrow{obj} Y$	Presenting a cheque is synonymous to writing a cheque.
$X \xleftarrow{subj} \text{write} \xrightarrow{obj} Y$	$X \xleftarrow{subj} \text{issue} \xrightarrow{obj} Y$	Issuing a report (e.g. by a policeman) is synonymous to writing a report.
$X \xleftarrow{subj} \text{produce} \xrightarrow{obj} Y$	$X \xleftarrow{subj} \text{obtain} \xrightarrow{obj} Y$	A telescope producing images is synonymous to a telescope obtaining images.

Glickman

- Acquiring Lexical Entailment Relations
  - identify lexical paraphrases of verbs

subject	secretary_general_boutros_boutros_ghali	subject	iraqi_force
object	implementation_of_deal	object	kurdish_rebel
modifier	after	pp-on	august_31
(A) verb: delay		(B) verb: attack	

Figure 6.2: Extracted verb instances for sentence "But U.N. Secretary-General Boutros Boutros-Ghali delayed implementation of the deal after Iraqi forces attacked Kurdish rebels on August 31."

DFKI

- Rui Wang, Günter Neumann
- Based on:
  - H is usually textually shorter than T
  - not all information in **T** is relevant to make decisions for the entailment
  - the dissimilarity of relations among the same topics between T and H are of great importance.
- Process
  - starting from H to T
  - exclude irrelevante information from T
  - represent the structural differences between T and H by means of a set of Closed-Class Symbols (Entailment Patterns – EPs)
  - classification using subsequence kernels

COGEX

COGEX, Tatu 2006, 2007

- Combination of LEX, COGEXd, COGEXc
- EXtended WordNet Knowledge Base (XWN-KB)
  - XWN Lexical Chains
  - coarse-grained sense inventory for WordNet 2.1 released for Task #7 in SemEval-2007. This clustering was created automatically with the aid of a methodology described in (Navigli, 2006).
- NLP Axioms
  - links a NE to its set of aliases
- Named Entity Check
  - deducts points for each pair whose H contains at least one *named entity* not-derivable from T

Groundhog

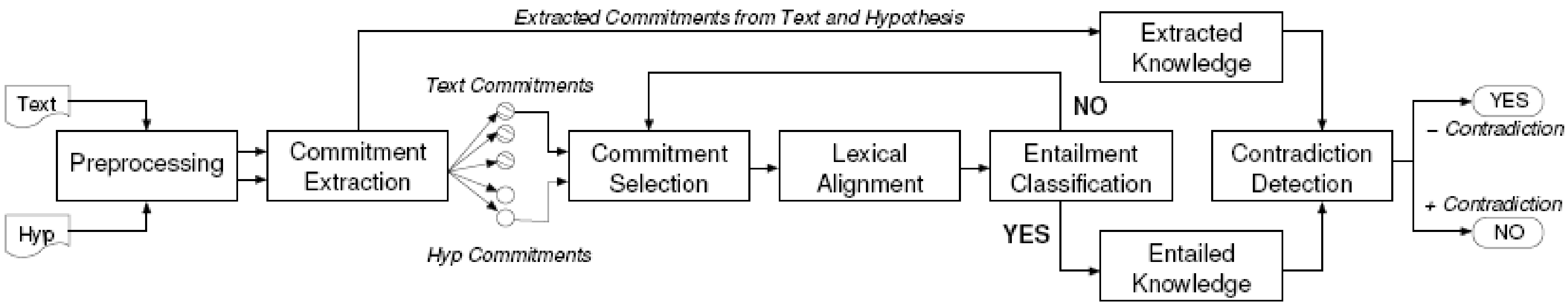
LCC Groundhog, Hickl, 2006

- Lexical Alignment
  - Maximum Entropy classifier to compute the probability that an element selected from a text corresponds to – or can be **aligned with** – an element selected from a hypothesis.
  - Three-step Process:
    - **First,** sentences were decomposed into a set of “alignable chunks” that were derived from the output of a chunk parser and a collocation detection system.
    - **Next,** chunks from the text (C_t) and hypothesis (C_h) were assembled into an alignment matrix ($C_t \times C_h$).
    - **Finally,** each pair of chunks were then submitted to a classifier which output the probability that the pair represented a positive example of alignment.

LCC Hickl

LCC, Hickl, 2007

- Discourse Commitment-based Framework



LCC Hickl

Text: A Revenue Cutter, the ship was named for Harriet Lane, niece of President James Buchanan, who served as Buchanan's White House hostess.

T1. A Revenue Cutter is a ship.
T2. The ship was named for Harriet Lane.
T3. Harriet Lane was the niece of President James Buchanan.
T4. The niece of Buchanan served as Buchanan's White House hostess.
T5. A Revenue Cutter was named for Harriet Lane.
T6. A Revenue Cutter was named for the niece of President James Buchanan.
T7. A Revenue Cutter was named for Buchanan's White House hostess.
T8. A Revenue Cutter was named for a White House hostess.
T9. A Revenue Cutter was named for a hostess.
T10. The niece of a President served as Buchanan's White House hostess.
T11. The niece of a President served as Buchanan's hostess.
T12. The niece of a President served as a White House hostess.
T13. The niece of a President served at the White House.
T14. The niece of a President had occupation hostess.
T15. The niece of a President served as a hostess.

Selected Commitment

Hyp(34): Harriet Lane owned a Revenue Cutter.

Negative Instance of Textual Entailment

T16. Harriet Lane was related to President James Buchanan.
T17. Harriet Lane was the niece of a President.
T18. Harriet Lane was related to a President.
T19. Harriet Lane was related to James Buchanan.
T20. James Buchanan had title of President.
T21. James Buchanan had a White House hostess.
T22. James Buchanan had a hostess.
T23. James Buchanan was associated with the White House.
T24. James Buchanan had a niece.
T25. Harriet Lane served as Buchanan's White House hostess.
T26. Harriet Lane served as Buchanan's hostess.
T27. Harriet Lane served as a White House hostess.
T28. Harriet Lane served at the White House.
T29. Harriet Lane had occupation hostess.
T30. Harriet Lane served as a hostess..

Selected Commitment

Hyp(36): Harriet Lane worked at the White House.

Positive Instance of Textual Entailment

LCC Hickl

- Some of these commitments can be easily deduced from the text:
  - Harriet Lane is niece of James Buchanan
  - James Buchanan (is/was) president
  - A Revenue Cutter is a ship
  - …
- **Other commitments do not occur explicitly in the text and have to be extracted as World Knowledge**:
  - James Buchanan (is/was) a president of USA
  - USA presidents live at the White House
  - The White House is placed in Washington
  - Hostess is a profession
  - …

Stanford

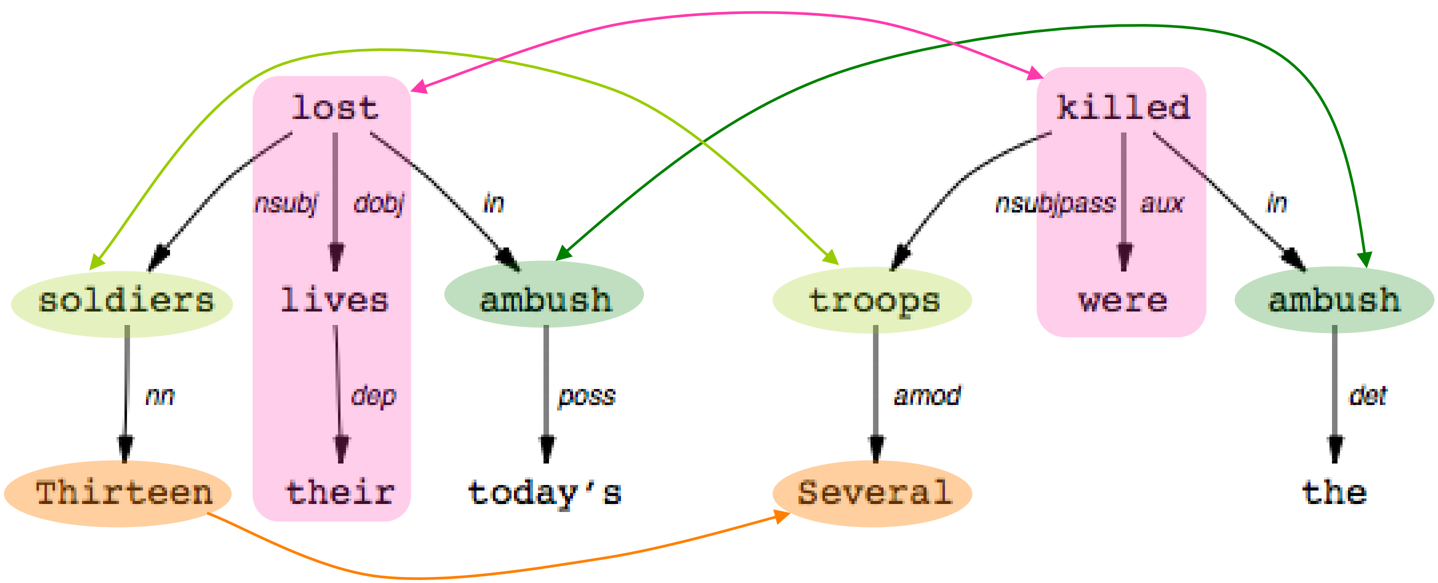
De Marneffe et al, 2005, 2006

Chambers et al, 2007

- Graph matching
  - Represent sentences as typed dependency trees
  - Find low-cost alignment (using lexical & structural match costs)

T: *Thirteen soldiers lost their lives in today's ambush*

H: *Several troops were killed in the ambush*

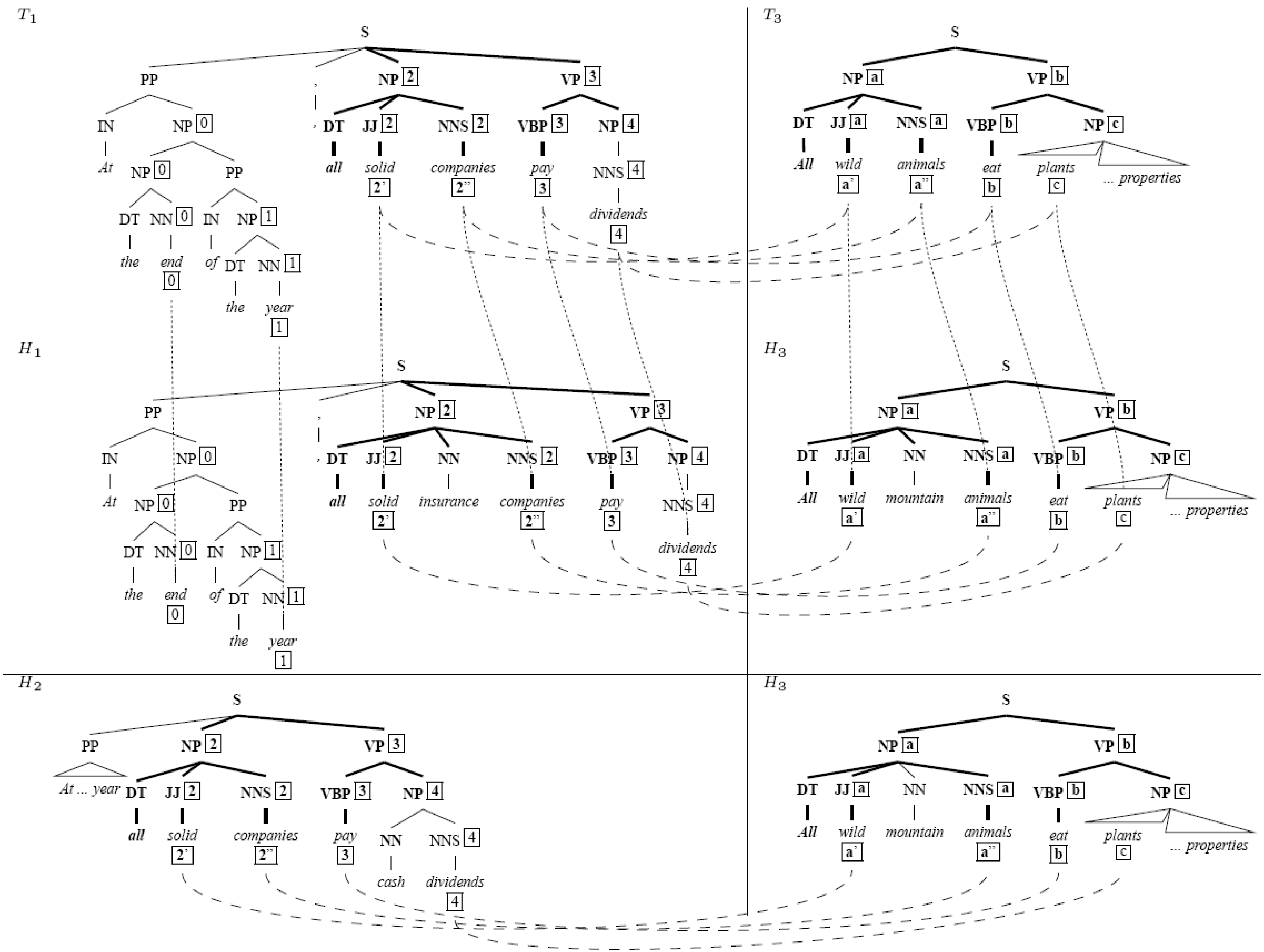


Tor Vergata

- Zanzotto et al, 2006, 2007
- *cross-pair similarity*
  - similarity measure aiming at capturing *rewrite rules* from training examples, computing a *cross-pair similarity* $K_S((T',H'), (T'',H''))$.
  - if two pairs are similar, it is extremely likely that they have the same entailment value. The key point is the use of *placeholders* to mark the relations between the sentence words. A *placeholder* co-indexes two substructures in the parse trees of text and hypothesis
  - a tree similarity measure $K_T(\tau_1, \tau_2)$ (Collins and Duffy, 2002) that counts the subtrees that τ_1 and τ_2 have in common
  - a substitution function $t(\cdot, c)$ that changes names of the placeholders in a tree according to a set of correspondences between placeholders c

Zanzotto, Moschitti

- Zanzotto, Moschitti, 2006
  - textual entailment pairs as pairs of syntactic trees with co-indexed nodes
  - consider both the structural similarity between syntactic tree pairs and the similarity between relations among sentences within a pair
  - similarities
    - cross-pair
      - K((T',H'), (T'',H''))
      - structural and lexical similarity between T', T'' and H', H''
      - intra-pair word movement compatibility between (T',H') and (T'',H'')
    - intra-pair
  - novel kernel function



T1
S
PP , NP 2 VP 3
IN NP 0 DT JJ 2 NNS 2 VBP 3 NP 4
At NP 0 PP all solid 2' companies 2" pay 3 NNS 4
DT NN 0 IN NP 1 dividends 4
the end 0 of DT NN 1
the year 1
T3
S
NP a VP b
DT JJ a NNS a VBP b NP c
All wild a' animals a" eat b plants c ... properties
H1
S
PP , NP 2 VP 3
IN NP 0 DT JJ 2 NN NNS 2 VBP 3 NP 4
At NP 0 PP all solid 2' insurance companies 2" pay 3 NNS 4
DT NN 0 IN NP 1 dividends 4
the end 0 of DT NN 1
the year 1
H3
S
NP a VP b
DT JJ a NN NNS a VBP b NP c
All wild a' mountain animals a" eat b plants c ... properties
H2
S
PP NP 2 VP 3
At ... year DT JJ 2 NNS 2 VBP 3 NP 4
all solid 2' companies 2" pay 3 NN NNS 4
cash dividends 4
H3
S
NP a VP b
DT JJ a NN NNS a VBP b NP c
All wild a' mountain animals a" eat b plants c ... properties

Zanzotto, Moschitti

- kernels
  - $K_l((T',H'), (T'',H''))$
    - based on the intra-pair lexical similarity siml(T,H) as defined in (Corley and Mihalcea, 2005).
    - siml(T',H') × siml(T'',H'').
  - K_l+K_s
    - combines our kernel with the lexical-similarity-based kernel
  - $K_l + K_t$
    - combines the lexical-similarity-based kernel with a basic tree kernel.
    - $K_t((T',H'), (T'',H'')) = K_T (T', T'')+K_T (H',H'')$

Zanzotto, Moschitti

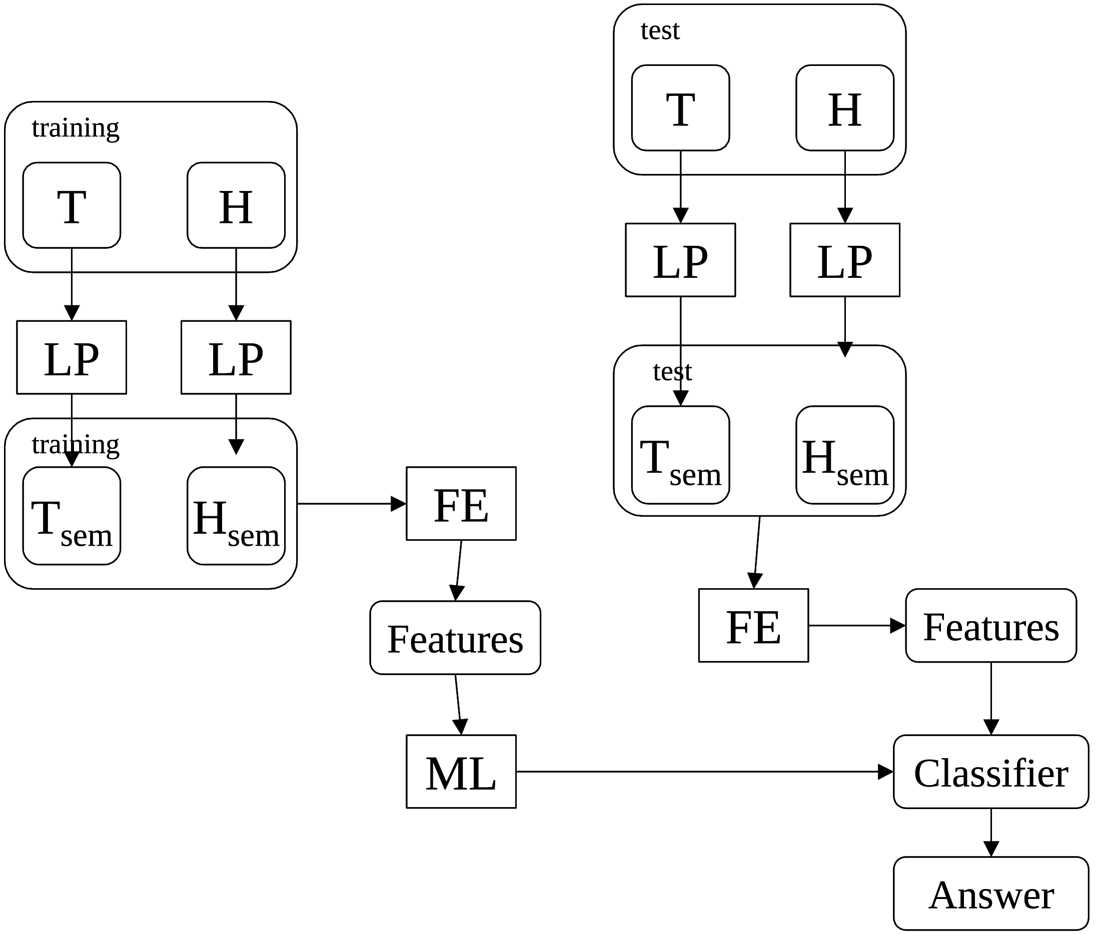
- Results

Datasets	K_l	$K_l + K_t$	$K_l + K_s$
Train: $D1$ Test: $T1$	0.5888	0.6213	**0.6300**
Train: $T1$ Test: $D1$	0.5644	0.5732	**0.5838**
Train: $D2(50\%)'$ Test: $D2(50\%)''$	0.6083	0.6156	**0.6350**
Train: $D2(50\%)''$ Test: $D2(50\%)'$	0.6272	0.5861	**0.6607**
Train: $D2$ Test: $T2$	0.6038	0.6238	**0.6388**
Mean	0.5985 ($\pm$ 0.0235)	0.6040 ($\pm$ 0.0229)	**0.6297** ($\pm$ 0.0282)

TALP UPC

Process

- Linguistic Processing
- Semantic-based distance measures
- Classifier
  - Adaboost
  - SVM



TALP UPC

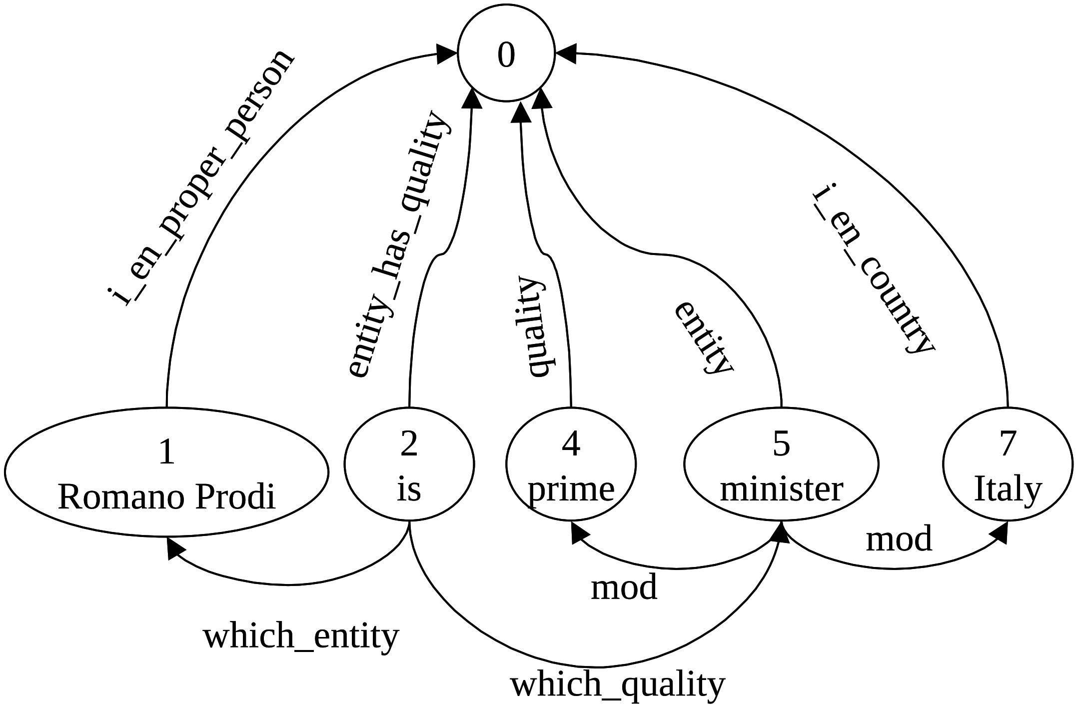
Linguistic Processing

- tokenization
- morphologic tagging
- lemmatization
- fine grained Named Entities Recognition and Classification
- syntactic parsing and robust detection of verbal predicate arguments
  - Spear parser (Surdeanu, 2005)
- semantic labeling, with WordNet synsets
- Magnini´s domain markers
- EuroWordNet Top Concept Ontology labels

TALP UPC

"Romano_Prodi $_1$ is $_2$ the $_3$ prime $_4$ minister $_5$ of $_6$ Italy $_7$"

i_en_proper_person(1),
entity_has_quality(2),
entity(5),
i_en_country(7),
quality(4),
which_entity(2,1),
which_quality(2,5),
mod(5,7),
mod(5,4).



TALP UPC

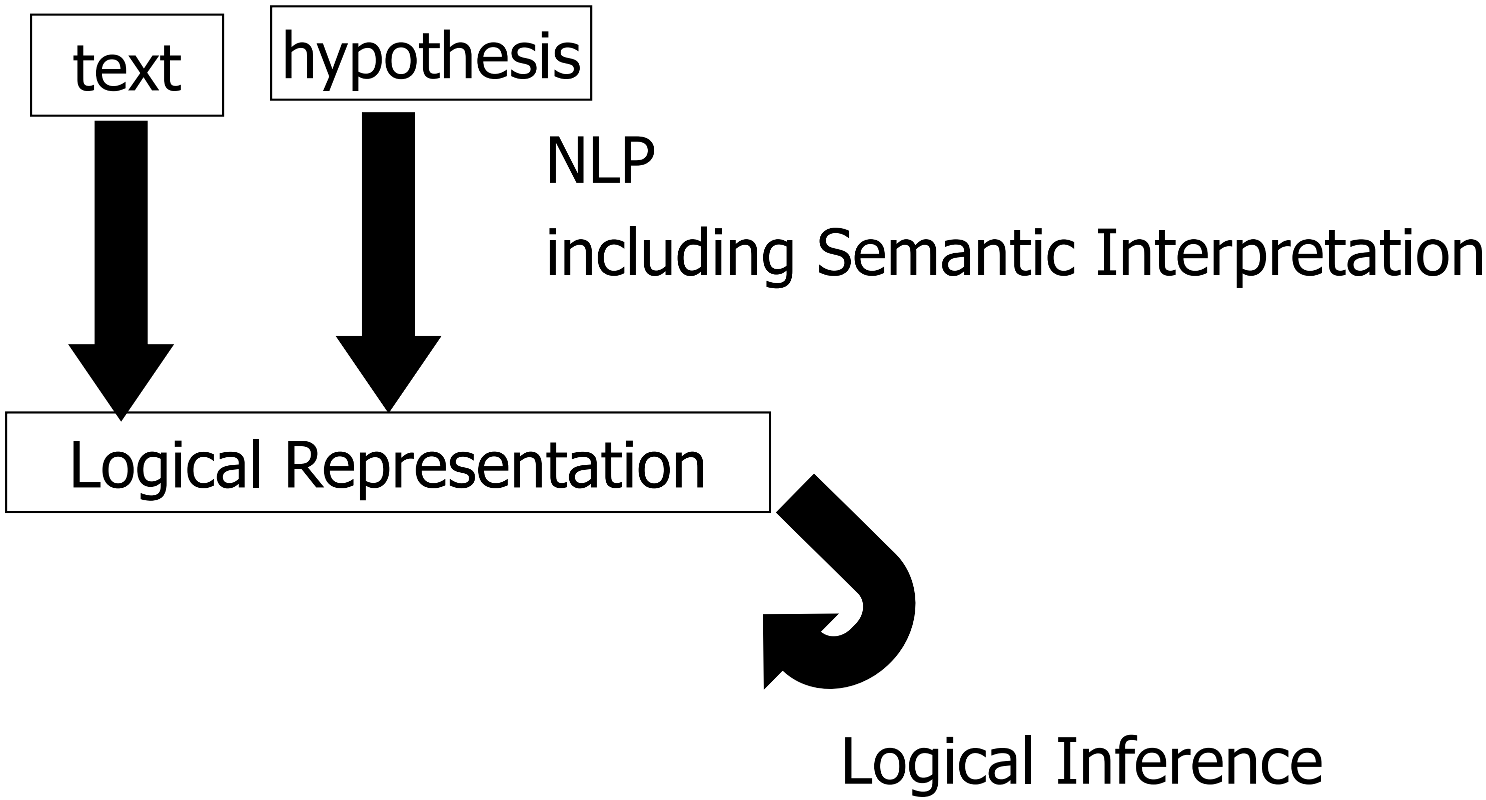
Semantic-based distance measures between T and H

- Strict overlapping of unary predicates.
- Strict overlapping of binary predicates.
- Loose overlapping of unary predicates.
- Loose overlapping of binary predicates.

TALP UPC

Type of feature	**# features**	**description**
semantic content of T	12	# locations, # persons, # dates, # actions, ...
semantic content of H	12	...
intersection of T and H	12	...
Strict overlapping of unary predicates	5	length of intersection score of intersection ratio of intersection related to shortest env ratio of intersection related to longest env ratio of intersection related to both (union of)
Strict overlapping of binary predicates	5	...
Loose overlapping of unary predicates	5	...
Loose overlapping of binary predicates	5	...
Verbal entailment (WordNet)	1	$V_1 \in T$, $V_2 \in H$, such that V_1 verbal_entails V_2
Antonymy	1	$A_1 \in T$, $A_2 \in H$, such that A_1 and A_2 are antonyms and no token compatible with A_2 occurs in H
Negation	1	Difference between # negation tokens in H and T

Approaching RTE from Logic Inference



Nutcracker

Nutcracker, Roma (La Sapienza)

- Components of Nutcracker:
  - The C&C parser for CCG
  - Boxer
  - Vampire, a FOL theorem prover
  - Paradox and Mace, FOL model builders
- Background knowledge
  - WordNet [hyponyms, synonyms]
  - NomLex [nominalisations]

Nutcracker

- Given a textual entailment pair T/H withtext T and hypothesis H:
  - Produce DRSs for T and H
  - Translate these DRSs into FOL
  - **Generate Background Knowledge in FOL**
- Use ATPs to determine the likelyhood of entailment

Nutcracker

- ...
  - Generate Background Knowledge in FOL
- ...

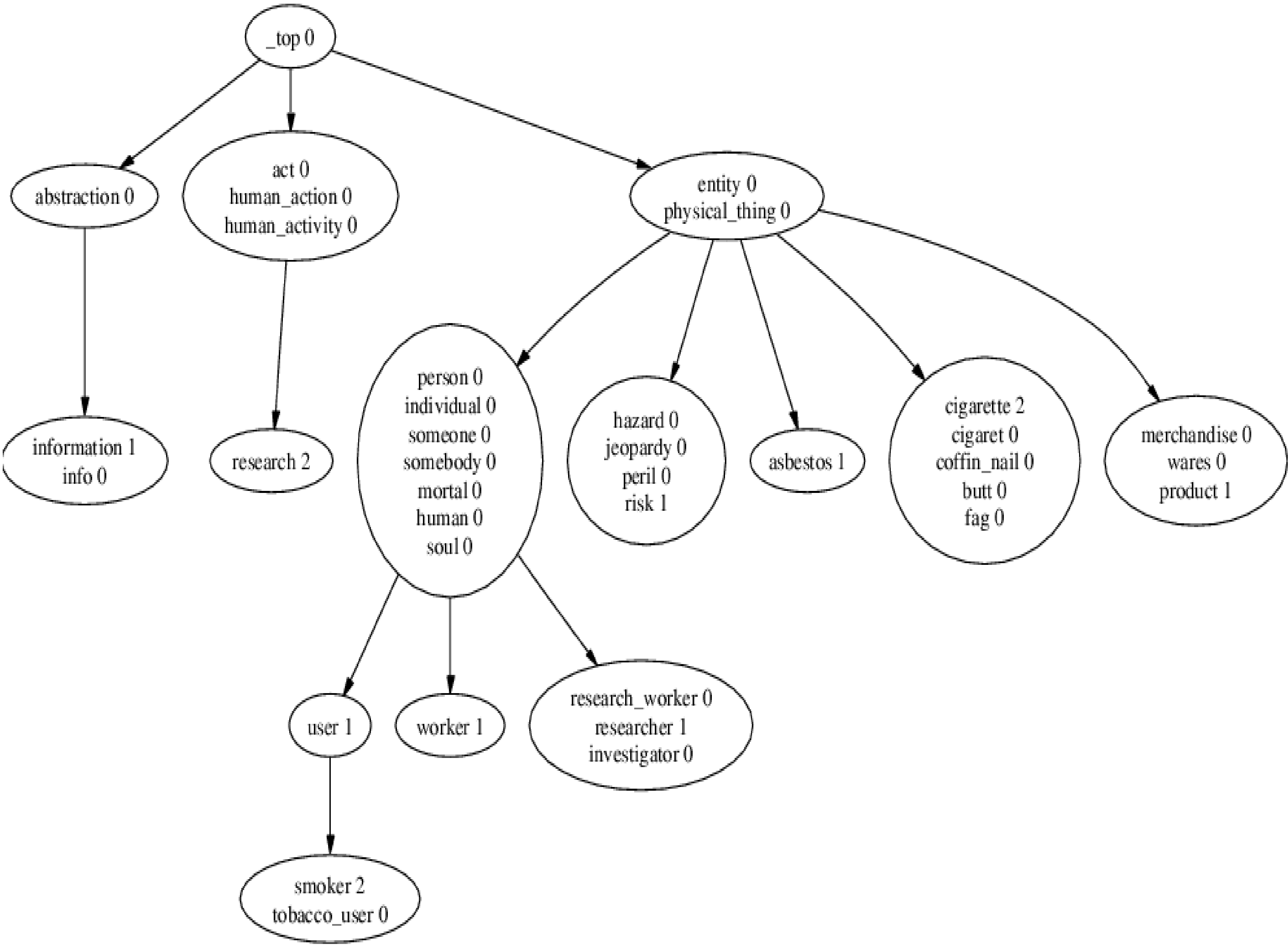
- MiniWordNets
  - Use hyponym relations from WordNet to build an ontology
  - Do this only for the relevant symbols
  - Convert the ontology into first-order axioms

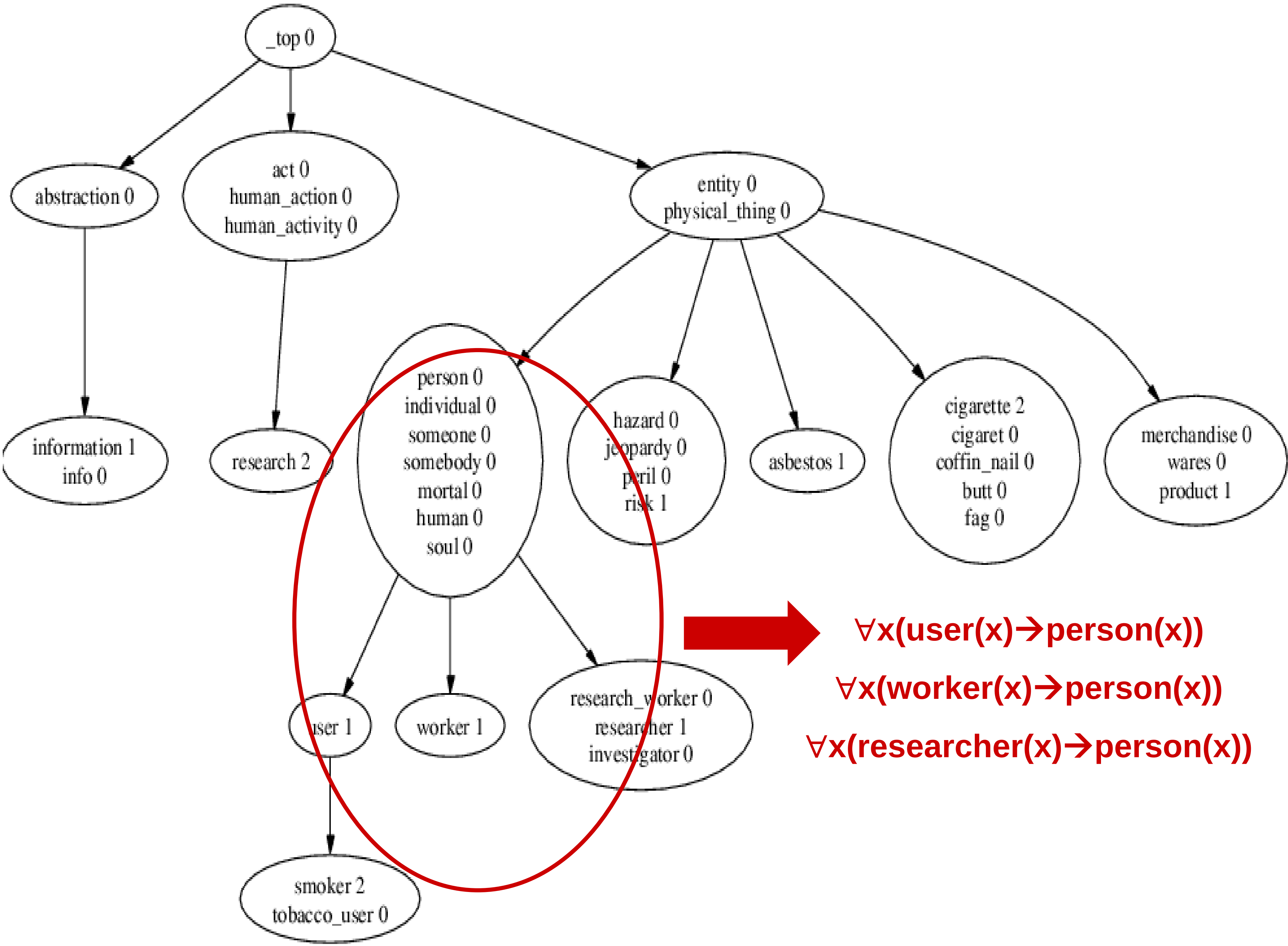
Nutcracker

- ...
  - MiniWordNets

- Example text:

There is no asbestos in our products now. Neither Lorillard nor the researchers who studied the workers were aware of any research on smokers of the Kent cigarettes.

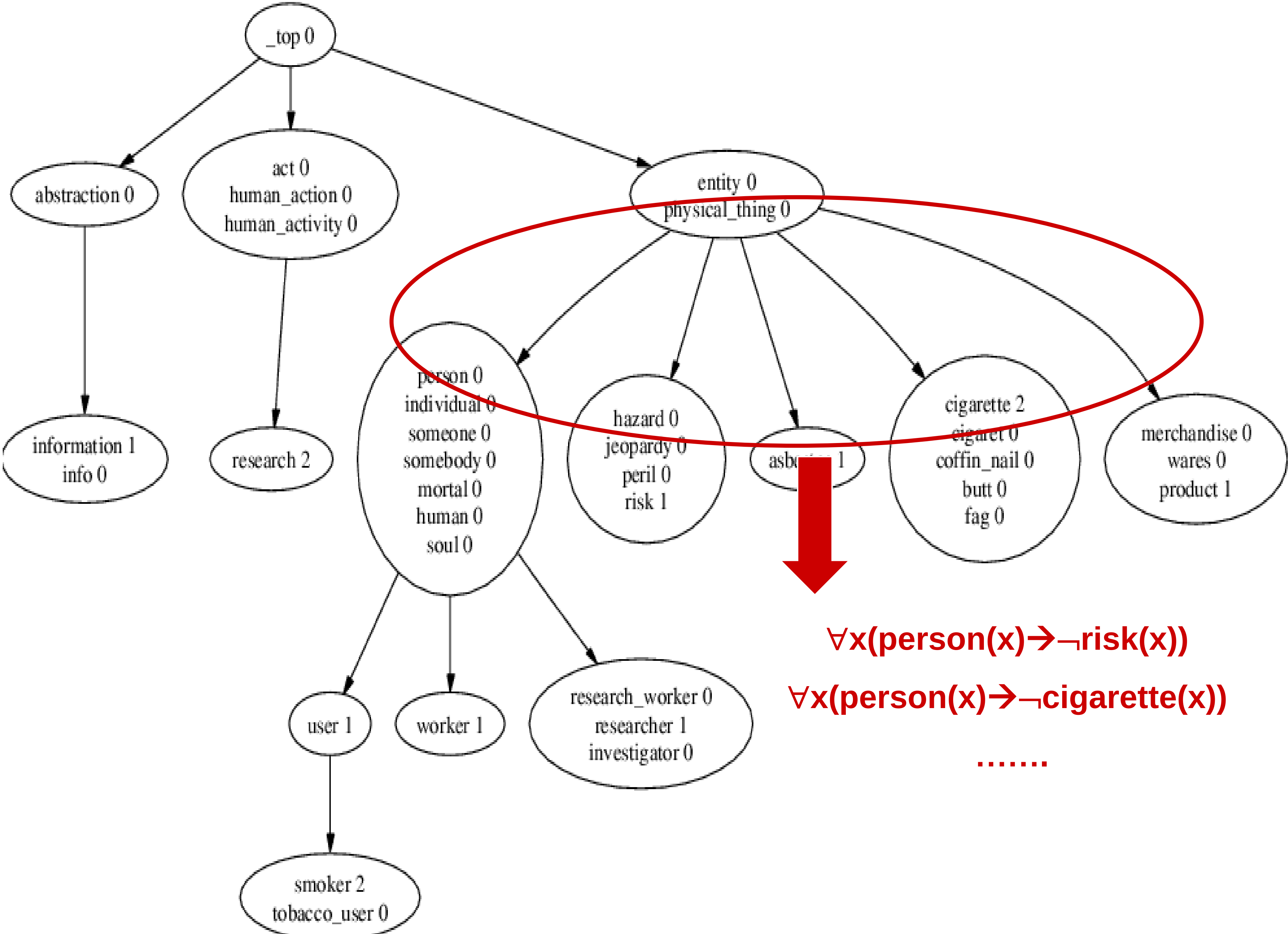

_top 0
abstraction 0
act 0
human_action 0
human_activity 0
entity 0
physical_thing 0
information 1
info 0
research 2
person 0
individual 0
someone 0
somebody 0
mortal 0
human 0
soul 0
hazard 0
jeopardy 0
peril 0
risk 1
asbestos 1
cigarette 2
cigaret 0
coffin_nail 0
butt 0
fag 0
merchandise 0
wares 0
product 1
user 1
worker 1
research_worker 0
researcher 1
investigator 0
smoker 2
tobacco_user 0



$\forall x(\text{user}(x) \rightarrow \text{person}(x))$

$\forall x(\text{worker}(x) \rightarrow \text{person}(x))$

$\forall x(\text{researcher}(x) \rightarrow \text{person}(x))$



$\forall x(\text{person}(x) \rightarrow \neg \text{risk}(x))$

$\forall x(\text{person}(x) \rightarrow \neg \text{cigarette}(x))$

....

Nutcracker

- ...
  - Use ATPs to determine the likelyhood of entailment

    - Create Background Knowledge for T&H
    - Give this to the theorem prover:
      - (BK & T’) → H’
    - If the theorem prover finds a proof, then we predict that T entails H

Nutcracker

- The basic problem of this approach is the use of BK
  - The results are excellent on precision but have a a low recall
  - WN is clearly not enough for representing BK
  - Other Knowledge Sources are needed

Paraphrases

- Paraphrases
  - alternative ways to convey the same information
  - they retain "approximate conceptual equivalence"
- Some Applications of paraphrases:
  - increase the expresive power of NLG systems
  - MDS
  - IE
  - Q&A
  - Language simplification
  - Generating artificial examples for ML
  - ...

Introduction 2

- Linguistic bases:
  - Generative Transformational Grammar (Chomsky)
    - transformational rules (e.g. active to passive voive transformation
  - Meaning Text Theory (Melcuk)
    - Explanatory Combinatorial Dictionary
    - 60 lexical functions
      - ex.
      - Magn(X) maps a word X into words that intensify it
      - Magn("condemn")="strongly" ...
      - Magn("shave")="clean" ...
    - 60 paraphrasing rules

Introduction 3

- Types of paraphrases:
  - lexical vs syntactic
  - granularity
    - word (D)
    - phrase (C,G)
    - sentence (A,B,E,F)
  - atomic vs compositional
    - compositional rules
      - represented as partly lexicalized dependency trees
        - » (NP1 VB1 NP2; NP2 was VBed1 by NP1)
  - meaning distorsion effects (Dras, 1999)
    - change of perspective
    - change of enphasis
    - change of relation
    - deletion
    - clause movement

Introduction

- Examples (from Barzilay, 2003)

A) Emma did not know how to waltz.
Emma had no clue about waltzing

B) The paper was hotly debated, causing a fine old uproar
The article was warmly discussed, which procured it a high reputation.

C) wooden frame
frame made of wood

D) debate
discuss

E) Eli planted a tomato bush.
a tomato bush was planted by Eli.

F) Louis sold the book to Noemie.
Noemie bought the book from Louis.

G) to aim the guns.
to get the best firing angles.

What to read about ...

- Regina Barzilay's thesis (2003)
- Marta Vila’s thesis (2015)
- Proceedings of the ACL 2003 Workshop on Paraphrasing
- Other thesis
  - Mark Dras (1999)
  - Florence Duclaye (2003)
- People:
  - Dekang Lin, Lillian Lee, Kevin Knight, Satoshi Sekine, Hua Wu

Related issues ...

- Similarity & distance measures...
- Looking for synonyms
  - Pereira et al (1993), Lin (1998), Wu, Zhou (2003)
- Looking for collocations, multiword decomposition
  - Baldwin et al (2003), Evert (2004), Pearce (2001,2002)
- Looking for terms
  - Vivaldi (2003), Jacquemin (1999)
- Text simplification
  - Chandrasekar et al, 2003, Carroll et al, (1999)
- Induction of IE patterns
  - Turmo (2003)
- Parallel corpus aligment
  - Melamed (2000), Giza,
- Analogy learning
  - Turney et al, 2003

More in depth ...

- Text-to-text generation vs Concept-to-text generation
- Transforming text satisfying specific constraints:
  - Summarization: length
  - Text simplification: style
  - Paraphrasing: ???
- lack of a formal model
- paraphrase within a particular context
- sense meaning vs reference meaning
- synonymy as a subclass of atomic paraphrase
- near-synonymy

Knowledge Sources

- monolingual vs multilingual
  - monolingual dictionaries
    - Wu, Zhou, 2003, Kaji et al, 2000
  - multilingual dictionaries
  - monolingual corpus
    - Barzilay (2003)
  - multilingual corpus
    - parallel
      - Pang et al (2003), Ibrahim et al (2003)
    - comparable
      - Barzilay, Elhadad (2003)
- Thesaurus
  - WN
    - synonymy relations,
    - other relations (direct or derived)
    - mapping WN relations into paraphrases
  - automatically built from distributional information

Approaches 1

- Depending on the granularity
  - Atomic
    - Learning synonyms: Pereira, Wu, Lin
  - Phrase-level
    - Barzilay, McKeown, Jacquemin
  - Structural
    - Lin, Pantel, Shinyama, Sekine
- and the KS
  - Parallel translations (mono-l or multi-l)
    - Barzilay, McKeown, Elhadad, Lee,
  - MI of word distribution
    - Lin, Pantel
  - Text alignment + FSA
    - Pang, Knight
  - Pairs Q&A
    - Ravichandran, Hovy