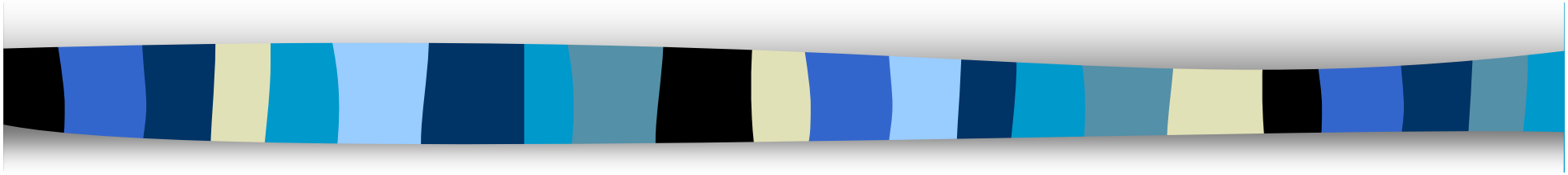
Relation Validation via Recognizing Textual Entailment



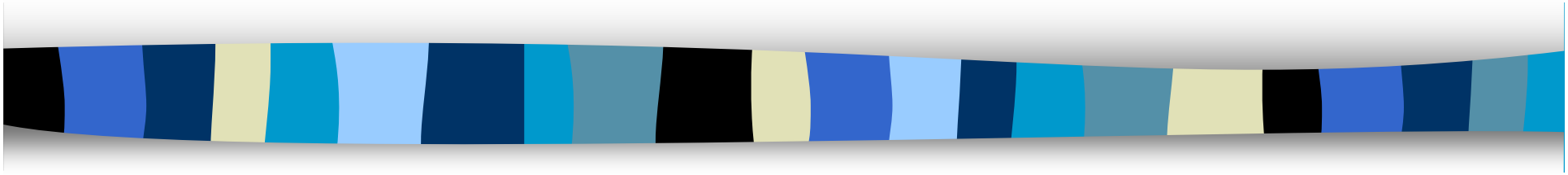
Rui Wang (*Saarland University*)
Günter Neumann (*DFKI*)

OBIES 2008 (*Kaiserslautern, 23.09.2008*)

Outline

- Introduction
  - Relation Validation (RV)
  - Recognizing Textual Entailment (RTE)
- Approach
  - RTE approaches
  - From RTE to RV
- Evaluation
  - The RTE system
  - The whole system
- Conclusion & Future Work

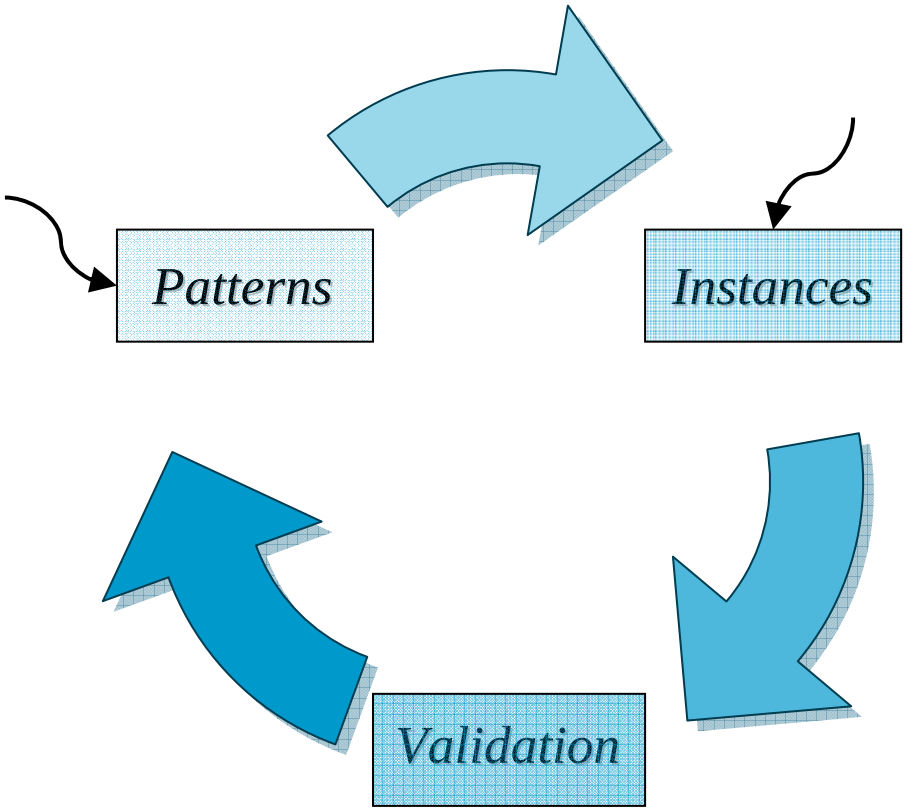
Introduction



Relation Extraction vs. Relation Validation

- An example
  - *Work_for* (person, company, location)
- Why *relation validation* is important/necessary
  - (Unsupervised) relation extraction
  - (Automatic) Ontology population

Unsupervised Relation Extraction



Conventional Validation Approaches

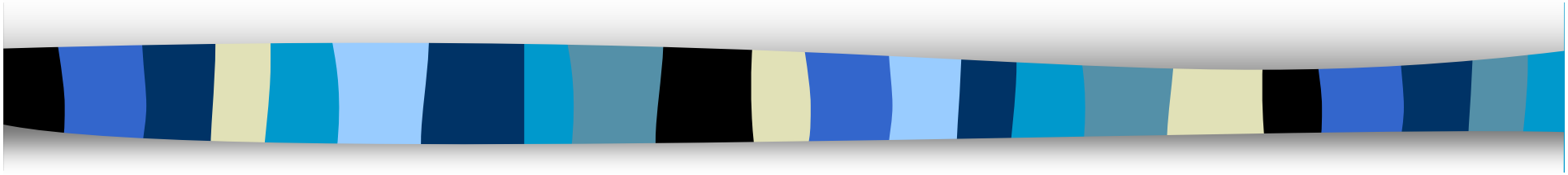
- KnowItAll (Etzioni et al., 2005)
  - Expectation-Maximization
- SnowBall (Agichtein, 2006)
  - PMI (Turney, 2001)
- REALM (Downey et al., 2007)
  - Language models

Recognizing Textual Entailment (RTE)

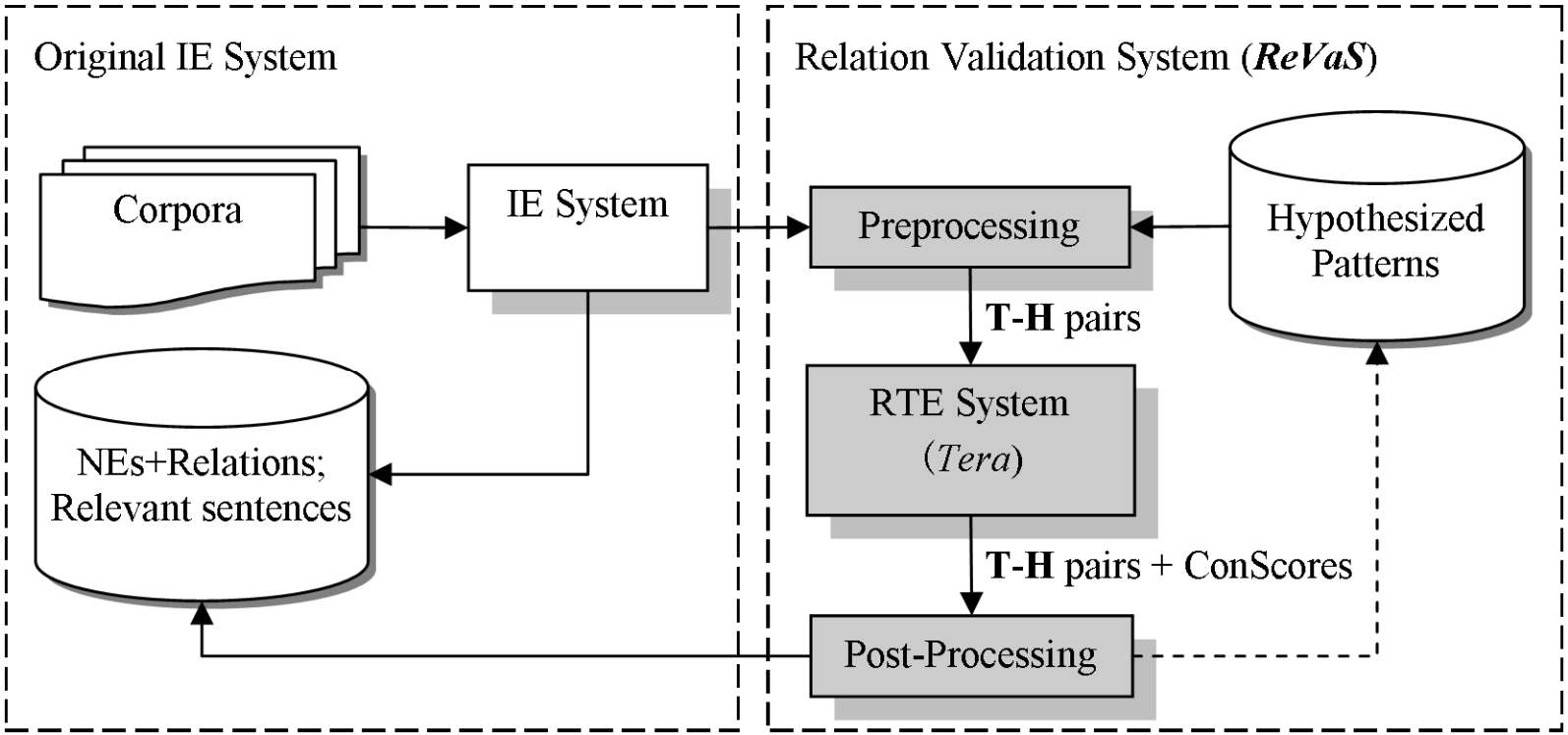
- It is defined as recognizing, given two text fragments, whether the meaning of one text (*Hypothesis* – **H**) can be inferred (entailed) from the other (*Text* – **T**).

- Applications of RTE
  - Answer Validation (Wang and Neumann, 2007 & 2008)
  - Relation Validation (This work)

Approach



System Architecture



The RTE System (*Tera*)

- The Main approach
  - Tree Skeleton extraction
  - Subsequence kernel
- Backup strategies
  - BoW similarity
  - Triple similarity

An Example

- Pair: id=*“61"* entailment=*“**YES**"* task=*“**IE**“* source=*“RTE”*
  - Text:

    *Although they were born on different planets, Oscar-winning actor **Nicolas Cage**'s new **son** and Superman have something in common, both were named **Kal-el**.*

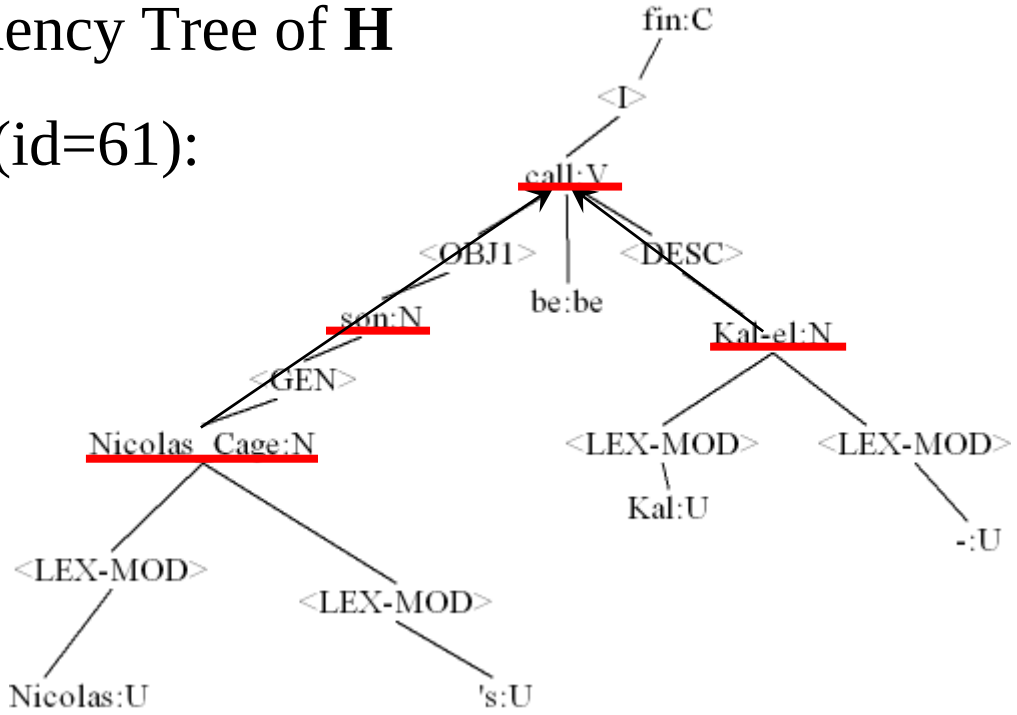
  - Hypothesis:

    ***Nicolas Cage**'s **son** is called **Kal-el**.*

Tree Skeleton

Dependency Tree of **H**

of pair (id=61):

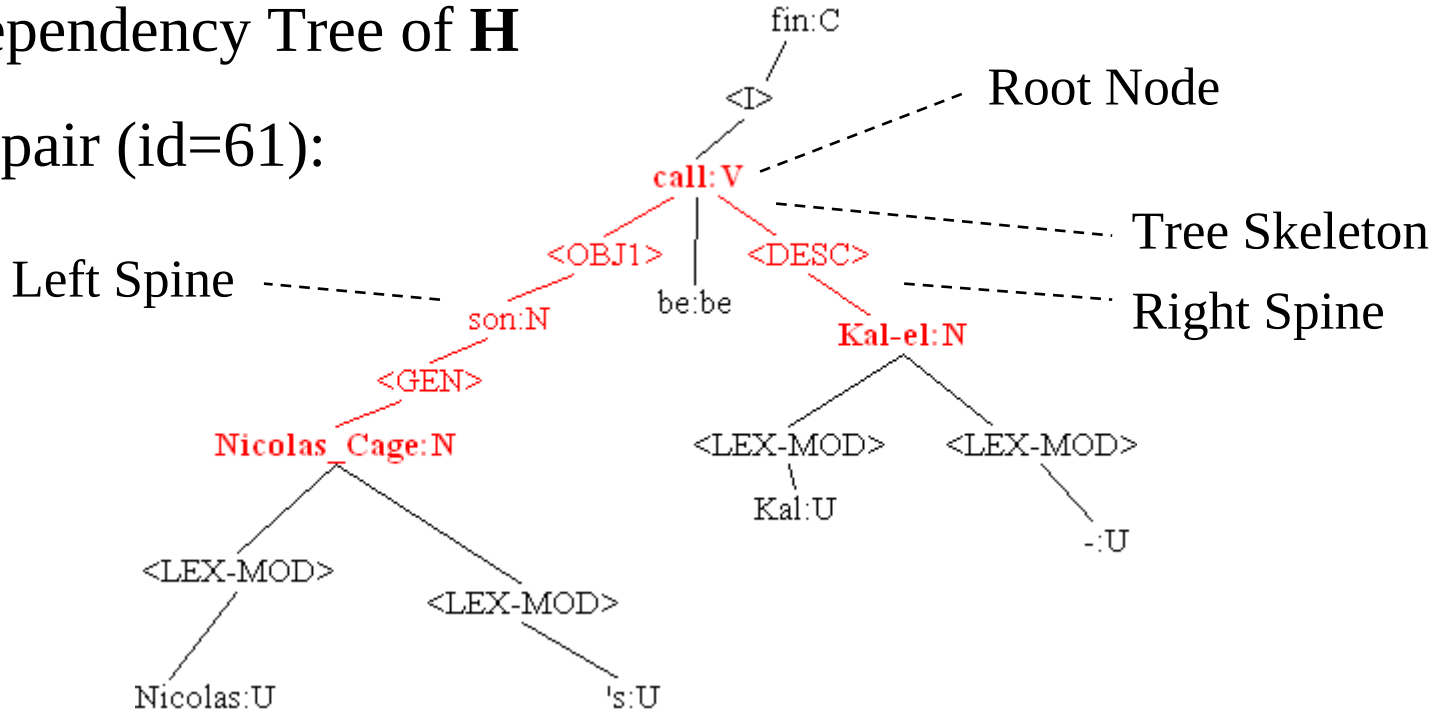


- Text: ***Nicolas Cage's son*** *is called* ***Kal-el.***

Tree Skeleton

Dependency Tree of **H**

of pair (id=61):

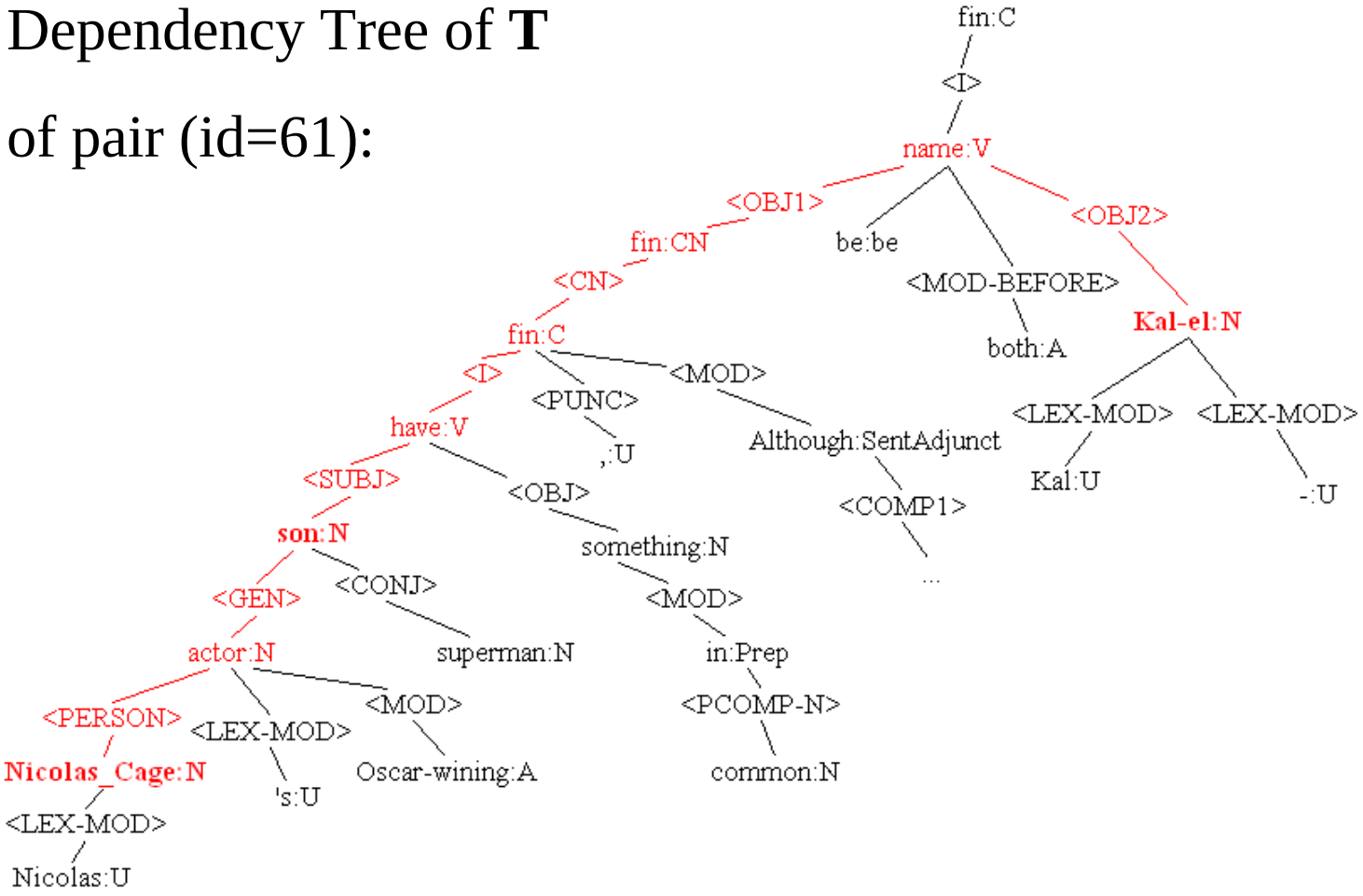


- Text: ***Nicolas Cage**'s **son** is called **Kal-el**.*

Tree Skeleton (cont.)

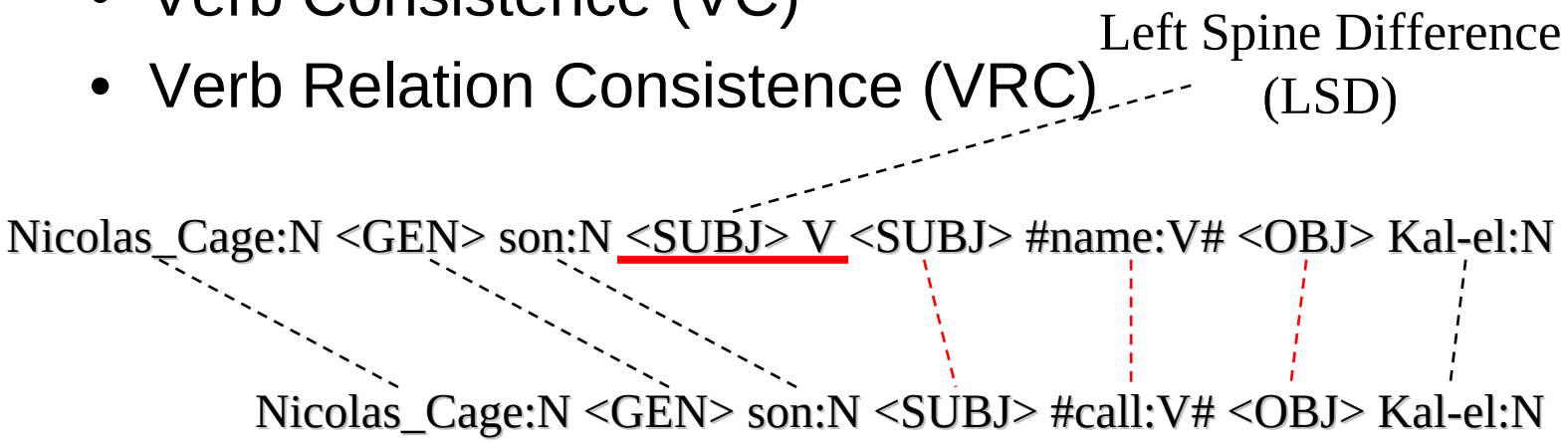
Dependency Tree of **T**

of pair (id=61):



Spine Merging

- Merging
  - Left Spines: exclude Longest Common Prefixes
  - Right Spines: exclude Longest Common Suffixes
- RootNode Comparison
  - Verb Consistence (VC)
  - Verb Relation Consistence (VRC)



Testing Phase

- Pair: id="*247*" entailment="*YES*" task="*IE*" source="*BinRel*"
  - Text:

    Author Jim Moore was invited to argue his viewpoint that Oswald , acting alone , killed Kennedy.
  - Hypothesis:

    Oswald killed Kennedy.

Testing Phase (cont.)

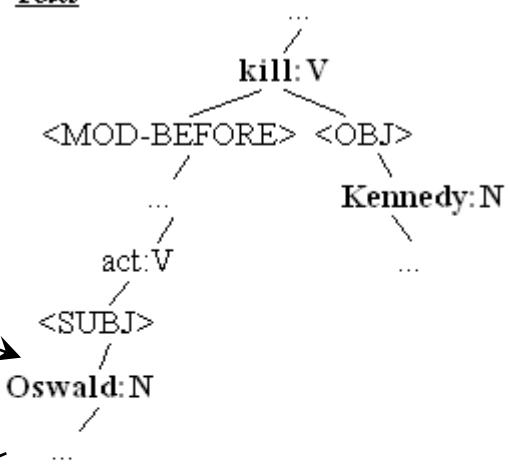
Text

```
.....
<triple left="17" right="E0">kill:V mod-before vpsc:C</triple>
<triple left="17" right="16">kill:V punc ,:U</triple>
<triple left="17" right="E8">kill:V subj Oswald:N</triple>
<triple left="17" right="18">kill:V obj Kennedy:N</triple>
.....
```

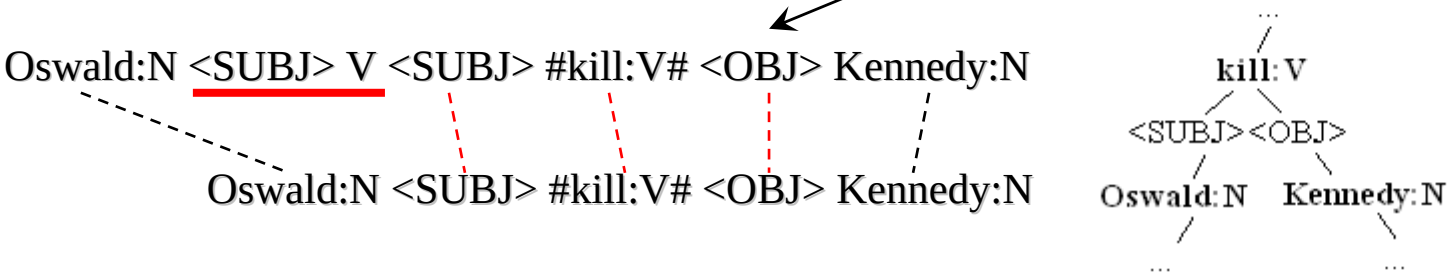
Hypothesis

```
<triple left="E0" right="2">fin:C i kill:V</triple>
<triple left="2" right="1">kill:V s Oswald:N</triple>
<triple left="2" right="E2">kill:V subj Oswald:N</triple>
.....
```

Text



Hypothesis



<"SUBJ V", "", 1, 1> → YES

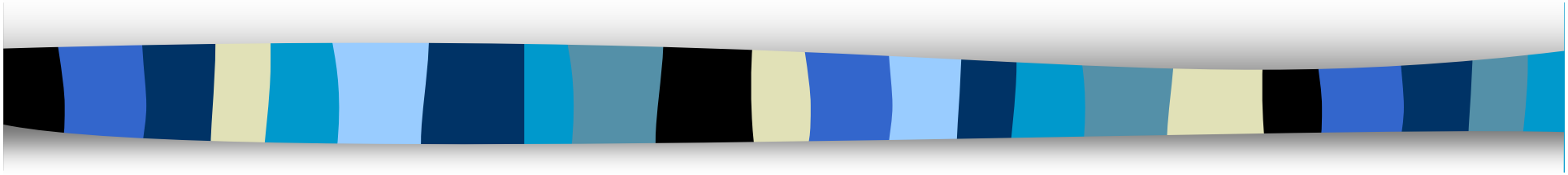
From RTE to RV

- "*The union has hired a number of professional consultants in its battle with the company, including **Ray Rogers** of Corporate Campaign Inc., the **New York** labor consultant who developed the strategy at **Geo. A. Hormel** & Co.'s **Austin, Minn.**, meatpacking plant last year. That campaign, which included a strike, faltered when the company hired new workers and the International Meatpacking Union wrested control of the local union from **Rogers'** supporters.*"

From RTE to RV (cont.)

- Person names
  - “*Ray Rogers*”, “*Rogers*”
- Location names
  - “*New York*”, “*Austin*”, “*Minn.*”
- Consequently, the possible <PN, LN> NE pairs with *birthplace relation* are
  - <“*Ray Rogers*”, “*New York*”>, <“*Rogers*”, “*Austin*”>, …
- Hypotheses
  - “*Ray Rogers* ***is born in*** *New York.*”
  - “***The birthplace of*** *Rogers* ***is*** *Austin.*”

Evaluation



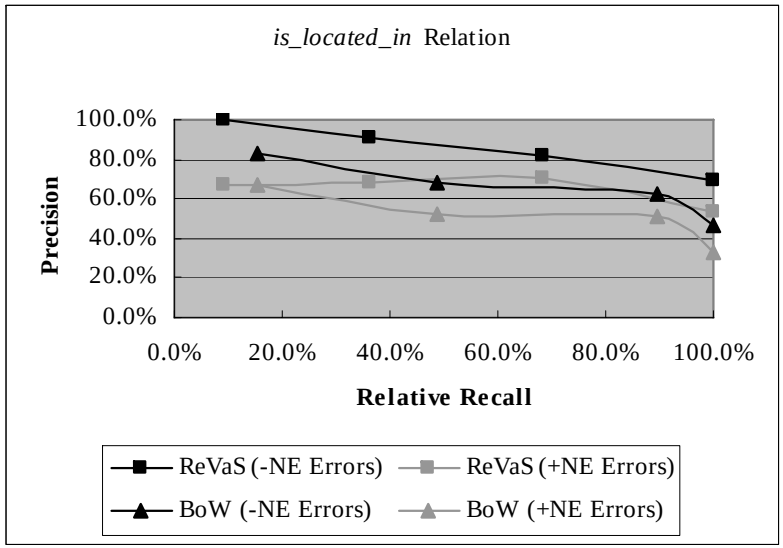
Experiments on Annotated Data

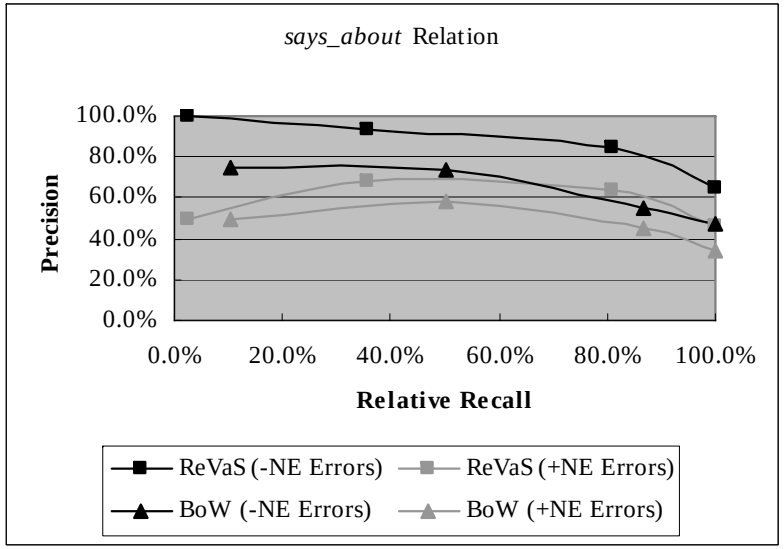
- On annotated data
  - The BinRel corpus (Roth and Yih, 2004)
  - The *birthplace* relation (268 pairs)
  - The *kill* relation (199 pairs)
  - The negative (650 pairs)

Systems	*Kill* relation	*Birthplace* relation
BoW (Baseline1)	72.0%	75.0%
Triple (Baseline2)	70.3%	76.4%
Main+Backups	84.1%	86.5%

Experiments on Web Data

- On web data
  - Collected by IDEX (Eichler et al., 2008)
  - For example,
    - Located<Berlin, Germany>
    - Working<Tricon, Bangkok>
    - Say<Britons, Slovakians>





Discussions

- Advantages of using RTE
  - *…**The City Partner Hotel am Gendarmenmarkt** offers our guests a personal home in the heart of **Berlin**.*
  - The implicit relation

  - ***Randall Lewis**, a spokesman for the Squamish First Nation, said **CN** ...*
  - Non-local dependency relation

Discussions (cont.)

- NE Recognition Errors
  - *CCNB office and core facility The CCNB Core Facility will be centrally located in a designated building on the Campus of the Charité in Berlin **Mitte**.*
  - *She received her **PhD** from the University of Wisconsin-Madison in 1997.*

Discussions (cont.)

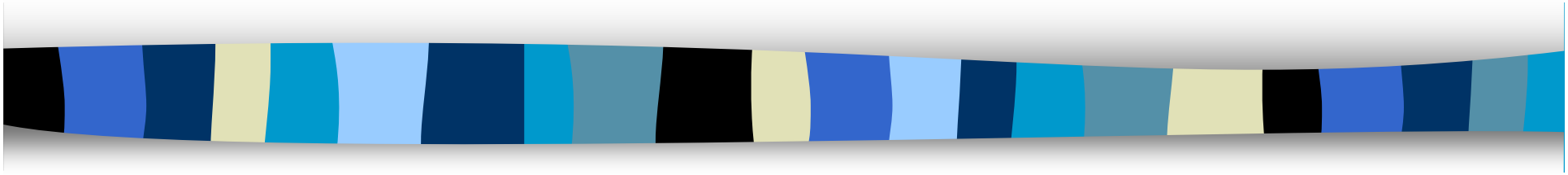
- Relation Errors
  - *…**David W. Roubik**, a staff scientist with the Smithsonian Tropical Research Institute in **Balboa, Panama**.*
  - Birthplace vs. workplace

  - *Geography Setting **Berlin** is located in eastern **Germany**, about 110 kilometers (65 miles) west of the border with **Poland**.*
  - Depth of analyzing modifiers

Discussions (cont.)

- *But the end of **Zoo Station** is the end of yet another era in **Berlin**, the '60s through the '80s, and one can only wonder where the junkies in west Berlin will congregate after it's gone. posted by **Ed Ward** @ 1:22 AM 2 comments 2 Comments: At 3:08 PM, **Daniel Rubin** said... First time I saw the **Hamburg Bahnhof** it was like a scene from a horror movie - - all these grizzled creatures staggering around as the loudspeakers blasted **Mozart**…*

Conclusion & Future Work



Conclusion

- Relation validation is important and necessary
- A precision-oriented approach is suitable for this task

Future Work

- More complex validation tasks
  - Binary relations → N-ary relations
- From validation to extraction (ongoing)
  - Hypothesized relations
- Evaluation metrics
  - Semantic similarity