Temporal Information Extraction

Xiao Ling
University of Washington
Joint work with Dan Weld

Motivation

- Traditional Relation Extraction

Headquartered_in(Boeing, Seattle)

True?

This talk: Temporal Information Extraction

- **Input**: raw text, e.g.

  Steve Jobs revealed the iPhone in 2007.

- **Output**:

  events annotated with bounds on endpoints

$$^{\triangleleft}\texttt{Year-2007} < {}^{\triangleleft}\texttt{Reveal(Jobs, iPhone)} < \texttt{Year-2007}^{\triangleright}$$

$$^{\triangleleft}\texttt{Year-2007} < \texttt{Reveal(Jobs, iPhone)}^{\triangleright} < \texttt{Year-2007}^{\triangleright}$$

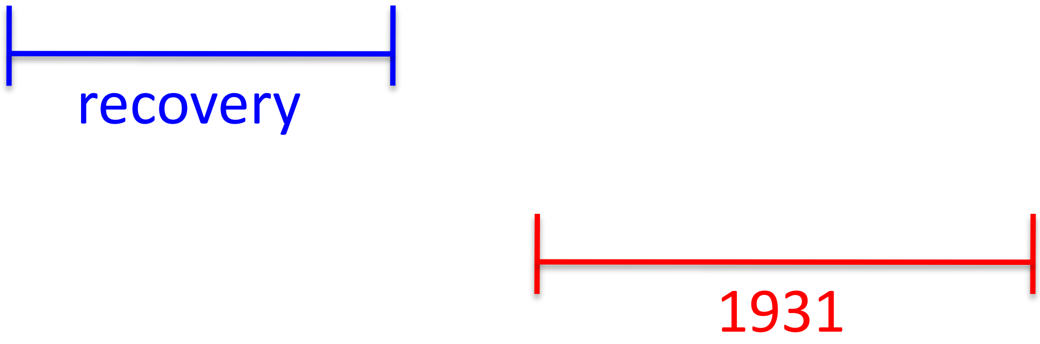
* This work focuses on one sentence at a time

Outline

- Motivation
- Previous Work
- TIE
- Experiments
- Conclusion

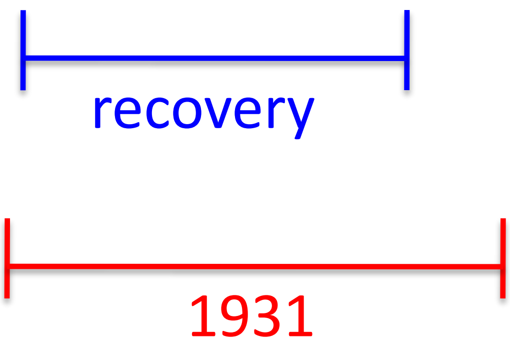
TempEval [Verhagen et al, 2007]

In most countries of the world ***recovery*** *from the Great Depression began between late* ***1931*** *…*



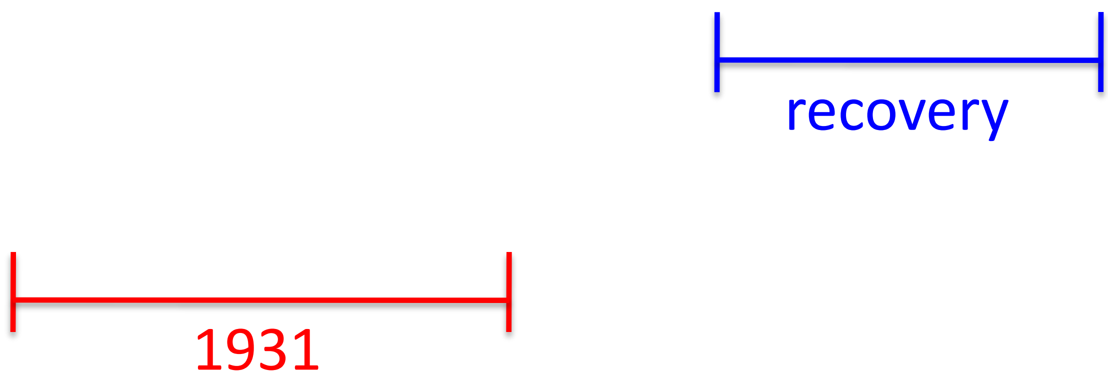
TempEval [Verhagen et al, 2007]

*In most countries of the world **recovery** from the Great Depression began between late **1931** …*



TempEval [Verhagen et al, 2007]

In most countries of the world ***recovery*** *from the Great Depression began between late* ***1931*** *…*



Transitivity (Yoshikawa et al., 2009)

- Restricted to the elements in test data

In most countries of the world, recovery from the Great Depression began between late 1931 and early 1933 .

In the United States recovery began in the spring of 1933.

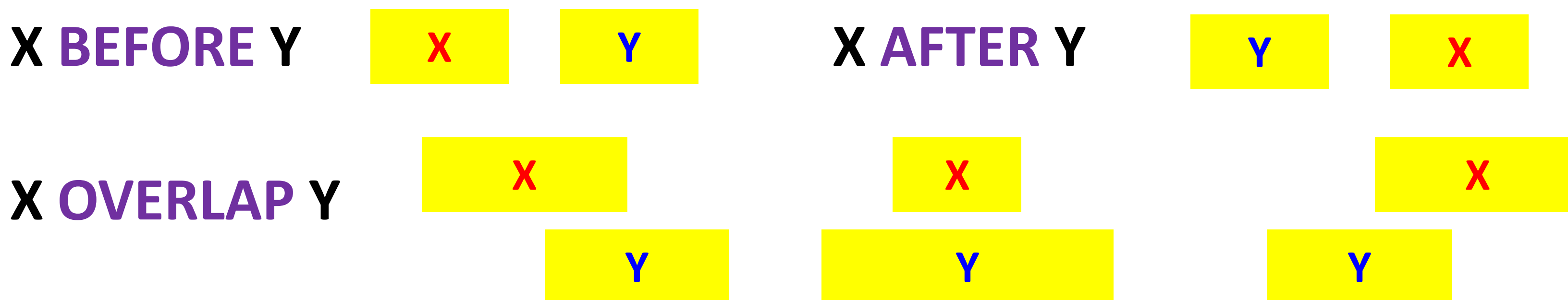
- **Intra-sentence** transitivity not fully exploited

Transitivity cont.

(Verhagen et al, 2007), (Yoshikawa et al., 2009)

- Restrict the relation set to

  {BEFORE, AFTER, OVERLAP}



OVERLAP is ambiguous!

A **OVERLAP** B, B **OVERLAP** C => A ? C

- Point-based relations:

  p1 < p2, p2 < p3 => p1 < p3

Contributions

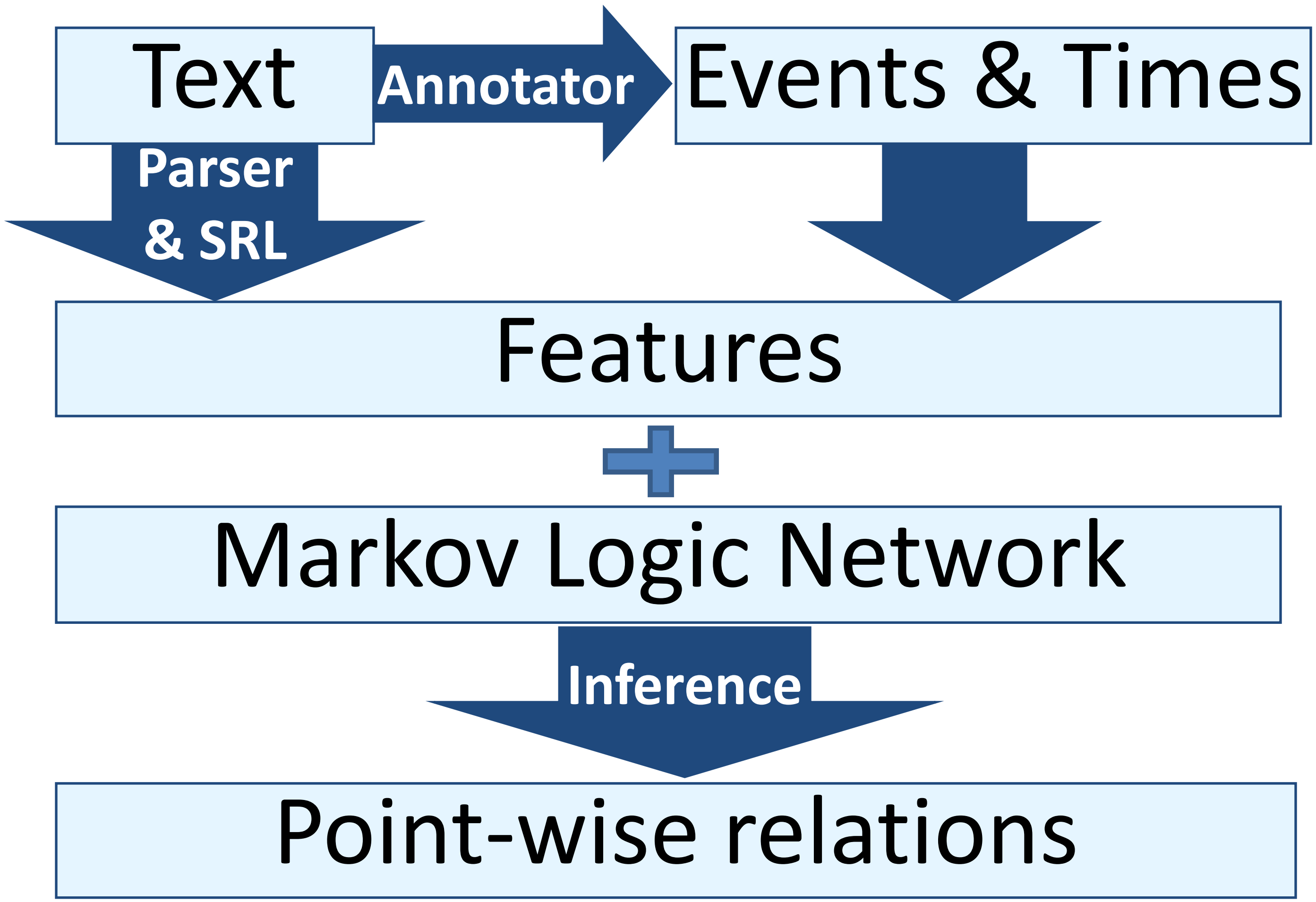
- System TIE (Temporal Information Extractor)
  - Intra-sentence transitivity
  - high level features
- Temporal Entropy

  a new metric for measuring tightness of the bounds

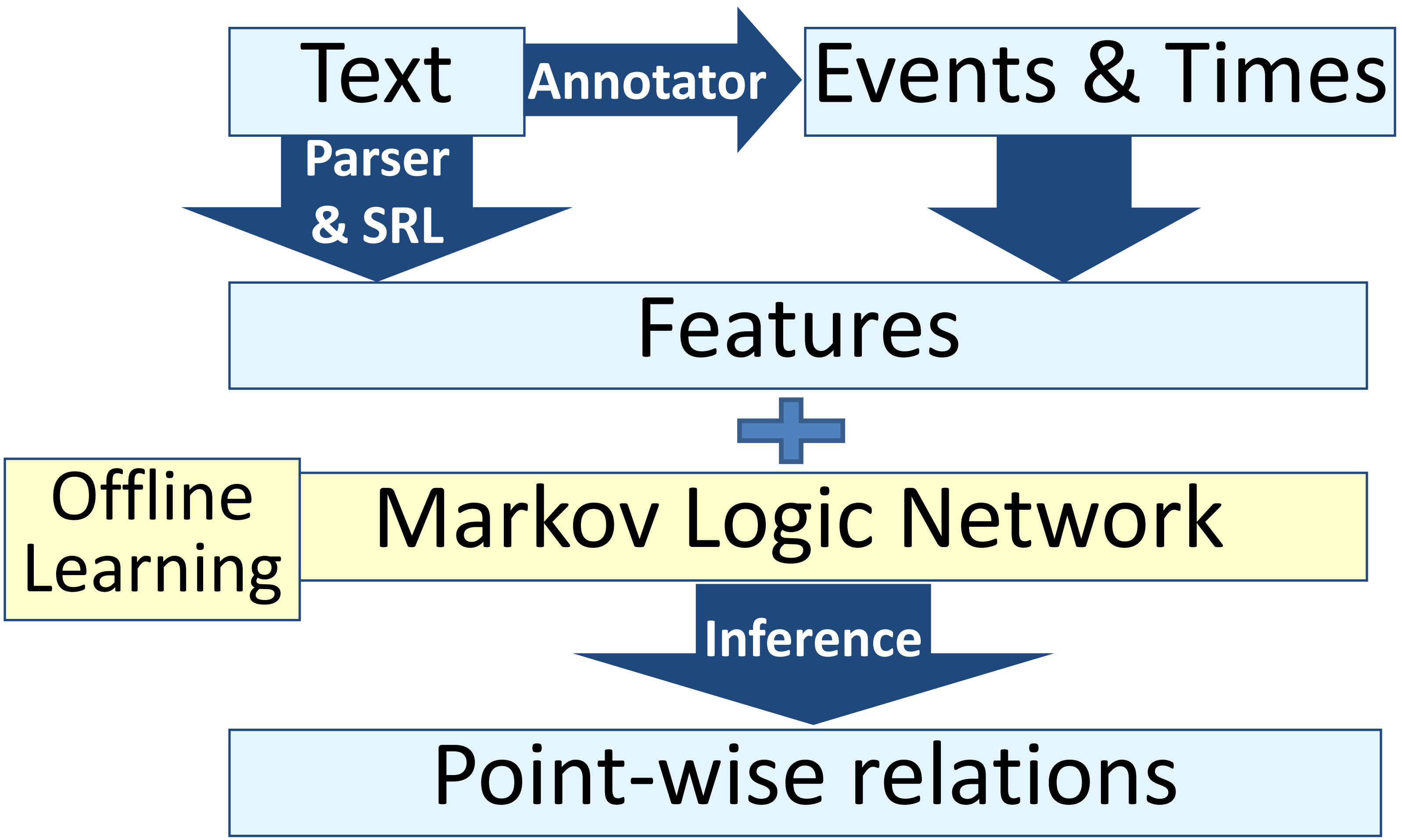
Outline

- Motivation
- Previous Work
- TIE
  - Architecture
  - Learning
  - Inference
- Experiments
- Conclusion

TIE: System Overview



TIE: System Overview



Markov Logic Networks

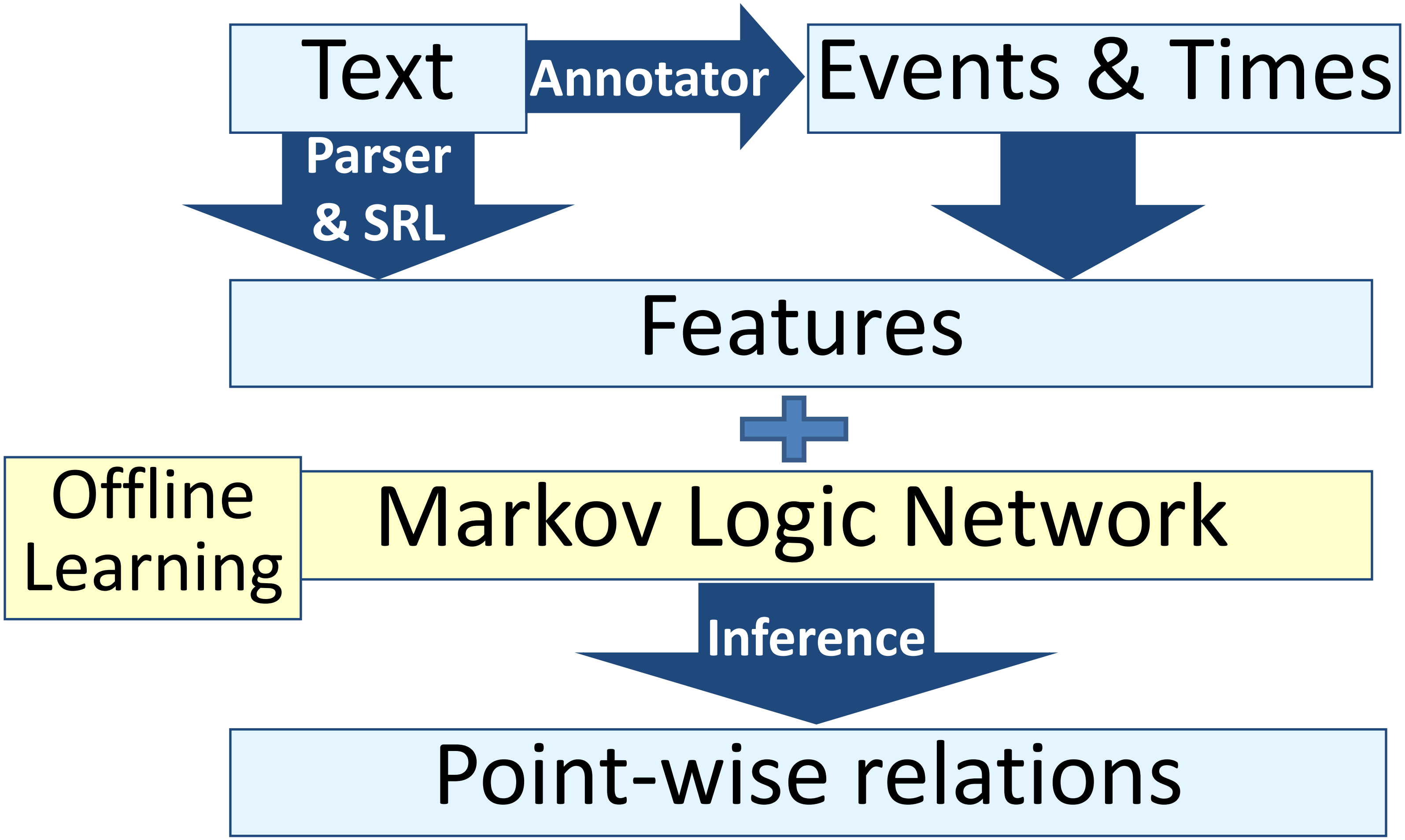
- A **Markov Logic Network (MLN)** is a set of pairs (F_i, w_i) where
  - $\mathbf{F_i}$ is a formula in first-order logic
  - $\mathbf{w_i}$ is a real number as the w

# of true groundings of F_i

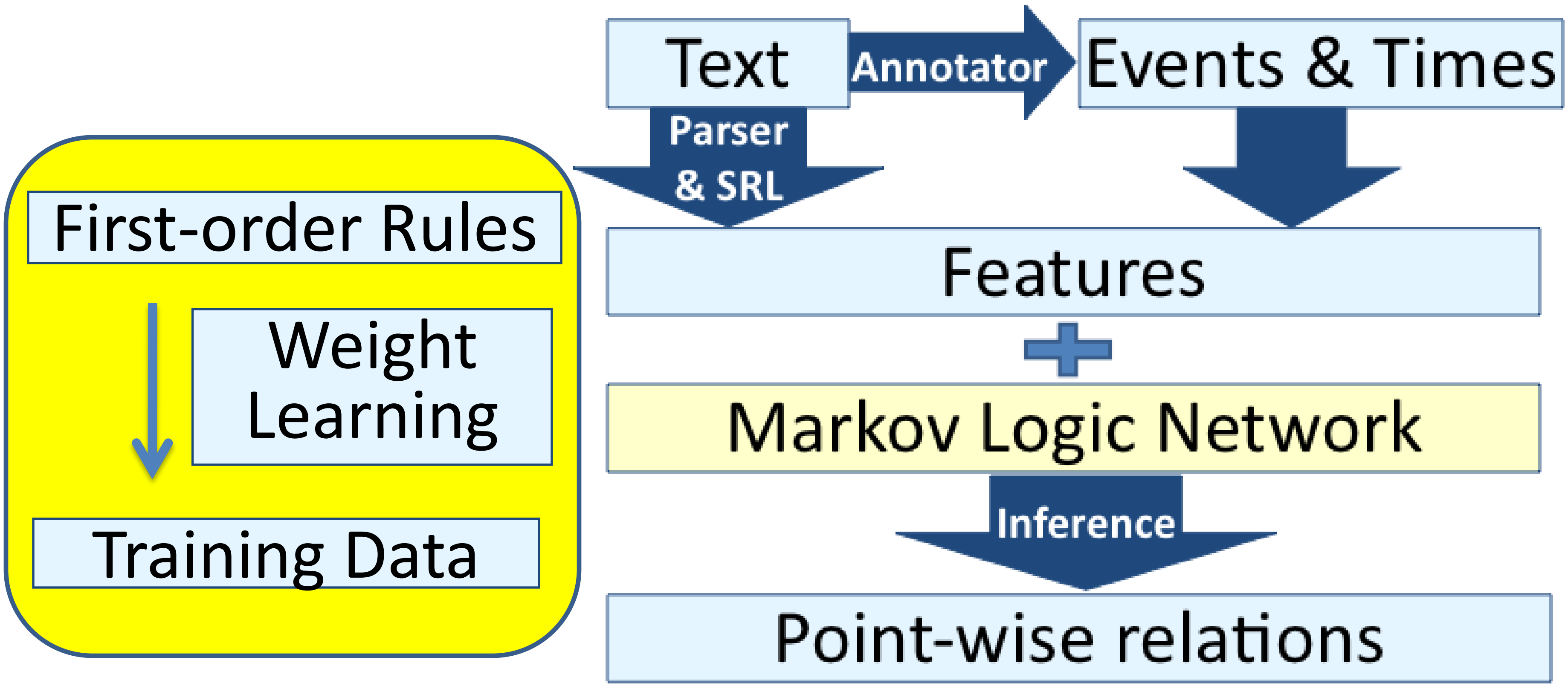
$$P(x) = \frac{1}{Z} \exp\left(\sum_i w_i \cdot N_i(x) \right)$$

*Borrowed from Hoifung's slides

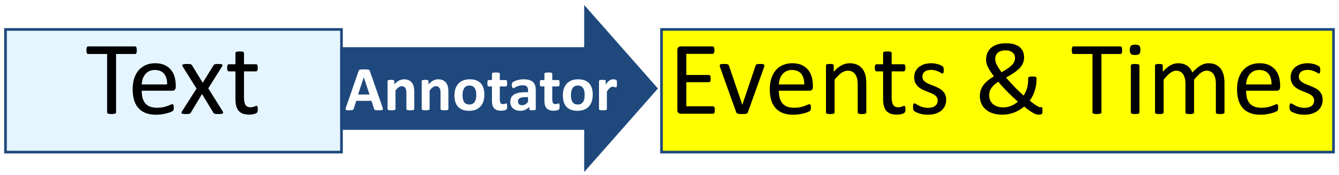
TIE: System Overview



TIE: System Overview



TIE: System Overview



Annotating Event and Time

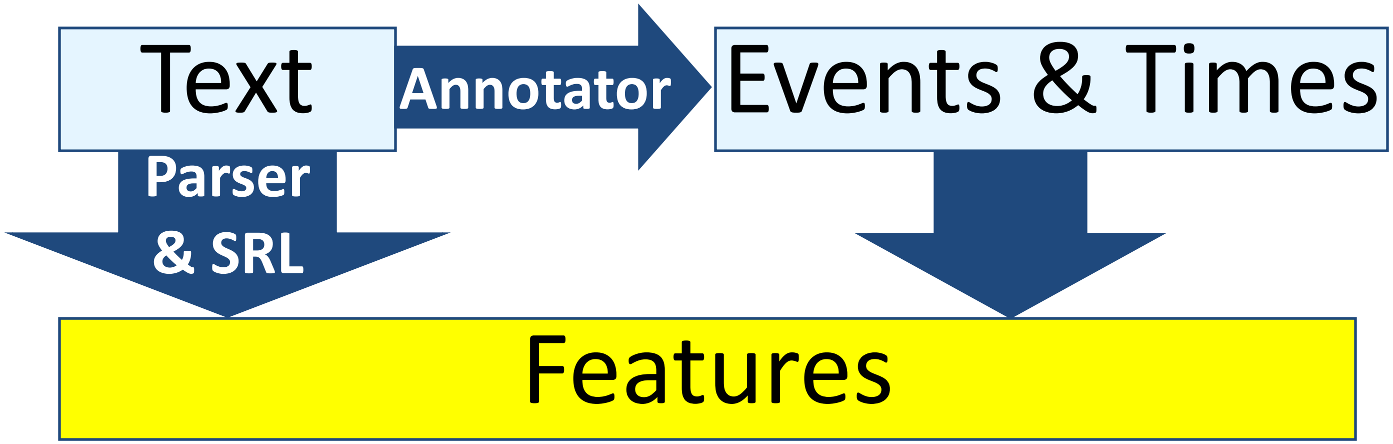
[Verhagen et al,2005]

In most countries of the world, recovery from the Great Depression began between late 1931 and early 1933 .

TARSQI

In most countries of the world, recovery^{e1} from the Great Depression^{e2} began^{e3} between late 1931^{t1} and early 1933^{t2}.

TIE: System Overview



Dependency Parsing

[De Marneffe et al,2006]

In most countries of the world, recovery^{e1} from the Great Depression^{e2} began^{e3} between late 1931^{t1} and early 1933^{t2}.

Syntactic Dependency: prep_between(began, 1931)

Feature: **prep_between(e3, t1)**

Semantic Role Labeling

[Koomen et al,2005]

In most countries of the world, recovery^{e1} from the Great Depression^{e2} **began** between late 1931^{t1} and early 1933^{t2}.

In most countries of the world, recovery^{e1} from the Great Depression^{e2} began^{e3} between late 1931^{t1} and early 1933^{t2}.

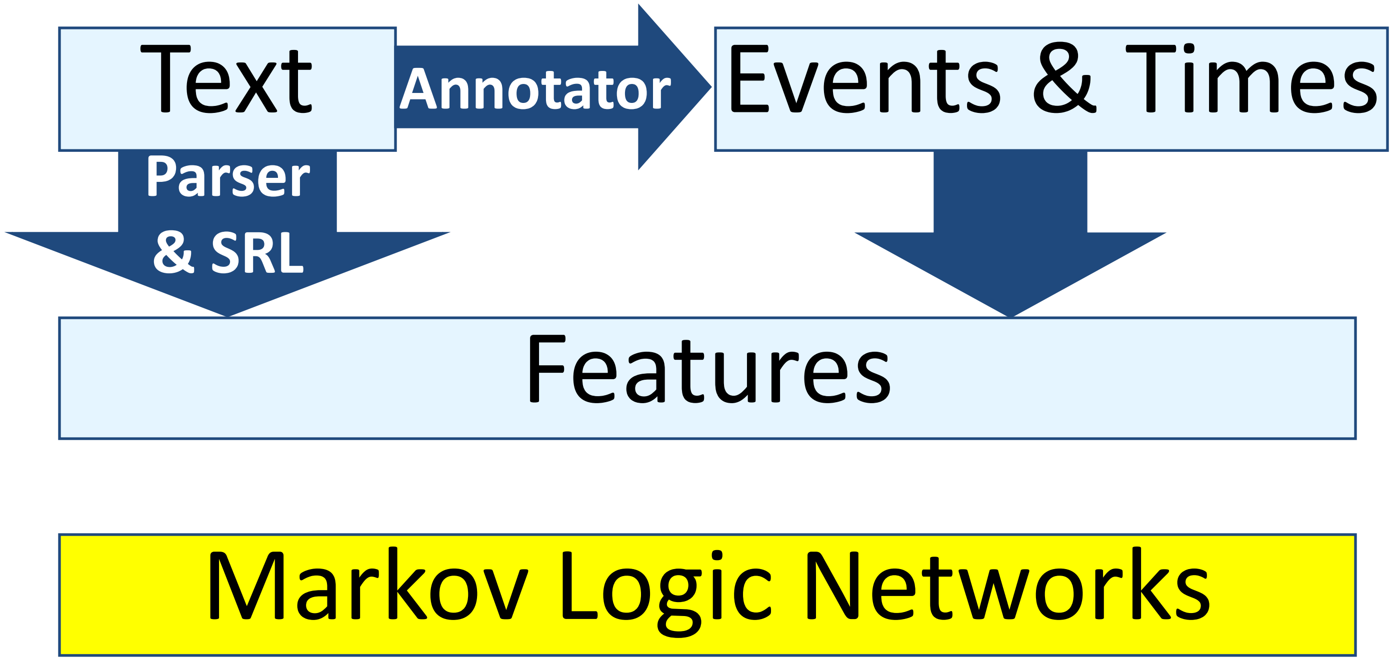
Temporal Argument

Feature: **srl_after(◁e3, ◁t1), srl_after(t2▷, ◁e3)**

Summary of Features

- Event and Time attributes
  - value(t1, “1933”), tense(e1, “PAST”)
- Syntactic Dependency
  - prep_between(e3, t1)
- SRL Features
  - srl_after(◁e3, ◁t1)

TIE: System Overview



4 MLN formula templates

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

point(x) $\in \{ {}^{\triangleleft}x, x^{\triangleright} \}$

4 MLN formula templates

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

point(x) $\in \{ ^{\triangleleft}x, x^{\triangleright} \}$

4 MLN formula templates

$\text{prep\_before}(e_1, t_1) \Rightarrow \text{after}(\lhd t_1, e_1 \rhd)$

John left[e1] before 6 pm[t1].

$\text{point}(x) \in \{\lhd x, x \rhd\}$

4 MLN formula templates

$\text{prep\_before}(e_1, t_1) \Rightarrow \text{after}(\triangleleft t_1, e_1 \triangleright)$

$\text{prep\_before}(e_1, t_1) \Rightarrow \text{after}(\triangleleft t_1, \triangleleft e_1)$

$\text{prep\_before}(e_1, t_1) \Rightarrow \text{after}(t_1 \triangleright, e_1 \triangleright)$

……

John left^{e1} before 6 pm^{t1}.

$\text{point}(x) \in \{\triangleleft x, x \triangleright\}$

4 MLN formula templates

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

value(t_1, “future”) ^ tense(e_1, “past”) => after($\triangleleft t_1, e_1 \triangleright$)

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

point(x) $\in \{\triangleleft x, x \triangleright\}$

4 MLN formula templates

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

$\text{point}(x) \in \{ {}^{\triangleleft}x, x^{\triangleright} \}$

4 MLN formula templates

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

$\text{point}(x) \in \{ {}^{\triangleleft}x, x^{\triangleright} \}$

MLN: Learning

- Training set: TimeBank
  - manually labeled news articles
  - 1456 pairs of temporal elements

dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

MLN: Learning

- Training set: TimeBank
  - manually labeled news articles
  - 1456 pairs of temporal elements

Learned weights for: dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

MLN: Learning

- Training set: TimeBank
  - manually labeled news articles
  - 1456 pairs of temporal elements

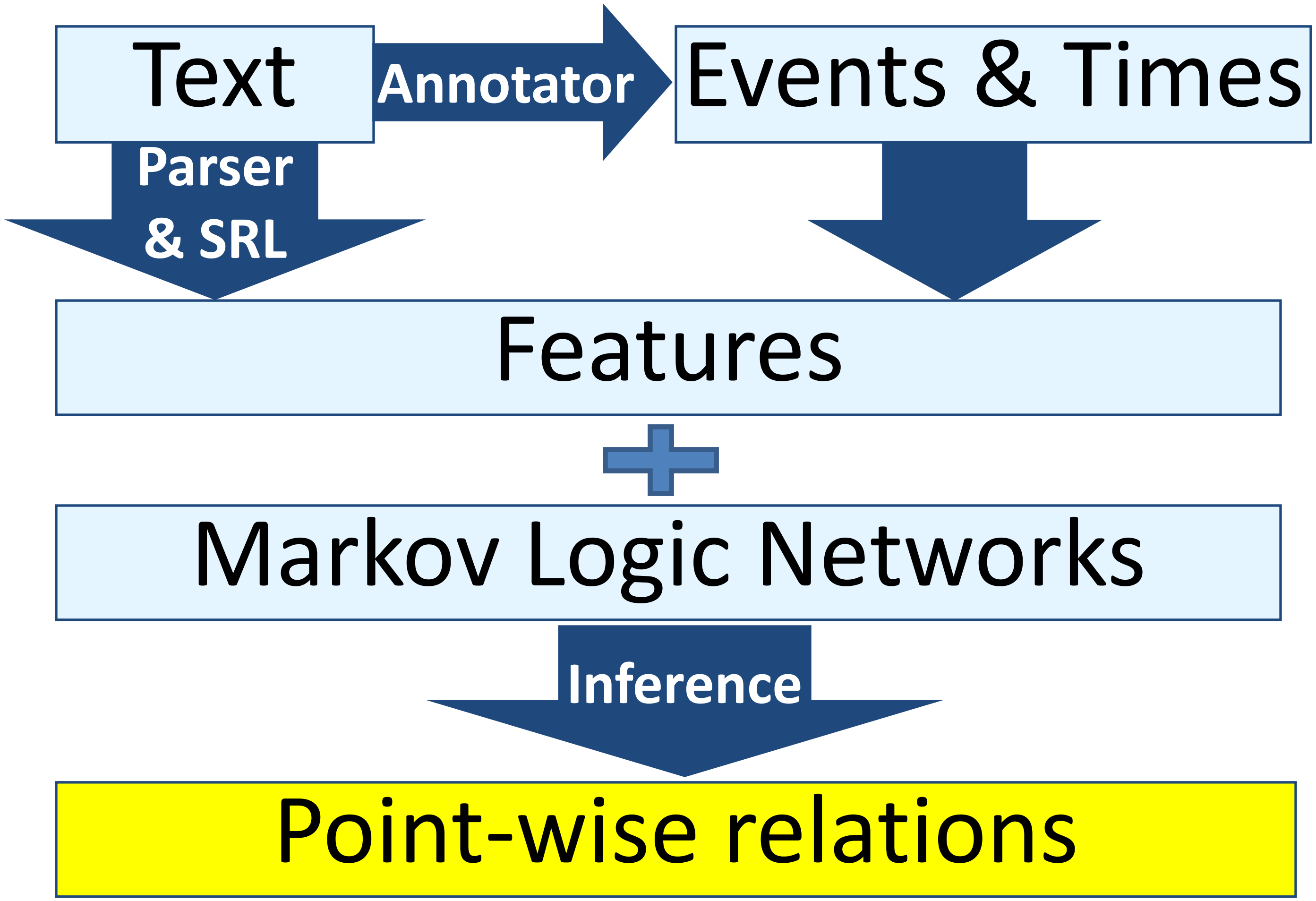
dep(x,y) => after(point(x),point(y))

value(t,+v) ^ tense(e, +s) => after(point(e),point(t))

manually srl_after(p1, p2) => after(p1,p2)

after(p1 ,p2) ^ after(p2, p3) => after(p1, p3)

TIE: System Overview



MLN: Inference

- MC-SAT (Poon et al, 2006):
  - marginal probabilities

    over relations of all possible point pairs
  - predictions

    by thresholding the probabilities

Outline

- Motivation
- Previous Work
- TIE
- Experiments
- Conclusion

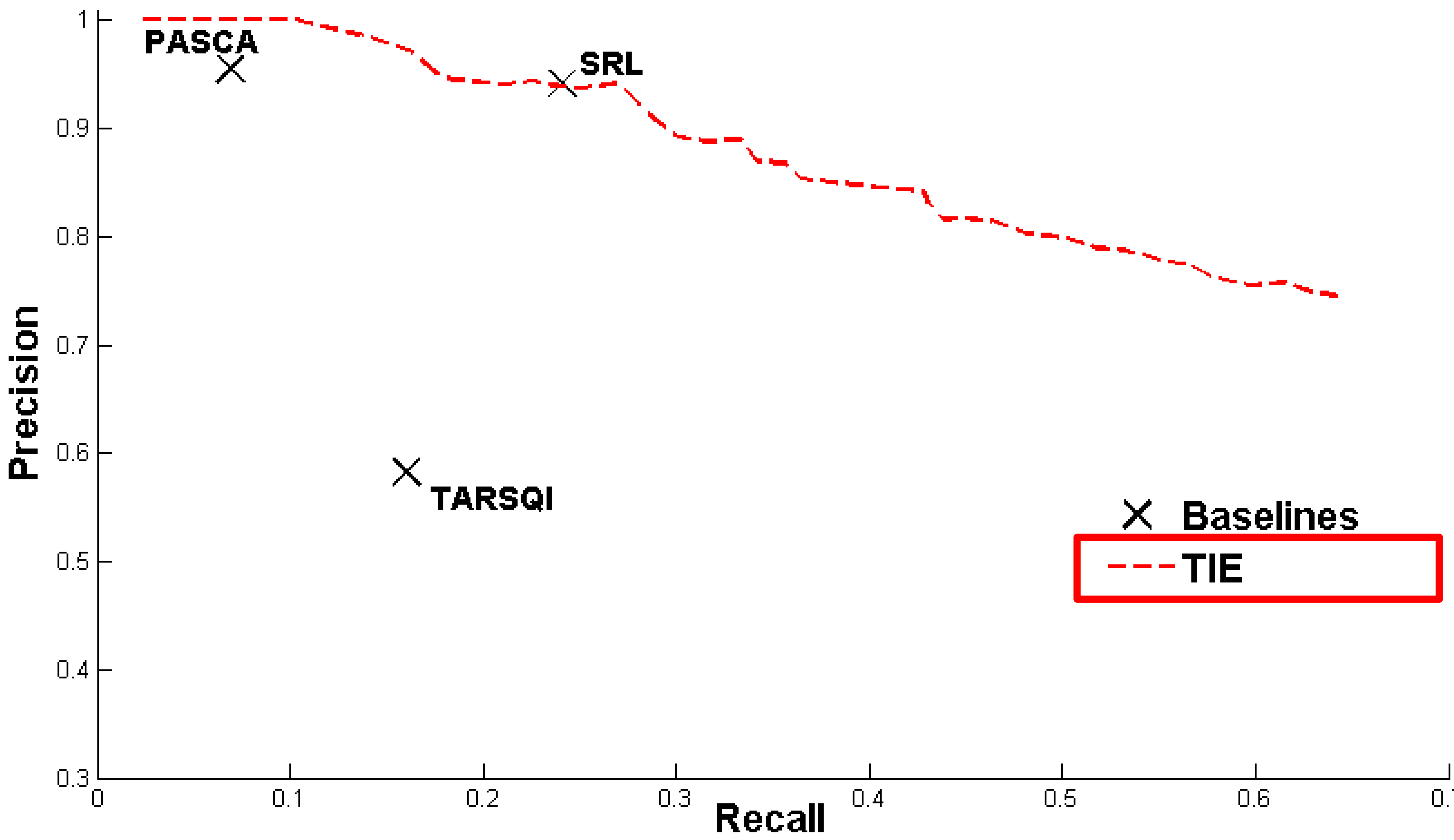
Experiments

- Dataset (From Wikipedia)
  - 45 sentences:
    - 151 events and 56 times
    - 644 point pairs in total
- Labeling all point-wise constraints
  - 2 people and a 3rd person to resolve conflicts

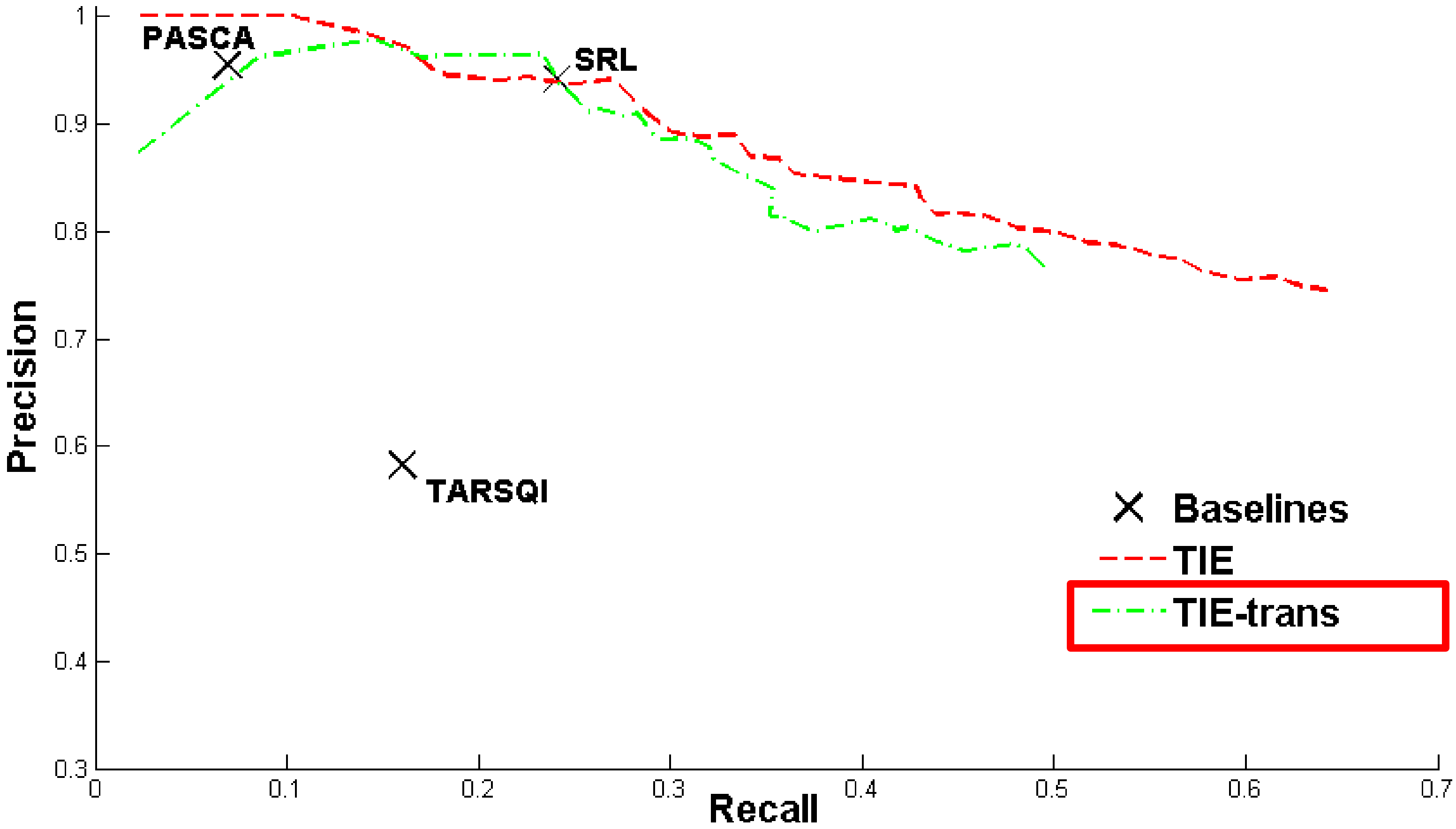
Comparison Systems

- (Pasca, 2008): lexico-syntactic patterns
- TARSQI: hand-code rules + maxent classifier
- SRL: interpreting tmp args based on the preps

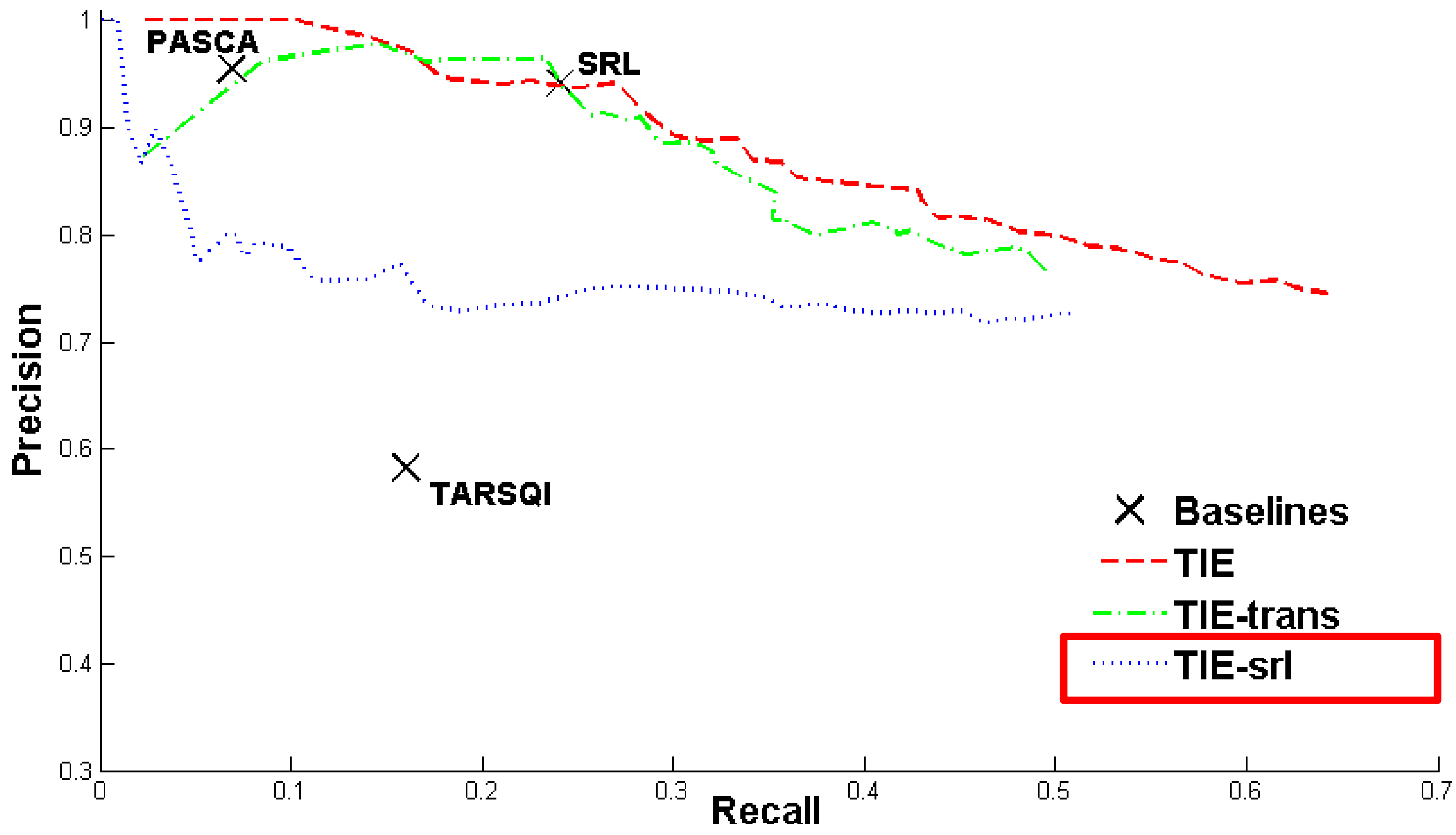
Experiments – PR Curves



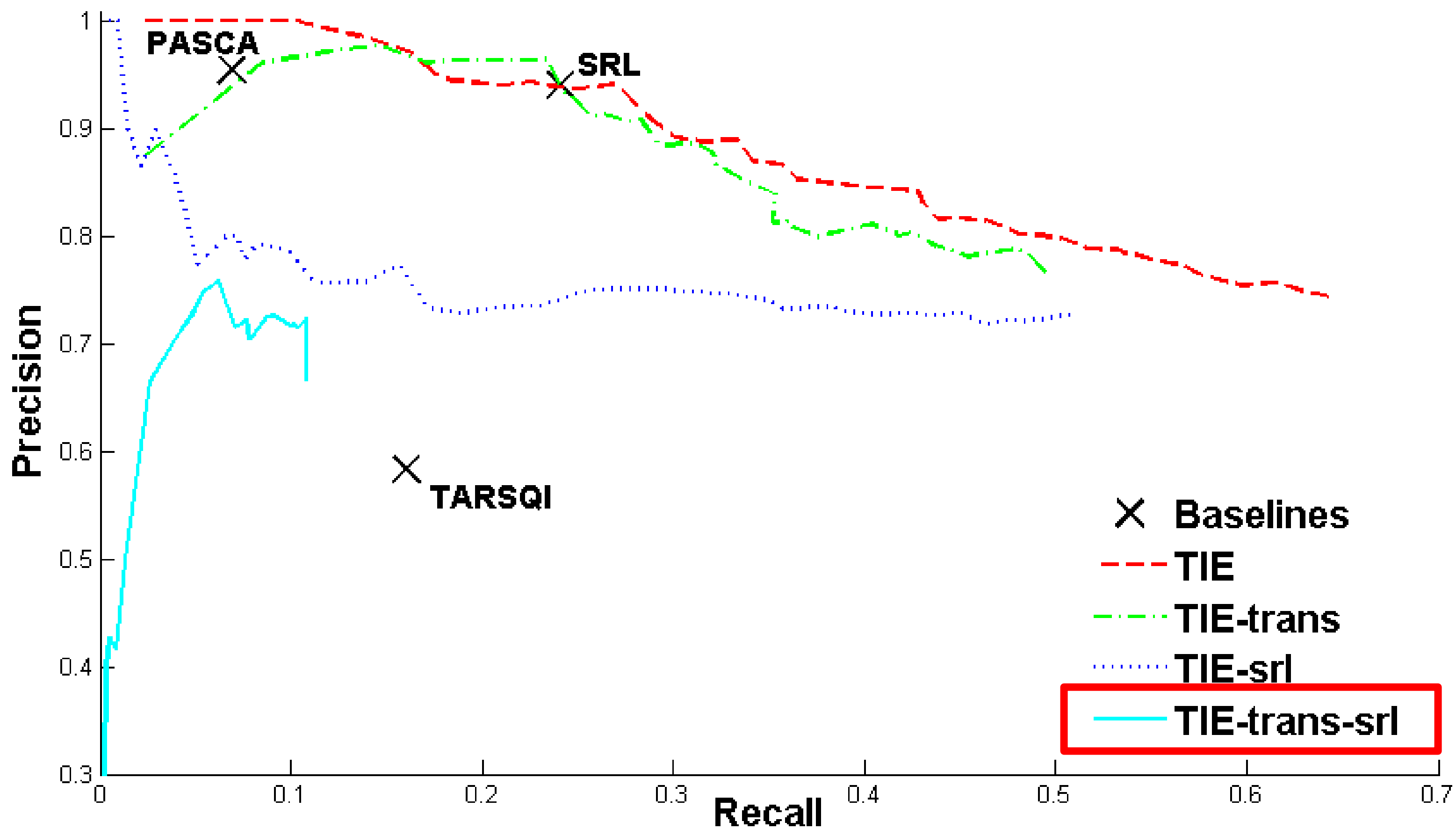
Ablation test: Transitivity [TIE-trans]



Ablation test: SRL [TIE-srl]



Ablation test: Both [TIE-srl-trans]



Temporal Entropy

- PR: are they predicted?

  after(p_1, p_2), after(p_3, p_2), after(p_2, p_4), after(p_1, p_3)

Temporal Entropy

- PR:

  after(p_1, p_2), after(p_3, p_2), after(p_2, p_4), after(p_1, p_3)

  v.s.

  after(p_1, p_2), after(p_3, p_2), after(p_2, p_4), after(p_1, p_3)

Temporal Entropy

- PR:

$after(p_1, p_2), after(p_3, p_2), after(p_2, p_4), after(p_1, p_3)$

$$p_4 < p_2 < p_1$$

v.s.

$after(p_1, p_2), after(p_3, p_2), after(p_2, p_4), after(p_1, p_3)$

$$p_4 < p_2 < p_3$$

Which is tighter?

Temporal Entropy

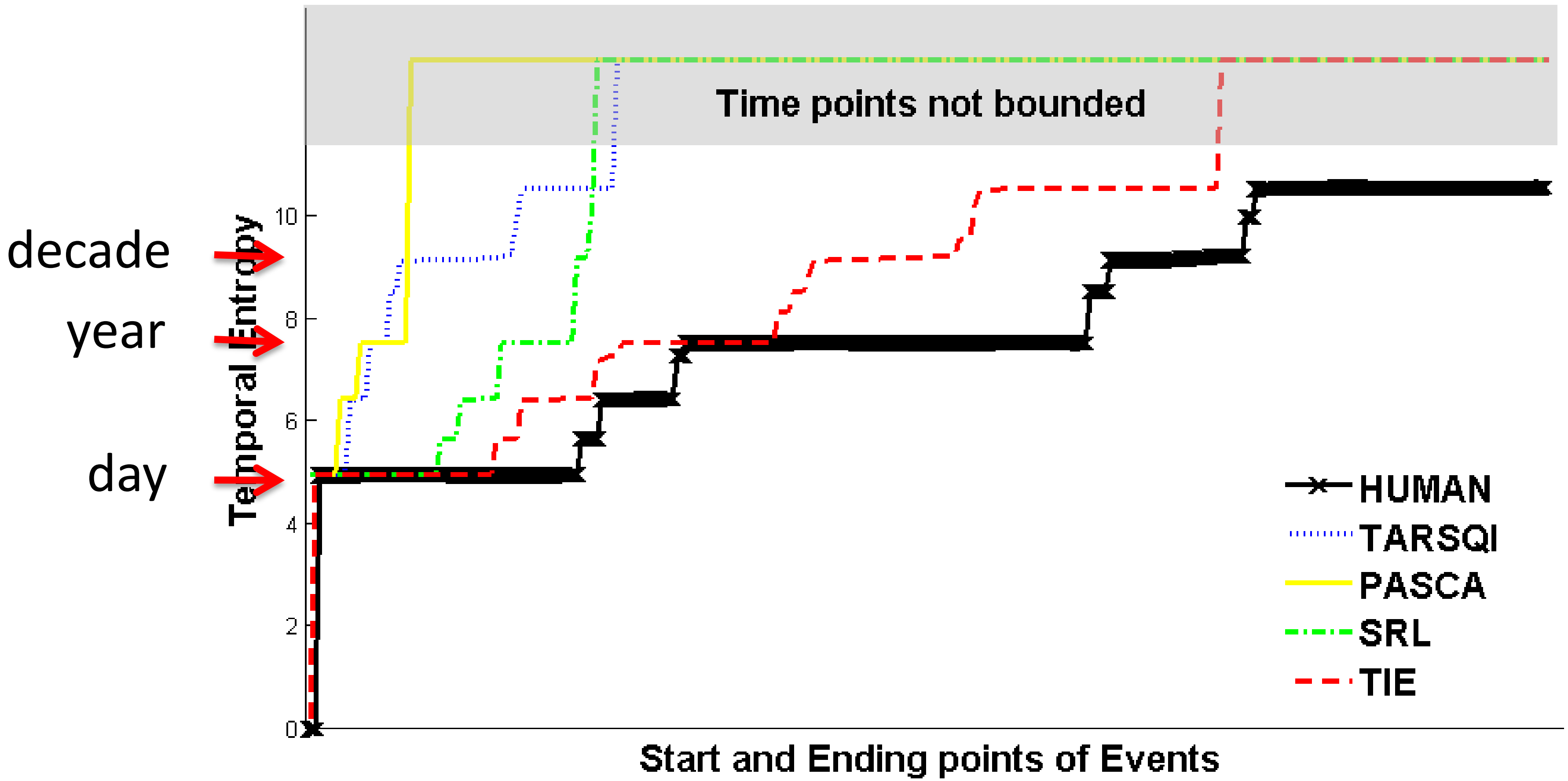
$$p \in [p^L, p^U]$$

◁1931 < ◁recovery < 1933 ▷

$$TE(p) = \log(p^U - p^L)$$

TE (◁recovery) = log(3 years in seconds)

Temporal Entropy



Conclusion

- TIE
  - **Input**: raw text
  - **Output**: events annotated with bounds on endpoints
- Exploits transitivity & high-level features
- Outperforms alternative state-of-the-art systems
- Temporal Entropy

  a new measure for tightness of the bounds

Thanks! Questions?

Future directions

- Improve Event-Event predictions
  - hard to predict w/o knowing the semantics

    e.g. The meeting has been cut off for two years.

    The meeting has been running for two years.
- Inter-sentence inference
  - e.g. adjacent sentence transitions, event coref, etc.