

# Co-dependent Information Extraction

David McClosky  
IBM Research

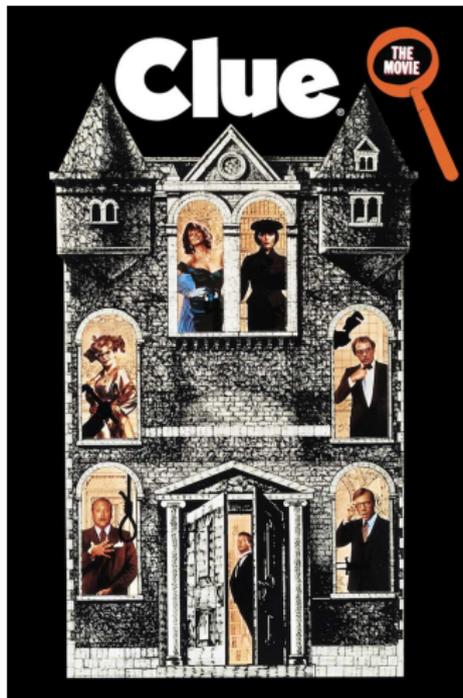
Joint work with Chris Manning and Mihai Surdeanu



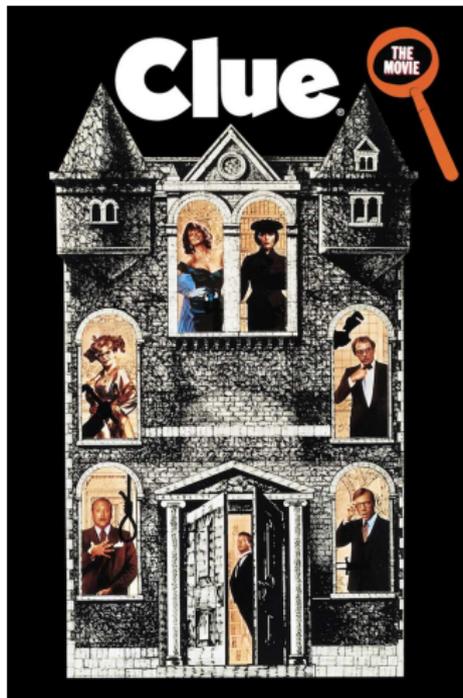
4.15.2013



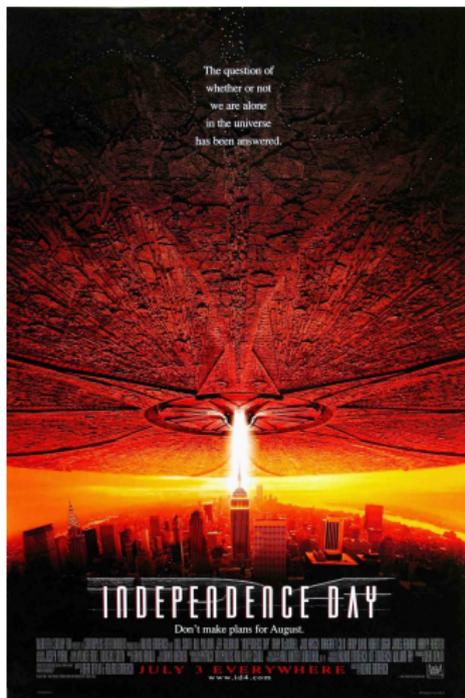
“A talk should not be a mystery...”



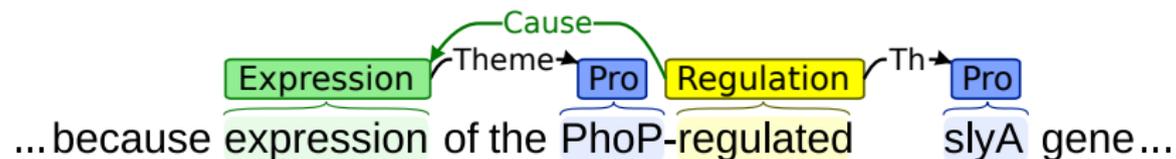
More like an action movie...



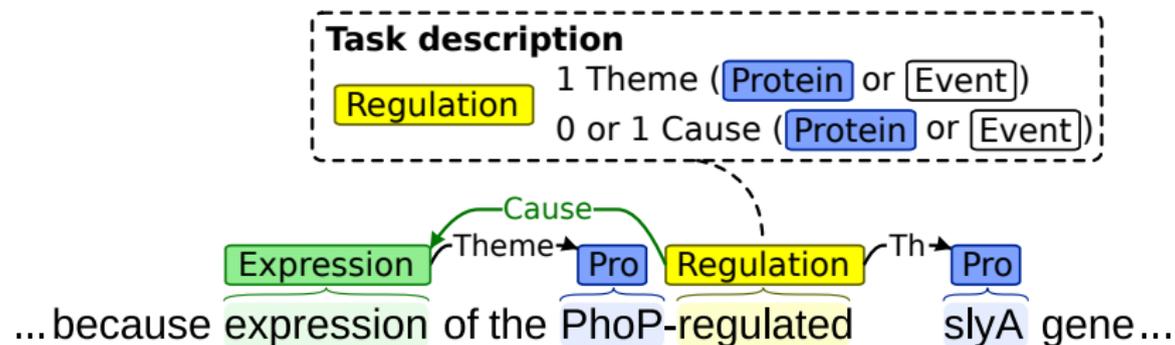
# Independence Assumptions: The Movie



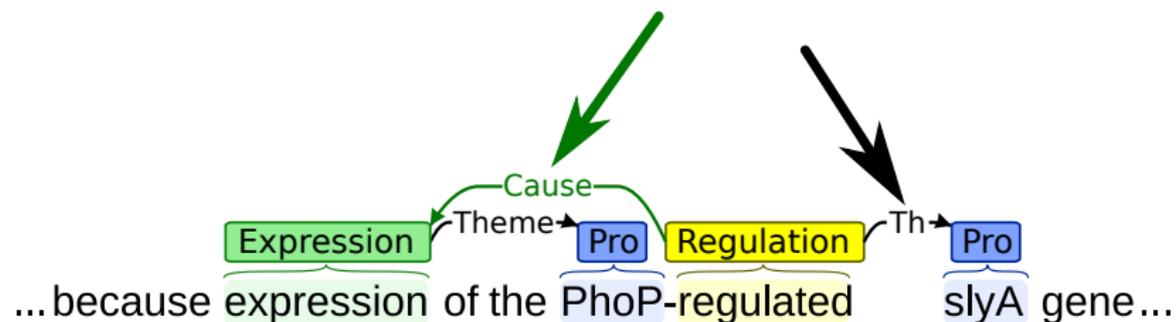
## Independent extractions



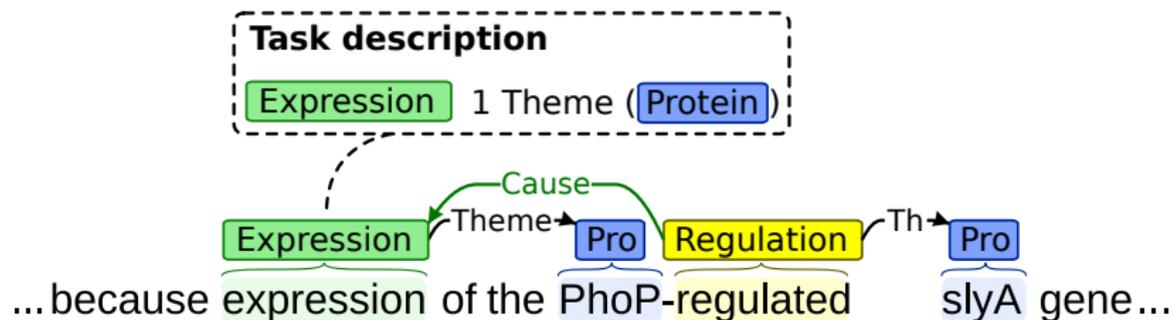
## Independent extractions



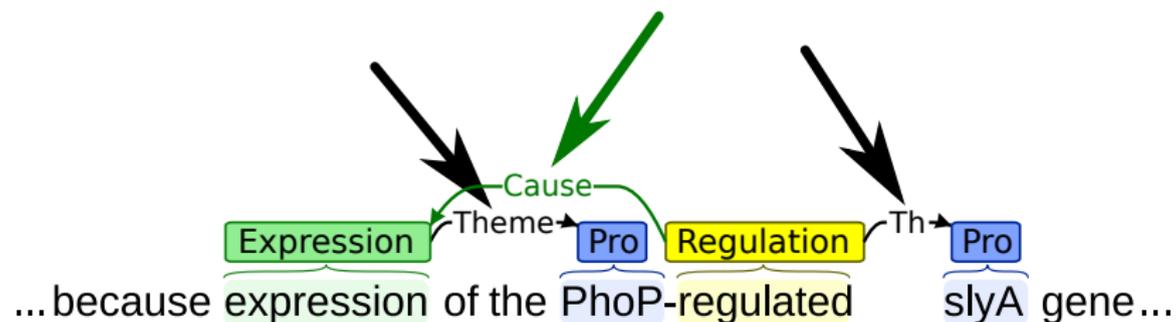
## Decisions, decisions...



## Decisions, decisions...



## Decisions, decisions...



## Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues
  - ▶ Model tractability issues

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues
  - ▶ Model tractability issues
- ▶ Use simpler models:  
(pipelined systems, independence assumptions)

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues
  - ▶ Model tractability issues
- ▶ Use simpler models:  
(pipelined systems, independence assumptions)
  - ▶ Fast, tractable inference

# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues
  - ▶ Model tractability issues
- ▶ Use simpler models:  
(pipelined systems, independence assumptions)
  - ▶ Fast, tractable inference
  - ▶ Easier to train (less data sparsity)

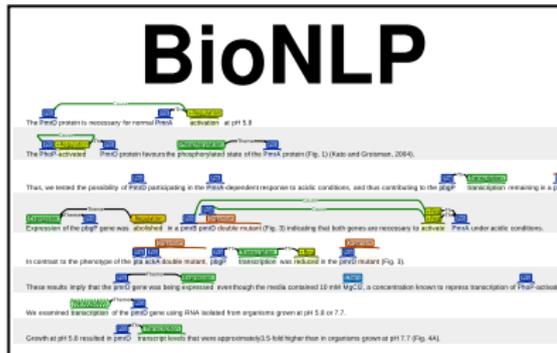
# Modeling domain semantics

- ▶ Modeling complete semantics probably AI-complete
  - ▶ Model design issues
  - ▶ Data availability issues
  - ▶ Model tractability issues
- ▶ Use simpler models:  
(pipelined systems, independence assumptions)
  - ▶ Fast, tractable inference
  - ▶ Easier to train (less data sparsity)
  - ▶ Easier to implement, evaluate

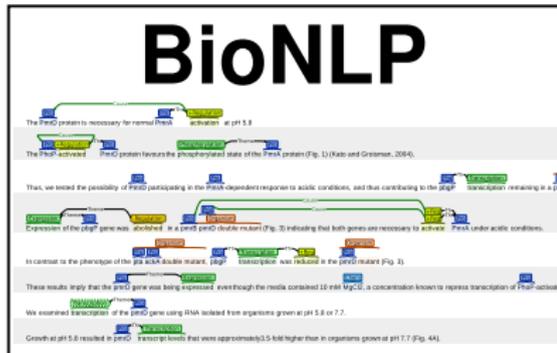
## Related approaches

- ▶ Dual decomposition  
[Riedel and McCallum, EMNLP 2011]
- ▶ Markov logic networks  
[Poon and Vanderwende, NAACL 2010], [Ling and Weld, AAAI 2010]
- ▶ Integer linear programming  
[Chang *et al.*, ACL 2007]
- ▶ Gibbs sampling  
[Finkel *et al.*, ACL 2005]
- ▶ Discriminatively constrained generative models  
[Druck and McCallum, ICML 2010]

# Incorporating domain semantics in information extraction

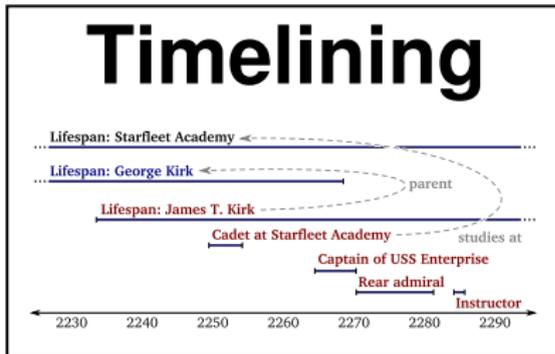
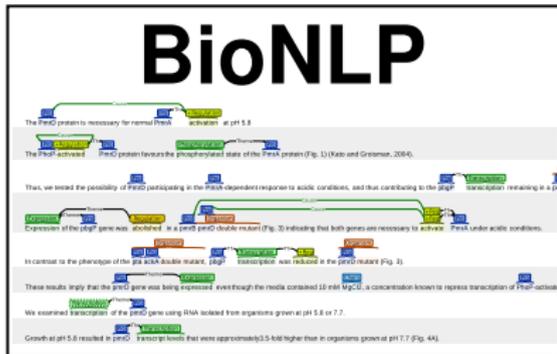


# Incorporating domain semantics in information extraction



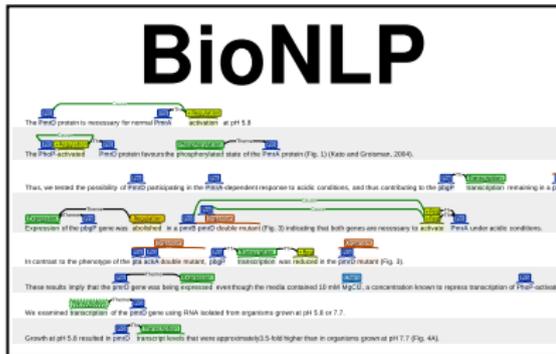
Cast as structured learning problem

# Incorporating domain semantics in information extraction

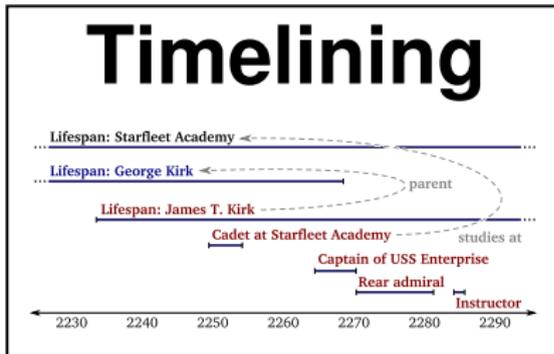


Cast as structured learning problem

# Incorporating domain semantics in information extraction



Cast as structured learning problem



Incorporate global consistency constraints

# Incorporating domain semantics in information extraction

## BioNLP

The PntA protein is necessary for normal PntA activation at pH 5.8

The PntA2 mutation favours the phosphorylated state of the PntA protein (Fig. 1) (Kato and Groisman, 2004).

Thus, we tested the possibility of PntA2 participating in the pH-dependent response to acidic conditions, and thus contributing to the PntA transcription remaining in a PntA state.

Expression of the *pntA2* gene was discovered in a PntA2 double mutant (Fig. 3) indicating that both genes are necessary to activate PntA under acidic conditions.

In contrast to the phenotype of the *pntA* double mutant, *pntA2* transcription was reduced in the PntA mutant (Fig. 3).

These results imply that the PntA gene was being repressed even though the media contained 10 mM bicCO<sub>2</sub>, a concentration known to repress transcription of PntA-activated genes.

We measured transcription of the PntA gene using RNA isolated from organisms grown at pH 5.8 or 7.7.

Growth at pH 5.8 resulted in PntA transcript levels that were approximately 3.5-fold higher than in organisms grown at pH 7.7 (Fig. 6A).

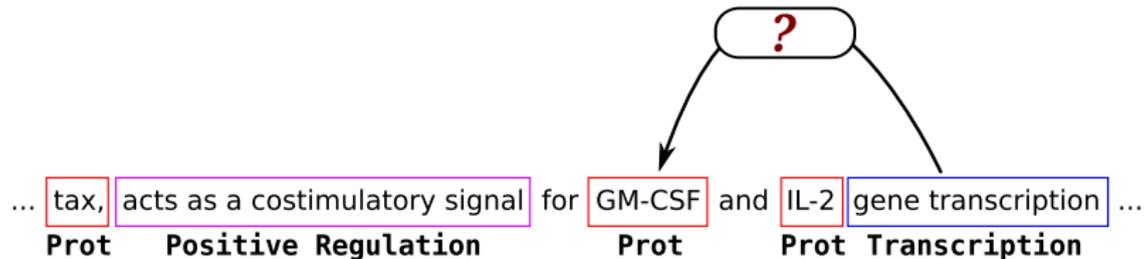
## One independent approach

... tax, acts as a costimulatory signal for GM-CSF and IL-2 gene transcription ...

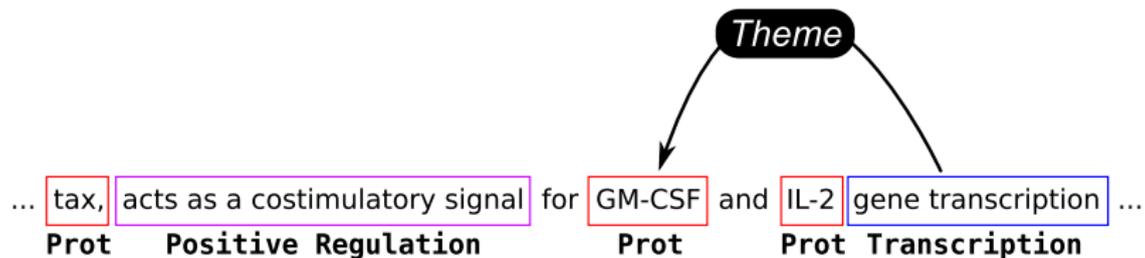
## One independent approach

... **tax,** **acts as a costimulatory signal** for **GM-CSF** and **IL-2** **gene transcription** ...  
**Prot**      **Positive Regulation**                      **Prot**                      **Prot**      **Transcription**

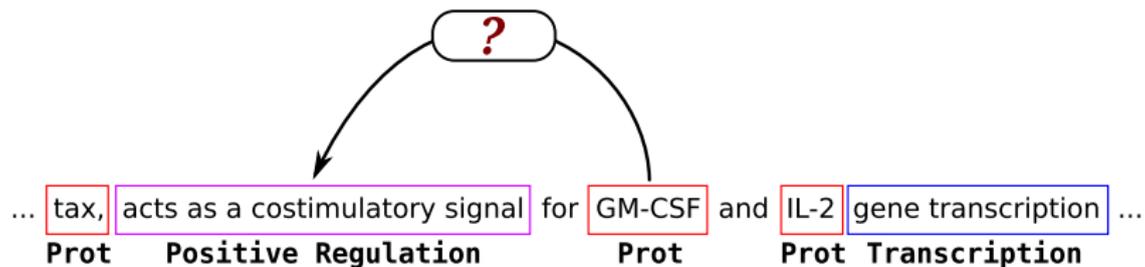
## One independent approach



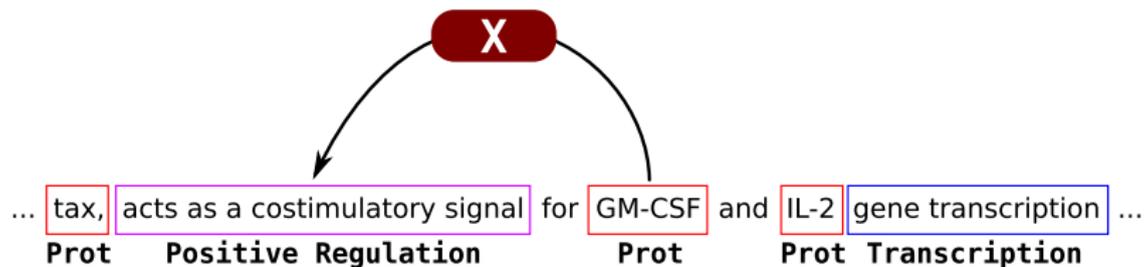
## One independent approach



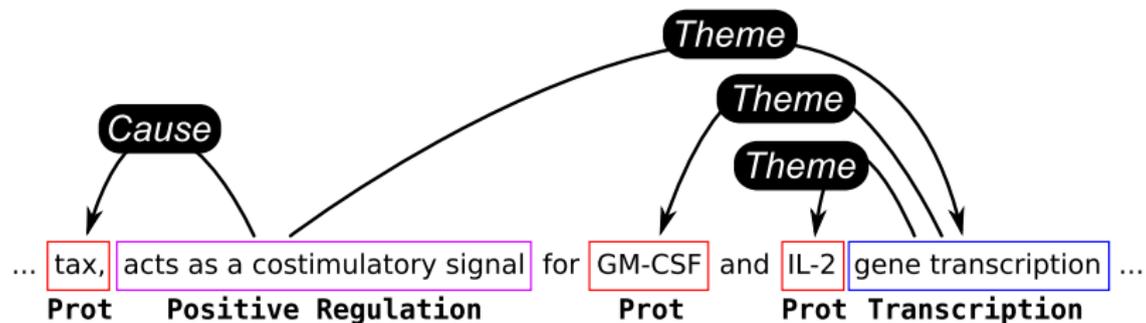
## One independent approach



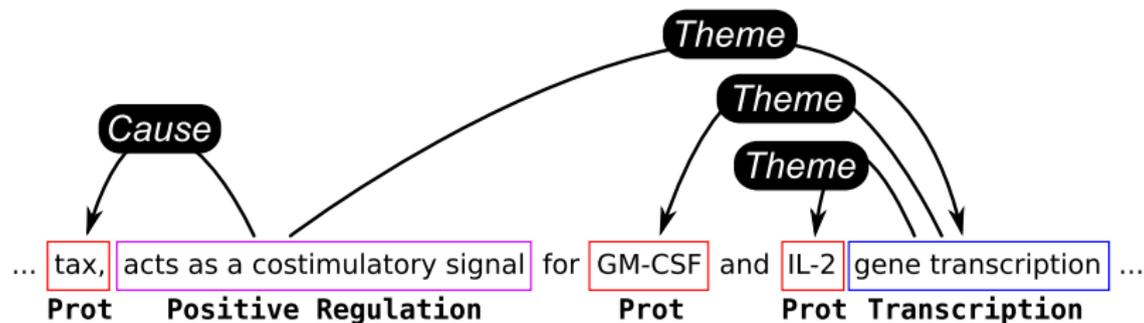
## One independent approach



## One independent approach

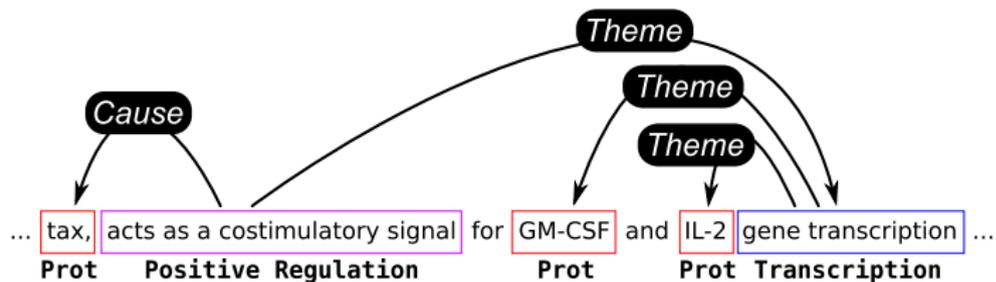


## One independent approach

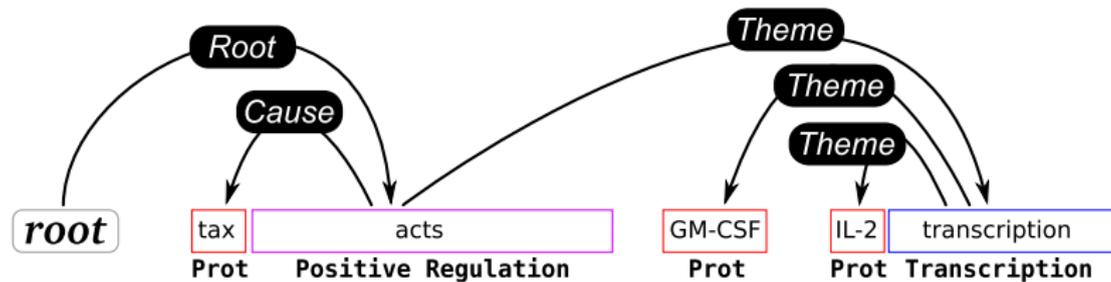


- ▶ UTurku: Björne *et al.* (2009), 1st place in BioNLP 2009

## Event structures → dependency graph

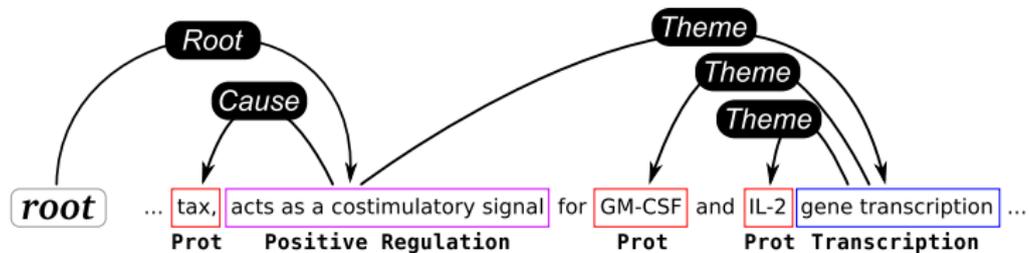


## Event structures → dependency graph



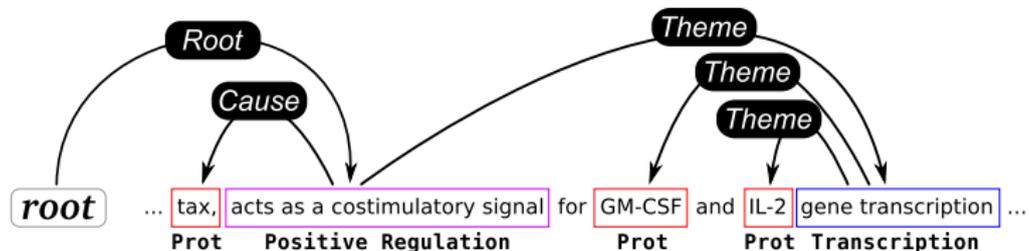
# Parsing event structures

[McDonald *et al.*, EMNLP 2005]



# Parsing event structures

[McDonald *et al.*, EMNLP 2005]



root •

• tax

transcription •

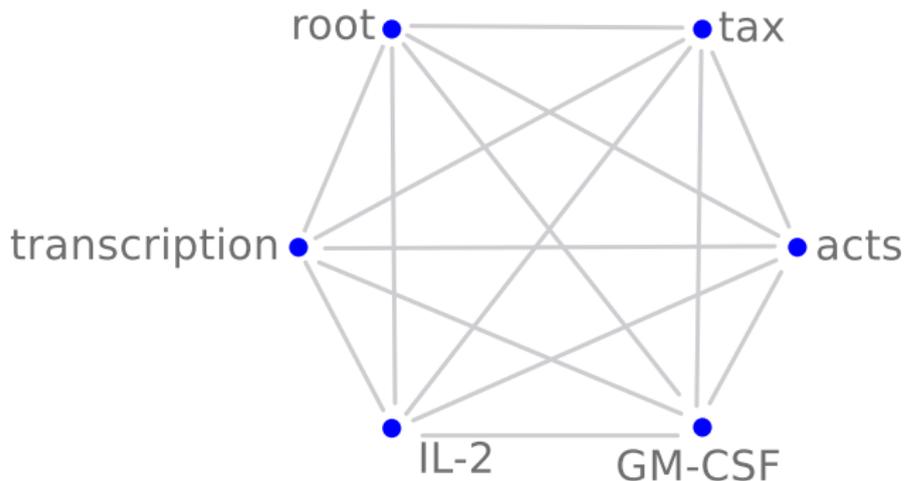
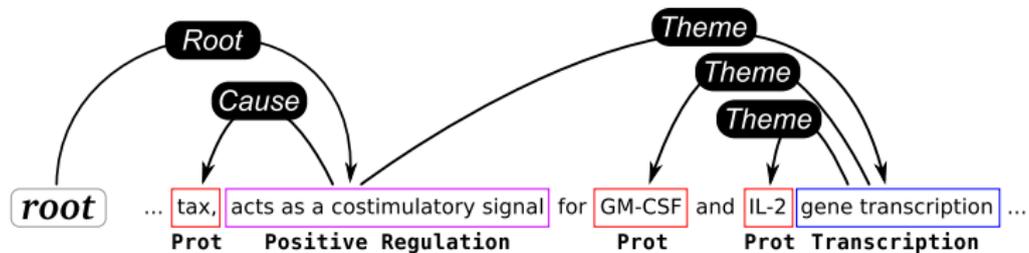
• acts

• IL-2

• GM-CSF

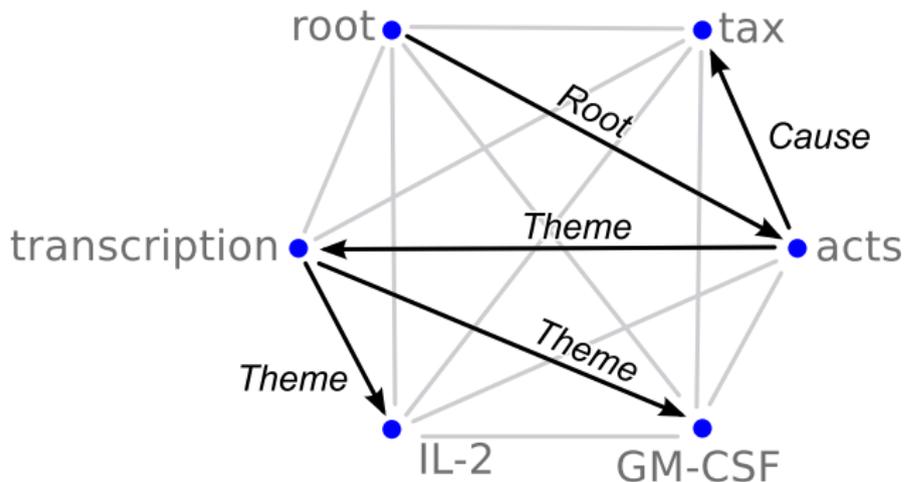
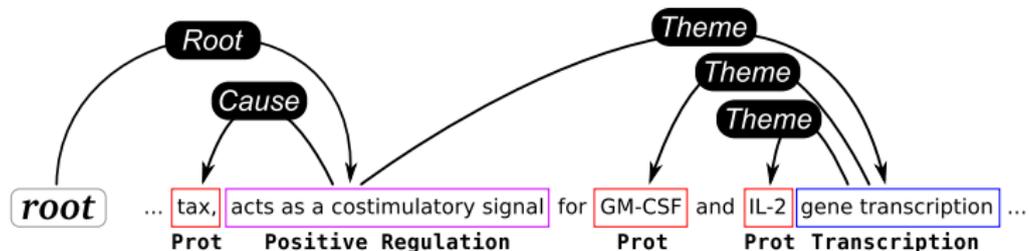
# Parsing event structures

[McDonald *et al.*, EMNLP 2005]



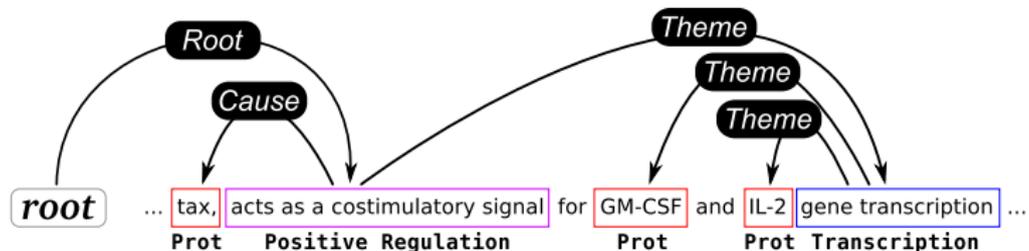
# Parsing event structures

[McDonald *et al.*, EMNLP 2005]

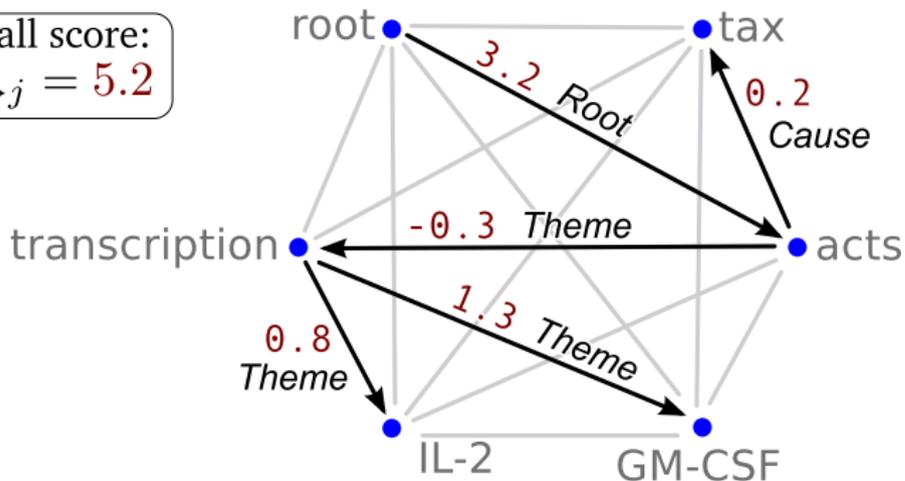


# Parsing event structures

[McDonald *et al.*, EMNLP 2005]

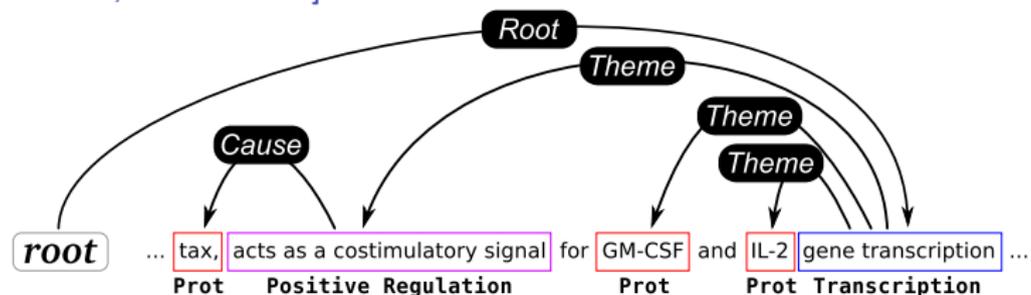


overall score:  
 $\sum w_{i \rightarrow j} = 5.2$

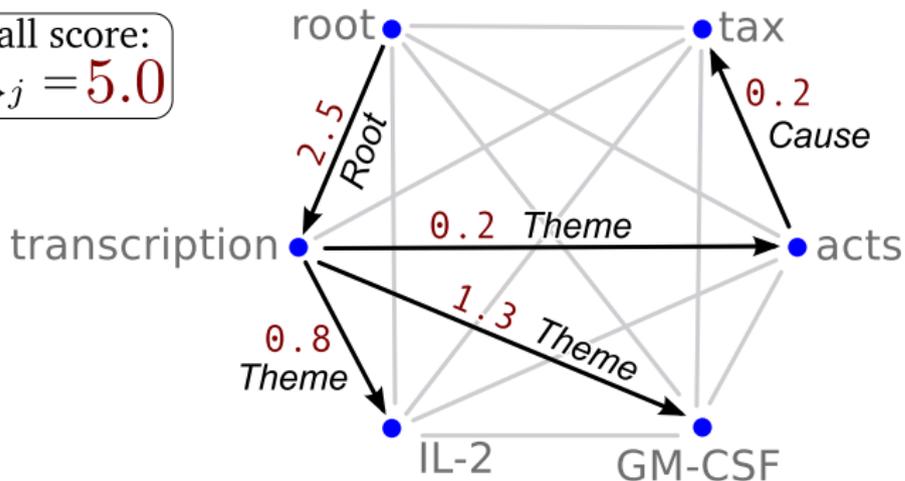


# Parsing event structures

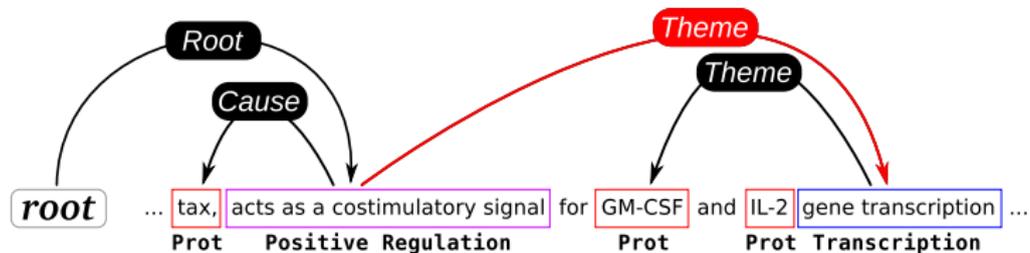
[McDonald *et al.*, EMNLP 2005]



overall score:  
 $\sum w_{i \rightarrow j} = 5.0$

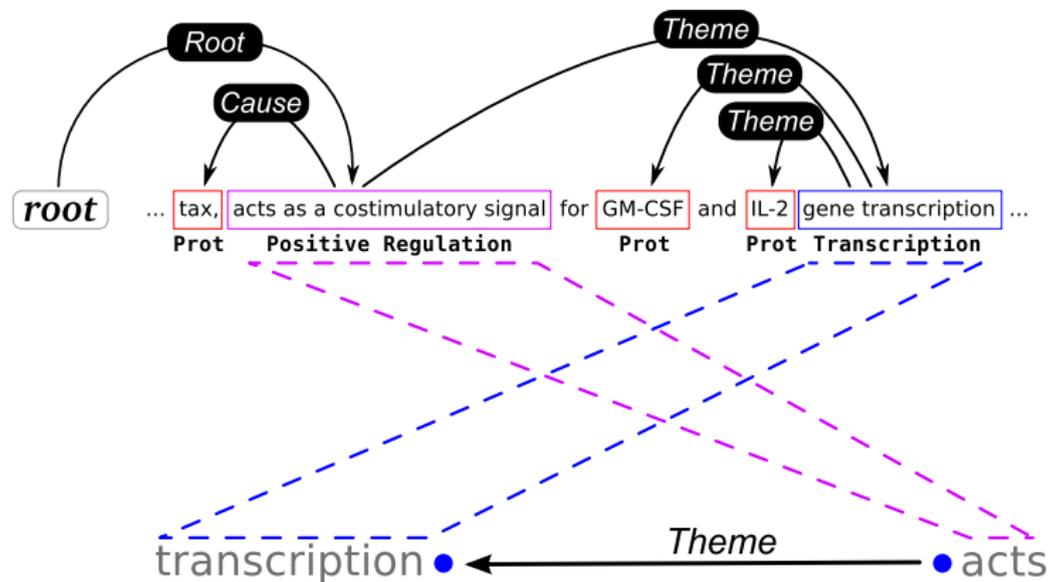


# Features for event parsing

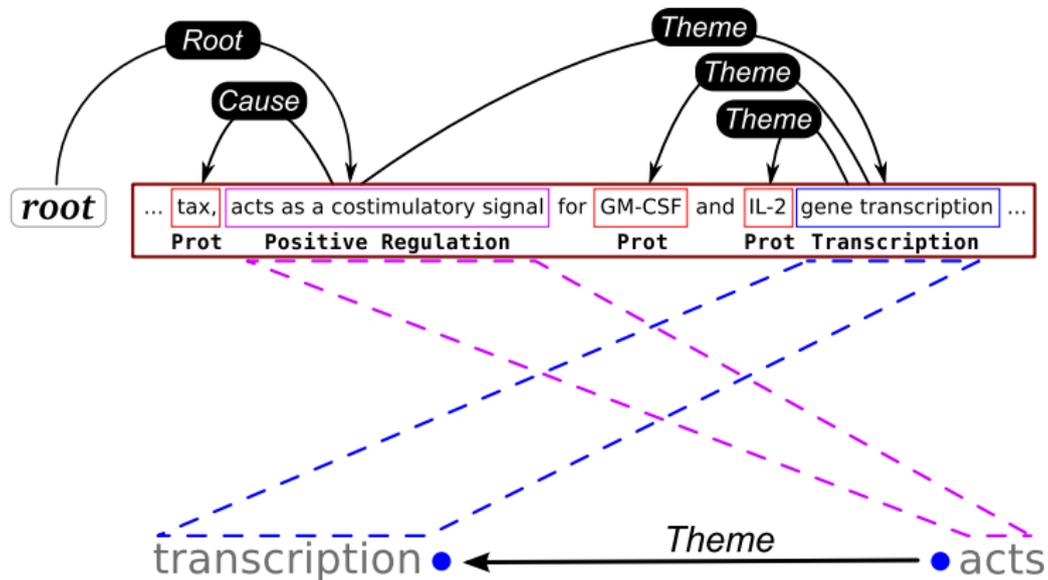


transcription ● ← *Theme* ● acts

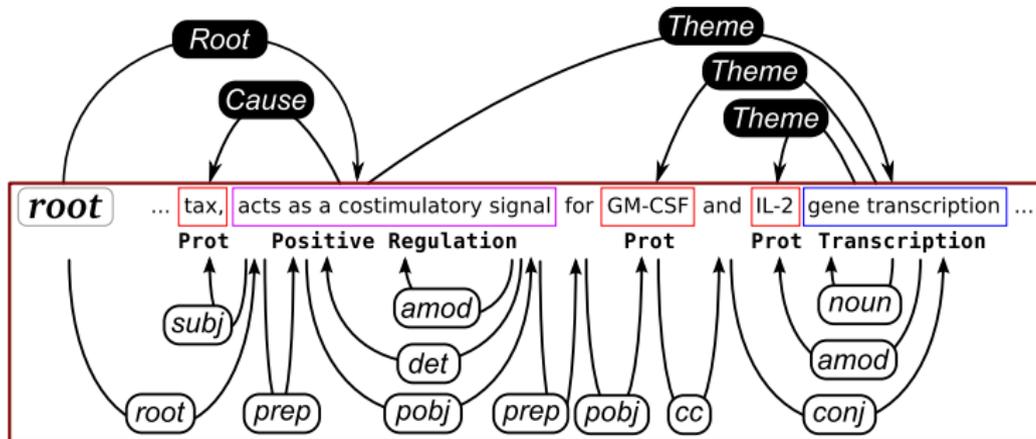
## Features for event parsing



# Features for event parsing

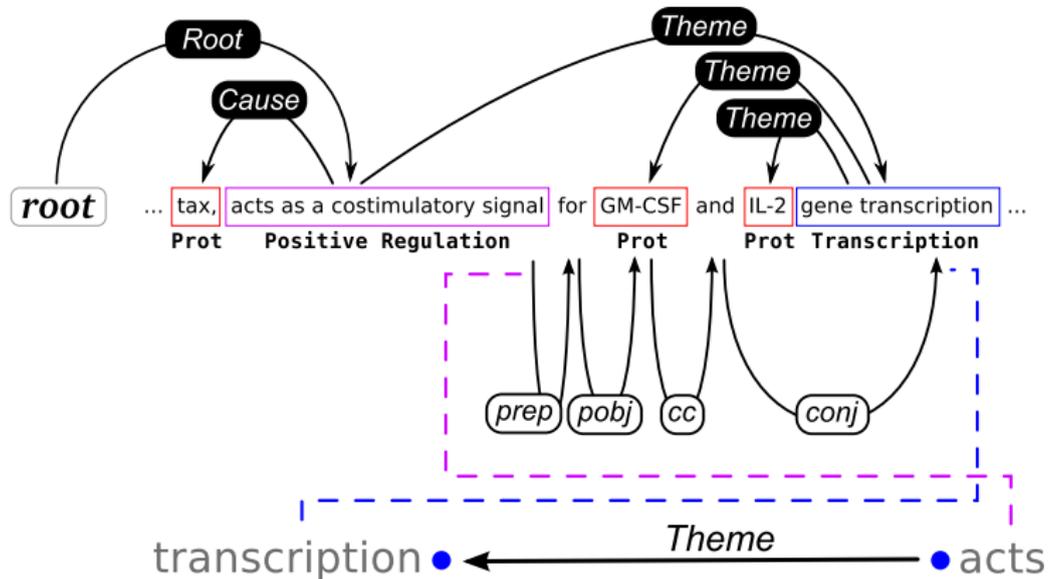


# Features for event parsing

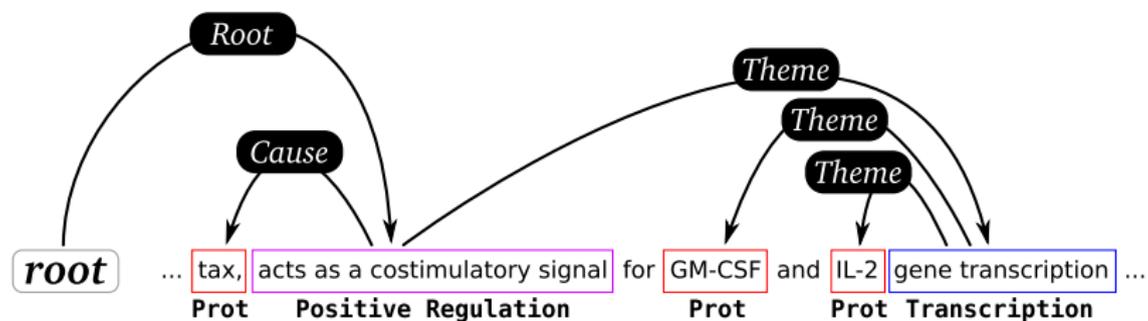


transcription ● ← *Theme* ● acts

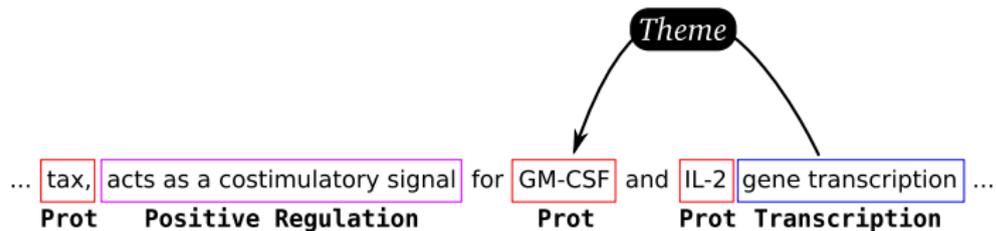
# Features for event parsing



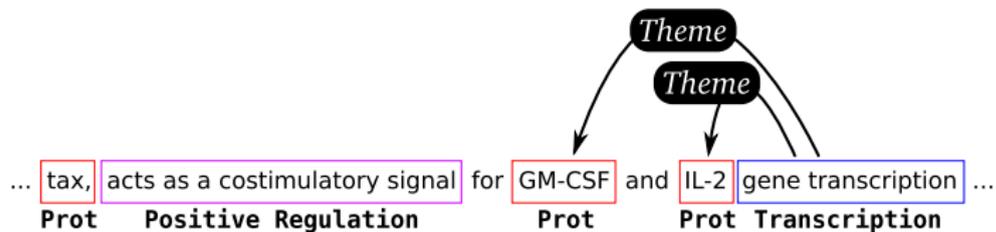
## Edge-factored features



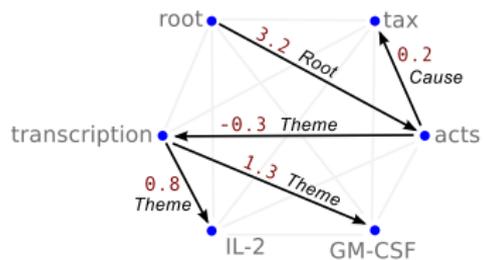
## Edge-factored features



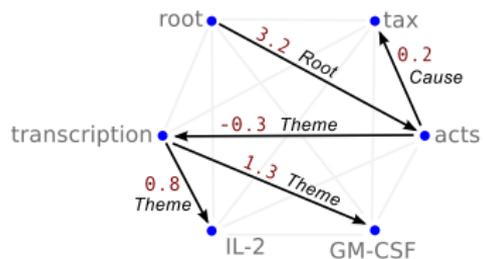
## Second-order edge-factored features



## Reranker allows global features

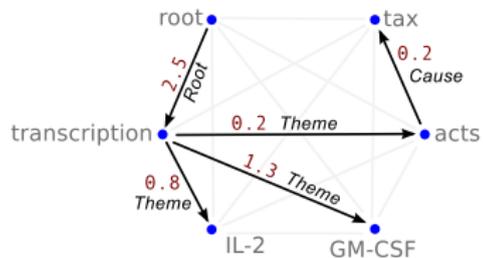


# Reranker allows global features

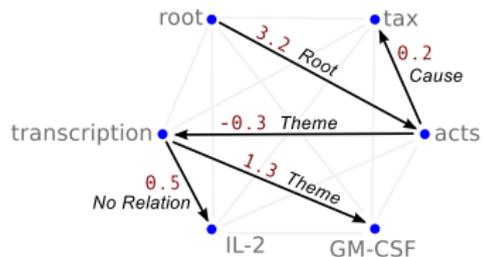


*parser score*

5.2

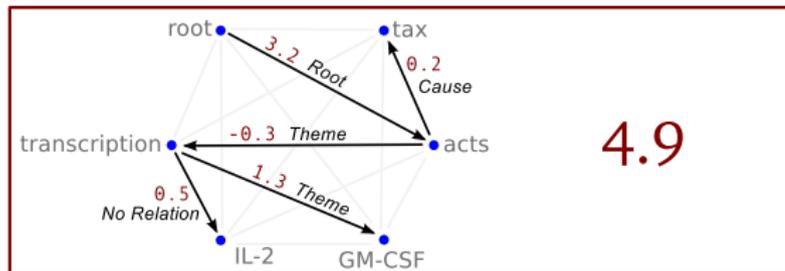
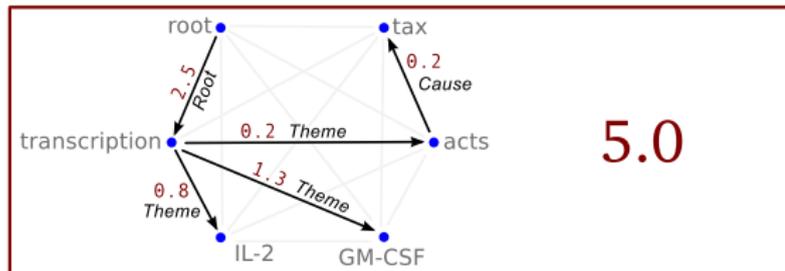
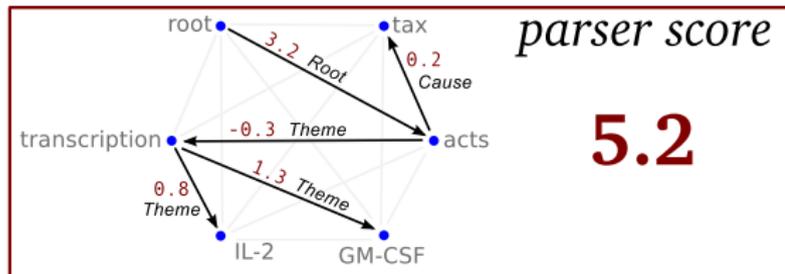


5.0

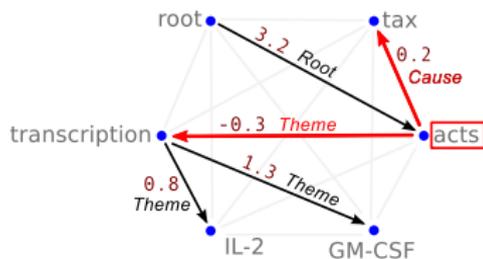


4.9

## Reranker allows global features

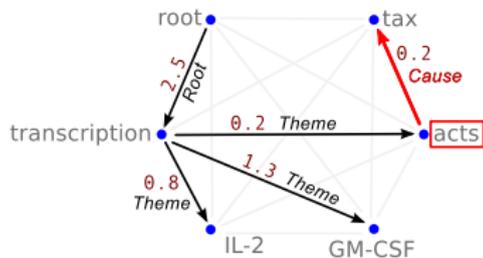


# Reranker allows global features

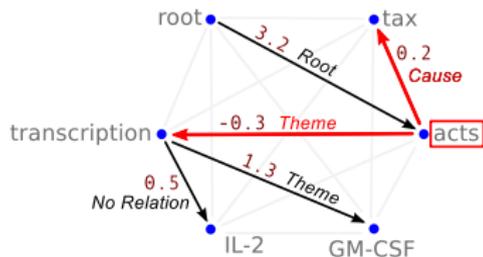


*parser score*

5.2

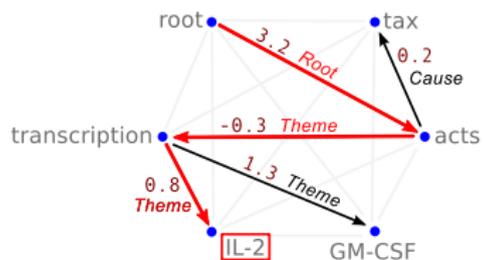


5.0



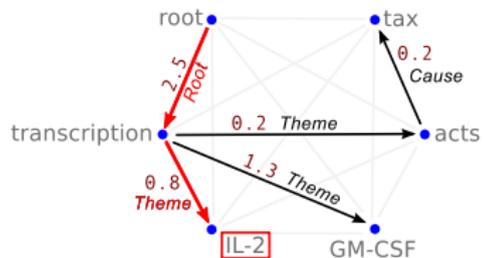
4.9

# Reranker allows global features

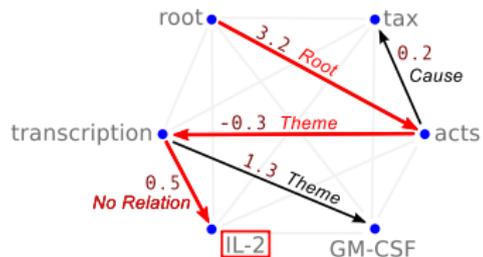


*parser score*

5.2

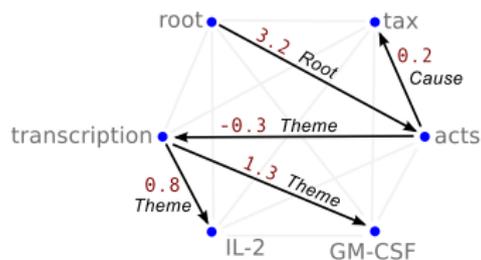


5.0



4.9

## Reranker allows global features

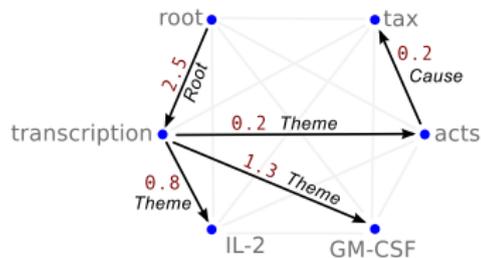


*parser score*

**5.2**

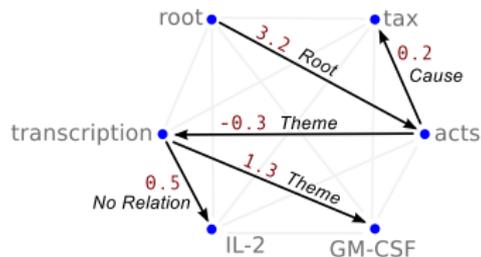
*reranker score*

**-3.87**



**5.0**

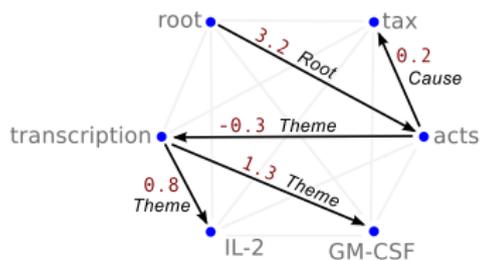
**-3.14**



**4.9**

**-3.44**

## Reranker allows global features

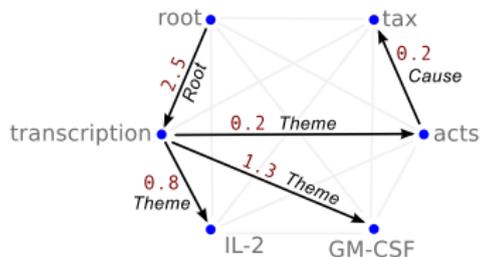


*parser scores*    *reranker score*

*parser 1* **5.2**

*parser 2* **4.6**

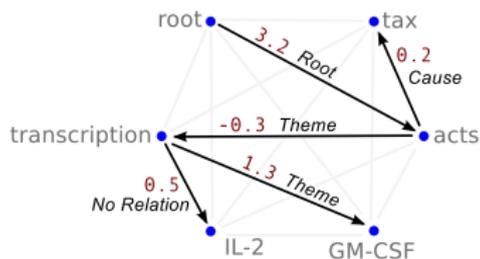
**-3.86**



*parser 1* **5.0**

*parser 2* **4.8**

**-2.71**



*parser 1* **4.9**

*parser 2* **4.2**

**-4.51**

## Event extraction evaluation

| Method             | <i>f</i> -score |
|--------------------|-----------------|
| Event Parser alone | 49.9            |
| <hr/>              |                 |
|                    |                 |

*f*-score on GENIA development section (task 1)

## Event extraction evaluation

| Method                      | <i>f</i> -score |
|-----------------------------|-----------------|
| Event Parser alone          | 49.9            |
| Reranker with single parser | 50.2            |
| -----                       |                 |
|                             |                 |

*f*-score on GENIA development section (task 1)

## Event extraction evaluation

| Method                         | <i>f</i> -score |
|--------------------------------|-----------------|
| Event Parser alone             | 49.9            |
| Reranker with single parser    | 50.2            |
| Reranker with multiple parsers | <b>50.7</b>     |
| -----                          |                 |
|                                |                 |

*f*-score on GENIA development section (task 1)

## Event extraction evaluation

| Method                         | <i>f</i> -score |
|--------------------------------|-----------------|
| Event Parser alone             | 49.9            |
| Reranker with single parser    | 50.2            |
| Reranker with multiple parsers | 50.7            |
| Riedel and McCallum (2011)     | 54.8            |

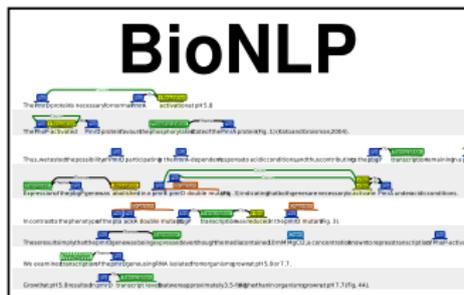
*f*-score on GENIA development section (task 1)

## Event extraction evaluation

| Method   | <i>f</i> -score            |
|--|----------------------------|
| Event Parser alone                                     | 49.9                       |
| Reranker with single parser                            | 50.2                       |
| Reranker with multiple parsers                         | 50.7                       |
| Riedel and McCallum (2011)<br>+ stacking event parsers | <b>54.8</b><br><b>55.9</b> |

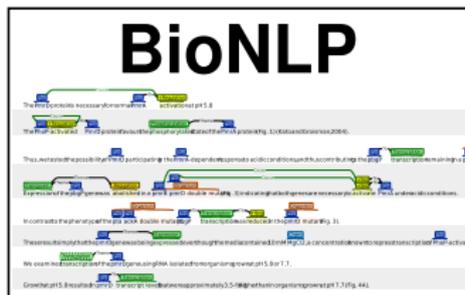
*f*-score on GENIA development section (task 1)

# Act I: Summary



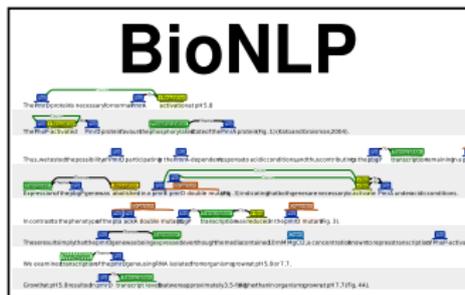
- ▶ Event parsing incorporates domain semantics in two ways

# Act I: Summary



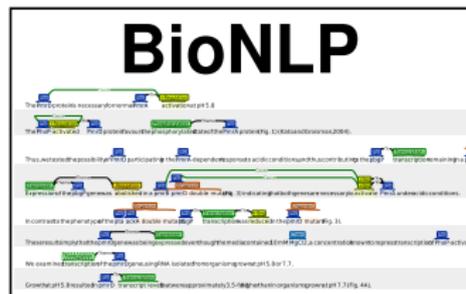
- ▶ Event parsing incorporates domain semantics in two ways:
  1. Parser weighs different extraction decisions

# Act I: Summary



- ▶ Event parsing incorporates domain semantics in two ways:
  1. Parser weighs different extraction decisions
  2. Reranker enables global features

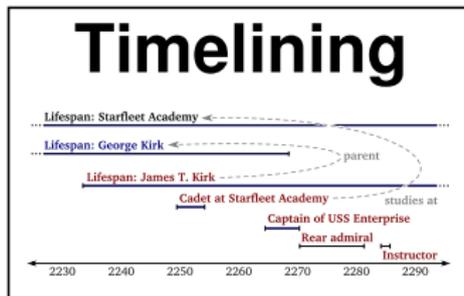
# Act I: Summary



- ▶ Event parsing incorporates domain semantics in two ways:
  1. Parser weighs different extraction decisions
  2. Reranker enables global features
- ▶ Works best with explicit dependencies between extraction items



# Act II: Modeling constraints on implicit dependencies



# Timelining input

[Log in / create account](#)

[Article](#) [Discussion](#)

[Read](#) [Edit](#) [View history](#)

**WIKIPEDIA**  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia

- Interaction
- Help
  - About Wikipedia
  - Community portal
  - Recent changes
  - Contact Wikipedia

- Toolbox
- Print/export
  - Create a book
  - Download as PDF
  - Printable version

- Languages
  - Български
  - Català
  - Česky
  - Dansk
  - Deutsch
  - Español

## James T. Kirk

From Wikipedia, the free encyclopedia  
(Redirected from **James Kirk**)

*"James Kirk" redirects here. For other uses, see [James Kirk \(disambiguation\)](#).*

**James Tiberius "Jim" Kirk** is a character in the *Star Trek* media franchise. Kirk was first played by William Shatner as the principal lead character in the original *Star Trek* series. Shatner voiced Kirk in the animated *Star Trek* series and appeared in the first seven *Star Trek* movies. Chris Pine portrayed a younger version of the character in the 2009 *Star Trek* film, with Jimmy Bennett playing Kirk as a child. Other actors have played the character in fan-created media, and the character has been the subject of multiple spoofs and satires. Kirk also appears in numerous books, comics, and video games. The character has been praised for his leadership traits, and also criticized for his relationships with women.

### Contents

- 1 Depiction
  - 1.1 Alternate timeline
- 2 Development
- 3 Reception
  - 3.1 Cultural impact
  - 3.2 Fan productions
- 4 References
- 5 External links

## Depiction

[[edit](#)]

James T. Kirk was born and raised in [Riverside, Iowa](#) starting in the year 2233.<sup>[1]</sup> [Diane Carey](#)'s novel *Best Destiny* identifies Kirk's parents as [George](#) and [Winona Kirk](#).<sup>[2]</sup> *Best Destiny* and [Carey's Final Frontier](#) novel describe George Kirk's adventures aboard the *USS Enterprise* under the command of Captain [Robert April](#). Although born on Earth, Kirk for a time lived on [Tarsus IV](#), where he was one of nine surviving witnesses to the massacre of 4,000 colonists by [Kodos the Executioner \(Arnold Moss\)](#).<sup>[1]</sup> James Kirk's brother [George Samuel Kirk](#) is first mentioned in "[What Are Little Girls Made Of?](#)" and introduced and killed in "[Operation: Annihilate!](#)", leaving behind three children.<sup>[1]</sup>

**James T. Kirk**

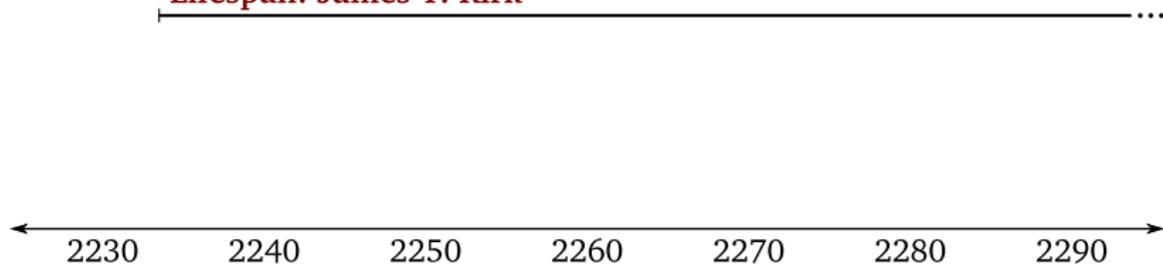
William Shatner as Kirk, in a publicity photograph for the original *Star Trek*

|                    |   |
|--------------------|---|
| <b>Species</b>     | Human                                     |
| <b>Home planet</b> | Earth                                     |
| <b>Affiliation</b> | United Federation of Planets<br>Starfleet |

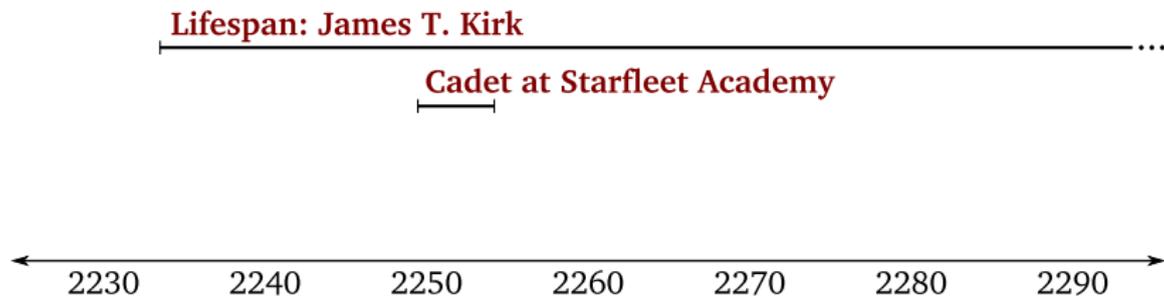
17

## Timelining output

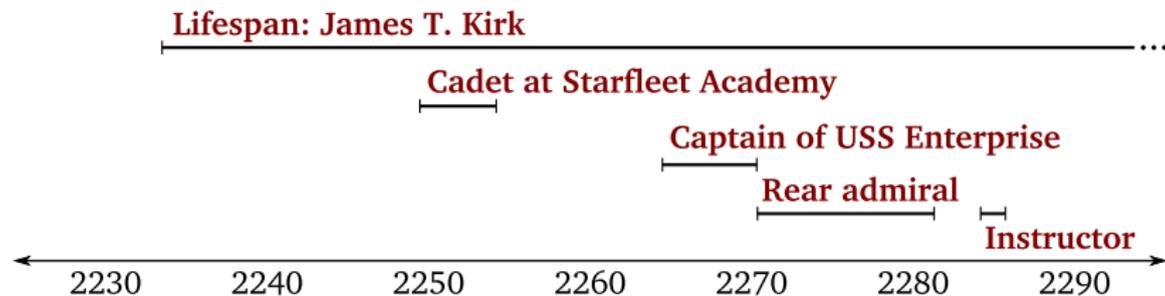
**Lifespan: James T. Kirk**



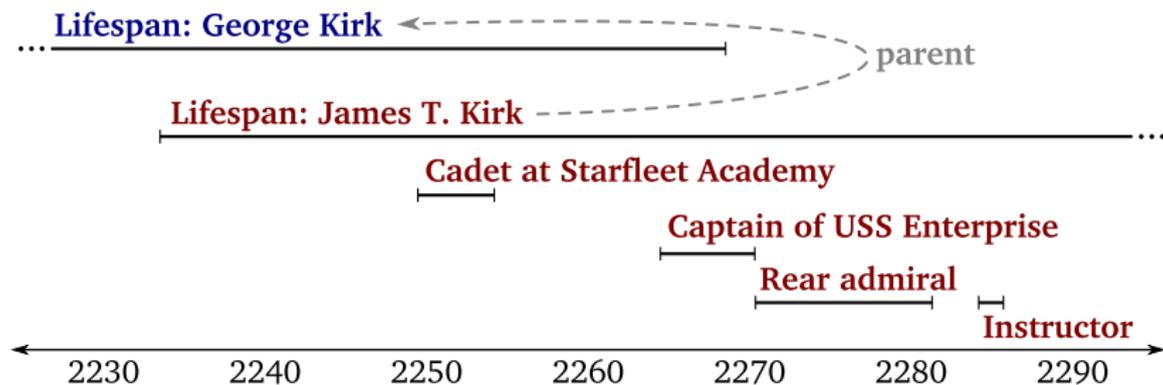
## Timelining output



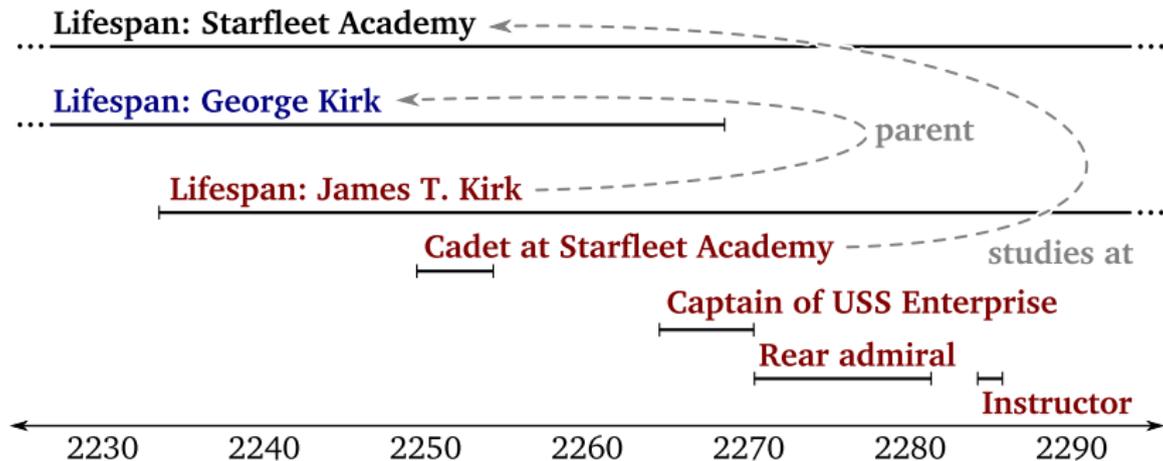
# Timelining output



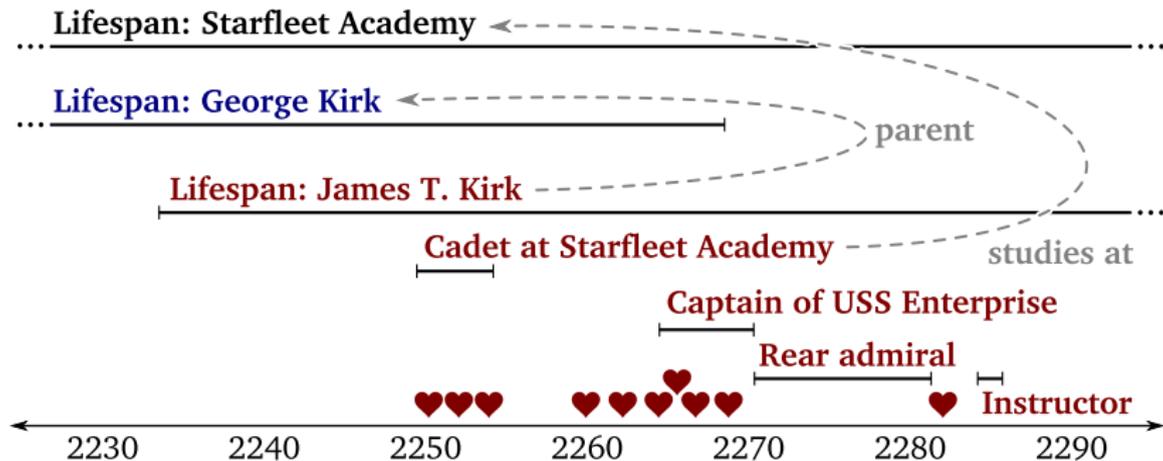
# Timelining output



# Timelining output



# Timelining output



## Task description

### Goal:

Given a set of entities, produce timelines of significant events

### Given:

- ▶ Example timelines (*Wikipedia infoboxes, Freebase entries*)
- ▶ Relations between entities (*Wikipedia infoboxes*)
- ▶ Lots of text (*newswire, blogs, Wikipedia articles, etc.*)

## Example: Kirk starts attending Starfleet

### Kirk article excerpt

#### **Starfleet Academy**

By 2250, **Kirk** returned to Earth to start his training at Starfleet Academy in San Francisco, thanks to some assistance from a man named Mallory. He later credited his father as his inspiration for joining Starfleet.

## Example: Kirk starts attending Starfleet

### Kirk article excerpt

#### Starfleet Academy

By 2250, **Kirk** returned to Earth to **start his training at Starfleet Academy** in San Francisco, thanks to some assistance from a man named Mallory. He later credited his father as his inspiration for joining Starfleet.

Given:

- ▶ relation  $r_1 = \textit{attendsSchool}(\textit{James T. Kirk}, \textit{Starfleet Academy})$

## Example: Kirk starts attending Starfleet

### Kirk article excerpt

#### Starfleet Academy

By 2250, Kirk returned to Earth to **start his training at Starfleet Academy** in San Francisco, thanks to some assistance from a man named Mallory. He later credited his father as his inspiration for joining Starfleet.

Given:

- ▶ relation  $r_1 = \textit{attendsSchool}(\textit{James T. Kirk}, \textit{Starfleet Academy})$
- ▶ time  $t_1 = 2250$

## Example: Kirk starts attending Starfleet

### Kirk article excerpt

#### Starfleet Academy

By 2250, Kirk returned to Earth to **start his training at Starfleet Academy** in San Francisco, thanks to some assistance from a man named Mallory. He later credited his father as his inspiration for joining Starfleet.

Given:

- ▶ relation  $r_1 = \textit{attendsSchool}(\textit{James T. Kirk}, \textit{Starfleet Academy})$
- ▶ time  $t_1 = 2250$

Learn temporal bounds:

- ▶  $\textit{classify}(r_1, t_1) = \text{STARTS}$

## Independent classifications

*classify*( $r_1$ ,  $t_1$ )

## Independent classifications

*classify*( $r_1$ ,  $t_1$ )

*classify*( $r_1$ ,  $t_2$ )

## Independent classifications

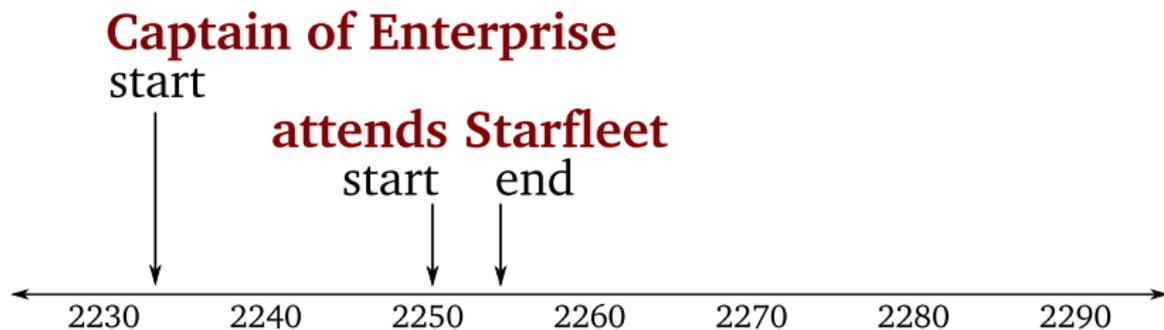
*classify*( $r_1, t_1$ )

*classify*( $r_1, t_2$ )

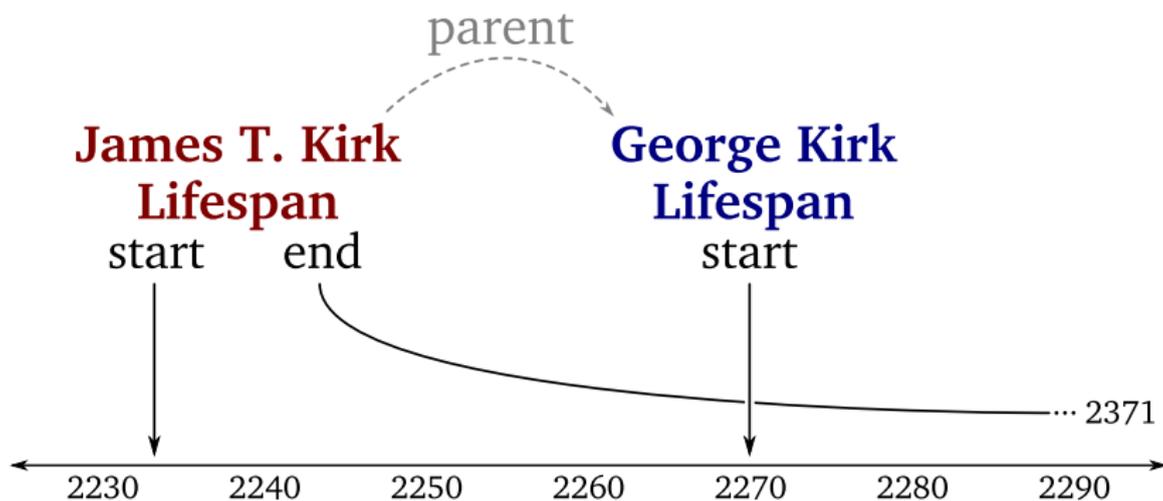
*classify*( $r_2, t_1$ )

⋮

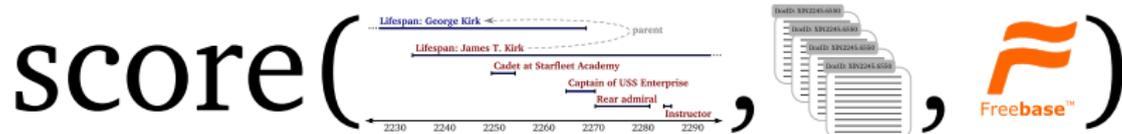
## Inconsistency: multiple relations



## Inconsistency: multiple relations, multiple entities

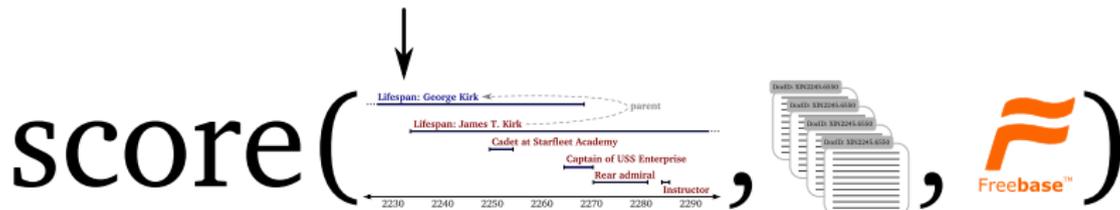


# Model overview



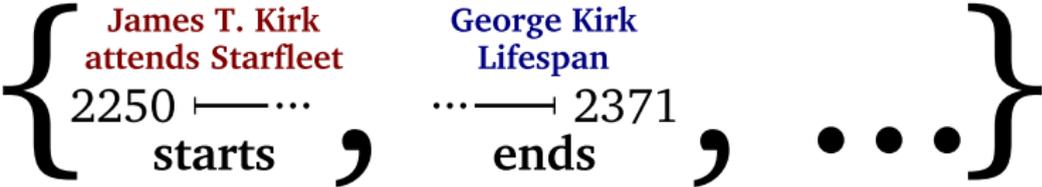
# Model overview

## timeline



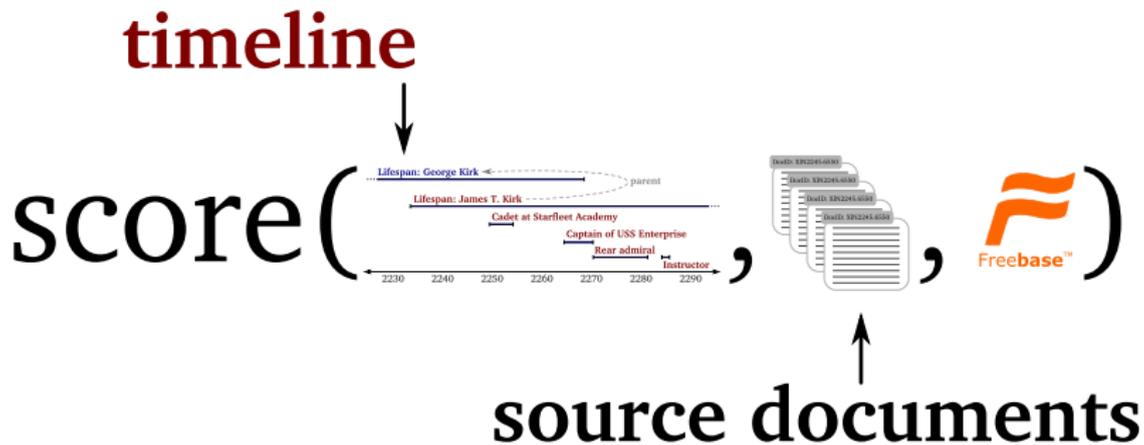
# Model overview

## timeline

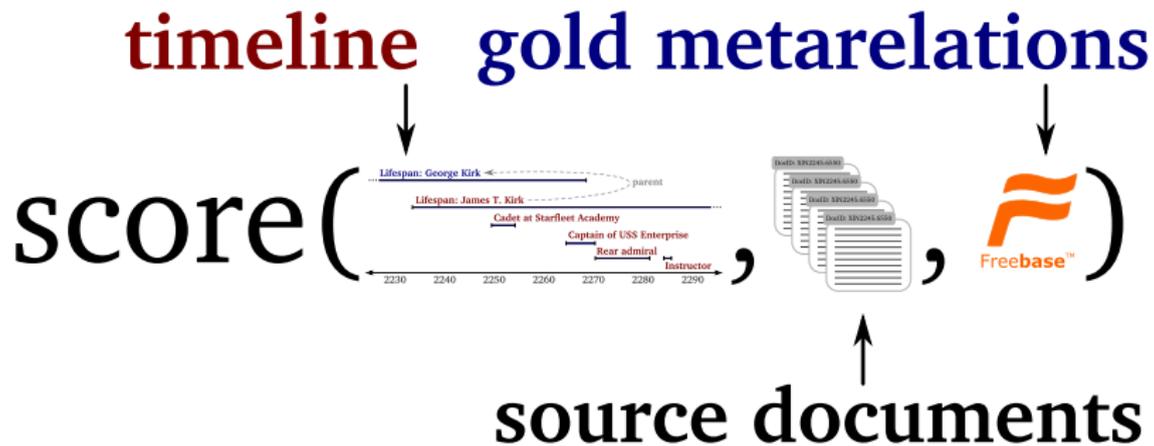


## set of metarelations

# Model overview



## Model overview



# Model overview

$$\text{score} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{DocID: SP12143.0181} \\ \text{DocID: SP12143.0182} \\ \text{DocID: SP12143.0183} \\ \text{DocID: SP12143.0184} \end{array} \right. , \text{Freebase}^{\text{TM}} \right)$$

$$= P_{\text{ext}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{DocID: SP12143.0181} \\ \text{DocID: SP12143.0182} \\ \text{DocID: SP12143.0183} \\ \text{DocID: SP12143.0184} \end{array} \right. \right)$$

$$\times P_{\text{cons}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \text{Freebase}^{\text{TM}} \right. \right)^{\beta}$$

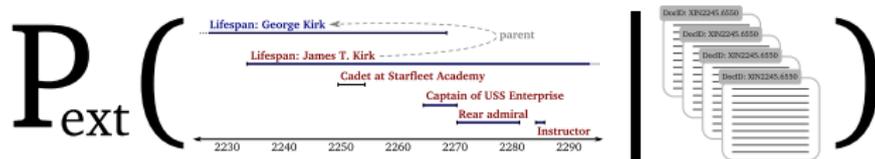
# Model overview

$$\text{score} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \Bigg| \begin{array}{c} \text{DocID: SP12143.0150} \\ \text{DocID: SP12143.0151} \\ \text{DocID: SP12143.0152} \\ \text{DocID: SP12143.0153} \end{array}, \text{Freebase}^{\text{TM}} \right)$$

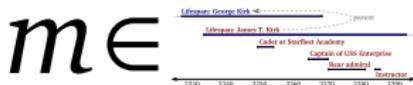
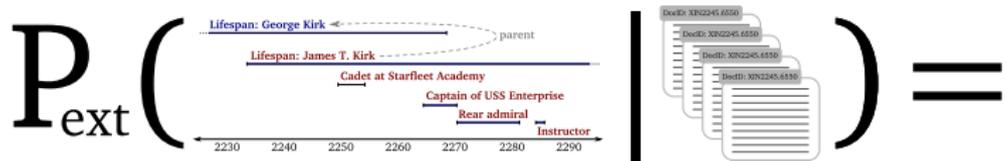
$$= P_{\text{ext}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \Bigg| \begin{array}{c} \text{DocID: SP12143.0150} \\ \text{DocID: SP12143.0151} \\ \text{DocID: SP12143.0152} \\ \text{DocID: SP12143.0153} \end{array} \right)$$

$$\times P_{\text{cons}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \Bigg| \text{Freebase}^{\text{TM}} \right)^{\beta}$$

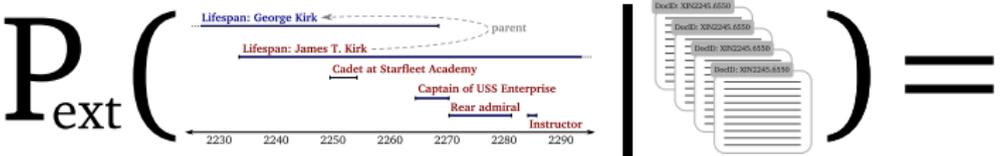
# Extraction model: Timelines



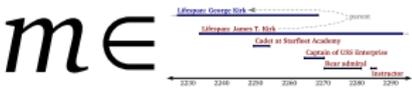
# Extraction model: Timelines



# Extraction model: Timelines

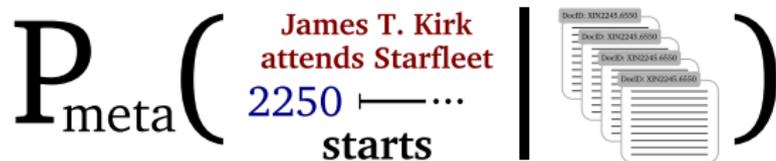


$$\prod P_{\text{meta}}(m \mid \text{documents})$$



$\left\{ \begin{array}{l} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \text{ --- } \dots \\ \text{starts} \end{array} \right\}, \left\{ \begin{array}{l} \text{George Kirk} \\ \text{Lifespan} \\ \dots \text{ --- } 2371 \\ \text{ends} \end{array} \right\}, \dots$

## Extraction model: Metarelations



# Extraction model: Metarelations

$$P_{\text{meta}} \left( \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \text{ ---} \dots \\ \text{starts} \end{array} \mid \begin{array}{c} \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \end{array} \right) =$$

$$P_C \left( \text{starts} \mid \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \end{array} , \begin{array}{c} \text{DocID: NYT2270.1005} \\ \text{Span: [10, 17]} \\ \text{...Kirk started at} \\ \text{Starfleet in 2250...} \end{array} \right)$$

$$\times P_C \left( \text{starts} \mid \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \end{array} , \begin{array}{c} \text{DocID: APW2266.0031} \\ \text{Span: [868, 890]} \\ \text{In September 2250,} \\ \text{James Kirk enrolled} \\ \text{in Starfleet...} \end{array} \right)$$

× ...

# Extraction model: Metarelations

$$P_{\text{meta}} \left( \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \text{ ---} \dots \\ \text{starts} \end{array} \mid \begin{array}{c} \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \\ \text{DocID: XN2245.0050} \end{array} \right) =$$

$$P_C \left( \text{starts} \mid \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \end{array} , \begin{array}{c} \text{DocID: NYT2270.1005} \\ \text{Span: [10, 17]} \\ \dots\text{Kirk started at} \\ \text{Starfleet in 2250...} \end{array} \right)$$

$$\times P_C \left( \text{starts} \mid \begin{array}{c} \text{James T. Kirk} \\ \text{attends Starfleet} \\ 2250 \end{array} , \begin{array}{c} \text{DocID: AP12266.0031} \\ \text{Span: [868, 890]} \\ \text{In September 2250,} \\ \text{James Kirk enrolled} \\ \text{in Starfleet...} \end{array} \right)$$

$\times \dots$  **mentions**

## Finding metarelation mentions

mentions( **James T. Kirk**  
**attends Starfleet**  
2250 ,  ) =

## Finding metarelation mentions

mentions( **James T. Kirk**  
**attends Starfleet**  
2250 ,  ) =

Search: “James T. Kirk” and “Starfleet”

# Finding metarelation mentions

mentions( **James T. Kirk**  
attends **Starfleet**  
2250 ,  ) =

Search: “James T. Kirk” and “Starfleet”

DocID: NYT2270.1005  
Span: [10, 17]

...Kirk started at  
Starfleet in 2250...

DocID: APW2266.0031  
Span: [868, 890]

In September 2250,  
James Kirk enrolled  
in Starfleet...

DocID: AFP2256.0112  
Span: [38, 55]

...Kirk...Later, he  
graduated from  
Starfleet in 2254...

# Finding metarelation mentions

mentions( **James T. Kirk**  
attends **Starfleet**  
2250 ,  ) =

Search: “James T. Kirk” and “Starfleet”

DocID: NYT2270.1005  
Span: [10, 17]

..**Kirk** started at  
Starfleet in 2250...

DocID: APW2266.0031  
Span: [868, 890]

In September 2250,  
**James Kirk** enrolled  
in **Starfleet**..

DocID: AFP2256.0112  
Span: [38, 55]

..**Kirk**..Later, **he**  
graduated from  
**Starfleet** in 2254...

# Finding metarelation mentions

mentions( **James T. Kirk**  
attends **Starfleet**  
2250 ,  ) =

Search: “James T. Kirk” and “Starfleet”

DocID: NYT2270.1005  
Span: [10, 17]

..**Kirk** started at  
Starfleet in **2250**..

DocID: APW2266.0031  
Span: [868, 890]

In **September 2250**,  
**James Kirk** enrolled  
in **Starfleet**..

DocID: AFP2256.0112  
Span: [38, 55]

..**Kirk**..Later, **he**  
graduated from  
**Starfleet** in **2254**..

# Finding metarelation mentions

mentions( **James T. Kirk**  
**attends Starfleet**  
2250 ,  ) =

Search: “James T. Kirk” and “Starfleet”

DocID: NYT2270.1005  
Span: [10, 17]

.. **Kirk** started at  
Starfleet in **2250**..

**temporal  
match**

DocID: APW2266.0031  
Span: [868, 890]

In **September 2250**,  
**James Kirk** enrolled  
in **Starfleet**..

**temporal match  
(approximate)**

DocID: AFP2256.0112  
Span: [10, 55]

.. **Kirk** later, **he**  
graduated from  
Starfleet in **2254**..

**temporal  
mismatch**

## Extraction model: Metarelation mentions



## Extraction model: Metarelation mentions

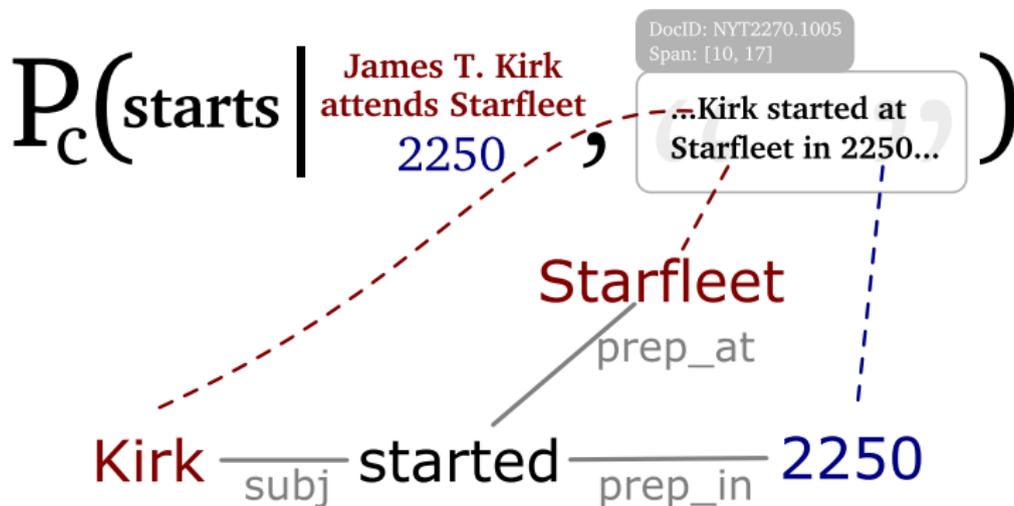


ends

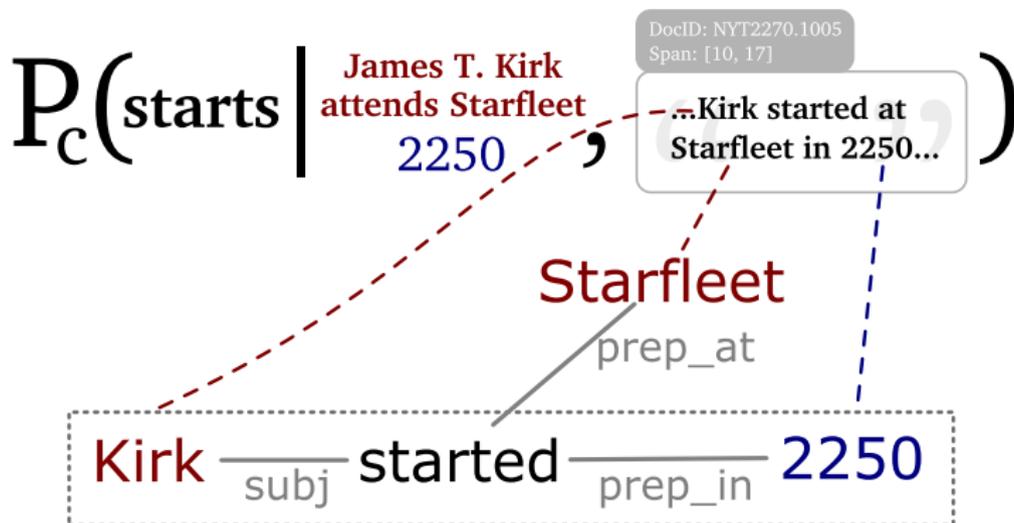
starts and ends

no relation

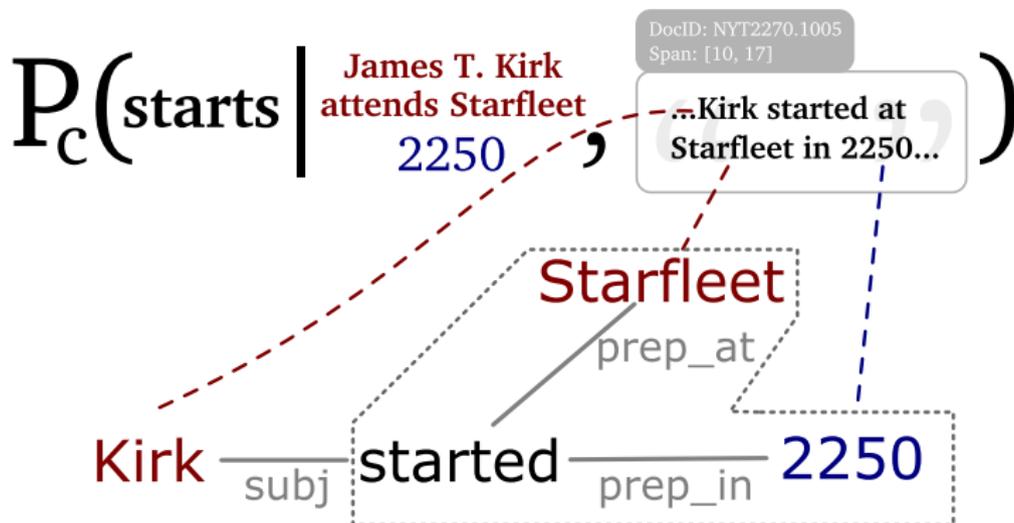
## Extraction model: Metarelation mentions



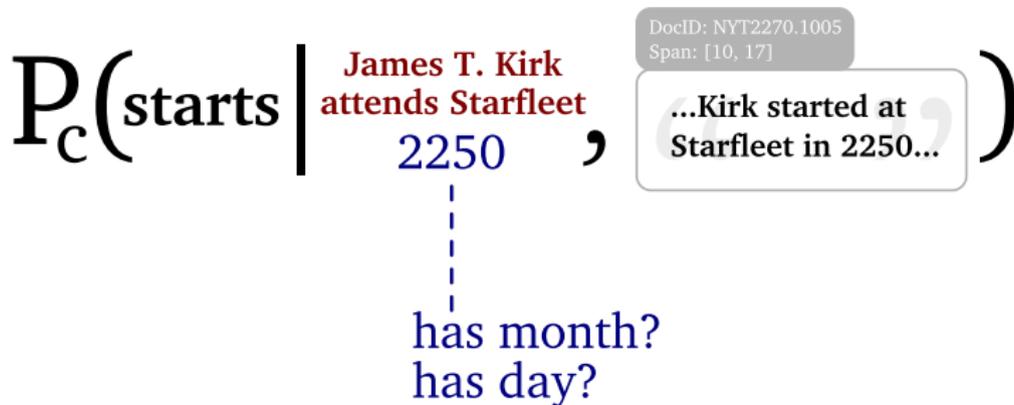
## Extraction model: Metarelation mentions



## Extraction model: Metarelation mentions



## Extraction model: Metarelation mentions



## Extraction model: Metarelation mentions



# Consistency model

$$\begin{aligned} & \text{score} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{Sheet 390241470} \\ \text{Sheet 390241470} \\ \text{Sheet 390241470} \\ \text{Sheet 390241470} \end{array} \right. , \text{Freebase}^{\text{TM}} \right) \\ &= P_{\text{ext}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{Sheet 390241470} \\ \text{Sheet 390241470} \\ \text{Sheet 390241470} \\ \text{Sheet 390241470} \end{array} \right. \right) \\ &\times P_{\text{cons}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \text{Freebase}^{\text{TM}} \right. \right)^{\beta} \end{aligned}$$

## Approaches to modeling consistency

- ▶ No reliable implicit negatives → can't use a classifier

## Approaches to modeling consistency

- ▶ No reliable implicit negatives → can't use a classifier
- ▶ Similar to language modeling (**density estimation**)

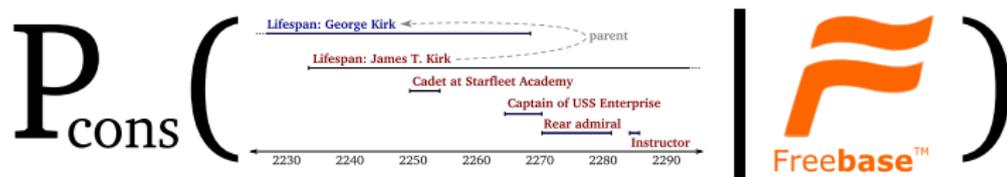
## Approaches to modeling consistency

- ▶ No reliable implicit negatives → can't use a classifier
- ▶ Similar to language modeling (**density estimation**)
- ▶ Decompose timelines into a series of **questions**

## Approaches to modeling consistency

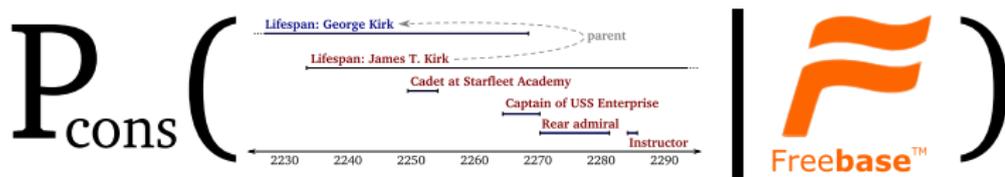
- ▶ No reliable implicit negatives → can't use a classifier
- ▶ Similar to language modeling (**density estimation**)
- ▶ Decompose timelines into a series of **questions**
- ▶ Estimate probabilities of answers from Freebase

# Consistency model: questions



$$= \prod_{(q,a) \in Q} P_q(a \mid q, \text{Freebase}^{\text{TM}})$$

# Consistency model: questions



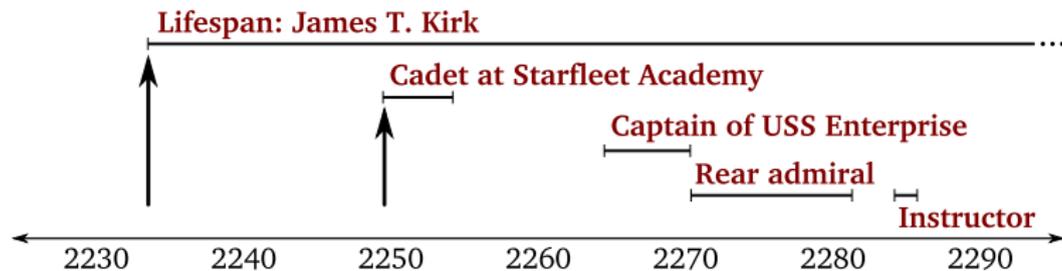
$$= \prod_{(q,a) \in Q} P_q(a | q, \text{Freebase})$$

$(q,a) \in Q$  (
 
 )


 all possible questions  
 about the timeline

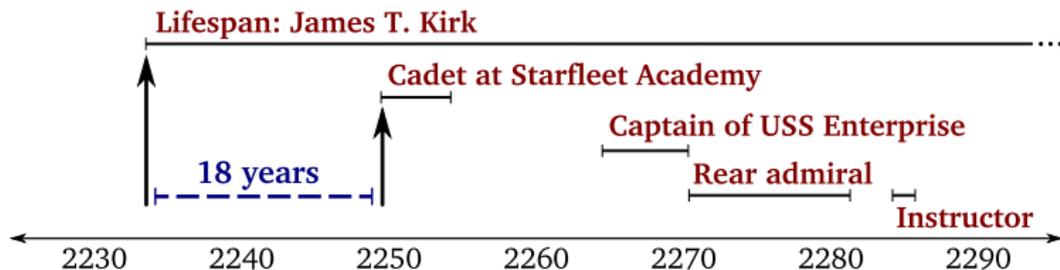
# Consistency model: questions

is start(**lifespan**) before start(**attended school**)?



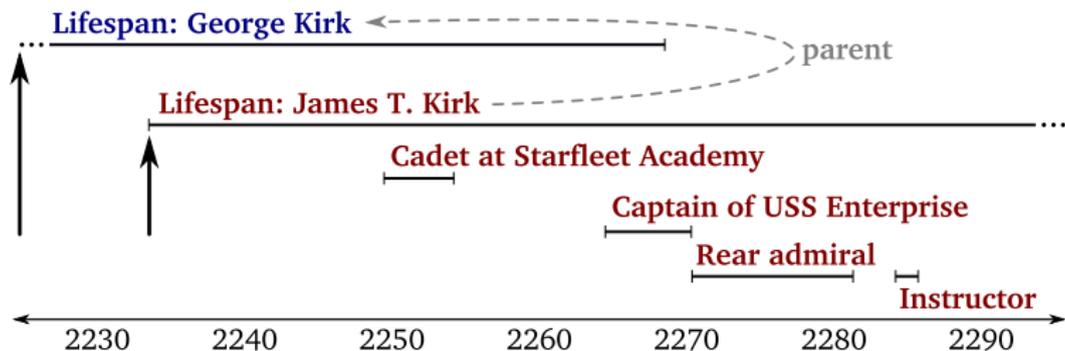
# Consistency model: questions

is start(**lifespan**) before start(**attended school**)?  
by how much?

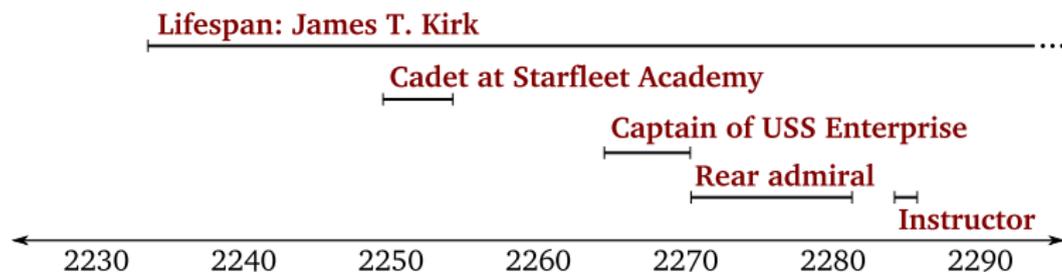


# Consistency model: questions

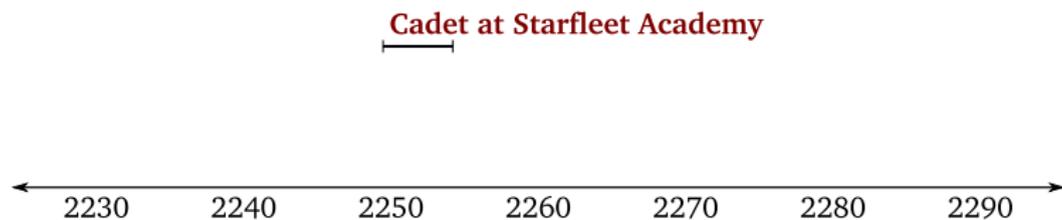
is start(Kirk's lifespan) before start(a parent's lifespan)?



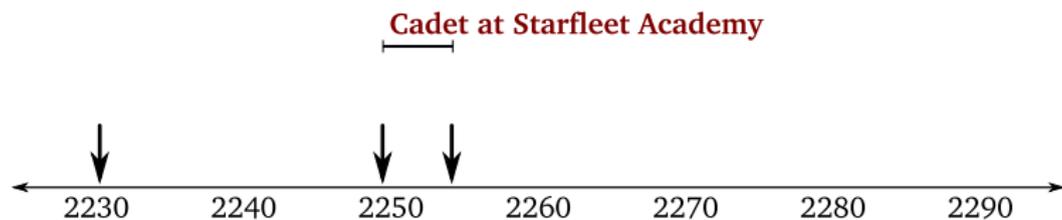
## Inference: Random restart hill climbing



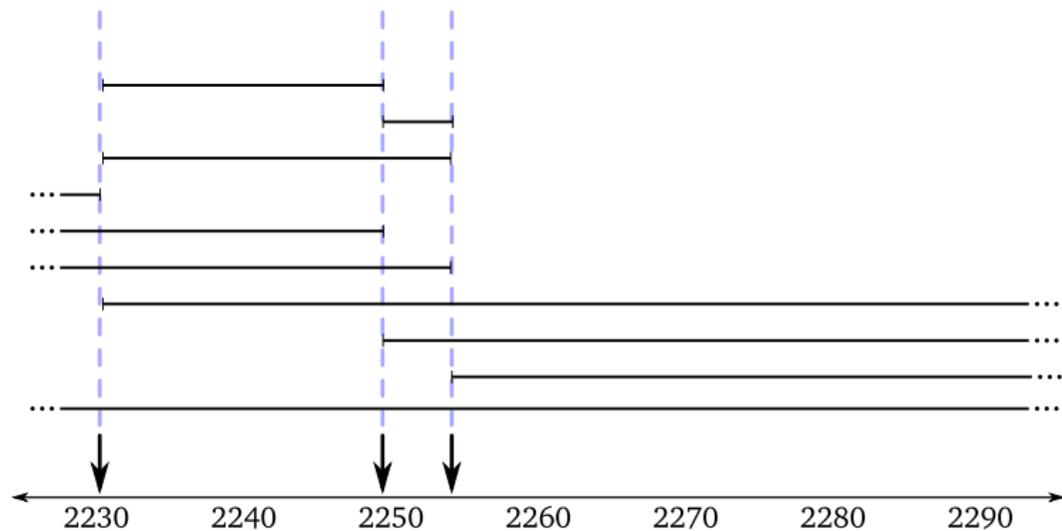
## Inference: Random restart hill climbing



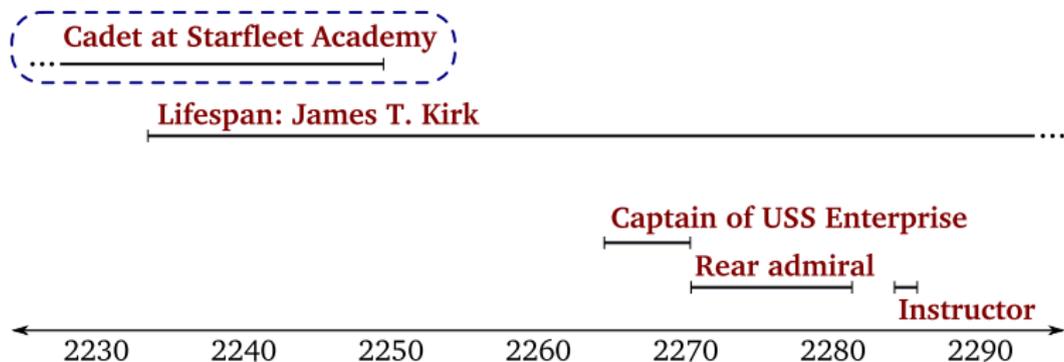
## Inference: Random restart hill climbing



## Inference: Random restart hill climbing

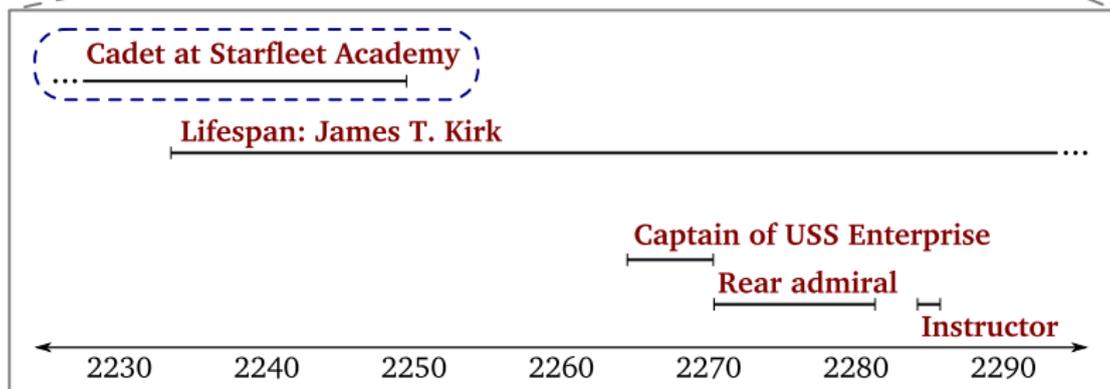


# Inference: Random restart hill climbing



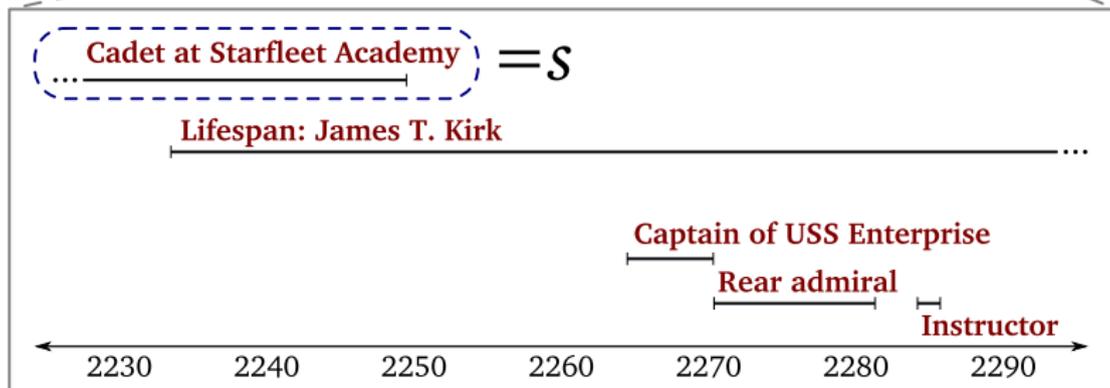
# Inference: Random restart hill climbing

score(  ,  ,  )



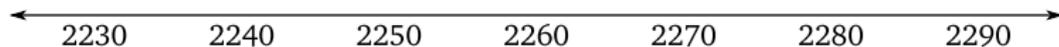
## Inference: Random restart hill climbing

$\operatorname{argmax}_{s \in \mathcal{S}} \operatorname{score}(\text{[ ]}, \text{[ ]}, \text{Freebase}^{\text{TM}})$

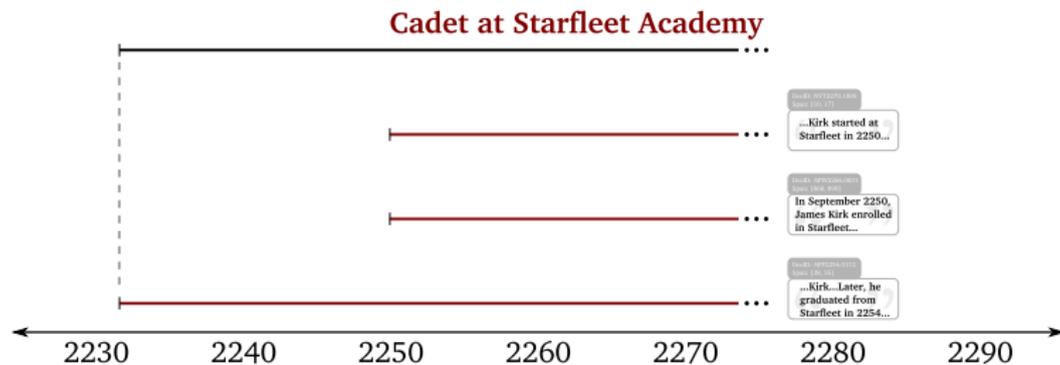


## *null* baseline

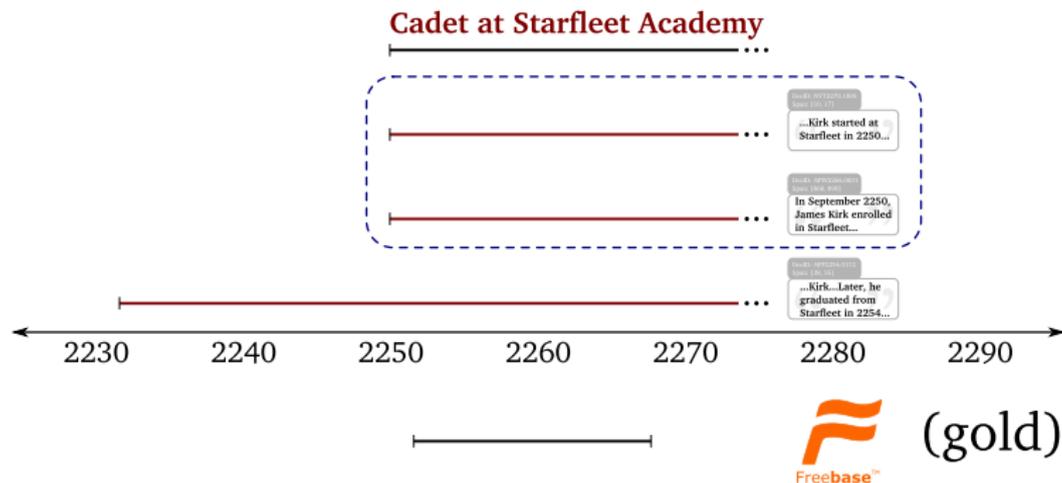
... **Cadet at Starfleet Academy** ...



## Pipeline



## Oracle



# Combined Classifier (CC)

$$\text{score} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \leftarrow \begin{array}{c} \text{DocID: SP12145.0150} \\ \text{DocID: SP12145.0151} \\ \text{DocID: SP12145.0152} \\ \text{DocID: SP12145.0153} \end{array}, \text{Freebase}^{\text{TM}} \right)$$
$$= P_{\text{ext}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \leftarrow \begin{array}{c} \text{DocID: SP12145.0150} \\ \text{DocID: SP12145.0151} \\ \text{DocID: SP12145.0152} \\ \text{DocID: SP12145.0153} \end{array} \right)$$

# Joint Classifier with Consistency (JCC)

$$\begin{aligned}
 & \text{score} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{DocID: SP12145.0150} \\ \text{DocID: SP12145.0151} \\ \text{DocID: SP12145.0152} \\ \text{DocID: SP12145.0153} \end{array} \right. , \text{Freebase}^{\text{TM}} \right) \\
 &= P_{\text{ext}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \begin{array}{c} \text{DocID: SP12145.0150} \\ \text{DocID: SP12145.0151} \\ \text{DocID: SP12145.0152} \\ \text{DocID: SP12145.0153} \end{array} \right. \right) \\
 &\times P_{\text{cons}} \left( \begin{array}{c} \text{Lifespan: George Kirk} \leftarrow \text{parent} \\ \text{Lifespan: James T. Kirk} \\ \text{Cadet at Starfleet Academy} \\ \text{Captain of USS Enterprise} \\ \text{Rear admiral} \\ \text{Instructor} \end{array} \left| \text{Freebase}^{\text{TM}} \right. \right)^{\beta}
 \end{aligned}$$

## Evaluation

| Model                                    | Test        |
|--|-------------|
| Oracle                                   | 75.2        |
| <b>Joint Classifier with Consistency</b> | <b>72.2</b> |
| Combined Classifier                      | 71.5        |
| Pipeline                                 | 70.5        |
| <i>null</i> baseline                     | 55.6        |

(metric is Temporal KBP-inspired)

## Timelining: Future work

- ▶ More elaborate question types
  - ▶ Questions that involve more than two fluents
  - ▶ Non-binary questions (e.g., real valued)

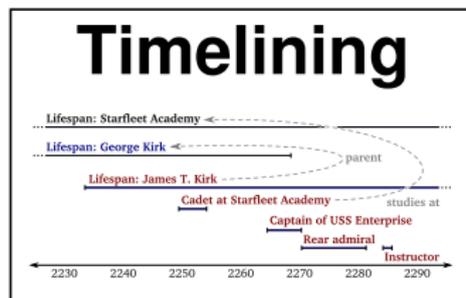
## Timelining: Future work

- ▶ More elaborate question types
  - ▶ Questions that involve more than two fluents
  - ▶ Non-binary questions (e.g., real valued)
- ▶ Increasing complexity of temporal representation
  - ▶ Ranges not recognized (“served as CEO for 7 years until 1993”)
  - ▶ Partial information unused (“died after 2265”, “was a student in 1994”)

## Timelining: Future work

- ▶ More elaborate question types
  - ▶ Questions that involve more than two fluents
  - ▶ Non-binary questions (e.g., real valued)
- ▶ Increasing complexity of temporal representation
  - ▶ Ranges not recognized (“served as CEO for 7 years until 1993”)
  - ▶ Partial information unused (“died after 2265”, “was a student in 1994”)
- ▶ Jointly performing relation extraction as well
  - ▶ Consistency model could capture non-temporal distributions too

## Act II: Conclusions



- ▶ Described a model to produce more consistent timelines
- ▶ 36% error reduction over pipelined baseline
- ▶ Random restart hill climbing: simple, fast, effective

## Summary

- ▶ Many IE tasks (e.g., ACE, KBP, BioNLP) have complex dependencies between extraction items
- ▶ Ignoring these can result in inconsistent output
- ▶ Showed two case studies to model these dependencies
- ▶ Biomedical event parser available at Stanford NLP:  
`http://nlp.stanford.edu/software/eventparser.shtml`
- ▶ Timelining dataset available on my website:  
`http://nlp.stanford.edu/~mcclosky/`

## Summary

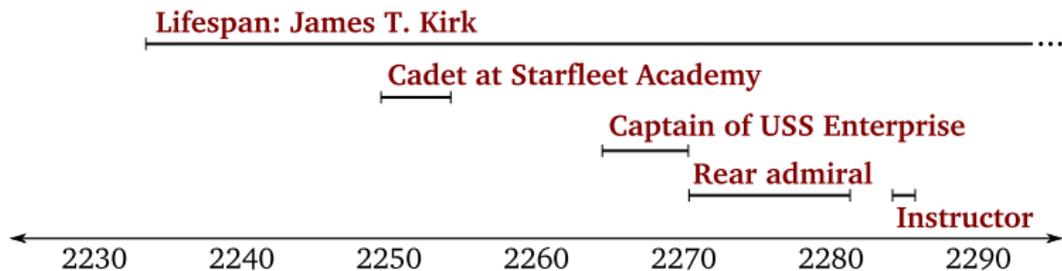
- ▶ Many IE tasks (e.g., ACE, KBP, BioNLP) have complex dependencies between extraction items
- ▶ Ignoring these can result in inconsistent output
- ▶ Showed two case studies to model these dependencies
- ▶ Biomedical event parser available at Stanford NLP:  
`http://nlp.stanford.edu/software/eventparser.shtml`
- ▶ Timelining dataset available on my website:  
`http://nlp.stanford.edu/~mcclosky/`

## Questions?

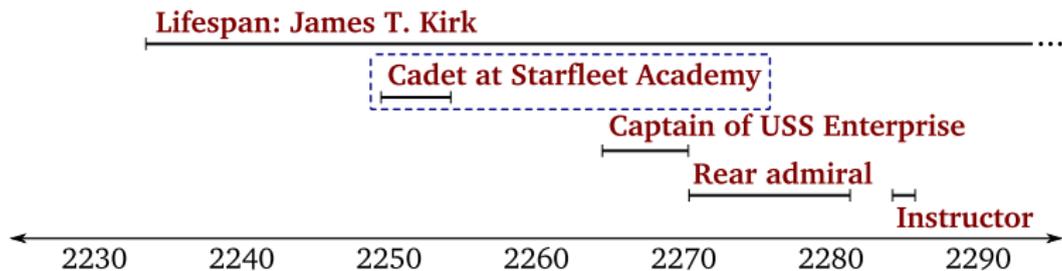
Thanks to the Stanford NLP group, Sebastian Riedel, Valentin Spitzkovsky, Micha Elsner, and our reviewers for their helpful feedback.

## Extra slides

## A (not so small) problem



## A (not so small) problem



## Handling differing numbers of fluents

- ▶ Compare with LM: shorter sentences are more probable

## Handling differing numbers of fluents

- ▶ Compare with LM: shorter sentences are more probable
- ▶ Possible solutions:
  - ▶ Regularization (penalize missing fluents)

# Handling differing numbers of fluents

- ▶ Compare with LM: shorter sentences are more probable
- ▶ Possible solutions:
  - ▶ Regularization (penalize missing fluents)
  - ▶ Averaging (normalize different numbers of fluents)

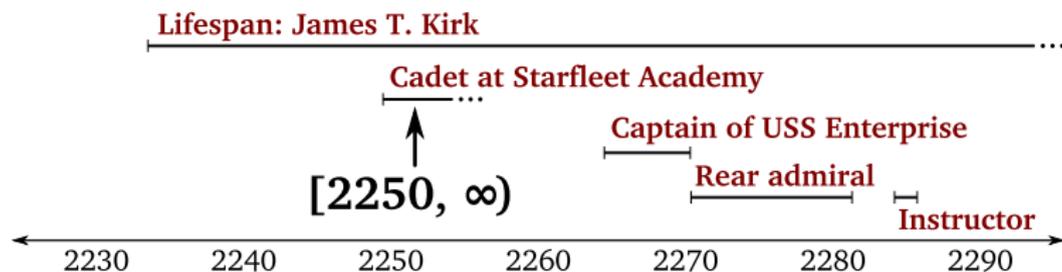
# Handling differing numbers of fluents

- ▶ Compare with LM: shorter sentences are more probable
- ▶ Possible solutions:
  - ▶ Regularization (penalize missing fluents)
  - ▶ Averaging (normalize different numbers of fluents)
  - ▶ Adjust weight between extraction and consistency models (tune  $\beta$ )

# Handling differing numbers of fluents

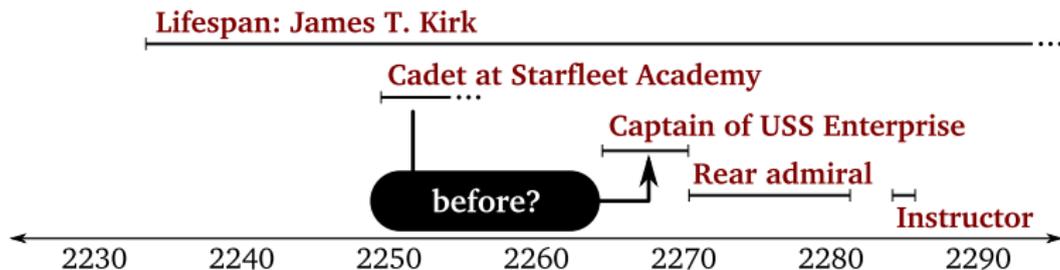
- ▶ Compare with LM: shorter sentences are more probable
- ▶ Possible solutions:
  - ▶ Regularization (penalize missing fluents)
  - ▶ Averaging (normalize different numbers of fluents)
  - ▶ Adjust weight between extraction and consistency models (tune  $\beta$ )
  - ▶ Add “unknown” as a possible answer

## Questions with unknown answers



## Questions with unknown answers

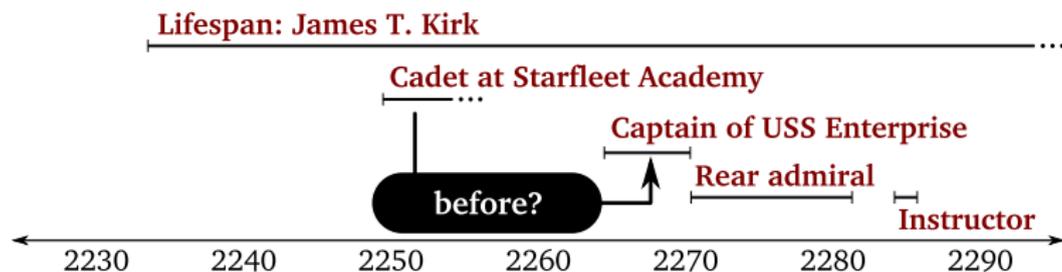
is **attends school** completely before **has title**?



## Questions with unknown answers

is **attends school** completely before **has title**?

**answer: maybe!**



# Corpora and Metric

- ▶ **Corpora:** Freebase and Wikipedia
  - ▶ Freebase and Wikipedia infoboxes include fluents
  - ▶ Scraped  $\approx 250,000$  entities from Freebase
  - ▶ Fluents from Freebase and Wikipedia are merged to form gold corpora
  - ▶ Restricted to entities linked to at least one other entity

# Corpora and Metric

- ▶ **Corpora:** Freebase and Wikipedia
  - ▶ Freebase and Wikipedia infoboxes include fluents
  - ▶ Scraped  $\approx 250,000$  entities from Freebase
  - ▶ Fluents from Freebase and Wikipedia are merged to form gold corpora
  - ▶ Restricted to entities linked to at least one other entity
- ▶ **Metric:** Temporal TAC-KBP inspired
  - ▶ Evaluate predicted temporal bounds on fluents, not fluent mentions
  - ▶ Punish vagueness and unnecessary constraints

# Metrics

- ▶ Evaluate predicted temporal bounds on fluents, not fluent mentions
- ▶ All times given at the day level
- ▶ Punish vagueness and unnecessary constraints:

$$\text{quality}(t_{\text{predicted}}, t_{\text{gold}}) = \begin{cases} 1 & t_{\text{predicted}} = t_{\text{gold}} \\ 0 & t_{\text{predicted}} = \pm\infty \oplus t_{\text{gold}} = \pm\infty \\ \frac{1}{1 + |t_{\text{predicted}} - t_{\text{gold}}|} & \textit{otherwise} \end{cases}$$

## Inspecting the Consistency Component

| <b>Question</b>              | <b>Yes</b> | <b>No</b>       |
|------------------------------|------------|-----------------|
| Born before getting married? | 14.8%      | $\approx 0.0\%$ |

## Inspecting the Consistency Component

| <b>Question</b>                      | <b>Yes</b> | <b>No</b>       |
|--------------------------------------|------------|-----------------|
| Born before getting married?         | 14.8%      | $\approx 0.0\%$ |
| Died before their parents were born? | 0.3%       | 53.7%           |

## Inspecting the Consistency Component

| <b>Question</b>                       | <b>Yes</b> | <b>No</b>       |
|---------------------------------------|------------|-----------------|
| Born before getting married?          | 14.8%      | $\approx 0.0\%$ |
| Died before their parents were born?  | 0.3%       | 53.7%           |
| Finished a job before starting a job? | 72.5%      | 20.5%           |

## Fluent-level results for people

| Fluent                      | Model       |       |      |      |        |
|-----------------------------|-------------|-------|------|------|--------|
|                             | <i>null</i> | Basic | CC   | JCC  | Oracle |
| <i>lifespan</i>             | 28.6        | 64.6  | 65.6 | 66.1 | 69.1   |
| <i>has spouse</i>           | 92.2        | 92.1  | 92.2 | 92.3 | 93.1   |
| <i>attends school</i>       | 97.7        | 97.7  | 97.7 | 98.1 | 98.1   |
| <i>has job title</i>        | 78.8        | 79.4  | 78.8 | 78.8 | 80.3   |
| <i>holds gov't position</i> | 16.7        | 19.7  | 19.7 | 19.7 | 25.1   |

## Fluent-level results for people

| Fluent                      | Model       |       |      |      |        |
|-----------------------------|-------------|-------|------|------|--------|
|                             | <i>null</i> | Basic | CC   | JCC  | Oracle |
| <i>lifespan</i>             | 28.6        | 64.6  | 65.6 | 66.1 | 69.1   |
| <i>has spouse</i>           | 92.2        | 92.1  | 92.2 | 92.3 | 93.1   |
| <i>attends school</i>       | 97.7        | 97.7  | 97.7 | 98.1 | 98.1   |
| <i>has job title</i>        | 78.8        | 79.4  | 78.8 | 78.8 | 80.3   |
| <i>holds gov't position</i> | 16.7        | 19.7  | 19.7 | 19.7 | 25.1   |

## Fluent-level results for people

| Fluent                      | Model       |       |      |      |        |
|-----------------------------|-------------|-------|------|------|--------|
|                             | <i>null</i> | Basic | CC   | JCC  | Oracle |
| <i>lifespan</i>             | 28.6        | 64.6  | 65.6 | 66.1 | 69.1   |
| <i>has spouse</i>           | 92.2        | 92.1  | 92.2 | 92.3 | 93.1   |
| <i>attends school</i>       | 97.7        | 97.7  | 97.7 | 98.1 | 98.1   |
| <i>has job title</i>        | 78.8        | 79.4  | 78.8 | 78.8 | 80.3   |
| <i>holds gov't position</i> | 16.7        | 19.7  | 19.7 | 19.7 | 25.1   |