Unsupervised Induction of Shallow Semantic Representations with Feature-Rich Models

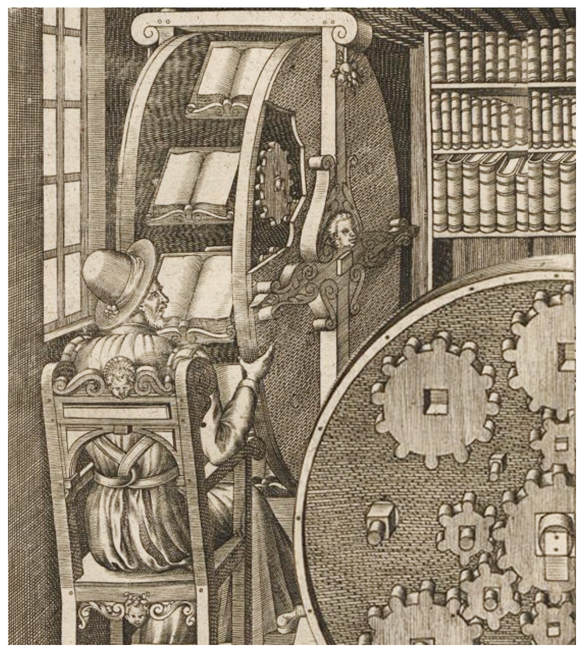
Diego Marcheggiani, Ivan Titov, and Ehsan Khoddam

University of Amsterdam

Natural language processing (NLP)

The key bottleneck: the lack of accurate methods for producing meaning representations of texts and reasoning with these representations

Machine translation

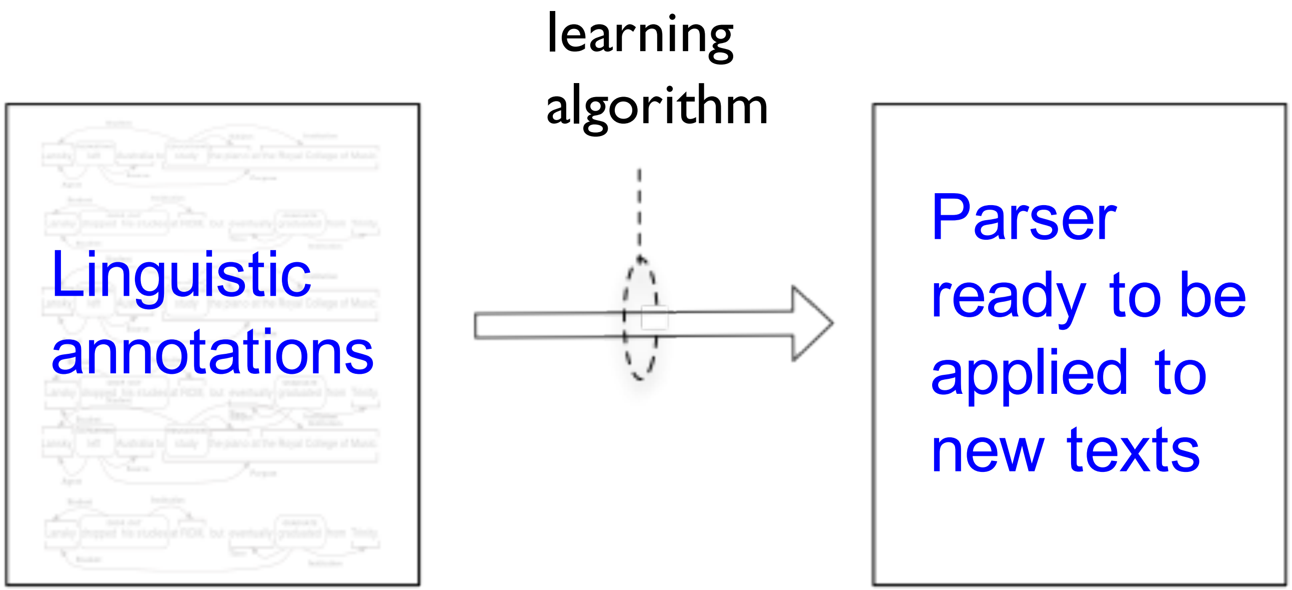


Question answering

Information retrieval

Modern semantics parsers

Modern frame-semantic parsers rely on supervised learning



Challenge #1

It is impossible to annotate enough data to estimate an effective broad-coverage semantic parser

Question Answering

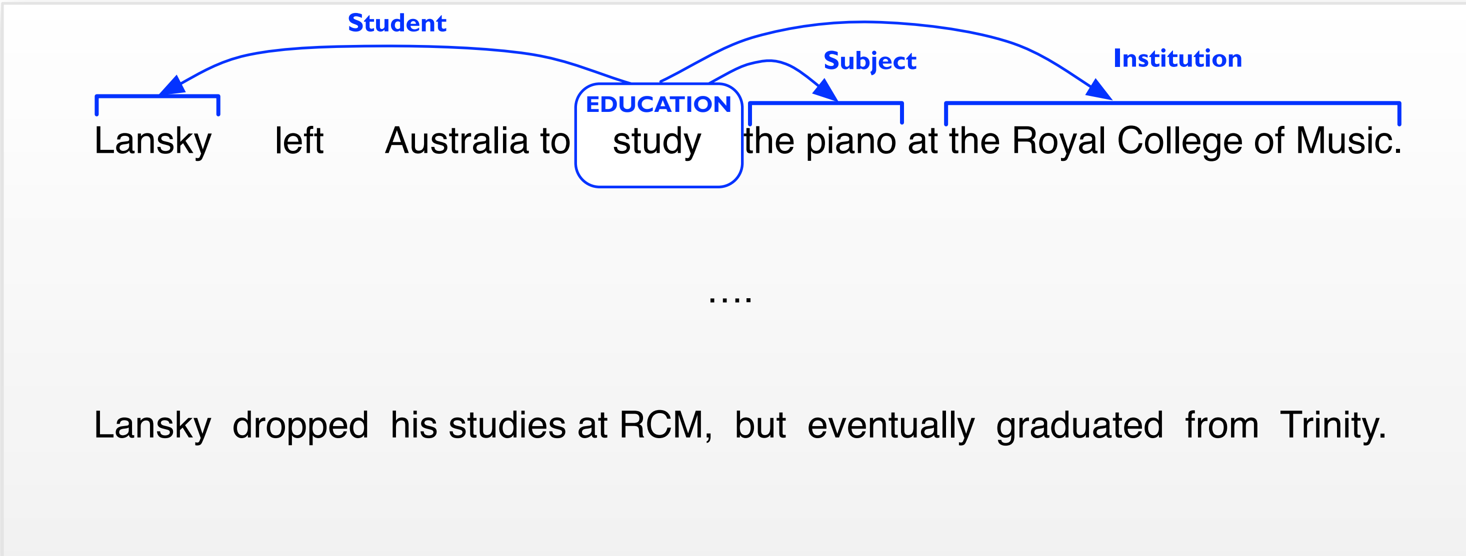
Lansky left Australia to study the piano at the Royal College of Music.

....

Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?

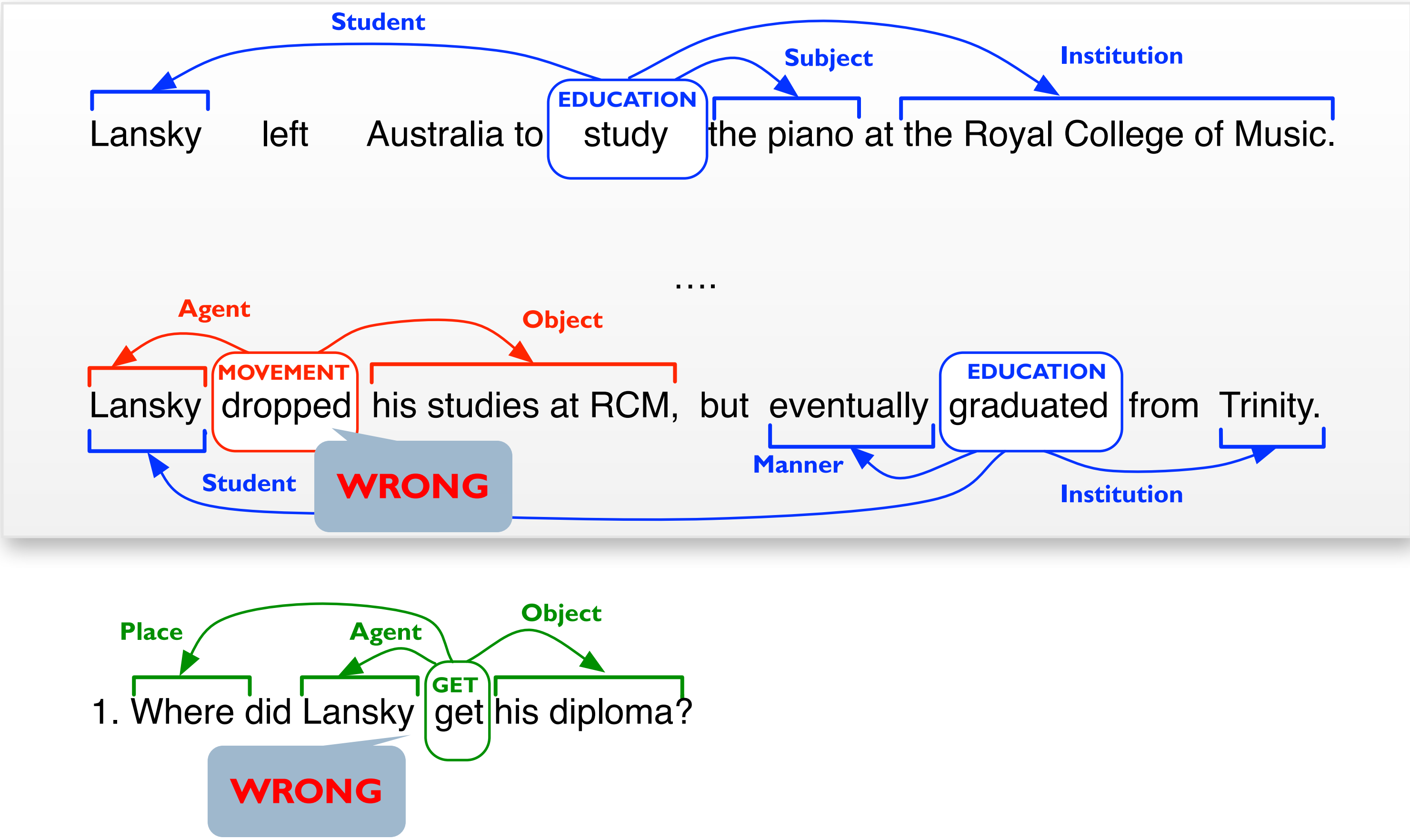
Question Answering



1. Where did Lansky get his diploma?

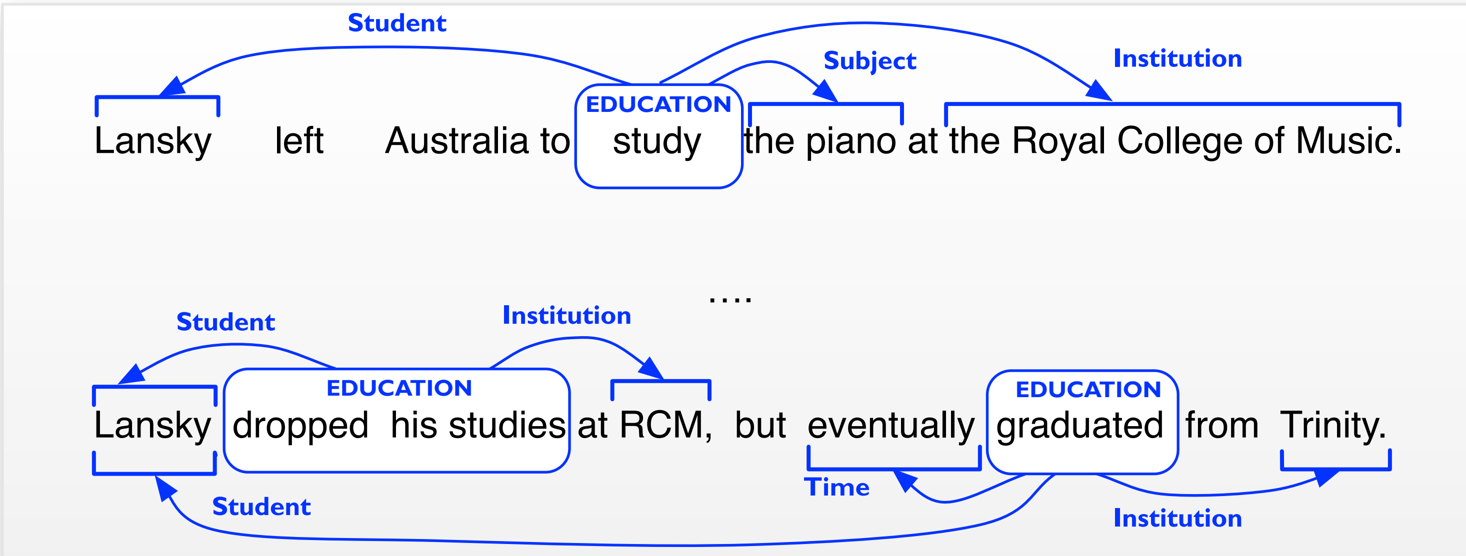
Output of a state-of-the-art parser

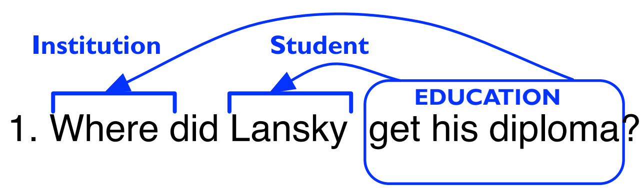
CMU's SEMAFOR [Das et al., 2012] trained on 100,000 sentences (FrameNet)



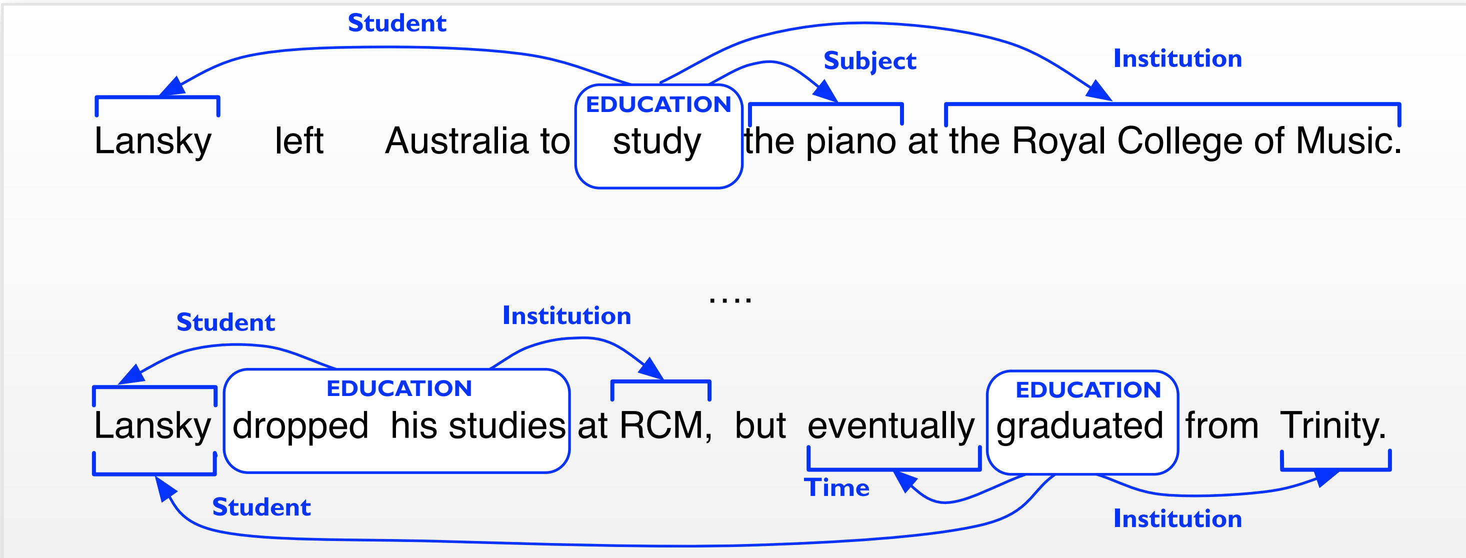
The parser's output does not let us answer even this simple question

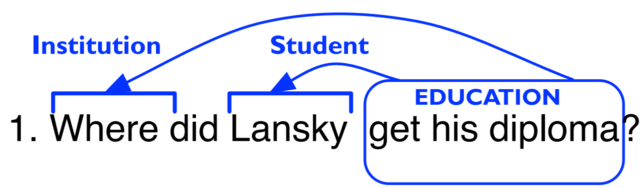
"Correct" semantics as imposed by linguists





"Correct" semantics as imposed by linguists



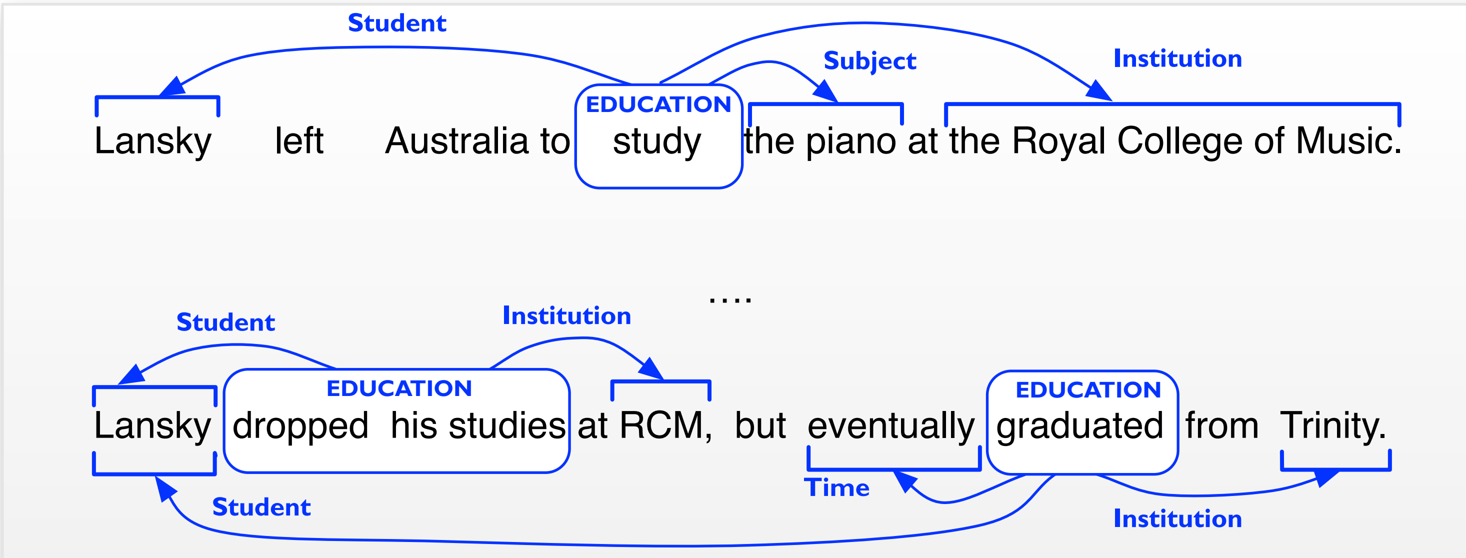


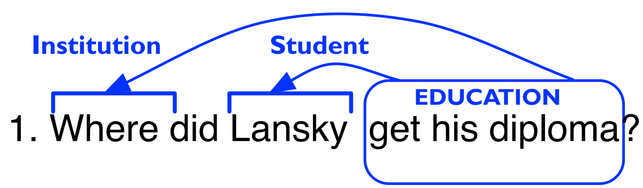
Trinity or RCM ????

Challenge #2

Representations defined by linguists are not appropriate for reasoning

"Correct" semantics as imposed by linguists





Trinity or RCM ????

Challenge #2

Representations defined by linguists are not always ~~appropriate for reasoning~~ helpful for solving specific tasks (i.e., QA)

Semantic frame and role labeling

- The challenges motivated research in unsupervised role / frame induction:
  - Role induction [Swier and Stevenson '04; Grenager and Manning '06; Lang and Lapata '10, '11, '14; Titov and Klementiev '12; Garg and Henderson '12;…]
  - Frame induction [Titov and Klementiev '11; O' Connor '12; Modi et al.'12; Materna '12; Kawahara et al. '13; Cheung et al. '13; Chambers et al., 14;…]

- These models rely on very restricted sets of features
- Not (quite) appropriate for inference (i.e., QA task)

Contributions

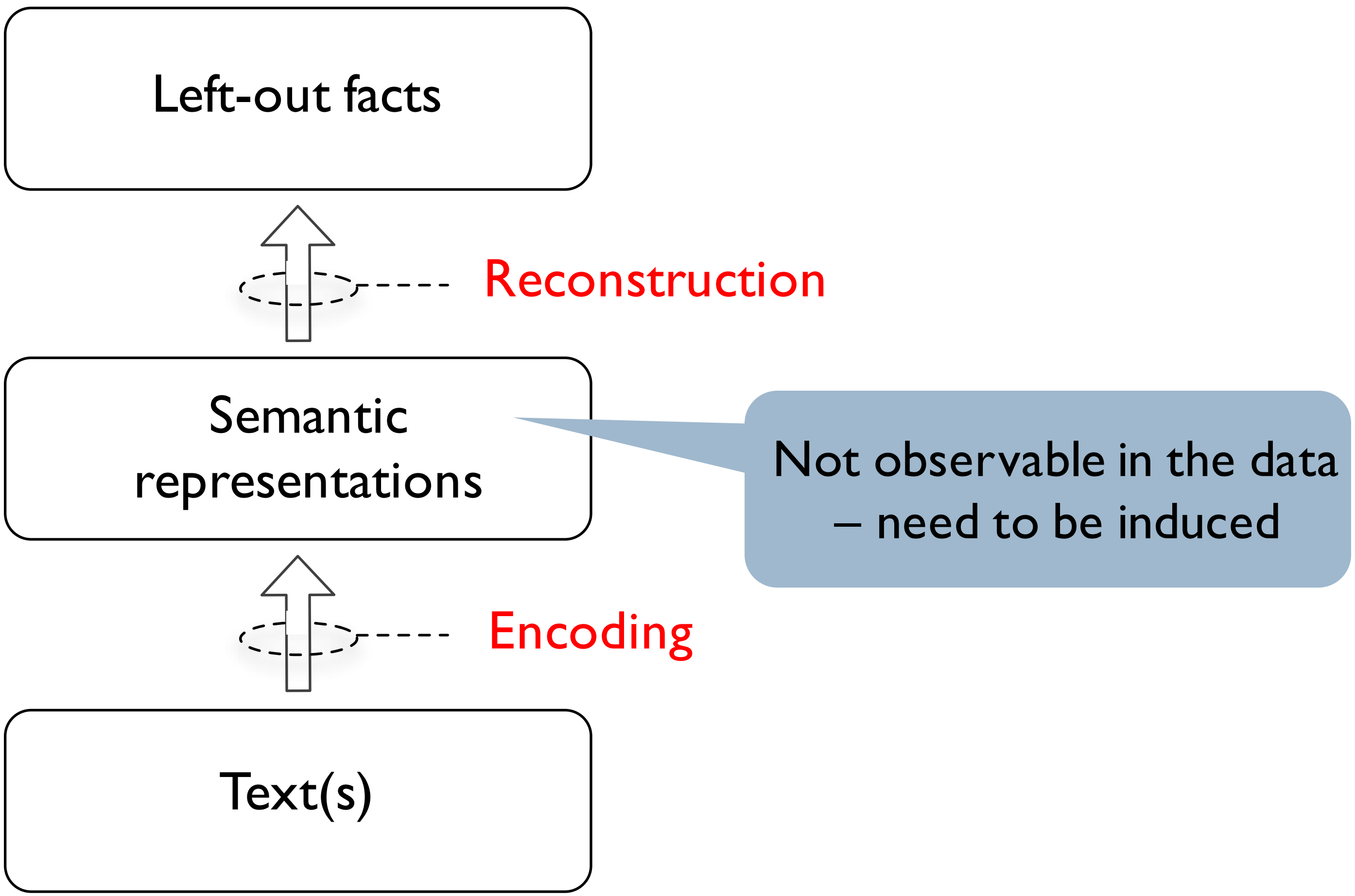
- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - language-independent

- The framework naturally supports:
  - Integration of prior linguistic knowledge
  - Semi-supervised learning

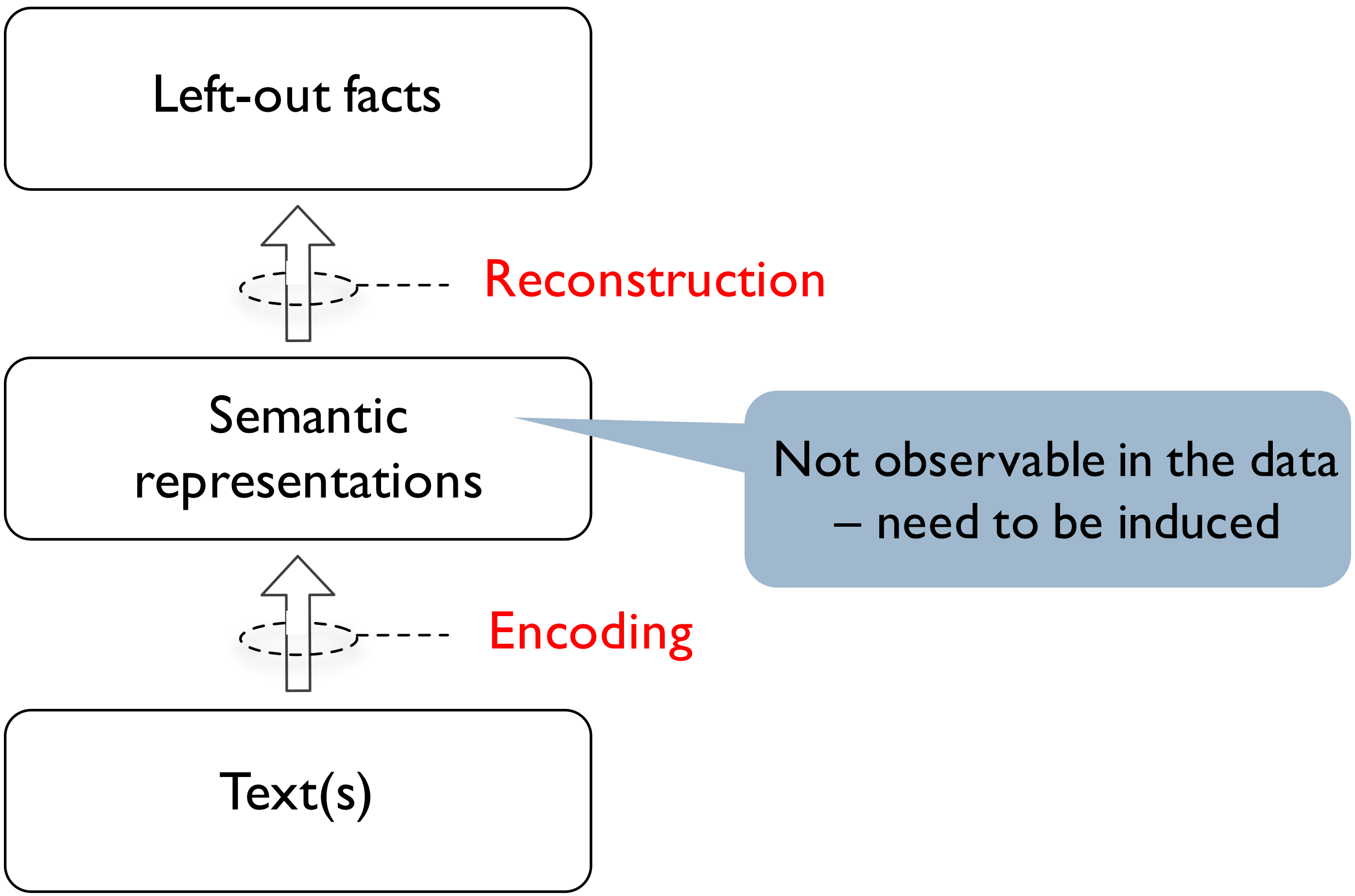
Outline

- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: role induction, frame induction

General framework

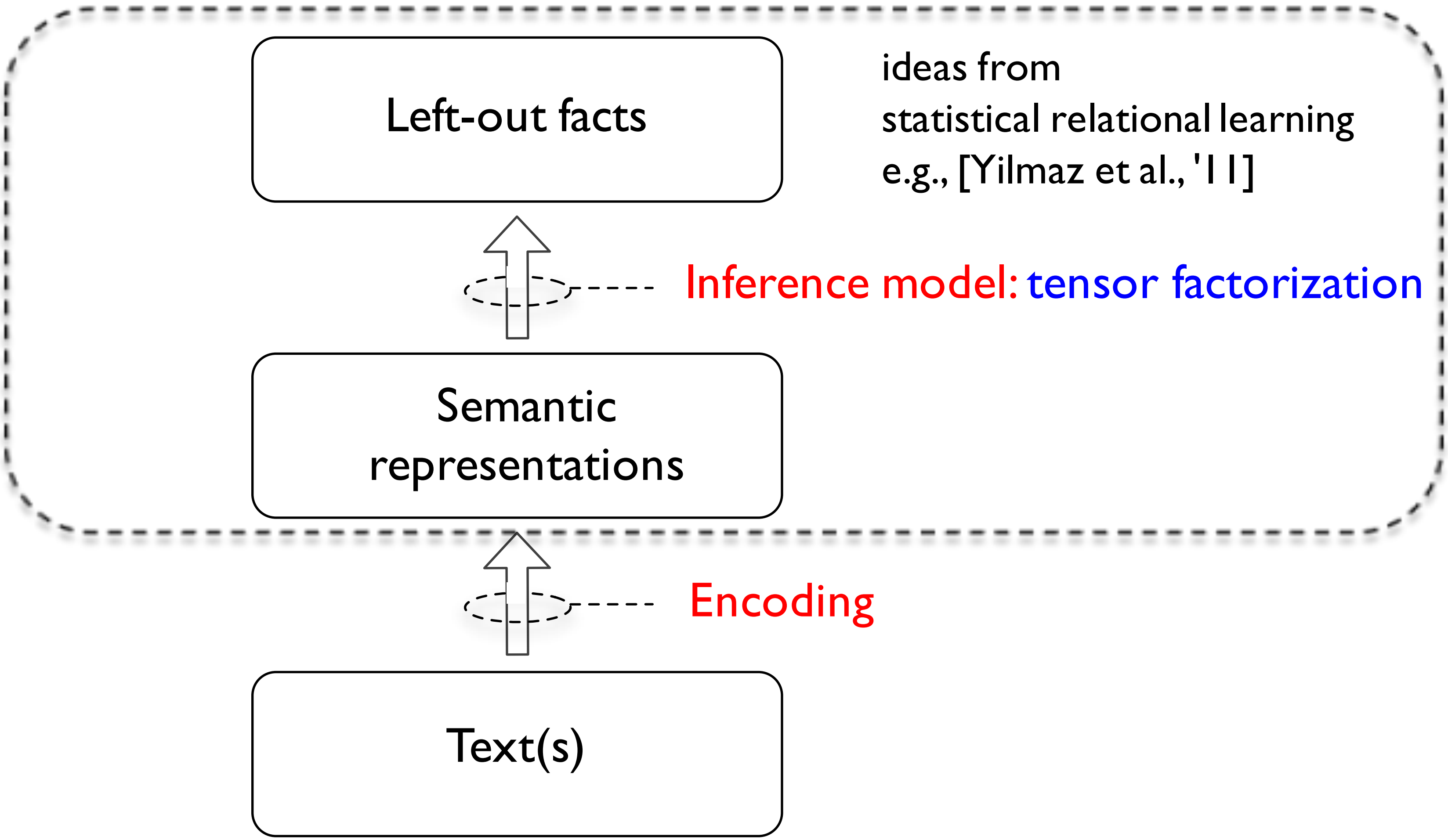


General framework

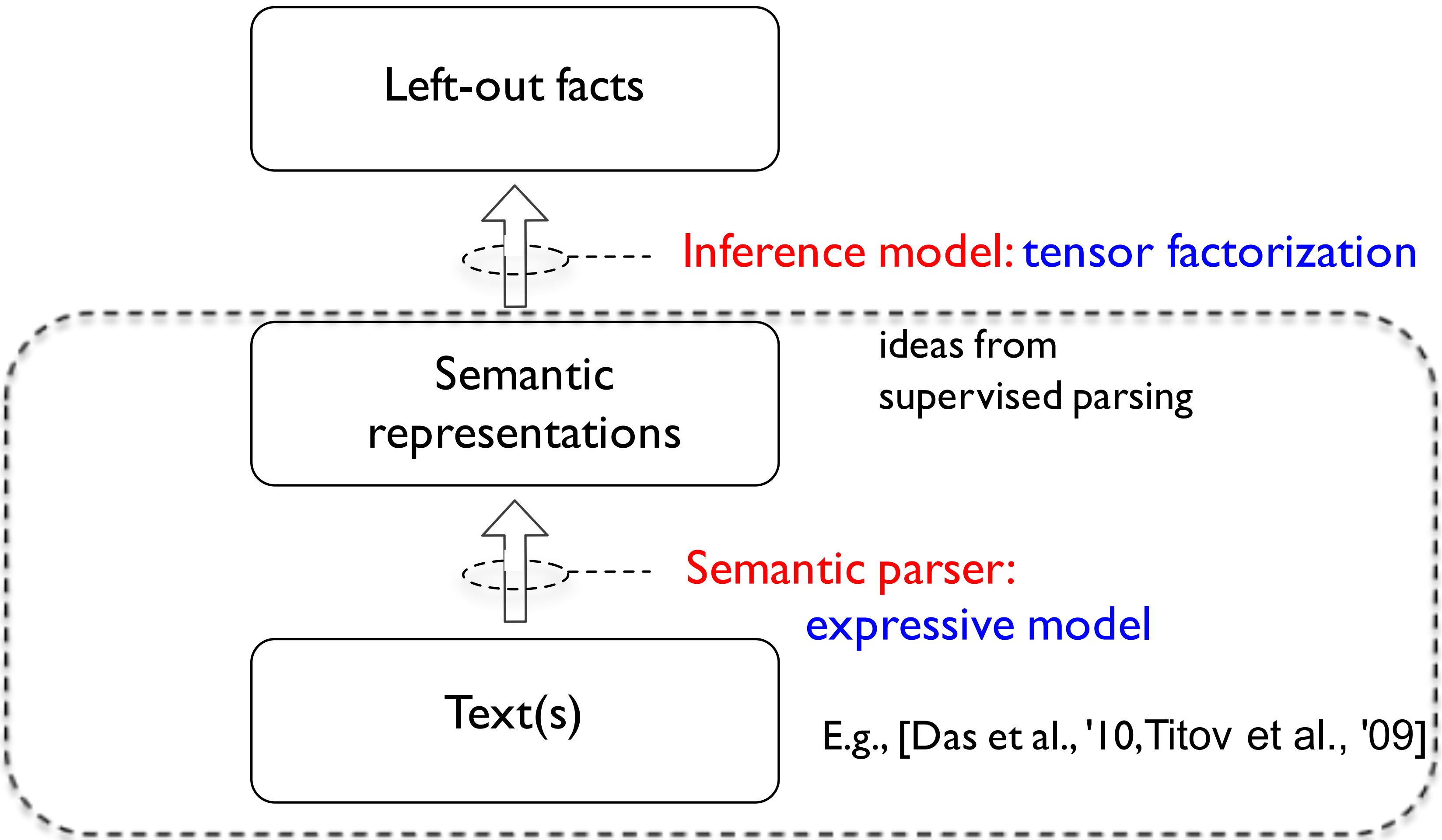


Instead of using annotated data, induce representations beneficial for inferring left-out facts

General framework

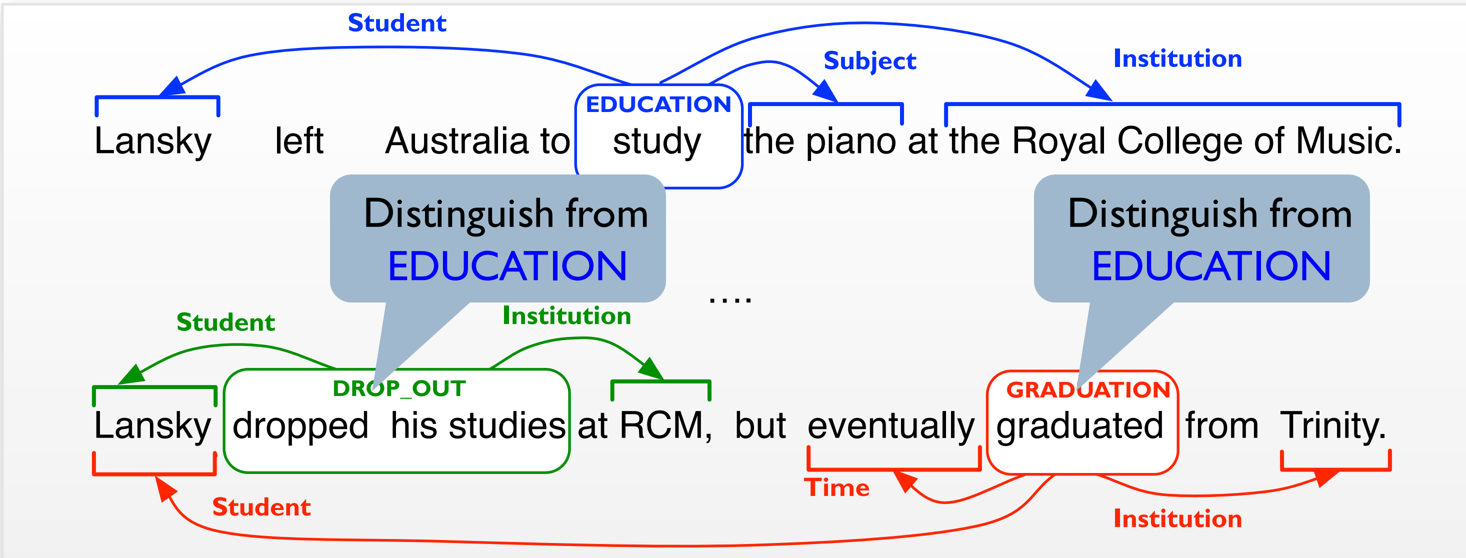


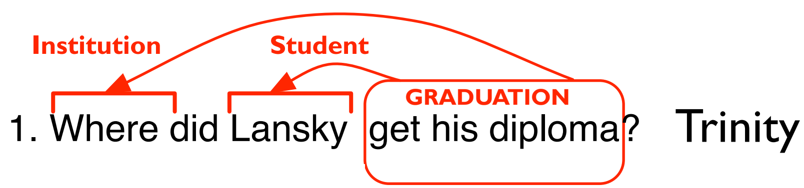
General framework



Inference model and semantic parser are jointly estimated from unlabeled data

Learning for reasoning





The learning objective can ensure that the representations are informative for reasoning

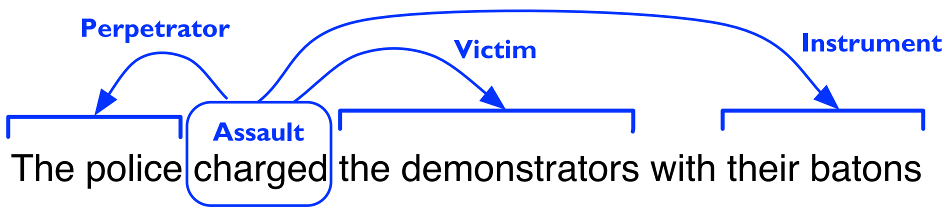
Outline

- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: role induction, frame induction

[Titov and Khoddam, ’15]

Feature-rich models of semantic frames

Consider a frame realization



Observable — $\mathbf{a} = (a_1, \ldots, a_n)$ - arguments (*police, the demonstrators, their batons*)

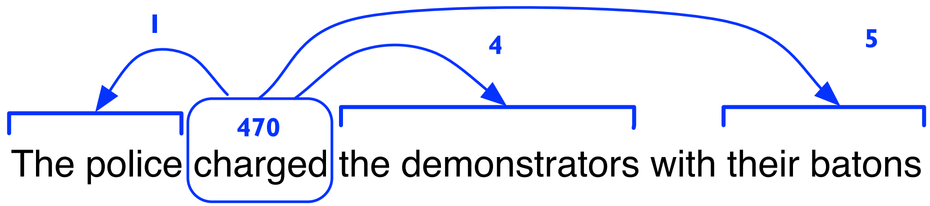
Latent — $\mathbf{r} = (r_1, \ldots, r_n)$ - roles (Perpetrator, Victim, Instrument)

f - frame (Assault)

[Titov and Khoddam, '15]

Feature-rich models of semantic frames

Consider a frame realization



Observable — $\mathbf{a} = (a_1, \ldots, a_n)$ - arguments (*police, the demonstrators, their batons*)

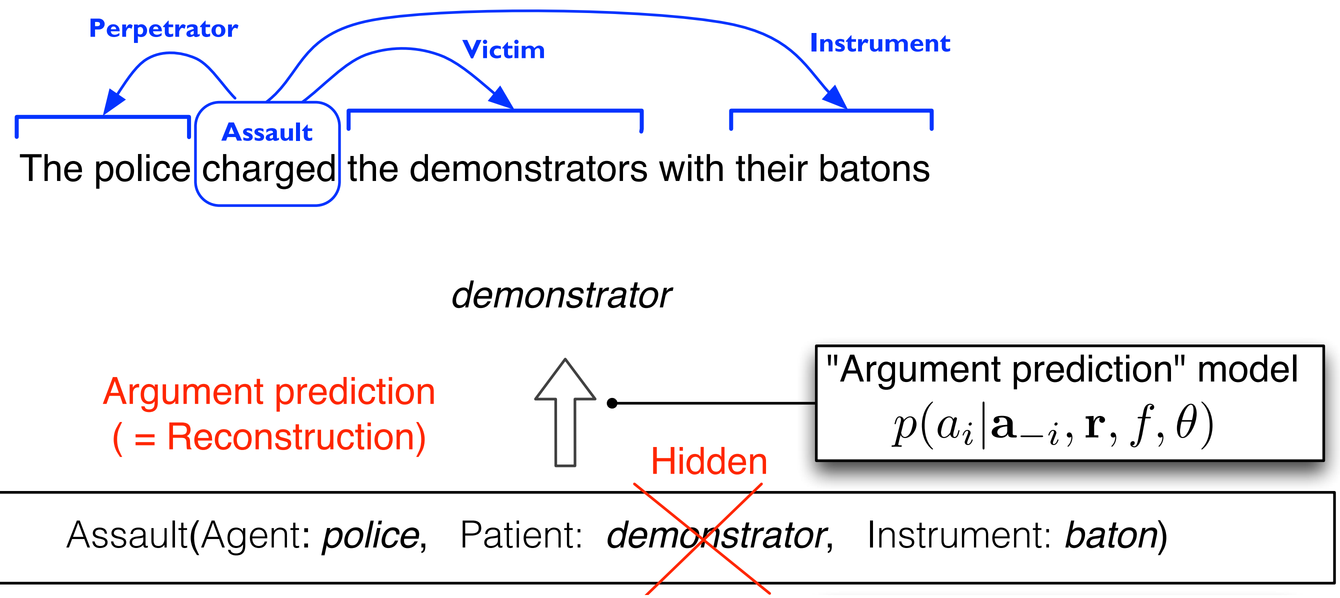
Latent — $\mathbf{r} = (r_1, \ldots, r_n)$ - roles (Perpetrator, Victim, Instrument)

f - frame (Assault)

[Titov and Khoddam, '15]

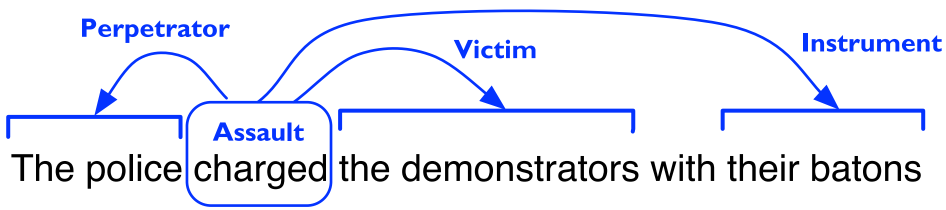
Argument reconstruction

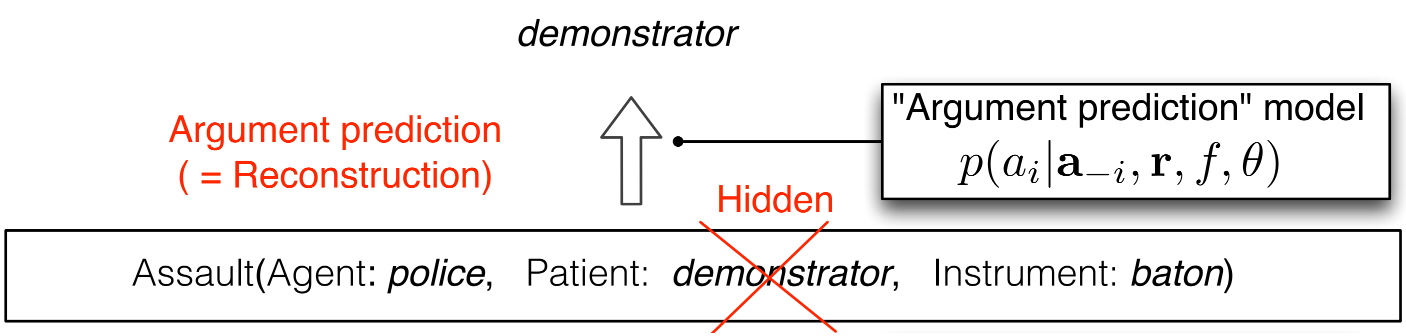
Consider a frame realization



Argument reconstruction

Consider a frame realization



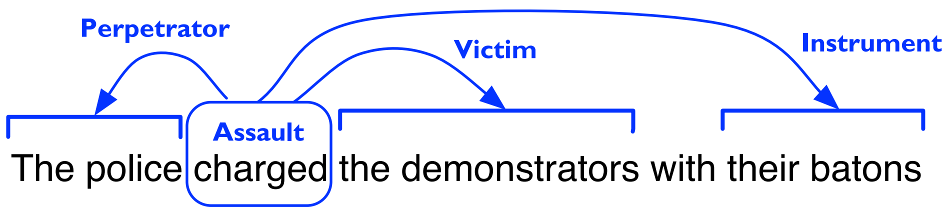


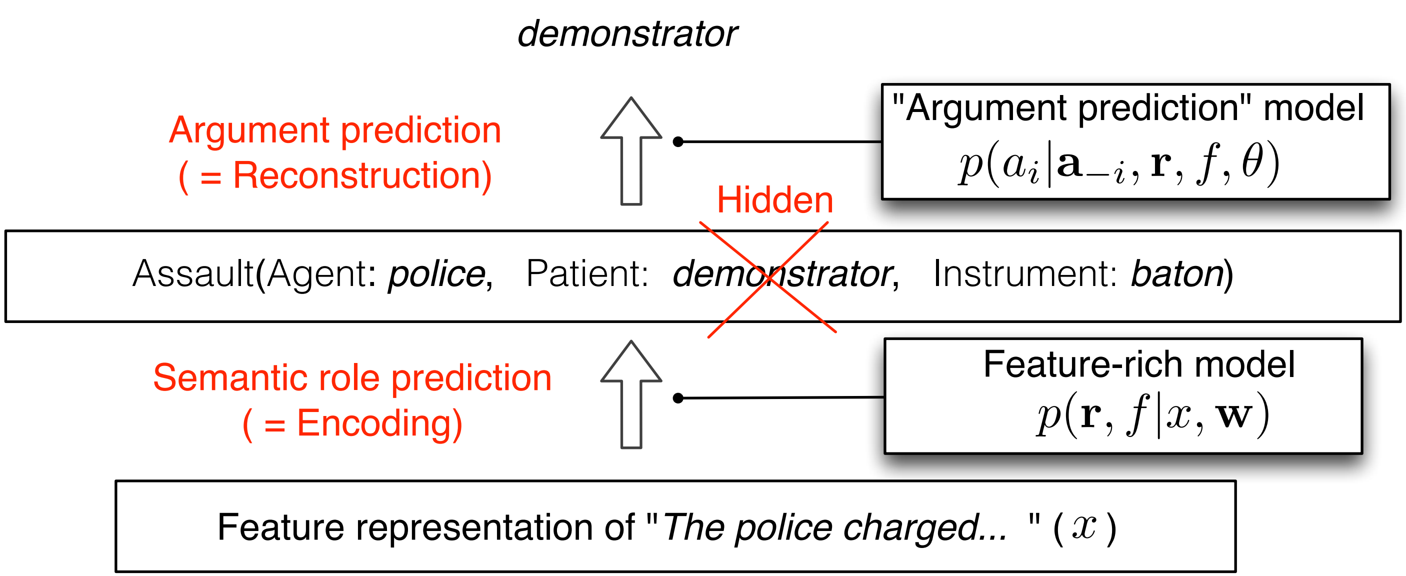
Hypothesis: semantic roles and frames are the latent representation which helps to reconstruct arguments

[Titov and Khoddam, ’15]

Argument reconstruction

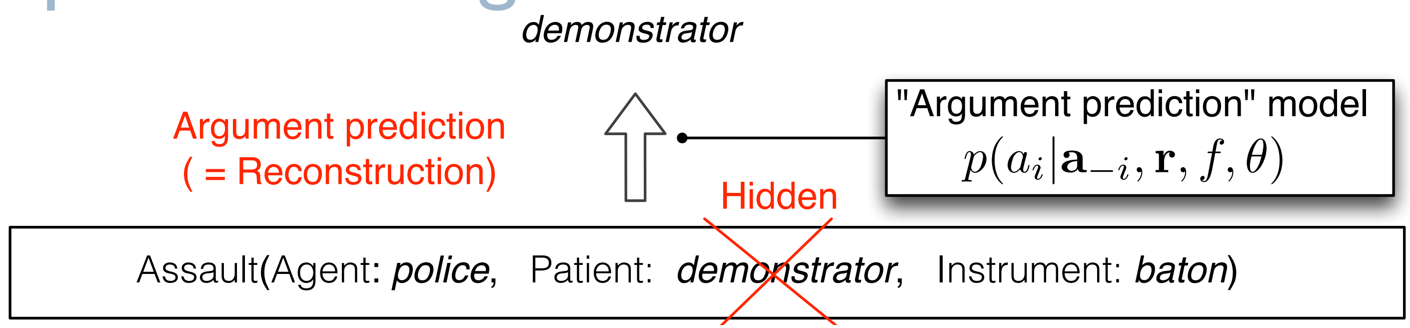
Consider a frame realization





How do the components look like and how do we estimate them jointly?

Component 1: argument reconstruction



Distributed vectors:

$\mathbf{u}_a \in \mathbb{R}^d$ - encode semantic properties of argument *a*

Component I: argument reconstruction

demonstrator

Argument prediction
(= Reconstruction)

"Argument prediction" model
$p(a_i | \mathbf{a}_{-i}, \mathbf{r}, f, \theta)$

Hidden

Assault(Agent: *police*, Patient: ~~*demonstrator*~~, Instrument: *baton*)

Distributed vectors:

$\mathbf{u}_a \in \mathbb{R}^d$ - encode semantic properties of argument *a*

$C_{f,r}\mathbf{u}_a \in \mathbb{R}^k$ - encode expectations about other arguments

A role-specific projection matrix

If Agent of Assault is the *police*, then Patient can be *demonstrators* or *protestors*

Component 1: argument reconstruction

demonstrator

Argument prediction
(= Reconstruction)

"Argument prediction" model
$p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, \theta)$

Hidden

Assault(Agent: *police*, Patient: ~~*demonstrator*~~, Instrument: *baton*)

Distributed vectors:

$\mathbf{u}_a \in \mathbb{R}^d$ - encode semantic properties of argument *a*

$C_{f,r}\mathbf{u}_a \in \mathbb{R}^k$ - encode expectations about other arguments

A role-specific projection matrix

If Agent of Assault is the *police*, then Patient can be *demonstrators* or *protestors*

$$p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, C, \mathbf{u}) = \frac{\exp(\mathbf{u}_{a_i}^T C_{f,r_i}^T \sum_{j \neq i} C_{f,r_j} \mathbf{u}_{a_j})}{Z(\mathbf{r}, f, i)}$$

Component I: argument reconstruction

demonstrator

Argument prediction
(= Reconstruction)

"Argument prediction" model
$p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, \theta)$

Hidden

Assault(Agent: *police*, Patient: ~~*demonstrator*~~, Instrument: *baton*)

Distributed vectors:

$\mathbf{u}_a \in \mathbb{R}^d$ - encode semantic properties of argument *a*

$C_{f,r}\mathbf{u}_a \in \mathbb{R}^k$ - encode expectations about other arguments

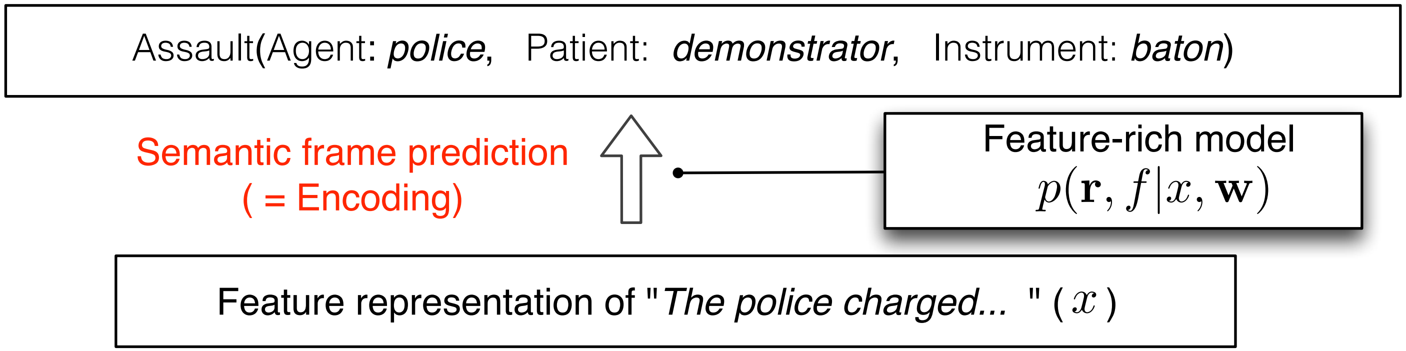
A role-specific projection matrix

If Agent of Assault is the *police*, then Patient can be *demonstrators* or *protestors*

$$p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, C, \mathbf{u}) = \frac{\exp(\boxed{\mathbf{u}_{a_i}^T C_{f,r_i}^T \sum_{j \neq i} C_{f,r_j} \mathbf{u}_{a_j}})}{Z(\mathbf{r}, f, i)}$$

Factorization function: scoring arguments tuples for a given frame and role assignment

Component 2: frame + role prediction



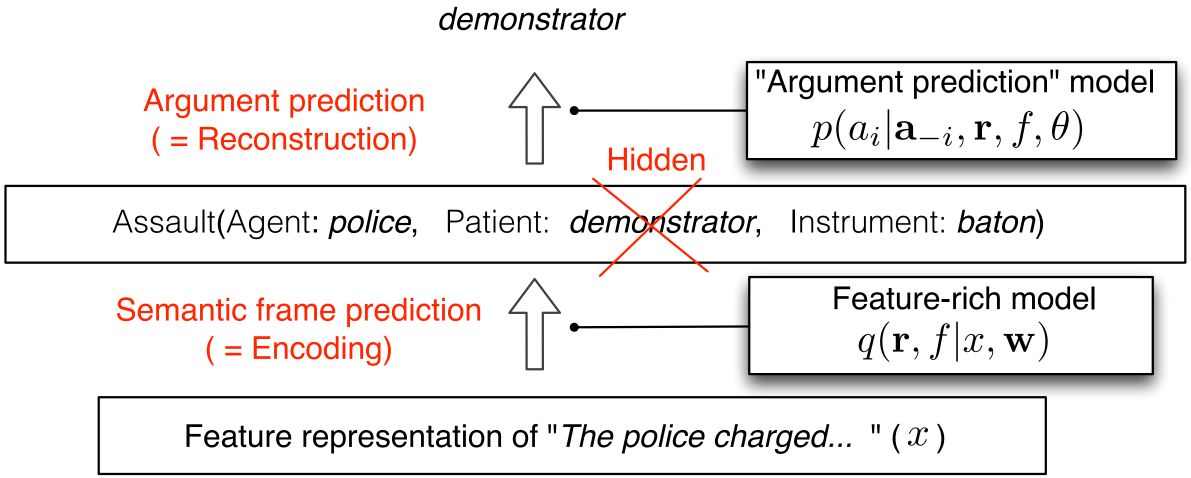
The role and frame labeling model:

$$p(\mathbf{r}, f | x, \mathbf{w}) \propto \exp(\mathbf{w}^T \mathbf{g}(x, f, \mathbf{r}))$$

A feature representation of text

It can be any model as long as role and frame posteriors $p(r_i | x, \mathbf{w})$ and $p(f | x, \mathbf{w})$ can be computed (or approximated)

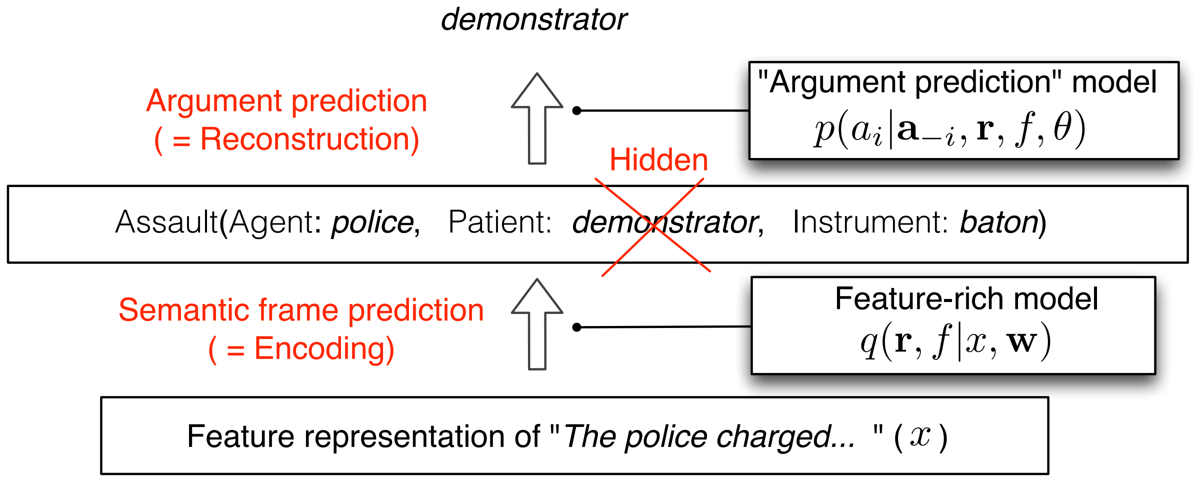
Joint learning



- For every structure, we aim to optimize the argument prediction quality given roles and frames:

$$\sum_{i=1}^{N} \sum_{\mathbf{r},f} q(\mathbf{r}, f|x, \mathbf{w}) \log p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, C, \mathbf{u})$$

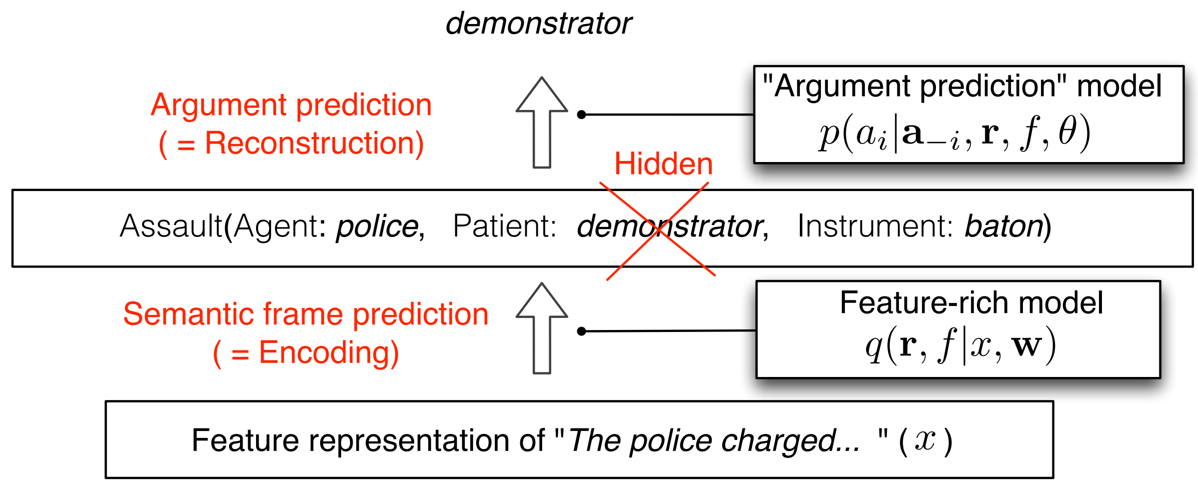
Joint learning



- For every structure, we aim to optimize the argument prediction quality given roles and frames:

$$\sum_{i=1}^{N} \sum_{\mathbf{r},f} q(\mathbf{r}, f|x, \mathbf{w}) \underbrace{\log p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, C, \mathbf{u})}_{\text{reconstruction}}$$

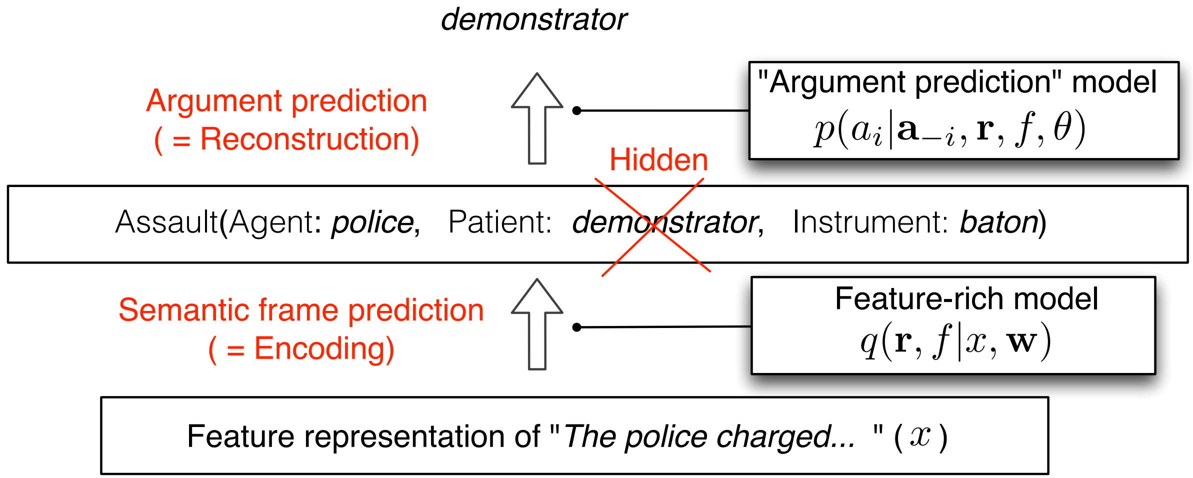
Joint learning



- For every structure, we aim to optimize the argument prediction quality given roles and frames:

$$\sum_{i=1}^{N} \underbrace{\sum_{\mathbf{r},f} q(\mathbf{r}, f|x, \mathbf{w})}_{\text{semantic parser}} \underbrace{\log p(a_i|\mathbf{a}_{-i}, \mathbf{r}, f, C, \mathbf{u})}_{\text{reconstruction}}$$

Joint learning



- For every structure, we aim to optimize the argument prediction quality given roles and frames:

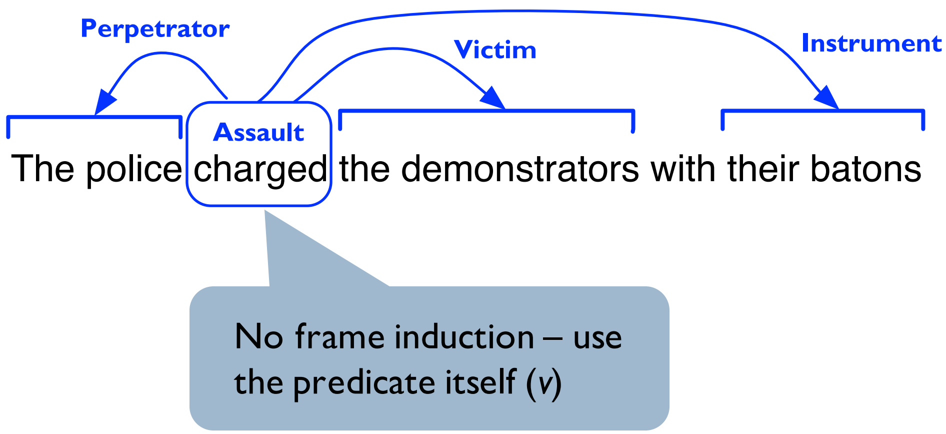
$$\sum_{i=1}^{N} \sum_{\mathbf{r},f} \underbrace{q(\mathbf{r},f|x,\mathbf{w})}_{\text{semantic parser}} \underbrace{\log p(a_i|\mathbf{a}_{-i},\mathbf{r},f,C,\mathbf{u})}_{\text{reconstruction}}$$

Technical details in Marcheggiani and Titov (2016)

Outline

- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: role induction, frame induction

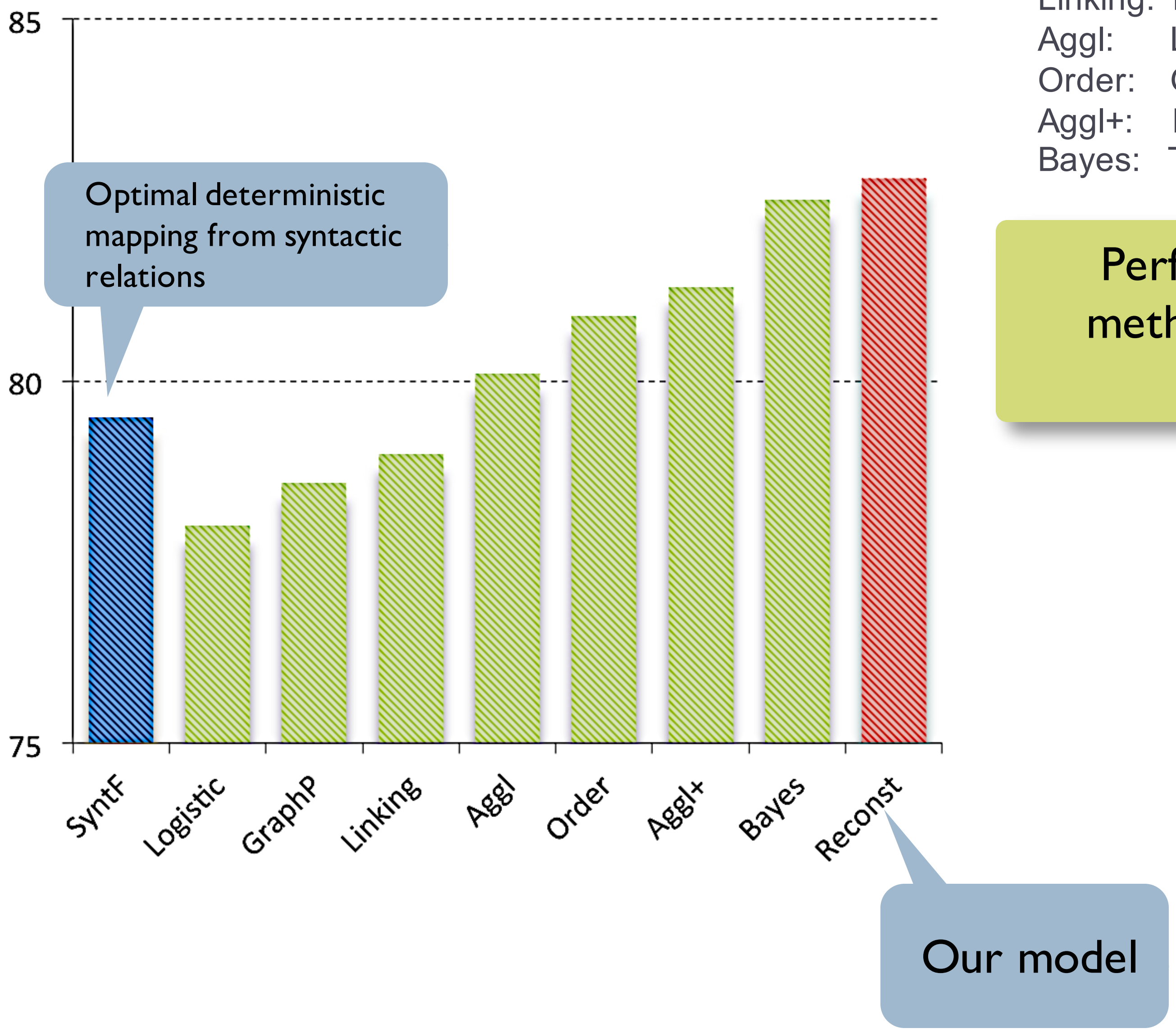
Semantic roles induction



- Task: induce for each argument the appropriate semantic role
- Evaluate on a dataset annotated with roles (PropBank for En, SALSA for De)
- Compare against previous models evaluated in this set-up
  - use clustering evaluation measures (purity, collocation, F1)

Results English (F1)

Logistic: Lang and Lapata ('10)
GraphP: Lang and Lapata ('11a)
Linking: Fürstenau and Rambow ('12)
Aggl: Lang and Lapata ('11b)
Order: Garg and Henderson ('12)
Aggl+: Lang and Lapata ('14)
Bayes: Titov and Klementiev ('12)

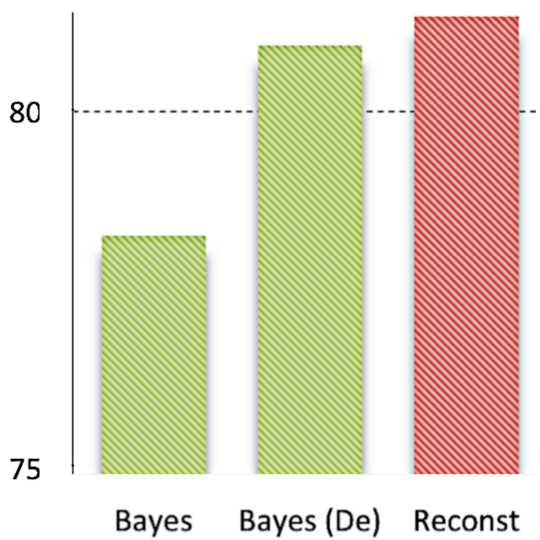


Performs on par with best methods (without language-specific priors)

[Titov and Khoddam, ’15]

Results German (F1)

Bayes: Titov and Klementiev ('12a)
Bayes (De): Titov and Klementiev ('12b)

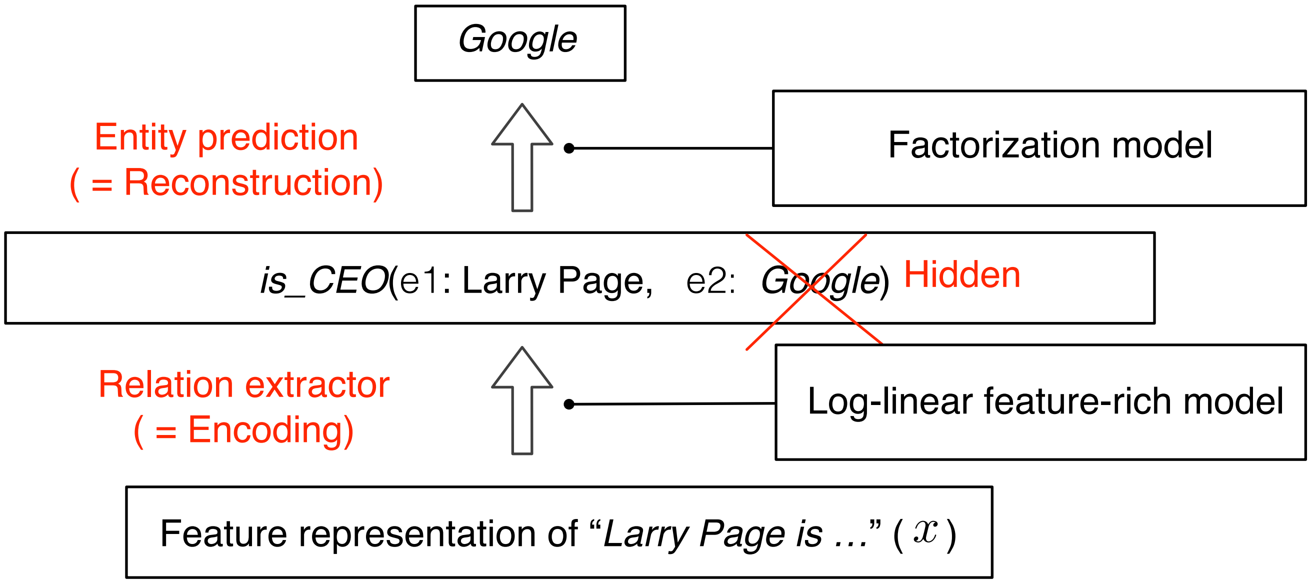


Performs on par with the best method without language-specific engineering

Bayes modified for German

The feature rich model (the same as for En)

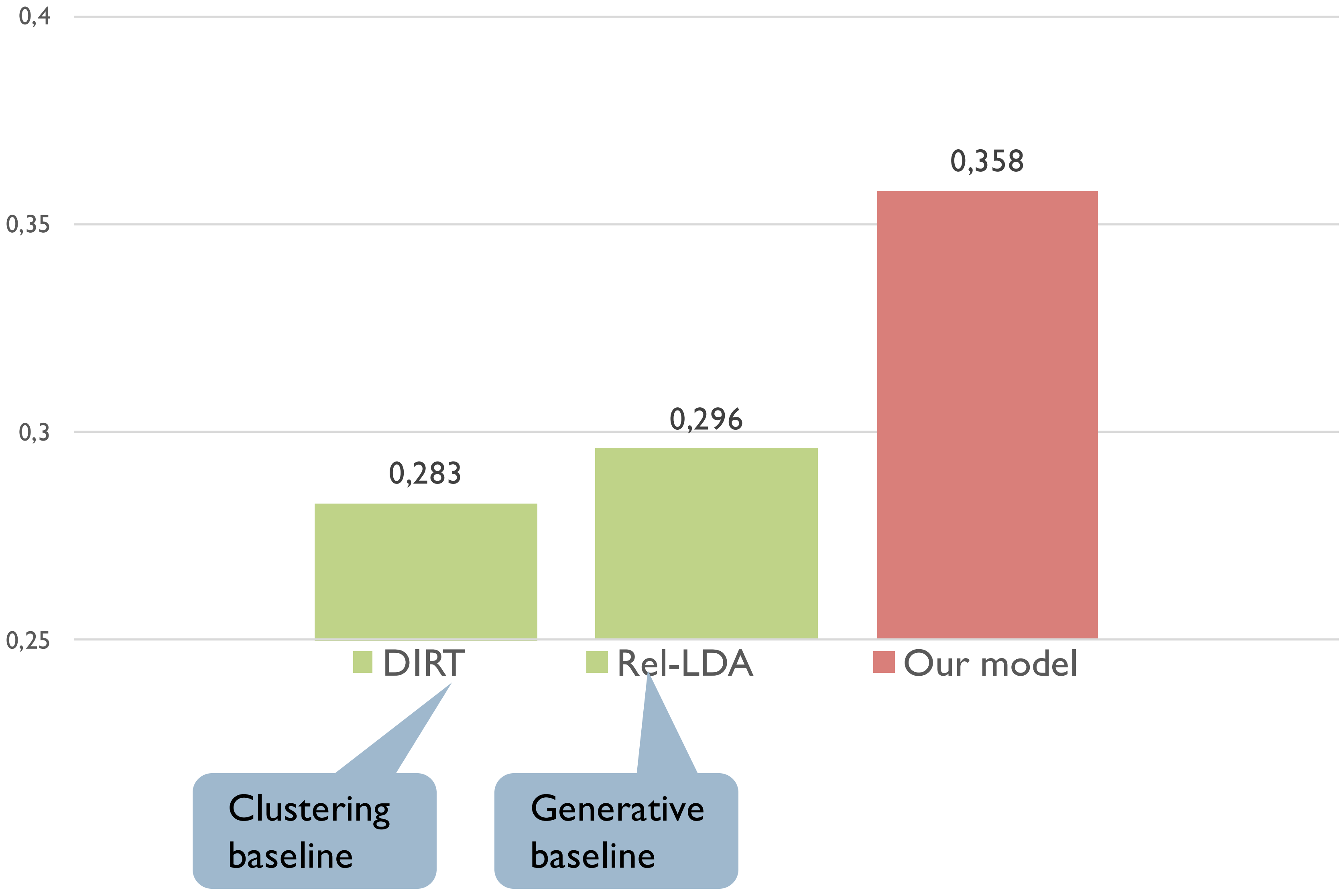
Relation discovery (Frame induction)



- Task: induce semantic relations between two given arguments
- Data: New York Times corpus
- Evaluation against Freebase

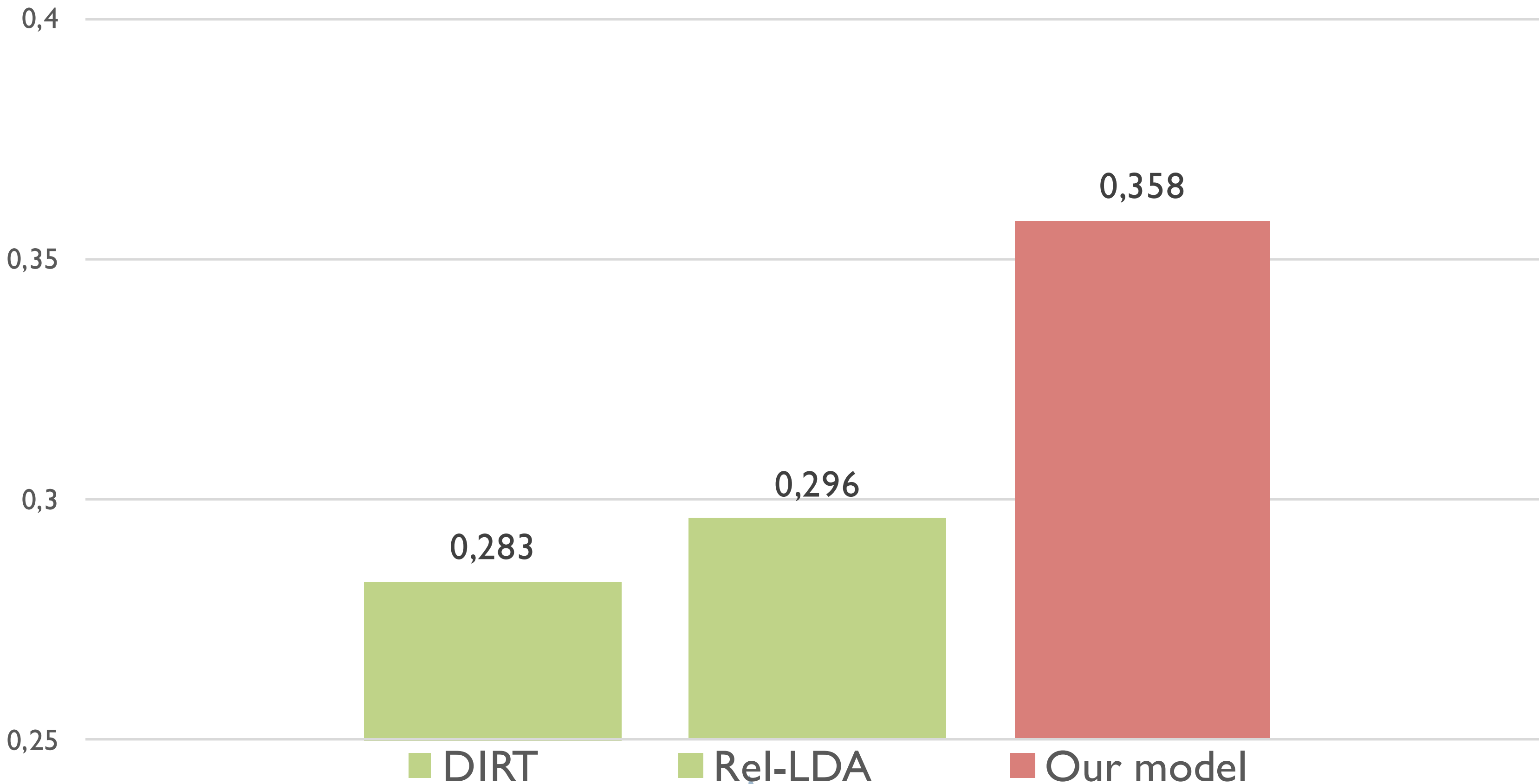
[Marcheggiani and Titov, TACL '16]

Results (F1)



[Marcheggiani and Titov, TACL '16]

Results (F1)



Clustering baseline

Generative baseline

6.2% more accurate than the Rel-LDA baseline.

Conclusion

- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - expressive features
  - supports semi-supervised learning

Conclusion

- A new framework for inducing shallow semantics
  - combining ideas from relation modeling and semantic parsing
  - expressive features
  - supports semi-supervised learning

Code available

Thank you!

Funding:

NWO VIDI grant

Google Focused Award on Natural Language Understanding