Inducing Shallow Semantic Representations from Text

Ivan Titov

University of Amsterdam

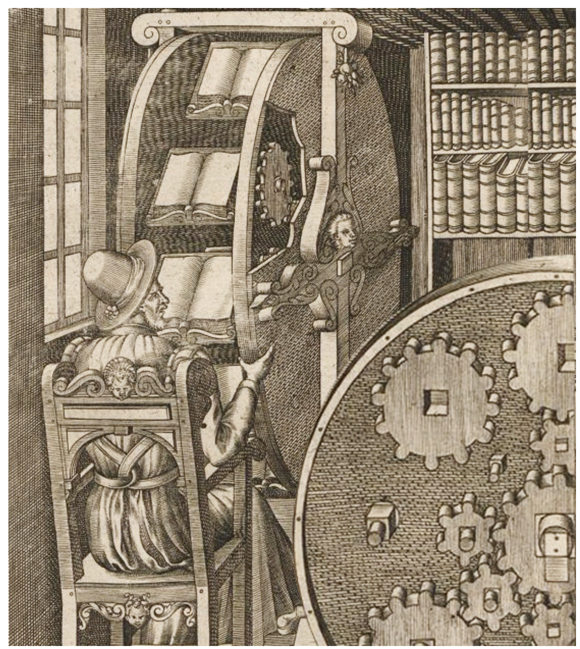
Joint work with Diego Marcheggiani and Ehsan Khoddam

UNIVERSITEIT VAN 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



Machine reading

Information retrieval

Machine reading

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

....

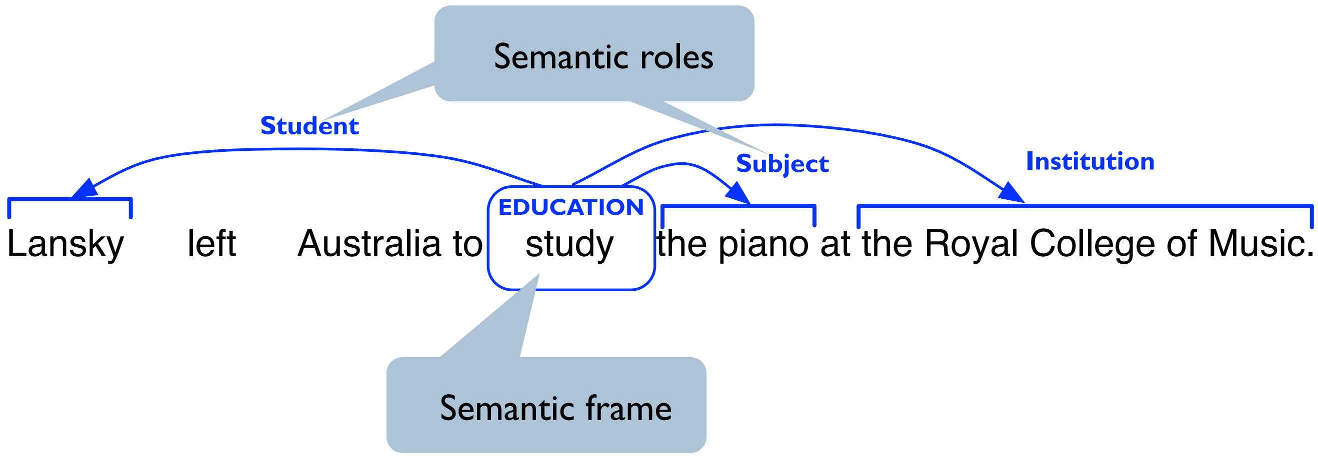
Lansky dropped his studies at RCM, but eventually graduated from Trinity.

1. Where did Lansky get his diploma?
2. Where did he live?
3. What does he do?

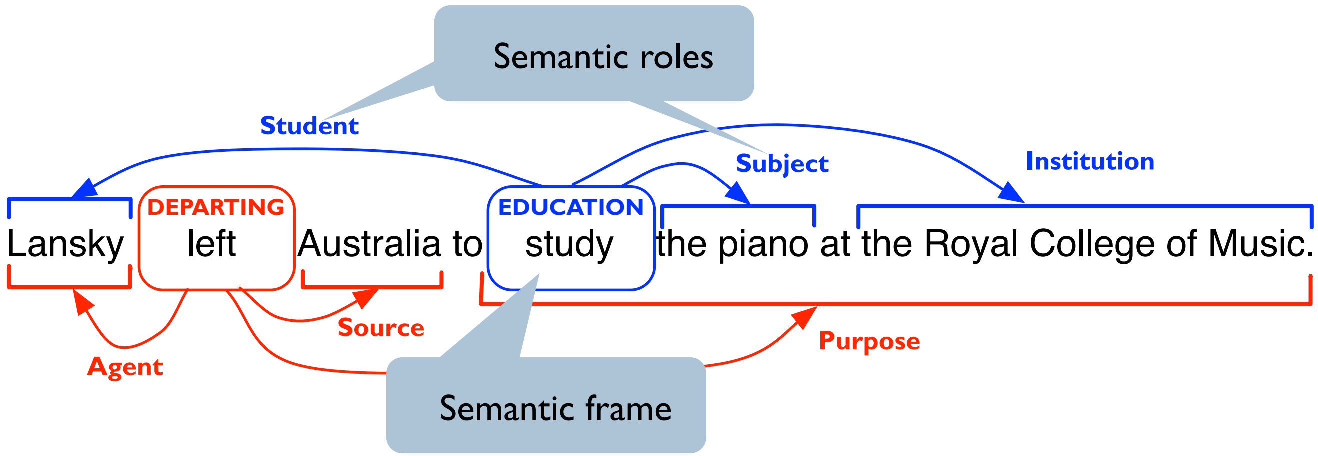
Frame-semantic parsing

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

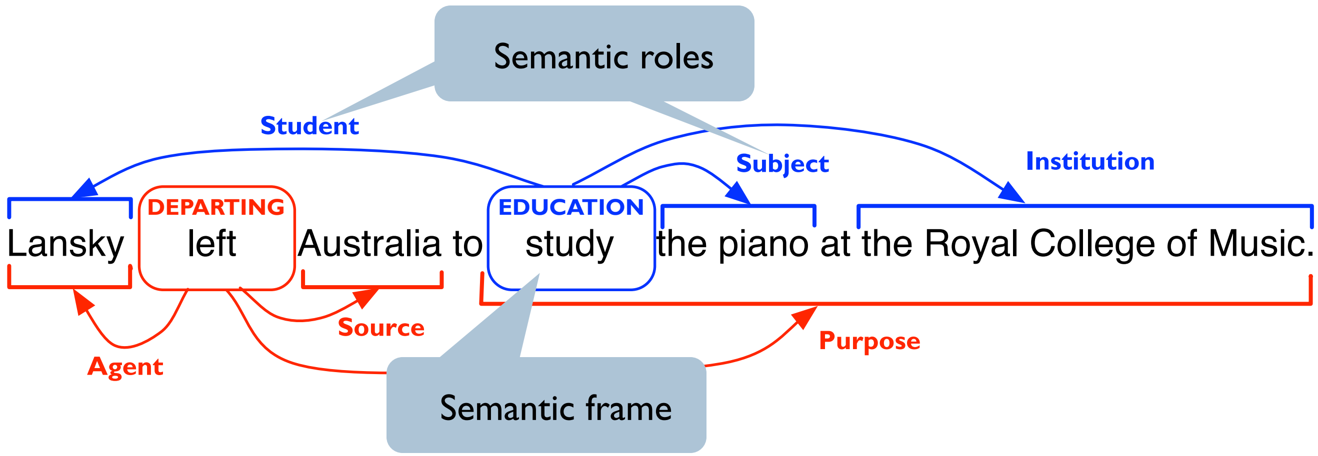
Frame-semantic parsing



Frame-semantic parsing



Frame-semantic parsing



- Intuitively, a frame-semantic parser extracts knowledge from text into a relational database

Frames are tables, roles are attributes

DEPARTING

Object	Source	Purpose	...

EDUCATION

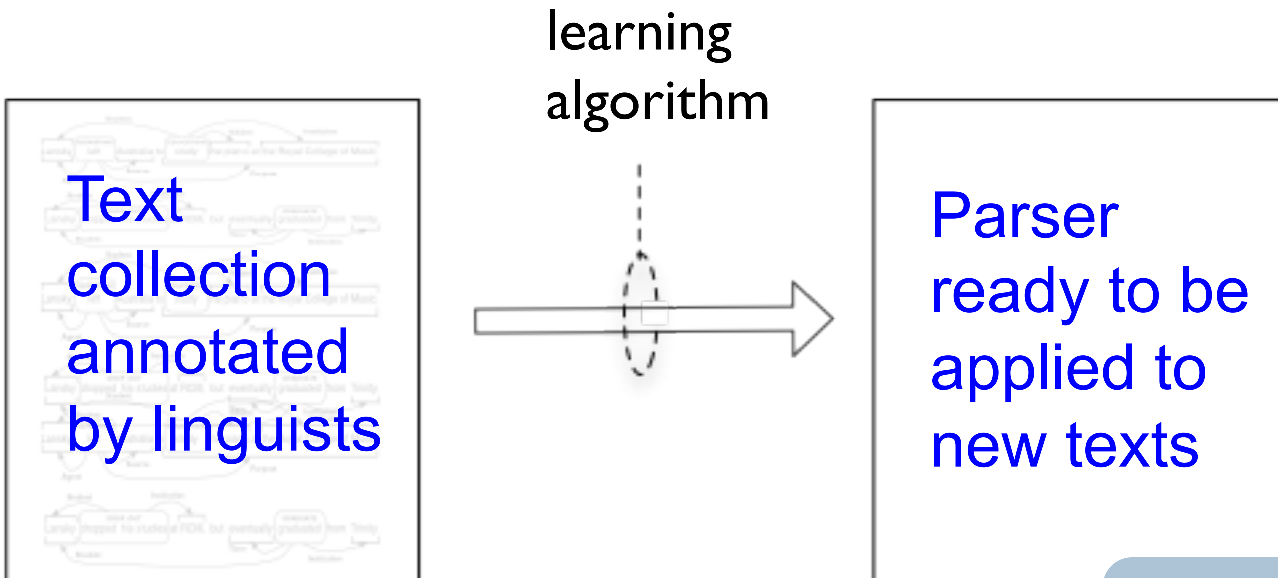
Student	Institution	Subject	...
Lansky	Royal College of Music	piano	...
...	...	...	...

Outline

- Motivation: why we need unsupervised feature-rich models and learning for inference
- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: preliminary experiments, insights, future work

Modern semantics parsers

Modern frame-semantic parsers rely on supervised learning

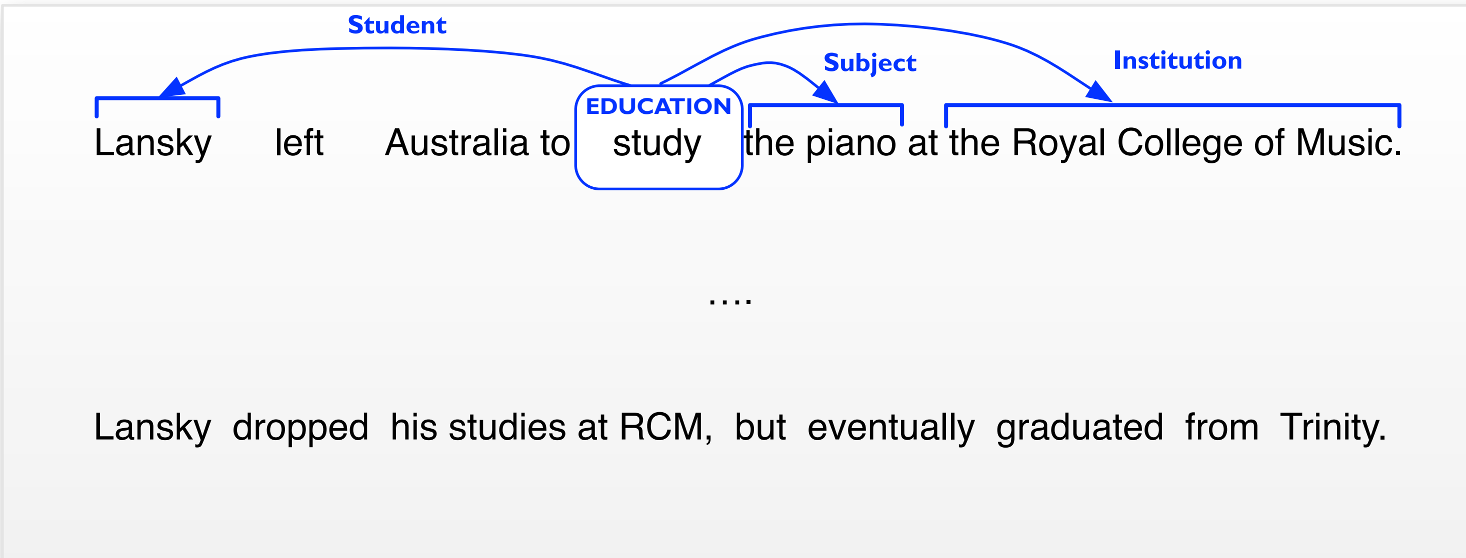


Especially problematic for tail phenomena

Challenge #1

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

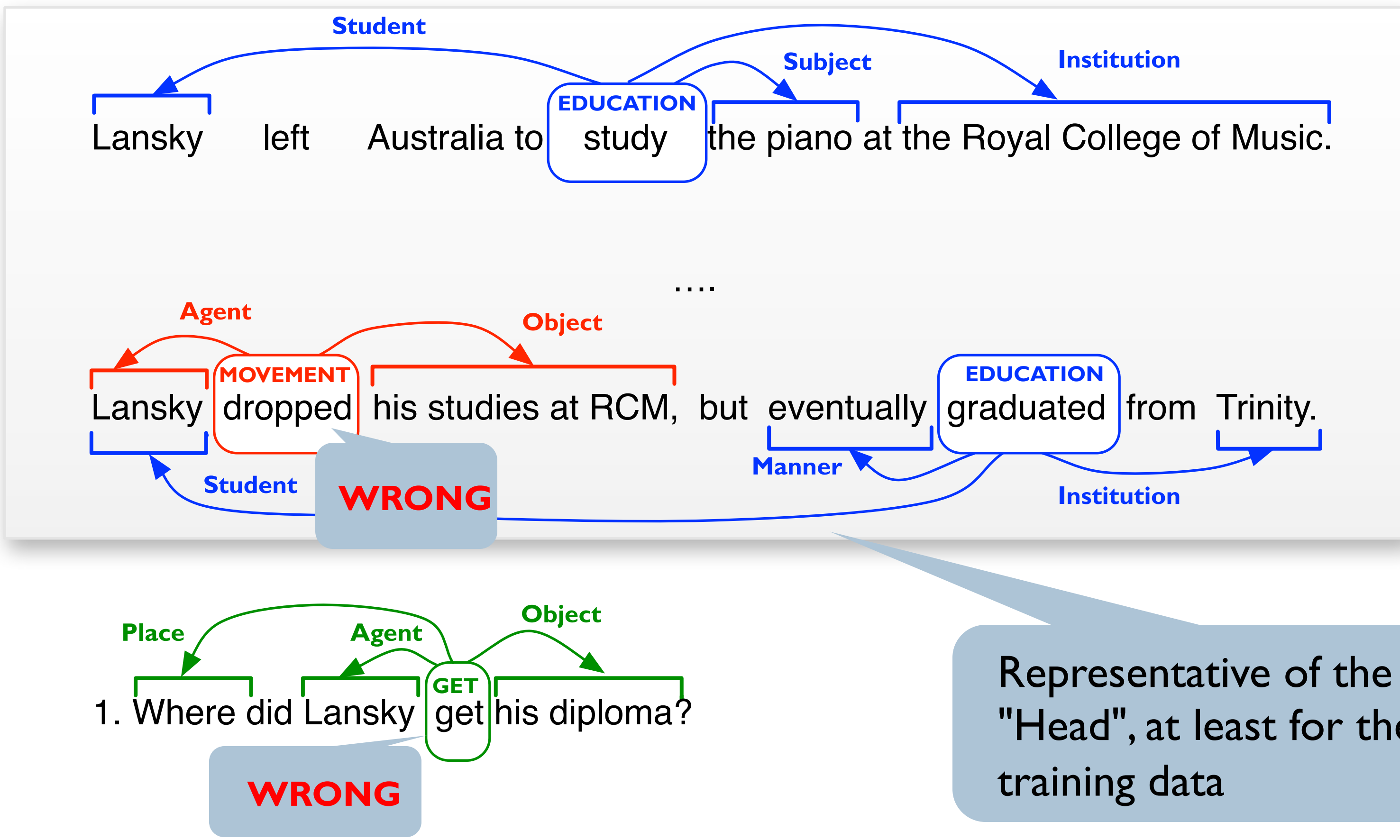
Machine reading



1. Where did Lansky get his diploma?

Output of a state-of-the-art parser

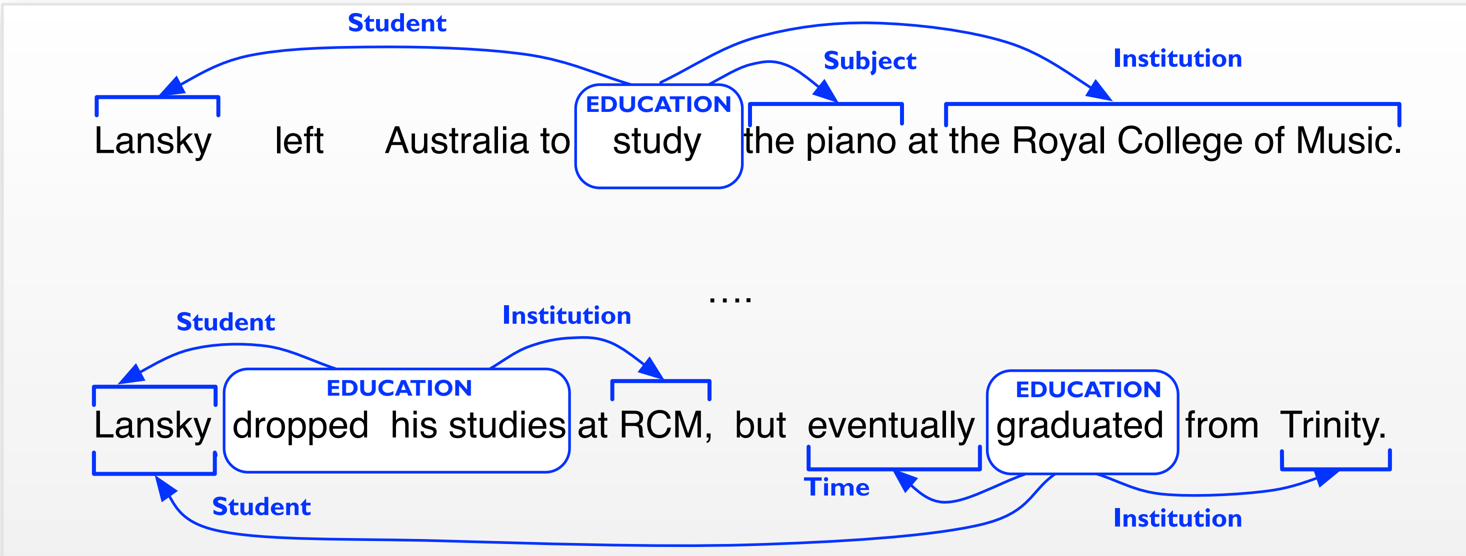
CMU's SEMAFOR [Das et al., 2012] trained on 100,000 sentences (FrameNet)

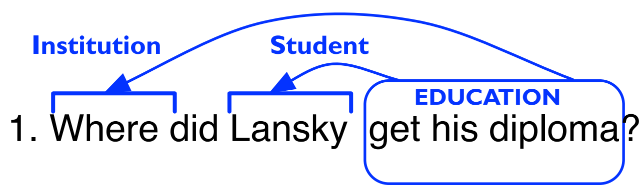


Representative of the "Head", at least for the training data

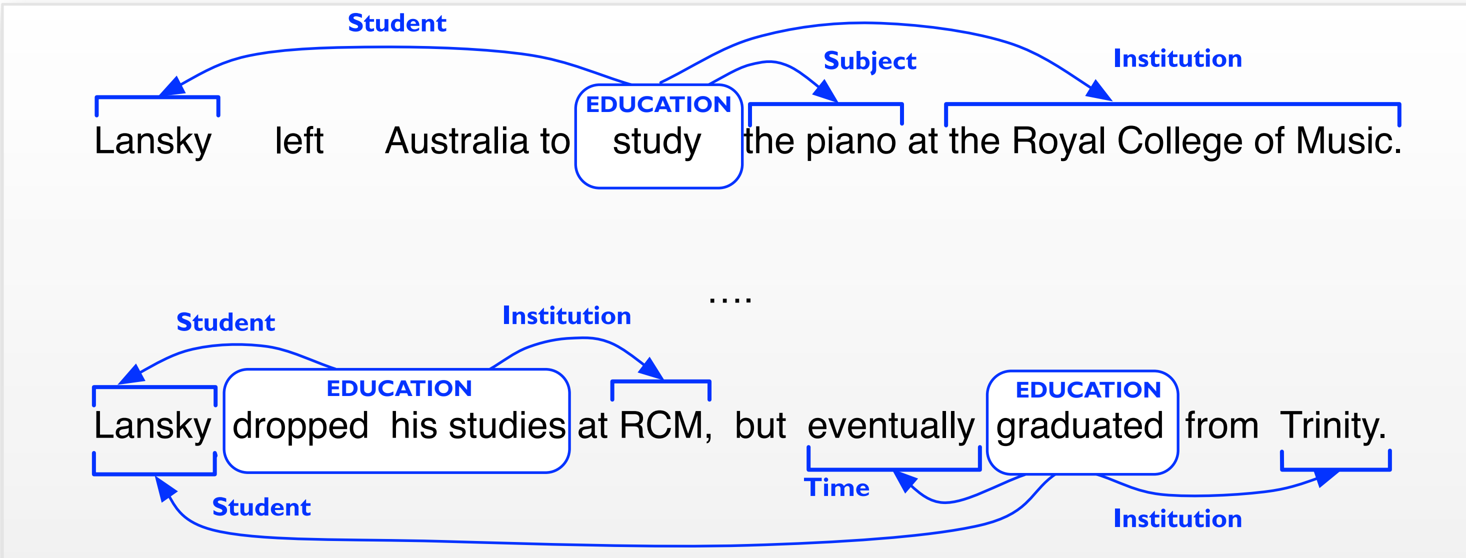
The parser's output does not let us answer even this simple question

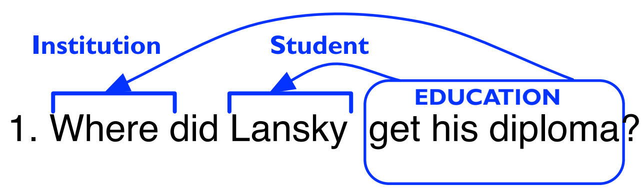
"Correct" semantics as imposed by linguists





"Correct" semantics as imposed by linguists





Trinity or RCM ????

Challenge #2

Representations defined by linguists are not appropriate for reasoning (i.e. inference)

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; Fürstenau and Rambow, '12;…]
  - Frame induction [Titov and Klementiev '11; O' Connor '12; Modi et al.'12; Materna '12; Lorenzo and Cerisara '12; Kawahara et al. '13; Cheung et al. '13; Chambers et al., 14; …]

Unsupervised frame and role induc

In contrast to supervised methods to frame-semantic parsing / semantic role labeling

- The models rely on very restricted sets of features
  - not very effective in the semi-supervised set-up, and not very appropriate for languages with freer order than English
- … over-rely on syntax
  - not going to induce, e.g., "X *sent* Y = Y *is a shipment from* X"
- … use language-specific priors
  - a substantial drop in performance if no adaptation
- … not (quite) appropriate for inference
  - not only no inference models but also opposites and antonyms (e.g., increase + decrease) are typically grouped together; induced granularity is often problematic; …

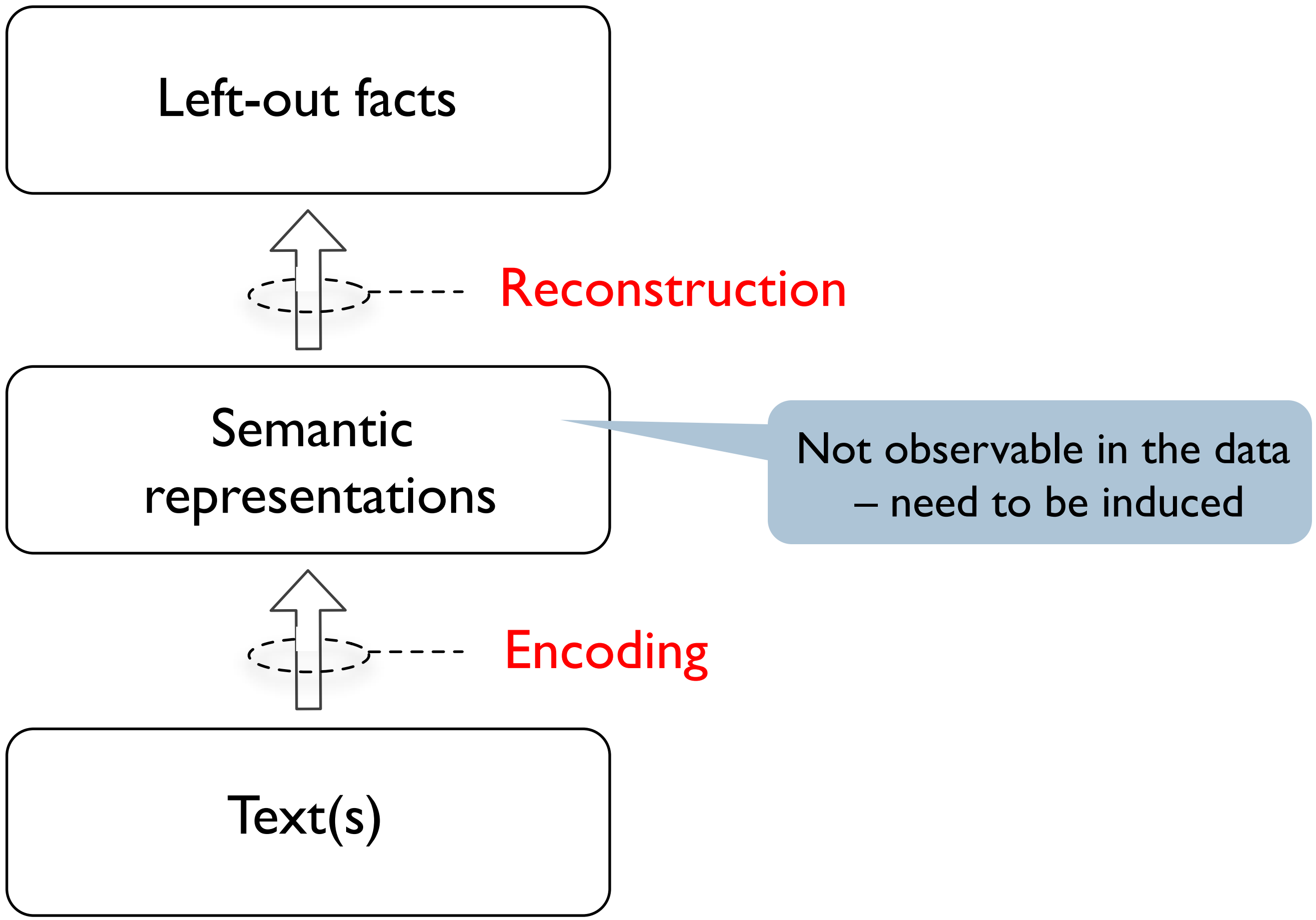
How can we induce frames in a less restrictive feature-rich framework and tackle other challenges along the way?

Need expressive models for dealing with less frequent and "direct" realizations of relations / frames

Outline

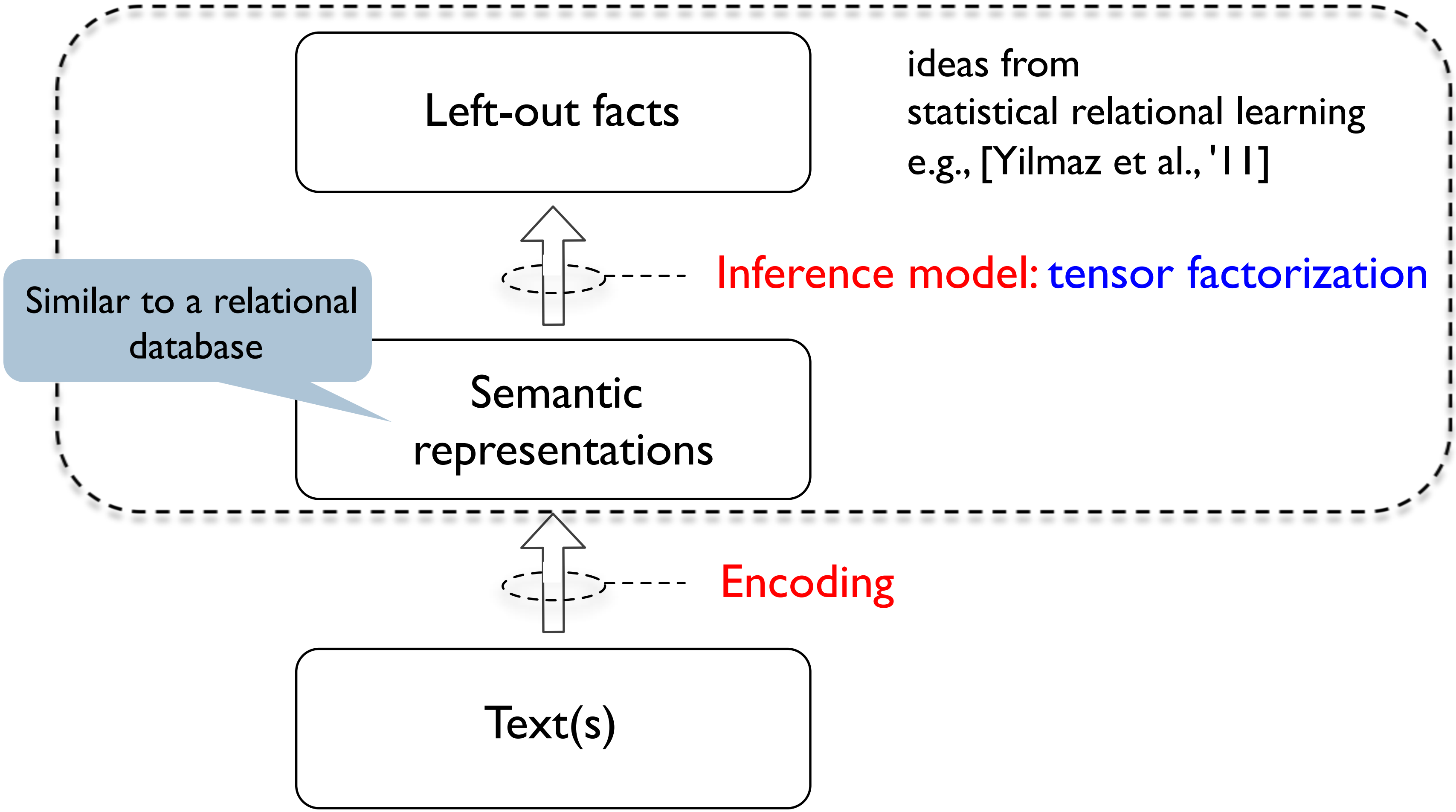
- Motivation: why we need unsupervised feature-rich models and learning for inference
- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: preliminary experiments, insights, future work

Idea: estimating the model

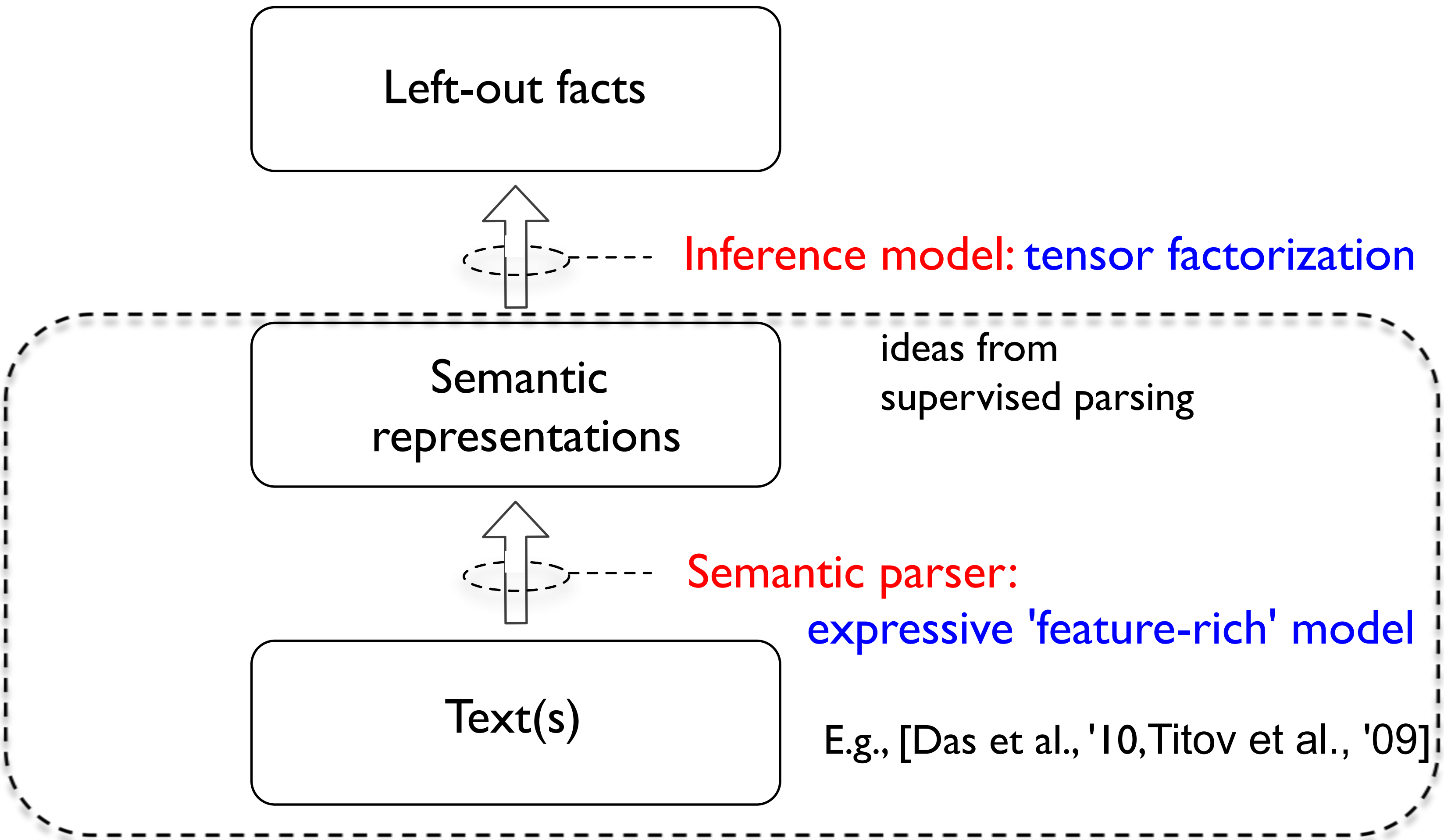


Instead of using annotated data, induce representations beneficial for inferring left-out facts

Idea: estimating the model

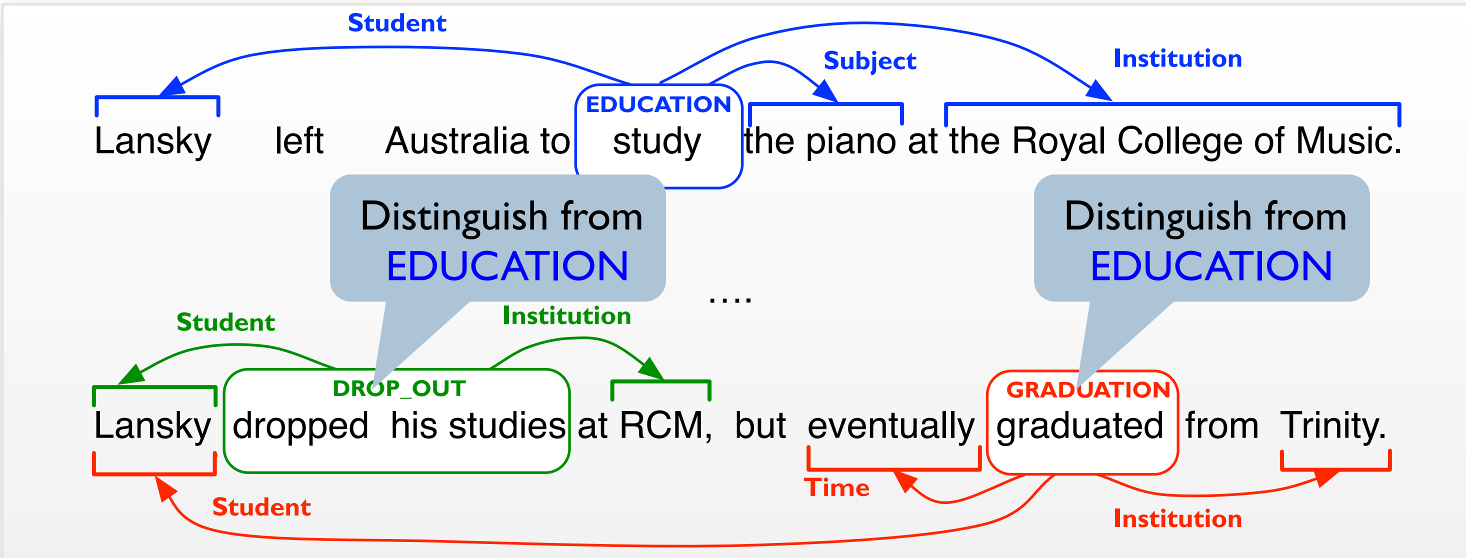


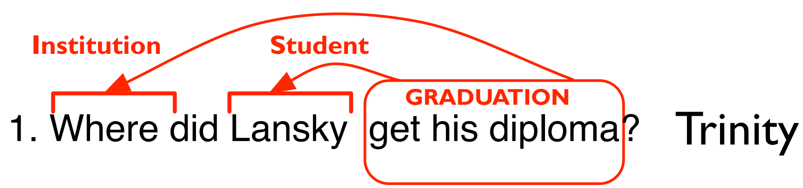
Idea: estimating the model



Inference model and semantic parser are jointly estimated from unannotated data

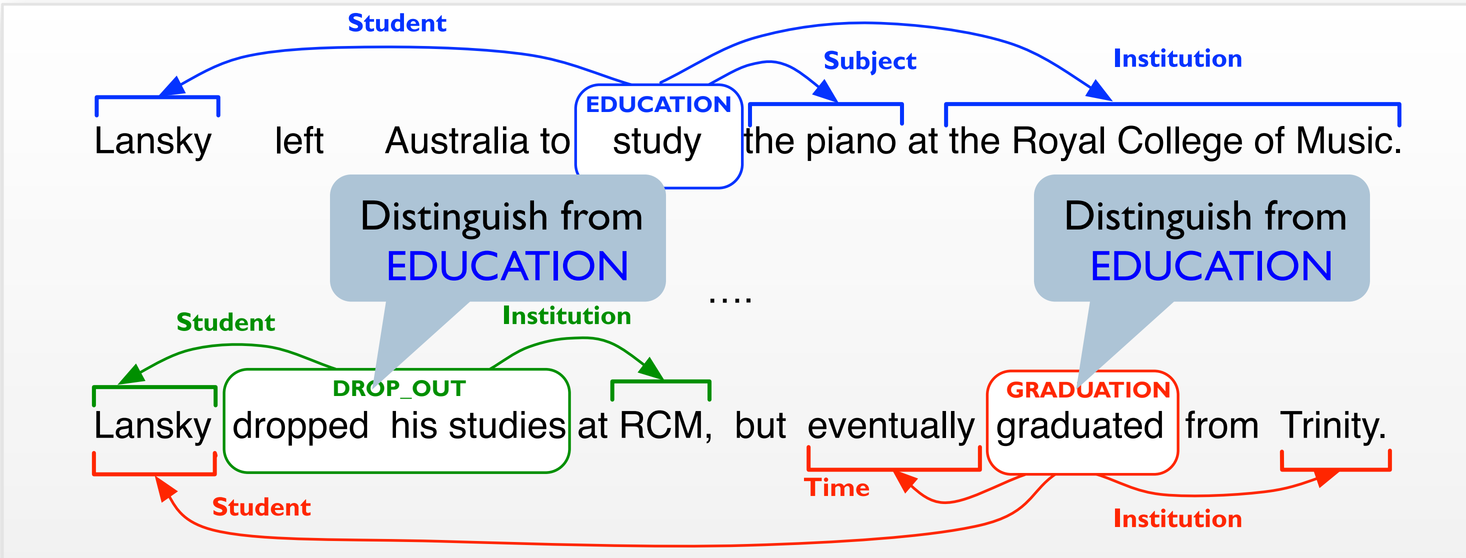
When learning for reasoning





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

When learning for reasoning

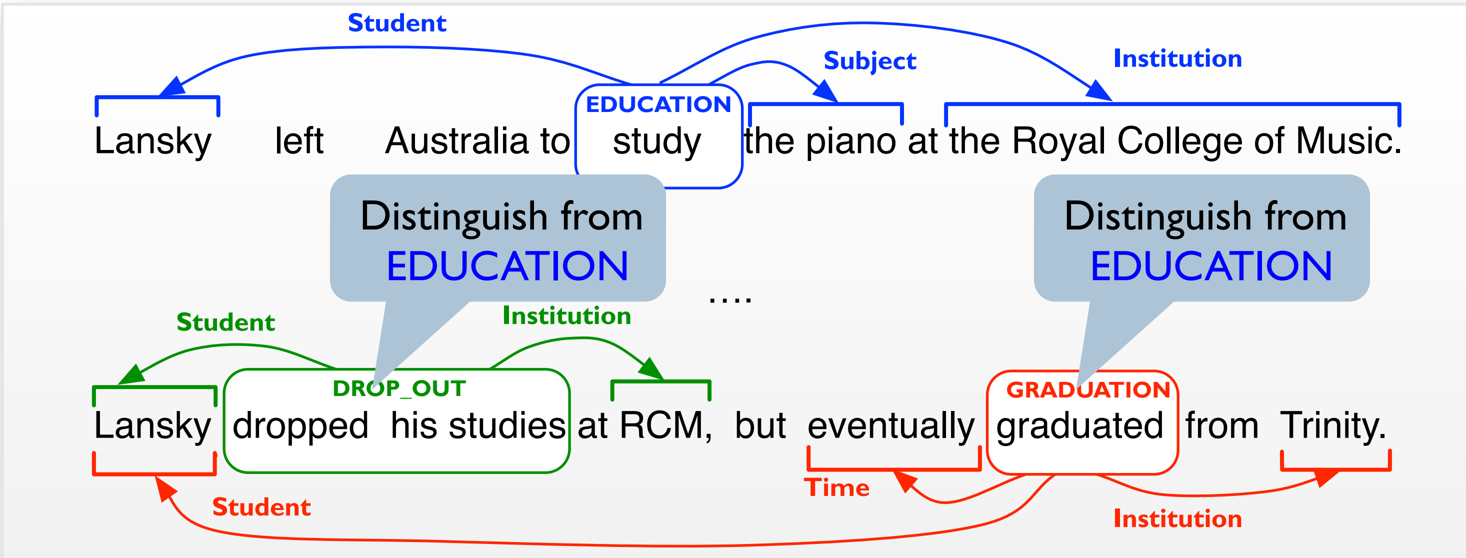


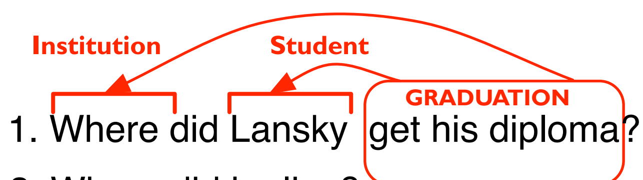
Institution · Student · GRADUATION

1. Where did Lansky get his diploma? Trinity
2. Where did he live? Australia and United Kingdom
3. What does he do?

Inference component can support 'reading between the lines'

When learning for reasoning





Trinity

2. Where did he live? Australia and United Kingdom

3. What does he do? He is a pianist (??)

Inference component can support 'reading between the lines'

Outline

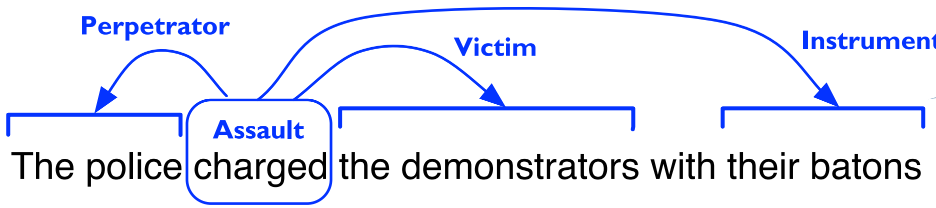
- Motivation: why we need unsupervised feature-rich models and learning for inference
- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: preliminary experiments, insights, future work

[Titov and Khoddam, '14]

Feature-rich models of semantic frames

Consider a frame realization

For simplicity: focus on frame and role labeling (no identification + one frame per sentence)



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)

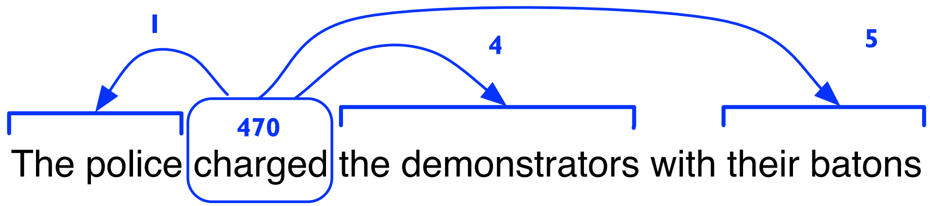
f - frame (Assault)

[Titov and Khoddam, '14]

Feature-rich models of semantic frames

Consider a frame realization

For simplicity: focus on frame and role labeling (no identification + one frame per sentence)



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)

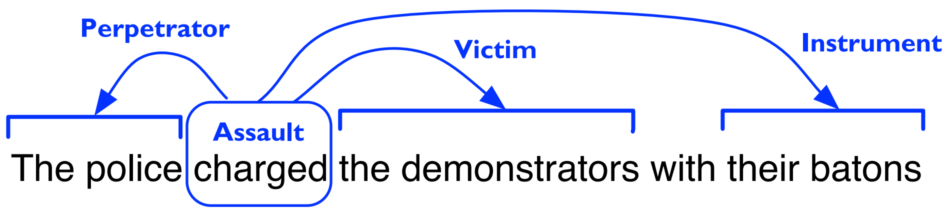
f - frame (Assault)

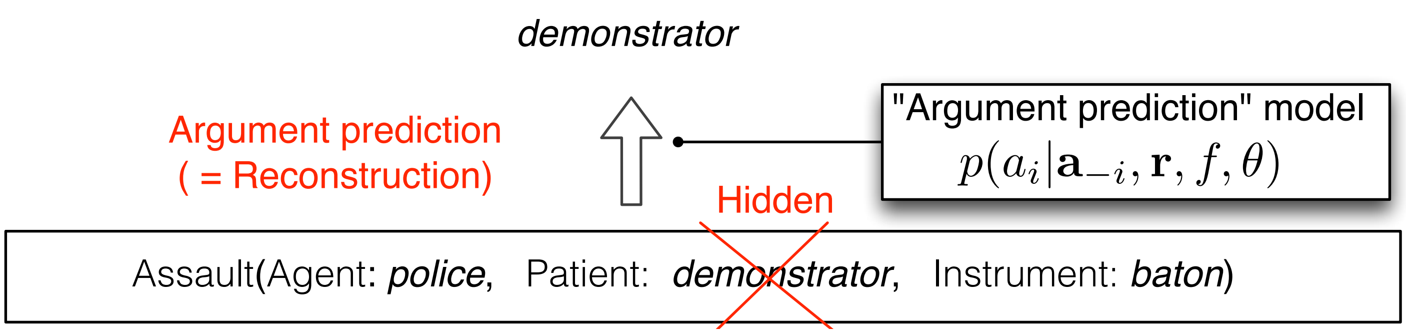
How can we define a feature-rich model for unsupervised induction of roles and frames?

[Titov and Khoddam, '14]

Argument reconstruction

Consider a frame realization



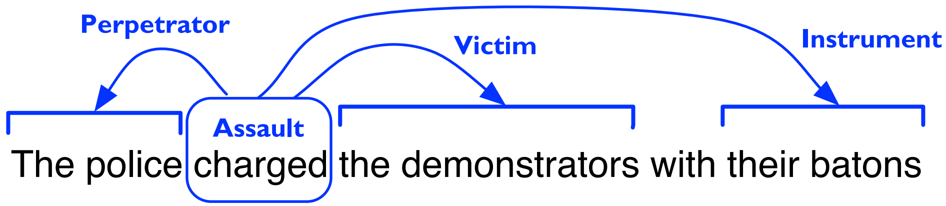


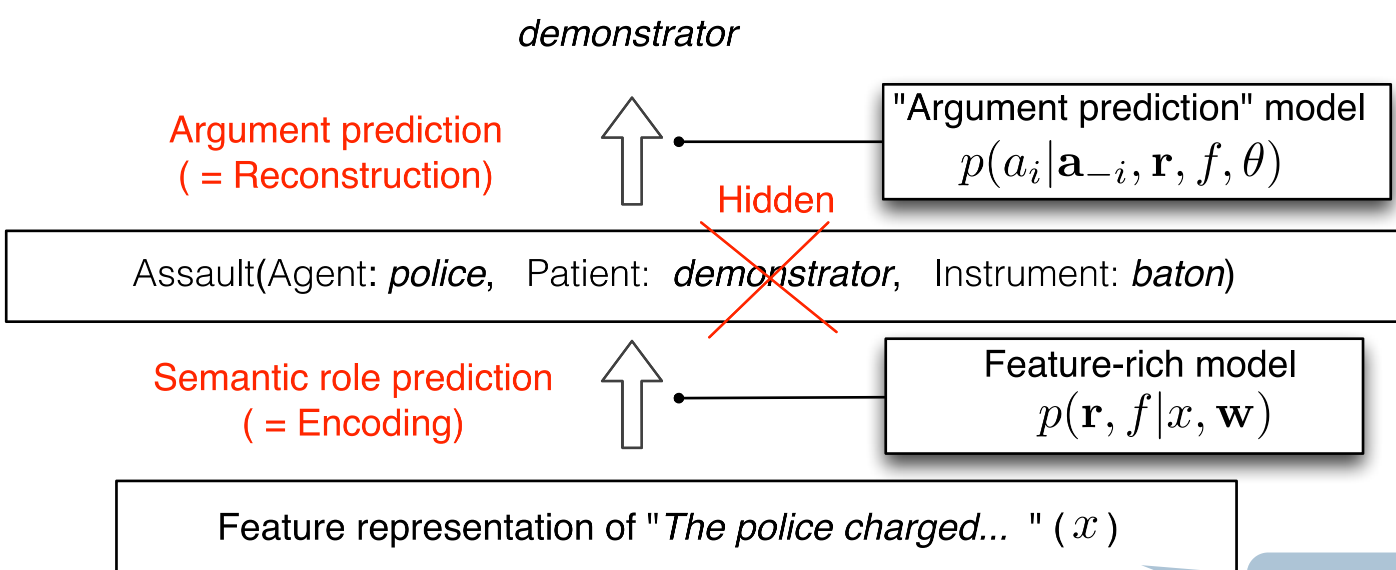
Hypothesis: semantic roles and frames are the latent representation which helps to reconstruct arguments

[Titov and Khoddam, '14]

Argument reconstruction

Consider a frame realization



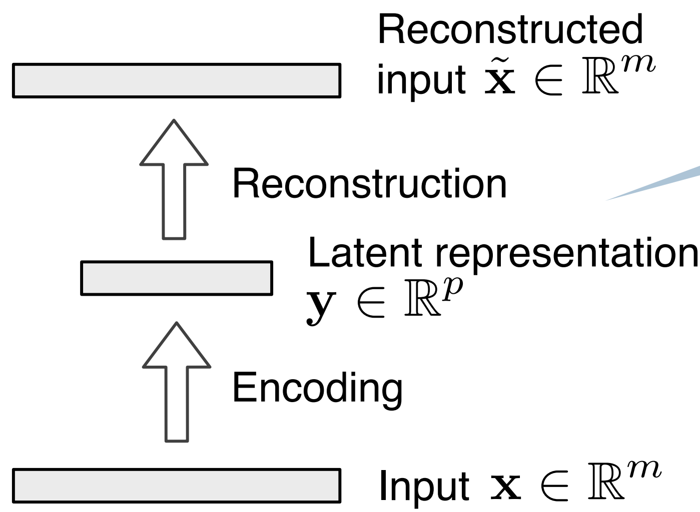


Any existing supervised role labeler would do

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

Reconstruction-error minimization

Neural autoencoders [Hinton '99, Vincent et al. 08]:



Trained to minimize the reconstruction error, for example, e.g., $||\mathbf{x} - \tilde{\mathbf{x}}||_2$

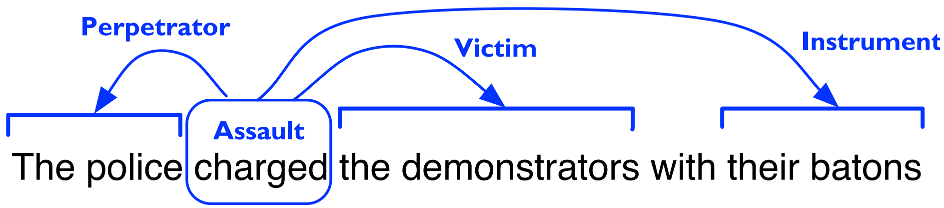
but

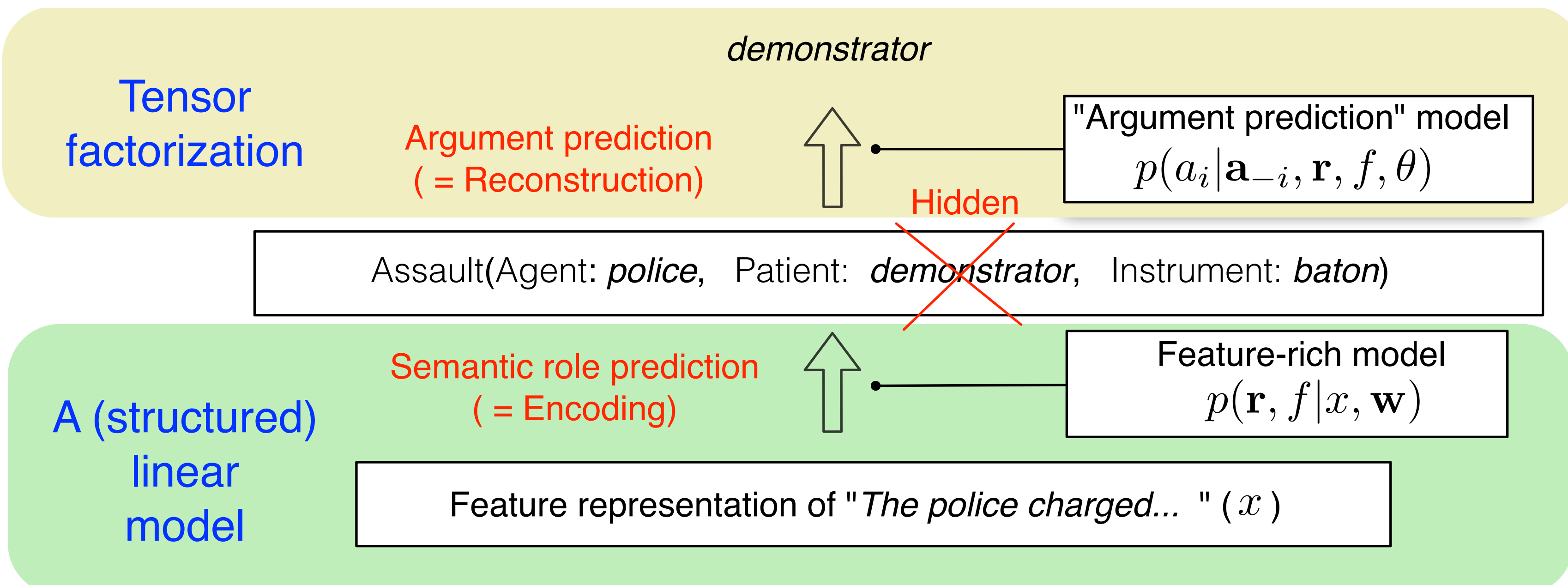
- … applicable not only to neural models
- … reconstruction and encoding components can belong to different model families
- … no need to reconstruct the entire input

See Titov and Khoddam ('14), Ammar et al. ('14) and also Daumé ('09)

Argument reconstruction

Consider a frame realization





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*)

May encode that *demonstrators* are similar to *protestors*

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 argument given that *a* is assigned to role *r* of frame *f*

A role-specific projection matrix

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

The reconstruction model:

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

Component 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*)

May encode that *demonstrators* are similar to *protestors*

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 argument given that *a* is assigned to role *r* of frame *f*

A role-specific projection matrix

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

Intuitively, score argument tuples according to the factorization:

$$\sum_{i \neq j} \mathbf{u}_{a_i}^T C_{f,r_i}^T C_{f,r_j} \mathbf{u}_{a_j}$$

Parallels to work on relation modeling (e.g., Bordes et al.,'11), distributional semantics (e.g., Mikolov et al., '13) or (coupled) tensor factorization (e.g.,Yilmiz et al., '11)

Component 2: frame + role prediction

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

Semantic frame prediction
(= Encoding)

Feature-rich model
$p(\mathbf{r}, f | x, \mathbf{w})$

Feature representation of "*The police charged...* " (x)

The role and frame labeling model:

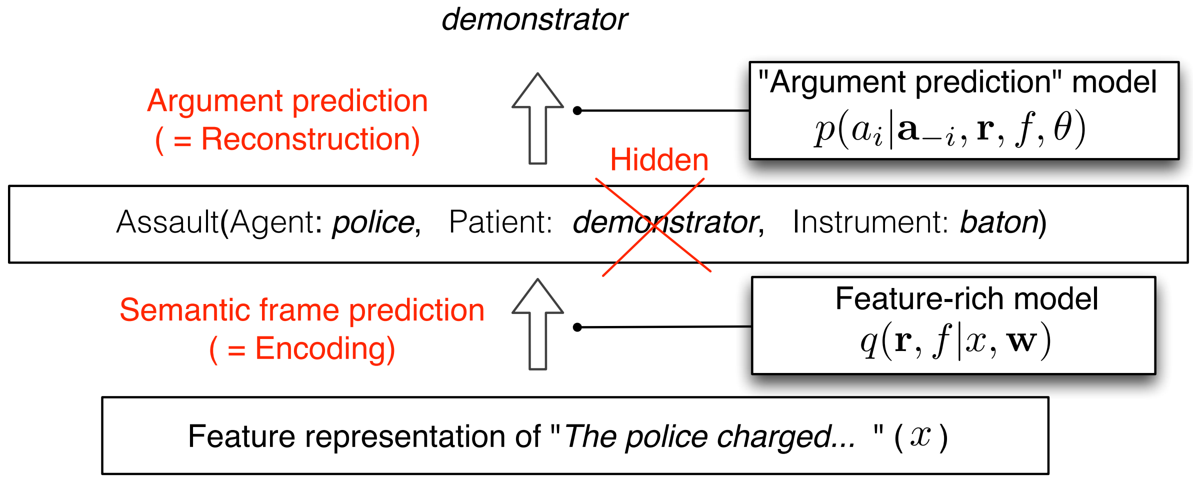
$$p(\mathbf{r}, f | x, \mathbf{w}) \propto \exp(\mathbf{w}^T \mathbf{g}(x, f, \mathbf{r}))$$

A feature-rich representation encoding syntax-semantics interface

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)

The majority of supervised SRL models; we used a simplified version of Johansson and Nugues ('08) "MATE tools"

Joint learning



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

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

A variational lower bound on (pseudo-) likelihood

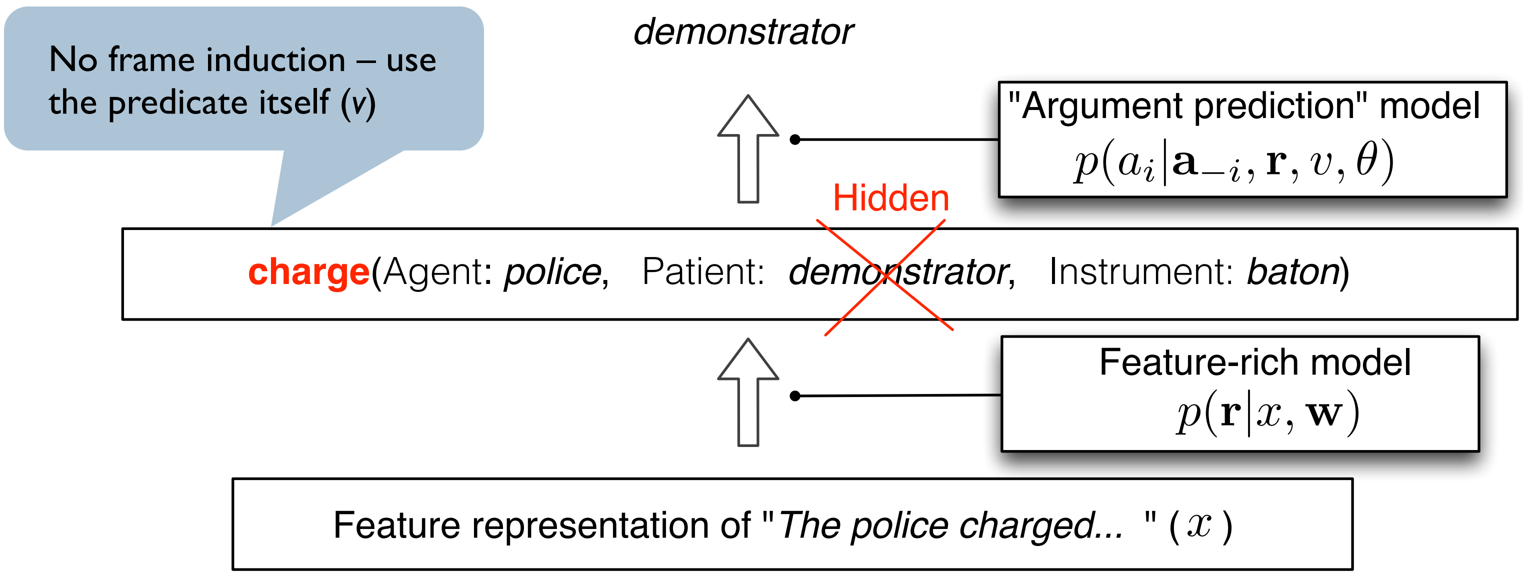
- Not very tractable in its exact form but standard 'tricks' can be used
  - negative sampling (as, e.g., in Mikolov et al '13) instead of 'softmax'

Training can be quite efficient as all models are linear (or bilinear)

Outline

- Motivation: why we need unsupervised feature-rich models and learning for inference
- Framework: reconstruction error minimization for semantics
- Special case: inferring missing arguments
- Empirical evaluation: preliminary experiments, insights, future work

Experiments: only role induction



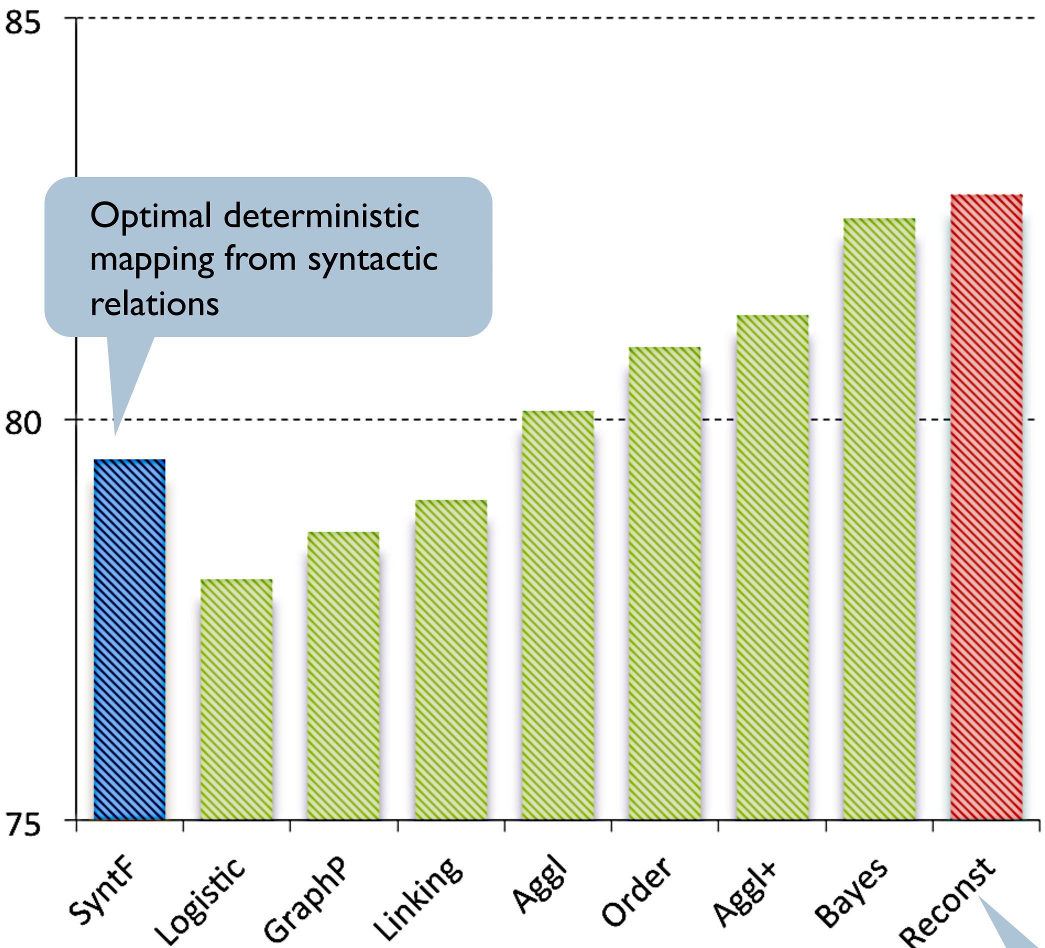
- Evaluate on a dataset annotated with roles (PropBank for En, SALSA for De)
- Compare against previous models evaluated in this set-up
  - use clustering evaluation measures (purity, collocation, F1)

May not be the optimal set-up for our expressive model

We replicate previous evaluation: datasets are fairly small (e.g., ~ 90,000 predicate-argument structures for English)

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)



Optimal deterministic mapping from syntactic relations

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

Induces fewer roles than most other approaches but under certain regimes, roles start to capture verb senses

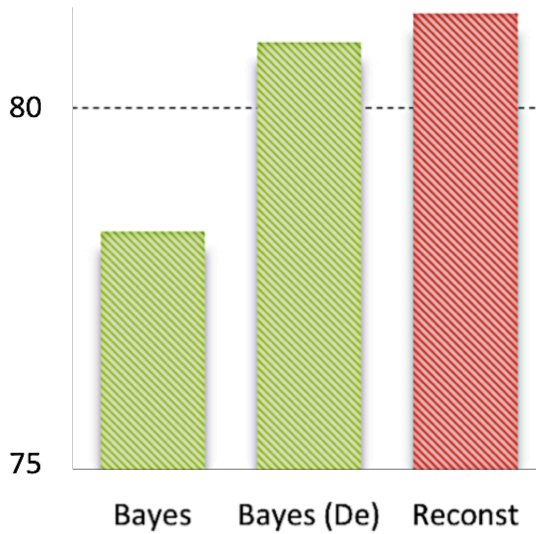
Previous approaches evaluated in the same setting

The feature-rich model

[NAACL '15]

German (F1)

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

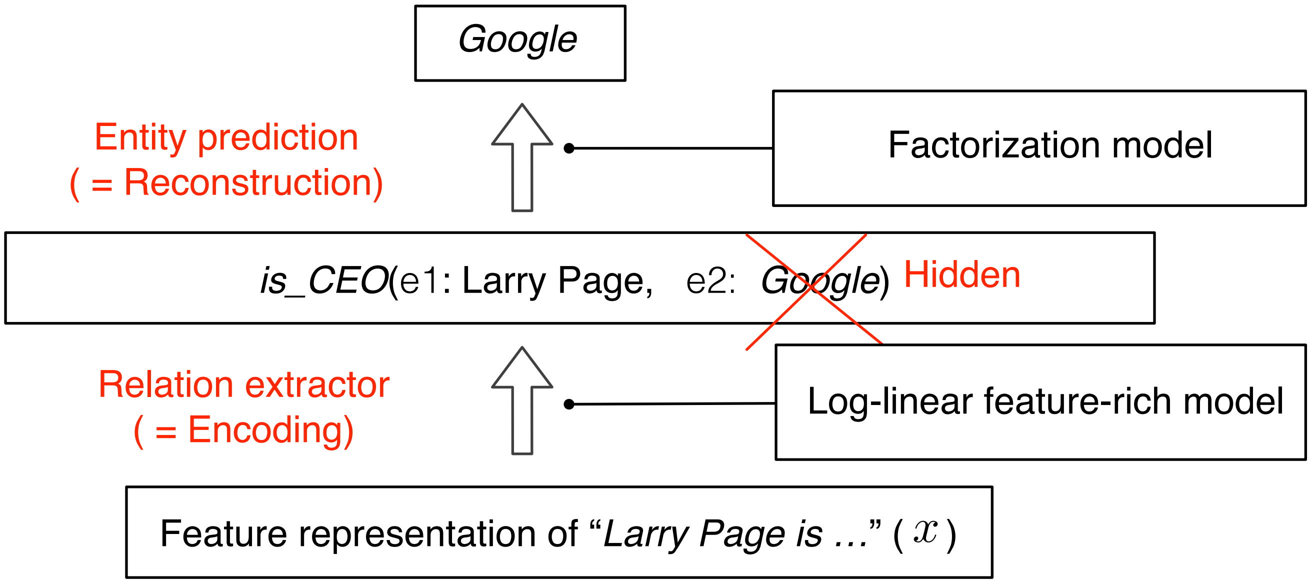


Performs on par with the best method without language-specific engineering

Bayes modified for German

The feature rich model (the same as for En)

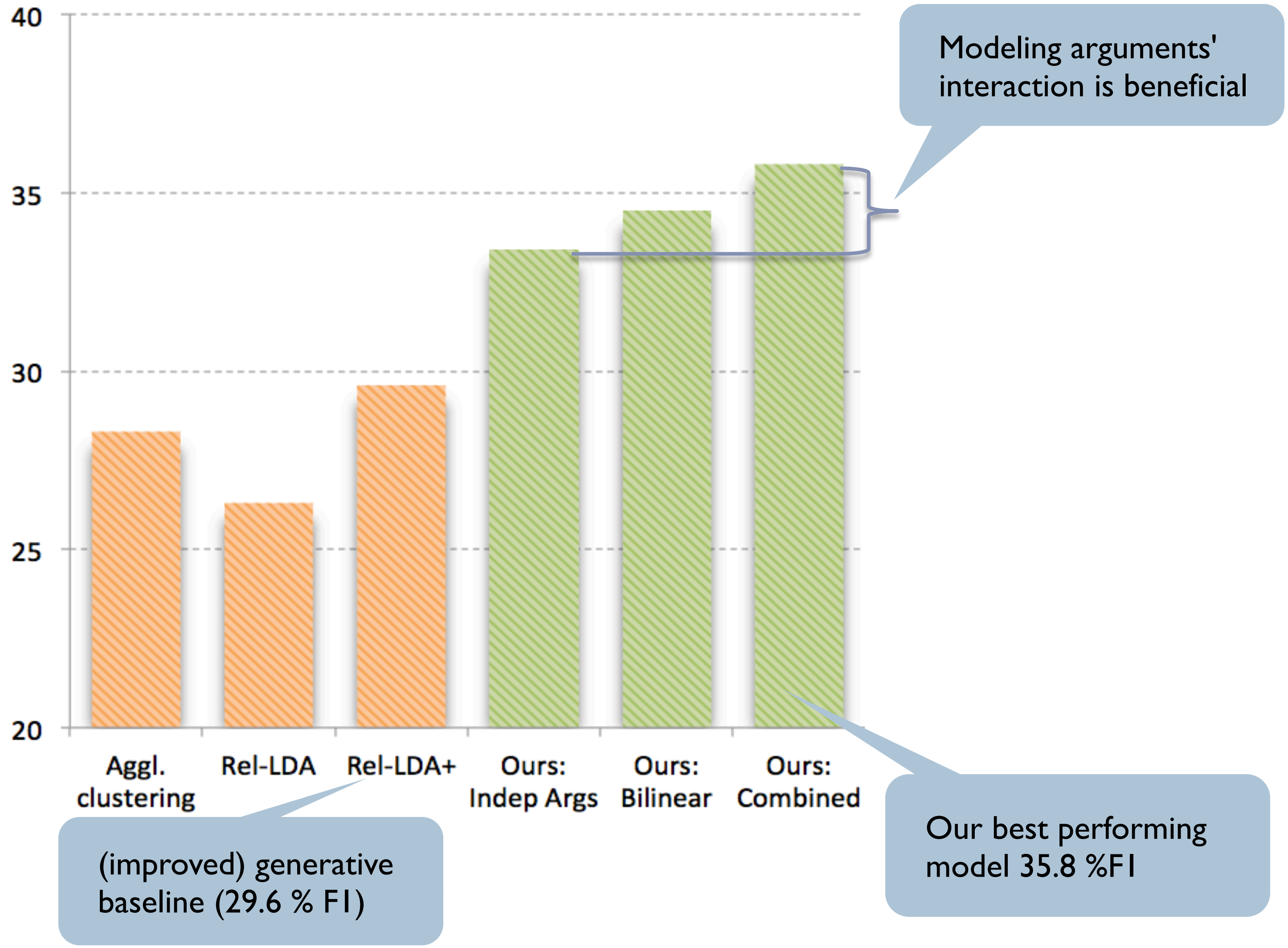
Only frame induction: relation discovery



- We simultaneously induce relations and learn their factorizations
- Data: New York Times corpus
- Evaluation against Freebase

In this way we induce relations which helps us to perform inference (i.e. fill gaps in the corresponding "knowledge base")

Evaluation (F1)



All our models outperform the generative and clustering baselines

REM framework

- A new framework for inducing shallow semantics
  - allowing for combining ideas from relation modeling and semantic parsing
  - language-independent
- Exploiting unlabeled data with expressive models – promising for the tail?
- The framework naturally supports:
  - Integration of prior linguistic knowledge
  - Semi-supervised learning
  - Learning for inference
  - Tighter integration with knowledge bases

Thank you!

- Special thanks to Dipanjan Das, Alex Klementiev, Alexis Palmer, Manfred Pinkal, …

- Funding:
  - Google Focused Award on Natural Language Understanding 2013-16
  - Google Faculty Research Award 2011
  - New: NWO VIDI 2015 + ERC StG 2015 + Amazon AWS grant