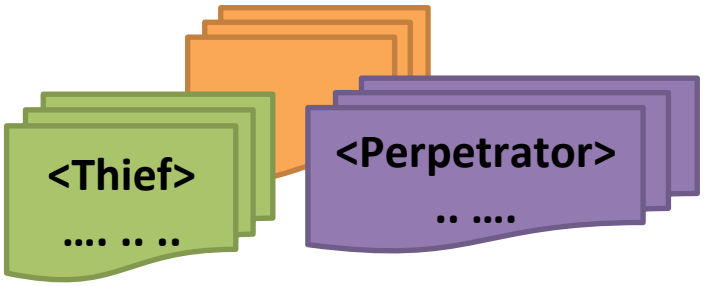
Frame-Semantic Role Labeling with Heterogeneous Annotations

Meghana Kshirsagar Sam Thomson Nathan Schneider
Jaime Carbonell Noah A. Smith Chris Dyer



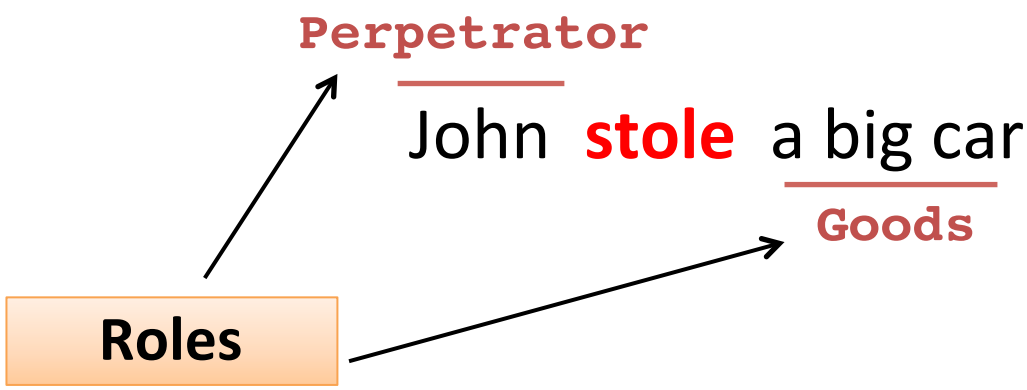
Semantic role labeling (SRL)

Input: a sentence
Output: representation of meaning

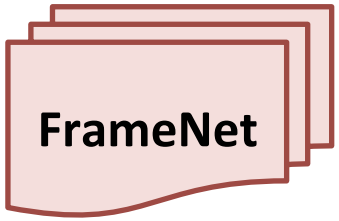
John stole a big car

Semantic role labeling (SRL)

Input: a sentence
Output: representation of meaning (using “roles”)



Frame SRL

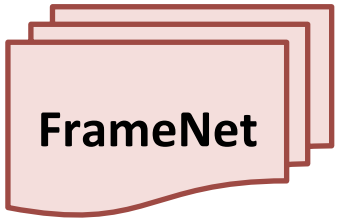


In the case of organic pollutions, the analysis itself **took** no more than five days

He sat up and **took** a piece of mud-coloured rag …

Predicate: “take”

Frame SRL



Activity

In the case of organic pollutions, the analysis itself **took** no more than five days

Time

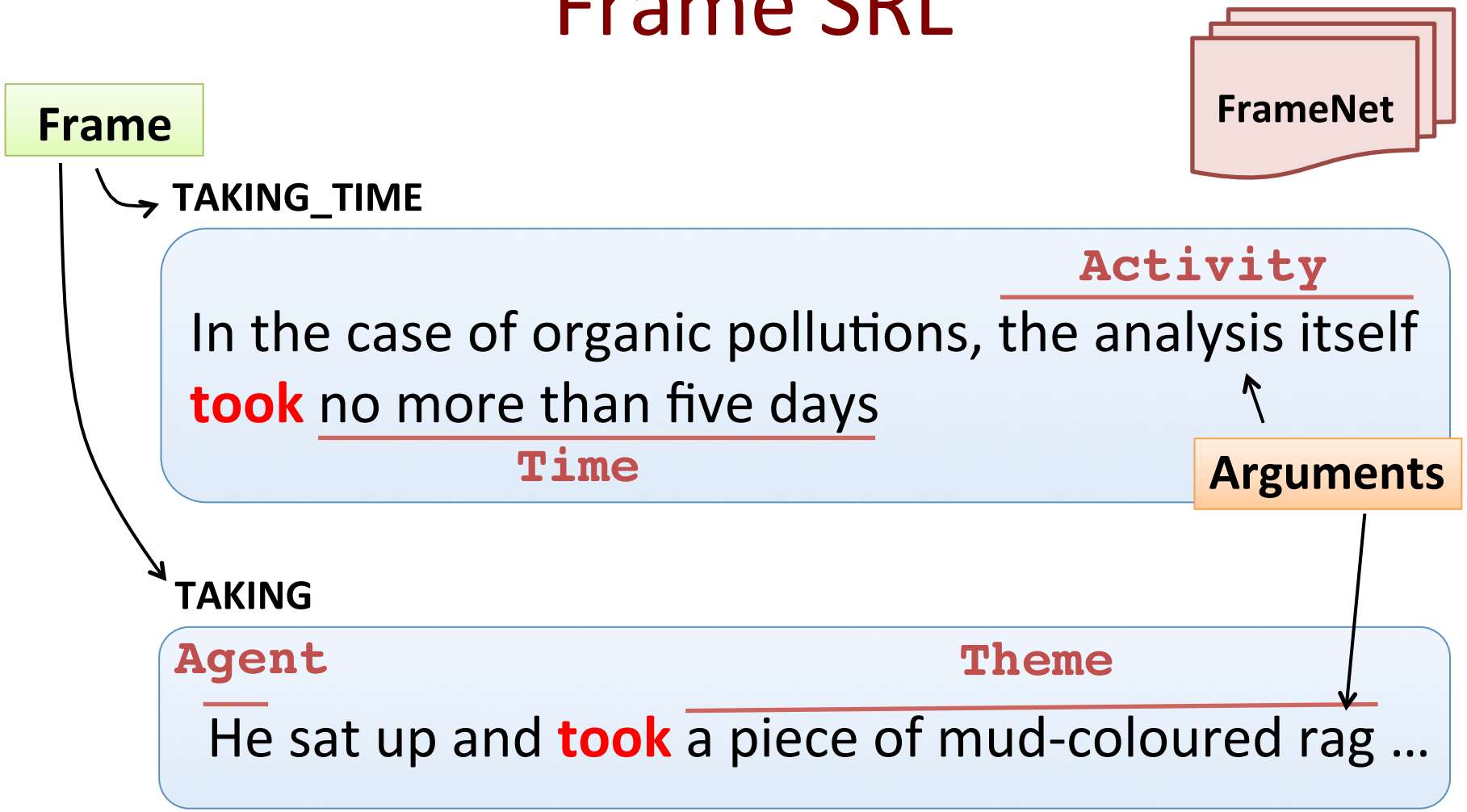
TAKING

Agent Theme

He sat up and **took** a piece of mud-coloured rag …

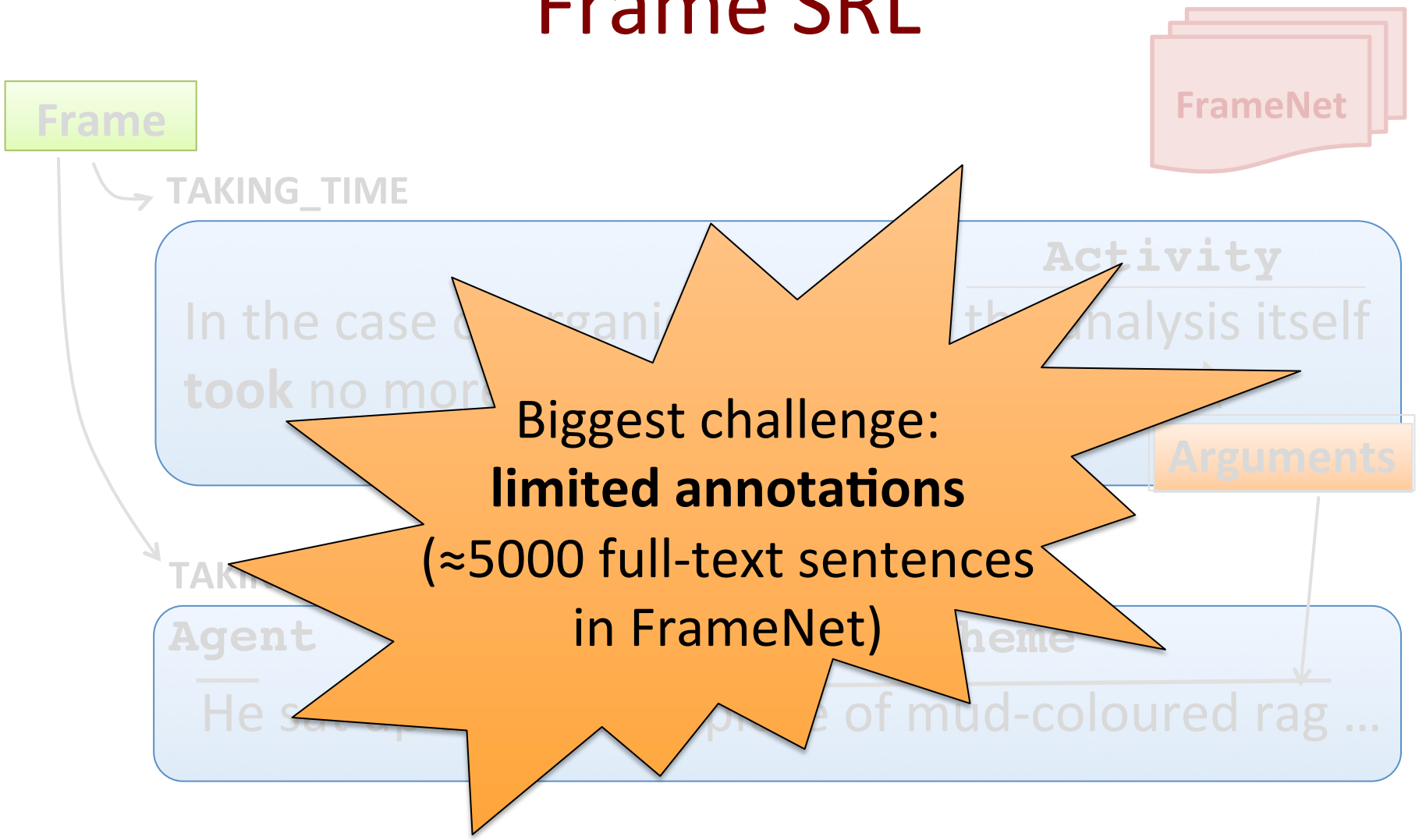
Predicate: “take”

Frame SRL



Predicate: “take”

Frame SRL



Biggest challenge:
limited annotations
(≈5000 full-text sentences
in FrameNet)

Many other resources for SRL

take.02

A0 A1
John couldn't **take** the heat, so he got out of the kitchen.

Many other resources for SRL

take.02

A0 A1
John couldn't **take** the heat, so he got out of the kitchen.

PropBank has many predicates, that are not in FrameNet
Ex: attest, involve, nominate …

Many other resources for SRL

take.02

A0 A1
John couldn't **take** the heat, so he got out of the kitchen.

PropBank has many predicates, that are not in FrameNet
Ex: attest, involve, nominate …

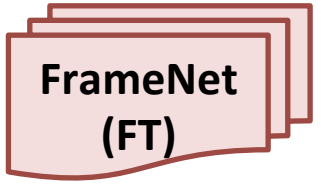
Goal:

- Improve semantic role labeling on FrameNet using other resources

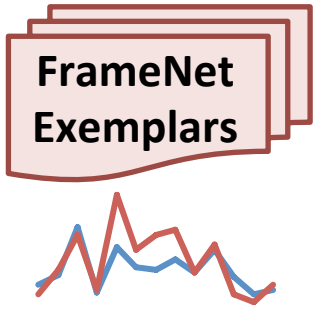
"Target" task

Many other resources for SRL

- FrameNet full-text (FT) ≈ 5,000
  - document annotations: newswire, emails, transcripts of phone conversations etc.



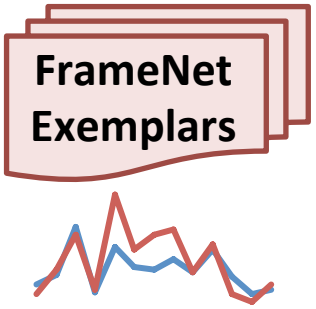
- FrameNet Exemplars ≈ 140,000
  - single sentences, primarily British National Corpus
  - distribution of roles is "artificial"



- PropBank ≈ 110,000
  - WSJ data, generally coarser sense distinctions
  - different annotation scheme

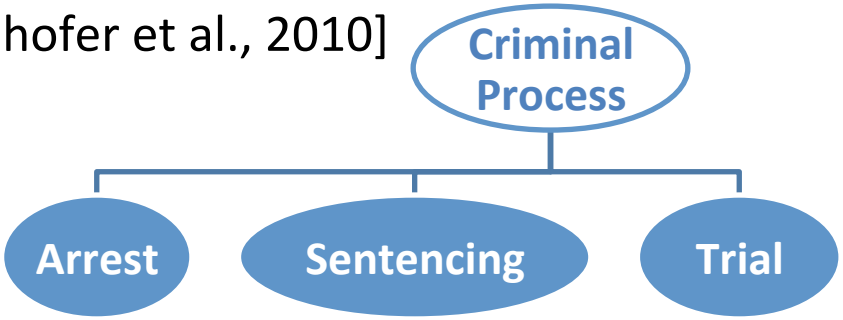
This work incorporates these resources..

- FrameNet Exemplars ≈ 140,000
  - single sentences, primarily British National Corpus
  - distribution of roles is "artificial"



- PropBank ≈ 110,000
  - WSJ data, generally coarser sense distinctions
  - different annotation scheme

- FrameNet Hierarchy [Ruppenhofer et al., 2010]
  - relationships such as *inheritance* between roles



A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

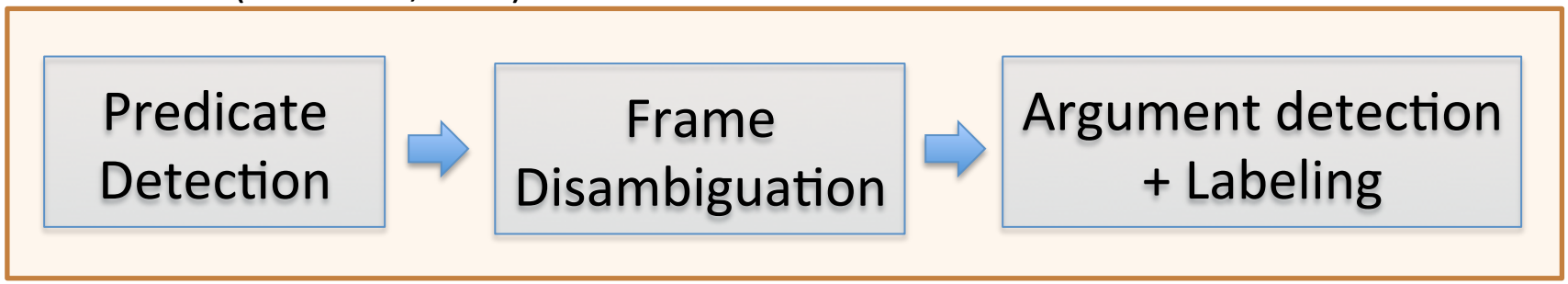
Output: a set of frame, <argument spans, role label>

A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

Output: a set of frame, <argument spans, role label>

SEMAFOR (Das et al.,2010)

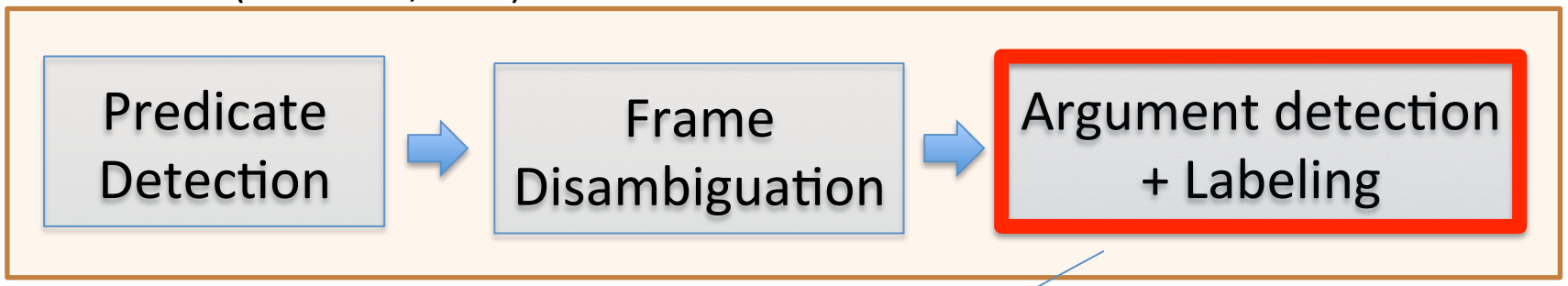


A model for Frame SRL

Given: a sentence, context features, POS tags, dependency parse

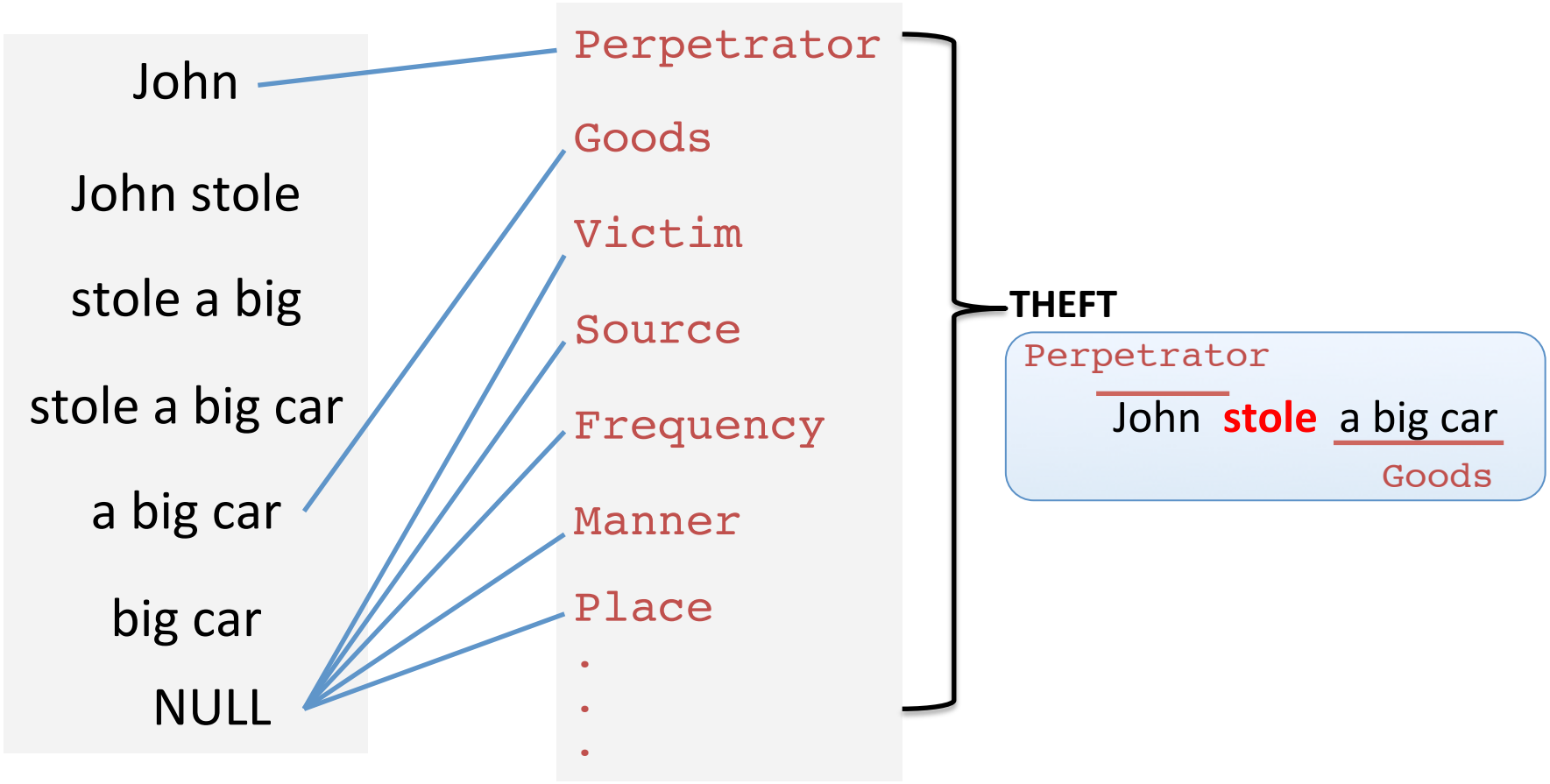
Output: a set of frame, <argument spans, role label>

SEMAFOR (Das et al.,2010)



Objective function

- Goal: Match text-spans with role labels



Objective function

- Goal: Match text-spans with role labels
- Score of a span '*a*'

$$score_{\mathbf{w}}(a \mid \mathbf{x}, p, f, r) = \mathbf{w}^{\top} \phi(a, \mathbf{x}, p, f, r)$$

sentence
predicate
frame
role

Objective function

- Goal: Match text-spans with role labels
- Score of a span '*a*'

$$score_{\mathbf{w}}(a \mid \mathbf{x}, p, f, r) = \mathbf{w}^{\top} \phi(a, \mathbf{x}, p, f, r)$$

sentence, predicate, frame, role

- Squared hinge loss for i^{th} example

$$SqHinge_{\mathbf{w}}(i) = \left(\max_{a'} \{\mathbf{w}^{\top} \phi(a', \mathbf{x}, p, f, r) + \text{cost}(a', a)\} - \mathbf{w}^{\top} \phi(a, \mathbf{x}, p, f, r)\right)^2$$

Adadelta for optimization

Objective function

- Match text-spans with role labels
- Score of a span '*a*'

$$score(a, \mathbf{x}, p, f, r)$$

role

sentence

predicate

frame

- Squared hinge loss with l_2 regularization over all training example

$$SqHinge_{\mathbf{w}} = \left(\max_{a'} \{\mathbf{w}^\top \phi(a', \mathbf{x}, p, f, r) + \text{cost}(a', a)\} - \mathbf{w}^\top \phi(a, \mathbf{x}, p, f, r)\right)^2$$

Significant benefits in run-time over prior work (1 week -> 9 hours)

Adadelta for optimization

Approaches to incorporate other resources

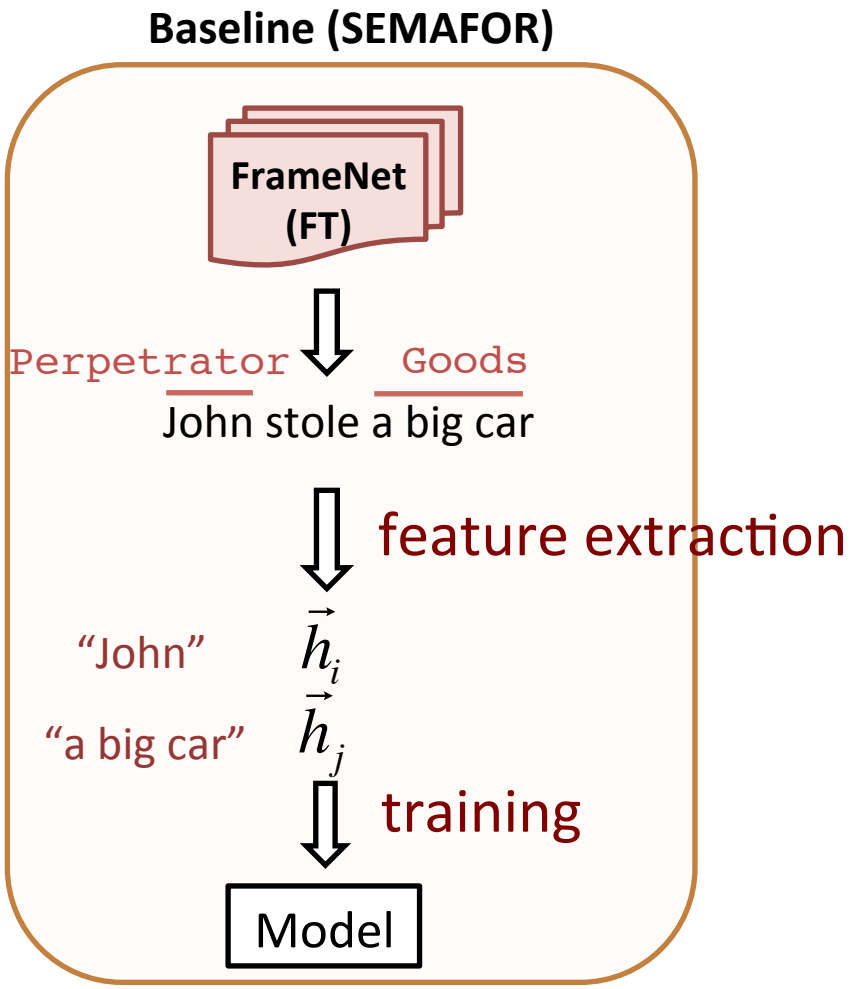
- Use as additional training data

- Via additional features (feature augmentation)
  - Frustratingly easy domain adaptation [Daumé, 09]
  - Defining "guide features" [Johansson, '13]

- Parameter sharing

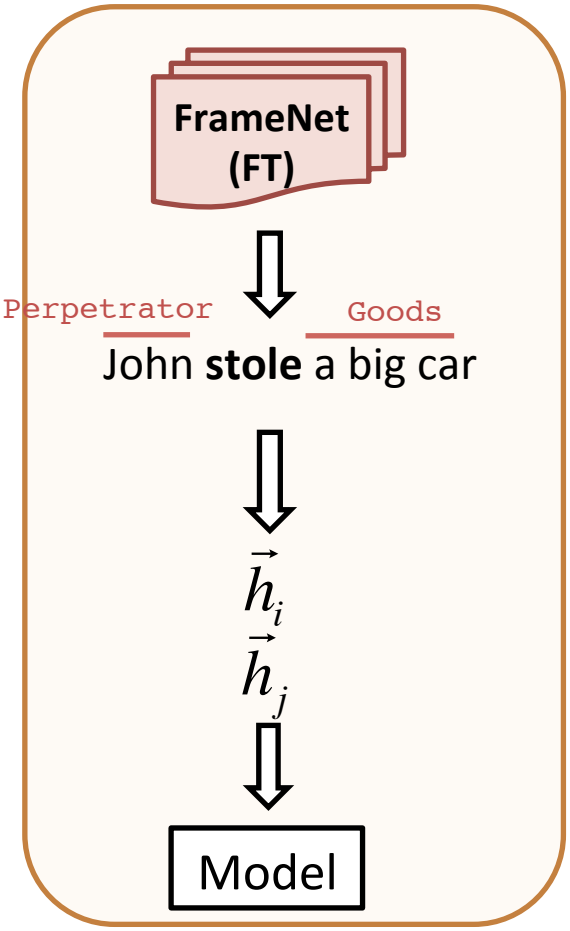
Feature augmentation using "guides"

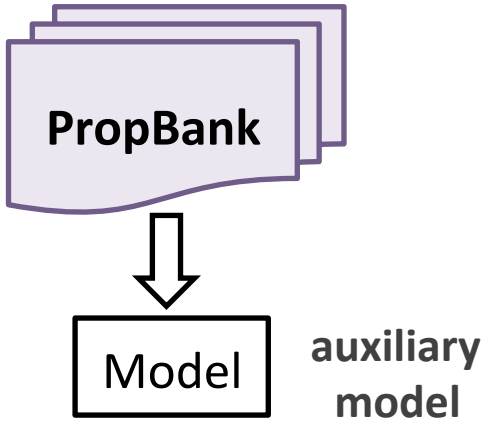
[Johansson, '13]



Feature augmentation using "guides"

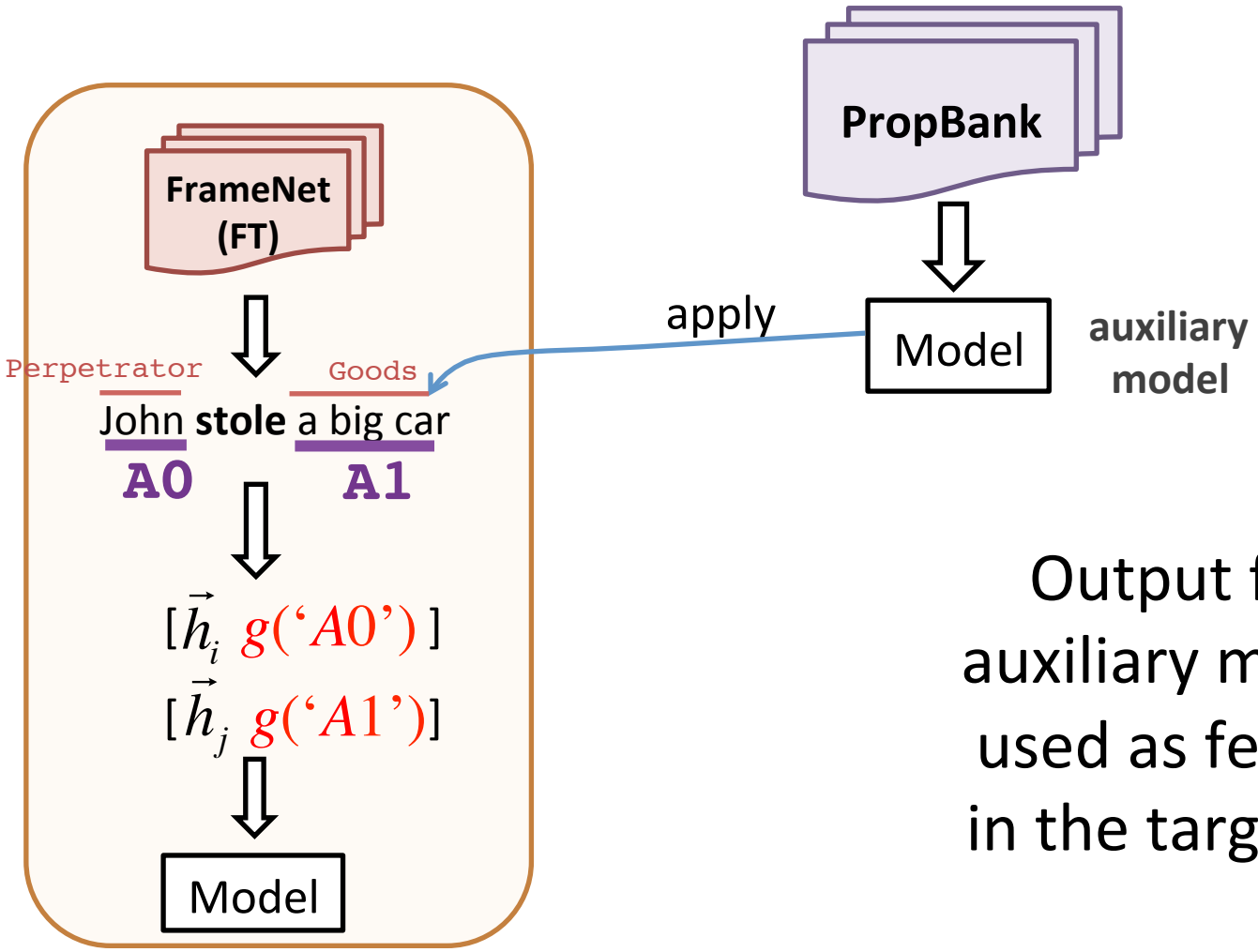
[Johansson, '13]





Feature augmentation using "guides"

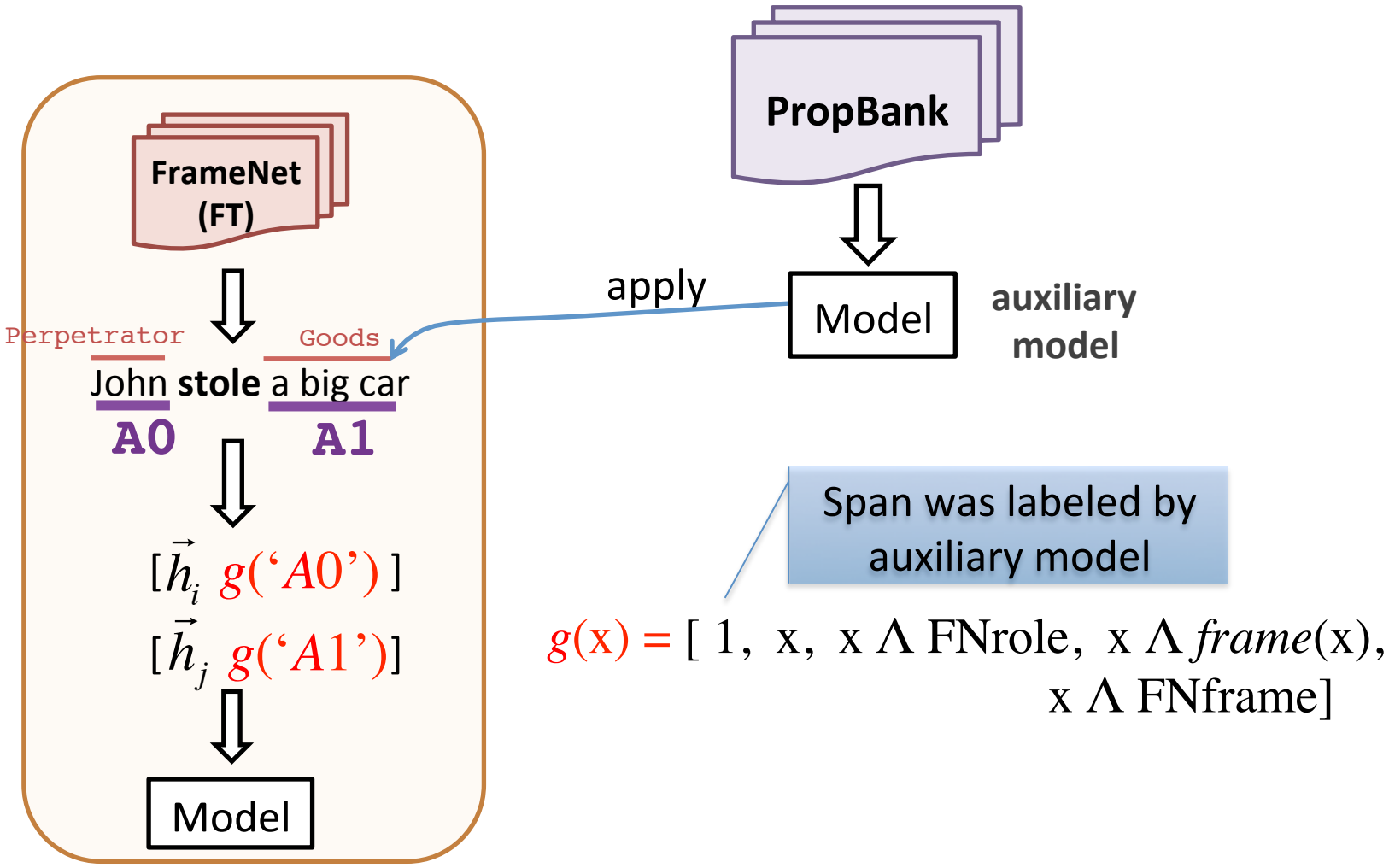
[Johansson, '13]



Output from auxiliary model is used as features in the target task

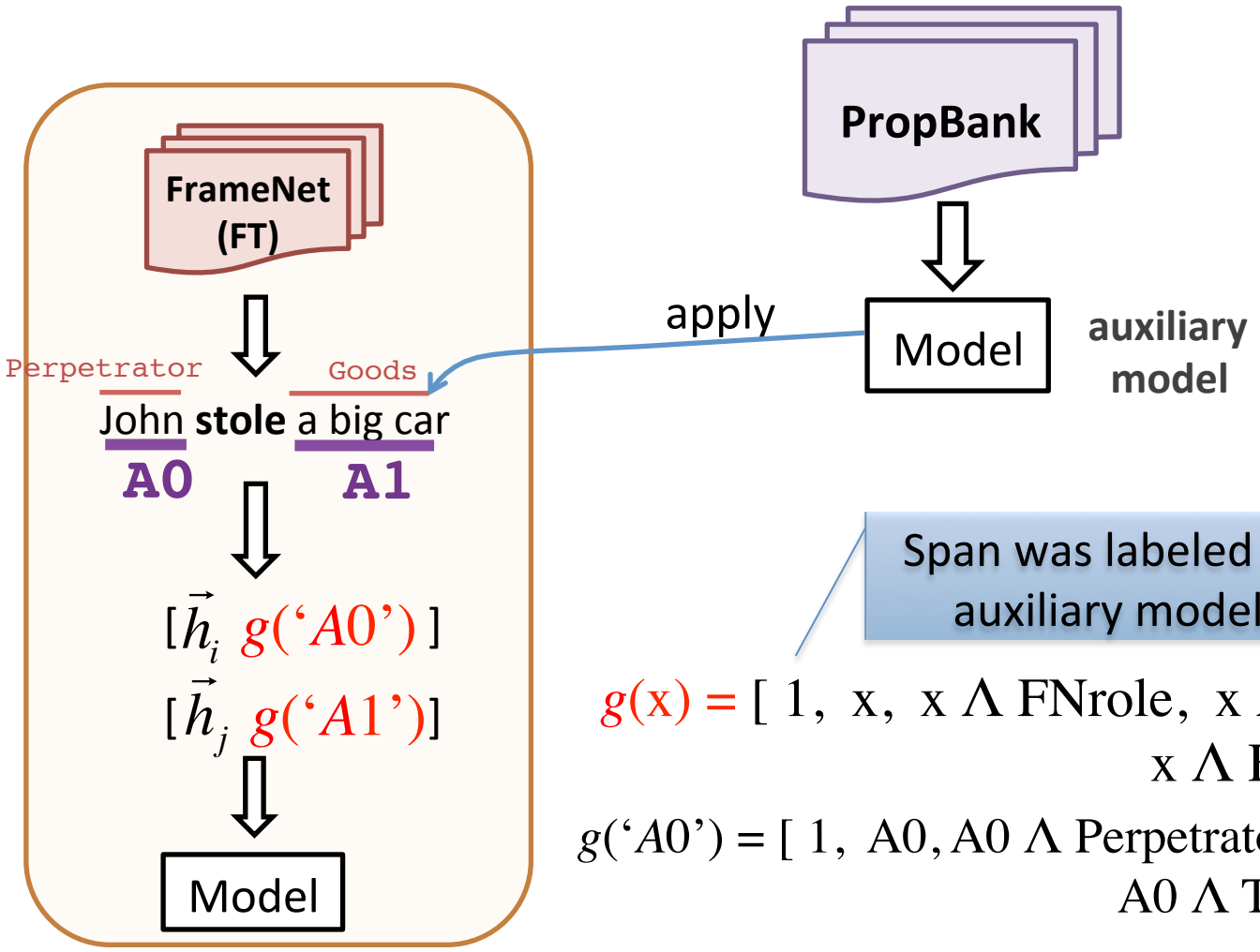
Feature augmentation using "guides"

[Johansson, '13]

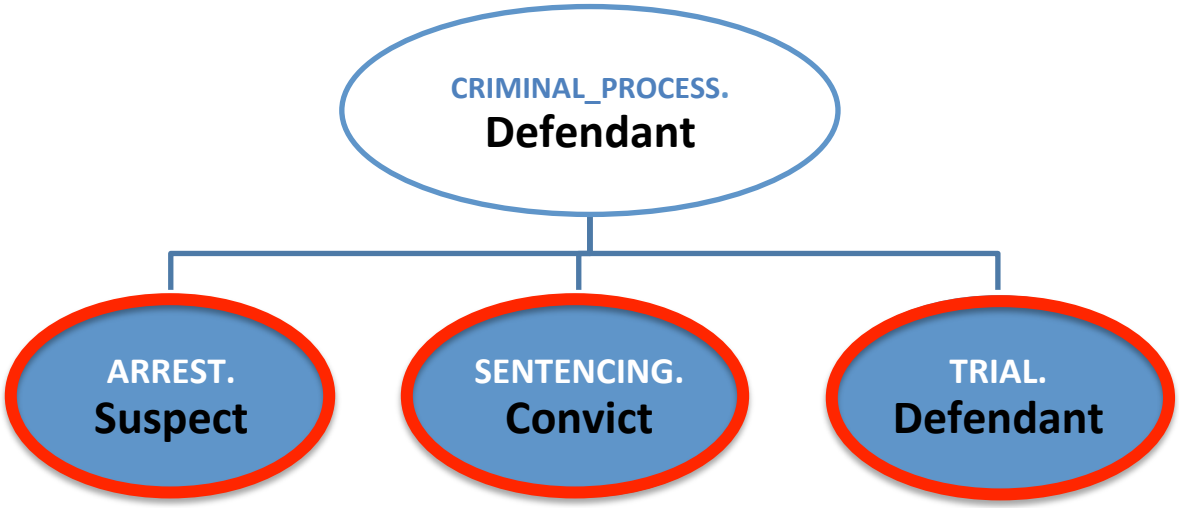


Feature augmentation using "guides"

[Johansson, '13]

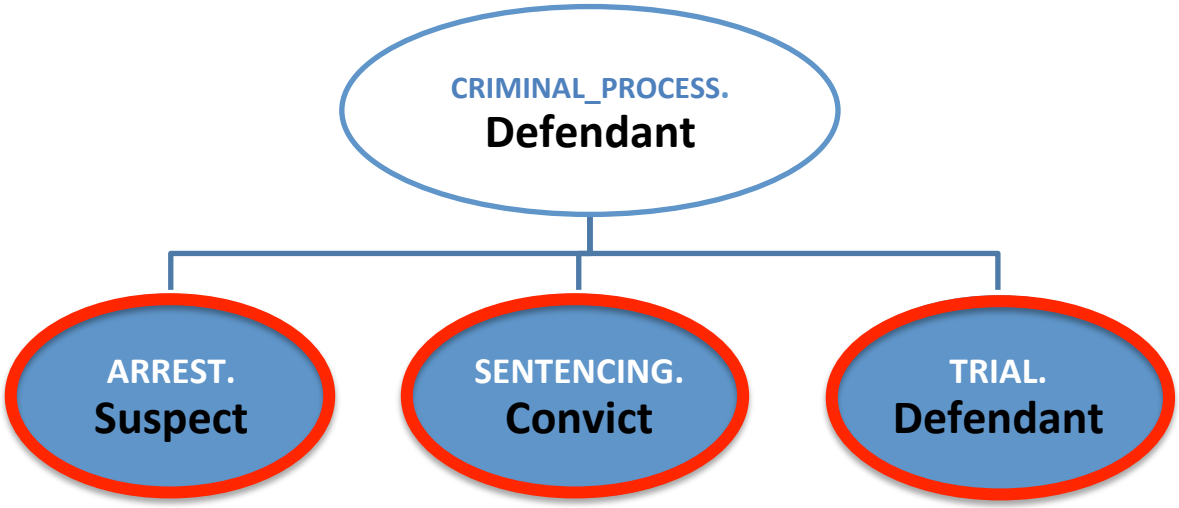


Parameter sharing using the FrameNet hierarchy



Please refer to paper for details!

Parameter sharing using the FrameNet hierarchy



Parameters are shared between all siblings

Sharing involving higher levels did not work as well

Please refer to paper for details!

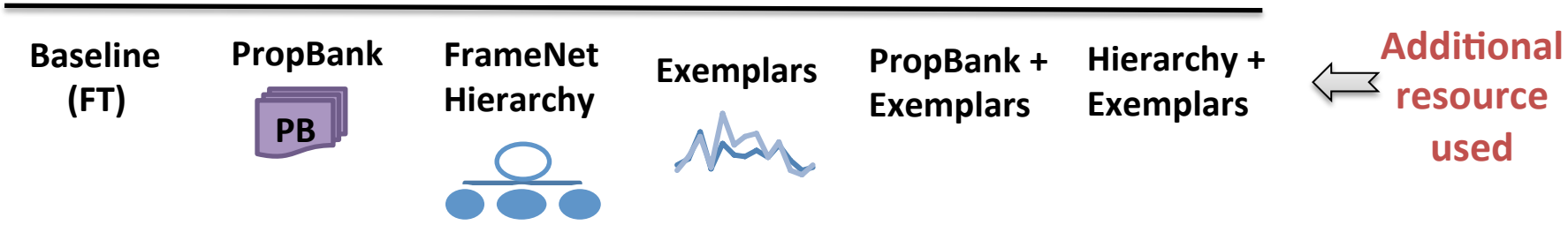
Evaluation

- FrameNet 1.5
  - test set from Das et al. 2010
  - 2420 sentences, 7210 overt arguments
- For frame:
  - assume gold frame is known
  - use frames from SOTA frame-identification

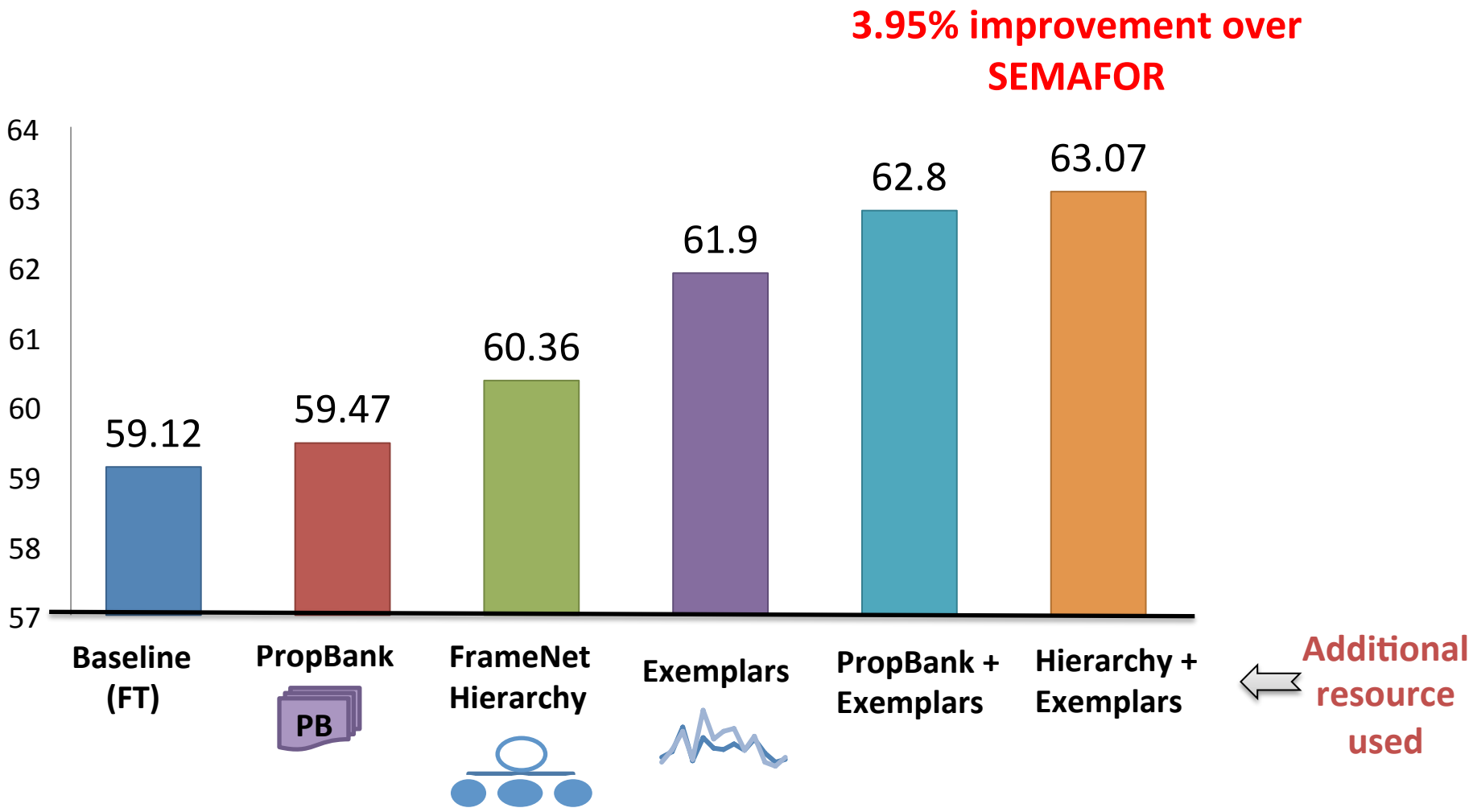
    [Hermann et al 2014]

Full system performance

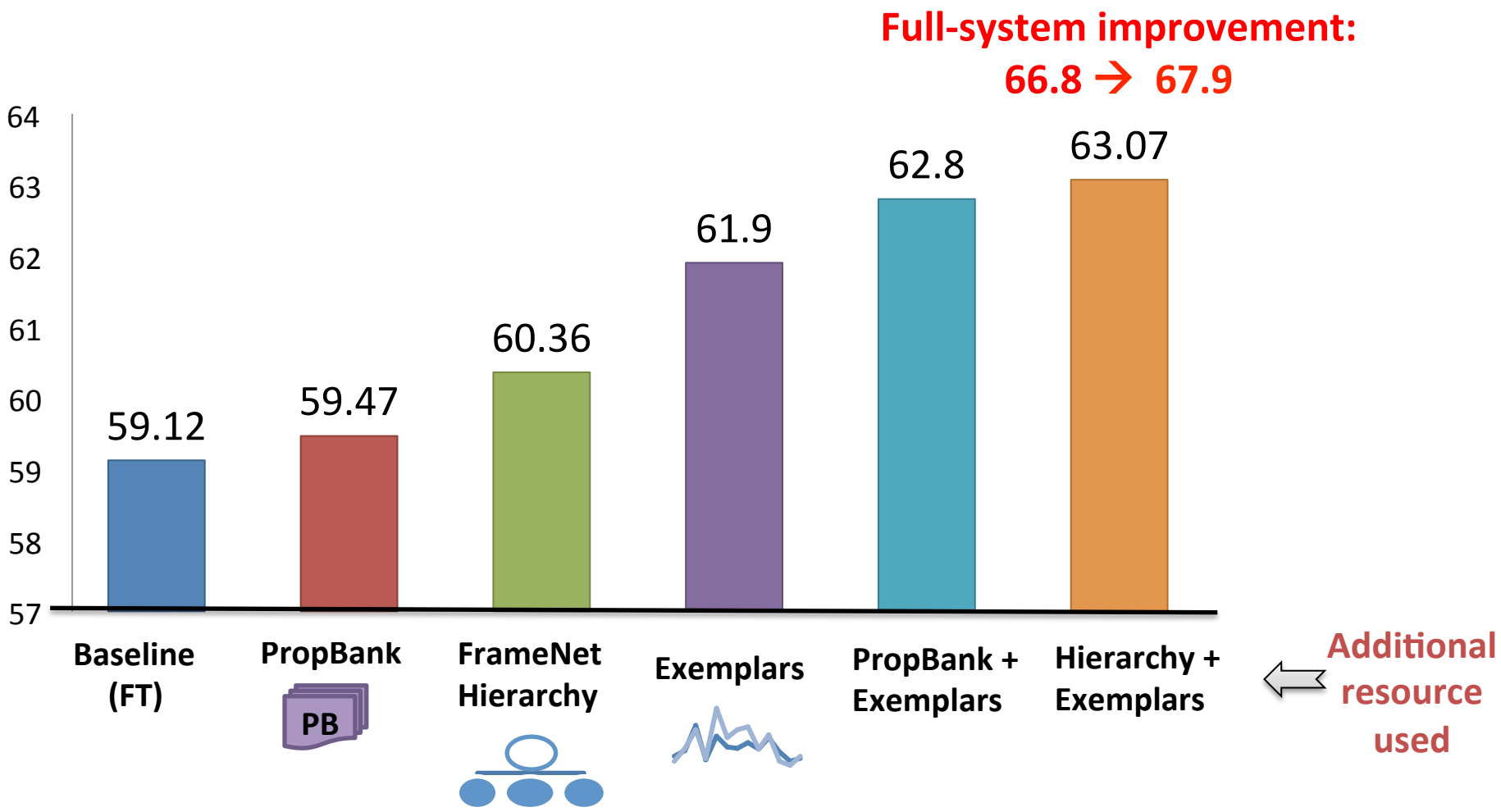
F_1 on test set (given gold frame)



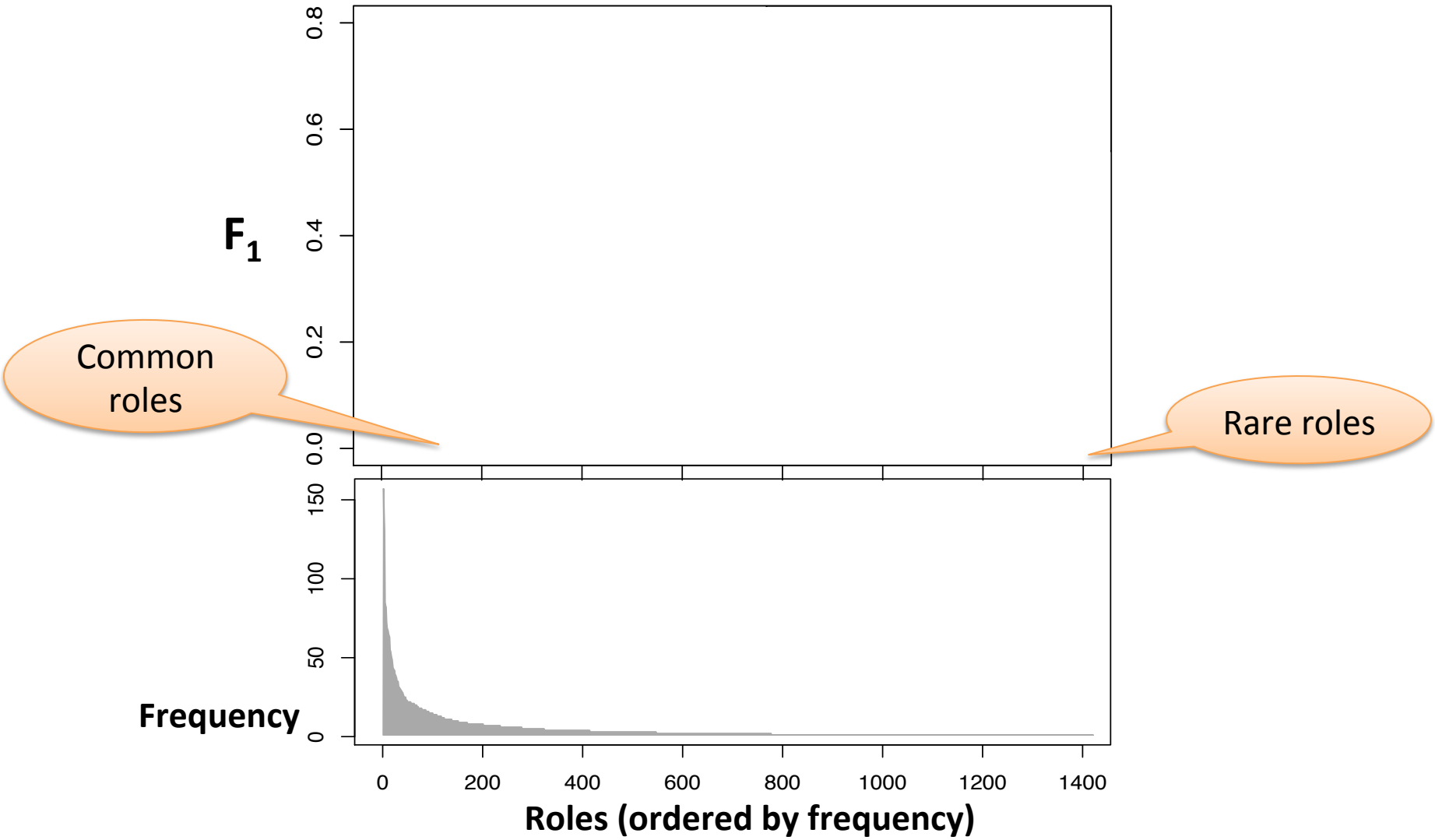
F_1 on test set (given gold frame)



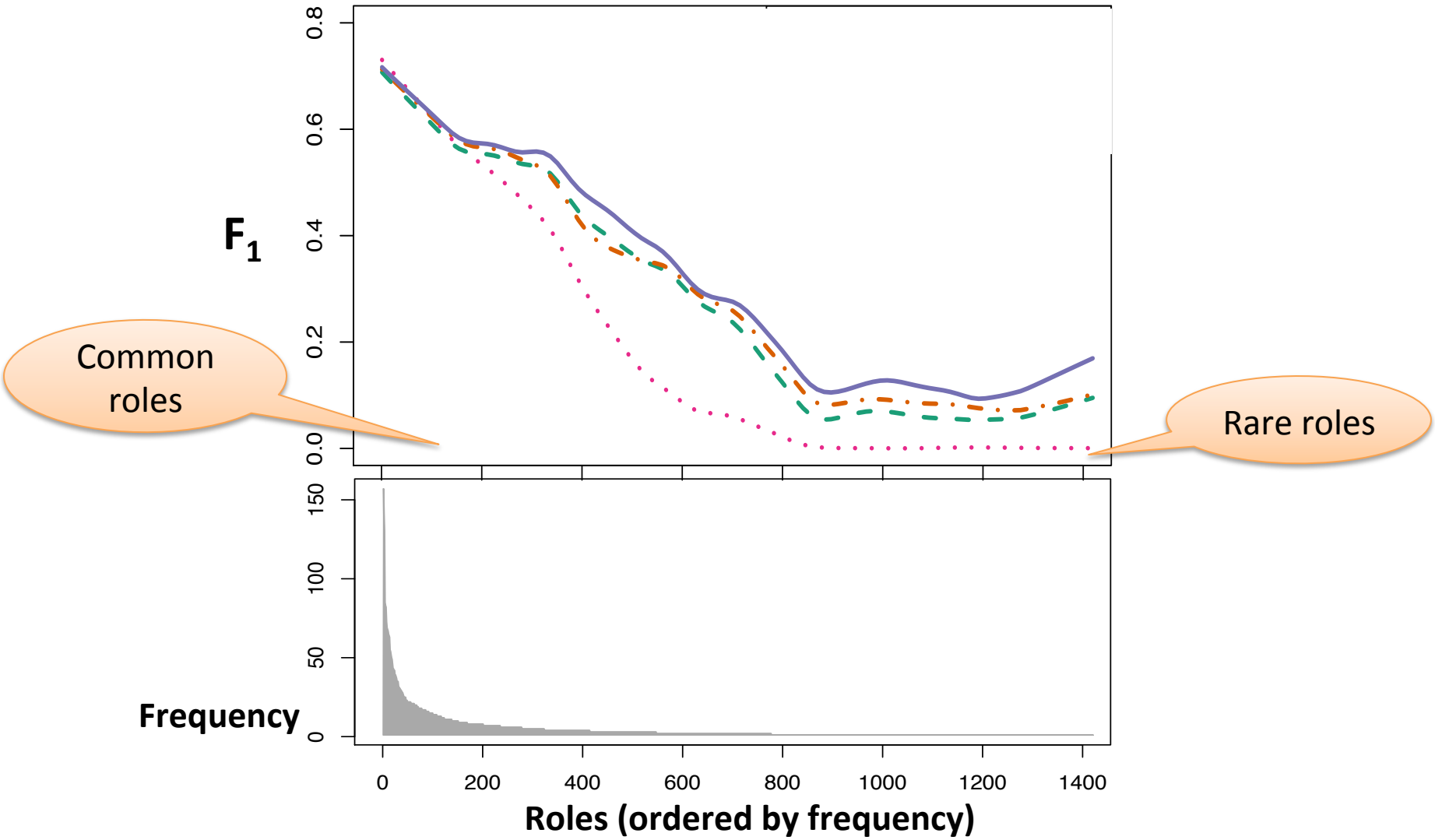
F_1 on test set (given gold frame)



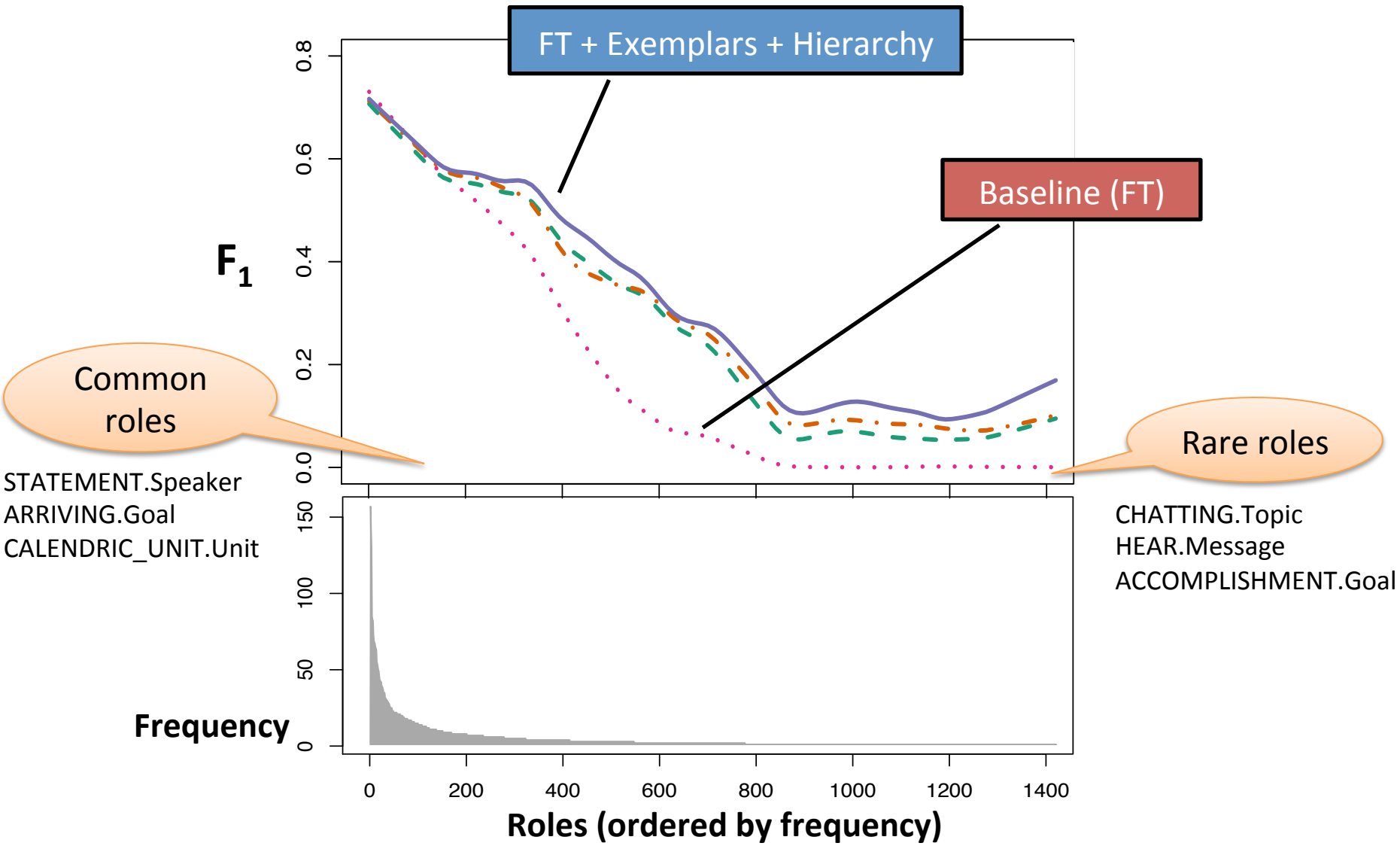
Role-wise F_1


F1
0.0
0.2
0.4
0.6
0.8
Common roles
Rare roles
Frequency
0
50
100
150
0
200
400
600
800
1000
1200
1400
Roles (ordered by frequency)

Role-wise F_1


F1
0.0
0.2
0.4
0.6
0.8
Common roles
Rare roles
Frequency
0
50
100
150
0
200
400
600
800
1000
1200
1400
Roles (ordered by frequency)

Role-wise F_1


FT + Exemplars + Hierarchy
Baseline (FT)
F1
0.0
0.2
0.4
0.6
0.8
Common roles
STATEMENT.Speaker
ARRIVING.Goal
CALENDRIC_UNIT.Unit
Rare roles
CHATTING.Topic
HEAR.Message
ACCOMPLISHMENT.Goal
Frequency
0
50
100
150
0
200
400
600
800
1000
1200
1400
Roles (ordered by frequency)

Test sentences with gains

BOARD_VEHICLE

Vehicle

Can he just **get** on a plane and fly to Paris ?

Traveller

BODY_MOVEMENT

Agent **Body_part** **Purpose**

Passengers **crane** their necks for dizzying glimpses of the harbor

Arguments in blue colour are missed by the baseline, but found by our model

Conclusion

- Contributions:
  - we exploit multiple diverse resources for better coverage
  - side-effect: faster training using hinge loss

- Future work:
  - incorporate additional resources
  - combine with other models as the baseline [Tackstorm et al., 2015]

Prior work

- Using FrameNet hierarchy
  - Matsubayashi et al., '09
  - Johansson '12
- Other directions
  - Pavlick, '15
  - Fezabadi & Pado, '15

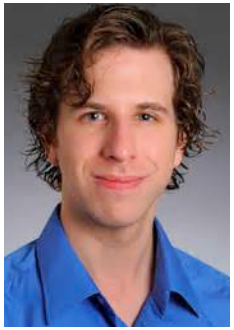
Co-authors

Jaime Carbonell

Nathan Schneider

Sam Thomson

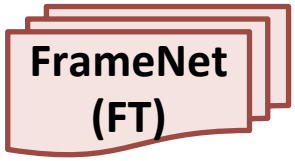
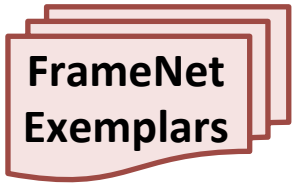
Chris Dyer



Noah Smith

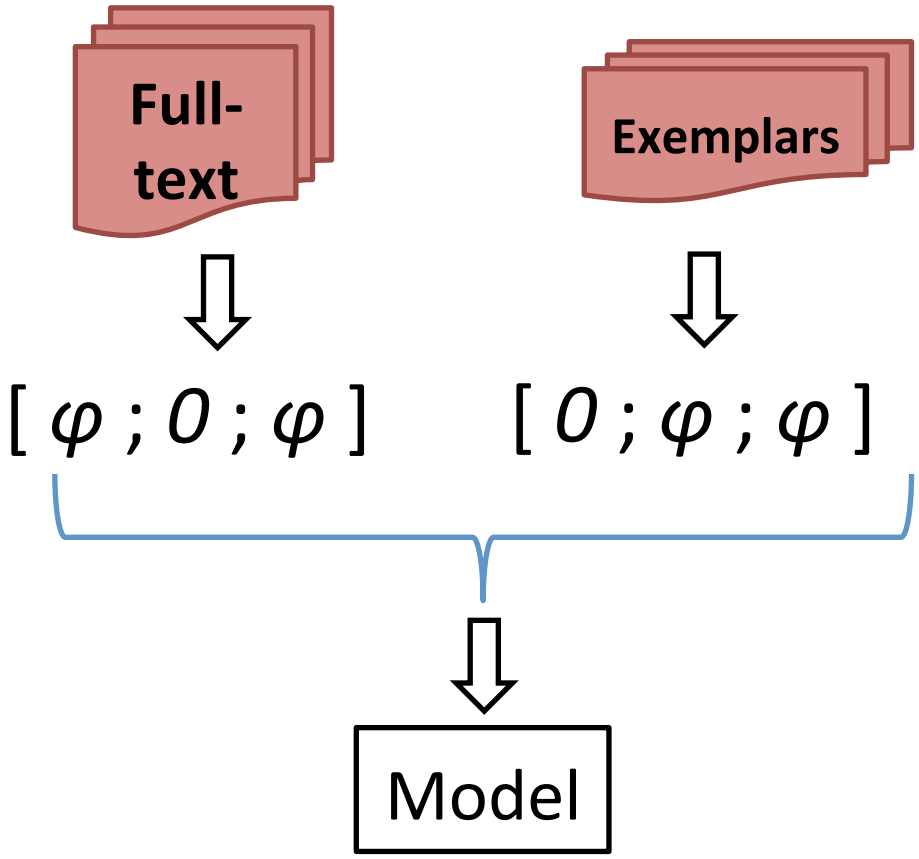
Thank you!

Sizes of the resources

Resource	Number of sentences	Number of overt arguments
	2,780	25,918
	137,515	278,985
	112,831	541,759

Frustratingly easy domain adaptation

[Daume et al., 2009]



Maintain task-specific and general copies of features