Probabilistic Frame-Semantic Parsing

Dipanjan Das   Nathan Schneider   Desai Chen   Noah A. Smith

School of Computer Science
Carnegie Mellon University

NAACL-HLT
June 4, 2010

Carnegie Mellon

In a Nutshell

- Most models for semantics are very local

  (cascades of classifiers)

- This work: towards more global modeling for rich semantic processing

  (feature sharing among all semantic classes)

  (*just* two probabilistic models)

- Our model outperforms the state of the art

- Our framework lends itself to extensions and improvements

Outline

- Introduction
- Background and Datasets
- Models and Results
- Conclusion

Outline

- Introduction
- Background and Datasets
- Models and Results
- Conclusion

Overview

- Annotate English sentences with semantic representations
- Combination of:
  - semantic frame (word sense) disambiguation
  - semantic role labeling
- Frame and role repository: FrameNet (Fillmore et al., 2003)

Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a *frame* of semantic knowledge

Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a *frame* of semantic knowledge

the 1995 book by John Grisham

Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a *frame* of semantic knowledge

the 1995 *book* by John Grisham

TEXT

Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a *frame* of semantic knowledge

the 1995 *book* by John Grisham

TEXT

  - a frame encodes a gestalt event or scenario

Frame Semantics

- Theory developed by Fillmore (1982)
  - a word evokes a *frame* of semantic knowledge

the 1995 *book* by John Grisham

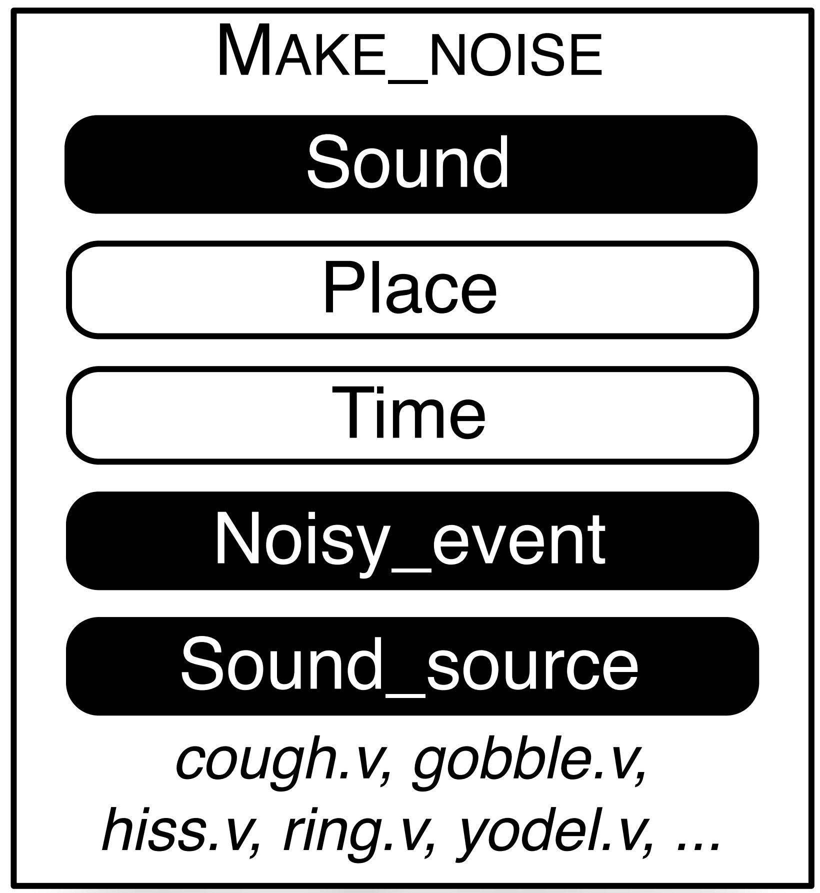
TEXT

Time_of_creation

Author

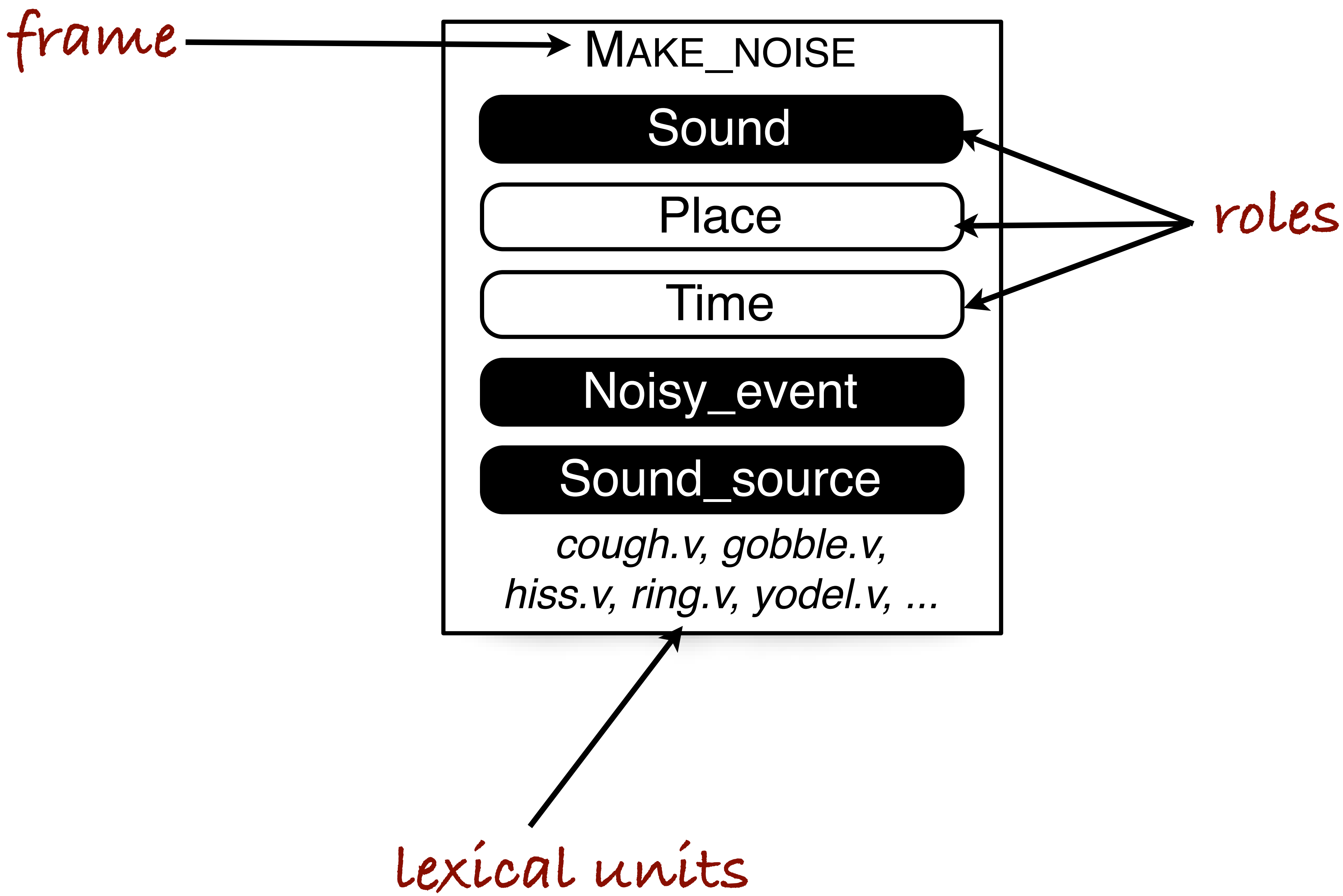
  - a frame encodes a gestalt event or scenario
  - it has conceptual dependents filling *roles* elaborating the frame instance

FrameNet



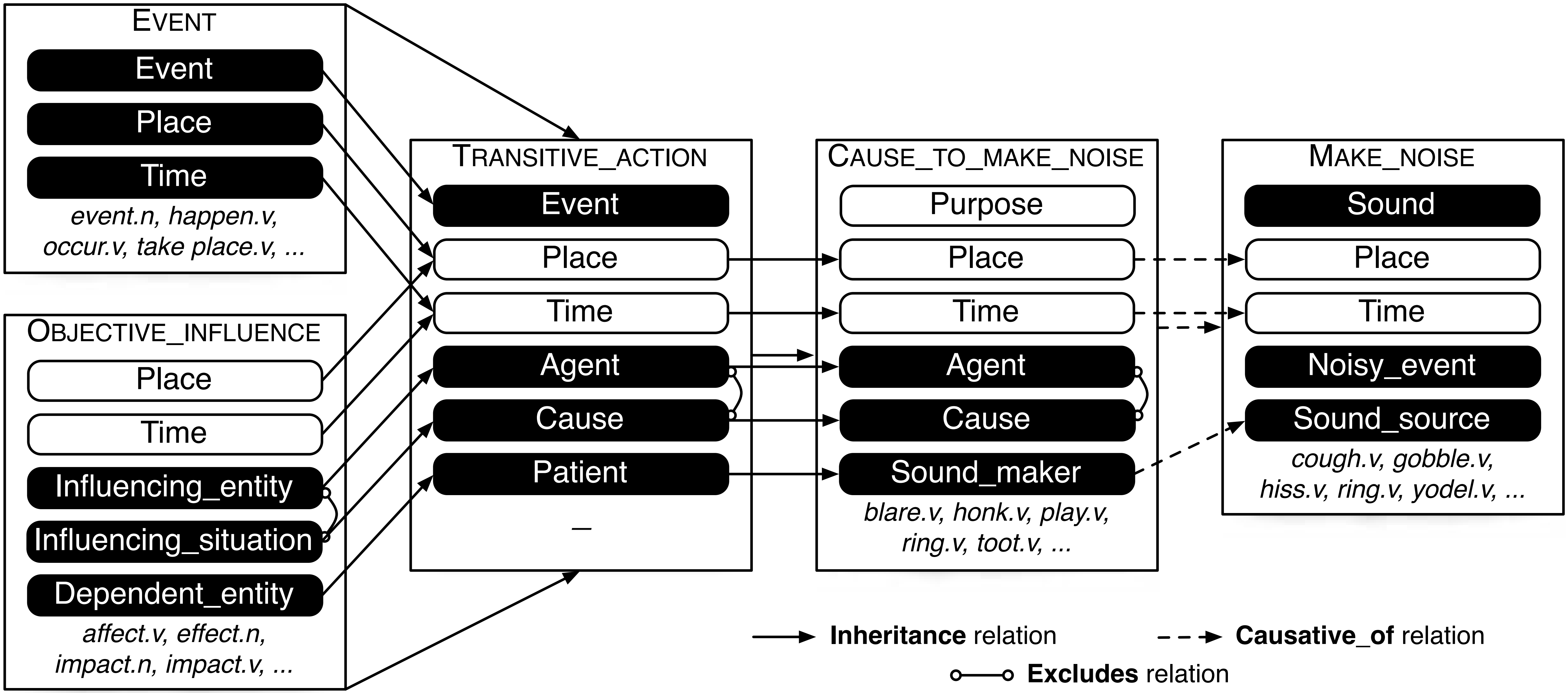
(Fillmore et al., 2003)

FrameNet



(Fillmore et al., 2003)

FrameNet



relationships between frames and between roles

(Fillmore et al., 2003)

FrameNet

- Statistics:
  - 795 semantic frames
  - 7124 roles
  - 8379 lexical units (predicates)
- 139,000 *exemplar* sentences containing one frame annotation per sentence

A Frame-Semantic Parse

Marco Polo wrote an account of Asian society during the 13th century .

TEXT

Author

Topic

A Frame-Semantic Parse

Marco Polo wrote an account of Asian society during the 13th century .

TEXT

Author

Topic

here, the ambiguous word evokes the TEXT frame

A Frame-Semantic Parse

Marco Polo wrote an account of Asian society during the 13th century .

TEXT

Author

Topic

participants in the event or scenario

A Frame-Semantic Parse

Marco Polo wrote an account of Asian society during the 13th century .

TEXT

Author

Topic

participants in the event or scenario

frame-specific

Why Frame-Semantic Parsing?

- Combines lexical and predicate-argument semantics
- Exploits meaningful primitives developed by experts
  - the FrameNet lexicon
- Richer representation than PropBank style SRL
- No inconsistent symbolic tags (ARG2-ARG5)
  (Yi et al. 2007, Matsubayashi et al. 2009)
- Patterns generalizing *across* frames and roles can be learned
  (Matsubayashi et al. 2009)

Outline

- Introduction
- Background and Datasets
- Models and Results
- Conclusion

Early Work

- Gildea and Jurafsky (2002)
  - Much smaller version of FrameNet
  - exemplar sentences

SemEval 2007

- Baker et al. (2007) organized the SemEval task on frame structure extraction
  - first set of *full* text annotations available
  - released a corpus of ~2000 sentences with full frame-semantic parses
- Johansson and Nugues (2007) submitted the best performing system
  - *our baseline for comparison* (J&N'07)

SemEval 2007

- SemEval 2007 dataset:
  - training set: 1941 sentences
  - test set: 120 sentences
- Three domains
  - American National Corpus (travel)
  - Nuclear Threat Initiative (bureaucratic)
  - PropBank (news)

SemEval 2007

- Evaluation is done using the official SemEval script
  - Measures precision, recall and F_1 score for frames and arguments
  - Features a partial matching criterion for frame identification
    - assigns score between 0 and 1 to closely related frames in the FrameNet hierarchy

Outline

- Introduction
- Background and Datasets
- Models and Results
- Conclusion

Challenges

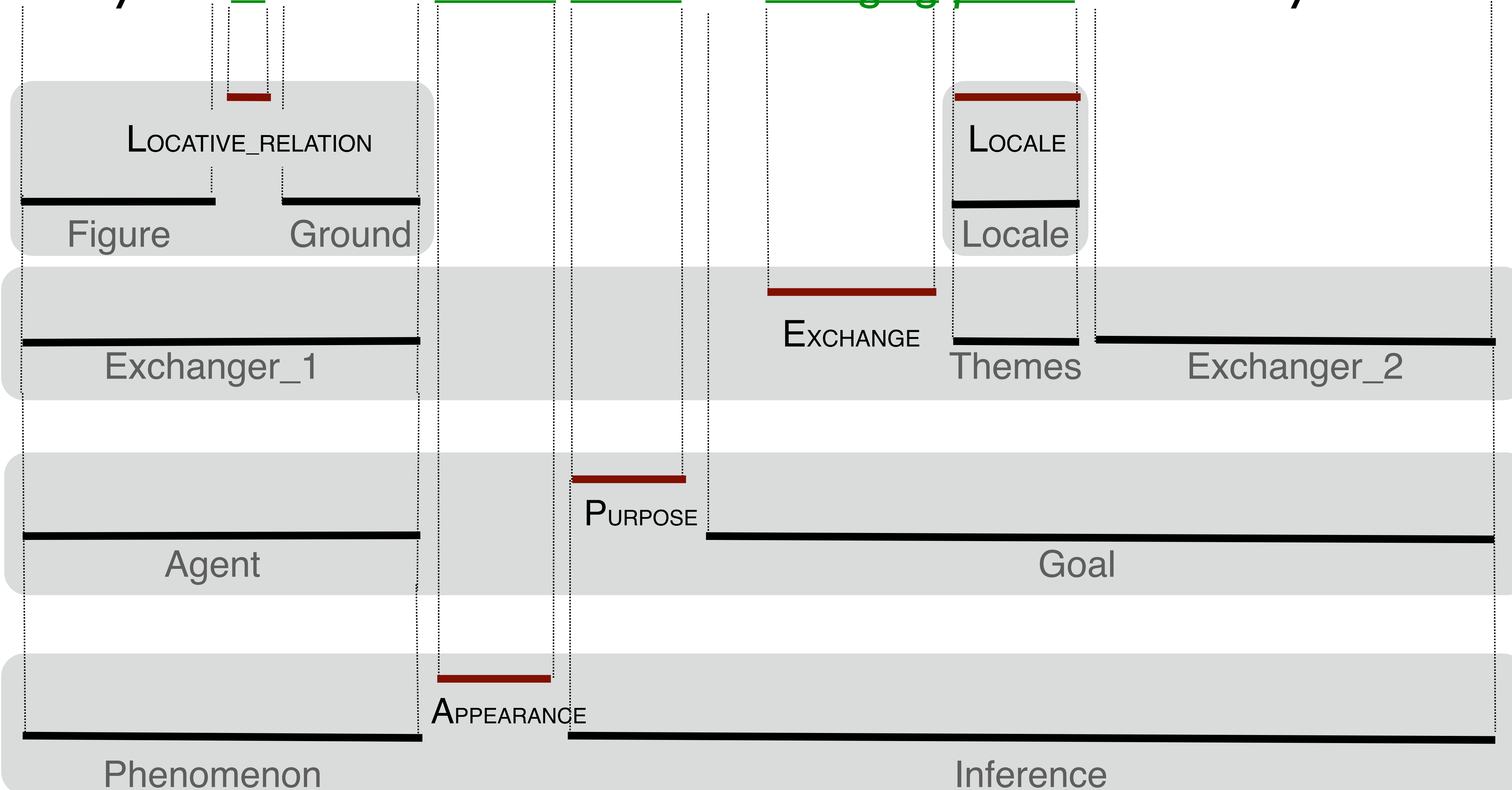
- Several times more labels than traditional shallow semantic parsing
- Annotated data does not have gold syntactic annotation
- Very little labeled data
  - Identifying semantic frames for unknown lexical units
  - Very sparse features

Desired Structure

Everyone in Dublin seems intent on changing places with everyone else .

Desired Structure

Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

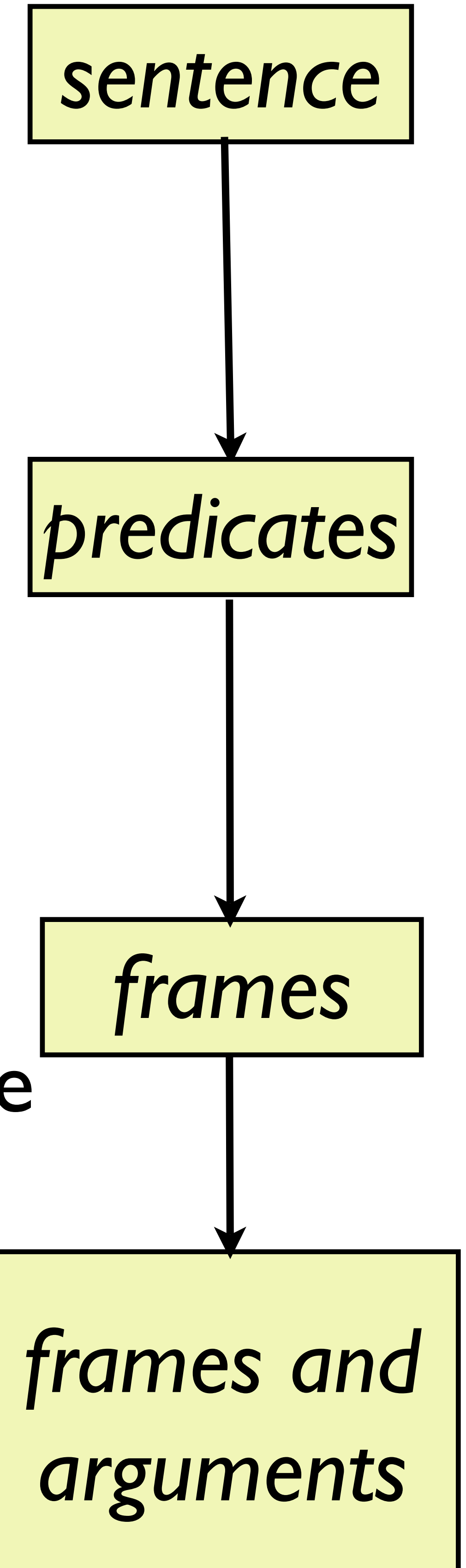


Three Subtasks:

- **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)
- **Frame identification**
  - Labeling each target with a frame type (795 possibilities; ~WSD)
- **Argument identification**
  - Finding each frame's arguments (~SRL; roleset is frame-specific)

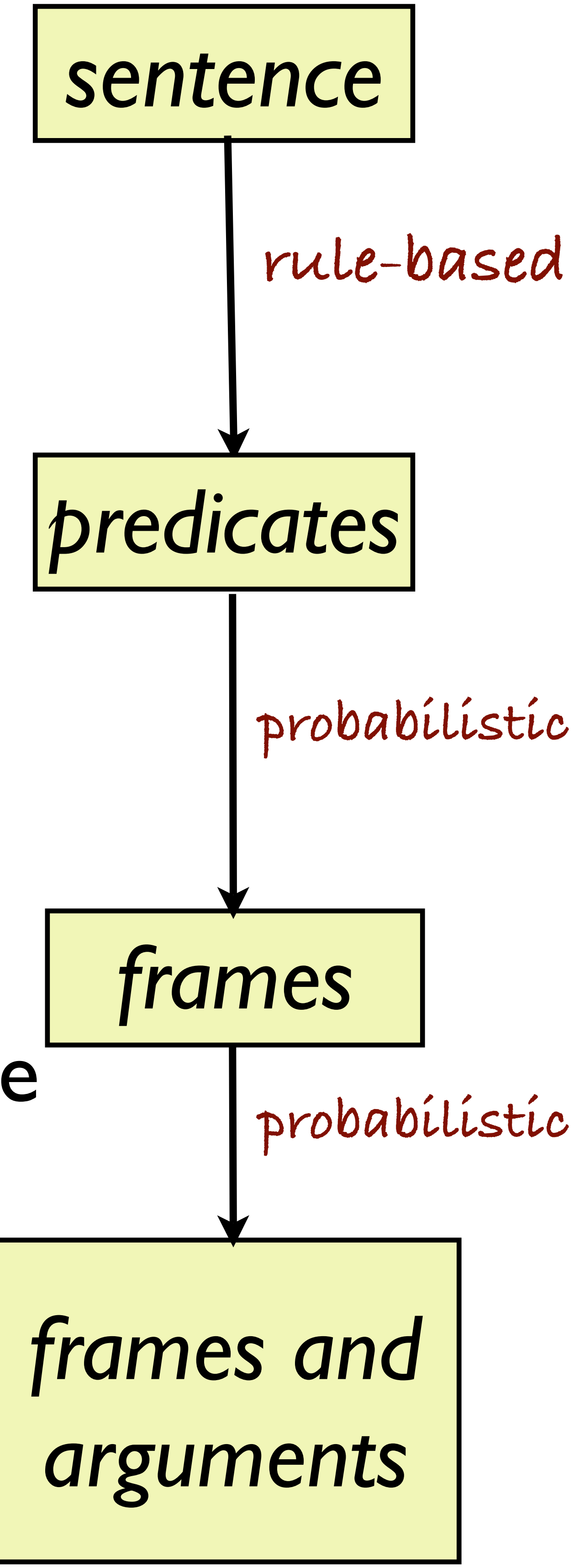
Three Subtasks:

- **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)
- **Frame identification**
  - Labeling each target with a frame type (795 possibilities; ~WSD)
- **Argument identification**
  - Finding each frame's arguments (~SRL; roleset is frame-specific)



Three Subtasks:

- **Target identification**
  - Identifying frame-evoking predicates (nontrivial!)
- **Frame identification**
  - Labeling each target with a frame type (795 possibilities; ~WSD)
- **Argument identification**
  - Finding each frame's arguments (~SRL; roleset is frame-specific)



Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion

Target Identification

Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

Target Identification

Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

- Rule-based identification
  - list of all morphological variants of predicates in the lexicon
  - all prepositions filtered
  - support verbs were not identified
    - J&N'07 filtered these

Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion

Frame Identification

Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

LOCATIVE_RELATION APPEARANCE PURPOSE EXCHANGE LOCALE

Frame Identification

Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

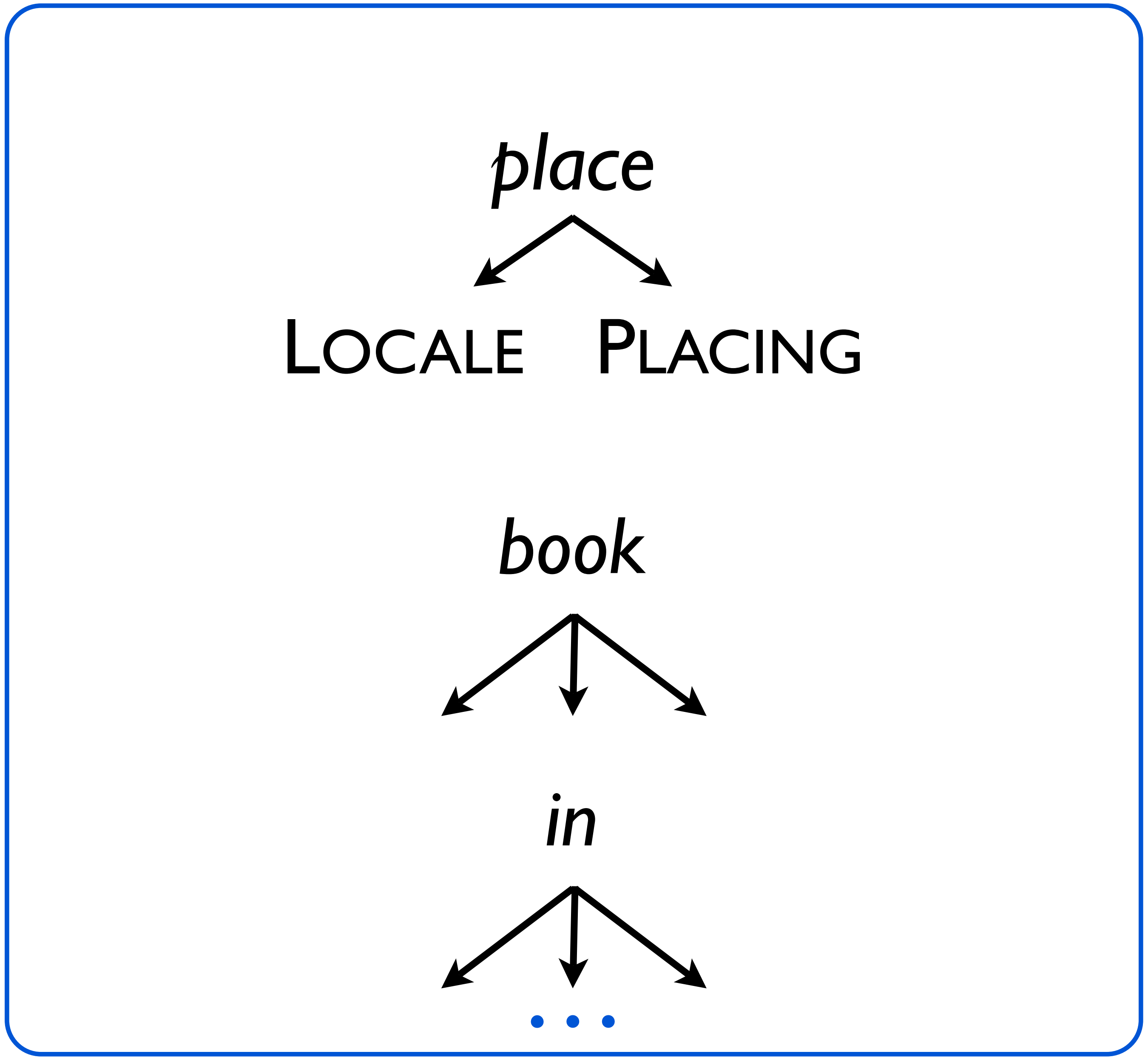
LOCATIVE_RELATION APPEARANCE PURPOSE EXCHANGE LOCALE

J&N'07 used several classifiers for this subtask

Frame Identification

(Johansson and Nugues, 2007)

Seen LUs

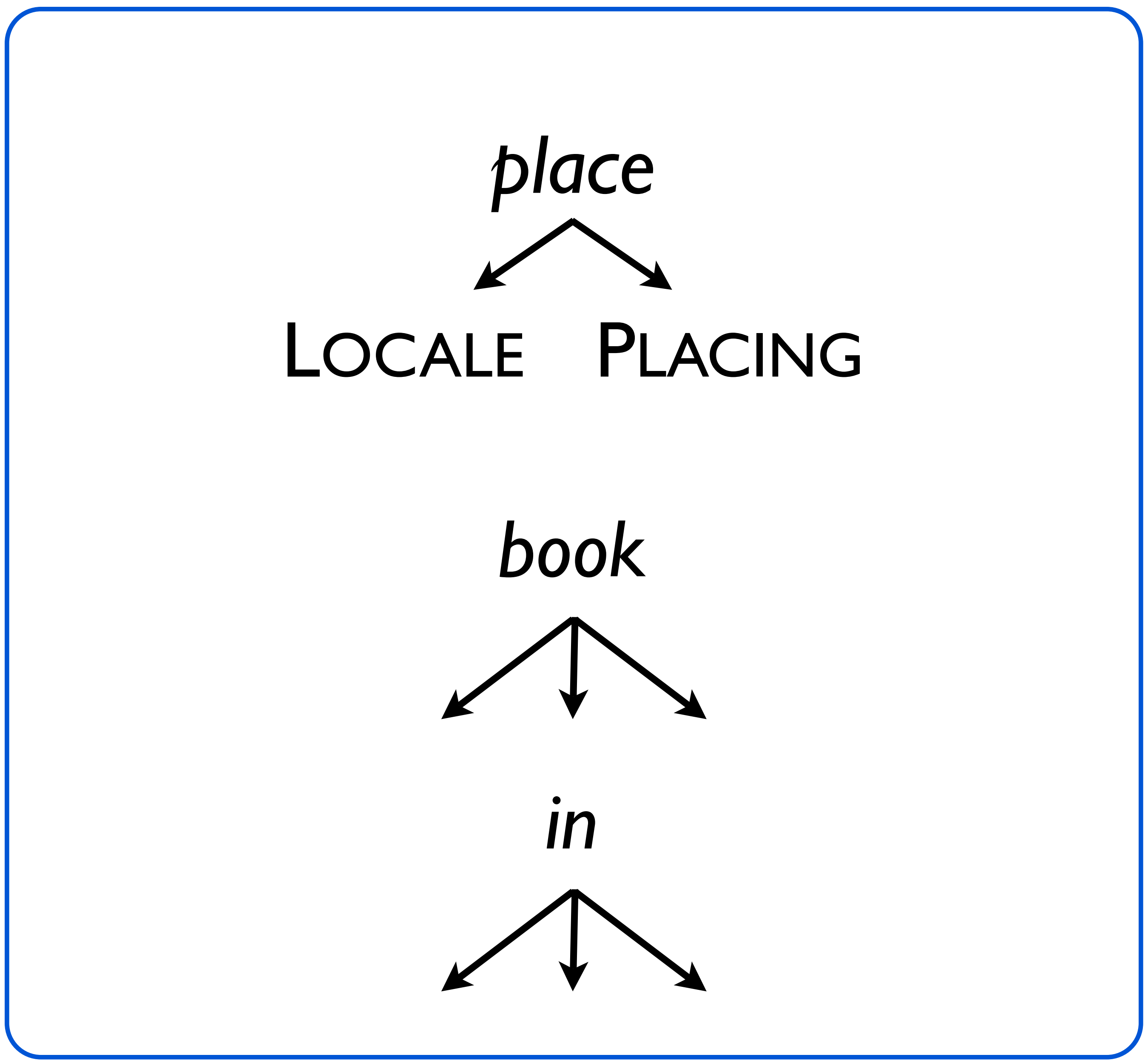


1 classifier

Frame Identification

(Johansson and Nugues, 2007)

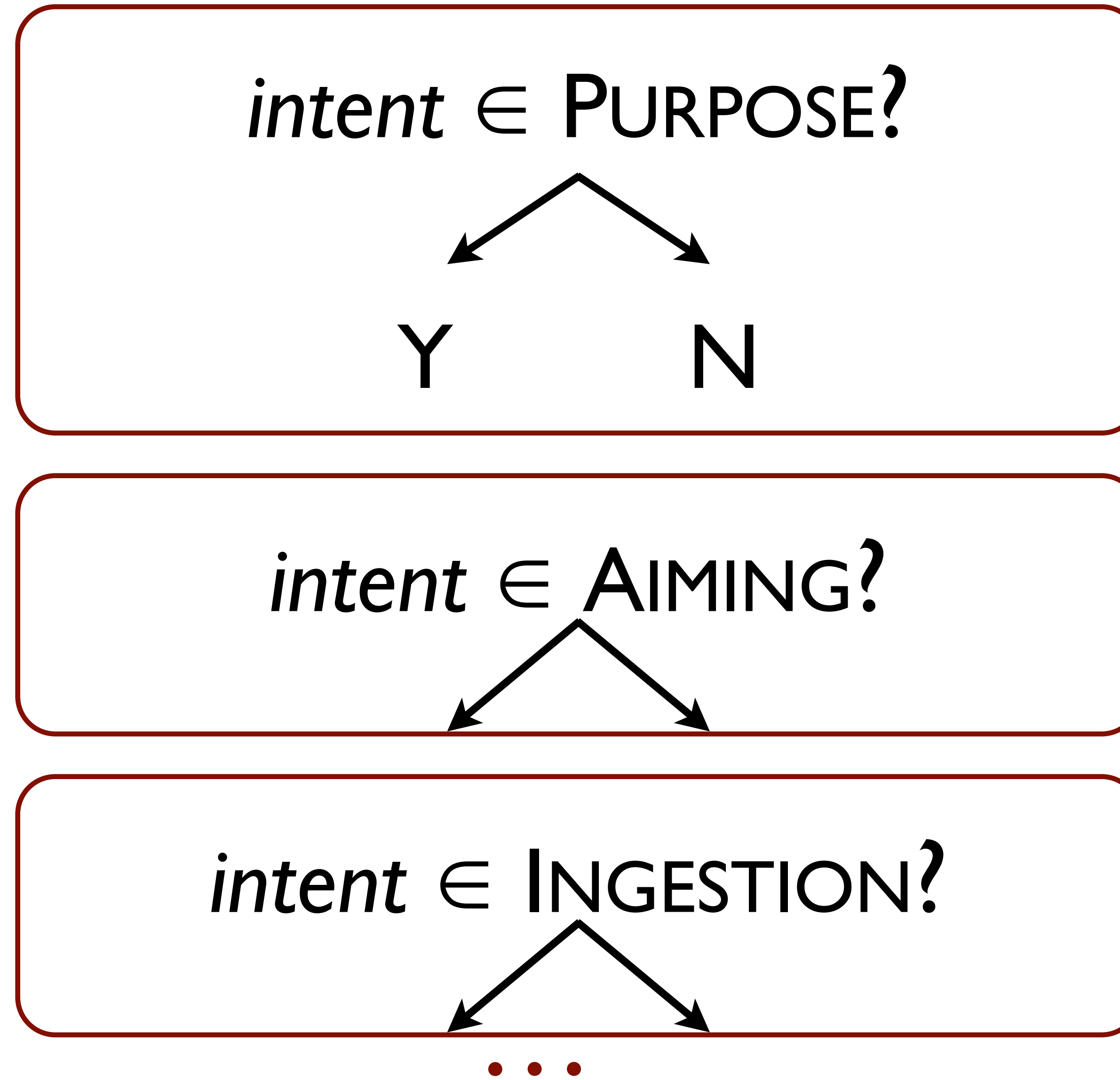
Seen LUs



1 classifier

Unseen LUs

from WordNet-extended set



795 classifiers

Frame Identification

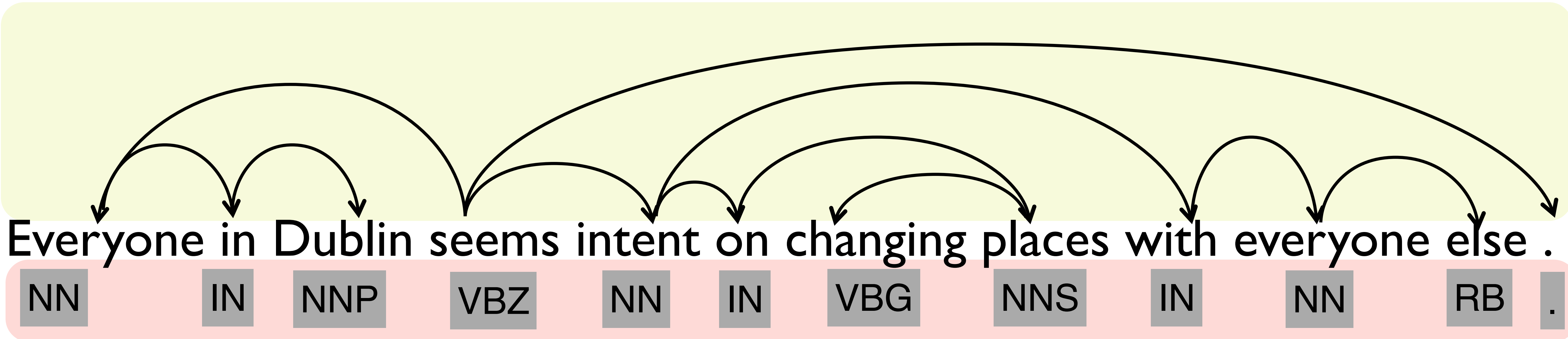
Everyone *in* Dublin *seems* *intent* on *changing* *places* with everyone else .

LOCATIVE_RELATION APPEARANCE PURPOSE EXCHANGE LOCALE

Our approach:

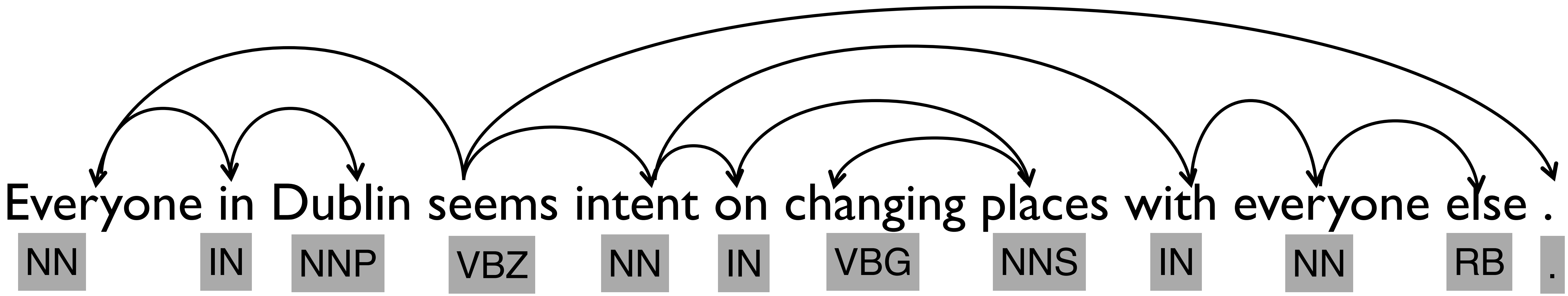
One single model for frame identification

Frame Identification

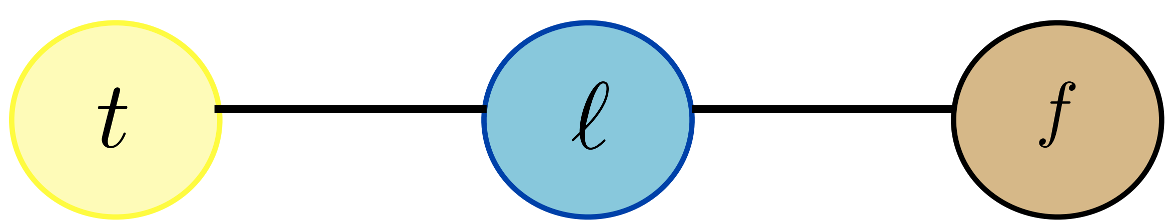


Assume POS tags and dependency trees to be given

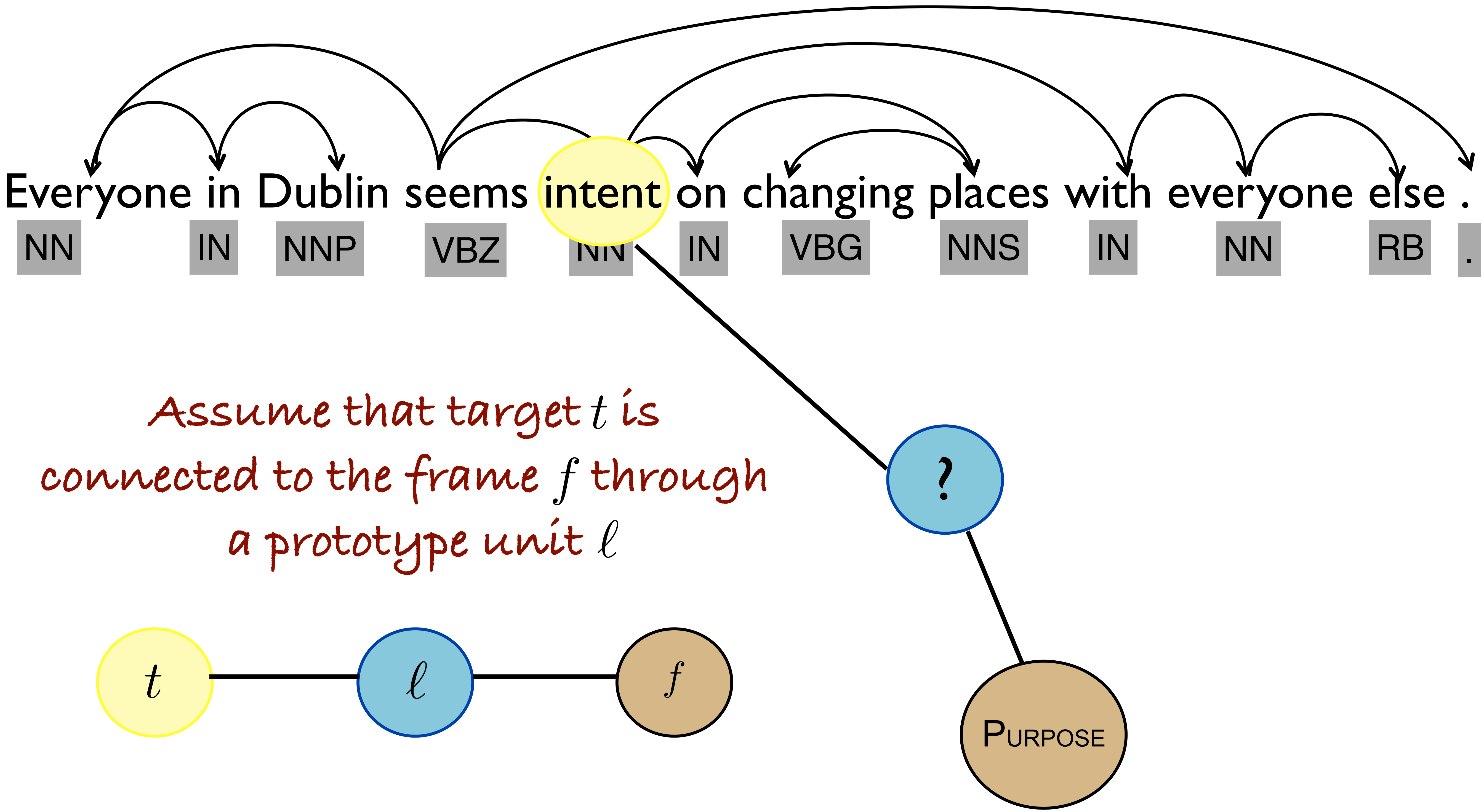
Frame Identification



Assume that target t is connected to the frame f through a prototype unit ℓ



Frame Identification



Assume that target t is connected to the frame f through a prototype unit ℓ

Frame Identification

- Consider the PURPOSE frame

PURPOSE

- Agent
- Goal
- Means
- Attribute
- Value

aim.n, goal.n, object.n, objective.n, purpose.n, target.n

Frame Identification

- Consider the Purpose frame

PURPOSE

Agent

Goal

Means

Attribute

Value

$\ell \in \{$ *aim.n, goal.n, object.n, objective.n, purpose.n, target.n* $\}$

Frame Identification

- Consider the PURPOSE frame

PURPOSE

Agent

Goal

Means

Attribute

Value

$$\ell \in \left\{ \begin{array}{c} \textit{aim.n, goal.n, object.n,} \\ \textit{objective.n, purpose.n,} \\ \textit{target.n} \end{array} \right\}$$

note that the target <u>intent</u> is unseen

Frame Identification

- Consider the Purpose frame

Purpose

Agent

Goal

Means

Attribute

Value

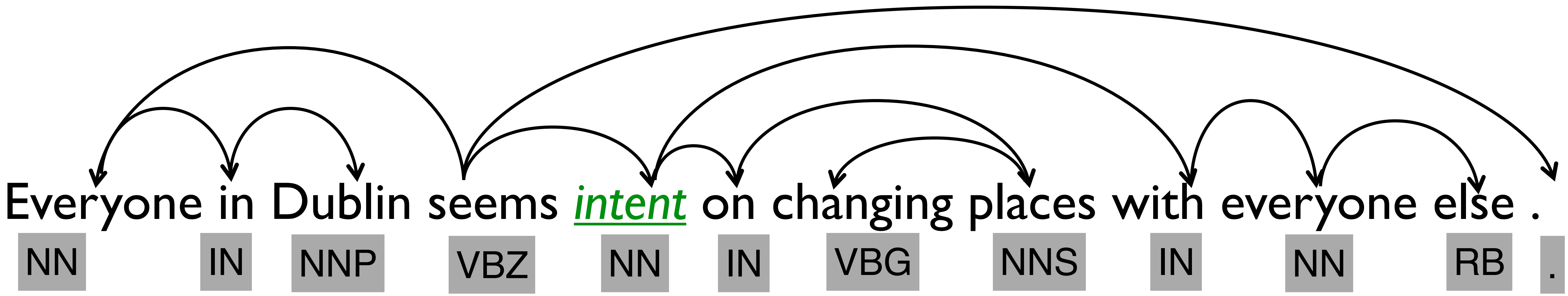
$\ell \in \{$ *aim.n, goal.n, object.n, objective.n, purpose.n, target.n* $\}$

note that the target <u>intent</u> is unseen

but lexical semantic relationships between some ℓ and the target exist

purpose ≈ intent

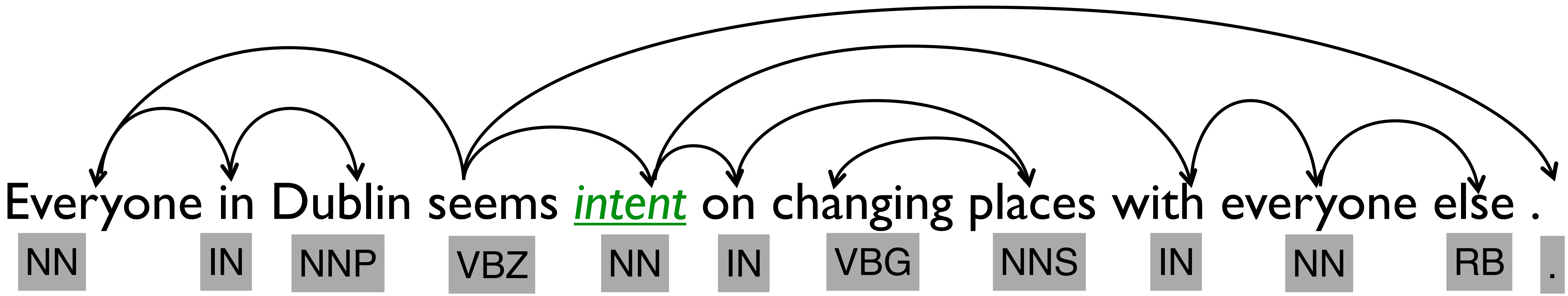
Frame Identification



Thus, we define a probabilistic model:

$$p_\theta(f, \ell \mid t, \mathbf{x}) \propto \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

Frame Identification



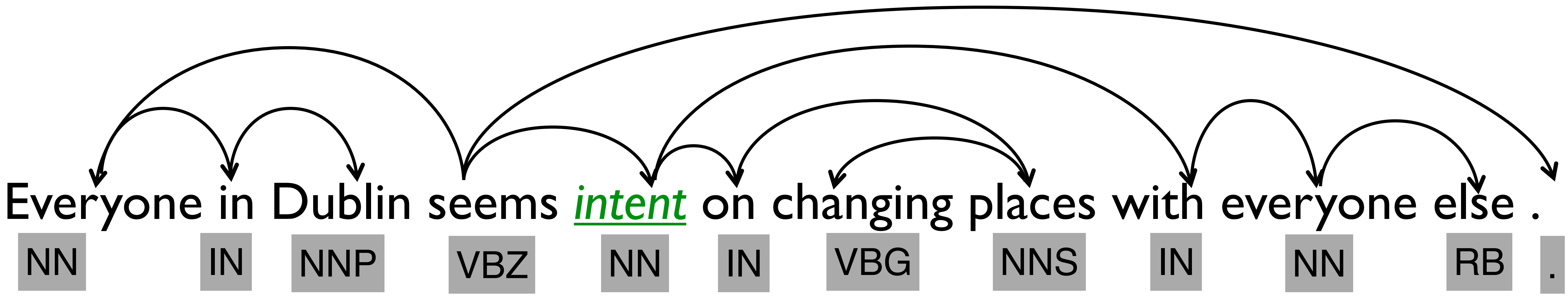
Thus, we define a probabilistic model:

$$p_\theta(f, \ell \mid t, \mathbf{x}) \propto \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

some features looking at the lexical and semantic relationships between ℓ and f

WordNet relationships!

Frame Identification

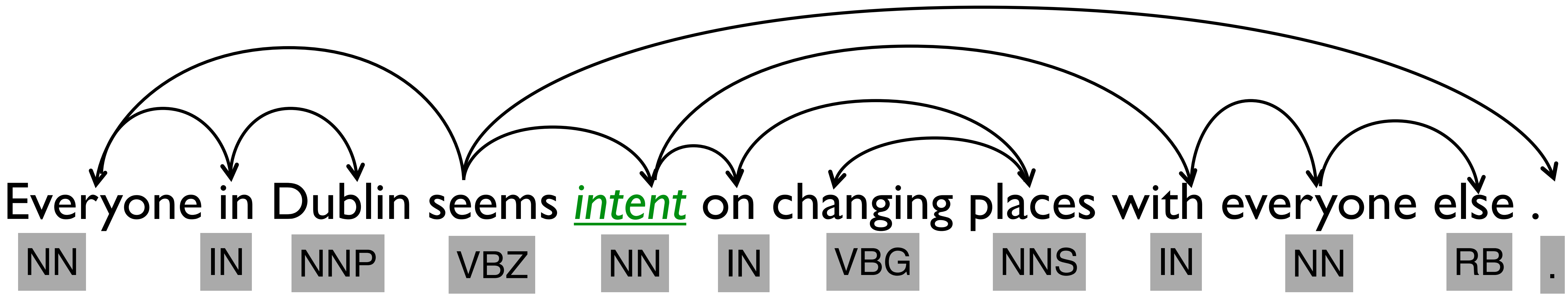


Thus, we define a probabilistic model:

$$p_\theta(f, \ell \mid t, \mathbf{x}) \propto \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

other features looking at the whole sentence structure $\mathbf{x}$

Frame Identification

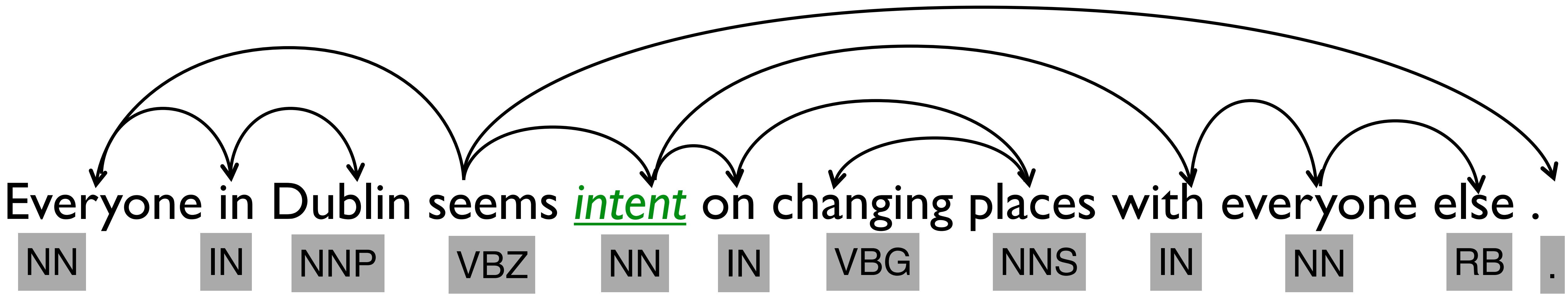


Thus, we define a probabilistic model:

$$p_\theta(f, \ell \mid t, \mathbf{x}) \propto \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

Note that ℓ is unknown

Frame Identification



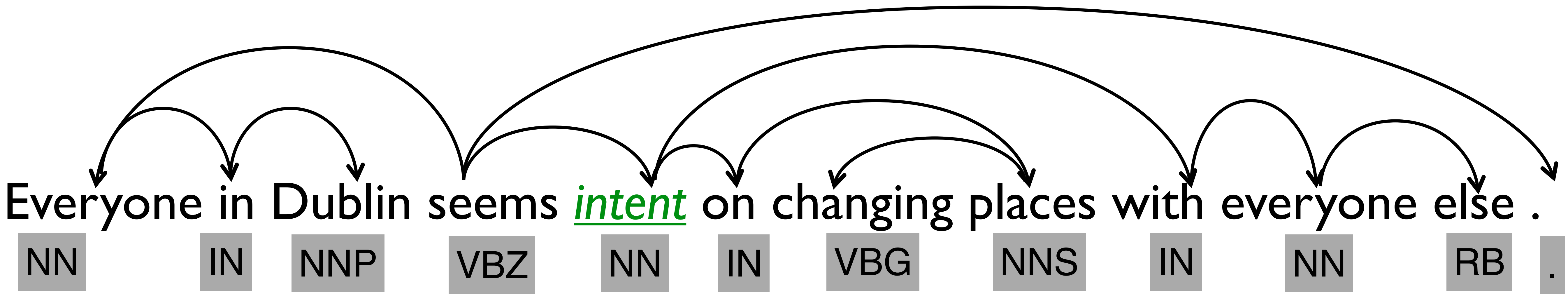
Thus, we define a probabilistic model:

$$p_\theta(f, \ell \mid t, \mathbf{x}) \propto \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

Marginalization of latent variable:

$$p_\theta(f \mid t, \mathbf{x}) \propto \sum_{\ell} \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

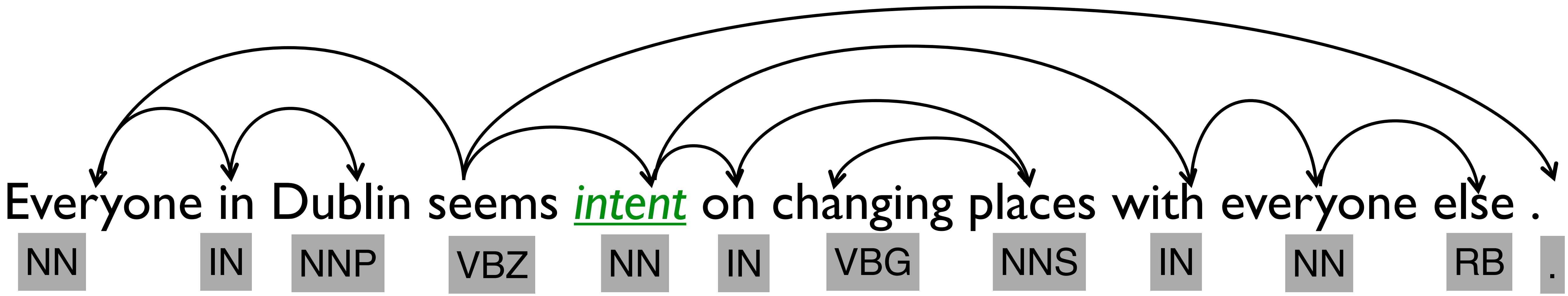
Frame Identification



Inference:

$$\hat{f} \leftarrow \text{argmax}_f \sum_{\ell} \exp \theta^{\top} \mathbf{g}(f, \ell, t, \mathbf{x})$$

Frame Identification



Inference:

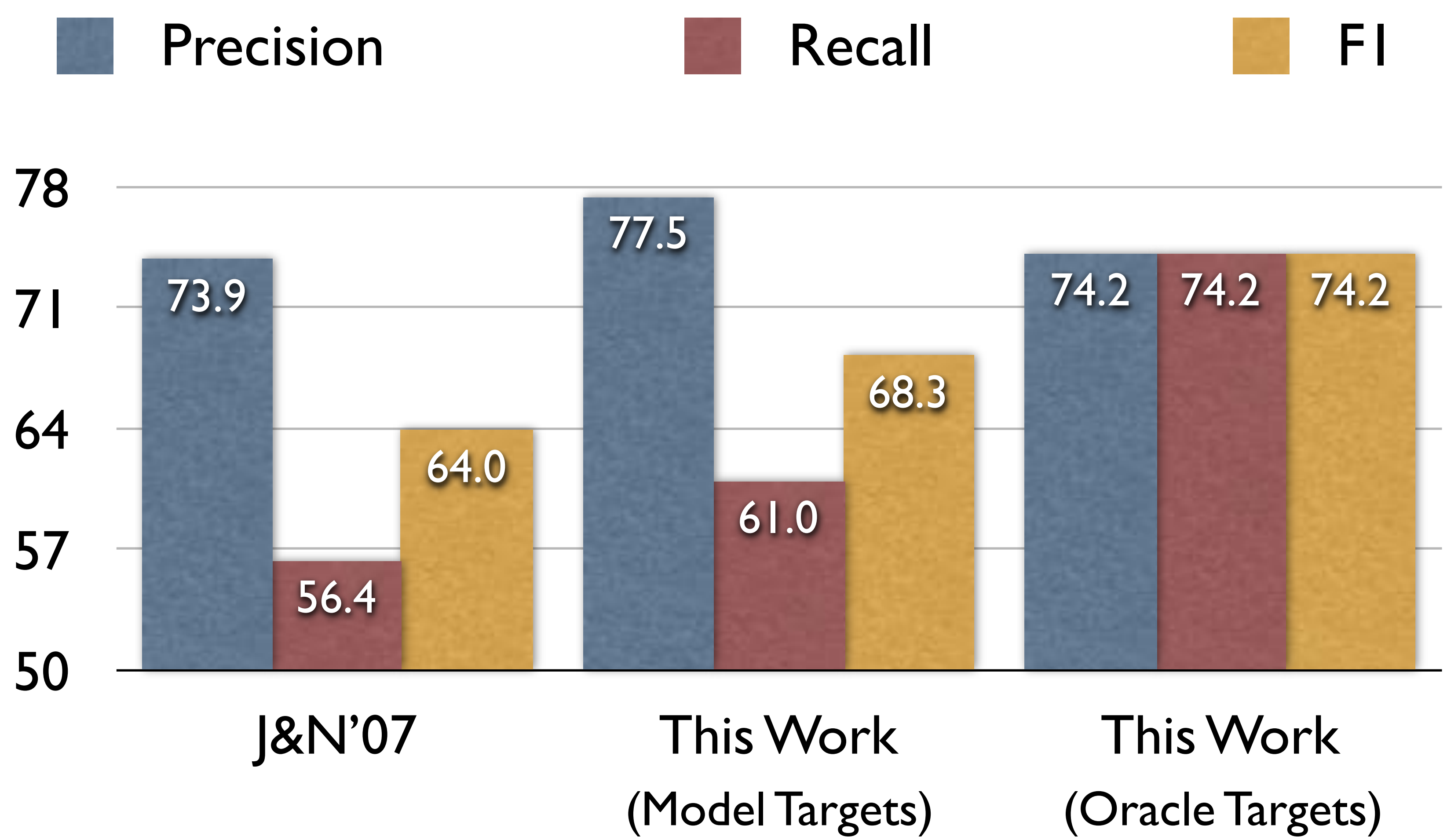
$$\hat{f} \leftarrow \text{argmax}_f \sum_{\ell} \exp \theta^\top \mathbf{g}(f, \ell, t, \mathbf{x})$$

Training:

maximum conditional likelihood

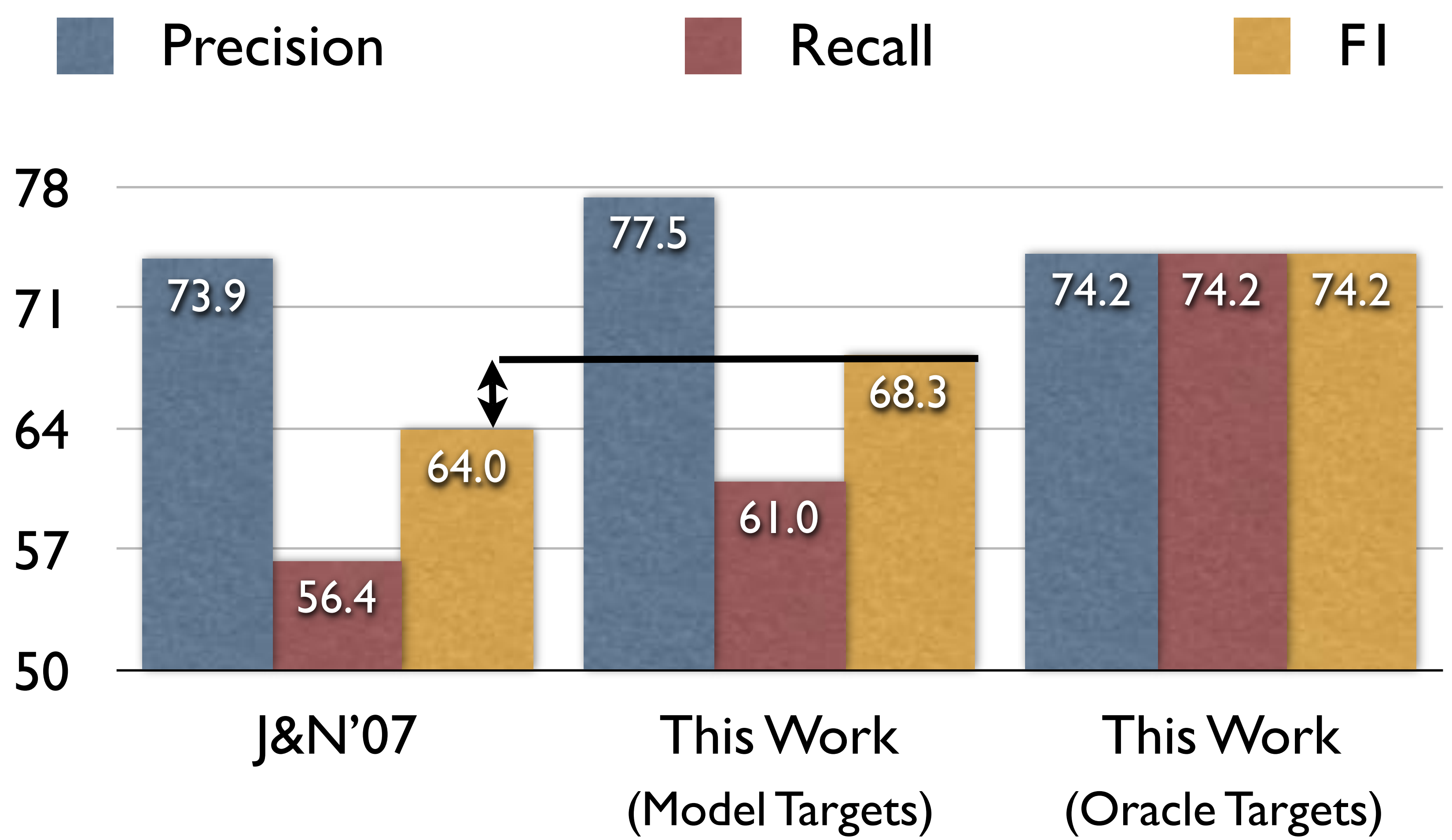
Frame Identification

Results



Frame Identification

Results



significant improvement

Frame Identification

- For gold standard targets, 210 out of 1058 lemmas were unseen
  - 190 of these get some positive score for partial frame matching
  - 4 of these exactly match
  - 44 get 0.5 or more, indicating close match

Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion

Argument Identification

Everyone in Dublin seems intent on *changing* places with everyone else .

Exchanger_1 EXCHANGE Themes Exchanger_2

Argument Identification: The traditional approach

Candidate spans

Everyone in Dublin

in Dublin

on changing places

Two steps:

changing places

with everyone else

places

everyone

......

Argument Identification: The traditional approach

Two steps:

Candidate spans

Everyone in Dublin	✔
in Dublin	✗
on changing places	✗
changing places	✗
with everyone else	✔
places	✔
everyone	✗
……	

binary filtering potential arguments

Argument Identification: The traditional approach

Candidate spans

Two steps:

Everyone in Dublin	✔	Exchanger_1
in Dublin	✗	
on changing places	✗	
changing places	✗	
with everyone else	✔	Exchanger_2
places	✔	Themes
everyone	✗	

......

classification of arguments into different roles

Argument Identification: The traditional approach

Candidate spans

Two steps:

Everyone in Dublin	✔	Exchanger_1
in Dublin	✗	
on changing places	✗	
changing places	✗	
with everyone else	✔	Exchanger_2
places	✔	Themes
everyone	✗	

......

Two steps unnecessary

Argument Identification: Our approach

Roleset for EXCHANGE

Exchanger_1

Exchanger_2

Themes

Exchangers

Theme_1

Theme_2

Manner

Means

...

Candidate spans

Everyone in Dublin

in Dublin

on changing places

changing places

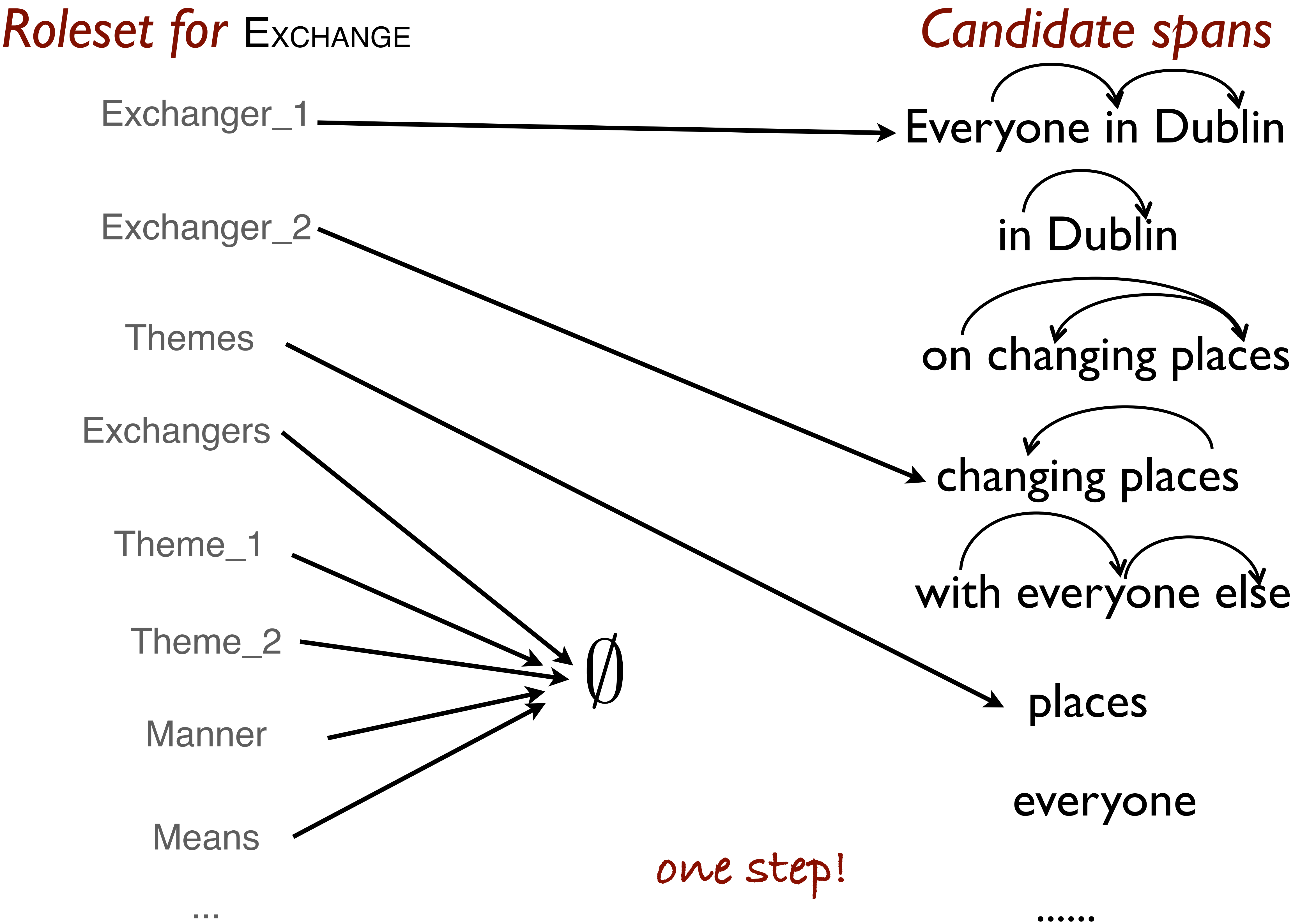
with everyone else

places

everyone

......

Argument Identification: Our approach



Argument Identification

A probabilistic model:

$$p_\psi(r \rightarrow s \mid f, t, \mathbf{x}) \propto \exp \psi^\top \mathbf{h}(r, s, f, t, \mathbf{x})$$

Argument Identification

A probabilistic model:

$$p_\psi(r \rightarrow s \mid f, t, \mathbf{x}) \propto \exp \psi^\top \mathbf{h}(r, s, f, t, \mathbf{x})$$

features looking at the span, the frame, the role and the observed sentence structure

Argument Identification

A probabilistic model:

$$p_\psi(r \rightarrow s \mid f, t, \mathbf{x}) \propto \exp \psi^\top \mathbf{h}(r, s, f, t, \mathbf{x})$$

Decoding:

Best span for each role is selected

For each frame, the best set of non-overlapping arguments is decoded together

Argument Identification

A probabilistic model:

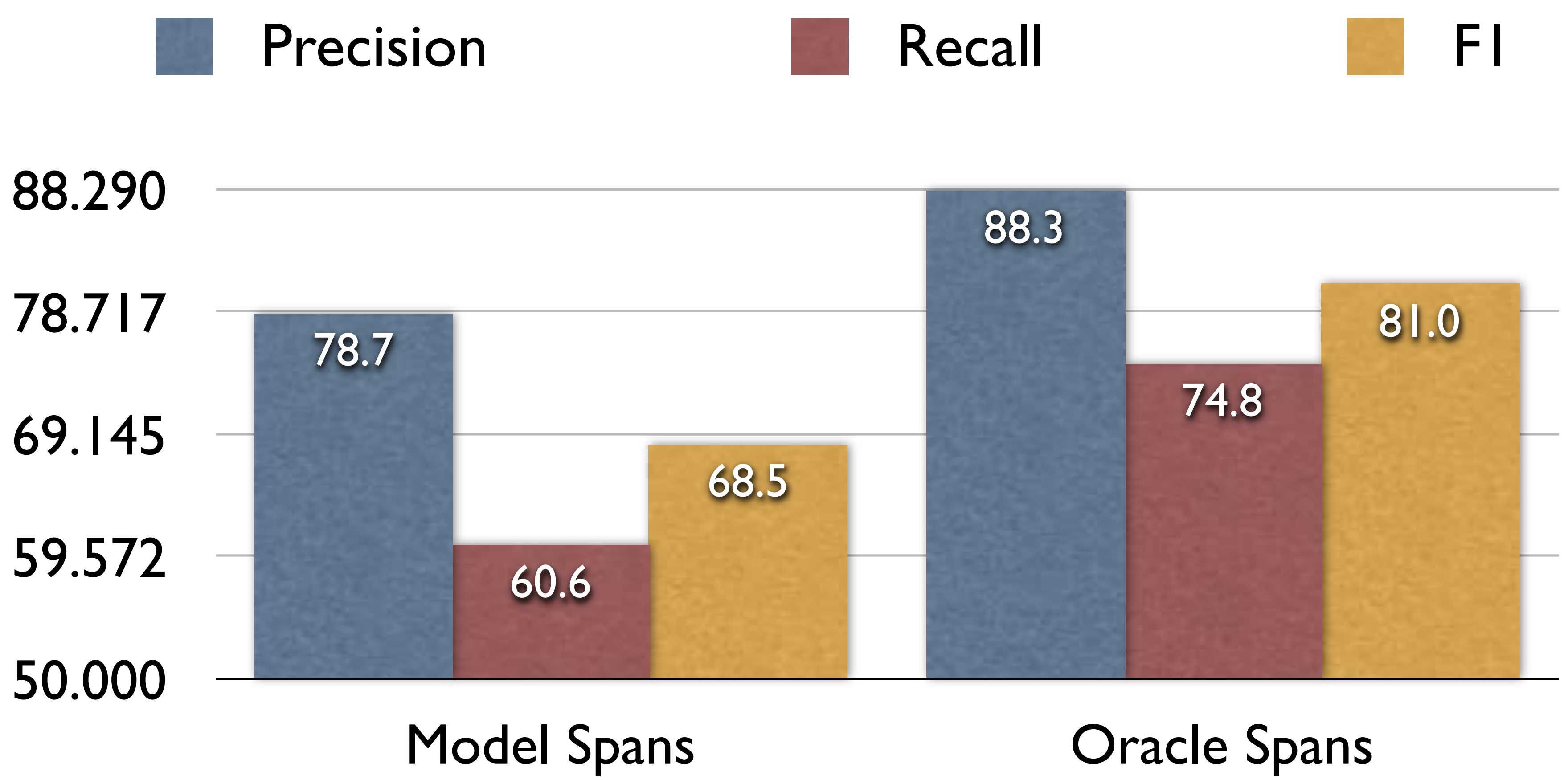
$$p_\psi(r \to s \mid f, t, \mathbf{x}) \propto \exp \psi^\top \mathbf{h}(r, s, f, t, \mathbf{x})$$

Training:

Maximum conditional likelihood

Argument Identification

Results



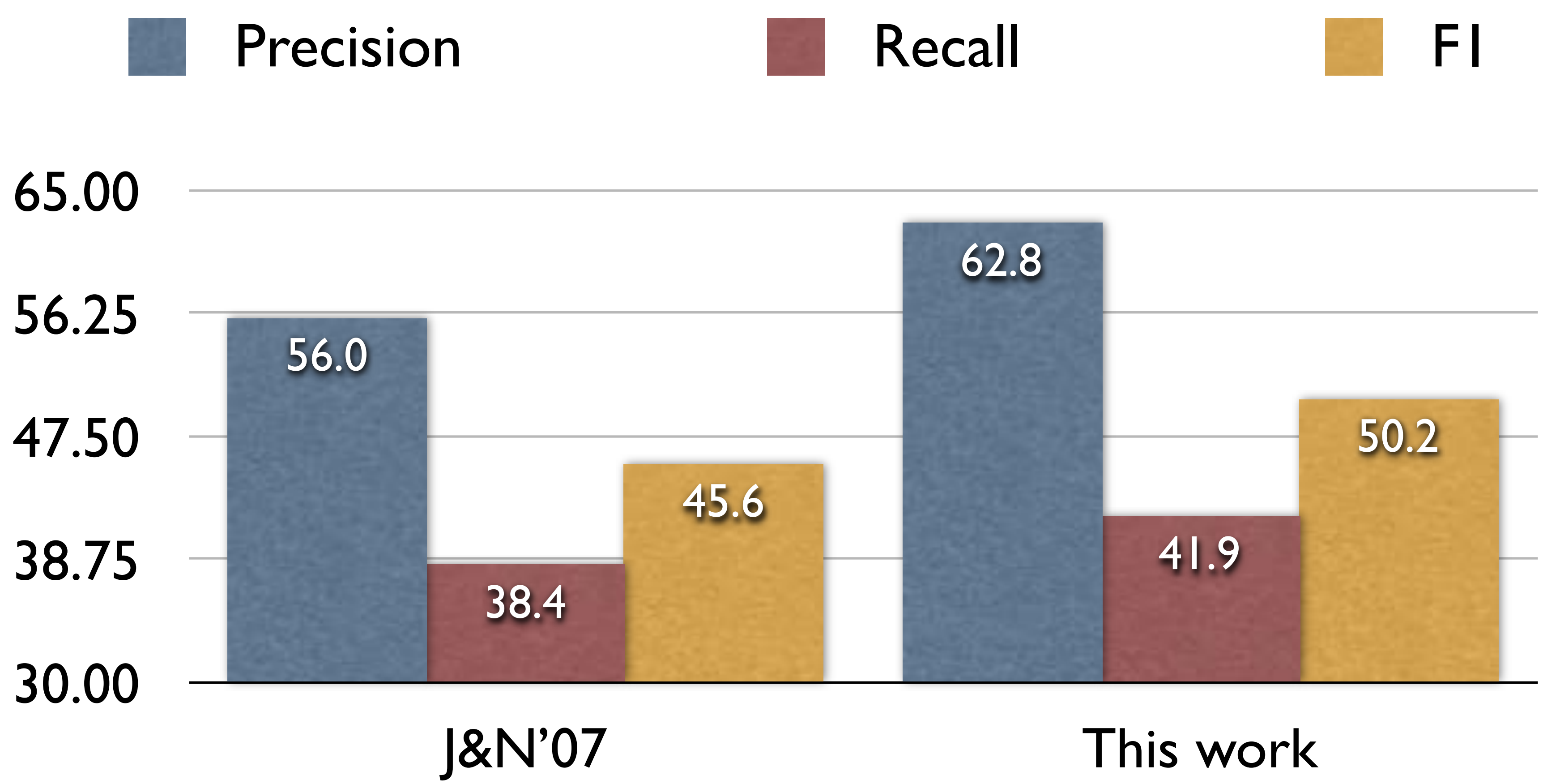
Argument identification only, with gold targets and frames

Outline

- Introduction
- Background and Datasets
- Models and Results
  - Target Identification
  - Frame Identification
  - Argument Identification
  - Final Results
- Conclusion

Full Frame-Semantic Parsing

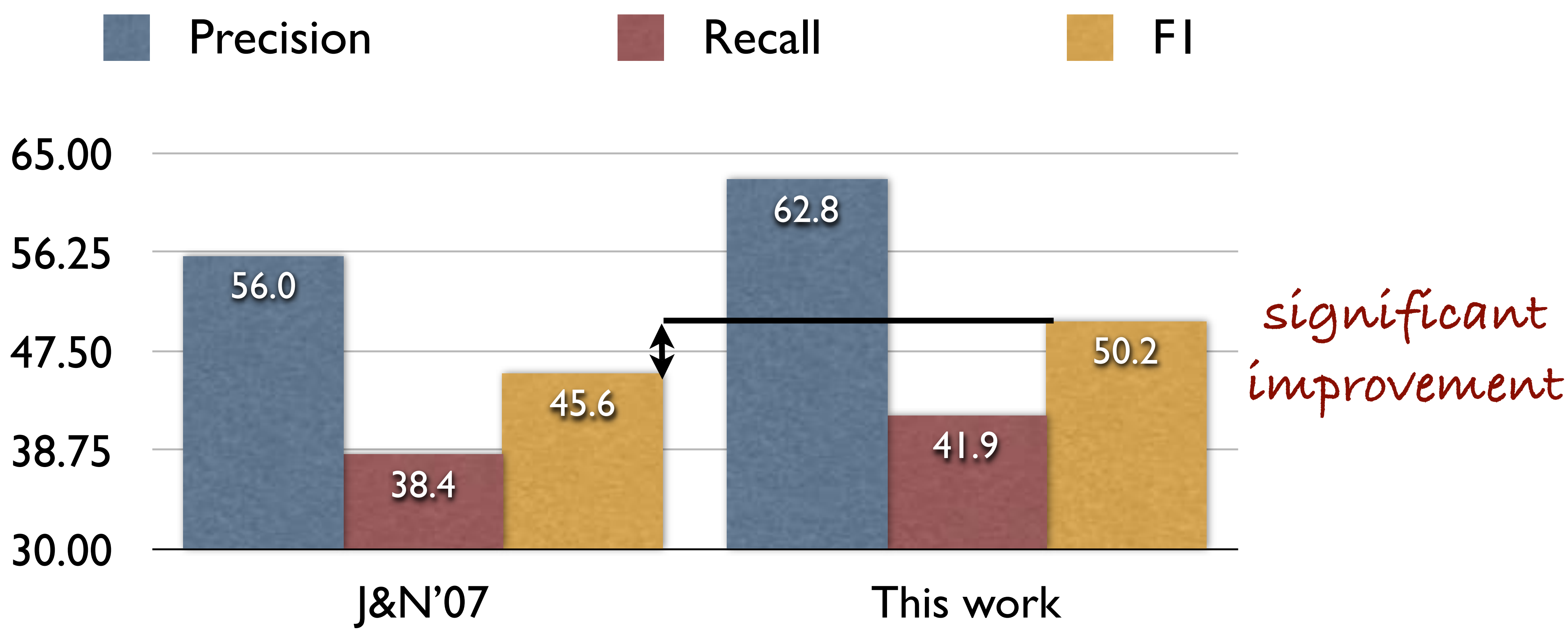
Results



full frame-semantic parsing

Full Frame-Semantic Parsing

Results



full frame-semantic parsing

Conclusion

- Best results to date on frame-semantic parsing
- Only two probabilistic models instead of a cascade of classifiers for the frame-semantic parsing task
- Latent variable model for frame identification
- Better modeling of the argument identification (SRL) stage using only one model instead of two
- Publicly available software: `http://www.ark.cs.cmu.edu/SEMAFOR`

Thanks!

http://www.ark.cs.cmu.edu/SEMAFOR

Thanks!

JUDGMENT_DIRECT_ADDRESS

http://www.ark.cs.cmu.edu/SEMAFOR