SEMAFOR: Frame Argument Resolution with Log-Linear Models

or, The Case of the Missing Arguments

Desai Chen   Nathan Schneider   Dipanjan Das   Noah A. Smith

School of Computer Science
Carnegie Mellon University

SemEval
July 16, 2010

SemEval Task 10: Linking Events and their Participants in Discourse

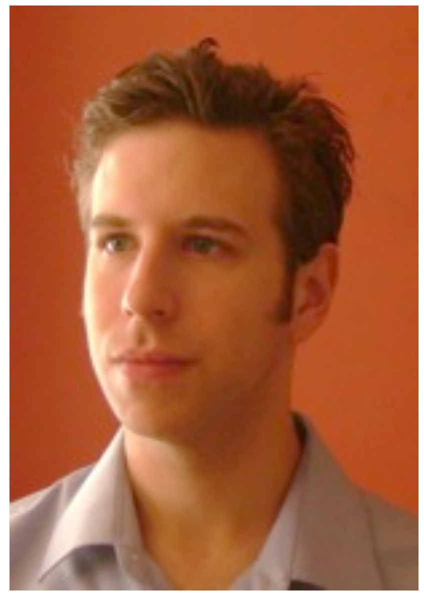
We describe an approach to frame-semantic role labeling and evaluate it on data from this task.

SEMAFOR: Frame Argument Resolution with Log-Linear Models

or, The Case of the Missing Arguments

Desai Chen Nathan Schneider Dipanjan Das Noah A. Smith

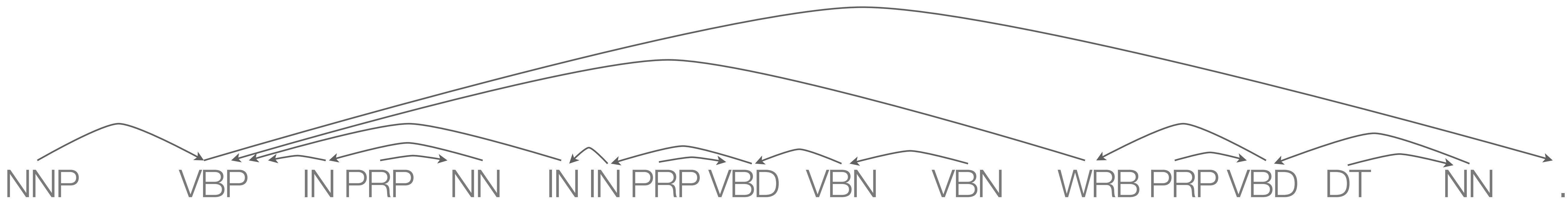
(guy in the front of the room)



ARK lti Carnegie Mellon

We describe an approach to frame-semantic role labeling and evaluate it on data from this task.

Frame SRL



Holmes sprang in his chair as if he had been stung when I read the headline.

(SemEval 2010 trial data)

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn’t encode, for instance, that “as if he had been stung” is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.

Frame SRL

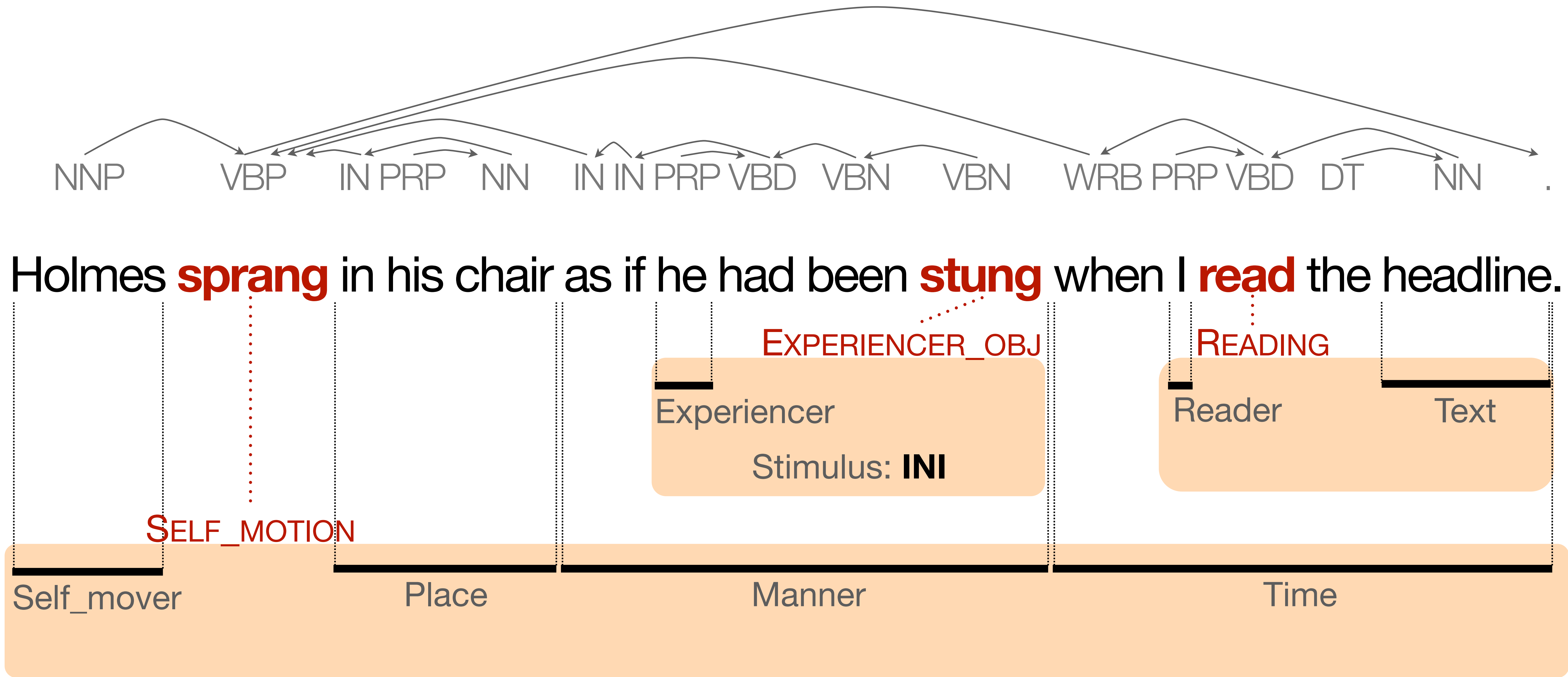
(SemEval 2010 trial data)

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn't encode, for instance, that "as if he had been stung" is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.

Frame SRL



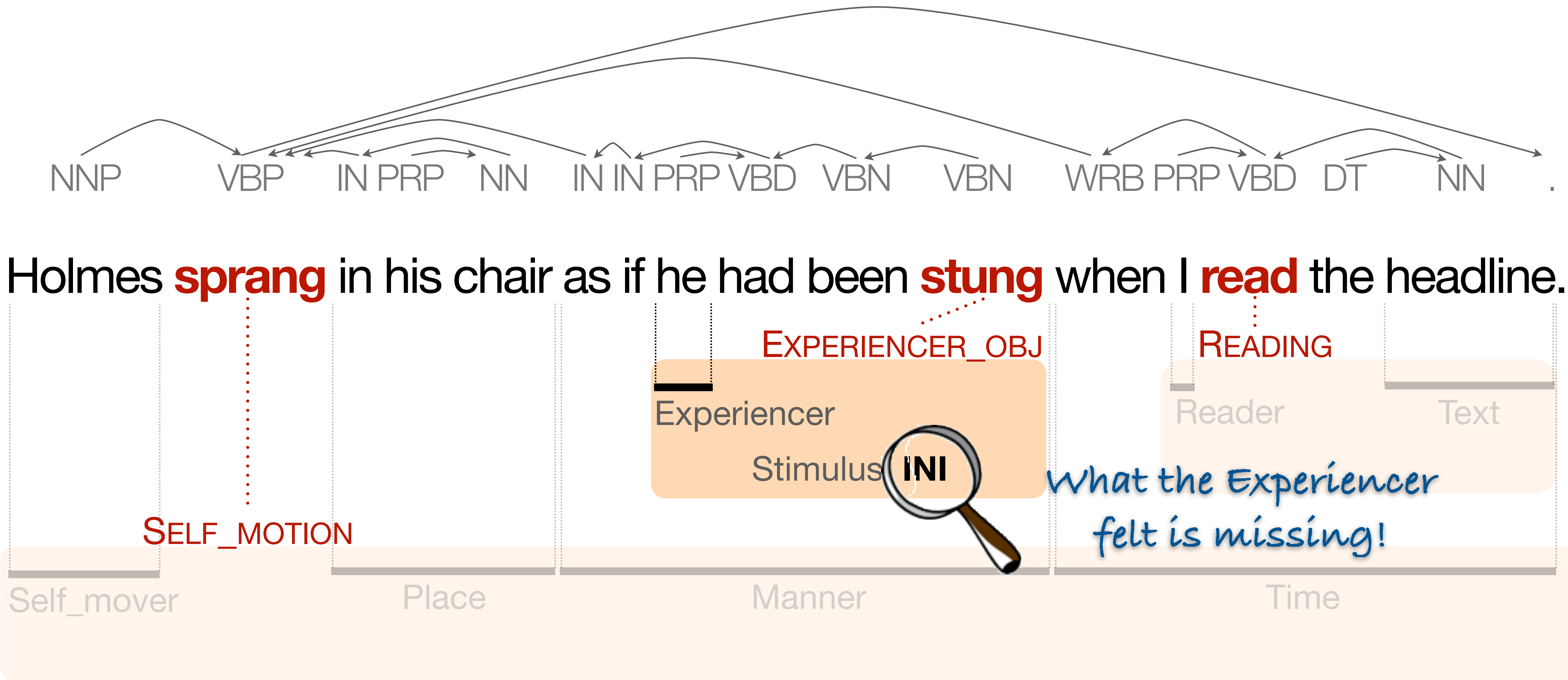
(SemEval 2010 trial data)

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn't encode, for instance, that "as if he had been stung" is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.

Frame SRL



(SemEval 2010 trial data)

This is a full annotation of a sentence in terms of its frames/arguments. Note that this is a *partial* semantic representation: it shows a certain amount of relational meaning but doesn't encode, for instance, that "as if he had been stung" is a hypothetical used to provide imagery for the manner of motion (we infer that it must have been rapid and brought upon by a shocking stimulus).

The SRL task: Given a sentence with POS tags, syntactic dependencies, predicates, and frame names, predict the arguments for each frame role.

New wrinkle in this version of the task: classifying and resolving missing arguments.

Contributions

- Evaluate frame SRL on **new data**
- Experiment with a classifier for **null instantiations** (NIs)
    - implicit interactions in a discourse

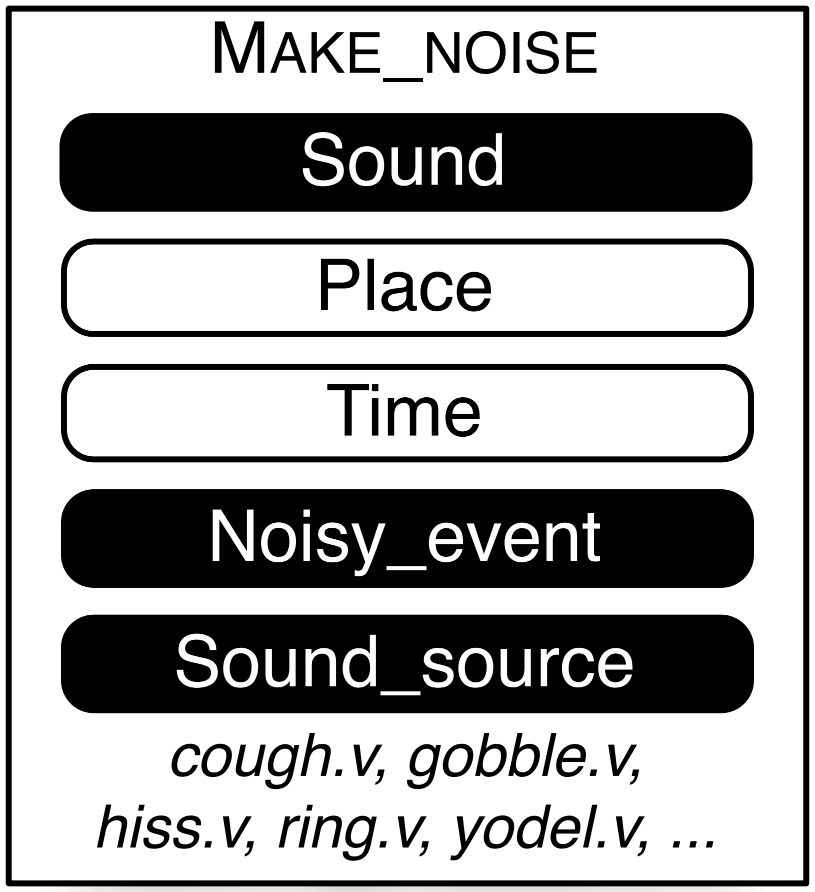
Overview

- ➡ Background: frame SRL
- Overt argument identification
- Null instantiation resolution
- Conclusion

FrameNet

- **FrameNet** (Fillmore et al., 2003) defines semantic frames, roles, and associated predicates
  - provides a linguistically **rich** representation for predicate-argument structures based on the theory of **frame semantics** (Fillmore, 1982)

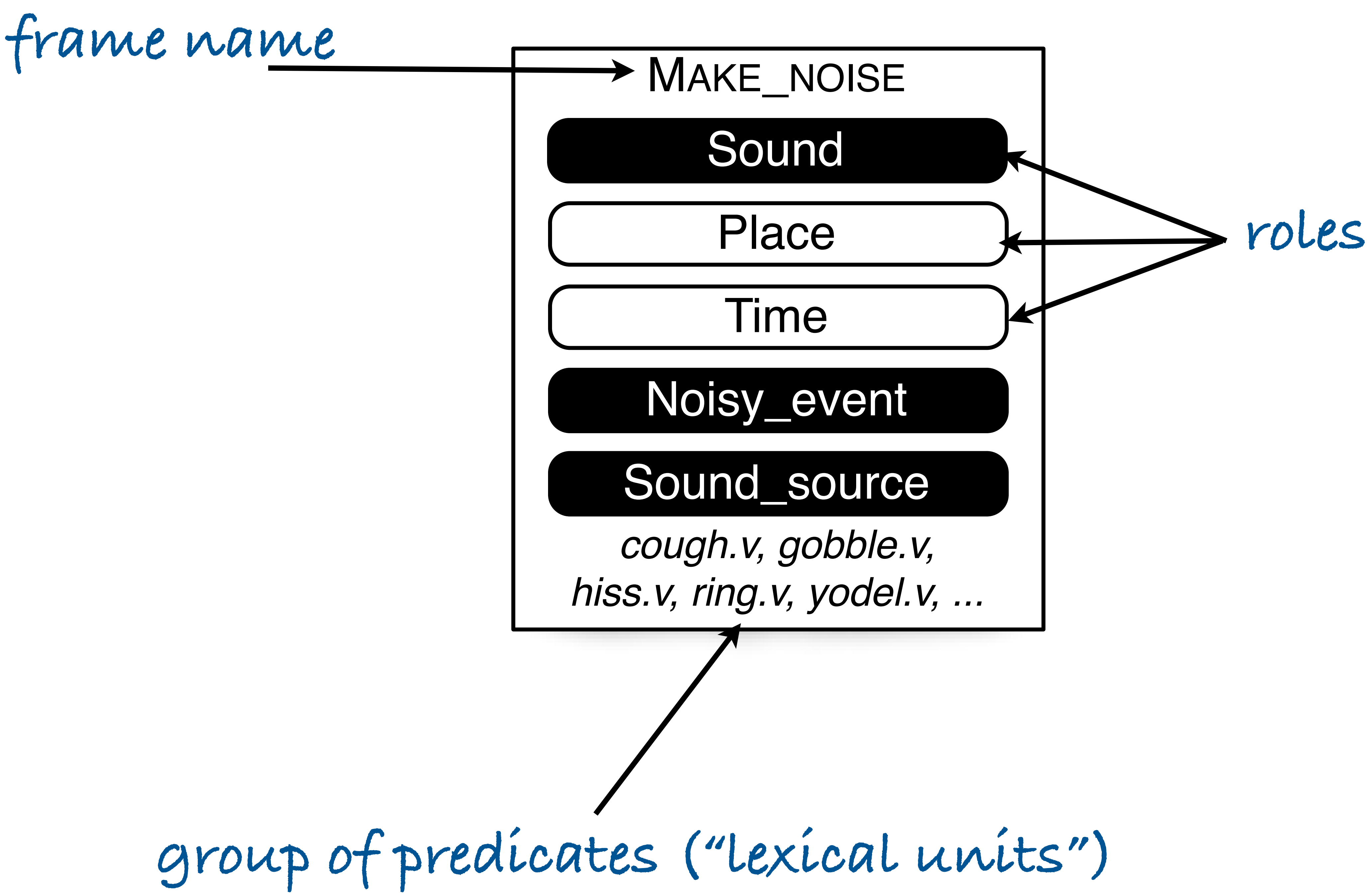
FrameNet



http://framenet.icsi.berkeley.edu

The FrameNet lexicon is a repository of expert information, storing the semantic frames and a number of (frame-specific) roles. Each frame represents a holistic event or scenario, generalizing over specific predicates. It also defines roles for the participants, props, and attributes of the scenario.

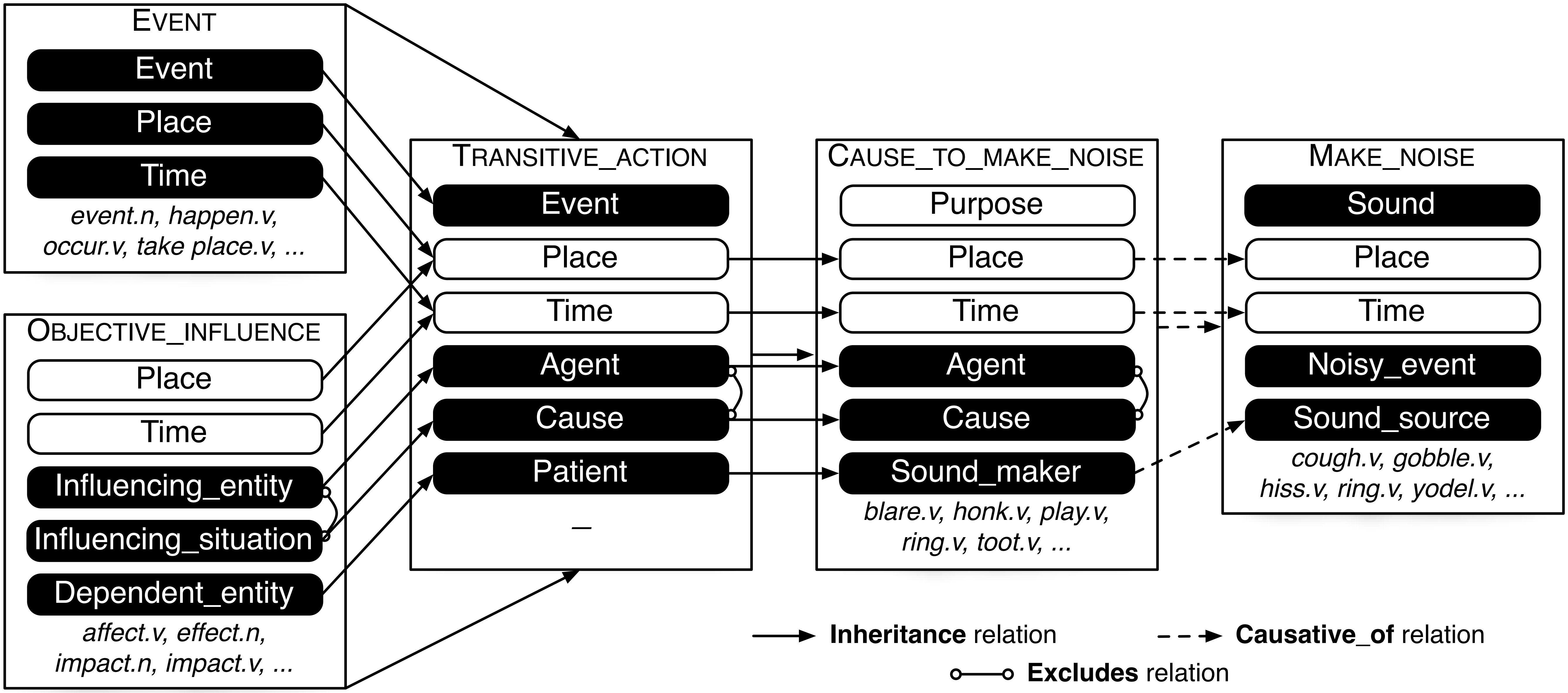
FrameNet



http://framenet.icsi.berkeley.edu

For example, here we show the Make_noise frame that has several roles such as Sound, Noisy_event, Sound_Source, etc. FrameNet also lists some possible lexical units which could evoke these frames. Examples for this frame are cough, gobble, hiss, ring, and so on.

FrameNet



relationships between frames and between roles

http://framenet.icsi.berkeley.edu

The FrameNet lexicon also provides relationships between frames and between roles

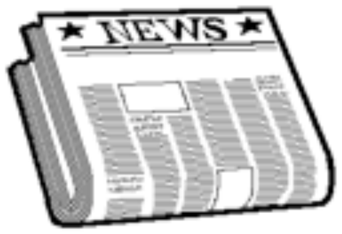
Annotated Data

[SE'07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles.
Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE'10] also has coreference, though we do not make use of this information.

Annotated Data

- Full-text annotations: all frames + arguments
  - [SE'07] SemEval 2007 task data: **news**, **popular nonfiction**, **bureaucratic**



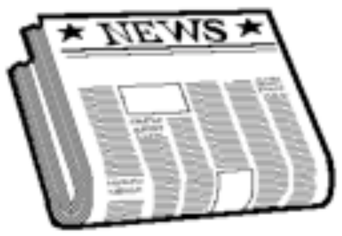
2000 sentences,
50K words

[SE'07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles.
Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE'10] also has coreference, though we do not make use of this information.

Annotated Data

- Full-text annotations: all frames + arguments
  - [SE'07] SemEval 2007 task data: **news**, **popular nonfiction**, **bureaucratic**




2000 sentences,
50K words

  - [SE'10] New SemEval 2010 data: **fiction**

1000 sentences,
17K words
½ train, ½ test

[SE'07] has ANC travel guides, PropBank news, and (mostly) NTI reports on weapons stockpiles.
Unlike other participants, we do not use the 139,000 lexicographic exemplar sentences (except indirectly through features) because the annotations are partial (only 1 frame) and the sample of sentences is biased (they were chosen manually to illustrate variation of arguments).

[SE'10] also has coreference, though we do not make use of this information.

Overview

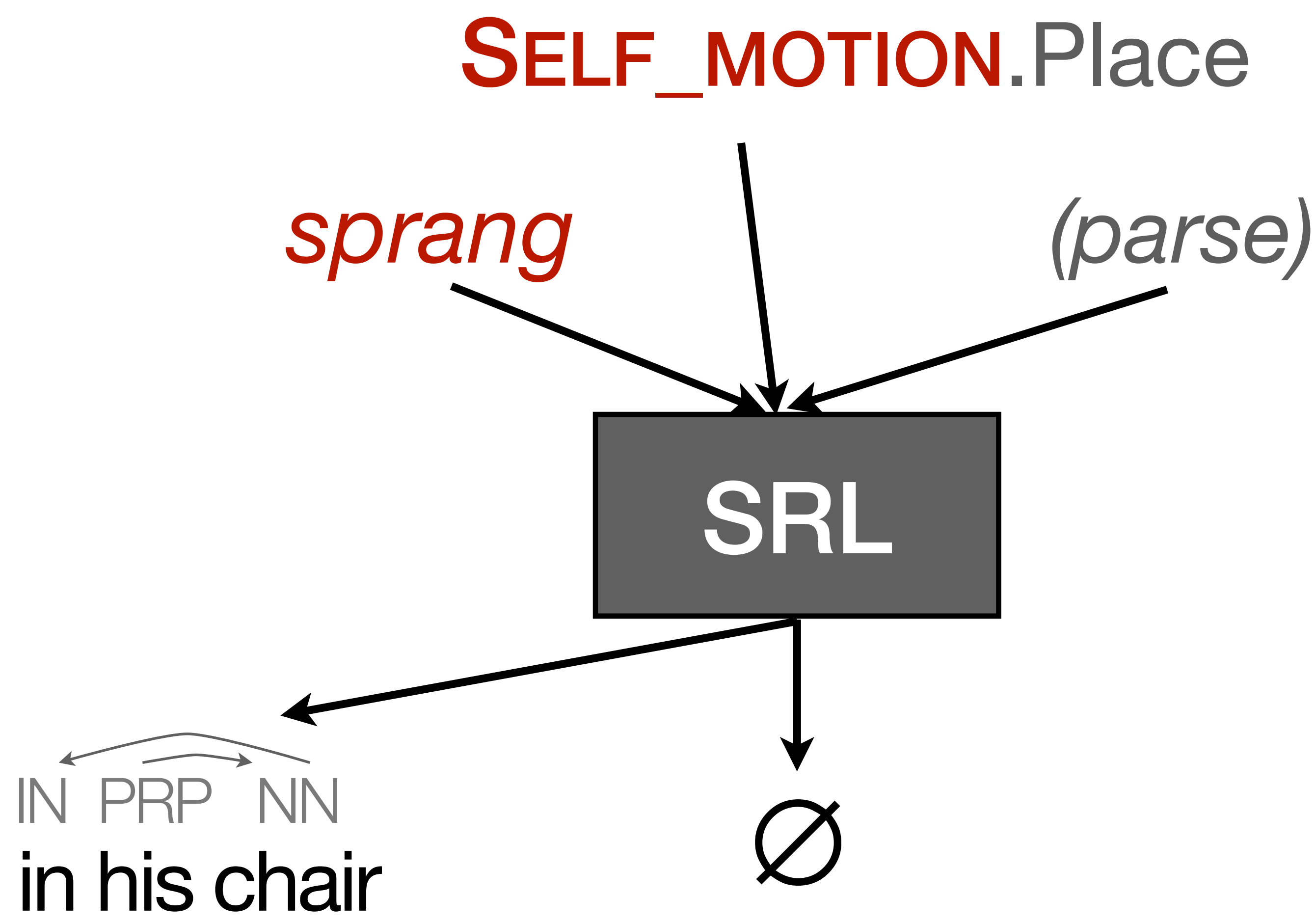
✓ Background: frame SRL

➡ Overt argument identification

- Null instantiation resolution
- Conclusion

Frame SRL: Overt Arguments

We train a **classifier** to pick an argument for each role of each frame.

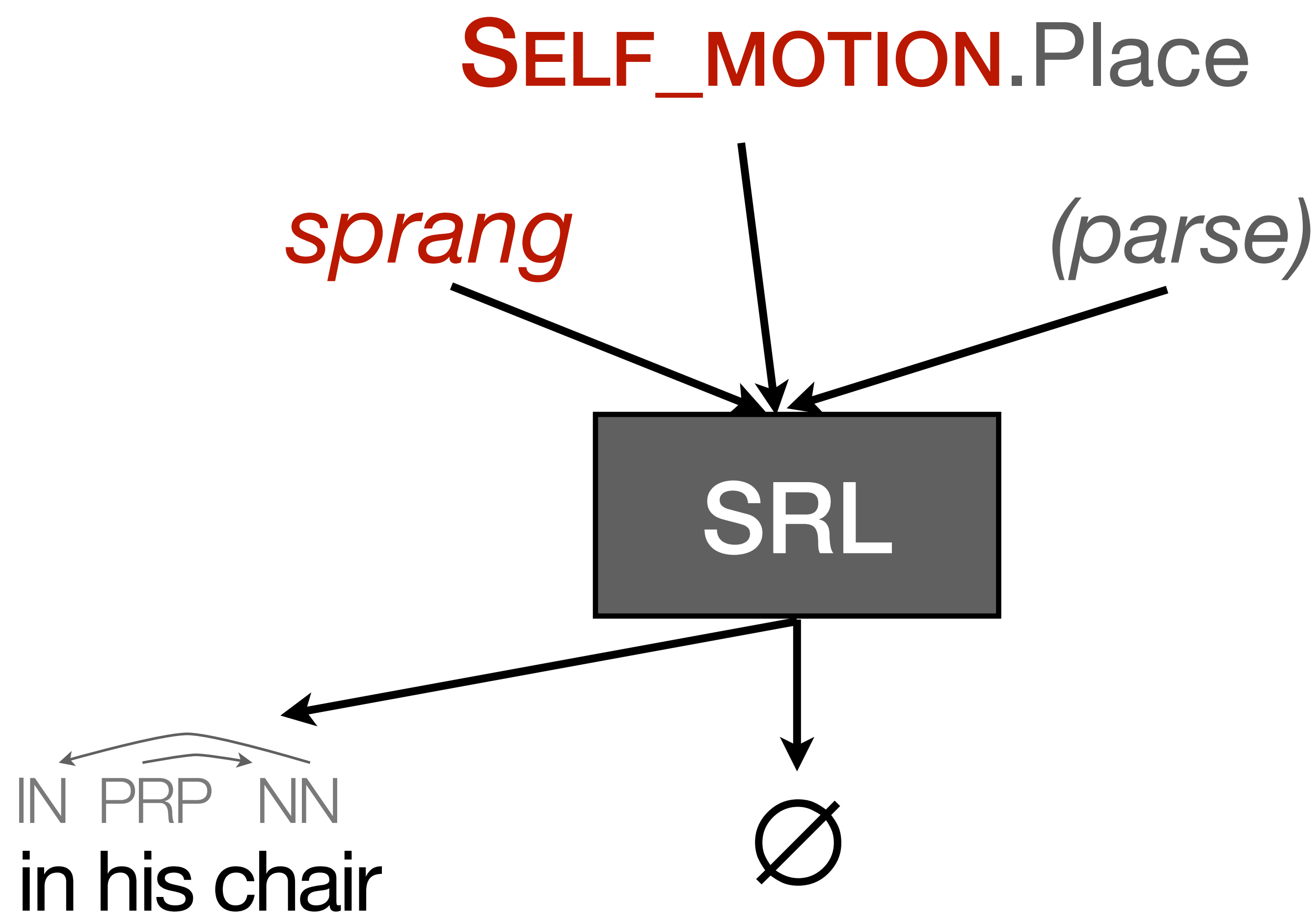


(Das et al., 2010)

See NAACL 2010 paper

Frame SRL: Overt Arguments

We train a **classifier** to pick an argument for each role of each frame.



a probabilistic model with features looking at the span, the frame, the role, and the observed sentence structure

(Das et al., 2010)

See NAACL 2010 paper

Frame SRL: Overt Arguments

sprang ~ SELF_MOTION

An example of the desired mapping. For the predicate 'sprang', each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments

sprang ~ **SELF_MOTION**

Self_mover

Place

Path

Goal

Time

Manner

...

An example of the desired mapping. For the predicate 'sprang', each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments

sprang ~ **SELF_MOTION**

Self_mover

Place

Path

Goal

Time

Manner

...

NNP
Holmes

IN PRP NN
in his chair

PRP NN
his chair

IN IN PRP VBD VBN VBN
as if he had been **stung**

PRP VBD VBN VBN
he had been **stung**

PRP
he

WRB PRP VBD DT NN
when I **read** the headline

PRP
I

DT NN
the headline

...

An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments

sprang ~ **SELF_MOTION**

Self_mover	NNP Holmes IN PRP NN in his chair PRP NN his chair
Place	IN IN PRP VBD VBN VBN as if he had been **stung** PRP VBD VBN VBN he had been **stung**
Path	PRP he
Goal	WRB PRP VBD DT NN when I **read** the headline
Time	PRP I
Manner	DT NN the headline
...	...

An example of the desired mapping. For the predicate 'sprang', each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments

sprang ~ **SELF_MOTION**

Self_mover → Holmes

Place

Path

Goal

Time

Manner

...

NNP
Holmes

IN PRP NN
in his chair

PRP NN
his chair

IN IN PRP VBD VBN VBN
as if he had been **stung**

PRP VBD VBN VBN
he had been **stung**

PRP
he

WRB PRP VBD DT NN
when I **read** the headline

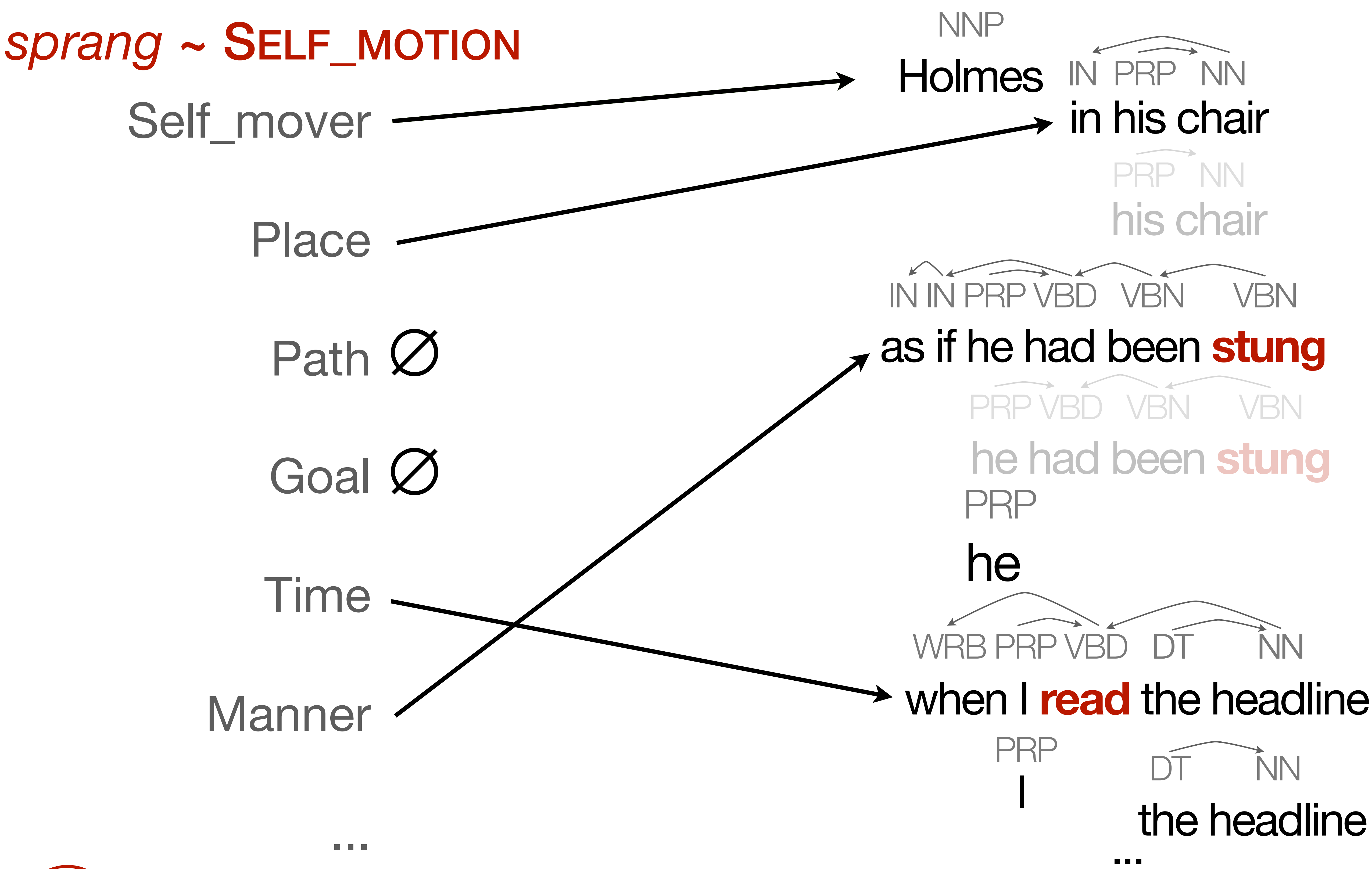
PRP
I

DT NN
the headline

...

An example of the desired mapping. For the predicate ‘sprang’, each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments



An example of the desired mapping. For the predicate 'sprang', each role of the evoked frame is considered separately, and filled with a phrase in the sentence or left empty.

Frame SRL: Overt Arguments

stung ~ EXPERIENCER_OBJ

	NNP Holmes IN PRP NN in his chair
Experiencer	PRP NN his chair
Stimulus	IN IN PRP VBD VBN VBN as if he had been **stung**
Degree	PRP VBD VBN VBN he had been **stung**
Time	PRP he
Manner	WRB PRP VBD DT NN when I **read** the headline
...	PRP I DT NN the headline
	...

...and likewise for ‘stung’, etc.

Frame SRL: Overt Arguments

stung ~ EXPERIENCER_OBJ

Experiencer → he

Stimulus ∅

Degree ∅

Time ∅

Manner ∅

...

NNP
Holmes

IN PRP NN
in his chair

PRP NN
his chair

IN IN PRP VBD VBN VBN
as if he had been **stung**

PRP VBD VBN VBN
he had been **stung**

PRP
he

WRB PRP VBD DT NN
when I **read** the headline

PRP
I

DT NN
the headline

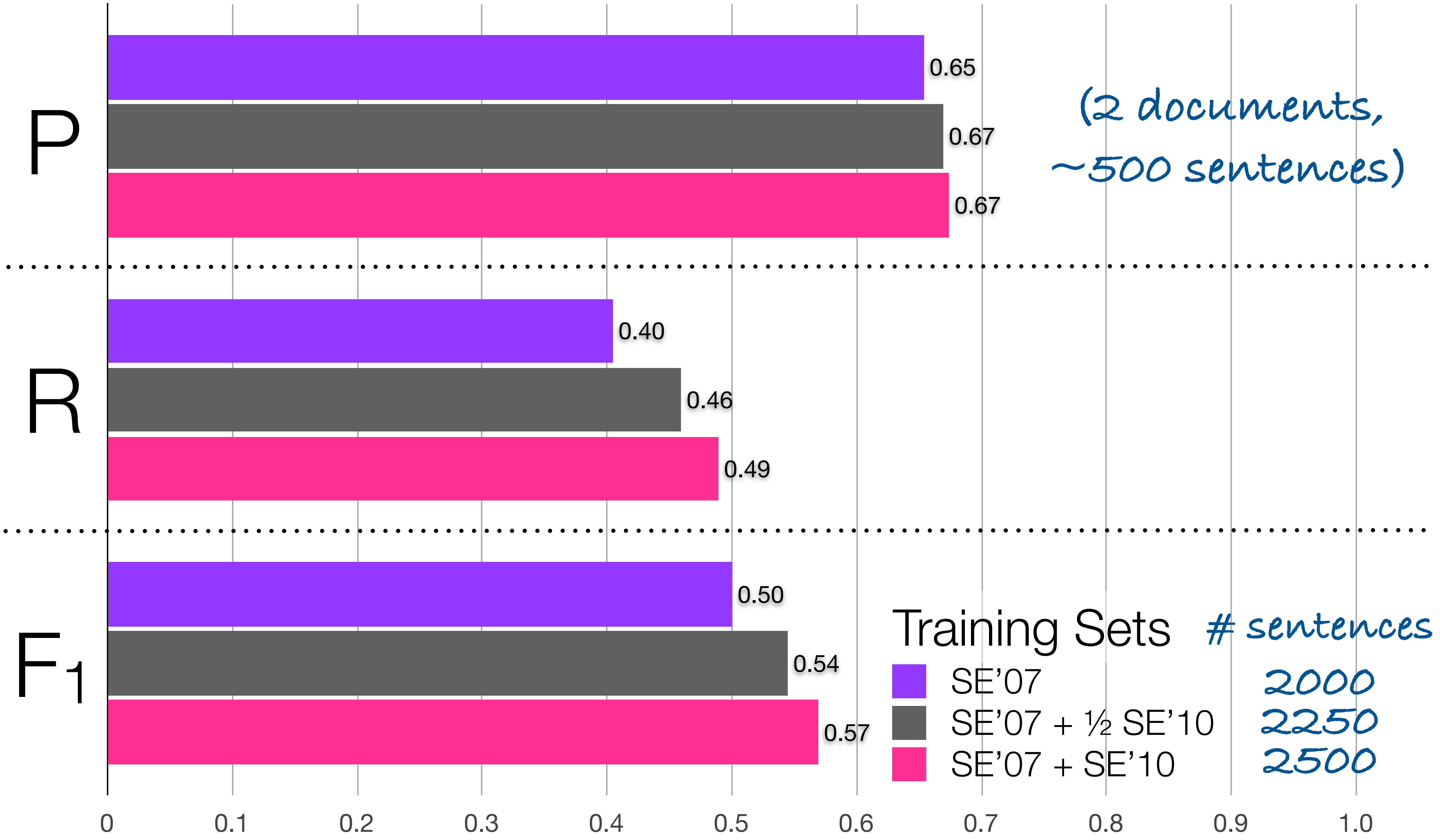
...

...and likewise for ‘stung’, etc.

Frame SRL: Experimental Setup

- SRL component of SEMAFOR 1.0
  (Das et al., 2010; `http://www.ark.cs.cmu.edu/SEMAFOR`)
- Task scoring script for overt argument precision, recall, F_1 on test set
  - Strict matching criterion: argument spans must be exact

SRL on SE'10 Test Data



SE'07: SEMAFOR trained only on old data (different domain from test set)
SE'10: new training data included (same domain as test set)
Adding a small amount of new data helps a lot: (~7% F1): domain issue + so little data to begin with. Suggests even more data might yield substantial improvements!

Scores are microaveraged according to the number of frames in each of the 2 test documents.

Overview

✓ Background: frame SRL

✓ Overt argument identification

➡ Null instantiation resolution

- Conclusion

Null Instantiations

- New this year: classification and resolution of **null instantiations** (NIs), arguments that are nonlocal or implicit in the discourse
  - a role is said to be *null-instantiated* if it has no (overt) argument in the sentence, but has an implicit contextual filler
  - see also (Gerber & Chai, 2010), which considers implicit argument resolution for several (nominal) predicates

(Fillmore, 1986; Ruppenhofer, 2005)

Null Instantiations

- **indefinite null instantiation** (INI): the referent is vague/deemphasized
  - *<u>We</u>* ***ate*** ∅Thing_eaten .
  - *<u>He</u> was* ***stung*** ∅Stimulus .

(Fillmore, 1986; Ruppenhofer, 2005)

Null Instantiations

- **indefinite null instantiation** (INI): the referent is vague/deemphasized
  - *We* ***ate*** ∅Thing_eaten .
  - *He* was ***stung*** ∅Stimulus .
- **definite null instantiation** (DNI): a *specific* referent is obvious from the discourse
  - *They*'ll ***arrive*** *soon* ∅Goal .
    (the goal is implicitly the speaker's location)

(Fillmore, 1986; Ruppenhofer, 2005)

DNI Example: overt nonlocal referent

"I **think** I shall be in a position to **make** the situation rather more **clear** to you before long. It has been an exceedingly **difficult** and most complicated business .

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.

DNI Example: overt nonlocal referent

"I **think** I shall be in a position to **make** the situation rather more **clear** to you before long. It has been an exceedingly **difficult** and most complicated business ∅Experiencer.

Degree DIFFICULTY

Activity

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.

DNI Example: overt nonlocal referent

“I **think** <u>I</u> Experiencer shall be in a position to **make** the situation rather more **clear** to you before long. It has been an <u>exceedingly</u> **difficult** and most complicated <u>business</u> ∅Experiencer.

Degree

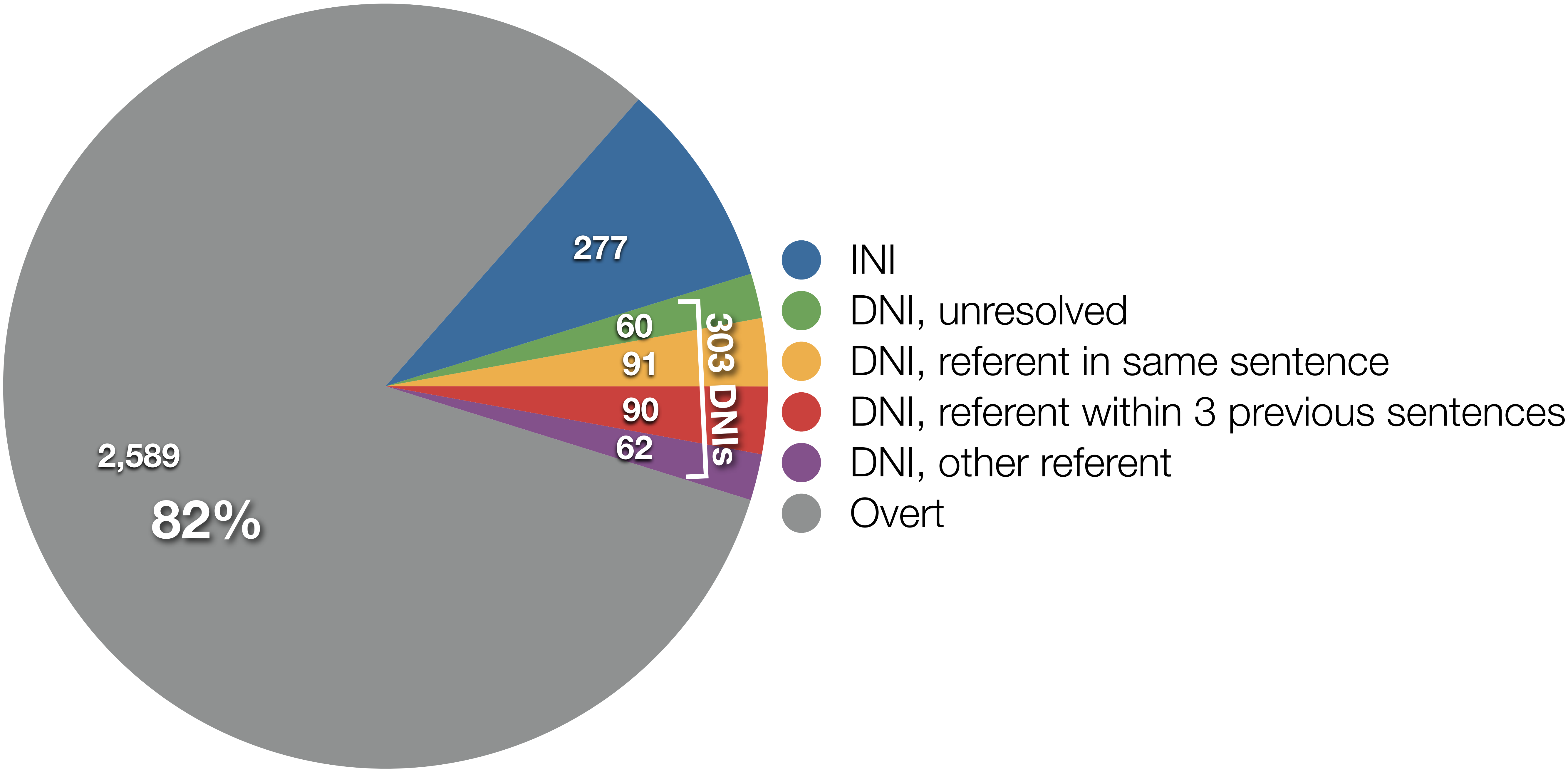
DIFFICULTY

Activity

(SemEval 2010 test data)

The other frame-evoking words are bolded, but their arguments are not shown.

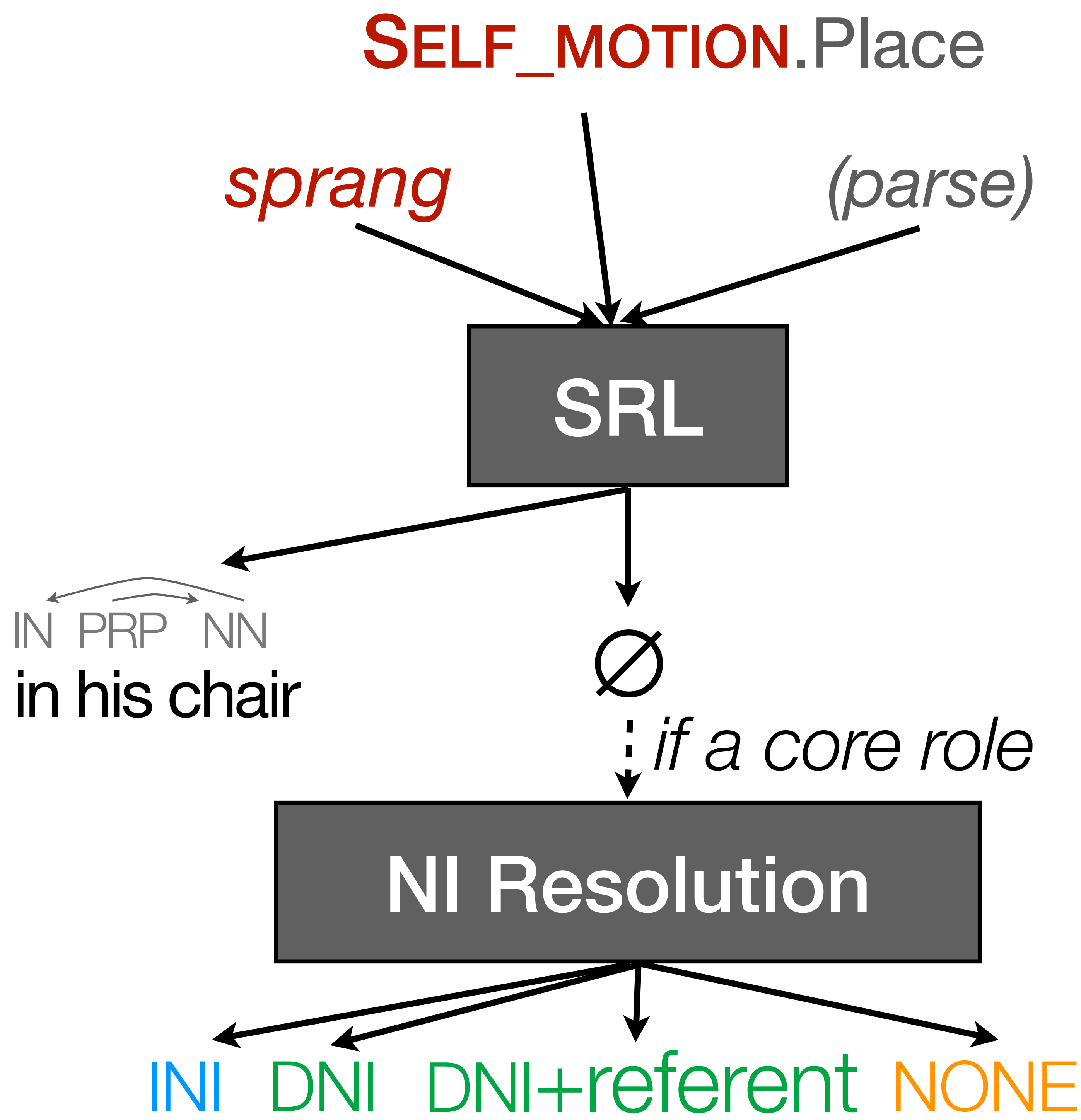
Prevalence of NIs



(SemEval 2010 new training data)

These numbers may be approximate. They show how few NIs there are compared to overt args, and why the DNI resolution task is so hard.

Modeling Approach for NIs

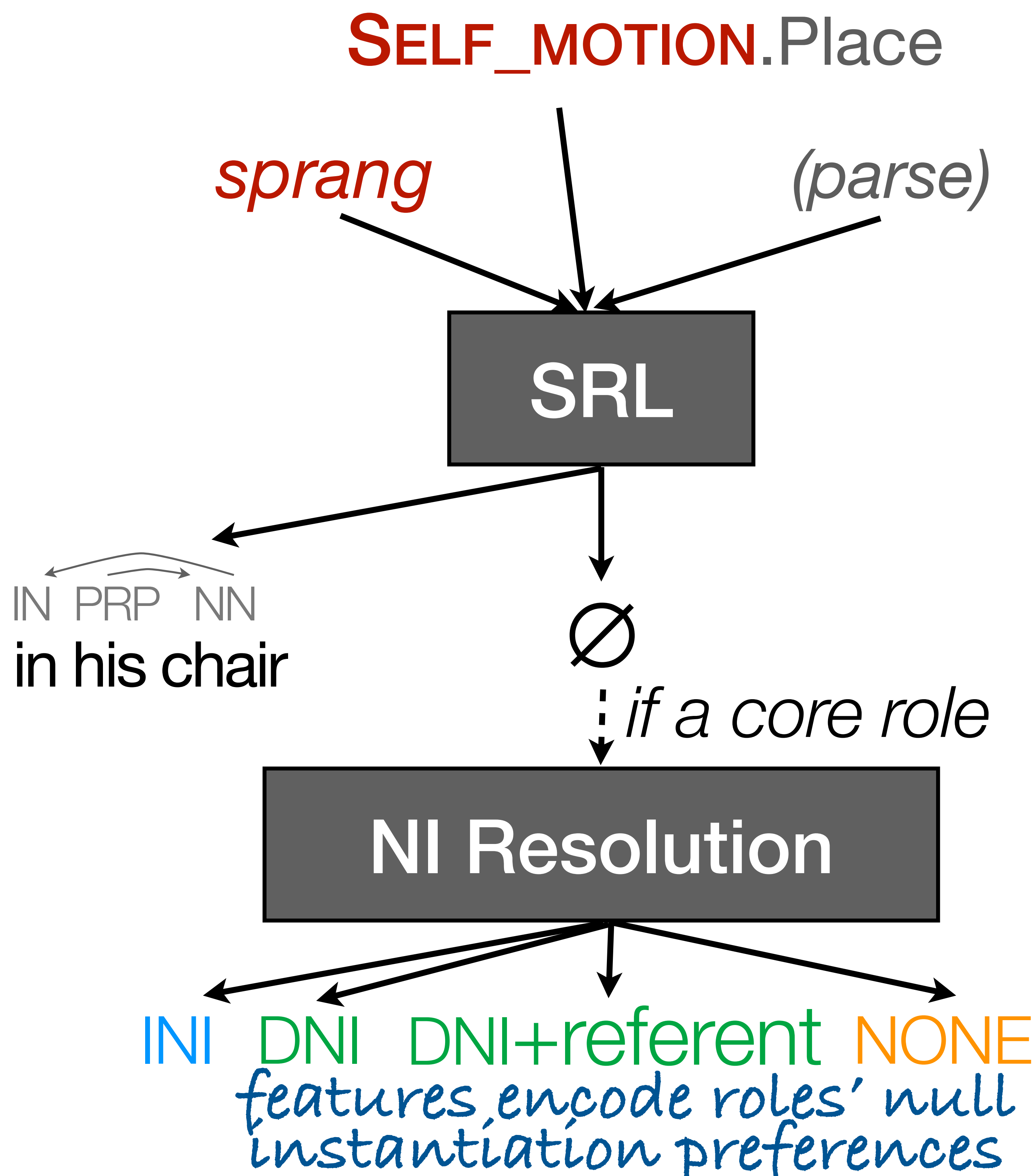


We try a straightforward approach for null instantiations: **a second classifier**

The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs.
Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNIs with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.

Modeling Approach for NIs

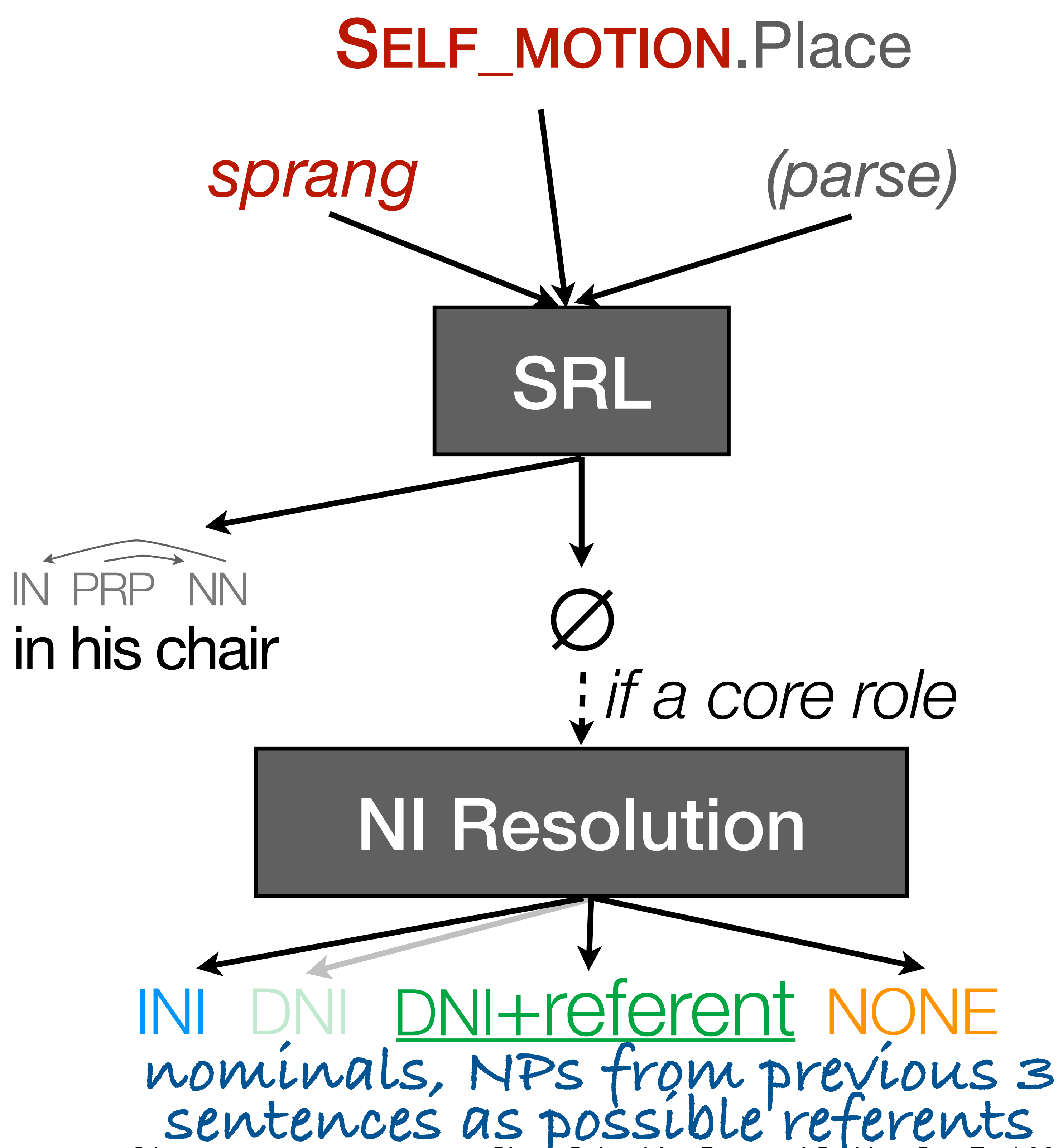
We try a straightforward approach for null instantiations: **a second classifier**



The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs.
Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNIs with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.

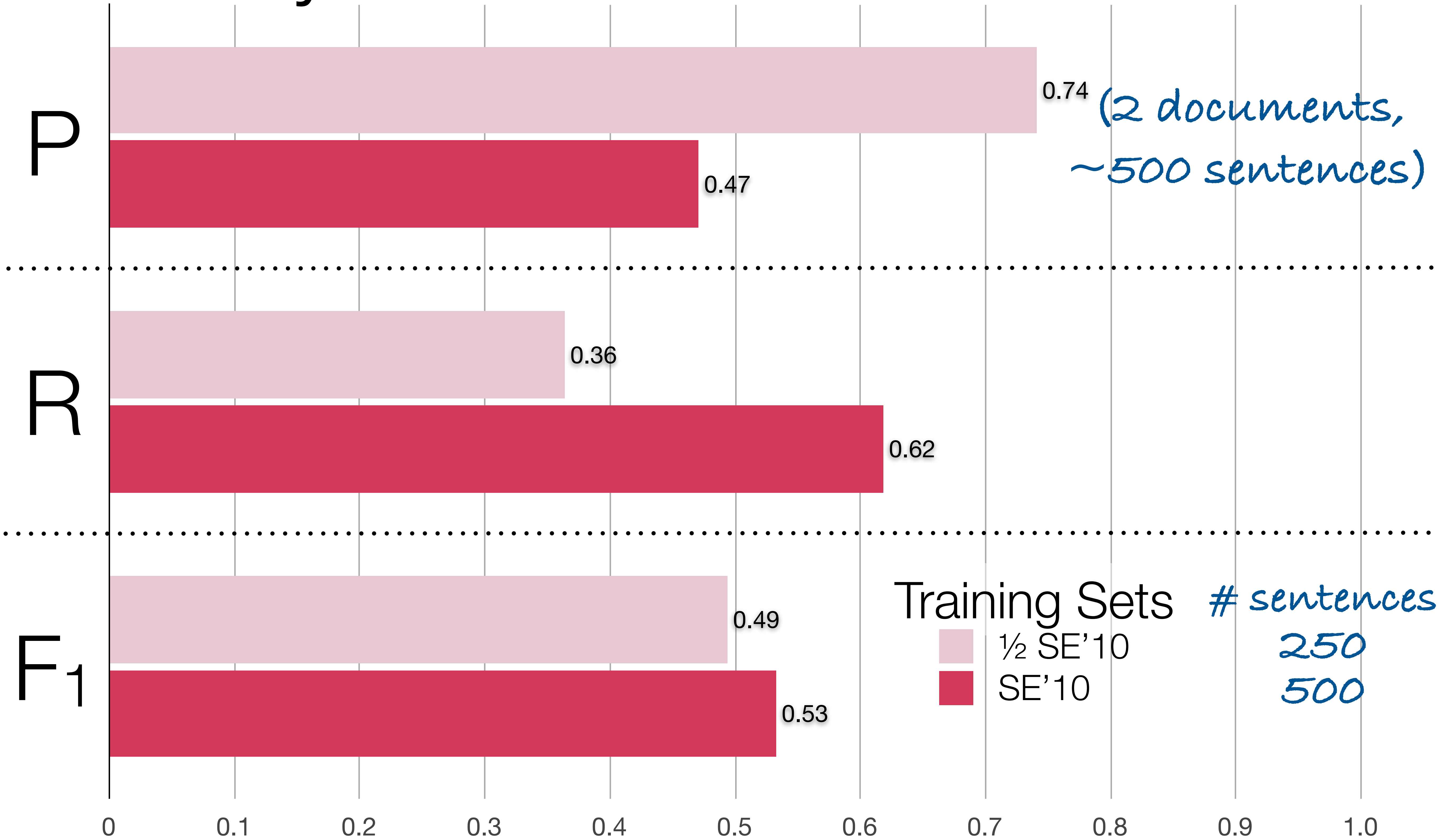
Modeling Approach for NIs

We try a straightforward approach for null instantiations: **a second classifier**



The SRL module selects an argument span or none for each role. For core roles, we then build a second classifier for disambiguating types of null elements. This uses the same mathematical techniques to predict a different kind of outputs.
Ideally, the NI module would be able to predict whether each core role was INI, DNI + its referent, if applicable, or not NI. Our system only considers DNIs with referents in the previous 3 sentences. Experiments show that a large search space, while leading to high *oracle* recall, confuses the model in practice.

NI-only results on SE'10 Test Data



Also: NI subtask confusion matrix

NIs only, oracle overt args

Evaluating NI performance only. We train only on the new SemEval 2010 data because the SemEval 2007 data used different annotation practices for null instantiations.

The evaluation criterion actually doesn't distinguish between INIs and unresolved DNIs. We predicted only the former.

Overview

✓ Background: frame SRL

✓ Overt argument identification

✓ Null instantiation resolution

➡ Conclusion

Contributions & Claims

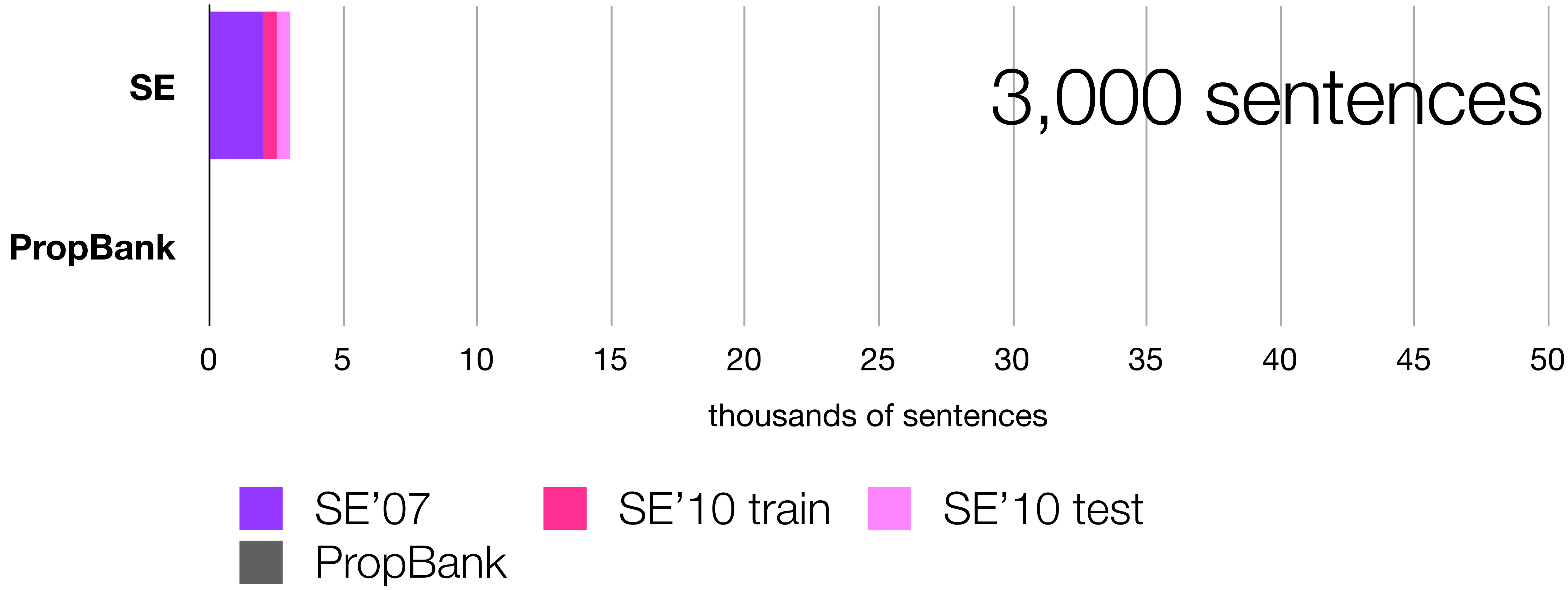
1. Evaluated frame SRL on **new data**
   - Amount of training data makes a big difference
   - Still lots of room for improvement
2. Experimented with a classifier for **null instantiations**, with mixed success
   - Resolving nonlocal referents is much harder than classifying the instantiation type
3. Learned models achieve higher **recall**, and consequently F_1, than custom heuristics used by other teams
   - Our modeling framework is **extensible**: it should allow us to incorporate many of these in a soft way as features

Size of Data

SE'07 SE'10 train SE'10 test
PropBank

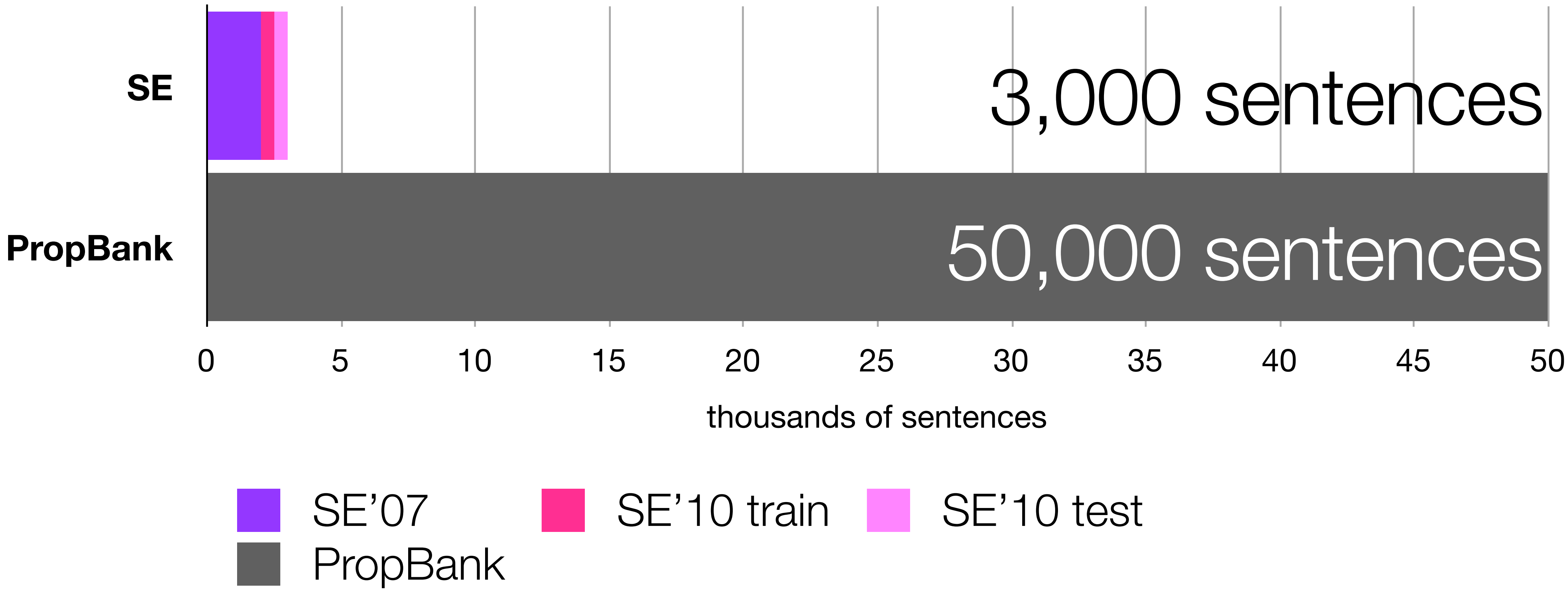
Sizes of frame-annotated data provided for SemEval '07 and '10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.

Size of Data



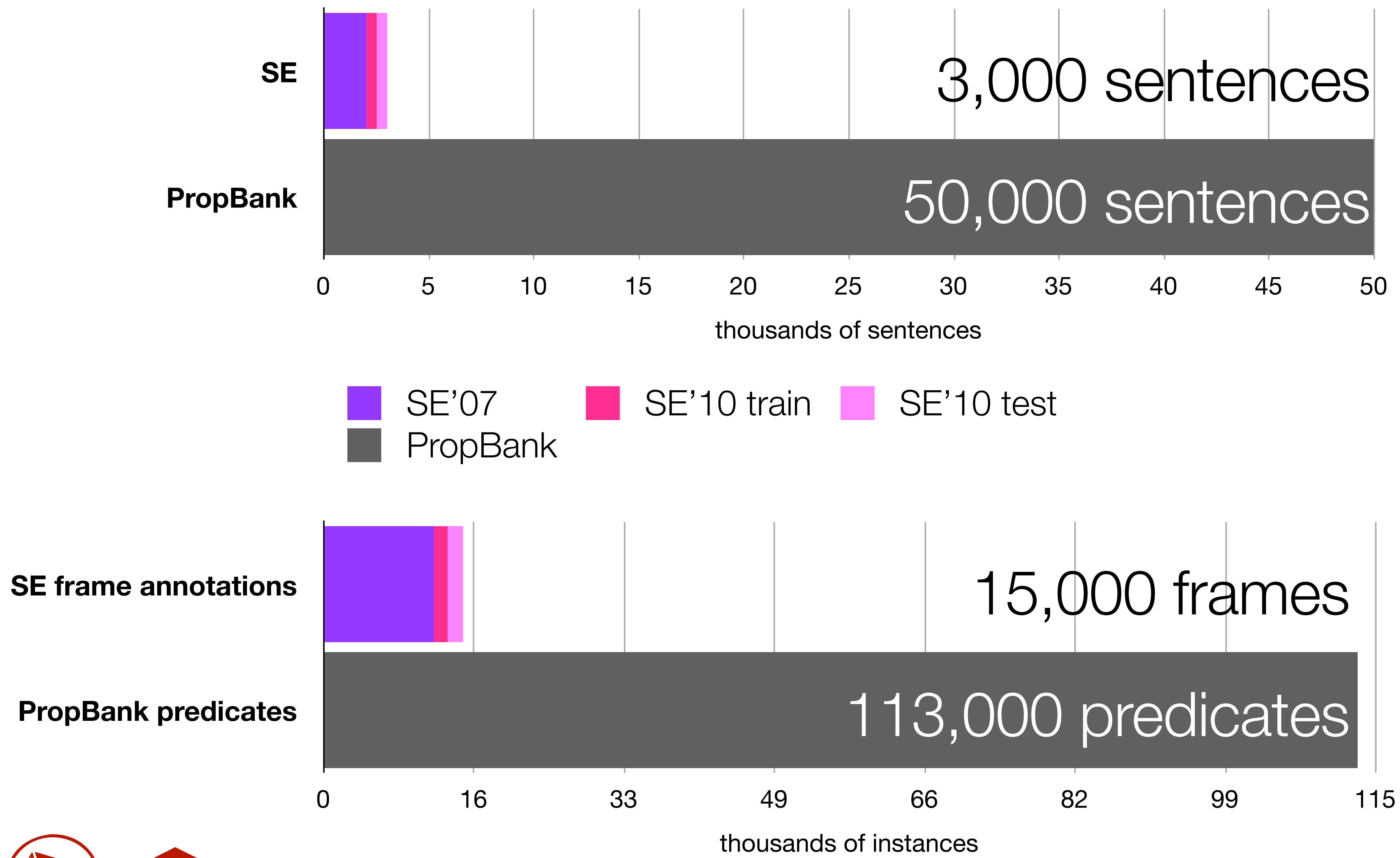
Sizes of frame-annotated data provided for SemEval '07 and '10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.

Size of Data



Sizes of frame-annotated data provided for SemEval ’07 and ’10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.

Size of Data



Sizes of frame-annotated data provided for SemEval '07 and '10 tasks, as compared to PropBank. The bottom graph is in terms of tokens. Whereas FrameNet provides a linguistically rich representation, PropBank has much higher coverage/annotated data.

Conclusion

- Next challenge: **data sparseness** in frame SRL
  - obtaining quality frame annotations from experts is expensive
  - opportunity for semi-supervised learning
  - additional knowledge/constraints in modeling
  - non-expert annotations?
  - bridging across lexical-semantic resources (FrameNet, WordNet, PropBank, VerbNet, NomBank, …)

Task 10 (Frame SRL) Posters

(101) CLR: Linking Events and Their Participants in Discourse Using a Comprehensive FrameNet Dictionary

Ken Litkowski

(102) VENSES++: Adapting a deep semantic processing system to the identification of null instantiations

Sara Tonelli & Rodolfo Delmonte

if you're interested in this task...

Thank you !

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

ARK lti Carnegie Mellon

Image from http://commons.wikimedia.org/wiki/File:SherlockHolmes.jpg

Thank you !

JUDGMENT_DIRECT_ADDRESS

Addressee

Communicator: **DNI**
Reason: **DNI**

Carnegie Mellon

Image from http://commons.wikimedia.org/wiki/File:SherlockHolmes.jpg

Extra Slides

- NI subtask confusion matrix
- NI-only and full results table

NI-only Subtask: Confusion Matrix

		Predicted					
		overt	**DNI**	**INI**	**masked**	**inc.**	**total**
Gold	**overt**	**2068 (1630)**	5	362	327	0	2762
	DNI	64	**12 (3)**	182	90	0	348
	INI	41	2	**214**	96	0	353
	masked	73	0	240	**1394**	0	1707
	inc.	12	2	55	2	**0**	71
	total	2258	21	1053	1909	0	**3688 correct**

from the paper

Results Table: NI-only and Full

	Training Data	Chapter 13 Prec.	Chapter 13 Rec.	Chapter 13 F_1	Chapter 14 Prec.	Chapter 14 Rec.	Chapter 14 F_1
NI-only	SemEval 2010 new: 100%	0.40	0.64	0.50	0.53	0.60	0.56
	SemEval 2010 new: 75%	0.66	0.37	0.50	0.70	0.37	0.48
	SemEval 2010 new: 50%	0.73	0.38	0.51	0.75	0.35	0.48
Full	All	0.35	0.55	0.43	0.56	0.49	0.52