Exploiting unlabeled and ontological data for frame-semantic parsing

Nathan Schneider

Oct. 29, 2009

Reading the Web


RESOLVE_PROBLEM
SLEEPING
CONTACTING
Medium
Communicator
Communication

3

When Gottlieb returns to his office that evening, Driftwood is sitting in his chair, and Tomasso is pouring drinks with his feet for the stowaways. When Gottlieb objects and attempts to phone the police to have them arrested , Tomasso strikes the Managing Director on the head, leaving him unconscious.

http://www.filmsite.org/night3.html

When Gottlieb returns to his office that evening, Driftwood is sitting in his chair, and Tomasso is pouring drinks with his feet for the stowaways. When Gottlieb objects and attempts [to phone the police] [to have them arrested], Tomasso strikes the Managing Director on the head, leaving him unconscious.

http://www.filmsite.org/night3.html

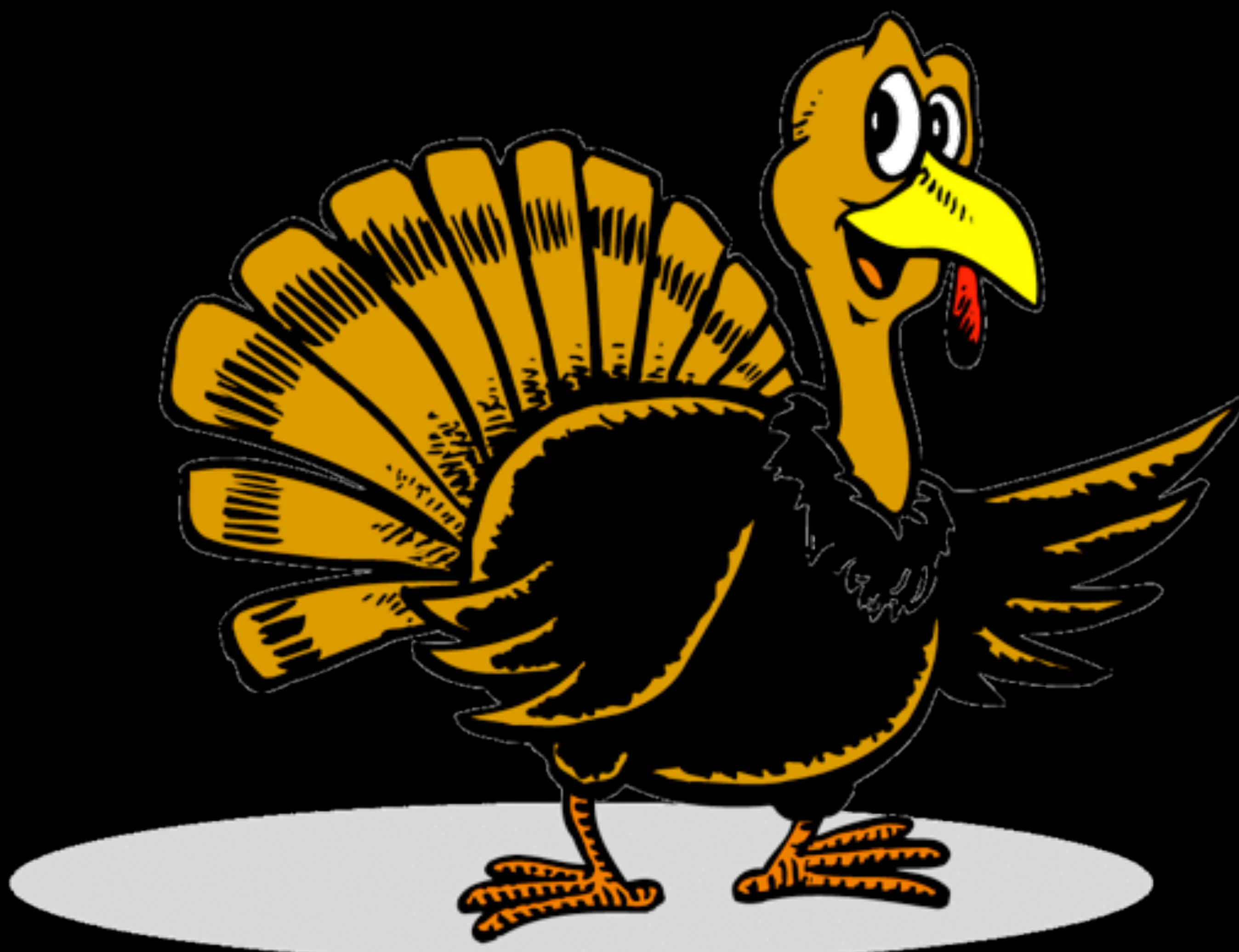
When Gottlieb returns to his office that evening, Driftwood is sitting in his chair, and Tomasso is pouring drinks with his feet for the stowaways. When Gottlieb objects and attempts [to phone the police] [to have them arrested], Tomasso strikes the Managing Director on the head, leaving him unconscious.

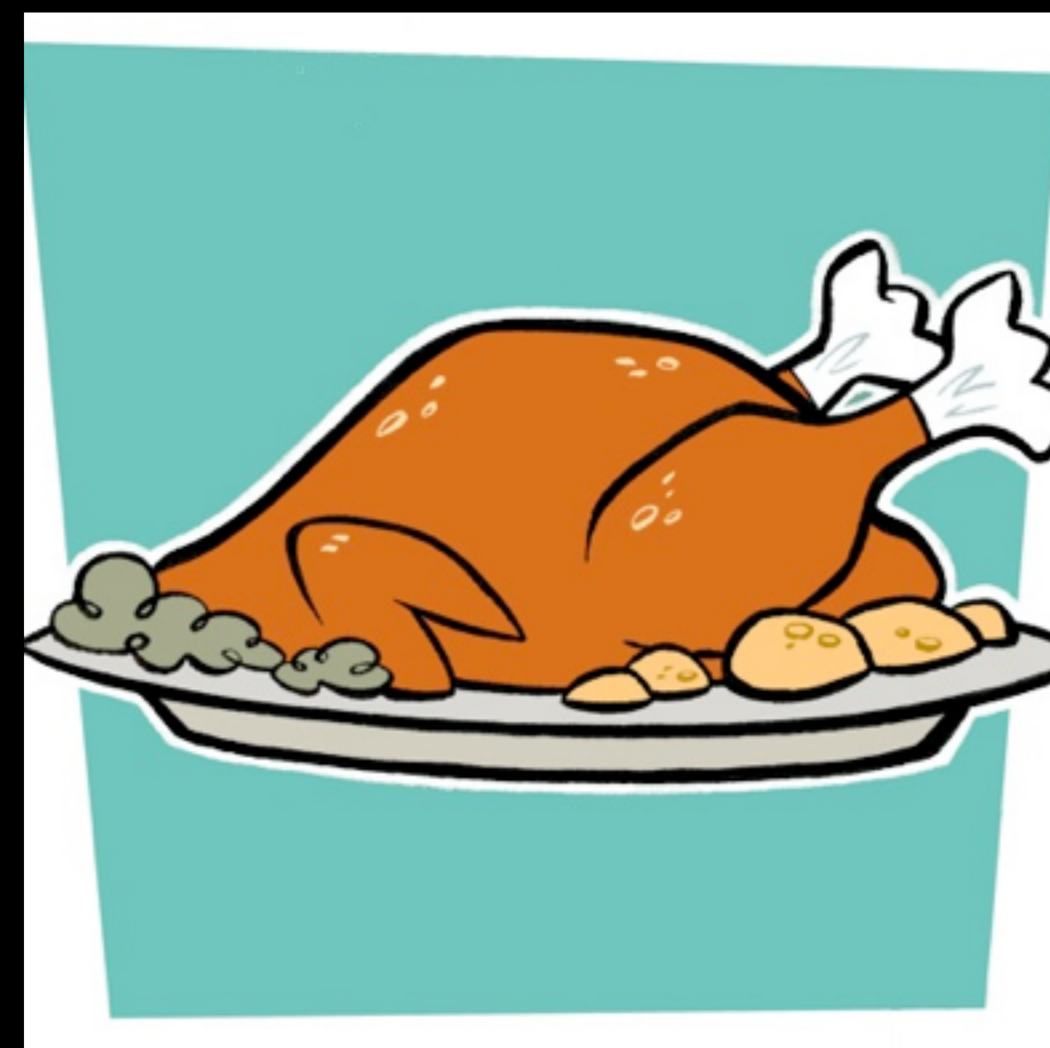
http://www.filmsite.org/night3.html

4

gobble

(verb)





5

Bella called Aunt Minnie to make a request .

Structural characteristics of semantics include:

6

Bella called Aunt Minnie to make a request .

call forth.v	Cause_to_start
call in.v	Contacting
call to mind.v	Evoking
call up.v	Contacting
call.n	Visiting
call.n	Contacting
call.v	Simple_naming
call.v	Being_named
call.v	Claim_ownership
call.v	Referring_by_name
call.v	Labeling
call.v	Contacting
call.v	Deserving
call.v	Request
called.a	Being_named

Structural characteristics of semantics include:

- Ambiguity

CONTACTING

Communicator Addressee Reason

Bella called Aunt Minnie to make a request .

Structural characteristics of semantics include:

- Ambiguity

Structural characteristics of semantics include:

- Ambiguity
- Diversity

CONTACTING

Communicator — Addressee — Reason

Bella called Aunt Minnie to make a request.

Speaker — Addressee — REQUEST

Structural characteristics of semantics include:

- Ambiguity
- Diversity
- Sharing

CONTACTING

Communicator | Addressee | Reason

Bella called Aunt Minnie to make a request.

Speaker | Addressee | REQUEST | Message

Structural characteristics of semantics include:

- Ambiguity
- Diversity
- Sharing
- Omission

CONTACTING

Communicator

Addressee

Communication

Reason

Time

Place

REQUEST

Speaker

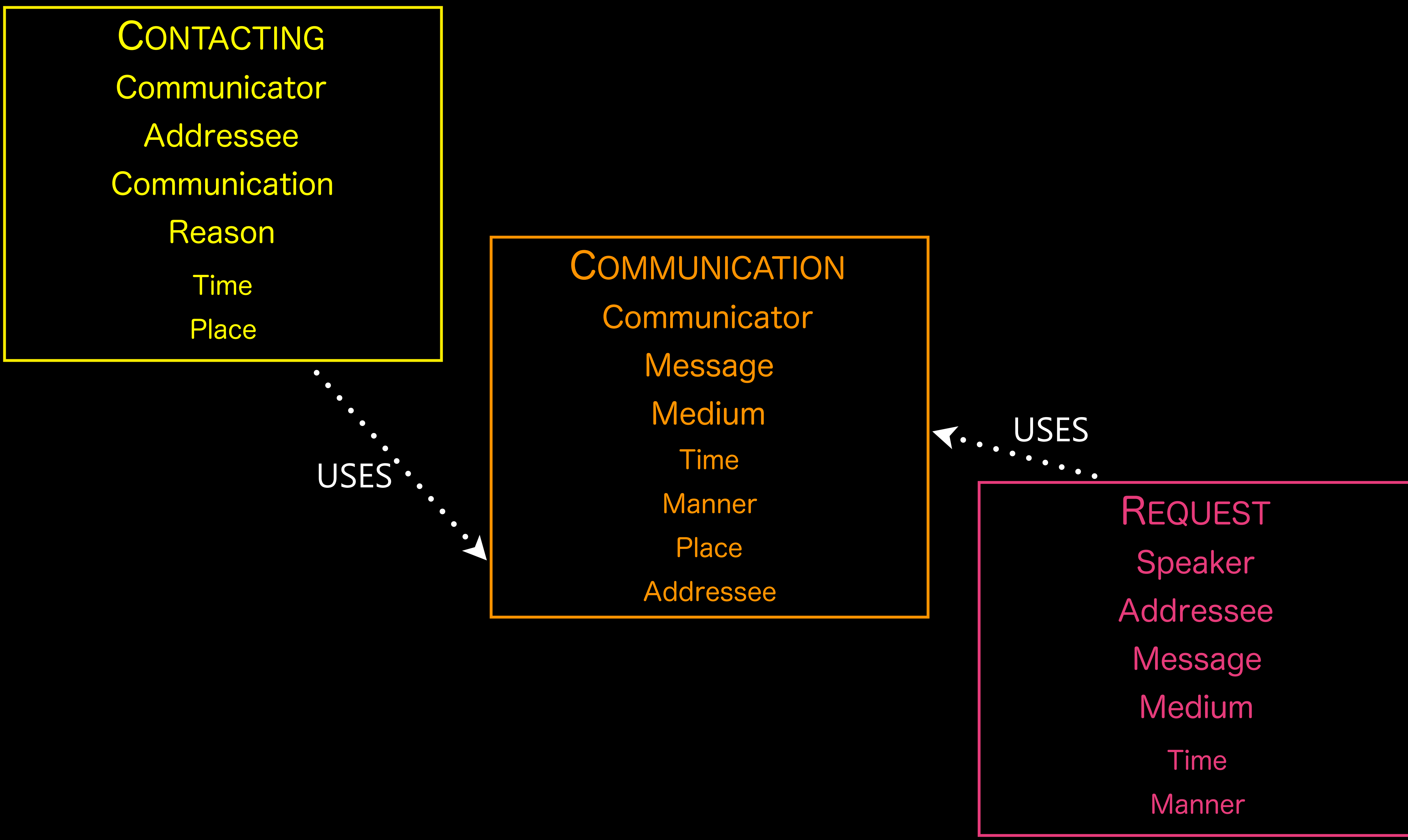
Addressee

Message

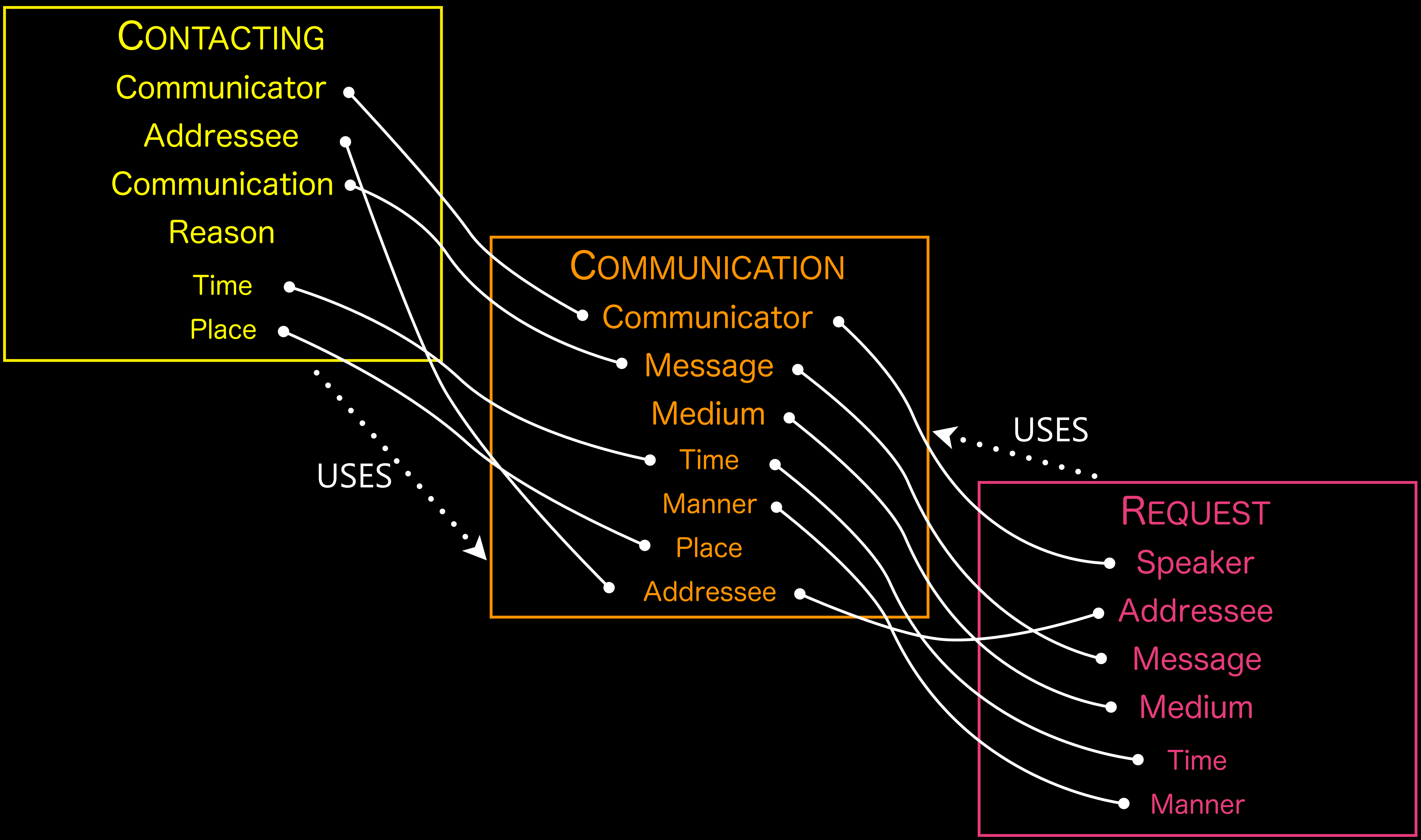
Medium

Time

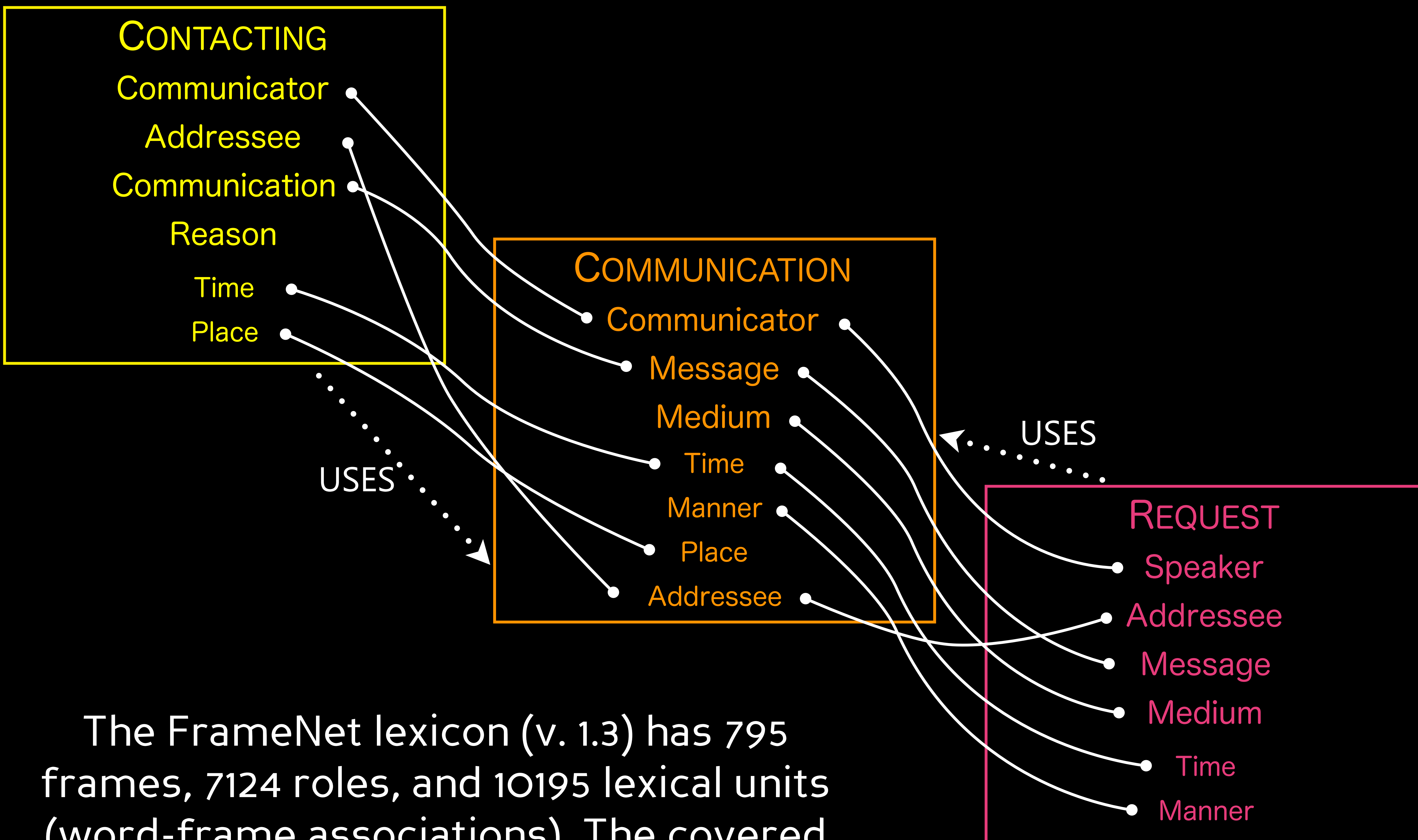
Manner


CONTACTING
Communicator
Addressee
Communication
Reason
Time
Place
USES
COMMUNICATION
Communicator
Message
Medium
Time
Manner
Place
Addressee
USES
REQUEST
Speaker
Addressee
Message
Medium
Time
Manner

7


CONTACTING
Communicator
Addressee
Communication
Reason
Time
Place
USES
COMMUNICATION
Communicator
Message
Medium
Time
Manner
Place
Addressee
USES
REQUEST
Speaker
Addressee
Message
Medium
Time
Manner

7



The FrameNet lexicon (v. 1.3) has 795 frames, 7124 roles, and 10195 lexical units (word-frame associations). The covered frames tend to be frequent, structurally complex, and domain-general.

Gildea & Jurafsky 2002

- Introduced semantic role labeling (SRL) as a task
  - Assume the frame, target word are given
- Train a supervised probability model for SRL with FrameNet
  - Requires careful smoothing
- Argument phrases are selected from among the constituents of the sentence parse

Gildea & Jurafsky: Arg Classification Features

- Linear position of argument with respect to target
- Syntactic features from (constituency) parse
  - Parse tree path from the target to the argument
- Voice: *the board changed the ruling*
  vs. *the ruling was changed (by the board)*
- Lexical features, e.g. head word of arg. phrase

9

Frame-semantic Data as of 2007

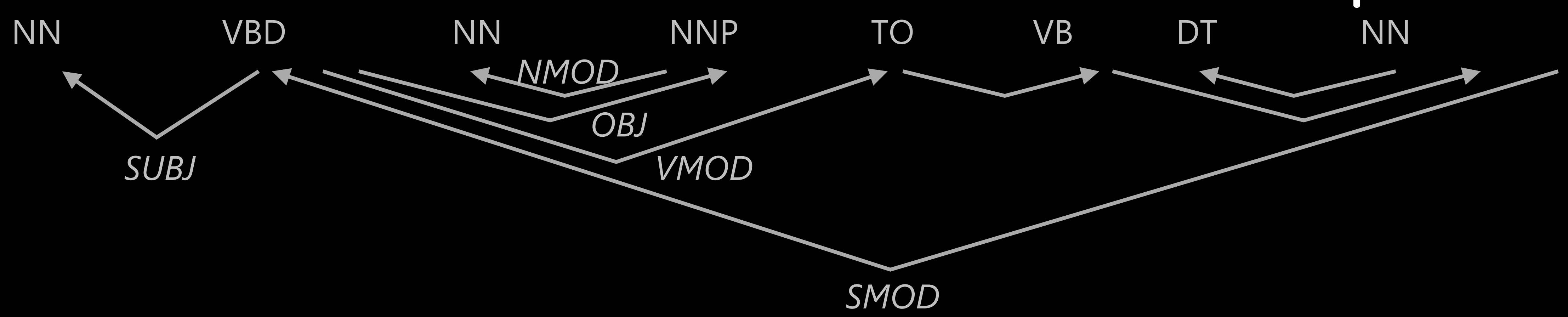
- The FrameNet lexicon gives the inventory of frames/roles/frame relations, as well as some sparsely-annotated exemplar sentences
- For training/test: a **small** corpus (29 documents, ~50,000 words) of articles which are fully annotated (albeit somewhat noisily) for their FrameNet frames. This corpus comprised the data set for a SemEval 2007 task on predicting frame-semantic structure.

10

Johansson & Nugues 2007

- Full frame-semantic parsing: identifying all predicates (targets), their frames, and their arguments
  - Pipeline of SVMs
  - Explores the use of syntactic dependency parses for features
  - Winning system of the SemEval 2007 task

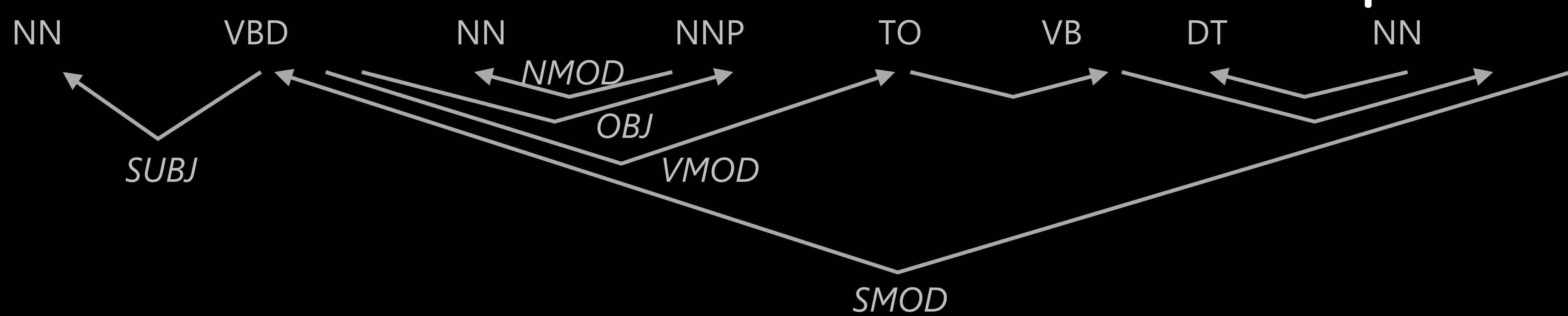
Input Bella called Aunt Minnie to make a request .



12

Targets Bella called Aunt Minnie to make a request .

Input Bella called Aunt Minnie to make a request .



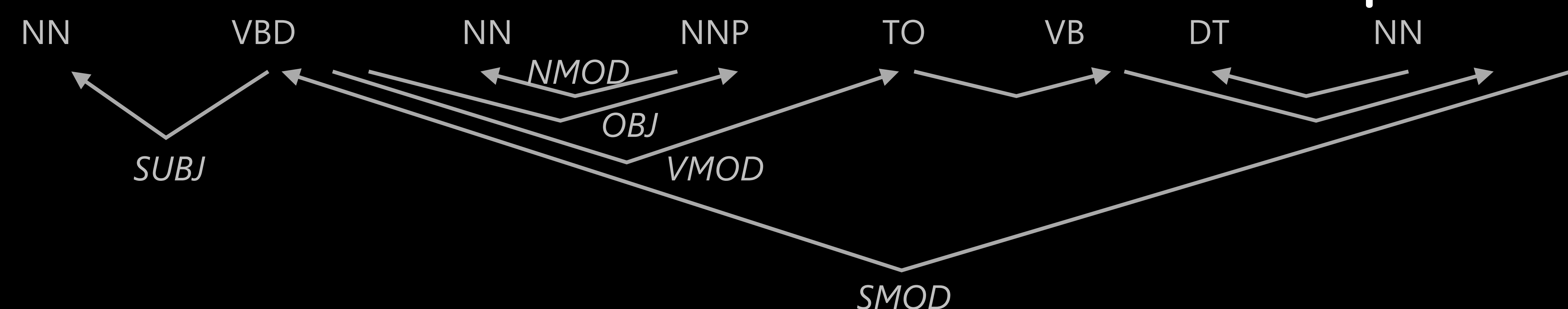
CONTACTING

Frames Bella called Aunt Minnie to make a request .

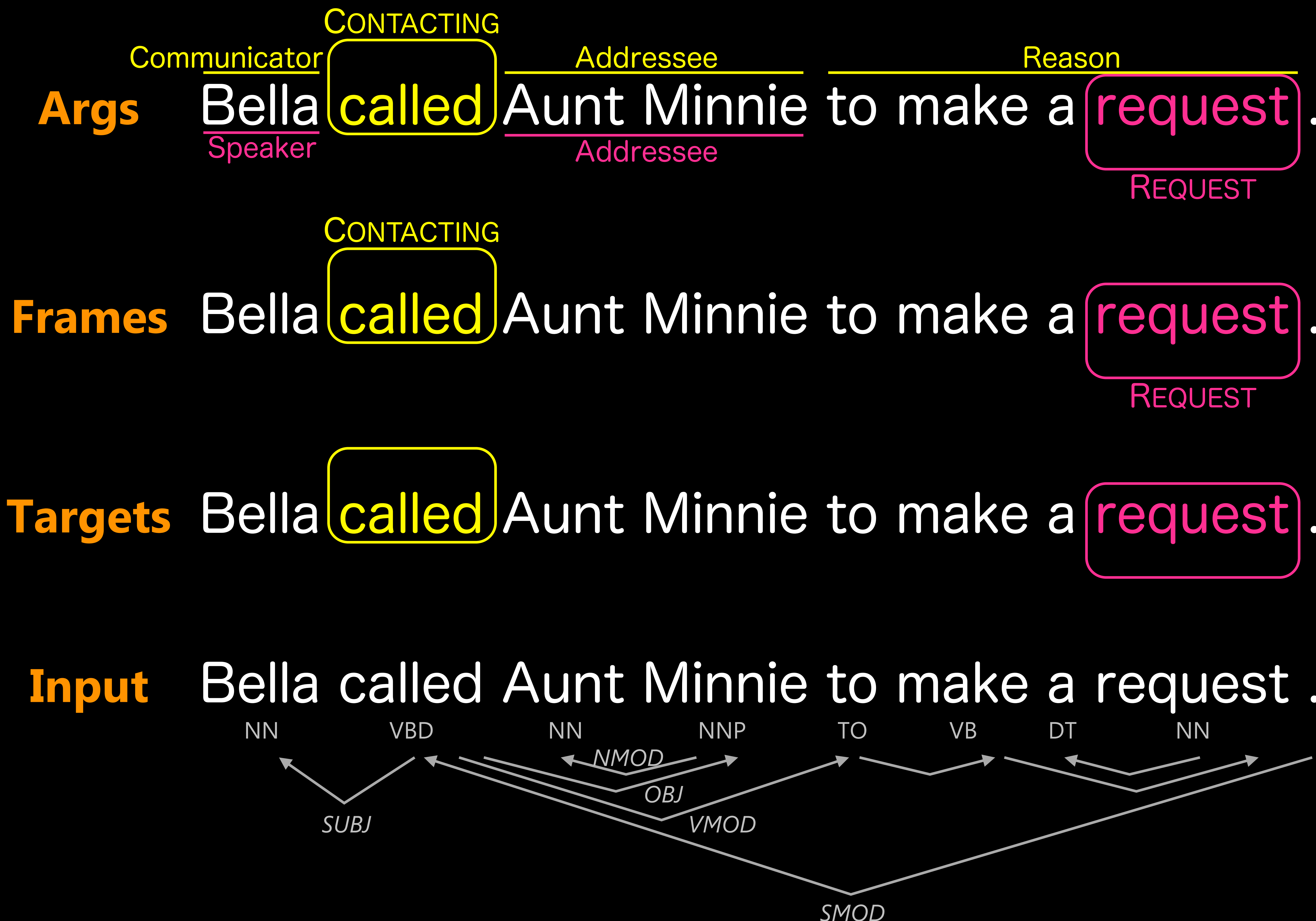
REQUEST

Targets Bella called Aunt Minnie to make a request .

Input Bella called Aunt Minnie to make a request .



12


CONTACTING
Communicator
Addressee
Reason
Args
Bella called Aunt Minnie to make a request .
Speaker
Addressee
REQUEST
CONTACTING
Frames
Bella called Aunt Minnie to make a request .
REQUEST
Targets
Bella called Aunt Minnie to make a request .
Input
Bella called Aunt Minnie to make a request .
NN VBD NN NNP TO VB DT NN
NMOD
OBJ
SUBJ
VMOD
SMOD

Results from Johansson & Nugues

- Partial credit for related frames
- Exact labeling of target/arg spans
- No use of NER features

Frames

Table 1: Results for frame detection.

Setting		Recall	Precision	$F1$
E	L	0.528	0.688	0.597
P	L	0.581	0.758	0.657
E	D	0.549	0.715	0.621
P	D	0.601	0.784	0.681

13

Results from Johansson & Nugues

Frames + Args

Table 3: Results for frame and FE detection.

Setting			Recall	Precision	$F1$
E	L	Y	0.372	0.532	0.438
P	L	Y	0.398	0.570	0.468
E	D	Y	0.389	0.557	0.458
P	D	Y	0.414	0.594	0.488
E	L	N	0.364	0.530	0.432
P	L	N	0.391	0.570	0.464
E	D	N	0.384	0.561	0.456
P	D	N	0.411	0.600	0.488

Frames

Table 1: Results for frame detection.

Setting		Recall	Precision	$F1$
E	L	0.528	0.688	0.597
P	L	0.581	0.758	0.657
E	D	0.549	0.715	0.621
P	D	0.601	0.784	0.681

- Partial credit for related frames
- Exact labeling of target/arg spans
- No use of NER features

13

Joint work with Dipanjan Das, Desai Chen, Noah Smith

Our Approach

Targets rule-based segmenter

14

Joint work with Dipanjan Das, Desai Chen, Noah Smith

Our Approach

Frames

for each frame target t_i, choose a frame label f independently of other targets

$$p_{\boldsymbol{\theta}}(f, \ell \mid t_i, \mathbf{x}) = \frac{\exp \boldsymbol{\theta}^\top \mathbf{g}(f, \ell, t_i, \mathbf{x})}{\sum_{f' \in \mathcal{F}_{\mathbf{x}_{t_i}}} \sum_{\ell' \in \mathcal{L}_{f'}} \exp \boldsymbol{\theta}^\top \mathbf{g}(f', \ell', t_i, \mathbf{x})}$$

Targets

rule-based segmenter

Joint work with Dipanjan Das, Desai Chen, Noah Smith

Our Approach

Args

for each role r_j of f_i, choose an argument filler span s independently of other roles

$$p_{\boldsymbol{\psi}}(s \mid \mathbf{x}, t_i, f_i) = \frac{\exp \boldsymbol{\psi}^\top \mathbf{h}(\mathbf{x}, t_i, f_i, r_j, s)}{\sum_{s' \in \mathcal{S}} \exp \boldsymbol{\psi}^\top \mathbf{h}(\mathbf{x}, t_i, f_i, r_j, s')}$$

Frames

for each frame target t_i, choose a frame label f independently of other targets

$$p_{\boldsymbol{\theta}}(f, \ell \mid t_i, \mathbf{x}) = \frac{\exp \boldsymbol{\theta}^\top \mathbf{g}(f, \ell, t_i, \mathbf{x})}{\sum_{f' \in \mathcal{F}_{\mathbf{x}_{t_i}}} \sum_{\ell' \in \mathcal{L}_{f'}} \exp \boldsymbol{\theta}^\top \mathbf{g}(f', \ell', t_i, \mathbf{x})}$$

Targets

rule-based segmenter

Joint work with Dipanjan Das, Desai Chen, Noah Smith

Our Approach: Differences from Johansson & Nugues

- We use log-linear models in order to formulate a full probability model in a discriminative setting
  - Latent variable provides smoothing for unseen targets
  - Enables us to consider joint inference techniques to break independence assumptions, e.g. between arguments of a frame

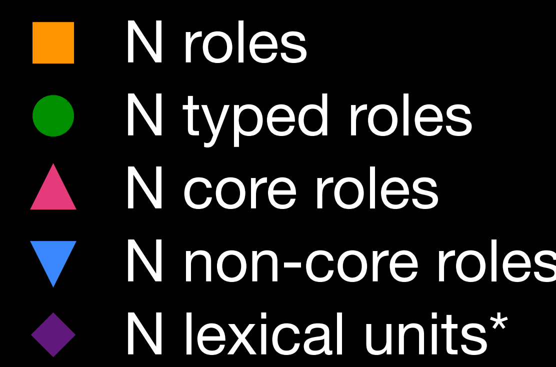
Joint work with Dipanjan Das, Desai Chen, Noah Smith

Our Approach: Differences from Johansson & Nugues

- Our argument identification model is a single role-filling model rather than a sequence of argument-finding + -classification models
  - Beam search at the end to ensure there are no overlapping arguments
- State-of-the-art results (numbers still preliminary, but we win on all stages). ≈50 F_1 means there's room for improvement!

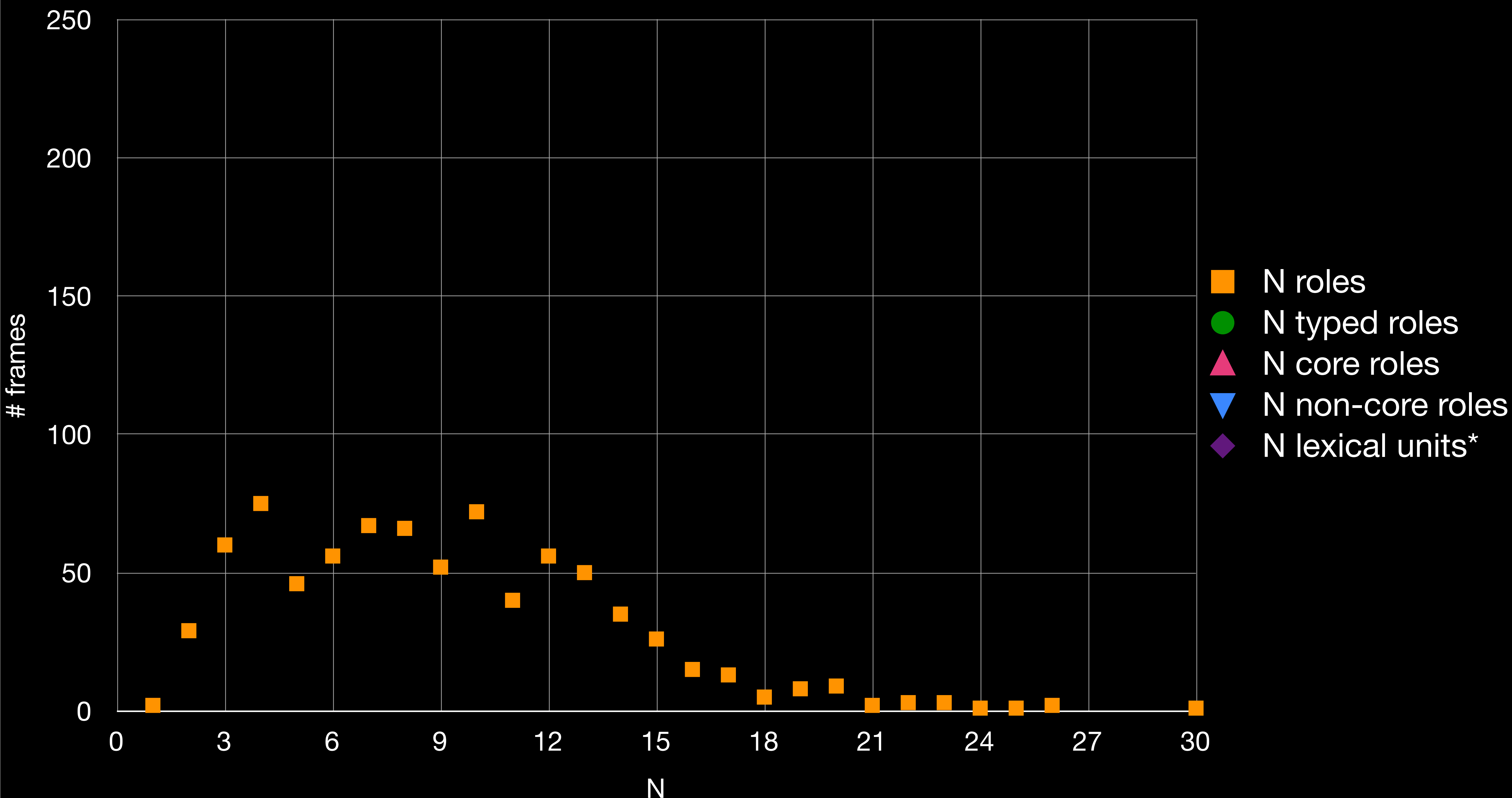
My Questions

- Can the argument identification subtask be improved by exploiting
  - features based on selectional restrictions (semantic type annotations on roles)?
  - sparsely annotated exemplars from the lexicon?
  - learned RTW instances/patterns via a mapping from RTW ontology types to FrameNet frames, roles, or semantic types?



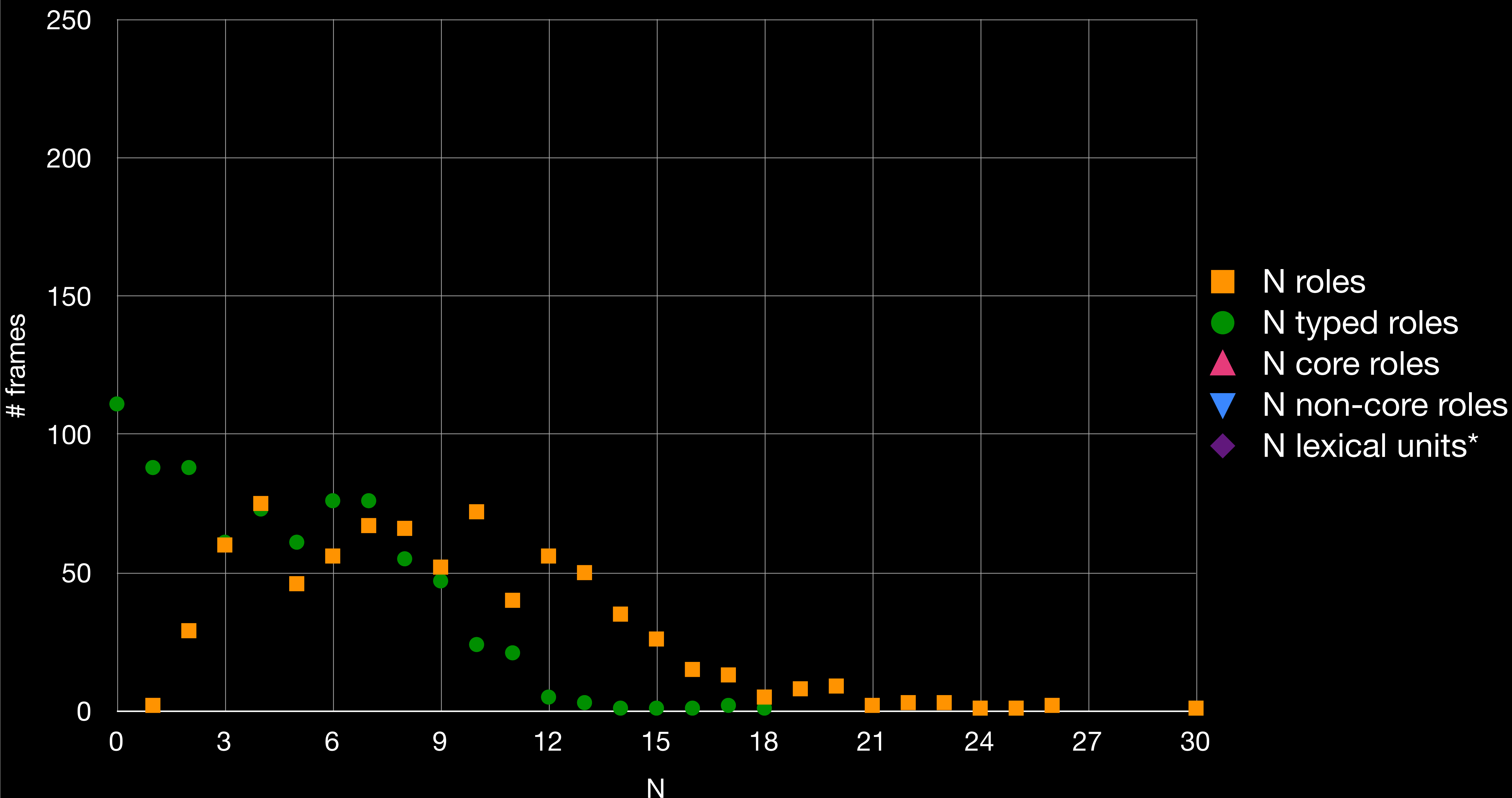
Histogram of role counts by frame in the lexicon. For instance, the dot at (1, 88) means that 88 frames have exactly 1 typed role. Weighting all frames equally, the average frame has 9.0 roles, 4.5 typed roles, 3.0 core roles, and 12.8 lexical units.

* Not depicted here are 75 frames (9.4%) which have over 30 lexical units. The EMOTION_DIRECTED frame has the largest number of lexical units (179).



Histogram of role counts by frame in the lexicon. For instance, the dot at (1, 88) means that 88 frames have exactly 1 typed role. Weighting all frames equally, the average frame has 9.0 roles, 4.5 typed roles, 3.0 core roles, and 12.8 lexical units.

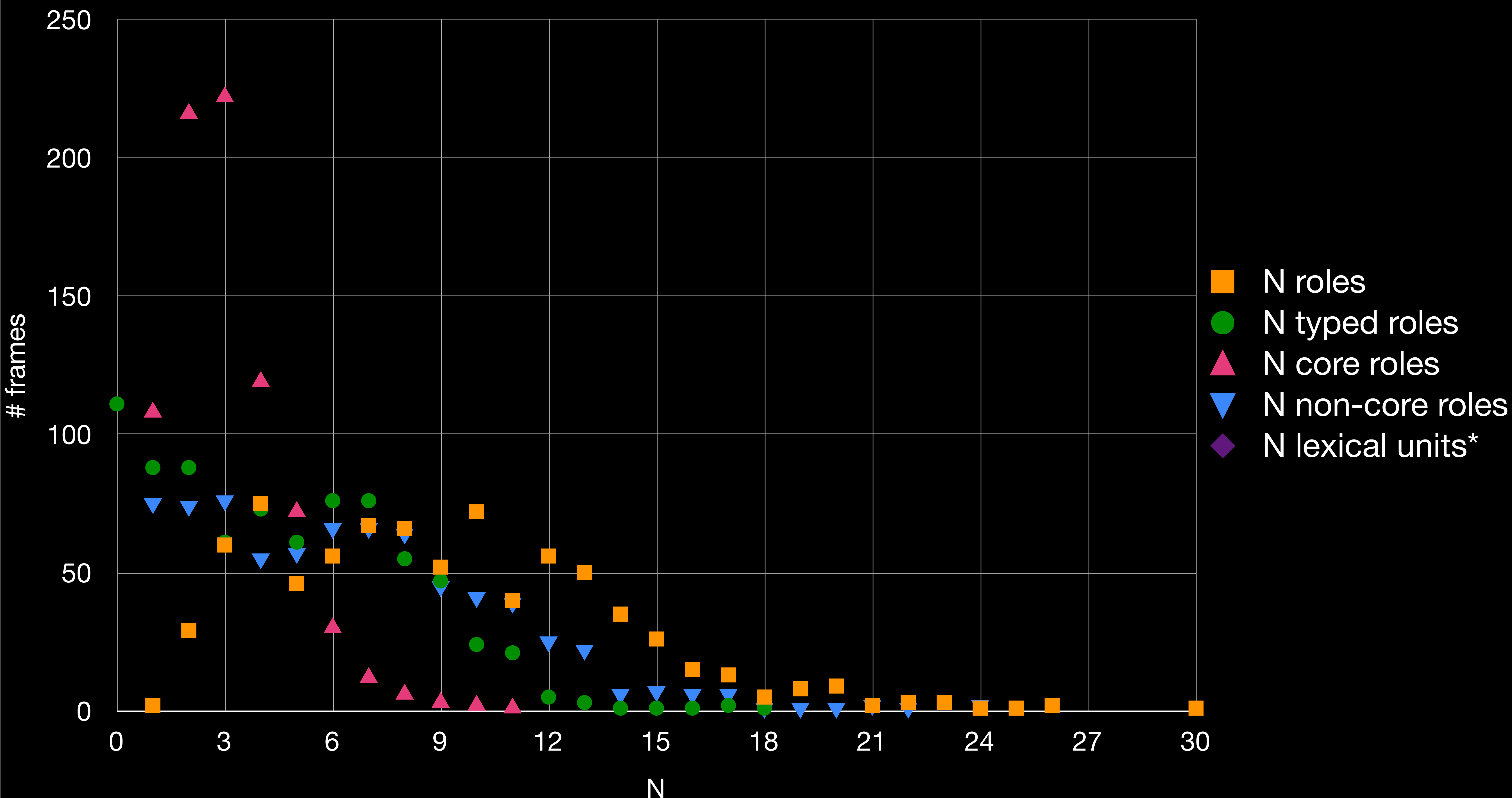
* Not depicted here are 75 frames (9.4%) which have over 30 lexical units. The EMOTION_DIRECTED frame has the largest number of lexical units (179).



Histogram of role counts by frame in the lexicon. For instance, the dot at (1, 88) means that 88 frames have exactly 1 typed role. Weighting all frames equally, the average frame has 9.0 roles, 4.5 typed roles, 3.0 core roles, and 12.8 lexical units.

* Not depicted here are 75 frames (9.4%) which have over 30 lexical units. The EMOTION_DIRECTED frame has the largest number of lexical units (179).

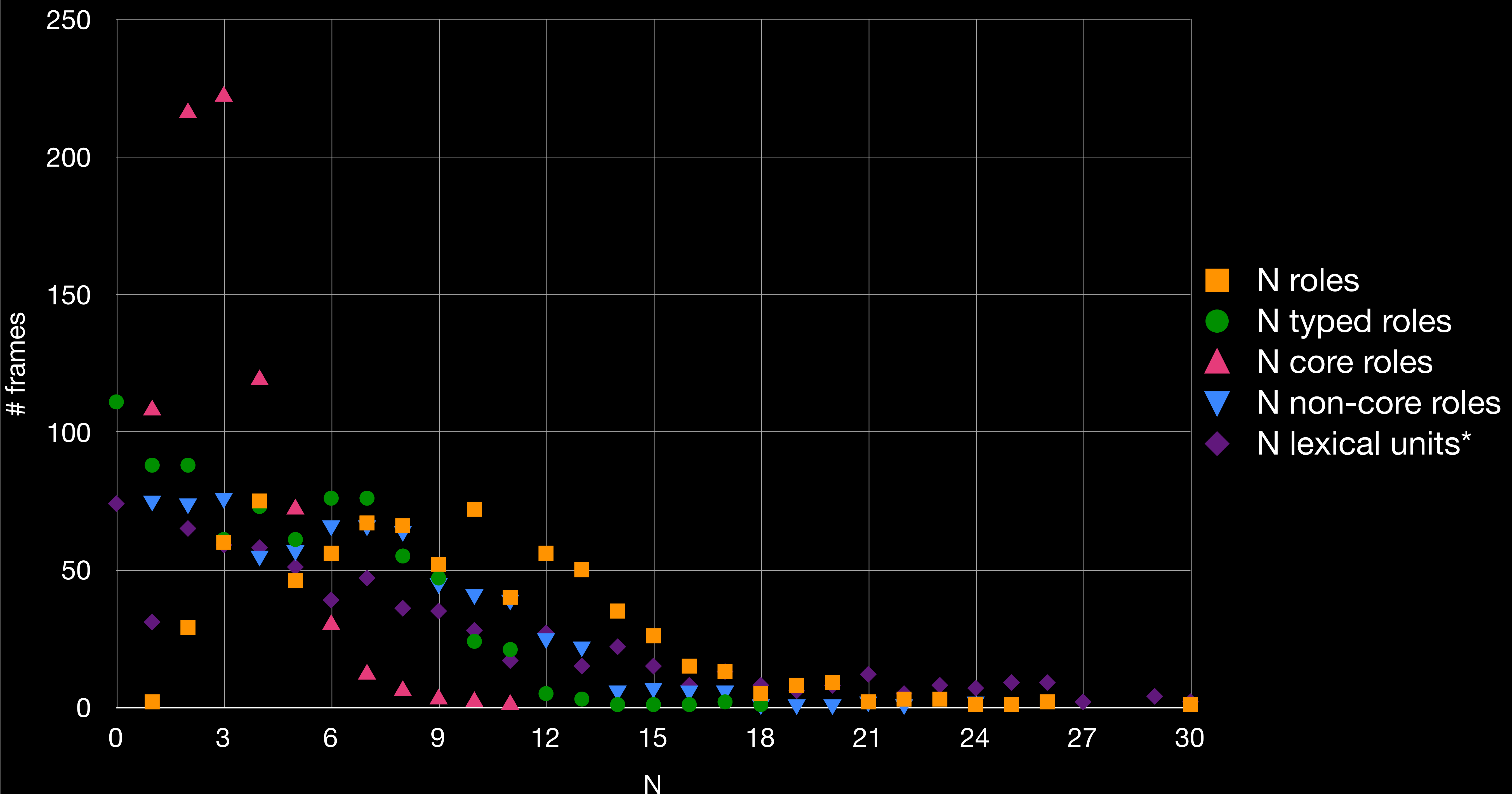
18



Histogram of role counts by frame in the lexicon. For instance, the dot at (1, 88) means that 88 frames have exactly 1 typed role. Weighting all frames equally, the average frame has 9.0 roles, 4.5 typed roles, 3.0 core roles, and 12.8 lexical units.

* Not depicted here are 75 frames (9.4%) which have over 30 lexical units. The EMOTION_DIRECTED frame has the largest number of lexical units (179).

18



Histogram of role counts by frame in the lexicon. For instance, the dot at (1, 88) means that 88 frames have exactly 1 typed role. Weighting all frames equally, the average frame has 9.0 roles, 4.5 typed roles, 3.0 core roles, and 12.8 lexical units.

* Not depicted here are 75 frames (9.4%) which have over 30 lexical units. The EMOTION_DIRECTED frame has the largest number of lexical units (179).

CONTACTING

Communicator	Sentient
Communication	
Reason	State_of_affairs
Addressee	Sentient
Place	Locative_relation
Time	Time

COMMUNICATION

Communicator	Sentient
Message	Message
Medium	
Time	Time
Manner	Manner
Place	Locative_relation
Addressee	Sentient

REQUEST

Speaker	Sentient
Message	Message
Medium	
Addressee	Sentient
Manner	Manner
Time	Time

Semantic types are specified for some roles of some frames in the lexicon.

19

Sentient	1912	Manner	198
Artifact	871	Locative_relation	192
State_of_affairs	693	Degree	187
Location	638	Quantity	171
Time	540	Content	171
Physical_object	423	Human	156
Physical_entity	408	Goal	147
Message	292	Source	80

- Above: semantic types most likely to be associated with roles filled by arguments in the SemEval 2007 training data (and their counts)
  - These 16 types capture 42% of arguments (to roles defined in the lexicon)!
  - If these few can be mapped to types in another ontology covering a lot of data, it is likely to help

Possible features leveraging unstructured text

In unsupervised data	In test sentence being FN parsed	Other possible constraints
CooccurInSentence (w_1,w_2) is large	w_1, w_2 are involved in the same frame instance (as arg* or target) w_1 heads an argument to the frame evoked by w_2 w_1 and w_2 are arguments to the same frame instance*	w_1, w_2 are linearly ordered or syntactically linked the same way in unsupervised and test sentence
w_1 and w_2 often occur in the same (word or syntactic) contexts	w_2 fills role r which is often filled by w_1 in the training data	

w_1, w_2 might refer to words heading disjoint NPs.

* using for features would require joint decision about a frame's arguments

How to factor out topical coocurrence? What about e.g. "president" and "politics" cooccurring? Topic model or domain classification?

21

Training with the Lexicon

- Due to biases in the choice of exemplar sentences, including these in training data hurts if evaluated for full-text frame parsing
  - Almost 2 orders of magnitude more exemplars than SemEval training sentences
  - Exemplars were chosen because they were lexicographically interesting; not IID
  - Is there a way to exclude or downweight certain data points w.r.t. specific features?

139K lexicon exemplar sentences (**3,100,00** words)

1.7K SemEval training sentences (**43,300** words)