Squashing

Computational Linguistics

Noah A. Smith

Paul G. Allen School of Computer Science & Engineering

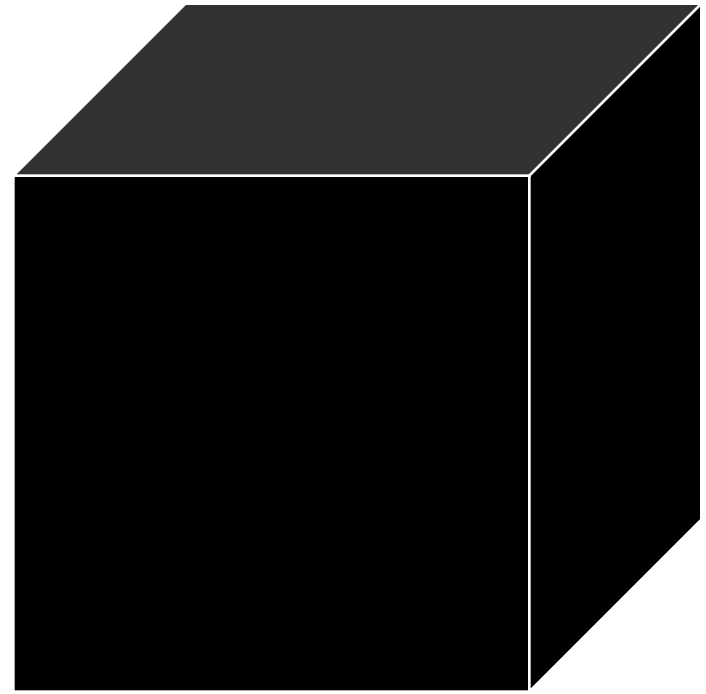
University of Washington

Seattle, USA

@nlpnoah

Research supported in part by: NSF, DARPA DEFT, DARPA CWC, Facebook, Google, Samsung, University of Washington.

data

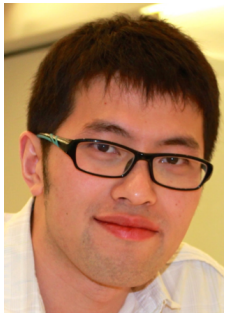


Applications of NLP in 2017

- Conversation, IE, MT, QA, summarization, text categorization

Applications of NLP in 2017

- Conversation, IE, MT, QA, summarization, text categorization
- Machine-in-the-loop tools for (human) **authors**



Chenhao Tan

Revise your message with help from NLP

tremoloop.com

Elizabeth Clark

Collaborate with an NLP model through an "exquisite corpse" storytelling game

Applications of NLP in 2017

- Conversation, IE, MT, QA, summarization, text categorization
- Machine-in-the-loop tools for (human) **authors**
- **Analysis** tools for measuring social phenomena

Lucy Lin

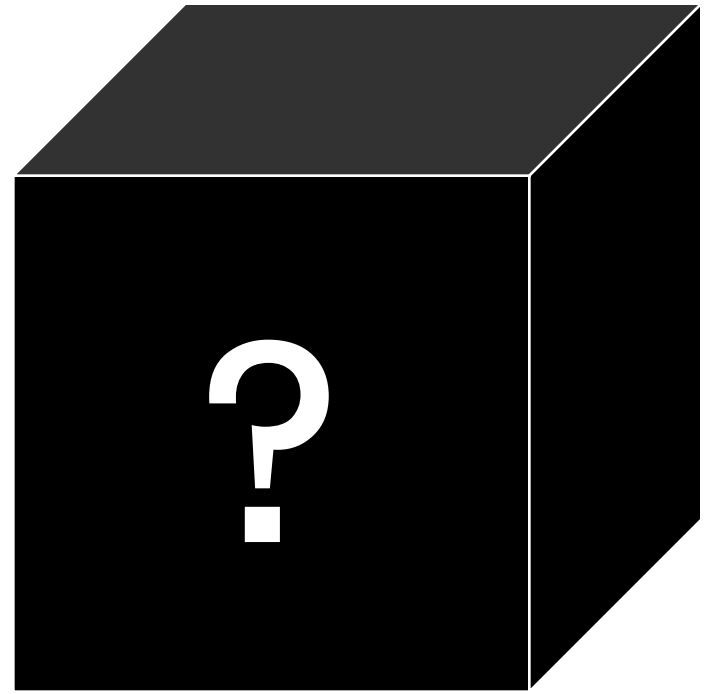
Sensationalism in science news

bit.ly/sensational-news
… bookmark this survey!

Dallas Card

Track ideas, propositions, frames in discourse over time

data



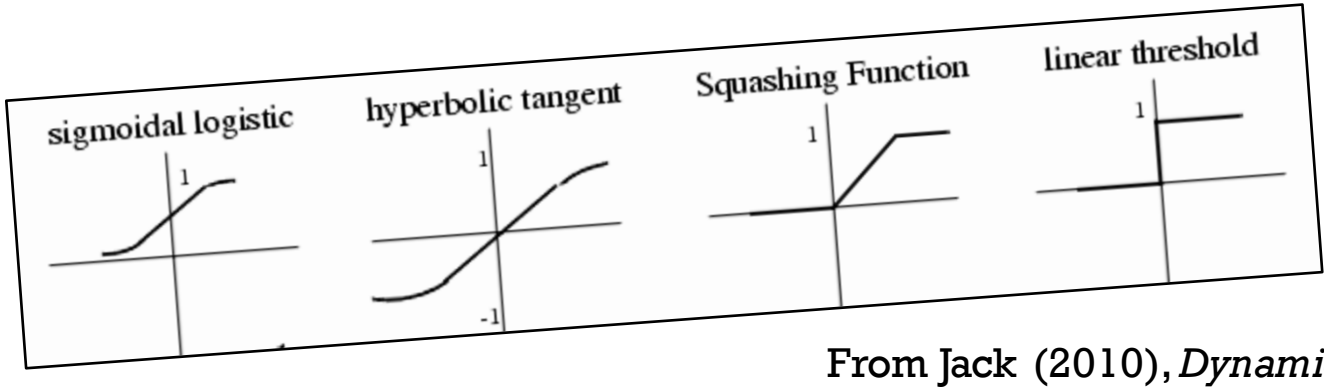
Squash

Squash Networks

- Parameterized differentiable functions composed out of simpler parameterized differentiable functions, some nonlinear

Squash Networks

- Parameterized differentiable functions composed out of simpler parameterized differentiable functions, some nonlinear



From Jack (2010), *Dynamic System Modeling and Control*, goo.gl/pGvJPS

*Yes, rectified linear units (relus) are only half-squash; hat-tip Martha White.

Squash Networks

- Parameterized differentiable functions composed out of simpler parameterized differentiable functions, some nonlinear

From existentialcomics.com

- Estimate parameters using Leibniz (1676)

Who wants an all-squash diet?

very *Curcurbita*

much festive

many dropout

wow

Linguistic Structure Prediction

output (structure)

input (text)

Linguistic Structure Prediction

output (structure)

input (text)

Linguistic Structure Prediction

output (structure)

input (text)

Linguistic Structure Prediction

output (structure)

input representation

input (text)

Linguistic Structure Prediction

output (structure)

clusters, lexicons,
embeddings, …

input representation

input (text)

Linguistic Structure Prediction

training objective

"gold" output

sequences, trees, graphs, …

output (structure)

clusters, lexicons, embeddings, …

input representation

input (text)

Linguistic Structure Prediction

training objective

output (structure)

input representation

input (text)

Linguistic Structure Prediction

training objective

output (structure)

part representations

clusters, lexicons,
embeddings, …

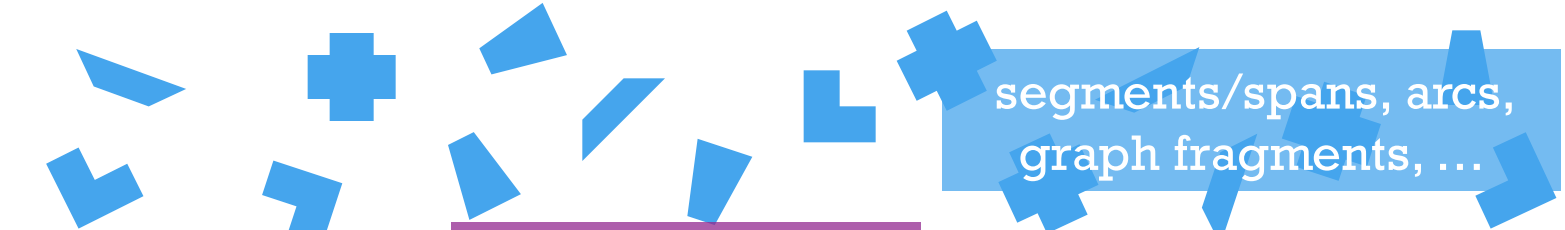
input representation

input (text)

Linguistic Structure Prediction

training objective

output (structure)



part representations

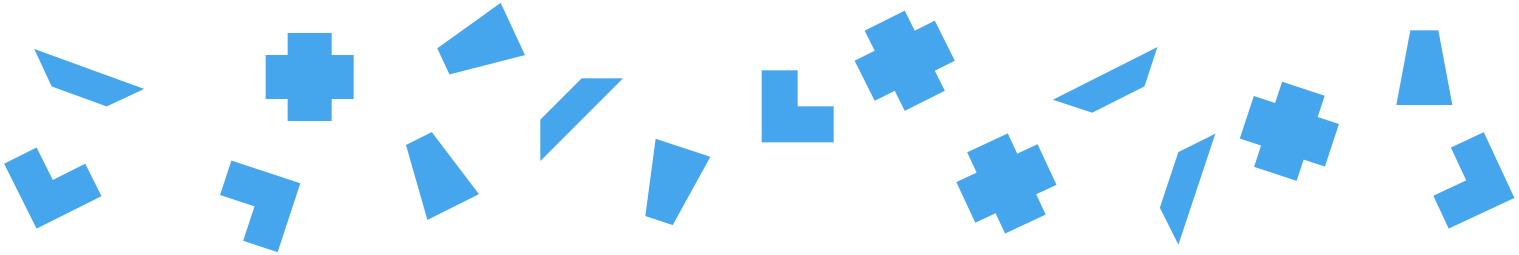
input representation

input (text)

Linguistic Structure Prediction

training objective

output (structure)

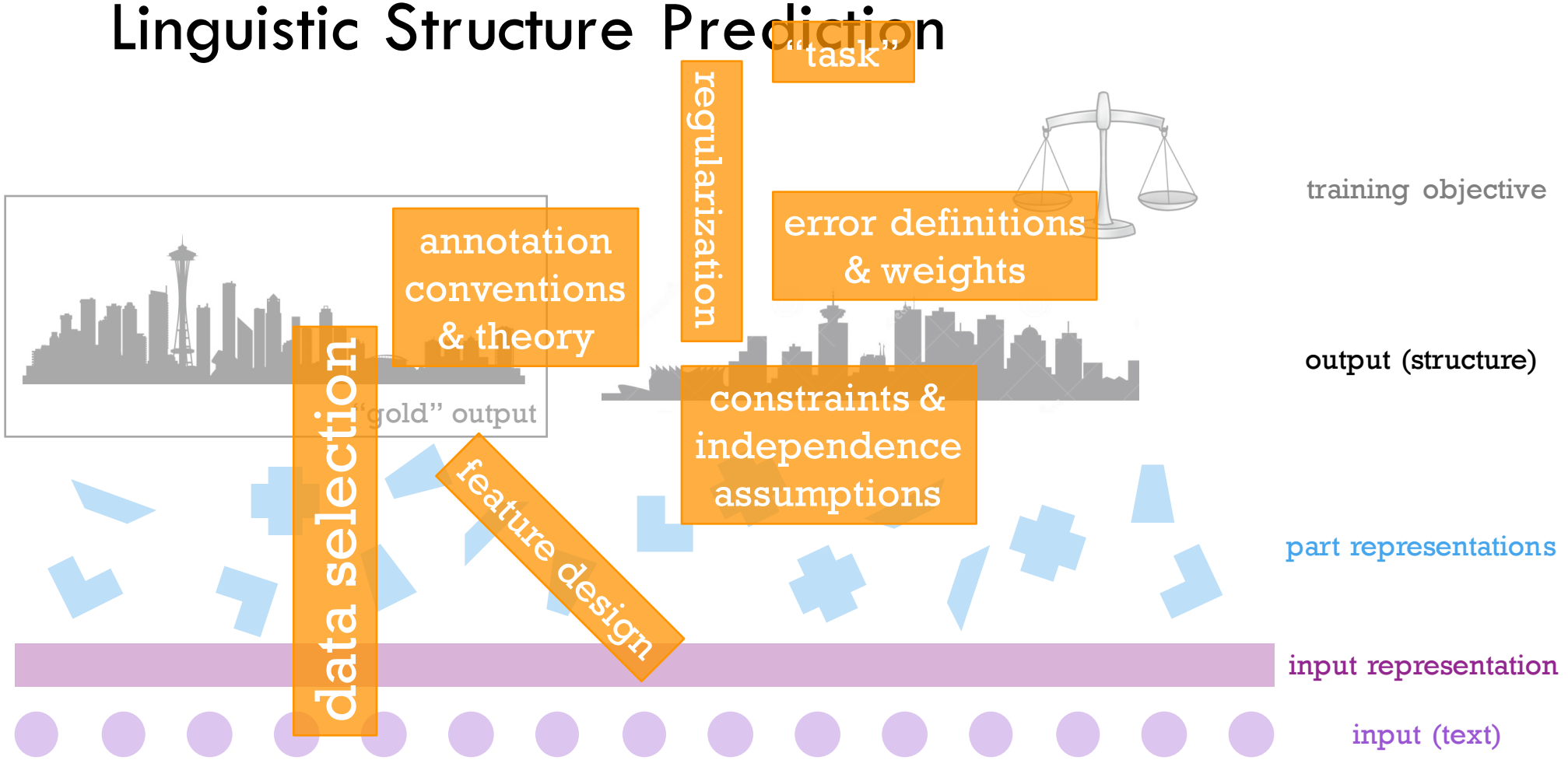


part representations

input representation

input (text)

Linguistic Structure Prediction



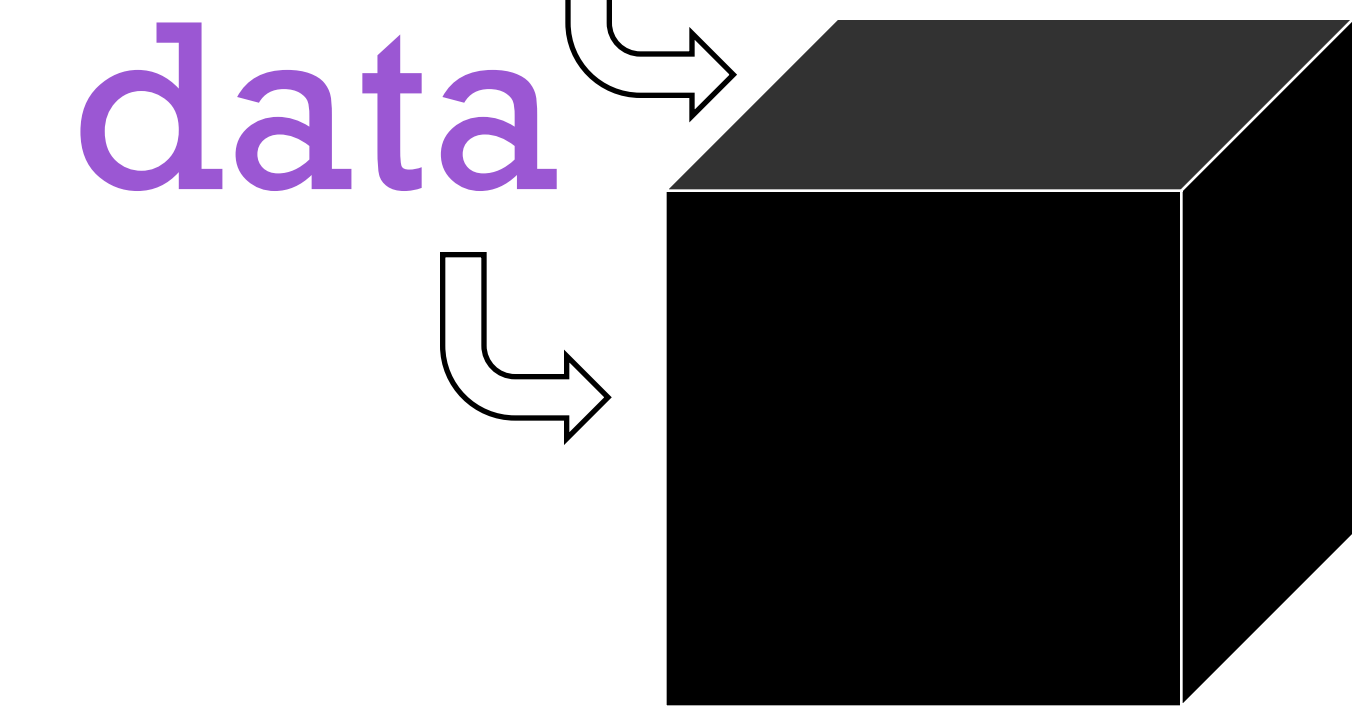
Inductive Bias

- What does your learning algorithm assume?
- How will it choose among good predictive functions?

See also:
No Free Lunch Theorem
(Mitchell, 1980; Wolpert, 1996)

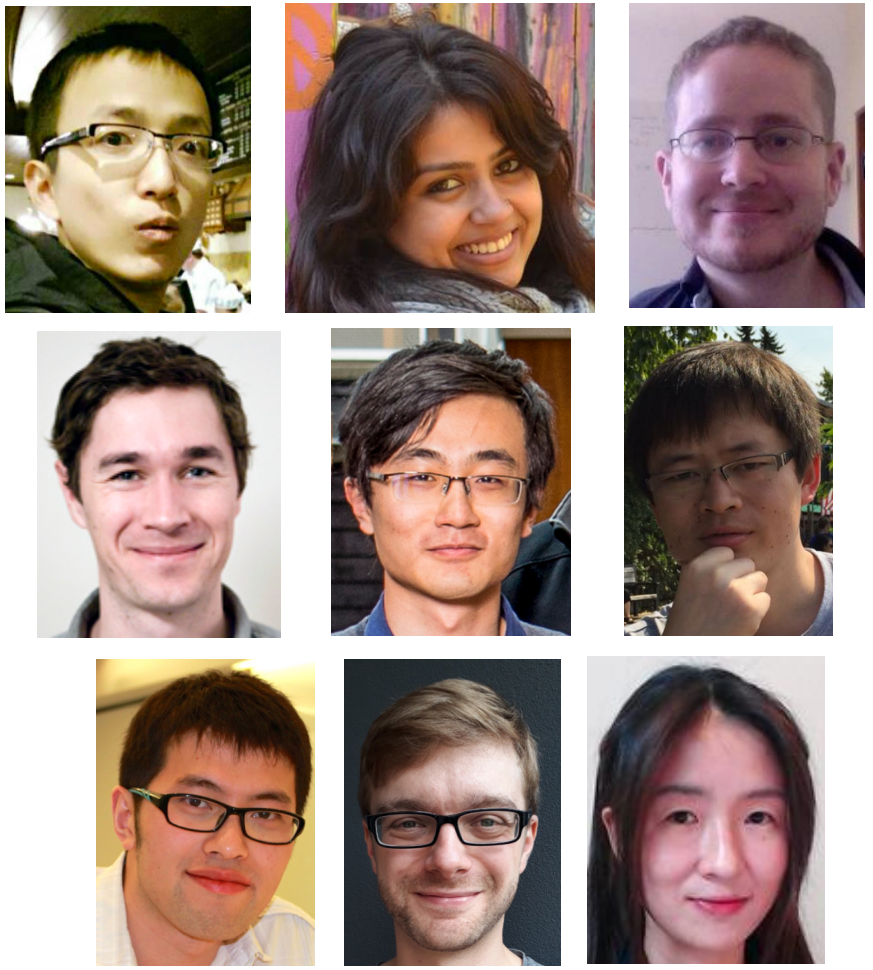
bias

data



Three New Models

- Parsing sentences into predicate-argument structures
  - Fillmore frames
  - Semantic dependency graphs
- Language models that dynamically track entities



When Democrats wonder why there is so much resentment of Clinton, they don't need to look much further than the Big Lie about philandering that Stephanopoulos, Carville helped to put over in 1992.

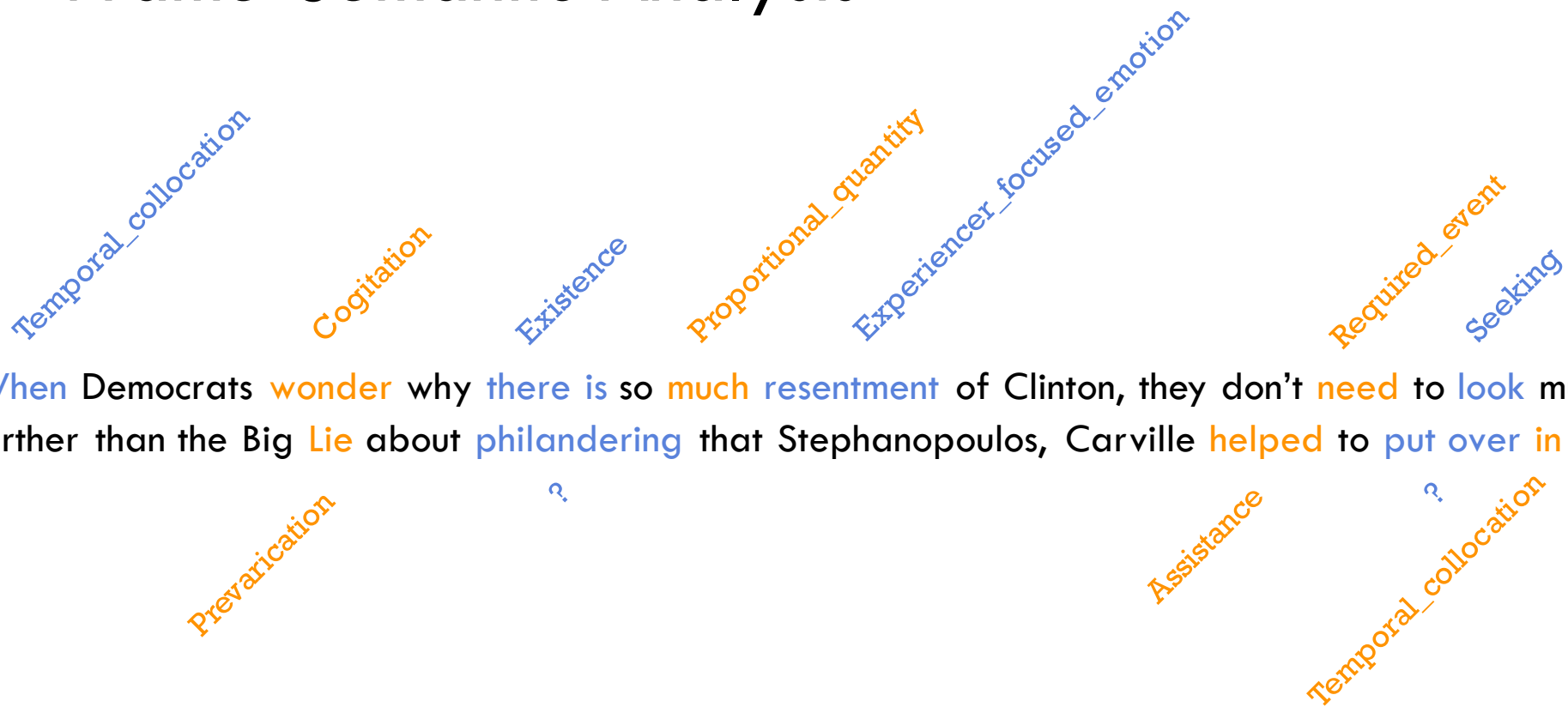
Original story on Slate.com: http://goo.gl/Hp89tD

Frame-Semantic Analysis

When Democrats wonder why there is so much resentment of Clinton, they don't need to look much further than the Big Lie about philandering that Stephanopoulos, Carville helped to put over in 1992.

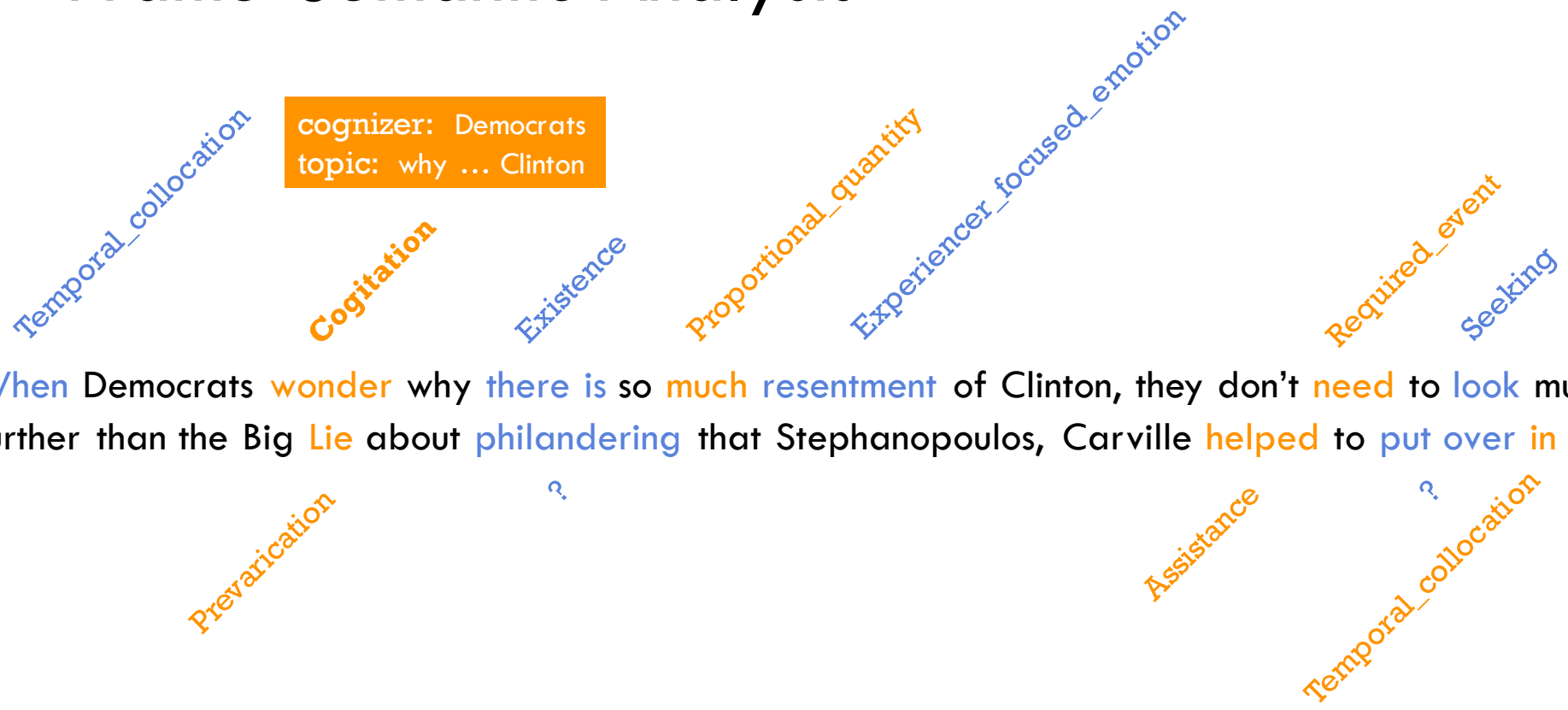
FrameNet: https://framenet.icsi.berkeley.edu

Frame-Semantic Analysis



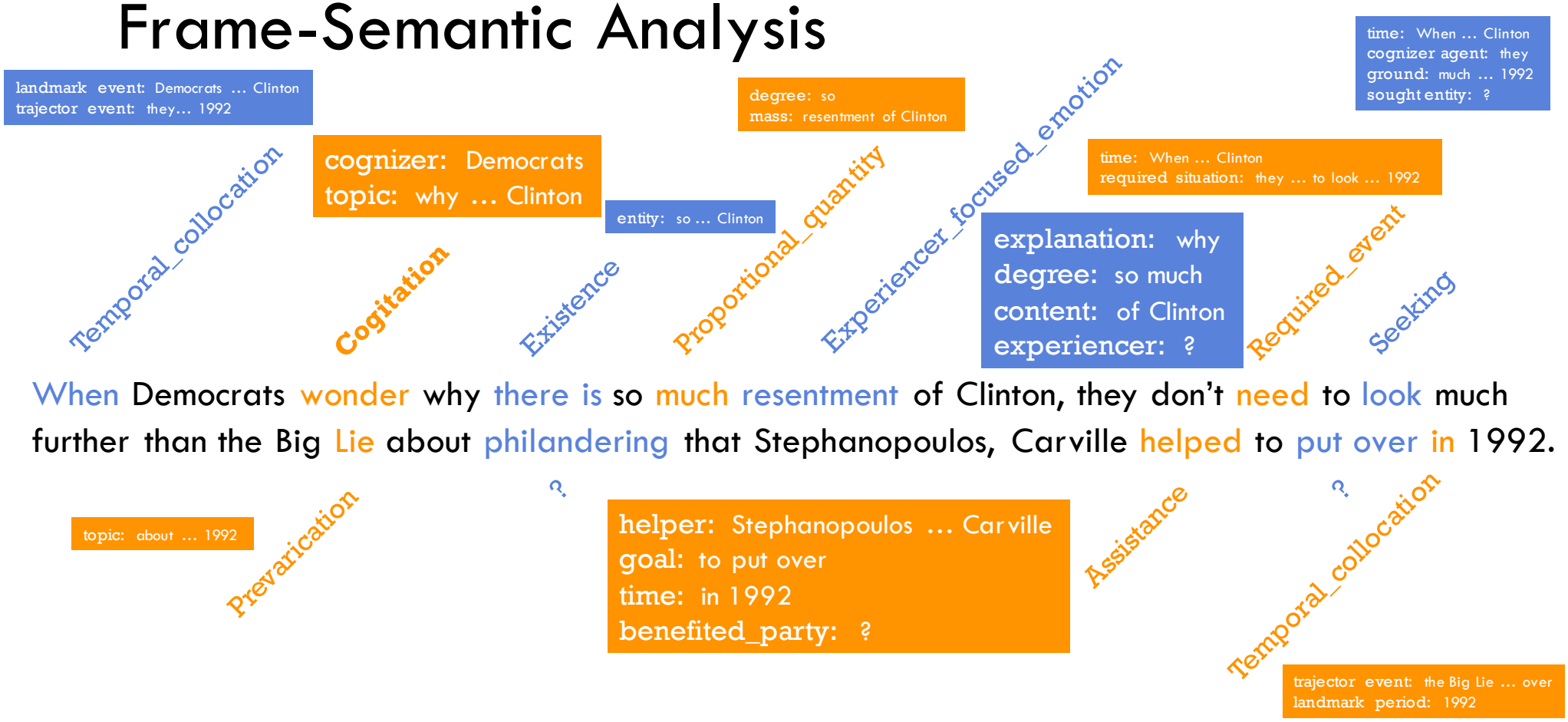
FrameNet: https://framenet.icsi.berkeley.edu

Frame-Semantic Analysis

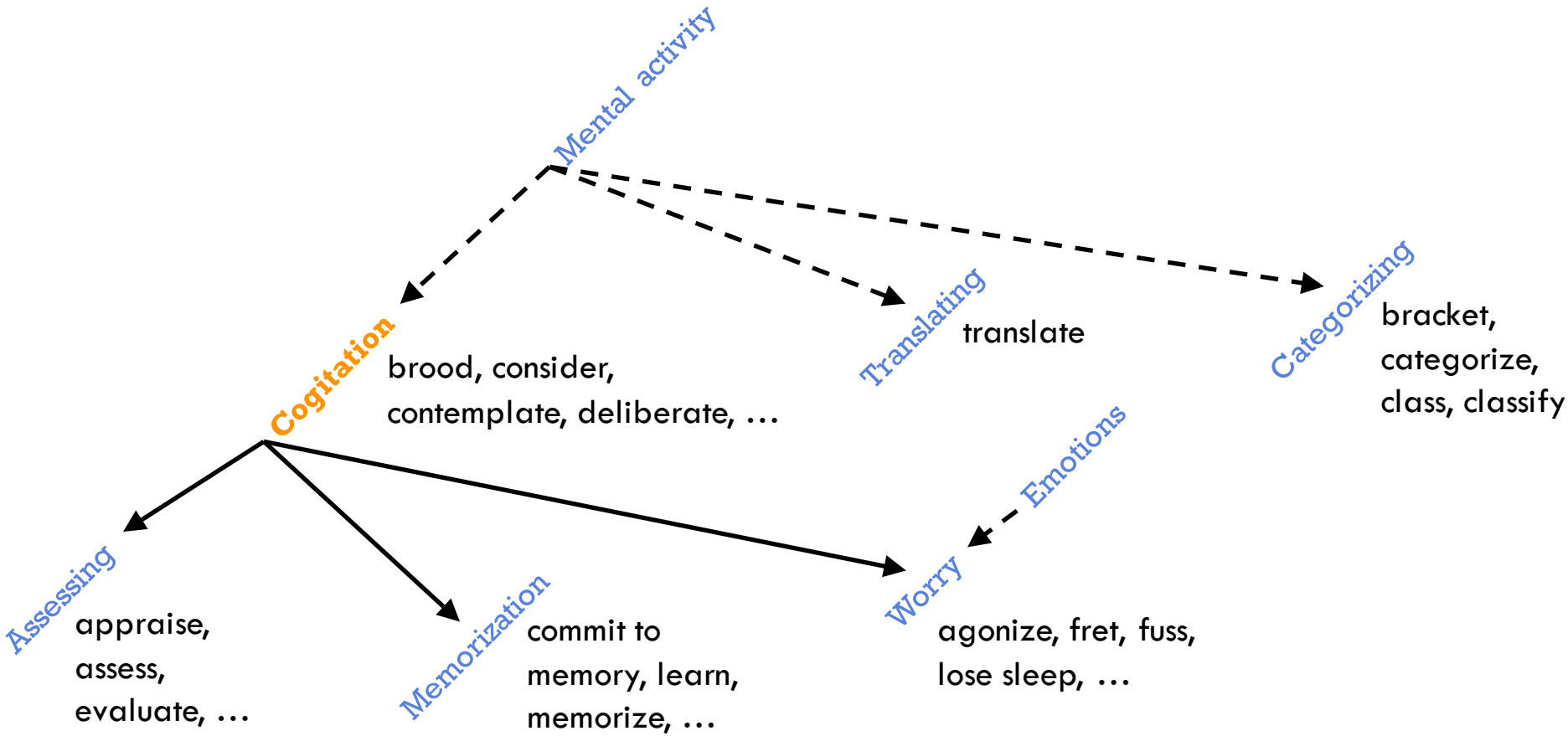


FrameNet: https://framenet.icsi.berkeley.edu

Frame-Semantic Analysis

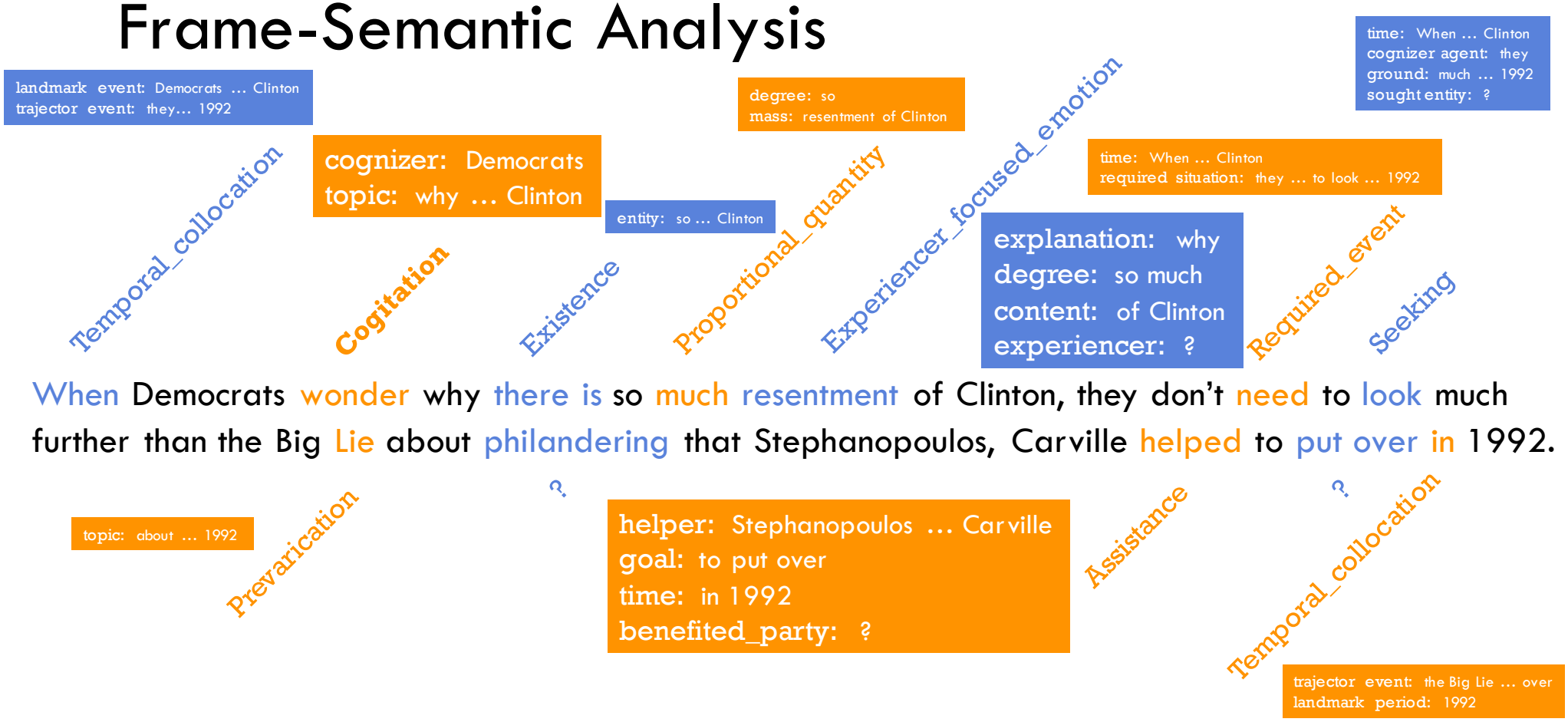


FrameNet: https://framenet.icsi.berkeley.edu



FrameNet: https://framenet.icsi.berkeley.edu

Frame-Semantic Analysis



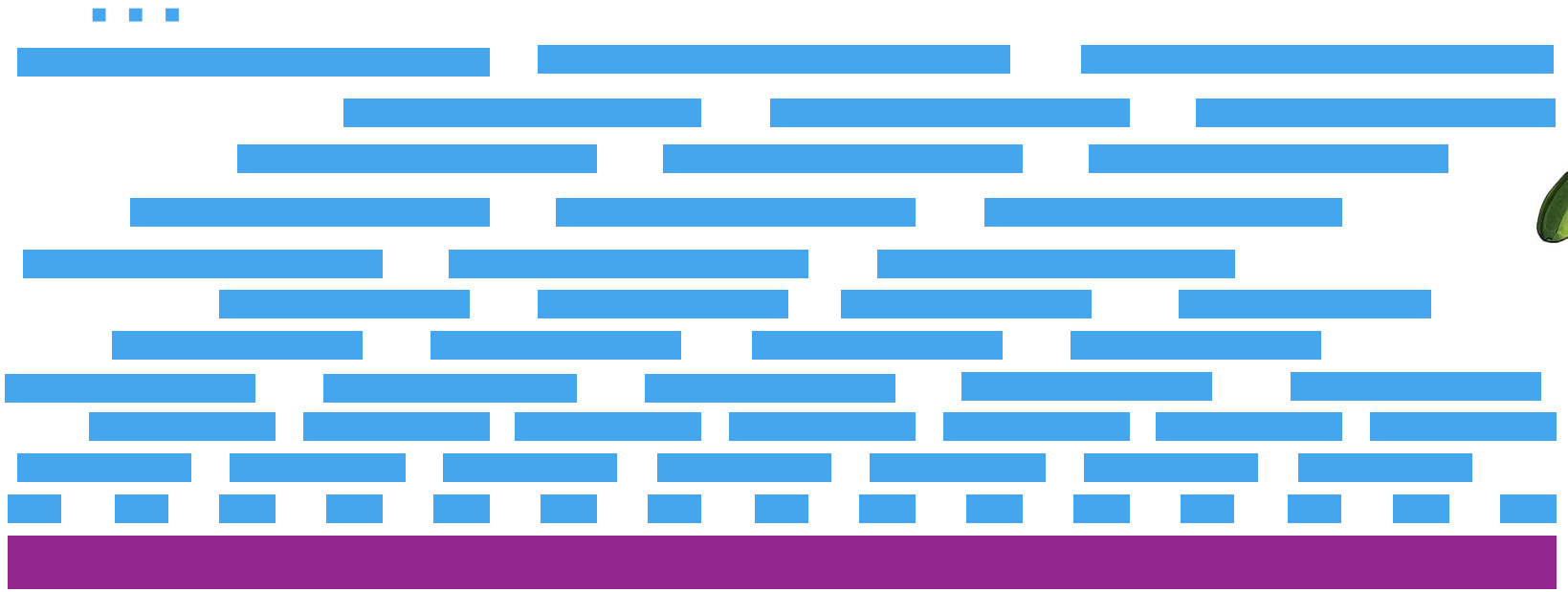
FrameNet: https://framenet.icsi.berkeley.edu

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need … words + frame

biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

words + frame



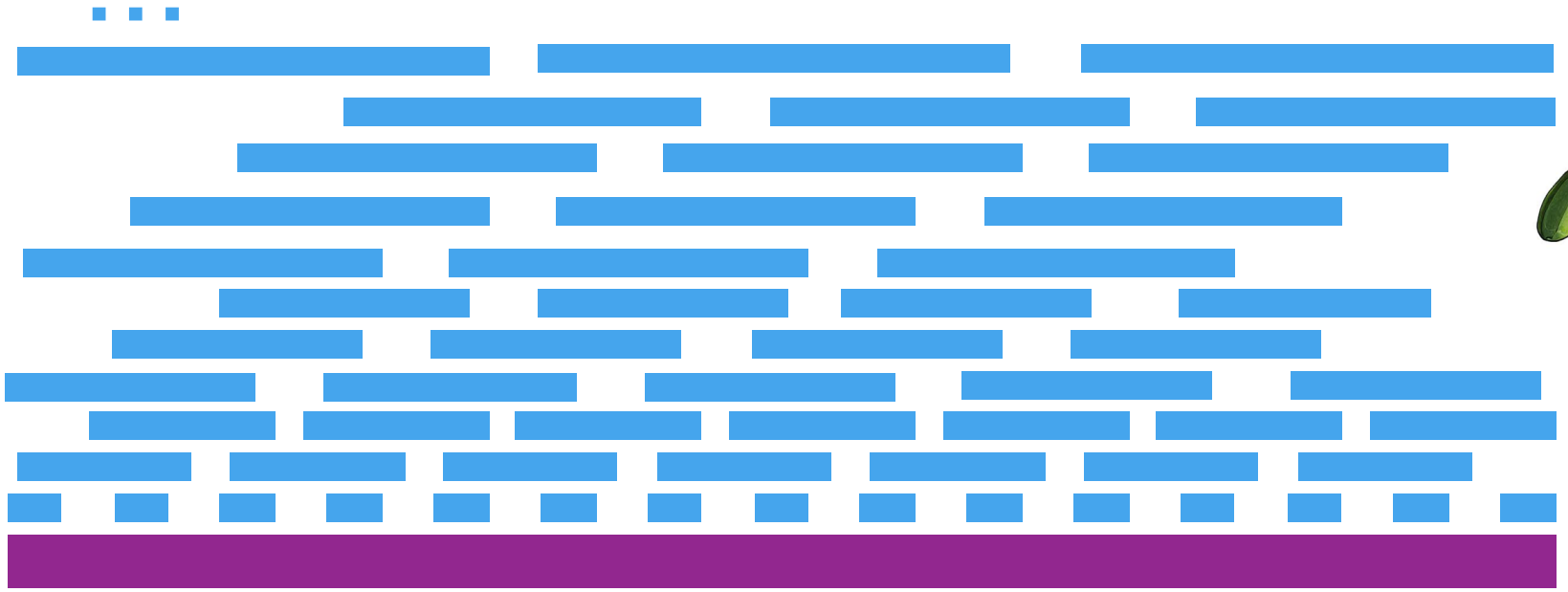
parts: segments up to length *d* scored by another biLSTM, with labels

biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

words + frame

output: covering sequence of nonoverlapping segments



parts: segments up to length *d* scored by another biLSTM, with labels

biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

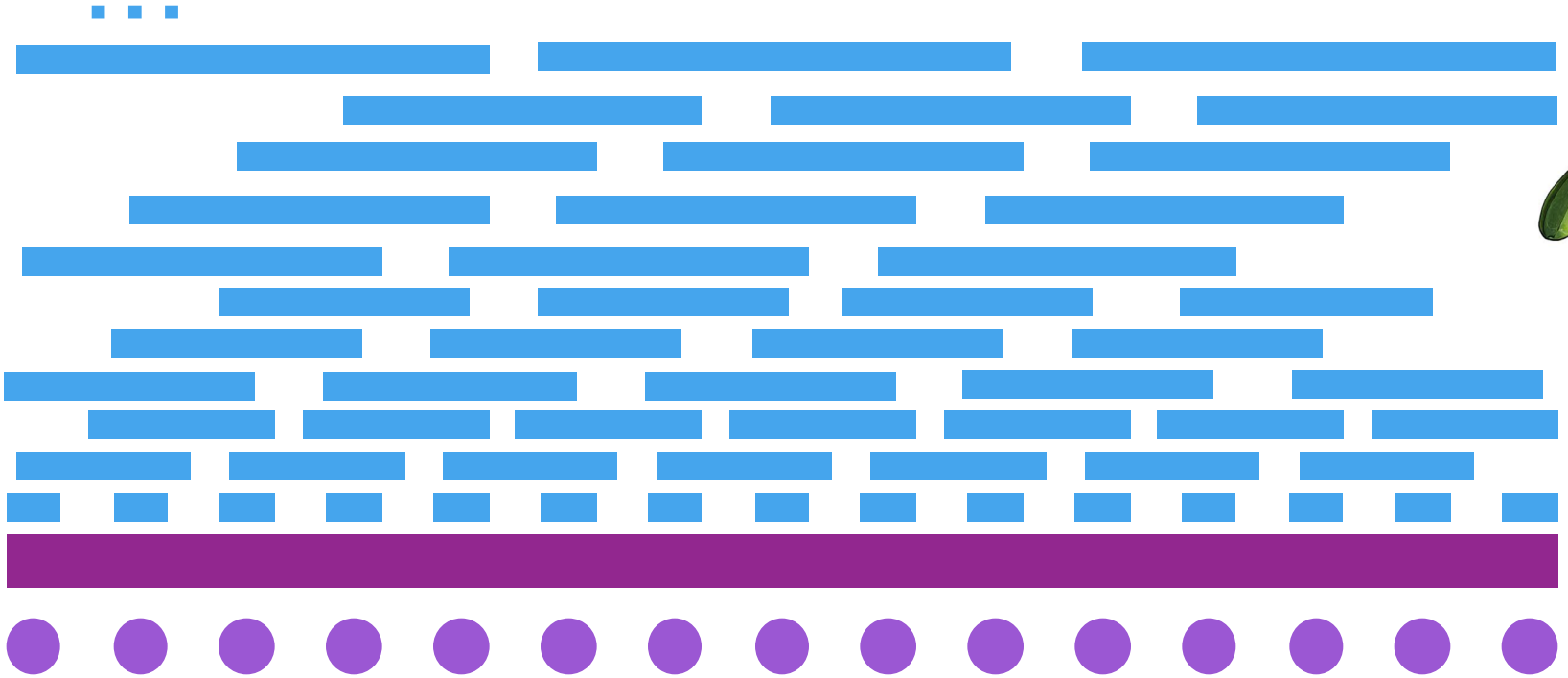
words + frame

Segmental RNN

(Lingpeng Kong, Chris Dyer, N.A.S., ICLR 2016)

training objective: log loss

output: covering sequence of nonoverlapping segments, recovered in $O(Ldn)$; see Sarawagi & Cohen, 2004

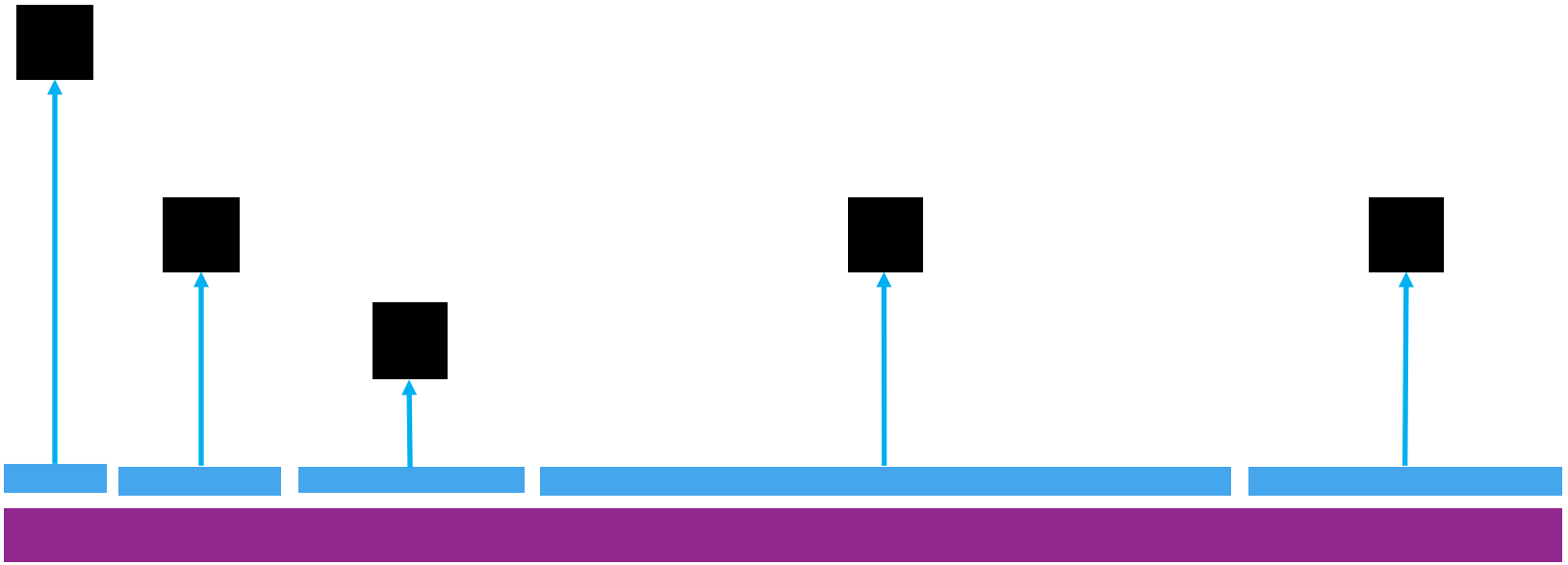


parts: segments up to length d scored by another biLSTM, with labels

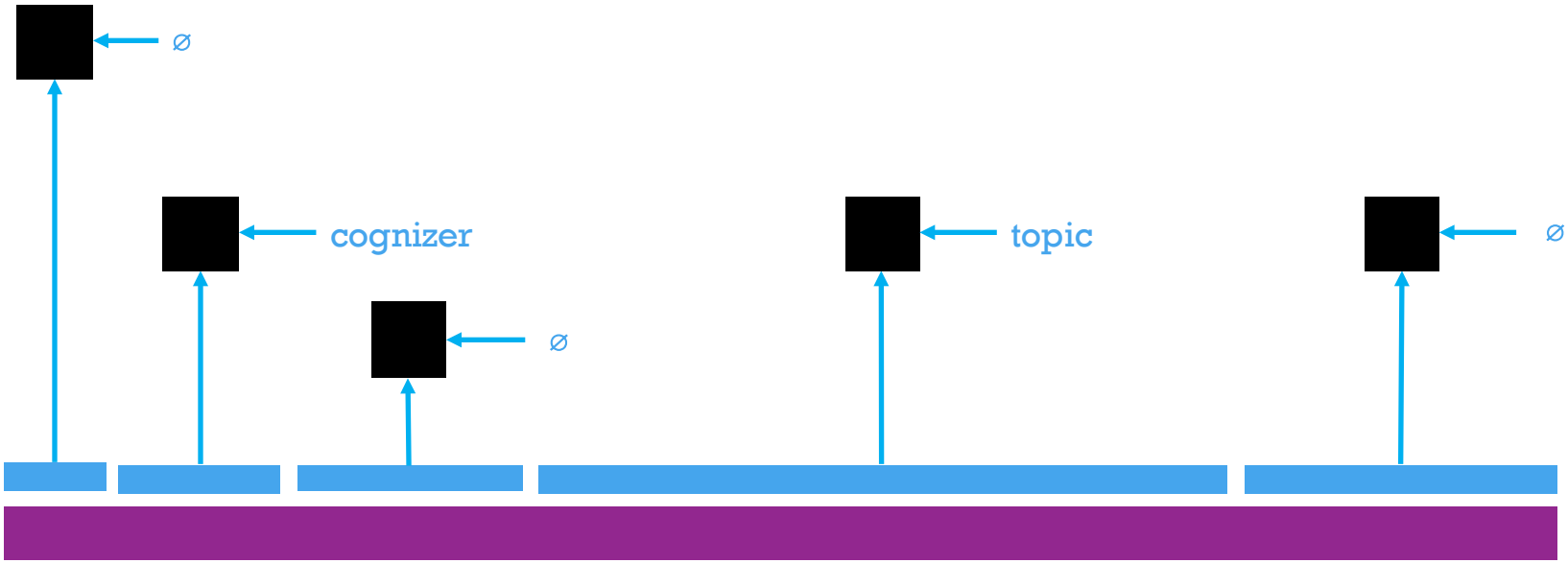
biLSTM (contextualized word vectors)

input sequence

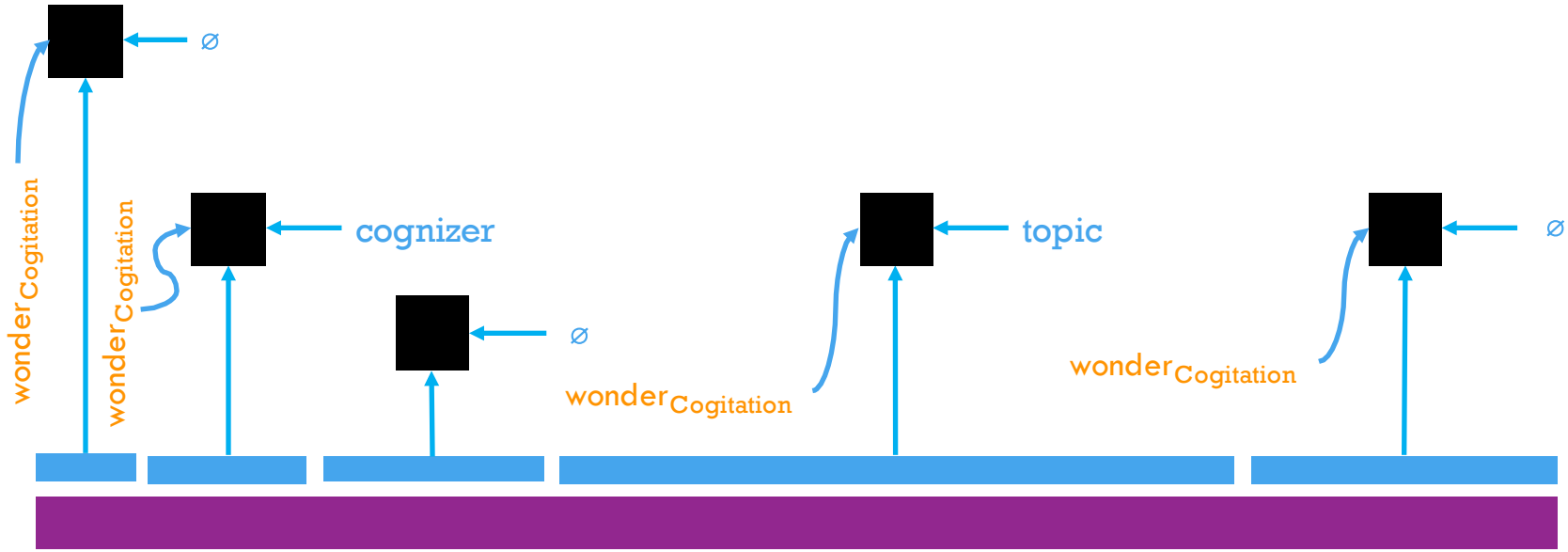
When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …



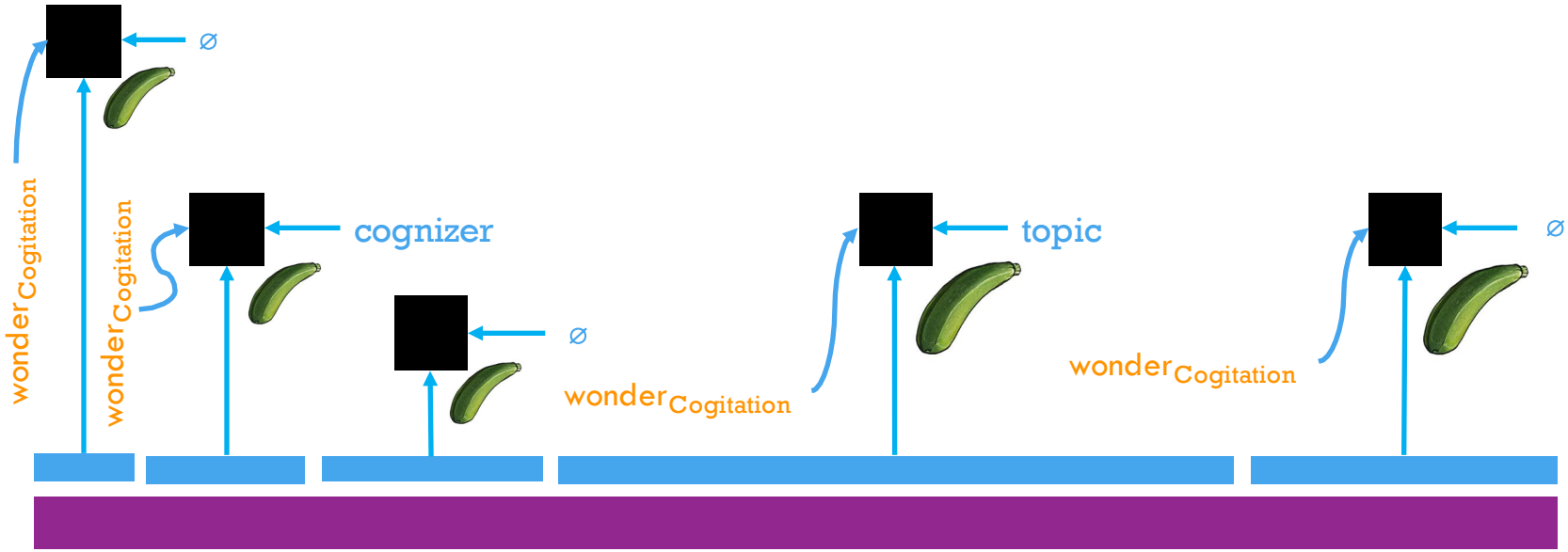
When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …



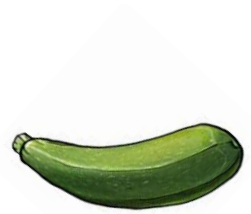
When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …



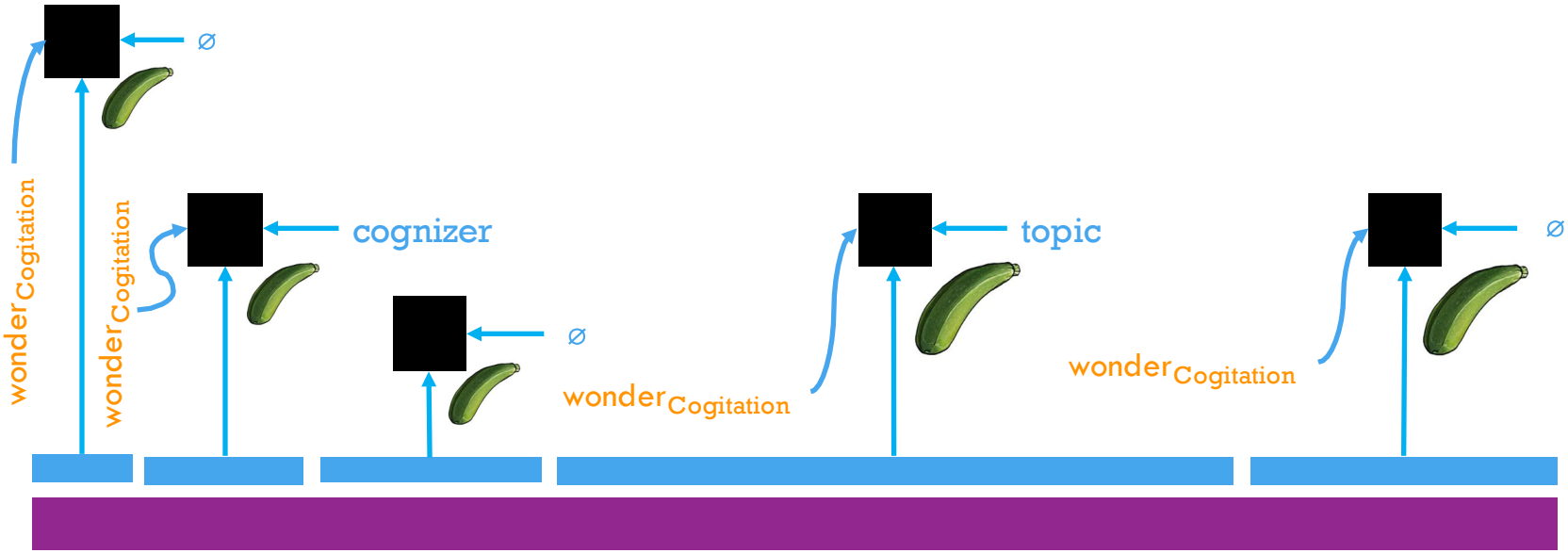
When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …



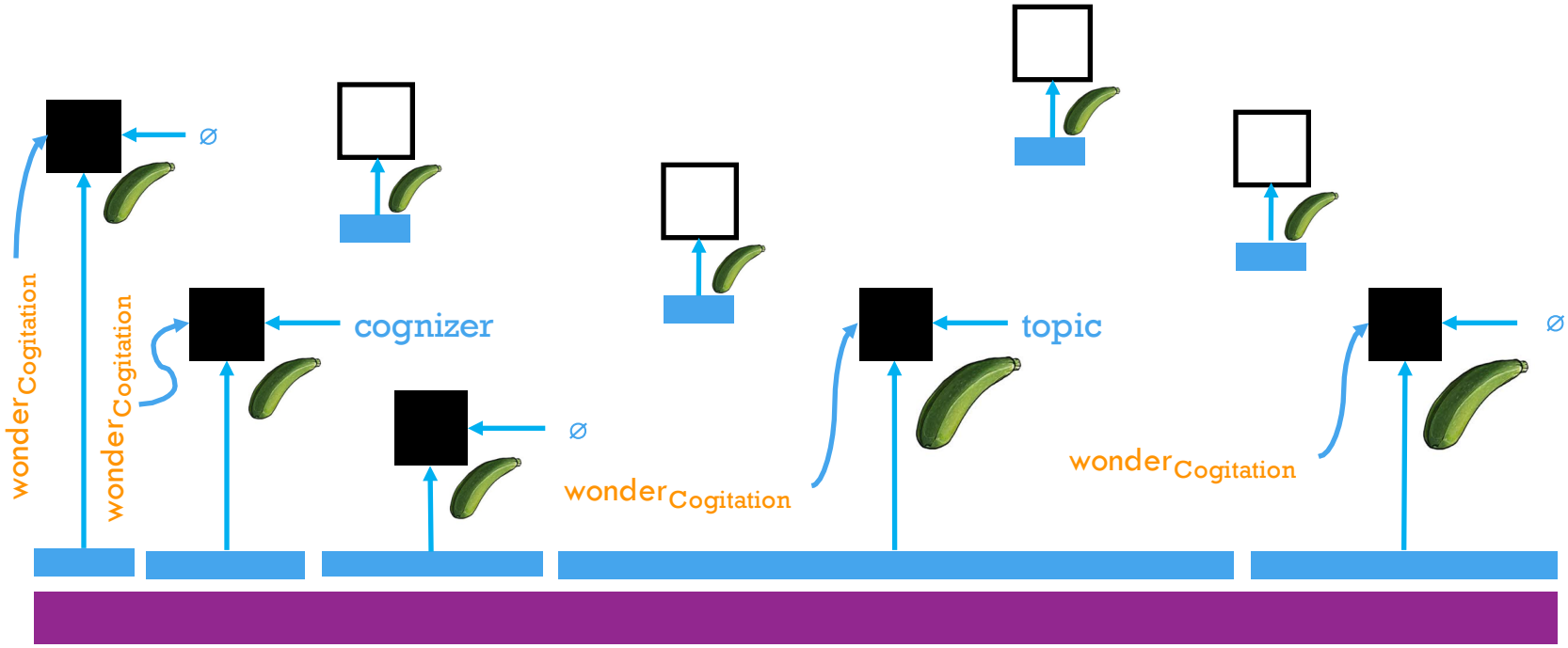
When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …



$$\mathbf{w}_2 \cdot \mathrm{relu}(\mathbf{W}_1[\boldsymbol{h}_{\text{why:Clinton}}; \mathbf{v}_{\text{topic}}; \boldsymbol{h}'_{\text{wonder, Cogitation}}])$$

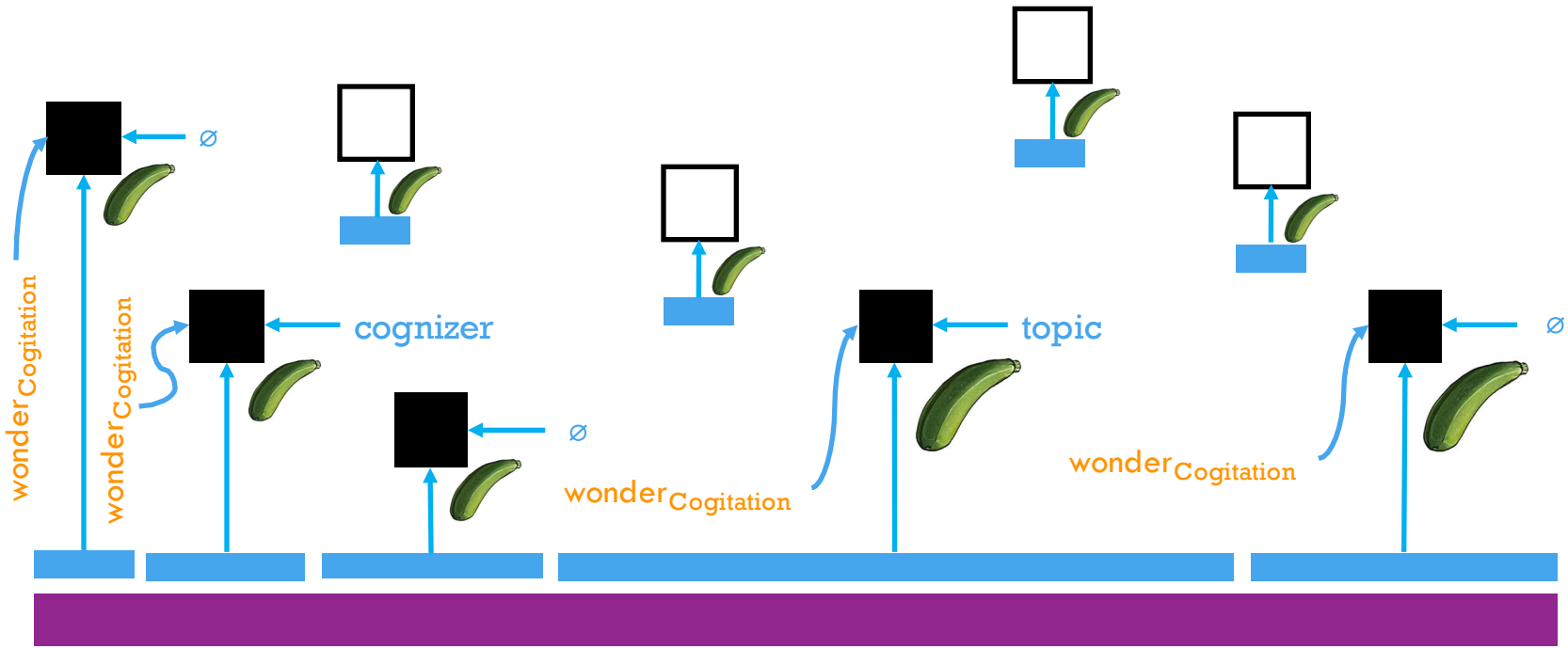


When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …



When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …

Inference via dynamic programming in O(*Ldn*)



When Democrats wonder Cogitation why there is so much resentment of Clinton, they don't need …

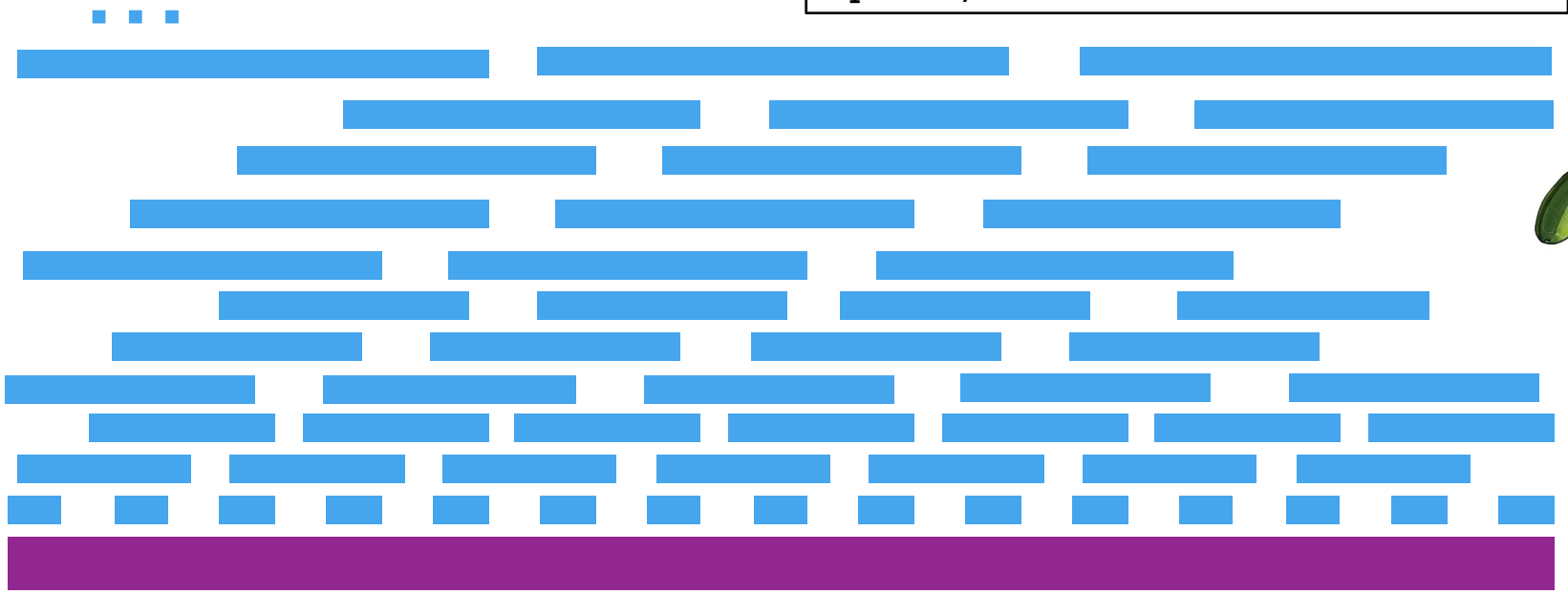
Open-SESAME

(Swabha Swayamdipta, Sam Thomson, Chris Dyer, N.A.S., arXiv:1706.09528)

training objective: recall-oriented softmax margin (Gimpel et al., 2010)

cognizer: Democrats
topic: why there is so much resentment of Clinton

output: labeled argument spans

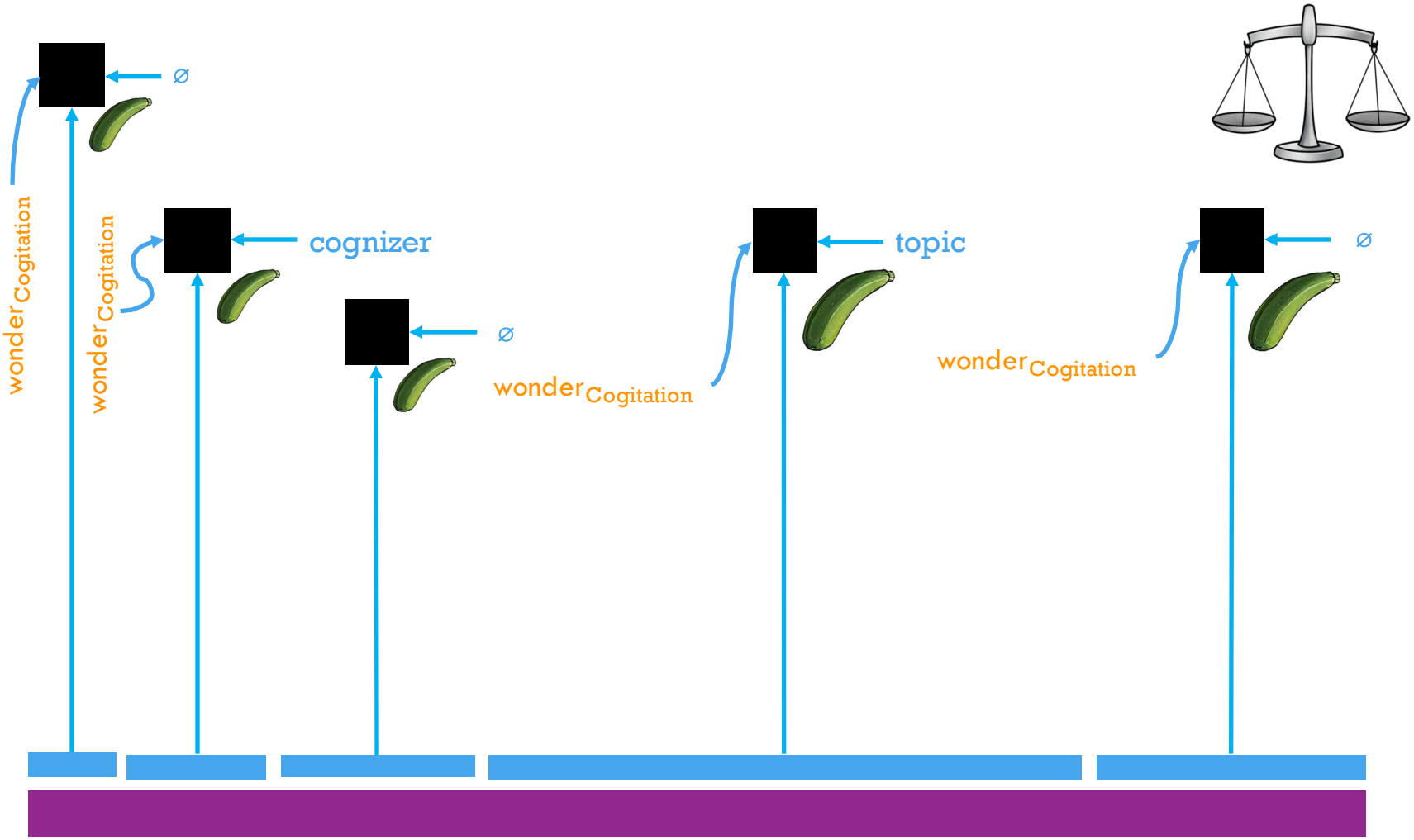


parts: segments with role labels, scored by another biLSTM

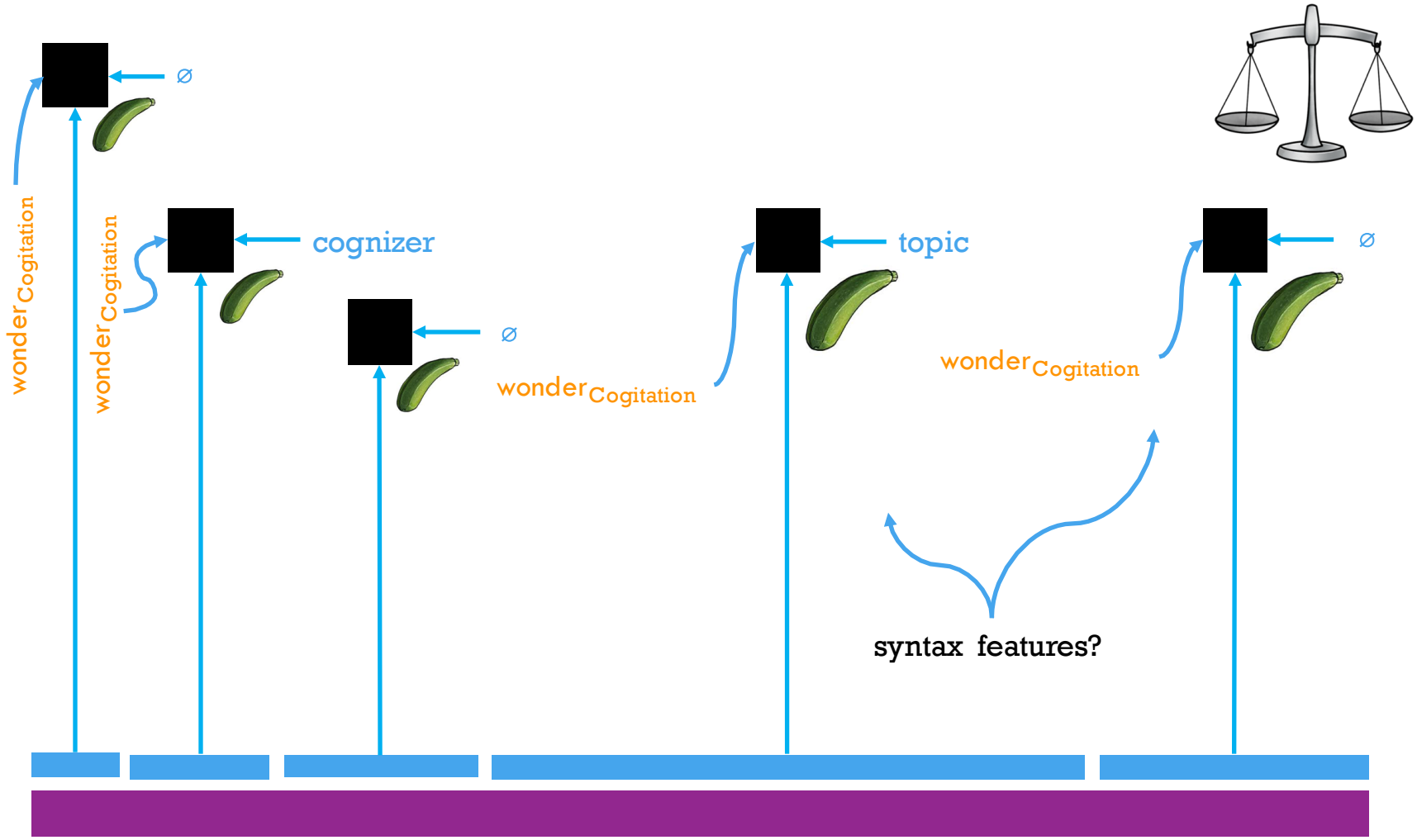
biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

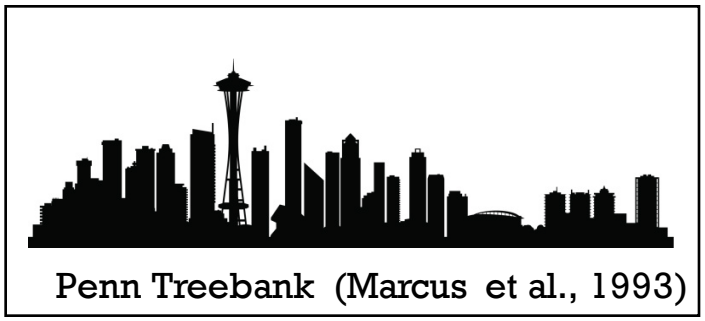
words + frame

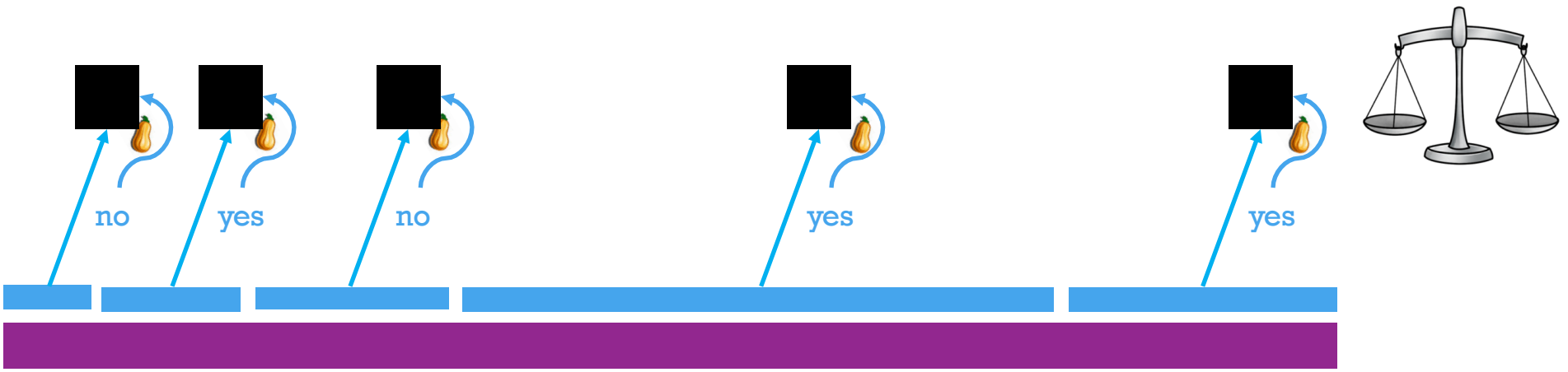


When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …

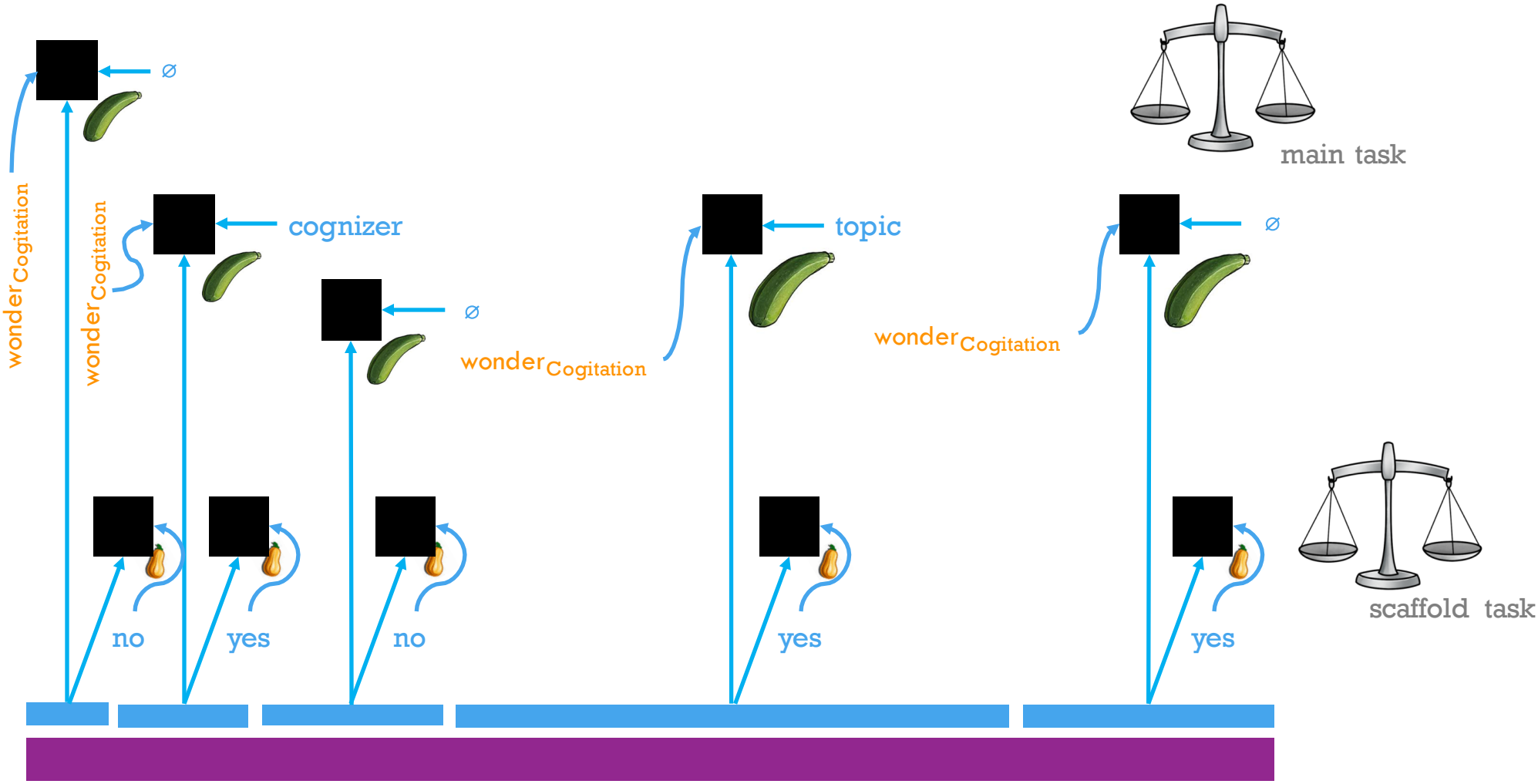


When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …





When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …



When Democrats wonder$_{Cogitation}$ why there is so much resentment of Clinton, they don't need …

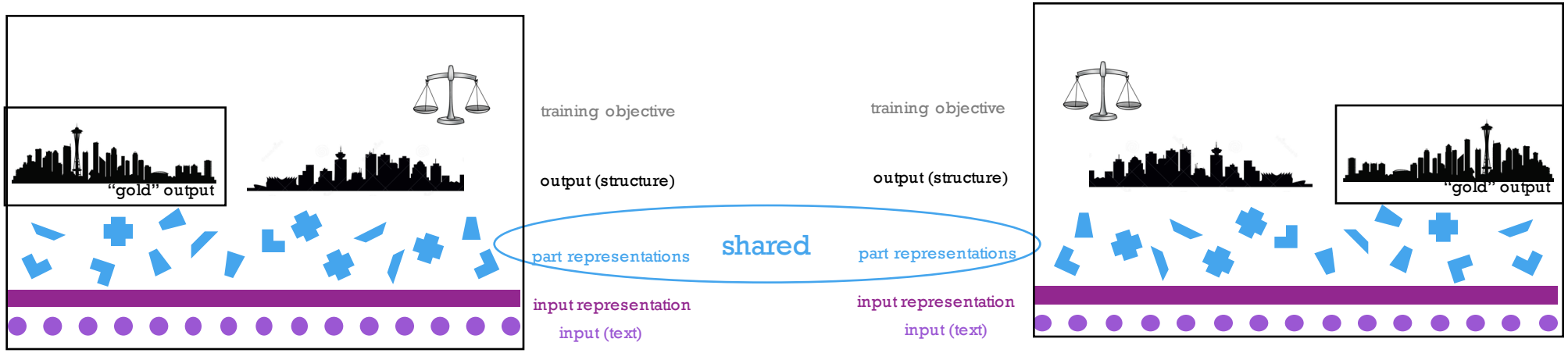
Multitask Representation Learning

(Caruana, 1997)

main task: find and label semantic arguments

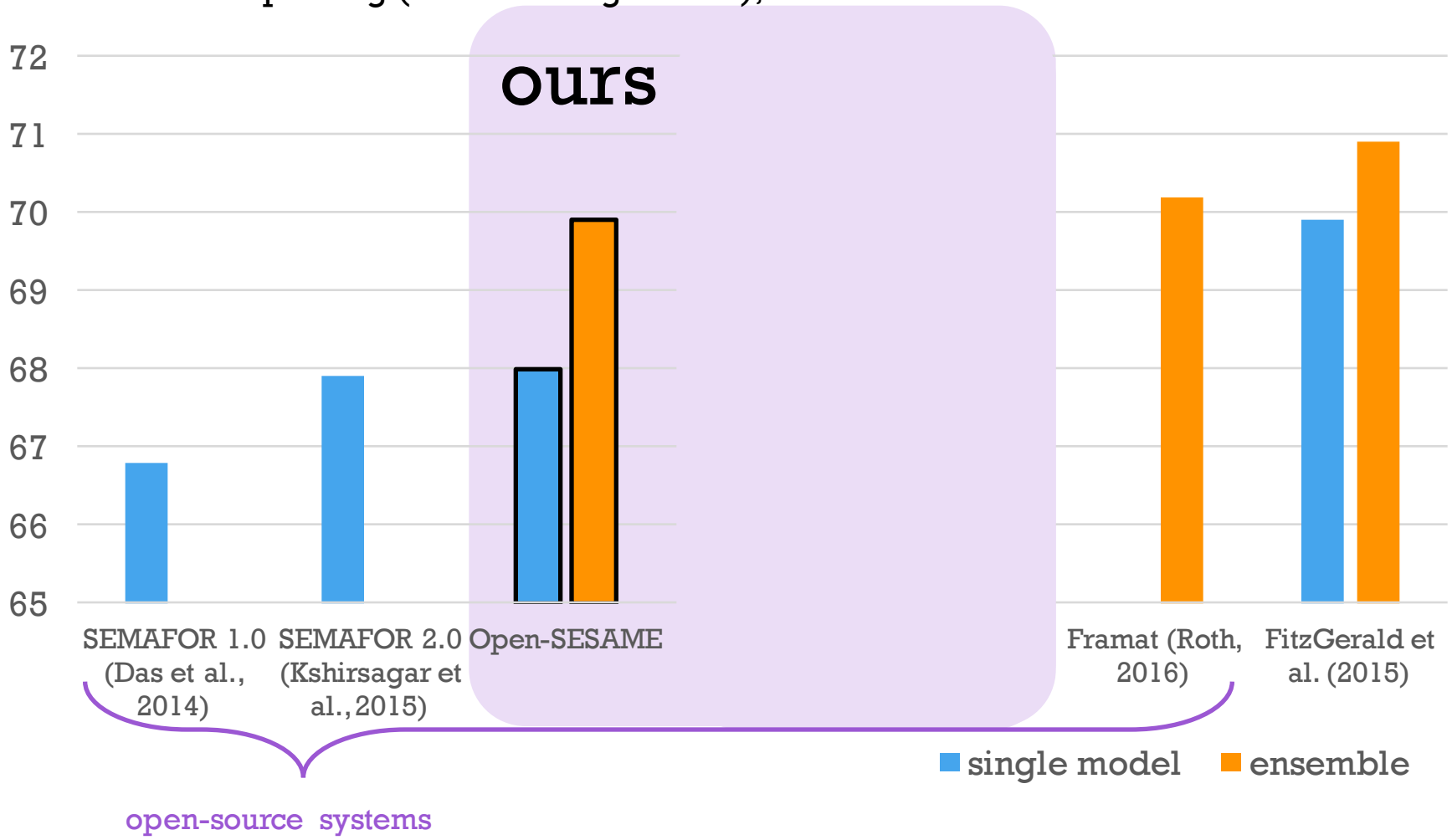
output structures need not be consistent

scaffold task: predict syntactic constituents

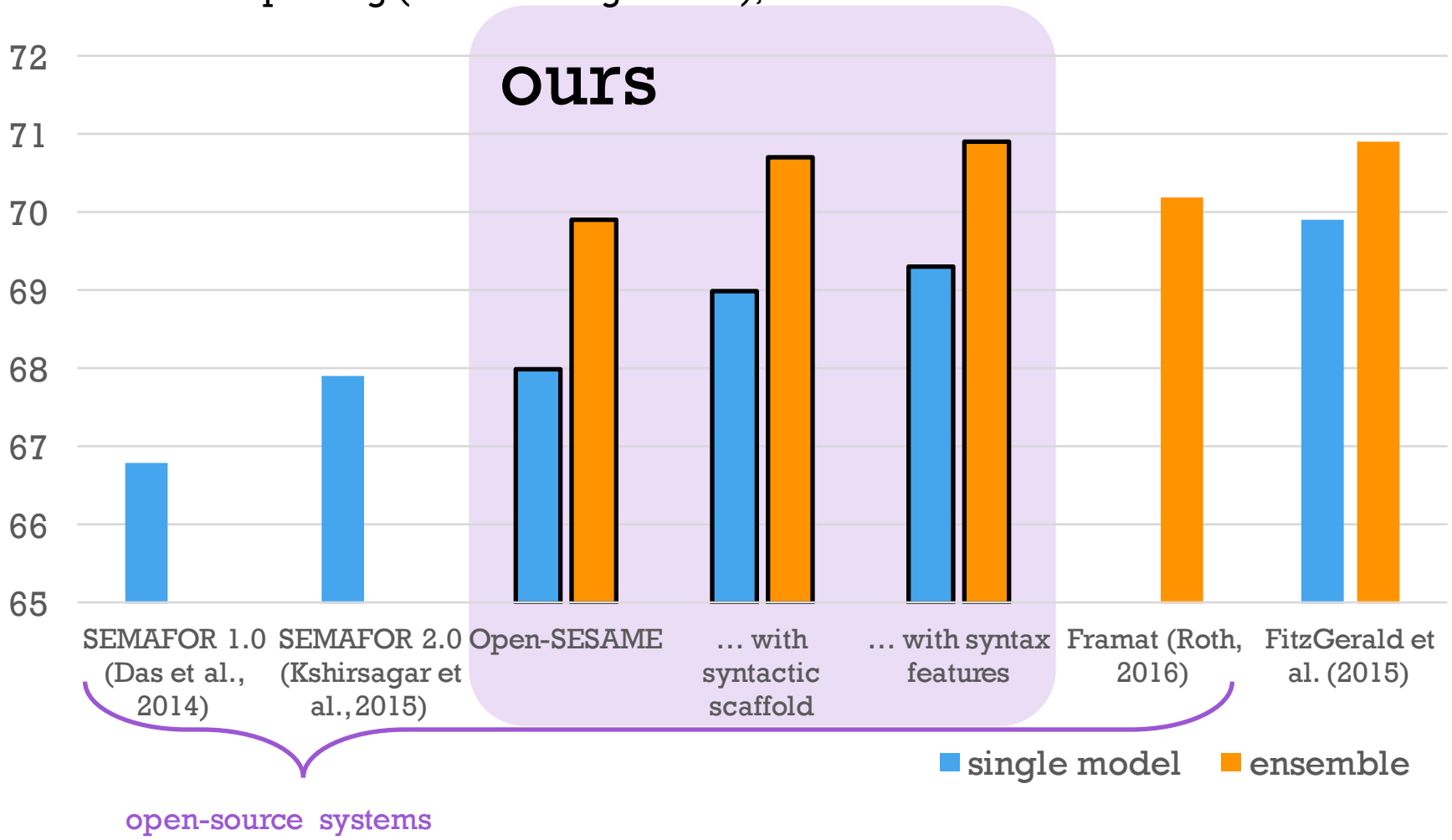


training datasets need not overlap

F_1 on frame-semantic parsing (frames & arguments), FrameNet 1.5 test set.



F_1 on frame-semantic parsing (frames & arguments), FrameNet 1.5 test set.



Bias?

training objective: recall-oriented softmax margin (Gimpel et al., 2010)

cognizer: Democrats
topic: why there is so much resentment of Clinton

output: labeled argument spans

segments get scores

parts: segments with role labels, scored by another biLSTM

biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

words + frame

Bias?

training objective: recall-oriented softmax margin (Gimpel et al., 2010)

syntactic scaffold

cognizer: Democrats
topic: why there is so much resentment of Clinton

output: labeled argument spans

segments get scores

parts: segments with role labels, scored by another biLSTM

biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

words + frame

Bias?

recall-oriented cost

training objective: recall-oriented softmax margin (Gimpel et al., 2010)

syntactic scaffold

cognizer: Democrats
topic: why there is so much resentment of Clinton

output: labeled argument spans

segments get scores

parts: segments with role labels, scored by another biLSTM

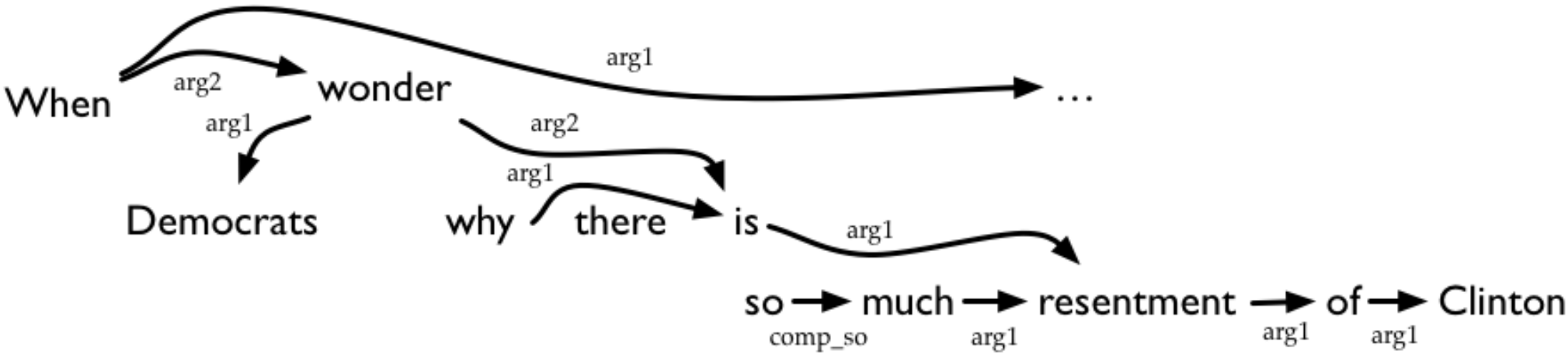
biLSTM (contextualized word vectors)

When Democrats wonder$_{\text{Cogitation}}$ why there is so much resentment of Clinton, they don't need …

words + frame

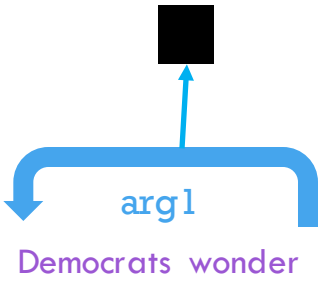
Semantic Dependency Graphs

(DELPH-IN minimal recursion semantics-derived representation; "DM")

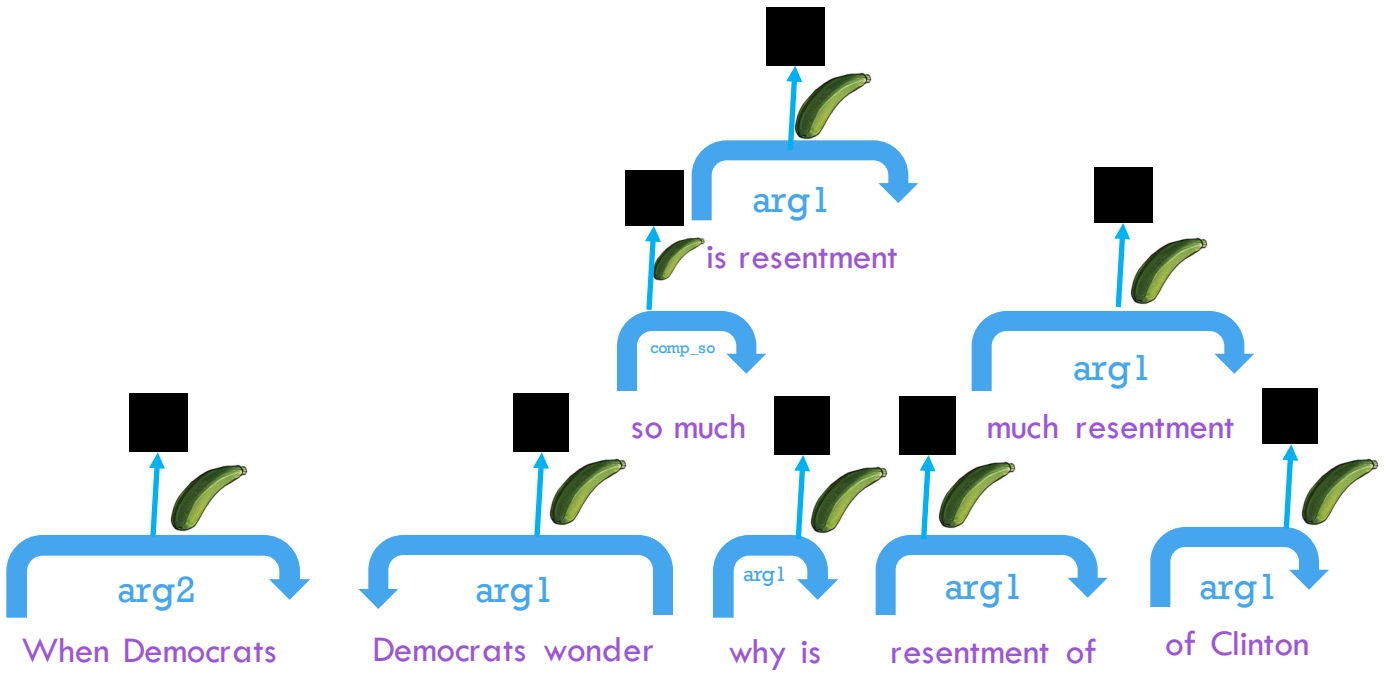


Oepen et al. (SemEval 2014; 2015), see also http://sdp.delph-in.net

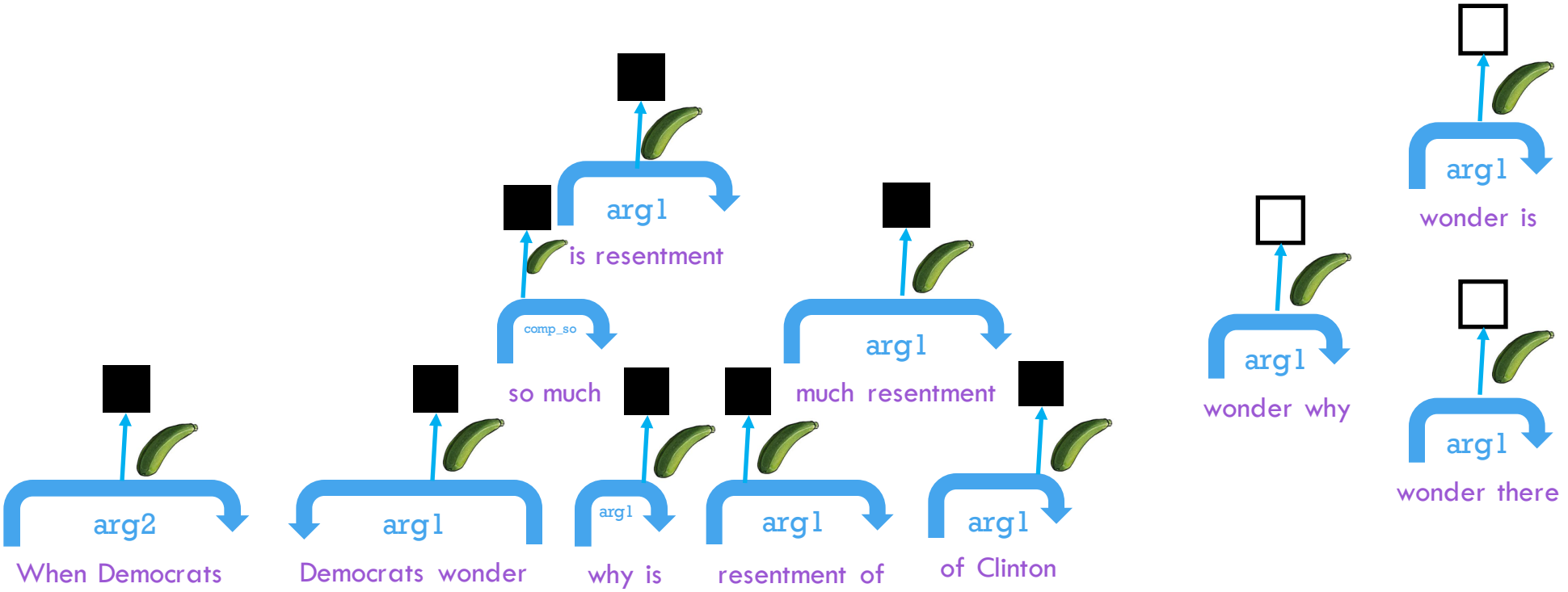
$$\tanh(\mathbf{C}[\boldsymbol{h}_{\text{wonder}}; \boldsymbol{h}_{\text{Democrats}}] + \mathbf{b}) \cdot \boldsymbol{\psi}_{\text{arg1}}$$



When Democrats wonder why there is so much resentment of Clinton, they don't need …



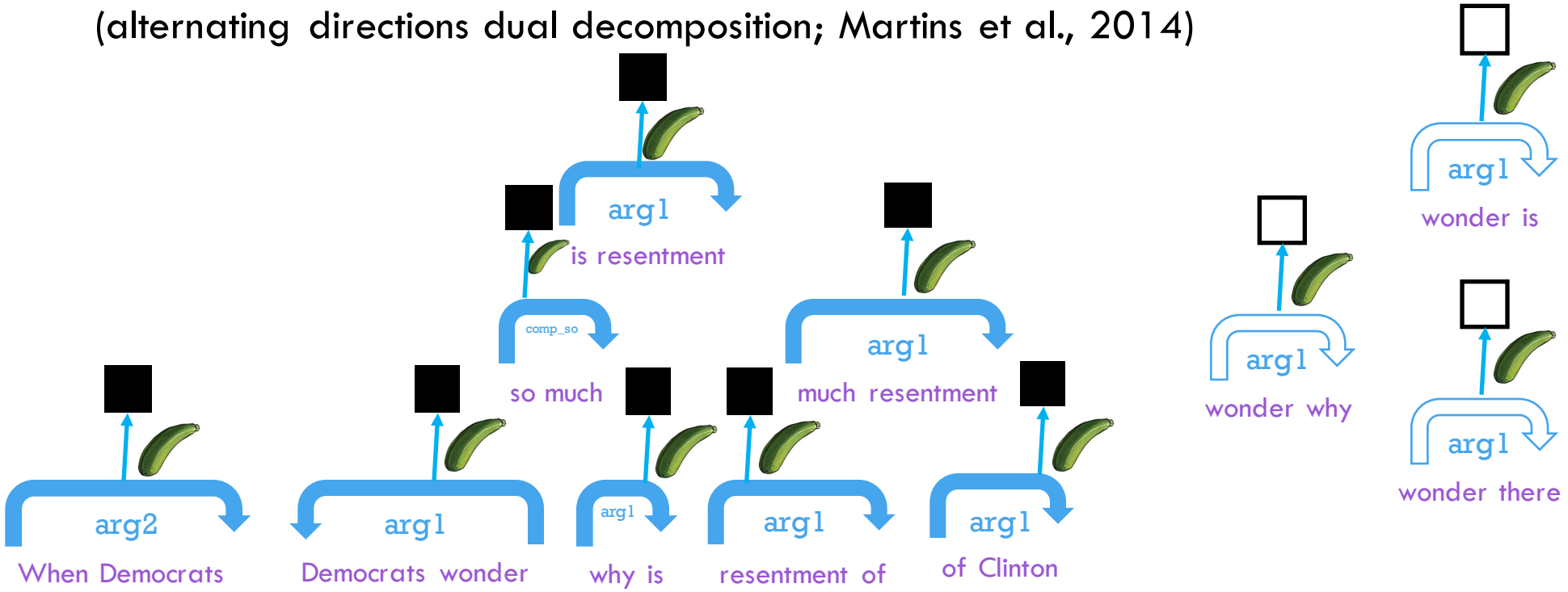
When Democrats wonder why there is so much resentment of Clinton, they don't need …



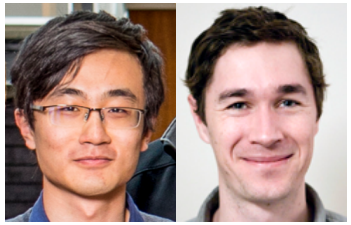
When Democrats wonder why there is so much resentment of Clinton, they don't need …

Inference via AD^3

(alternating directions dual decomposition; Martins et al., 2014)



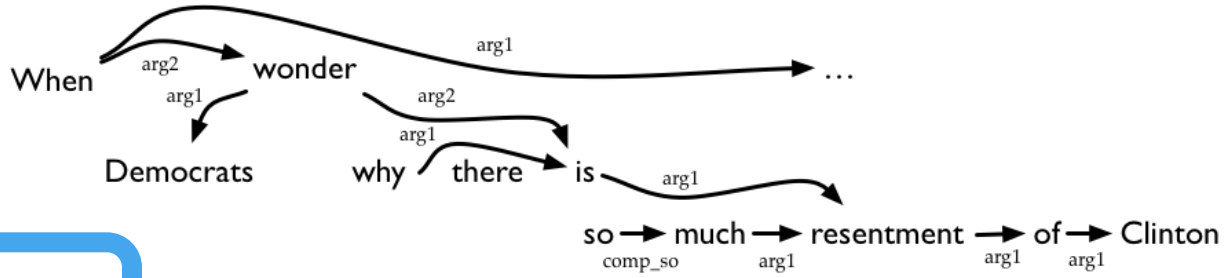
When Democrats wonder why there is so much resentment of Clinton, they don't need …



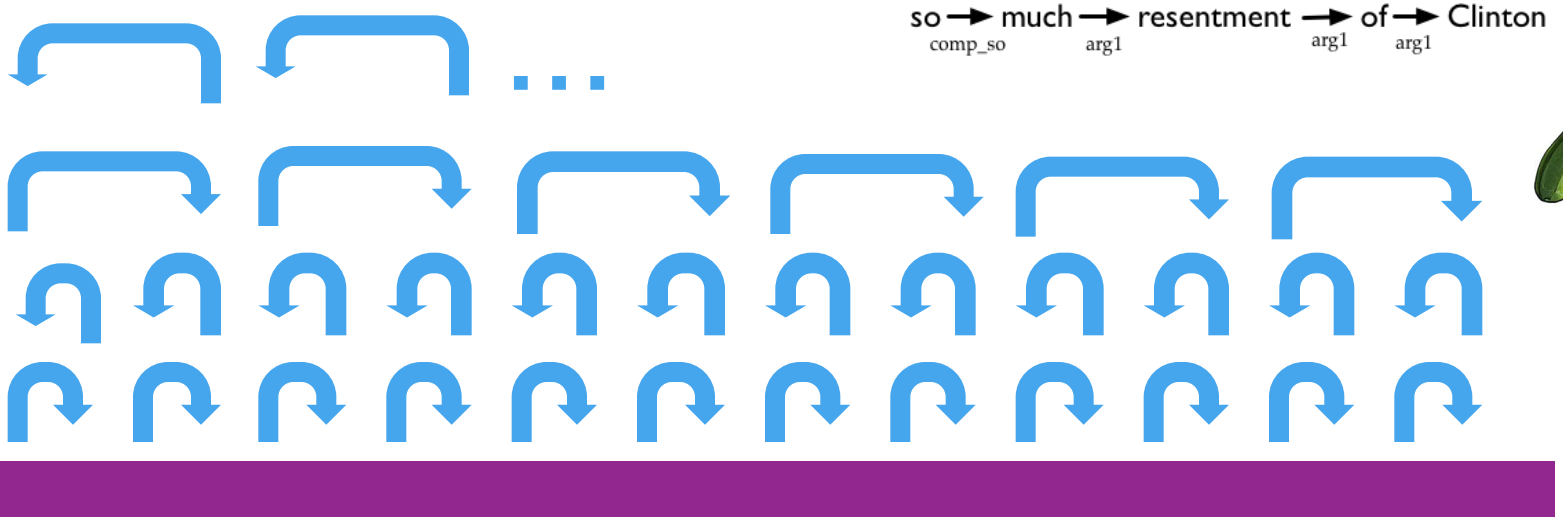
Neurboparser

(Hao Peng, Sam Thomson, N.A.S., ACL 2017)

training objective: structured hinge loss



output: labeled semantic dependency graph with constraints



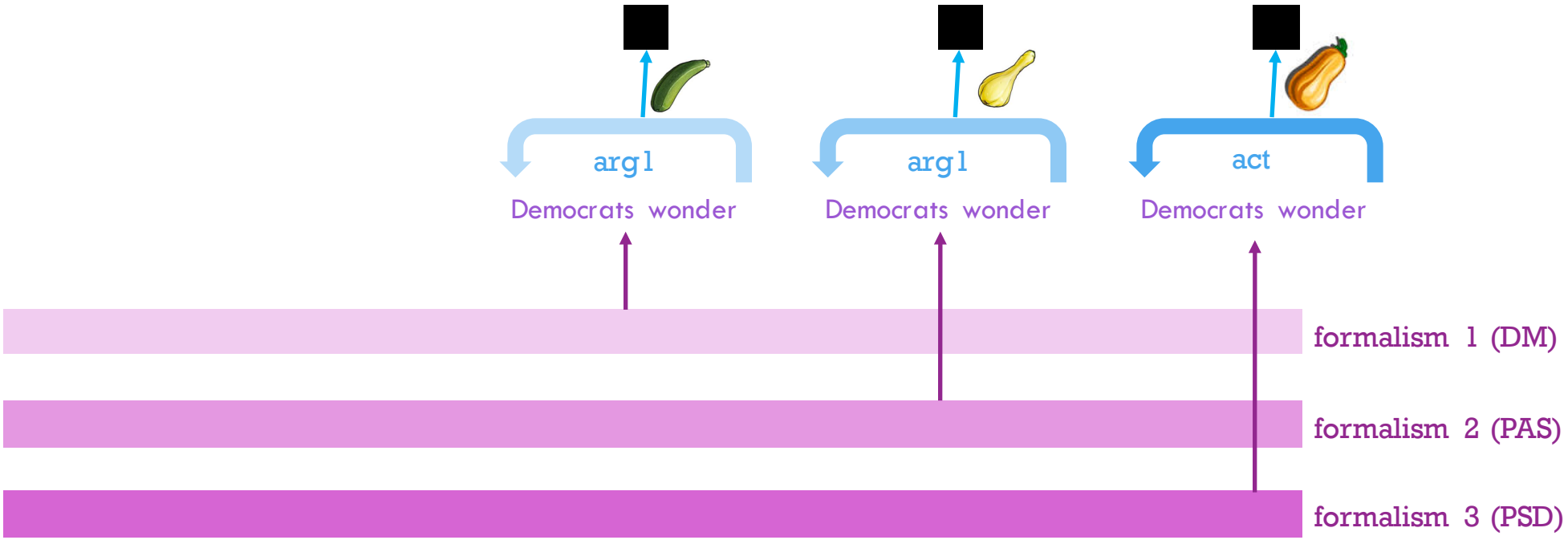
parts: labeled bilexical dependencies

biLSTM (contextualized word vectors)

When Democrats wonder why there is so much resentment of Clinton, they don't need …

words

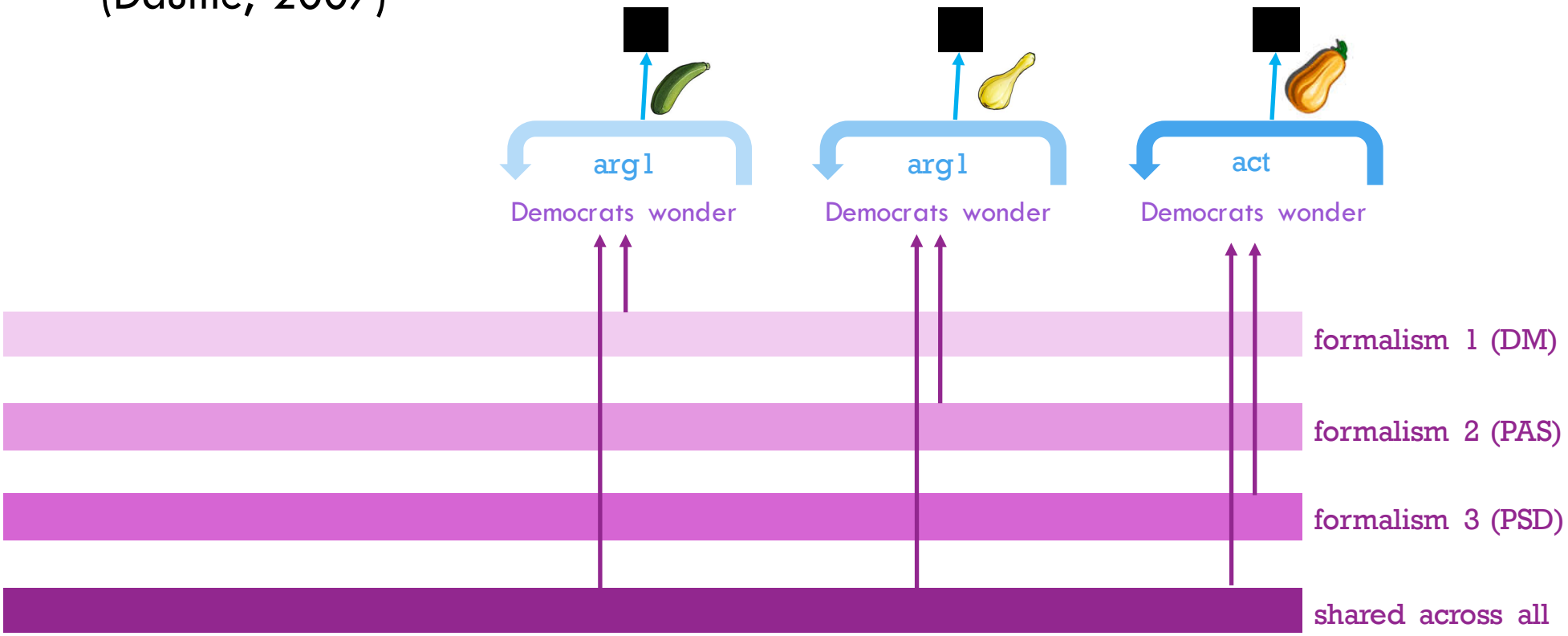
Three Formalisms, Three Separate Parsers



When Democrats wonder why there is so much resentment of Clinton, they don't need …

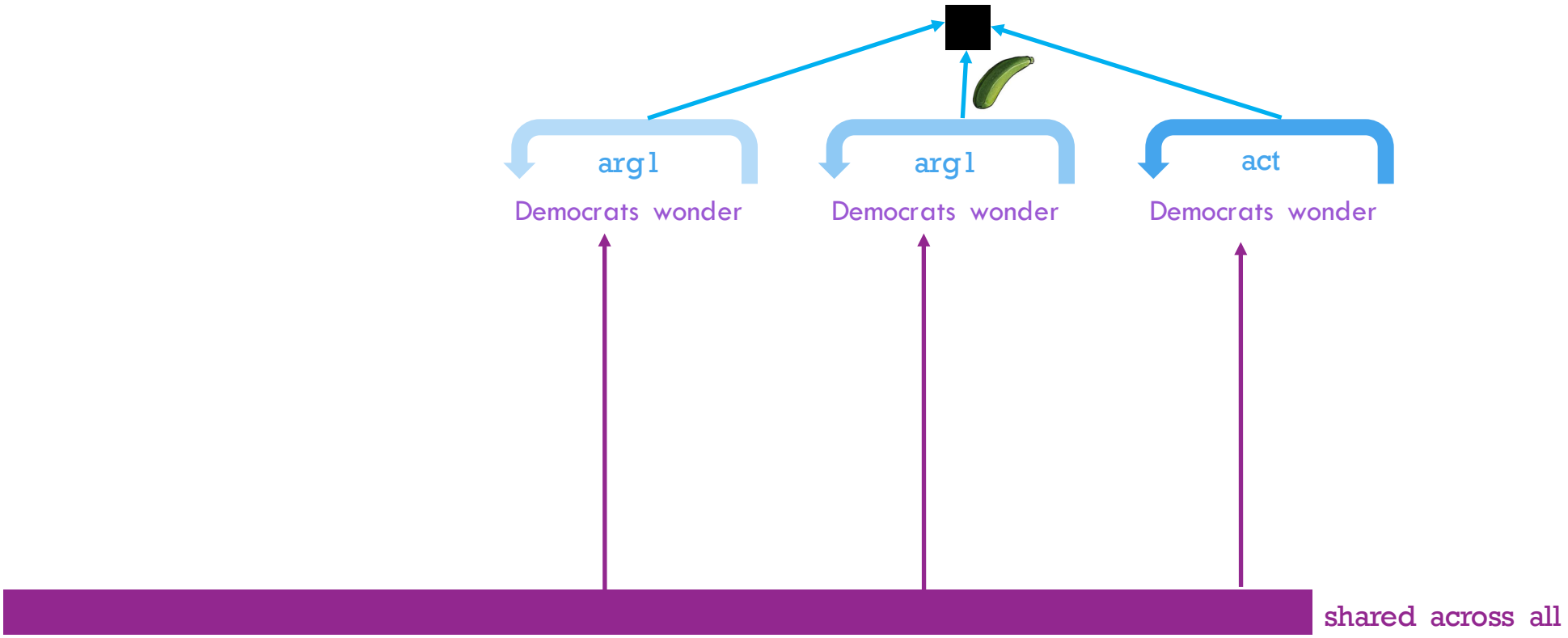
Shared Input Representations

(Daumé, 2007)



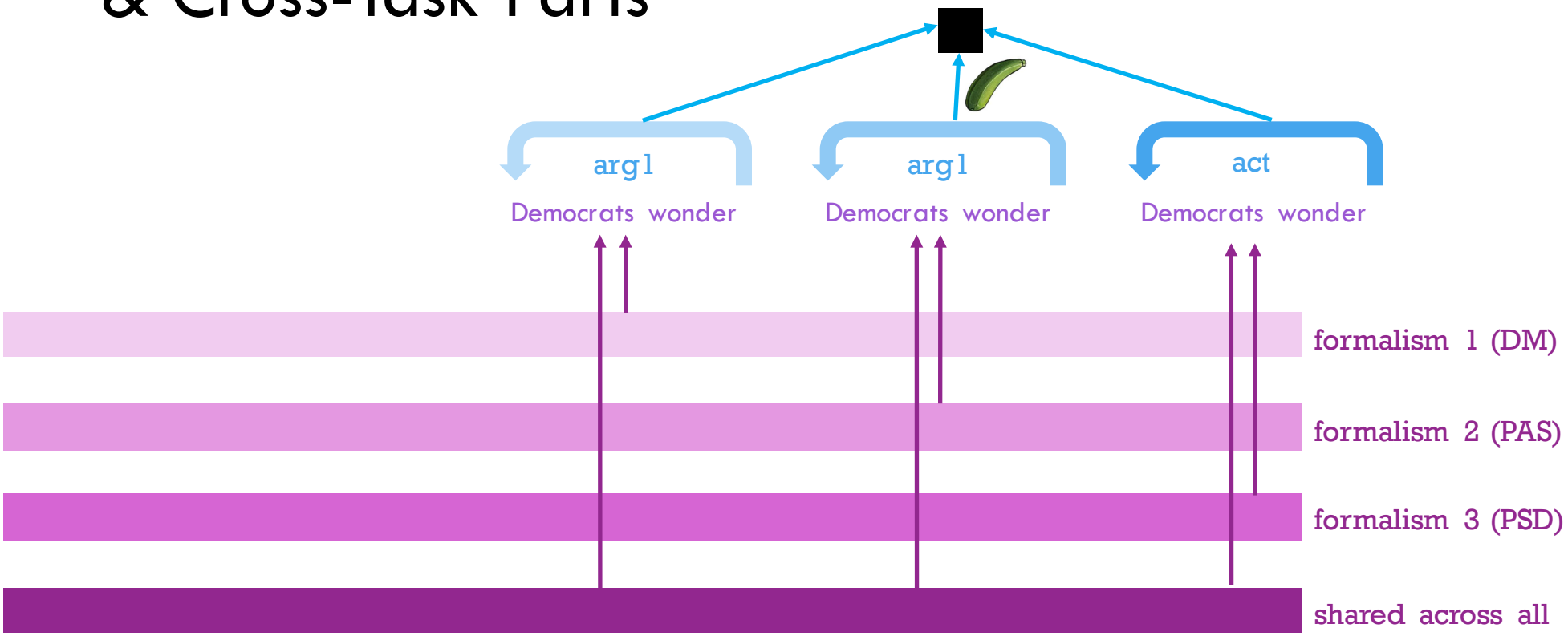
When Democrats wonder why there is so much resentment of Clinton, they don't need …

Cross-Task Parts



When Democrats wonder why there is so much resentment of Clinton, they don't need …

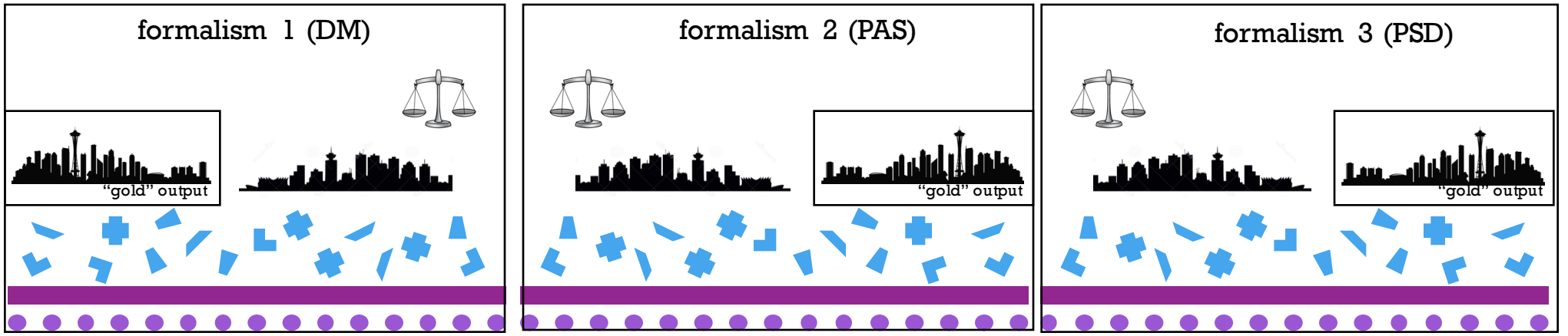
Both: Shared Input Representations & Cross-Task Parts



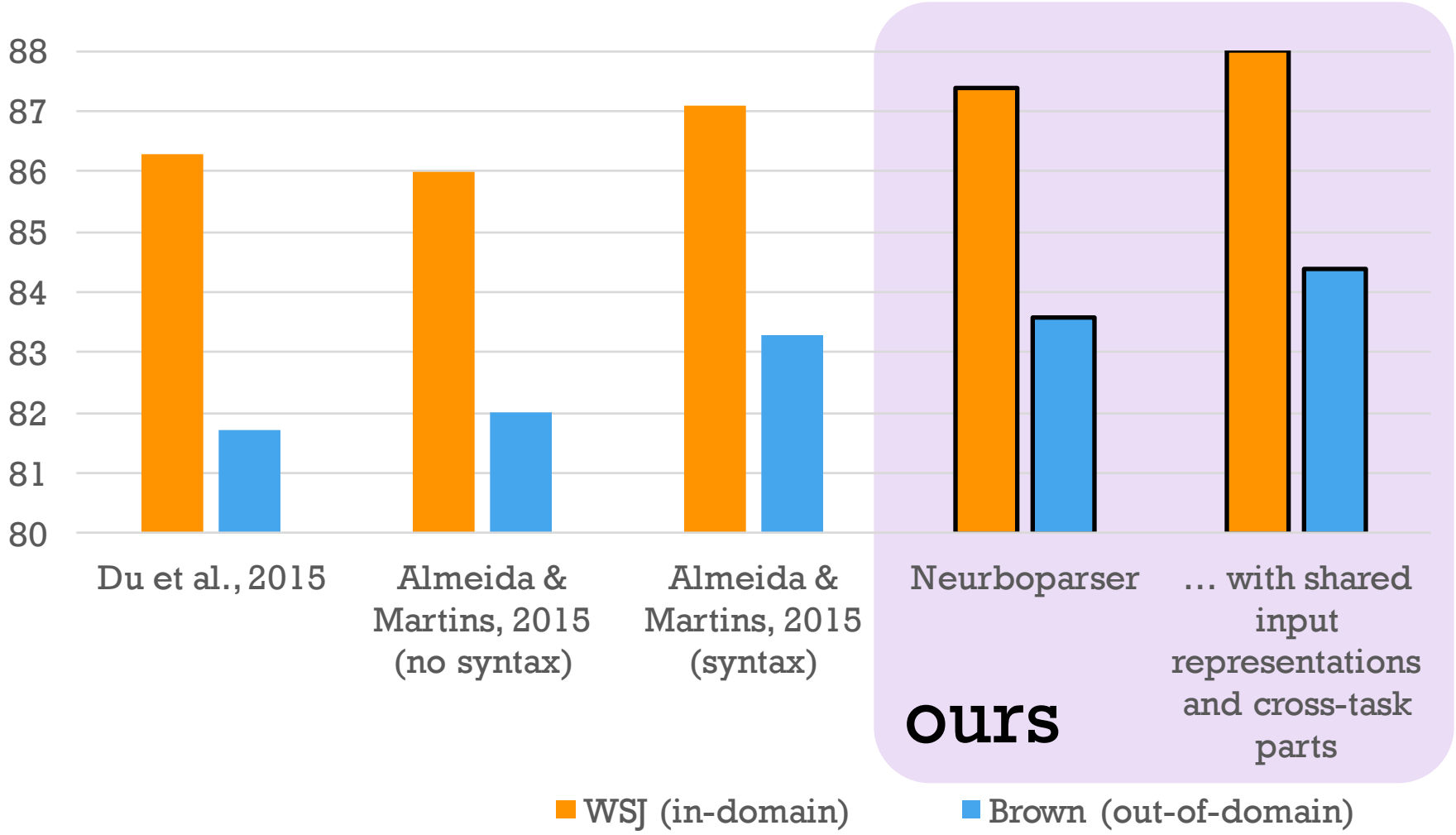
When Democrats wonder why there is so much resentment of Clinton, they don't need …

Multitask Learning: Many Possibilities

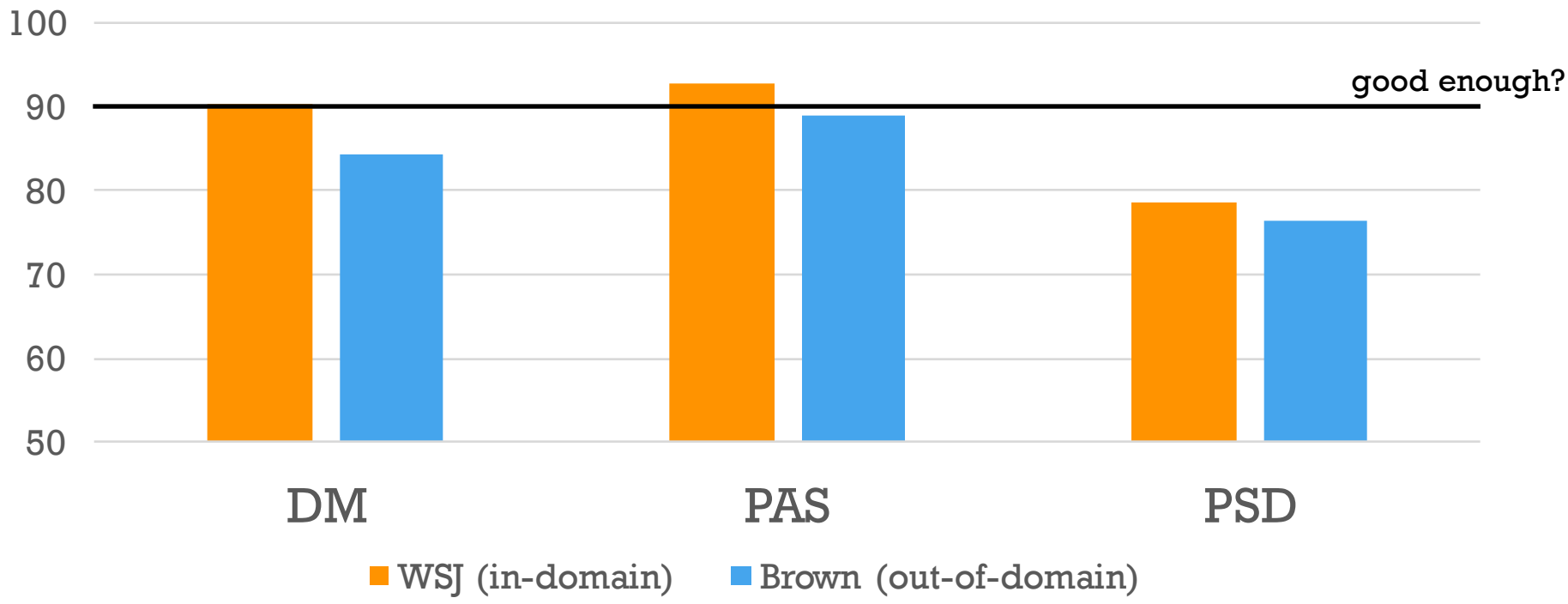
- Shared input representations, parts? *Which* parts?
- Joint decoding?
- Overlapping training data?
- Scaffold tasks?



F_1 averaged on three semantic dependency parsing formalisms, SemEval 2015 test set.

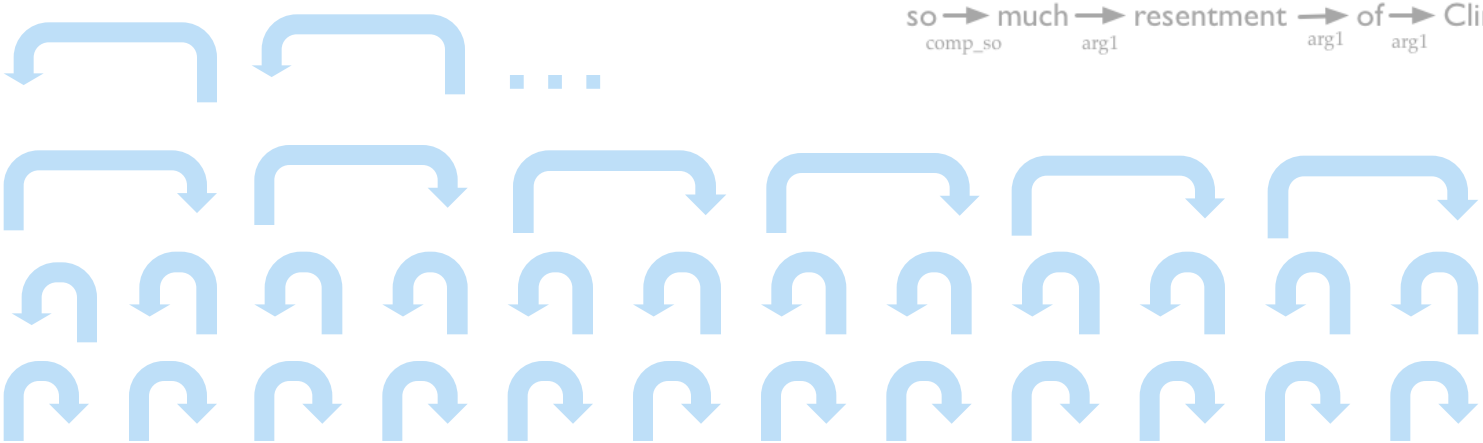


Neurboparser F_1 on three semantic dependency parsing formalisms, SemEval 2015 test set.



Bias?

cross-formalism sharing



training objective: structured hinge loss

output: labeled semantic dependency graph with constraints

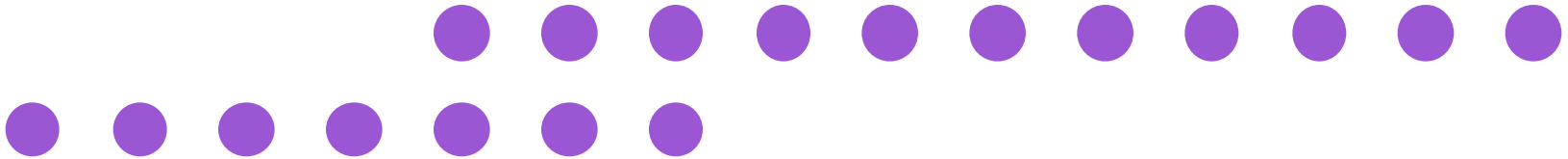
parts: labeled bilexical dependencies

biLSTM (contextualized word vectors)

words

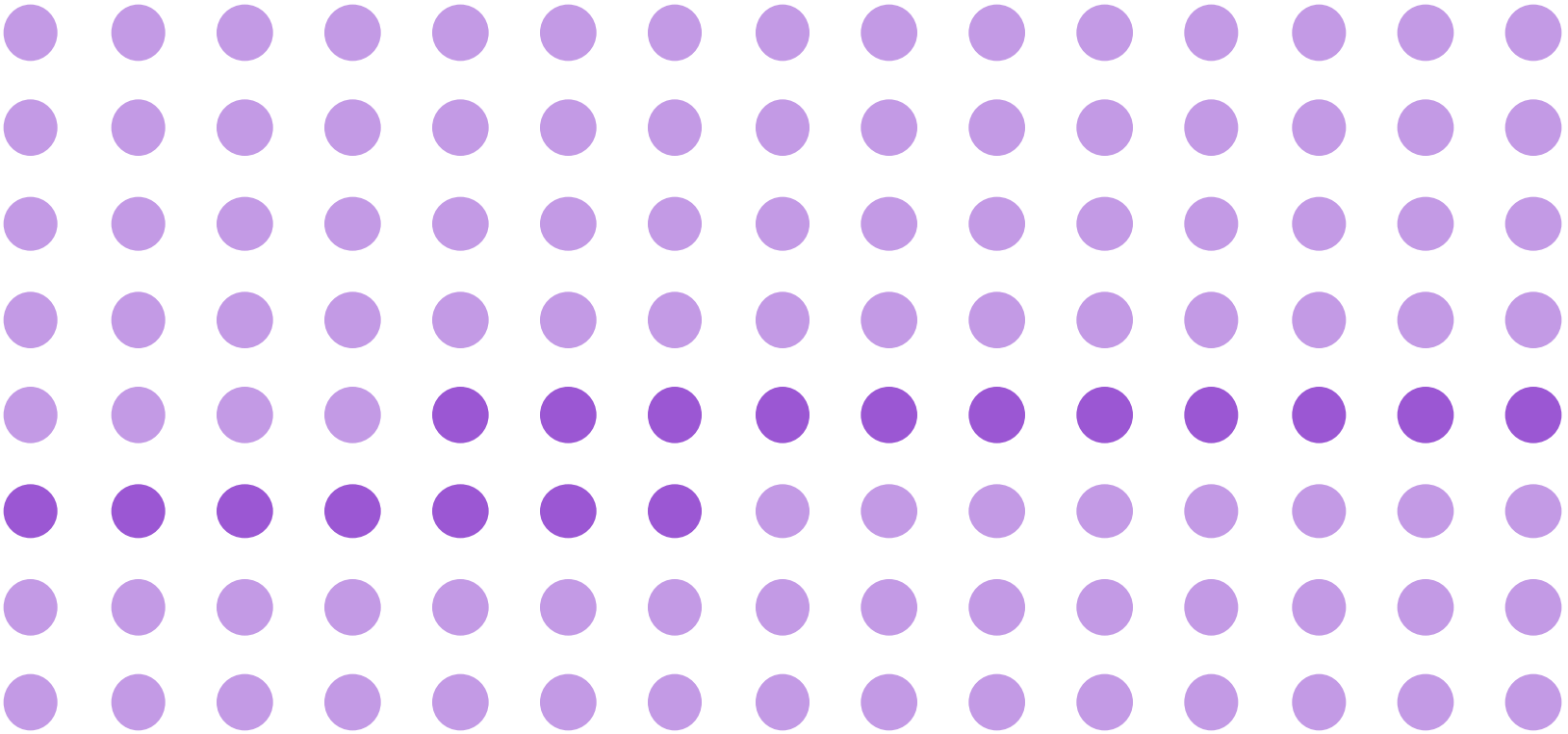
When Democrats wonder why there is so much resentment of Clinton, they don't need …

Text

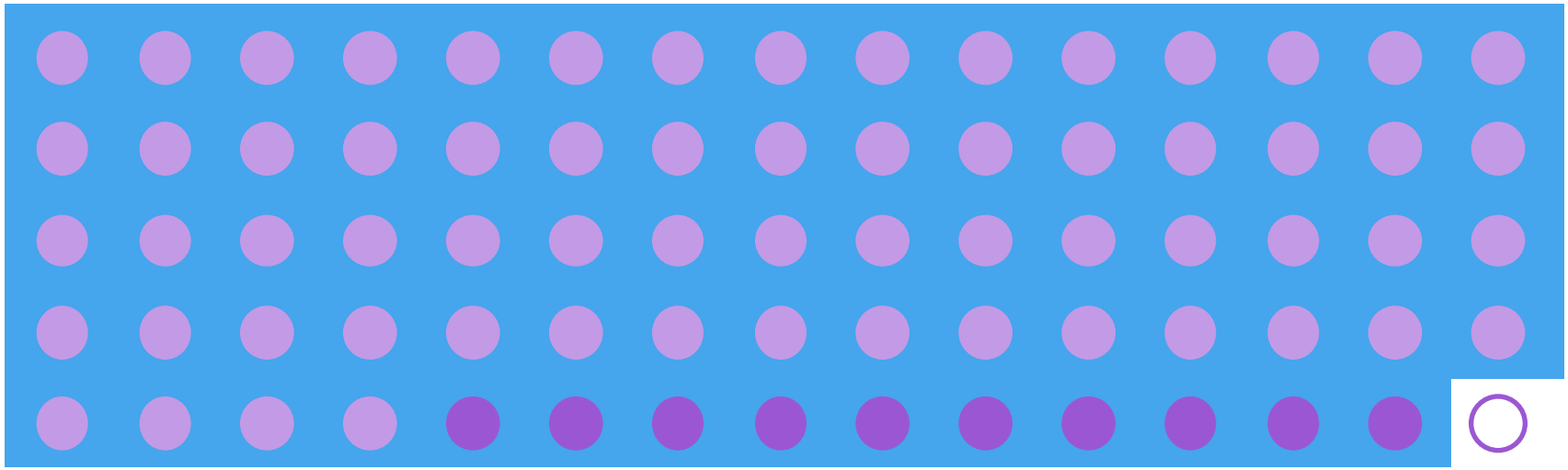


Text ≠ Sentences

larger context



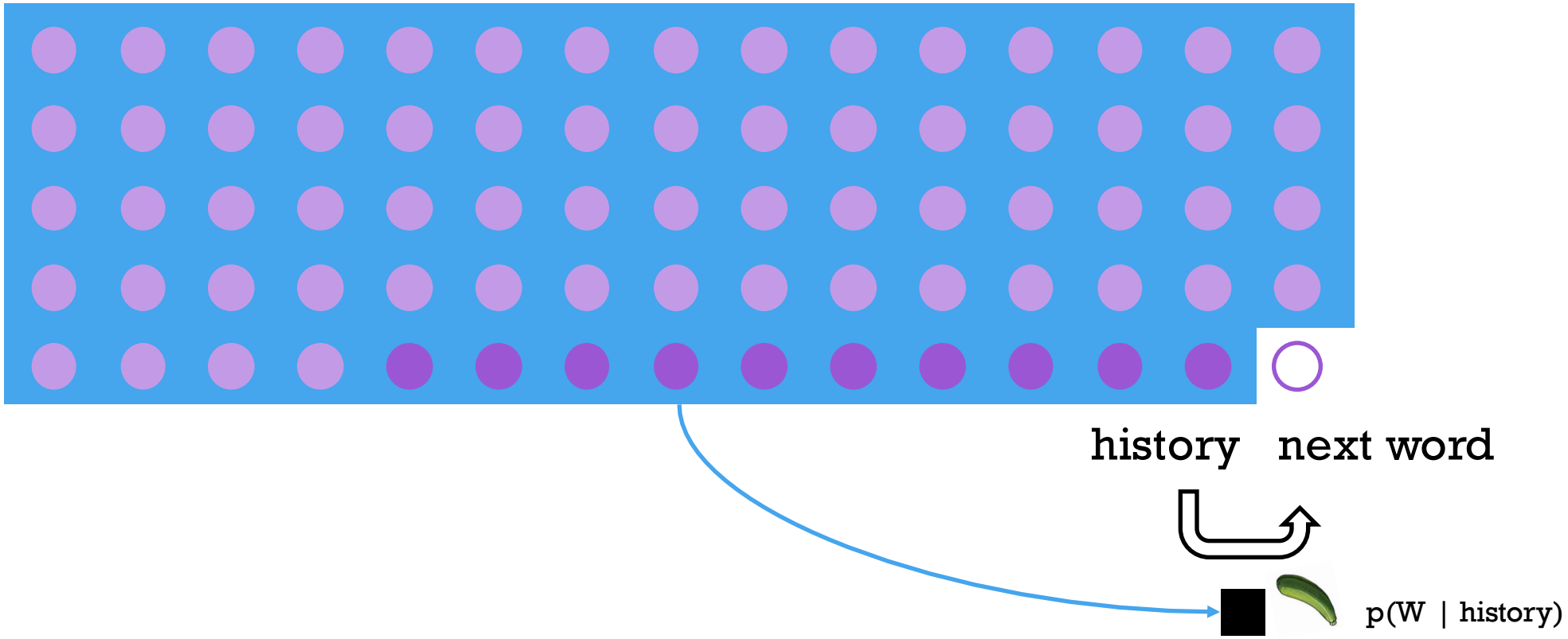
Generative Language Models

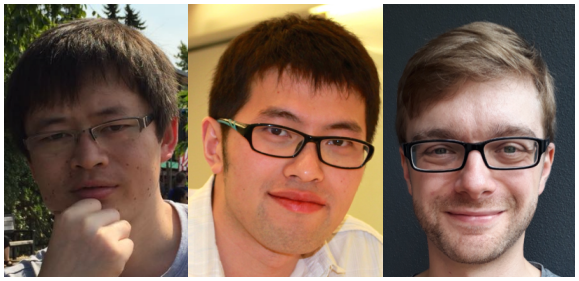


history next word

p(W | history)

Generative Language Models

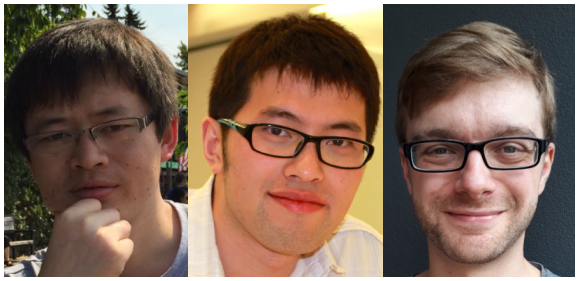




Entity Language Model

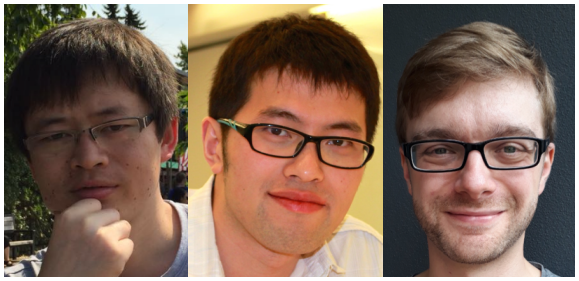
(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)

When Democrats wonder why there is so much resentment of ◯



Entity Language Model

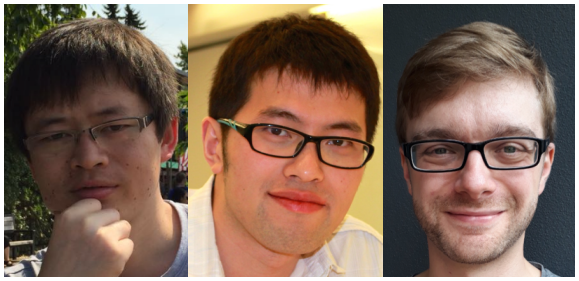
(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)



Entity Language Model

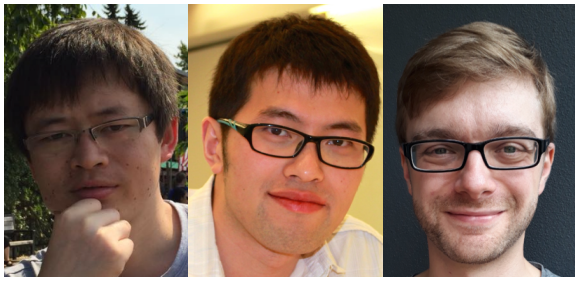
(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)

1. new entity with new vector
2. mention word will be "Clinton"



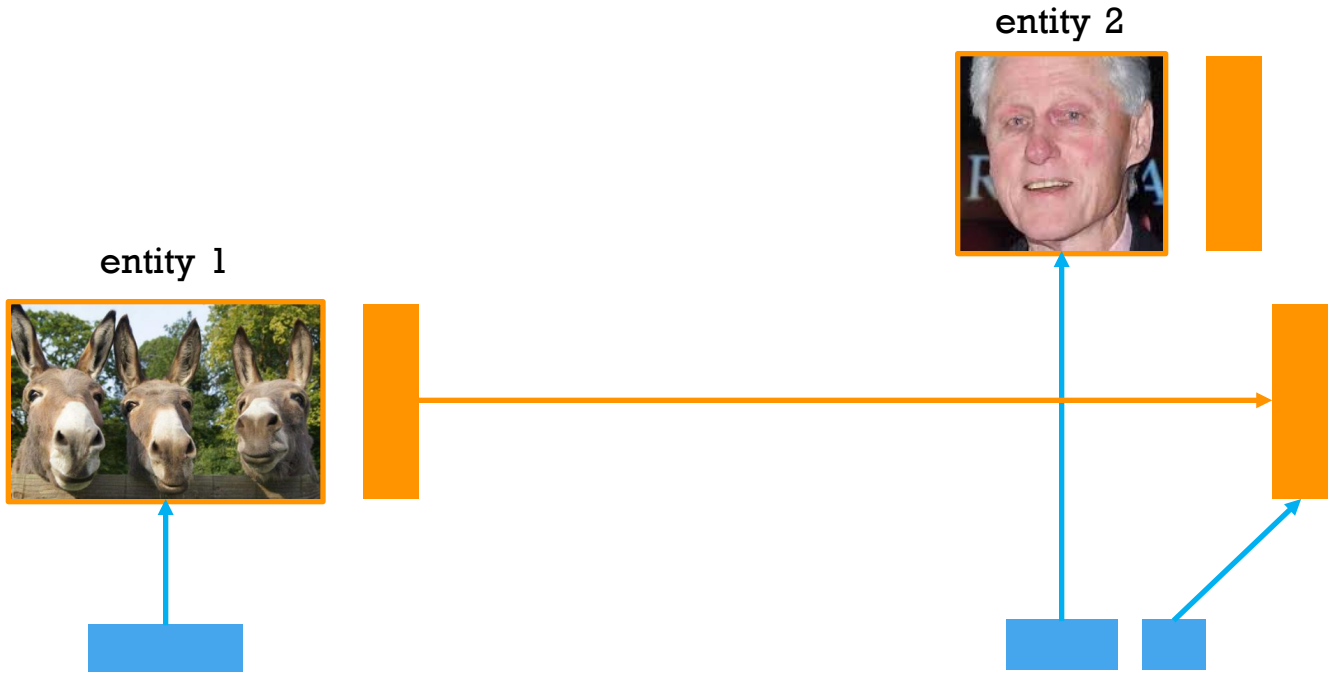
Entity Language Model

(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)

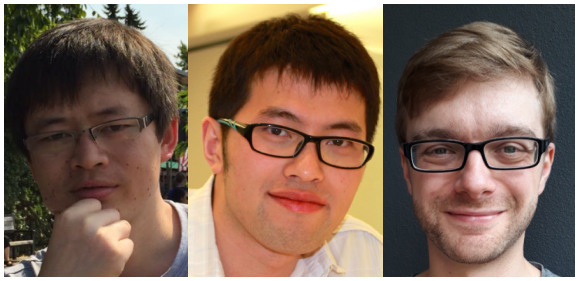


Entity Language Model

(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)

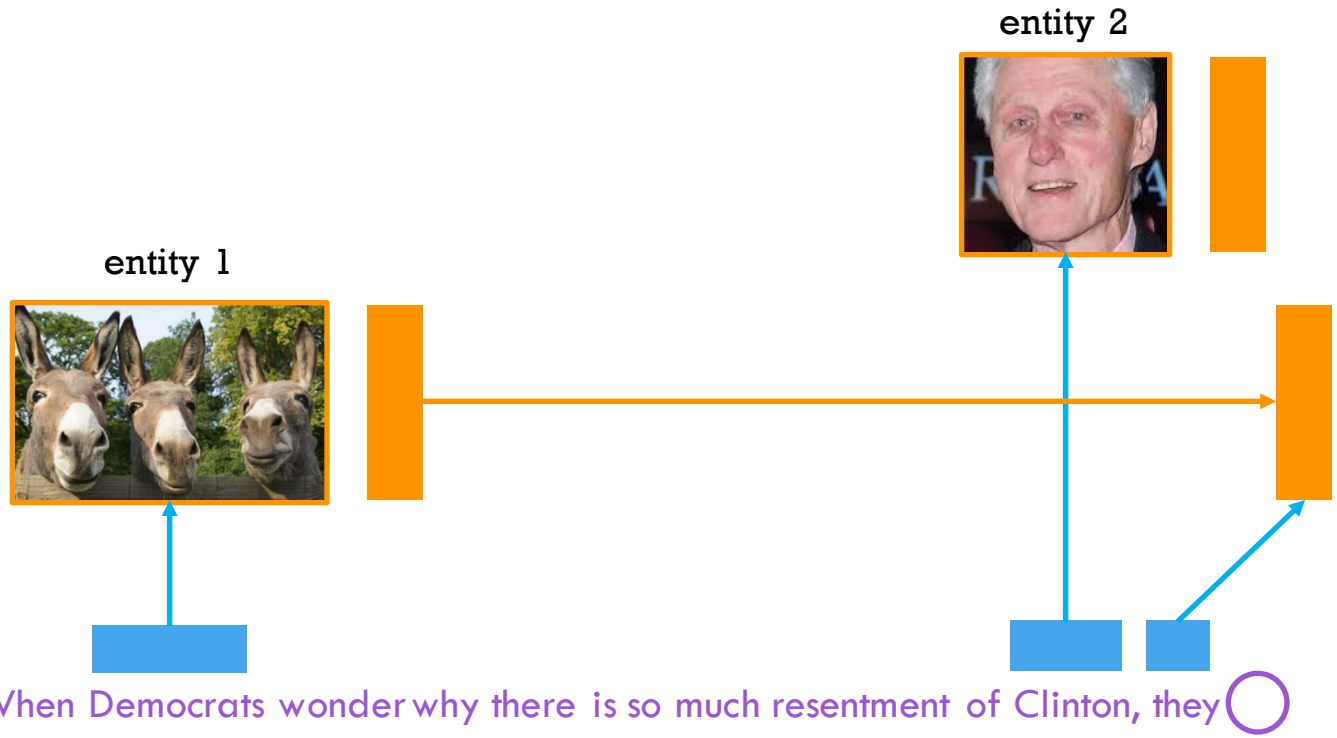


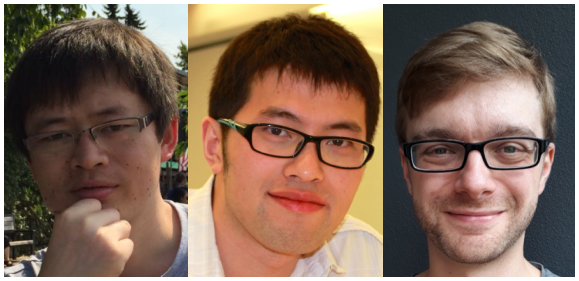
1. coreferent of entity 1 (previously known as "Democrats")
2. mention word will be "they"
3. embedding of entity 1 will be updated



Entity Language Model

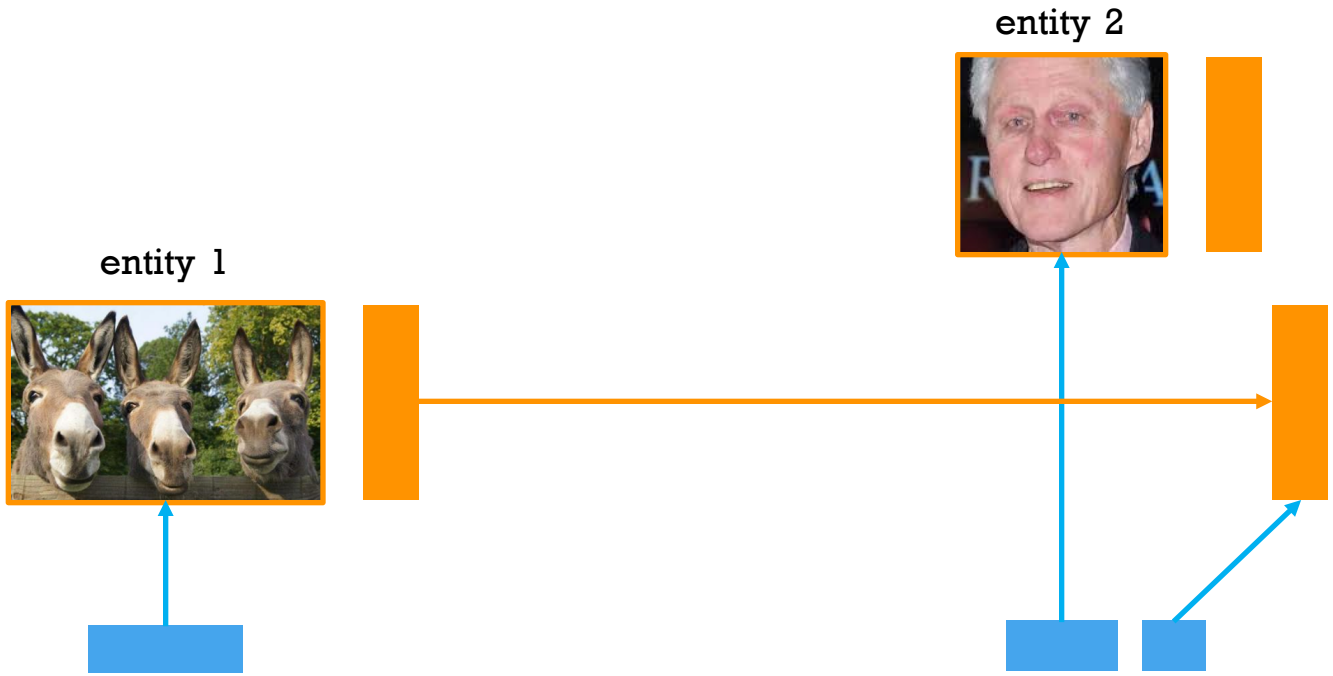
(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)



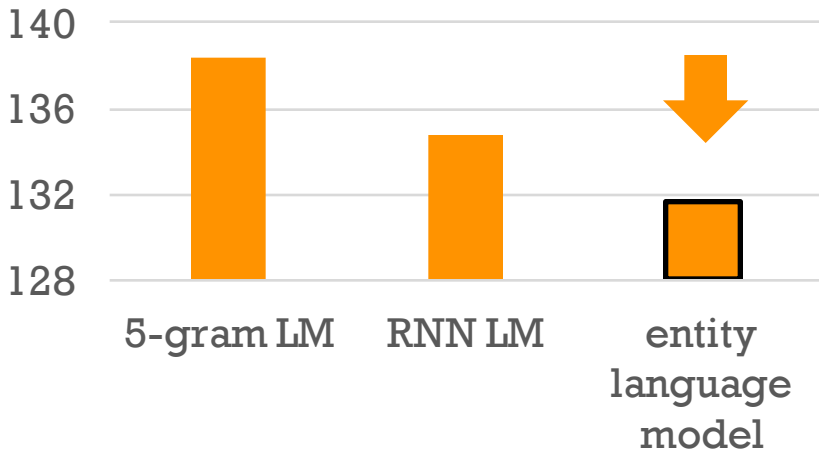


Entity Language Model

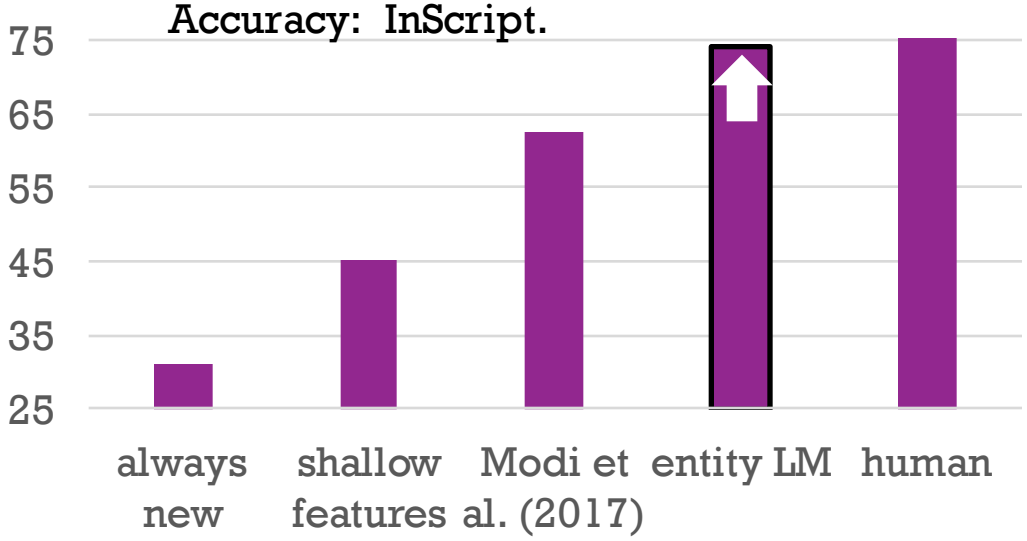
(Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, N.A.S., EMNLP 2017)



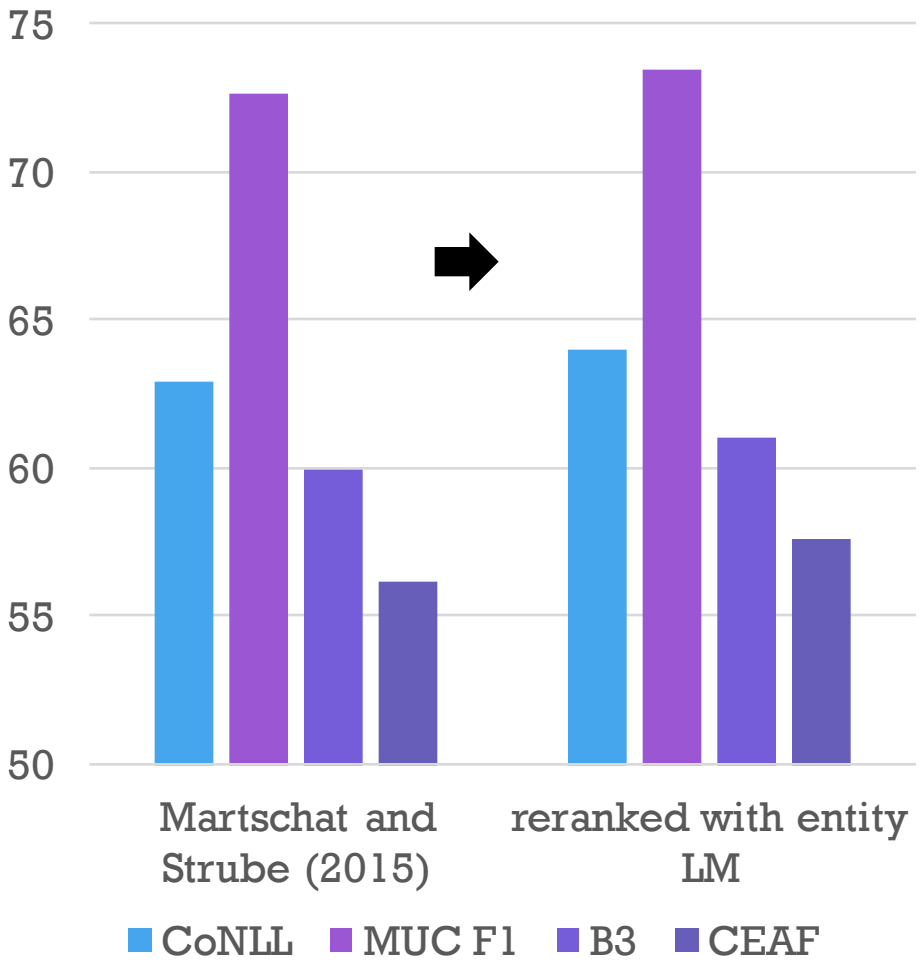
Perplexity on CoNLL 2012 test set.



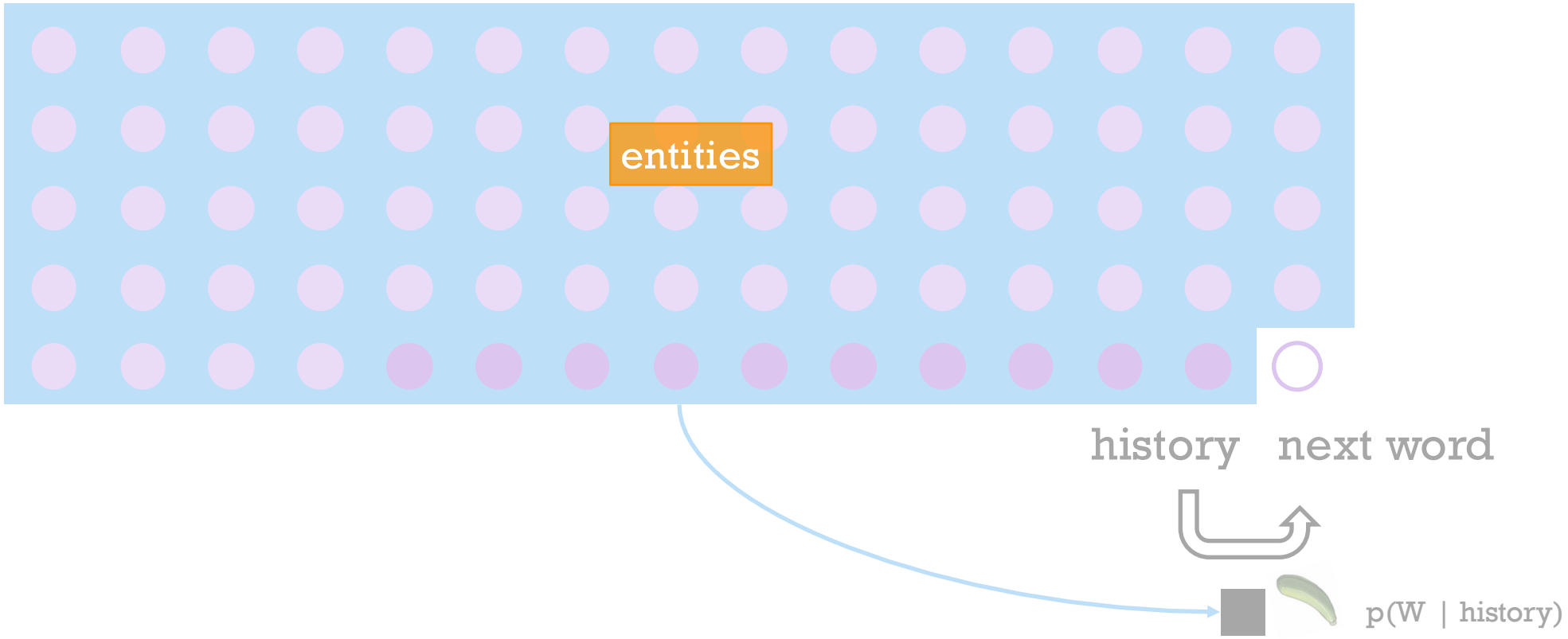
Accuracy: InScript.



CoNLL 2012 coreference evaluation.



Bias?



Bias in the Future?

- Linguistic **scaffold tasks.**

Bias in the Future?

- Linguistic **scaffold tasks.**
- Language is by and about **people.**

Bias in the Future?

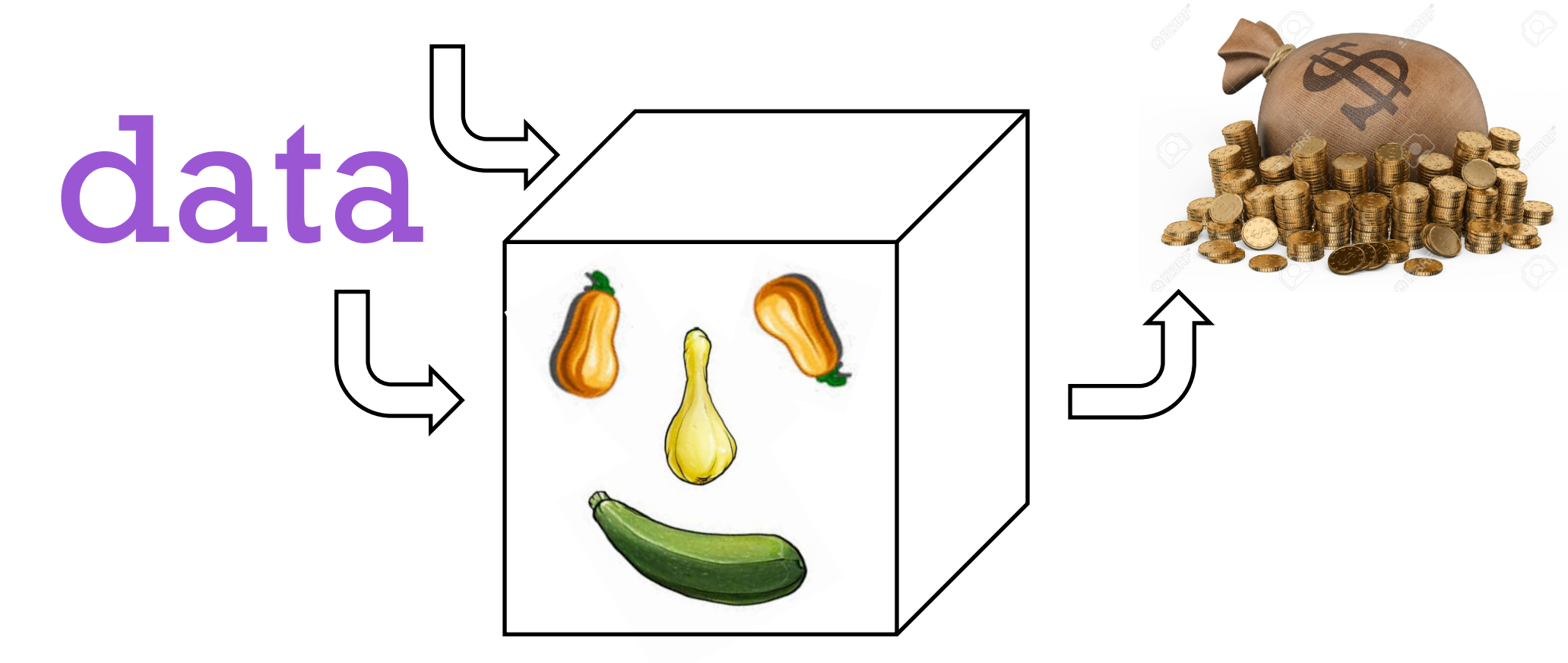
- Linguistic **scaffold tasks.**
- Language is by and about **people.**
- NLP is needed when texts are **costly to read.**

Bias in the Future?

- Linguistic **scaffold tasks.**
- Language is by and about **people.**
- NLP is needed when texts are **costly to read.**
- **Polyglot** learning.

bias

data



Thank you!