Microsoft Research

Faculty Summit

2014 15TH ANNUAL

Microsoft Research Faculty Summit 2014 15TH ANNUAL

Towards Understanding

Christopher Manning
Stanford University

1980s Natural Language Processing

VP →{ V (NP:(↑ OBJ)=↓ (NP:(↑ OBJ2)=↓))
(XP:(↑ XCOMP)=↓)
|@(COORD VP VP)}.

salmon N IRR @(CN SALMON)
(↑ PERSON)=3
{ (↑ NUM)=SG|(↑ NUM)=PL}.

$$\begin{bmatrix} \text{SUBJ} & \begin{bmatrix}\text{PRED} & \text{“I”}\end{bmatrix} \\ \text{OBJ}_{\text{go}} & \begin{bmatrix}\text{PRED} & \text{“you”}\end{bmatrix} \\ \text{PRED} & \text{‘perfect}\langle\text{XCOMP}\rangle\text{SUBJ’} \\ \text{XCOMP} & \begin{bmatrix} \text{SUBJ} & \text{—} \\ \text{PRED} & \text{‘bring}\langle\text{SUBJ, OBJ, OBJ}_{\text{go}}\rangle\text{’} \\ \text{OBJ} & \begin{bmatrix}\text{PRED} & \text{“candy”}\end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Learning language

WRB VBZ DT NN VB TO VB DT
How does a project get to be a
NN JJ . : CD NN IN DT NN .
year late ? ... One day at a time .

$$P(\text{late} \mid \text{a, year}) = 0.0087$$

$$P(\text{NN} \mid \text{DT, a, project}) = 0.9$$

The traditional word representation

motel

[0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]

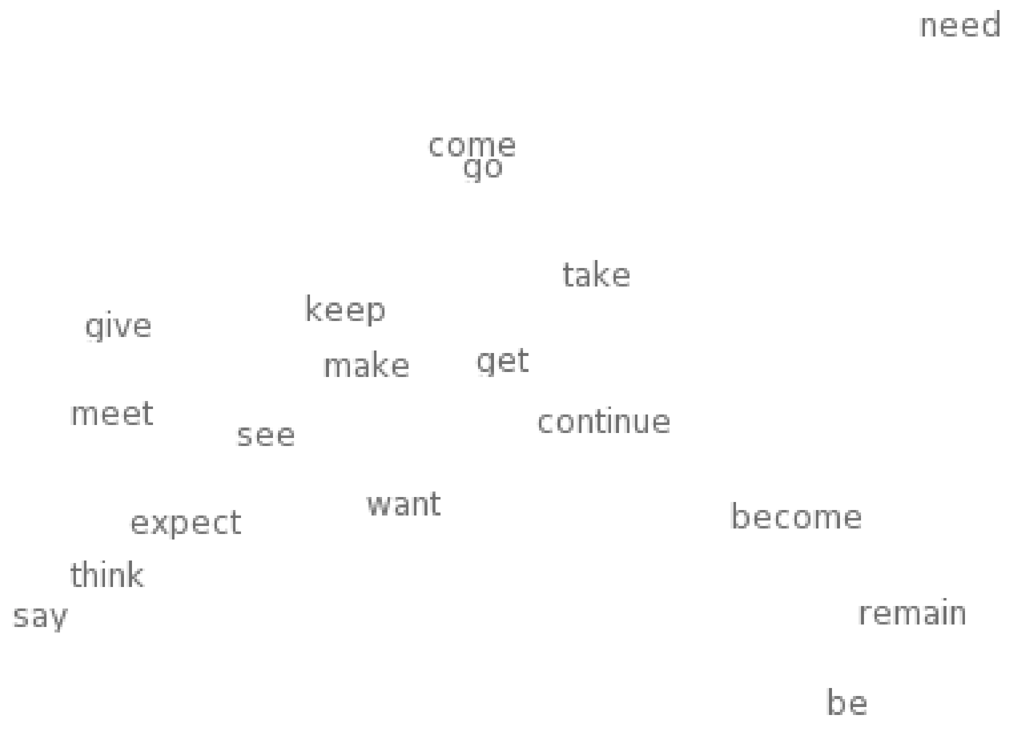
Dimensionality: 50K (small domain – speech/PTB) – 13M (web – Google 1T)

motel [0 0 0 0 0 0 0 0 0 0 1 0 0 0 0] AND
hotel [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0] = 0

Word distributions → word representations

Through corpus **linguistics**, large chunks
the study of language and **linguistics**.
The field of **linguistics** is concerned
Written like a **linguistics** text book
Phonology is the branch of **linguistics** that

$$\textit{linguistics} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \\ 0.487 \end{pmatrix}$$



[Bengio et al. 2003, Collobert & Weston 2008, Turian 2010, Mikolov 2013, etc.]

Encoding meaning in vector differences

[Pennington et al., to appear 2014]

Crucial insight: Ratios of co-occurrence probabilities can encode meaning components

	x = solid	x = gas	x = water	x = random
$P(x \mid \text{ice})$	large	small	large	small
$P(x \mid \text{steam})$	small	large	large	small
$\frac{P(x \mid \text{ice})}{P(x \mid \text{steam})}$	large	small	~1	~1

Encoding meaning in vector differences

[Pennington et al., to appear 2014]

Crucial insight: Ratios of co-occurrence probabilities can encode meaning components

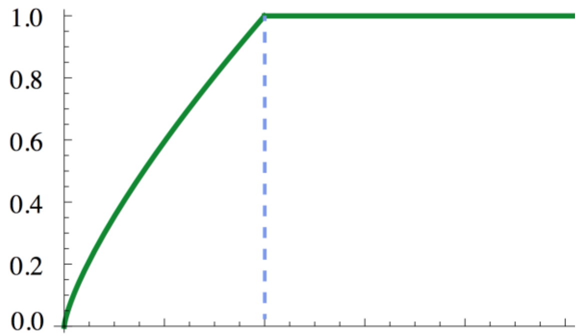
	x = solid	x = gas	x = water	x = fashion
$P(x\|\text{ice})$	1.9 x 10^{-4}	6.6 x 10^{-5}	3.0 x 10^{-3}	1.7 x 10^{-5}
$P(x\|\text{steam})$	2.2 x 10^{-5}	7.8 x 10^{-4}	2.2 x 10^{-3}	1.8 x 10^{-5}
$\frac{P(x\|\text{ice})}{P(x\|\text{steam})}$	8.9	8.5 x 10^{-2}	1.36	0.96

GloVe: A new model for learning word representations [Pennington et al., to appear 2014]

$$w_i \cdot w_j = \log P(i|j)$$

$$w_x \cdot (w_a - w_b) = \log \frac{P(x|a)}{P(x|b)}$$

$$J = \frac{1}{2} \sum_{ij} f(P_{ij}) \left(w_i \cdot \tilde{w}_j - \log P_{ij}\right)^2 \qquad f \sim$$



Word similarities

Nearest words to frog:

1. frogs
2. toad
3. litoria
4. leptodactylidae
5. rana
6. lizard
7. eleutherodactylus

litoria

leptodactylidae

rana

eleutherodactylus

Word analogy task [Mikolov, Yih & Zweig 2013a]

Model	Dimensions	Corpus size	Performance (Syn + Sem)
CBOW (Mikolov et al. 2013b)	300	1.6 billion	36.1

Machine translation with bilingual neural language models

[Devlin et al., ACL 2014]

S: 我 [6 就] [4 取] [3 钱] [5 给] [7 了] 她们

i *will* *get* *money* *to* perf. *them*

T: [2 i] [1 will] [0 get] [the] money to them

P(the | get, will, i, 钱, 取, 给, 就, 了)

Machine translation with bilingual neural language models [Devlin et al., ACL 2014]

NIST 2012 Open MT Arabic Results

	Arabic
1st Place (BBN)	49.5
2nd Place	47.5
…	…
9th Place	44.0
10th Place	41.2

NNJM on best system

	Arabic
Previous best BBN system	49.8
+ NNJM	52.8

+ 3.0 BLEU

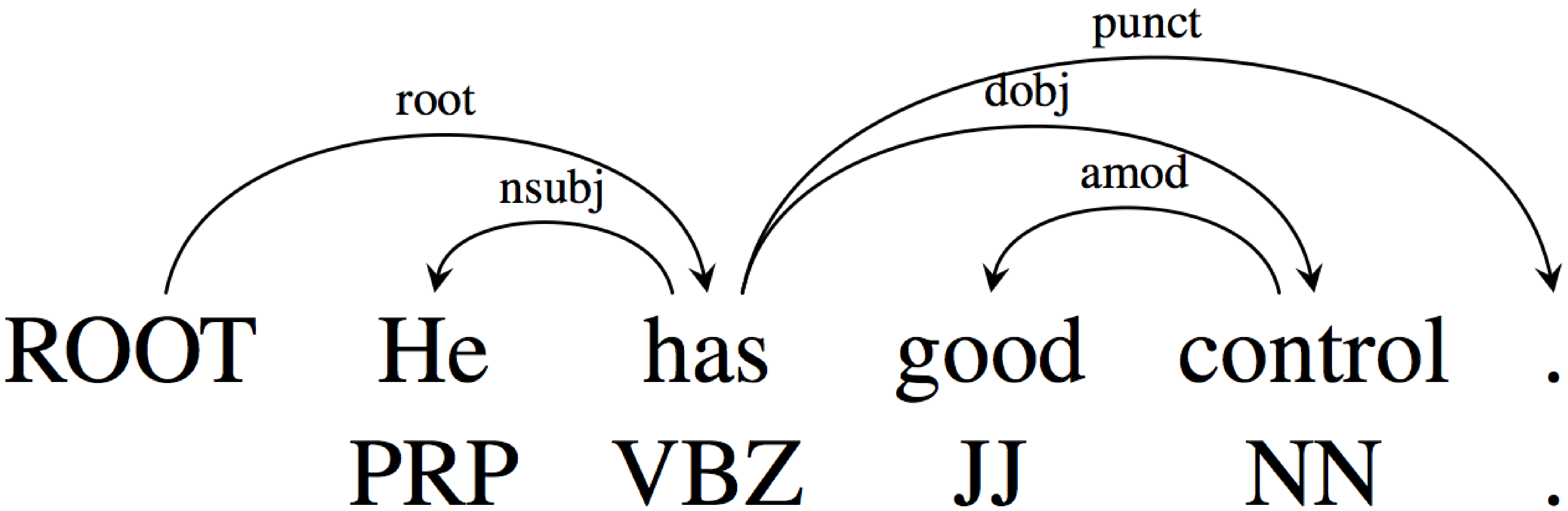
NNJM on "Baseline"

	Arabic
"Baseline Hiero"	43.4
+ NNJM	49.7

+ 6.3 BLEU

"Baseline Hiero" Features: (1) Rule probs, (2) lexical smoothing, (3) KN LM, (4) word penalty, (5) concat penalty

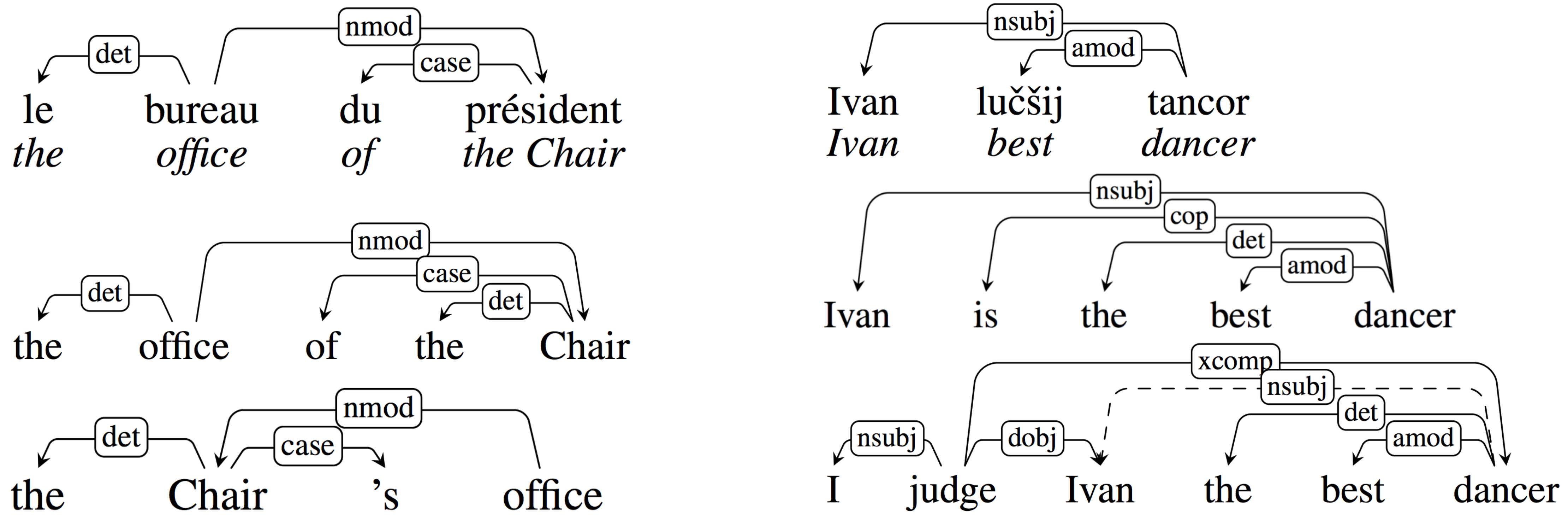
Sentence structure: Dependency parsing



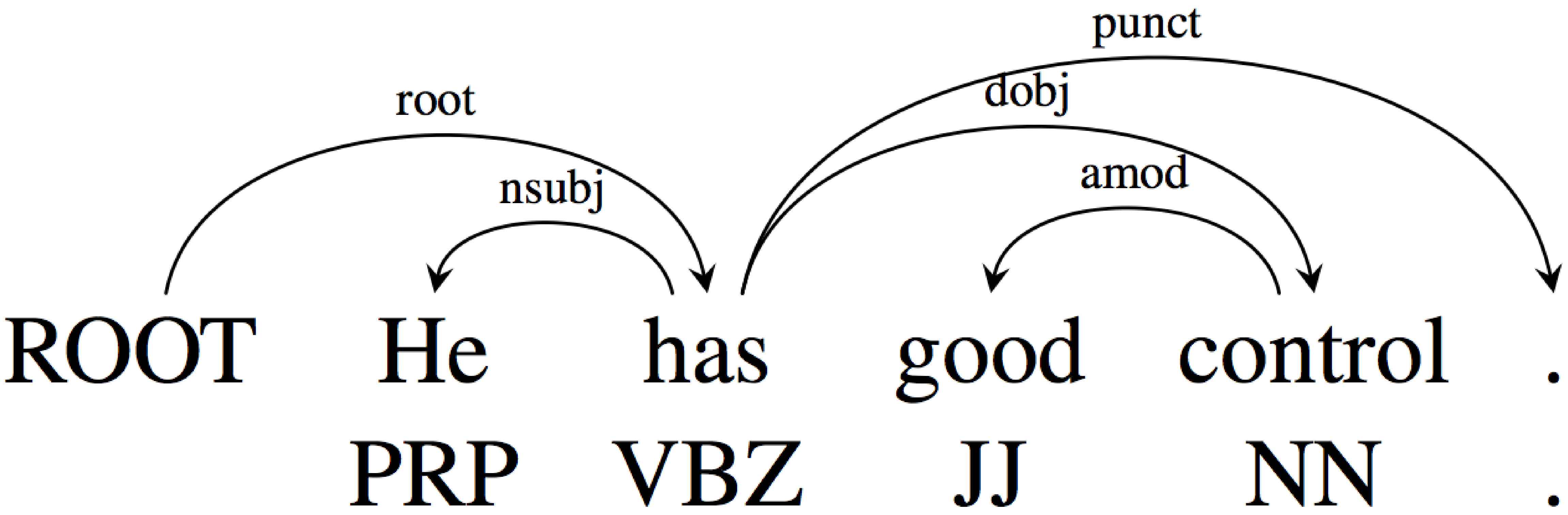
Universal Stanford Dependencies

[de Marneffe et al., LREC 2014]

A common dependency representation and label set applicable across languages – http://universaldependencies.github.io/docs/



Sentence structure: Dependency parsing



Deep Learning Dependency Parser

[Chen & Manning, forthcoming 2014]

Softmax layer:

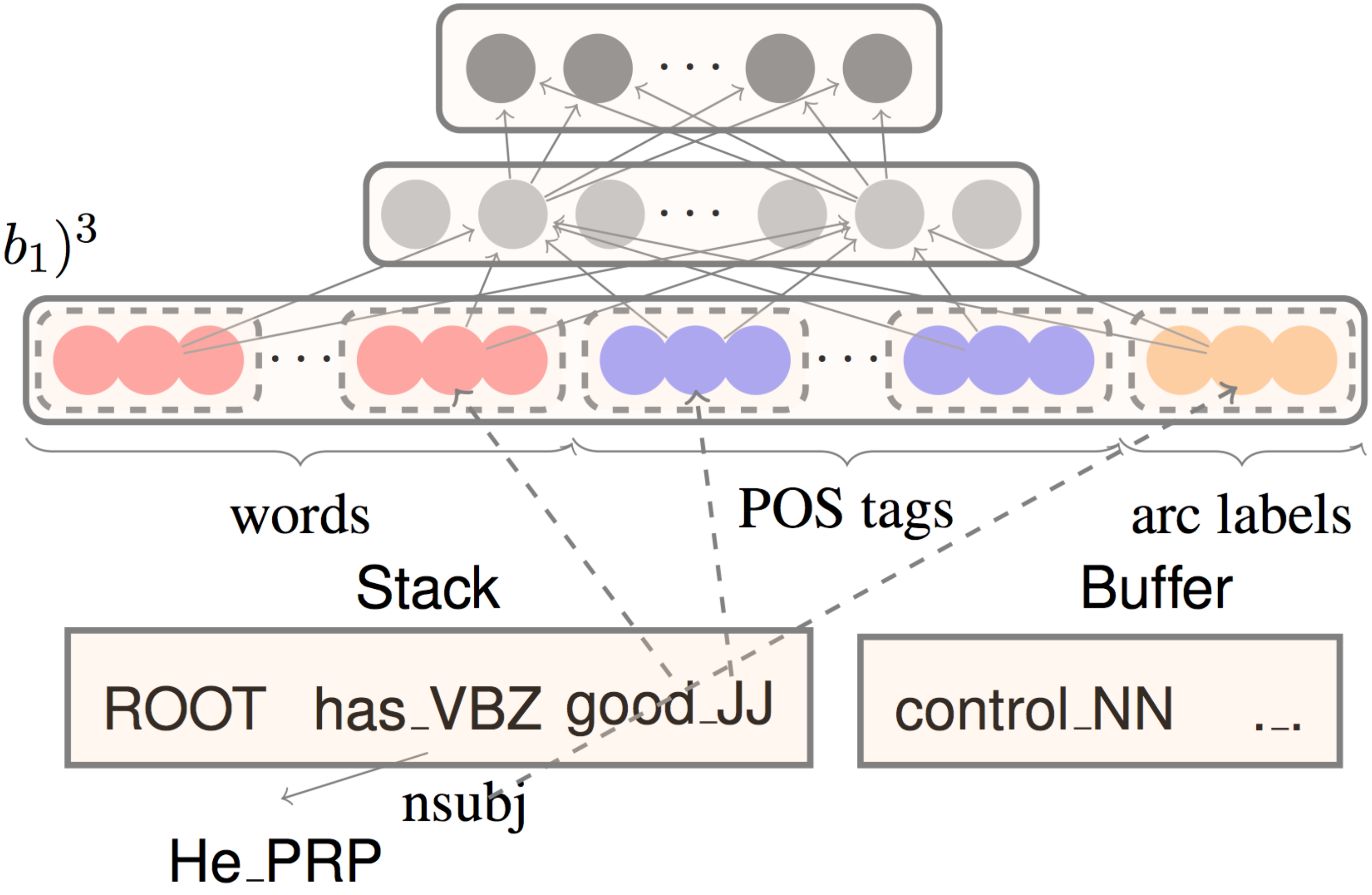
$$p = \texttt{softmax}(W_2 h)$$

Hidden layer:

$$h = (W_1^w x^w + W_1^t x^t + W_1^l x^l + b_1)^3$$

Input layer: $[x^w, x^t, x^l]$

Configuration



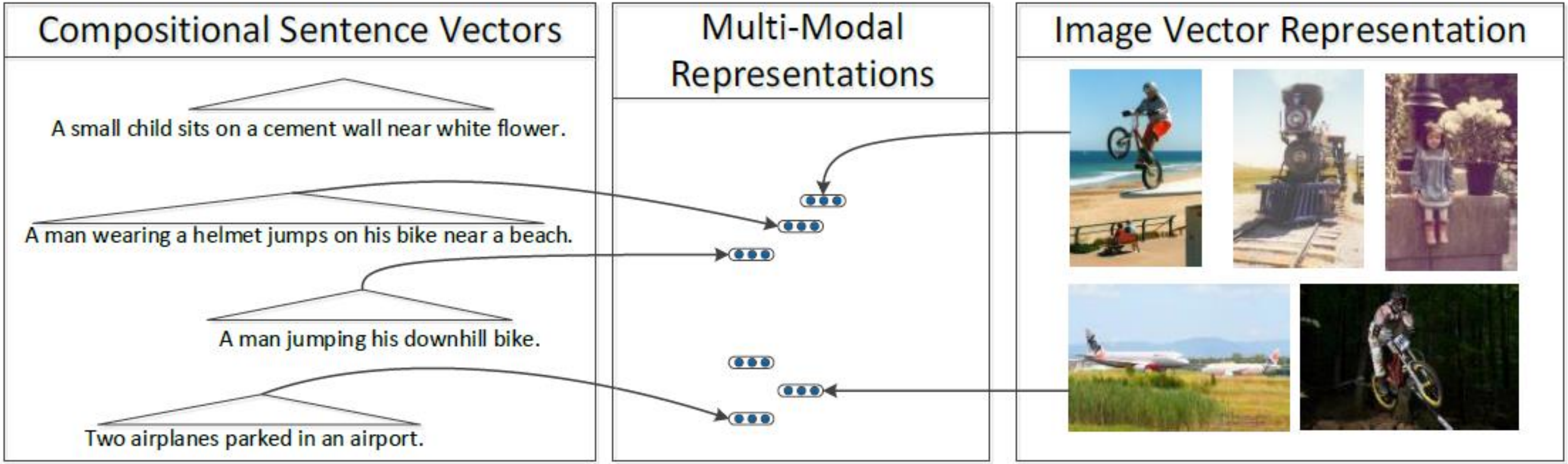
Deep Learning Dependency Parser

[Chen & Manning, forthcoming 2014]

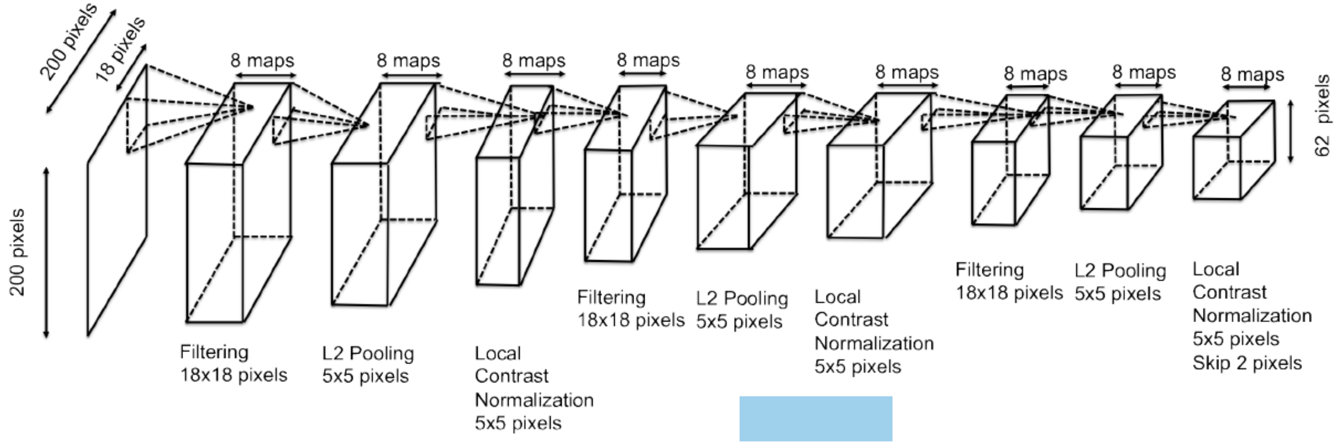
Parser type	Parser	LAS (Label & Attach)	Sentences / sec
Transition-based	MaltParser (stackproj)	86.9	469
Graph-based	MSTParser	87.6	10
	TurboParser (full)	89.7	8

Grounding language meaning with images

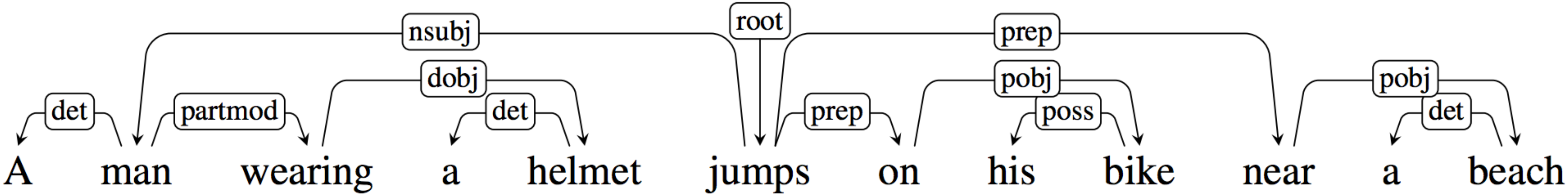
[Socher, Karpathy, Le, Manning & Ng, TACL 2014]



Example dependency tree and image

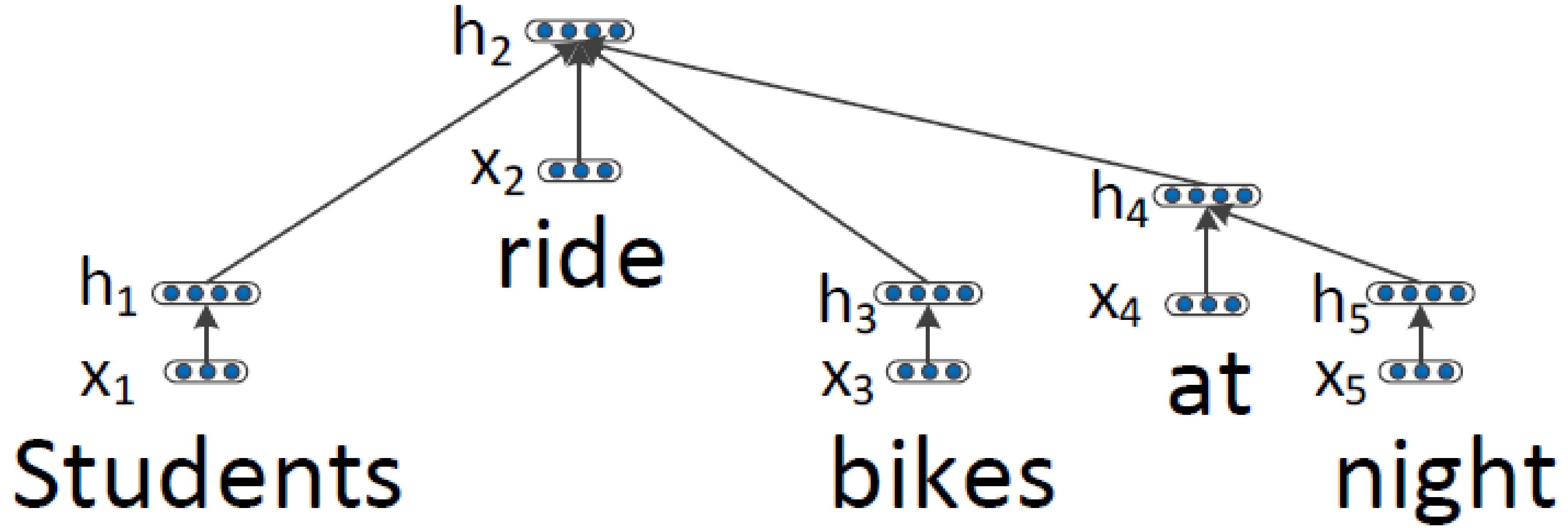


$$h_i = f\left(\frac{1}{\ell(i)}\left(W_v x_i + \sum_{j\in C(i)} \ell(j) W_{\mathrm{dep}(i,j)} h_j\right)\right)$$



Recursive computation of dependency tree

$$h_i = f\left(\frac{1}{\ell(i)}\left(W_v x_i + \sum_{j \in C(i)} \ell(j) W_{\mathrm{dep}\,(i,j)} h_j\right)\right)$$



Evaluation

Data of [Rashtchian, Young, Hodosh & Hockenmaier 2010]

1. A woman and her dog watch the cameraman in their living with wooden floors.
2. A woman sitting on the couch while a black faced dog runs across the floor.
3. A woman wearing a backpack sits on a couch while a small dog runs on the hardwood floor next to her.
4. A women sitting on a sofa while a small Jack Russell walks towards the camera.
5. White and black small dog walks toward the camera while woman sits on couch, desk and computer seen in the background as well as a pillow, teddy bear and moggie toy on the wood floor.

A gray convertible sports car is parked in front of the trees. ✓
A close-up view of the headlights of a blue old-fashioned car. ✗
Black shiny sports car parked on concrete driveway. ✓
Five cows grazing on a patch of grass between two roadways. ✗

A jockey rides a brown and white horse in a dirt corral. ✓
A young woman is riding a Bay hose in a dirt riding-ring. ✗
A white bird pushes a miniature teal shopping cart. ✗
A person rides a brown horse. ✓

1000 images,
5 descriptions each;
used as 800 train,
100 dev, 100 test

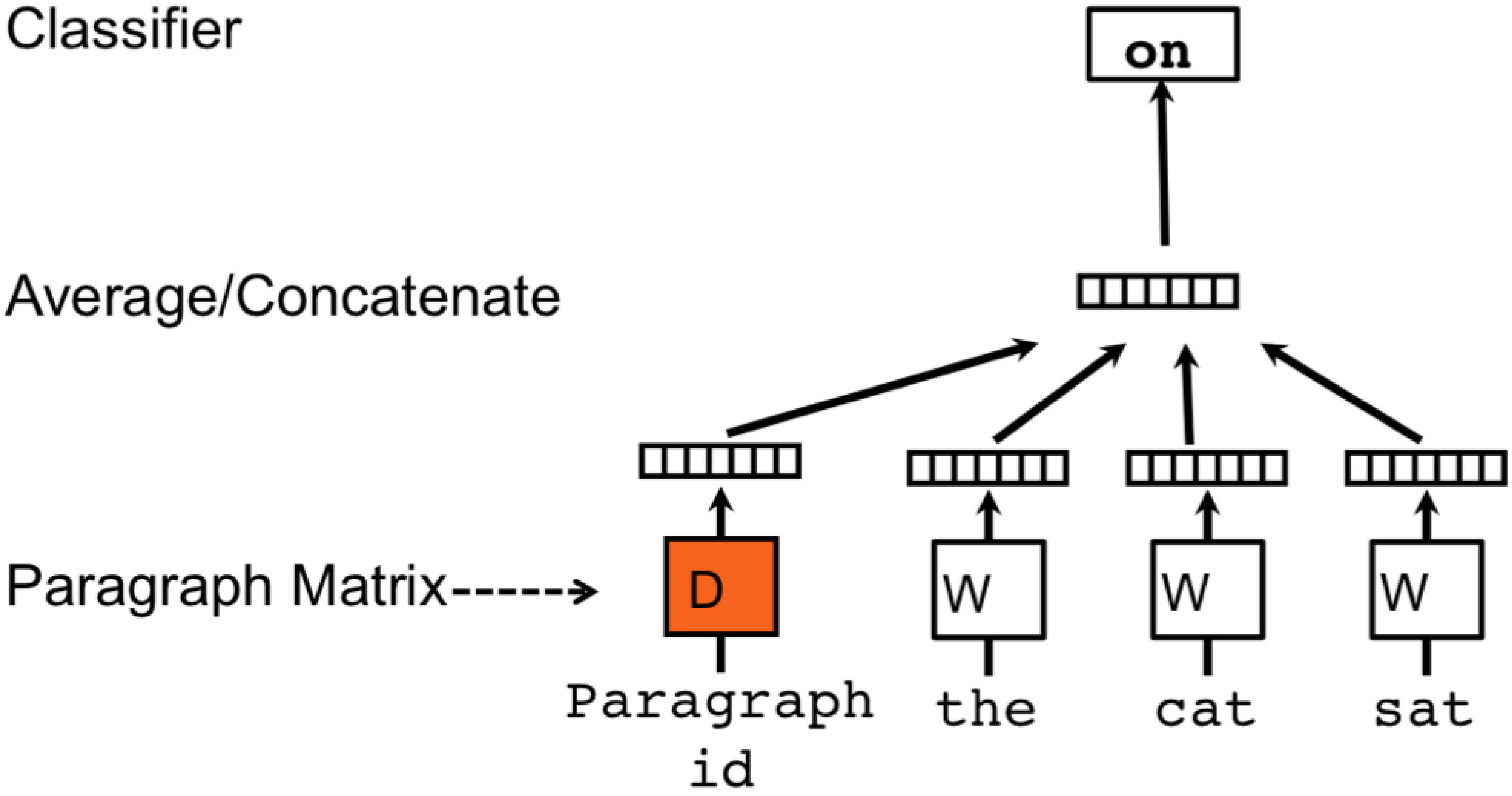
Results for image search

Model	Mean rank
Random	52.1
Recurrent NN	19.2
Constituency Tree Recursive NN	16.1
kCCA	15.9
Bag of Words	14.6
Dependency Tree Recursive NN	**12.5**

Lower is better!

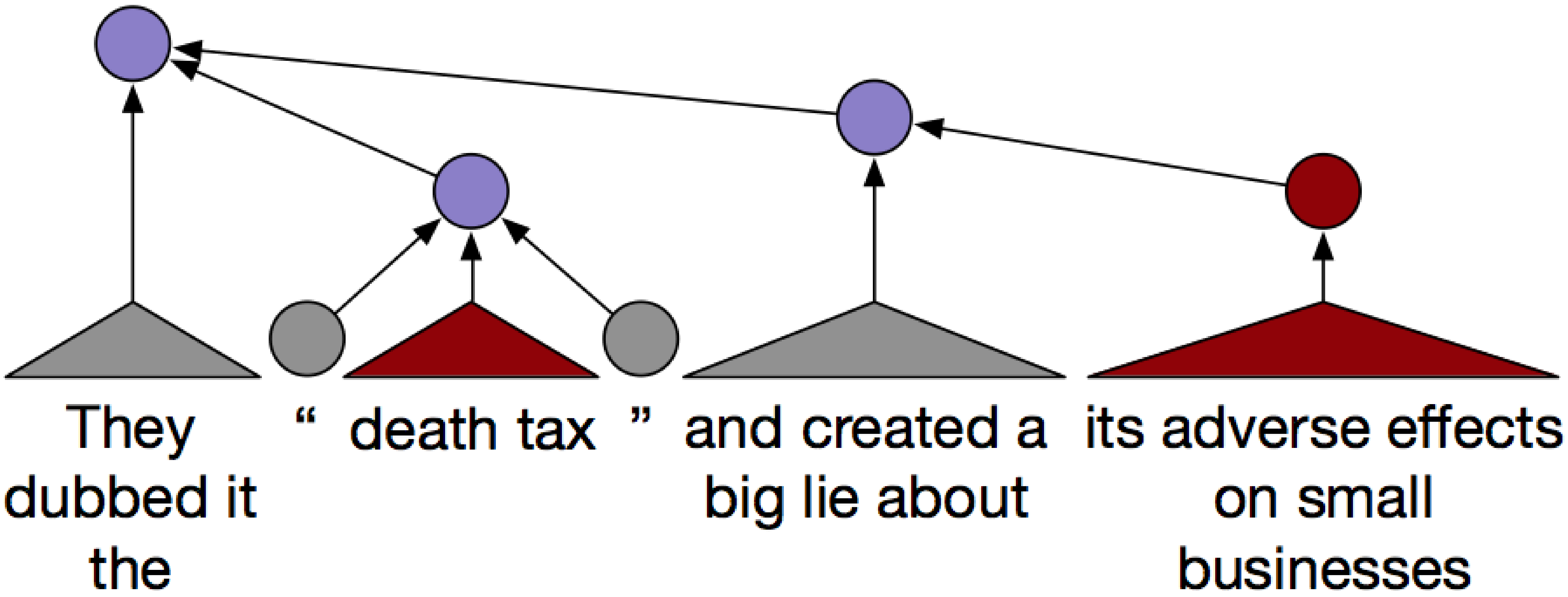
How to represent the meaning of texts

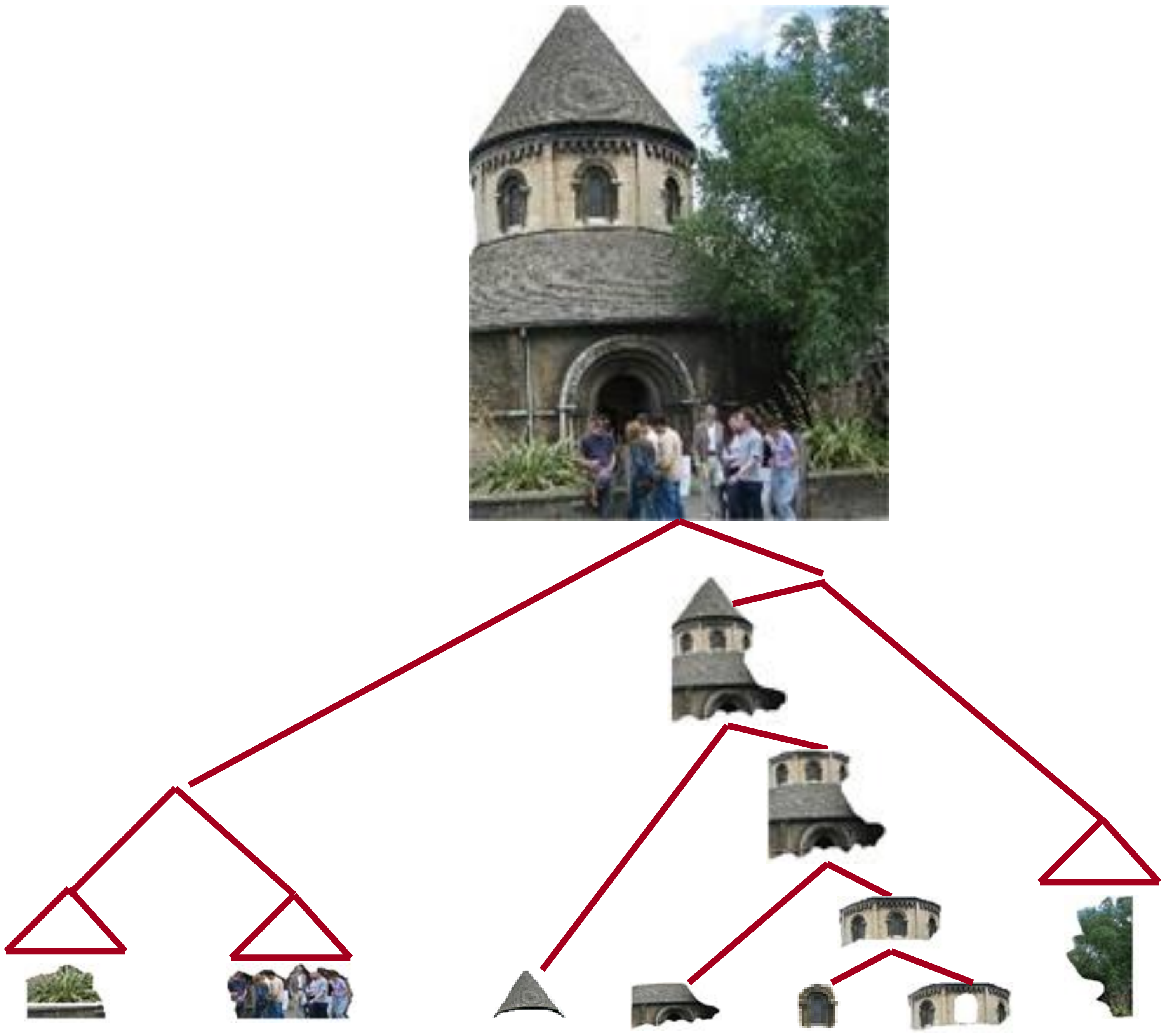
[Le and Mikolov, ICML 2014, Paragraph Vector]



Political Ideology Detection Using Recursive Neural Networks

[Iyyer, Enns, Boyd-Graber & Resnik 2014]

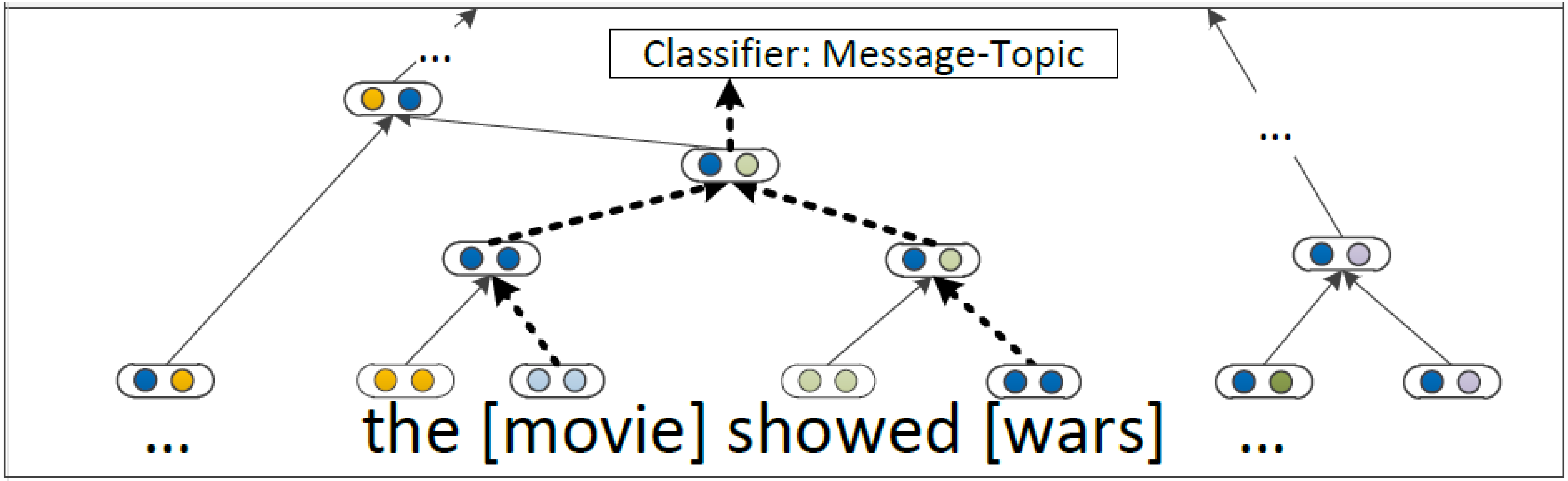




Extracting Semantic Relationships

[Socher, Huval, Manning & Ng, EMNLP 2012]

My [apartment]$_{e1}$ has a pretty large [kitchen] $_{e2}$
→ component-whole relationship (e2,e1)




DD
REGIS

Save the planet and return your name badge before you leave (on Tuesday)

Microsoft Research

Faculty Summit

2014 15TH ANNUAL

Image credits and permissions

- Slide 2 and 27: Trinity College Dublin library Free use picture from http://www.freegreatpicture.com/city-impression/trinity-college-dublin-the-old-library-14885
- Slide 3 From a paper of the author (Manning 1992, Romance is so complex) http://nlp.stanford.edu/manning/papers/romance.pdf
- Slides 4: PR2 robot reading CC BY-SA 3.0 by Troy Straszheim from http://commons.wikimedia.org/wiki/File:PR2_Robot_reads_the_Mythical_Man-Month_2.jpg

Image credits and permissions

Slide 10 images from Wikipedia (i) Public domain http://en.wikipedia.org/wiki/Litoria#mediaviewer/File:Caerulea_cropped(2).jpg
(ii) Brian Gratwick http://en.wikipedia.org/wiki/Pristimantis_cruentus#mediaviewer/File:Pristimantis_cruentus_studio.jpg
(iii) Grand-Duc, Niabot http://en.wikipedia.org/wiki/Edible_frog#mediaviewer/File:Rana_esculenta_on_Nymphaea_edit.JPG
(iv) Public domain http://en.wikipedia.org/wiki/Eleutherodactylus#mediaviewer/File:Eleutherodactylus_mimus.jpg

Slides 19, 20, 21, 25 Images from the PASCAL Visual Object Challenge 2008 data set http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/

Results on Gao et al. (2014) Dataset

	CW08	RNN		CBOW		GloVe	GloVe (840B)
Subtask	dim=50	dim=640	dim=1600	dim=100	dim=300	dim=300	dim=300
All capital cities	0.62%	1.23%	1.81%	6.62%	11.28%	54.79%	59.95%
Currency	0.25%	0.66%	0.87%	3.13%	4.32%	4.00%	3.57%
City-in-state	0.67%	3.14%	3.38%	1.55%	2.25%	7.94%	9.65%
Man-Woman	4.83%	18.46%	20.82%	25.89%	28.60%	38.38%	38.99%
Adjective to adverb	1.40%	1.17%	2.01%	3.45%	3.23%	8.84%	43.72%
Comparative	1.55%	34.92%	40.28%	33.41%	42.53%	53.18%	74.65%
Superlative	1.94%	25.33%	26.21%	23.56%	29.07%	33.08%	78.63%
Present participle	1.53%	20.03%	23.26%	8.20%	11.75%	20.21%	45.13%
Nationality adjective	3.07%	3.15%	3.76%	23.66%	47.44%	78.76%	90.35%
Past tense	1.84%	19.51%	22.77%	15.51%	24.15%	31.38%	49.58%
Plural nouns	3.21%	14.42%	18.28%	23.95%	38.82%	45.04%	75.15%
Plural verbs	2.44%	22.41%	26.62%	17.28%	31.82%	36.30%	60.39%
Total	2.36%	14.69%	17.85%	16.70%	27.10 %	33.44%	56.53%

Results on Gao et al. (2014) Dataset

	CW08	RNN		CBOW		GloVe	GloVe (840B)
Subtask	dim=50	dim=640	dim=1600	dim=100	dim=300	dim=300	dim=300
Antonym	0.28%	2.88%	3.12%	2.74%	4.57%	6.93%	6.18%
Attribute	0.22%	0.24%	0.42%	0.68%	1.18%	2.55%	6.41%
Causes	0.00%	0.00%	0.00%	0.15%	1.08%	1.38%	0.31%
DerivedFrom	0.05%	0.16%	0.18%	0.33%	0.63%	1.05%	2.51%
Entails	0.05%	0.05%	0.07%	0.26%	0.38%	0.61%	0.95%
HasContext	0.12%	0.16%	0.19%	0.28%	0.35%	0.75%	0.54%
InstanceOf	0.08%	0.81%	0.64%	0.48%	0.58%	1.56%	0.82%
IsA	0.07%	0.42%	0.47%	0.42%	0.67%	0.77%	0.90%
MadeOf	0.03%	0.10%	0.13%	0.33%	0.72%	1.77%	1.66%
MemberOf	0.08%	0.11%	0.13%	0.58%	1.06%	5.59%	2.83%
PartOf	0.31%	0.55%	0.60%	1.17%	1.27%	7.85%	7.10%
RelatedTo	0.00%	0.02%	0.00%	0.20%	0.05%	0.10%	0.24%
SimilarTo	0.02%	0.14%	0.18%	0.14%	0.29%	0.40%	0.62%
Total	0.06%	0.35%	0.40%	0.40 %	0.66%	0.90%	1.31%