UNIVERSITY OF CAMBRIDGE

Integrating Distributional and Compositional Semantics

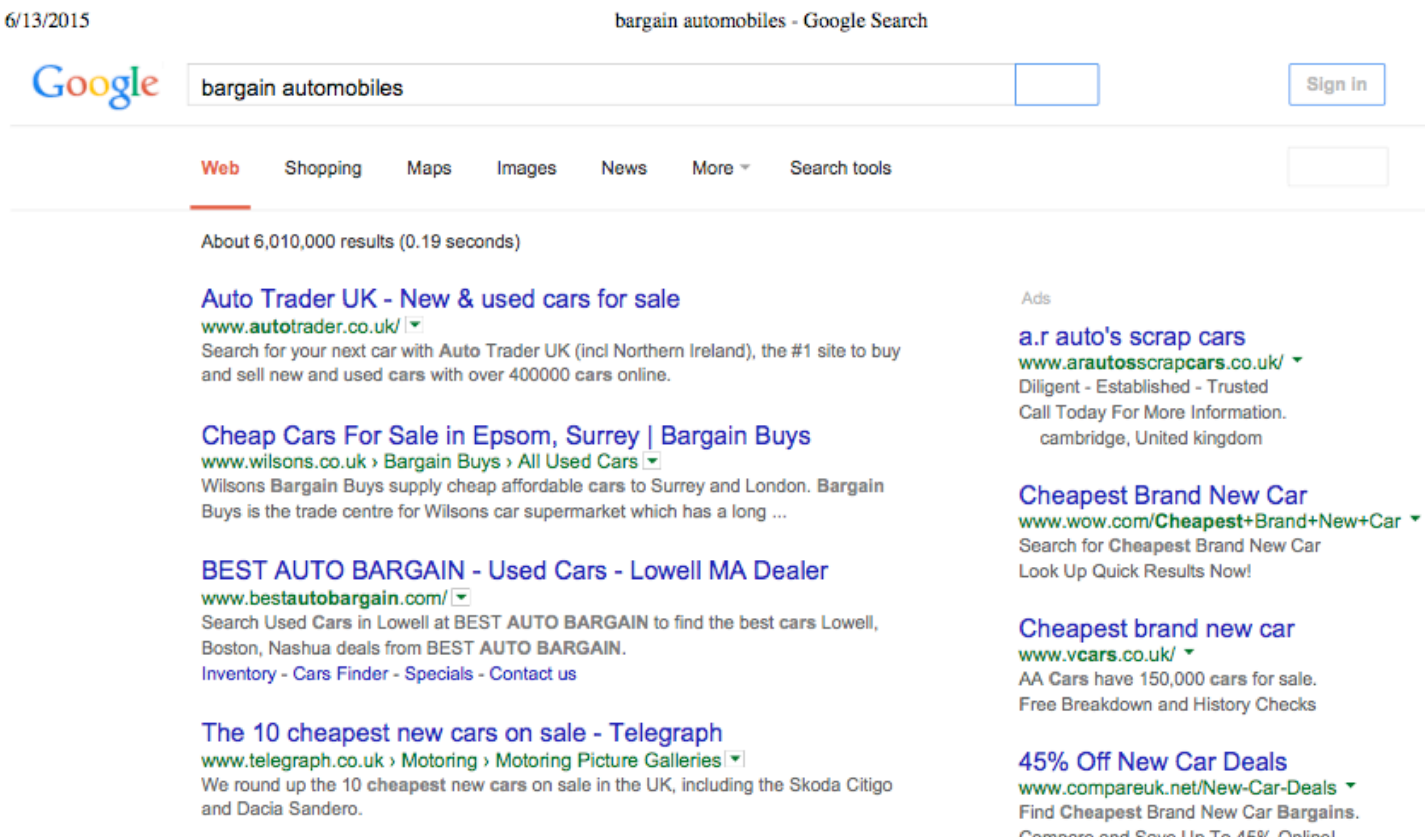
Stephen Clark, with Laura Rimell and Tamara Polajnar

Advanced Topics in Natural Language Processing

MPhil ACS, Lent Term

University of Cambridge Computer Laboratory

Words in Google



Sentences in Google

6/13/2015 man kills dog with rifle - Google Search

Google

man kills dog with rifle

Sign in

Web News Videos Images Shopping More Search tools

About 7,960,000 results (0.31 seconds)

Dog shoots and kills man in freak hunting accident - Daily Mail
www.dailymail.co.uk/.../Dog-shoots-kills-man-freak-hunting-accident.ht...
8 Jan 2008 - Dog shoots and kills man in freak hunting accident ... Price, 46, then set the gun in the back of his truck and was about to open the tailgate to ...

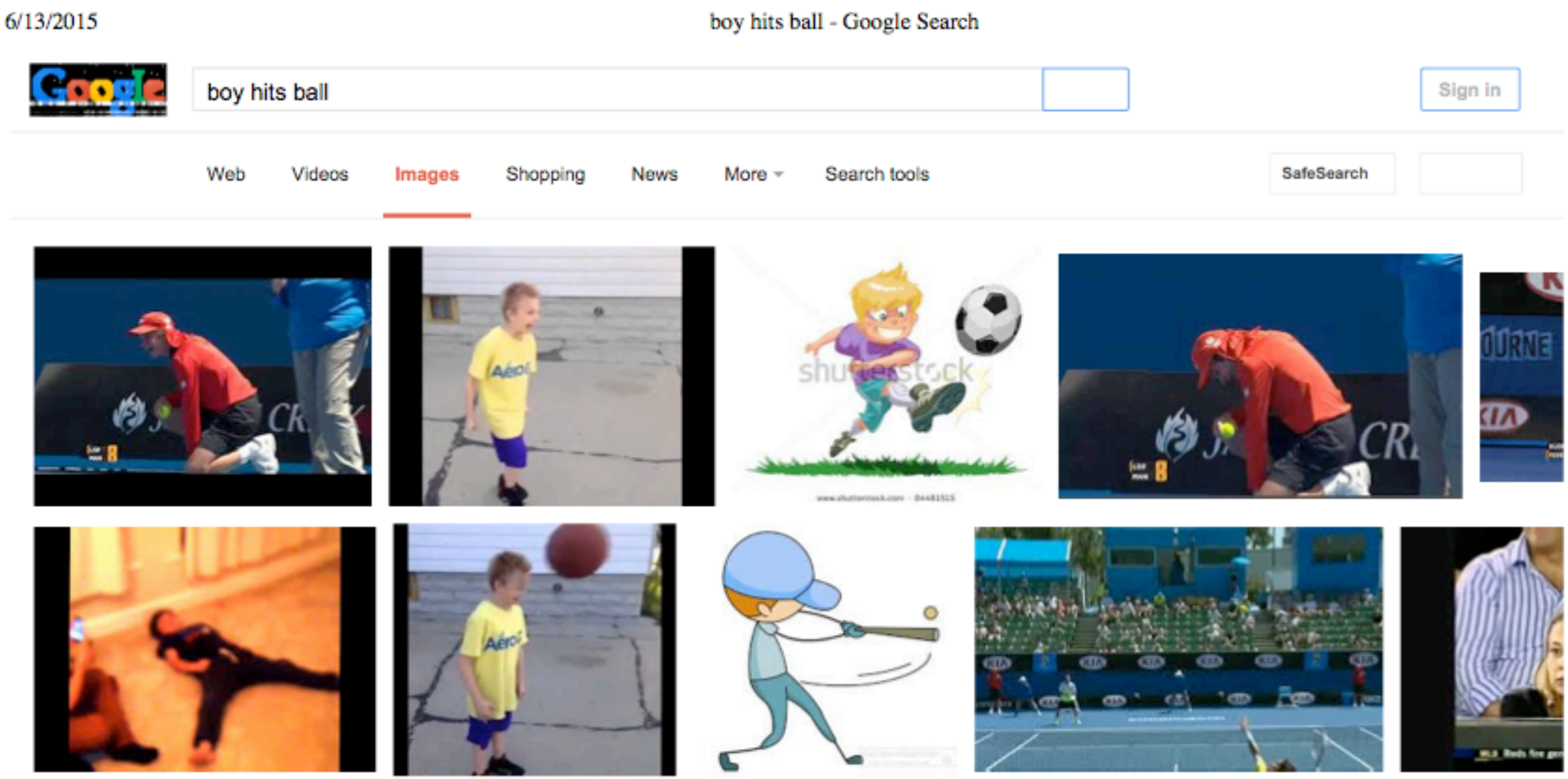
Man's Worst Enemy: 6 Negligent Gun Owners Who Were ...
www.alternet.org/.../mans-worst-enemy-6-negligent-gun-owners-who-w...
30 Dec 2014 - Guns don't kill people; dogs with guns kill people—or so it would seem from the recent rash of ... Dog Steps on Rifle and Shoots Wyoming Man.

Guns Don't Kill People, Dogs Kill People | Louis Klarevas
www.huffingtonpost.com/louis.../dog-shooting-accidents_b_4110822.ht...
17 Oct 2013 - Guns don't shoot and kill people. ... was shot in the leg when his dog jumped into his boat, landing on the man's shotgun and discharging it.

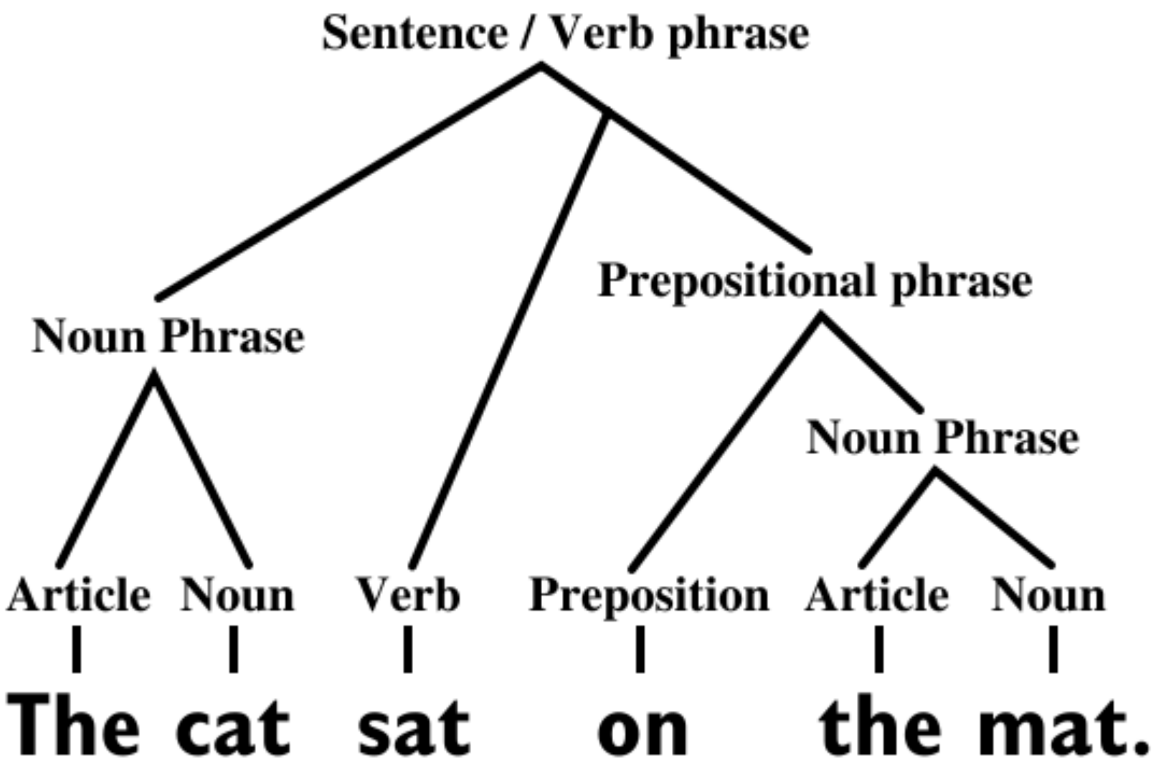
Friend with gun saves dog breeder from robber, kills thief
www.usatoday.com/story/news/nation/2015/01/30/dog.../22597495/
30 Jan 2015 - STONE MOUNTAIN, Ga. — A man who answered an online ad to buy a dog was killed Friday after attempting to rob the sellers, police said.

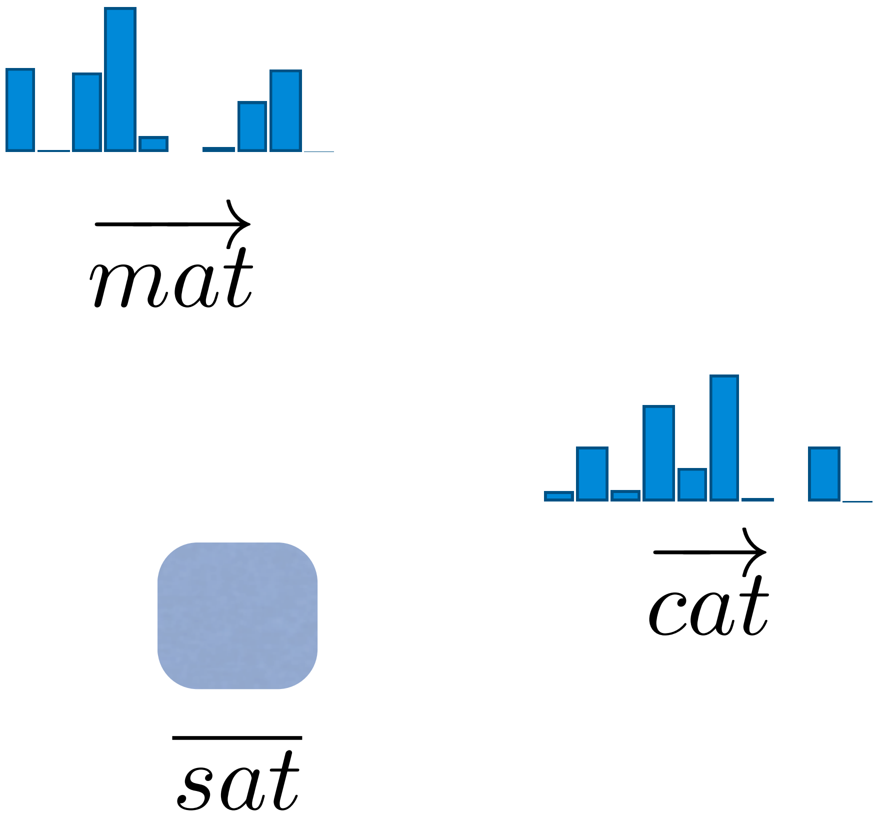
Rochester man allegedly shoots and kills dog - WMUR.com
www.wmur.com/news/rochester-man-allegedly...kills-dog/31597440
3 Mar 2015 - A Rochester man was arrested Monday after he allegedly shot and killed a dog with a high caliber hunting rifle.

Image Search

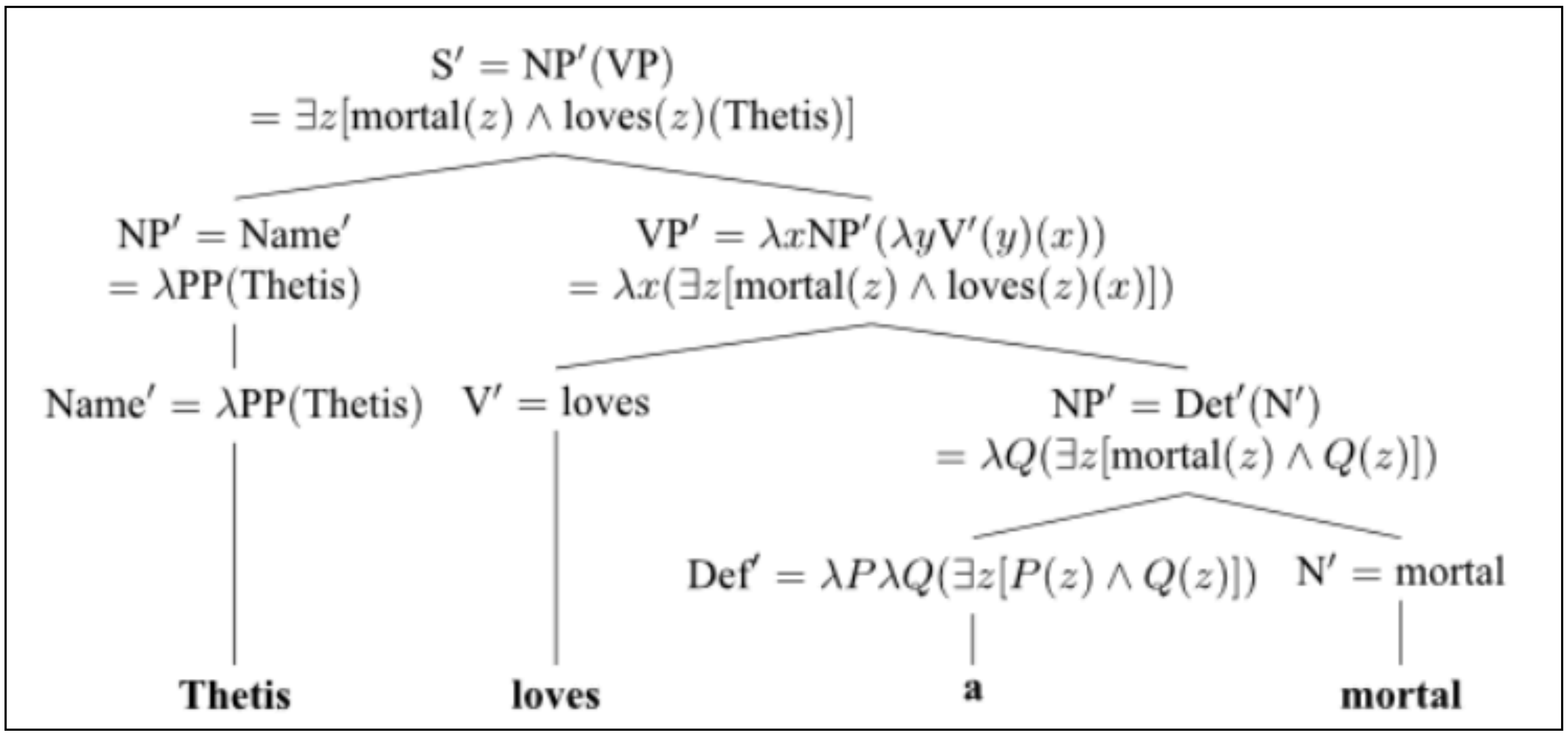


Compositional + Distributional ?

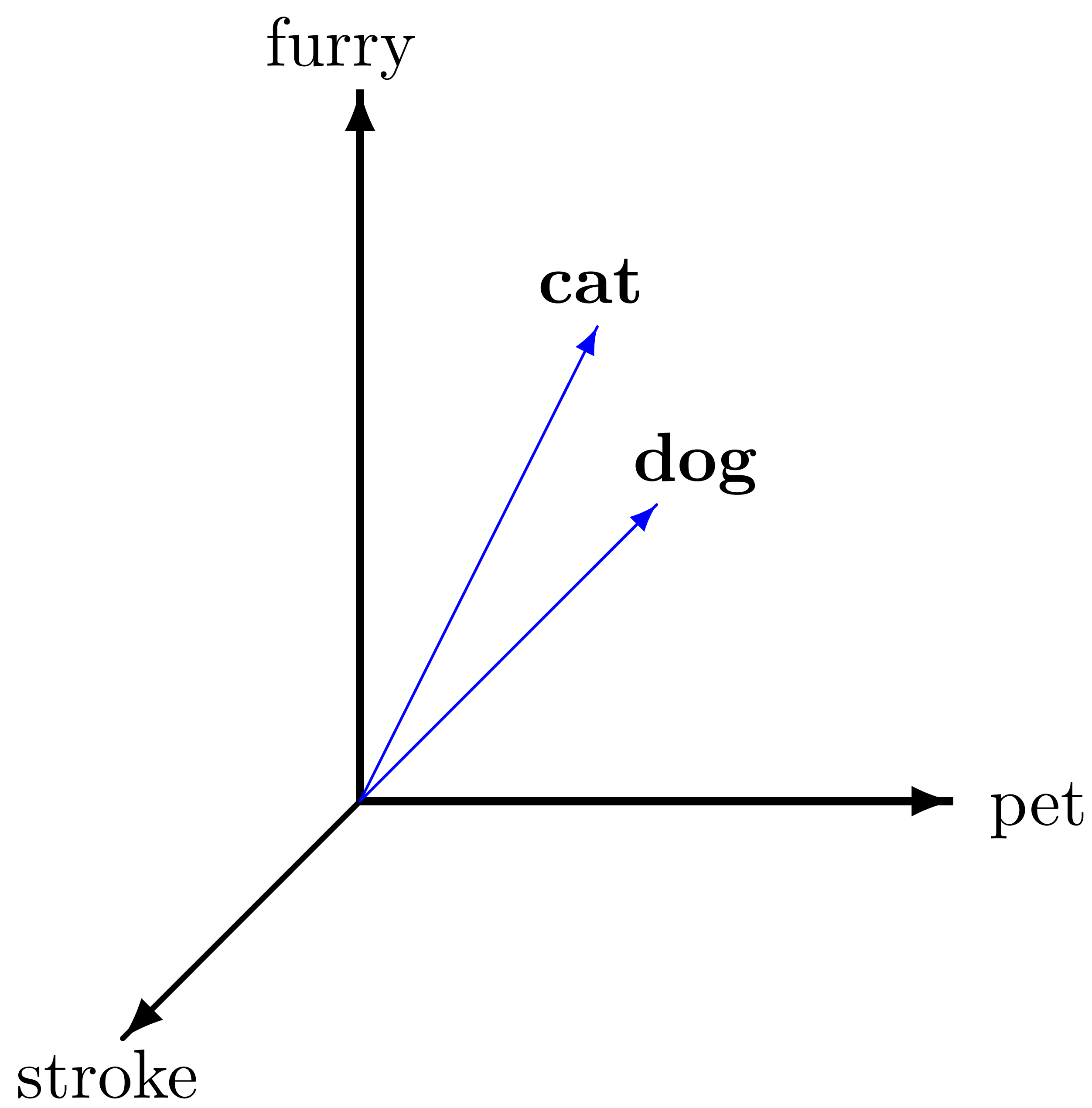




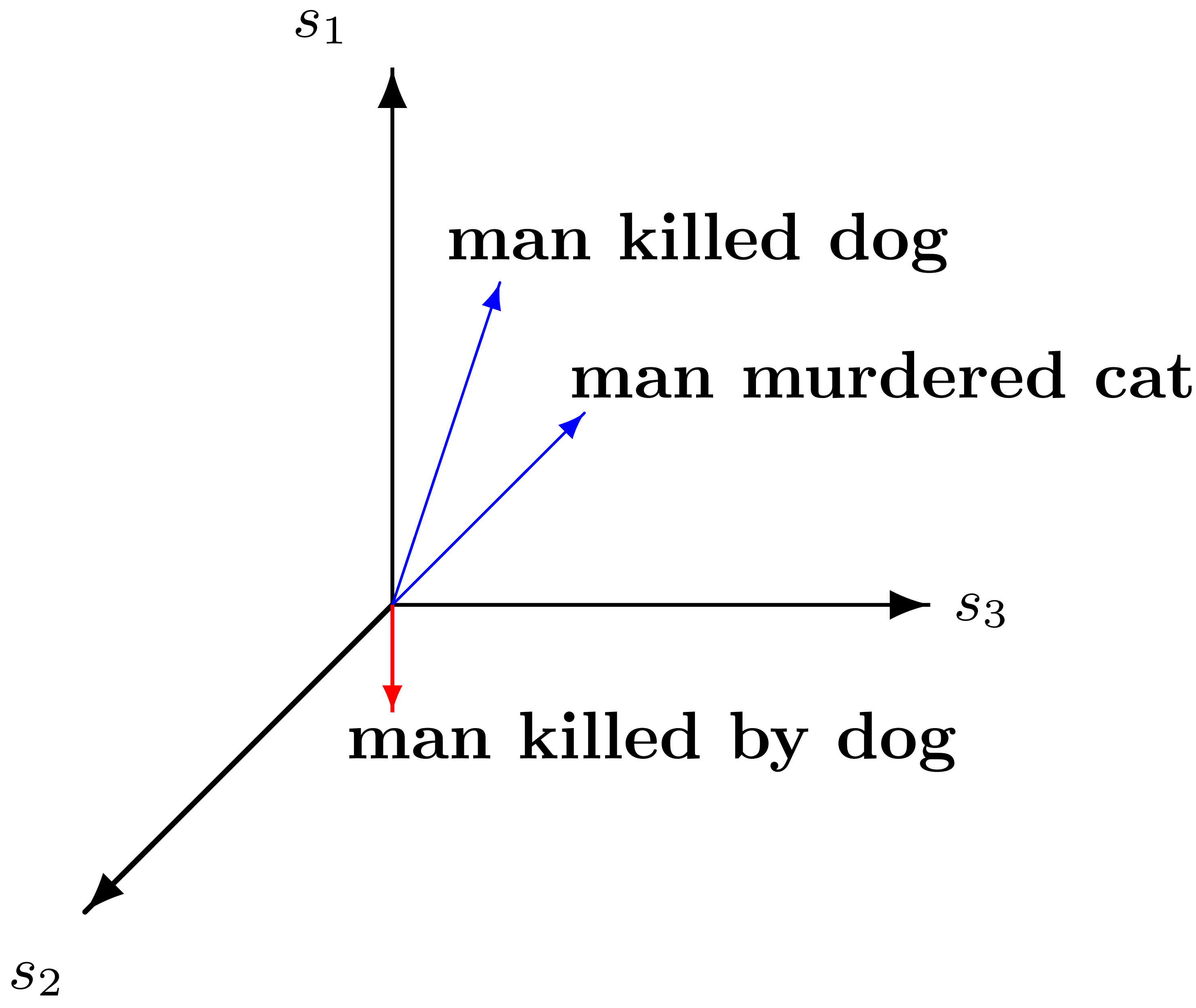
Formal Semantics



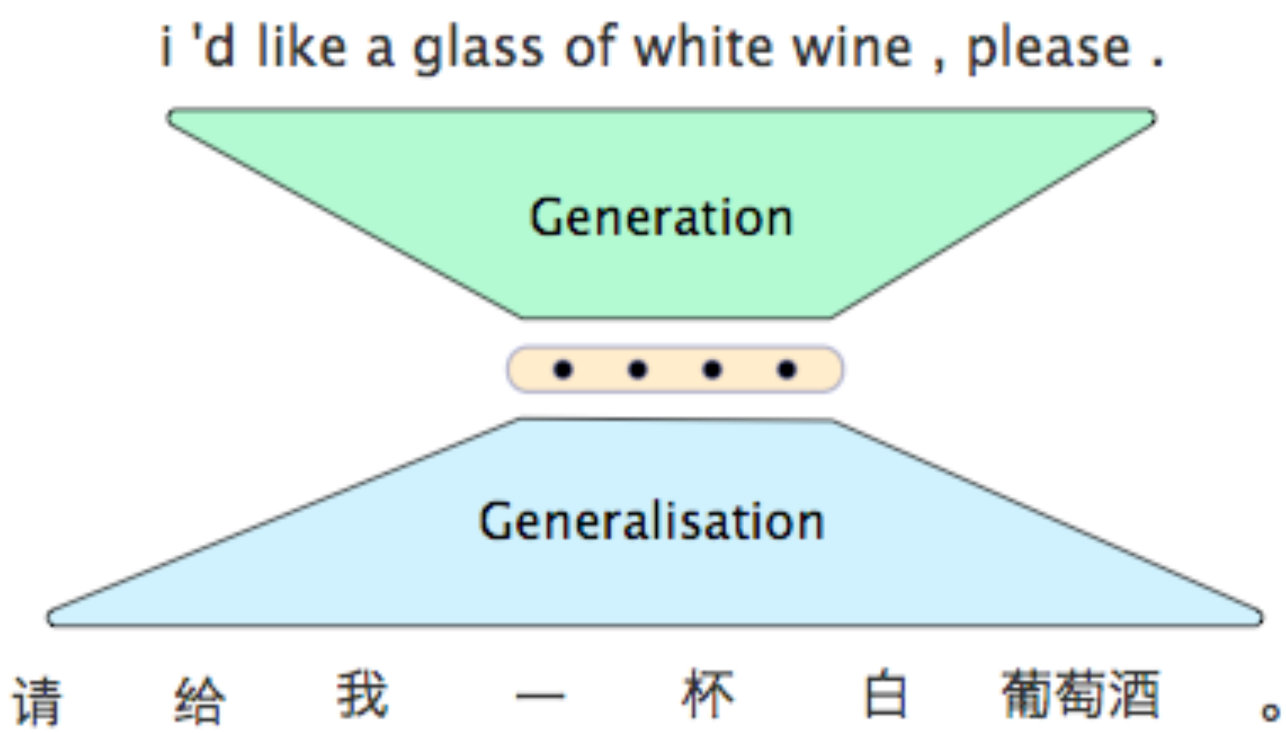
Vector Space Semantics



From Words to Sentences



Vector-Based Models of Sentences



Grefenstette et. al, New Directions in Vector Space Models of Meaning (ACL, 2014)

Lecture Outline

- Arguments against sentence vectors
- Vector addition
- Recurrent (recursive) neural networks
- Type-driven compositional distributional framework

Vectors are “Too Small”

- ★“You can’t cram the meaning of a whole %&!$# sentence into a single $&!#* vector!” (Ray Mooney)

Arguments Against Sentence Vectors

- A fixed-size vector can’t hold enough information (languages are infinite)
  - are languages really infinite? (not in practice, and maybe not in theory*)
  - the sentence vector could be a structured object (e.g. density matrix)
  - the sentence space doesn’t have to solve all of semantics (necessarily)
  - (and wouldn’t this argument apply to lexical semantics as well?)

*Recursion and the Infinitude Claim (Pullum and Scholz, 2010)

Arguments Against Sentence Vectors

- A fixed-size vector can't hold enough information (languages are infinite)
  - are languages really infinite? (not in practice, and maybe not in theory*)
  - the sentence vector could be a structured object (e.g. density matrix)
  - the sentence space doesn't have to solve all of semantics (necessarily)
  - (and wouldn't this argument apply to lexical semantics as well?)
- What about (formal) semantics?
  - compositionality, inference, logical operators, quantification, ...

*Recursion and the Infinitude Claim (Pullum and Scholz, 2010)

Talk Outline

- Arguments against sentence vectors
- **Vector addition**
- Recurrent (recursive) neural networks
- Type-driven compositional distributional framework

Element-wise Operators on Context Vectors

black	0.34	0.64	...	-0.06	...

+

cat	0.15	0.29	...	-0.03	...

=

black + cat	0.49	0.93	...	-0.09	...

black	0.34	0.64	...	-0.06	...

$\odot$

cat	0.15	0.29	...	0.03	...

=

black o cat	0.05	0.19	...	-0.002	...

Circular Convolution

	1	2	3
cat	0.0032	0.0025	-0.0085
black	0.0006	0.0005	-0.0017

$black^T$ x cat

1.90E-06	1.60E-06	5.00E-07
1.50E-06	1.25E-06	-4.25E-07
-5.1E-06	-4.25E-06	-1.15E-06

black ⊛ cat =

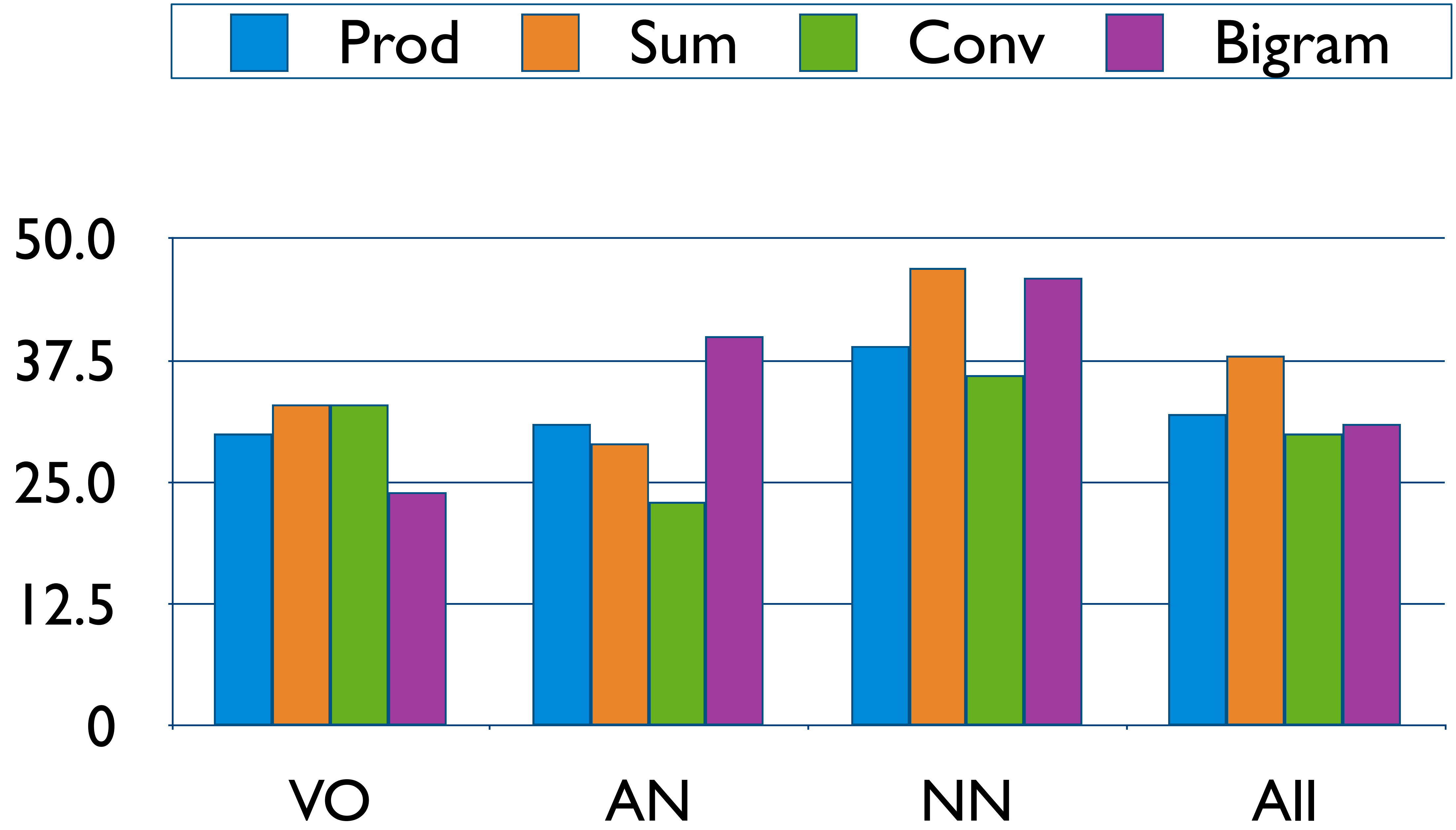
-2.76E-06	**4.55E-06**	**-4.39E-06**

Phrase Similarity Data

Mitchell & Lapata 2010 Dataset:

AN:	national government	cold air	**1**
	new information	further evidence	**6**
NN:	environment secretary	party leader	**5**
	telephone number	future development	**2**
VO:	offer support	provide help	**7**
	fight war	win battle	**5**

Phrase Similarity Results



Sentence Similarity Data

- Semantic Textual Similarity (STS) datasets from SEMEVAL
- MSR Par dataset (1,500 pairs):

The fines are part of failed Republican efforts to force or entice the Democrats to return.

Perry said he backs the Senates efforts, including fines, to force the Democrats to return. 2.8

The bill says that a woman who undergoes such an abortion couldn't be prosecuted.

A woman who underwent such an abortion could not be prosecuted under the bill. 5.0

Addition for Sentence Vectors?

"I know of no pressure," said Mr. Feith, the under secretary of defense for policy.

"I know of nobody who pressured anybody," Douglas Feith, undersecretary of defense for policy, said at a Pentagon briefing.

[Similarity 3.8/5]

Agirre et al. (Semeval STS); Polajnar, Rimell and Clark (LREC 2014)

Addition for Sentence Vectors?

"I know of no pressure," said Mr. Feith, the under secretary of defense for policy.

"I know of nobody who pressured anybody," Douglas Feith, undersecretary of defense for policy, said at a Pentagon briefing.

[Similarity 3.8/5]

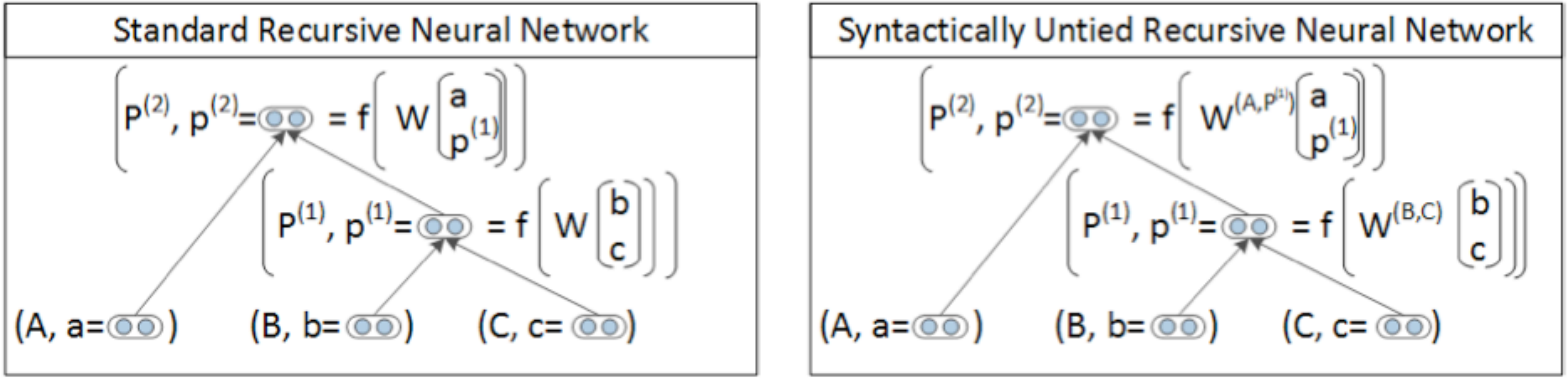
- Lexical overlap baseline is hard to beat
- Out of the vector-space methods, addition is hard to beat
- "Is God trying to tell us something?"

Agirre et al. (Semeval STS); Polajnar, Rimell and Clark (LREC 2014)

Lecture Outline

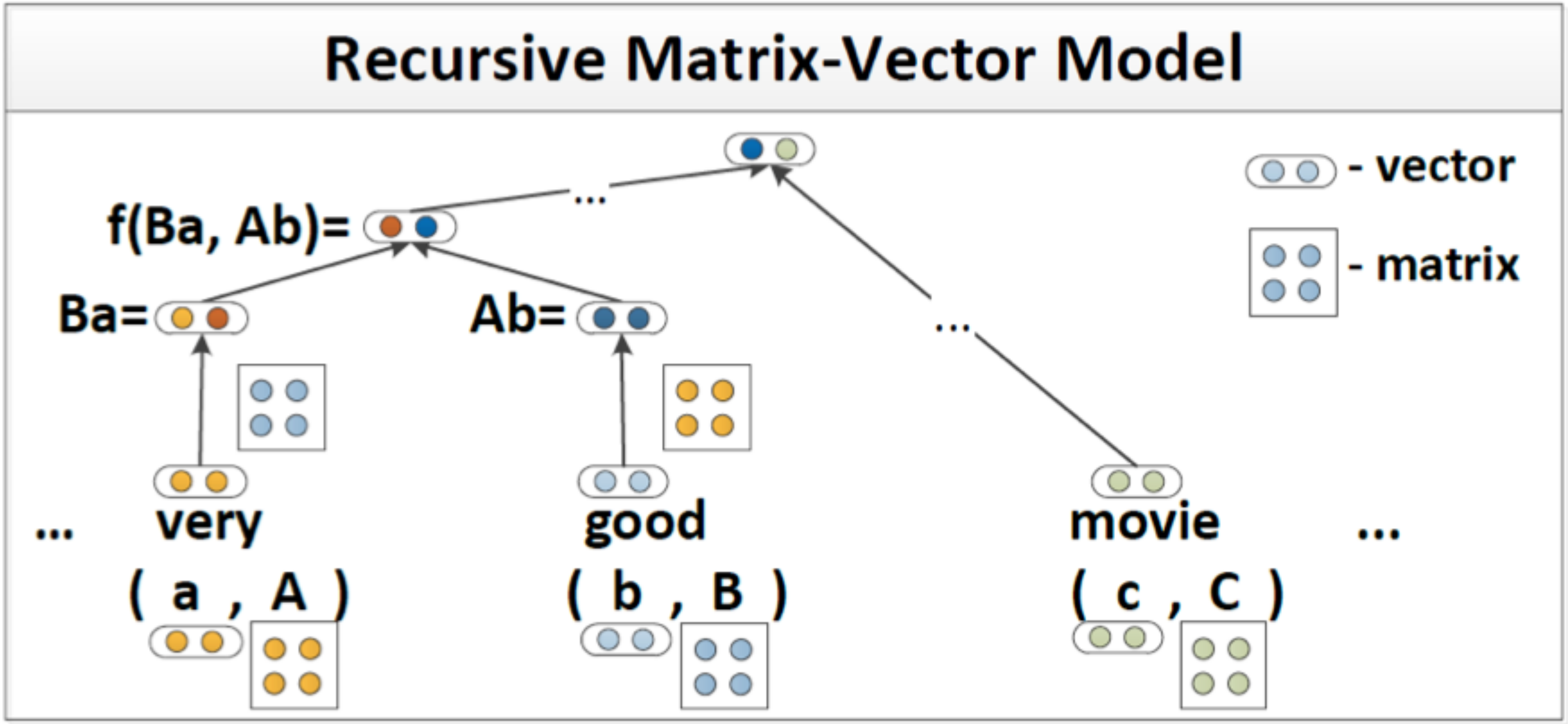
- Arguments against sentence vectors
- Vector addition
- **Recurrent (recursive) neural networks**
- Type-driven compositional distributional framework

Composition in Neural Models



Deep Learning for NLP (Socher et al., 2013)

Composition in Neural Models



Socher et al. (EMNLP 2013)

Lecture Outline

- Arguments against sentence vectors
- Vector addition
- Recurrent (recursive) neural networks
- **Type-driven compositional distributional framework**

Categorial Grammar

$$\dfrac{cat}{NP} \quad \dfrac{chases}{(S\backslash NP)/NP} \quad \dfrac{dog}{NP}$$

Categorial Grammar

$$
\begin{array}{ccc}
cat & chases & dog \\
\hline NP & (S\backslash NP)/{\color{blue}NP} & {\color{blue}NP} \\
 & \multicolumn{2}{c}{\underline{\hspace{10em}}>} \\
 & \multicolumn{2}{c}{S\backslash NP}
\end{array}
$$

Function application = “cancellation”

Categorial Grammar

$$
\begin{array}{ccc}
\mathit{cat} & \mathit{chases} & \mathit{dog} \\
\overline{} & \overline{} & \overline{} \\
{\color{blue}NP} & (S\backslash NP)/NP & NP \\
 & \multicolumn{2}{c}{\overline{}>} \\
 & \multicolumn{2}{c}{S\backslash {\color{blue}NP}} \\
\multicolumn{3}{c}{\overline{}<} \\
\multicolumn{3}{c}{S}
\end{array}
$$

Predicate-Argument Semantics

$$
\begin{array}{ccc}
cat & chases & dog \\
\hline
NP & (S\backslash NP)/NP & NP \\
cat' & \lambda x.\lambda y\ chases'(x,y) & dog'
\end{array}
$$

Predicate-Argument Semantics

$$
\begin{array}{ccc}
cat & chases & dog \\
\overline{\quad\quad} & \overline{\quad\quad\quad\quad\quad\quad\quad\quad} & \overline{\quad\quad} \\
NP & (S\backslash NP)/NP & NP \\
cat' & \lambda x.\lambda y\ chases'(x, y) & dog' \\
 & \multicolumn{2}{c}{\overline{\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad}} \\
 & \multicolumn{2}{c}{S\backslash NP} \\
 & \multicolumn{2}{c}{\lambda y\ chases'(dog', y)}
\end{array}
$$

Function application = substitution

Predicate-Argument Semantics

$$
\dfrac{\dfrac{\dfrac{cat}{\begin{array}{c}NP\\ cat'\end{array}} \quad \dfrac{\dfrac{chases}{\begin{array}{c}(S\backslash NP)/NP\\ \lambda x.\lambda y\ chases'(x,y)\end{array}} \quad \dfrac{dog}{\begin{array}{c}NP\\ dog'\end{array}}}{\begin{array}{c}S\backslash NP\\ \lambda y\ chases'(dog',y)\end{array}}}{}}{\begin{array}{c}S\\ chases'(dog',cat')\end{array}}
$$

Vector Space Semantics?

$$
\begin{array}{ccc}
cat & chases & dog \\
\hline
NP & (S\backslash NP)/NP & NP \\
cat' & \lambda x.\lambda y\ chases'(x, y) & dog' \\
\hline
 & \multicolumn{2}{c}{S\backslash NP} \\
 & \multicolumn{2}{c}{\lambda y\ chases'(dog', y)}
\end{array}
$$

- What are the semantic types of the vectors?
- What is the equivalent of function application?

Adjective-Noun Combinations

$$\dfrac{\dfrac{red}{N/\color{blue}{N}} \quad \dfrac{car}{\color{blue}{N}}}{N}$$

Adjective-Noun Combinations

$$\frac{\dfrac{red}{N/N} \qquad \dfrac{car}{N}}{N}$$

- Functions are matrices (linear maps) in linear algebra
- Functions combine with arguments using matrix multiplication (Baroni and Zamparelli, 2010)

Matrix Multiplication

$$RED \qquad \overrightarrow{car} \qquad \overrightarrow{red\ car}$$

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} rc_1 \\ rc_2 \\ rc_3 \\ rc_4 \\ rc_5 \end{pmatrix}$$

Matrix Multiplication

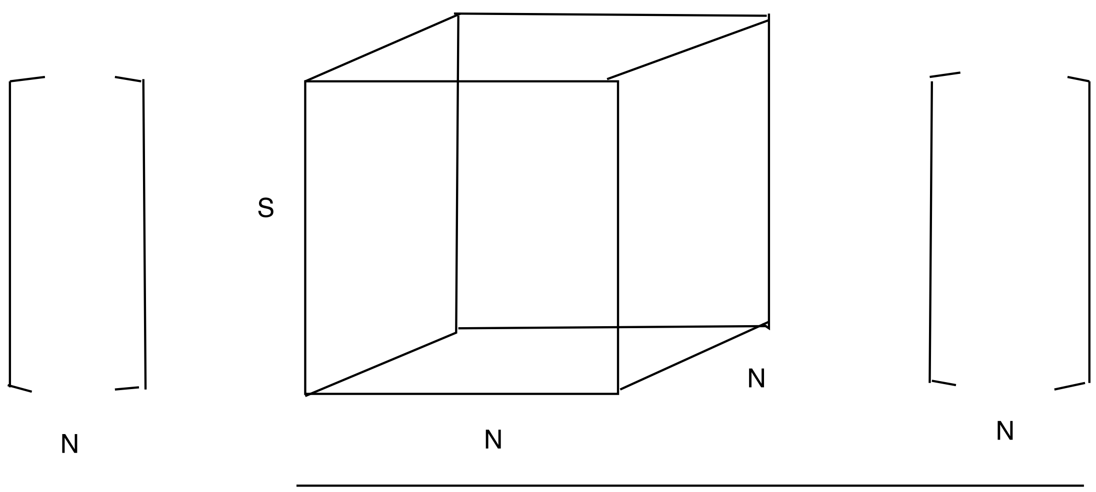
$$\begin{array}{ccc} RED & \overrightarrow{car} & \overrightarrow{red\ \ car} \\ \mathbf{N} \otimes \mathbf{N} & \mathbf{N} & \mathbf{N} \end{array}$$

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} & R_{14} & R_{15} \\ R_{21} & R_{22} & R_{23} & R_{24} & R_{25} \\ R_{31} & R_{32} & R_{33} & R_{34} & R_{35} \\ R_{41} & R_{42} & R_{43} & R_{44} & R_{45} \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix} = \begin{pmatrix} rc_1 \\ rc_2 \\ rc_3 \\ rc_4 \\ rc_5 \end{pmatrix}$$

Syntactic Types to Tensors

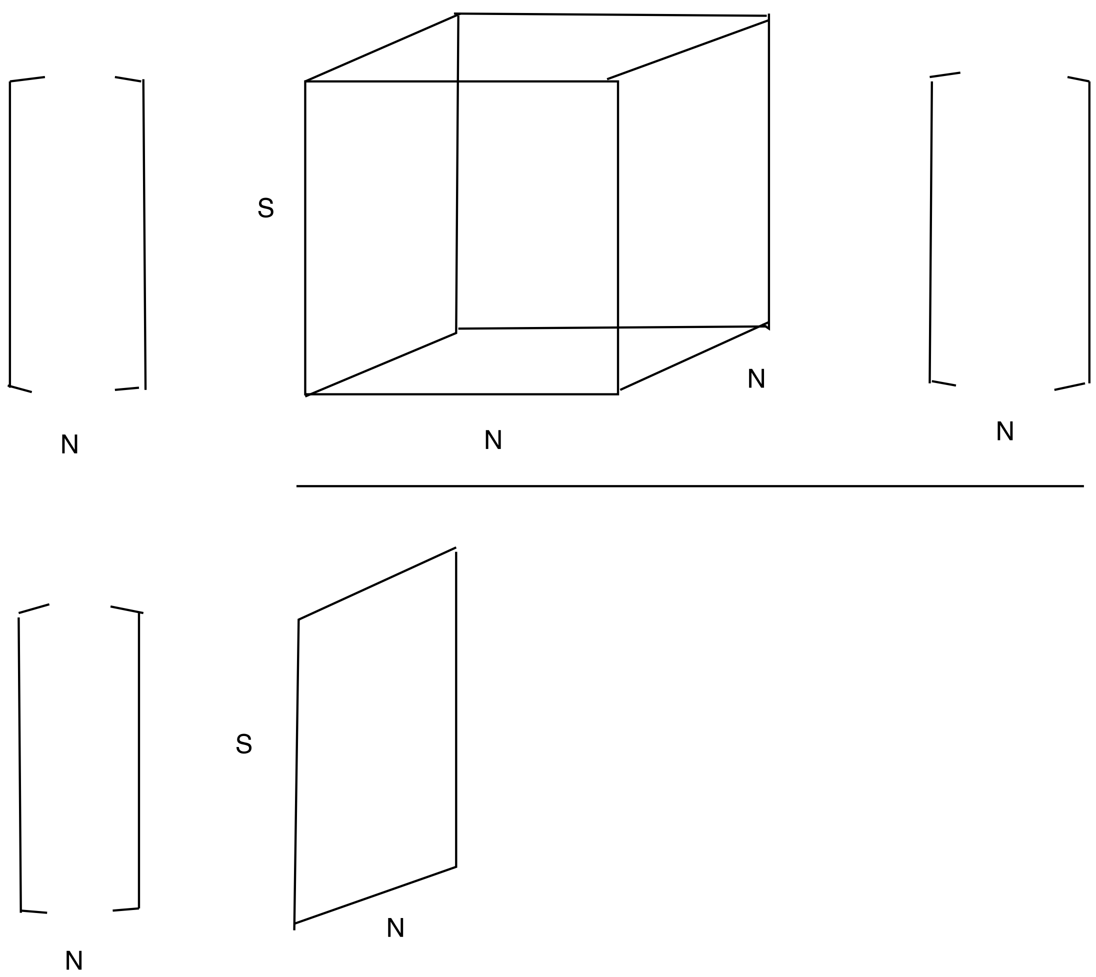
$$\begin{array}{ccc} cat & chases & dog \\ \overline{NP} & \overline{(S\backslash NP)/NP} & \overline{NP} \\ \mathbf{N} & \mathbf{S} \otimes \mathbf{N} \otimes \mathbf{N} & \mathbf{N} \end{array}$$

Syntactic Types to Tensors



$$\begin{array}{ccc} \overline{cat} & \overline{chases} & \overline{dog} \\ NP & (S\backslash NP)/NP & NP \\ \mathbf{N} & \mathbf{S} \otimes \mathbf{N} \otimes \mathbf{N} & \mathbf{N} \end{array}$$

Type and Tensor Reductions



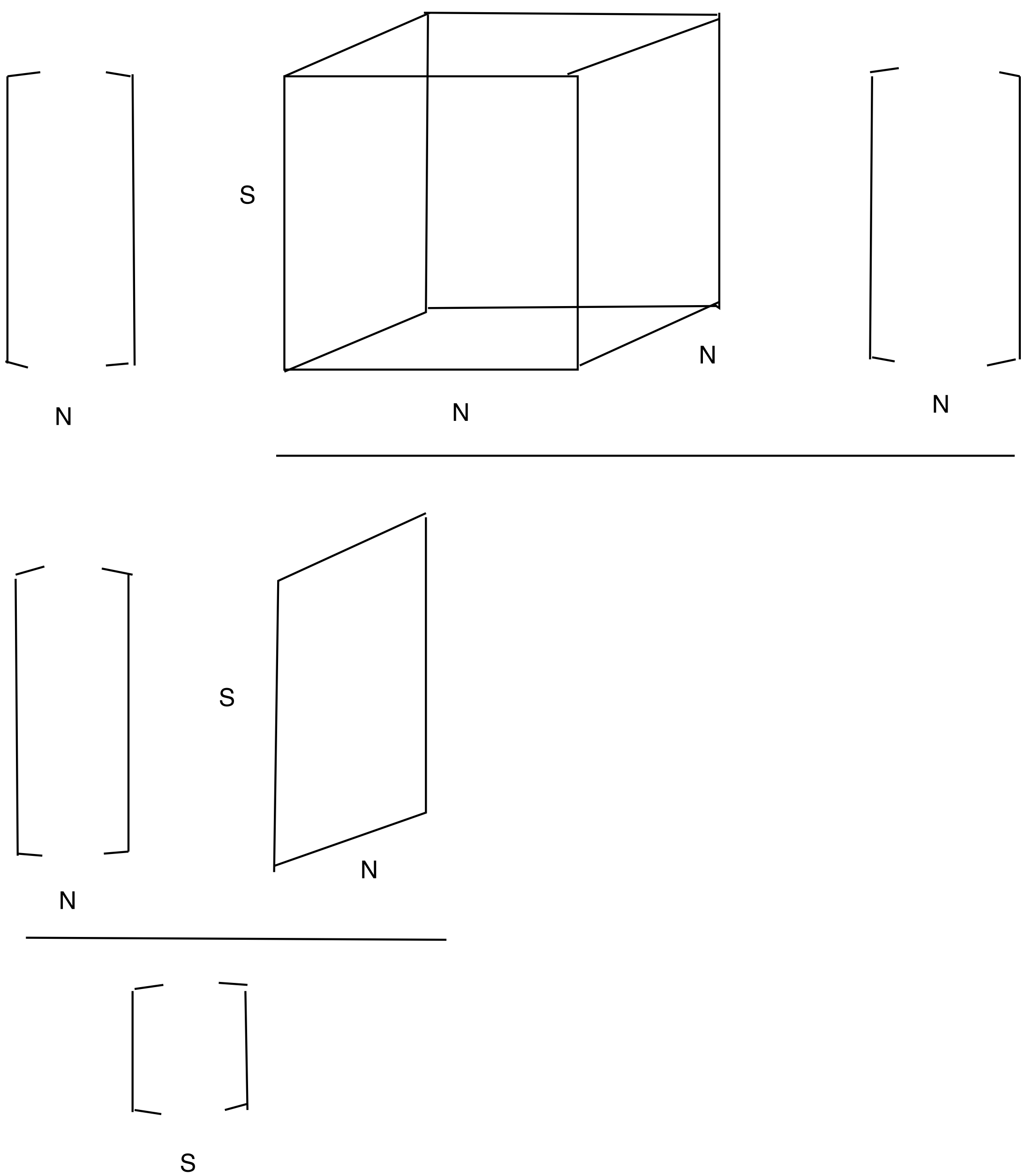
$$\begin{array}{ccc} cat & chases & dog \\ \overline{NP} & \overline{(S\backslash NP)/\color{blue}{NP}} & \overline{\color{blue}{NP}} \\ \mathbf{N} & \mathbf{S} \otimes \mathbf{N} \otimes \color{blue}{\mathbf{N}} & \color{blue}{\mathbf{N}} \end{array}$$

$$\dfrac{(S\backslash NP)/\color{blue}{NP} \quad \color{blue}{NP}}{S\backslash NP}$$

$$\mathbf{S} \otimes \mathbf{N}$$

Function application = taking inner products

Type and Tensor Reductions



$$
\begin{array}{ccc}
cat & chases & dog \\
\hline
\color{blue}{NP} & (S\backslash NP)/NP & NP \\
\color{blue}{\mathbf{N}} & \mathbf{S} \otimes \mathbf{N} \otimes \mathbf{N} & \mathbf{N}
\end{array}
$$

$$
\frac{\qquad}{\begin{array}{c} S\backslash \color{blue}{NP} \\ \mathbf{S} \otimes \color{blue}{\mathbf{N}} \end{array}}
$$

$$
\frac{\qquad}{\begin{array}{c} S \\ \mathbf{S} \end{array}}
$$

Function Composition in CCG

$$\dfrac{Pat}{NP} \quad \dfrac{might}{(S\backslash NP)/(S\backslash NP)} \quad \dfrac{kiss}{(S\backslash NP)/NP} \quad \dfrac{Sandy}{NP}$$

UNIVERSITY OF CAMBRIDGE

Function Composition in CCG

$$
\dfrac{\dfrac{}{\mathit{Pat}}}{NP} \quad \dfrac{\mathit{might}}{(S\backslash NP)/\color{blue}{(S\backslash NP)}} \quad \dfrac{\mathit{kiss}}{\color{blue}{(S\backslash NP)}/NP} \quad \dfrac{\mathit{Sandy}}{NP}
$$

$$
\dfrac{(S\backslash NP)/\color{blue}{(S\backslash NP)} \quad \color{blue}{(S\backslash NP)}/NP}{(S\backslash NP)/NP} >\mathbf{B}
$$

Function composition = “cancellation”

Function Composition in CCG

$$
\begin{array}{cccc}
\mathit{Pat} & \mathit{might} & \mathit{kiss} & \mathit{Sandy} \\
\hline
NP & (S\backslash NP)/(S\backslash NP) & (S\backslash NP)/NP & NP \\
\mathbf{N} & \mathbf{S} \otimes \mathbf{N} \otimes \mathbf{S} \otimes \mathbf{N} & \mathbf{S} \otimes \mathbf{N} \otimes \mathbf{N} & \mathbf{N}
\end{array}
$$

$$
\dfrac{(S\backslash NP)/(S\backslash NP) \quad (S\backslash NP)/NP}{(S\backslash NP)/NP} > \mathbf{B}
$$

$$\mathbf{S} \otimes \mathbf{N} \otimes \mathbf{N}$$

Function composition = taking inner products

Maillard, Clark, Grefenstette (2014)

From Theory to Implementation

- We have wide-coverage CCG parsers, and syntax determines semantics
- But, two crucial questions the framework does not answer:
  1. what is the sentence space?
  2. how do we learn the tensors?
- The tensors can get very large:

$$\frac{Revitalized}{N/N} \quad \frac{Classics}{N} \quad \frac{Take}{(S[dcl]\backslash NP)/NP} \quad \frac{the}{NP[nb]/N} \quad \frac{Stage}{N} \quad \frac{in}{((S\backslash NP)\backslash(S\backslash NP))/NP} \quad \frac{Windy}{N/N}$$

Contextual Sentence Spaces

- Two contenders for the sentence space in the current literature:
  - a space automatically induced by a (un-)supervised learning criterion
  - a contextual sentence space (extending the distributional hypothesis*)

*Baroni, Bernardi, Zamparelli; Grefenstette et al.

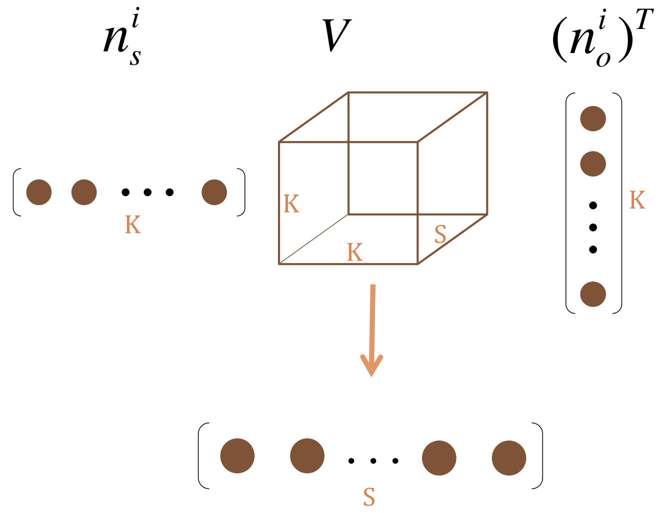
Contextual Sentence Spaces

- Two contenders for the sentence space in the current literature:
  - a space automatically induced by a (un-)supervised learning criterion
  - a contextual sentence space (extending the distributional hypothesis*)
    - should contextual noun and sentence spaces be the same?

*Baroni, Bernardi, Zamparelli; Grefenstette et al.

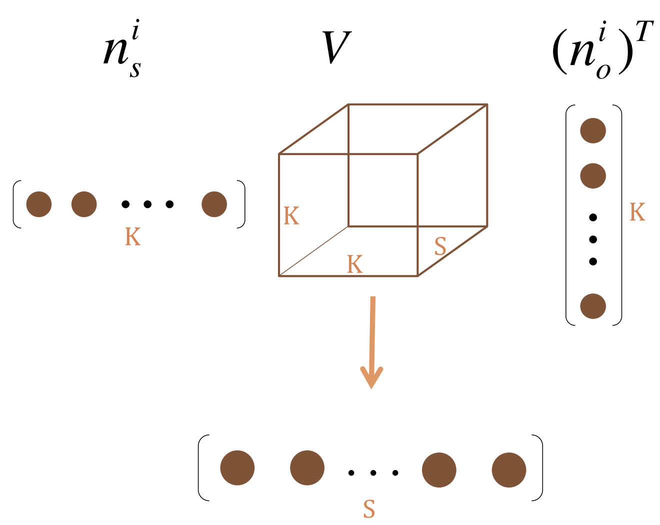
Context-based Sentence Vectors

$\mathbf{S}_{t-2}$: M. Atget captured the old Paris in his pictures. $\mathbf{S}_{t-1}$: His photographs show the city in its various facets. $\mathbf{S}_t$: He photographed stairwells and architectural details. $\mathbf{S}_{t+1}$: His interests also extended to the environs of Paris. $\mathbf{S}_{t+2}$: He also photographed street-hawkers and small tradesmen, as well as popular amusements.



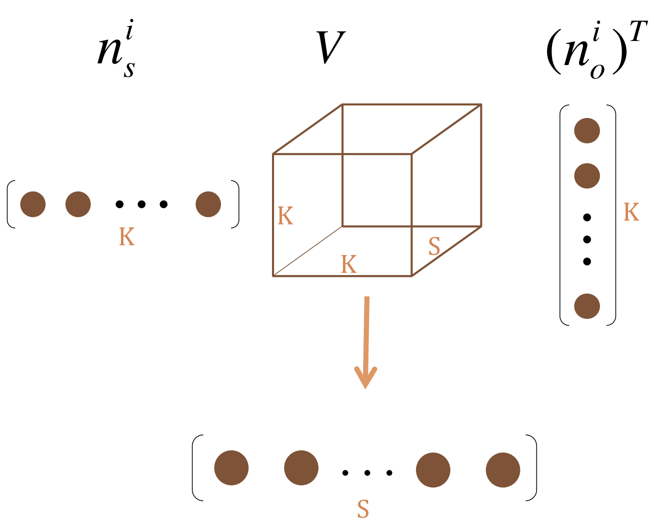
Context-based Sentence Vectors

$\mathbf{S}_{t-2}$: M. Atget captured the old Paris in his pictures. $\mathbf{S}_{t-1}$: His photographs show the city in its various facets. $\mathbf{S}_{t}$: He photographed stairwells and architectural details. $\mathbf{S}_{t+1}$: His interests also extended to the environs of Paris. $\mathbf{S}_{t+2}$: He also photographed street-hawkers and small tradesmen, as well as popular amusements.
IDist: stairwell, architectural, detail



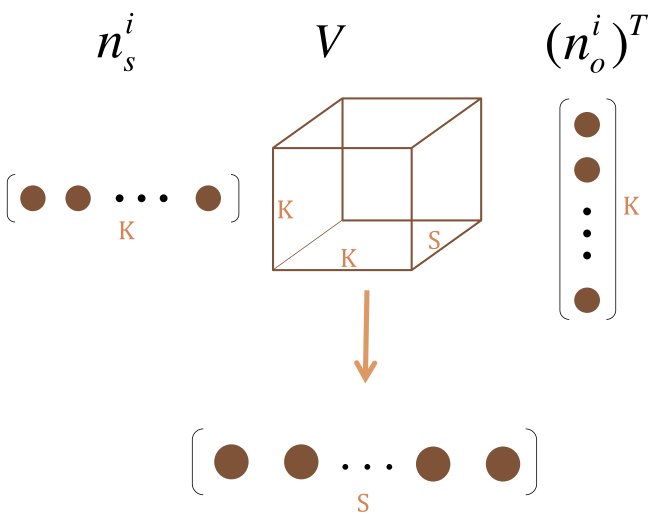
Context-based Sentence Vectors

$\mathbf{S}_{t-2}$: M. Atget captured the old Paris in his pictures. $\mathbf{S}_{t-1}$: His photographs show the city in its various facets. $\mathbf{S}_t$: He photographed stairwells and architectural details. $\mathbf{S}_{t+1}$: His interests also extended to the environs of Paris. $\mathbf{S}_{t+2}$: He also photographed street-hawkers and small tradesmen, as well as popular amusements.
IDist: stairwell, architectural, detail
DDist: capture, old, paris, picture, photograph, show, city, various, interest, extend, popular, amusement

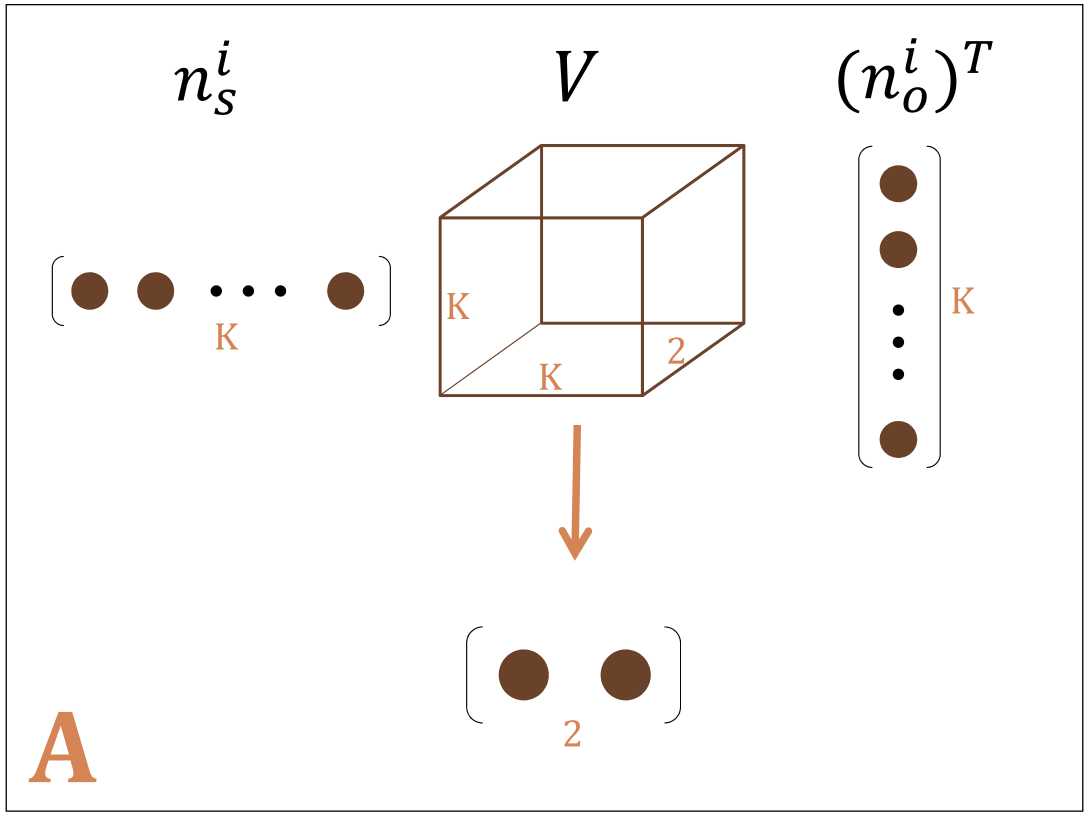


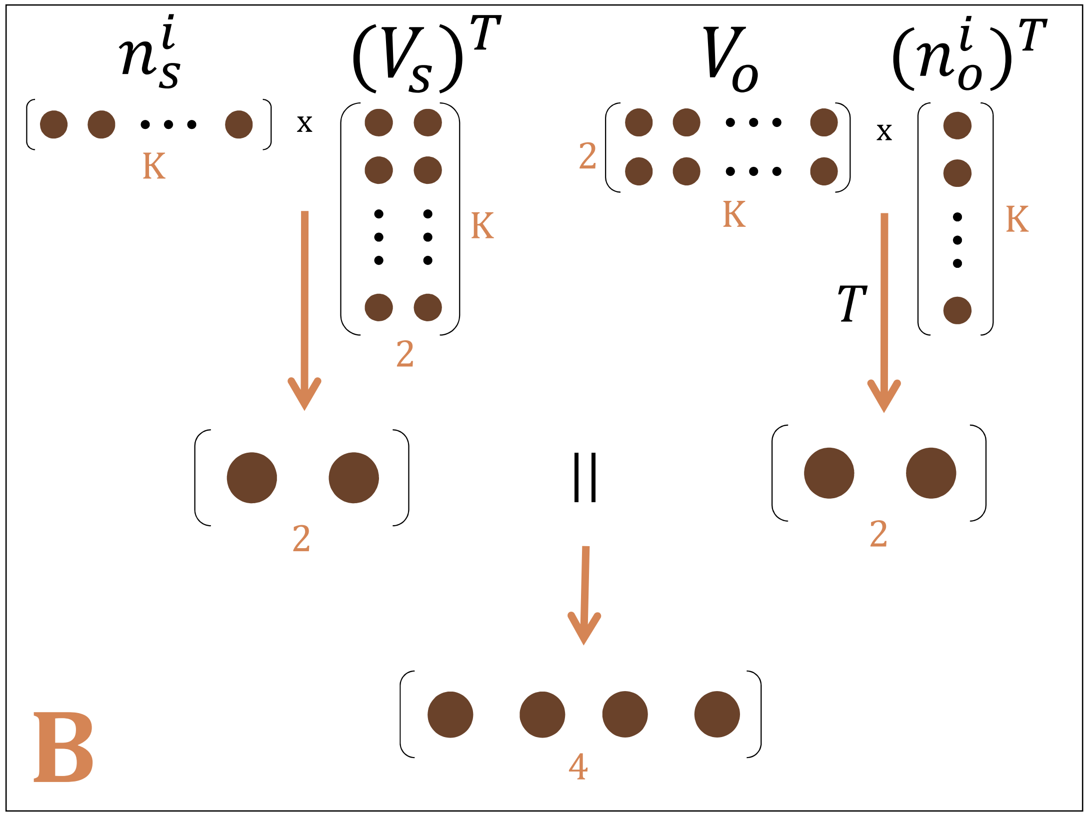
Context-based Sentence Vectors

$\mathbf{S}_{t-2}$: M. Atget captured the old Paris in his pictures. $\mathbf{S}_{t-1}$: His photographs show the city in its various facets. $\mathbf{S}_t$: He photographed stairwells and architectural details. $\mathbf{S}_{t+1}$: His interests also extended to the environs of Paris. $\mathbf{S}_{t+2}$: He also photographed street-hawkers and small tradesmen, as well as popular amusements.
IDist: stairwell, architectural, detail
DDist: capture, old, paris, picture, photograph, show, city, various, interest, extend, popular, amusement
DVerb: capture, show, extend, photograph



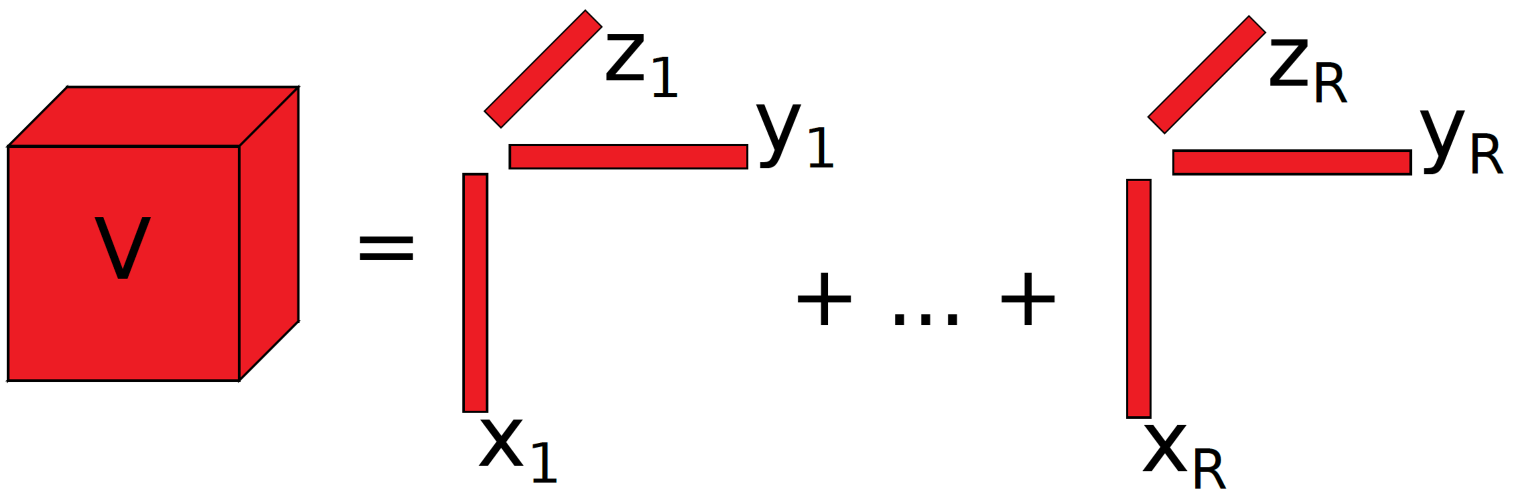
Simpler Matrix Networks



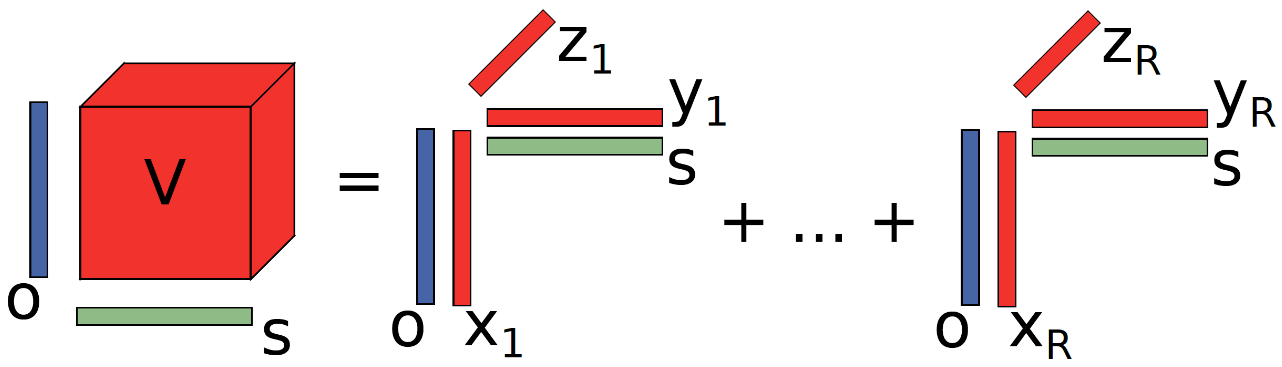


Polajnar, Fagarasan, Clark (EMNLP 2014); Paperno, Pham, Baroni (ACL 2014)

Low-Rank Approximations



$$\mathcal{V} = \sum_{r=1}^{R} \mathbf{x}_r \otimes \mathbf{y}_r \otimes \mathbf{z}_r$$



Fried, Polajnar, Clark (ACL 2015)

Summary

- Sentence vectors are here to stay
- Evaluation is problematic
- Recursive neural networks provide one solution
- Type-driven compositional framework is more linguistically motivated, but problematic in practice
- Maybe there is an ideal middle ground
- Composition has a role to play in other modalities (what does a red car look like, a bike with no wheels, a sleeveless dress, ...?)