The Effect of Polysemy in Compositional Semantics

Siva Reddy
Collaborators: Ioannis P. Klapaftis, Diana McCarthy, Spandana Gella, Suresh Manandhar

Artificial Intelligence Group,
Department of Computer Science,
University of York, UK

Lexical Computing Ltd., UK

Dept. of Intelligent Computer Systems
University of Malta
October 31, 2011

THE UNIVERSITY *of York*

Distributional Model: Meaning as a distributional vector

Distributional Hypothesis (Harris, 1954)

Words that occur in similar contexts tend to have similar meanings i.e. meaning of a word can be defined in terms of its context.

Word Space Model (WSM)

Meaning of a word is represented as a co-occurrence vector built from a corpus

	vector dimensions					
	animal	buy	apartment	price	rent	kill
House	⟨ 30	60	90	55	45	10 ⟩
Hunting	⟨ 90	15	12	20	33	90 ⟩

THE UNIVERSITY of York

Semantic Composition

How to compose the meaning of **house hunting** without using corpus instances of **house hunting**?

Semantic Composition

How to compose the meaning of **house hunting** without using corpus instances of **house hunting**?

Semantic Composition

The meaning of a complex expression can be defined by a function of the meanings of its constituents and its structure.

Semantic Composition Functions

- Several semantic composition functions are proposed to compose meaning of a phrase from its constituents (Mitchell and Lapata, 2008; Widdows, 2008; Erk and Padó, 2008)
- **House**$\oplus$**Hunting** is the meaning composed from **House** and **Hunting**
- $\oplus$ is the composition function
- Most successful $\oplus$s are simple addition (+) and simple multiplication (*) (Mitchell and Lapata, 2008; Vecchi et al., 2011)

	vector dimensions					
	animal	buy	apartment	price	rent	kill
House	⟨ 30	60	90	55	45	10 ⟩
Hunting	⟨ 90	15	12	20	33	90 ⟩
a.**House** + b.**Hunting**	⟨ 120	75	102	75	78	100 ⟩
House * Hunting	⟨ 2700	900	1080	1100	1485	900 ⟩

Polysemy of constituents is a problem

	vector dimensions					
	animal	buy	apartment	price	rent	kill
House	⟨ 30	60	90	55	45	10 ⟩
Hunting	⟨ 90	15	12	20	33	90 ⟩
a.**House** + b.**Hunting**	⟨ 120	75	102	75	78	100 ⟩
House * Hunting	⟨ 2700	900	1080	1100	1485	900 ⟩

Polysemy of constituents is a problem

Hunting-n have 3 senses in WordNet

1. Killing or capture of wild animals regarded as a sport
2. The activity of looking thoroughly in order to find something or someone
3. Killing or capturing animals for food or pelts

Static Prototype Vectors

- Existing compositional methods represent each word as a single vector, a prototype (Mitchell and Lapata, 2008; Widdows, 2008; Guevara, 2011)
- This vector conflates all the senses of a word

.

THE UNIVERSITY of York

Noisy Composition

Polysemy of constituents leads to
noisy composition away from true composition

THE UNIVERSITY of York

Solution to Noisy Composition

Sense Specific Prototype based Composition

1. Prototype vector for each of the senses of *house* and *hunting*
2. Sense specific prototypes to perform semantic composition

THE UNIVERSITY of York

Solution to Noisy Composition

Sense Specific Prototype based Composition

1. Prototype vector for each of the senses of *house* and *hunting*
2. Sense specific prototypes to perform semantic composition

Sense Specific Prototype vectors

- Static Multi Prototypes
- Dynamic Prototypes

THE UNIVERSITY *of York*

Solution to Noisy Composition

Sense Specific Prototype based Composition

1. Prototype vector for each of the senses of *house* and *hunting*
2. Sense specific prototypes to perform semantic composition

Sense Specific Prototype vectors

- Static Multi Prototypes
- Dynamic Prototypes

- We focus on Compound Nouns containing two nouns.

THE UNIVERSITY of York

Static Multi Prototypes (Klapaftis and Manandhar, 2010; Reisinger and Mooney, 2010)

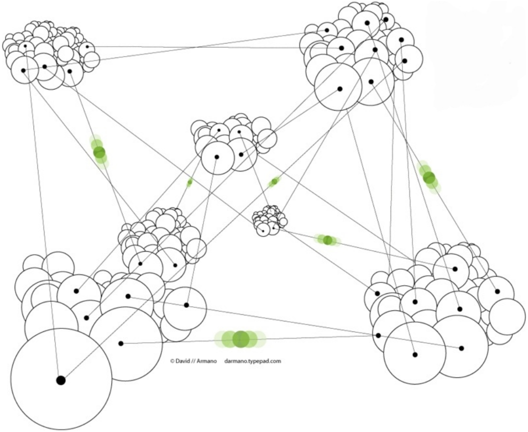
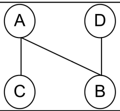


Figure: Word Sense Induction in a Graph based setting

THE UNIVERSITY of York

Static Multi Prototypes: Corpus Preprocessing

- Each target word's sentence (window of size ± 100) is
  - Tokenized
  - POS-tagged
  - lemmatized
- Nouns, verbs are kept
- Context words are weighted using log-likelihood (Dunning, 1993)
- Filtering out words < threshold
- Upper left of the figure

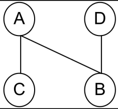
Target word: *mouse*	
Extracted nouns & verbs	**Extracted collocations**
A: device, windows, move, pc, **mouse**,...	A: {1, 2, 3, 4, 5, 6}
B: animal, move, cat, tail, **mouse**,.......	B: {7, 8, 9, 10, 11, 12}
C: computer, pc, windows, **mouse**,.......	C: {5, 13, 14}
D: mousetrap, catch, tail, **mouse**,.......	D: {15, 16, 17}
Collocations index	
1:device_windows, 2:device_move, 3:device_pc, 4:windows_move, 5:windows_pc, 6:move_pc, 7:animal_move, 8:animal_cat, 9:animal_tail, 10:move_tail, 11:move_cat, 12:cat_tail, 13: computer_pc, 14:computer_windows, 15: mousetrap_catch, 16:mousetrap_tail, 17:catch_tail	
Graph	
	

THE UNIVERSITY of York

Static Multi Prototypes: Graph Creation (1/3)

- Graph vertices:
  - Every target word's sentence as a vertex
- Graph Edges:
  - Given two vertices A & B
  - Collocational similarity
  - Bag of Words similarity

Target word: *mouse*	
Extracted nouns & verbs	**Extracted collocations**
A: device, windows, move, pc, **mouse**,...	A: {1, 2, 3, 4, 5, 6}
B: animal, move, cat, tail, **mouse**,.......	B: {7, 8, 9, 10, 11, 12}
C: computer, pc, windows, **mouse**,.......	C: {5, 13, 14}
D: mousetrap, catch, tail, **mouse**,.......	D: {15, 16, 17}
Collocations index	
1:device_windows, 2:device_move, 3:device_pc, 4:windows_move, 5:windows_pc, 6:move_pc, 7:animal_move, 8:animal_cat, 9:animal_tail, 10:move_tail, 11:move_cat, 12:cat_tail, 13: computer_pc, 14:computer_windows, 15: mousetrap_catch, 16:mousetrap_tail, 17:catch_tail	
Graph	

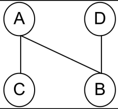


THE UNIVERSITY *of York*

Static Multi Prototypes: Graph Creation (2/3)

- In a sentence, each word is combined with every other word
  - Yielding collocations
  - Middle section of Figure
- Collocations weighted using log-likelihood (Dunning, 1993)
- Each sentence associated with a set of collocations
  - Upper-right of Figure

Target word: *mouse*	
Extracted nouns & verbs	**Extracted collocations**
A: device, windows, move, pc, **mouse**,...	A: {1, 2, 3, 4, 5, 6}
B: animal, move, cat, tail, **mouse**,.......	B: {7, 8, 9, 10, 11, 12}
C: computer, pc, windows, **mouse**,.......	C: {5, 13, 14}
D: mousetrap, catch, tail, **mouse**,.......	D: {15, 16, 17}
Collocations index	
1:device_windows, 2:device_move, 3:device_pc, 4:windows_move, 5:windows_pc, 6:move_pc, 7:animal_move, 8:animal_cat, 9:animal_tail, 10:move_tail, 11:move_cat, 12:cat_tail, 13: computer_pc, 14:computer_windows, 15: mousetrap_catch, 16:mousetrap_tail, 17:catch_tail	
Graph	



Static Multi Prototypes: Graph Creation (3/3)

- Bag of Words Weight of edge A-B
  - Jaccard Similarity between context word sets of A B
  - Upper-left of the figure
- Collocational weight of edge A-B
  - Jaccard Similarity between collocation sets of A B
  - Upper-right of the figure
- Sum of the above weights as edge weight

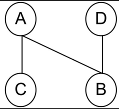
Target word: *mouse*

Extracted nouns & verbs	**Extracted collocations**
A: device, windows, move, pc, **mouse**,...	A: {1, 2, 3, 4, 5, 6}
B: animal, move, cat, tail, **mouse**,.......	B: {7, 8, 9, 10, 11, 12}
C: computer, pc, windows, **mouse**,.......	C: {5, 13, 14}
D: mousetrap, catch, tail, **mouse**,.......	D: {15, 16, 17}

Collocations index

1:device_windows, 2:device_move, 3:device_pc, 4:windows_move, 5:windows_pc, 6:move_pc, 7:animal_move, 8:animal_cat, 9:animal_tail, 10:move_tail, 11:move_cat, 12:cat_tail, 13: computer_pc, 14:computer_windows, 15: mousetrap_catch, 16:mousetrap_tail, 17:catch_tail

Graph



Static Multi Prototypes: Final Stage

- Chinese Whispers
  - Linear graph clustering method
  - Automatically identifies the number of clusters
- Parameter settings
  - Optimised to give best performance on SemEval 2007 (Agirre and Soroa, 2007)
- Final output is a set of clusters (senses) for a target word
  - Each cluster is a set of sentences

THE UNIVERSITY of York

Static Multi Prototypes

Exemplar

- Each sentence of a target word is represented as a vector, an exemplar
- Exemplar of hunting in the sentence *the-x purpose-n of-i autumn-n hunting-n be-v in-i part-n to-x cull-v the-x number-n of-i young-j autumn-n fox-n* is $\langle$ purpose-n:1; autumn-n:2; part-n:1; cull-v; number-n:1; young-j:1; fox-n:1 $\rangle$

Static Multi Prototypes

- A prototype is defined for each sense
  - Centroid of all the exemplars in a sense cluster
- Multiple prototypes per word
- Static because multiple prototypes are always fixed for a word

THE UNIVERSITY *of York*

Static Multi Prototype Based Composition

Which Prototypes to select for composition?

- *house* -> m senses
- *hunting* -> n senses
- Which one to choose from *house* and *hunting* for composition

We tried many variations

- Choose the most similar senses from each other
  - Similar to Lesk algorithm (Lesk, 1986)
  - Drawback: Always preferred small sized clusters
- Choose most similar senses from 5/10 large clusters of each
  - Better than the previous
- Choose a word sense most similar to compound's distributional vector
  - Guided selection since compound corpus instances are used
  - Idea of upper bound performance

Dynamic Prototypes

Dynamic Prototype of a word

- Is not based on a fixed sense inventory
  - Static Multi Prototypes have a fixed sense inventory
- Sense inventories fail to capture multi shades of senses
- *I don't believe in word senses* (Kilgarriff, 1997)
  - We don't believe in fixed sense inventories
- On-the-fly sense representation relevant to a given context
  - In *house hunting*, the context of *hunting* is *house* and vice-versa

Dynamic Prototype of a word

- Exemplar-based (memory-based) modeling (Erk and Padó, 2010; Smith and Medin, 1981)
  - Represent a word by all its exemplars (sentences) rather than a single prototype
- Select only the relevant exemplars of a target word based on its context
- Build a prototype vector of the target word from the refined exemplar set
- Dynamic because prototype vector of the target word changes with change in context

Exemplars of *hunting*

both factions to enjoy the social side of	**hunting**	with no obvious detrimental effects. The
Greefswald region was primarily used for	**hunting**	from around the turn of the century until
they are now hunting the traditional drag	**hunting**	in the traditional way. No matter how this
keep their horses exclusively for going fox	**hunting**	. Publicans and their staff welcome the
ride horses which were bred locally for	**hunting**	and the owners also breed replacements.
country houses and in the popularity of	**hunting**	. 3.4. Concerning design, the original country
nmunities were able to repeat the bond that	**hunting**	and farming have the world would be a far
everyone loves the countryside and hates	**hunting**	." (4) The suggested job losses in associated
about the advantages and disadvantages of	**hunting**	with dogs in terms of agriculture and pest
ything up to 15 miles with the dog working (	**hunting**	) ground and cover in front of the guns.

Figure: A random concordance of *hunting* from ukWaC (Ferraresi et al., 2008)

Exemplars of *hunting*

both factions to enjoy the social side of	**hunting**	with no obvious detrimental effects. The
Greefswald region was primarily used for	**hunting**	from around the turn of the century until
they are now hunting the traditional drag	**hunting**	in the traditional way. No matter how this
keep their horses exclusively for going fox	**hunting**	. Publicans and their staff welcome the
ride horses which were bred locally for	**hunting**	and the owners also breed replacements.
country houses and in the popularity of	**hunting**	. 3.4. Concerning design, the original country
nmunities were able to repeat the bond that	**hunting**	and farming have the world would be a far
everyone loves the countryside and hates	**hunting**	." (4) The suggested job losses in associated
ıbout the advantages and disadvantages of	**hunting**	with dogs in terms of agriculture and pest
ything up to 15 miles with the dog working (	**hunting**	) ground and cover in front of the guns.

Figure: A random concordance of *hunting* from ukWaC (Ferraresi et al., 2008)

- None of the exemplars are related to sense of *hunting* in *house hunting*
- Skewed by most frequent sense of *hunting*

THE UNIVERSITY of York

Dynamic Prototype vector **Hunting$^{\textbf{House}}$**

Hunting$^{\textbf{House}}$: The prototype vector of *hunting* in the presence of *house*

- Choose only the exemplars of *hunting* which have context words related to *house*
  - Reason: Distributional vector of *house hunting* is likely to have words related to both *house* and *hunting*
- We rank each exemplar of *hunting* using
  - Collocations of *house*
  - Distributionally similar words of *house*

THE UNIVERSITY *of York*

Collocations of *house*

object_of	97056	2.3
terrace	1729	9.1
build	8408	9.03
detach	1759	8.99
buy	3960	8.33
board	846	7.95
rent	929	7.95
sell	2470	7.87
situate	1050	7.86
demolish	644	7.58
own	1281	7.53
move	2444	7.51
occupy	789	7.32
leave	2926	7.13
enter	1307	7.07
decorate	471	6.95
destroy	592	6.76

subject_of	59167	2.5
belong	316	6.72
stand	736	6.65
overlook	243	6.35
date	266	5.89
rebuild	131	5.69
front	88	5.48
burn	115	5.17
sit	226	5.12
occupy	131	5.11
line	84	5.03
consist	125	4.96
lie	178	4.94
boast	76	4.82
comprise	118	4.75
survive	105	4.75
collapse	60	4.75

adj_subject_of	8329	2.3
uninhabited	70	7.86
adjoining	126	7.79
repossessed	34	6.95
unoccupied	38	6.84
empty	194	6.76
habitable	28	6.53
adjacent	77	6.08
tidy	35	6.01
clean	136	5.68
vacant	24	5.49
worth	222	5.49
uninhabitable	12	5.45
spotless	12	5.36
situate	11	5.34
semi-detached	13	5.32
spacious	35	5.27

modifier	160373	1.7
manor	2330	8.74
guest	2485	8.18
publishing	1416	7.86
Victorian	1330	7.79
public	4559	7.76
bedroom	1715	7.71
dwelling	1196	7.67
old	3959	7.42
Georgian	848	7.36
semi-detached	816	7.36
auction	919	7.32
historic	1011	7.2
private	1798	7.15
opera	768	7.07
coffee	942	6.98

House^{colloc}: Collocational vector of *house*

- Computed using logDice (Curran, 2003)
- *terrace, build, rent* ... occur with *house hunting*

Distributional similar words of *house*

Lemma	Score	Freq
building	0.534	363768
home	0.483	675005
room	0.461	364176
garden	0.44	171248
church	0.432	253000
shop	0.421	171029
town	0.413	260679
property	0.412	329119
area	0.409	1103121
office	0.407	289728
village	0.398	169340
car	0.397	419404
hotel	0.396	131472
centre	0.395	334158
site	0.393	915103

House$^{\text{similar}}$: Distributional neighbors of hunting

- Computed using (Rychlý and Kilgarriff, 2007)
- Provide more evidence - home hunting, room hunting, flat hunting etc

Dynamic Prototype **Hunting**$^{\textbf{House}}$

both factions to enjoy the social side of	**hunting**	with no obvious detrimental effects. The
Greefswald region was primarily used for	**hunting**	from around the turn of the century until
they are now hunting the traditional drag	**hunting**	in the traditional way. No matter how this
keep their horses exclusively for going fox	**hunting**	. Publicans and their staff welcome the
ride horses which were bred locally for	**hunting**	and the owners also breed replacements.
country houses and in the popularity of	**hunting**	. 3.4. Concerning design, the original country
nmunities were able to repeat the bond that	**hunting**	and farming have the world would be a far
everyone loves the countryside and hates	**hunting**	." (4) The suggested job losses in associated
ıbout the advantages and disadvantages of	**hunting**	with dogs in terms of agriculture and pest
ything up to 15 miles with the dog working (	**hunting**	) ground and cover in front of the guns.

THE UNIVERSITY of York

Dynamic Prototype **Hunting**$^{\textbf{House}}$

both factions to enjoy the social side of	**hunting**	with no obvious detrimental effects. The
Greefswald region was primarily used for	**hunting**	from around the turn of the century until
they are now hunting the traditional drag	**hunting**	in the traditional way. No matter how this
keep their horses exclusively for going fox	**hunting**	. Publicans and their staff welcome the
ride horses which were bred locally for	**hunting**	and the owners also breed replacements.
country houses and in the popularity of	**hunting**	. 3.4. Concerning design, the original country
nmunities were able to repeat the bond that	**hunting**	and farming have the world would be a far
everyone loves the countryside and hates	**hunting**	." (4) The suggested job losses in associated
ıbout the advantages and disadvantages of	**hunting**	with dogs in terms of agriculture and pest
ything up to 15 miles with the dog working (	**hunting**	) ground and cover in front of the guns.

Rank each exemplar **e** of *hunting* using *house*

- $sim(\mathbf{e}, \mathbf{House}^{\mathbf{colloc}}) + sim(\mathbf{e}, \mathbf{House}^{\mathbf{similar}})$
- sim is Cosine similarity

THE UNIVERSITY *of York*

{'search-n': 1.0, 'week-n': 1.0, 'document-n': 1.0, 'property-n': 2.0, 'translation-n': 1.0}
{'locate-v': 1.0, 'area-n': 2.0, 'build-v': 1.0, 'town-n': 1.0, 'home-n': 1.0, 'fishing-n': 1.0}
{'area-n': 2.0, 'mountain-n': 1.0, 'sale-n': 1.0, 'town-n': 1.0, 'km-n': 1.0, 'home-n': 1.0, 'fishing-n': 1.0}
{'situate-v': 1.0, 'area-n': 2.0, 'town-n': 1.0, 'countryside-n': 1.0, 'village-n': 1.0, 'nice-j': 1.0, 'property-n': 1.0}
{'boost-v': 1.0, 'home-n': 2.0, 'buyer-n': 1.0, 'lack-n': 1.0, 'price-n': 2.0, 'drive-v': 1.0, 'house-n': 1.0}
{'land-n': 1.0, 'market-n': 1.0, 'country-n': 1.0, 'enthusiast-n': 1.0, 'live-v': 1.0}
{'locate-v': 1.0, 'area-n': 1.0, 'mountain-n': 1.0, 'town-n': 1.0, 'lovely-j': 1.0, 'highway-n': 1.0}
{'village-n': 1.0, 'house-n': 1.0, 'area-n': 1.0, 'manor-n': 1.0, 'control-v': 1.0}
{'area-n': 1.0, 'home-n': 1.0, 'spring-n': 1.0, 'sale-n': 2.0, 'sell-v': 1.0, 'property-n': 2.0, 'water-n': 1.0}

Figure: Ranked exemplars of *hunting-n* w.r.t. *house-n*

{'search-n': 1.0, 'week-n': 1.0, 'document-n': 1.0, 'property-n': 2.0, 'translation-n': 1.0}
{'locate-v': 1.0, 'area-n': 2.0, 'build-v': 1.0, 'town-n': 1.0, 'home-n': 1.0, 'fishing-n': 1.0}
{'area-n': 2.0, 'mountain-n': 1.0, 'sale-n': 1.0, 'town-n': 1.0, 'km-n': 1.0, 'home-n': 1.0, 'fishing-n': 1.0}
{'situate-v': 1.0, 'area-n': 2.0, 'town-n': 1.0, 'countryside-n': 1.0, 'village-n': 1.0, 'nice-j': 1.0, 'property-n': 1.0}
{'boost-v': 1.0, 'home-n': 2.0, 'buyer-n': 1.0, 'lack-n': 1.0, 'price-n': 2.0, 'drive-v': 1.0, 'house-n': 1.0}
{'land-n': 1.0, 'market-n': 1.0, 'country-n': 1.0, 'enthusiast-n': 1.0, 'live-v': 1.0}
{'locate-v': 1.0, 'area-n': 1.0, 'mountain-n': 1.0, 'town-n': 1.0, 'lovely-j': 1.0, 'highway-n': 1.0}
{'village-n': 1.0, 'house-n': 1.0, 'area-n': 1.0, 'manor-n': 1.0, 'control-v': 1.0}
{'area-n': 1.0, 'home-n': 1.0, 'spring-n': 1.0, 'sale-n': 2.0, 'sell-v': 1.0, 'property-n': 2.0, 'water-n': 1.0}

Figure: Ranked exemplars of *hunting-n* w.r.t. *house-n*

Hunting^{House}

- Select top n% ranked exemplars
- Centroid of all the selected exemplars
- Prototype of *hunting* in the presence of *house*

THE UNIVERSITY *of York*

Dynamic Prototype Vector based Composition

$$\mathbf{House}^{\mathbf{Hunting}} \oplus \mathbf{Hunting}^{\mathbf{House}}$$

THE UNIVERSITY of York

Composition Functions $\oplus$

$$
\begin{aligned}
\text{ADD:} \quad & \oplus(\mathbf{N}) = \alpha\ \mathbf{n} + \beta\ \mathbf{n}' \\
& \text{i.e. } \oplus(\mathbf{N})_i = \alpha\ \mathrm{n}_i + \beta\ \mathrm{n}'_i \\
\\
\text{MULT:} \quad & \oplus(\mathbf{N}) = \mathbf{n} * \mathbf{n}' \\
& \text{i.e. } \oplus(\mathbf{N})_i = \mathrm{n}_i * \mathrm{n}'_i
\end{aligned}
\qquad (1)
$$

- N is a compound noun with constituent n and $n^{'}$
- **n** represents a sense prototype vector for n
- n_i the value of n^{th} cooccurrence in the vector **n**

THE UNIVERSITY *of York*

Evaluation Setting: Phrase Similarity Task (Mitchell and Lapata, 2010)

Annotator	N	N'	rating
4	phone call	committee meeting	2
25	phone call	committee meeting	7
11	football club	league match	6
11	health service	bus company	1
14	company director	assistant manager	7

Table: Evaluation dataset of (Mitchell and Lapata, 2010)

- 108 compound noun pairs
- 7 annotators judge each pair for phrase similarity
- Score range: 0-7

Evaluation Setting: Phrase Similarity Task

- Model's phrase similarity prediction $\mathrm{sim}(\oplus(N), \oplus(N'))$
  - i.e. the similarity between composed vectors
  - sim is Cosine similarity
- Correlation between model prediction scores and mean of human judgments

THE UNIVERSITY of York

	ADD	MULT
Static Prototypes (not sense based)		
	0.5173	0.6104
Static Multi Prototypes		
Top 5 clusters	0.1171	0.4150
Top 10 clusters	0.0663	0.2655
Static Multi Prototypes with Guided Selection		
Top 5 clusters	0.2290	0.4187
Top 10 clusters	0.2710	0.4140

Table: Spearman Correlation of Model predictions with Human Judgments

- Static Multi Prototypes worse than normal composition
- Reasons: Is it because of Sense Selection process?
  - But guided is the upper bound
- Is it because of Clustering algorithm
  - Not possibly. May be in our graph setting (verbs are highly polysemous)
- Selecting multiple senses rather than single sense may help

	ADD	MULT
Static Prototypes (not sense based)		
	0.5173	0.6104
Dynamic Prototypes		
Top 2 % exemplars	0.6261	0.6552
Top 5 % exemplars	0.6326	0.6478
Top 10 % exemplars	0.6402	0.6515
Top 20 % exemplars	0.6273	0.6359
Top 50 % exemplars	0.5948	0.6340
Distributional Prototype of the Compound		
0.4152		

Table: Spearman Correlation of Model predictions with Human Judgments

- Dynamic Prototypes show clear upper hand
  - Sense disambiguation is useful for semantic composition
- Better than distributional prototype of the compound
  - Composition solves data sparsity
- Word sense can be modelled with very few exemplars
  - With increase in exemplars, noise increases

THE UNIVERSITY of York

Take-away Message

- Sense disambiguation helps Semantic Composition
- Dynamic Prototypes are better than Static Prototypes
- Dynamic Prototypes capture context sensitive meaning
- Semantic Composition solves data sparsity problem

Bibliography I

Agirre, E. and Soroa, A. (2007). Semeval-2007 task 02: Evaluating word sense induction and discrimination systems. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 7–12, Prague, Czech Republic. Association for Computational Linguistics.

Curran, J. R. (2003). From distributional to semantic similarity. Technical report, PhD Thesis, University of Edinburgh.

Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1):61–74.

Erk, K. and Padó, S. (2008). A structured vector space model for word meaning in context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 897–906, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bibliography II

Erk, K. and Padó, S. (2010). Exemplar-based models for word meaning in context. In *Proceedings of the ACL 2010 Conference Short Papers*, ACLShort '10, pages 92–97, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ferraresi, A., Zanchetta, E., Baroni, M., and Bernardini, S. (2008). Introducing and evaluating ukWaC, a very large web-derived corpus of english. In *Proceedings of the WAC4 Workshop at LREC 2008*, Marrakesh, Morocco.

Guevara, E. R. (2011). Computing semantic compositionality in distributional semantics. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '2011.

Harris, Z. S. (1954). Distributional structure. *Word*, 10:146–162.

Kilgarriff, A. (1997). I don't believe in word senses. In *Computers and the Humanities, 31(2):91-113.*

Bibliography III

Klapaftis, I. and Manandhar, S. (2010). Word sense induction & disambiguation using hierarchical random graphs. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 745–755, Cambridge, MA. Association for Computational Linguistics.

Lesk, M. (1986). Automated Sense Disambiguation Using Machine-readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In *Proceedings of the AAAI Fall Symposium Series*, pages 98–107.

Mitchell, J. and Lapata, M. (2008). Vector-based Models of Semantic Composition. In *Proceedings of ACL-08: HLT*, pages 236–244, Columbus, Ohio. Association for Computational Linguistics.

Mitchell, J. and Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science*.

Reisinger, J. and Mooney, R. J. (2010). Multi-prototype vector-space models of word meaning. In *HLT-NAACL*, pages 109–117.

THE UNIVERSITY of York

Bibliography IV

Rychlý, P. and Kilgarriff, A. (2007). An efficient algorithm for building a distributional thesaurus (and other sketch engine developments). In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, ACL '07, pages 41–44, Stroudsburg, PA, USA. Association for Computational Linguistics.

Smith, E. E. and Medin, D. L. (1981). *Categories and concepts / Edward E. Smith and Douglas L. Medin*. Harvard University Press, Cambridge, Mass. :.

Vecchi, E. M., Baroni, M., and Zamparelli, R. (2011). (linear) maps of the impossible: Capturing semantic anomalies in distributional space. In *Proceedings of the Workshop on Distributional Semantics and Compositionality*, pages 1–9, Portland, Oregon, USA. Association for Computational Linguistics.

Widdows, D. (2008). Semantic vector products: Some initial investigations. In *Second AAAI Symposium on Quantum Interaction, Oxford*.