Generative Distributional Models

Siva Reddy

Lexical Computing Ltd, UK
http://sketchengine.co.uk

University of Edinburgh
May 22 2012

in collaboration with Diana McCarthy

and also with:
Spandana Gella, Ioannis Klapaftis, Suresh Manandhar

Caution

- This talk is not about
  - Probabilistic Distributional Models

Outline

1. Compositional Semantics
   - Generative Distributional Models
2. Compositionality in language
   - Analysis on the Data
   - Computational Models
3. Context Aware Composition
4. Generative Distributional Grammar

Outline

Compositional Semantics

- Can you infer meaning of an expression from its parts?
- Lexical Semantics - meaning of basic units like words
- Compositional Semantics - beyond words from words

Compositional Semantics

- Can you infer meaning of an expression from its parts?
- Lexical Semantics - meaning of basic units like words
- Compositional Semantics - beyond words from words

Compositional Semantics

The Principle of Semantic Compositionality (PoC, Partee, 1995)

The meaning of a complex expression is determined by the meanings of its constituents and its structure

Example

Compound Noun	swimming pool
Adjective Noun	blue sky
Subject Verb	flies fly
Verb Object	lose keys
Verb Particle	*climb up* the hill

Compositional Semantics

Two lines of research

1. Formal Semantics
2. Distributional Semantics
   - So far under the umbrella of lexical semantics

Formal Semantics

Foundation

Human language can be modeled within a mathematically precise theory (Montague, 1970).

Methodology

- Meaning of a word as a symbol
- Meaning of an expression as a interaction between symbols
- e.g. *every man walks* is represented as $\forall u[man(u) \implies walk(u)]$ (Montague, 1973)
- Pros: Generative model for describing language

Major drawback

Concerns only with truth and falsity of an expression. Cannot quantify similarity between two expressions

Formal Semantics

Foundation

Human language can be modeled within a mathematically precise theory (Montague, 1970).

Methodology

- Meaning of a word as a symbol
- Meaning of an expression as a interaction between symbols
- e.g. *every man walks* is represented as $\forall u[man(u) \implies walk(u)]$ (Montague, 1973)
- Pros: Generative model for describing language

Major drawback

Concerns only with truth and falsity of an expression. Cannot quantify similarity between two expressions

Formal Semantics

Foundation

Human language can be modeled within a mathematically precise theory (Montague, 1970).

Methodology

- Meaning of a word as a symbol
- Meaning of an expression as a interaction between symbols
- e.g. *every man walks* is represented as $\forall u[man(u) \implies walk(u)]$ (Montague, 1973)
- Pros: Generative model for describing language

Major drawback

Concerns only with truth and falsity of an expression. Cannot quantify similarity between two expressions

Distributional Semantics

Foundation: Distributional Hypothesis (Harris, 1954)

- The words that occur in similar contexts tend to have similar meaning
- Backbone for *Vector Space Model of Semantics*.

Firth (Firth, 1957)

- You shall know a word from its context - Firth's Principle

Vector Space Model of Semantics

Meaning as a vector (Schütze, 1998)

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

	police-n	photon-n	speed-n	car-n	soul-n
Traffic	142	0	293	347	1
Light	41	29	222	198	50
TrafficLight	5	0	13	48	0

Vector Space Model of Semantics

Meaning as a vector (Schütze, 1998)

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

	police-n	photon-n	speed-n	car-n	soul-n
Traffic	142	0	293	347	1
Light	41	29	222	198	50
TrafficLight	5	0	13	48	0

Outline

Generative Distributional Models (GDMs)

Making distributional models generative - beyond words from words

Variation of Principle of Compositionality (PoC)

Distributional representation of an expression can be composed from the distributional representation of its parts and its structure

Name inspired from Pustejovsky (1991)'s Generative Lexicon

Generative Distributional Models (GDMs)

Making distributional models generative - beyond words from words

Variation of Principle of Compositionality (PoC)

Distributional representation of an expression can be composed from the distributional representation of its parts and its structure

Name inspired from Pustejovsky (1991)'s Generative Lexicon

Generative Distributional Models (GDMs)

Making distributional models generative - beyond words from words

Variation of Principle of Compositionality (PoC)

Distributional representation of an expression can be composed from the distributional representation of its parts and its structure

Name inspired from Pustejovsky (1991)'s Generative Lexicon

Composition Function

`meaning(w1 w2) = f(w1, w2,` γ, ε`)` (Mitchell and Lapata, 2008)

- `w1 w2` is an expression with words `w1` and `w2`, and structure γ
- `f` is the composition function
- ε represents world knowledge (not mentioned by PoC)
- Goal of GDMs is to break the code behind f.

Challenges

- Structure
  - house rent
  - rent house
- Polysemy
  - football coach
  - air-conditioned coach
- Idiosyncrasy
  - melting pot
  - cooking pot
- Metonymy
  - everybody reads Shakespeare at school

Evaluation of GDMs

- Intrinsic evaluation
  - similarity ($\overrightarrow{\textbf{house hunting}}$, $\overrightarrow{\oplus(\textbf{house hunting})}$)
  - A combination of Distributional Hypothesis and Principle of Compositionality
- Extrinsic evaluation
  - Compositionality Detection
  - Phrasal Similarity

Outline

Compositionality in language

How do humans detect compositionality?

- Do we look at each constituent separately?
- Study on human annotations
  - Contribution of each constituent to the phrasal semantics
  - Relation between constituent contribution to the phrase compositionality
- Computational Models

Noun-noun compounds

with Diana McCarthy and Suresh Manandhar (Reddy et al., 2011b)

roast potato vs couch potato

- A unique dataset with:
  - compositionality judgment of phrase and both constituents in phrase
  - use data to examine relation in the gold-standard
- Two types of computational models
  - constituent based
  - composition function based

Existing Datasets

Resource	Phrase Types	# Anns	# Phrases	Jdgm
MKC	V+part	4	117	phr(1-10)
BBL	V+part	28	40	const(+-)
VJ	V+Obj	2	800	phr(1-6)
BG	V+{Obj,Subj} Adj+N	20	145	phr(1-10)
KS	NN	1	38	phr(+-)

- MKC McCarthy et al. (2003),
- BBL Bannard et al. (2003)
- VJ Venkatapathy and Joshi (2005)
- BG Biemann and Giesbrecht (2011),
- KS Korkontzelos and Manandhar (2009)

Data (rationale)

- compound nouns containing two words
  - no existing dataset with compositionality
  - relatively simple since no morphological or syntactic variations
- constituent scores with phrase level compositionality scores; examine the relation
- balance data; examine score distribution

Compound Noun Set

90 compounds from four different classes - extracted semi-automatically

1. Both words are literal
   - swimming pool
2. First word is literal and second is non-literal
   - night owl
3. First word is non-literal and second literal
   - zebra crossing
4. Both words non-literal
   - smoking gun

Experimental Setup

Three tasks per compound

1. is the phrase literal?
2. is the first constituent used literally in the given phrase?
3. is the second constituent used literally in the given phrase?

- Each task annotated by 30 random annotators out of 151 annotators
- Total 8100 annotations (90 * 3 * 30 = 8100)
- 5 random examples from ukWaC (Ferraresi et al., 2008)

How literal is this phrase?

Sample examples at http://tinyurl.com/is-it-lit

web site:

Definitions:

1. a computer connected to the internet that maintains a series of web pages on the World Wide Web

Examples:

1. can simply update the firmware and modem drivers by downloading patches from the modem manufacturers **web site** . It may be best to contact the manufacturers of your modem in the first
2. up with the Government position here (mainly pro-badger killing) , visit the DEFRA **web site** , and use the search function to trace papers about badgers and tuberculosis . Action
3. of galaxy formation and evolution and of the enrichment of the intergalactic medium . This **web site** is part of a research project by Graham Thurgood who is a senior lecturer .
4. of use represent the complete and only statement of the terms of use of this **web site** . 4 . My Portfolio within the Financial Organiser Friends Provident receives its data feed
5. Courts . If you require to contact us in regard to the content of this **web site** or with a view to obtaining consent from the University to use the material contained

Note: Please select the answers below carefully based on the definition which occurs frequently in the examples

Step 1: score of 0-5 for how literal is the use of **"web"** in the phrase **"web site"**

○0 ○1 ○2 ○3 ○4 ○5

Please provide any comments in case you want to tell us about your judgement or any other queries/suggestions! Not Mandatory but helpful.

Annotation

No. of turkers participated	260
No. of them qualified	151
'Spammers' $\rho <= 0$	21
Turkers with $\rho >= 0.6$	81
annotations rejected	383

Table: Amazon Mechanical Turk statistics

Compound	**Word1**	**Word2**	**Phrase**
climate change	4.90$\pm$0.30	4.83$\pm$0.38	4.97$\pm$0.18
search engine	4.62$\pm$0.96	2.25$\pm$1.70	3.32$\pm$1.16
face value	1.39$\pm$1.11	4.64$\pm$0.81	3.04$\pm$0.88
blame game	4.61$\pm$0.67	2.00$\pm$1.28	2.72$\pm$0.92
sitting duck	1.48$\pm$1.48	0.41$\pm$0.67	0.96$\pm$1.04

Table: Compounds with their constituent and phrase level mean$\pm$st. dev scores

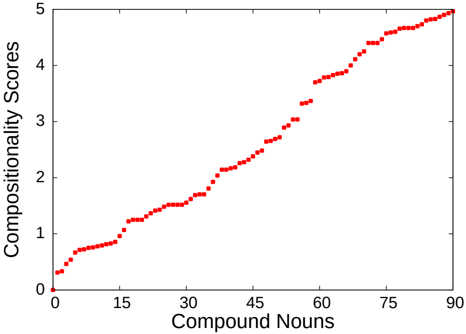
Agreement: Spearman's correlation

	highest ρ	avg. ρ
ρ for phrase compositionality	0.741	0.522
ρ for first word's literality	0.758	0.570
ρ for second word's literality	0.812	0.616
ρ all three tasks	0.788	0.589

Outline

Phrase Compositionality

a continuum



NB we targeted 4 classes

Relation between Constituent and Phrase Compositionality Scores

We tried various functions to model the human judgments

- ADD: $a.s1 + b.s2 = s3$
- MULT: $a.s1.s2 = s3$
- COMB: $a.s1 + b.s2 + c.s1.s2 = s3$
- WORD1: $a.s1 = s3$
- WORD2: $a.s2 = s3$
    - s1 and s2: contributions from first and second constituent resp.
    - s3: phrase compositionality score

Study on human judgments

Function f	ρ
ADD	0.966
MULT	0.965
COMB	0.971
WORD1	0.767
WORD2	0.720

Table: Spearman Correlation ρ between functions and phrase compositionality scores

- Both the words determine compositionality
- The phrase score can be predicted from the constituents scores

Study on human judgments

Function f	ρ
ADD	0.966
MULT	0.965
COMB	0.971
WORD1	0.767
WORD2	0.720

Table: Spearman Correlation ρ between functions and phrase compositionality scores

- Both the words determine compositionality
- The phrase score can be predicted from the constituents scores

Outline

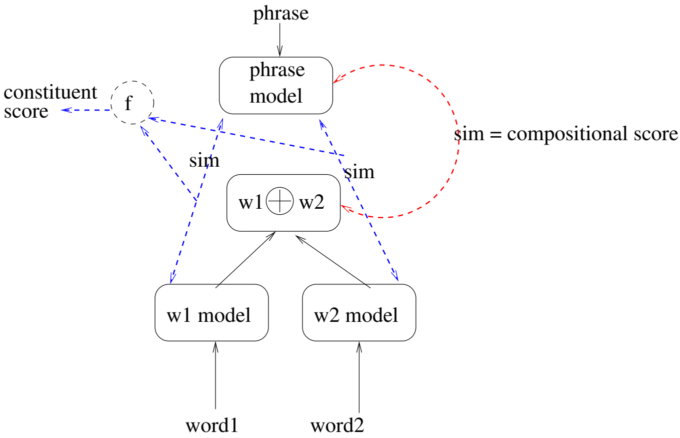
Computational Models for Compositionality

- Constituent based models
  - determine the literality of each constituent
  - use literality score of each constituent to predict phrase compositionality score
- Composition function based models
  - build a compositional model of a phrase using its constituents
  - similarity between the composed model and phrase model gives phrase compositionality score

Computational Models for Compositionality

- Constituent based models
  - determine the literality of each constituent
  - use literality score of each constituent to predict phrase compositionality score
- Composition function based models
  - build a compositional model of a phrase using its constituents
  - similarity between the composed model and phrase model gives phrase compositionality score

Computational Models for Compositionality



Constituent Based Models

s3 = f(s1, s2)

If a constituent word is used literally in a given compound it is likely that the compound and the constituent share common co-occurrences e.g. swimming in swimming pool.

Literality of a Constituent

- s1= sim(v1, v3); s2= sim(v2, v3)
- sim is Cosine Similarity.

	human judgments	
	first constituent	second constituent
s1	0.616	–
s2	–	0.707

Composition Function based models

s3= sim(v1 $\oplus$ v2, v3)

- Mitchell and Lapata (2008); Widdows (2008); Erk and Padó (2008)
- e.g. **Traffic$\oplus$Light** is the meaning composed from **Traffic** and **Light**
- $\oplus$ is the composition function
- simple addition and simple multiplication (Mitchell and Lapata, 2008)

	police-n	photon-n	speed-n	car-n	soul-n
v1 Traffic	142	0	293	347	1
v2 Light	41	29	222	198	50
v3 TrafficLight	5	0	13	48	0
a**Traffic** + b**Light**	183	29	515	545	51
Traffic * Light	5822	0	65046	68706	50

Results for Computational Models

Phrase level correlations

Model	ρ
Constituent Based Models s3 = f(s1,s2)	
ADD	**0.686**
MULT	0.670
COMB	0.682
WORD1	0.669
WORD2	0.515
Composition Function Based Models s3= sim(v1 $\oplus$ v2, v3)	
$a\mathbf{v1} + b\mathbf{v2}$	**0.714**
v1v2	0.650
RAND	0.002

Findings

- both types of models competitive
- additive composition models best
- Possible reasons
  - constituent based models use contextual information of each constituent *independently*
  - composition function models use contexts of both the constituents *simultaneously*
  - perhaps contexts salient to both the words are important?

Outline

Current composition methods

- Represent a word as a single prototype vector
- e.g. $\overrightarrow{\textbf{house}} \oplus \overrightarrow{\textbf{hunting}}$

	vector dimensions					
	animal	buy	apartment	price	rent	kill
$\overrightarrow{\textbf{house}}$	⟨ 30	60	90	55	45	10 ⟩
$\overrightarrow{\textbf{hunting}}$	⟨ 90	15	12	20	33	90 ⟩
$a.\overrightarrow{\textbf{house}} + b.\overrightarrow{\textbf{hunting}}$	⟨ 120	75	102	75	78	100 ⟩
$\overrightarrow{\textbf{house}} * \overrightarrow{\textbf{hunting}}$	⟨ 2700	900	1080	1100	1485	900 ⟩

Current composition methods

- Represent a word as a single prototype vector
- e.g. $\overrightarrow{\textbf{house}} \oplus \overrightarrow{\textbf{hunting}}$

	vector dimensions					
	animal	buy	apartment	price	rent	kill
$\overrightarrow{\textbf{house}}$	⟨ 30	60	90	55	45	10 ⟩
$\overrightarrow{\textbf{hunting}}$	⟨ 90	15	12	20	33	90 ⟩
$a.\overrightarrow{\textbf{house}} + b.\overrightarrow{\textbf{hunting}}$	⟨ 120	75	102	75	78	100 ⟩
$\overrightarrow{\textbf{house}} * \overrightarrow{\textbf{hunting}}$	⟨ 2700	900	1080	1100	1485	900 ⟩

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)

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

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*

Dynamic Prototypes

with Ioannis Klapaftis, Diana McCarthy, Suresh Manandhar (Reddy et al., 2011a)

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 meaning
- *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 vector $\overrightarrow{\textbf{hunting}}^{house}$

$\overrightarrow{\textbf{hunting}}^{house}$: 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*

Collocations of *house*

object_of	97056	2.3	subject_of	59167	2.5	adj_subject_of	8329	2.3	modifier	160373	1.7
terrace	1729	9.1	belong	316	6.72	uninhabited	70	7.86	manor	2330	8.74
build	8408	9.03	stand	736	6.65	adjoining	126	7.79	guest	2485	8.18
detach	1759	8.99	overlook	243	6.35	repossessed	34	6.95	publishing	1416	7.86
buy	3960	8.33	date	266	5.89	unoccupied	38	6.84	Victorian	1330	7.79
board	846	7.95	rebuild	131	5.69	empty	194	6.76	public	4559	7.76
rent	929	7.95	front	88	5.48	habitable	28	6.53	bedroom	1715	7.71
sell	2470	7.87	burn	115	5.17	adjacent	77	6.08	dwelling	1196	7.67
situate	1050	7.86	sit	226	5.12	tidy	35	6.01	old	3959	7.42
demolish	644	7.58	occupy	131	5.11	clean	136	5.68	Georgian	848	7.36
own	1281	7.53	line	84	5.03	vacant	24	5.49	semi-detached	816	7.36
move	2444	7.51	consist	125	4.96	worth	222	5.49	auction	919	7.32
occupy	789	7.32	lie	178	4.94	uninhabitable	12	5.45	historic	1011	7.2
leave	2926	7.13	boast	76	4.82	spotless	12	5.36	private	1798	7.15
enter	1307	7.07	comprise	118	4.75	situate	11	5.34	opera	768	7.07
decorate	471	6.95	survive	105	4.75	semi-detached	13	5.32	coffee	942	6.98
destroy	592	6.76	collapse	60	4.75	spacious	35	5.27			

$\overrightarrow{\textbf{house}}^{colloc}$: 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

$\overrightarrow{\textbf{house}}^{similar}$: Distributional neighbors of hunting

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

Dynamic Prototype $\overrightarrow{\textbf{hunting}}^{house}$

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*

- $sim(\vec{\mathbf{e}}, \overrightarrow{\textbf{house}}^{colloc}) + sim(\vec{\mathbf{e}}, \overrightarrow{\textbf{house}}^{similar})$
- sim is Cosine similarity

{'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}
{'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}
{'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*

$\overrightarrow{\text{hunting}}^{\textit{house}}$

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

{'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}
{'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}
{'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*

$\overrightarrow{\textbf{hunting}}^{house}$

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

Dynamic Prototype Vector based Composition

$$\overrightarrow{\mathbf{house}}^{hunting} \oplus \overrightarrow{\mathbf{hunting}}^{house}$$

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

	Simple Add	Simple 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 on Phrasal Similarity task (Mitchell and Lapata, 2010)

- Static Multi Prototypes worse than normal composition

	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 on Phrasal Similarity task (Mitchell and Lapata, 2010)

- Dynamic Prototypes show clear upper hand
- Better than distributional prototype of the compound
  - data sparsity
- Word sense can be modeled with very few exemplars
  - With increase in exemplars, noise increases

Dynamic Prototypes on Compositionality Detection

	Compounds	**Spearman ρ**	**Kendalls τ**
Dynamic Prototypes	169	**0.35**	**0.24**
Static Prototypes	169	0.33	0.23
SharedTaskNextBest	174	0.33	0.23
Rand-Base	174	0.02	0.02

Table: Correlation Scores on DisCo'2011 Shared Task Data

(Biemann and Giesbrecht, 2011) "... across different scoring mechanisms, **UoY is the most robust of the systems"**

Outline

Beyond Phrases: Sentential Semantics

Questions to be answered

- Which space does a sentence live in?
- Larger issues: entailment, intersection

Beyond Phrases: Sentential Semantics

(Clark and Pulman, 2007)

- *"the boy ate a juicy orange"*
- *boy* –subj→ *ate* ←obj– *orange* ←mod– *juicy*
- $\overrightarrow{\mathbf{boy}} \otimes \overrightarrow{\mathbf{subj}} \otimes (\overrightarrow{\mathbf{ate}} \otimes \overrightarrow{\mathbf{obj}} \otimes (\overrightarrow{\mathbf{juicy}} \otimes \overrightarrow{\mathbf{mod}} \otimes \overrightarrow{\mathbf{orange}}))$

(Grefenstette and Sadrzadeh, 2011)

- Words belong to different type-based categories (e.g. transitive verb, ditransitive, noun, . . .)
- $\overrightarrow{\mathbf{ate}} \odot (\overrightarrow{\mathbf{boy}} \otimes (\overrightarrow{\mathbf{juicy}} \odot \overrightarrow{\mathbf{orange}}))$
- $\otimes$ is accumulation operator and $\odot$ is filtering operator

How about distributional representation of semantic roles?

Semantic Roles as Functions, and Words as Arguments

- laser light, infrared light
- laser light, traffic light
- laser light, laser treatment
- very fine grained semantic roles

Generative Distributional Grammar

- Grammar consists of semantic roles (matrices) and phrases (vectors)
- *"the boy ate a juicy orange"*
- *boy* –A0⟶ *ate* ⟵A1– *orange* ⟵ A2– *juicy*
- $\mathbf{A0} \odot ((\mathbf{A1} \odot (\overrightarrow{\mathbf{ate}} \otimes (\mathbf{A2} \odot (\overrightarrow{\mathbf{orange}} \otimes \overrightarrow{\mathbf{juicy}})))) \otimes \overrightarrow{\mathbf{boy}})$

How about distributional representation of semantic roles?

Semantic Roles as Functions, and Words as Arguments

- laser light, infrared light
- laser light, traffic light
- laser light, laser treatment
- very fine grained semantic roles

Generative Distributional Grammar

- Grammar consists of semantic roles (matrices) and phrases (vectors)
- *"the boy ate a juicy orange"*
- *boy* –A0⟶ *ate* ⟵A1– *orange* ⟵ A2– *juicy*
- $\mathbf{A0} \odot ((\mathbf{A1} \odot (\overrightarrow{\mathbf{ate}} \otimes (\mathbf{A2} \odot (\overrightarrow{\mathbf{orange}} \otimes \overrightarrow{\mathbf{juicy}})))) \otimes \overrightarrow{\mathbf{boy}})$

Bibliography I

Bannard, C., Baldwin, T., and Lascarides, A. (2003). A statistical approach to the semantics of verb-particles. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*, MWE '03, pages 65–72, Stroudsburg, PA, USA. Association for Computational Linguistics.

Biemann, C. and Giesbrecht, E. (2011). Distributional semantics and compositionality 2011: Shared task description and results. In *Proceedings of DISCo-2011 in conjunction with ACL 2011*.

Clark, S. and Pulman, S. (2007). Combining symbolic and distributional models of meaning. In *Proceedings of the AAAI Spring Symposium on Quantum Interaction, Stanford, CA, 2007*, pages 52–55.

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

Bibliography II

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.

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.

Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, pages 1–32.

Grefenstette, E. and Sadrzadeh, M. (2011). Experimental Support for a Categorical Compositional Distributional Model of Meaning. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*.

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

Bibliography III

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

Korkontzelos, I. and Manandhar, S. (2009). Detecting compositionality in multi-word expressions. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, ACLShort '09, pages 65–68, Stroudsburg, PA, USA. Association for Computational Linguistics.

McCarthy, D., Keller, B., and Carroll, J. (2003). Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*, MWE '03, pages 73–80, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

Bibliography IV

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

Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.

Montague, R. (1973). The Proper Treatment of Quantification in Ordinary English. pages 221–242.

Partee, B. (1995). Lexical semantics and compositionality. *L. Gleitman and M. Liberman (eds.) Language, which is Volume 1 of D. Osherson (ed.) An Invitation to Cognitive Science (2nd Edition)*, pages 311–360.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17.

Reddy, S., Klapaftis, I. P., McCarthy, D., and Manandhar, S. (2011a). Dynamic and static prototype vectors for semantic composition. In *Proceedings of The 5th International Joint Conference on Natural Language Processing 2011 (IJCNLP 2011)*, Chiang Mai, Thailand.

Bibliography V

Reddy, S., McCarthy, D., and Manandhar, S. (2011b). An empirical study on compositionality in compound nouns. In *Proceedings of The 5th International Joint Conference on Natural Language Processing 2011 (IJCNLP 2011)*, Chiang Mai, Thailand.

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.

Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, 24(1):97–123.

Bibliography VI

Venkatapathy, S. and Joshi, A. K. (2005). Measuring the relative compositionality of verb-noun (v-n) collocations by integrating features. In *Proceedings of the joint conference on Human Language Technology and Empirical methods in Natural Language Processing*, pages 899–906, Vancouver, B.C., Canada.

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