

# Exemplar-based Word-Space Model for Compositionality Detection

Siva Reddy<sup>1,2</sup>, Diana McCarthy<sup>2</sup>, Suresh Manandhar<sup>1</sup> and  
Spandana Gella<sup>1</sup>

<sup>1</sup>Artificial Intelligence Group,  
Department of Computer Science,  
University of York, UK

<sup>2</sup>Lexical Computing Ltd., UK

University of Leeds

Jun 16 2011

# Multiword

- A sequence of two or more words describing a meaning together.
- Compound Nouns
  - credit card
  - leather jacket
- Phrasal Verbs
  - look up
  - get over
- Idiomatic expressions
  - kick the bucket
  - spill the beans

# Compositionality

Given the meanings of

- couch
- roast
- potato

# Compositionality

Given the meanings of

- couch
- roast
- potato

Can we interpret the meanings of

- couch potato
- roast potato

# Couch Potato



# Roast Potato



# Compositionality

## Compositional Multi-words

- $m(\text{"A B"}) = m(\text{A}) \oplus m(\text{B})$
- e.g. water tank, post man, roast potato
- caution: subjective task
- cracking " $\oplus$ " is a miracle

# Compositionality

## Compositional Multi-words

- $m(\text{"A B"}) = m(\text{A}) \oplus m(\text{B})$
- e.g. water tank, post man, roast potato
- caution: subjective task
- cracking " $\oplus$ " is a miracle

## Non-compositional multi-words

- think tank, smoking gun, apple polisher, couch potato

# Importance of Compositionality Detection

## Dictionary Building: Lexicography and Terminology

- A good dictionary
  - includes non-compositional multi-words
  - does not have redundant information

## Machine Translation: In sports domain

- goose egg  $\neq$  Gänseei
- goose egg  $\rightarrow$  unwichtig

## Word Tokenization

- Search engines
  - zebra
  - zebra crossing
  - zebra park

# Compositionality Detection Methods

- Non-compositional MWEs exhibit idiosyncratic properties - special properties.
- Detection methods exploit these special properties
- Methods can be classified into two types
  - exploiting **syntactic** properties
  - exploiting **semantic** properties

# Methods Exploiting Syntactic Properties

Lexical fixedness e.g. *red tape* vs *red car*

- high statistical association (Lin, 1999; Pedersen, 2011)
- poor results
- Pitfall: Institutionalized phrases e.g. *traffic light*

# Methods Exploiting Syntactic Properties

Lexical fixedness e.g. *red tape vs red car*

- high statistical association (Lin, 1999; Pedersen, 2011)
- poor results
- Pitfall: Institutionalized phrases e.g. *traffic light*

Syntactic rigidity e.g. “*kick the great buckets*”

- Cook et al. (2007); Fazly et al. (2009)
- Works mostly for verbal idioms

# Methods Exploiting Syntactic Properties

Lexical fixedness e.g. *red tape vs red car*

- high statistical association (Lin, 1999; Pedersen, 2011)
- poor results
- Pitfall: Institutionalized phrases e.g. *traffic light*

Syntactic rigidity e.g. *“kick the great buckets”*

- Cook et al. (2007); Fazly et al. (2009)
- Works mostly for verbal idioms

Selectional preferences e.g. *gravy train vs gravy recipe*

- McCarthy et al. (2007)
- Resource dependent: WordNet, selectional preferences
- Works mostly for verbal idioms

# Methods Exploiting Semantic Properties

*A Non-compositional MWE and its constituents  
have semantic differences*

# Methods Exploiting Semantic Properties

*A Non-compositional MWE and its constituents  
have semantic differences*

- low semantic similarity/relatedness  $\Rightarrow$  non-compositional
  - sim(smoking gun, smoking)
  - sim(smoking gun, gun)
- high semantic similarity/relatedness  $\Rightarrow$  compositional
  - sim(traffic light, traffic)
  - sim(traffic light, light)
- Better results reported (Baldwin et al., 2003; Bannard et al., 2003)

# Semantic Similarity Measures

How to measure **sim(traffic light, light)??**

# Semantic Similarity Measures

How to measure **sim(traffic light, light)**??

## Word Space Model (Sahlgren, 2006)

- Distributional hypothesis: Words that occur in similar contexts tend to have similar meanings
- Meaning of a word is represented as a co-occurrence vector built from a corpus

	police-n	photon-n	speed-n	car-n	router-n
Traffic	40	0	10	16	23
Light	5	27	20	12	3

## Semantic Similarity: Distributional Way

	police-n	photon-n	speed-n	car-n	soul-n
<b>Traffic</b>	40	0	10	16	3
<b>Light</b>	5	27	20	12	23
<b>TrafficLight</b>	25	0	15	14	0

- Distributional hypothesis holds for MWEs too.
- To construct the cooccurrence vector **TrafficLight** treat MWE as a single entity.
- Co-occurrence frequencies are converted to association scores like MI, PMI, conditional probabilities
- Cosine similarity  $\text{sim}(\mathbf{TrafficLight}, \mathbf{Light})$ 
  - $\frac{\mathbf{TrafficLight} \cdot \mathbf{Light}}{\|\mathbf{TrafficLight}\| \|\mathbf{Light}\|}$

## Baldwin et al. (2003)

- If  $\text{sim}(\mathbf{TrafficLight}, \mathbf{Light}) > \gamma$ , MWE is compositional (Baldwin et al., 2003)
- Any single threshold value  $\gamma$  did not hold
- As  $\gamma$  changes, performance varied arbitrarily
- 51% accuracy reported

## Baldwin et al. (2003)

- If  $\text{sim}(\mathbf{TrafficLight}, \mathbf{Light}) > \gamma$ , MWE is compositional (Baldwin et al., 2003)
- Any single threshold value  $\gamma$  did not hold
- As  $\gamma$  changes, performance varied arbitrarily
- 51% accuracy reported

### Reason

- All the senses of a word are represented by a single prototype vector
- Skewed nature of senses: **RiverBank** is not similar to **Bank**
- Irrelevant contexts create noise e.g. photon-n and soul-n in **Light**

# Compositionality Function Based approaches

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

The meaning of a complex expression is a function of the meaning of its parts and of the syntactic rules by which they are combined.

# Compositionality Function Based approaches

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

The meaning of a complex expression is a function of the meaning of its parts and of the syntactic rules by which they are combined.

- **Traffic** $\oplus$ **Light** is the meaning composed from **Traffic** and **Light**
- $\oplus$  is the compositionality function
- Several compositionality functions are proposed (Mitchell and Lapata, 2008; Widdows, 2008; Erk and Padó, 2008)
- Most popular are simple addition and simple multiplication (Guevara, 2010; Mitchell and Lapata, 2008)

# Compositionality Functions

	police-n	photon-n	speed-n	car-n	soul-n
<b>Traffic</b>	40	0	10	16	3
<b>Light</b>	5	27	20	12	23
<b>TrafficLight</b>	25	0	15	14	0
<b>aTraffic + bLight</b>	45	27	30	28	26
<b>Traffic * Light</b>	200	0	200	192	69

# Katz and Giesbrecht (2006); Giesbrecht (2009)

- Methods cash the failure of the compositionality principle

## Katz and Giesbrecht (2006); Giesbrecht (2009)

- Methods cash the failure of the compositionality principle
- if  $\text{sim}(\mathbf{TrafficLight}, \mathbf{Traffic} \oplus \mathbf{Light}) < \gamma$ , MWE is non-compositional
- Similar observations to (Baldwin et al., 2003)
- No single  $\gamma$  worked

## Katz and Giesbrecht (2006); Giesbrecht (2009)

- Methods cash the failure of the compositionality principle
- if  $\text{sim}(\mathbf{TrafficLight}, \mathbf{Traffic} \oplus \mathbf{Light}) < \gamma$ , MWE is non-compositional
- Similar observations to (Baldwin et al., 2003)
- No single  $\gamma$  worked

### Reason: Polysemy again

	police-n	photon-n	speed-n	car-n	soul-n
<b>Traffic</b>	40	0	10	16	3
<b>Light</b>	5	27	20	12	23
<b>TrafficLight</b>	25	0	15	14	0
<b>aTraffic + bLight</b>	45	27	30	28	26
<b>Traffic * Light</b>	200	0	200	192	69

# Words are polysemous

## Prototype Vectors are the problem

Currently most methods represent each word as a single vector i.e. a prototype vector for each word.

# Words are polysemous

## Prototype Vectors are the problem

Currently most methods represent each word as a single vector i.e. a prototype vector for each word.

## Light occur in many contexts

- like quantum theory, optics, bulbs and traffic domain
- Not all contexts of *light* are relevant for *traffic light*
- $\text{sim}(\mathbf{TrafficLight}, \mathbf{Light})$  in ukWaC is 0.27
- **Light** is noisy  $\Rightarrow$  **Traffic**  $\oplus$  **Light** is noisy

# Words are polysemous

## Prototype Vectors are the problem

Currently most methods represent each word as a single vector i.e. a prototype vector for each word.

## Light occur in many contexts

- like quantum theory, optics, bulbs and traffic domain
- Not all contexts of *light* are relevant for *traffic light*
- $\text{sim}(\mathbf{TrafficLight}, \mathbf{Light})$  in ukWaC is 0.27
- **Light** is noisy  $\Rightarrow$  **Traffic**  $\oplus$  **Light** is noisy

## Exemplars of *light* in ukWaC corpus

'interest-n': 1.0, 'round-n': 1.0, 'open-v': 1.0

'business-n': 1.0, 'bad-j': 1.0, 'put-v': 1.0

'framework-n': 1.0, 'generation-n': 1.0, 'technique-n': 1.0, 'follow-v': 1.0

'material-n': 1.0, 'day-n': 1.0, 'complete-j': 1.0

# Concordance of *light*

the idea of a much worse prison ; where No	<b>light</b>	, but rather darkness visible , Served
tabernacle-work over the stalls carved in a	<b>light</b>	and elegant manner . St. John 's , which
There was a general , unrelieved , dull	<b>light</b>	; so that , unless when looking at your
He half opened one of them , and as the	<b>light</b>	poured in , looked round with mournful
beauty heightened by the aid of brilliant	<b>lights</b>	, of costly jewels , and all the pride
use the cycle paths and have good bright	<b>lights</b>	, then you should have no problems . Bus
. I think it puts business in a very bad	<b>light</b>	. Alan Sugar does everyone a great disservice
framework , Tati became an influential guiding	<b>light</b>	for the generations of comedians and filmmakers
morning - it 's night it 's dark - it 's	<b>light</b>	It 's raining - it 's sunny life 's serious
or feeling low M Baird 167 The Northern	<b>lights</b>	and Mackie 's means home sweet home to
This investigation is intended to bring to	<b>light</b>	some reasons for connecting the notion
24 hours a day and the proprietors keep	<b>light</b>	security , perticularly a local rent a cop
form with the only pleasantness being the	<b>light</b>	white fluffy foam of the recently sumped
I thought his material had all seen the	<b>light</b>	of day . TK 's mentor , Henry Stone sent
1973 . I am amazed this has n't seen the	<b>light</b>	of day . It is wonderful , and definitely
legal issues , that it would never see the	<b>light</b>	of day . Frank has taken the reigns , as
quite rightly so . Bright and breezy and	<b>lit</b>	up a few dancefloors as well as receiving
requested . â € Ideal as a clip-on book	<b>light</b>	â € Reaches places other torches can
cover at the end of the arm . The Flexi	<b>Light</b>	requires two AAA batteries ( not included
and clips in for compact storage . Flexi	<b>Light</b>	FAQ 's : Q ) Hi , what bulb should I use

# Dynamic Prototypes using Exemplar based Models

*Static prototype vectors are noisy*  
*A need for refined vectors*

# Dynamic Prototypes using Exemplar based Models

*Static prototype vectors are noisy*  
*A need for refined vectors*

## Exemplar-based Models

- Select (examples) exemplars of *light* which have similar context to *traffic*
- Prunes out the irrelevant exemplars
- Use selected exemplars to build the **Dynamic Prototype Light<sub>Traffic</sub>**
- Exemplar-based Modelling for paraphrasing (Erk and Padó, 2010)

# Dynamic Prototypes using Exemplar based Models

*Static prototype vectors are noisy*  
*A need for refined vectors*

## Exemplar-based Models

- Select (examples) exemplars of *light* which have similar context to *traffic*
- Prunes out the irrelevant exemplars
- Use selected exemplars to build the **Dynamic Prototype**  $\text{Light}_{\text{Traffic}}$
- Exemplar-based Modelling for paraphrasing (Erk and Padó, 2010)

## Dynamic Prototype $\text{Light}_{\text{Traffic}}$

- $\text{Light}_{\text{Traffic}}$  represents the sense of *light* in the presence of *traffic*
- Benefit from the information from *traffic*: cooccurrences and distributionally similar neighbours
- Others: Static Multi Prototypes (Reisinger and Mooney, 2010)

# cooccurrences of *traffic*

<b>object of</b> 18950 1.4	<b>and/or</b> 10005 0.8	<b>pp_along-i</b> 95 5.2	<b>n_modifier</b> 23201 2.5
divert 198 7.78	pedestrian 157 7.56	road 25 0.92	freight 512 8.63
slow 212 7.58	congestion 134 6.92	route 14 0.83	road 5520 8.63
block 319 7.5	pollution 168 6.42	street 9 0.62	air 2248 8.09
generate 711 7.34	noise 188 5.8		passenger 708 7.54
speed 149 7.25	parking 173 5.57	<b>pp_onto-i</b> 64 4.6	rush 211 7.45
motorise 93 7.17	freight 39 5.45	road 25 0.92	commuter 134 7.09
encrypt 83 6.79	roadwork 14 5.26	route 9 0.2	motor 360 6.94
rout 78 6.76	traffic 190 5.17		rail 240 6.27
direct 218 6.65	transportation 32 5.09	<b>pp_off-i</b> 54 4.4	network 874 6.0
calm 75 6.5	highway 38 5.05	road 31 1.23	motorway 85 5.96
congest 55 6.43	passenger 99 4.87		good 350 5.85
wheel 55 6.28	fume 14 4.73	<b>pp_through-i</b> 305 2.8	lorry 73 5.85
redirect 52 6.24	pedestrianisation 8 4.63	firewall 10 4.35	Internet 492 5.83
monitor 260 6.21	lorry 20 4.53	village 46 2.74	coal 117 5.75
increase 985 6.18	commuter 13 4.49	port 18 2.7	multicast 37 5.63
drive 407 6.12	motorway 20 4.38	tunnel 8 2.69	tourist 106 5.48
stop 330 6.1	transport 130 4.33		barge 40 5.27
reduce 698 6.08	TCP 11 4.33		data 143 5.25
induce 83 6.01	road 240 4.15		container 72 5.23

*Sketch Engine (www.sketchengine.co.uk) is used to extract these cooccurrences*

## Distributionally similar words to *traffic*

Lemma	Score	Freq
<a href="#">transport</a>	0.362	134717
<a href="#">road</a>	0.339	324641
<a href="#">train</a>	0.336	114514
<a href="#">vehicle</a>	0.331	160671
<a href="#">bus</a>	0.322	131884
<a href="#">route</a>	0.312	168121
<a href="#">network</a>	0.311	262162
<a href="#">trade</a>	0.308	165216
<a href="#">market</a>	0.307	379176
<a href="#">travel</a>	0.306	115459
<a href="#">communication</a>	0.3	171501
<a href="#">flow</a>	0.299	77846
<a href="#">station</a>	0.295	175788
<a href="#">operation</a>	0.295	198053
<a href="#">car</a>	0.293	419404
<a href="#">access</a>	0.289	376109
<a href="#">business</a>	0.287	700710
<a href="#">speed</a>	0.286	146544
<a href="#">sale</a>	0.285	239489

*Not only context words of traffic but also similar words to traffic are useful*

# Dynamic Prototype **Light**<sub>Traffic</sub>

## Ranked exemplars of *light*

'speed-n': 4.0, 'create-v': 1.0, 'mass-n': 1.0

'road-n': 2.0, 'good-j': 1.0, 'white-j': 3.0

'street-n': 1.0, 'road-n': 2.0, 'limit-n': 1.0, 'sign-n': 1.0

'road-n': 2.0, 'side-n': 1.0, 'wrong-j': 1.0, 'drive-v': 1.0

'bright-j': 1.0, 'day-n': 1.0

# Dynamic Prototype **Light**<sub>Traffic</sub>

## Ranked exemplars of *light*

'speed-n': 4.0, 'create-v': 1.0, 'mass-n': 1.0

'road-n': 2.0, 'good-j': 1.0, 'white-j': 3.0

'street-n': 1.0, 'road-n': 2.0, 'limit-n': 1.0, 'sign-n': 1.0

'road-n': 2.0, 'side-n': 1.0, 'wrong-j': 1.0, 'drive-v': 1.0

'bright-j': 1.0, 'day-n': 1.0

- **Light**<sub>Traffic</sub> is built by summing up the top n % exemplars of *light*
- $\text{sim}(\text{TrafficLight}, \text{Light}_{\text{Traffic}})$  increased to 0.47 from an initial value of 0.27
- Just 2% of 316133 exemplars
- Similarly **Traffic**<sub>Light</sub> is built

# DisCo 2011 Shared Task (Biemann and Giesbrecht, 2011)

- Phrases consist of two lemmas and come in three grammatical relations:
  - ADJ\_NN: adjective modifying a noun
  - V\_SUBJ: noun as a subject of a verb
  - V\_OBJ: noun as an object of a verb
- For each phrase, 4 Amazon Mechanical Turkers annotate the data
  - Each score in the range 0-10 for compositionality
  - 4-5 random sentences are presented to the annotator
- Final compositionality score is averaged over all the workers
- 0-25 as non-compositional, 38-62 as medium and >75 as compositional
- 40% training, 10% validation and 50% test

# DisCo 2011 Shared Task (Biemann and Giesbrecht, 2011)

- ADJ\_NN: 144 (102 coarse)
  - blue chip: 11, non
  - great deal: 40, medium
  - stainless steel: 92, high
- V\_SUBJ: 74 (56)
  - interest lie: 40, medium
  - women want: 81, high
- V\_OBJ: 133 (96)
  - reinvent wheel: 5, non
  - put pressure: 44, medium
  - give advice: 86, high

# Compositionality Score

$$\begin{aligned}
 \text{Score } \alpha(V^{w_1}, V^{w_2}) = & a_0 + a_1 \cdot \text{sim}(V^{w_1 w_2}, V^{w_1}) \\
 & + a_2 \cdot \text{sim}(V^{w_1 w_2}, V^{w_2}) \\
 & + a_3 \cdot \text{sim}(V^{w_1 w_2}, V^{w_1} + V^{w_2}) \\
 & + a_4 \cdot \text{sim}(V^{w_1 w_2}, V^{w_1} * V^{w_2})
 \end{aligned}$$

where

- $V^{w_1 w_2} = \text{TrafficLight}$
- $V^{w_1} = \{\text{Traffic}, \text{Traffic}_{\text{Light}}\}$
- $V^{w_2} = \{\text{Light}, \text{Light}_{\text{Traffic}}\}$
- $a_i$  is a constant. All the  $a_i$ 's are estimated using linear regression on the training data
- All the combinations are tried

## Best Models on validation data

### V\_OBJ

- $\alpha(V_{OBJ}, OBJ_V)$
- Both the constituent words help each other in disambiguation

### V\_SUBJ

- $\alpha(V_{SUBJ}, SUBJ_V)$
- Both the constituent words help each other in disambiguation
- It is found  $a_3=0$ ,  $a_4=0$  i.e. composition doesn't help.

### ADJ\_NN

- $\alpha(ADJ_{NN}, NN)$
- Adjective fails in disambiguating the noun

## Dynamic weights in additive model

In the model simple addition  $a$ **Traffic** +  $b$ **Light**

- Mitchell and Lapata (2008) use static weights  $a =$  (say) 0.2,  $b =$  0.8
- Guevara (2010) also use static weights. But  $A$  and  $B$  are matrices
- We use Dynamic Weights  $a = \frac{\text{sim}(\text{TrafficLight}, \text{Traffic})}{\text{sim}(\text{TrafficLight}, \text{Traffic}) + \text{sim}(\text{TrafficLight}, \text{Light})}$  and  $b = \frac{\text{sim}(\text{TrafficLight}, \text{Light})}{\text{sim}(\text{TrafficLight}, \text{Traffic}) + \text{sim}(\text{TrafficLight}, \text{Light})}$

- $\text{sim}(\text{TrafficLight}, \text{Traffic}) = 0.54$
- $\text{sim}(\text{TrafficLight}, \text{Light}) = 0.27$
- **Traffic** contributes more towards the meaning of **TrafficLight**

## Dynamic weights in additive model

In the model simple addition  $a$ **Traffic** +  $b$ **Light**

- Mitchell and Lapata (2008) use static weights  $a =$  (say) 0.2,  $b =$  0.8
- Guevara (2010) also use static weights. But  $A$  and  $B$  are matrices
- We use Dynamic Weights  $a = \frac{\text{sim}(\text{TrafficLight}, \text{Traffic})}{\text{sim}(\text{TrafficLight}, \text{Traffic}) + \text{sim}(\text{TrafficLight}, \text{Light})}$  and  $b = \frac{\text{sim}(\text{TrafficLight}, \text{Light})}{\text{sim}(\text{TrafficLight}, \text{Traffic}) + \text{sim}(\text{TrafficLight}, \text{Light})}$

- $\text{sim}(\text{TrafficLight}, \text{Traffic}) = 0.54$
- $\text{sim}(\text{TrafficLight}, \text{Light}) = 0.27$
- **Traffic** contributes more towards the meaning of **TrafficLight**

- $\text{sim}(\text{Student}, \text{StudentNurse}_{\text{Dist}}) = 0.238$
- $\text{sim}(\text{Nurse}, \text{StudentNurse}_{\text{Dist}}) = 0.893$
- **Nurse** contributes more towards the meaning of **StudentNurse**

# Correlation Scores

	TotPrd	Spearman $\rho$	Kendalls $\tau$
Rand-Base	174	0.02	0.02
Exm-Best	169	<b>0.35</b>	<b>0.24</b>
Pro-Best	169	0.33	0.23
Exm	169	0.26	0.18
SharedTaskNextBest	174	0.33	0.23

Table: Correlation Scores

# Average Point Difference Scores

	en-all	en-ADJ-NN	en-SUBJ	en-OBJ
Rand-Base	32.82	34.57	29.83	32.34
Zero-Base	23.42	24.67	17.03	25.47
Exm-Best	<b>16.51</b>	15.19	15.72	18.6
Pro-Best	16.79	14.62	18.89	18.31
Exm	17.28	15.82	18.18	18.6
SharedTaskBest	<b>16.19</b>	14.93	21.64	14.66

Table: Average Point Difference Scores

# Coarse Grained Accuracy

	en-all	en-ADJ-NN	en-SUBJ	en-OBJ
Rand-Base	0.297	0.288	0.308	0.30
Zero-Base	0.356	0.288	0.654	0.25
Most-Freq-Base	0.593	0.673	0.346	0.65
Exm-Best	<b>0.576</b>	0.692	0.5	0.475
Pro-Best	0.567	0.731	0.346	0.5
Exm	0.542	0.692	0.346	0.475
SharedTaskBest	<b>0.585</b>	0.654	0.385	0.625

Table: Coarse Grained Accuracy

# Conclusions

- Biemann and Giesbrecht (2011) referred to our system Exm-Best as the **robust system** among all the participating systems.
- Polysemy is a problem in word space models for compositionality detection
- Exemplar-based Dynamic prototypes are encouraging to address polysemy
- Our ongoing work reveal
  - Dynamic prototypes are better than Static Multi Prototypes
  - Constituent based Compositionality Detection models are as better as Compositionality function based models.

## Bibliography I

- Baldwin, T., Bannard, C., Tanaka, T., and Widdows, D. (2003). An empirical model of multiword expression decomposability. In *Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment - Volume 18*, MWE '03, pages 89–96, Stroudsburg, PA, USA. Association for Computational Linguistics.
- 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*.

## Bibliography II

Cook, P., Fazly, A., and Stevenson, S. (2007). Pulling their weight: exploiting syntactic forms for the automatic identification of idiomatic expressions in context. In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, MWE '07, pages 41–48, Stroudsburg, PA, USA. Association for Computational Linguistics.

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.

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.

Fazly, A., Cook, P., and Stevenson, S. (2009). Unsupervised type and token identification of idiomatic expressions. *Comput. Linguist.*, 35:61–103.

## Bibliography III

- Giesbrecht, E. (2009). In search of semantic compositionality in vector spaces. In *Proceedings of the 17th International Conference on Conceptual Structures: Conceptual Structures: Leveraging Semantic Technologies*, ICCS '09, pages 173–184, Berlin, Heidelberg. Springer-Verlag.
- Guevara, E. (2010). A regression model of adjective-noun compositionality in distributional semantics. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, GEMS '10, pages 33–37, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Katz, G. and Giesbrecht, E. (2006). Automatic identification of non-compositional multi-word expressions using latent semantic analysis. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, MWE '06, pages 12–19, Stroudsburg, PA, USA. Association for Computational Linguistics.

## Bibliography IV

- Lin, D. (1999). Automatic identification of non-compositional phrases. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 317–324, Stroudsburg, PA, USA. Association for Computational Linguistics.
- McCarthy, D., Venkatapathy, S., and Joshi, A. K. (2007). Detecting compositionality of verb-object combinations using selectional preferences. In *EMNLP-CoNLL*, pages 369–379.
- 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.
- 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.

## Bibliography V

- Pedersen, T. (2011). Identifying collocations to measure compositionality : Shared task system description. In *Proceedings of DISCo-2011 in conjunction with ACL 2011*.
- Reisinger, J. and Mooney, R. J. (2010). Multi-prototype vector-space models of word meaning. In *HLT-NAACL*, pages 109–117.
- Sahlgren, M. (2006). *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. PhD thesis, Stockholm University.
- Widdows, D. (2008). Semantic vector products: Some initial investigations. In *Second AAAI Symposium on Quantum Interaction, Oxford*.