Compositionality Modelling and Non-Compositionality Detection with Distributional Semantics

Diana McCarthy

Visiting Scholar: 'Erasmus Mundus Masters Program'

Saarland May 2012

in collaboration with Siva Reddy

and also with:
John Carroll, Spandana Gella, Bill Keller,
Aravind Joshi, Suresh Manadhar, Sriram Venkatapathy

Outline

Semantic Compositionality

The Principle of Semantic Compositionality [Partee, 1995] The meaning of a complex expression is determined by the meanings of its constituents and its structure

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

Semantic Compositionality

The Principle of Semantic Compositionality [Partee, 1995] The meaning of a complex expression is determined by the meanings of its constituents and its structure

Compound Noun	*swimming pool*	*couch potato*
Adjective Noun	*blue sky*	*red tape*
Verb Object	*lose keys*	*take heart*
Verb Particle	*climb up the hill*	*blow up the bridge*

Compositionality: 2 current strands of research

distributional/vector space models

- from words to phrases
- additive vs multiplicative functions [Mitchell and Lapata, 2008]
- polysemy [Reddy et al., 2011a]
- evaluation:
  - distance between 2 phrases
  - GEMS 2011

Compositionality: 2 current strands of research

distributional/vector space models

- from words to phrases
- additive vs multiplicative functions [Mitchell and Lapata, 2008]
- polysemy [Reddy et al., 2011a]
- evaluation:
  - distance between 2 phrases
  - GEMS 2011

detecting semantic non-compositional

- also with distributional / vector space models
- and other tech-niques [Fazly and Stevenson, 2006, Melamed, 1997]
- some models here [Reddy et al., 2011b, Reddy et al., 2011c] also those for phrasal compositionality
- in future . . .

Compositionality: 2 current strands of research

distributional/vector space models

- from words to phrases
- additive vs multiplicative functions [Mitchell and Lapata, 2008]
- polysemy [Reddy et al., 2011a]
- evaluation:
  - distance between 2 phrases
  - GEMS 2011

detecting semantic non-compositional

- also with distributional / vector space models
- and other tech-niques [Fazly and Stevenson, 2006, Melamed, 1997]
- some models here [Reddy et al., 2011b, Reddy et al., 2011c] also those for phrasal compositionality
- in future . . . need these over the other side too

Vector Space Models and Distributional Similarity

	context	frequency		
	gram rel (or proximity)	*plant*	*tree*	*factory*
grow	verb object	52	60	10
weed	verb object	31	23	2
water	verb object	23	15	4
dead	adj modifier	10	12	0
operate	verb subject	16	2	22
demolish	verb object	11	5	15

Distributional Thesaurus (Neighbour) Output:
Word: <closest word> <score> <2nd closest > <score>...
plant: *tree* 0.17 *flower* 0.16 *factory* 0.15 *bush* 0.13

Multiword Expression: A Working Definition

A multiword expression is a combination of two or more words whose semantic, syntactic etc... properties cannot fully be predicted from those of its components, and which therefore has to be listed in a lexicon.
[Boleda and Evert, ESLLI 2009]

Approaches for Detecting MWEs

- statistical: e.g. pointwise mutual information
  [Church and Hanks, 1990, Dunning, 1993, Smadja, 1993, Krenn and Evert, 2001]
  $PMI = log \frac{p(chew,fat)}{p(chew)p(fat)}$
- translations in parallel text: [Melamed, 1997]
  chew the fat $\leftrightarrow$ *conversar*
- dictionaries: [Piao et al., 2006]
  listings, semantic codes and relationships
- lexical variation
  [Lin, 1999, Pearce, 2001, Fazly and Stevenson, 2006]
  couch potato: *sofa potato, couch onion*
- syntactic variation: [Fazly and Stevenson, 2006]
  take heart
- distributional similarity:
  [Schone and Jurafsky, 2001, Baldwin et al., 2003]

Outline

Distributional Similarity for Compositionality Detection

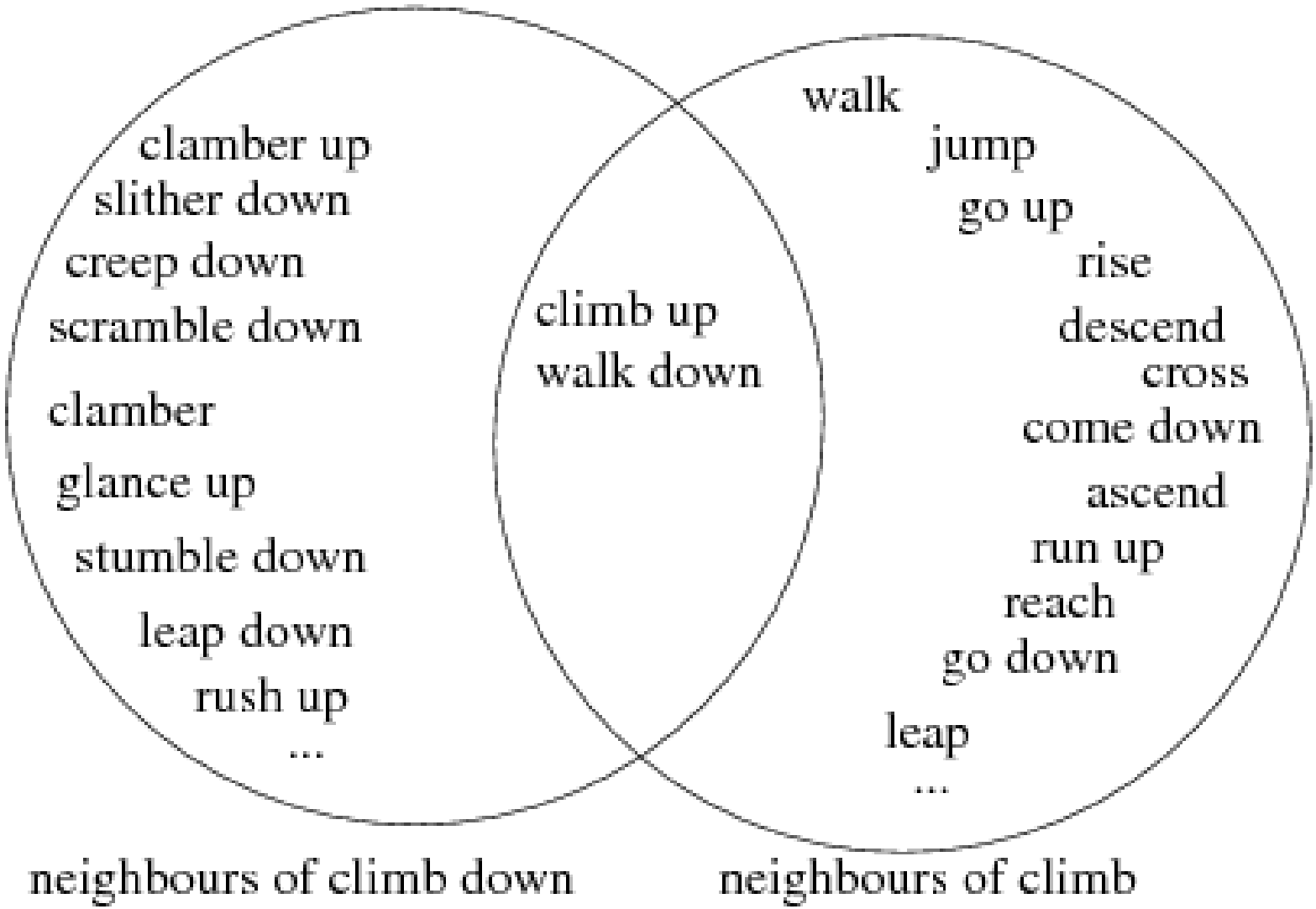
phrasal verbs: with Bill Keller and John Carroll [McCarthy et al., 2003]

e.g. *blow up* vs *eat up*

- intuition: the more compositional the phrasal, the closer the neighbours of the phrasal and the corresponding constituent verb
- also, the more likely that the verb will appear as a neighbour of the phrasal
- some measures control for particle

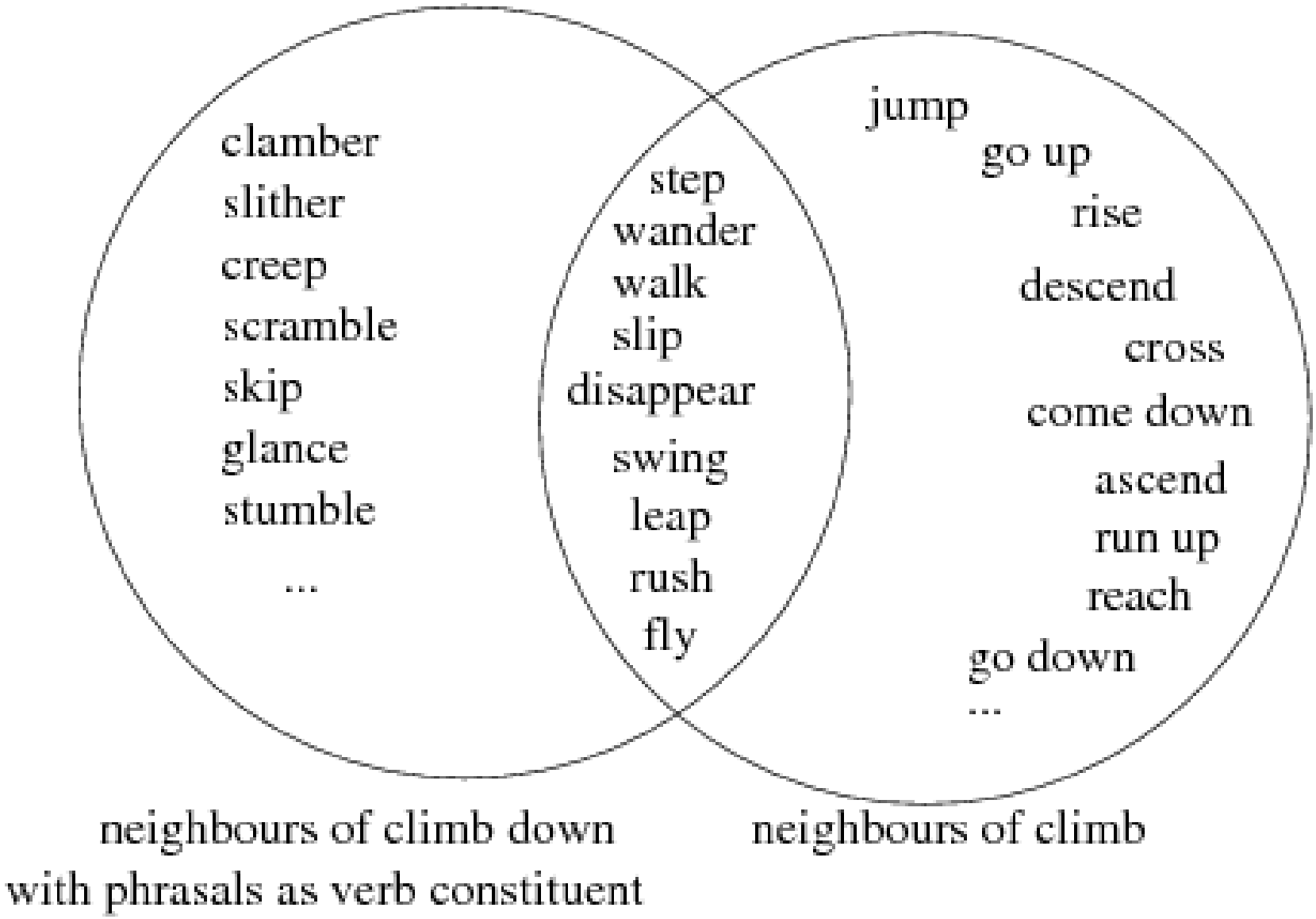
Distributional Similarity for Compositionality Detection

phrasal verbs



Distributional Similarity for Compositionality Detection

phrasal verbs



Results: correlated against human judgments (0-10)

Correlation with Measures Using the Thesaurus		
measure	correlation statistic	p under H_o
overlap	$\rho = 0.166$	0.04
overlapS	$\rho = 0.303$	$<$0.0007
sameparticle	$\rho = 0.414$	$<$0.00003
sameparticle-simplex	$\rho = 0.490$	$<$0.00003
Correlation with Statistics (used for multiword extraction)		
χ^2	$\rho =$ -0.213	0.0139
LLR	$\rho =$ -0.168	0.0392
MI	$\rho =$ -0.248	0.0047
phrasal Freq	$\rho =$ -0.096	0.156
simplex Freq	$\rho = 0.092$	0.169

Outline

Selectional Preferences for Compositionality: verb-object

with Sriram Venkatapathy and Aravind Joshi [McCarthy et al., 2007]

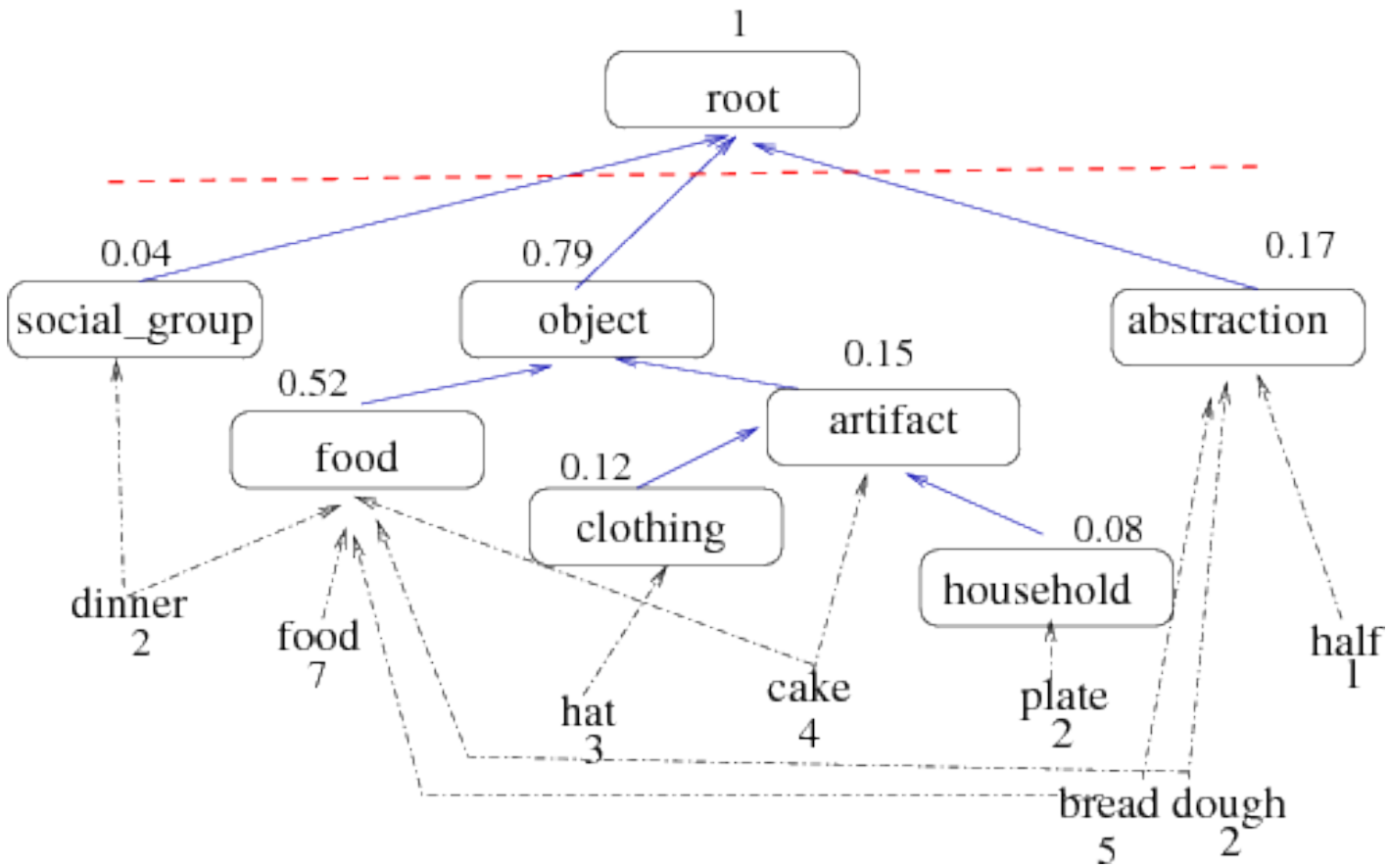
e.g. *shoot the breeze* vs *shoot the gun*

- measure likelihood of verb object combinations
- does the verb have a preference for this sort of object?
- compare WordNet and distributional similarity preference models

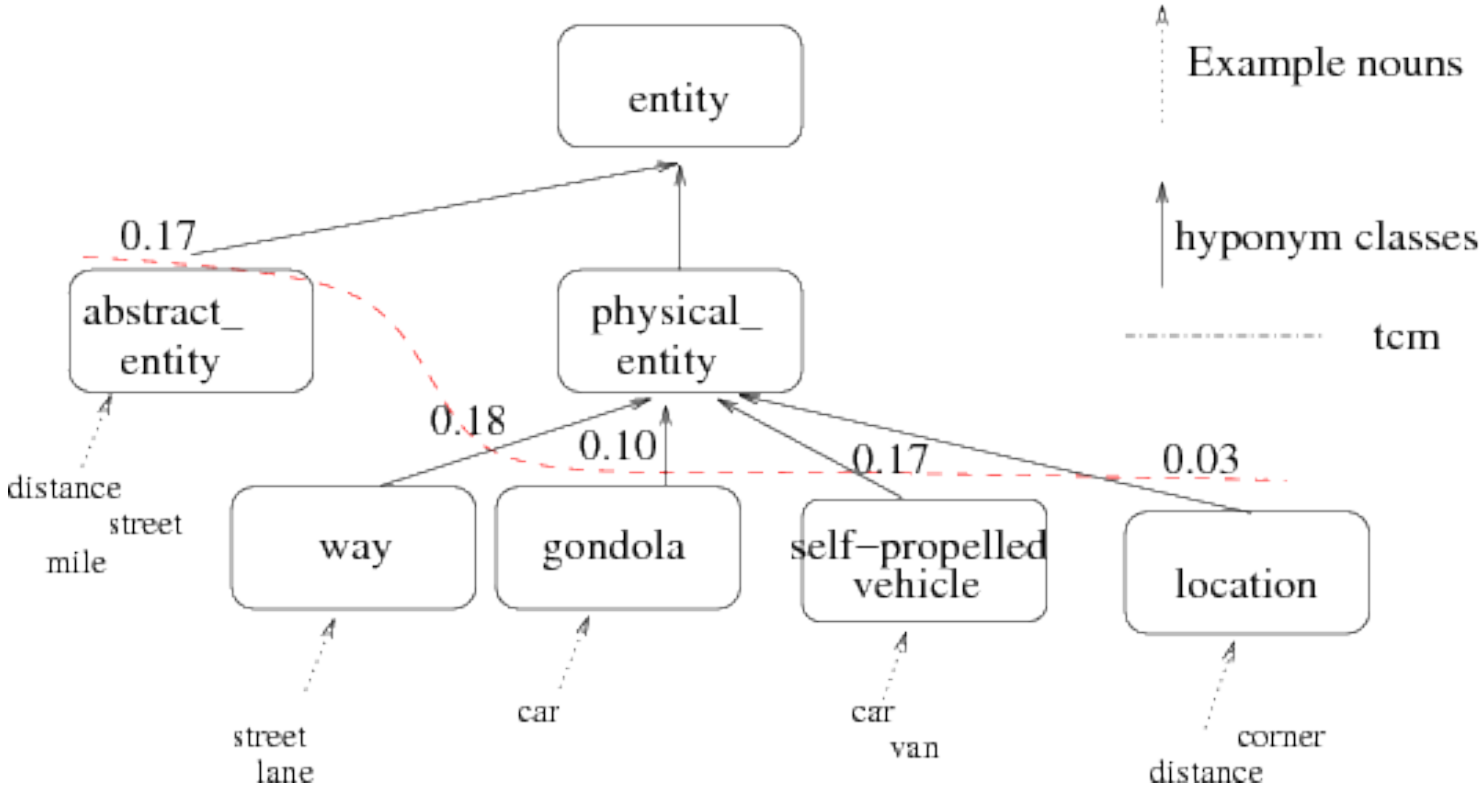
WordNet based Tree Cut Models (tcms) [Li and Abe, 1998]

example *eat*

food 7, bread 5, cake 4, hat 3, dinner 2, dough 2, plate 2, half 1



Portion of tcm for object of *park*



- Noise from *car* which occurs 174 times (out of 345).
- Contrast tokens (tcm) and type (wnproto) to obtain classes for representation, (tokens to estimate probability).

DSprotos [McCarthy et al., 2007]

- nouns are listed in thesaurus built from parses of the BNC
  van: truck 0.230, lorry 0.229, car 0.222, vehicle 0.196, ...
  bread: loaf 0.195, cheese 0.179, cake 0.169, potato 0.158, ...
- each listing is considered a grouping or "class"
- classes with at least 2 types
- argument head nouns are disambiguated by class with largest type ratio
- noun frequency to calculate probability over the classes in the model

DSproto for object slot of *park*

class ($p(c)$)	disambiguated objects (freq)
van (0.86)	car (174) van (11) vehicle (8) ...
mile (0.05)	street (5) distance (4) mile (1) ...
yard (0.03)	corner (4) lane (3) door (1)
backside (0.02)	backside (2) bum (1) butt (1) ...

Evaluating DSprotos

[Venkatapathy and Joshi, 2005] data

method	ρ	$p <$ (one tailed)
selectional preferences		
tcm	0.090	0.0119
wnproto	0.223	0.00003
DSproto	0.398	0.00003
features from V&J		
frequency (f1)	0.141	0.00023
MI (f2)	0.274	0.00003
Lin [Lin, 1999] (f3)	0.139	0.00023
LSA2 (f7)	0.209	0.00003
combination		
f2,3,7	0.413	0.00003
f1,2,3,7	0.419	0.00003
DSproto f1,2,3,7	0.454	0.00003

Outline

Noun-noun compounds

with Siva Reddy and Suresh Manadhar [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 exisiting 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
- No annotator worked on all three tasks of a compound
- Lower chance of bias to any annotator
- 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

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±0.30	4.83±0.38	4.97±0.18
search engine	4.62±0.96	2.25±1.70	3.32±1.16
face value	1.39±1.11	4.64±0.81	3.04±0.88
blame game	4.61±0.67	2.00±1.28	2.72±0.92
sitting duck	1.48±1.48	0.41±0.67	0.96±1.04

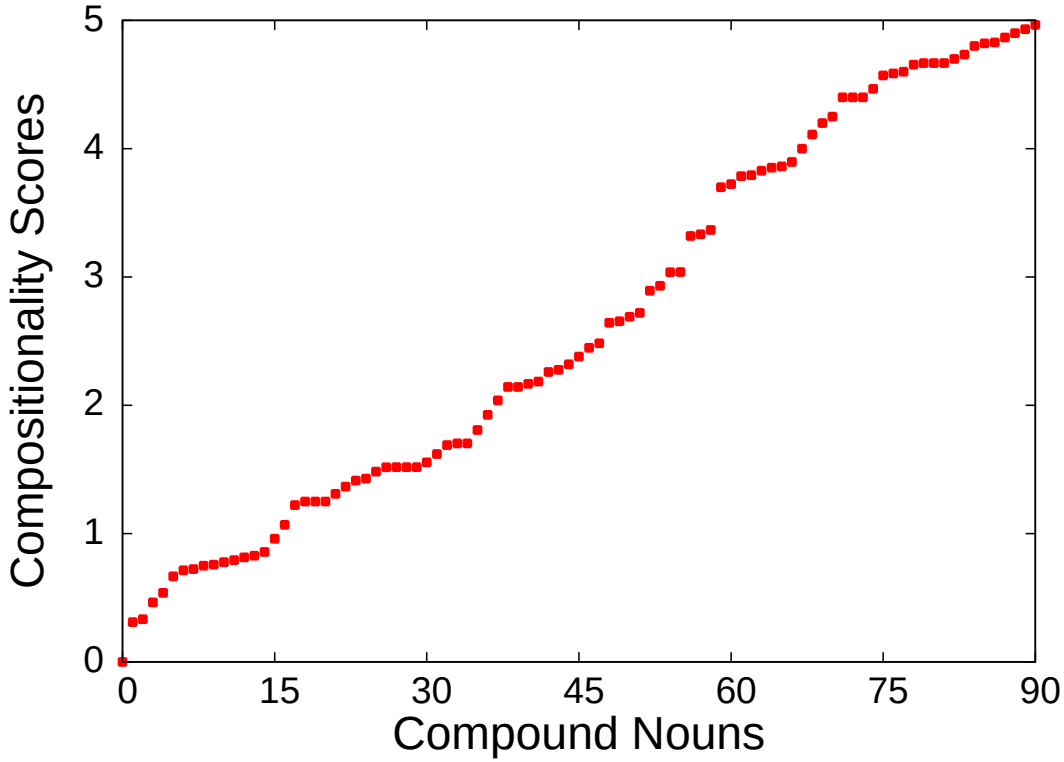
Table: Compounds with their constituent and phrase level mean±st. dev scores

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

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

- 3-fold cross validation to evaluate the above functions
- The coefficients of the functions are estimated using least square linear regression over the training samples

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

Study on human judgments

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

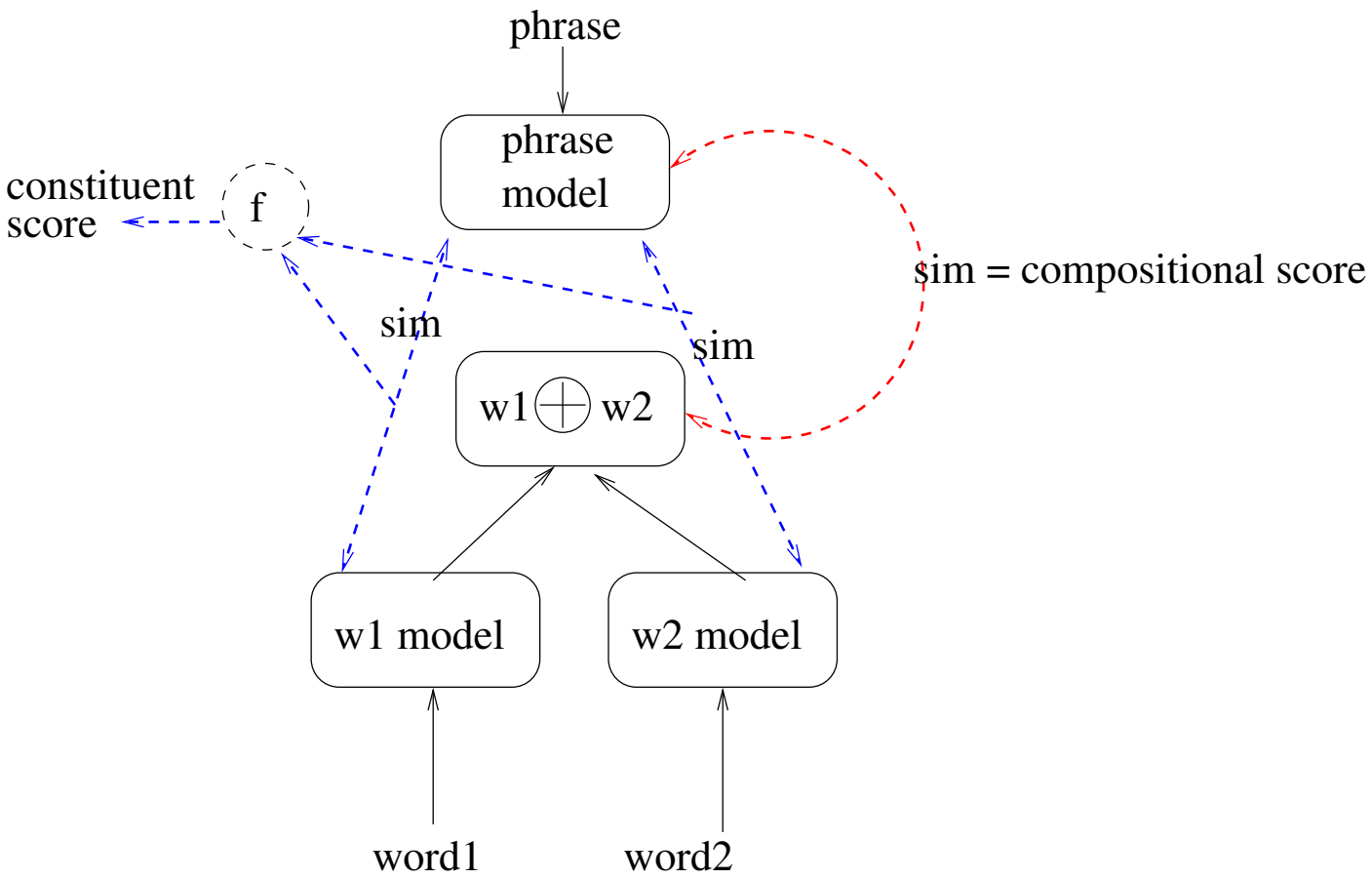
Table: Spearman Correlation ρ between functions and phrase compositionality scores

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

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
  - difference 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, Vecchi et al., 2011]

	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
aTraffic + bLight	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*
  - contexts salient to both the words are important. Foundations for our DisCo 2011 Shared Task System [Reddy et al., 2011c]

Conclusions (noun noun work)

- novel dataset for Compositionality judgments
  - contains constituent level contributions
  - continuum of compositionality
- study of relation between constituent contributions to phrase level contributions
- comparison of different computational models

The dataset is downloadable from `http://sivareddy.in`

Credits

Thank you for your attention!

Credits

Thank you for your attention!

Acknowledgments to my collaborators on these projects:
John Carroll, Spandana Gella, Bill Keller, Aravind Joshi
Suresh Manadhar, Siva Reddy, Sriram Venkatapathy

Baldwin, T., Bannard, C., Tanaka, T., and Widdows, D. (2003).
An empirical model of multiword expression decomposability.
In *Proceedings of the ACL Workshop on multiword expressions: analysis, acquisition and treatment*, pages 89–96.

Bannard, C., Baldwin, T., and Lascarides, A. (2003).
A statistical approach to the semantics of verb-particles.
In *Proceedings of the ACL Workshop on multiword expressions: analysis, acquisition and treatment*, pages 65–72.

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*.

Church, K. and Hanks, P. (1990).
Word association norms, mutual information and lexicography.

Computational Linguistics, 19(2):263–312.


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.


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. and Stevenson, S. (2006).

Automatically constructing a lexicon of verb phrase idiomatic combinations.
In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 337–344, Trento, Italy.


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 Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.


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.


Krenn, B. and Evert, S. (2001).
Can we do better than frequency? A case study on extracting PP-verb collocations.
In *Proceedings of the ACL Workshop on Collocations*, pages 39–46, Toulouse, France.


Li, H. and Abe, N. (1998).
Generalizing case frames using a thesaurus and the mdl principle.
Computational Linguistics, 24(2):217–244.


Lin, D. (1999).
Automatic identification of non-compositional phrases.
In *Proceedings of ACL-99*, pages 317–324, Univeristy of Maryland, College Park, Maryland.


McCarthy, D., Keller, B., and Carroll, J. (2003).
Detecting a continuum of compositionality in phrasal verbs.

In *Proceedings of the ACL 03 Workshop: Multiword expressions: analysis, acquisition and treatment*, pages 73–80.


McCarthy, D., Venkatapathy, S., and Joshi, A. (2007).
Detecting compositionality of verb-object combinations using selectional preferences.
In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 369–379.


Melamed, I. D. (1997).
Automatic discovery of non-compositional compounds in parallel data.
In *Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing (EMNLP 1997)*.


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.


Pearce, D. (2001).
Synonymy in collocation extraction.
In *Proc. of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, CMU.


Piao, S. S., Rayson, P., Mudraya, O., Wilson, A., and Garside, R. (2006).
Measuring mwe compositionality using semantic annotation.

In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 2–11, Sydney, Australia. Association for Computational Linguistics.

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.

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.


Reddy, S., McCarthy, D., Manandhar, S., and Gella, S. (2011c).
Exemplar-based word-space model for compositionality detection.
In *Proceedings of the ACL/HLT workshop: Disco Distributional Semantics and Compositionality*, Portland, USA. Association for Computational Linguistics.


Schone, P. and Jurafsky, D. (2001).
Is knowledge-free induction of multiword unit dictionary headwords a solved problem?
In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, pages 100–108, Hong Kong.


Smadja, F. (1993).
Retrieving collocations from text: Xtract.
Computational Linguistics. Special Issue on Using Large Corpora, 19(1):143.


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.


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*.