

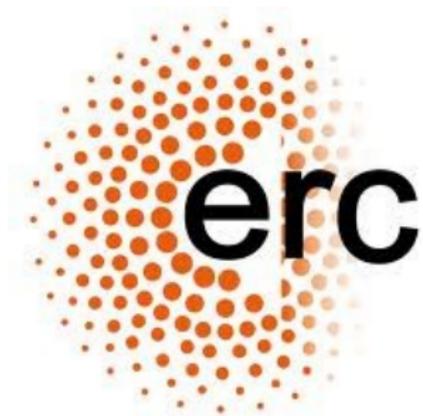
# Composition in distributional semantics

Marco Baroni

Center for Mind/Brain Sciences  
University of Trento

ABBY Open Seminar  
June 2014

COMPOSES:  
COMpositional Operations  
In SEMantic Space



In collaboration with:  
Roberto Zamparelli,  
Raffaella Bernardi,  
Georgiana Dinu,  
Marco Marelli,  
Denis Paperno,  
Jiming Li,  
Eva Maria Vecchi,  
Nghia The Pham,  
Germán Kruszewski,  
Angeliki Lazaridou,  
Edward Grefenstette,  
Louise McNally,  
Gemma Boleda  
...

# Outline

Distributional semantics

Compositionality

Experiments and extensions



# The distributional hypothesis

Harris, Charles and Miller, Firth, Wittgenstein? ...

The meaning of a word is (can be approximated by, derived from) the set of contexts in which it occurs in texts

We found a little, hairy **wampimuk** sleeping behind the tree

See also MacDonald & Ramscar CogSci 2001

# Distributional semantics

Landauer and Dumais PsychRev 1997, Turney and Pantel JAIR 2010, ...

he curtains open and the moon shining in on the barely  
ars and the cold , close moon " . And neither of the w  
rough the night with the moon shining so brightly , it  
made in the light of the moon . It all boils down , wr  
surely under a crescent moon , thrilled by ice-white  
sun , the seasons of the moon ? Home , alone , Jay pla  
m is dazzling snow , the moon has risen full and cold  
un and the temple of the moon , driving out of the hug  
in the dark and now the moon rises , full and amber a  
bird on the shape of the moon over the trees in front  
But I could n't see the moon or the stars , only the  
rning , with a sliver of moon hanging among the stars  
they love the sun , the moon and the stars . None of  
the light of an enormous moon . The splash of flowing w  
man 's first step on the moon ; various exhibits , aer  
the inevitable piece of moon rock . Housing The Airsh  
oud obscured part of the moon . The Allied guns behind

# Distributional semantics

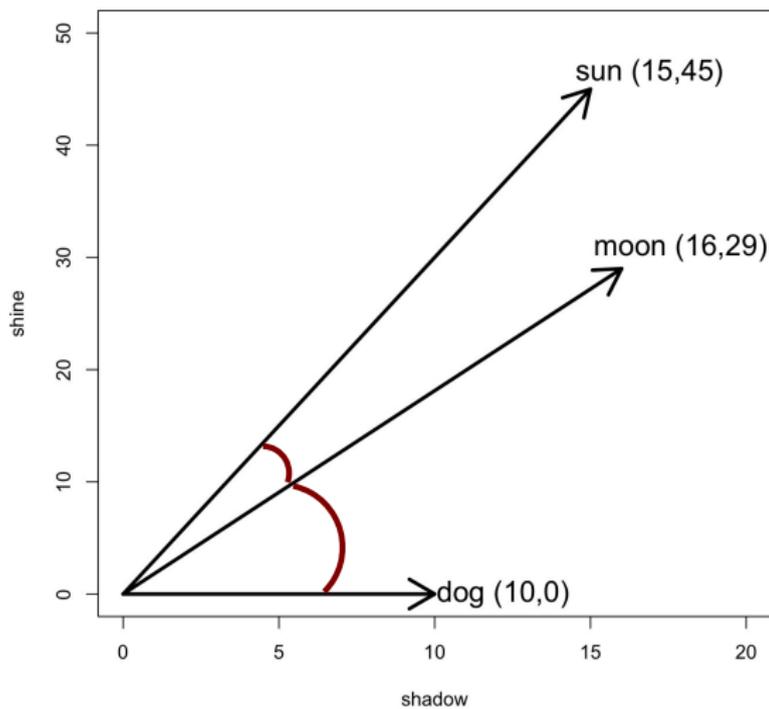
Distributional meaning encoded in co-occurrence vectors

	planet	night	full	shadow	shine	crescent
moon	10	22	43	16	29	12
sun	14	10	4	15	45	0
dog	0	4	2	10	0	0

# Distributional semantics

The geometry of meaning

	shadow	shine
moon	16	29
sun	15	45
dog	10	0



## Geometric neighbours $\approx$ semantic neighbours

rhino	fall	good	sing
woodpecker	rise	bad	dance
rhinoceros	increase	excellent	whistle
swan	fluctuation	superb	mime
whale	drop	poor	shout
ivory	decrease	improved	sound
plover	reduction	perfect	listen
elephant	logarithm	clever	recite
bear	decline	terrific	play
satin	cut	lucky	hear
sweatshirt	hike	smashing	hiss

# Distributional semantics: A general-purpose representation of lexical meaning

Baroni and Lenci 2010

- ▶ Similarity (*cord-string* vs. *cord-smile*)
- ▶ Synonymy (*zenith-pinnacle*)
- ▶ Concept categorization (*car* ISA *vehicle*; *banana* ISA *fruit*)
- ▶ Selectional preferences (*eat topinambur* vs. \**eat sympathy*)
- ▶ Analogy (*mason* is to *stone* like *carpenter* is to *wood*)
- ▶ Relation classification (*exam-anxiety* are in CAUSE-EFFECT relation)
- ▶ Qualia (TELIC ROLE of *novel* is *to entertain*)
- ▶ Salient properties (*car-wheels*, *dog-barking*)
- ▶ Argument alternations (*John broke the vase* - *the vase broke*, *John minces the meat* - \**the meat minced*)

# Selectional preferences in semantic space

Padó et al. EMNLP 2007

To kill. . .

<i>object</i>	<i>cosine</i>	<i>with</i>	<i>cosine</i>
kangaroo	0.51	hammer	0.26
person	0.45	stone	0.25
robot	0.15	brick	0.18
hate	0.11	smile	0.15
flower	0.11	flower	0.12
stone	0.05	antibiotic	0.12
fun	0.05	person	0.12
book	0.04	heroin	0.12
conversation	0.03	kindness	0.07
sympathy	0.01	graduation	0.04

# Outline

Distributional semantics

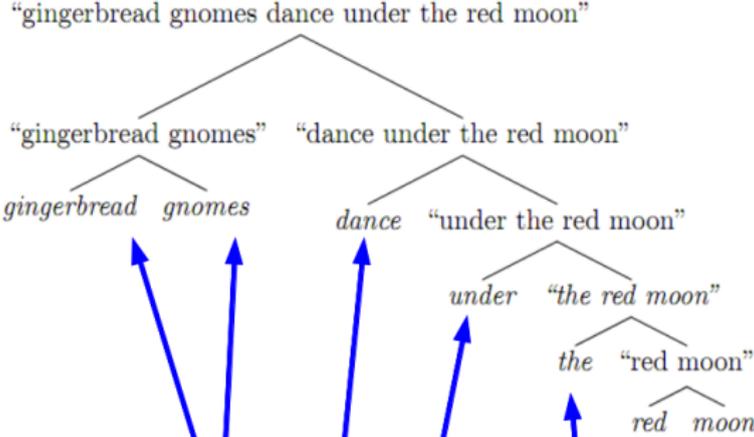
**Compositionality**

Experiments and extensions

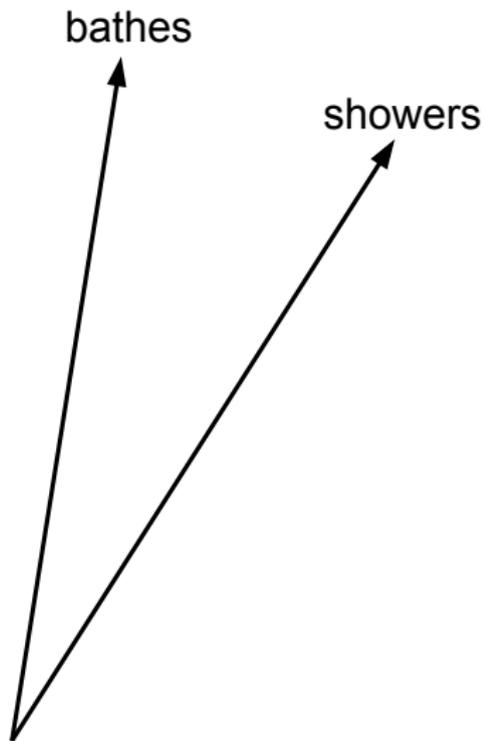


# Compositionality

The meaning of an utterance is a function of the meaning of its parts and their composition rules (Frege 1892)



# The unavoidability of distributional representations of phrases



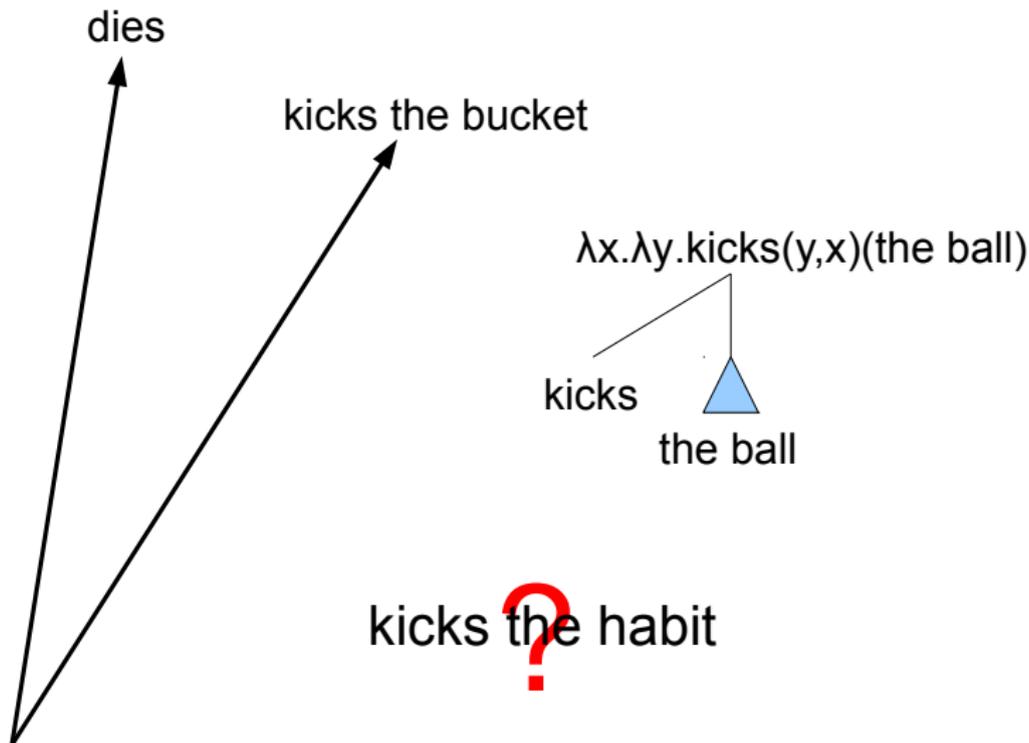
$\lambda x.\lambda y.\text{takes}(y,x)(\text{a shower})$

takes

a shower



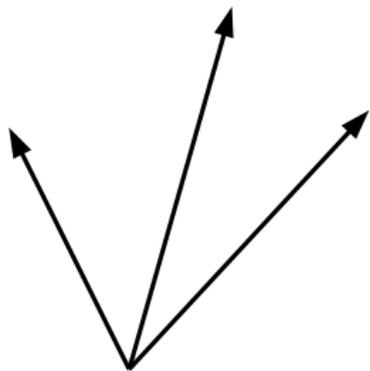
# The unavoidability of distributional representations of phrases



# What can you do with distributional representations of phrases and sentences?

**NOT** this:

every farmer who owns a donkey beats it



$\forall x \forall y (\text{FARMER}(x) \wedge \text{DONKEY}(y) \wedge \text{OWNS}(x, y) \rightarrow \text{BEAT}(x, y))$

What can you do with distributional representations of phrases and sentences?

**NOT** this:

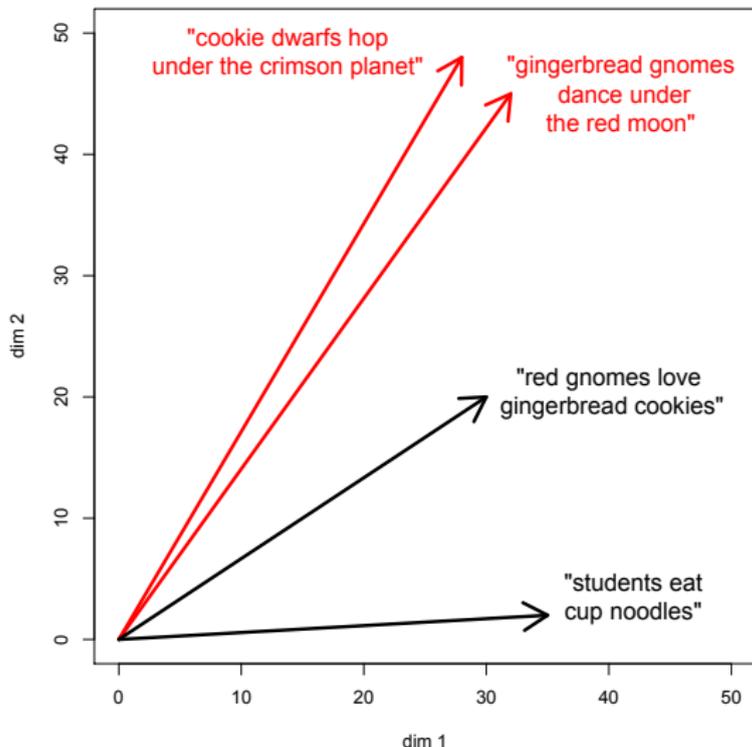


A dog **F** is barking

A dog **T** is flying

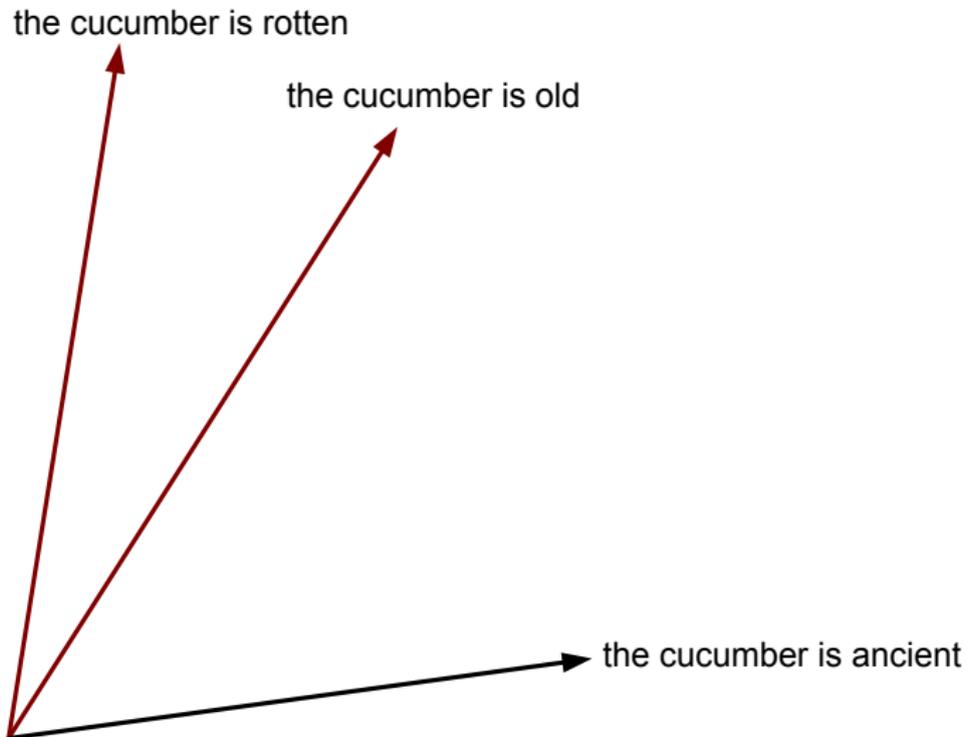
# What can you do with distributional representations of phrases and sentences?

## Paraphrasing



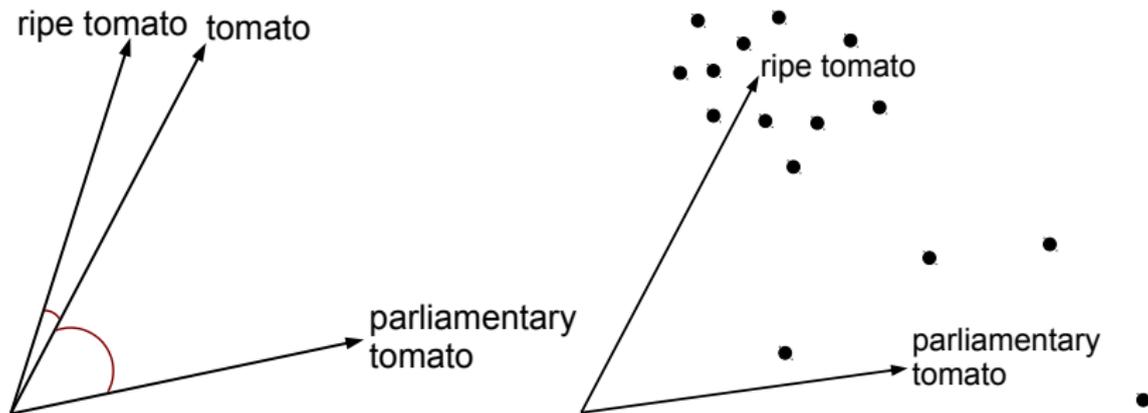
# What can you do with distributional representations of phrases and sentences?

## Contextual disambiguation



# What can you do with distributional representations of phrases and sentences?

## Semantic acceptability



# Compositional distributional semantics

How?

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red+moon	39.6	18.9	22.3	27.3	20.3	20.7
red⊙moon	371.8	56.2	44.2	72.9	22.9	10.1
red(moon)	24.6	19.3	12.4	22.6	23.9	7.1

# (Weighted) additive model

Mitchell and Lapata 2010

$$\vec{p} = \alpha \vec{u} + \beta \vec{v}$$

	music	solution	economy	craft	reasonable
practical	0	6	2	10	4
difficulty	1	8	4	4	0
practical + difficulty	1	14	6	14	4
0.4×practical + 0.6×difficulty	0.6	5.6	3.2	6.4	1.6

# Multiplicative model

Mitchell and Lapata 2010

$$\vec{p} = \vec{u} \odot \vec{v}$$

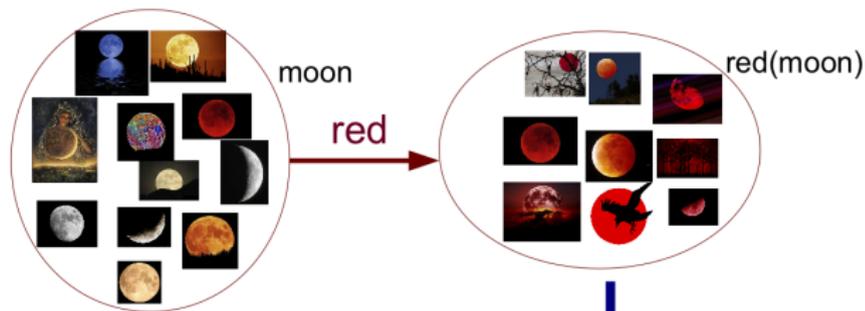
$$p_i = u_i v_i$$

	music	solution	economy	craft	reasonable
practical	0	6	2	10	4
difficulty	1	8	4	4	0
practical + difficulty	1	14	6	14	4
0.4×practical + 0.6×difficulty	0.6	5.6	3.2	6.4	1.6
practical $\odot$ difficulty	0	48	8	40	0

## Composition as dimension averaging/mixing

- ▶ red moon
- ▶ red face? fake gun?? buy guns??
- ▶ the car??? of the car????
- ▶ pandas eat bamboo  
vs. bamboo eats pandas???
- ▶ lice and dogs  
vs. lice on dogs???
- ▶ some children walk happily in the valley of the moon????

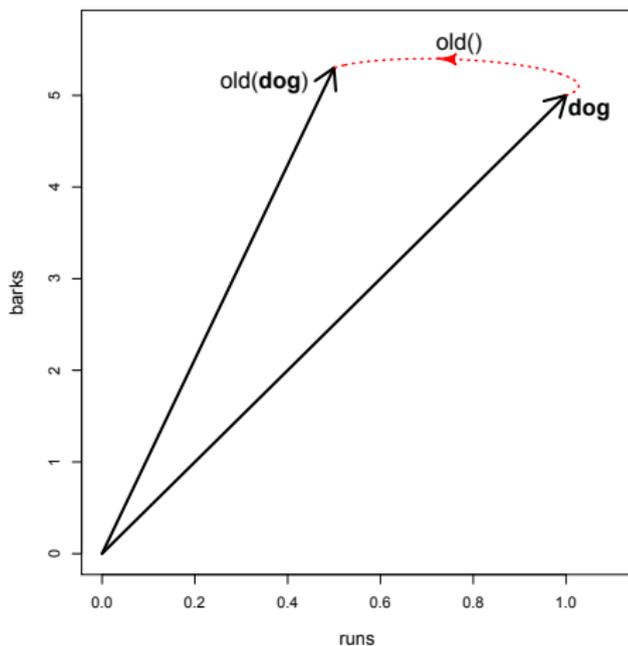
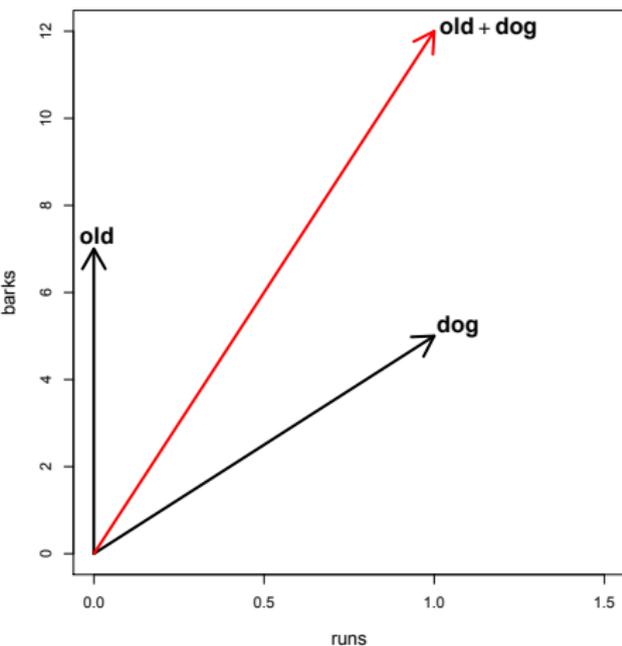
# Composition as (distributional) function application



$$\vec{\text{moon}} = \begin{bmatrix} \text{shine} & 301 \\ \text{eclipse} & 250 \\ \text{blood} & 93 \\ \dots & \dots \end{bmatrix} \xrightarrow{\text{red}} \vec{\text{red}(\text{moon})} = \begin{bmatrix} \text{shine} & 11 \\ \text{eclipse} & 245 \\ \text{blood} & 90 \\ \dots & \dots \end{bmatrix}$$

Coecke+Clark+Grefenstette+Sadrzadeh et al.,  
Baroni and Zamparelli  
Related: Guevara, Socher et al., Zanzotto et al.

# Composition as (distributional) function application



# Baroni and Zamparelli's EMNLP 2010 proposal

- ▶ How do we implement the idea of function application in a vector space?
  - ▶ Functions as **linear maps** between vector spaces
  - ▶ Functions are matrices, function application is function-by-vector multiplication:

$$F\vec{a}$$

- ▶ How do we induce a compositional function from corpus data?
  - ▶ The weights in the matrix/function are learned with **least-squares techniques** from corpus-extracted examples of how their input and output vectors look like (see also Guevara GEMS 2010)

# Learning distributional composition functions

n and the moon shining i  
with the moon shining s  
rainbowed moon . And the  
crescent moon , thrille  
in a blue moon only , wi  
now , the moon has risen  
d now the moon rises , f  
y at full moon , get up  
crescent moon . Mr Angu

f a large red moon , Campana  
, a blood red moon hung over  
glorious red moon turning t  
The round red moon , she 's  
l a blood red moon emerged f  
n rains , red moon blows , w  
monstrous red moon had climb  
. A very red moon rising is  
under the red moon a vampire

	shine	blood	Soviet
moon	301	93	1
red moon	11	90	0
army	2	454	20
red army	0	22	18

# Outline

Distributional semantics

Compositionality

Experiments and extensions

# Contextual disambiguation

- ▶ Composed sentence cosine distance correlates with human similarity judgments for ambiguous verbs in simple
  - ▶ intransitive sentences:  
*fire glows*  $\approx$  *fire burns*  $\not\approx$  *fire beams* (cf. *face glows/beams*)  
(Mitchell and Lapata ACL 2008)
  - ▶ transitive sentences:  
*table shows results*  $\approx$  *table expresses results*  $\not\approx$  *table pictures results* (cf. *map shows/pictures location*)  
(Grefenstette and Sadrzadeh EMNLP 2011)
- ▶ Best Spearman correlations with functional approach: .32 for both intransitives (Li et al. EACL 2014) and transitives (Grefenstette et al. IWCS 2013)

## Phrase paraphrasing

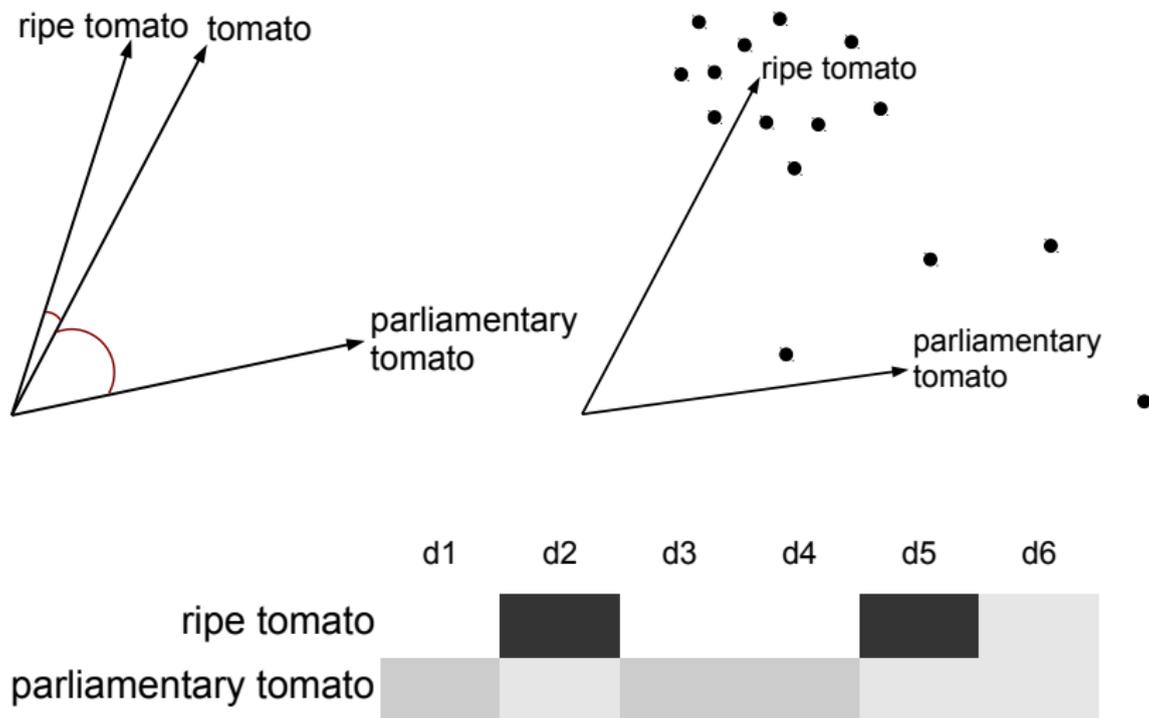
- ▶ Find right single-noun paraphrase of adjective-noun phrase (*false believe/fallacy*, *personal appeal/charisma*) in lexicon of 20K nouns (Turney JAIR 2012, Dinu et al. CVSC 2013)
  - ▶ Best median rank of correct noun: 11/20K, with syntactic-slot-based functional approach (Dinu and Baroni ACL 2014)
- ▶ Select right determiner-noun paraphrase of noun among 6 alternatives (*duel: two opponents or three opponents?* *homeless: no home or too few homes?*) (Bernardi et al. ACL 2013)
  - ▶ Best accuracy: 40% with functional approach (Li et al. EACL 2014)

## Sentence paraphrasing

- ▶ Semeval-2012 Semantic Textual Similarity (STS) MSR Video description data, with subject similarity ratings (very similar: *A man with a hard hat is dancing/A man wearing a hard hat is dancing*; not so similar: *A woman is playing the guitar/A man is playing guitar*)
  - ▶ Best Spearman correlation: .79 with functional approach (Paperno et al. ACL 2014), would have attained 25/89 ranking at STS 2012
- ▶ \*SEM-2013 STS gloss similarity ratings (very similar: *break up a solid, usually into a solution/ become or cause to become soft or liquid*; quite dissimilar: *a large truck used to haul material or goods/a long truck for carrying motor vehicles*)
  - ▶ Best Spearman correlation: .67 with functional approach (Paperno et al. 2014), would have attained 20/90 ranking at STS 2012

# Measuring semantic acceptability of phrases

Proximity, density and entropy



## Measuring adjective-noun phrase plausibility

- ▶ Distinguish, among phrases unattested in large corpus, the ones that make sense (*vulnerable gunman*, *huge joystick*, *academic crusade*) from those that do not (*academic bladder*, *parliamentary tomato*, *blind pronunciation*)
  - ▶ Functional composition model generates vectors that reside in significantly *denser* area of semantic space for acceptable vs. unacceptable nonce phrases (Vecchi et al. DISCO 2011)
- ▶ Composed-vector plausibility measures can be successfully used to predict right bracketing of noun phrases (*miracle [home run]* vs. *[miracle home] run*) (Lazaridou et al. EMNLP 2013)

# Predicting the recursive behaviour of adjectives

Vecchi et al. EMNLP 2013

- ▶ “Flexible” (*daily national newspaper/daily national newspaper*) vs. “rigid” (*rapid social change/\*social rapid change*) adjective-adjective-noun phrases
- ▶ Proximity of recursively-constructed phrase vector to components predict if adjective is flexible or rigid, and correct order for rigid phrases
- ▶ Adjective nearest to noun has strongest effect on phrase meaning

# Functional composition in morphology

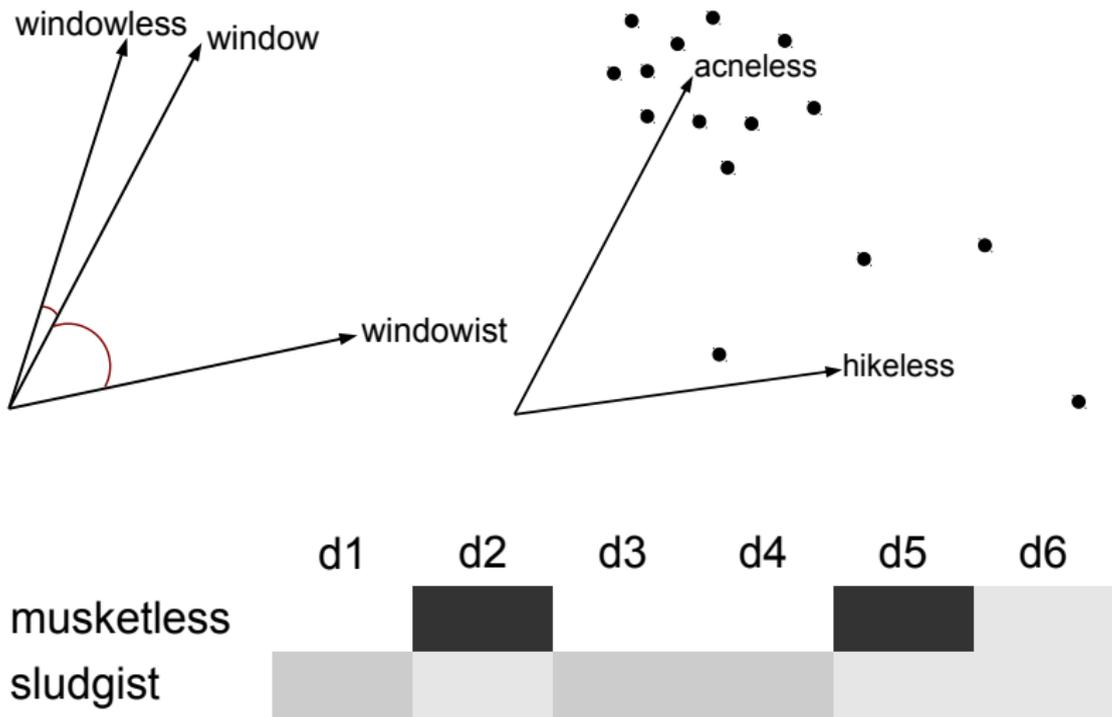
Lazaridou et al. ACL 2013, Marelli and Baroni *submitted*

- ▶ Affixes as functions from stems to derived forms:

$$\vec{redo} = \mathbf{RE} \times \vec{do}$$

- ▶ Affix matrices learned with least-squares techniques from corpus-extracted stem/derived vector pair examples (*try/retry*, *climb/reclimb*, *open/reopen*)

# Plausibility measures for morphology



# Predicting nonce derived-form acceptability

Marelli and Baroni *submitted*

- ▶ Subject ratings of unattested forms
  - ▶ Acceptable: *sketchable, hologramist*
  - ▶ Unacceptable: *happenable, windowist*
- ▶ Plausibility measures applied to vectors derived by functional composition
- ▶ Entropy and proximity (non-linear effect) predict subject intuitions (both  $p < 0.001$  in mixed model regression)

# Phrase generation in distributional semantics

Dinu and Baroni ACL 2014 (slides courtesy of Georgiana Dinu)

**From vectors  
to  
words**



gingerbread gnomes dance under the red moon

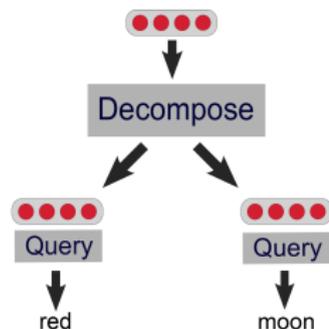
What for?

- ▶ Paraphrase generation, Machine translation
- ▶ Caption generation, Brain activation patterns to language?

# Paraphrasing

Noun phrase generation through:

- ▶ Syntax-aware decomposition functions
- ▶ Nearest neighbor queries



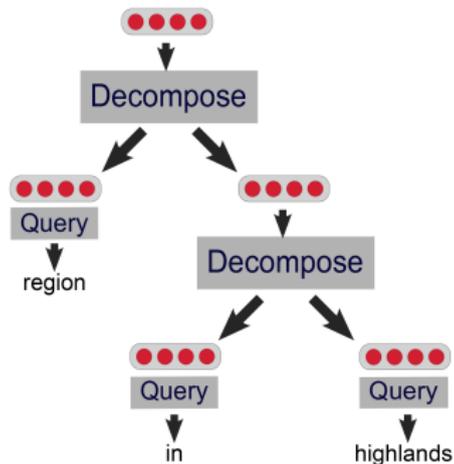
Examples

	Noun	→	Adj Noun
✓	reasoning		deductive thinking
✓	jurisdiction		legal authority
×	superstition		old-fashioned religion
×	zoom		fantastic camera

# Paraphrasing

Noun phrase generation through:

- ▶ Syntax-aware decomposition functions
- ▶ Nearest neighbor queries



Examples

Adj Noun

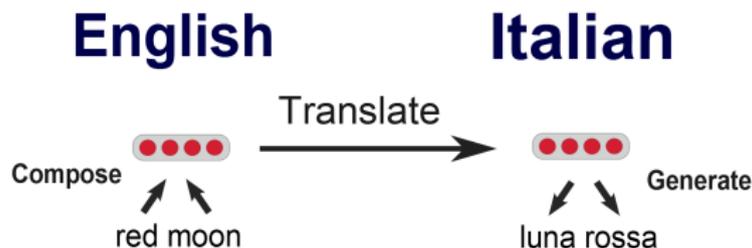
→

Noun Prep Noun

- ✓ mountainous region
- ✓ undersea cable
- ? inter-war years
- × superficial level

region in highlands  
cable through ocean  
years during 1930s  
level between levels

# Cross-lingually



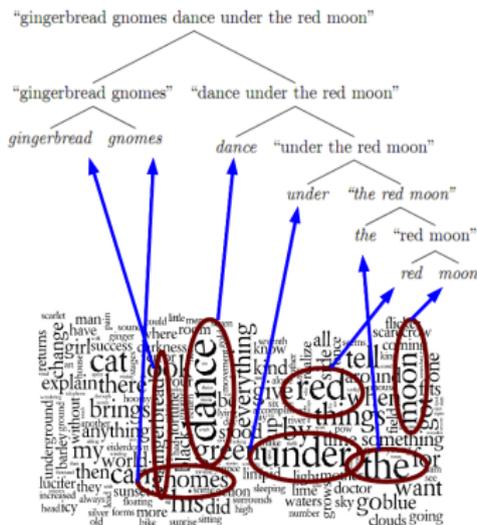
Noun phrase translation:

- ▶ Use small dictionary to learn cross-lingual mapping
- ▶ Translate source language space into target space
- ▶ Learn joint (de)composition functions
- ▶ Test on 3000 Adjective-Noun English-Italian phrases



That's all, folks!

# Thank you!



<http://clic.cimec.unitn.it/marco>