Distributional Semantics and Linguistic Theory

Gemma Boleda
Universitat Pompeu Fabra / ICREA

CLASP seminar
Gothenburg, Sweden, April 29 2020

Thanks

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 715154).

Outline

Outline

The distributional hypothesis in real life

Jurafsky & Martin, SNLP3, Chapter 6.2

What is **Ongchoi**?

The distributional hypothesis in real life

Jurafsky & Martin, SNLP3, Chapter 6.2

What is **Ongchoi**?

- Ongchoi is delicious sauteed with garlic.
- Ongchoi is superb over rice.
- . . . ongchoi leaves with salty sauces. . .

The distributional hypothesis in real life

Jurafsky & Martin, SNLP3, Chapter 6.2

What is **Ongchoi**?

- Ongchoi is delicious sauteed with garlic.
- Ongchoi is superb over rice.
- . . . ongchoi leaves with salty sauces. . .



top-bottom, left-right: debaird / DeusXFlorida (flickr) / Eric in SF (Wikicommons) CC BY-SA 4.0/2.0

The distributional hypothesis in real life

Jurafsky & Martin, SNLP3, Chapter 6.2

What is **Ongchoi**?

- Ongchoi is delicious sauteed with garlic.
- Ongchoi is superb over rice.
- . . . ongchoi leaves with salty sauces. . .



top-bottom, left-right: debaird / DeusXFlorida (flickr) / Eric in SF (Wikicommons) CC BY-SA 4.0/2.0

The distributional hypothesis

Jurafsky & Martin, SNLP3, Chapter 6.2

What is **Ongchoi**?

- Ongchoi is delicious sauteed with garlic.
- Ongchoi is superb over rice.
- . . . ongchoi leaves with salty sauces. . .

- . . . spinach sauteed with garlic over rice. . .
- . . . chard stems and leaves are delicious. . .
- . . . collard greens and other salty leafy greens. . .

The distributional hypothesis

- A word is defined by the environment or distribution it occurs in language use: the set of contexts in which it occurs
- Two words that have have related meanings are likely to have similar distributions
(Joos, 1950; Harris, 1954; Firth, 1957)

Slide by Carina Silberer

Distributional semantics in a nutshell

meaning

$\Downarrow$

distribution

Distributional semantics in a nutshell

meaning

$\Downarrow$

distribution

meaning

$\Uparrow$

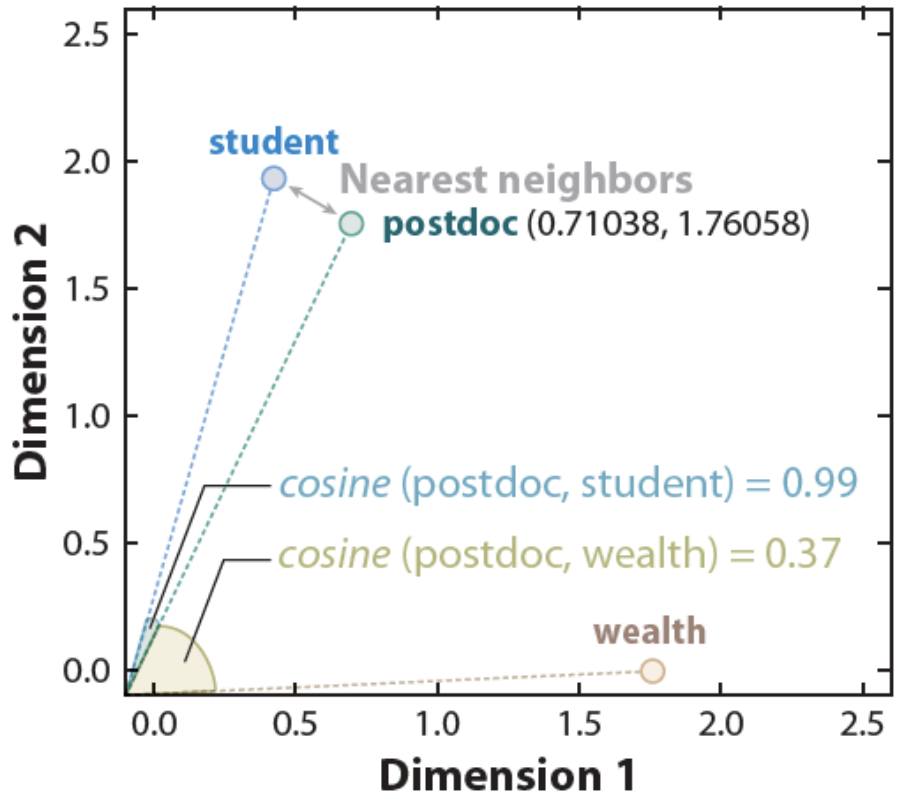
distribution

Distributional semantics in a nutshell

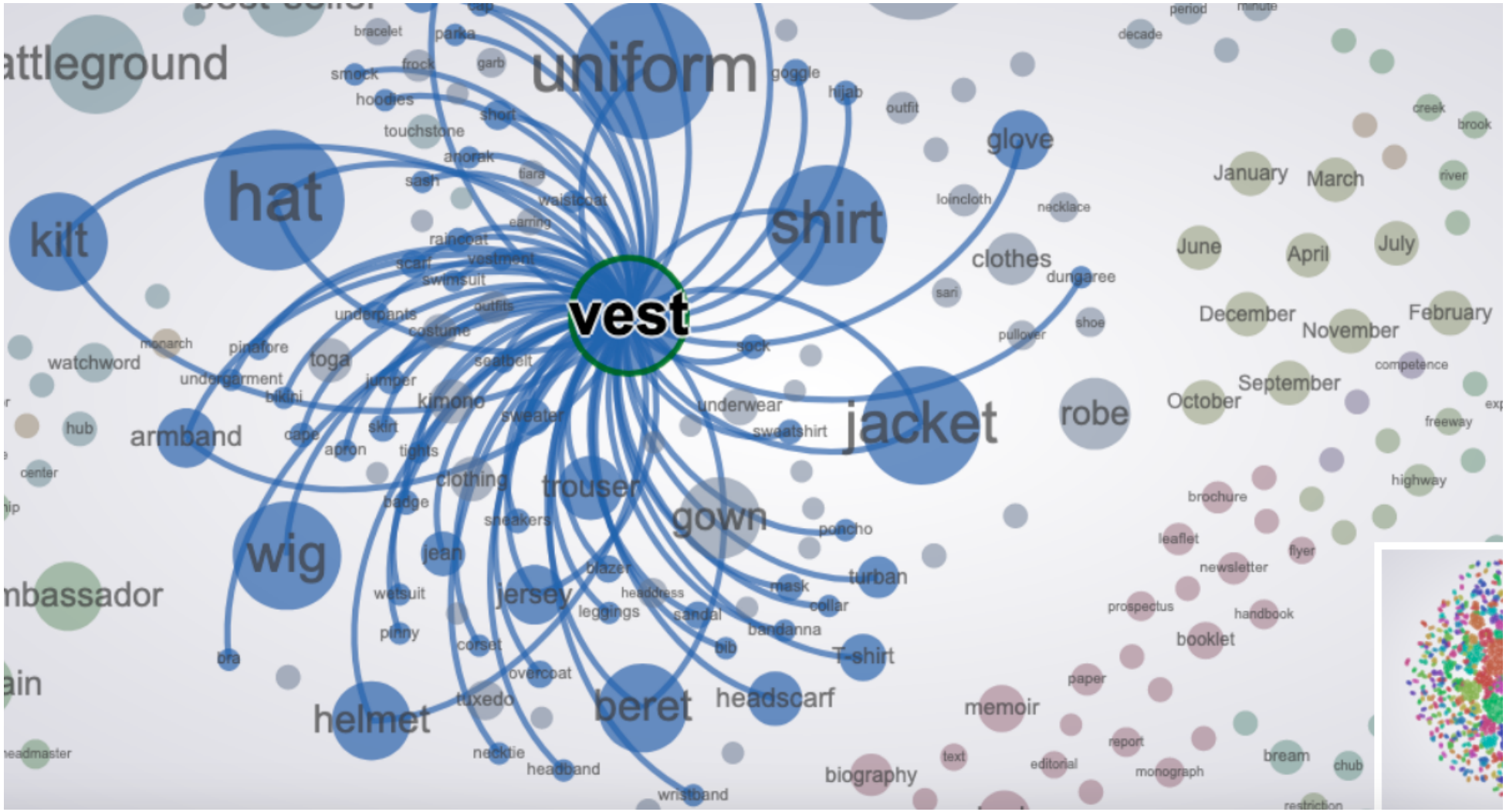
Boleda 2020, *Annu. Rev. Ling.* 6:213-23, Fig. 1

Any grad student or postdoc he'd have would be a clonal copy of himself. During that postdoc, I didn't publish much. . . .

	Dimension 1	Dimension 2
postdoc	0.71038	1.76058
student	0.43679	1.93841
wealth	1.77337	0.00012



"Vest" in a real semantic space



From http://colinglab.humnet.unipi.it/Demo/DistributionalMemoryNouns/

Outline

What does distributional semantics model?

- speaker meaning: what a given speaker communicates with the use of a specific expression in a given context
- expression meaning: what a linguistic expression signifies outside of any particular context

Distributional semantics models expression meaning

Westera and Boleda 2019

- models expression meaning, not speaker meaning
  - abstractions over contexts of use
    $\rightarrow$ context-independent representations
- very successful for lexical semantics
- and conceptual aspects of meaning more generally

Distributional semantics models expression meaning

Westera and Boleda 2019

- models expression meaning, not speaker meaning
  - abstractions over contexts of use
    $\rightarrow$ context-independent representations
- very successful for lexical semantics
- and conceptual aspects of meaning more generally
- we suggest:
  - distributional semantics: expression meaning
  - formal semantics: speaker meaning

Distributional semantics as a model of word meaning

Boleda and Erk 2015; Boleda 2020

- strong version

  "The meaning of a word is its use in the language" (Wittgenstein, 1953, PI 43)

  "the meaning of an expression is an abstraction over its uses" (Westera and Boleda 2019, p. 124)

Distributional semantics as a model of word meaning

Boleda and Erk 2015; Boleda 2020

- strong version

  "The meaning of a word is its use in the language" (Wittgenstein, 1953, PI 43)

  "the meaning of an expression is an abstraction over its uses" (Westera and Boleda 2019, p. 124)

- weak version
  learnt, multi-dimensional, graded

Distributional semantics captures semantic features. . .

Boleda and Erk 2015

man
woman
gentleman
gray-haired
boy
person
lad
men
girl

Words most similar to *man* in Baroni et al. (2014)

Distributional semantics captures semantic features. . .

Boleda and Erk 2015

man

---

woman

gentleman

gray-haired

boy

person

lad

men

girl

+HUMAN

Words most similar to *man* in Baroni et al. (2014)

Distributional semantics captures semantic features. . .

Boleda and Erk 2015

man

---

woman

gentleman

gray-haired

boy

person

lad

men

girl

+HUMAN +MALE

Words most similar to *man* in Baroni et al. (2014)

Distributional semantics captures semantic features. . .

Boleda and Erk 2015

man

---

woman

gentleman

gray-haired

boy

person

lad

men

girl

+HUMAN +MALE +ADULT

Words most similar to *man* in Baroni et al. (2014)

. . . and semantic nuances

Boleda and Herbelot 2016

man	chap	lad	dude	guy
woman	bloke	boy	freakin'	bloke
gentleman	guy	bloke	woah	chap
gray-haired	lad	scouser	dorky	doofus
boy	fella	lass	dumbass	dude
person	man	youngster	stoopid	fella

Words most similar to *man*, *chap*, *lad*, *dude*, *guy* in Baroni et al. (2014).

. . . and semantic nuances

Boleda and Herbelot 2016

man	chap	lad	dude	guy
woman	bloke	boy	freakin'	bloke
gentleman	guy	bloke	woah	chap
gray-haired	lad	scouser	dorky	doofus
boy	fella	lass	dumbass	dude
person	man	youngster	stoopid	fella

Words most similar to *man*, *chap*, *lad*, *dude*, *guy* in Baroni et al. (2014).

Outline

Outline

Semantic change: distributional approaches

Sagi et al. 2009, Kim et al. 2014, Hamilton et al. 2016, Del Tredici et al. 2019

1900 We assembled around the breakfast with spirits as gay and appetites as sharp as ever.

2000 . . . the expectation that effeminate men and masculine women are more likely to be seen as gay men and lesbians, respectively.

Semantic change: distributional approaches

Sagi et al. 2009, Kim et al. 2014, Hamilton et al. 2016, Del Tredici et al. 2019

1900 We assembled around the breakfast with spirits as gay and appetites as sharp as ever.

2000 . . . the expectation that effeminate men and masculine women are more likely to be seen as gay men and lesbians, respectively.

change in meaning

$\Downarrow$

change in context

Semantic change: distributional approaches

Sagi et al. 2009, Kim et al. 2014, Hamilton et al. 2016, Del Tredici et al. 2019

1900 We assembled around the breakfast with spirits as gay and appetites as sharp as ever.

2000 . . . the expectation that effeminate men and masculine women are more likely to be seen as gay men and lesbians, respectively.

change in meaning $\Downarrow$ change in context

change in meaning $\Uparrow$ change in context

Semantic change: distributional approaches

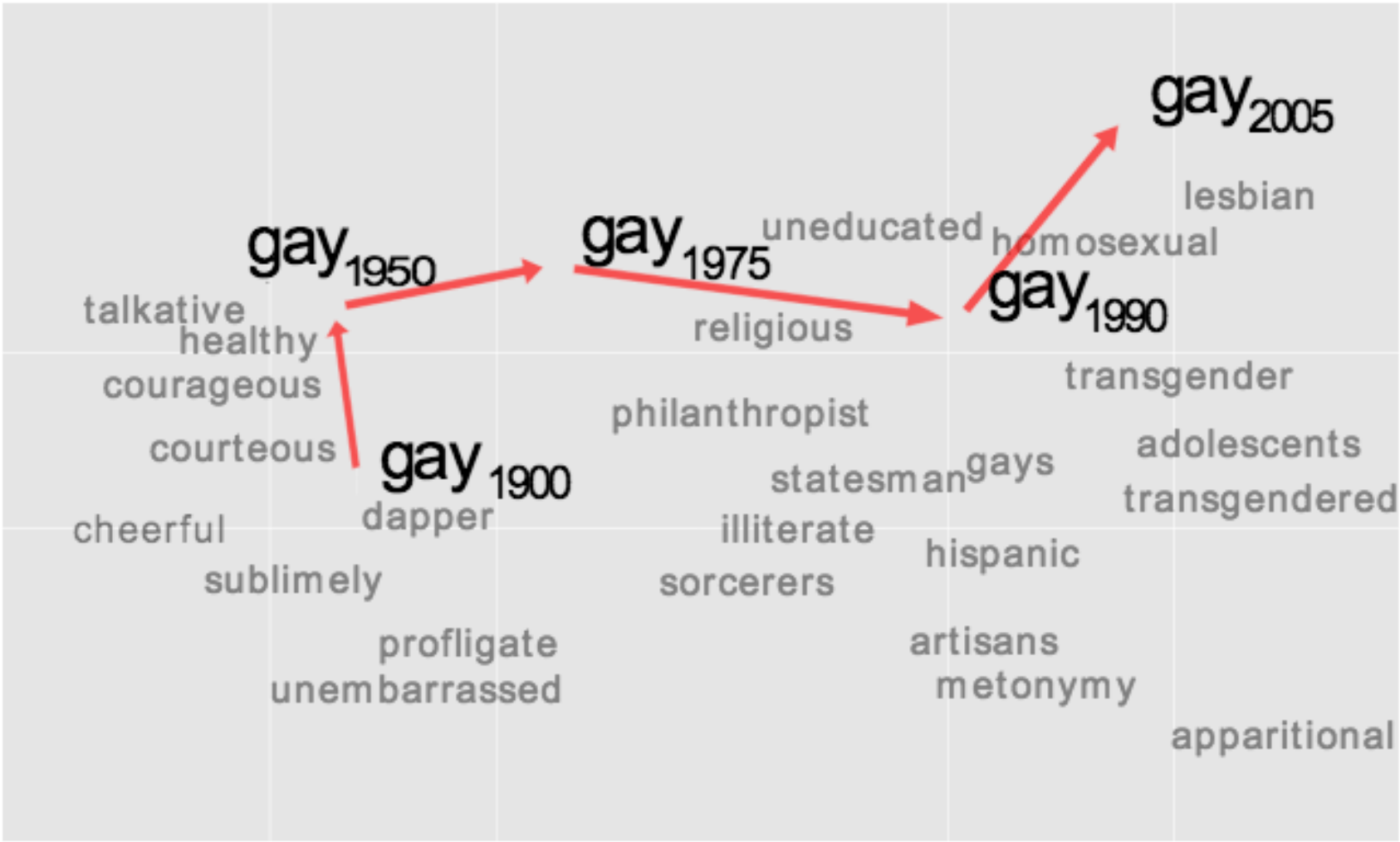
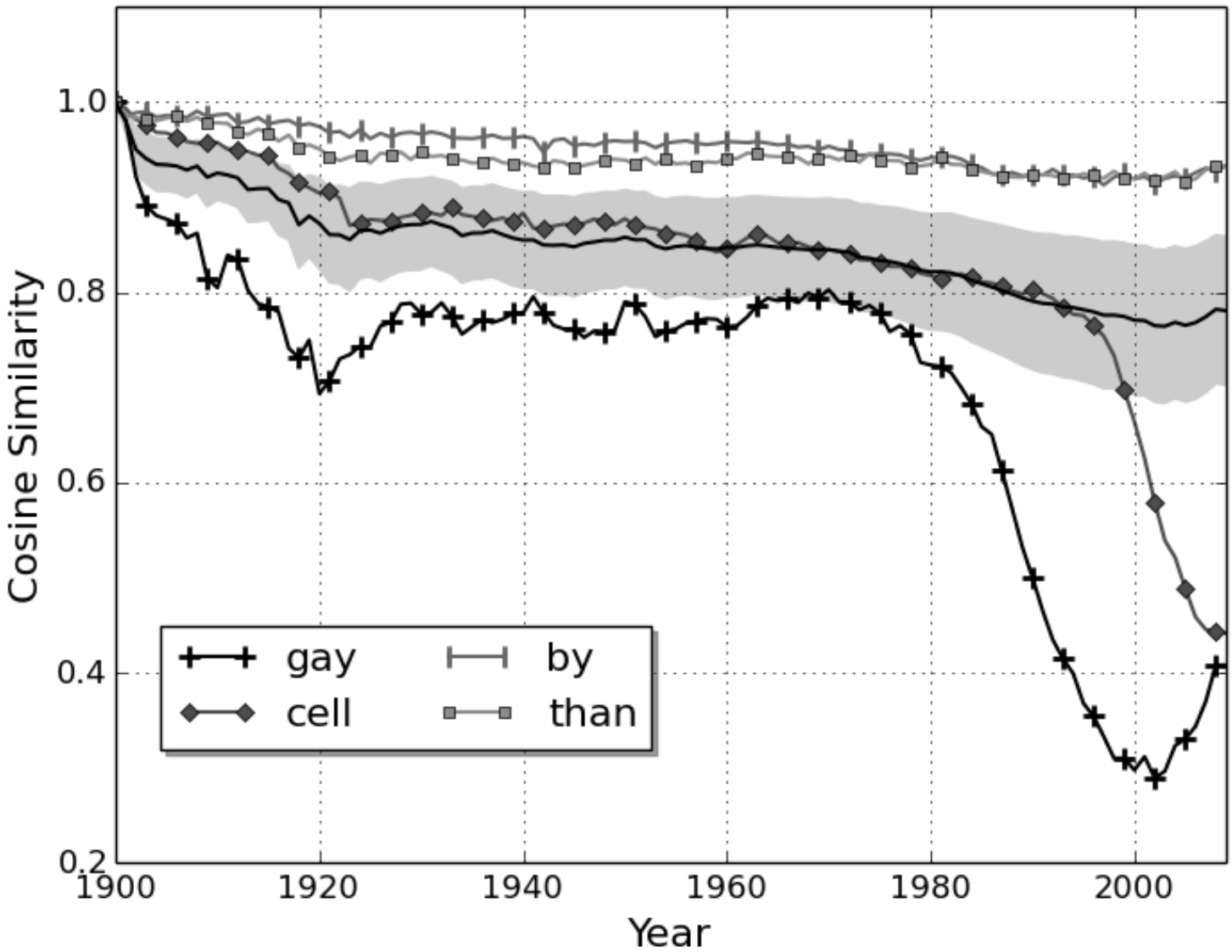


Figure from Kulkarni et al. 2015

Semantic change: distributional approaches



Outline

Polysemy

cut (Wiktionary, entry for *cut*)

1 To incise, to cut into the surface of something. *You must* **cut** *this flesh from off his breast.*

...

3 To separate, remove, reject or reduce. *They're going to* **cut** *salaries by fifteen percent.*

...

Polysemy

cut (Wiktionary, entry for *cut*)

1 To incise, to cut into the surface of something. *You must* **cut** *this flesh from off his breast.*

...

3 To separate, remove, reject or reduce. *They're going to* **cut** *salaries by fifteen percent.*

...

- "sense enumeration": how many senses? how to account for relationships between senses?
- Generative Lexicon and other approaches: Single representation, polysemy via composition.

Single representation, polysemy via composition

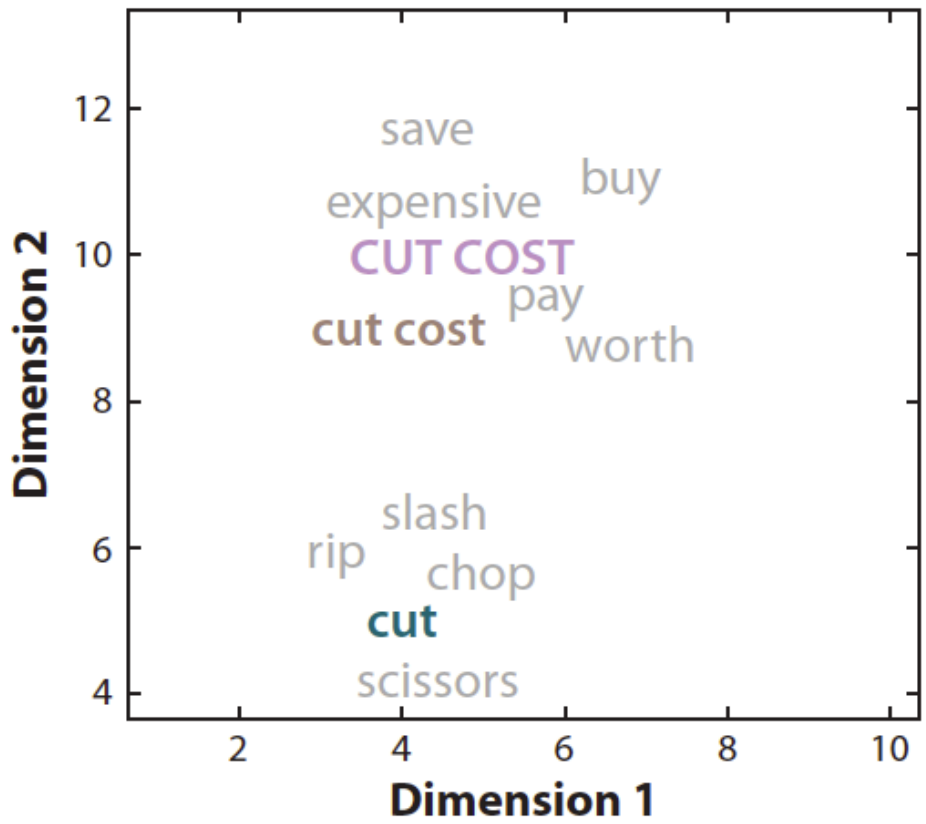
Boleda 2020, *Annu. Rev. Ling.* 6:213-23, Fig. 3

Corpus based

	Dimension 1	Dimension 2
cut	4	5
cost	1	5
cut cost	4	9

Synthetic

	Dimension 1	Dimension 2
CUT COST	5	10



Outline

Syntax-semantics interface and argument structure

Levin 1993, Grimshaw 1990, a.o.

"the behavior of a verb, particularly with respect to the expression of its arguments, is to a large extent determined by its meaning" (Levin 1993, p. 1).

Syntax-semantics interface and argument structure

Levin 1993, Grimshaw 1990, a.o.

"the behavior of a verb, particularly with respect to the expression of its arguments, is to a large extent determined by its meaning" (Levin 1993, p. 1).

Example verb alternation:

John broke the vase - The vase broke ✔
*John minced the meat - *The meat minced* ✘

Detecting verb alternations with distributional semantics

Merlo and Stevenson 2001, Schulte im Walde 2006, Baroni and Lenci 2010

John broke the vase - The vase broke ✔

*John minced the meat - *The meat minced* ✘

- DS: detect alternation from distributional verb representations
- Baroni and Lenci 2010: based on the similarity between (abstractions over) subjects and objects of the verbs[1]

  break 0.6

  mince 0.1
- (many other methods)

[1] simplified illustration

Outline

Morphology-semantics interface: Derivational morphology

Phenomenon	Word
Affix polysemy	carver
	broiler
Sense selection	column
	columnist
Differential effect of the affix	industrial
	industrious

Compositional DS for derivational morphology

Lazaridou et al. 2013, Marelli and Baroni 2015, Padó et al. 2016, Cotterell and Schütze 2018

- *carve*, *-er*: corpus-based distributional representations
- combine compositionally: obtain synthetic vector CARVER

Corpus based

	Dimension 1	Dimension 2
cut	4	5
cost	1	5
cut cost	4	9

Synthetic

	Dimension 1	Dimension 2
CUT COST	5	10

Compositional DS for derivational morphology

Marelli and Baroni 2015

Phenomenon	Word	Nearest neighbors (selection)
Affix polysemy	CARVER	potter, engraver, goldsmith
	BROILER	oven, stove, to cook, kebab, done
Sense selection	*column*	arch, pillar, bracket, numeric
	COLUMNIST	publicist, journalist, correspondent
Differential effect	INDUSTRIAL	environmental, land-use, agriculture
of the affix	INDUSTRIOUS	frugal, studious, hard-working

Outline

Distributional semantics and linguistic theory

- provides useful meaning representations on a large scale

$\rightarrow$ allows us to formulate and test predictions

Distributional semantics and linguistic theory

- provides useful meaning representations on a large scale

→ allows us to formulate and test predictions

- example: Boleda et al. IWCS 2013
  - formal semantics theory posits more complex types for adjectives like *alleged* than for adjectives like *round*
  - prediction: *alleged* should require a more sophisticated composition operation than *round*
  - test: 2 compositional distributional semantic models
  - (spoiler: the prediction is not borne out)

Distributional semantics and linguistic theory

However:

- most studies to date are within Computational Linguistics [and Cognitive Science]
- show that Distributional Semantics can do X:
  - spot semantic change
  - automatically determine whether given verbs participate in some alternation
  - ...
- very few studies using DS for **linguistically relevant research questions**
- (although this is changing fast)

Distributional semantics and linguistic theory

enormous potential: systematic

- exploration: distributional data (similarity scores, nearest neighbors)
- identification: specific instances of linguistic phenomena
  - e.g. words that undergo semantic change
- testbed for linguistic hypotheses
  - testing predictions in distributional terms
- actual discovery of linguistic phenomena

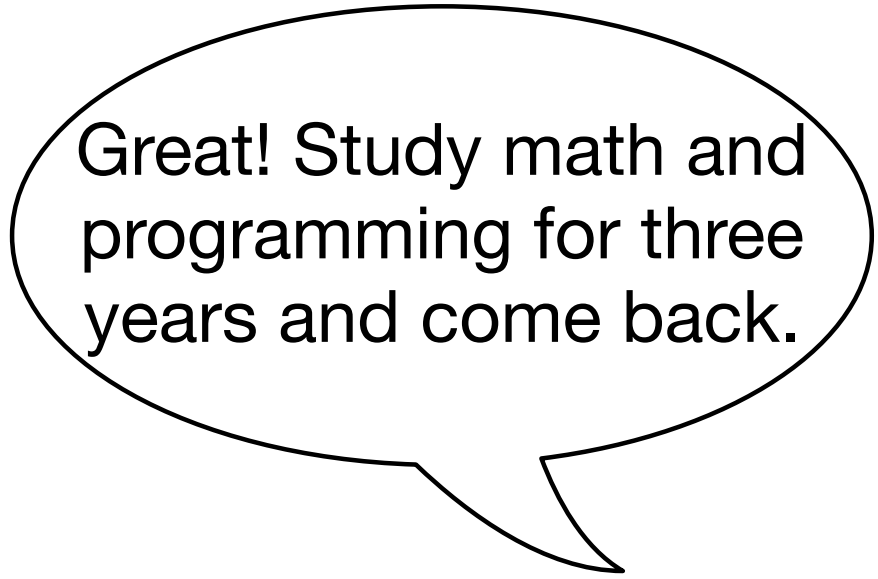
A real word vector for dog

0.006784000000000001 -0.08366999999999994 -0.0276 0.15977
-0.051539000000000001 0.2588099999999998 0.128604
0.043097000000000003 0.022886 0.1651209999999999
-0.1990580000000001 -0.1117559999999999 0.011864
-0.20073099999999999 0.168099 -0.146171 0.244815
-0.31515599999999999 0.012591 -0.09918899999999999
0.011284000000000001 0.1501929999999999
0.07532999999999994 -0.2389640000000001
0.032051999999999997 0.24129900000000001 0.058816
-0.38864799999999999 0.09967700000000002 0.183504
-0.018511 0.123728 0.19941200000000001 -0.191748
-0.019918000000000002 -0.101323 -0.029946
-0.005316999999999997 -0.007123 0.08295700000000003
-0.087373000000000006 0.272984 0.026393 0.124167 0.231517
-0.242756 -0.173259 -0.08976599999999999 0.204042 -0.017602
...

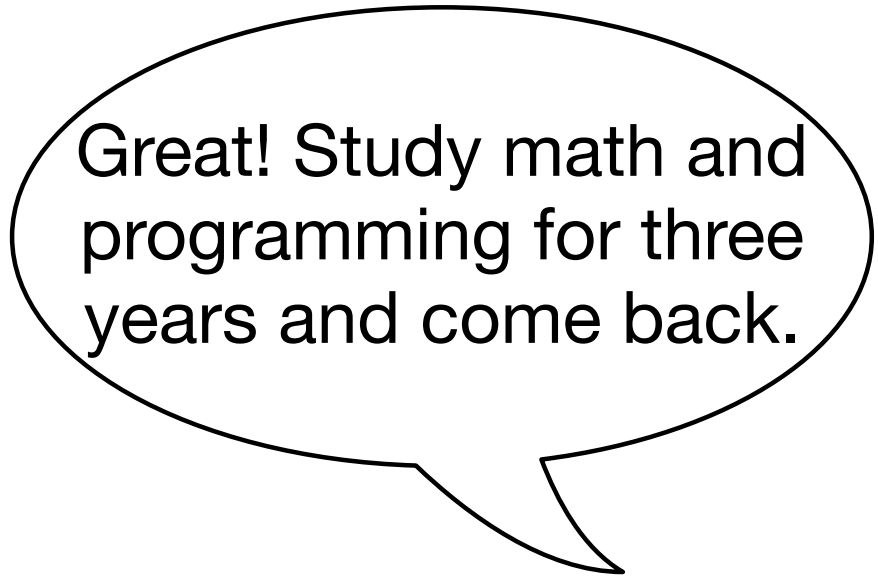
Representation of *dog* in the space of Baroni et al. (2014).

Challenge


What you do is cool! I want to do it, too!


Great! Study math and programming for three years and come back.

Challenge



- short- to mid-term: collaborate
- long-term: change training curriculum for Linguistics

Want to know more?

- Boleda, G. 2020. Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics*, Vol. 6: 213-23.
- web interface to an English space (has Dutch, too): http://meshugga.ugent.be/snaut-english
- web visualization tools: http://colinglab.humnet.unipi.it/Demo/

Distributional Semantics and Linguistic Theory

Gemma Boleda
Universitat Pompeu Fabra / ICREA

CLASP seminar
Gothenburg, Sweden, April 29 2020

Horizon 2020
European Union funding
for Research & Innovation

Appendix

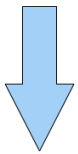
Distributional semantics in a nutshell

likely) mug of bourbon in hand. Some
stewed milk into a heavy mug, granules of
holding his coffee mug cupped in his hands.
drained his mug, dropping it over his
tablespoons of coffee and a single mug of
milk into the mug plus four spoons of sugar
placing the empty mug on the floor
picking up my mug with one hand and
followed by a very hot mug of tea into which
from time to time to drink a mug of tea. The
briefed, relax over a mug of tea and a
cake and cheese and a mug of strong, black
then we had a mug of cocoa and a gingerbread
and a white mug with a blurred inscription.
was carrying a mug of tea and

mug	0.984757	0.1098487	...
cup	0.9684626	0.2358760	...
dog	0.1640873	0.00123857	...

Distributional semantics in a nutshell

likely) mug of bourbon in hand. Some
stewed milk into a heavy mug, granules of
holding his coffee mug cupped in his hands.
drained his mug, dropping it over his
tablespoons of coffee and a single mug of
milk into the mug plus four spoons of sugar
placing the empty mug on the floor
picking up my mug with one hand and
followed by a very hot mug of tea into which
from time to time to drink a mug of tea. The
briefed, relax over a mug of tea and a
cake and cheese and a mug of strong, black
then we had a mug of cocoa and a gingerbread
and a white mug with a blurred inscription.
was carrying a mug of tea and



- word vectors, aka word embeddings
- semantic spaces, aka vector space models

mug	0.984757	0.1098487	...
cup	0.9684626	0.2358760	...
dog	0.1640873	0.00123857	...

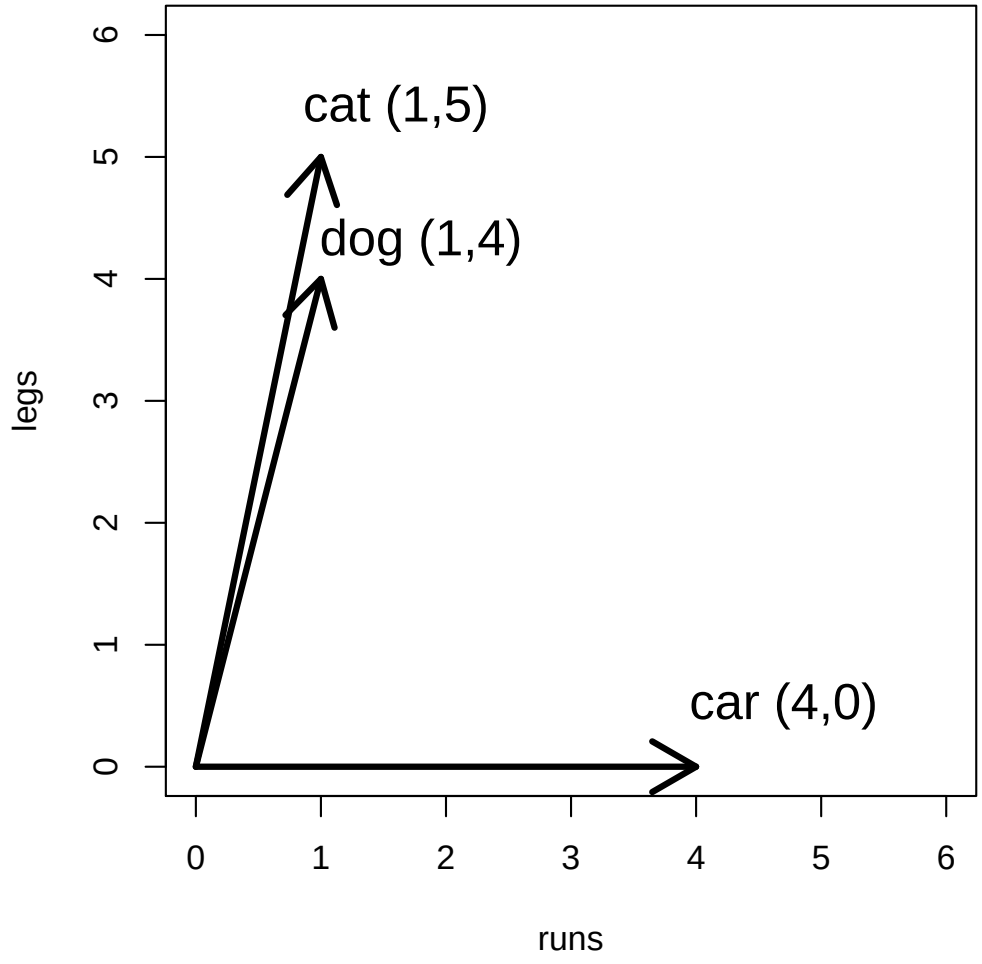
Words as vectors

	runs	*sleeps*
dog	1	4
cat	1	5
car	4	0

Based on material by Marco Baroni

Words as vectors

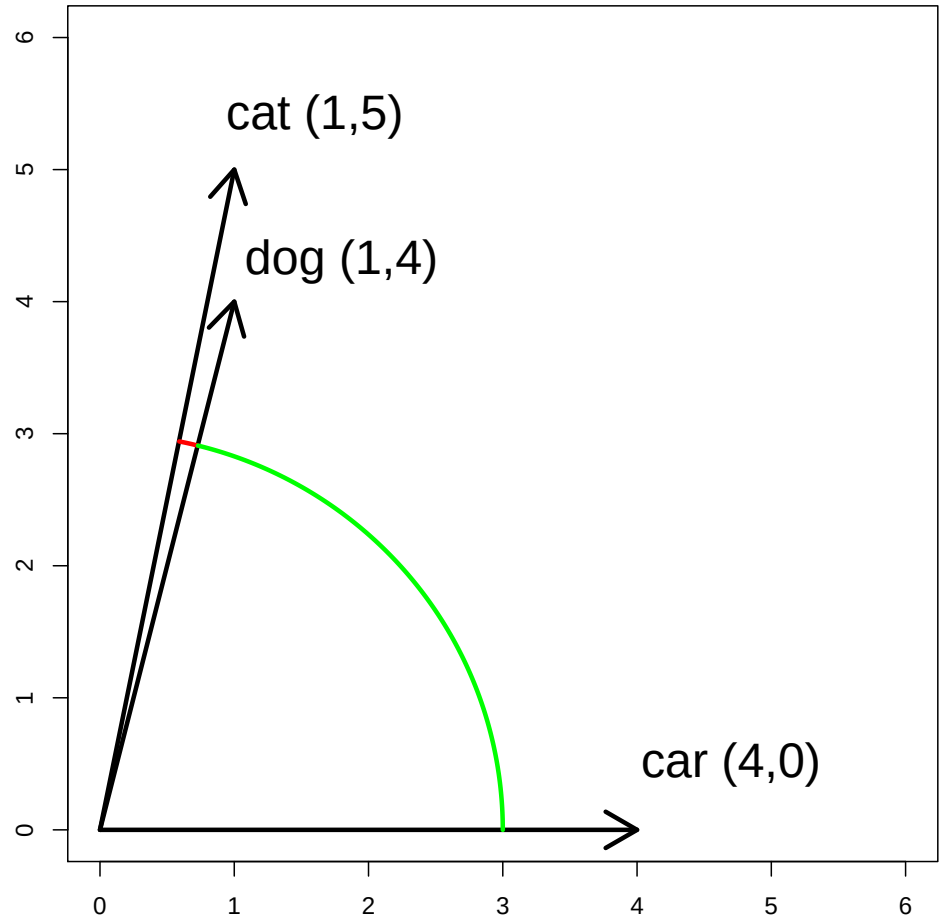
	runs	*sleeps*
dog	1	4
cat	1	5
car	4	0



Based on material by Marco Baroni

Words as vectors

	runs	*legs*
dog	1	4
cat	1	5
car	4	0



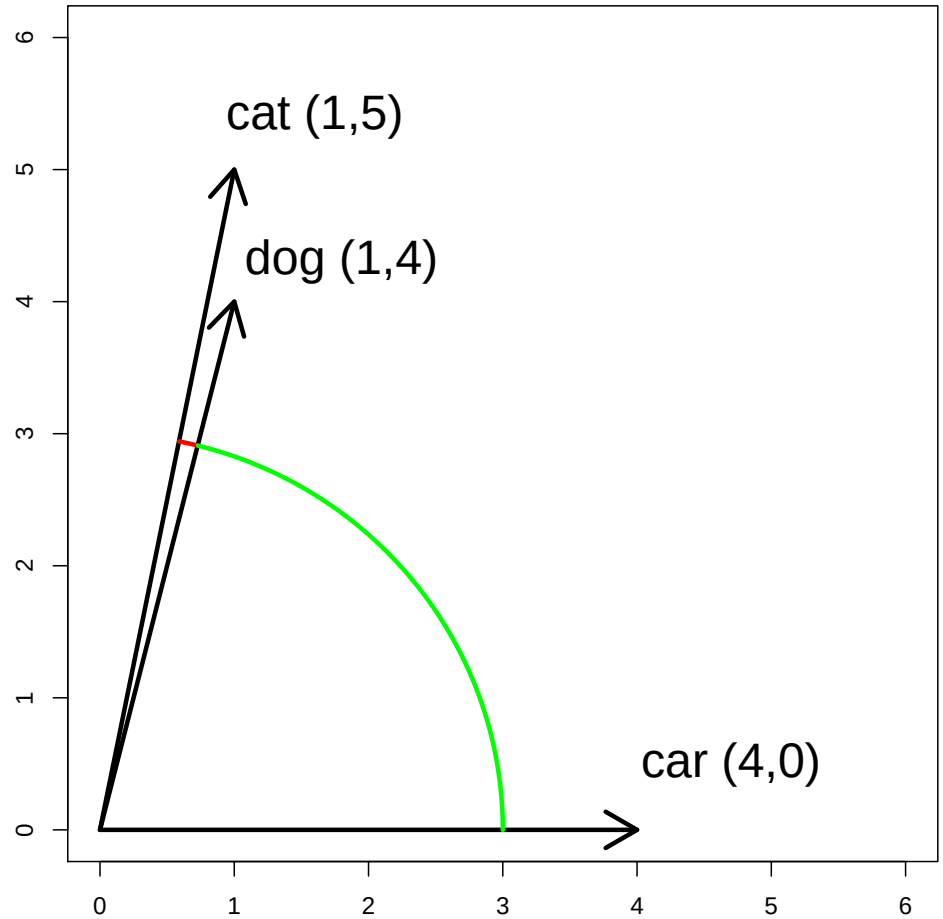
Based on material by Marco Baroni

Words as vectors

	runs	*legs*
dog	1	4
cat	1	5
car	4	0

cosine similarity:

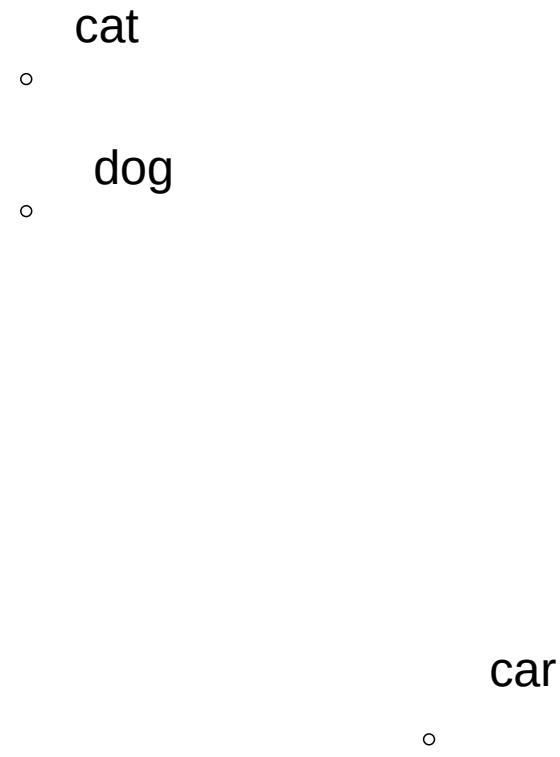
- *dog* - *cat*: 0.99
- *dog* - *car*: 0.20



Based on material by Marco Baroni

Words as vectors

	runs	*legs*
dog	1	4
cat	1	5
car	4	0



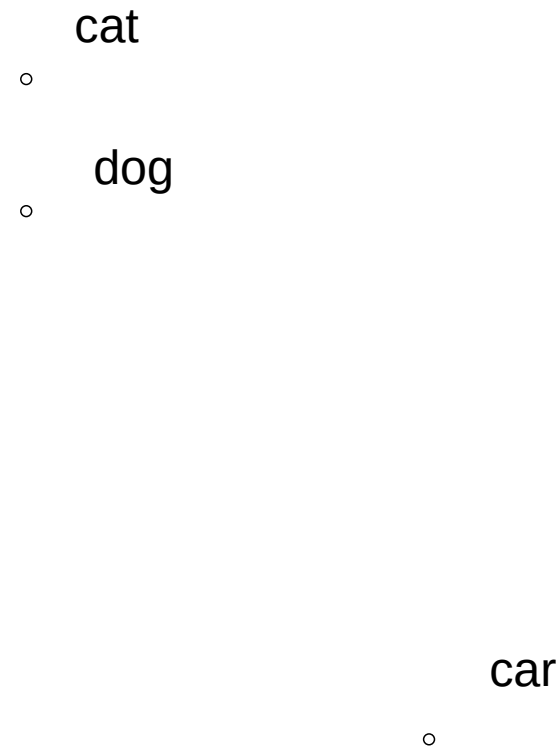
cosine similarity:

- *dog* - *cat*: 0.99
- *dog* - *car*: 0.20

Based on material by Marco Baroni

Words as vectors

	runs	*legs*
dog	1	4
cat	1	5
car	4	0



cosine similarity:

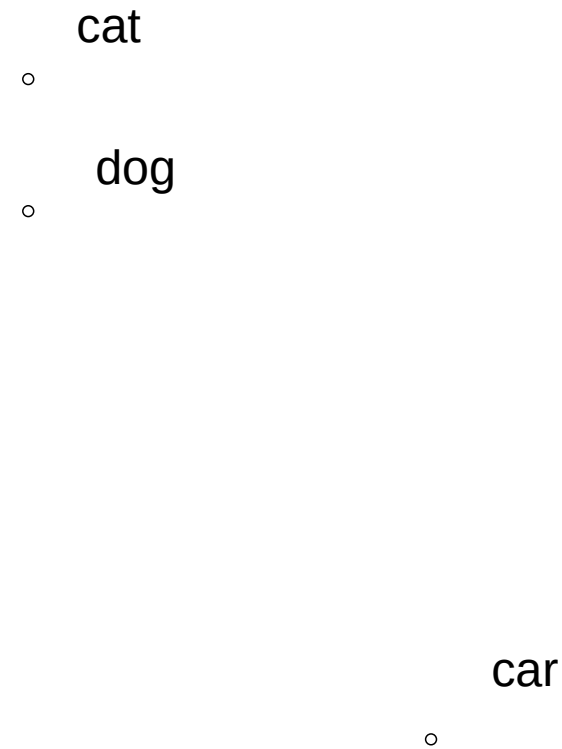
- *dog* - *cat*: 0.99
- *dog* - *car*: 0.20

nearest neighbor

Based on material by Marco Baroni

Words as vectors

	runs	*legs*
dog	1	4
cat	1	5
car	4	0



cosine similarity:

- *dog* - *cat*: 0.99
- *dog* - *car*: 0.20

nearest neighbor

Why is this approach to meaning unsatisfactory?

Based on material by Marco Baroni

What is "context"?

The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.

Based on material by Marco Baroni

What is "context"?

Content words in a sentence window

The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.

Based on material by Marco Baroni

What is "context"?

Morphologically coded content lemmas filtered by syntactic path, with the syntactic path encoded as part of the context

The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitter-v_subj although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.

Based on material by Marco Baroni

What is "context"?

Not only text!

The silhouette of the sun beyond a wide-open bay on the lake; the sun still glitters although evening has arrived in Kuhmo. It's midsummer; the living room has its instruments and other objects in each of its corners.

Based on material by Marco Baroni

Same corpus (BNC), different contexts (window sizes)

Nearest neighbours of *dog*

2-word window

- cat
- horse
- fox
- pet
- rabbit
- pig
- animal
- mongrel
- sheep
- pigeon

30-word window

- kennel
- puppy
- pet
- bitch
- terrier
- rottweiler
- canine
- cat
- to bark
- Alsatian

Slide by Marco Baroni

Selectional preferences

Model: Padó et al. (2007); implementation: Baroni and Lenci (2010)

Acceptability of some potential objects of *kill*

Selectional preferences

Model: Padó et al. (2007); implementation: Baroni and Lenci (2010)

Acceptability of some potential objects of *kill*

object	*cosine*
kangaroo	0.51
person	0.45
robot	0.15
hate	0.11
flower	0.11
stone	0.05
fun	0.05
book	0.04
conversation	0.03
sympathy	0.01

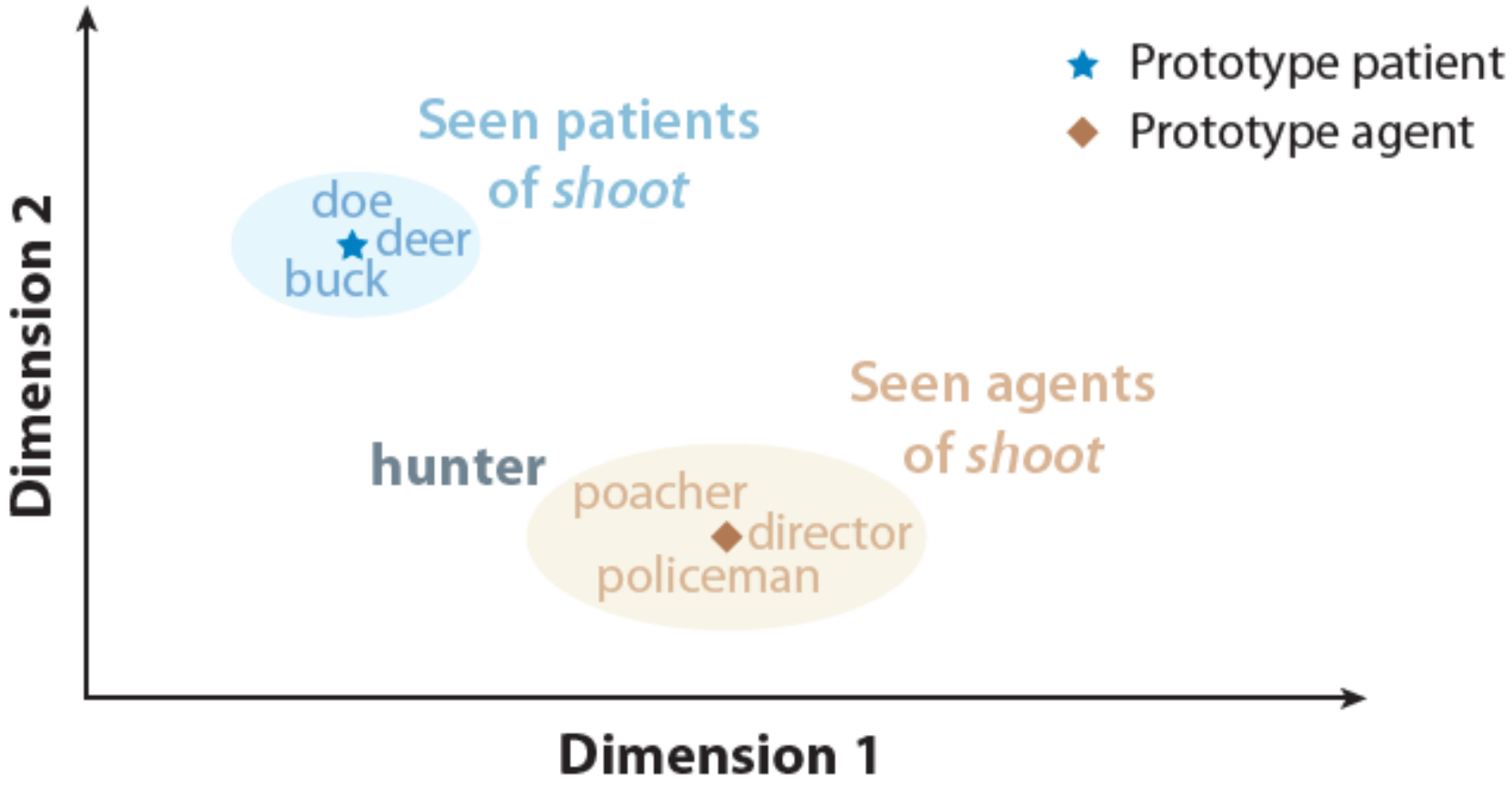
Selectional preferences

Model: Padó et al. (2007); implementation: Baroni and Lenci (2010)

Acceptability of some potential instruments of *kill*

with	*cosine*
hammer	0.26
stone	0.25
brick	0.18
smile	0.15
flower	0.12
antibiotic	0.12
person	0.12
heroin	0.12
kindness	0.07
graduation	0.04

Selectional preferences



Boleda 2020, Figure 4; adapted from Padó et al. 2007, Figure 1