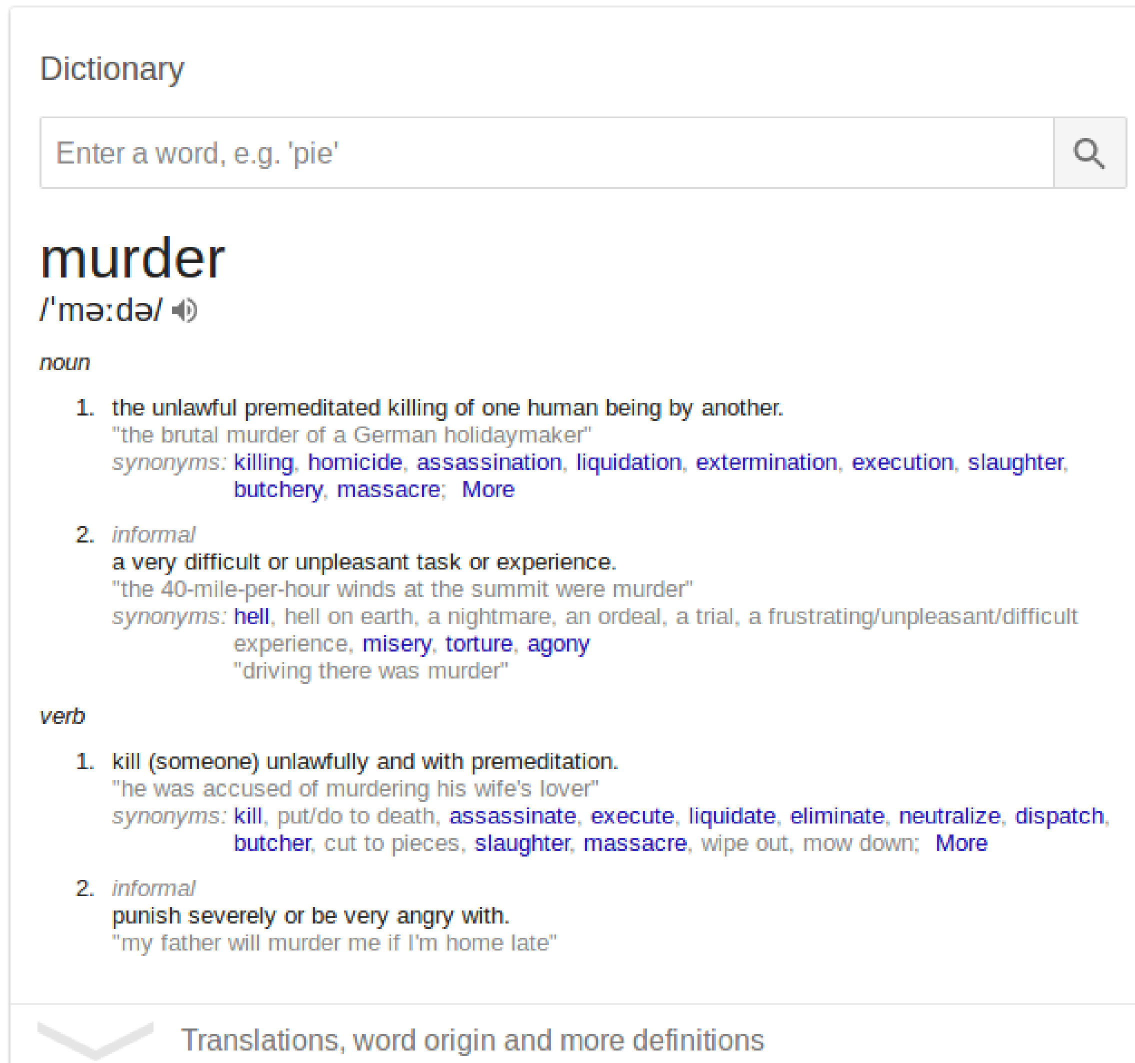
Deep Learning for NLP

Semantics

- How does a computer understands meaning?
  - Dictionary
  - Thesuras
  - WordNet



Semantics

- How does a computer understands meaning?

What is the meaning of “bardiwac”?

He handed her a glass of bardiwac.

Beef dishes are made to complement the bardiwacs.

Malbec, one of the lesser-known bardiwac grapes, responds well to Australia’s sunshine.

Word Sense Disambiguation

- What is the meaning of “Bank”?
  - he sat on the bank of the river and watched the currents --> sloping land
  - a huge bank of earth --> a long ridge or pile
  - The State Bank has allowed me the loan --> Financial Institution

Meaning and Neighbourhood

You shall know a word by the company it keeps!

Meaning of a word can be derived from the meaning of its contexts

Meaning and Neighbourhood

You shall know a word by the company it keeps!

Meaning of a word can be derived from the meaning of its contexts

AKA

The Distributional Hypothesis

Semantics

Meaning of a word can be derived from the meaning of its ***contexts***

Context??

He handed her a glass of bardiwac.

Beef dishes are made to complement the bardiwacs.

Malbec, one of the lesser-known bardiwac grapes, responds well to Australia’s sunshine.

Context??

- I love playing cricket
- Sachin is a cricketer
- Cricket is played using a bat and a ball
- Sourav plays cricket
- Sachin is the highest test scorer
- Maradona plays football
- ...
- ...

Context

- He handed her a glass of bardiwac.
- Beef dishes are made to complement the bardiwacs.
- Malbec, one of the lesser-known bardiwac grapes, responds well to Australia’s sunshine.

- I love playing cricket
- Sachin is a cricketer
- Cricket is played using a bat and a ball
- Sourav plays cricket
- Sachin is the highest test scorer
- Maradona plays football

Central/
Target Word

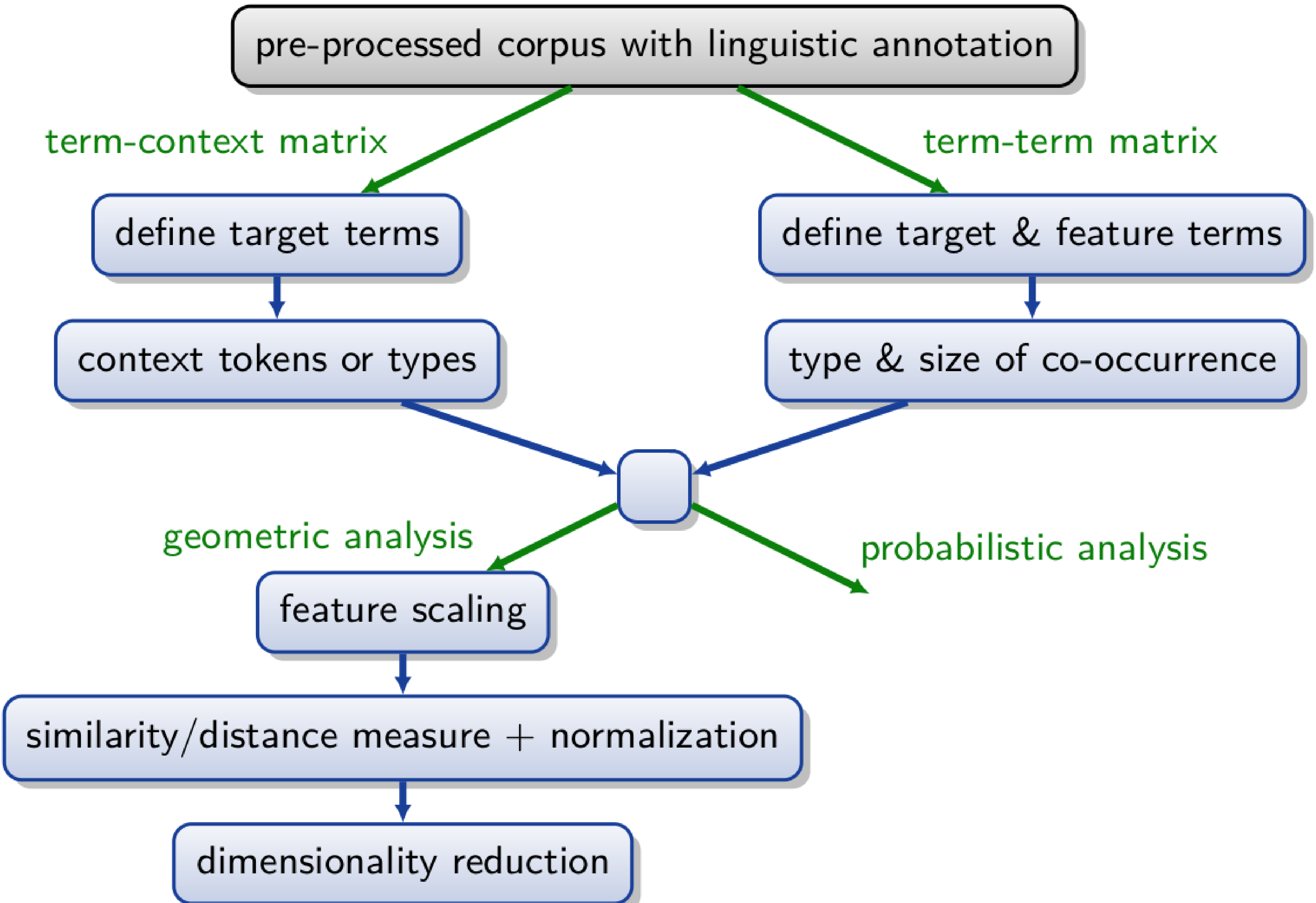
Context

- He handed her a glass of bardiwac.
- Beef dishes are made to complement the bardiwacs.
- Malbec, one of the lesser-known bardiwac grapes, responds well to Australia's sunshine.

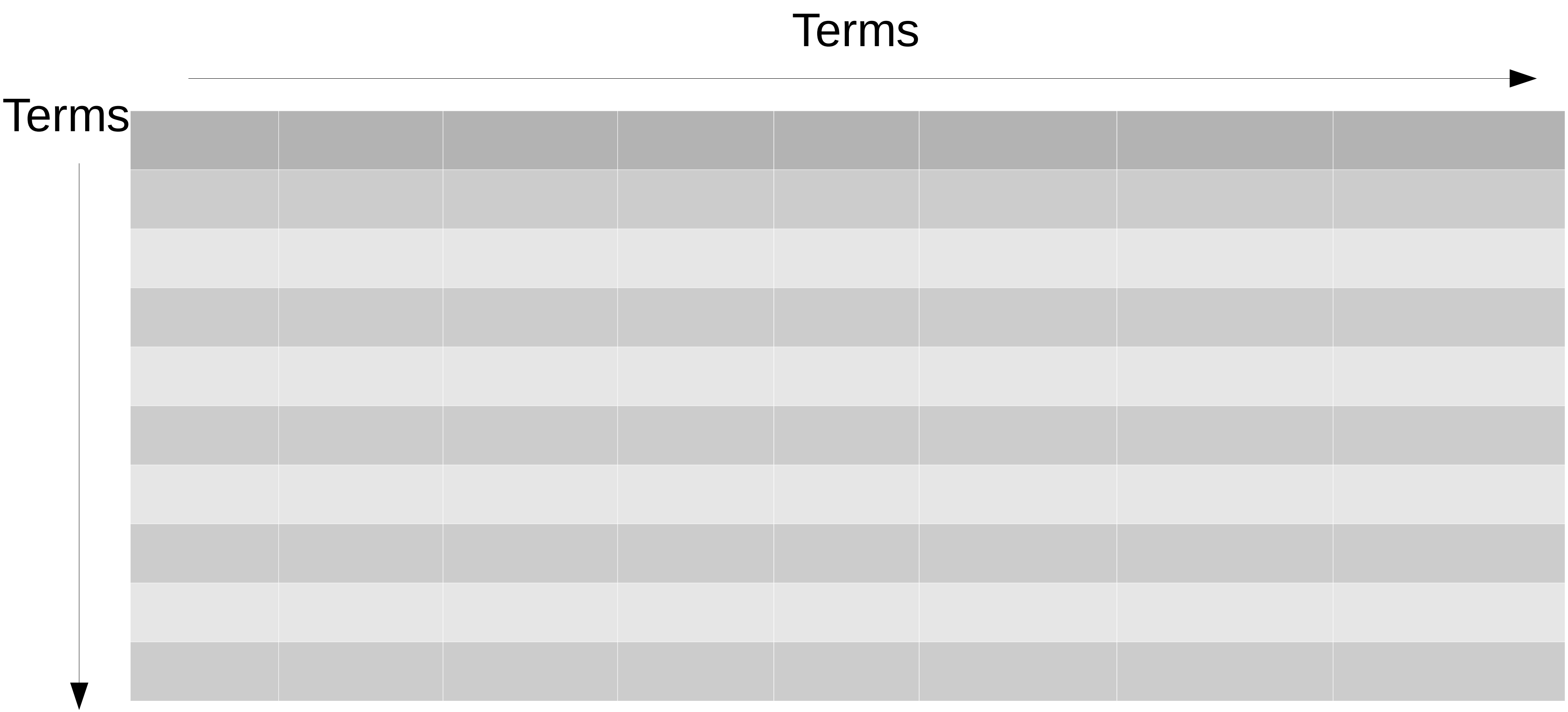
- I love playing cricket
- Sachin is a cricketer
- Cricket is played using a bat and a ball
- Sourav plays cricket
- Sachin is the highest test scorer
- Maradona plays football

Context Word

Building a distributional model



Construction of co-occurrence matrix



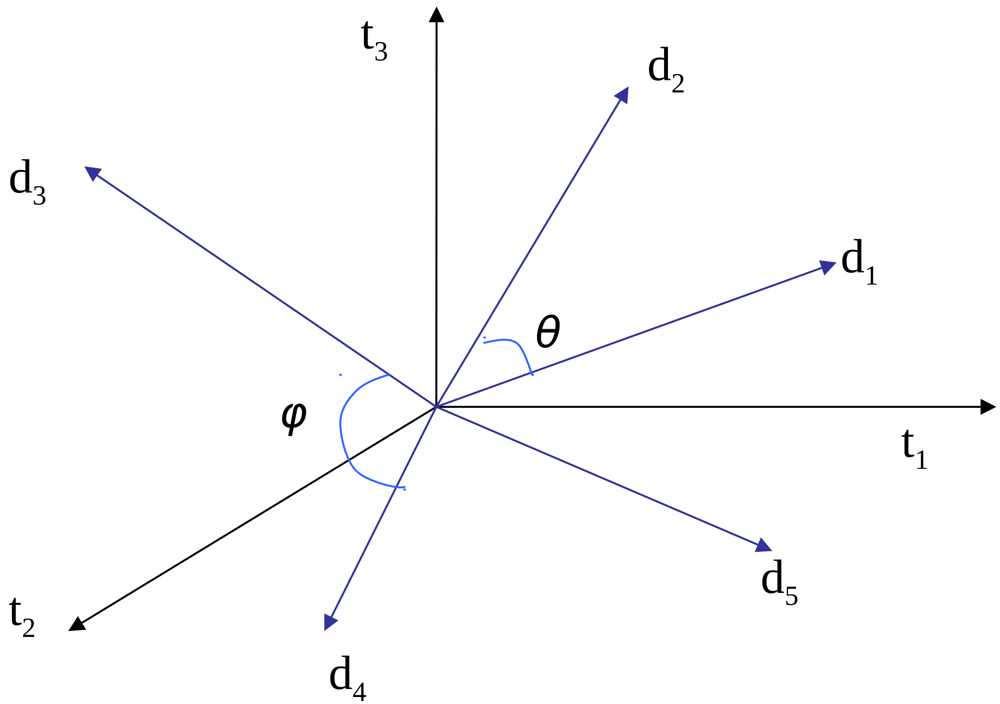
Geometric interpretation

- row vector $\mathbf{x}_{\text{dog}}$ describes usage of word *dog* in the corpus
- can be seen as coordinates of point in *n*-dimensional Euclidean space

	get	see	use	hear	eat	kill
knife	51	20	84	0	3	0
cat	52	58	4	4	6	26
dog	115	83	10	42	33	17
boat	59	39	23	4	0	0
cup	98	14	6	2	1	0
pig	12	17	3	2	9	27
banana	11	2	2	0	18	0

co-occurrence matrix M

Intuition



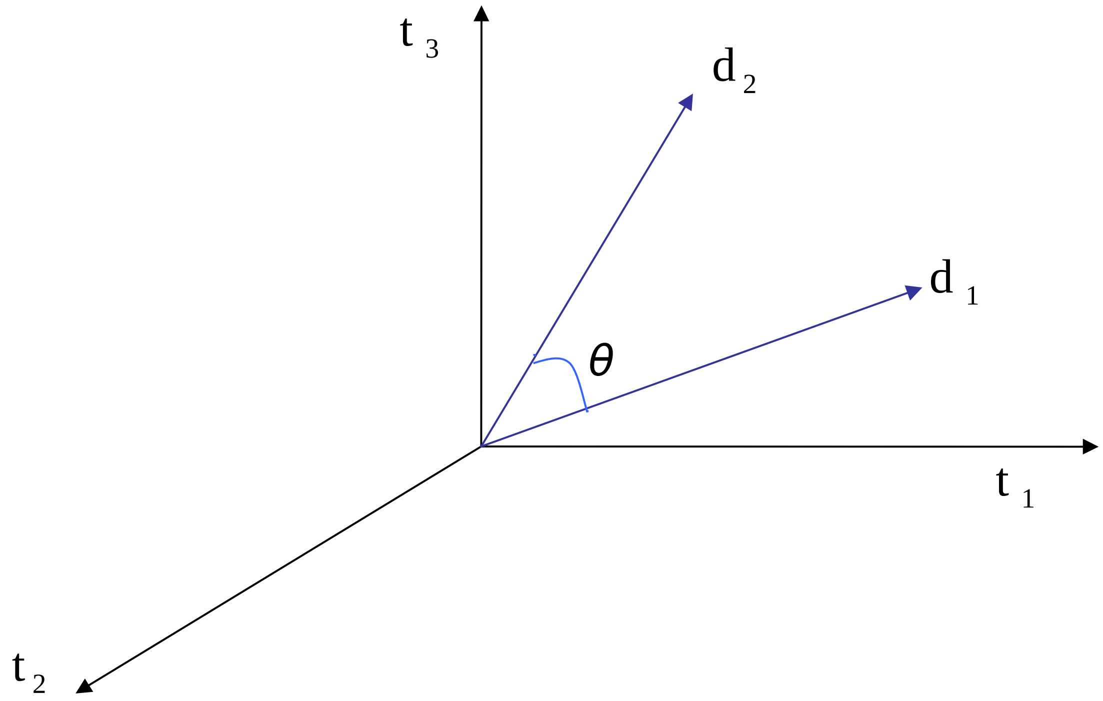
Postulate: Words that are "close together" in the vector space talk about the same things.

Desiderata for proximity

- If d_1 is near d_2, then d_2 is near d_1.
- If d_1 near d_2, and d_2 near d_3, then d_1 is not far from d_3.
- No word is closer to d than d itself.

Cosine similarity

- Distance between vectors d_1 and d_2 *captured* by the cosine of the angle *x* between them.



Similarity measures

- There are many different ways to measure how similar two documents are, or how similar a document is to a query
- The cosine measure is a very common similarity measure
- Using a similarity measure, a set of documents can be compared to a query and the most similar document returned

The cosine measure

- For two vectors d and d’ the cosine similarity between d and d’ is given by:

$$\frac{d \times d'}{|d||d'|}$$

- Here d X d’ is the vector product of d and d’, calculated by multiplying corresponding frequencies together
- The cosine measure calculates the angle between the vectors in a high-dimensional virtual space

Example

- Let d = (2,1,1,1,0) and d' = (0,0,0,1,0)
  - dXd' = 2X0 + 1X0 + 1X0 + 1X1 + 0X0=1
  - $|d| = \sqrt{(2^2+1^2+1^2+1^2+0^2)} = \sqrt{7}=2.646$
  - $|d'| = \sqrt{(0^2+0^2+0^2+1^2+0^2)} = \sqrt{1}=1$
  - Similarity = 1/(1 X 2.646) = 0.378
- Let d = (1,0,0,0,1) and d' = (0,0,0,1,0)
  - Similarity =

Class Assignments

- Let d = (1,0,0,0,1) and d’ = (0,0,0,1,0)
  - Similarity =
- Let d = (21,10,25,30,1) and d’ = (10,20,50,11,54)
  - Similarity =
- Let d1 = (21,10,25,30,1), d2 = (10,20,50,11,54) and d3= (13, 16,1,10,5)
  - Similarity (d1,d2) =
  - Similarity (d2,d3) =
  - Similarity (d3,d1) =

Which pair is most similar??

Cosine similarity

$$sim(d_j, d_k) = \frac{\vec{d_j} \cdot \vec{d_k}}{\left|\vec{d_j}\right| \left|\vec{d_k}\right|} = \frac{\sum_{i=1}^{n} w_{i,j} w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2} \sqrt{\sum_{i=1}^{n} w_{i,k}^2}}$$

- Cosine of angle between two vectors
- The denominator involves the lengths of the vectors.

Normalization

Geometric interpretation

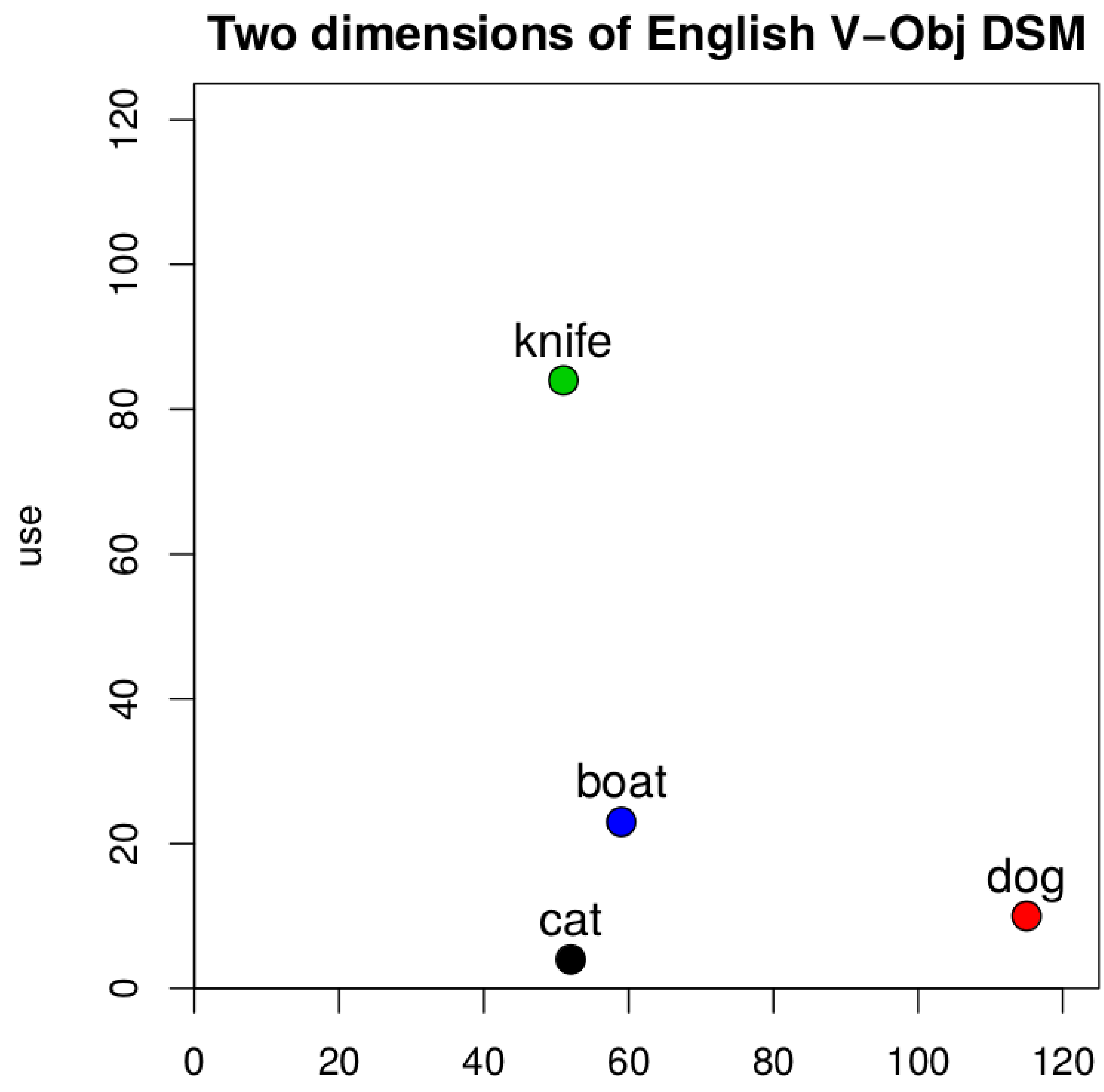
- row vector $\mathbf{x}_{\text{dog}}$ describes usage of word *dog* in the corpus
- can be seen as coordinates of point in n-dimensional Euclidean space

	get	see	use	hear	eat	kill
knife	51	20	84	0	3	0
cat	52	58	4	4	6	26
dog	115	83	10	42	33	17
boat	59	39	23	4	0	0
cup	98	14	6	2	1	0
pig	12	17	3	2	9	27
banana	11	2	2	0	18	0

co-occurrence matrix $\mathbf{M}$

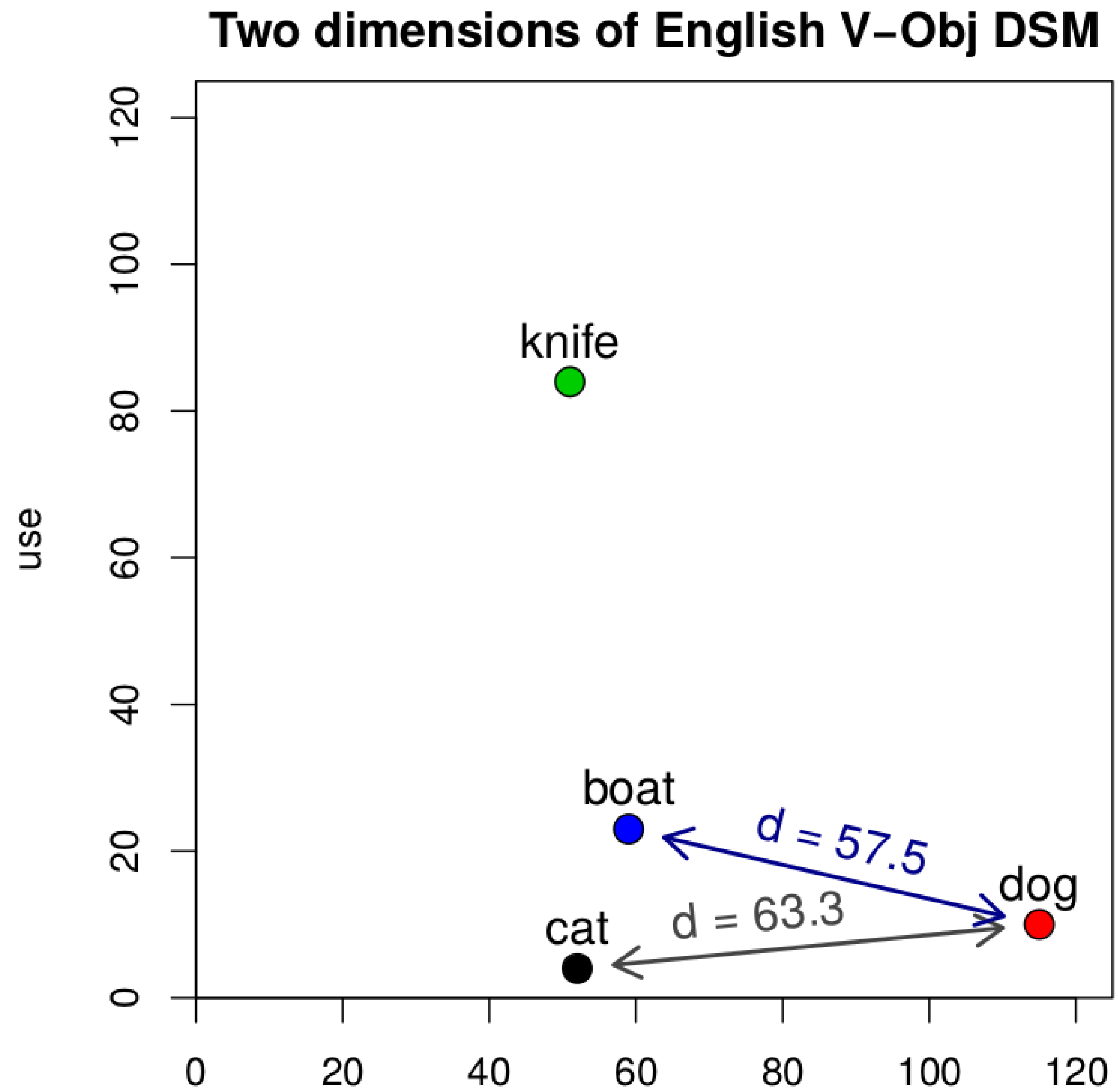
Geometric interpretation

- row vector $\mathbf{x}_{\text{dog}}$ describes usage of word *dog* in the corpus
- can be seen as coordinates of point in n-dimensional Euclidean space
- illustrated for two dimensions: *get* and *use*
- $\mathbf{x}_{\text{dog}} = (115, 10)$



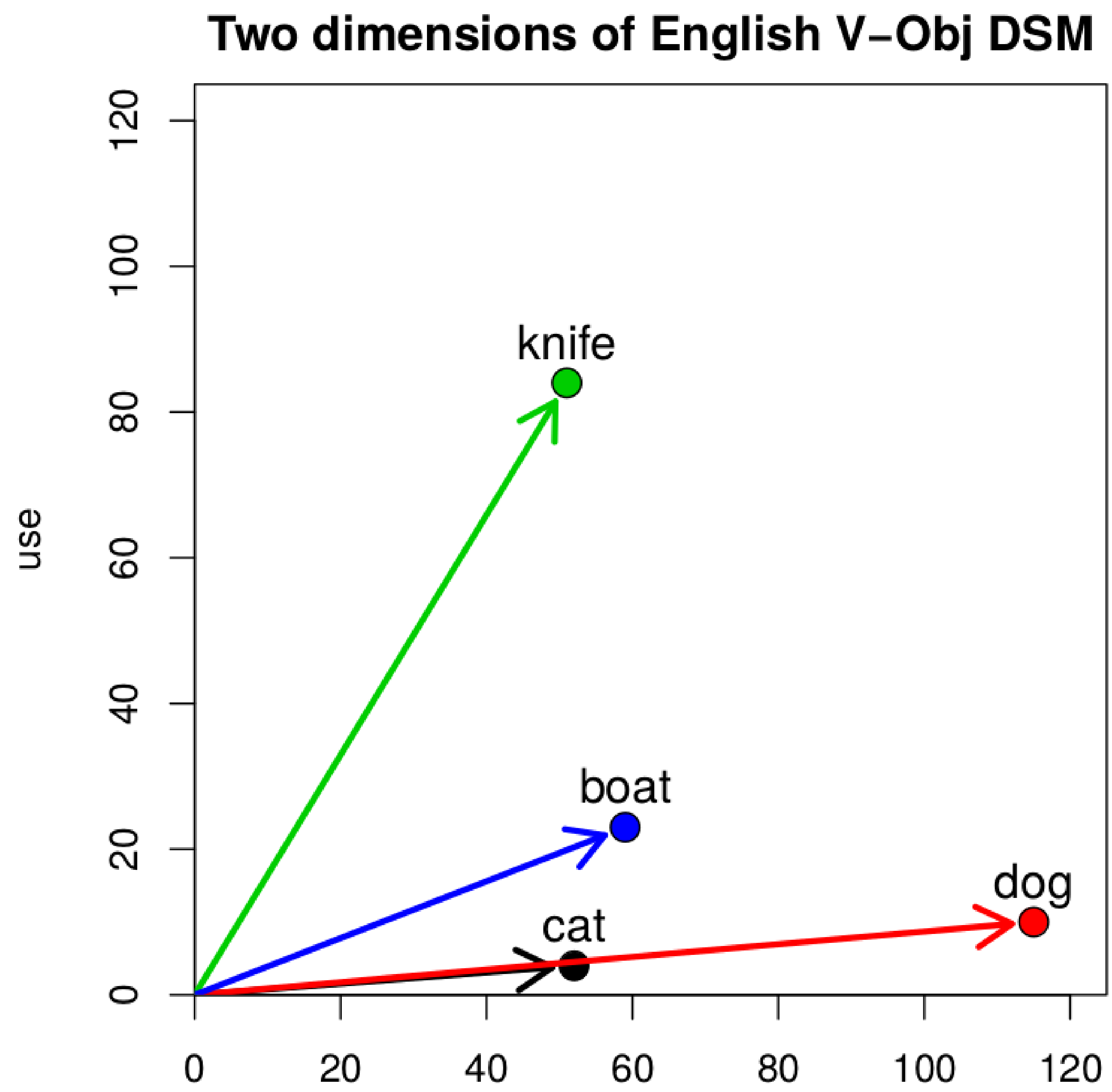
Geometric interpretation

- similarity = spatial proximity (Euclidean dist.)
- location depends on frequency of noun ($f_{\text{dog}} \approx 2.7 \cdot f_{\text{cat}}$)



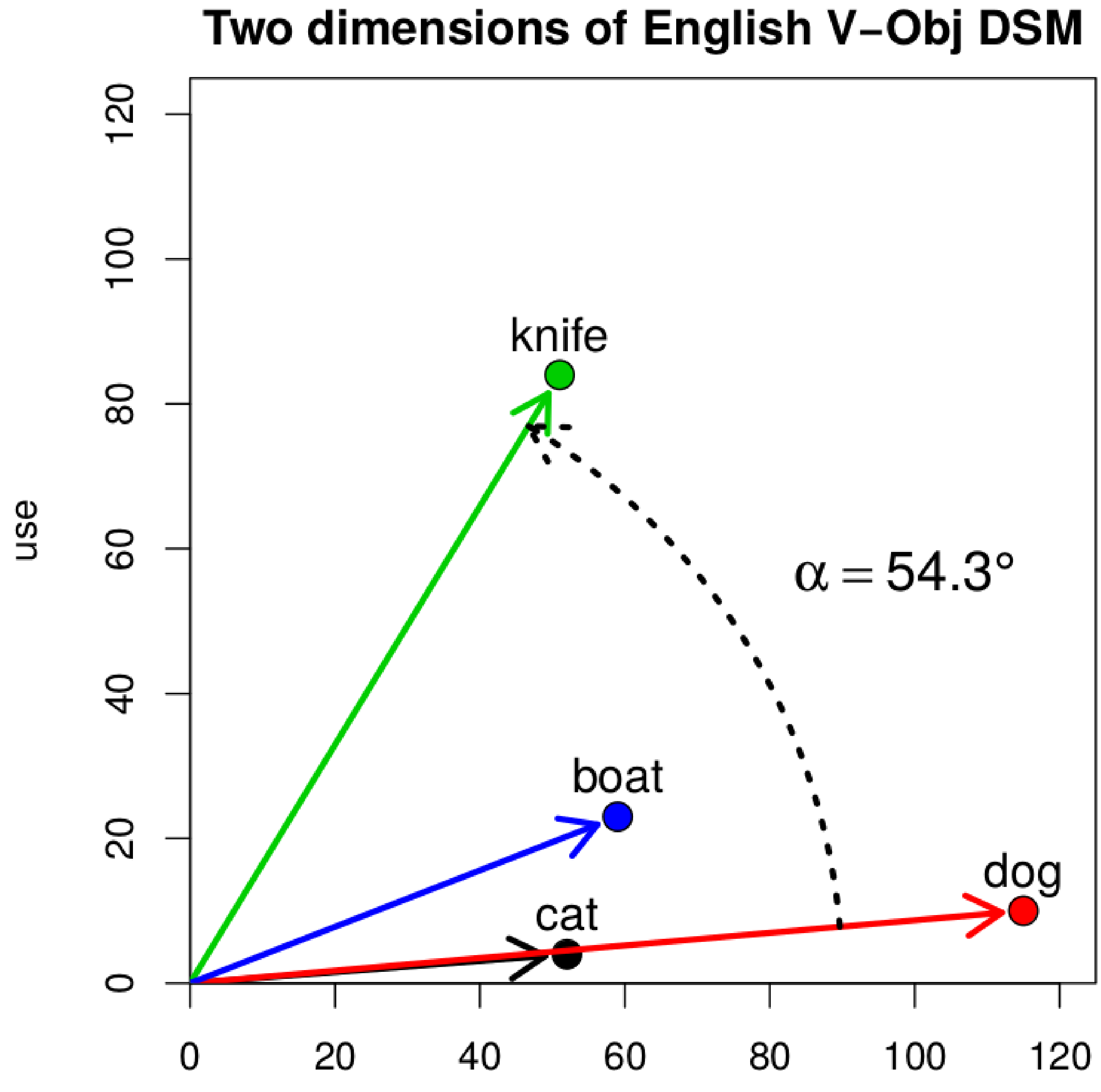
Geometric interpretation

- vector can also be understood as arrow from origin
- direction more important than location



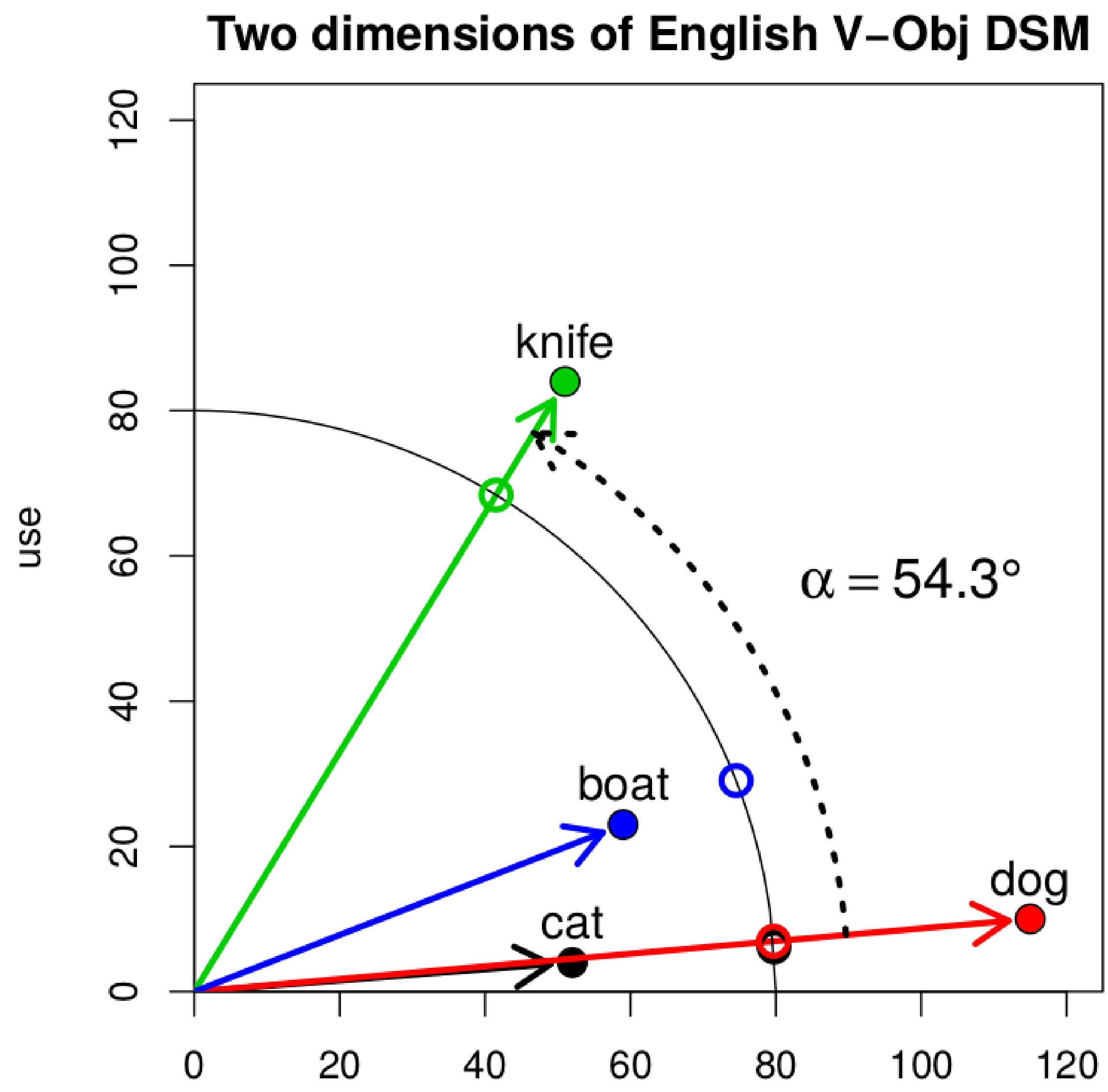
Geometric interpretation

- vector can also be understood as arrow from origin
- direction more important than location
- use angle α as distance measure



Geometric interpretation

- vector can also be understood as arrow from origin
- direction more important than location
- use angle α as distance measure
- or normalise length $\|\mathbf{x}_{\text{dog}}\|$ of arrow



General definition of DSMs

A **distributional semantic model** (DSM) is a scaled and/or transformed co-occurrence matrix $\mathbf{M}$, such that each row $\mathbf{x}$ represents the distribution of a target term across contexts.

	get	see	use	hear	eat	kill
knife	0.027	-0.024	0.206	-0.022	-0.044	-0.042
cat	0.031	0.143	-0.243	-0.015	-0.009	0.131
dog	-0.026	0.021	-0.212	0.064	0.013	0.014
boat	-0.022	0.009	-0.044	-0.040	-0.074	-0.042
cup	-0.014	-0.173	-0.249	-0.099	-0.119	-0.042
pig	-0.069	0.094	-0.158	0.000	0.094	0.265
banana	0.047	-0.139	-0.104	-0.022	0.267	-0.042

Term = word, lemma, phrase, morpheme, word pair, . . .

A thought experiment: deciphering hieroglyphs

(knife)	51	20	84	0	3	0
(cat)	52	58	4	4	6	26
???	115	83	10	42	33	17
(boat)	59	39	23	4	0	0
(cup)	98	14	6	2	1	0
(pig)	12	17	3	2	9	27
(banana)	11	2	2	0	18	0

A thought experiment: deciphering hieroglyphs

(knife)	**51**	**20**	**84**	**0**	**3**	**0**
(cat)	52	58	4	4	6	26
???	115	83	10	42	33	17
(boat)	59	39	23	4	0	0
(cup)	98	14	6	2	1	0
(pig)	12	17	3	2	9	27
(banana)	11	2	2	0	18	0

$$\mathrm{sim}(\;,\;) = 0.770$$

A thought experiment: deciphering hieroglyphs

(knife)	51	20	84	0	3	0
(cat)	52	58	4	4	6	26
???	115	83	10	42	33	17
(boat)	59	39	23	4	0	0
(cup)	98	14	6	2	1	0
(pig)	**12**	**17**	**3**	**2**	**9**	**27**
(banana)	11	2	2	0	18	0

sim(,) = 0.939

A thought experiment: deciphering hieroglyphs

(knife)	51	20	84	0	3	0
(cat)	**52**	**58**	**4**	**4**	**6**	**26**
???	115	83	10	42	33	17
(boat)	59	39	23	4	0	0
(cup)	98	14	6	2	1	0
(pig)	12	17	3	2	9	27
(banana)	11	2	2	0	18	0

$$\mathrm{sim}(\;,\;) = 0.961$$

English as seen by the computer . . .

	get	see	use	hear	eat	kill
knife	51	20	84	0	3	0
cat	52	58	4	4	6	26
dog	115	83	10	42	33	17
boat	59	39	23	4	0	0
cup	98	14	6	2	1	0
pig	12	17	3	2	9	27
banana	11	2	2	0	18	0

Nearest neighbours

DSM based on verb-object relations from BNC, reduced to 100 dim. with SVD

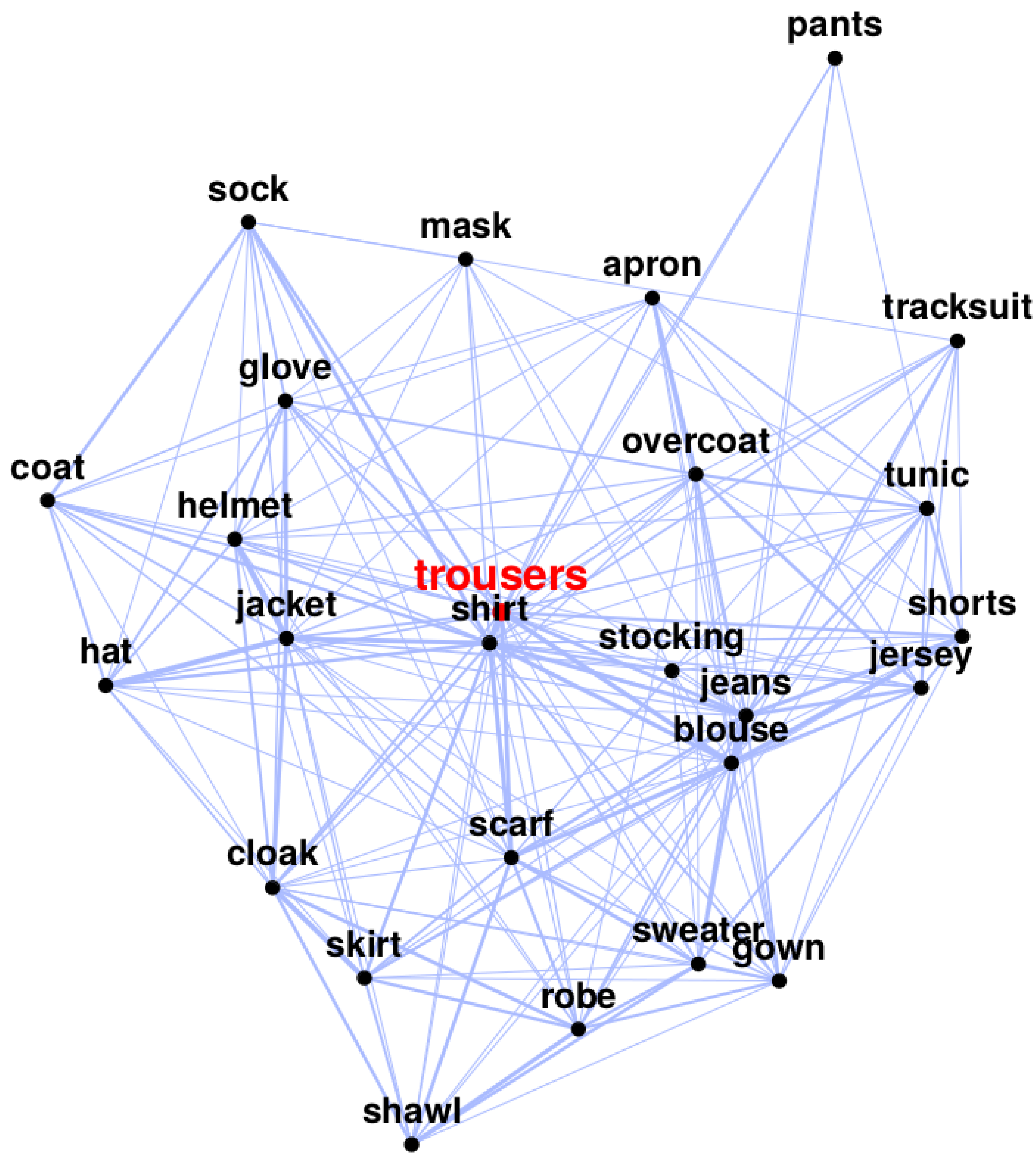
Neighbours of **trousers** (cosine angle):

☞ shirt (18.5), blouse (21.9), scarf (23.4), jeans (24.7), skirt (25.9), sock (26.2), shorts (26.3), jacket (27.8), glove (28.1), coat (28.8), cloak (28.9), hat (29.1), tunic (29.3), overcoat (29.4), pants (29.8), helmet (30.4), apron (30.5), robe (30.6), mask (30.8), tracksuit (31.0), jersey (31.6), shawl (31.6), . . .

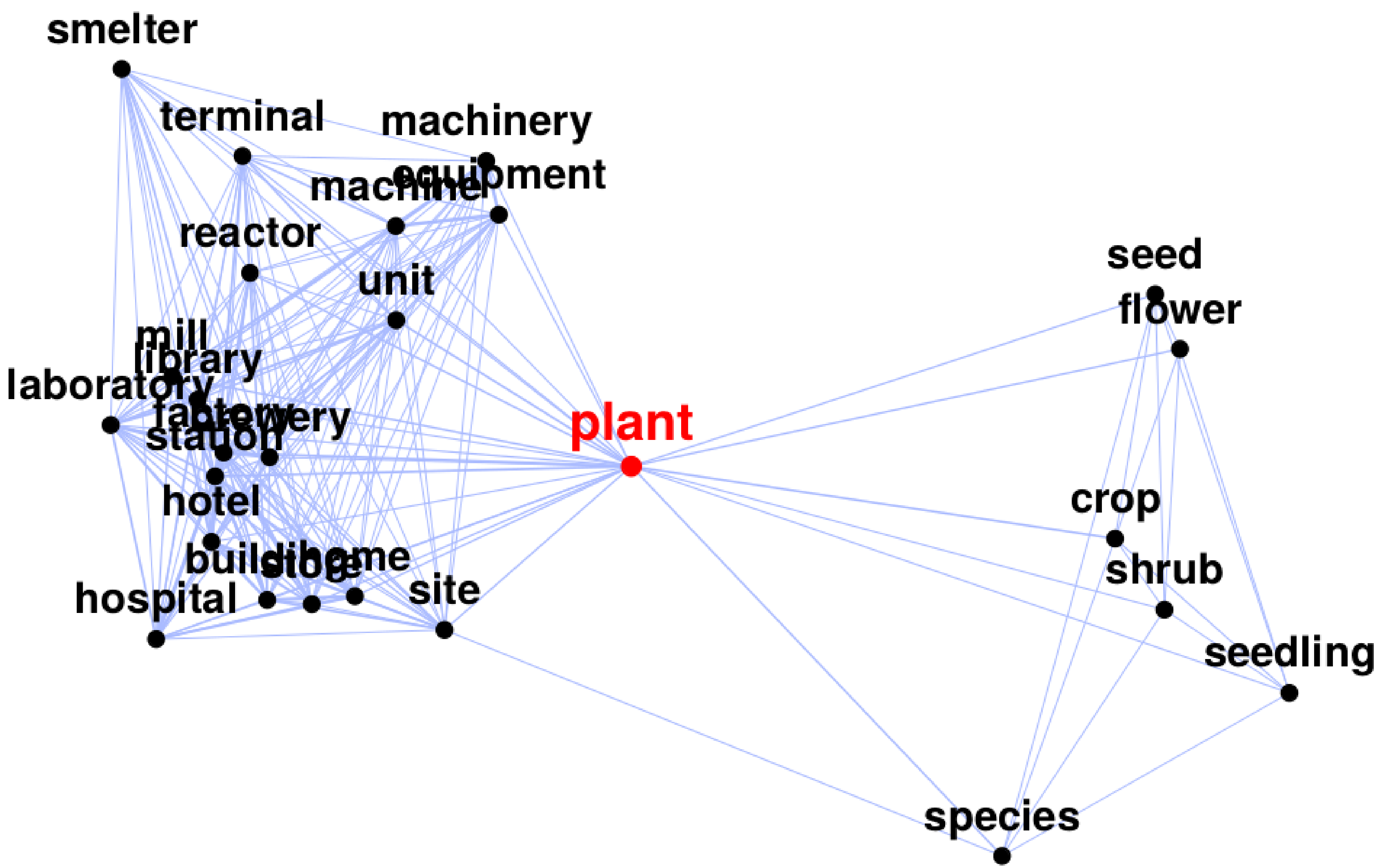
Neighbours of **rage** (cosine angle):

☞ anger (28.5), fury (32.5), sadness (37.0), disgust (37.4), emotion (39.0), jealousy (40.0), grief (40.4), irritation (40.7), revulsion (40.7), scorn (40.7), panic (40.8), bitterness (41.6), resentment (41.8), indignation (41.9), excitement (42.0), hatred (42.5), envy (42.8), disappointment (42.9), . . .

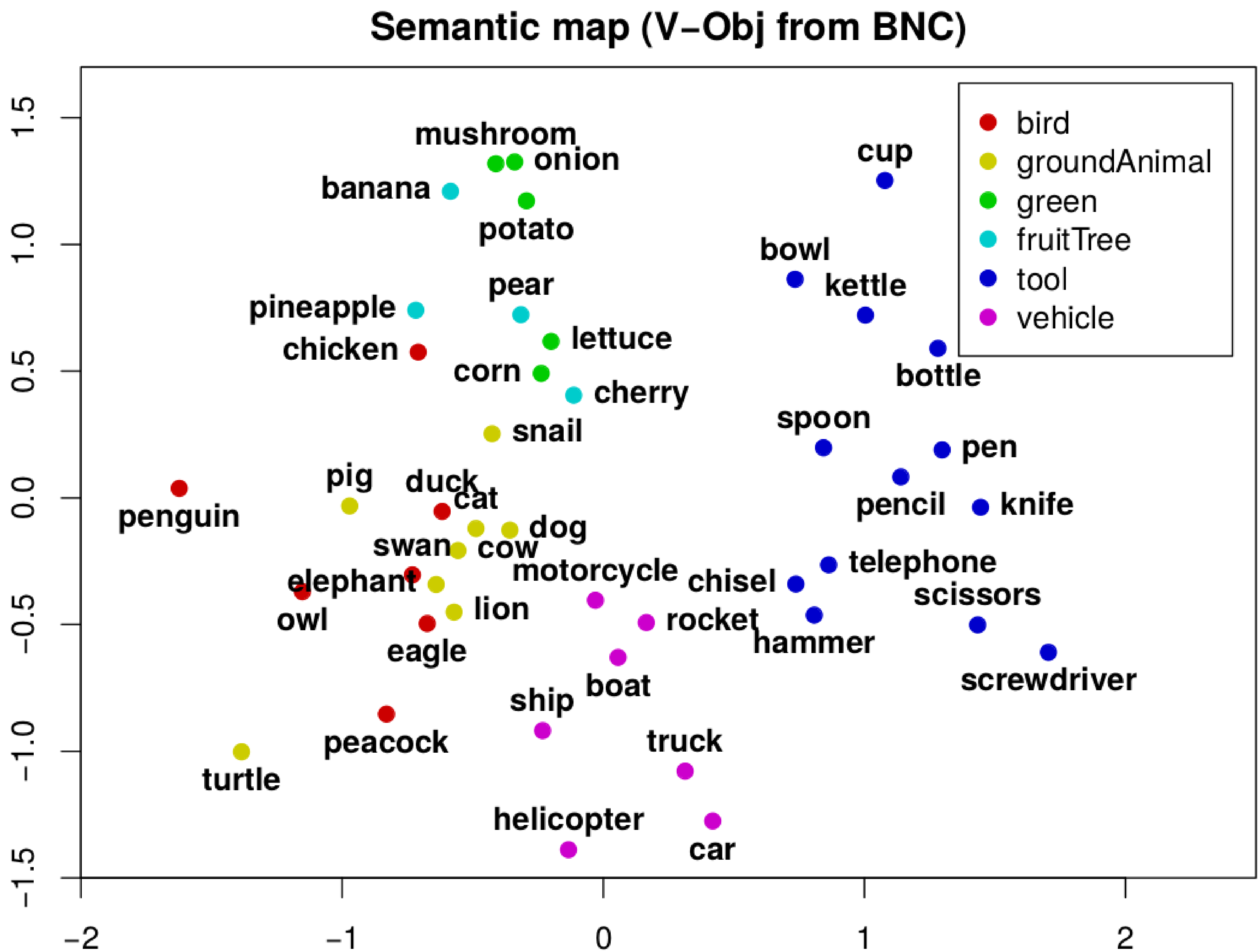
Nearest neighbours with similarity graph



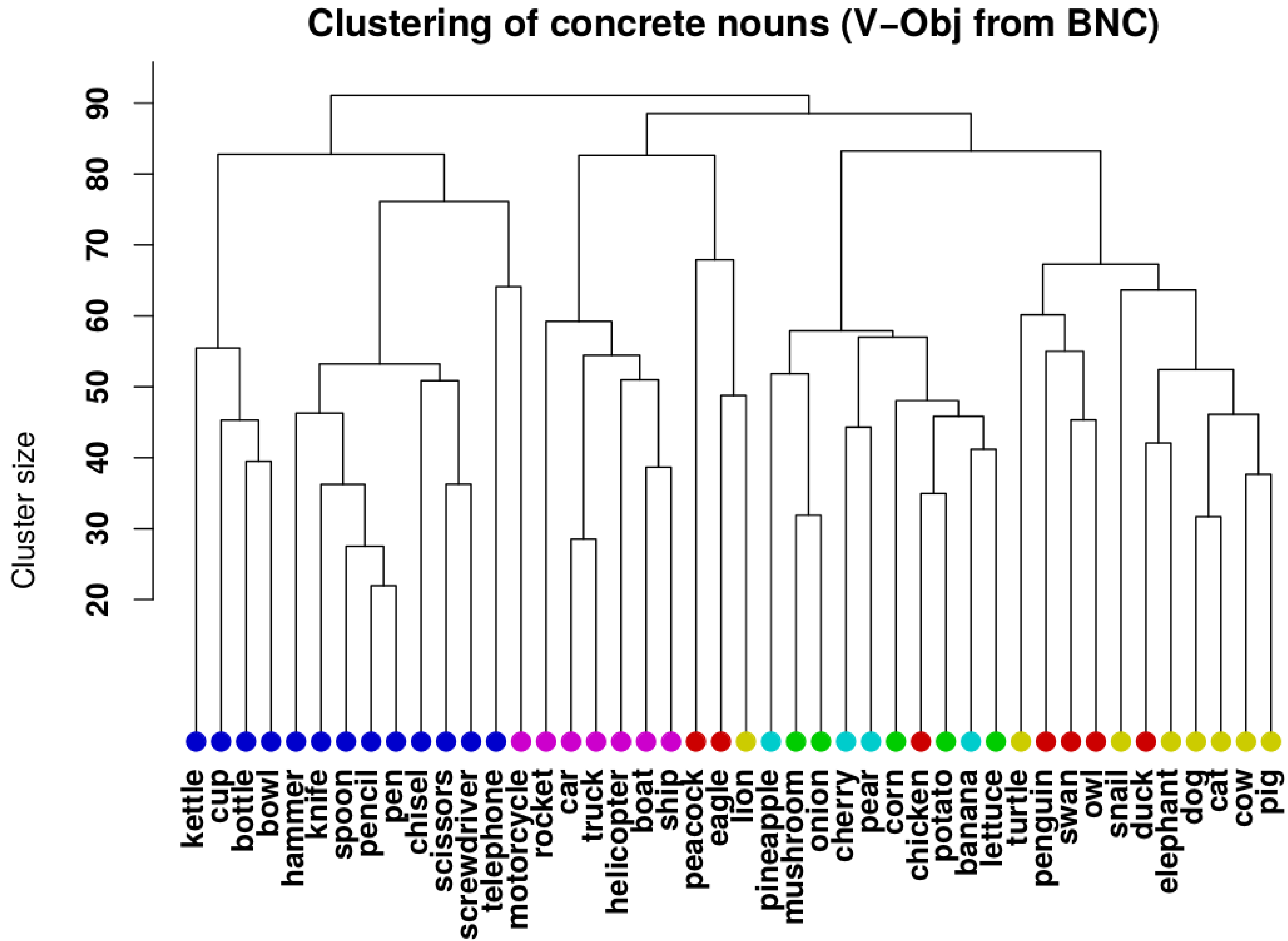
Nearest neighbours with similarity graph



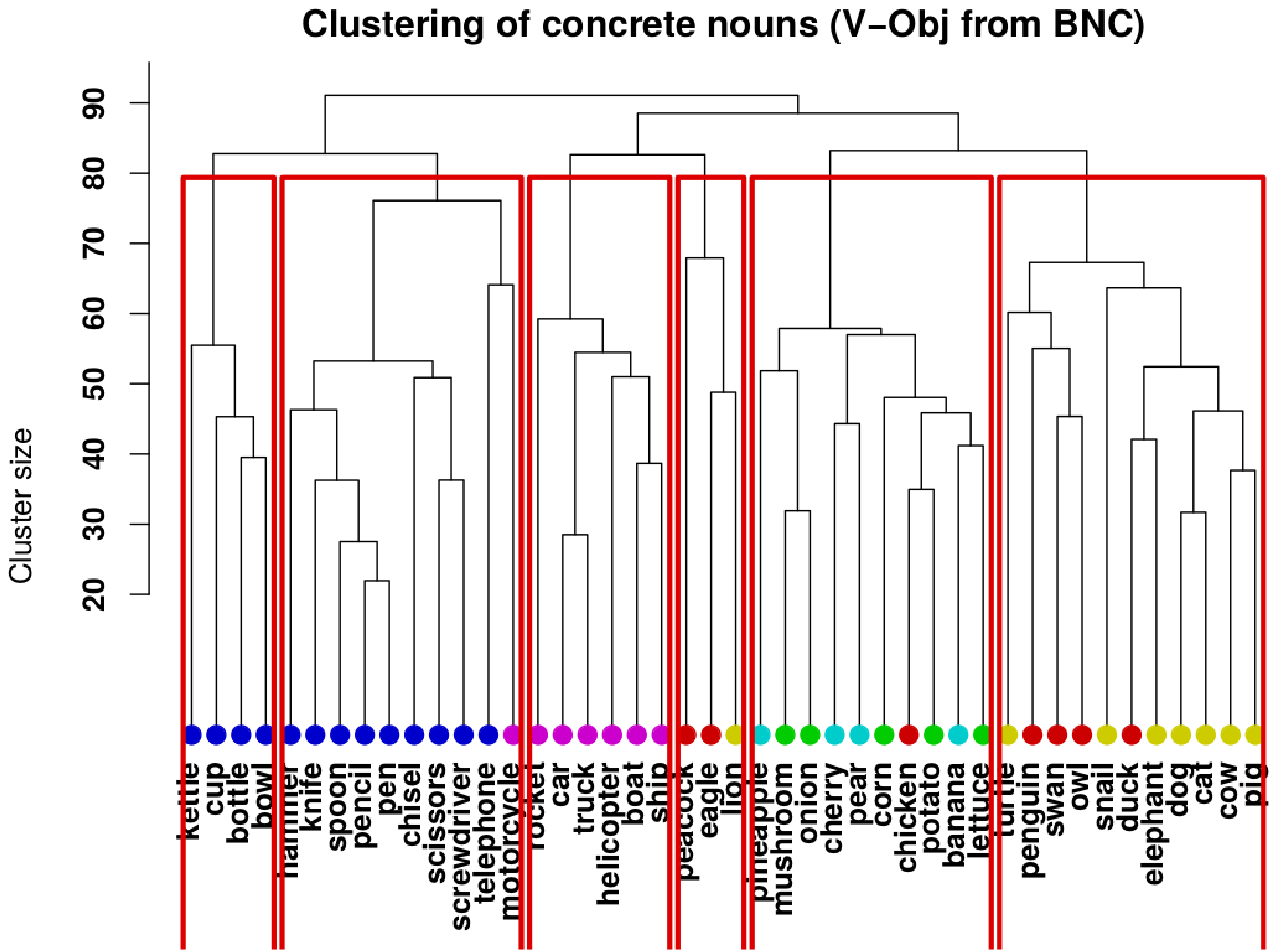
Semantic maps



Clustering


Clustering of concrete nouns (V−Obj from BNC)
Cluster size
90
80
70
60
50
40
30
20
kettle
cup
bottle
bowl
hammer
knife
spoon
pencil
pen
chisel
scissors
screwdriver
telephone
motorcycle
rocket
car
truck
helicopter
boat
ship
peacock
eagle
lion
pineapple
mushroom
onion
cherry
pear
corn
chicken
potato
banana
lettuce
turtle
penguin
swan
owl
snail
duck
elephant
dog
cat
cow
pig

Clustering



Clustering of concrete nouns (V−Obj from BNC)
Cluster size
20
30
40
50
60
70
80
90
kettle
cup
bottle
bowl
hammer
knife
spoon
pencil
pen
chisel
scissors
screwdriver
telephone
motorcycle
rocket
car
truck
helicopter
boat
ship
peacock
eagle
lion
pineapple
mushroom
onion
cherry
pear
corn
chicken
potato
banana
lettuce
turtle
penguin
swan
owl
snail
duck
elephant
dog
cat
cow
pig

Further information

- Handouts & other materials available from wordspace wiki at

  http://wordspace.collocations.de/

  ☞ based on joint work with Marco Baroni and Alessandro Lenci

- Tutorial is open source (CC), and can be downloaded from

  http://r-forge.r-project.org/projects/wordspace/

- Review paper on distributional semantics:

  Turney, Peter D. and Pantel, Patrick (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, **37**, 141–188.

Assignment 1.1

a) Given an English corpus construct a term-term co-occurrence matrix

b) Compute and display top 10 similar words in the corpus

(The corpus will be provided to you. You only need to construct the cooccurrence matrix)

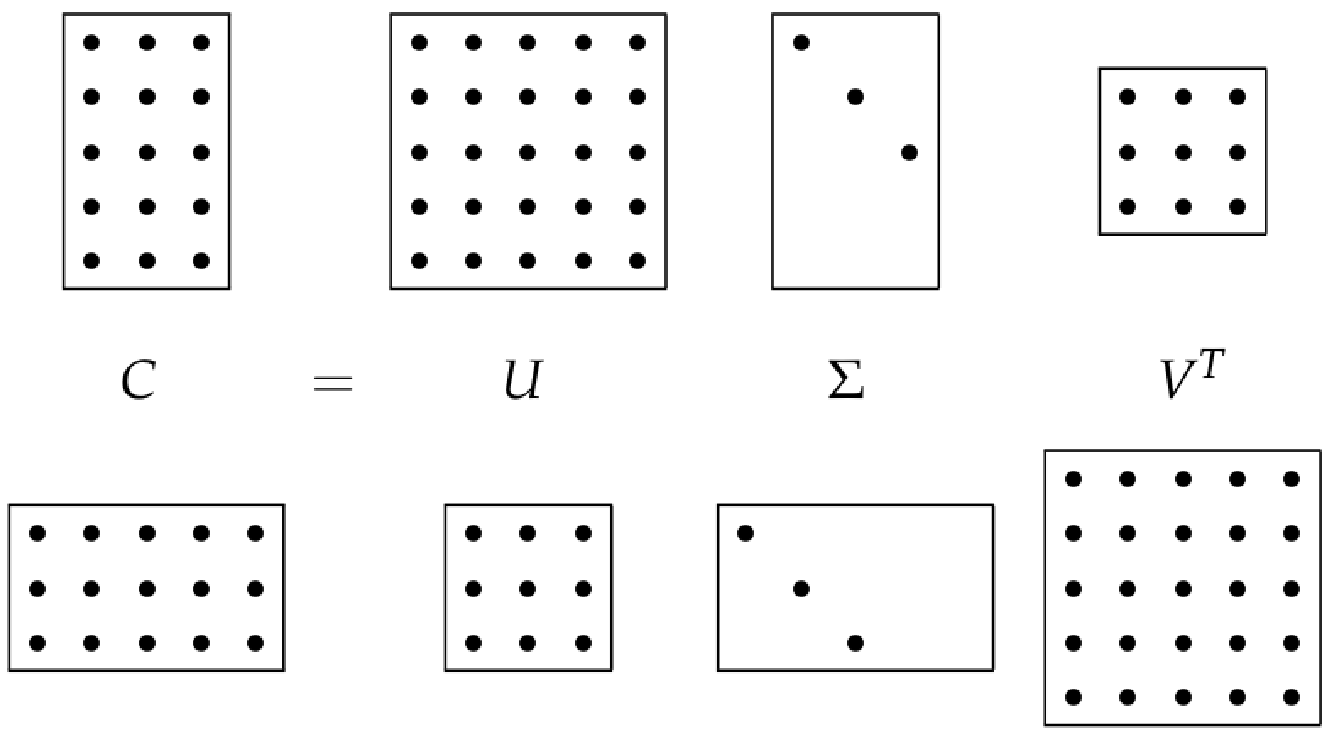
Real world scenerio: Sparce matrix

	i	love	playing	cricket	sachin	is	a	cricketer		played	using	bat	and	ball	sourav	plays	sachine	the	highest	test	scorer	maradona	football
i	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
love	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
playing	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
cricket	1	1	1	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	
sachin	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
is	0	0	0	1	1	0	2	1	1	1	0	0	0	0	0	1	1	1	1	0	0	0	
a	0	0	0	0	1	2	2	1	1	1	2	2	1	0	0	0	0	0	0	0	0	0	
cricketer		0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
played	0	0	0	1	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	
using	0	0	0	1	0	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	
bat	0	0	0	0	0	0	2	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	
and	0	0	0	0	0	0	2	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	
ball	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	
sourav	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
plays	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	
sachine	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	
the	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	
highest	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	
test	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	
scorer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	
maradona		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1
football		0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0

SVD: Intuition

- Singular Value Decomposition

$$C = U\Sigma V^T$$



Semantics

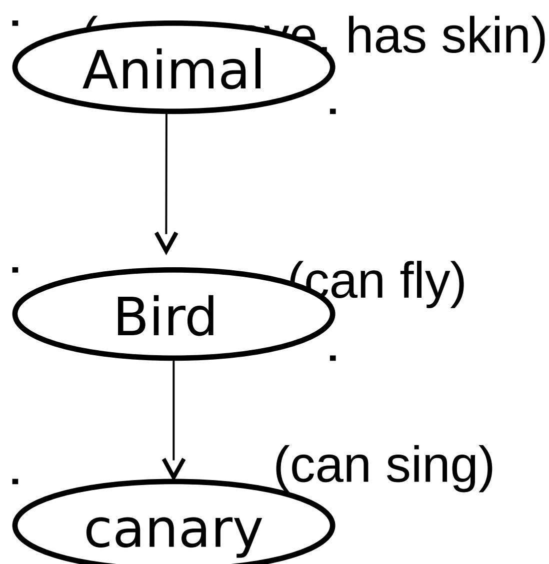
Wordnet

Wordnet

- **A lexical knowledgebase based on conceptual lookup**
    - Organizing concepts in a semantic network.
- **Organize lexical information in terms of word meaning, rather than word form**
    - Wordnet can also be used as a thesaurus.

Psycholinguistic Theory

- Human lexical memory for nouns as a hierarchy.
- *Can canary sing? - Pretty fast response.*
- *Can canary fly? - Slower response.*
- *Does canary have skin? – Slowest response.*

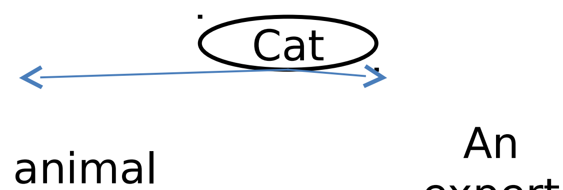


Wordnet - a lexical reference system based on psycholinguistic theories of human lexical memory.

Relational and Componential Semantics

Relational Semantics (Words can disambiguate each other) **vs.** Componential Semantics (Words need features for disambiguation)

Example



Possible Features: Animate, Human, Carnivorous, Small, Moving

<u>Componential Semantics</u>

Semantic Feature Vector for

cat (animal): <1,0,1,1,1>

cat (expert): <1,1,U,U,1>

<u>Relational Semantics</u>

cat (animal): {cat, feline}

cat (expert): {cat, expert}

Componential Semantics

- Consider *cat* and *tiger.* Decide on *componential attributes.*

Furry	Carnivorous	Heavy	Domesticable

- For *cat* (Y, Y, N, Y)
- For *tiger* (Y,Y,Y,N)

  Complete and correct Attributes are difficult to design.

Semantic relations in wordnet

1. Synonymy
2. Hypernymy / Hyponymy
3. Antonymy
4. Meronymy / Holonymy.