FEATURE SPACES

Alexander Huth
@ UBA ECI Course
2017-07-27

REMINDER

* Download slides from

  http://github.com/alexhuth/eci-2017

AGENDA

* Feature spaces & linearized models
  * Language
  * Vision
* Handling stimulus timeseries

SYSTEM IDENTIFICATION

$$Y = f(X)$$

* What kind of a function is f?

SYSTEM IDENTIFICATION

READ THIS PAPER:

Complete Functional Characterization of Sensory Neurons by System Identification

Michael C.-K. Wu,[1] Stephen V. David,[2] and Jack L. Gallant[3,4]

http://dx.doi.org/10.1146/annurev.neuro.29.051605.113024

SYSTEM IDENTIFICATION

* Linear model

$$Y = X\beta$$

* Linearized model

$$Y = \mathbb{L}(X)\beta$$

* Nonlinear model

$$Y = \Theta(X)$$

LINEAR MODELS

$$Y = X\beta$$

image pixels

X1, Y=0.7

X2, Y=0.3

X3, Y=0.0

LINEAR MODELS

$$Y = X\beta$$

image pixels

X

14	100	120	121
12	58	103	107
8	32	78	99
10	14	62	102
3	32	56	81

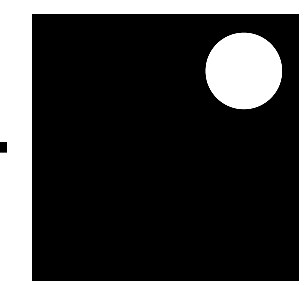
… ⋮

unravel

Y = X • B

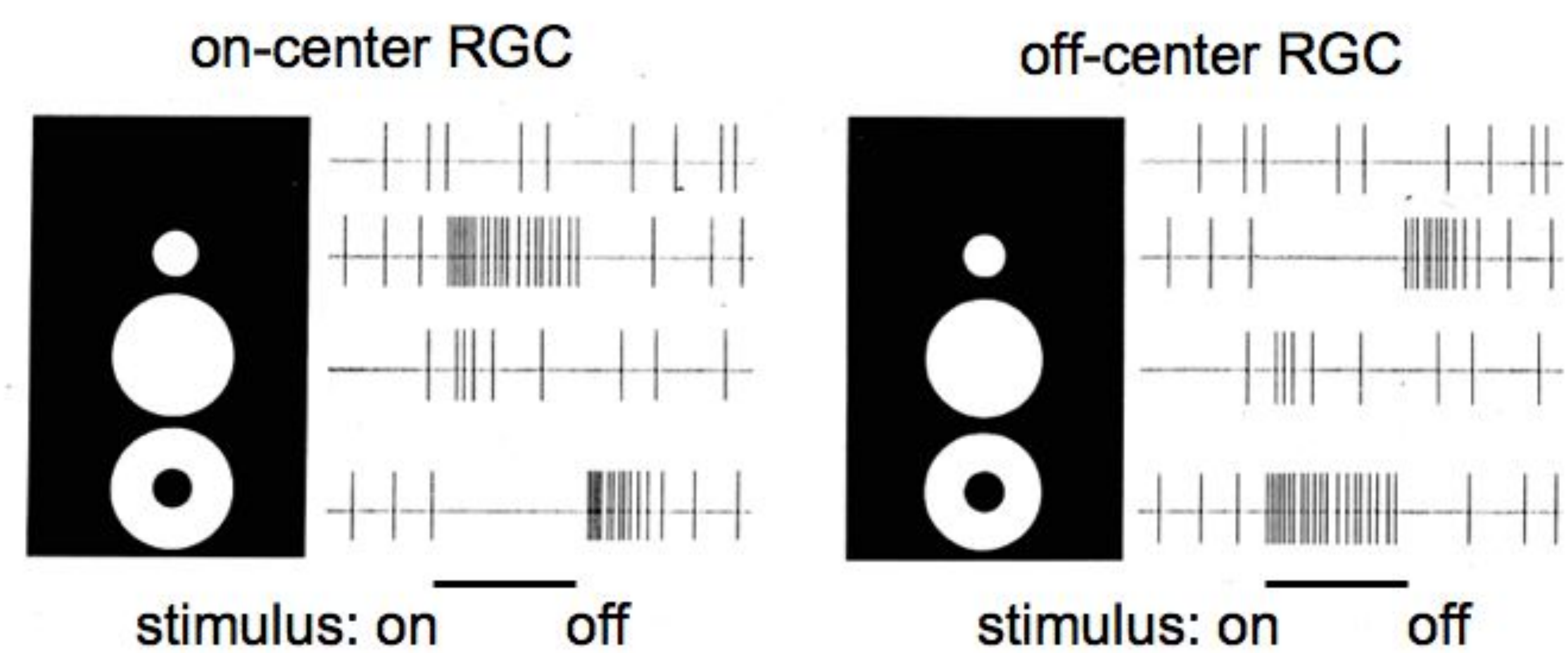
X	B
14	1
12	1
8	1
10	1
3	1
100	0
58	0
32	0
14	0
32	0
120	0
103	0
78	0
62	0
56	0
121	0
107	0
99	0
102	0
81	0
⋮	⋮

filter

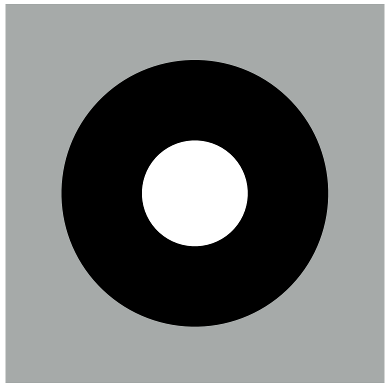


LINEAR MODELS

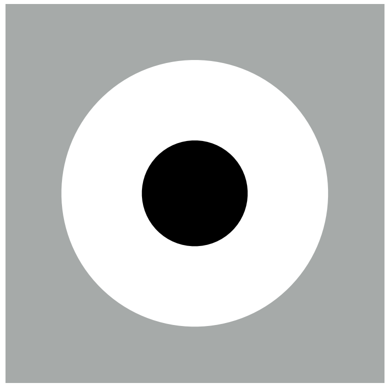
Retinal ganglion cell responses



Beta
(on-center)



Beta
(off-center)



http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/ganglion/ganglion.html

LINEARIZED MODELS

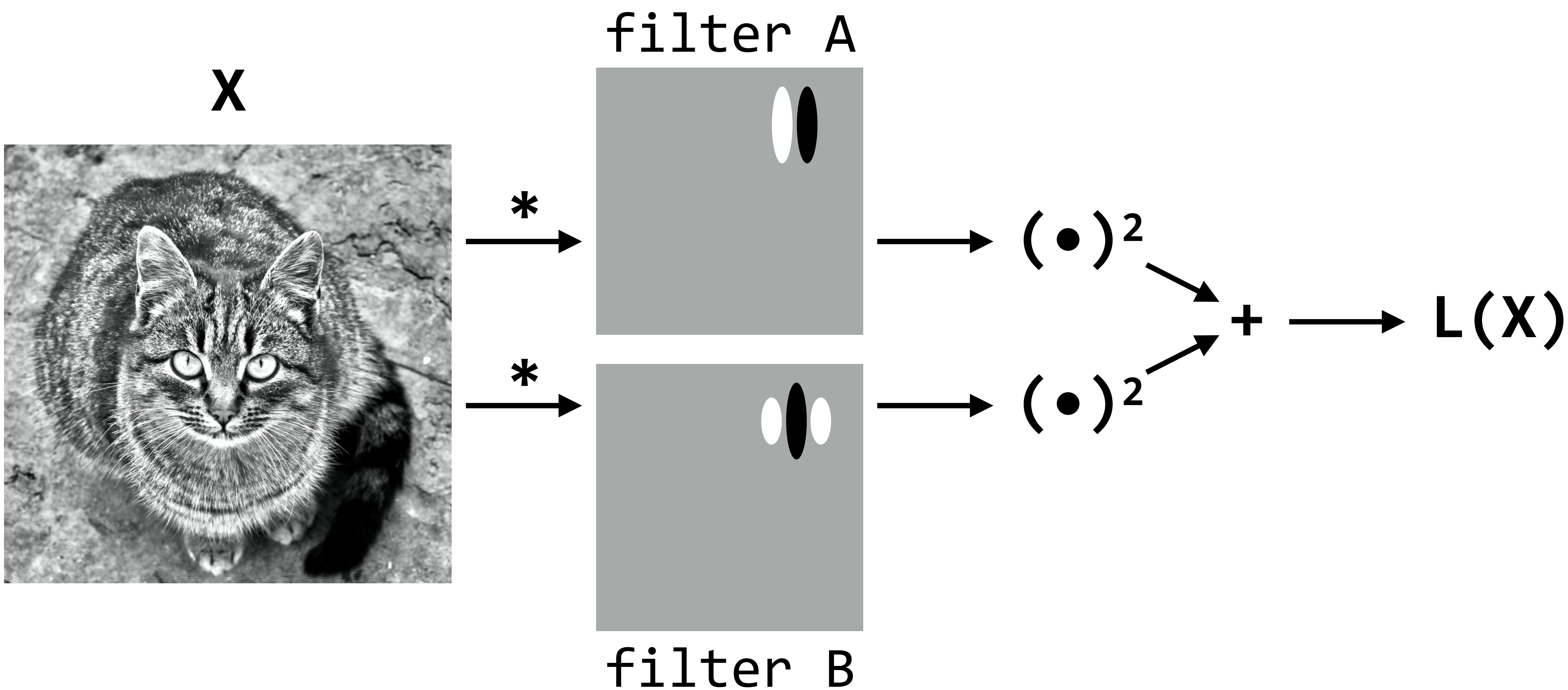
$$Y = \mathbb{L}(X)\beta$$

* **L** is some non-linear function of the stimulus ***X*** that gives us *features*

* ***Beta*** is a linear weighting of the *features* that gives us the response ***Y***

LINEARIZED MODELS

$$Y = \mathbb{L}(X)\beta$$



LINEARIZED MODELS

$$Y = \mathbb{L}(X)\beta$$

story

word matrix

time

word				
	now	1	0	0
	this	1	0	0
	is	0	1	0	…
	a	0	1	0
	story	0	0	1

⋮

NONLINEAR MODELS

$$Y = \Theta(X)$$

X1, Y="cat"

X2, Y="dog"

X3, Y="mate"

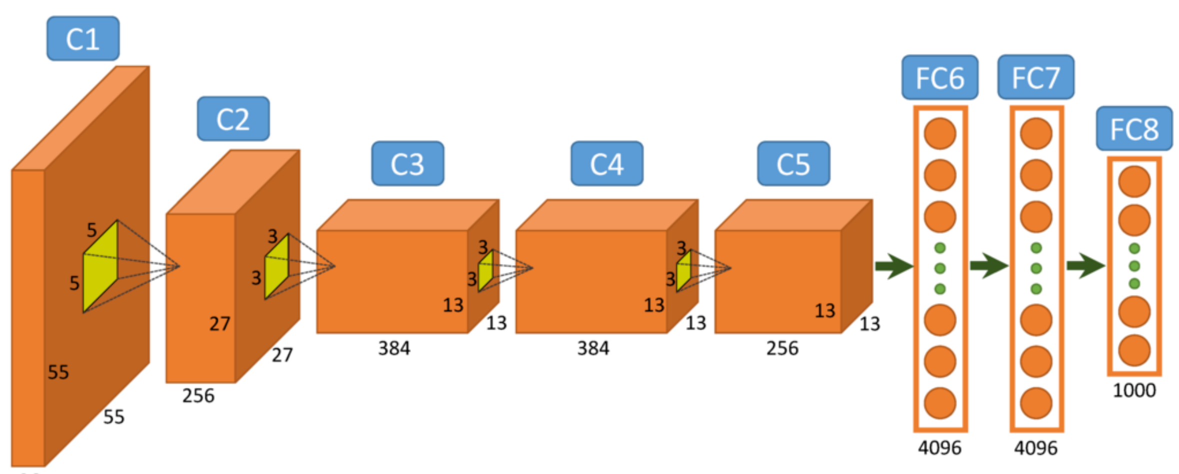
NONLINEAR MODELS

$$Y = \Theta(X)$$

image pixels

X **Theta** **Y**

(deep neural network)



→ “cat”

https://www.saagie.com/blog/object-detection-part1

SYSTEM IDENTIFICATION

* **Linear model**
  * cheap, pointless
* **Linearized model**
  * sweet spot, but requires **hypothesis**!
* **Nonlinear model**
  * wildly expensive, difficult

LINEARIZING TRANSFORMATION

=

FEATURE SPACE

=

HYPOTHESIS

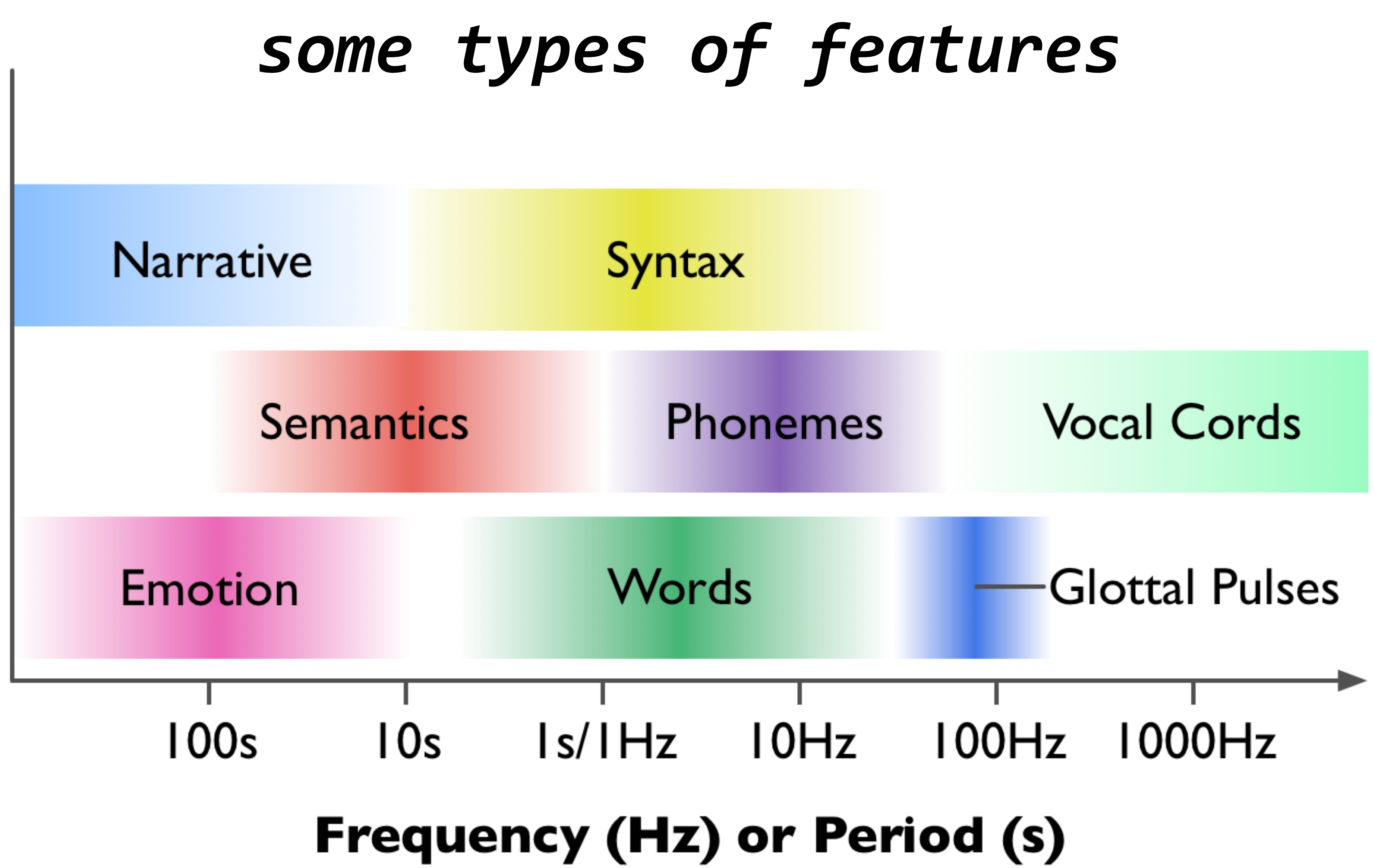
LINEARIZED MODELS

LANGUAGE

"Now this is a story all about how my life got flipped-turned upside down…"

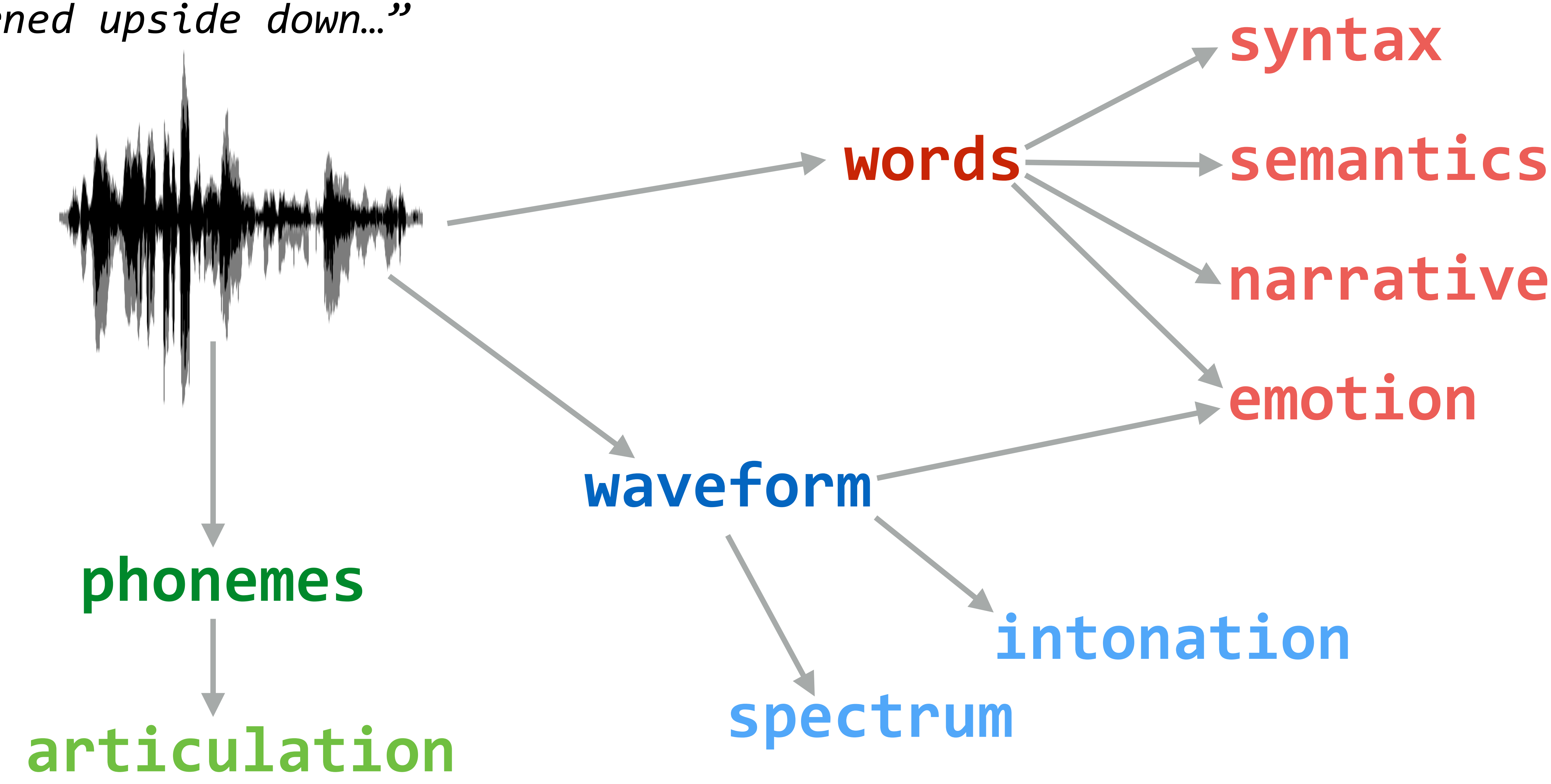
what kinds of features?

LANGUAGE


some types of features
Narrative
Syntax
Semantics
Phonemes
Vocal Cords
Emotion
Words
Glottal Pulses
100s
10s
1s/1Hz
10Hz
100Hz
1000Hz
Frequency (Hz) or Period (s)

LANGUAGE

“Now this is a story all about how my life got flipped-turned upside down…”



SYNTAX -
PART OF SPEECH

"Now this is a story all about how my life

{adv} {pn} {v} {dt} {n} {adj} {prep} {adv} {pn} {n}

got flipped-turned upside down…"

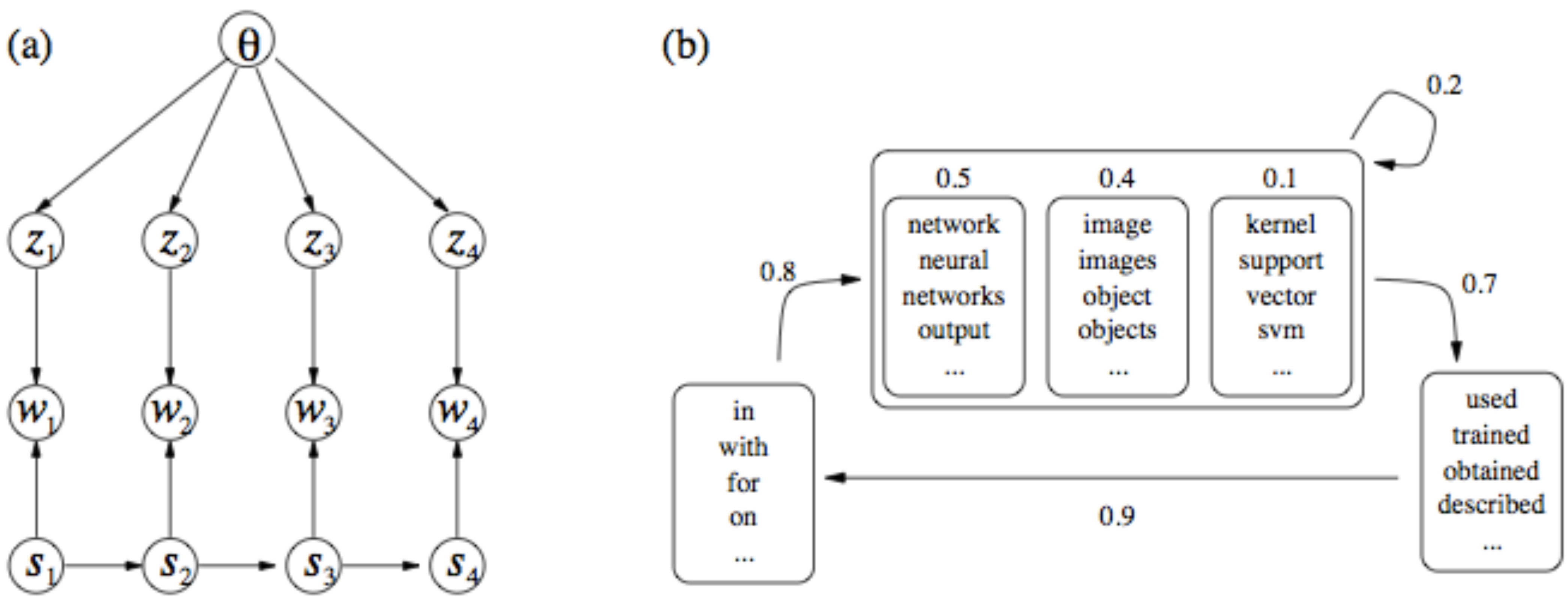
{v-p} {v-p} {v-p} {prep phrase}

time

parts of speech

adv	1	0	0	0	0	0	1	0	0	…
pn	0	1	0	0	0	0	0	1	0	…
v	0	0	1	0	0	0	0	0	0	…
dt	0	0	0	1	0	0	0	0	0	…
n	0	0	0	0	1	0	0	0	1	…
⋮										

SYNTAX - HMM+LDA

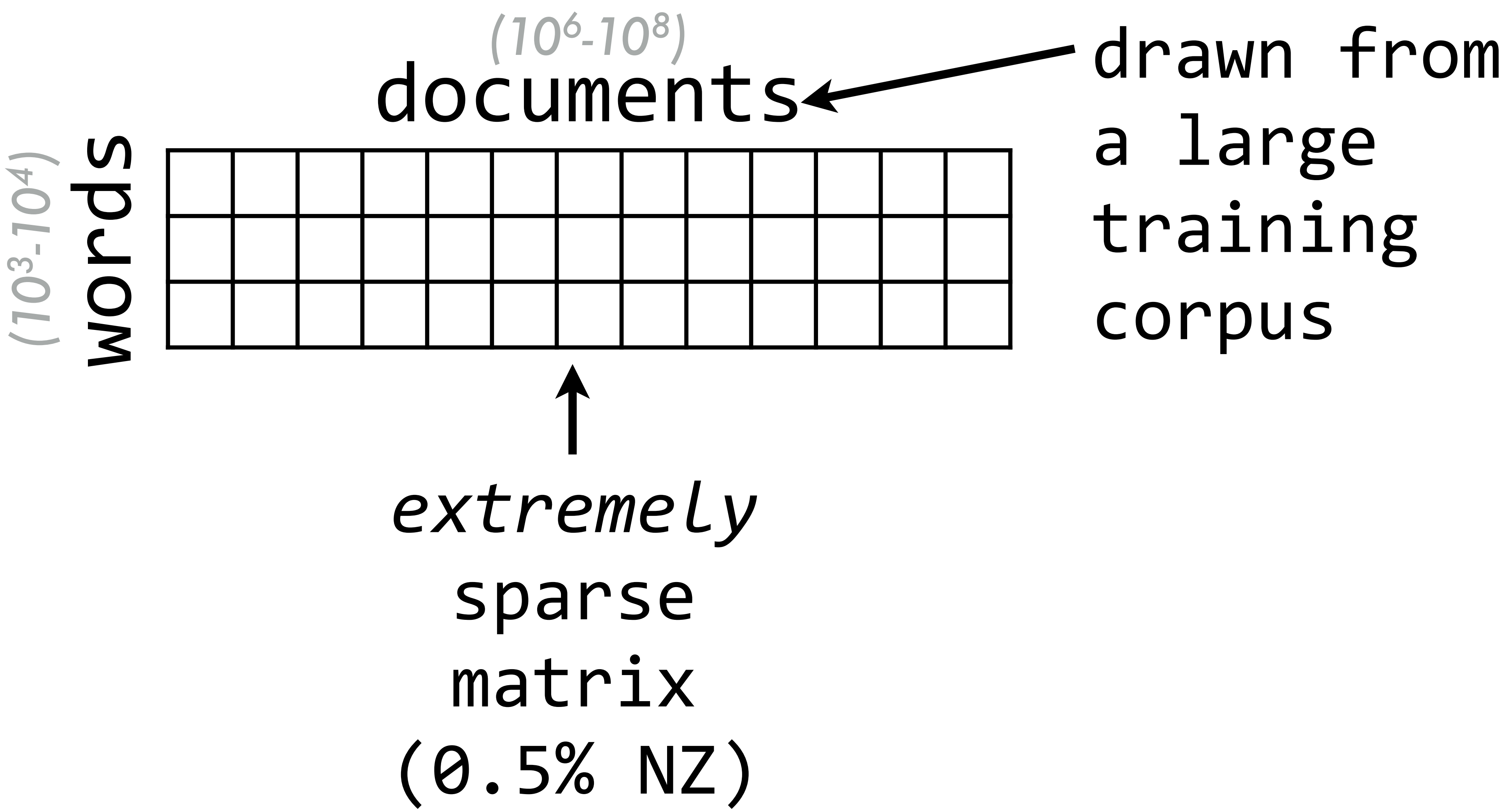


Griffiths et al, NIPS 2005

PAUSA DE 10 MINUTOS

SEMANTICS - LSA

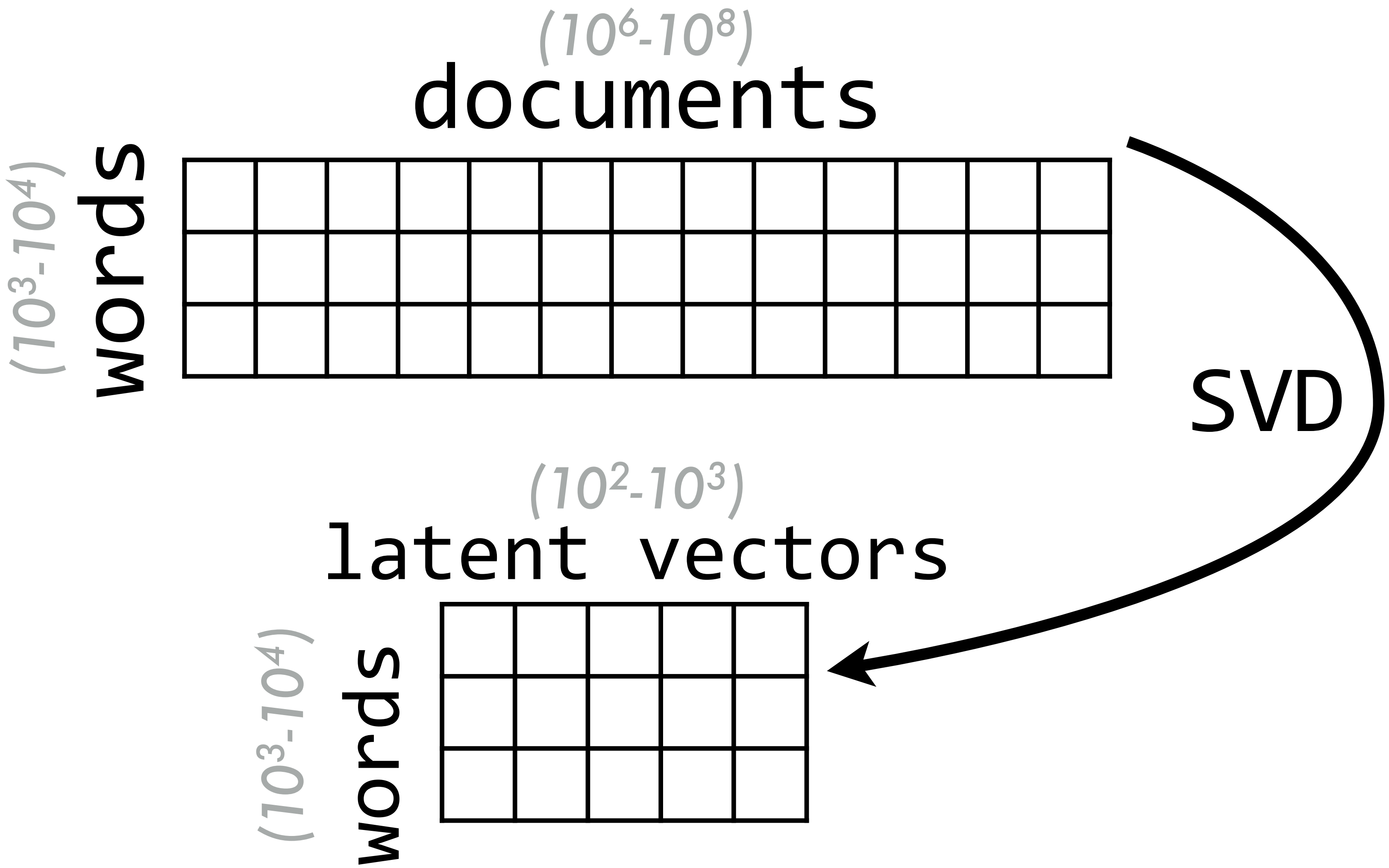
* Latent Semantic Analysis (LSA)



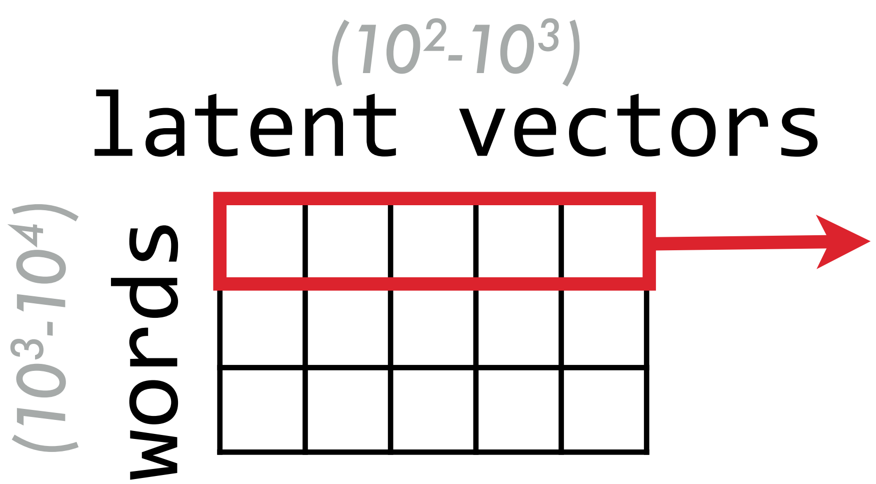
Deerwester et al. (1988)

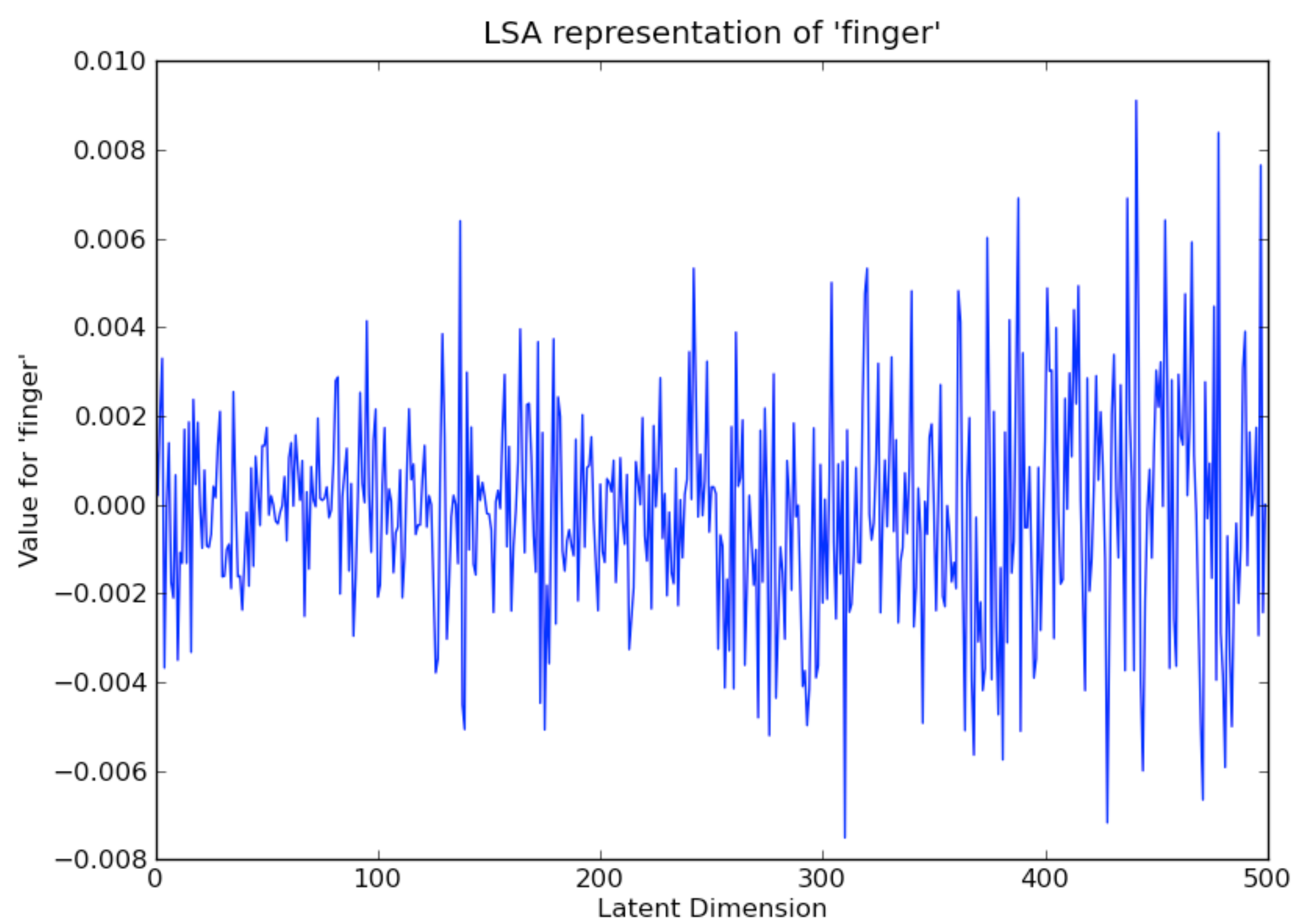
SEMANTICS - LSA

* Latent Semantic Analysis (LSA)

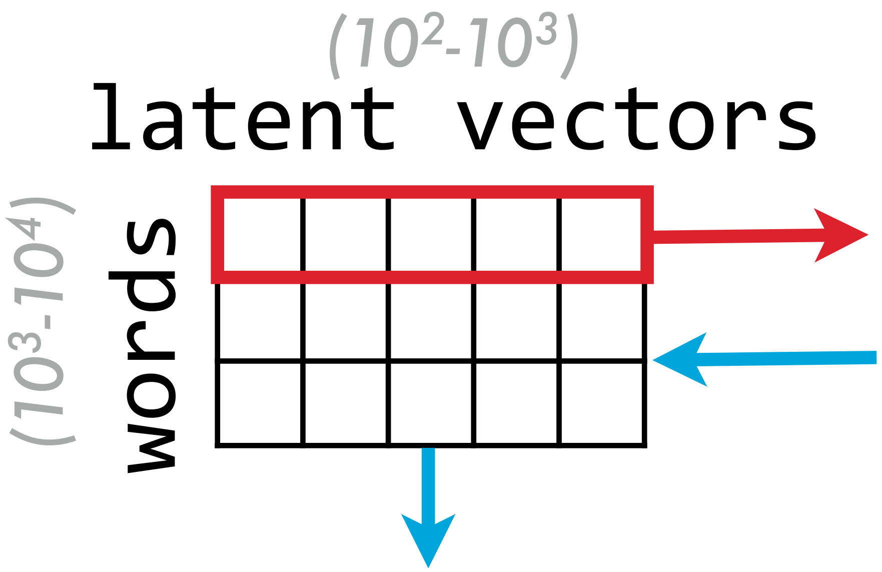


SEMANTICS - LSA

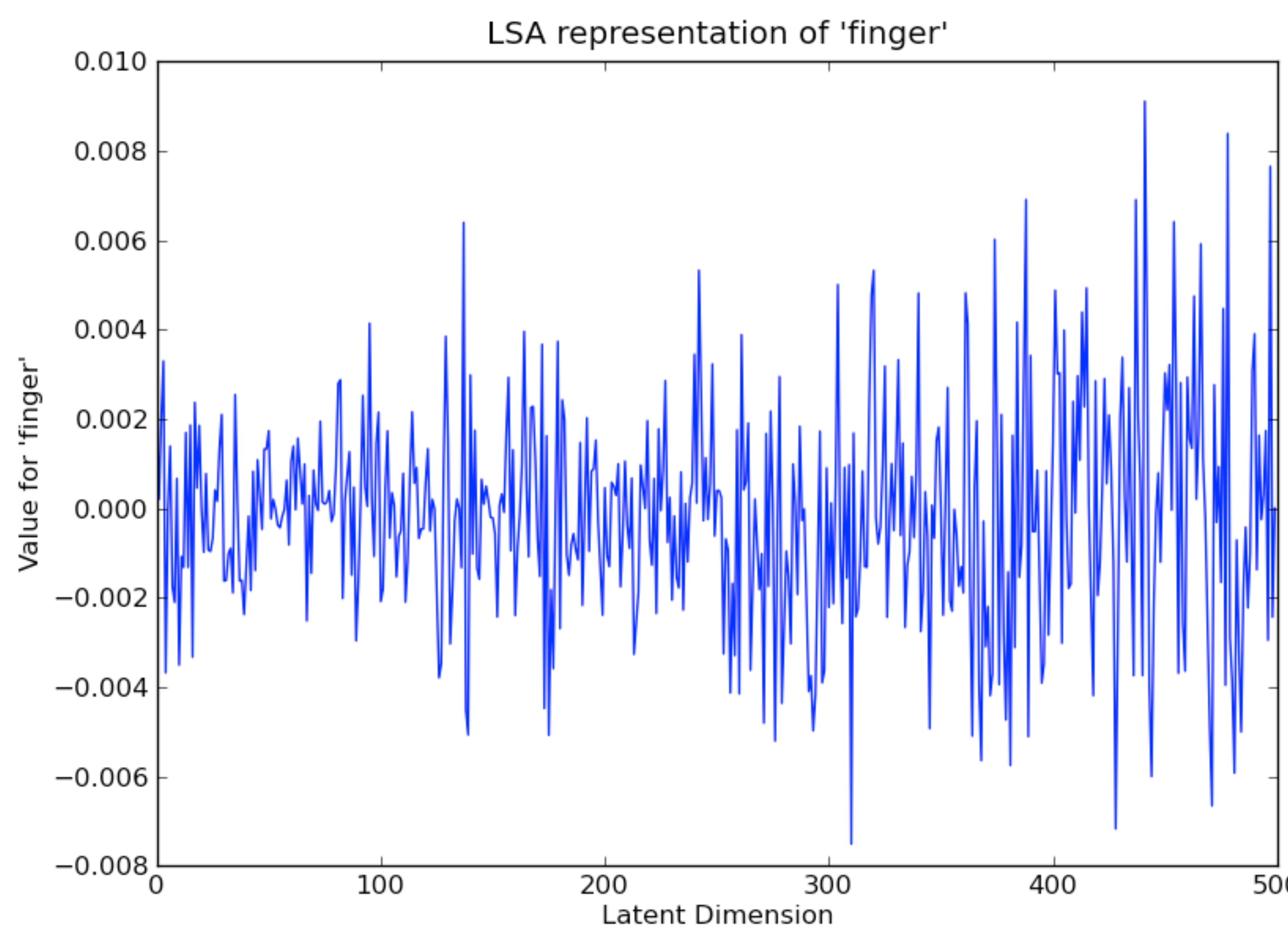




SEMANTICS - LSA

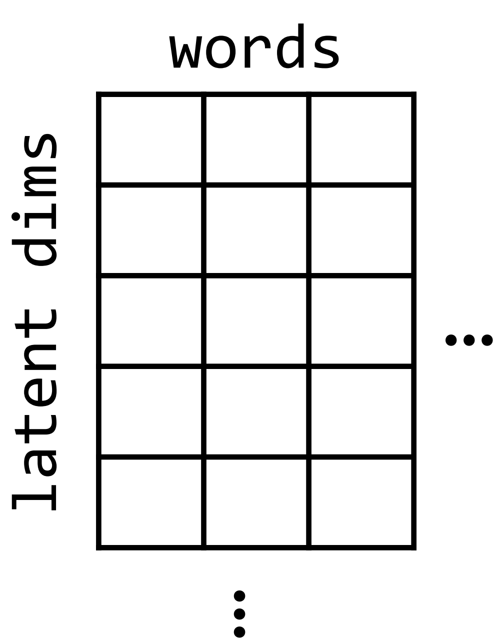


0.61- thumb
0.52- fingers
0.51- forefinger
0.49- tip
0.47- wrist
0.41- elbow
0.41- throat
0.41- lips
0.40- string

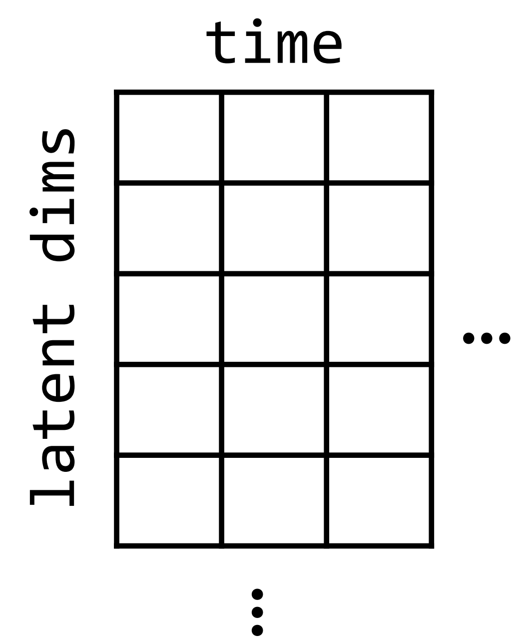


SEMANTICS - LSA

embedding matrix * word matrix = semantic stimulus matrix



words	time		
now	1	0	0
this	1	0	0
is	0	1	0
a	0	1	0
story	0	0	1
⋮			…



REMINDER FROM YESTERDAY...

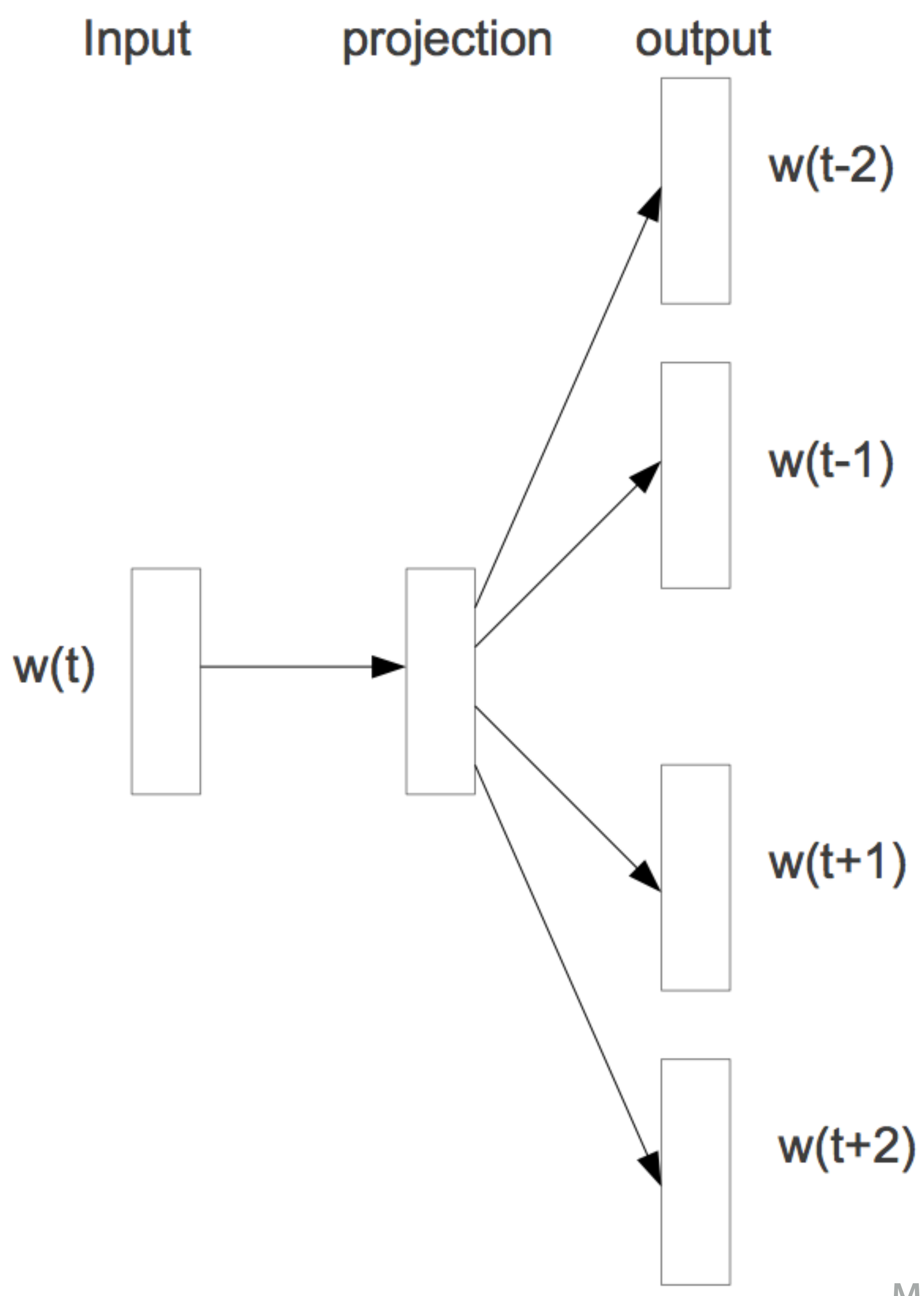
TIKHONOV REGRESSION

* this is equivalent to **TIKHONOV REGRESSION** on the **WORDS** with a prior determined by the **WORD EMBEDDING**

$$\underbrace{\frac{1}{\sigma^2}\Sigma_\beta}_{\text{PRIOR COVARIANCE}} = \underbrace{(C^T C)^{-1}}_{\text{INVERSE OF PENALTY INNER PRODUCT}} = \underbrace{E^T E}_{\text{EMBEDDING INNER PRODUCT}}$$

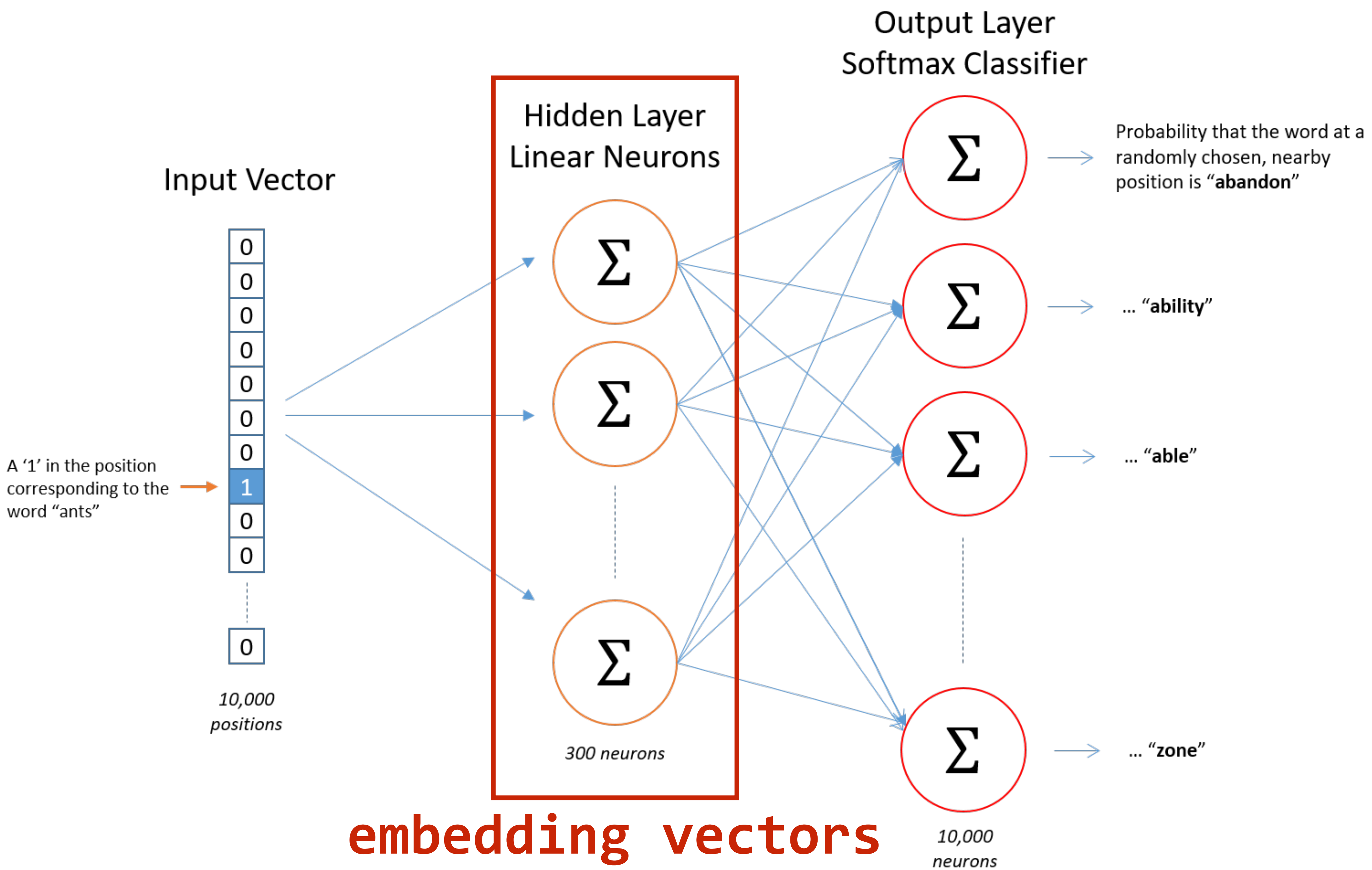
* **i.e. the prior covariance between two words' weights is equal to the dot product of their embedding vectors**

SEMANTICS - WORD2VEC



Mikolov et al. (2013)

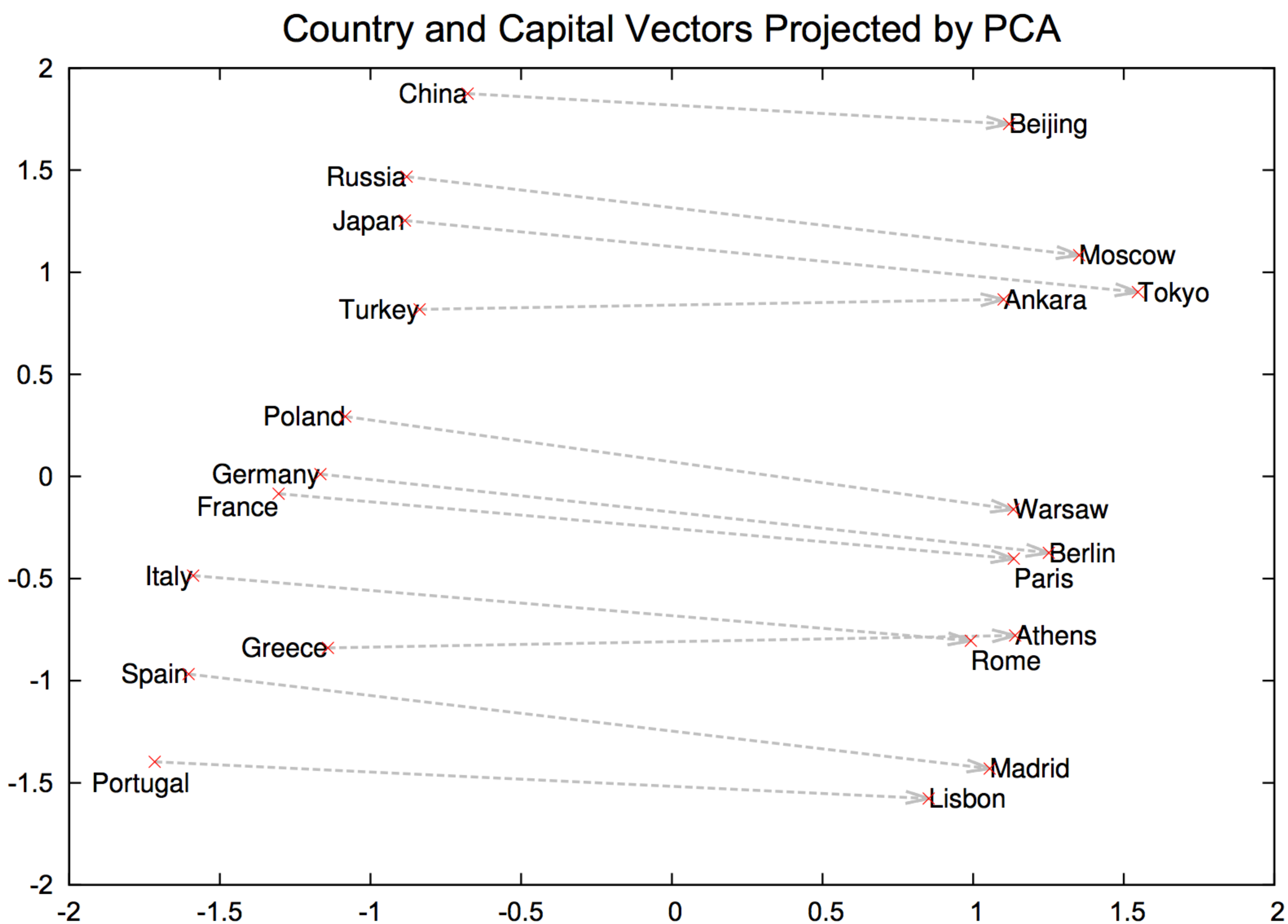
SEMANTICS - WORD2VEC



embedding vectors

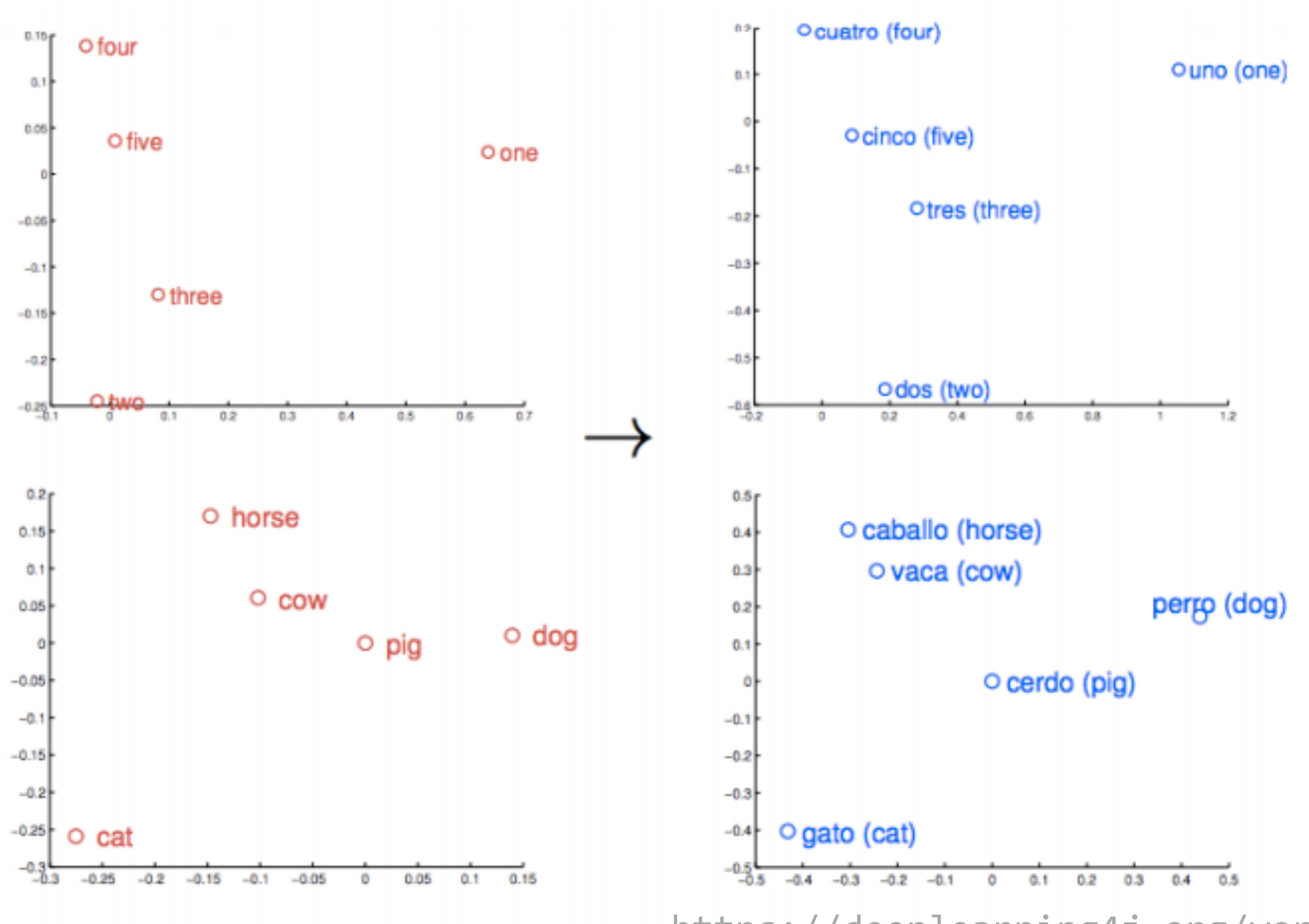
http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

SEMANTICS - WORD2VEC



Mikolov et al. (2013)

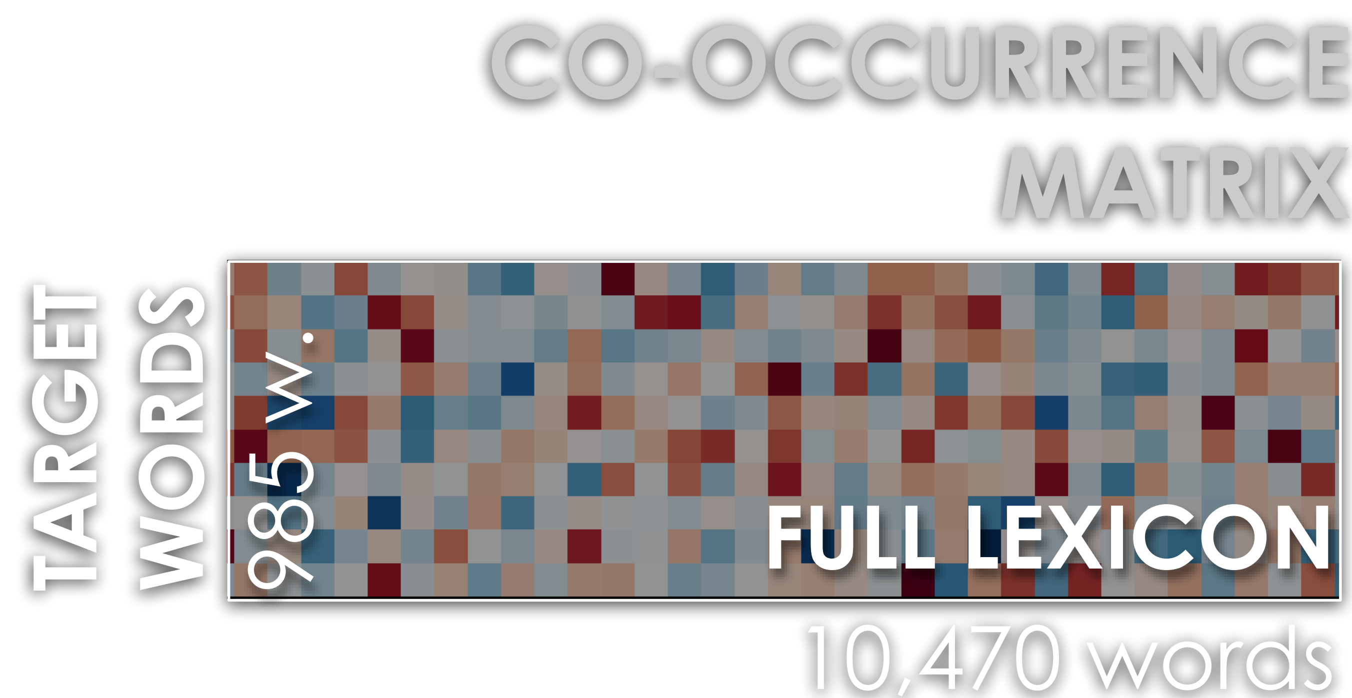
SEMANTICS - WORD2VEC



https://deeplearning4j.org/word2vec

SEMANTICS - ENGLISH-1000

…
difficult
…
husband
…
potato
…
remember
…



SEMANTICS - ENGLISH-1000

* The corpus was used to build a 985 x 10,470 matrix M

  * $M_{i,j}$ is the number of times target i occurs within 15 words of word j

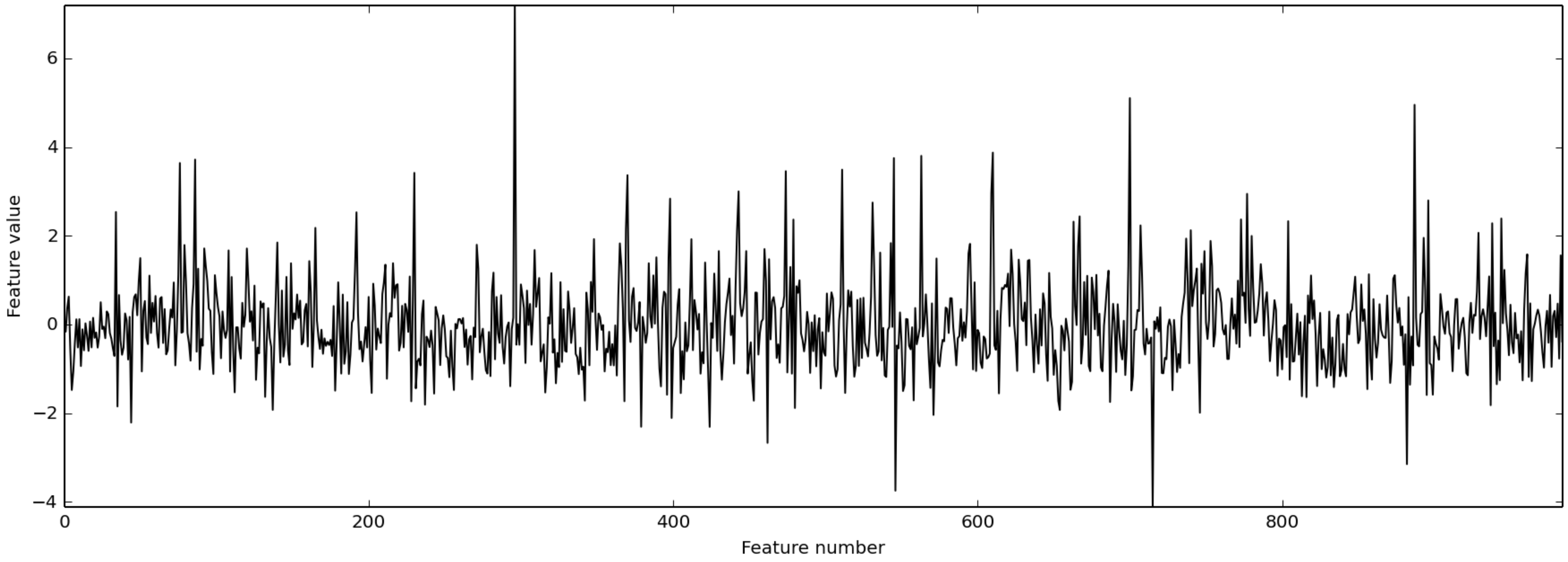
* Then log-transform: $M^*_{i,j} = \log(M_{i,j}+1)$

* Then z-score each row, then each column

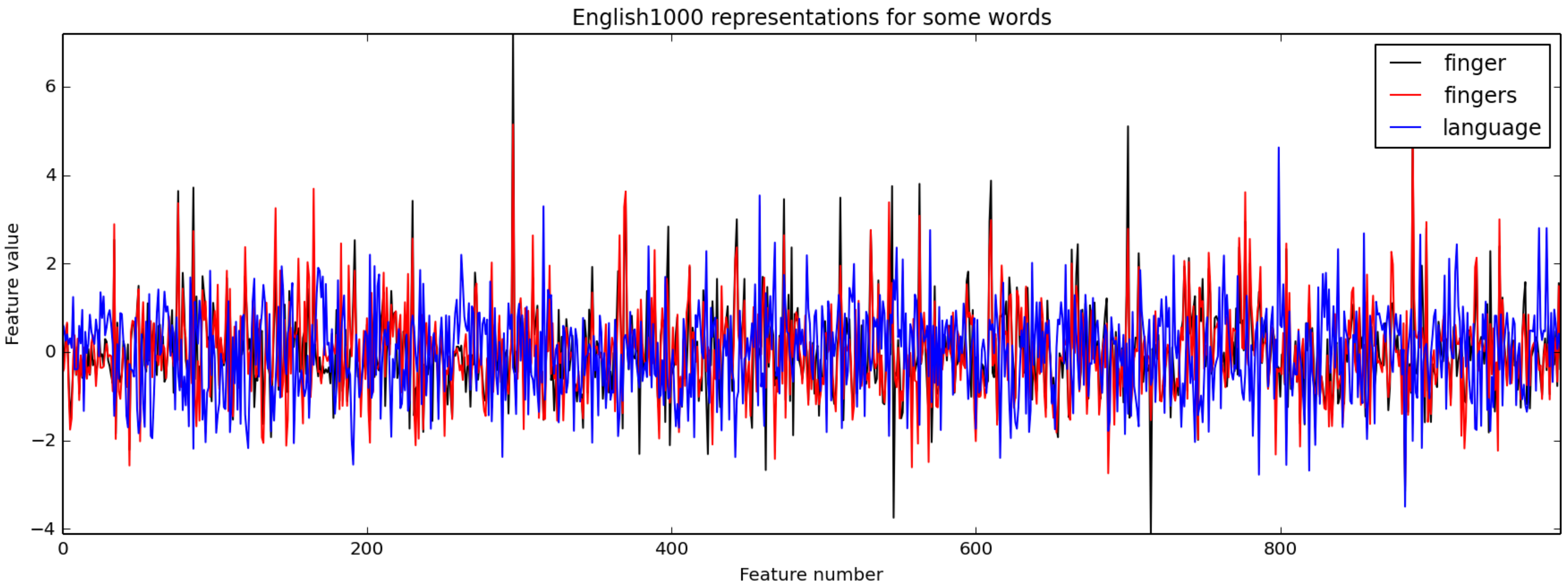
* ... yielding 985-D vector representation of each word in the lexicon

SEMANTICS - ENGLISH-1000

“finger” in english-1000



SEMANTICS - ENGLISH-1000



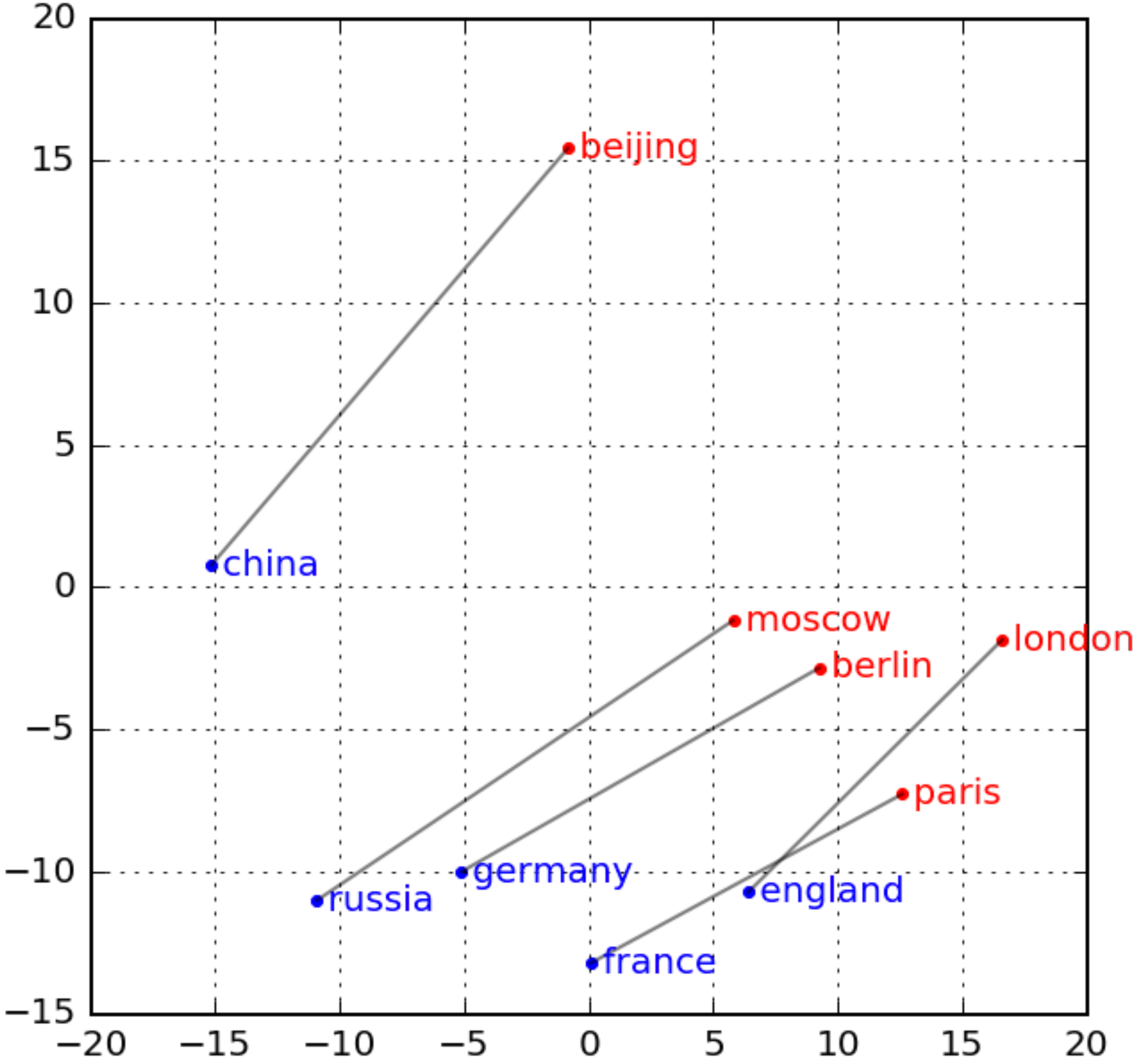
SEMANTICS - ENGLISH-1000

correlation
with “finger”

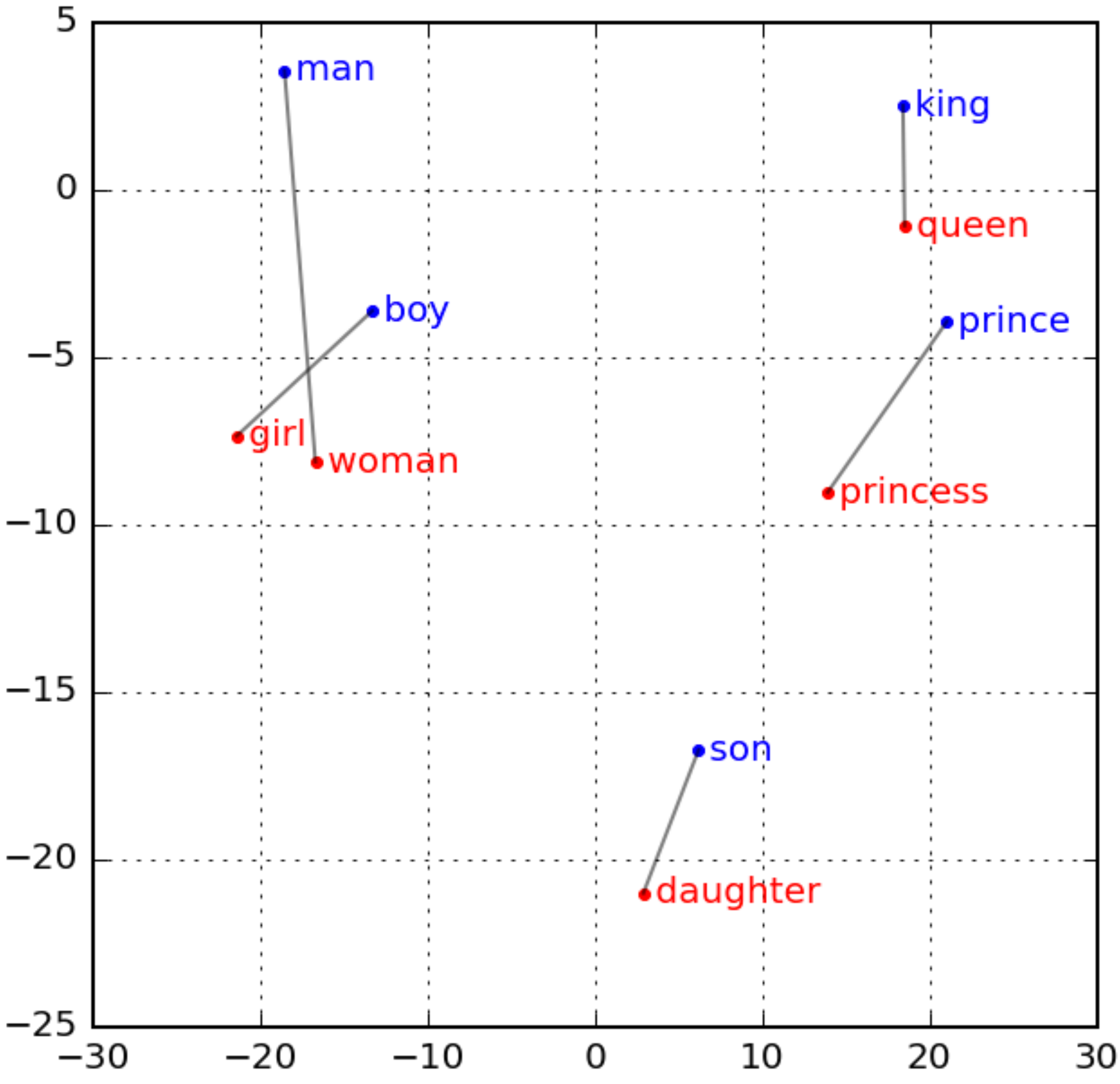
word

1.00, 'finger'
0.81, 'fingers'
0.67, 'hand'
0.67, 'nose'
0.66, 'arm'
0.64, 'mouth'
0.64, 'stick'
0.63, 'neck'
0.63, 'forehead'
0.62, 'tongue'

SEMANTICS - ENGLISH-1000

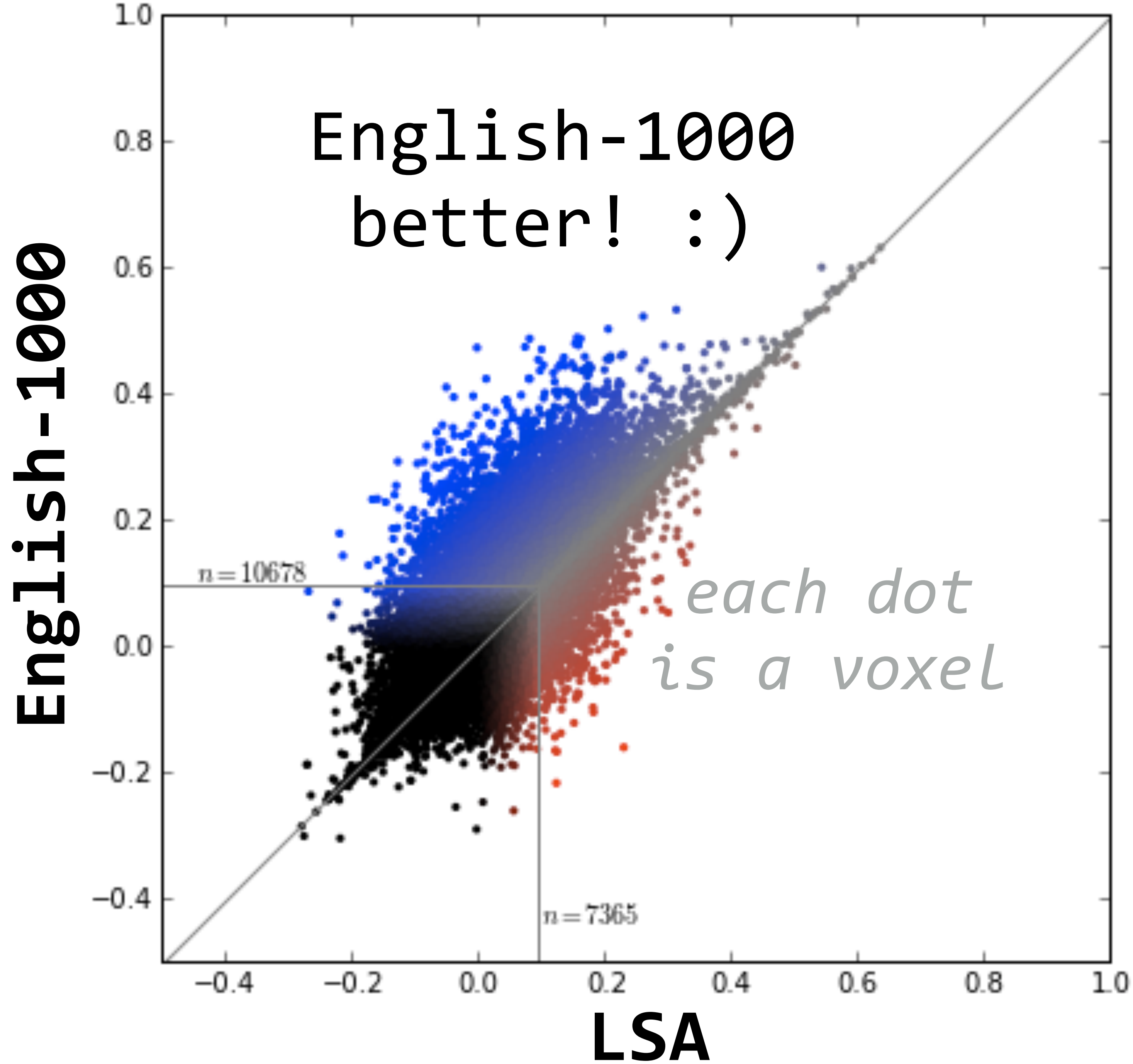


SEMANTICS - ENGLISH-1000



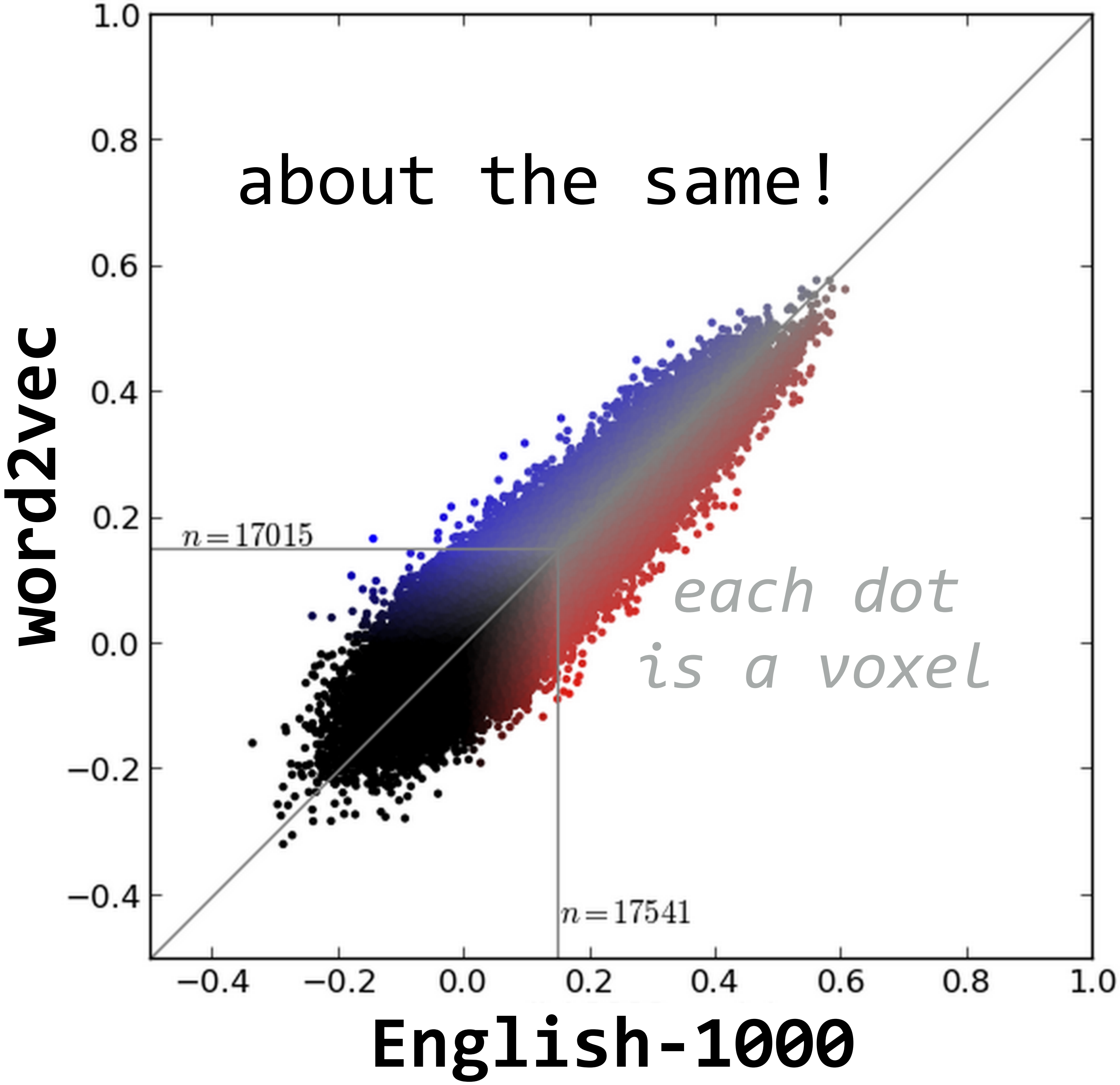
ENGLISH-1000 VS LSA

model performance on held-out data



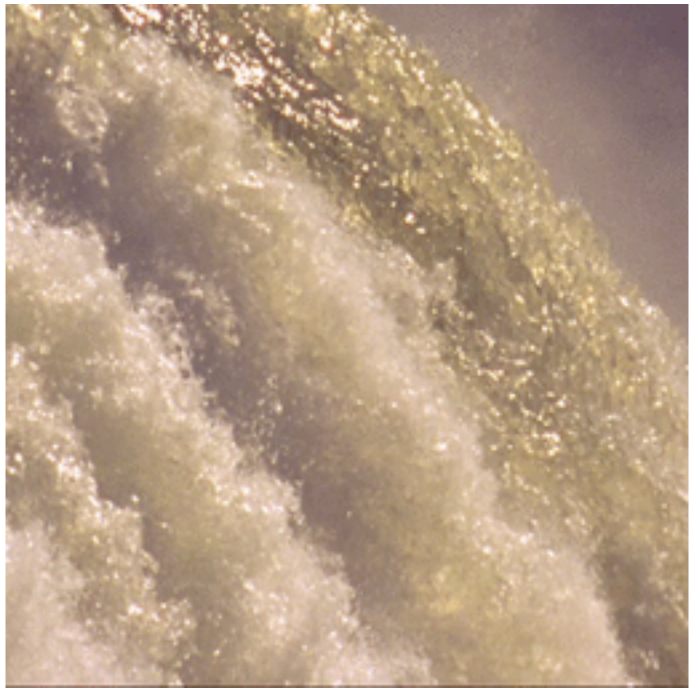
ENGLISH-1000 VS WORD2VEC

model performance on held-out data

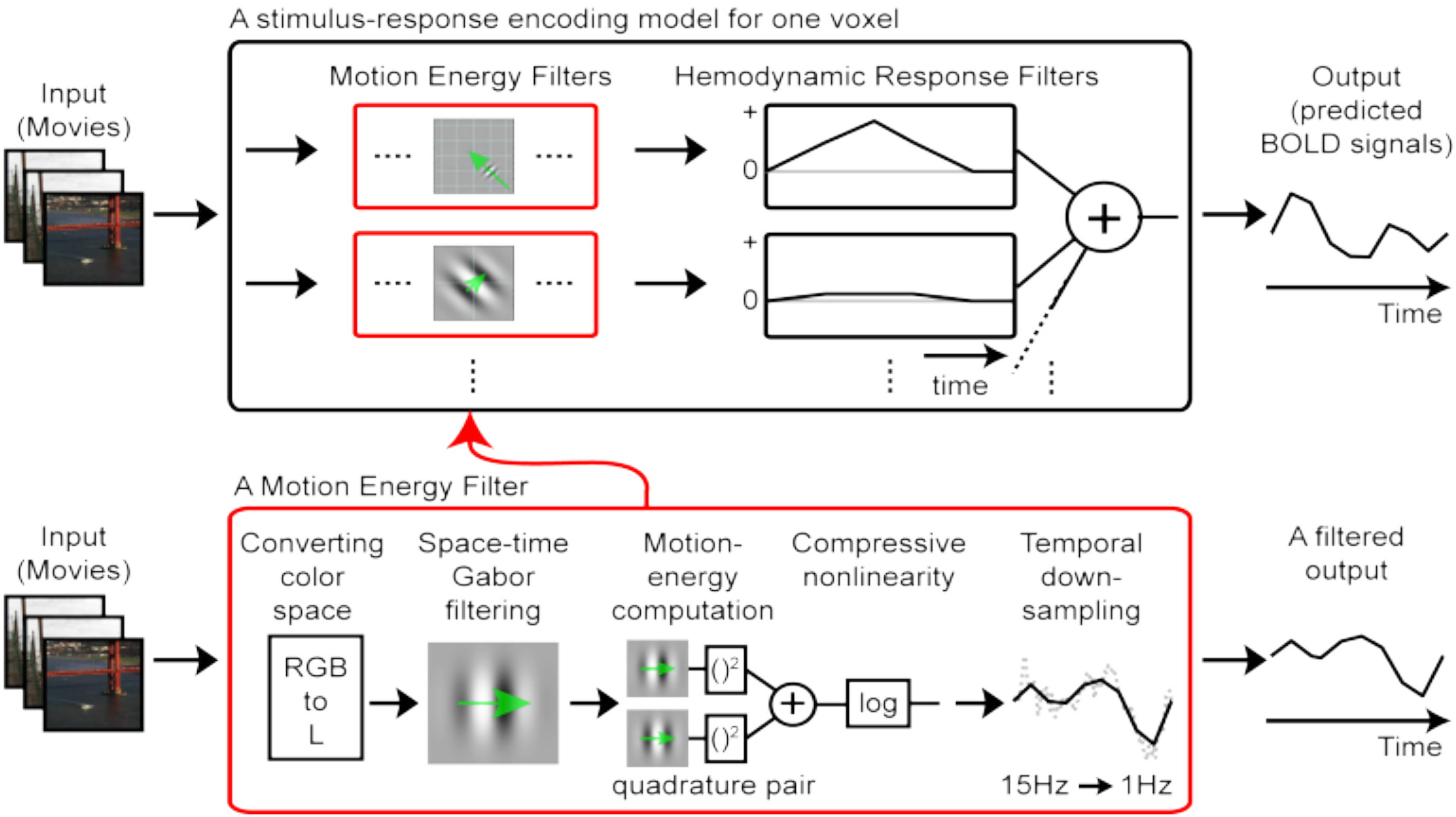


VISION



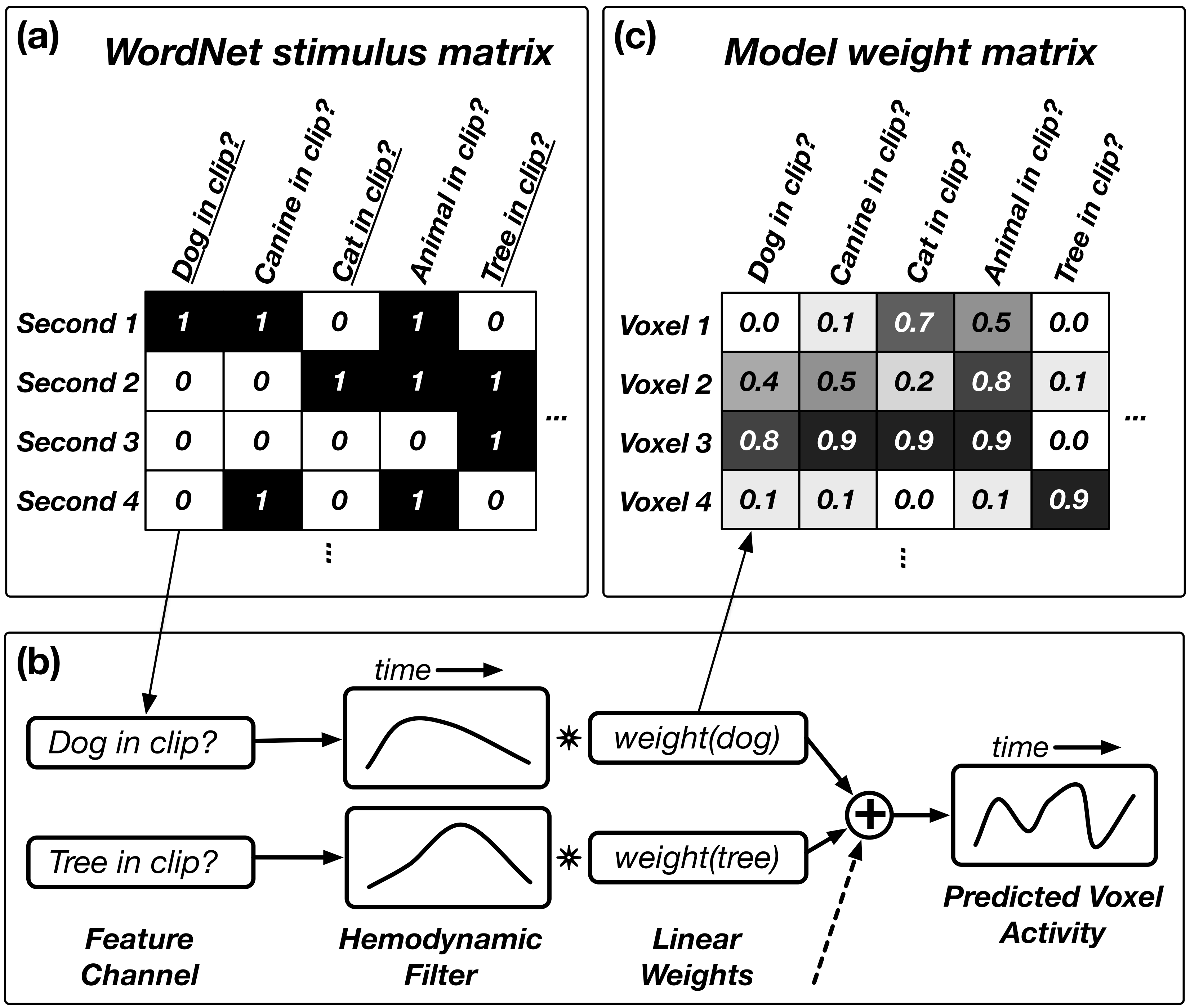



VISION - MOTION ENERGY



Nishimoto et al. (2011)

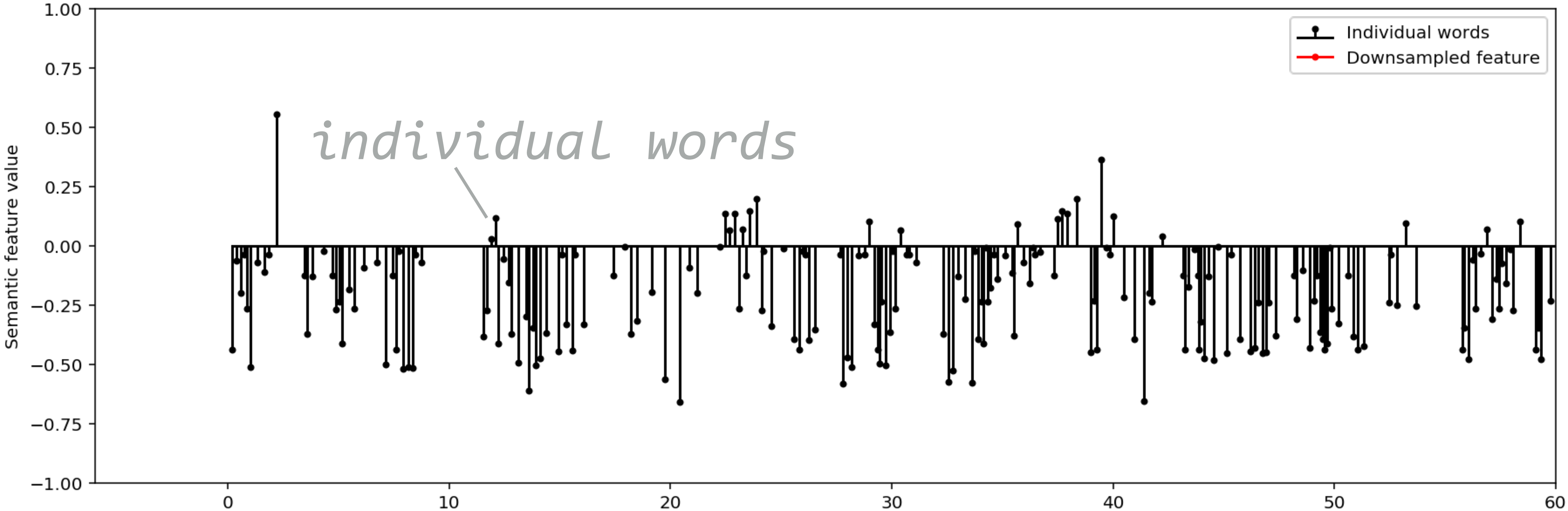
VISION - CATEGORIES



Huth et al. (2012), Huth et al. (2016)

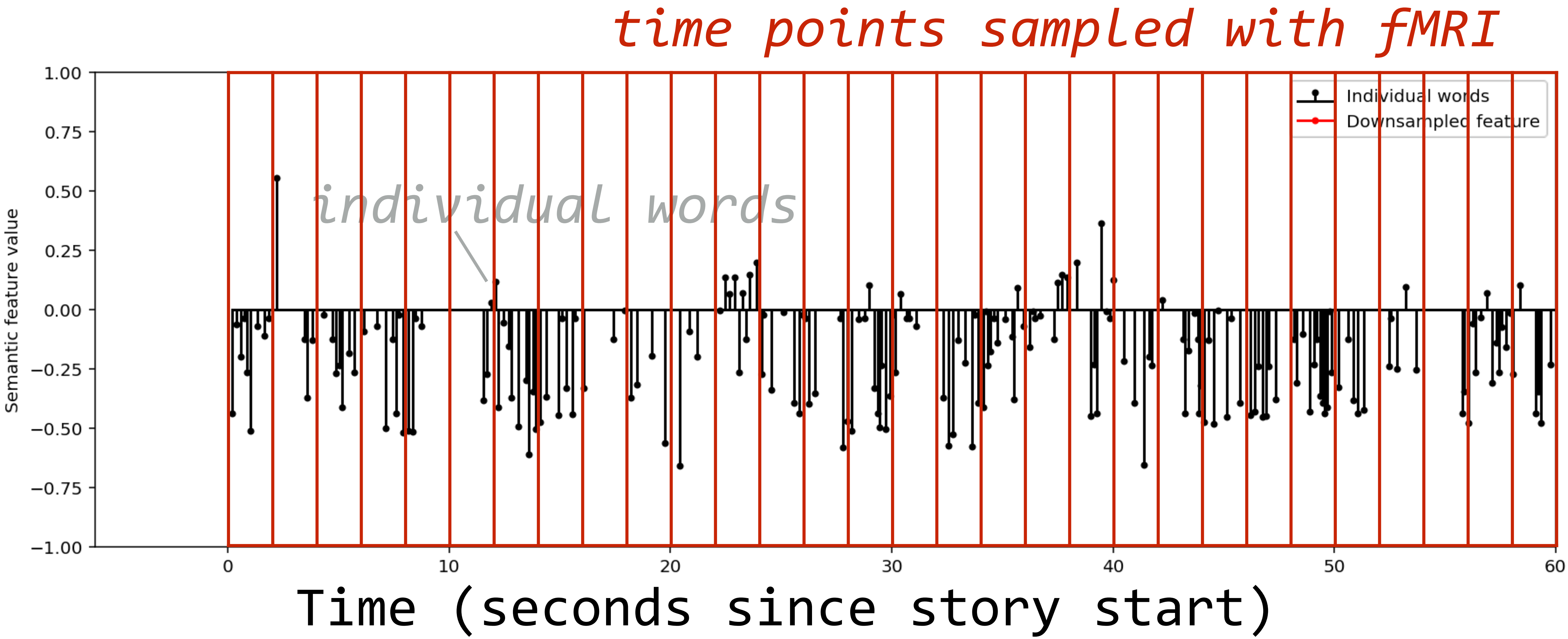
VISION - COMPARISON

STIMULUS TIMESERIES



Time (seconds since story start)

STIMULUS TIMESERIES

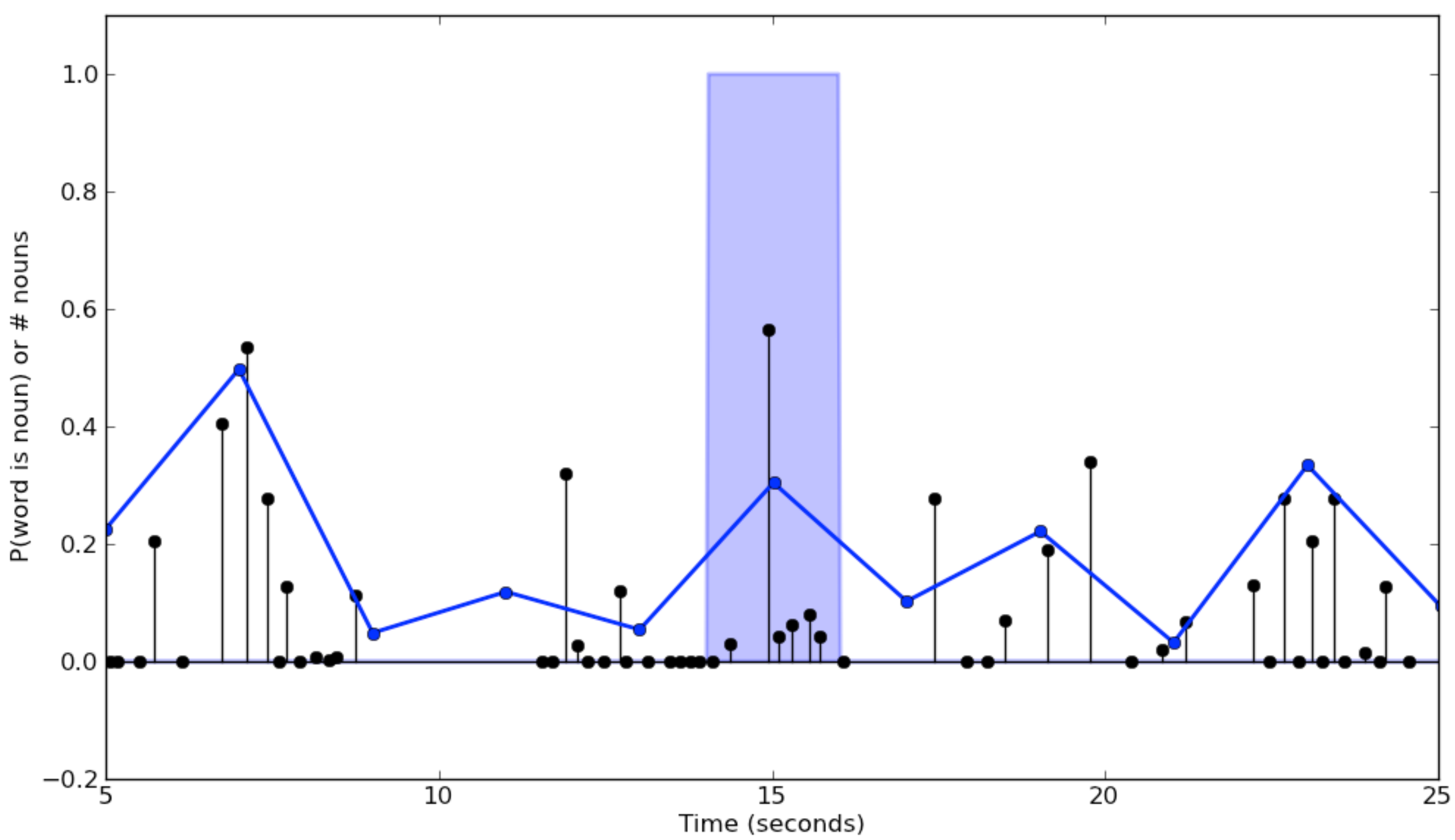


STIMULUS TIMESERIES

* How do we go from fine timescale (fast) stimulus timecourse to coarse timescale (slow) response timecourse?

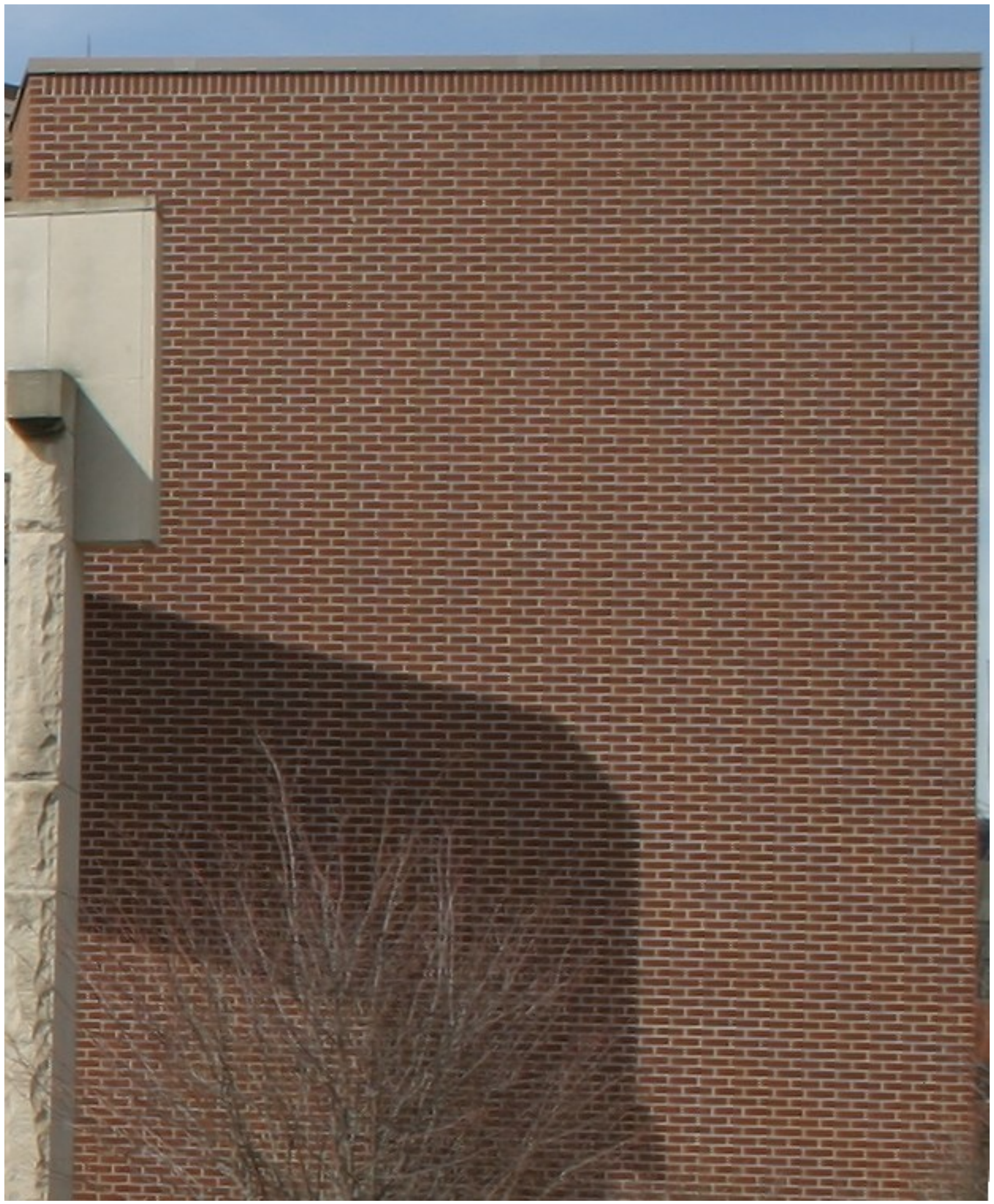
    * Average all the points within each 2-second period? ***NO!***

SAMPLING

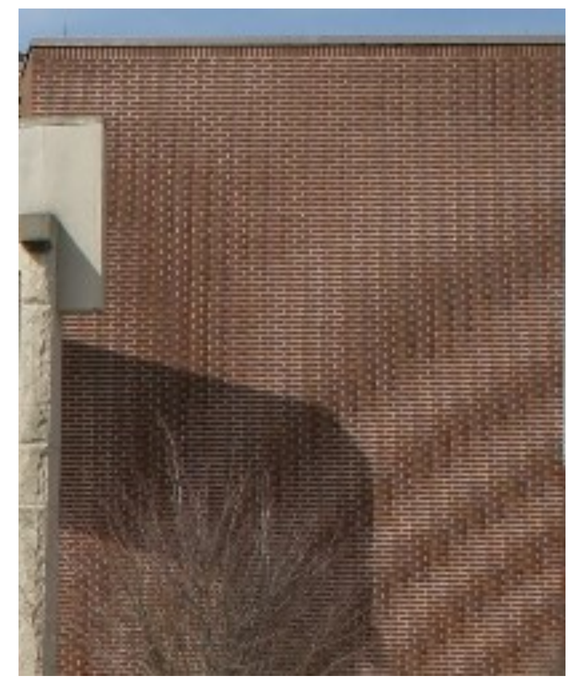


Averaging within each time period =
downsampling with rectangular window

Original Image



Poorly downsampled

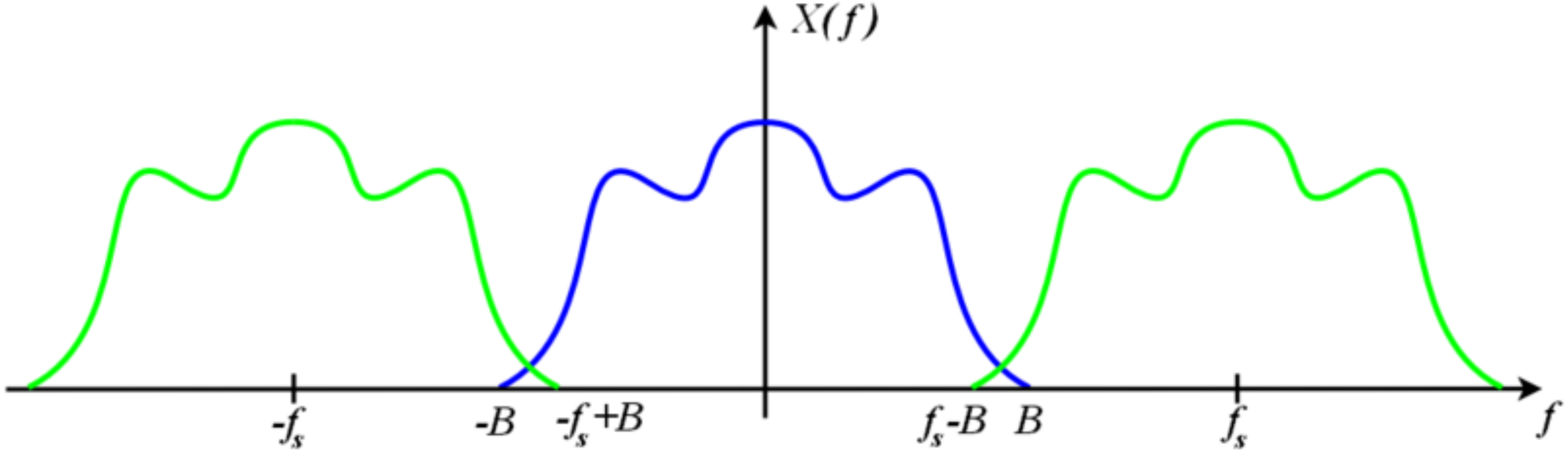


(using rectangular window!)

SAMPLING

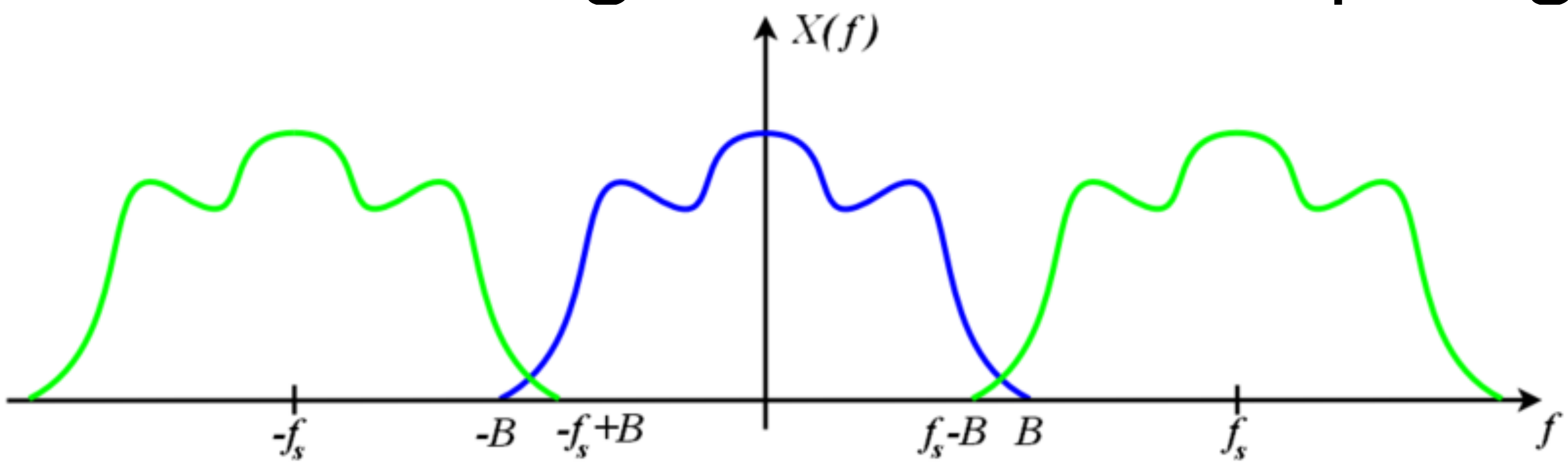
* If the **sampling rate** is too low, the spectrum of the signal can overlap with itself

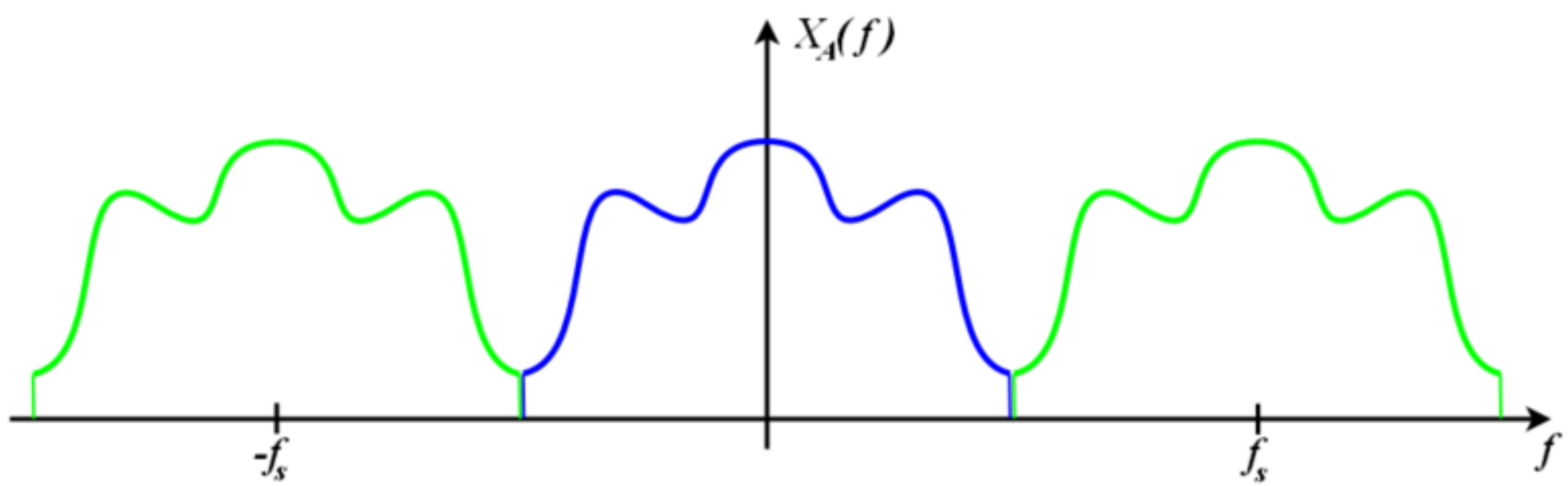
* This is called **folding** or **aliasing**



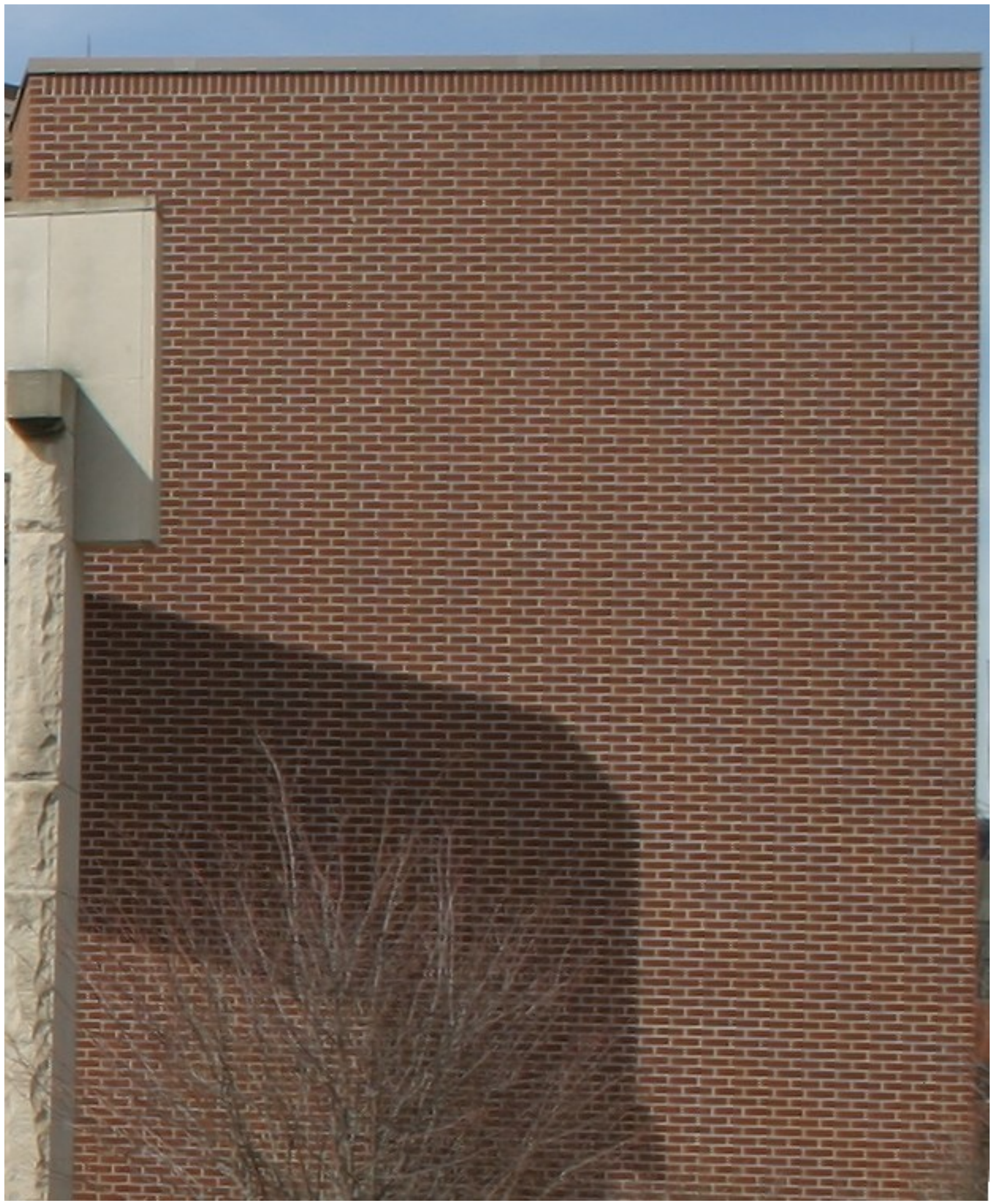
SAMPLING

* To avoid aliasing we can **low-pass filter** our signal before sampling





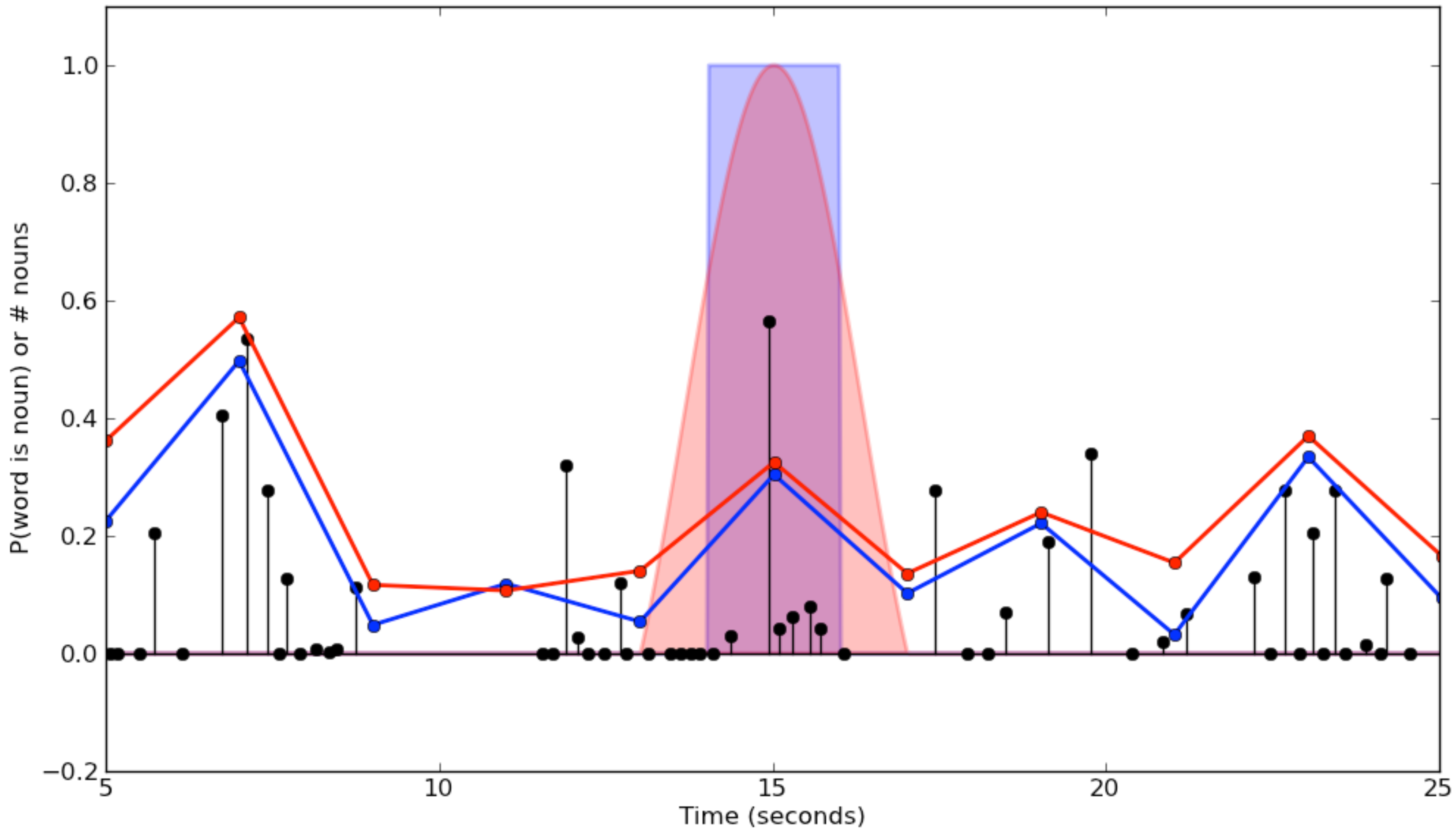
Original Image



Poorly downsampled

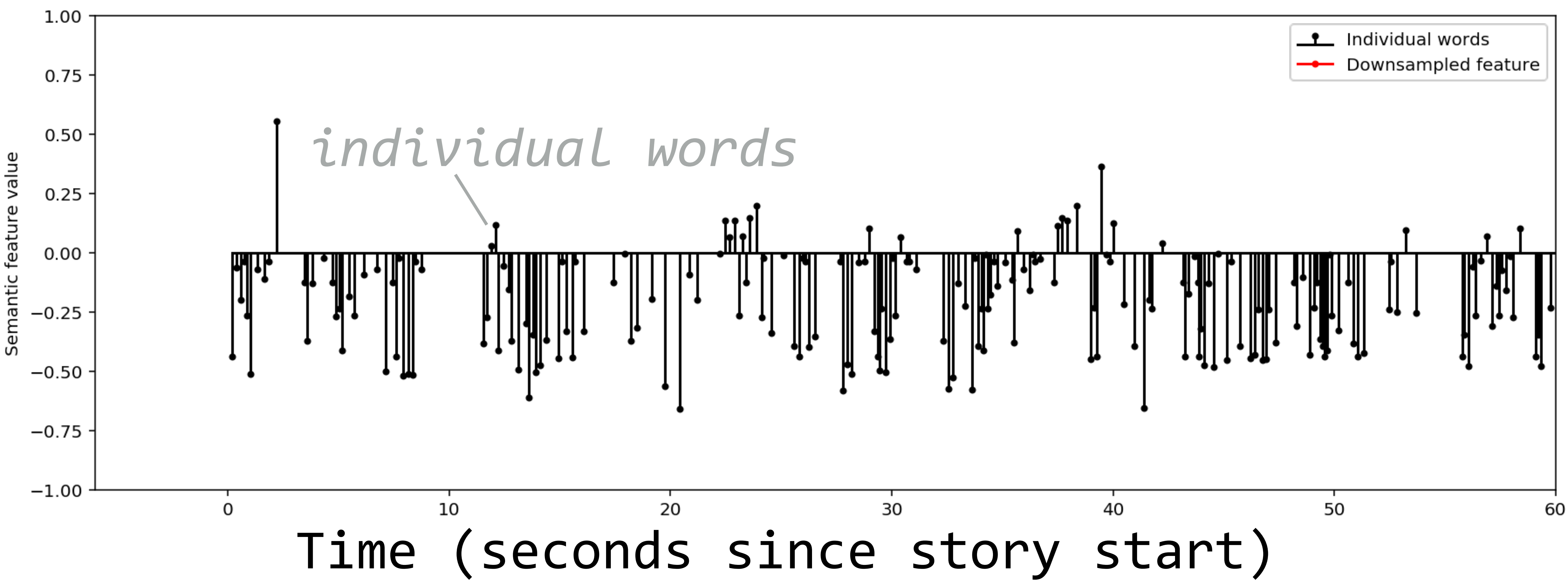
Properly downsampled

SAMPLING

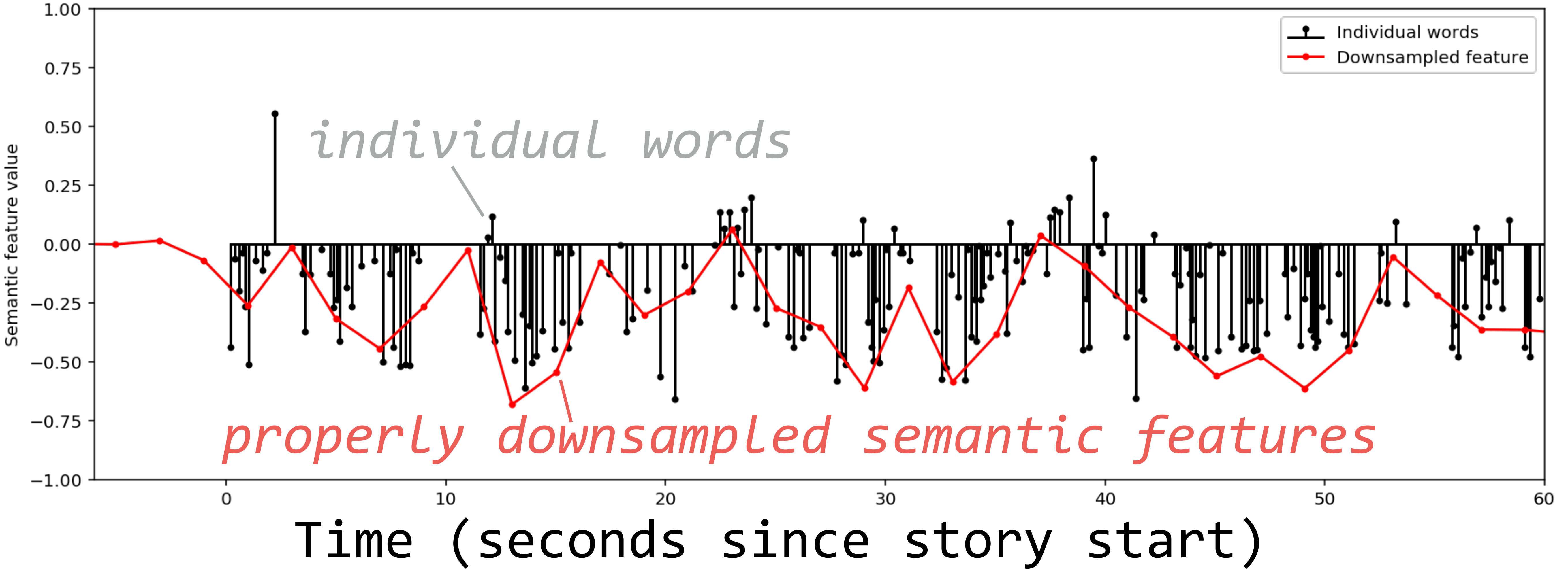


Instead, downsample
using a 1-lobe sinc filter
(AKA a Lanczos filter)

STIMULUS TIMESERIES



STIMULUS TIMESERIES



RECAP

* System identification - linear, linearized, nonlinear

* ***LINEARIZED MODELS***

* How to downsample stimuli that are too fast

HASTA MAÑANA, CHICOS!