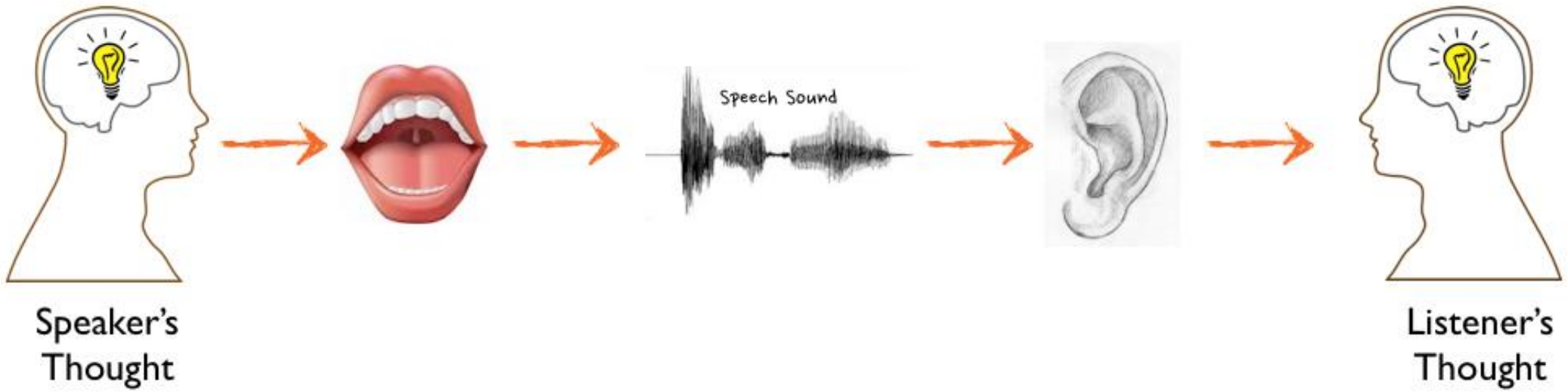
Language and Cognition

Amitabha Mukerjee
IIT Kanpur

amit@cse.iitk.ac.in

Language and Cognition



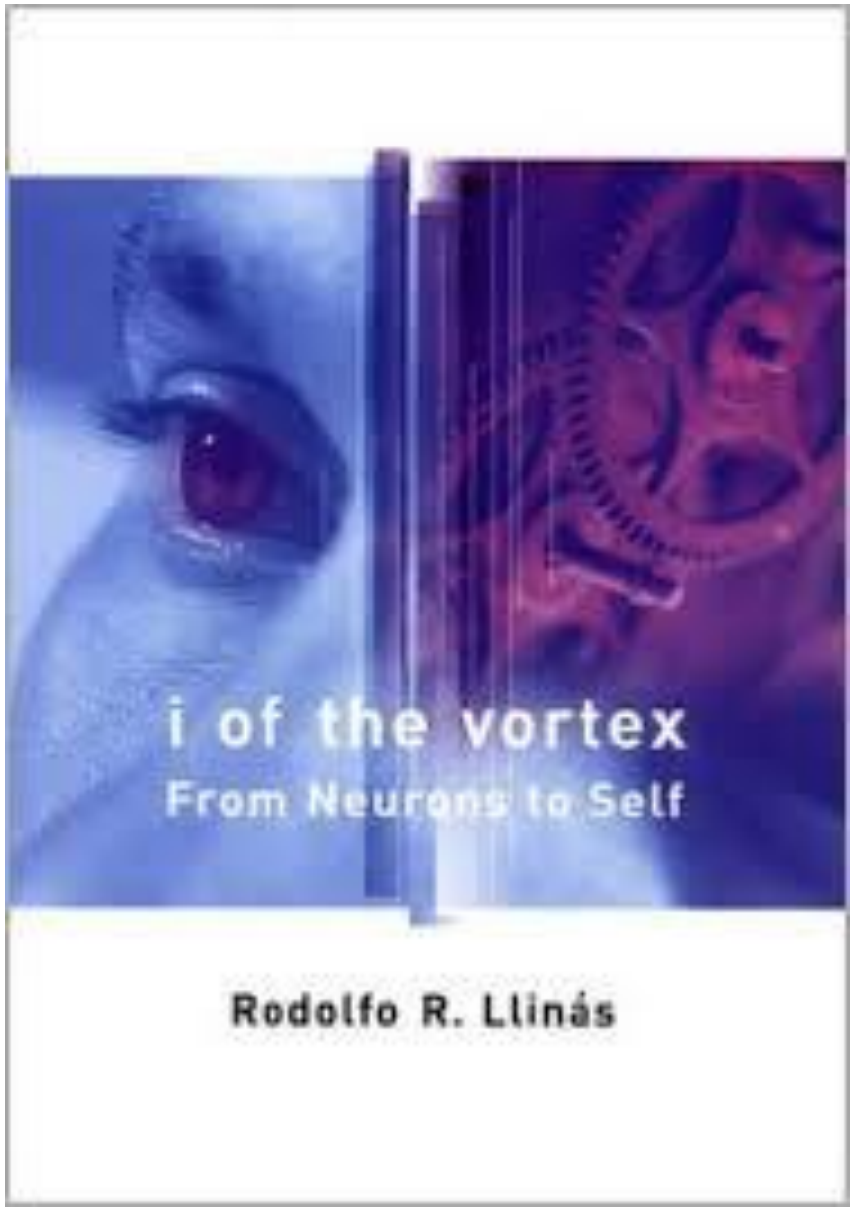
source: http://www.mimicmethod.com/flow-101-day-1.html

Biological Origins of Mind and Language

Motor origins of the mind

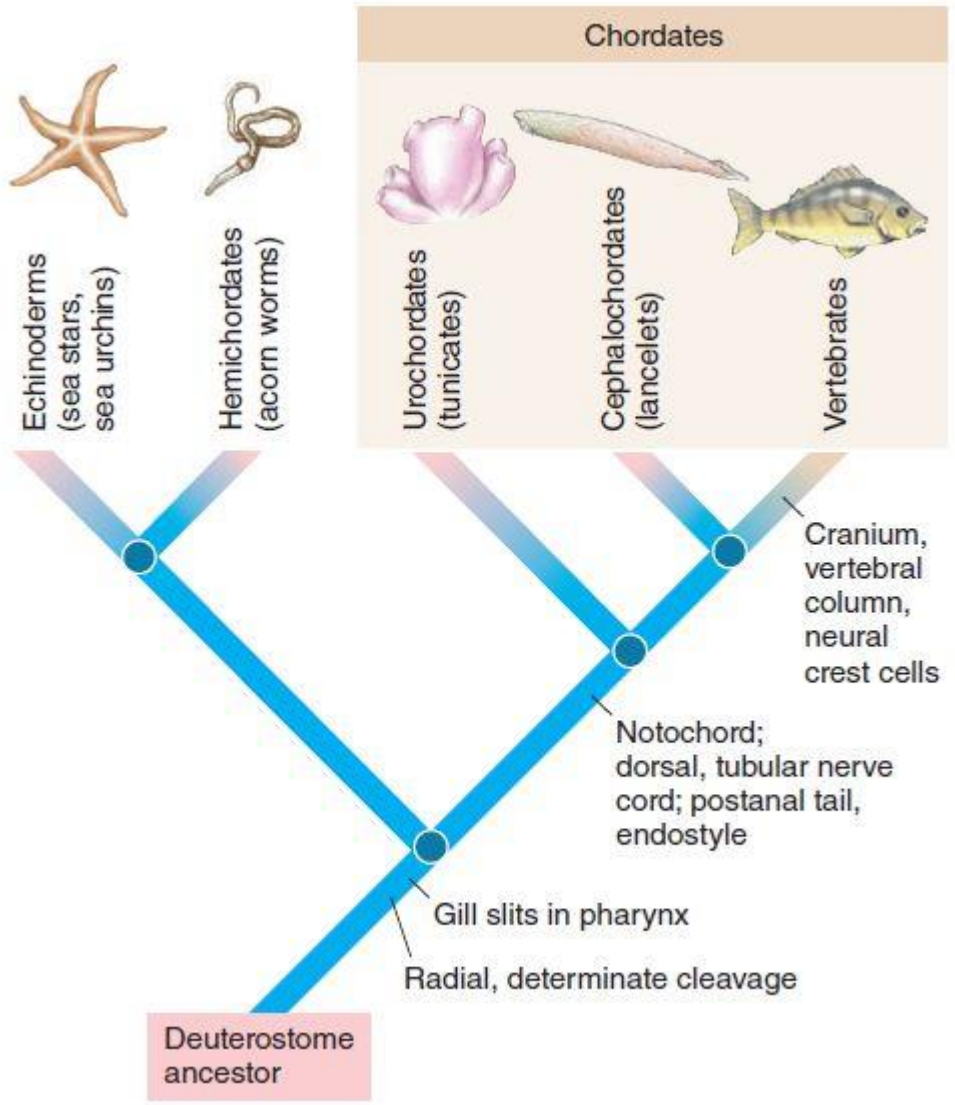
Rodolfo Llinas

I of the Vortex, 2002



Motricity → Nervous system

Tunicates (sea squirts) :
notochord + ganglion:
stage before evolution of
vertebrates



[Delsuc et al 06]

Motricity → Nervous system

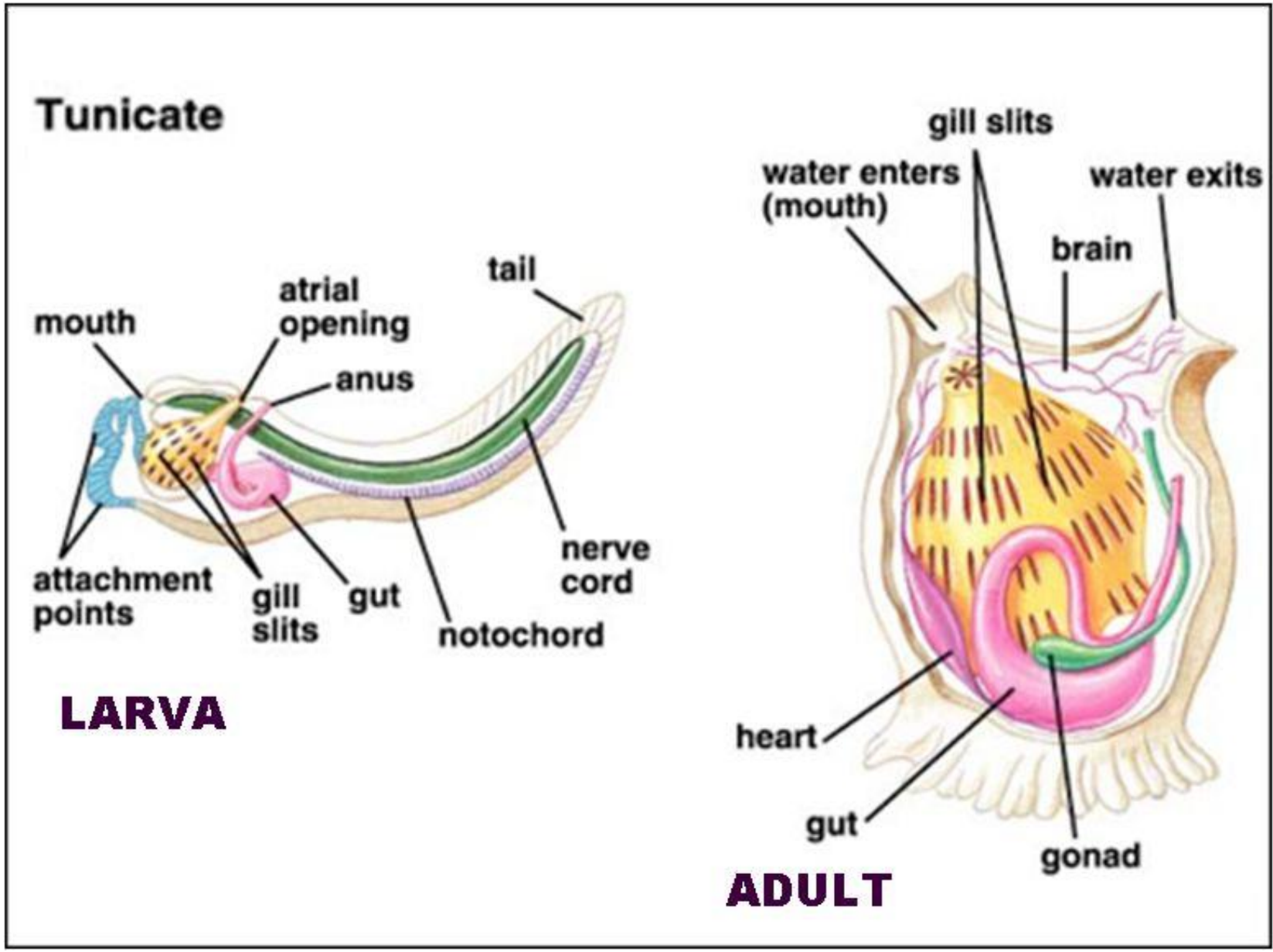
Tunicates (sea squirts) : sessile adults

adult - immobile (sessile)

larval form - briefly free swimming

larva has 300 cell ganglion + notochord

(digested after it finds and attaches to a site)



http://finstofeet.com/2010/03/14/1-2-behold-the-tunicates/

Nervous system:
Evolved for planning motions

planning ← prediction

Predicting → Planning

panther chameleon tongue

The capacity to predict the outcome of future events—critical to successful movement— is, most likely, the ultimate and most common of all global brain functions.

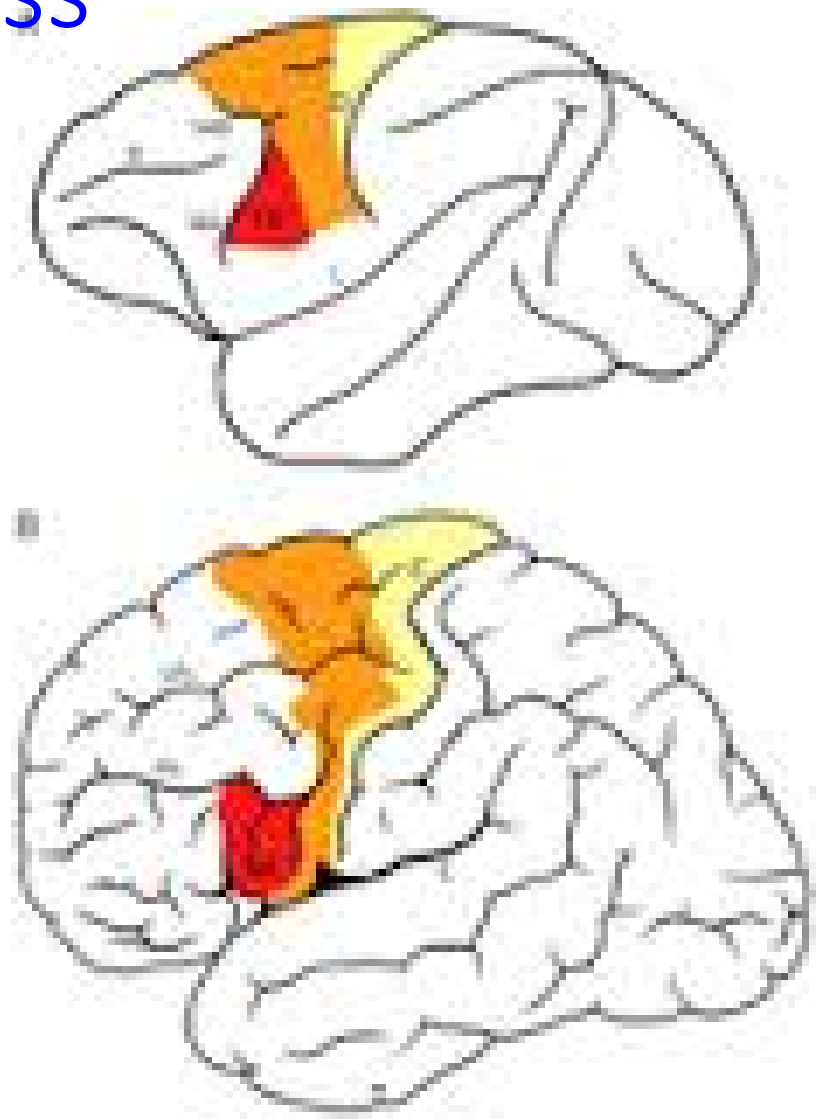
- Rodolfo Llinas

Motor knowledge → Mindness

predictive / intentional interactions

- requires internal image of world
- requires models for consequence of actions

organized motricity: cephalization



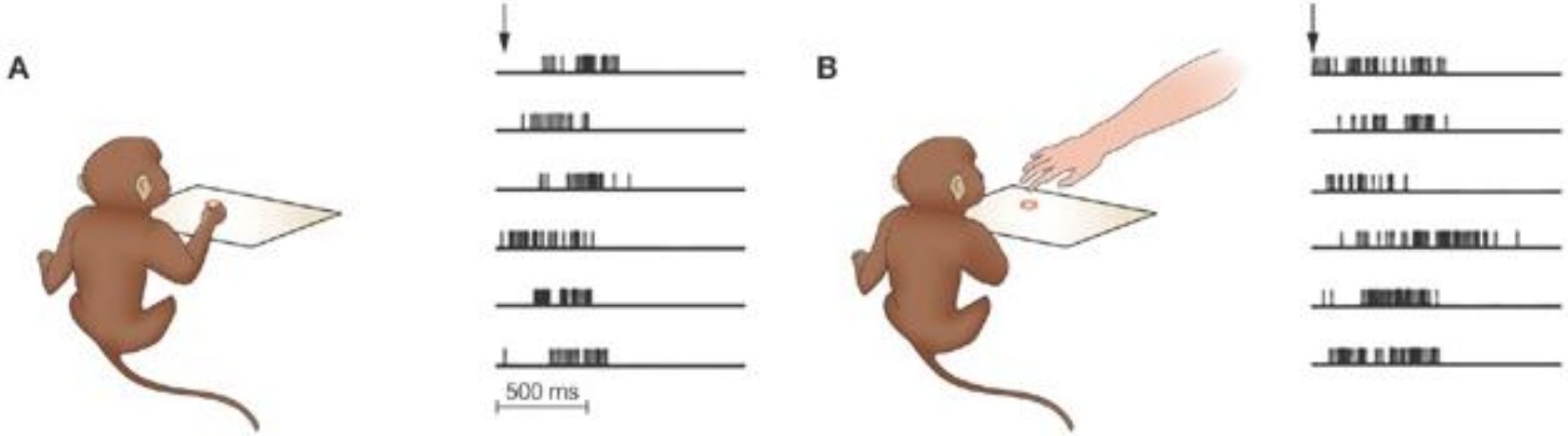
sensory-motor areas in
macaque and human cortex

The Complexity of Language:

Unifying multiple sensory + motor modalities

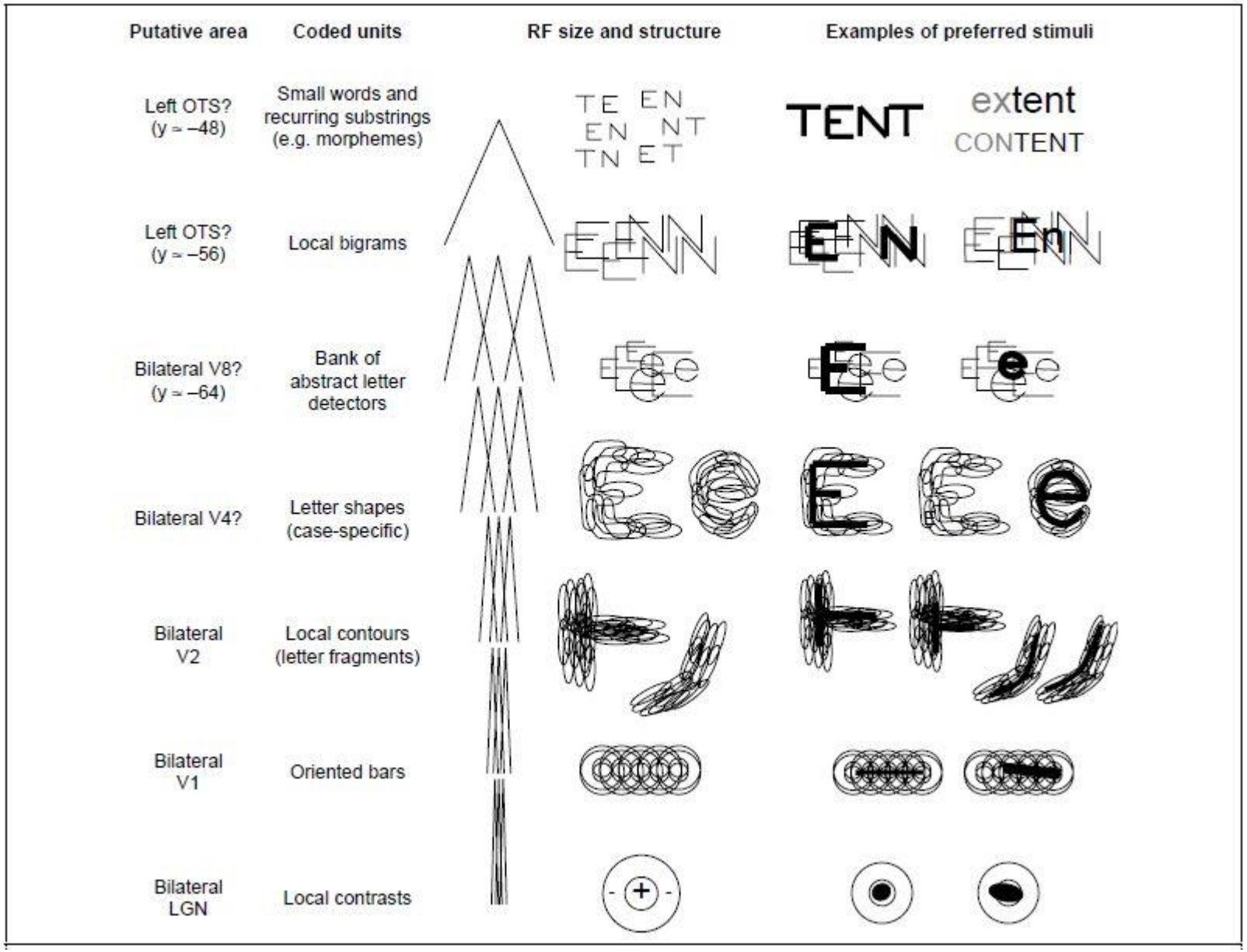
Meaning : Unifying Modalities

Mirror Neurons



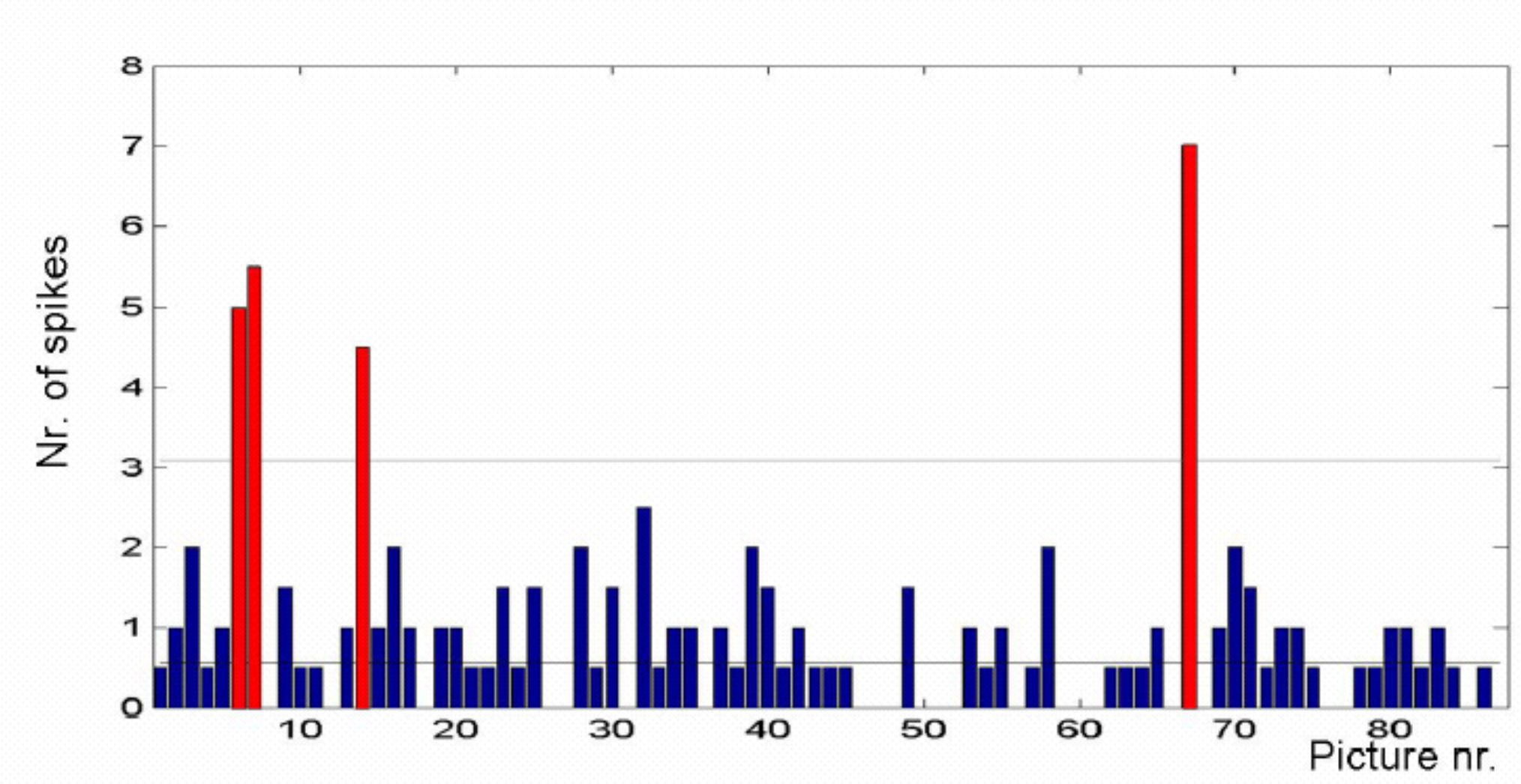
[Rizzolatti G and Fabbri-Destro M 08]

Reading in the brain



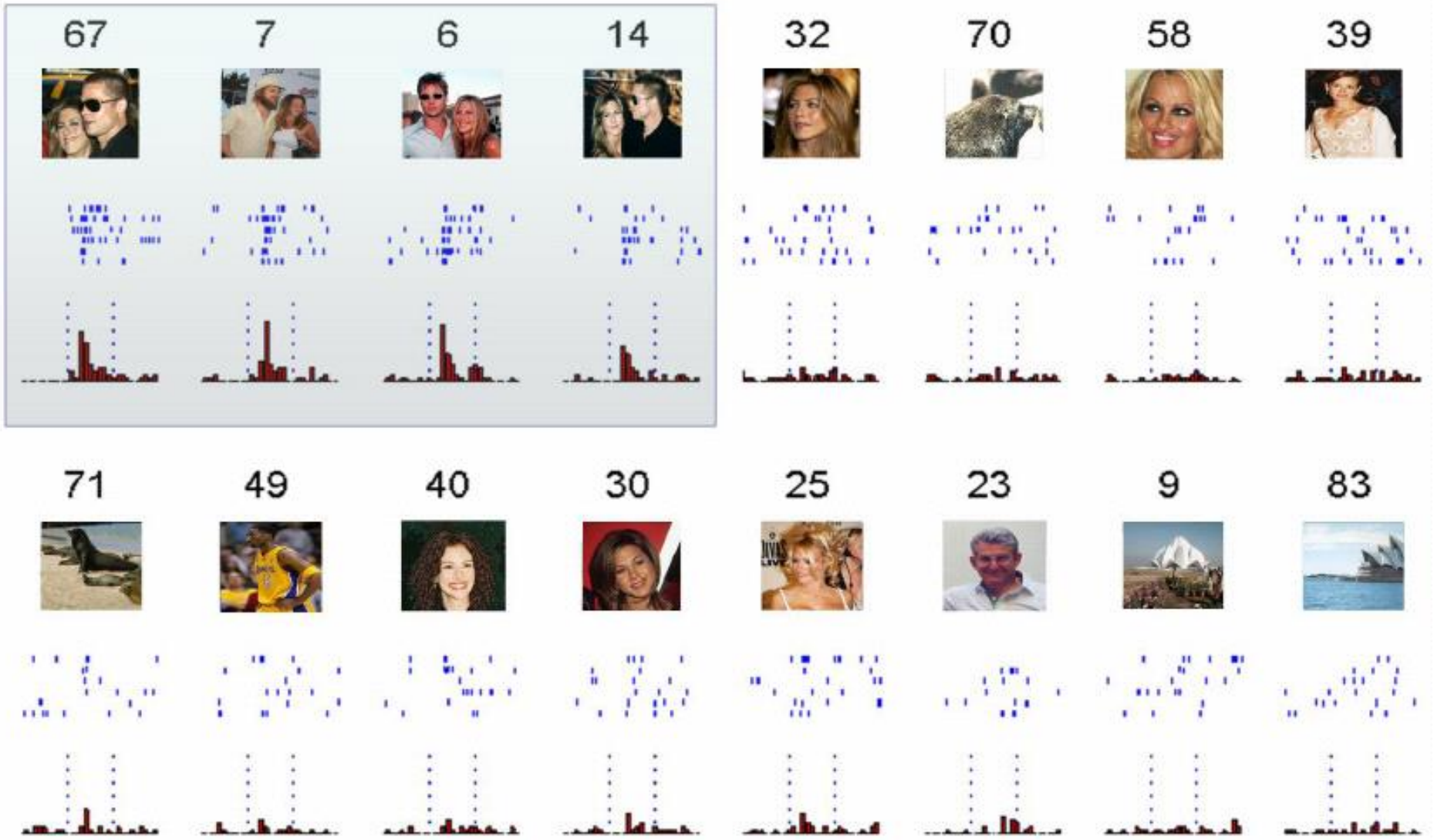
[dehaene-cohen-05_neural-code-for-written-words-proposal]

Visual Recognition: IT Cortex

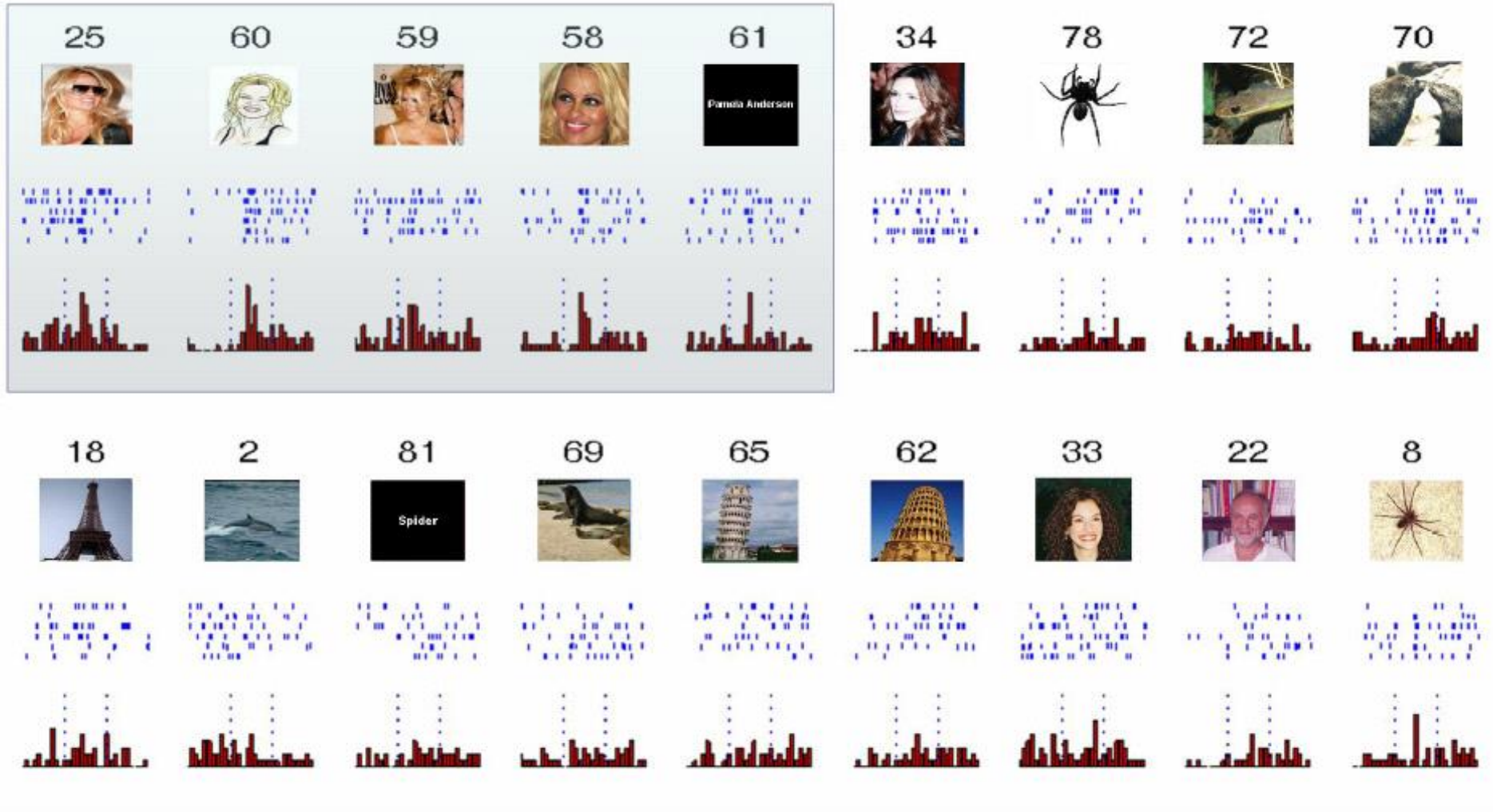


Quian Quiroga, Reddy, Kreiman, Koch and Fried, *Nature* (2005)

What is *s*?



Higher Neurons



Grammar and Cognition: A history

Empiricism vs Rationalism - Pendulum

pAniNi, aristotle – empiricist

plato – mystical / rationalist

port royal grammarians 17th c. – “mental” aspects – language is universal

wundt / james – introspective – [ebbinghaus]

behaviourism – empiricist – rejected mentalism

chomsky – rationalist – “mental” – innate - universal

Behaviourism

In teaching the young child to talk, the formal specifications upon which reinforcement is contingent are at first greatly relaxed. Any response which vaguely resembles the standard behavior is reinforced. When these begin to appear more frequently, a closer approximation is insisted upon. In this manner, very complex verbal forms may be reached.

BF Skinner, *Verbal Behaviour* 1956, (p.29–30)

Chomsky

Mostly, sentences have never been seen before

(e.g. “Colourless green ideas sleep furiously”)

Hence cannot have been learned via reinforcement

Also – grammar requires long distance dependencies

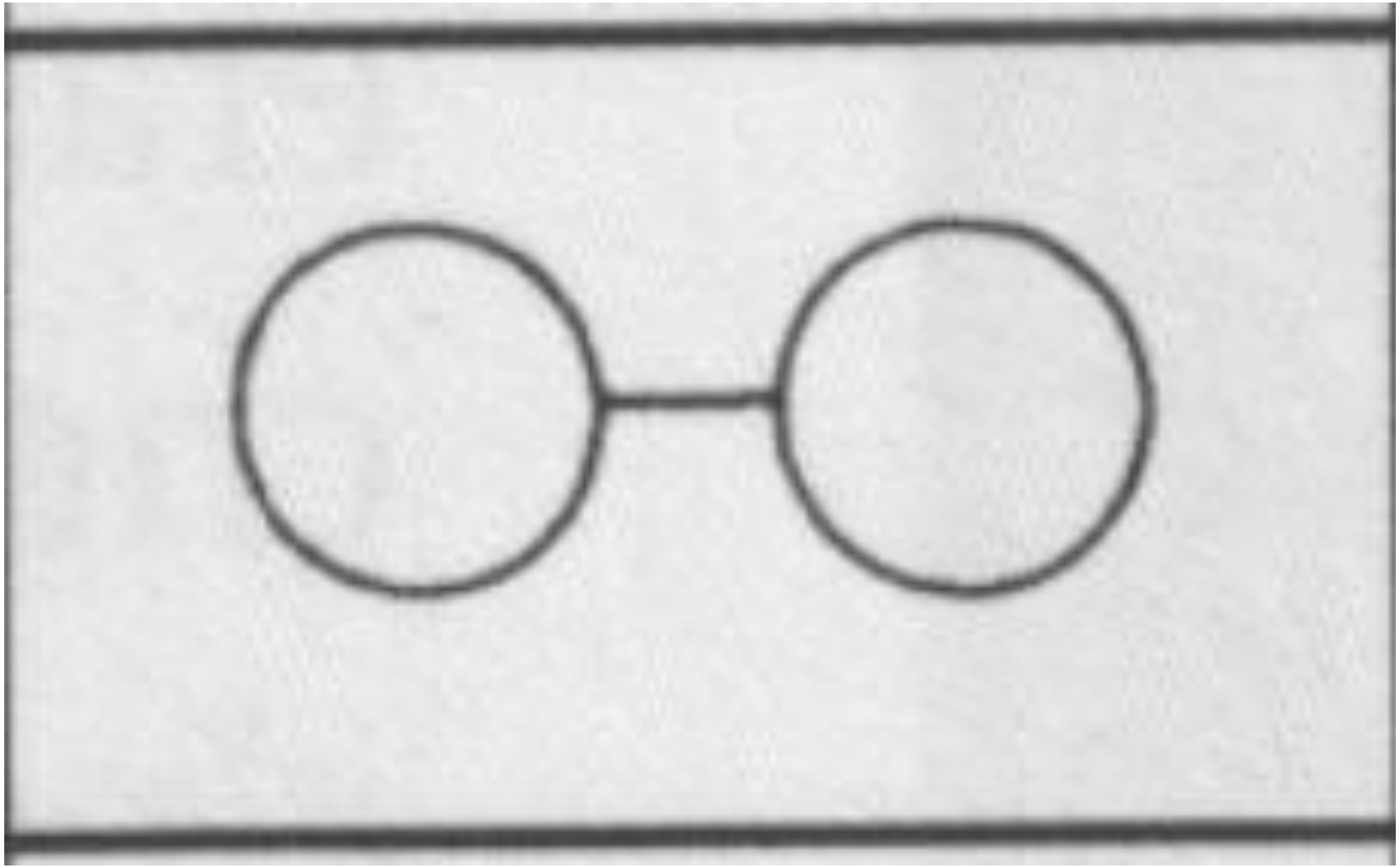
Also probabilities are not possible since various word combinations may not have been seen before.

Language - amodal?
Multimodal?

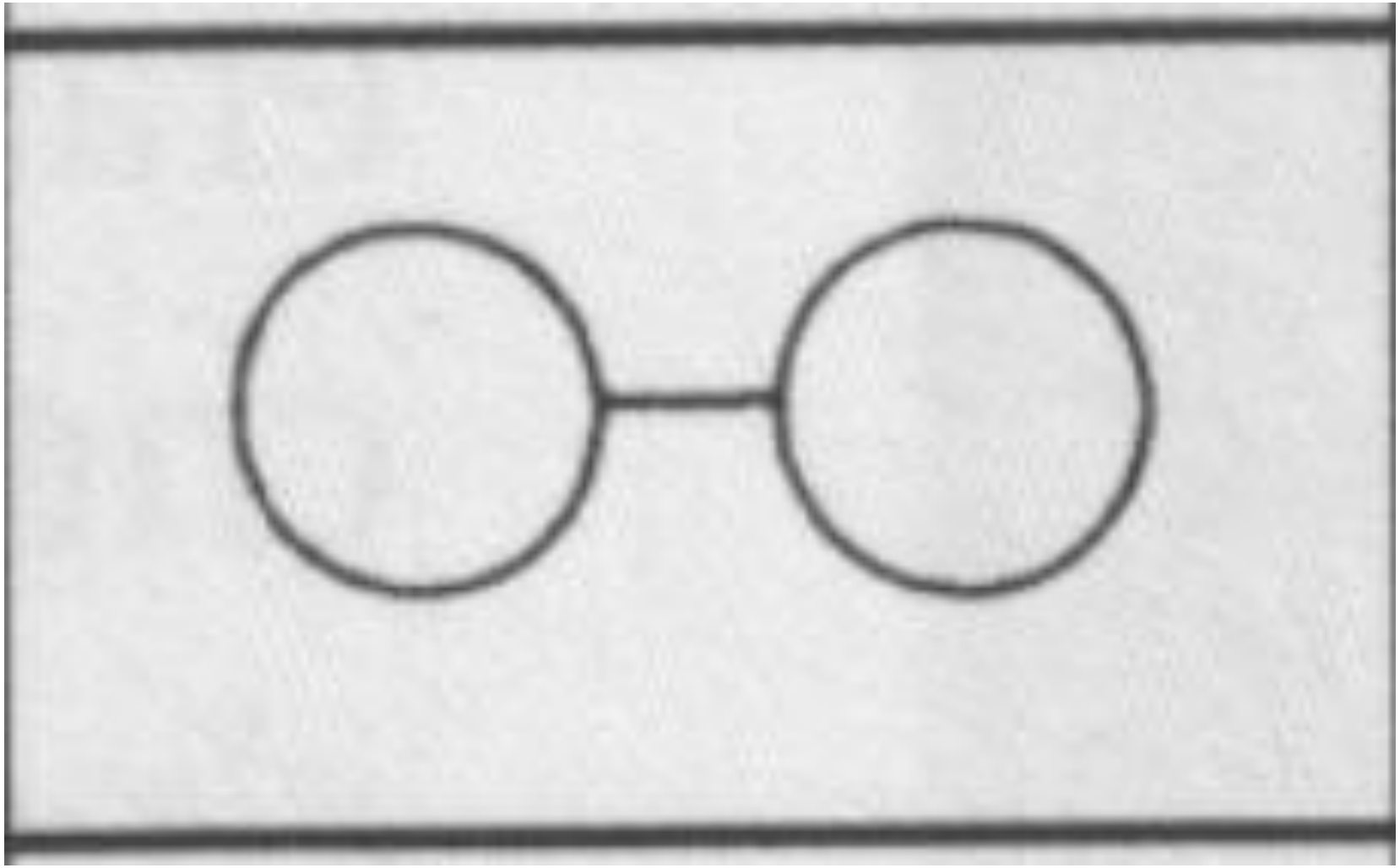
Please read the title and look at the picture

Try to remember both

Eye-glass



Dumb-bell



Perception and Language affect each other

Structure in Language

Structure in language : Word

पांच फिरंगी अफसरों __ फांसी पर ___ दिया

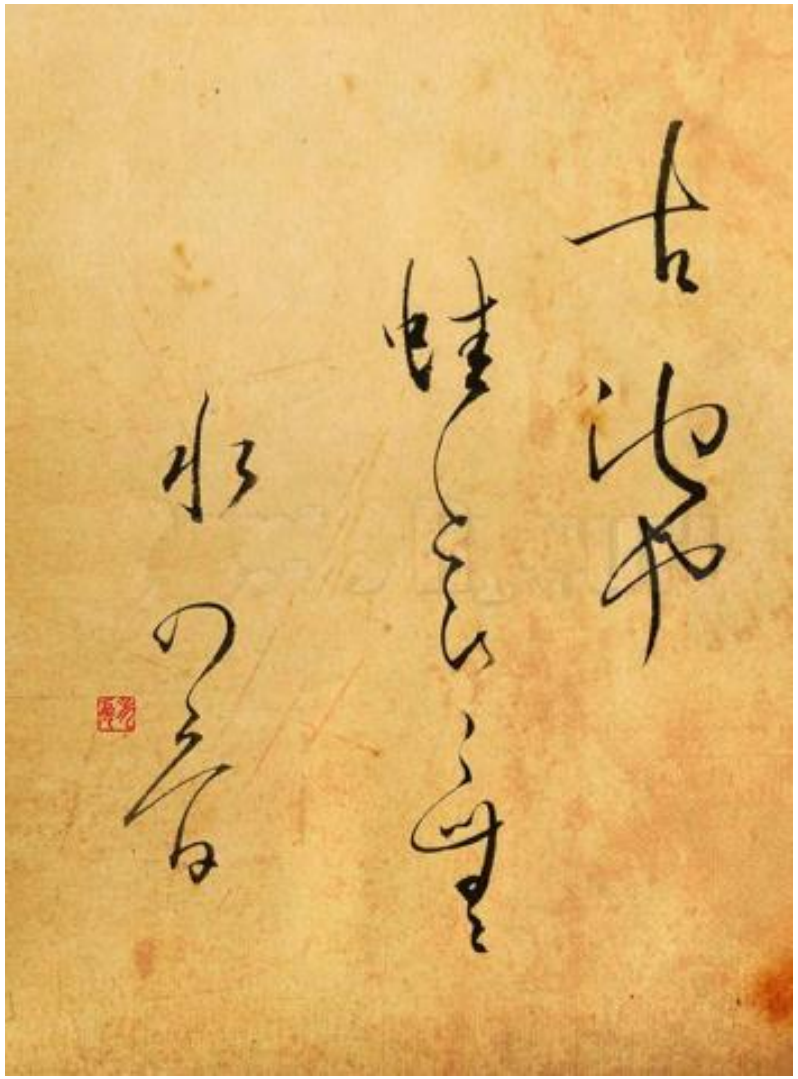
what can go in the blanks?

Structure in language : Syllable

पां च फि रं गी अफ स रों को फां सी
पर लट का दि या

Which syllables follow which others?

Word?　haiku　古池や蛙飛こむ水のおと



古池や蛙飛こむ水のおと
Matsuo Basho, (1644-94)

Furuike ya
kawazu tobikomu
mizu no oto

ancient pond
frog jumps in
sound of water

古old 池 pond や –prtcl, “a”
蛙 frog 飛こむ jump-3p-trml
水 water の -gen おと sound

Word?

旅夜書懷 – Du Fu

Du Fu 712-770

旅夜書懷

細草微風岸
危檣獨夜舟
星垂平野闊
月湧大江流
名豈文章著
官應老病休
飄飄何所似
天地一沙鷗

Thoughts While Travelling at Night

Light breeze on the fine grass.
I stand alone at the mast.
    Stars lean on the vast wild plain.
    Moon bobs in the Great River's spate.
Letters have brought no fame.
Office? Too old to obtain.
    Drifting, what am I like?
A gull between earth and sky.

[tr. Vikram Seth]

Word? Thai Khlong

Stanza from Lilit Phra Lo (ลิลิตพระลอ) :

เสียงฦๅเสียงเล่าอ้าง อันใด พี่เอย
เสียงย่อมยอยศใคร ทั่วหล้า
สองเขือพี่หลับไหล ลืมตื่น ฤๅพี่
สองพี่คิดเองอ้า อย่าได้ถามเผือ

What tales, what rumours, you ask?
Of whom is this praise being broadcast?
Were you two sleeping, have you forgotten waking up?
Figure it out yourselves; don't ask me.

Shannon Entropy

- Predict the next word/ letter / syllable, given (*n-1)* previous letters or words
- Surprisal on hearing “x” : $- \log \mathrm{P}(\mathrm{x})$
- Entropy : Expectation of surprise

$$\mathrm{H}(X) = -\sum_i \mathrm{P}(x_i) \log_b \mathrm{P}(x_i)$$

Claude E. Shannon. “Prediction and Entropy of Printed English”, 1951.

Shannon Entropy : Human

- Ask human to guess the next letter:

```
THE ROOM WAS NOT VERY LIGHT A SMALL OBLONG
----ROO------NOT-V-----I------SM----OBL---

READING LAMP ON THE DESK SHED GLOW ON
REA----------O------D----SHED-OLD--O-

POLISHED WOOD BUT LESS ON THE SHABBY RED CARPET
P-L-S-----O---BU--L-S—O-------SH-----RE—-C-----
```

- 69% guessed on 1st attempt ["-" = 1st attempt]

Claude E. Shannon. "Prediction and Entropy of Printed English", *Bell System Technical Journal* 30:50-64. 1951.

The Shannon Generation Method

- Choose a random bigram

  (<s>, w) according to its probability

- Now choose a random bigram (w, x) according to its probability

- And so on until we choose </s>
- Then string the words together

```
<s> I
    I want
      want to
           to eat
              eat Chinese
                  Chinese food
                           food
  </s>
I want to eat Chinese food
```

Shannon generation: English

- Word Model: Second-Order (bigram)
- THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED T

Claude E. Shannon. *A Mathematical Theory of Communication*, 1948.

The Corpus matters

- What corpus was used to generate these:

Bigram

What means, sir. I confess she? then all sorts, he is trim, captain.

Why dost stand forth thy canopy, forsooth; he is this palpable hit the King Henry. Live king. Follow.

What we, hath got so she that I rest and sent to scold and nature bankrupt, nor the first gentleman?

Trigram

Sweet prince, Falstaff shall die. Harry of Monmouth's grave.

This shall forbid it should be branded, if renown made it empty.

Indeed the duke; and had a very good friend.

Fly, and will rid me these news of price. Therefore the sadness of parting, as they say, 'tis done.

Quadrigram

King Henry.What! I will go seek the traitor Gloucester. Exeunt some of the watch. A great banquet serv'd in;

Will you not tell me who I am?

It cannot be but so.

Indeed the short and the long. Marry, 'tis a noble Lepidus.

Surprisal in Syntax

The horse raced past the barn fell

Bever (1970)

The prime number few

Milne (1982)

gardenpath sentences

Formal Models

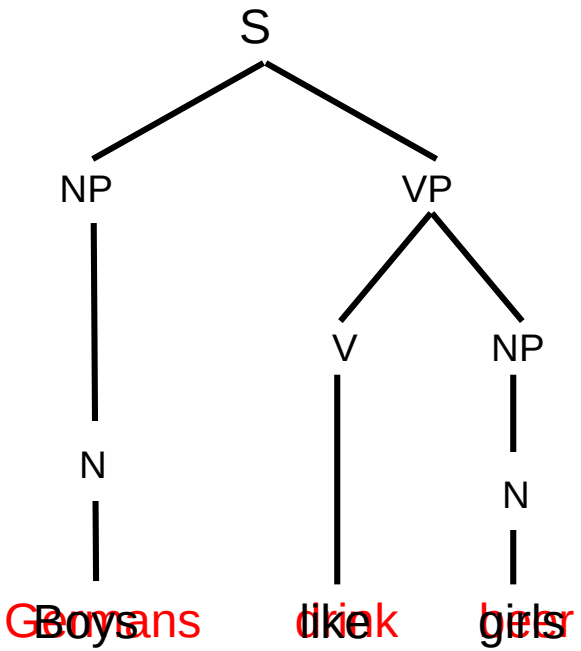
Formal Semantics

- Declarative Sentences: Assign Truth Values
- Non-Declarative: inferential connections
- Interpretation function: Semantics of Words -> composition → semantics for complex expressions
  - Model-Theoretic: Map phrases / words → model
    - [Montague PTQ]
  - Truth-Theoretic: Conditions under which sentence is true. [Tarski, Davidson]

Model Theory

- Montague grammar :
  - Handles FRAGMENT of language
  - Syntax – define expression structure
  - Translation – into logical structure
  - Model-Theory : meanings as sets / individuals (PN) → Denotata
- Modern versions of Montague grammar – avoid “translation”

Syntactic Analysis



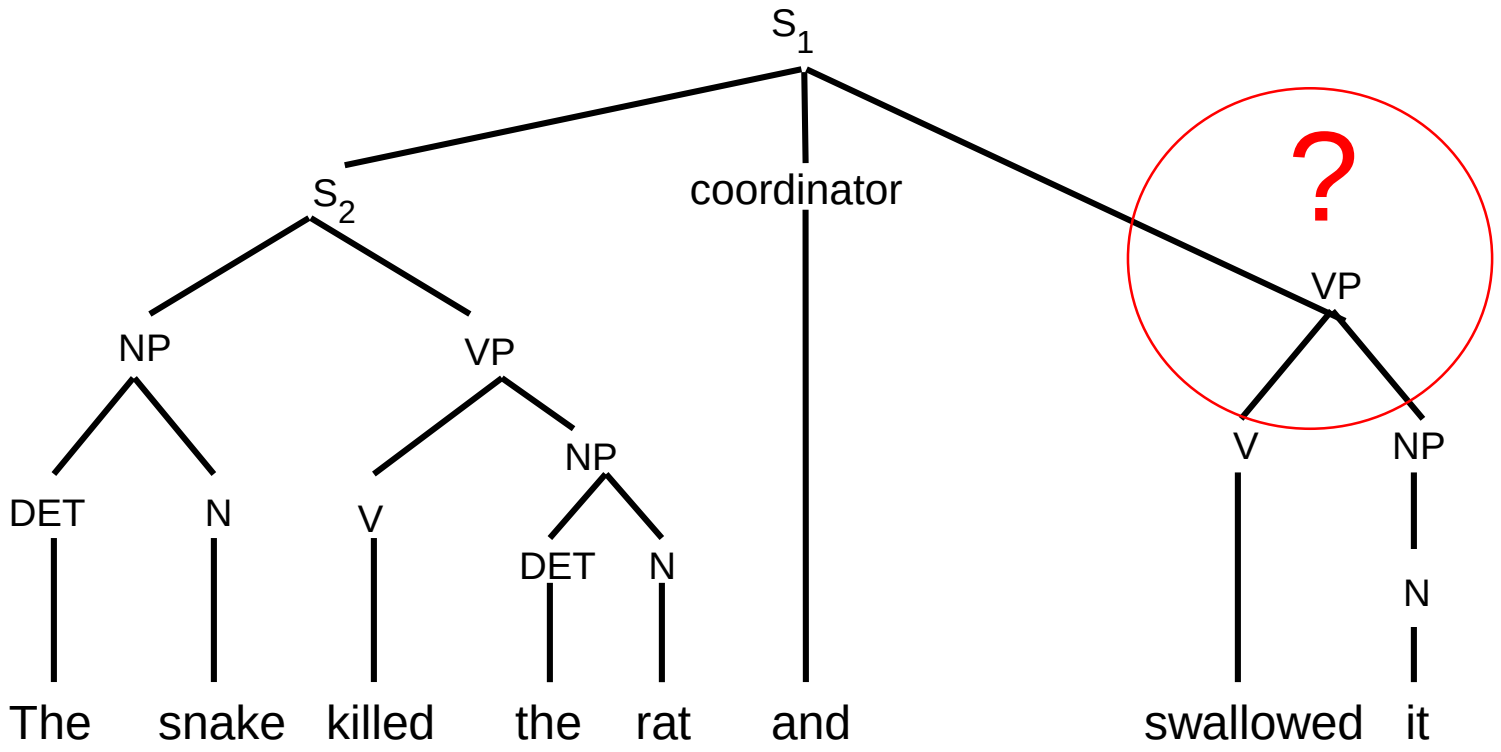
Phrase structure rules

S → NP VP
NP → N
VP → V NP
NP → det N

Lexicon

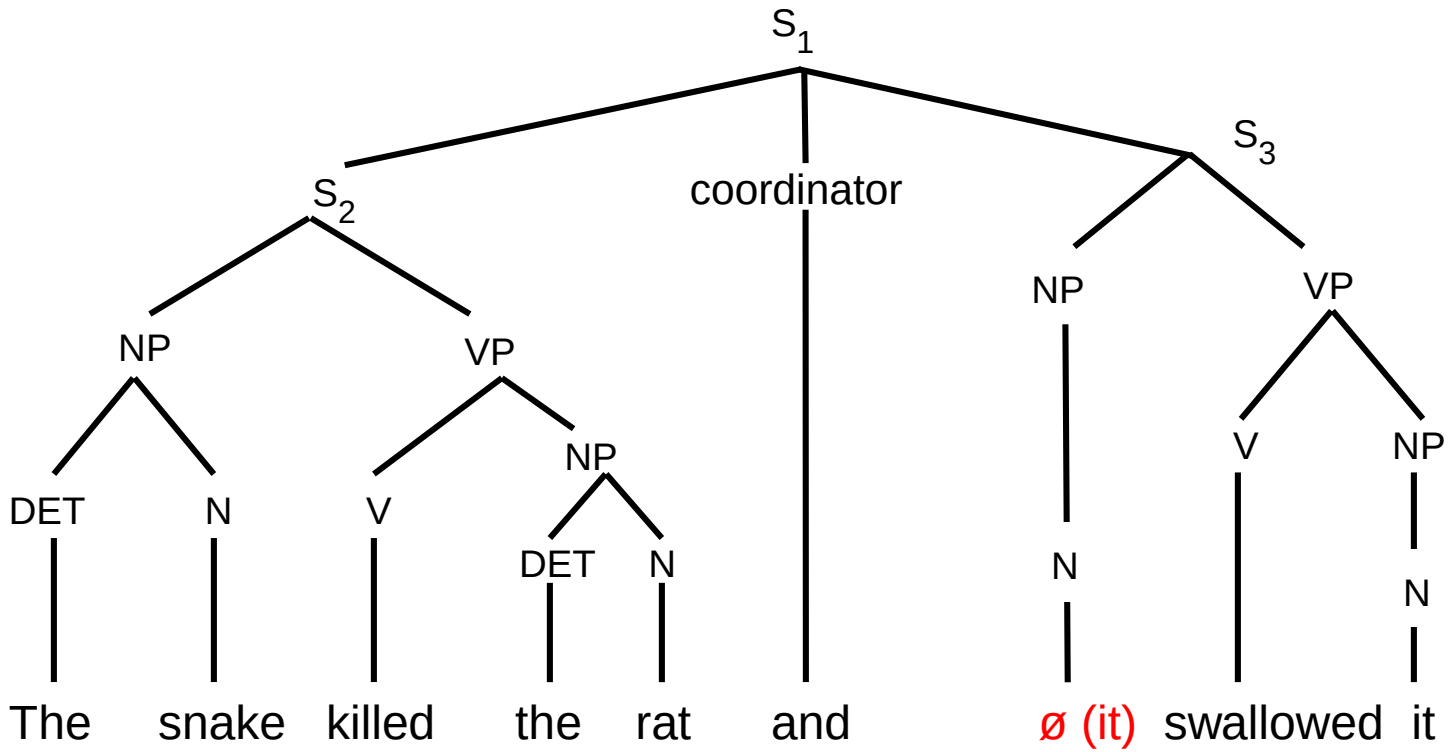
N → german[s], boy[s], girl[s], beer
V → like, drink

Missing Elements?



[haegeman wekker 03] modern course in english syntax

Missing Elements : Ellipsis



ellipsis

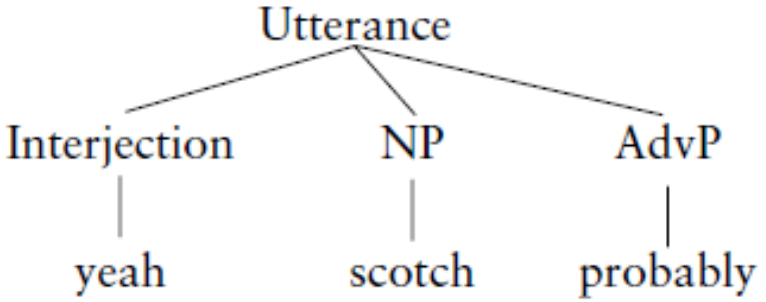
[haegeman wekker 03] modern course in english syntax

Bare argument ellipsis (BAE)

A: I hear Harriet's been drinking again.
B: *Yeah, scotch, probably*

Generative Grammar analysis (ellipsis):
B: *Yeah, [Harriet has been drinking] scotch probably*
[$_{ADVP}$ *Yeah*] [$_{NP}$ e] [$_{VP}$ e *scotch*]] [$_{ADVP}$ *probably*]

Culicover / Jackendoff 02*:*
Accept fragment as is
use semantics / pragmatics
to judge grammaticality



Language and Meaning

Montague Translation [1973]

A student sleeps

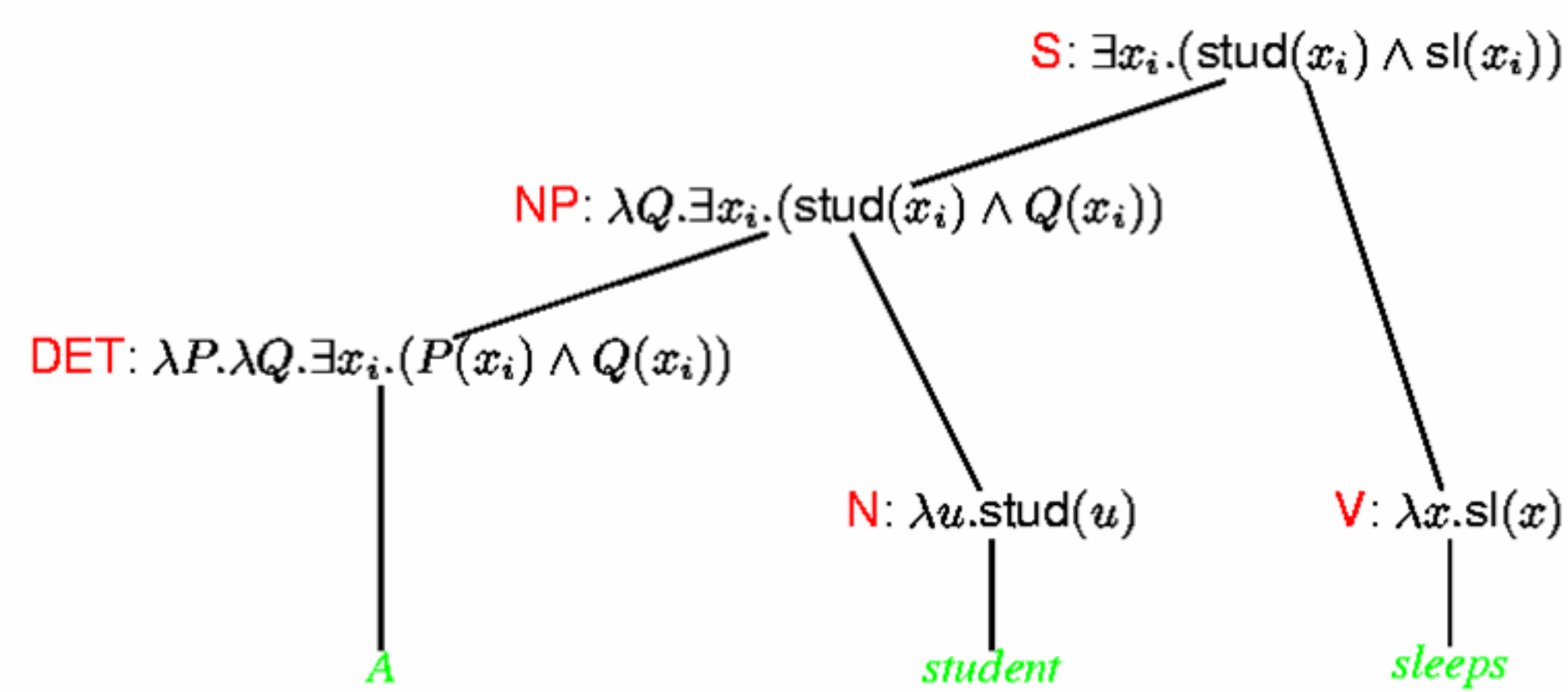
Lexicon:

student, N: $\lambda u.\mathsf{stud}(u)$

sleep, V : $\lambda x.\mathsf{sl}(x)$

a, DET : $\lambda P.\lambda Q.\exists x_i.(P(x_i) \wedge Q(x_i))$

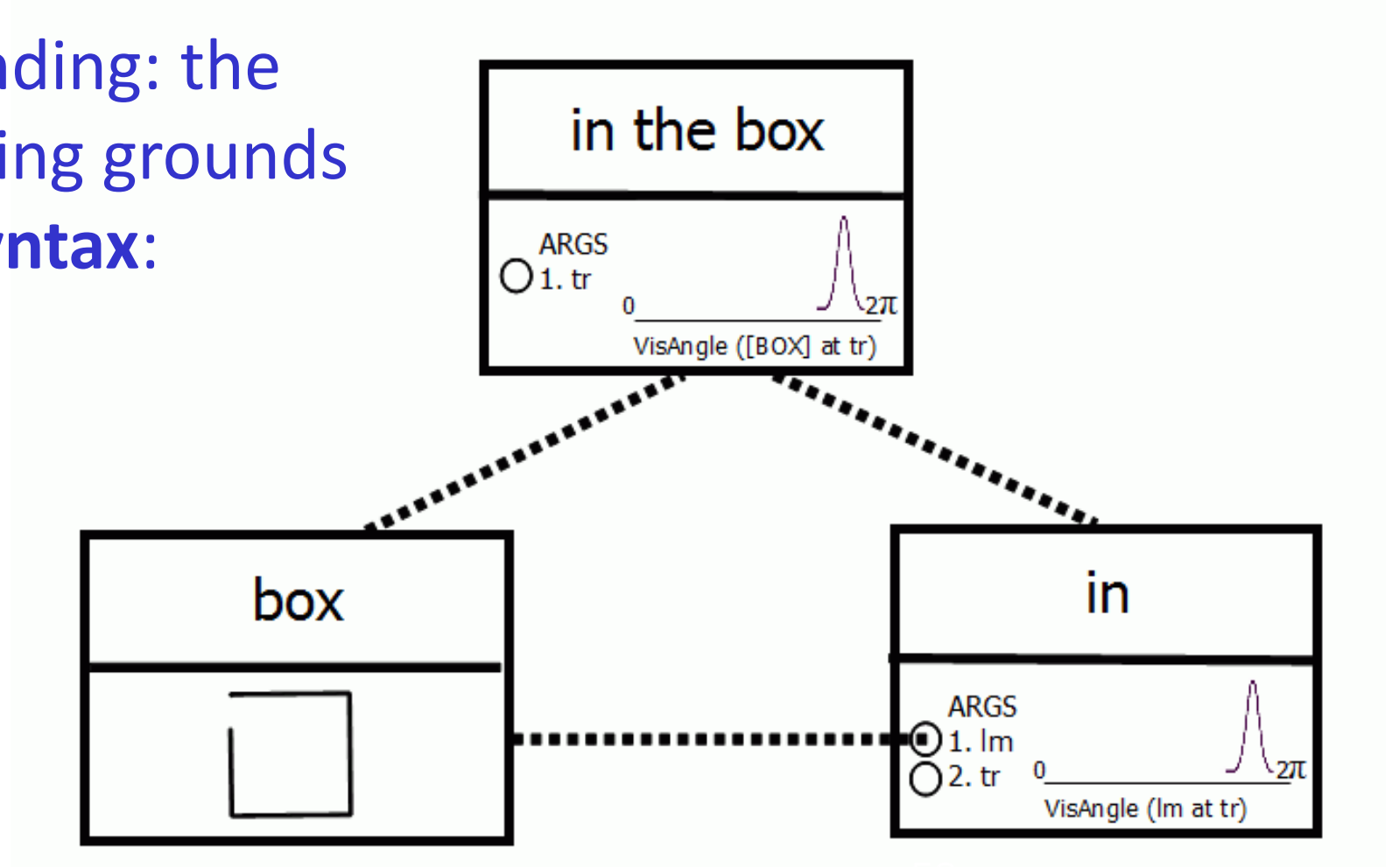
Montagovian Translation [1973]



[Kohlhase]

Cognitive Grammar (Langacker)

- Grounding: the meaning grounds the **syntax**:

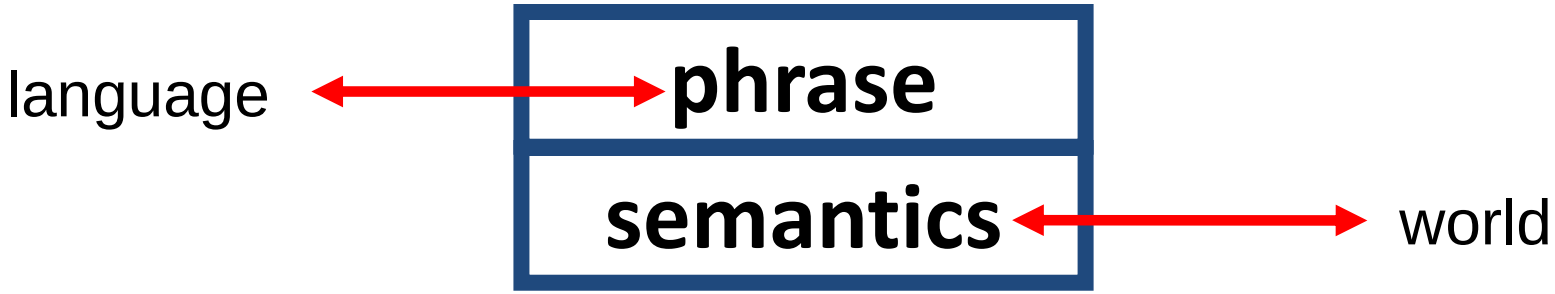


The role of Context

- Charles Morris and Rudolf Carnap: 3-fold division of the theory of language:
  - syntax : relations between expressions
  - semantics: relations between expressions and what they stand for
  - pragmatics: relations between expressions and those who use it
- [Peregrin 98]
  - Internal Challenge (deictic - demonstrative/ anaphora)
  - External Challenge (function rather than designation)

Symbol = Form-Meaning pair

- Symbols = (form) label + **meanings**.



symbol = label + semantics

[langacker 87]

- Semantics : not static: evolves with language use
- *image schema* : map in perceptual space
- Linguistic label acts as index to concept
- Earliest image schemas = pattern on sensory data (chunk)

Grounded Language

- grounded **lexicon**:
  relation between sounds and sensorimotor patterns

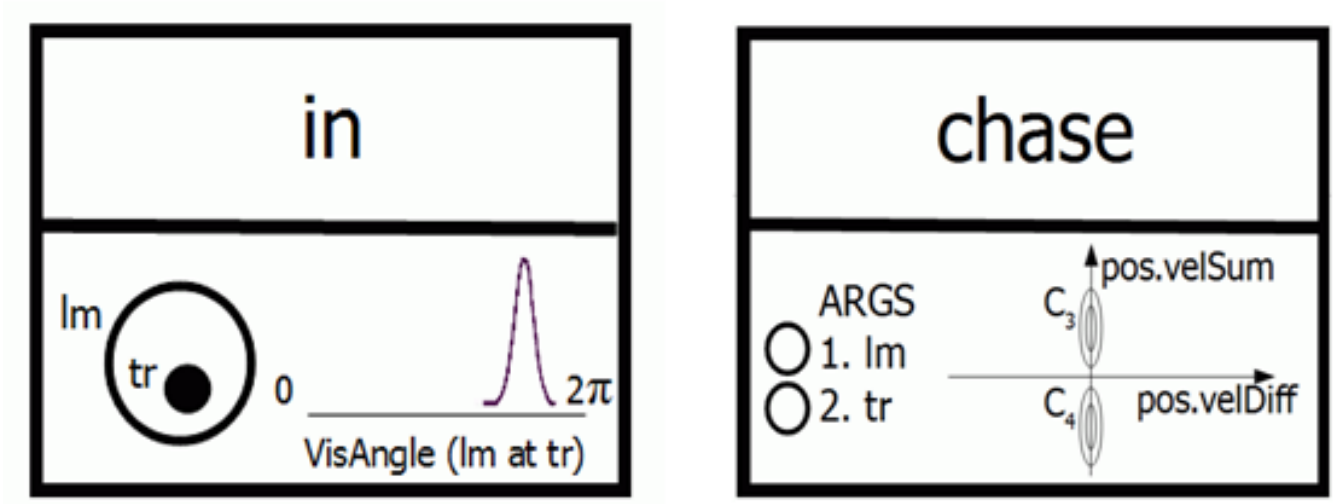
- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Lexicon

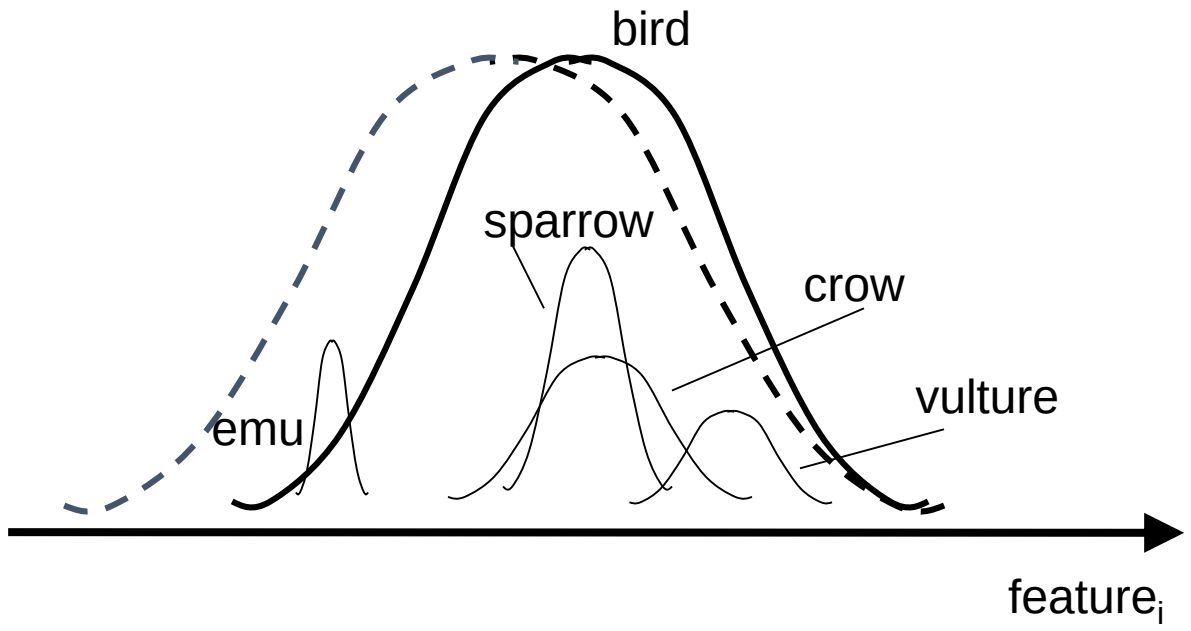
- grounded **lexicon**:



- semantic pole : perceptual patterns (image schemas)
  → probabilistic predicate + arguments

Evolving Semantics

Conceptual Space



Lexicon vs Grammar

- lexicon = mental inventory of units
  = set of all lexemes

- Is “cats” a lexeme?

  cook → cooks : grammatical (rule-driven, inflection)
  → cooker : cook + er (not fully a rule; derivation)

Older thinking : lexicon is separate from grammar
cog L: lexicon - grammar is a continuum

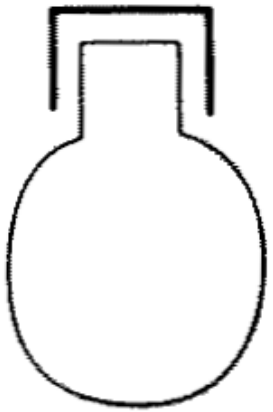
Commitment of Grammar

Cognitive Grammar:

- Try to make sense of
  - polysemy (systematically related linguistic forms),
  - inference,
  - historical change,
  - gesture,
  - language acquisition
  - iconicity in signed languages.

[Lakoff/Johnson p.80]

Try to explain all aspects

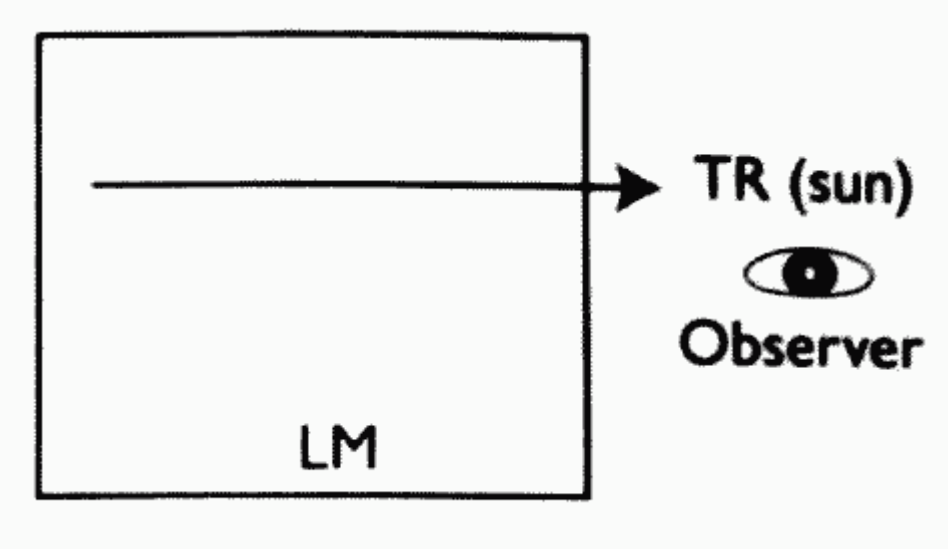


The bulb in the socket

* The jar in the lid

[lee 01] ch.2

Perspective? Idiom?

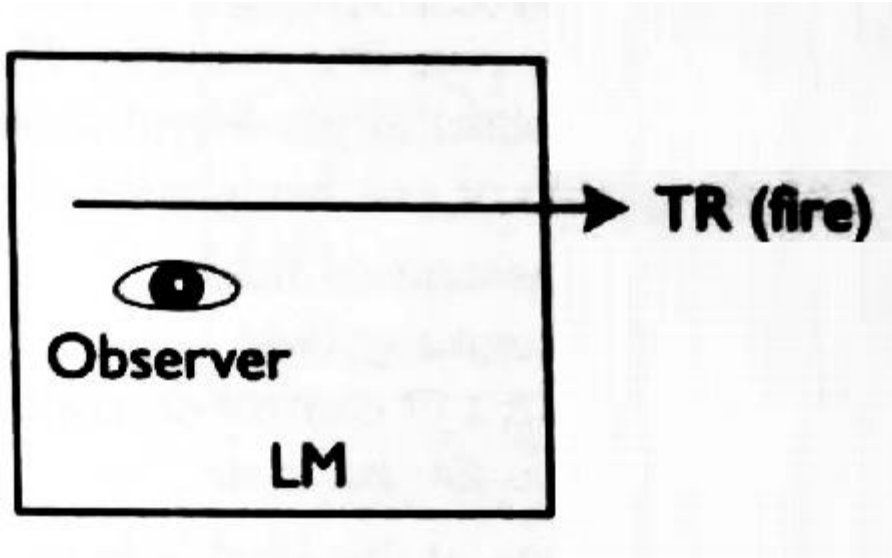


The sun came out.

The secret is out

The fire went out.

The music was
drowned out by noise.



Cognitive Grammar View:

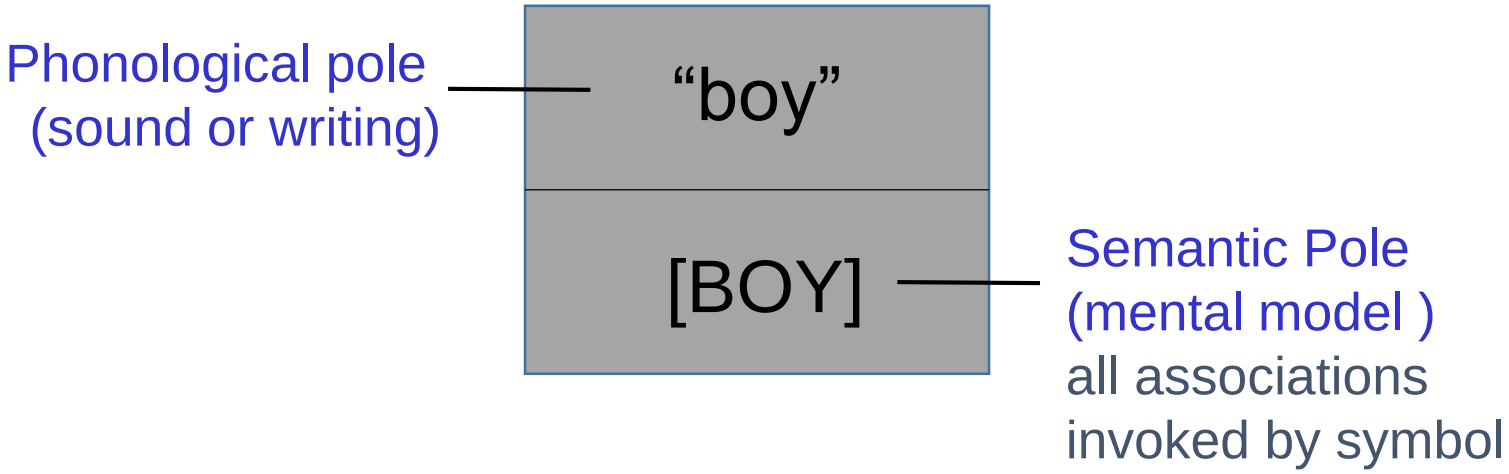
Lexicon vs Grammar

Lexicon / Grammar is a *graded* distinction – more of a continuum than a sharp difference

There are rule-like *schemas*, but they apply in differing degrees for different instances

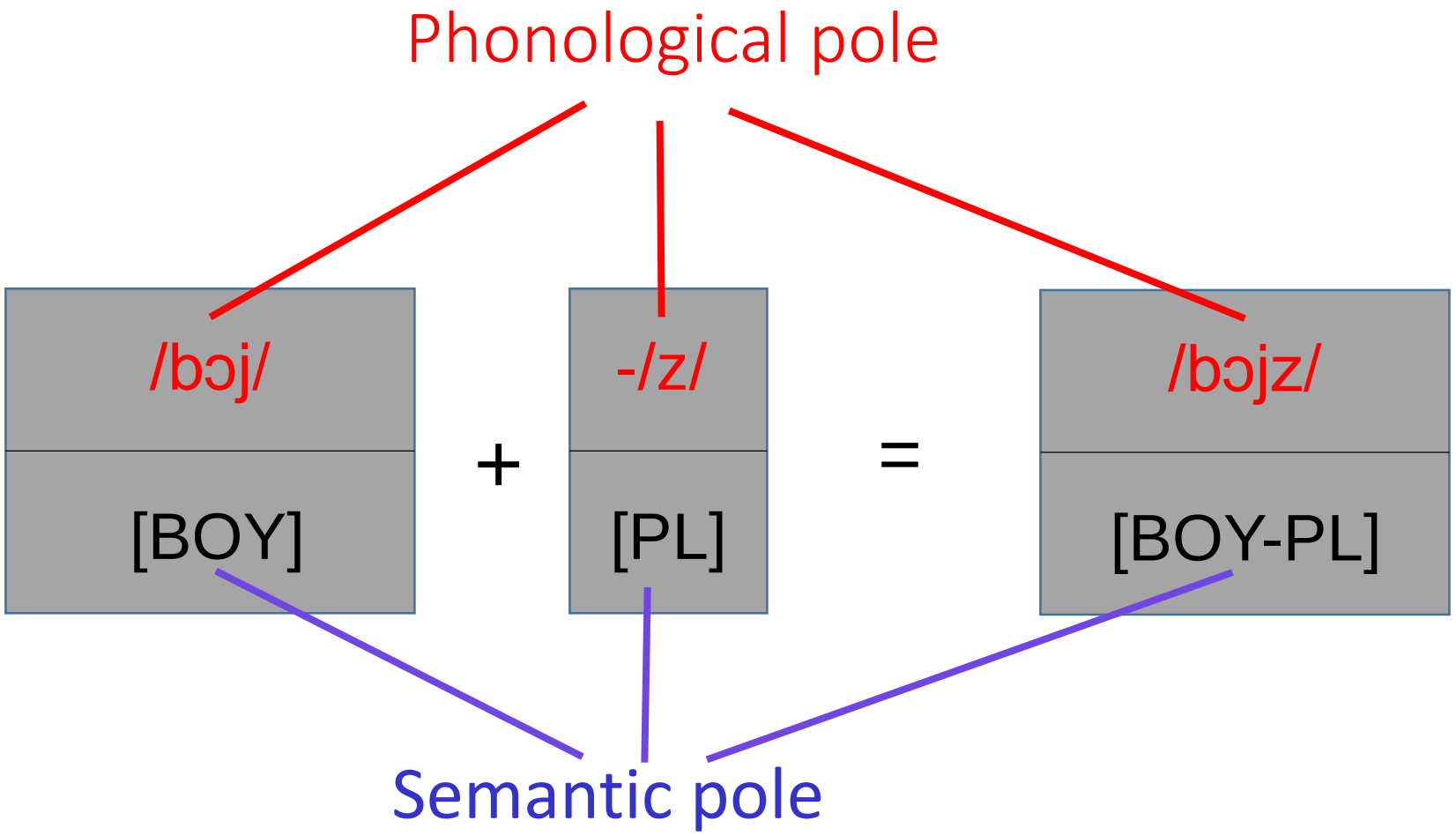
Cognitive Grammar View:

Symbolic Unit

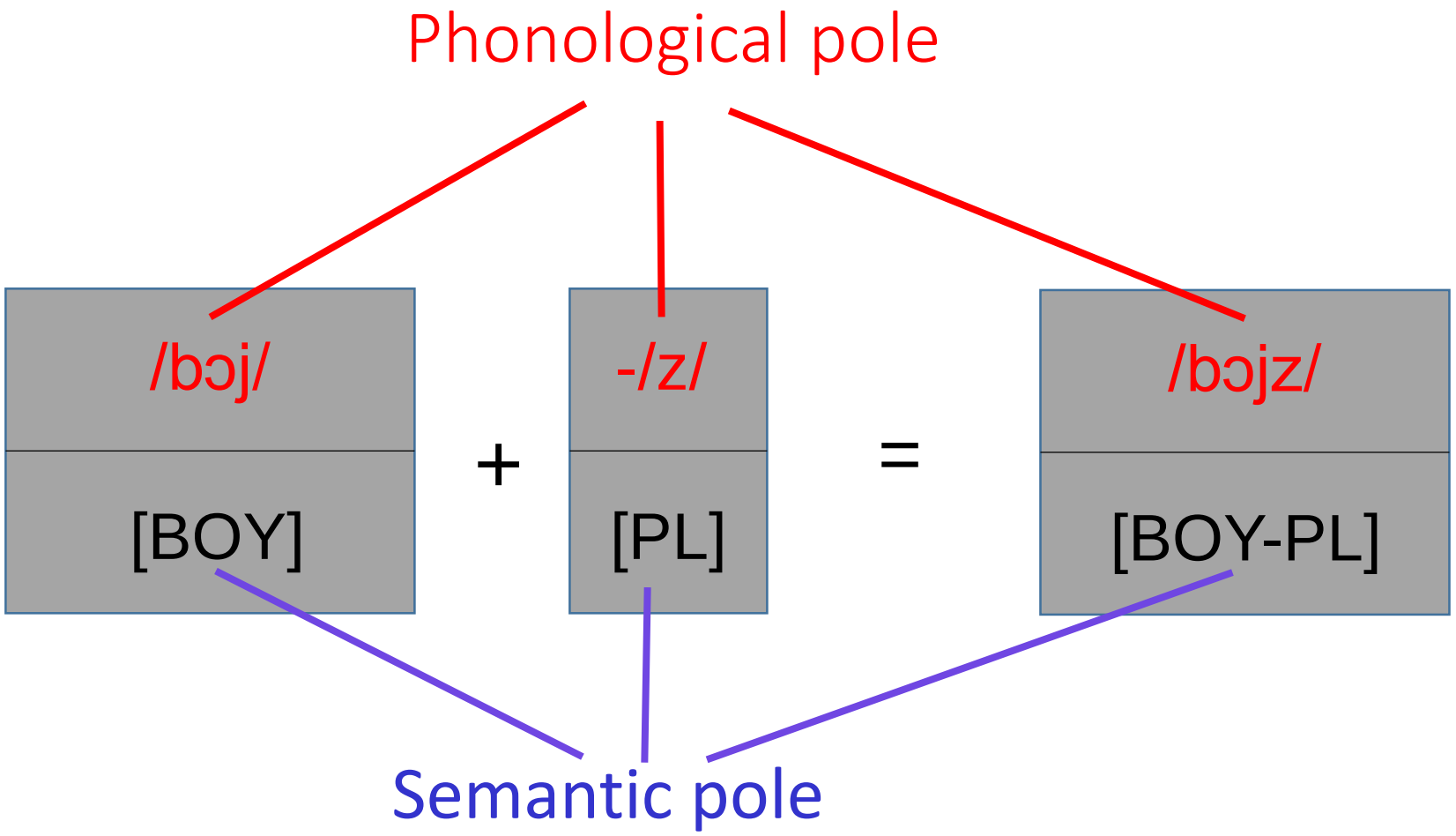


symbol: interrelation between thought, meaning, and linguistic structure

Combining phonemes

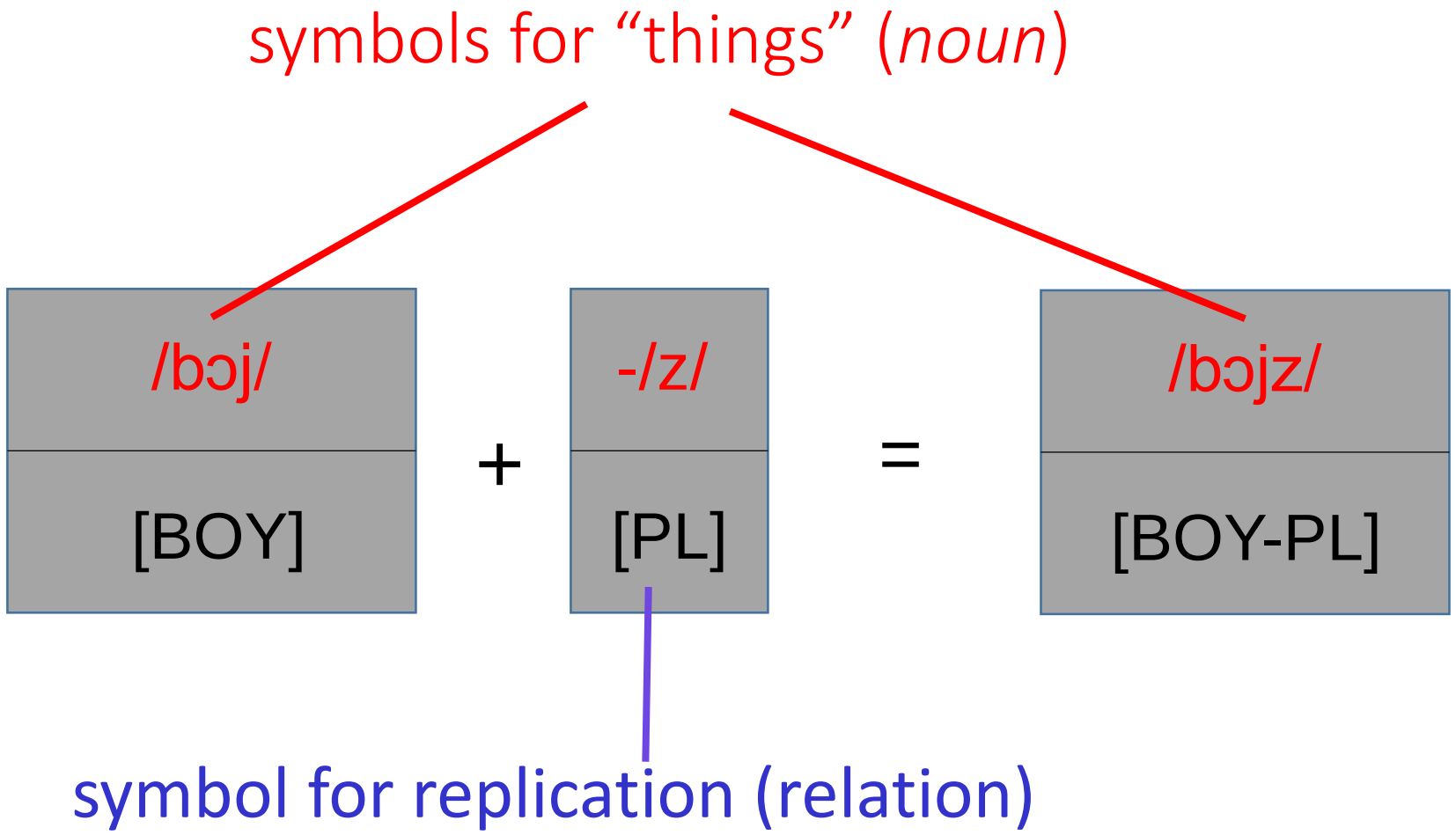


Cognitive Grammar : Inflection



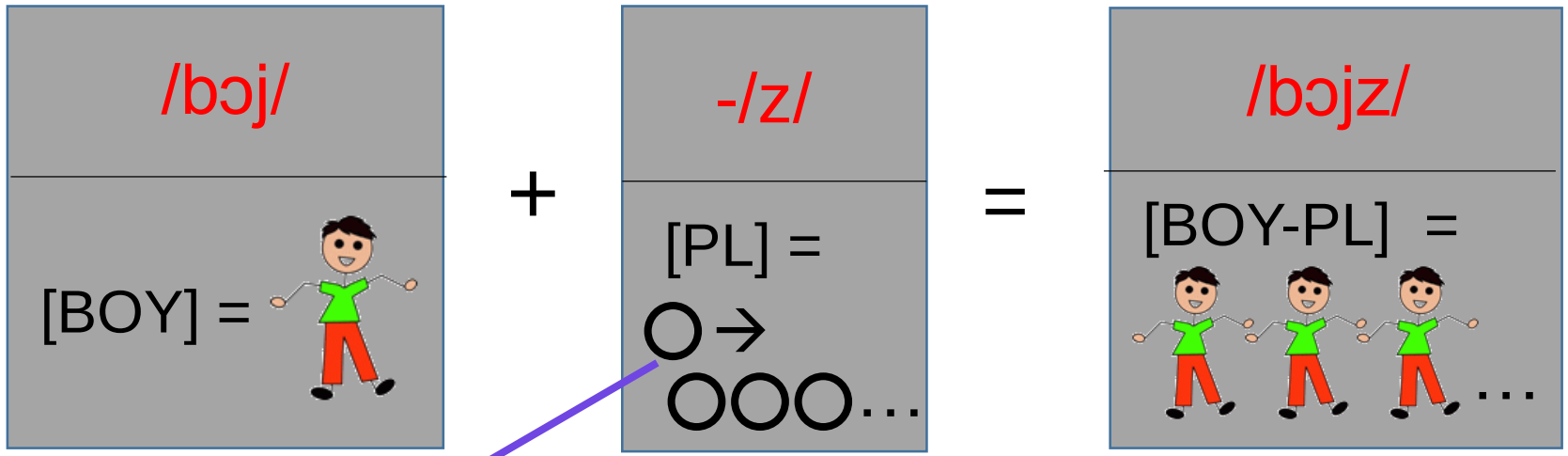
symbolic complex [langacker 87]

Cognitive Grammar : Inflection

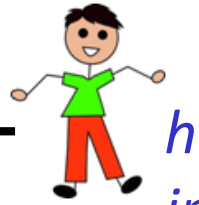


symbolic complex [langacker 87]

Semantic Pole: Image Schemas

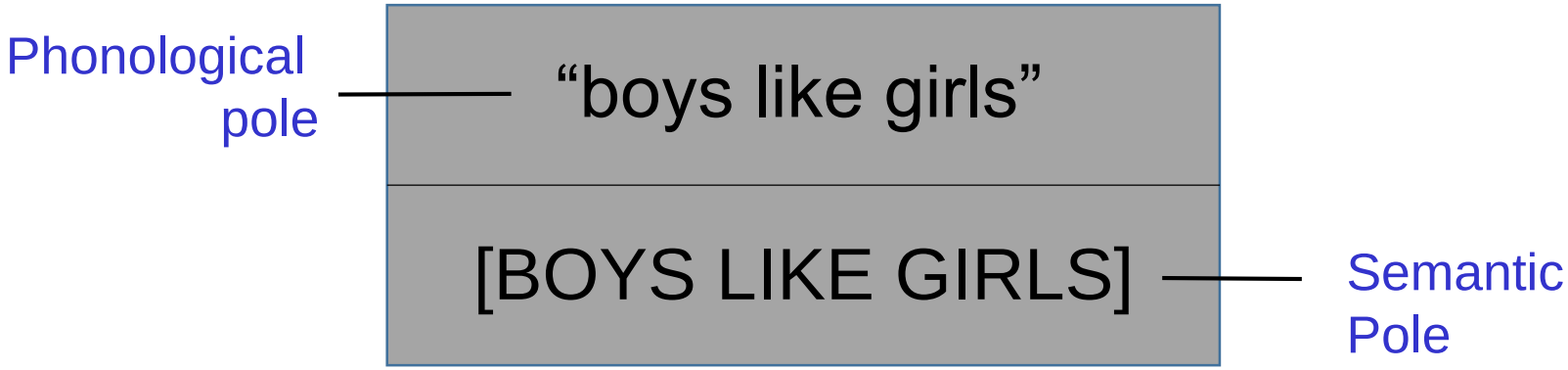


must be countable

O empty slot ←  *has discrete instances*

Countable = has distinct instances. Not a continuum.
another pen; but
more water

Cognitive Grammar View:
All Language is Symbolic

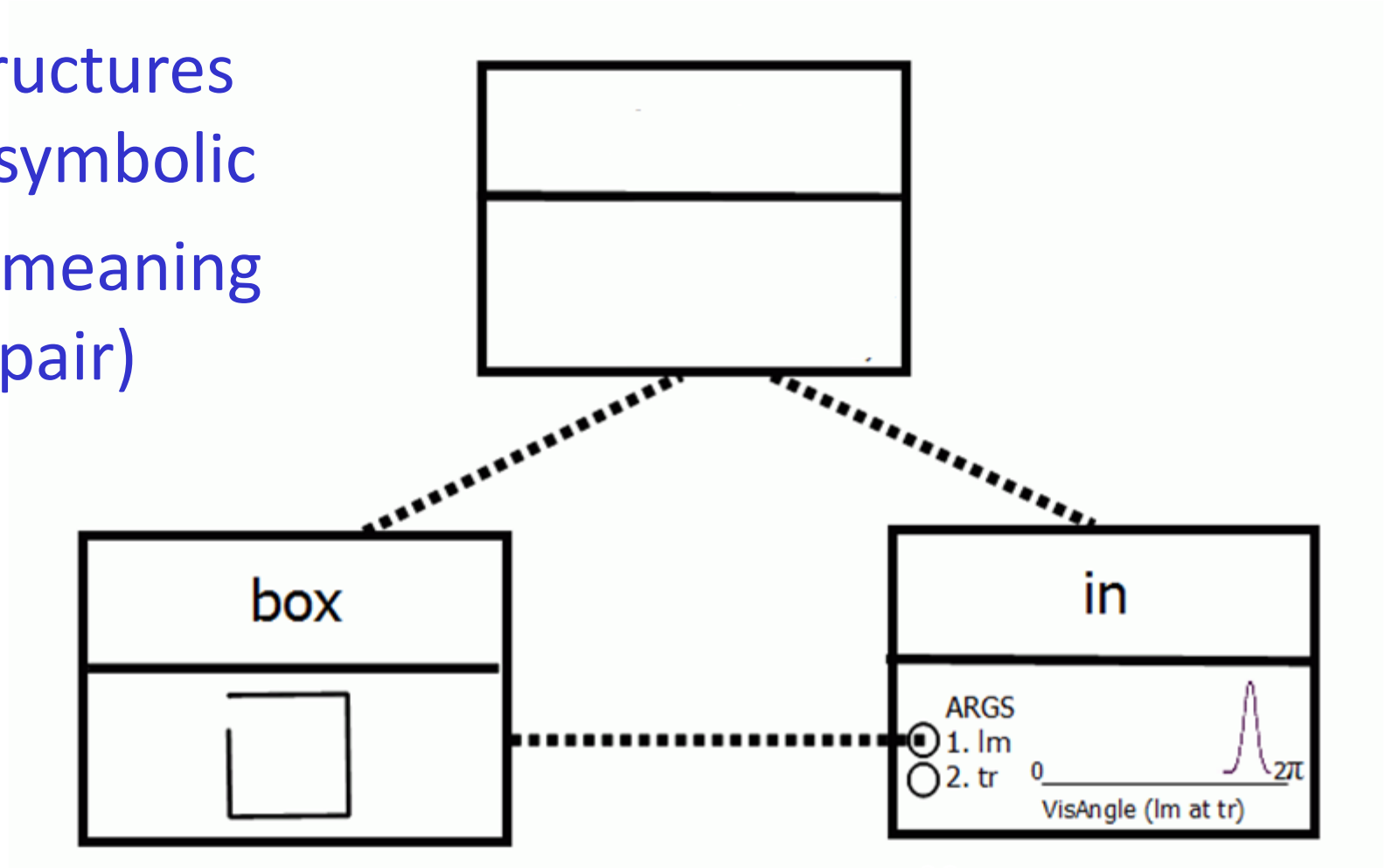


Grammar: applies to the composition of both phonological pole (surface form) and semantic pole (meaning)

Cognitive Grammar (Langacker)

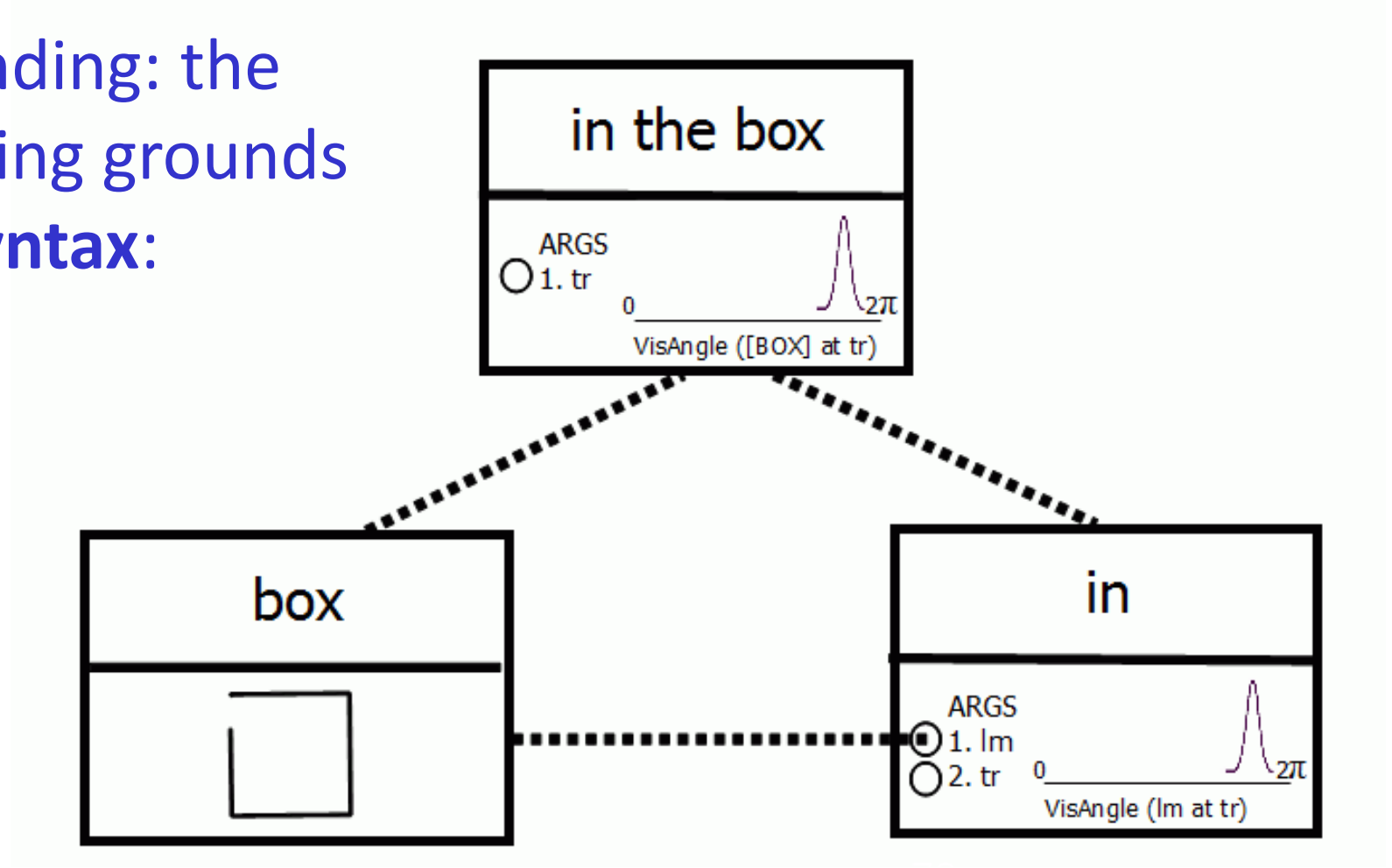
All structures are symbolic

(form-meaning pair)



Cognitive Grammar (Langacker)

- Grounding: the meaning grounds the **syntax**:

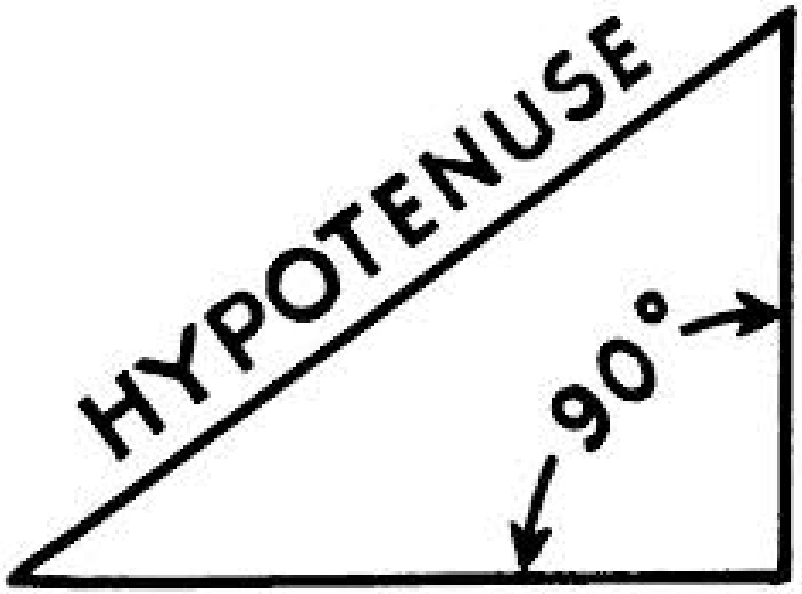


Language is Symbolic

- “boy” = sound (or written form) of language
- [BOY] = all possible mental associations that may be invoked. Meaning is *encyclopedic*

- Selecting from encyclopedic associations
  - *construal*：Constructed against a background or *frame*
  - takes a particular *perspective*
  - *subjective*: Differences owing to individual experiences and goals.
  - *relativism*: Language Structures can influence other parts of cognition

Frame (background knowledge)



“hypotenuse”
[hypotenuse]

[hypotenuse]: frame = right-angled triangles

The side opposite the right-angle is foregrounded or profiled

Frame (background knowledge)

“wicket ”
[wicket]

Semantic Pole
can be understood only
with a background [frame]
of cricket knowledge

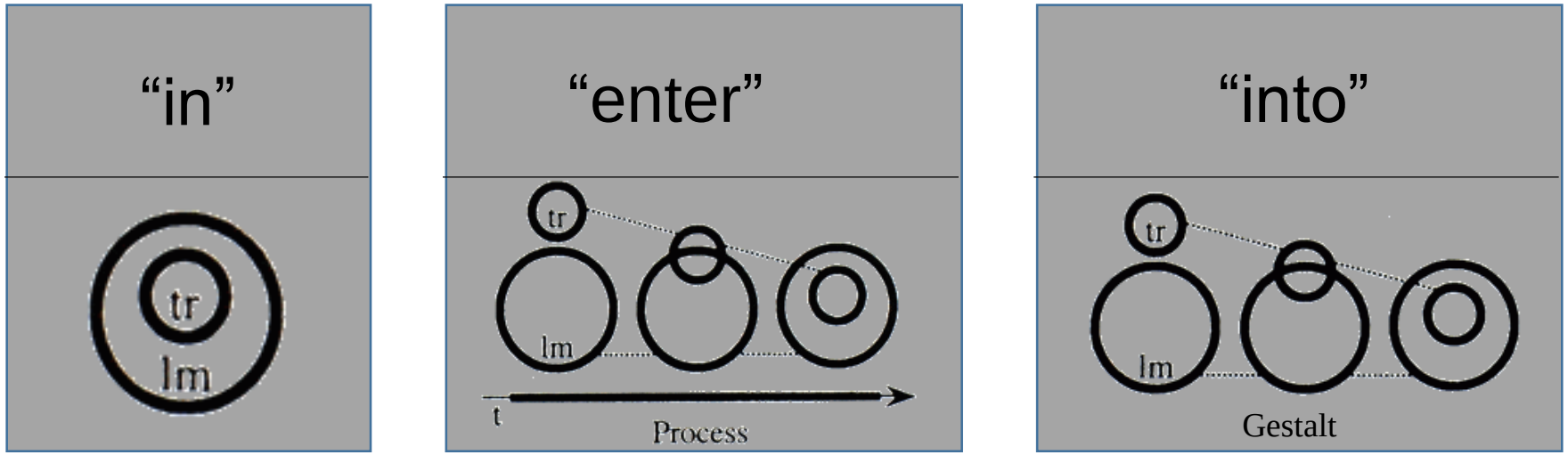
[wicket]: frame =
game of cricket

Language is Symbolic

- “boy” = sound (or written form) of language
- [BOY] = all possible mental associations that may be invoked. Meaning is *encyclopedic*

- Selecting from encyclopedic associations
  - *construal*：Constructed against a background or *frame*
  - takes a particular *perspective*
  - *subjective*: Differences owing to individual experiences and goals.
  - *relativism*: Language Structures can influence other parts of cognition

Semantics as Image Schema



- Image schemas differ in what is foregrounded
- Process view: Time is part of the frame
- Non-Processual : no temporality
  - Simplex: Just a state (e.g. IN)
  - Complex: Summary or Gestalt (whole) of an aggregate (e.g. Temporal or Spatial)

Langacker 1987/2008

Grounded Language

- grounded **lexicon**:
  relation between sounds and sensorimotor patterns

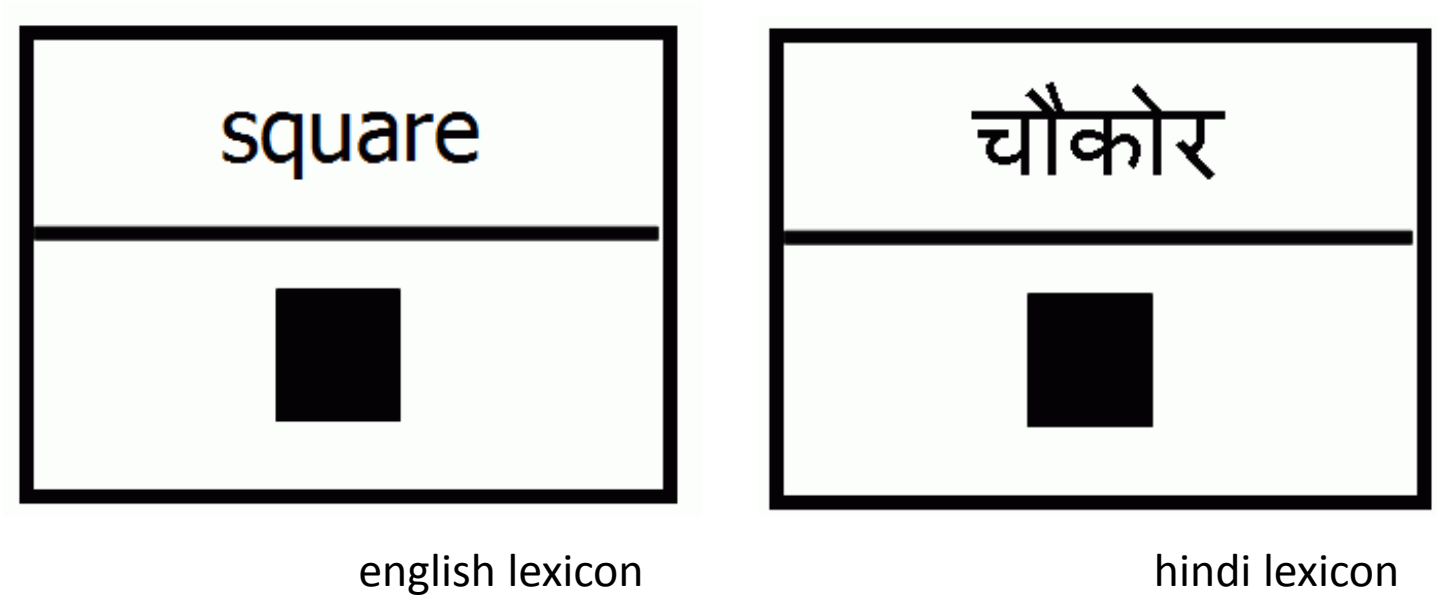
- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Lexicon

- grounded **lexicon**:



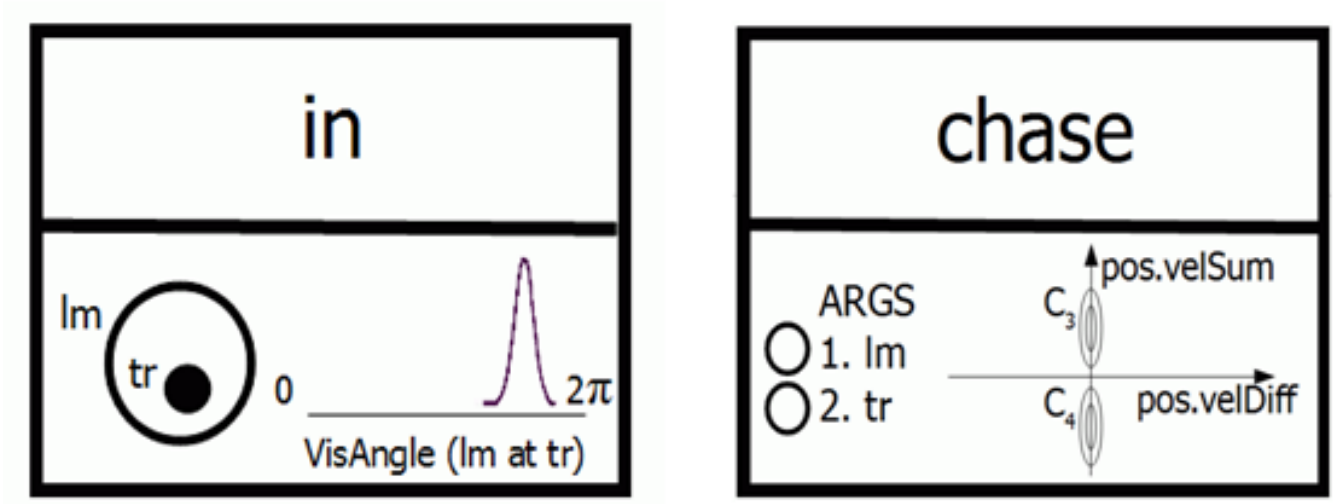
english lexicon

hindi lexicon

[langacker 87]

Lexicon

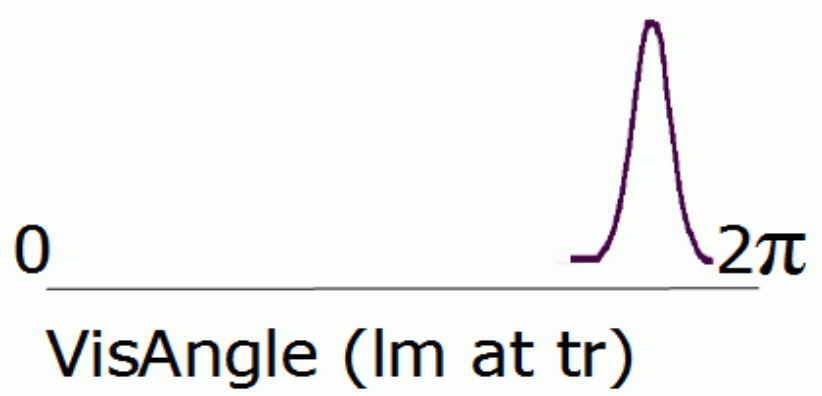
- grounded **lexicon**:



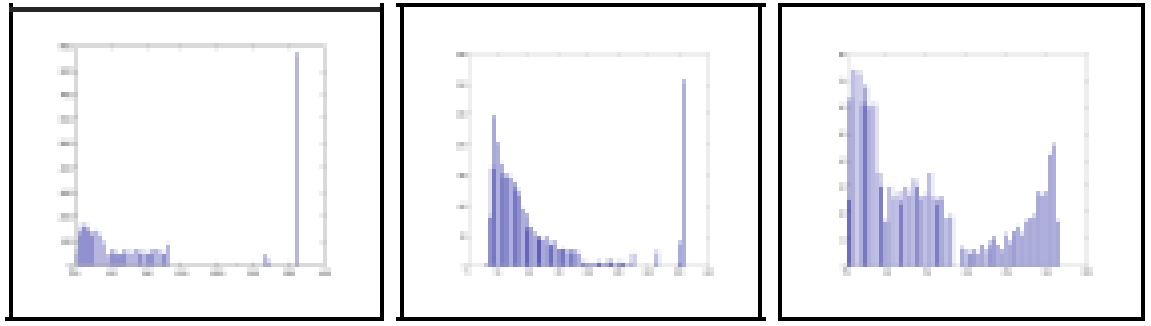
- semantic pole : perceptual patterns (image schemas)
  → probabilistic predicate + arguments

Clustering spatial relations

IN cluster
(emergent)



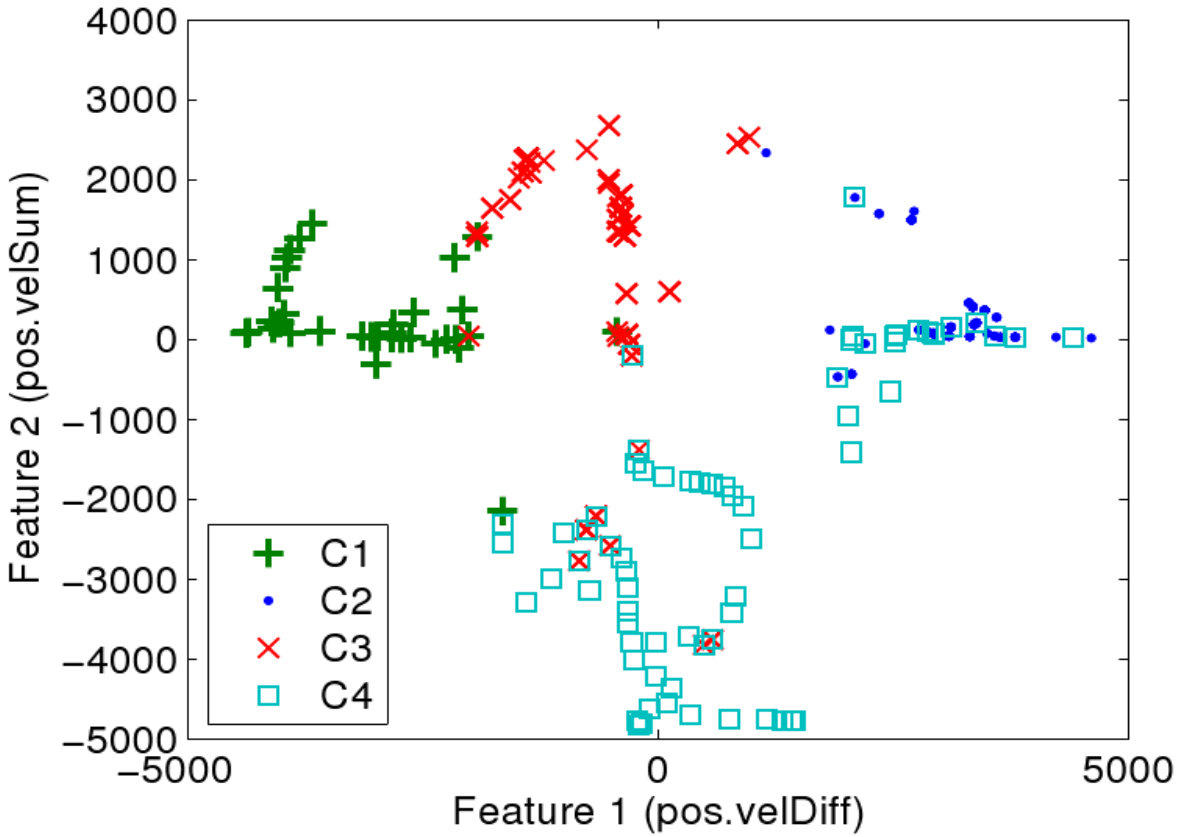
Histogram of visual subtended angle for the 3 shapes



[Sarkar/Mukerjee 07; Nayak/Mukerjee 12]

Perceptual Discovery: 2-agent actions

- Static time-shots of feature space trajectories



Similarity : Word vector models

"**fast**" is similar to "**rapid**"

"**tall**" is similar to "**height**"

Question answering:

*Q: "How **tall** is Mt. Everest?"*
*Candidate A: "The official **height** of Mount Everest is 29029 feet"*

Rule-based Syntax

Syntax-driven view of Language

- *Compositionality* Assumption: Larger phrases built up from smaller ones

- Construct rules for how words compose into phrases and sentences = Grammar
  - may also apply to morphemes

- Map to semantics:
  - Assumption: words have meaning
  - Syntax : Composes words into new composite meaning

Why is Syntax Important?

- Grammar checkers
- Question answering
- Word sense Disambiguation

- Information retrieval (?)
- Machine translation

- Map to semantics

Theories of Syntax?

- Unfortunately, no consensus on a theory of grammar - aggressive debates :
  - Chomskyan – formalist, autonomous from semantics, we are born with syntax
  - Cognitive linguistics – semantics has a role, language is learned by discovering patterns in usage
  - Computational : Use what works

Syntax : Composability

- Are sentences constructed by combining words? [decomposability]
- Or are words obtained by breaking up sentences? [holism]
- At least some times, while learning a language, babies understand the sentence *before* the words

Chomskyan (Generative) view

- Syntax is independent of meaning.
  Perception, action, etc. are not relevant to grammar

- Of course, language is compositional

- Lexicon = list of words → *arbitrary*

- Syntax: Words are composed via deterministic, formal rules → *systematic*

Chomskyan Language Acquisition

- Babies acquire language with very little guidance. (Poverty of Stimulus)
- Possible only if we have an innate *Language Faculty* with a built-in *Universal Grammar* (Nativism)
- Language learning = filling language-specific parameters in the UG

Autonomous Syntax

- Are grammaticality judgments based on form alone?

colourless green ideas sleep furiously

vs

furiously sleep ideas green colorless

→ ***autonomy of syntax*** argument

[chomsky 57]: syntactic structures

Autonomous Syntax : Assumptions

- Rules determining the syntax (form) of language are formulated without reference to meaning, or language use.

- Related : **Grammar is not statistical**

  “There appears to be no particular relation between statistical relations and grammaticalness” p.17

[chomsky 57]: syntactic structures

see P. Norvig: On Chomsky and the Two Cultures of Statistical Learning [http://norvig.com/chomsky.html]

Ambiguity : Newspaper headlines

- Ban on Nude Dancing on Governor's Desk
- Kids Make Nutritious Snacks
- Iraqi Head Seeks Arms
- Juvenile Court to Try Shooting Defendant
- Stolen Painting Found by Tree
- Local High School Dropouts Cut in Half
- Red Tape Holds Up New Bridges

Semantic Lexicons

Frame Elements for frame Ingestion

Frame Elements	Type
Degree	Peripheral
Ingestibles	Core
Ingestor	Core
Instrument	Peripheral
Manner	Peripheral
Means	Peripheral
Place	Peripheral
Source	Peripheral
Time	Peripheral

Lexical Units in : *Ingestion*

Lexical Units for Ingestion

English	Hindi	Bangla
breakfast.v	नाश्ता	prAtarAsh v
Consume.v	भोग करना	bhog k.v
drink.v	पी	khA.v
eat.v	खा	khA.v
feast.v	भोज करना	bhoj k .v
feed.v	खिला	khAoyA.v
gulp.v	निगल	gelA.v
have.v	ले	Neo.v
munch.v	चबा	chebA.v
nibble.v	कुतर	ThokrA.v
sip.n	घूँट	chumuk.n
sip.v	घूँट लेना	Chumuk de.v

Distributional LANGUAGE MODELS

Distributional models of meaning
= vector-space models of meaning
= word vector models

Intuitions: Zellig Harris (1954):

- "oculist and eye-doctor … occur in almost the same environments"
- "If A and B have almost identical environments we say that they are synonyms."

Firth (1957):

- "You shall know a word by the company it keeps!"

Intuition of distributional word similarity

- Sentences in Corpus. Q. what is ***tesgüino***?

```
A bottle of tesgüino is on the table
Everybody likes tesgüino
Tesgüino makes you drunk
We make tesgüino out of corn.
```

- From context words humans can guess ***tesgüino*** means
- Intuition for algorithm:
  - **Two words are similar if they have similar word contexts.**

Distributional Hypothesis

- Bhartrihari (6th c.) : Words by themselves may have no meaning –
  meaning = contexts of use (holism)

- Wittgenstein (1953): The meaning of a word is its usage in language

- J. R. Firth (1957) : Word is known by the company it keeps (*Modes of Meaning, 1965*)

- Word meaning= set of contexts in which it may be used.

POS Tagging : Problematic

Inter-annotator disagreement on Penn Treebank: 7.2%.

Disagreement on correcting the output of an automatic tagger 4.1% (3.5% if one text omitted)

Best POS Taggers:

- 97.3% token accuracy
- Sentence level accuracy = 56%

- e.g. “Marketing” - can be a Noun, Verb, or Adjective

Word and Phrase categories

Part of Speech categories:

Debates over how many POS tags are needed.

“Marketing” - can be a Noun, Verb, or Adjective

Using syntax to define a word's context

- Zellig Harris (1968)

  "The meaning of entities, and the meaning of grammatical relations among them, is related to the restriction of combinations of these entities relative to other entities"

- **Two words are similar if they have similar syntactic contexts**

Duty and **responsibility** have similar syntactic distribution:

Modified by adjectives	additional, administrative, assumed, collective, congressional, constitutional …
Objects of verbs	assert, assign, assume, attend to, avoid, become, breach..

The word-word or word-context matrix

- A word is now defined by a vector over counts of *context* words
- Each vector is big - of length |V|

- The word-word matrix is |V|x|V| - huge but sparse

Word-Word co-occurrence matrix

sample context : ± 7 words

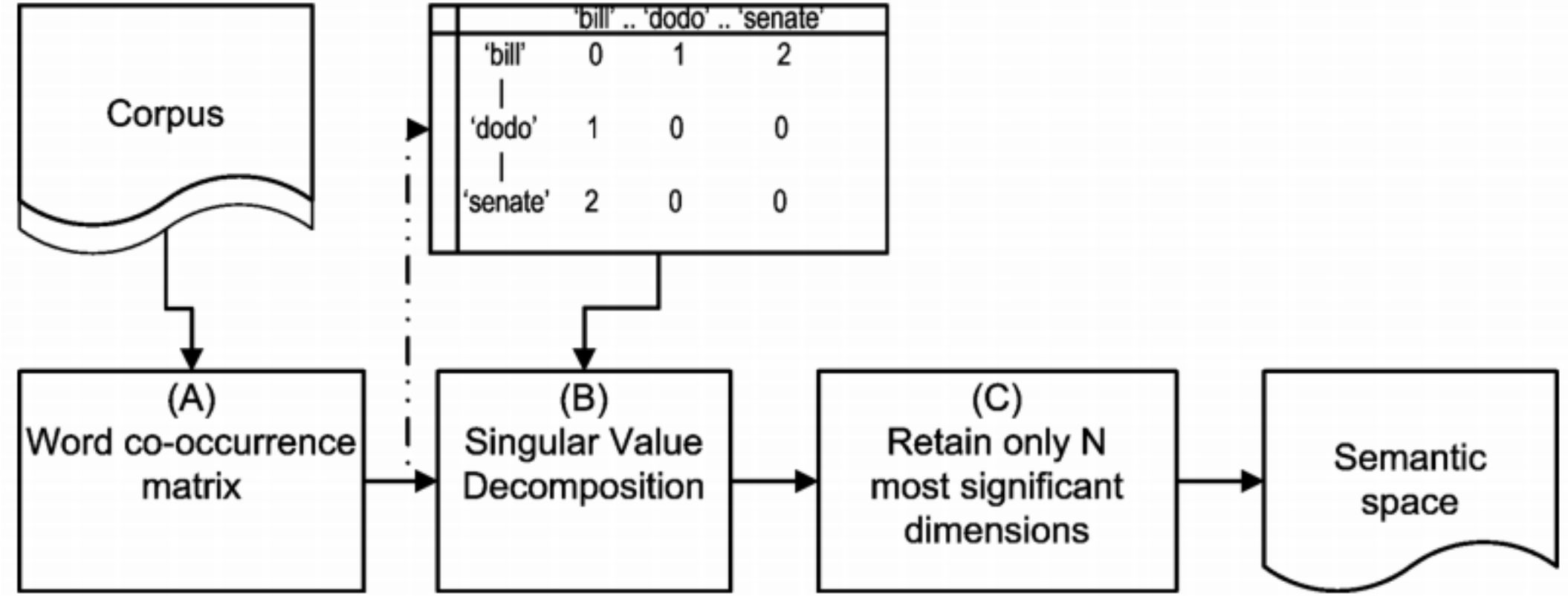
sugar, a sliced lemon, a tablespoonful of **apricot** preserve or jam, a pinch each of,
their enjoyment. Cautiously she sampled her first **pineapple** and another fruit whose taste she likened
well suited to programming on the digital **computer**. In finding the optimal R-stage policy from
for the purpose of gathering data and **information** necessary for the study authorized in the

	aardvark	computer	data	pinch	result	sugar	…
apricot	0	0	0	1	0	1	
pineapple	0	0	0	1	0	1	
digital	0	2	1	0	1	0	
information	0	1	6	0	4	0	
…	…						

Problem with raw counts

- Raw word frequency – very skewed : “the” “of” etc. are very frequent, but maybe not very discriminative
- Emphasize context words that are more **informative** about the target word.
  - tf-idf : Term-frequency * inverse-document frequency
  - PPMI : Positive Pointwise Mutual Information
  - T-test : t statistic of difference between means

Word Vectors : WORDSPACE



sagi-diermeier-13_identifying-issue-frames-in-text

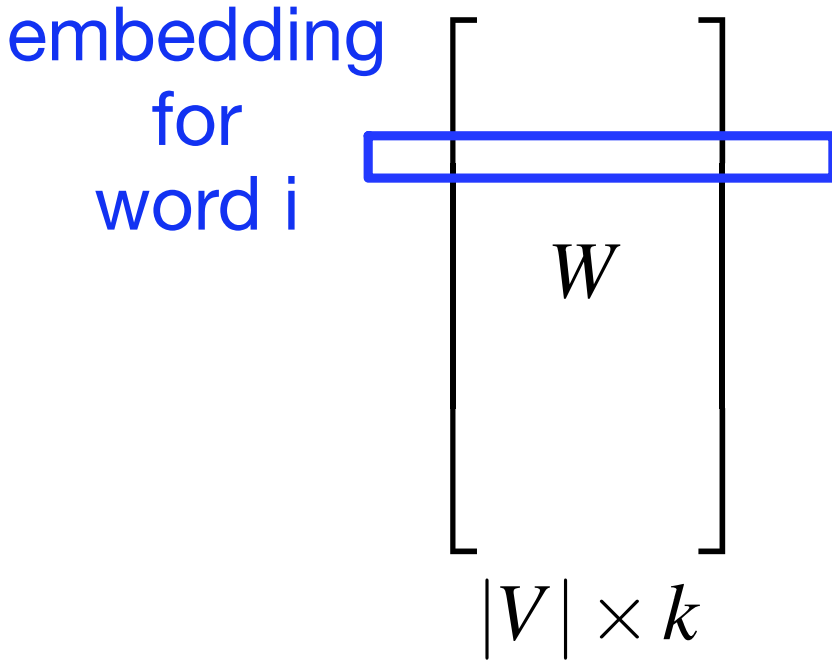
Truncated SVD on co-occurrence matrix

$$\underset{|V| \times |V|}{\begin{bmatrix} \\ X \\ \\ \\ \end{bmatrix}} = \underset{|V| \times k}{\begin{bmatrix} \\ W \\ \\ \\ \end{bmatrix}} \underset{k \times k}{\begin{bmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_k \end{bmatrix}} \underset{k \times |V|}{\begin{bmatrix} \\ C \\ \\ \end{bmatrix}}$$

word vectors (length k)

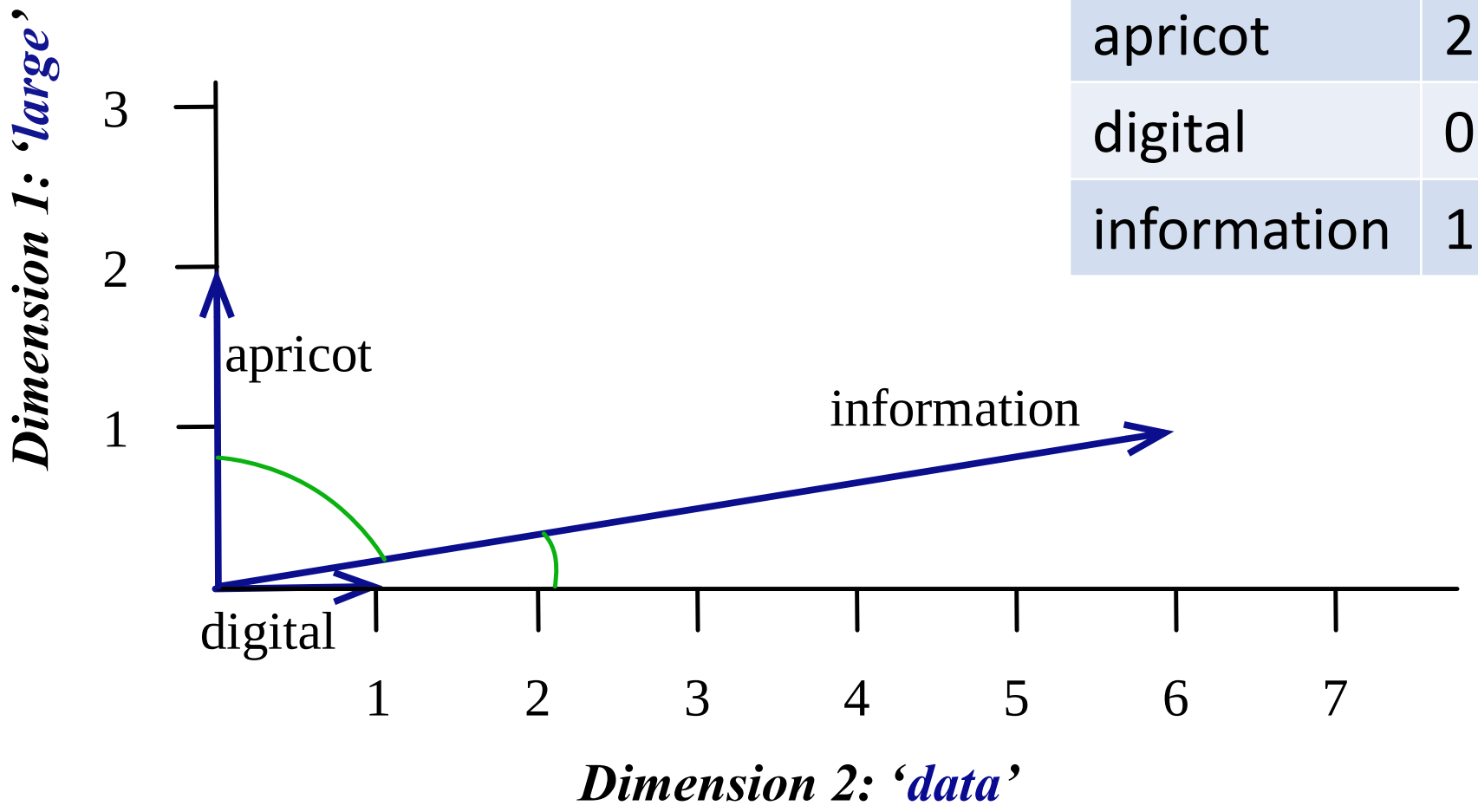
Truncated SVD produces embeddings

- Word vectors = each row of W matrix
- k ~= from 50 to 1000
- Generally we keep the top k dimensions, but some experiments suggest that getting rid of the top 1 dimension or even the top 50 dimensions is helpful (Lapesa and Evert 2014).
- Pre-weighting that sparsifies vectors is helpful [ ]



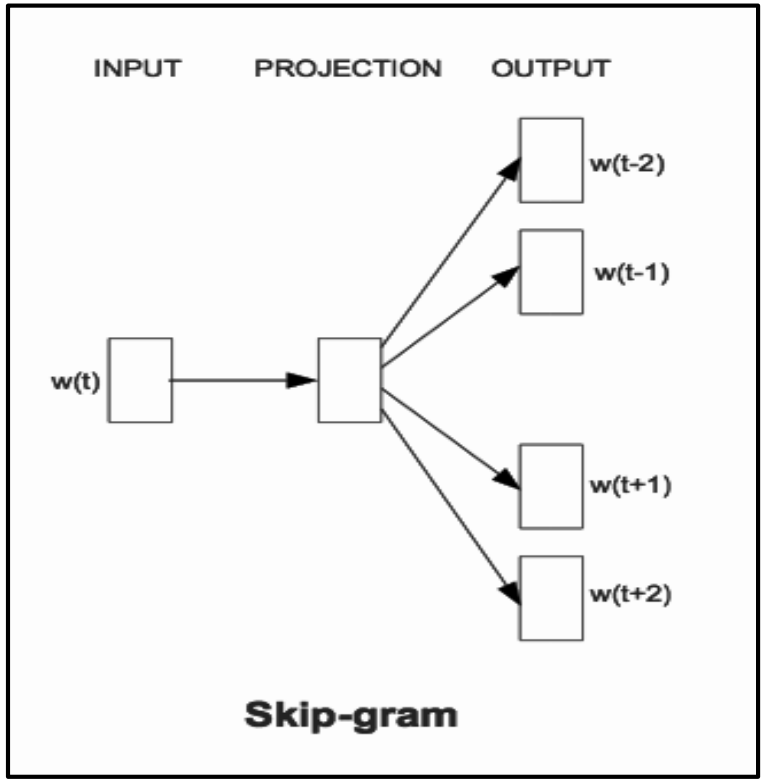
Similarity metric : cosine

	large	data
apricot	2	0
digital	0	1
information	1	6



Skip-gram Model [Mikolov 13]

- Single “projection” layer; no other hidden layers
- Projection layer shared for all words
- All words get projected into the same position (vectors are averaged).
- Skip-gram : Given w in a phrase, attempt to predict left and right context (k words each) from projection layer.
- Efficient: Softmax replaced by Hierarchical softmax

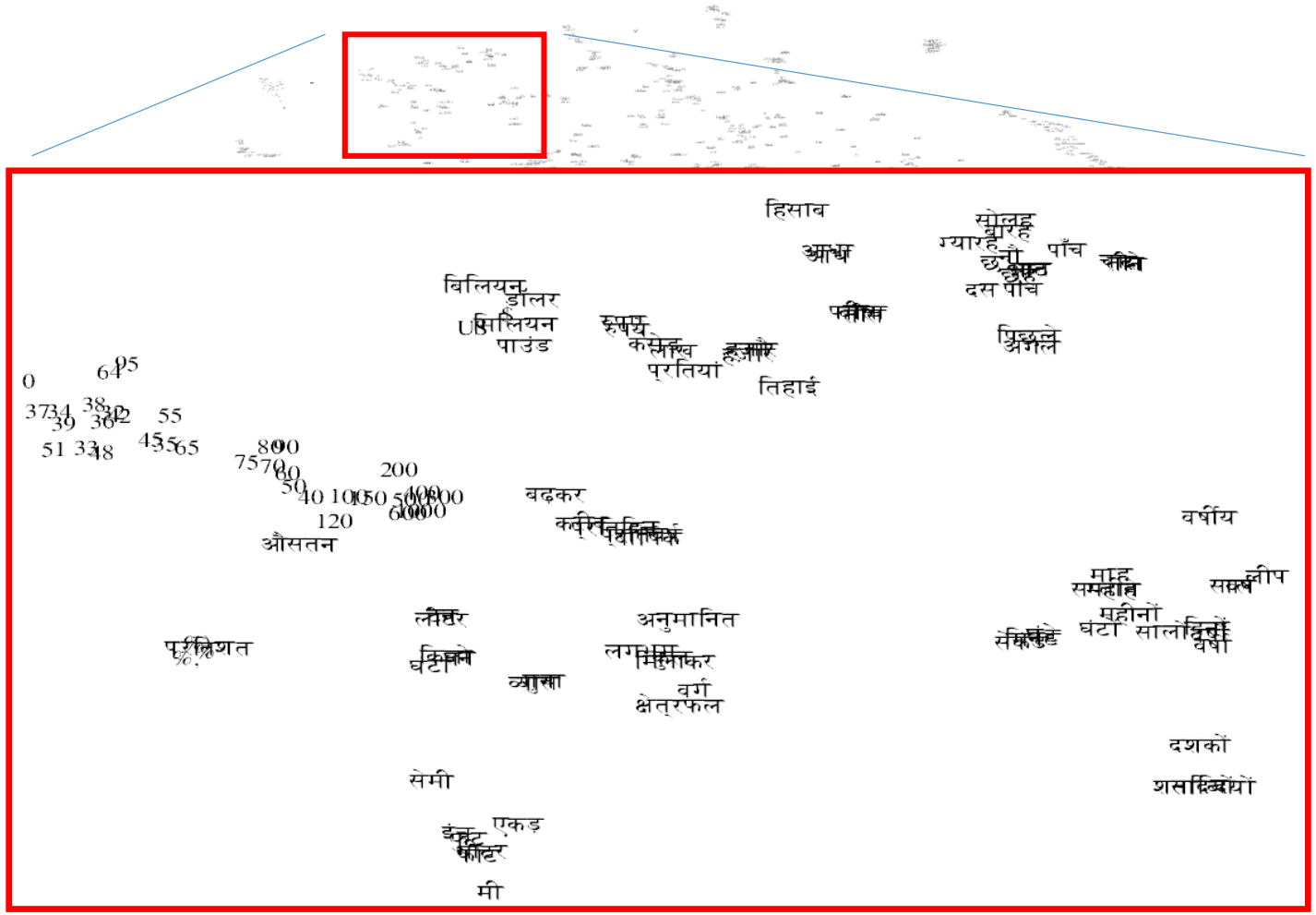


Try to predicts context words given target word

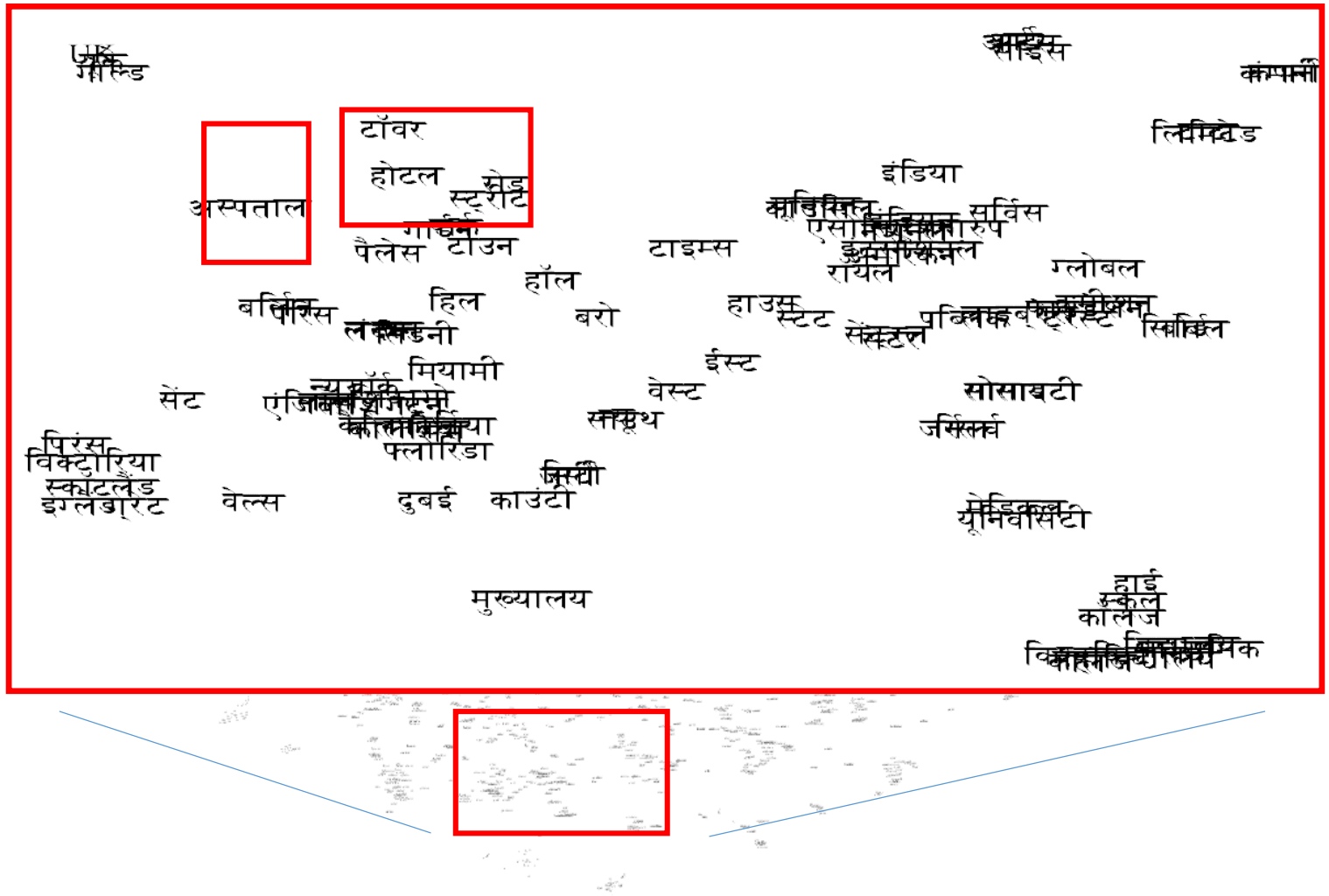
Word Vector Space: Hindi (top 5000)

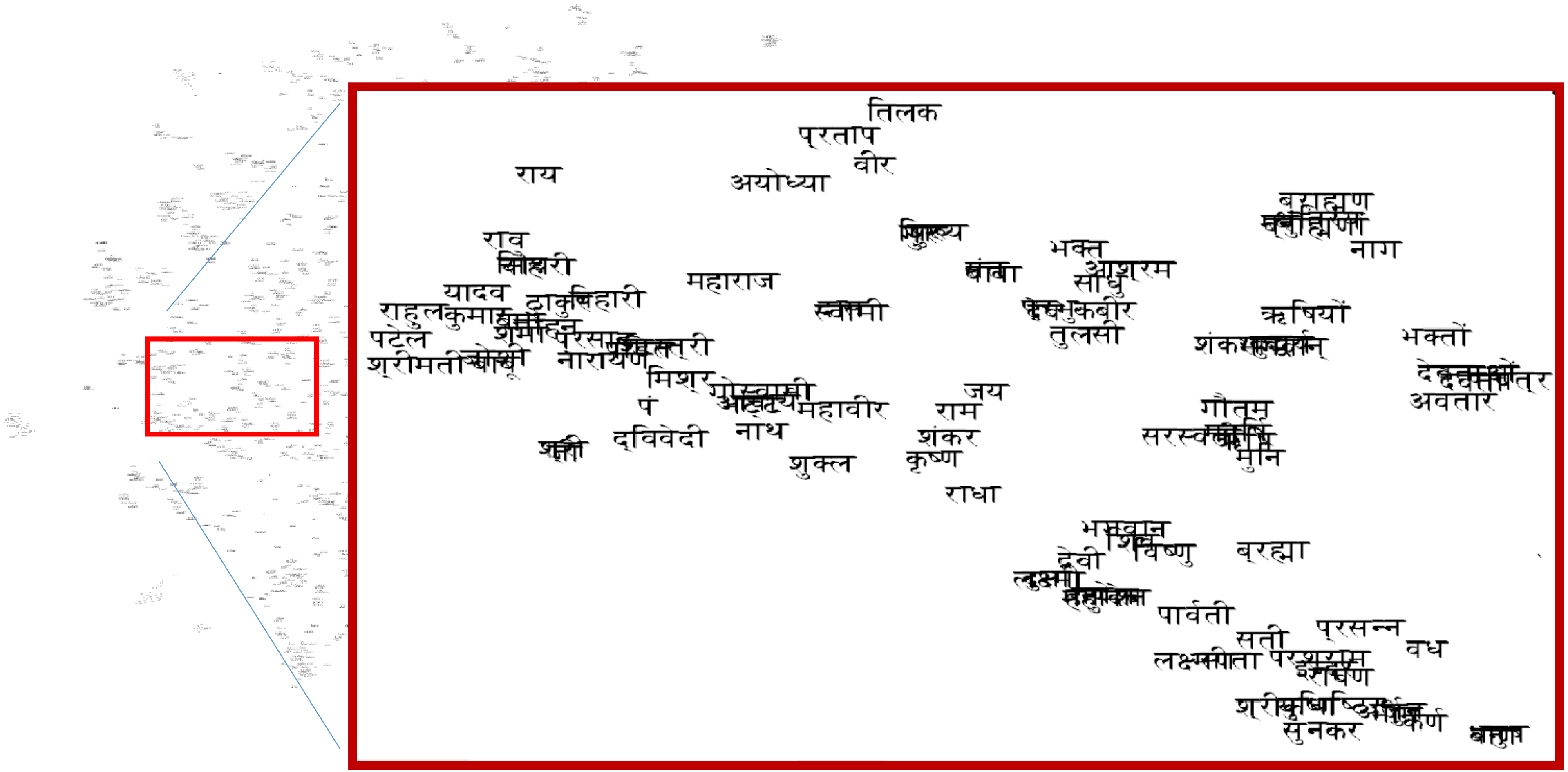
Word Vector Space: Hindi (top 5000)

Word Vector Space: Hindi (top 5000)

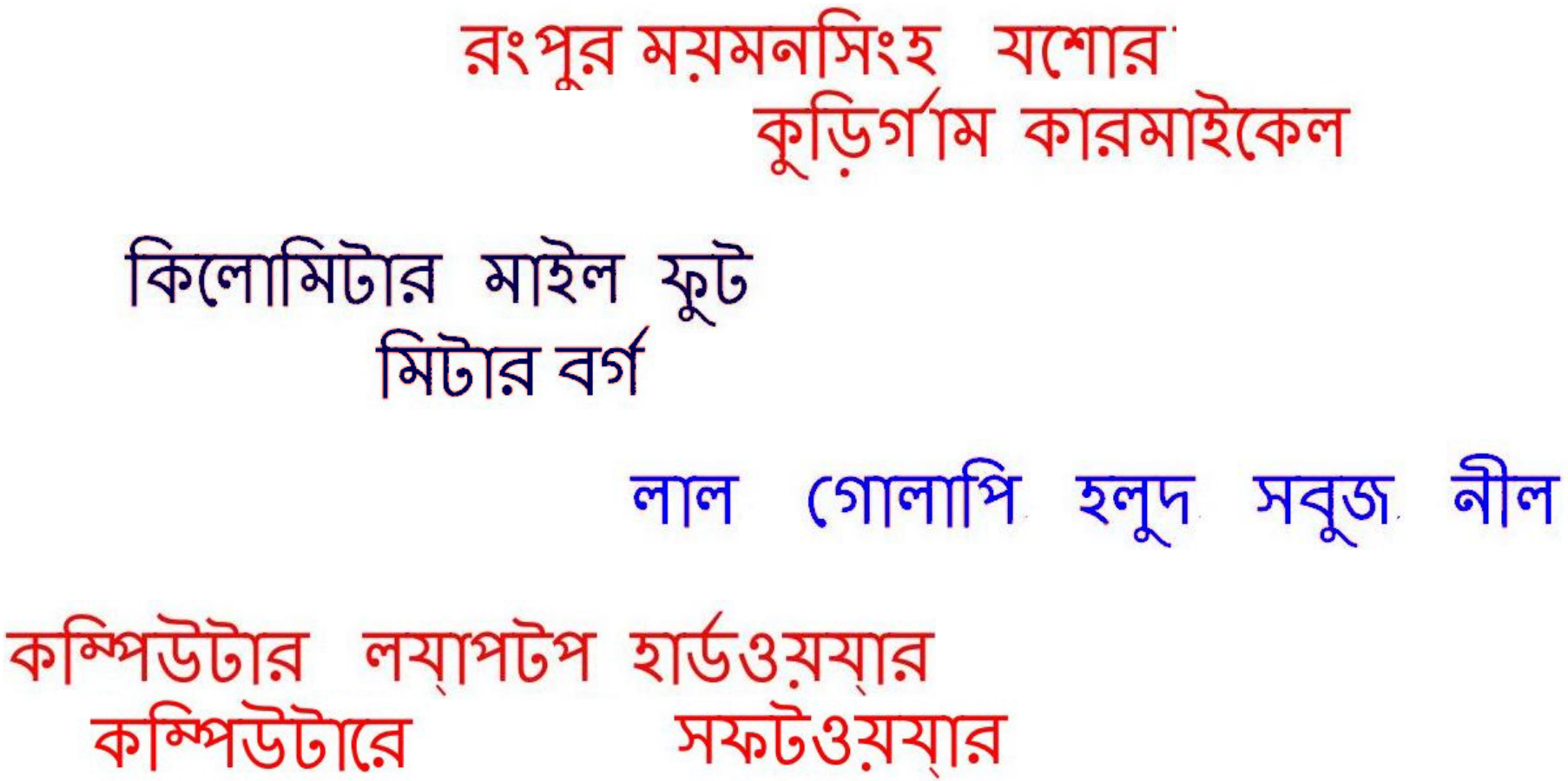


Word Vector Space: Hindi (top 5000)

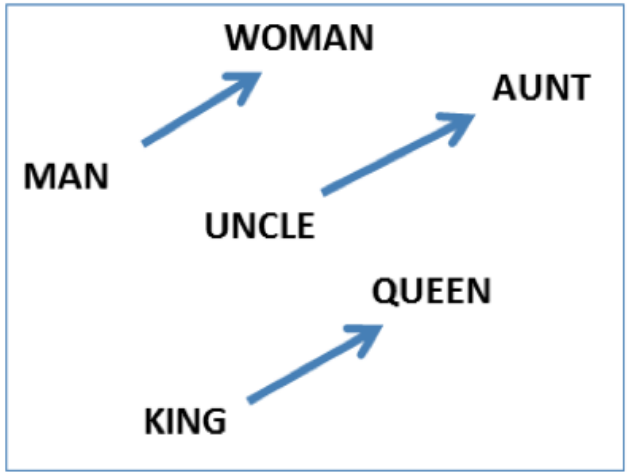


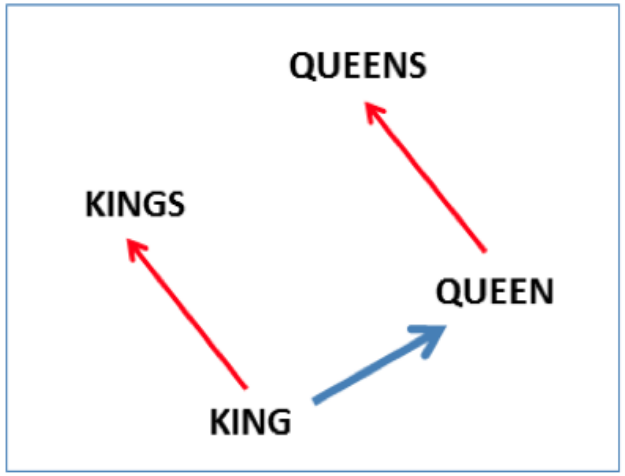

तिलक
प्रताप
वीर
राय
अयोध्या
राव
महाराज
यादव
बिहारी
राहुल
कुमार
पटेल
श्रीमती
नारायण
मिश्र
पं
द्विवेदी
नाथ
महावीर
शुक्ल
स्वामी
भक्त
आश्रम
साधु
कबीर
तुलसी
जय
राम
शंकर
कृष्ण
राधा
ब्राह्मण
नाग
ऋषियों
भक्तों
अवतार
गौतम
सरस्वती
मुनि
भगवान
शिव
विष्णु
ब्रह्मा
देवी
लक्ष्मी
पार्वती
सती
प्रसन्न
वध
परशुराम
सुनकर

Word Vector visualization : Bangla

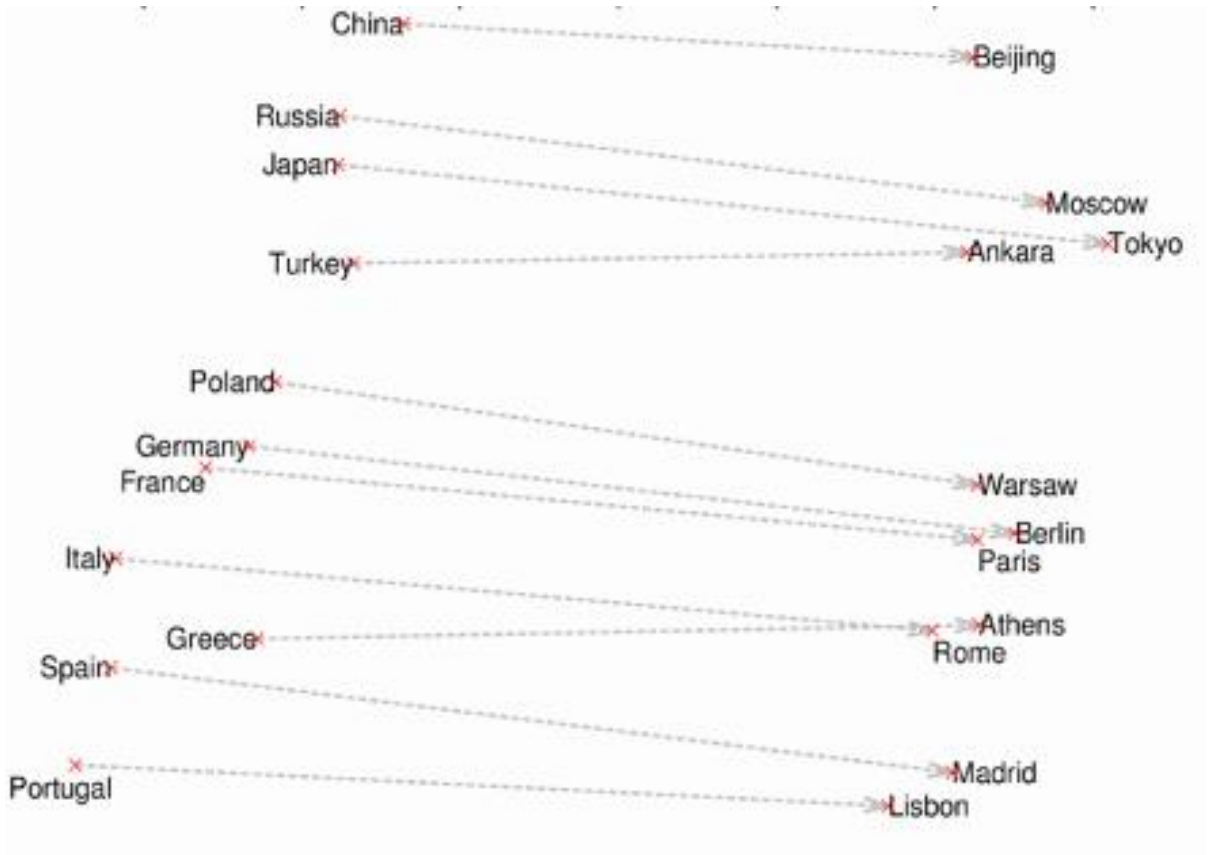


Gender and Number Relations





Ontological Relations



AK Zehady, Purdue U

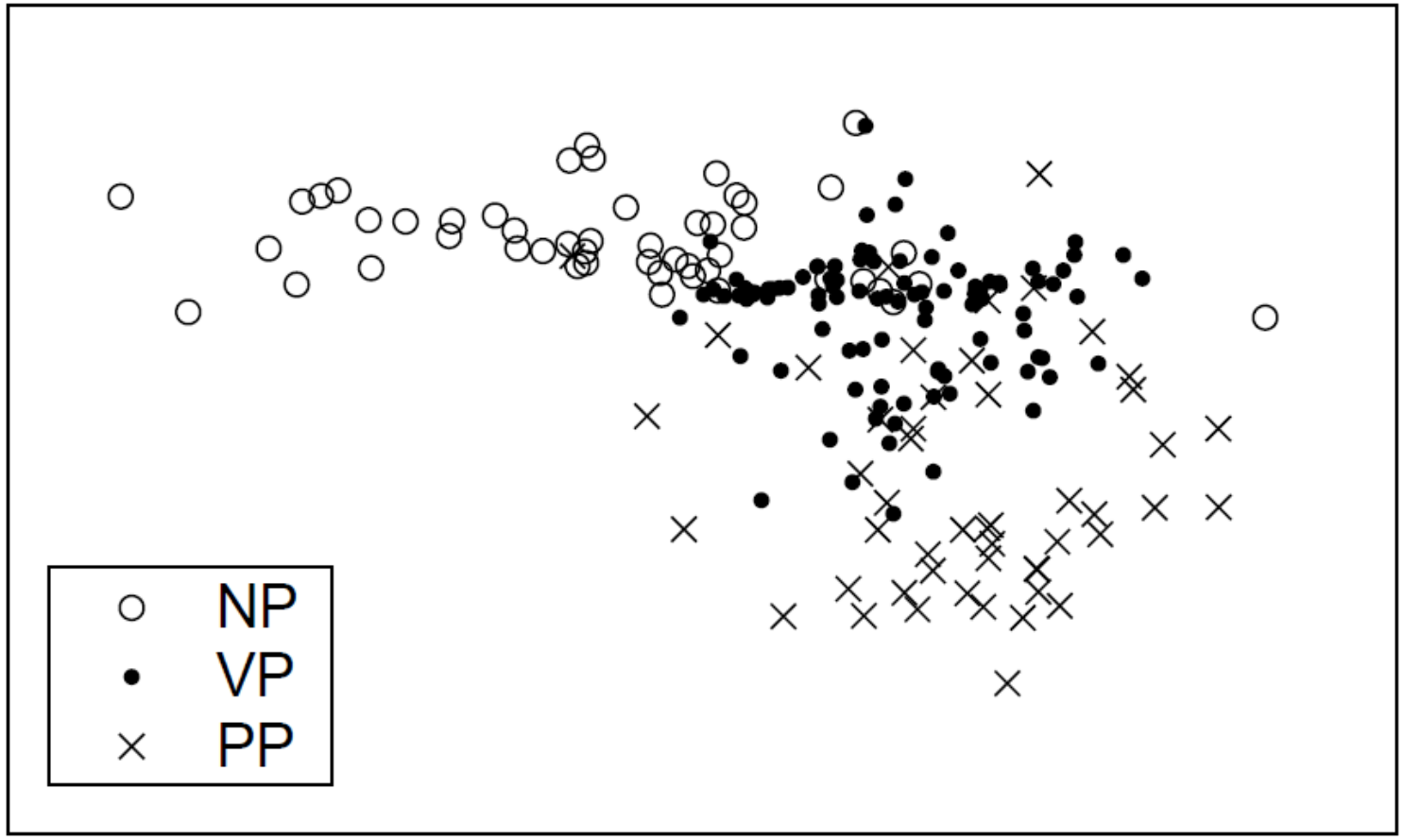
Sequence Models (syntax)

Compositionality in LSTM

context vectors for three types of phrases
→ PCA → space of first two principal components

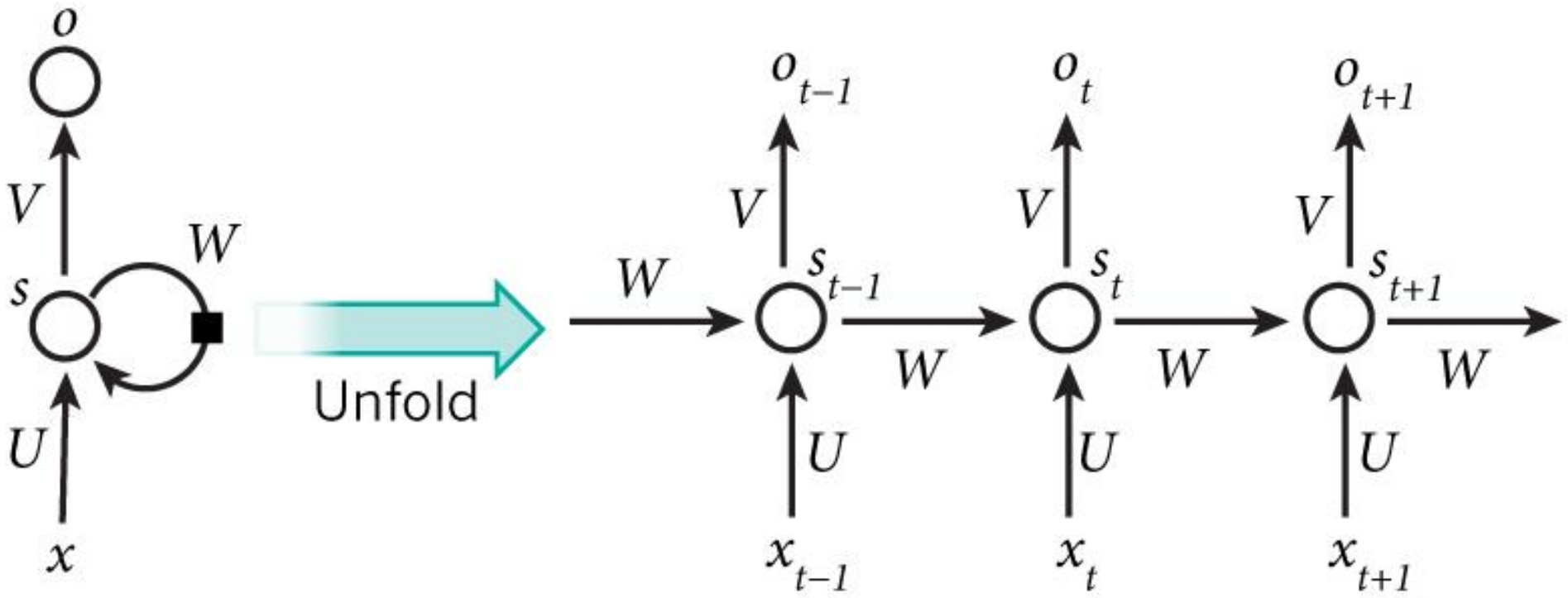
Syntax as Dimensionality Reduction

context vectors for three types of phrases
→ PCA → space of first two principal components



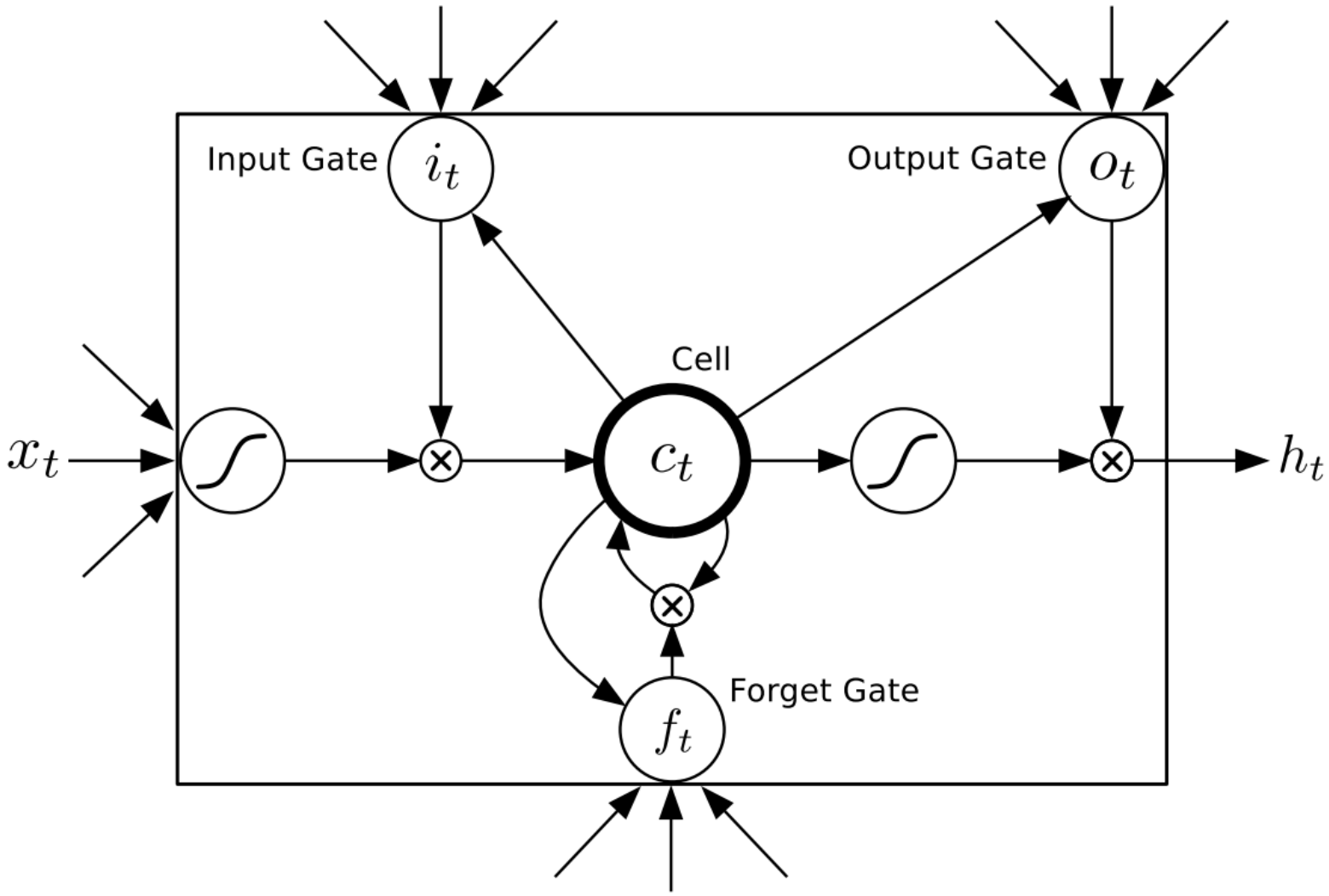
Recurrent Neural Networks (RNN)

Recurrent Network : Latent vars from (t-1) are fed into time t; Recursively encode Past data



Long Short-Term Memory

Recurrent Network : Latent vars from (t-1) are fed into time t; Recursively encode Past data

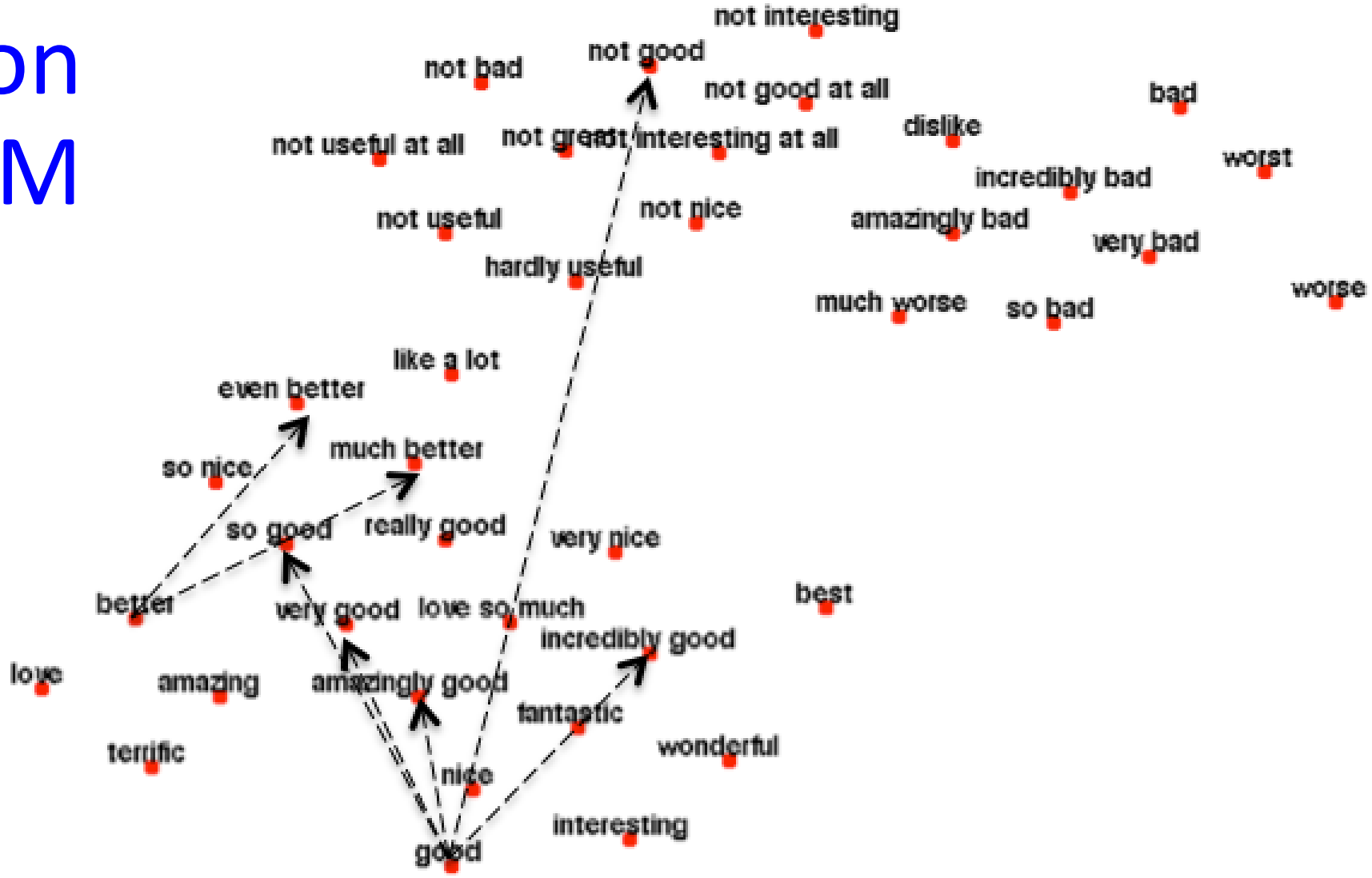


Long Short-Term Memory

Recurrent
Network :
Latent vars from
(t-1) are fed into
time t;
Recursively encode
Past data

$$i_t = \sigma(W_i \cdot e_t + V_i \cdot h_{t-1})$$
$$f_t = \sigma(W_f \cdot e_t + V_f \cdot h_{t-1})$$
$$o_t = \sigma(W_o \cdot e_t + V_o \cdot h_{t-1})$$
$$l_t = \tanh(W_l \cdot e_t + V_l \cdot h_{t-1})$$
$$c_t = f_t \cdot c_{t-1} + i_t \times l_t$$
$$h_t = o_t \cdot m_t$$

Composition in LSTM



Trained on Stanford Sentiment Treebank dataset (Socher 13)

li-chen-hovy-jurafsky-15-ACL_visualizing-neural-models-in-NLP

Translation without Parsing

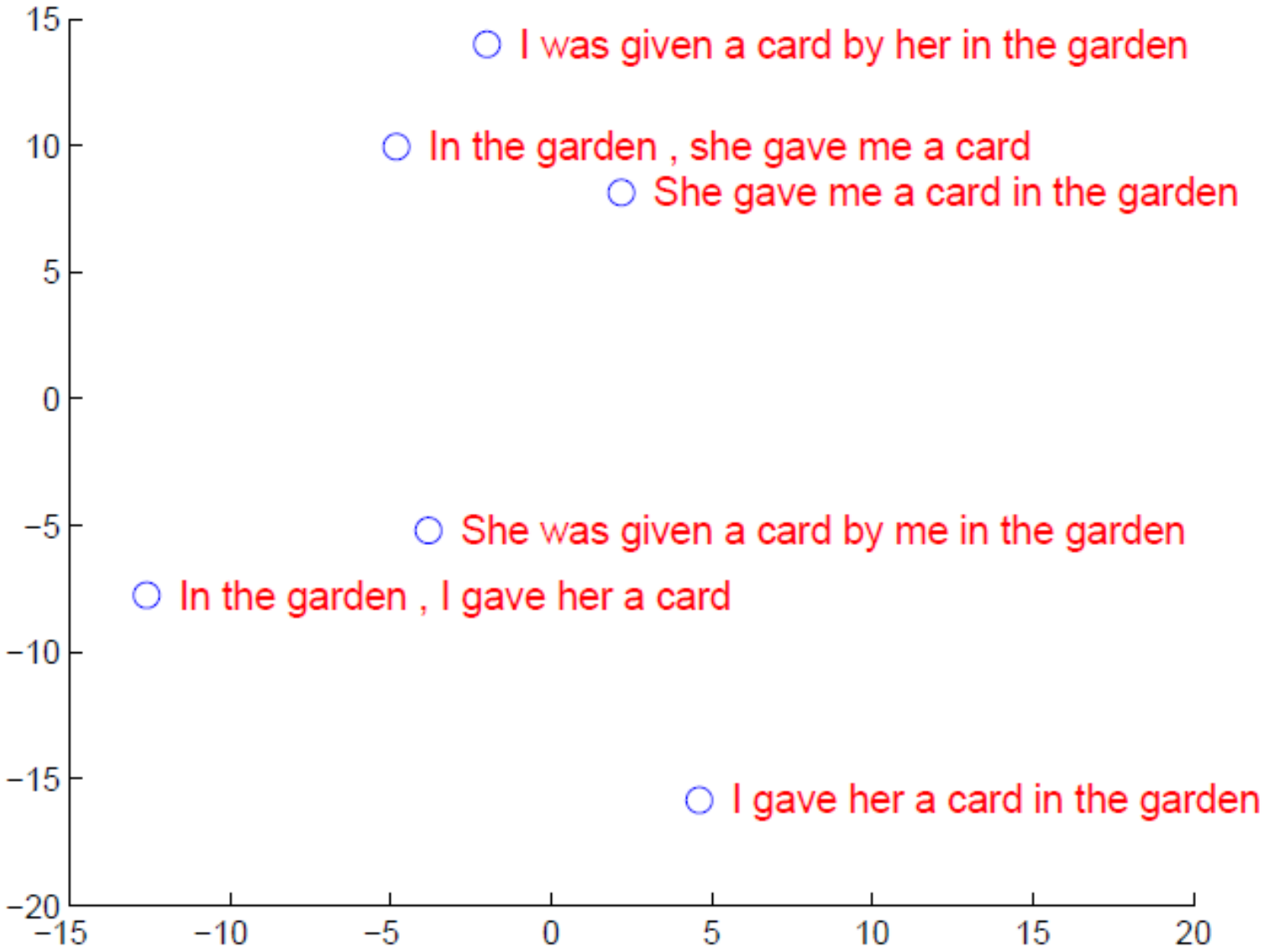
RNN models for Translation

- Source L1 : LSTM based sentence vector model trained on input.
- Target L2 : LSTM based generative model
  - Auli Galley Quirk & Zqeig 13 : joint language and translation w RNN
  - Sutskever Vinyals & Le 14 : sequence to sequence LSTM translation
  - 2015 : translation from image inputs (caption generation); translation to images (image search) etc.

Active / Passive Voice

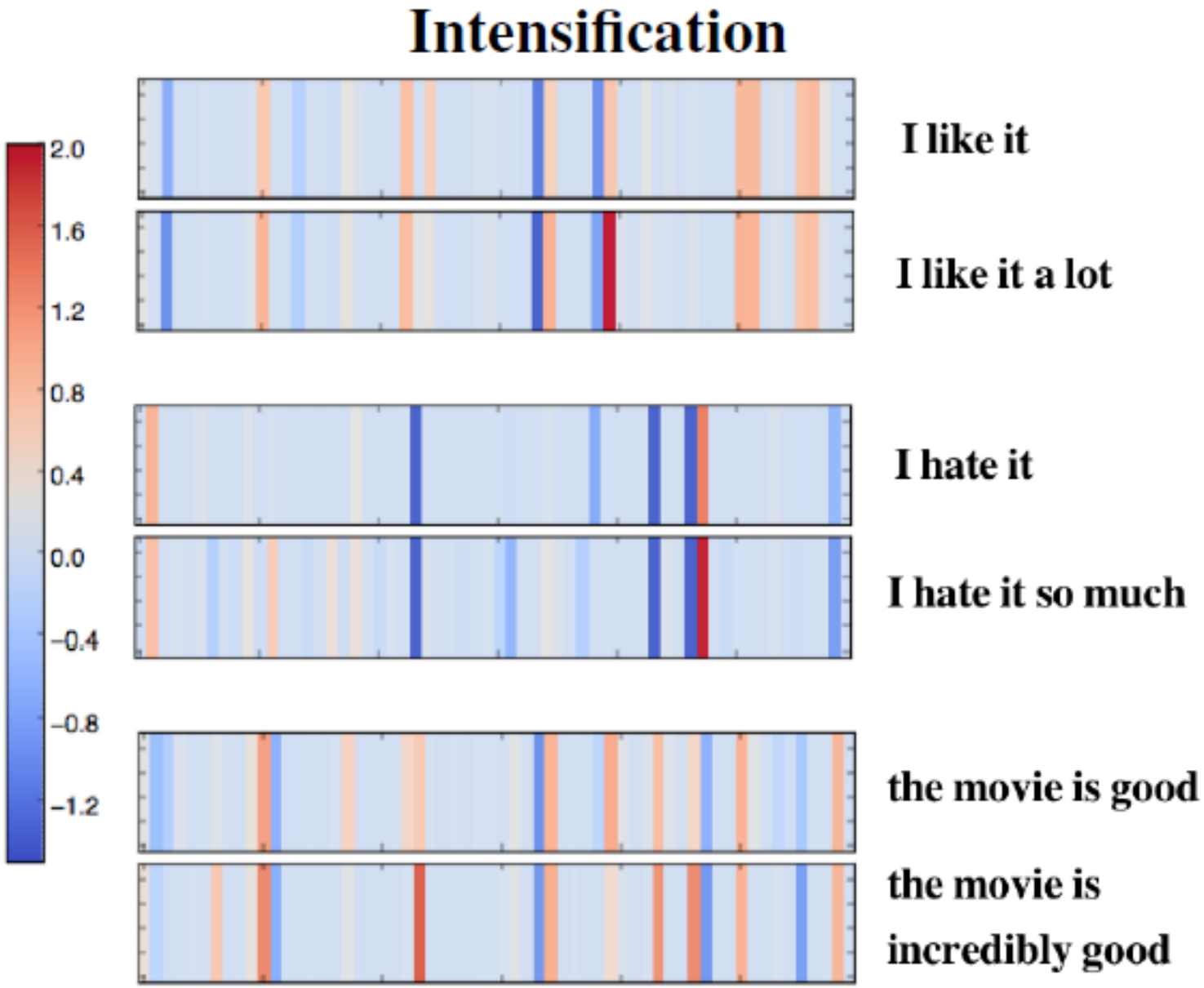
LSTM hidden states 2D PCA visualizn

English to French Translation task



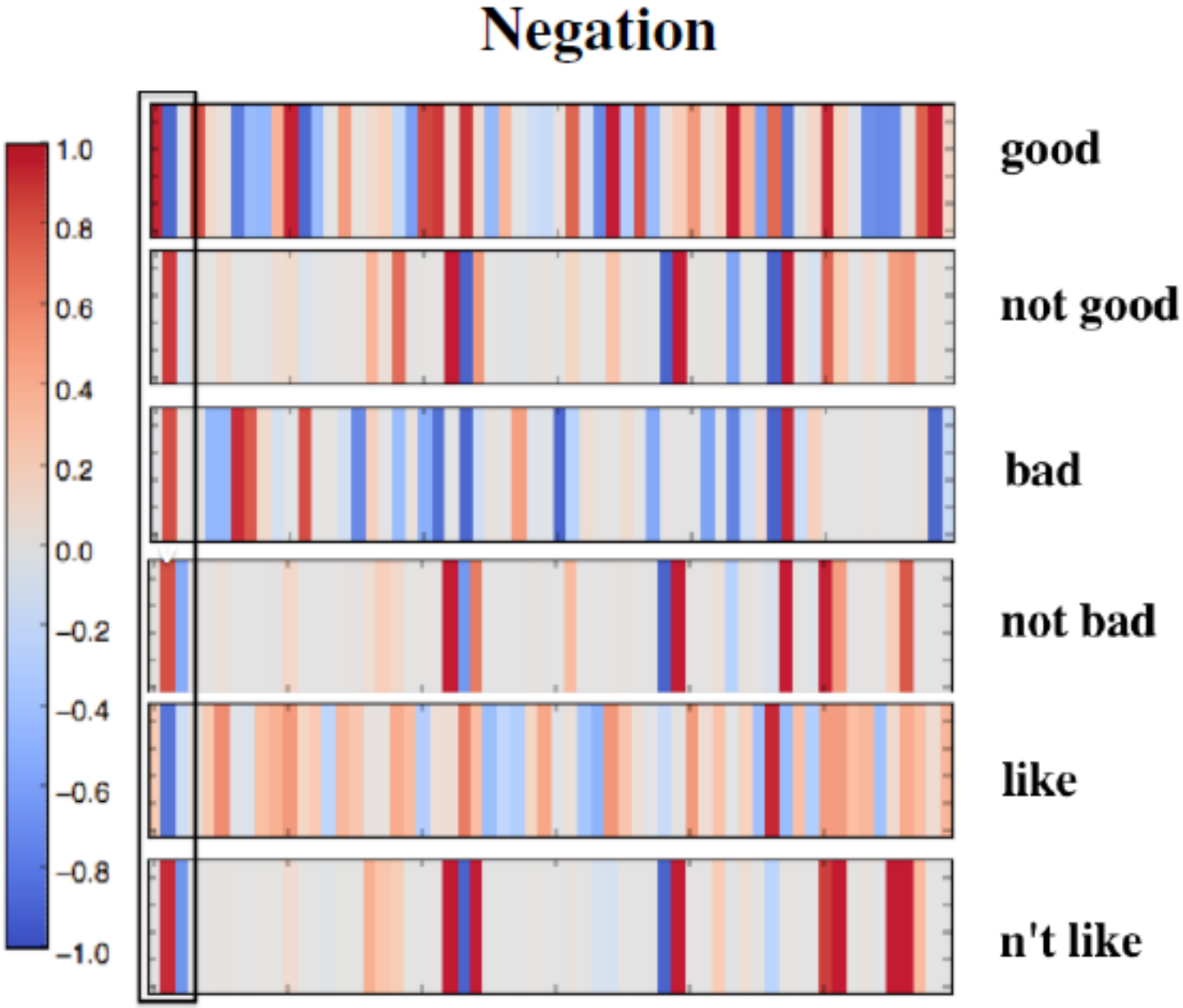
sutskever-vinyals-14_sequence-to-sequence-LSTM_translation

Composition in LSTM



li-chen-hovy-jurafsky-15-ACL_visualizing-neural-models-in-NLP

Composition in LSTM



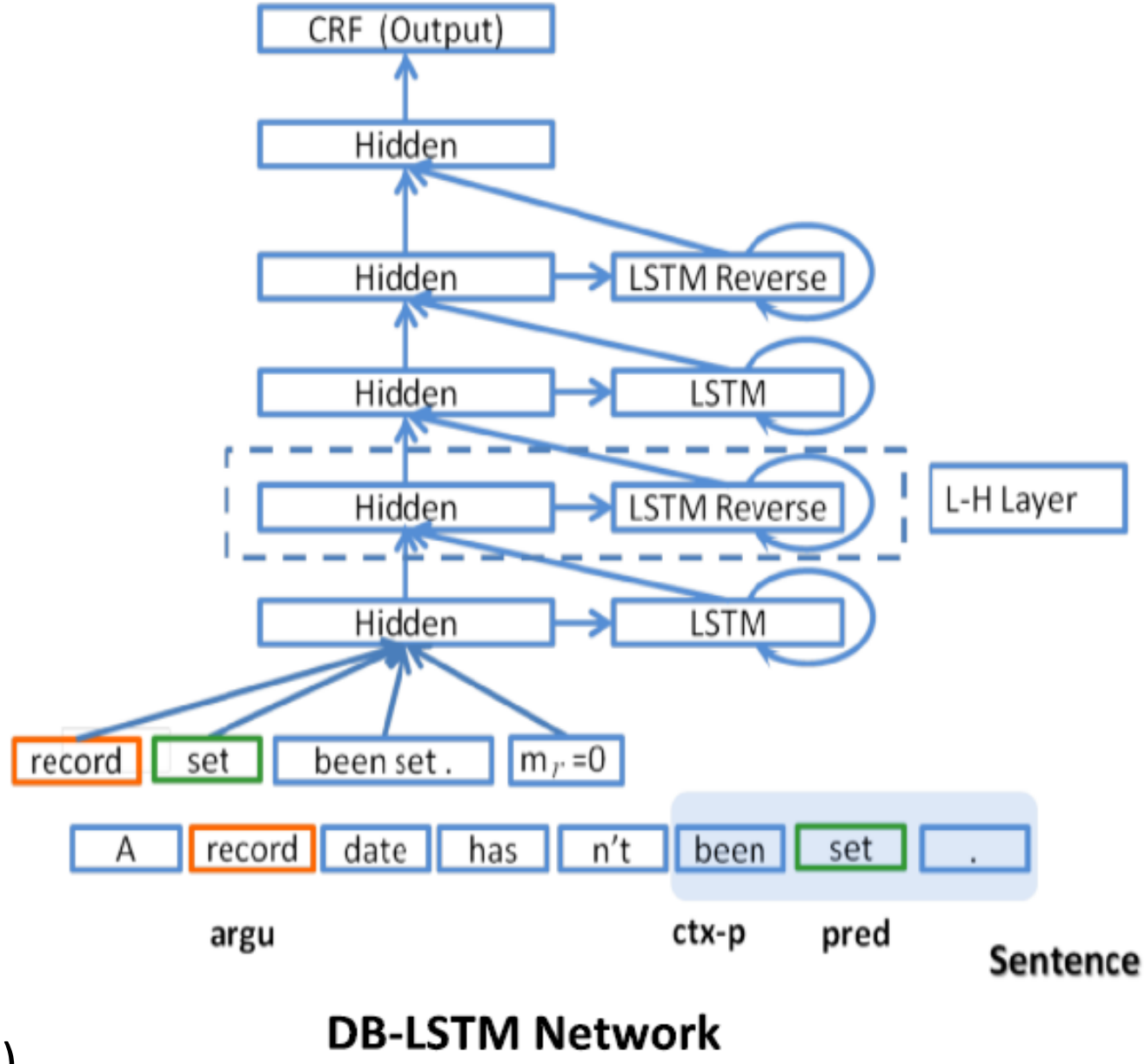
li-chen-hovy-jurafsky-15-ACL_visualizing-neural-models-in-NLP

Four Features:

- Predicate (verb)
- Arguments
- Context
- Region

- F-score: 81

Outperforms
parsing based models
and also CNNs (F1 ~ 50s)



DB-LSTM Network

zhou-xu-15_end-to-end-semantic-role-labeling-w-RNN

A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A skateboarder does a trick on a ramp.

A dog is jumping to catch a frisbee.

A group of young people playing a game of frisbee.

Two hockey players are fighting over the puck.

A little girl in a pink hat is blowing bubbles.

A refrigerator filled with lots of food and drinks.

A herd of elephants walking across a dry grass field.

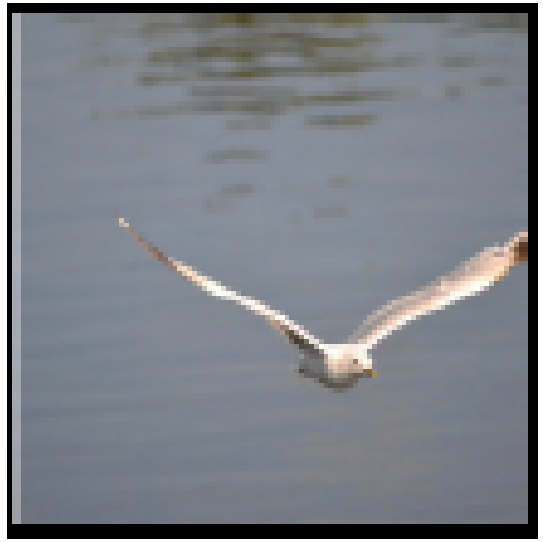
A close up of a cat laying on a couch.

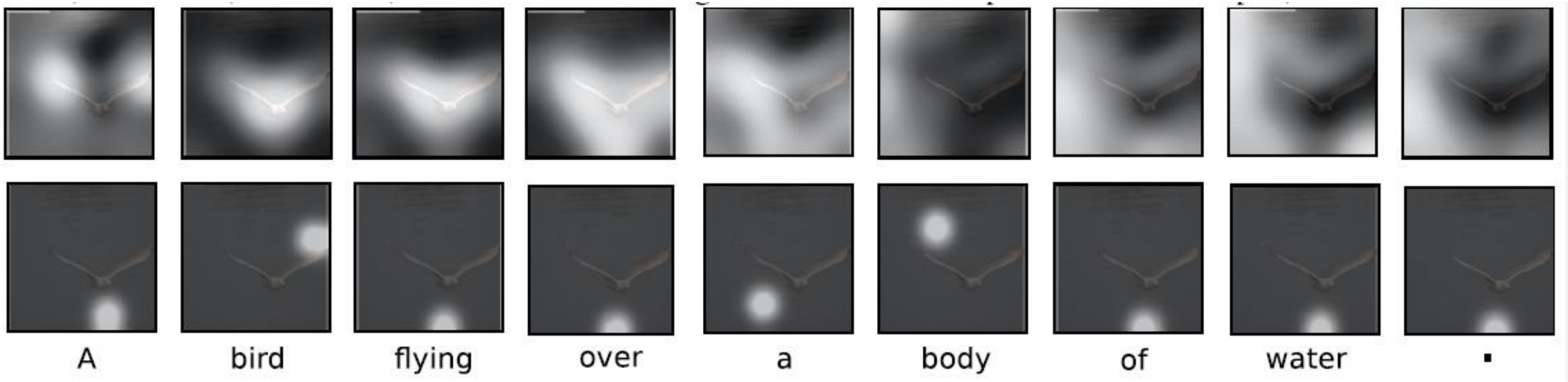
A red motorcycle parked on the side of the road.

A yellow school bus parked in a parking lot.

Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image

Describing images with attention





xu-K-ba-J-kiros-15_show-attend-and-tell_image-captioning

Describing images with attention

A woman is throwing a frisbee in a park.

A dog is standing on a hardwood floor.

A stop sign is on a road with a mountain in the background.

A little girl sitting on a bed with a teddy bear.

A group of people sitting on a boat in the water.

A giraffe standing in a forest with trees in the background.

xu-K-ba-J-kiros-15_show-attend-and-tell_image-captioning

Describing images with attention

Errors :

Can be analyzed by looking at attention window

A large white bird standing in a forest.

A woman holding a clock in her hand.

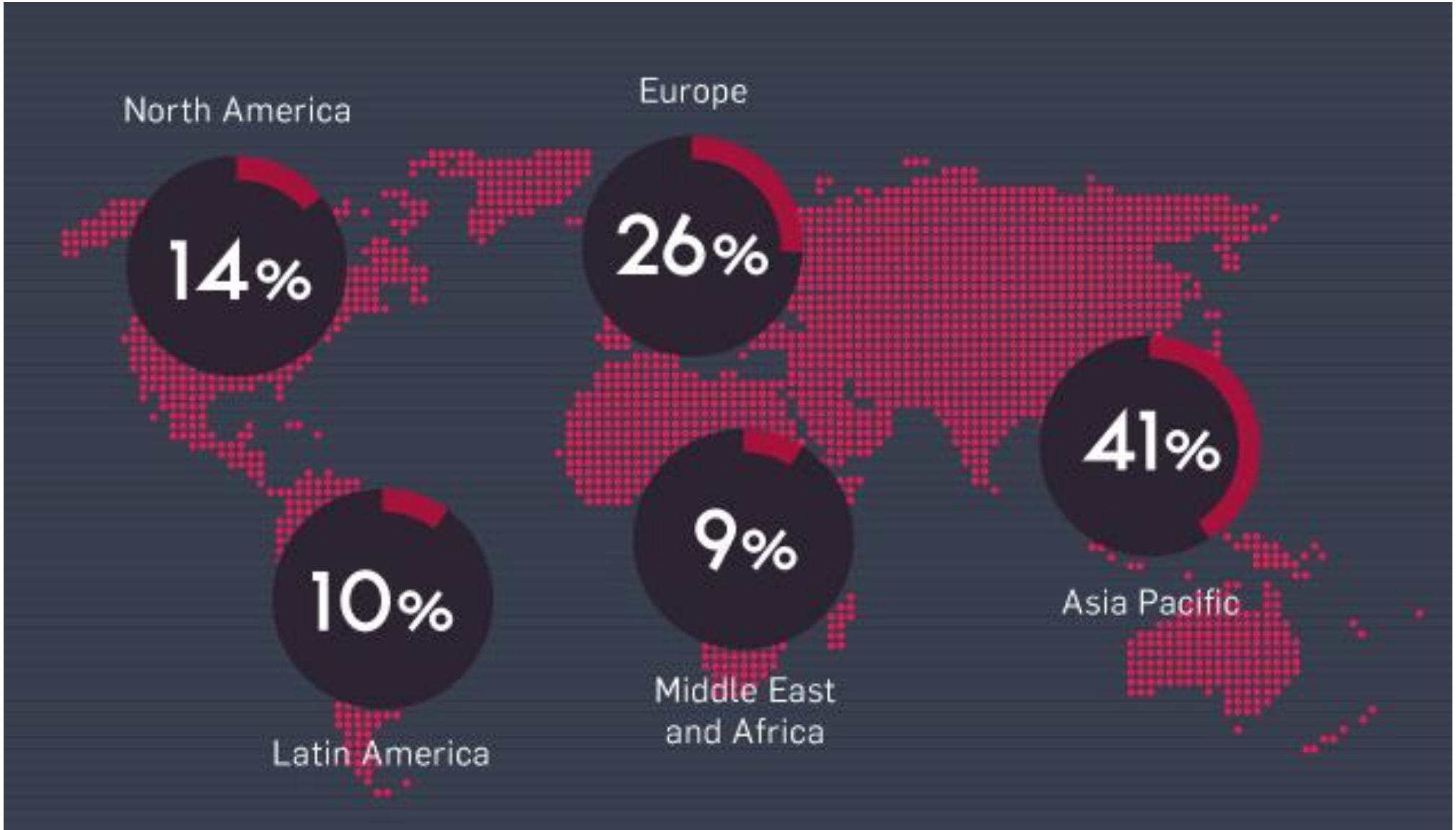
A person is standing on a beach with a surfboard.

A woman is sitting at a table with a large pizza.

xu-K-ba-J-kiros-15_show-attend-and-tell_image-captioning

Generative Lexicon

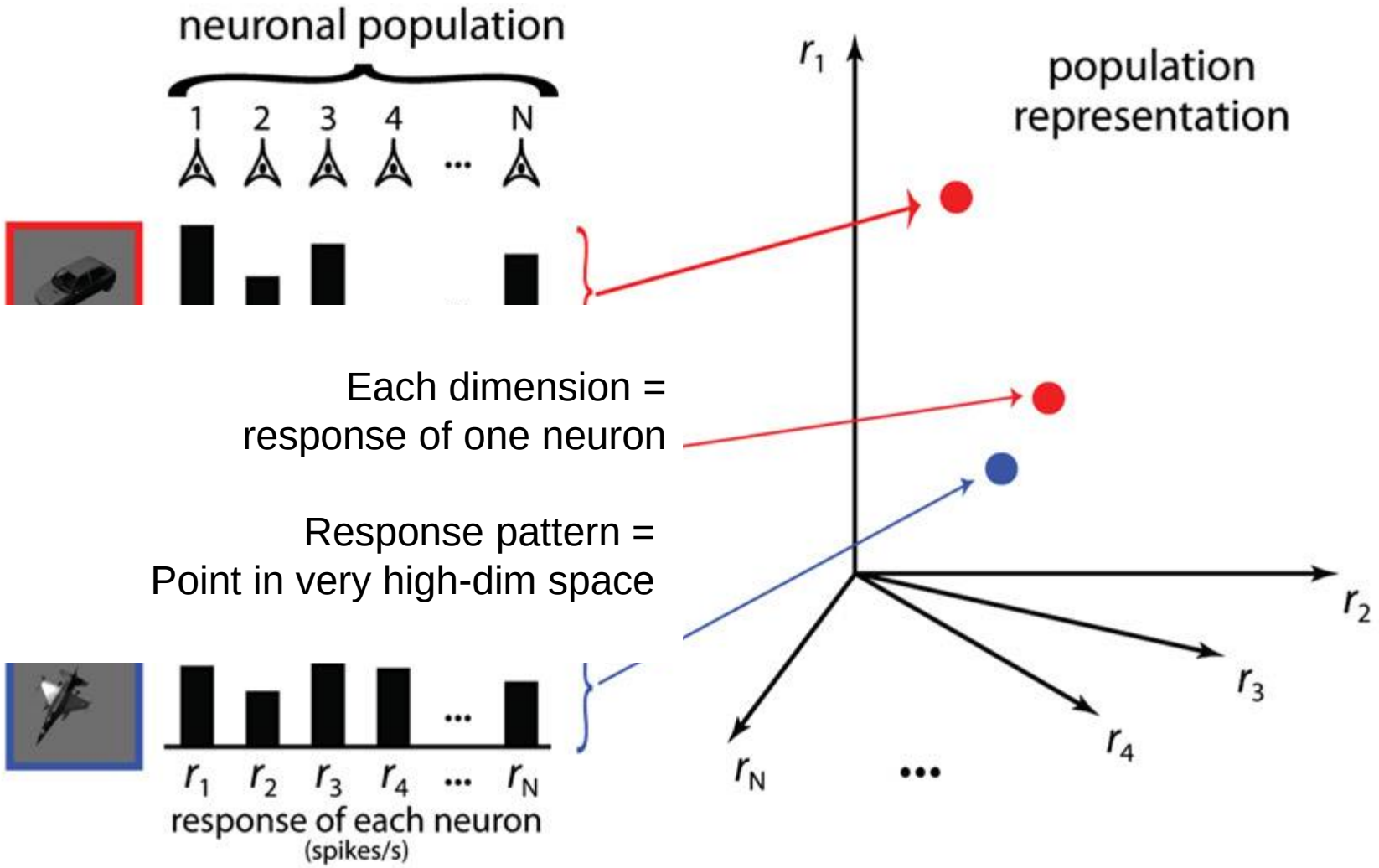
Web Users Map- 2014



- http://www.statista.com

Population Code (vector models)

Population Codes



Each dimension =
response of one neuron

Response pattern =
Point in very high-dim space

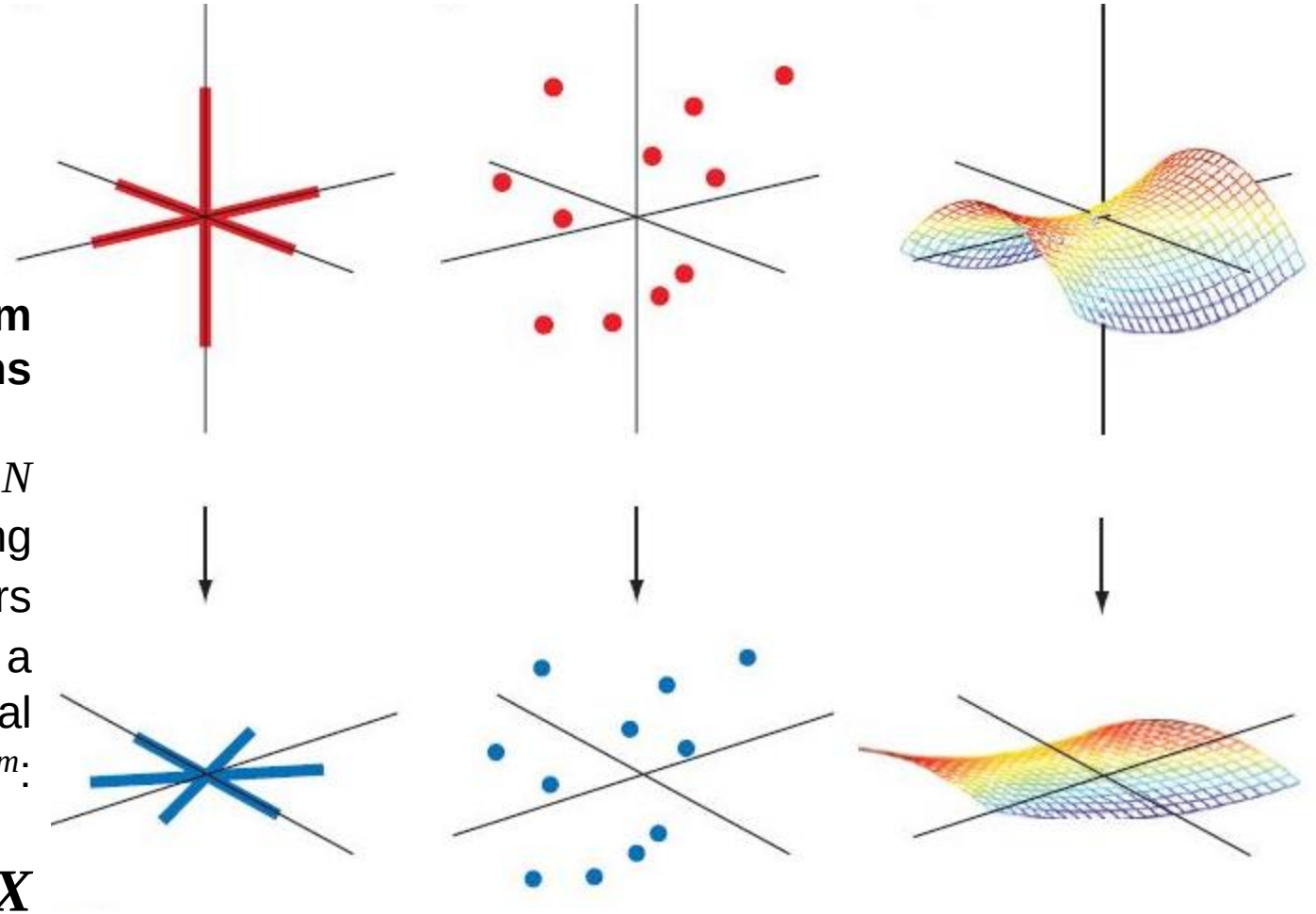
diCarlo-cox-2007trics,

Sparse Coding

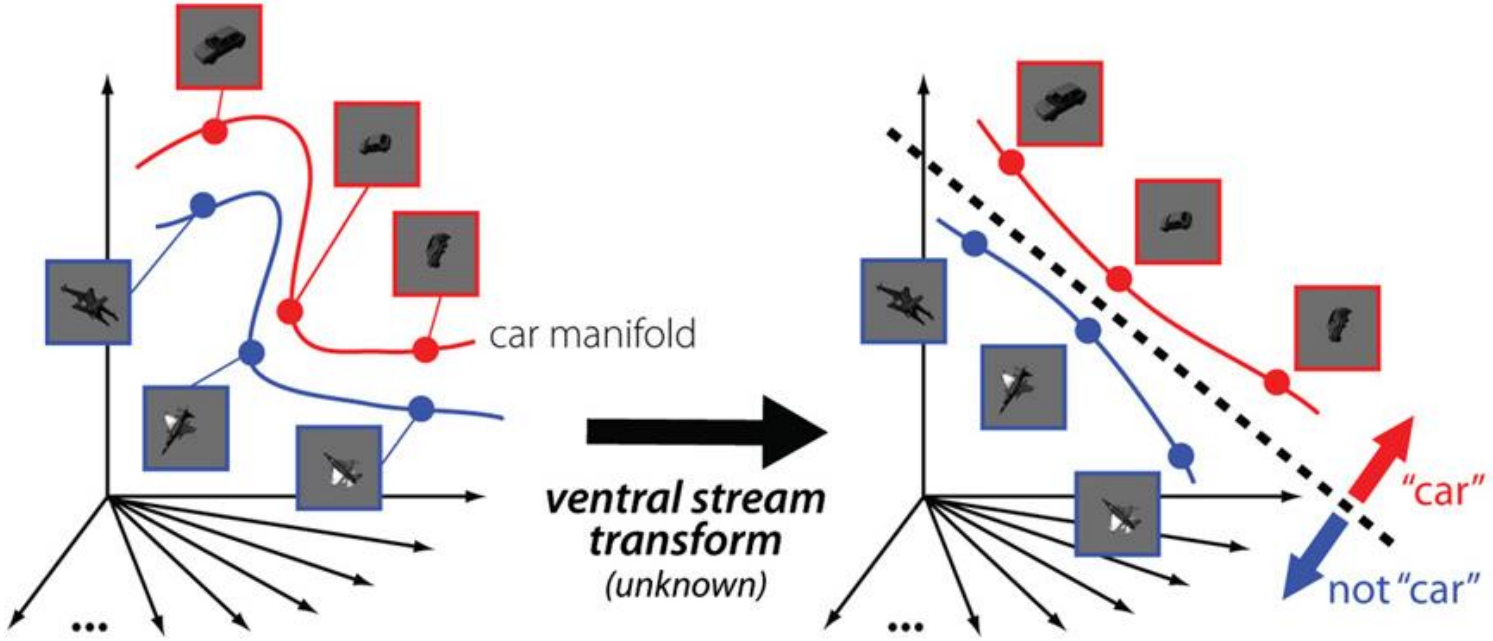
Random Projections

project each of the N vectors $\boldsymbol{x}$ along m random vectors ***[rp]*** to obtain a lower-dimensional projection in R^m:

$$\boldsymbol{y = [rp] \,.\, X}$$

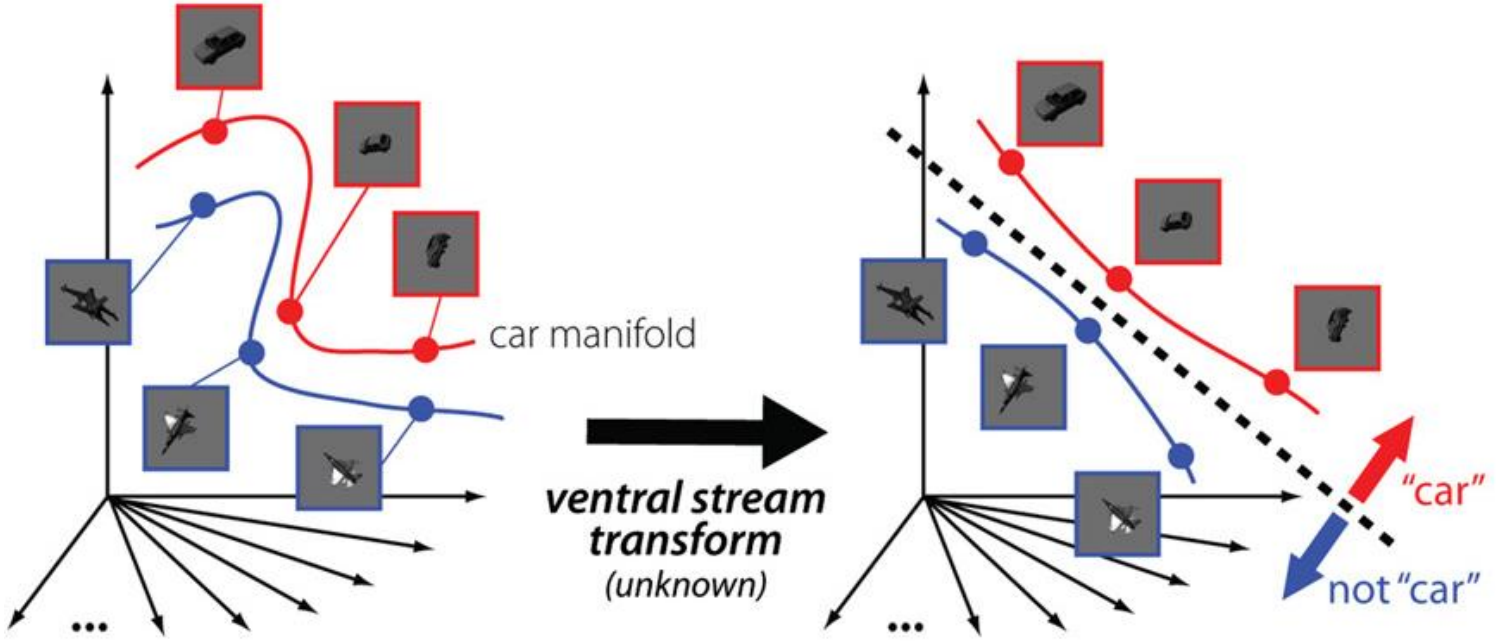


Object Manifolds

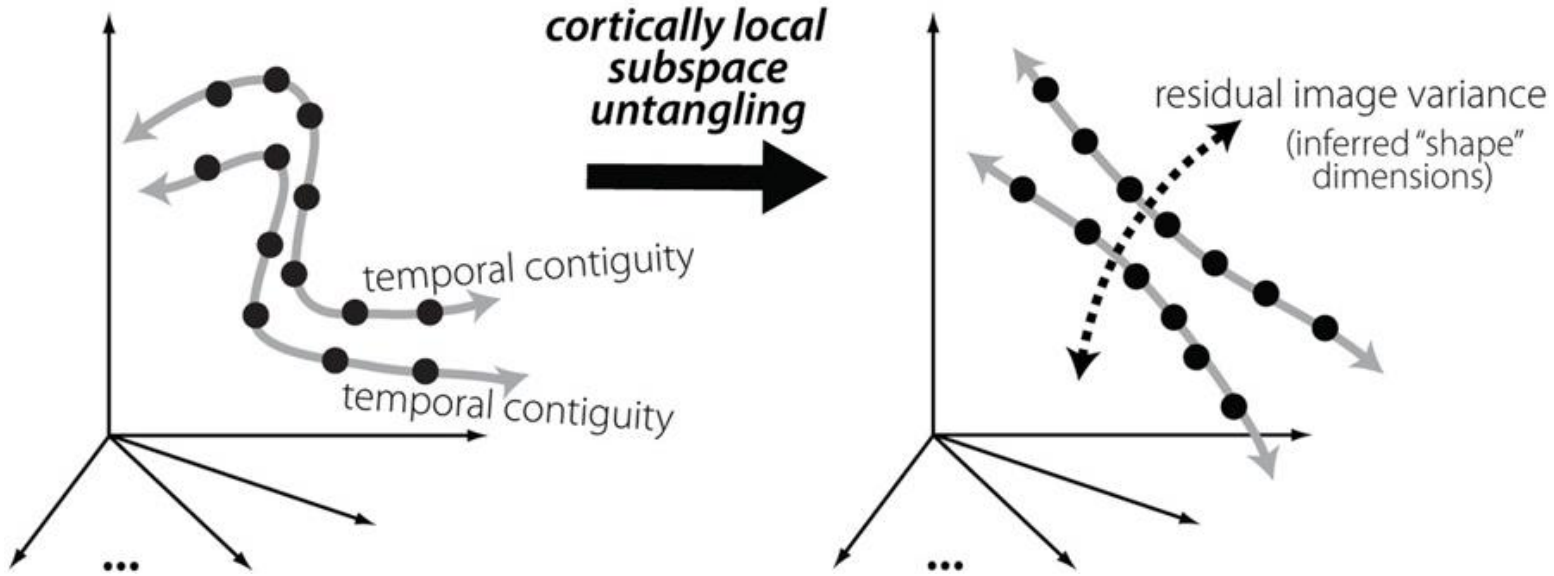


diCarlo-cox-2007trics,

Object Manifolds

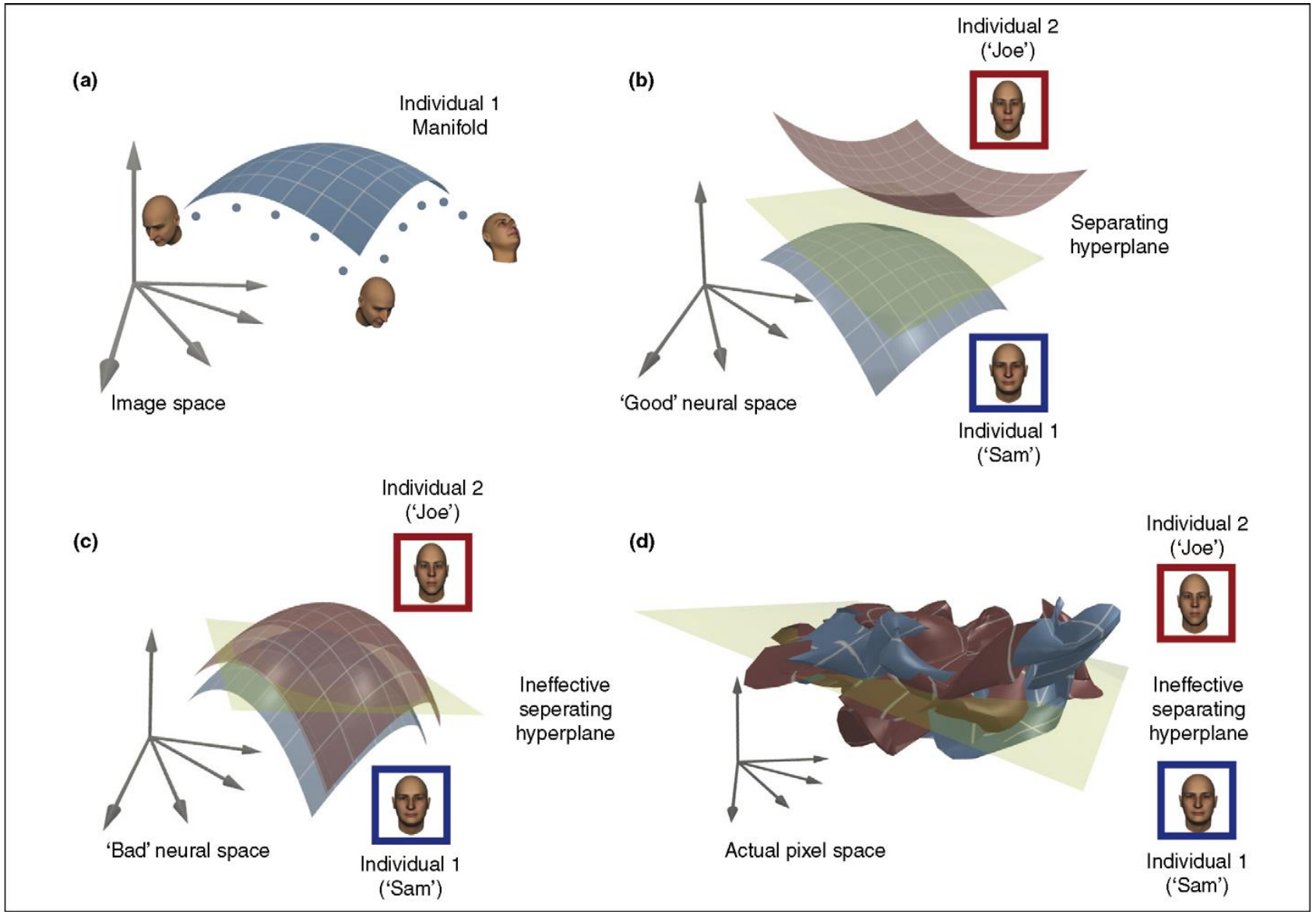


Object Recog-nition



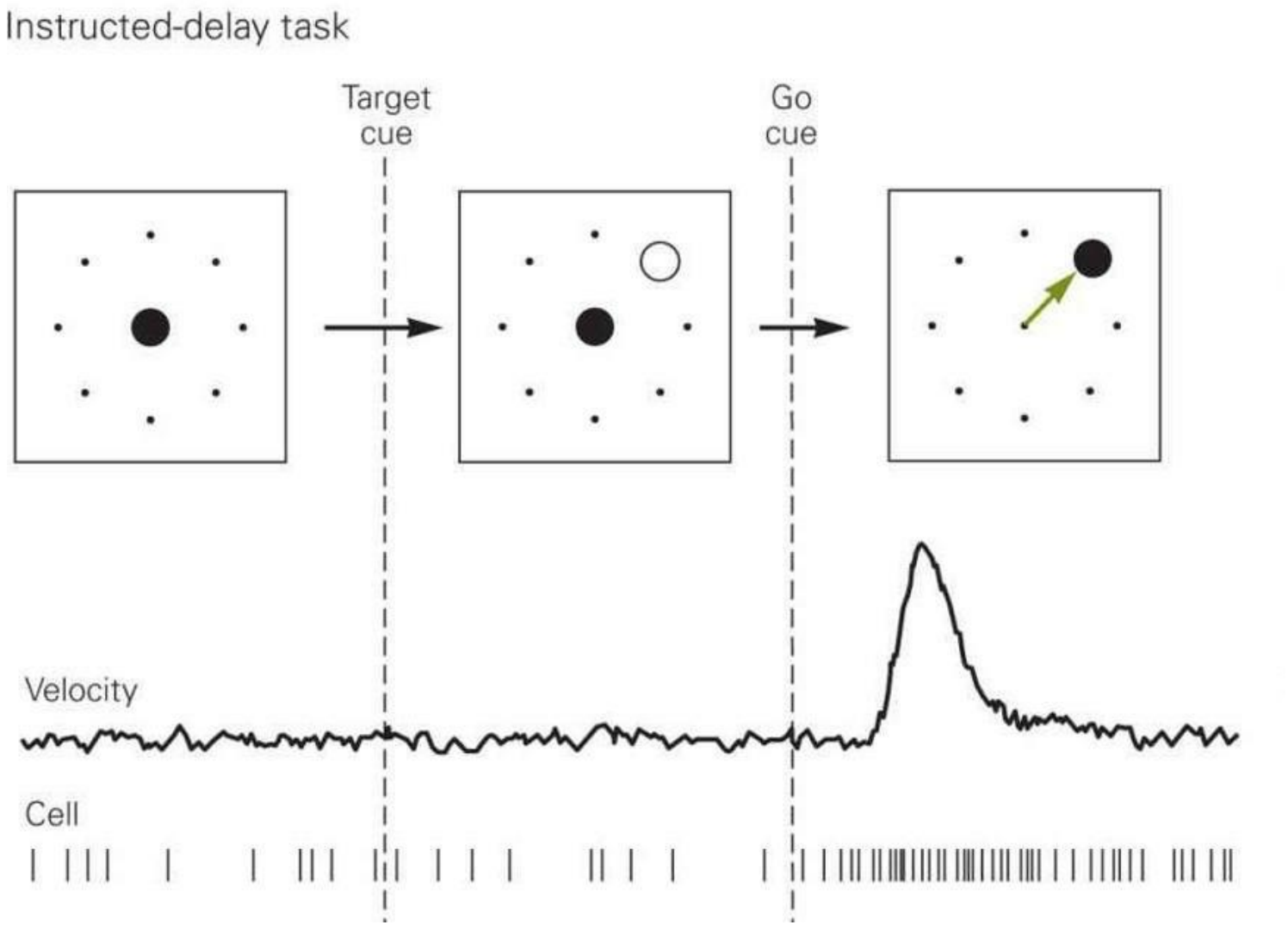
diCarlo-cox-2007trics,

Face Recognition: Manifolds in the brain



dicarlo-cox-07_untangling-invariant-object-recog

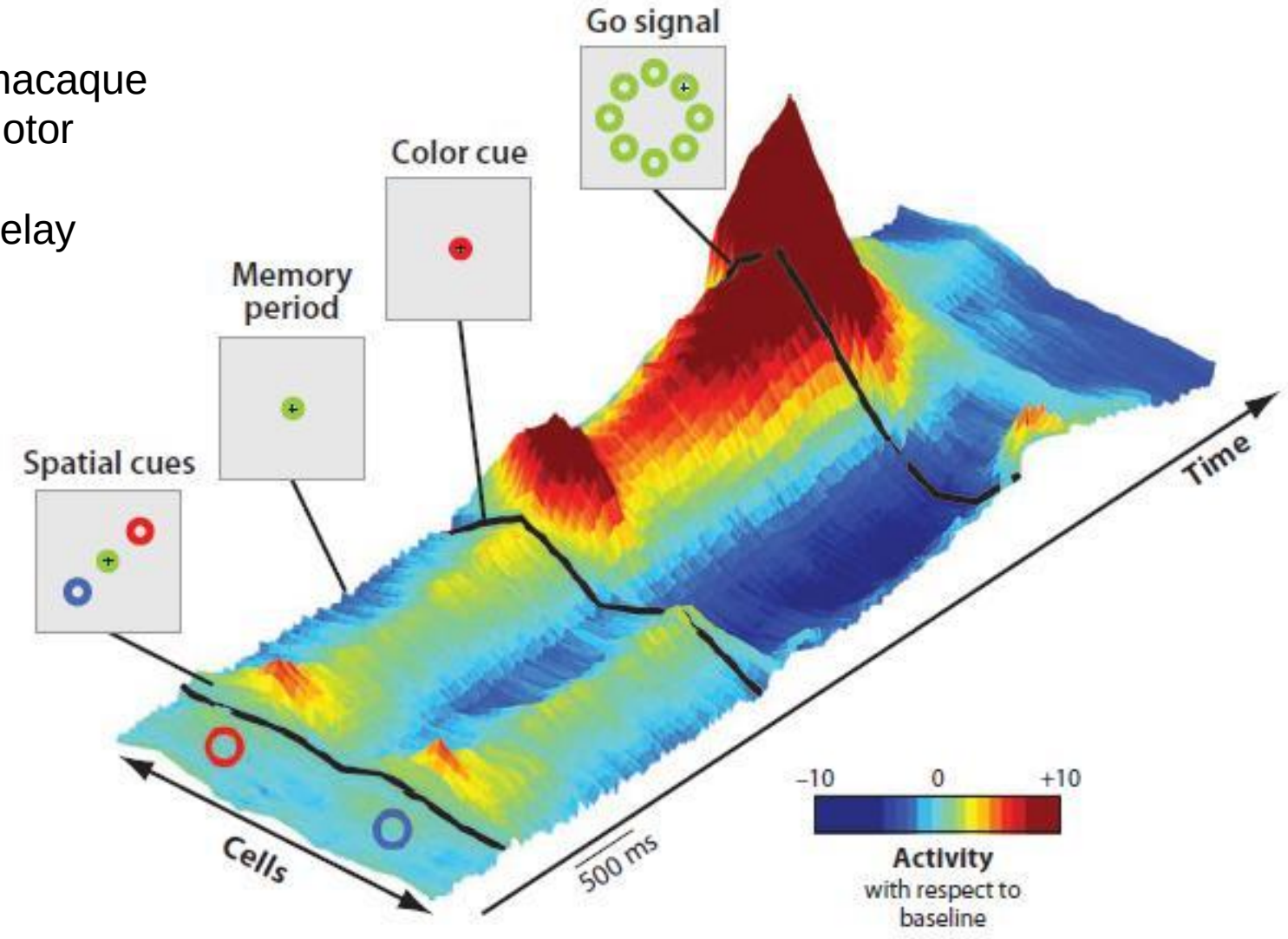
Instructed delay tasks



[Kandel 12]

Unified decision-making

Data from macaque dorsal premotor cortex in instructed-delay reaching tasks



[Cisek and Kalaska 2010]