NLP and Cognitive Science

Translation from Images

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

Image caption variability

"people lined up in terminal"
"people lined up at train station"
"long line at a station"
"people waiting for train

"alleyway in a small town"
"People sitting and walking"
"man walking in shopping area with others selling products"

jas-parikh-15_image-specificity

Image caption variability

Specificity = 0.89

There is a lot of snow on the mountain.
There is a snow covered mountain.
A snow covered mountain.
A mountain with snow.
A snowy mountain.

Specificity = 0.59

Children play racing games in an arcade.
A group of kids playing games.
A few kids playing arcade games.
some kids in an arcade.
Kids are playing racing games.

Specificity = 0.37

A house with a porch.
There is a railing around the porch of the house.
House with really green grass.
A view of a small white and blue house.
a house shown from outside.

Specificity = 0.11

People waiting at an airport.
The interior of a building with a sloped roof.
the inside of airport.
A decadent room with people walking arour
A large bowling rink.

jas-parikh-15_image-specificity

Image caption variability

- N sentences describing each image
- M human subjects rate the similarity of pairs of sentences sa and sb

$$spec_{\text{hum}} = \frac{1}{M\binom{N}{2}} \sum_{\forall\{s_a,s_b\}\subset S} \sum_{m=1}^{M} sim^{m}_{\text{hum}}(s_a, s_b)$$

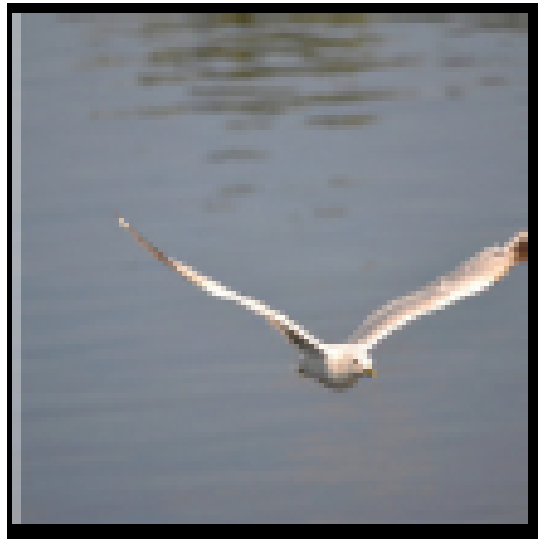
- Automatic Similarity: choose WordNet synset with maximum similarity in context. tf-idf weighted sum

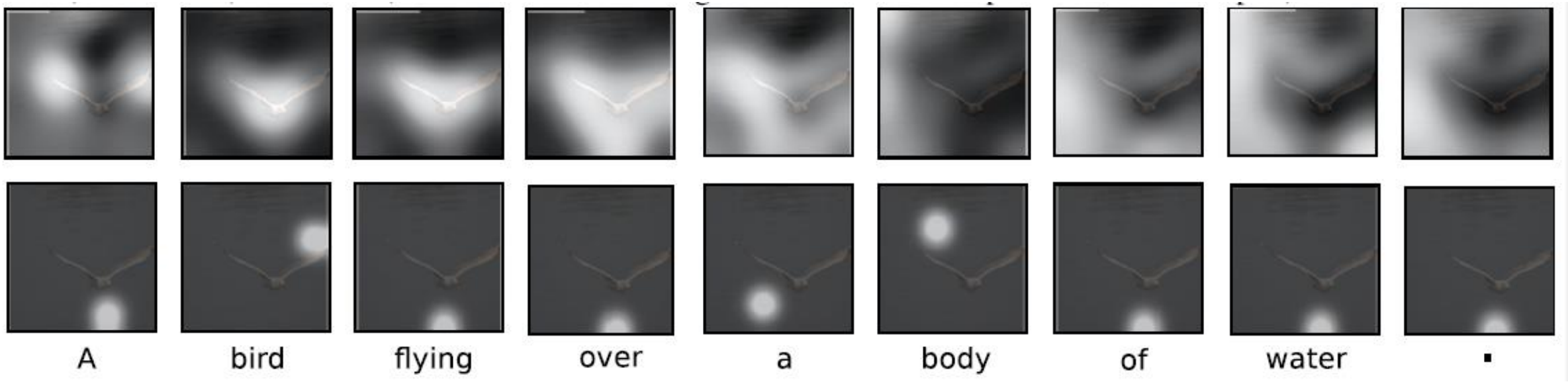
$$sim_{\text{auto}}(s_a, s_b) = \frac{\sum_u t_{au} c_{au} + \sum_v t_{bv} c_{bv}}{\sum_u t_{au} + \sum_v t_{bv}}$$

jas-parikh-15_image-specificity

Attention in Image Captioning

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

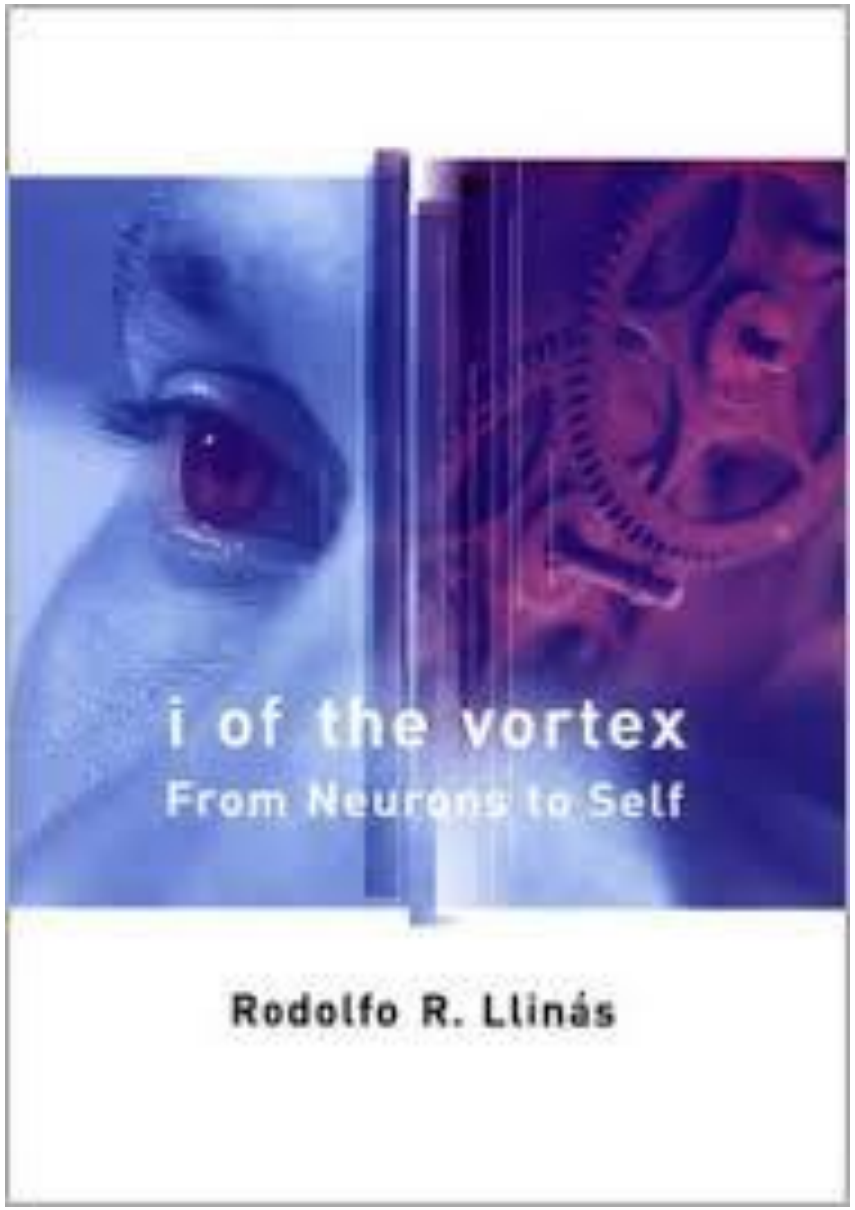
What is “attention”?

Biological Origins of Mind and Language

Motor origins of the mind

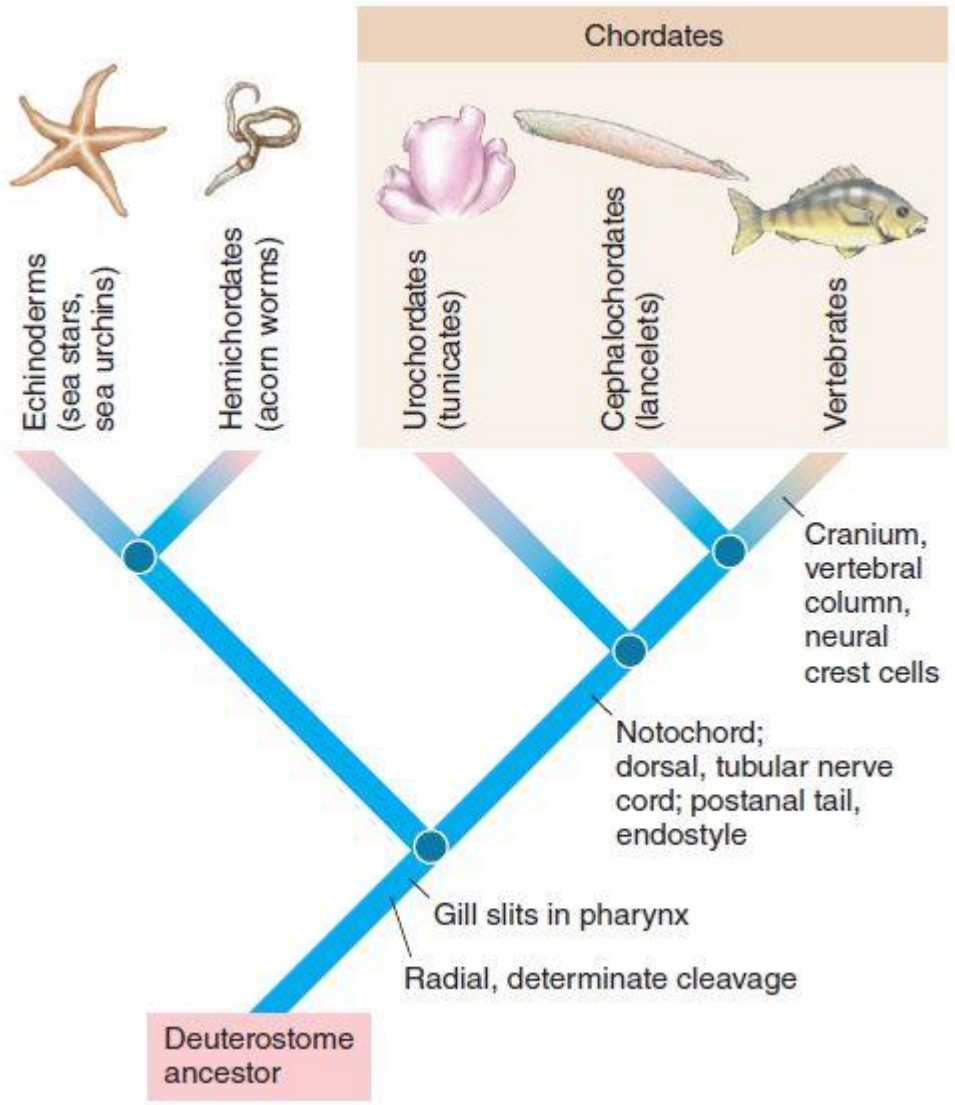
Rodolfo Llinas

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

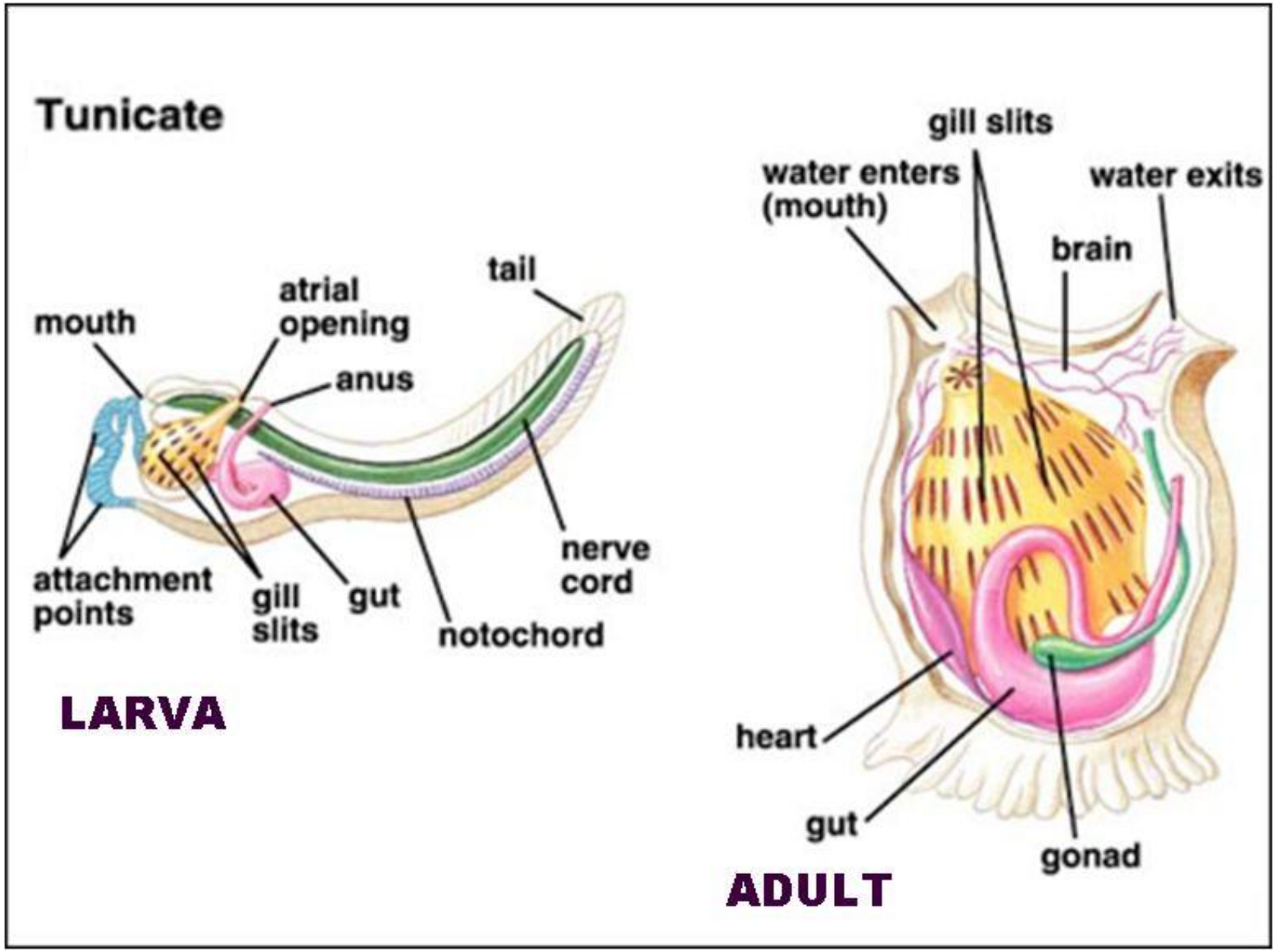
Tunicates (sea squirts) : sessile adults

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.

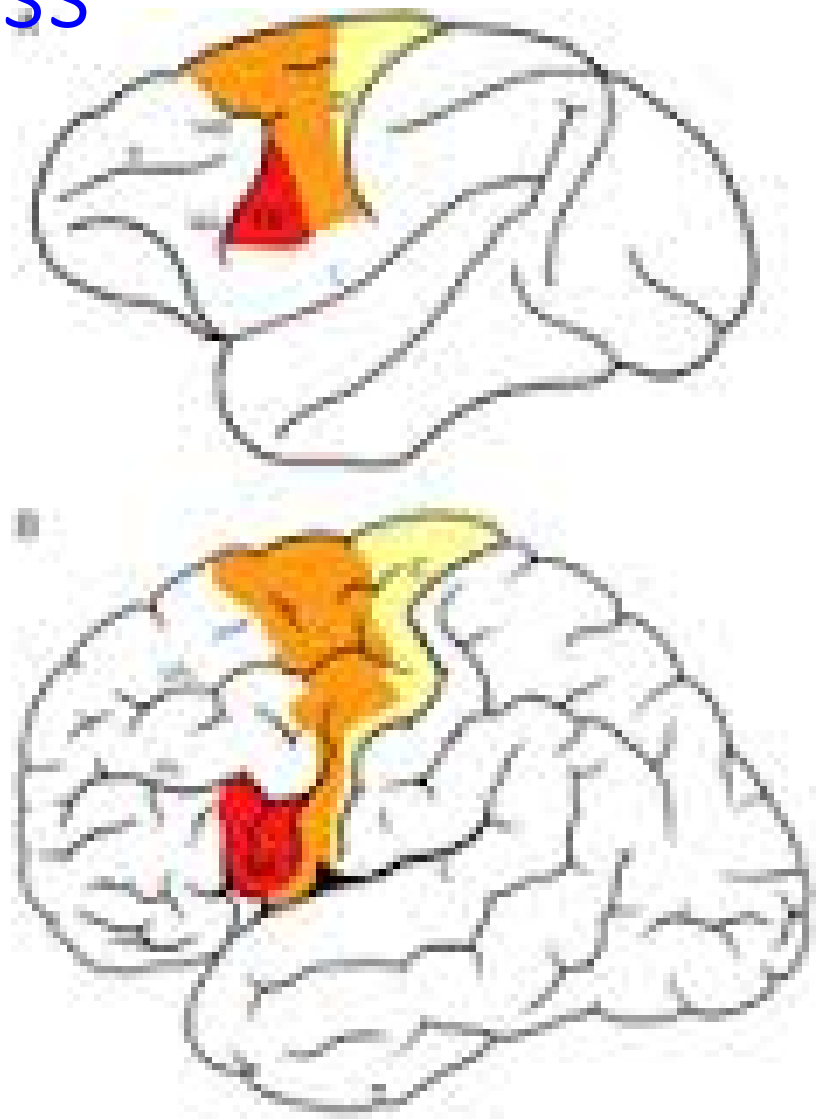
- Rodolfo Llinas

Motor knowledge → Mindness

predictive / intentional interactions

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

organized motricity: cephalization



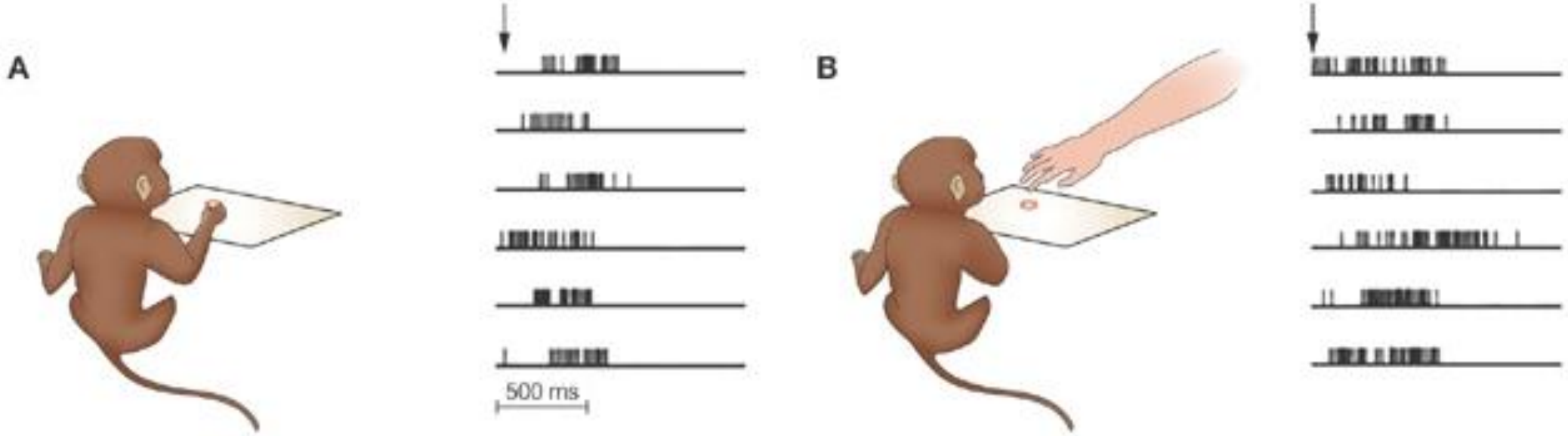
sensory-motor areas in
macaque and human cortex

The Complexity of Language:

Unifying multiple sensory + motor modalities

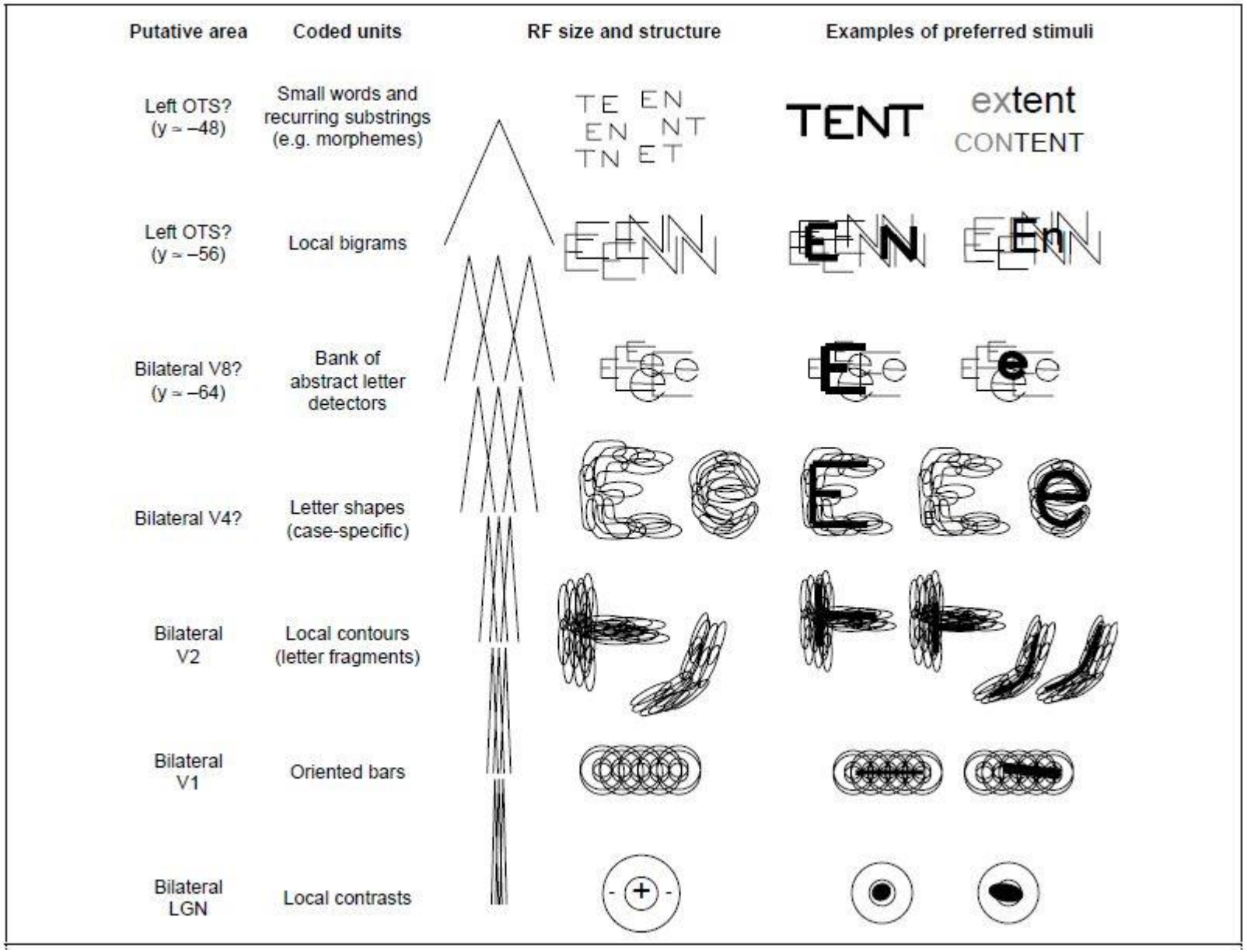
Meaning : Unifying Modalities

Mirror Neurons



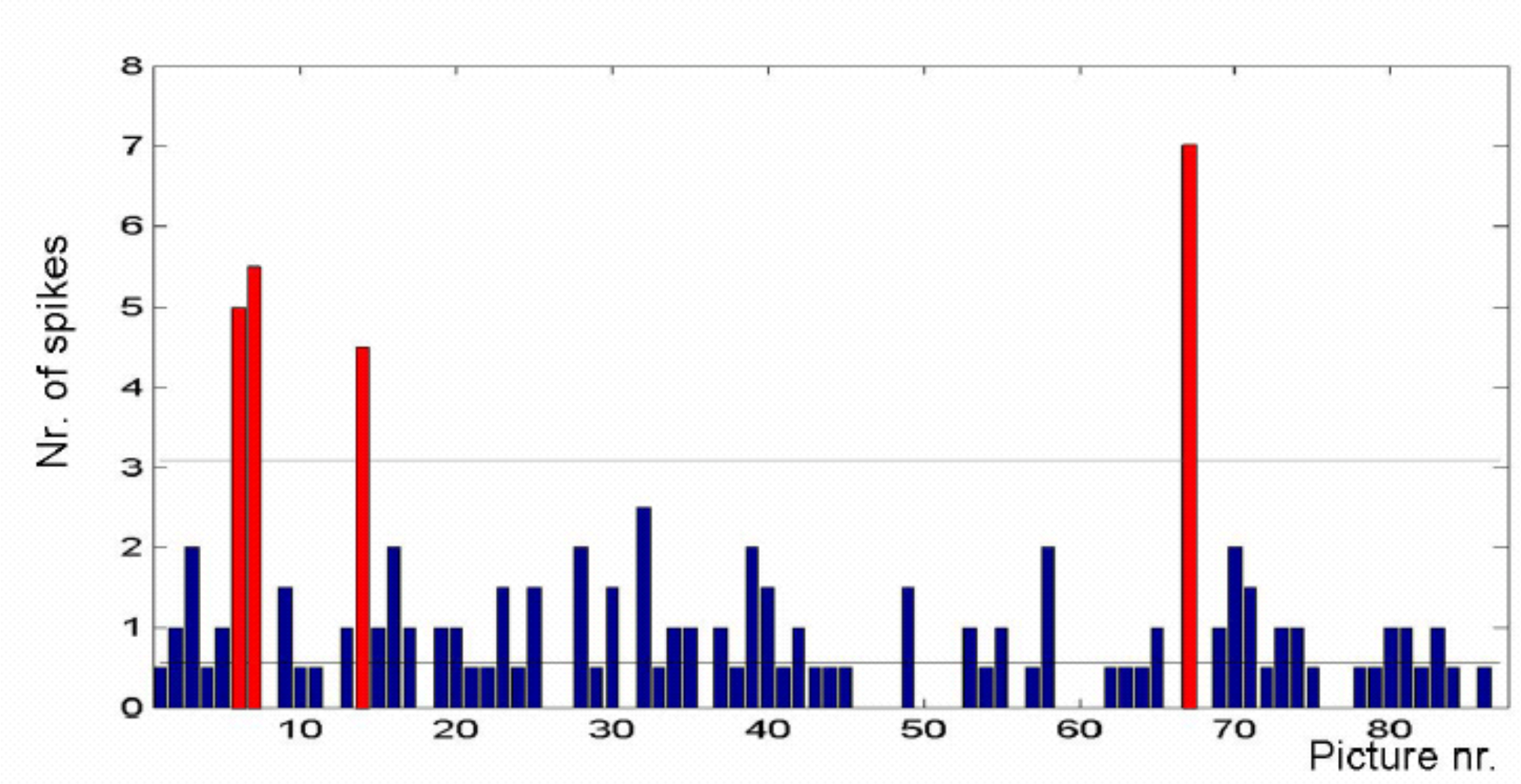
[Rizzolatti G and Fabbri-Destro M 08]

Reading in the brain



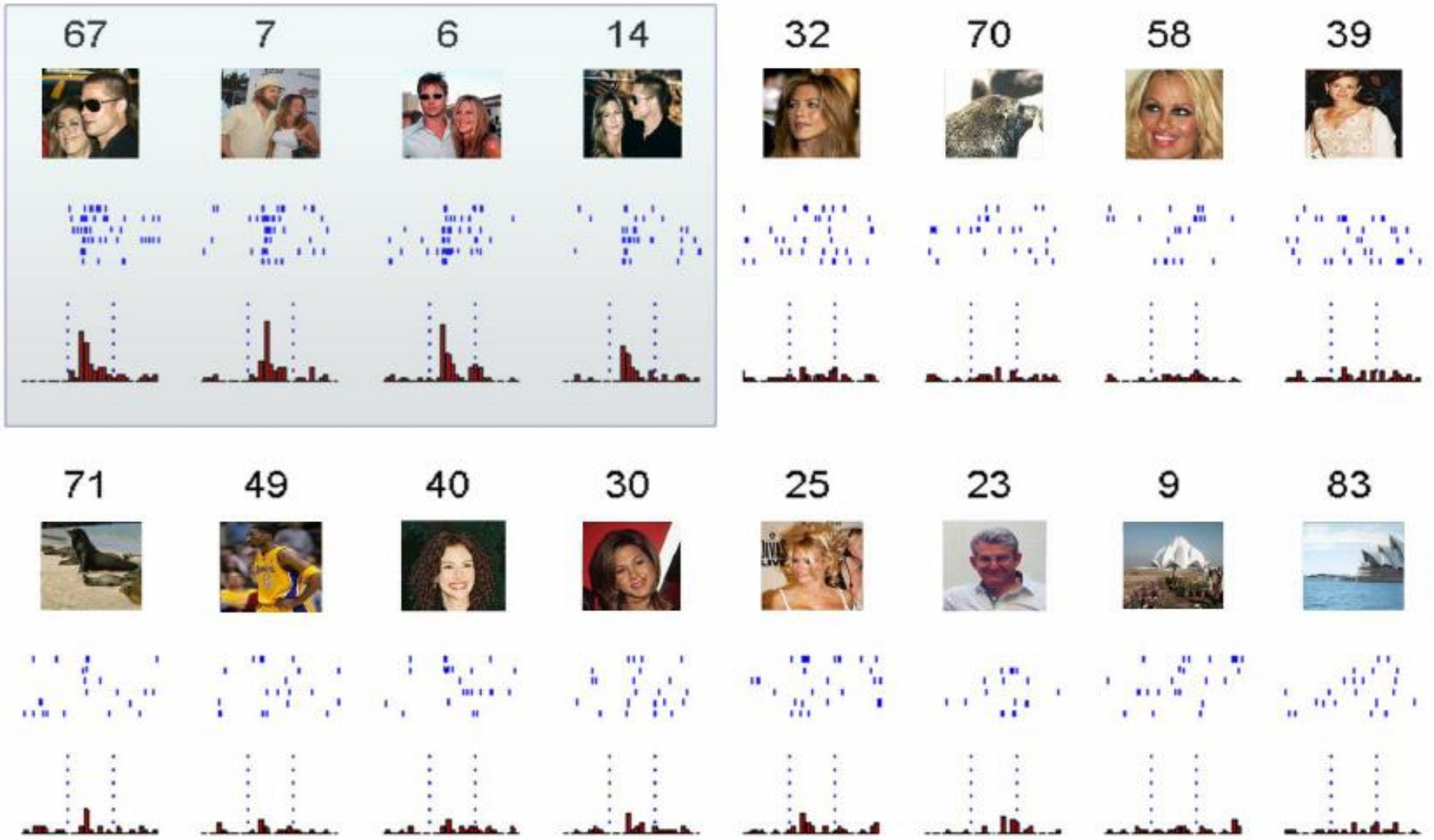
[dehaene-cohen-05_neural-code-for-written-words-proposal]

Visual Recognition: IT Cortex

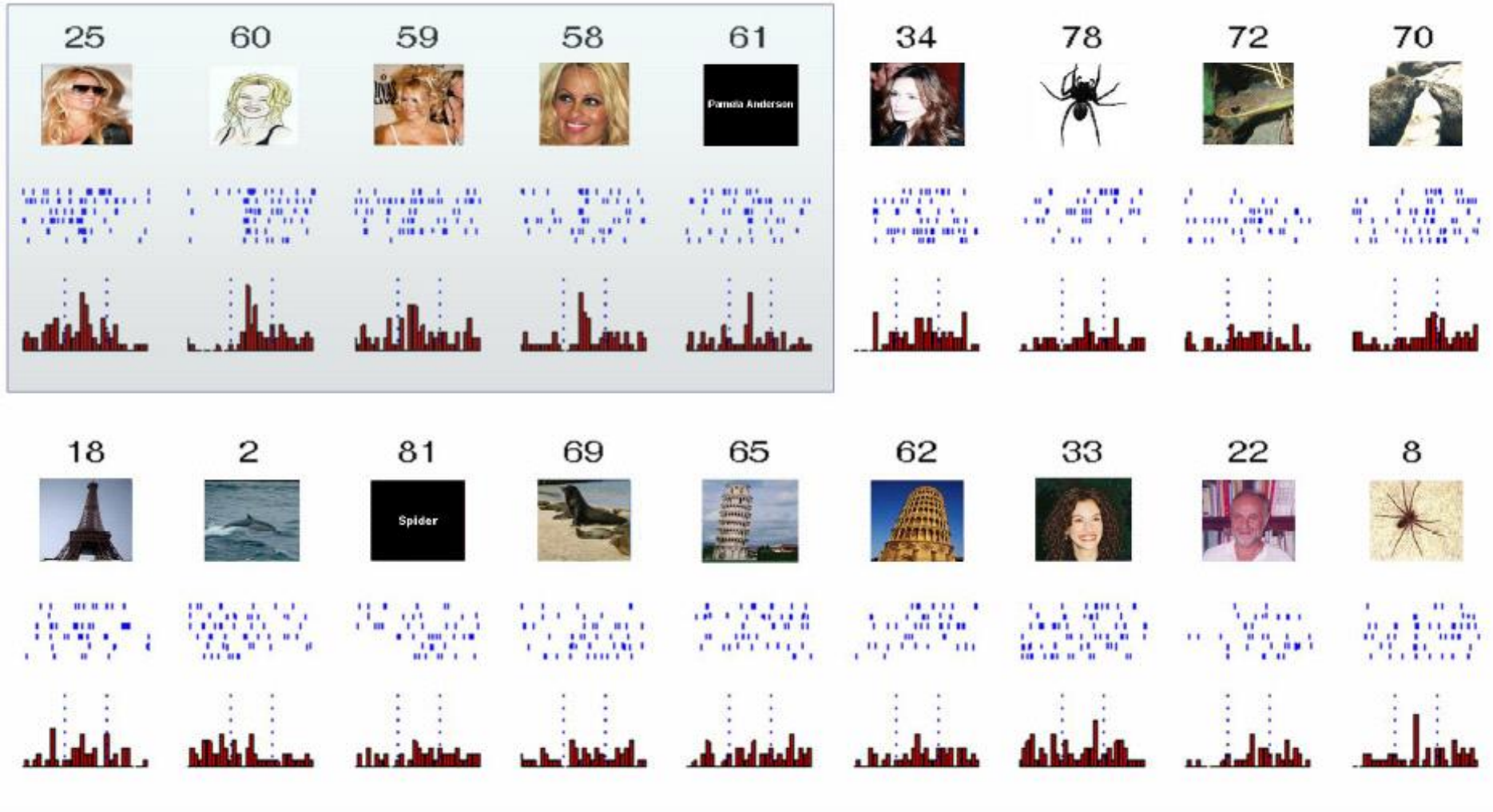


Quian Quiroga, Reddy, Kreiman, Koch and Fried, *Nature* (2005)

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 for reinforcement 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)

Behaviourism

"could i have some water"

→ someone brings him a glass of water.

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

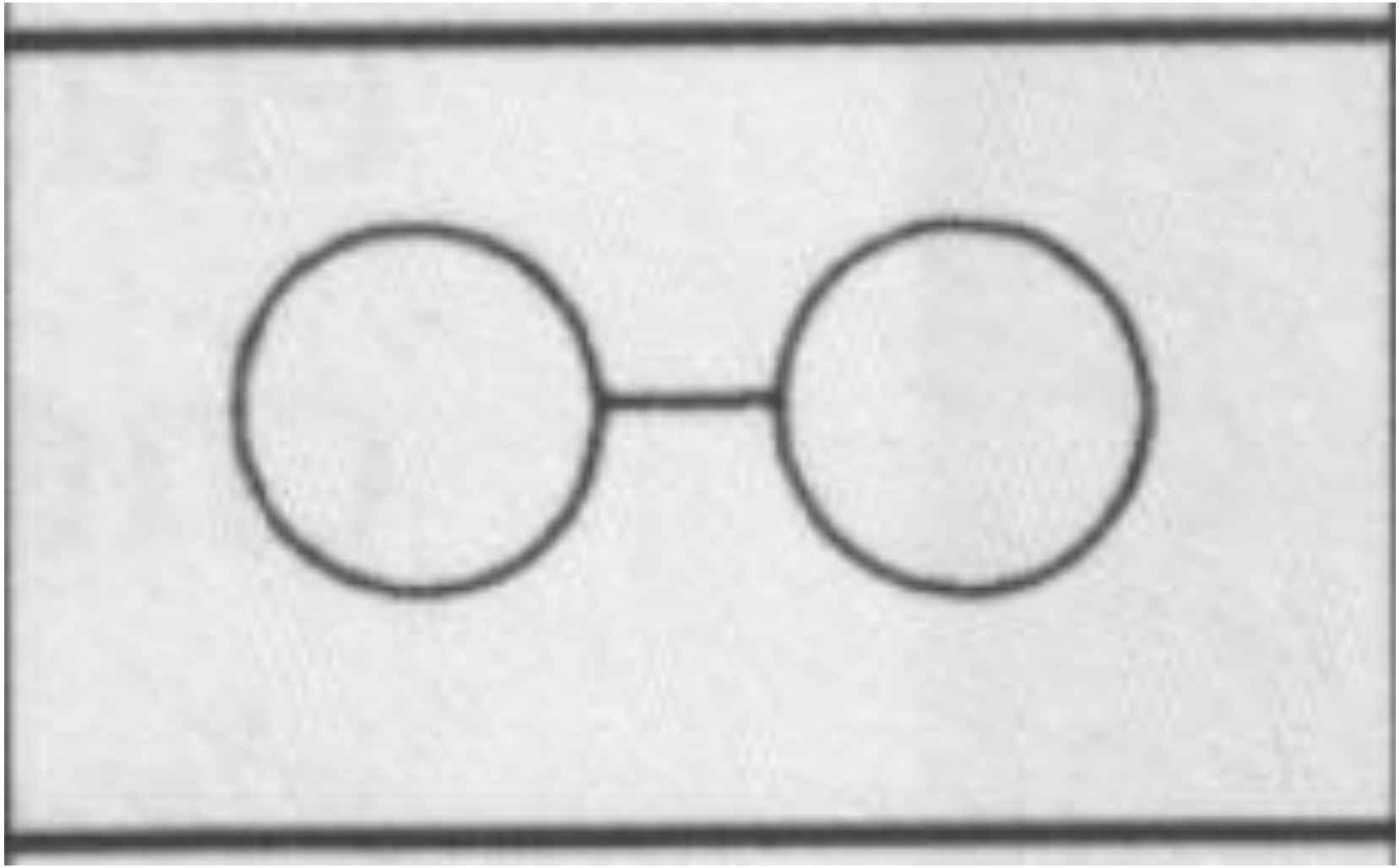
Also probabilities are not possible since various word combinations may not have been seen before.

Language - amodal?
Multimodal?

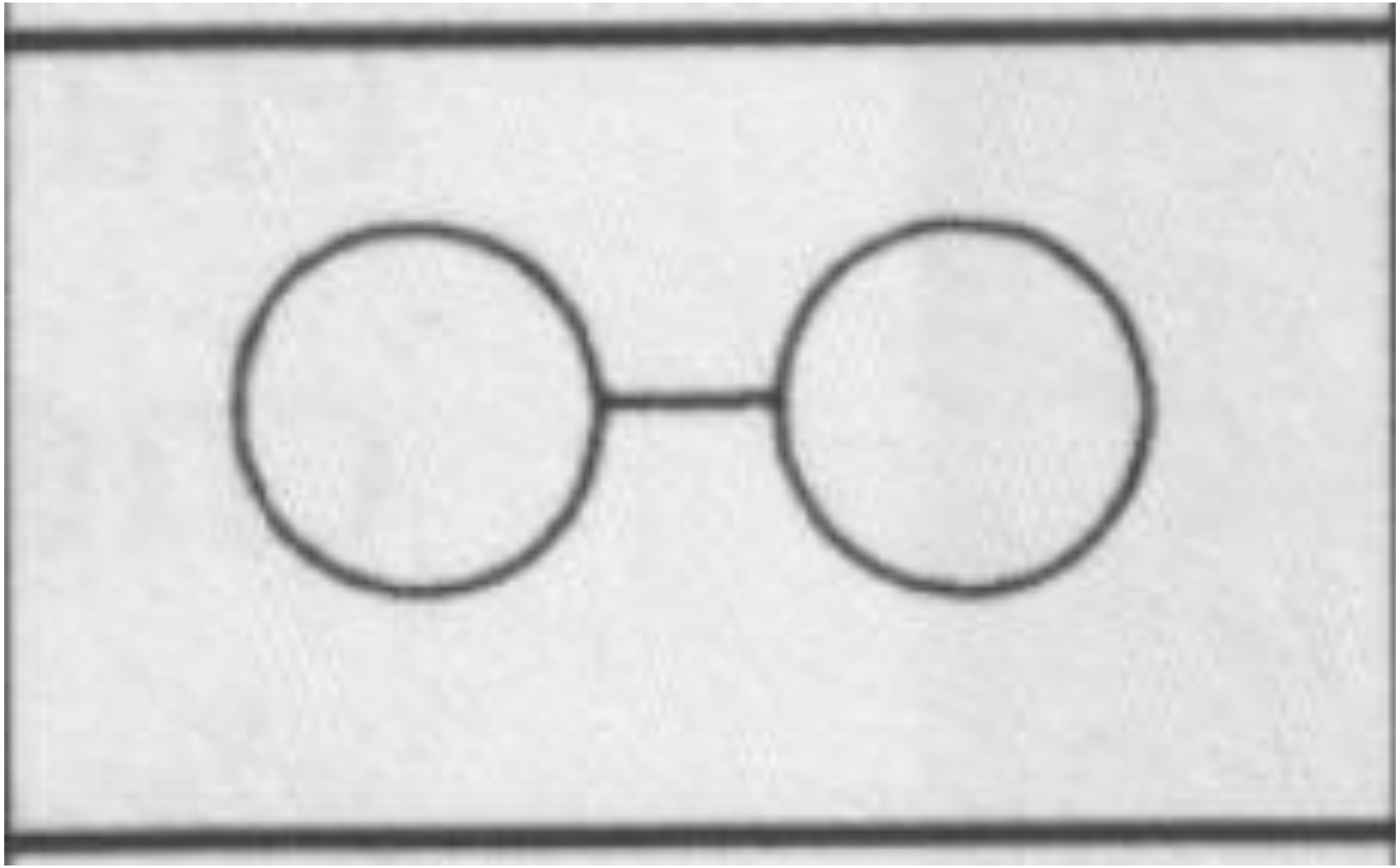
Please read the title and look at the picture

Try to remember both

Eye-glass



Dumb-bell



Perception and Language affect each other

Language and Meaning

Montague Translation [1973]

A student sleeps

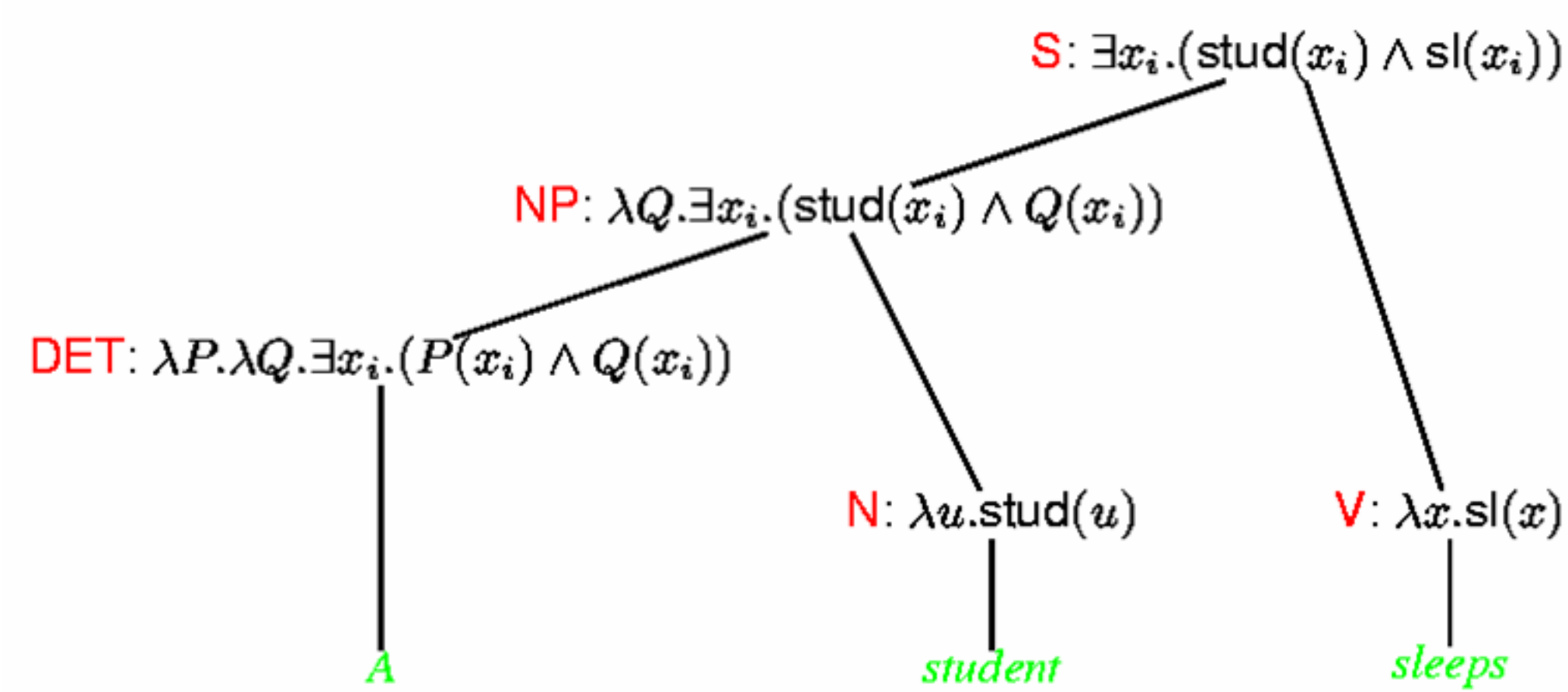
Lexicon:

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

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

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

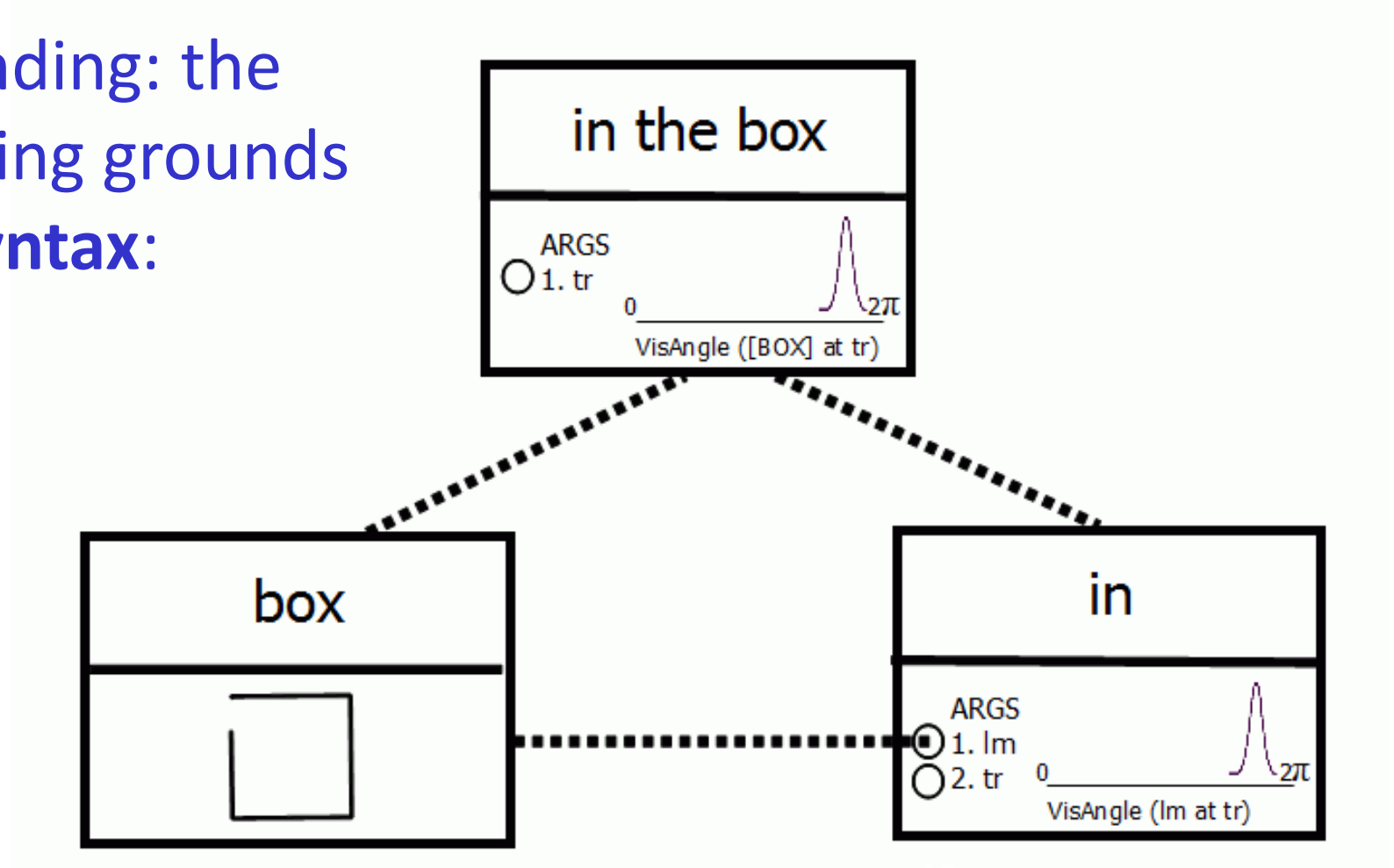
Montagovian Translation [1973]



[Kohlhase]

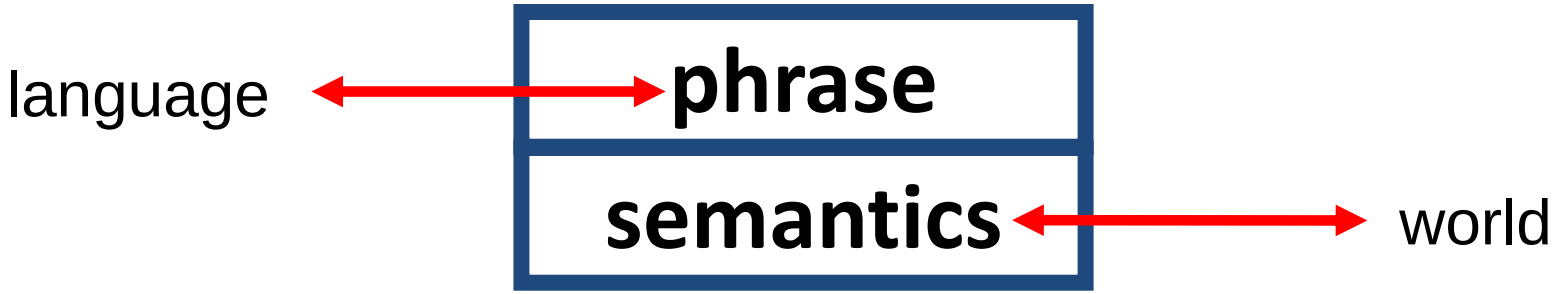
Cognitive Grammar (Langacker)

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



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

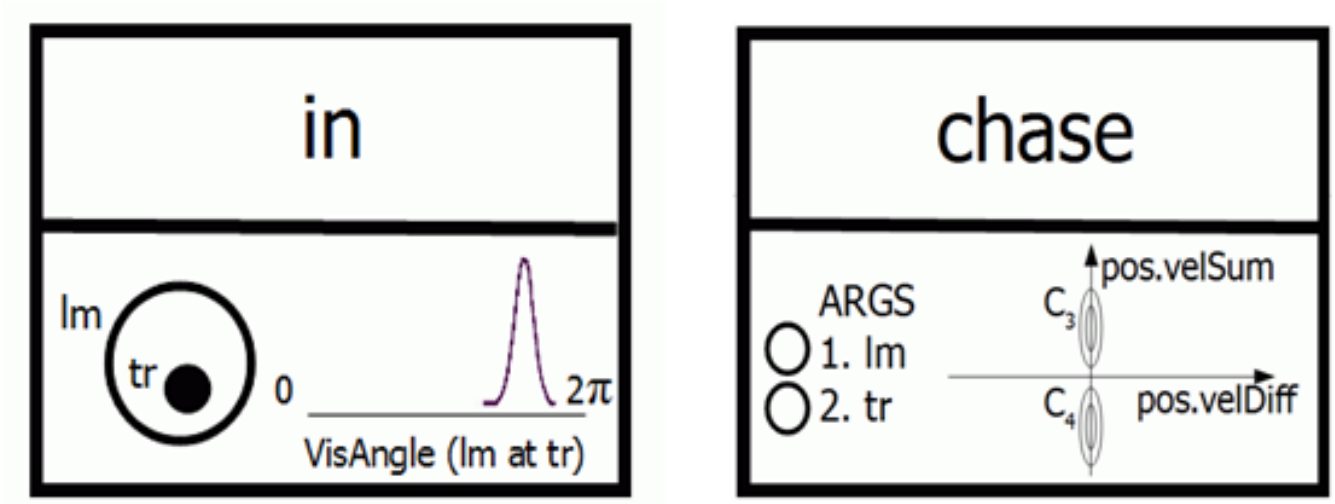
- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Lexicon

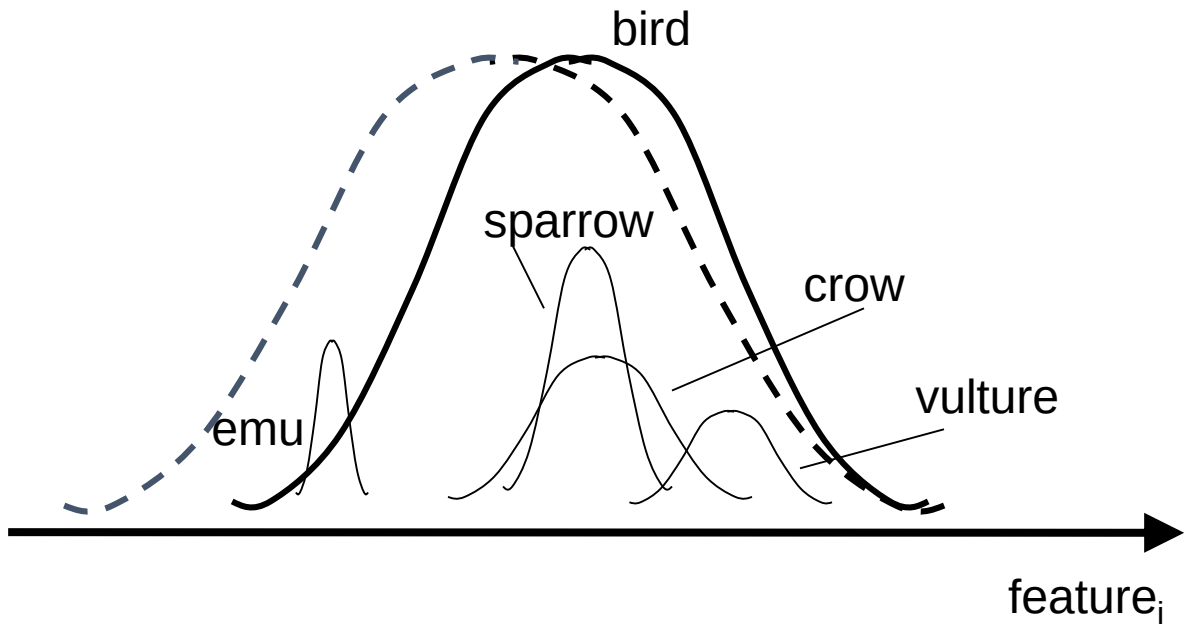
- grounded **lexicon**:



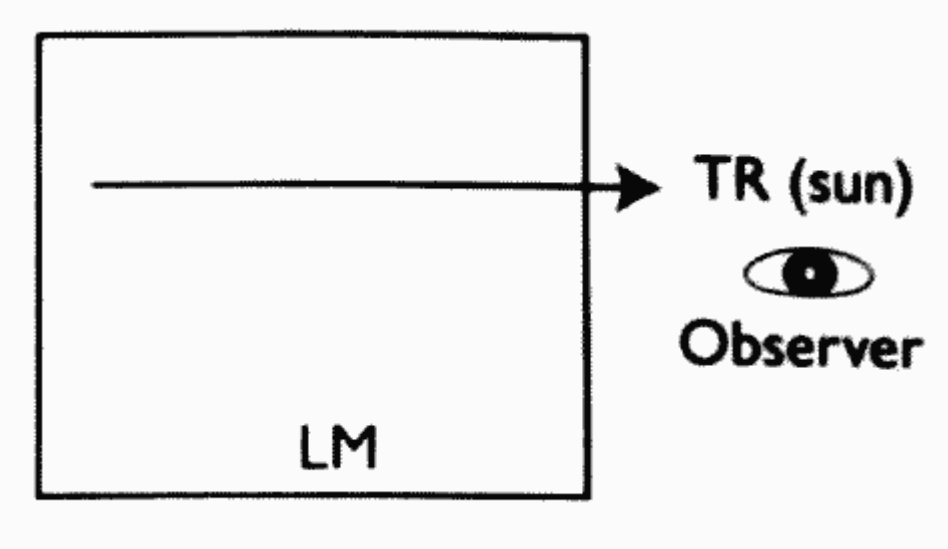
- semantic pole : perceptual patterns (image schemas)
  → probabilistic predicate + arguments

Evolving Semantics

Conceptual Space



Perspective? Idiom?

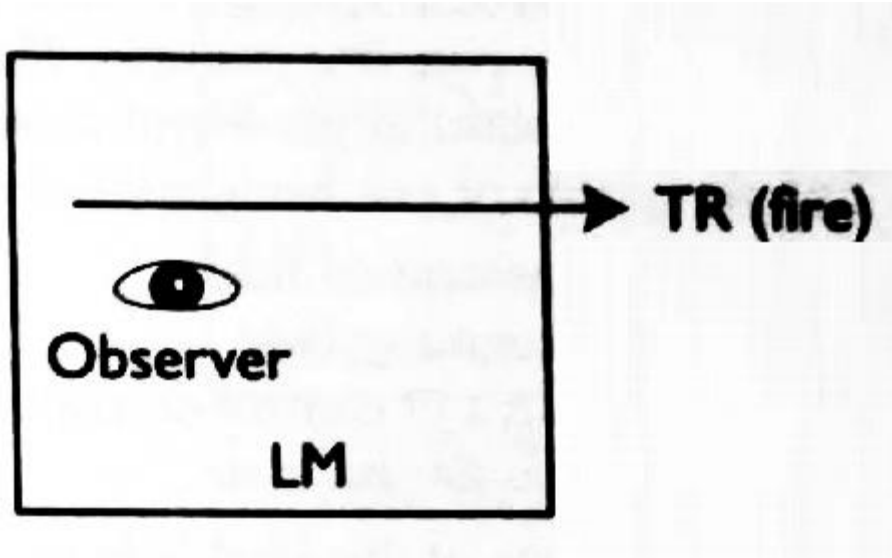


The sun came out.

The secret is out

The fire went out.

The music was
drowned out by noise.



Cognitive Grammar View:

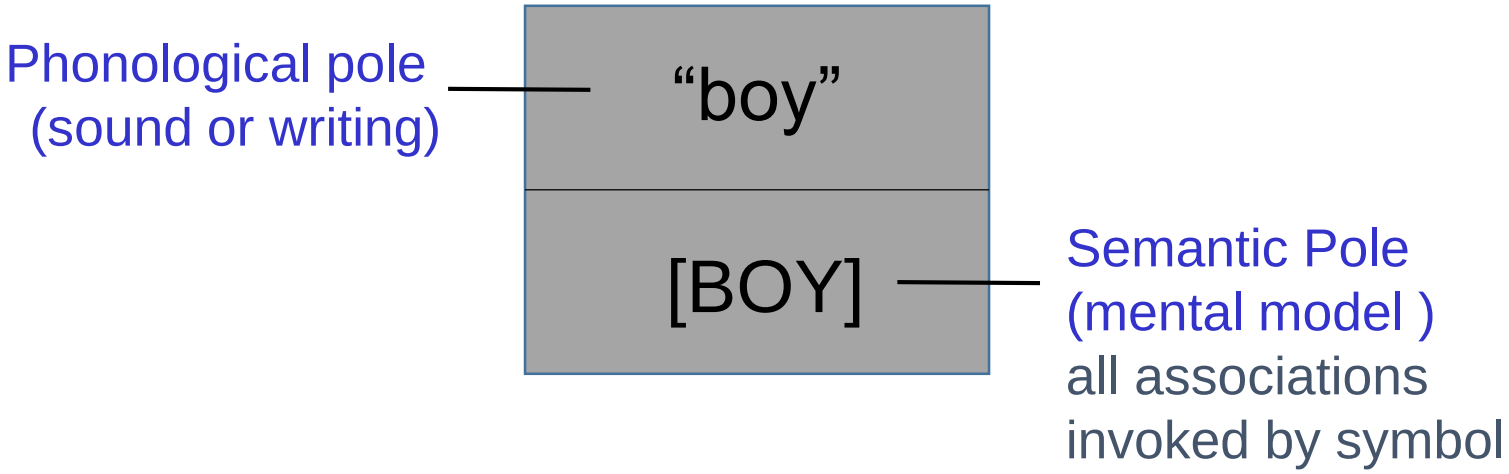
Lexicon vs Grammar

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

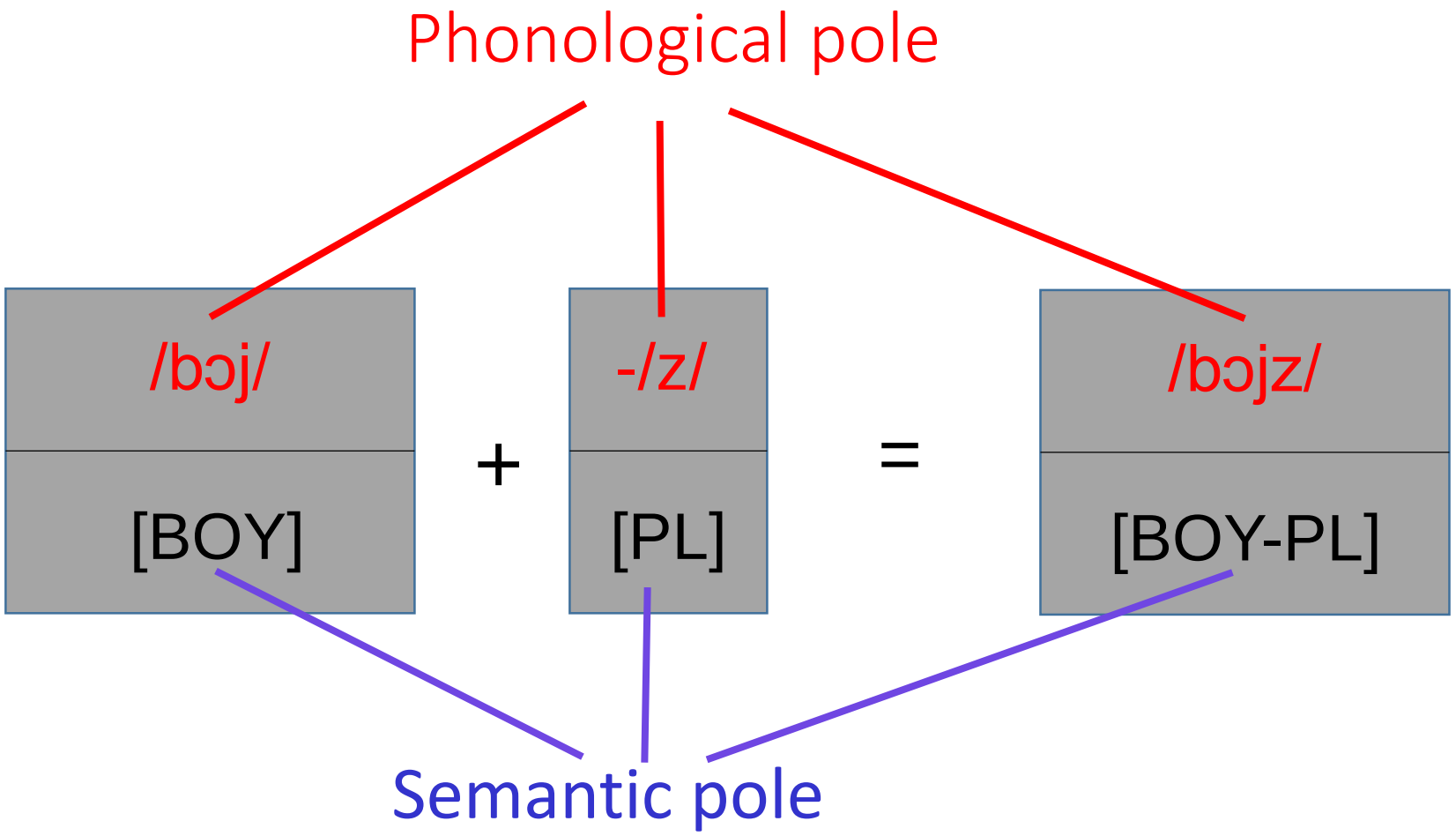
Cognitive Grammar View:

Symbolic Unit

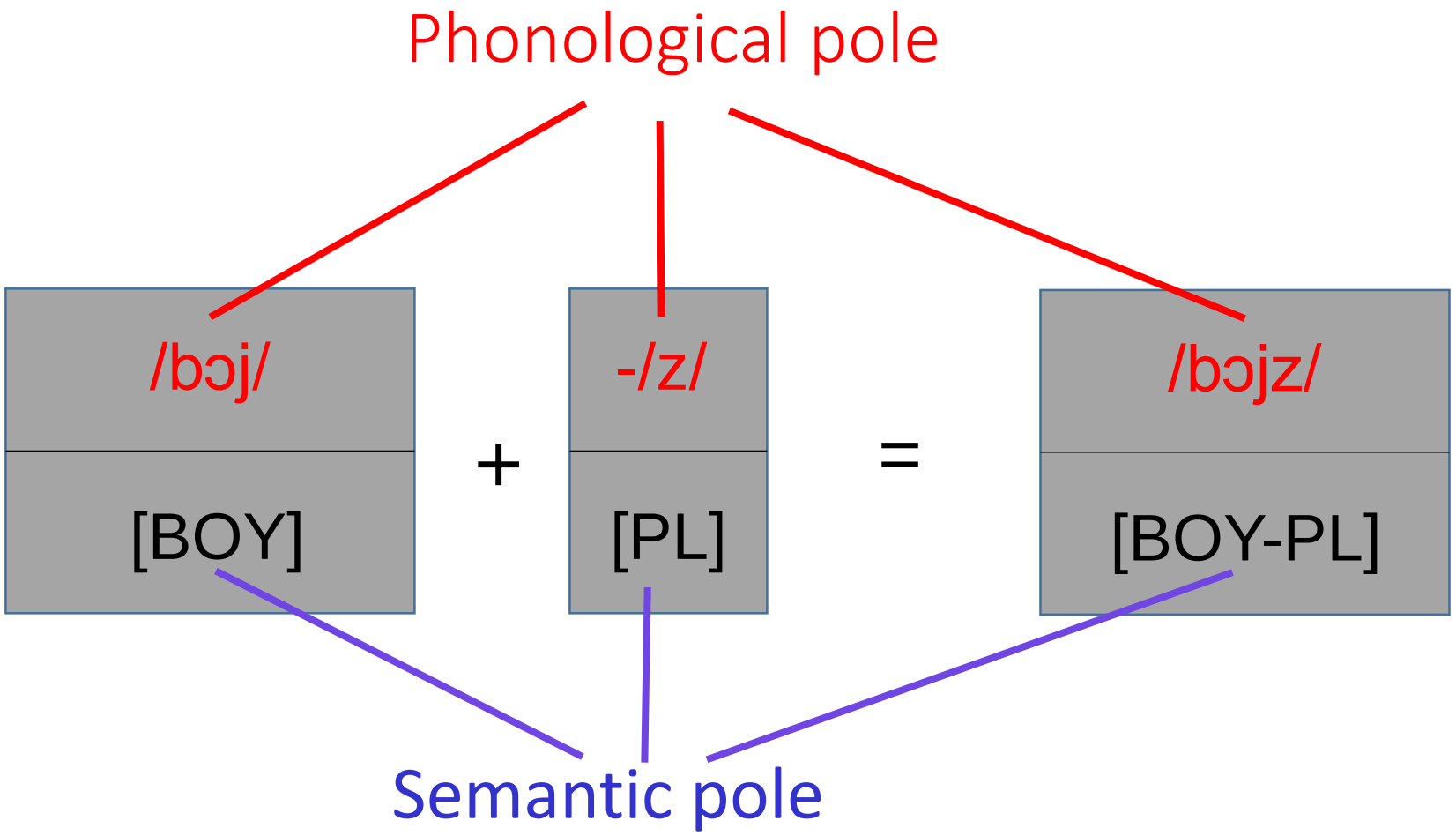


symbol: interrelation between thought, meaning, and linguistic structure

Combining phonemes

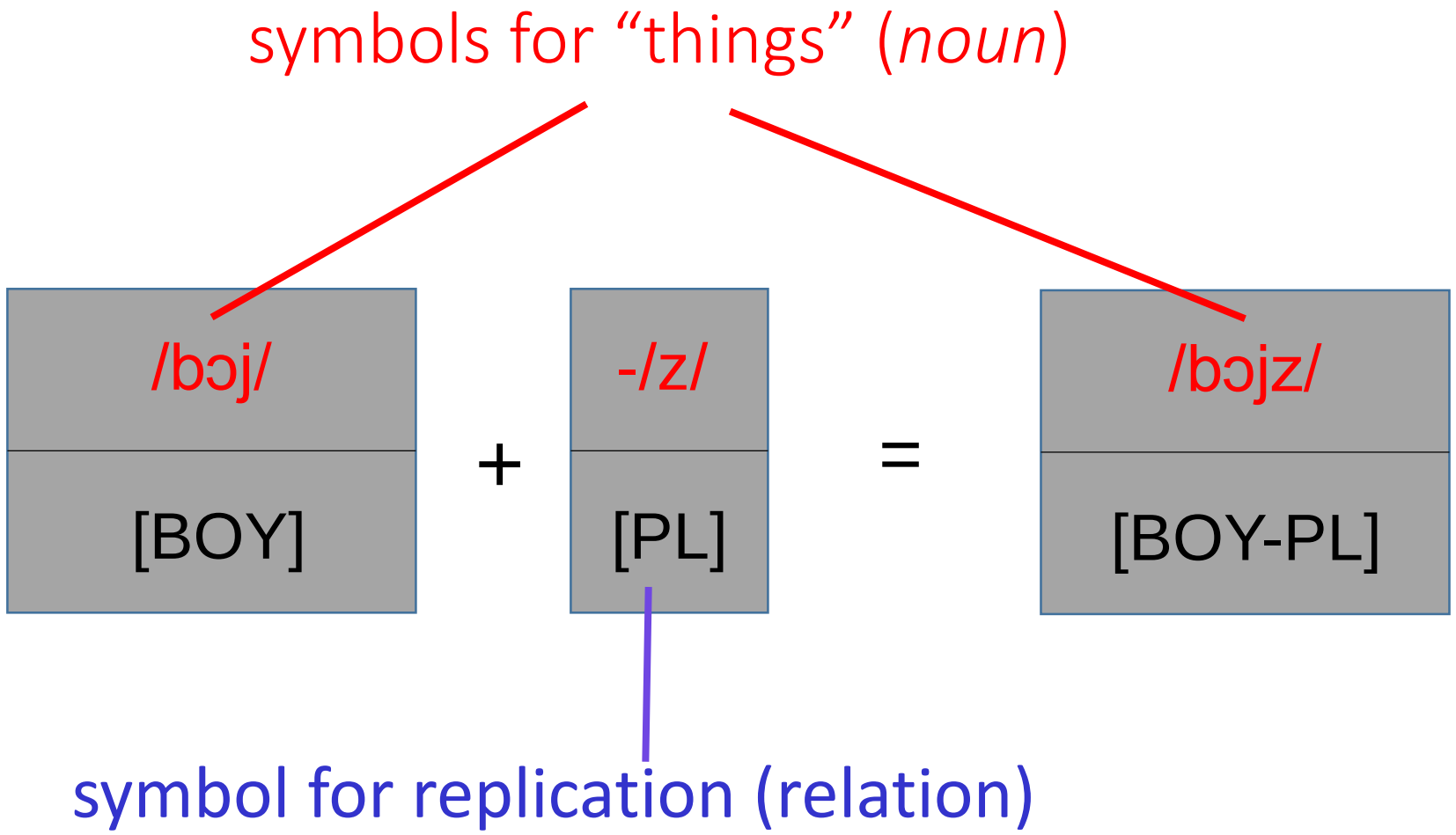


Cognitive Grammar : Inflection



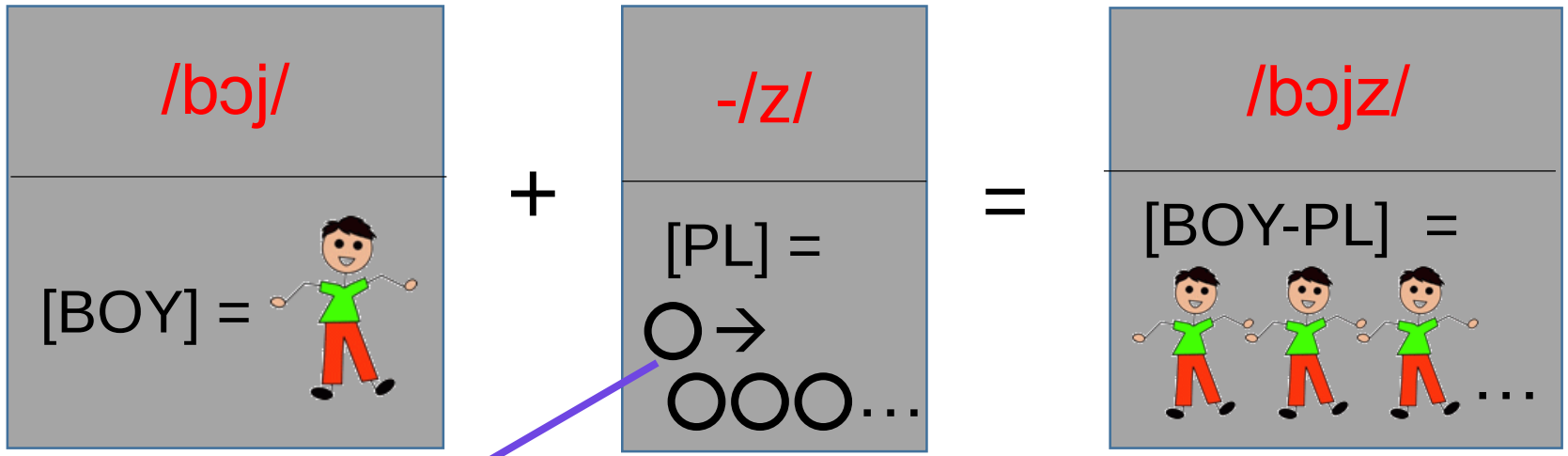
symbolic complex [langacker 87]

Cognitive Grammar : Inflection

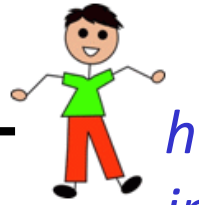


symbolic complex [langacker 87]

Semantic Pole: Image Schemas

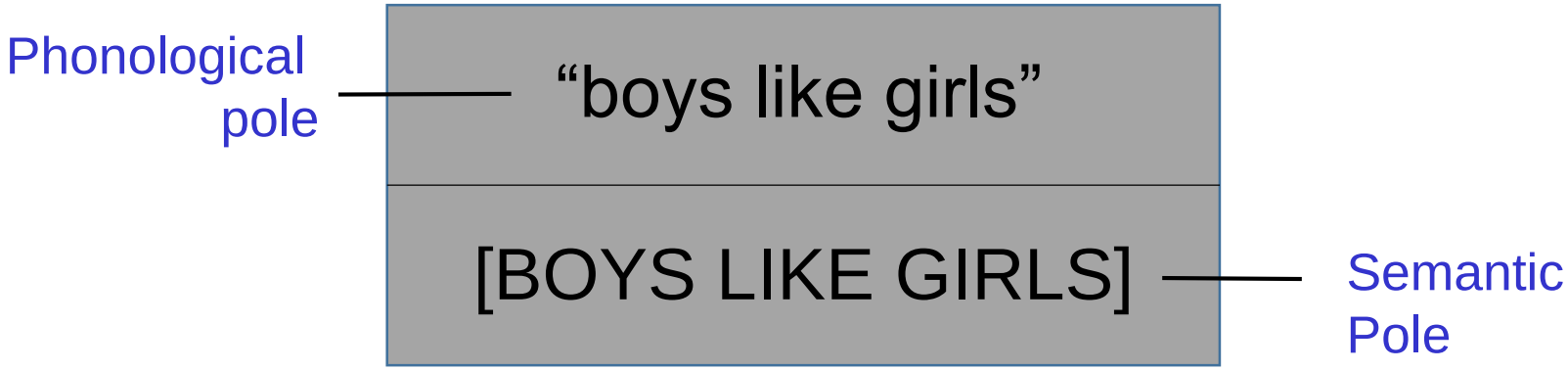


must be countable

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

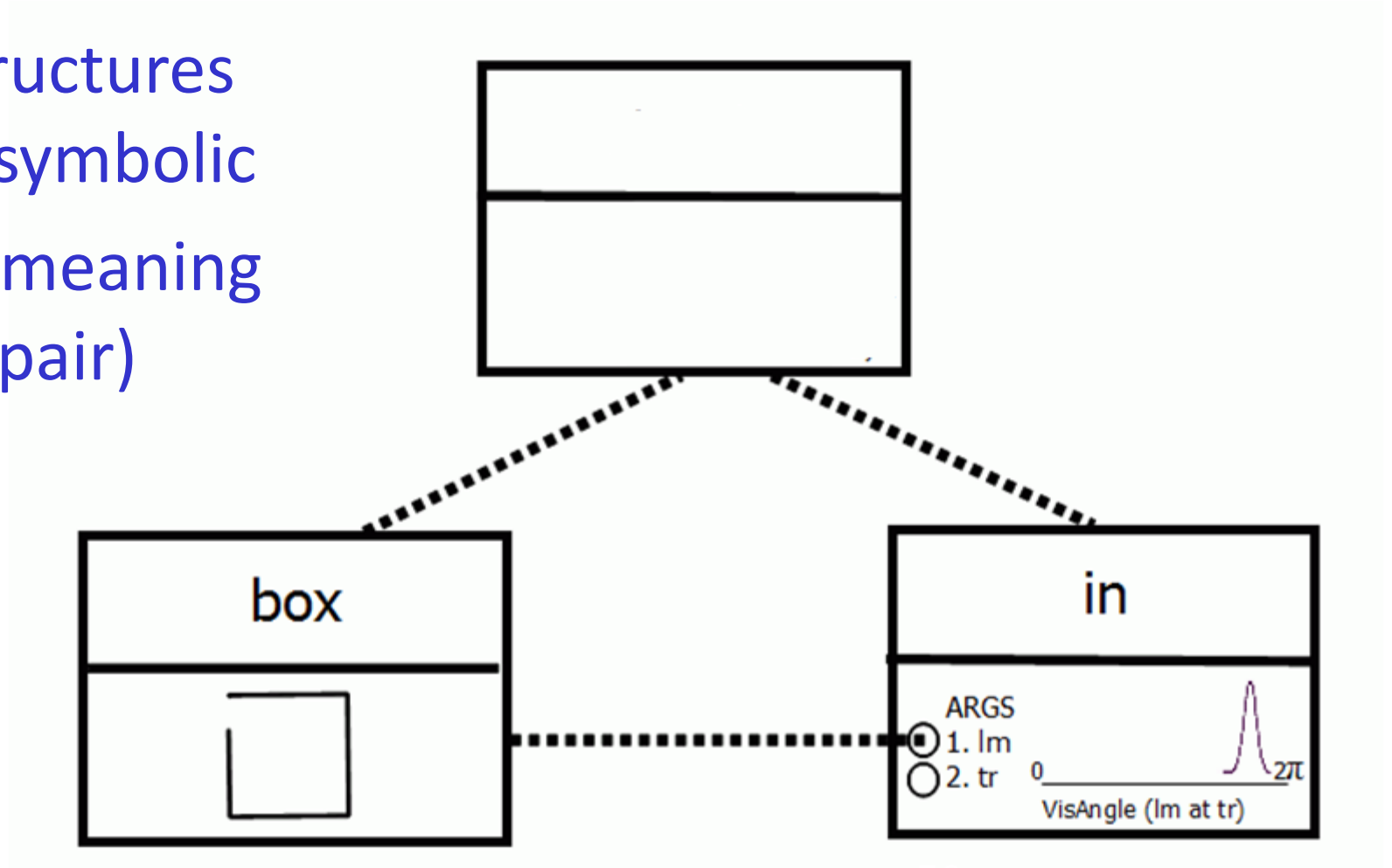


Grammar: applies to the composition of both phonological pole (surface form) and semantic pole (meaning)

Cognitive Grammar (Langacker)

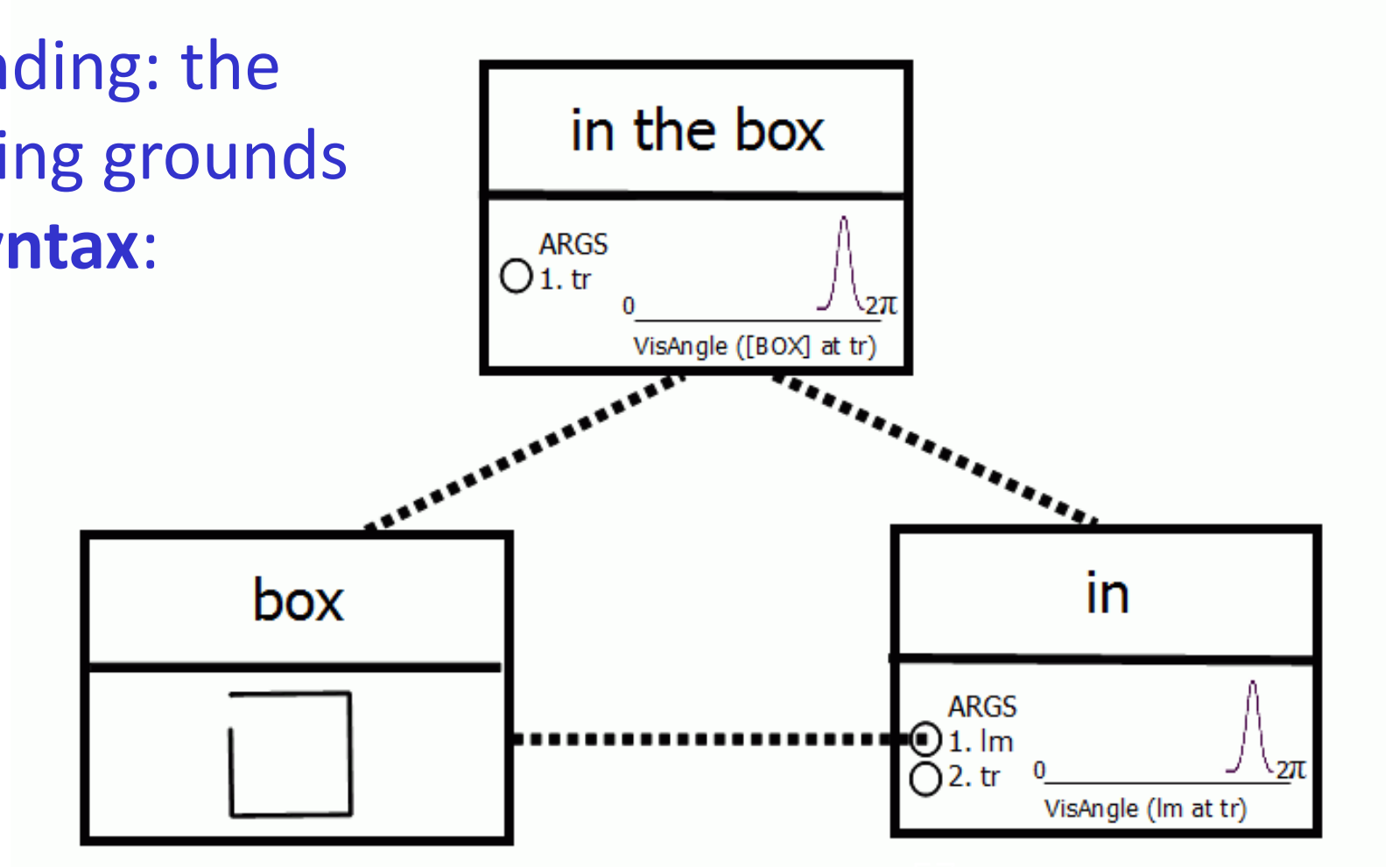
All structures are symbolic

(form-meaning pair)



Cognitive Grammar (Langacker)

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

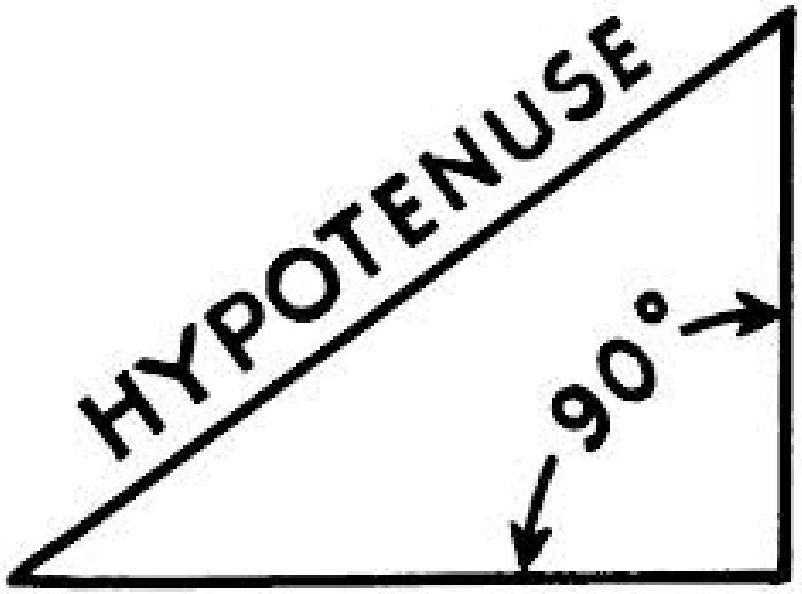


Language is Symbolic

- “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

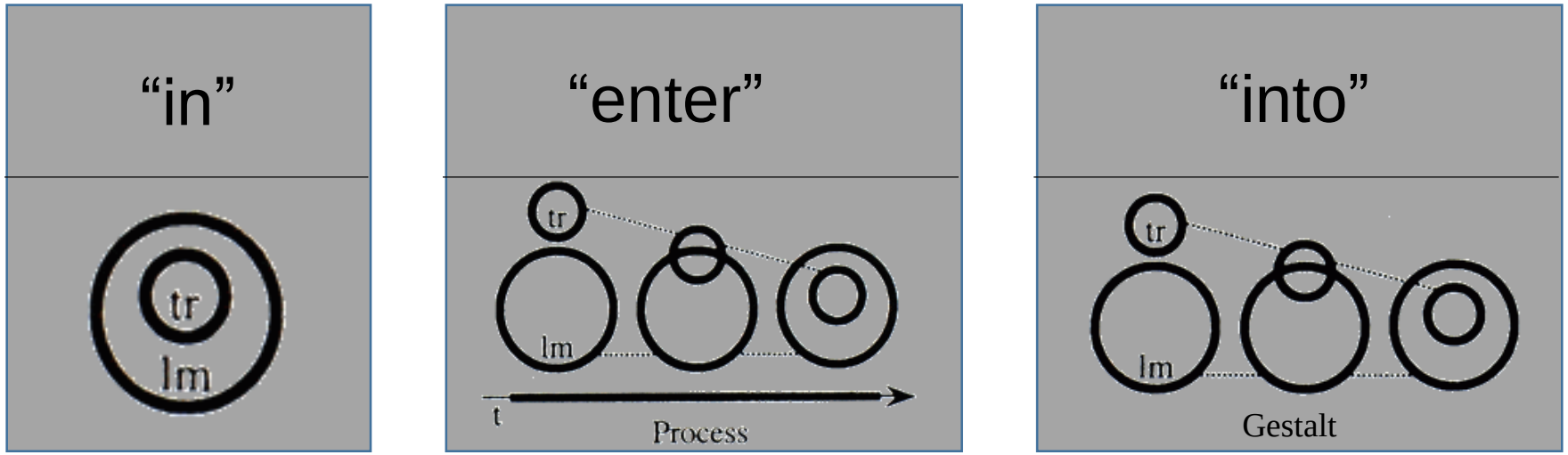
[wicket]: frame =
game of cricket

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

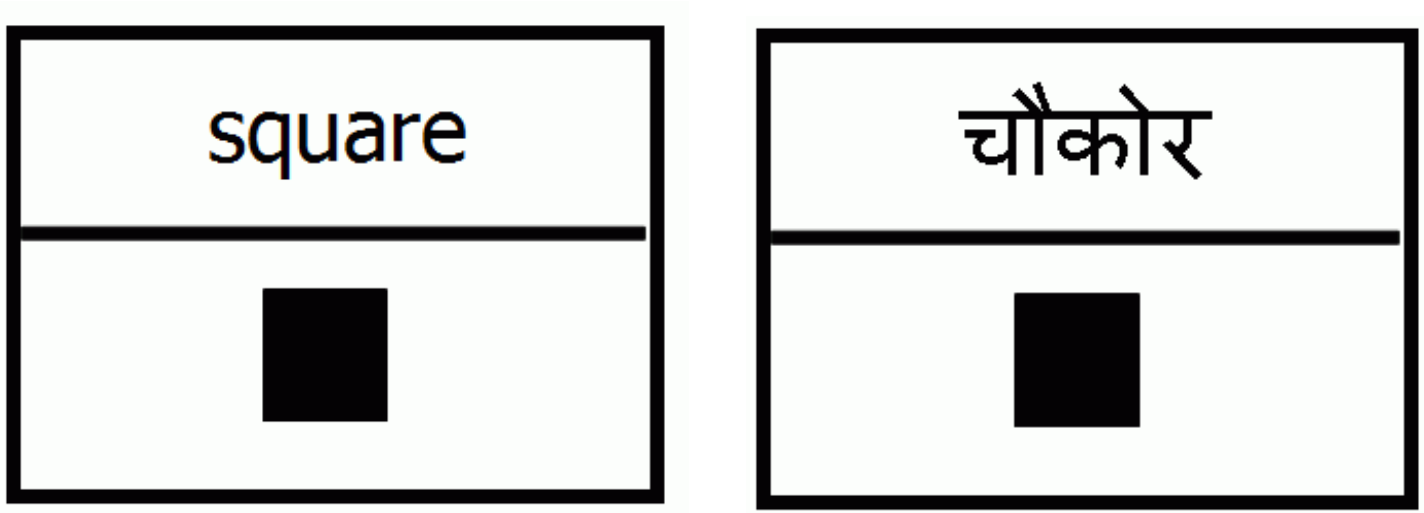
- grounded **syntax**:
  mapping from syntactic patterns to objects, relations or events in perceptual space

- Units for language = form-meaning pairs

[langacker 87] [bergen etal 04]

Lexicon

- grounded **lexicon**:



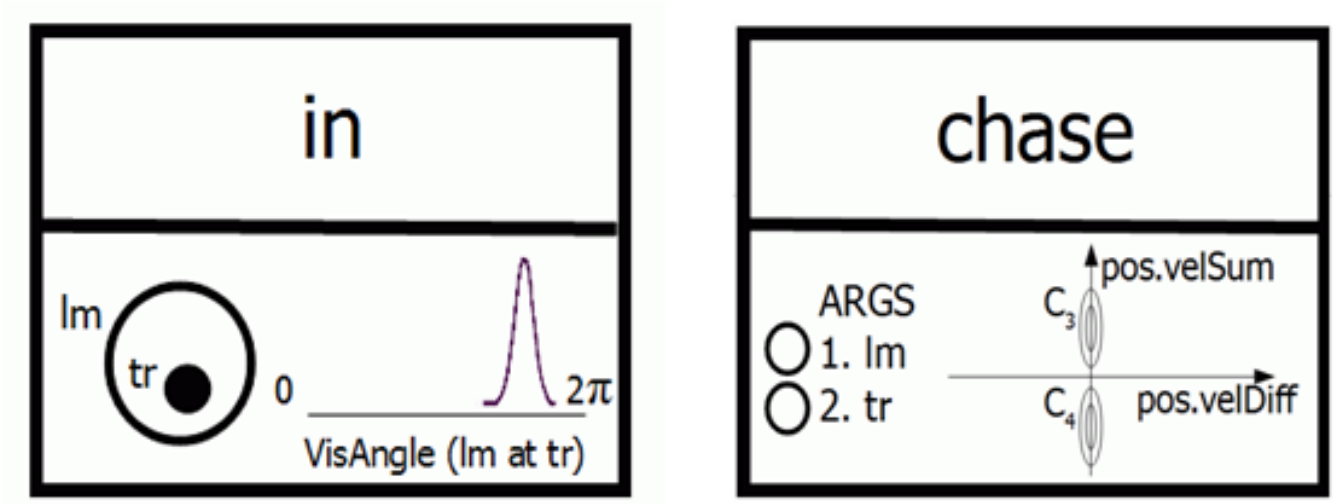
english lexicon

hindi lexicon

[langacker 87]

Lexicon

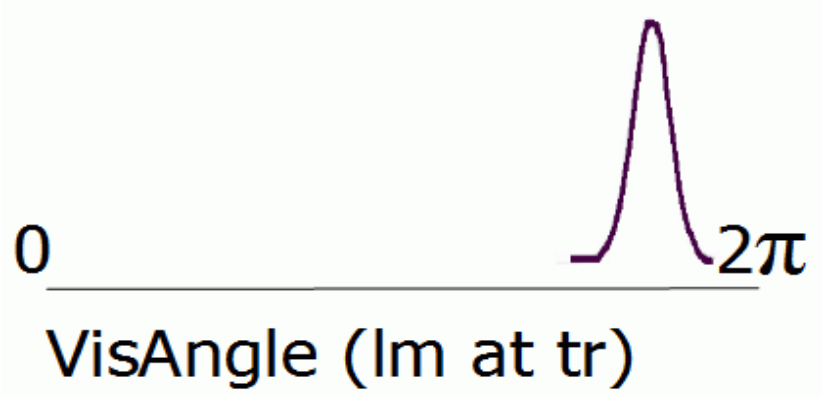
- grounded **lexicon**:



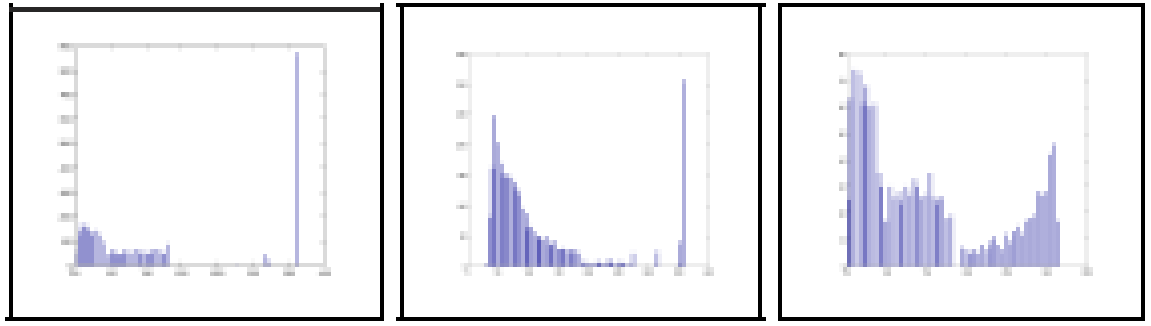
- semantic pole : perceptual patterns (image schemas)
  → probabilistic predicate + arguments

Clustering spatial relations

IN cluster
(emergent)



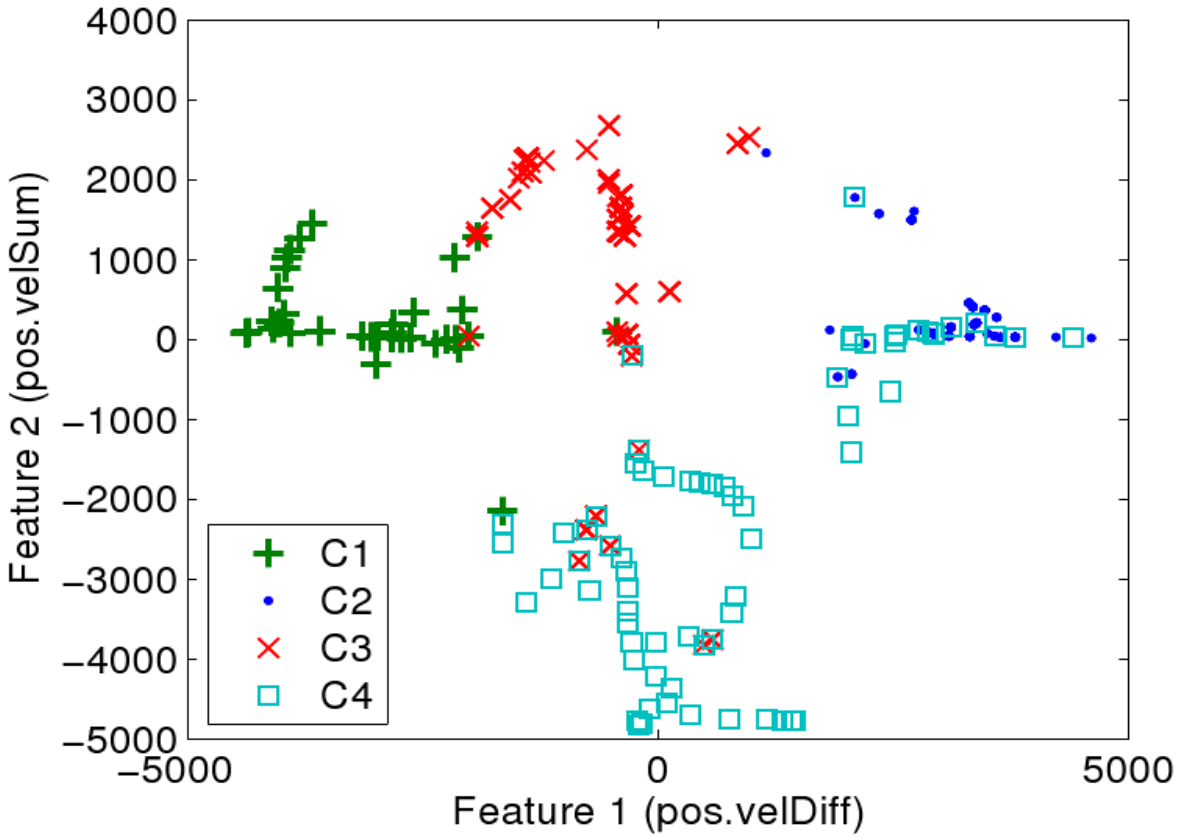
Histogram of visual subtended angle for the 3 shapes



[Sarkar/Mukerjee 07; Nayak/Mukerjee 12]

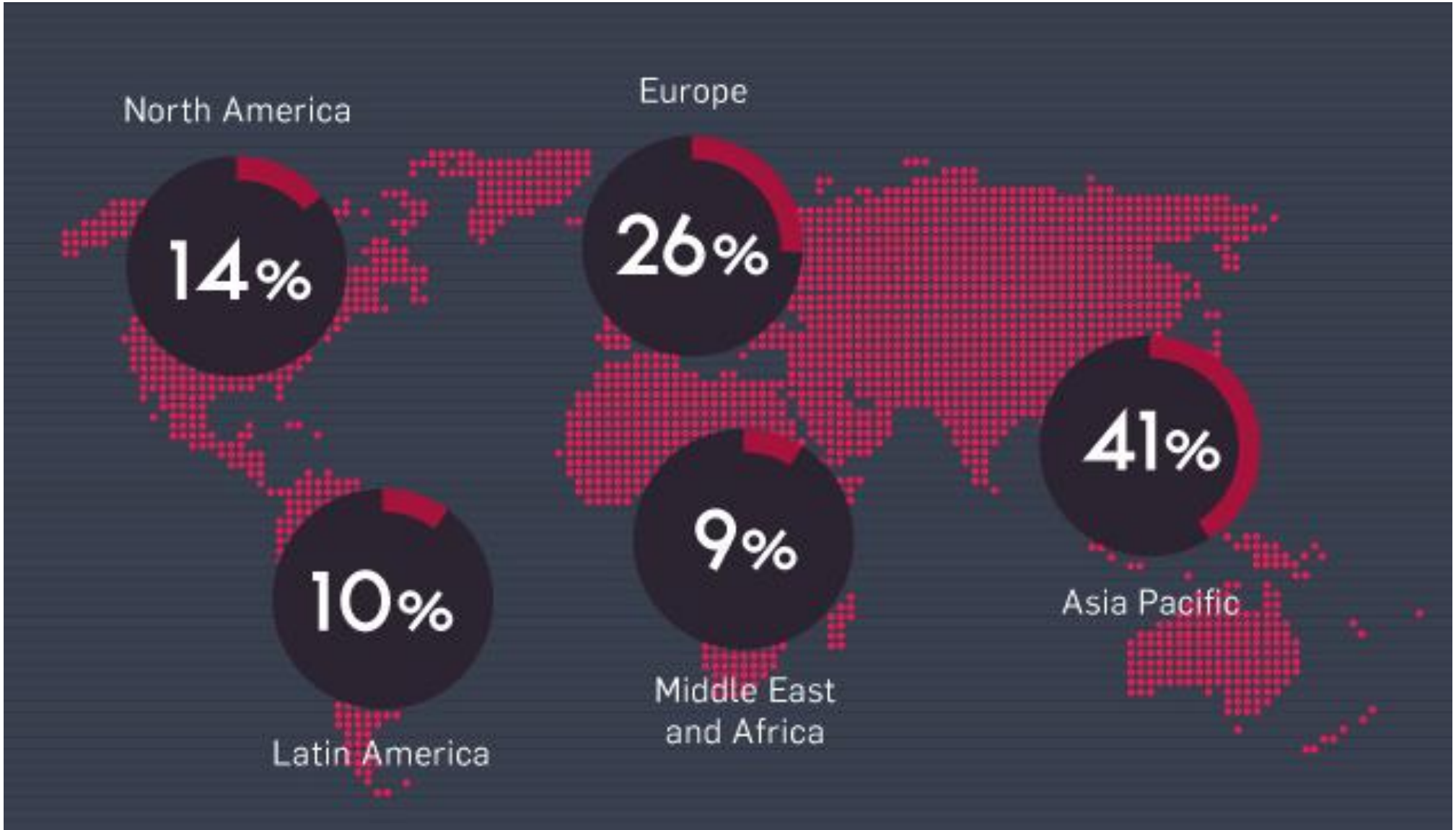
Perceptual Discovery: 2-agent actions

- Static time-shots of feature space trajectories



Language Use Patterns

Web Users Map- 2014



- http://www.statista.com

Perception and Language affect each other

Structure in Language

words may not be space-separated

Structure in language : Word

पांच फिरंगी अफसरों __ फांसी पर ___ दिया

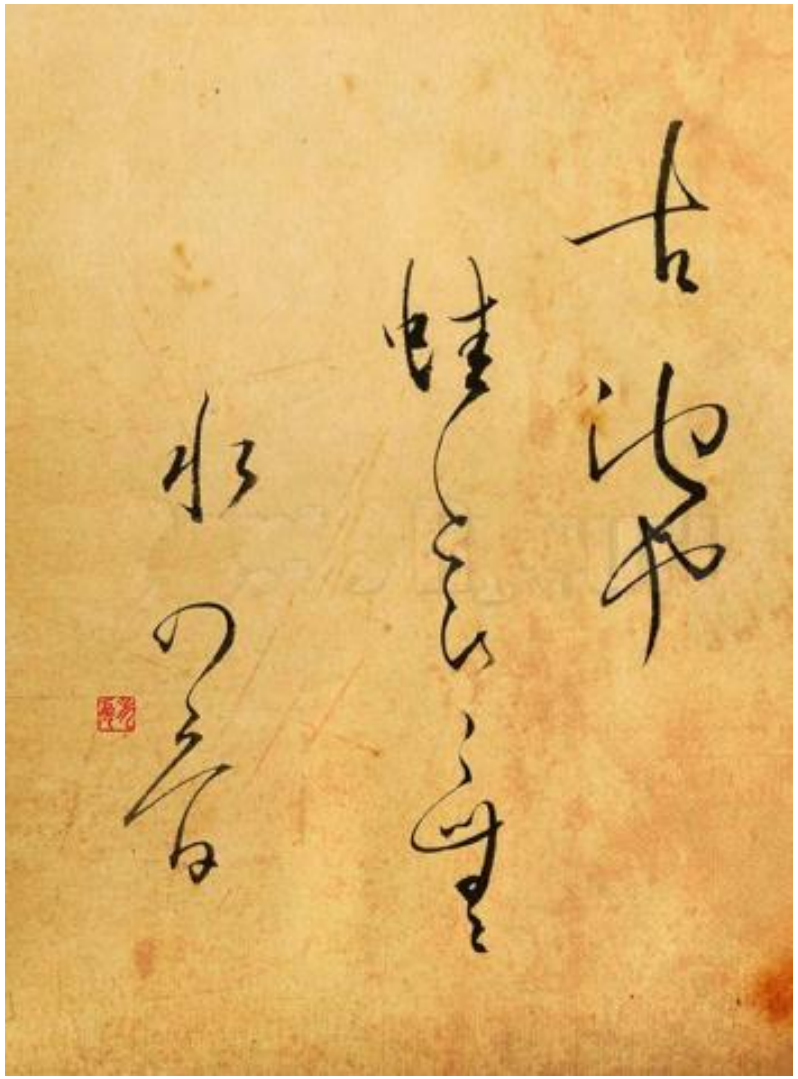
what can go in the blanks?

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.