NLP and Cognitive Science
Translation from Images
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"
Image caption variability

- Specificity = 0.89
  - There is a lot of snow on the mountain.
  - There is a snow covered mountain.
  - 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 around.
  - A large bowling rink.
Image caption variability

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

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

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

\[
sim_{\text{auto}}(s_a, s_b) = \frac{\sum_u t_{au} c_{au} + \sum_v t_{bv} c_{bv}}{\sum_u t_{au} + \sum_v t_{bv}}
\]
Attention in Image Captioning
Describing images with attention
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”?
Motor origins of the mind

Rodolfo Llinas

I of the Vortex, 2002
Motricity $\rightarrow$ 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)
Nervous system:
Evolved for planning motions

planning ← prediction
Predicting ➔ Planning
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

<table>
<thead>
<tr>
<th>Putative area</th>
<th>Coded units</th>
<th>RF size and structure</th>
<th>Examples of preferred stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left OTS? (y = -48)</td>
<td>Small words and recurring substrings (e.g., morphemes)</td>
<td>T E N</td>
<td>TENT extent</td>
</tr>
<tr>
<td>Left OTS? (y = -56)</td>
<td>Local bigrams</td>
<td>E N N</td>
<td>E N N E N N</td>
</tr>
<tr>
<td>Bilateral V8? (y = -64)</td>
<td>Bank of abstract letter detectors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral V2</td>
<td>Local contours (letter fragments)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral V1</td>
<td>Oriented bars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bilateral LGN</td>
<td>Local contrasts</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[dehaene-cohen-05_neural-code-for-written-words-proposal]
Visual Recognition: IT Cortex

What is $s$?

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>7</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>70</td>
<td>58</td>
<td>39</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>49</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>23</td>
<td>9</td>
<td>83</td>
</tr>
</tbody>
</table>
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
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)
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: \( \lambda u. \text{stud}(u) \)
sleep, V: \( \lambda x. \text{sl}(x) \)
a, DET: \( \lambda P. \lambda Q. \exists x_i. (P(x_i) \land Q(x_i)) \)
Montagovian Translation [1973]

Kohlhase
Cognitive Grammar (Langacker)

• Grounding: the meaning grounds the **syntax**:
Symbol = Form-Meaning pair

- Symbols = (form) label + **meanings**.
- 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 et al 04]
Lexicon

• grounded lexicon:

• semantic pole: perceptual patterns (image schemas) → probabilistic predicate + arguments
Evolving Semantics

Conceptual Space

feature_i

emu
sparrow
crow
vulture

bird
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

Phonological pole

Semantic pole

/boj/ + -/z/ = /bojz/

[BOY] + [PL] = [BOY-PL]
Cognitive Grammar: Inflection

Semantic pole

Phonological pole

/bɔj/  +  /-z/  =  /bɔjz/

[BOY]  +  [PL]  =  [BOY-PL]

symbolic complex [langacker 87]
Cognitive Grammar : Inflection

symbols for “things” (noun)

symbol for replication (relation)

symbolic complex [langacker 87]
Semantic Pole: Image Schemas

must be countable

Countable = has distinct instances. Not a continuum.

another pen; but more water

has discrete instances empty slot
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)

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 et al 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

![Feature Space Trajectories Diagram]
Language Use Patterns
Web Users Map- 2014

- North America: 14%
- Europe: 26%
- Latin America: 10%
- Middle East and Africa: 9%
- Asia Pacific: 41%

http://www.statista.com
Perception and Language affect each other
Structure in Language

words may not be space-separated
पांच फिरंगी अफसरों ___ फांसी पर ___ दिया

what can go in the blanks?
Structure in language : Syllable

पांच फिरंगी अफसरों को फांसी पर लट का दिया

Which syllables follow which others?
Matsuo Basho, (1644-94)

Furuike ya
kawazu tobikomu
mizu no oto

ancient pond
frog jumps in
sound of water
旅夜書懷

思緒而旅行在晚上

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]
Stanza from Lilit Phra Lo (ลิลิตพระ珞) :

เสียงพวกเสียงเล่าอ้าง อันใด พี่เยอะ
เสียงยอดยอดย่อมศใคร ทั่วหล้า
สองเขือพี่หลับใหล 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.