Embodied Models of Language Learning and Use

Session 2: Embodied representations for simulative inference

Srini Narayanan and Nancy Chang
{snarayan,nchang}@icsi.berkeley.edu
UC Berkeley / International Computer Science Institute

Course Overview

- Session 1: Foundations of embodied language
  - Introduction to NTL: language, neural computation
- Session 2: Embodied representations
  - Modeling actions and perception
  - Simulative inference
- Session 3: Language understanding
  - Construction Grammar
  - Metaphor, aspect, perspective
- Session 4: Grammar learning
  - Modeling child language acquisition

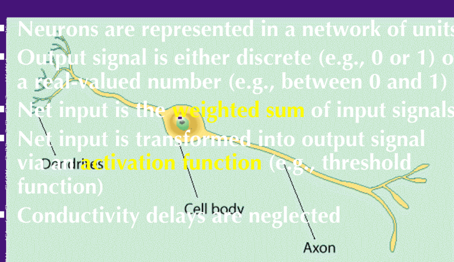
Session 1 recap

1. Introduction to NTL
   - Goal: computationally precise, biologically motivated theories of language structure, use and acquisition
   - Layered methodology
2. Cognition and language
   - Language: challenges of ambiguity, context, creativity
   - parallel activation/integration of (and competition among) multiple kinds of information
3. Neural computation
   - Large-scale functional structure
   - Nature (genetically specified connection patterns) and nurture (activity-dependent tuning/pruning)
   - Hebbian learning: co-activation -> strengthening

Session 1 overflow

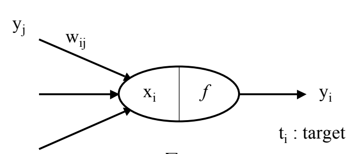
1. Introduction to NTL
2. Cognition and language
3. Neural computation
4. Computational modeling
   - Abstract neuron models
   - Triangle nodes
   - Recruitment leaning

Abstract neuron models



- Neurons are represented in a network of units
- Output signal is either discrete (e.g., 0 or 1) or a real-valued number (e.g., between 0 and 1)
- Net input is the **weighted sum** of input signals
- Net input is transformed into output signal via an **activation function** (e.g., threshold function)
- Conductivity delays are neglected

The McCullough-Pitts Neuron



$$x_i = \sum_j w_{ij} y_j$$

$$y_i = f(x_i - \theta i)$$

Threshold

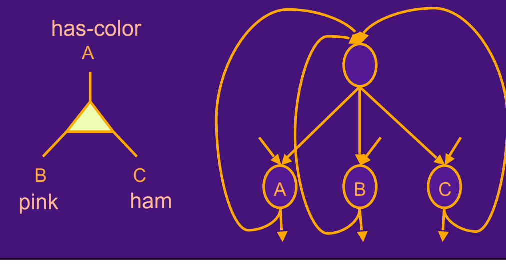
y_j: output from unit j
W_{ij}: weight on connection from j to i
x_i: weighted sum of input to unit i

Networks of neurons

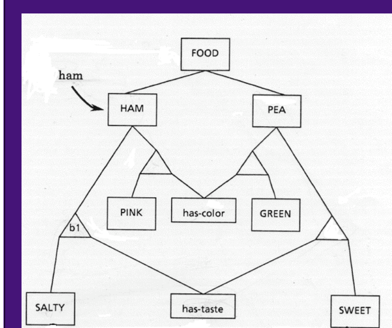
- Parameters of variation
  - Activation function (threshold, linear, sigmoid, Gaussian / radial basis functions)
  - discrete/continuous input/output
  - Network architecture: # nodes, # hidden layers
- General function approximators
  - logical functions (AND, NOT, OR)
  - decision hyperplanes
  - pattern encoding/recognition, self-organizing maps
  - Applications to speech, vision, etc.

Triangle nodes (2/3 node)

- When any 2 of inputs fire, fire all 3
- Can represent features and relations



Triangle nodes for concepts



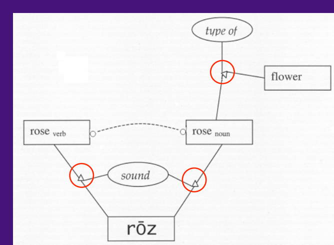
…for associating attributes with values:

2 of 3 input units fire

-->

3rd input unit fires

"They all rose"

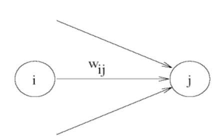


Triangle nodes and inhibition can be used to model priming and spreading activation

Learning in neural networks

- Hebbian ~ coincidence
  - Strengthening co-active connections
- Recruitment ~ one-shot
  - Recruiting "new" connections
- Supervised ~ correction (backprop)
- Reinforcement ~ delayed reward
- Unsupervised ~ similarity

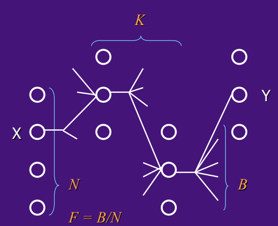
A possible interpretation of Hebb's rule



How often when unit j was firing, was unit i also firing?

$$w_{ij} = \frac{\text{the number of times both } i \text{ and } j \text{ fire}}{\text{the number of times } j \text{ fires}}$$

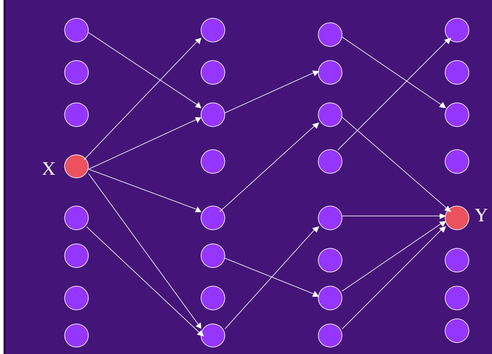
Recruitment Learning

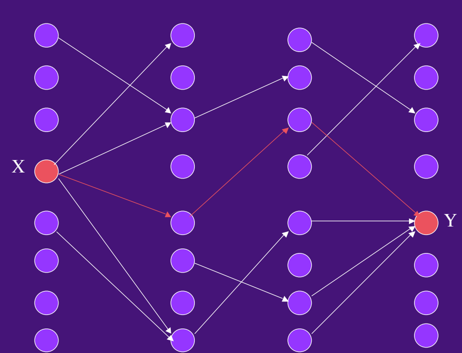


- Suppose we want to link up node X to node Y
- The idea is to pick the two nodes in the middle to link them up
- Can we be sure that we can find a path to get from X to Y?

$$P_{\text{no link}} = (1-F)^{B^K}$$

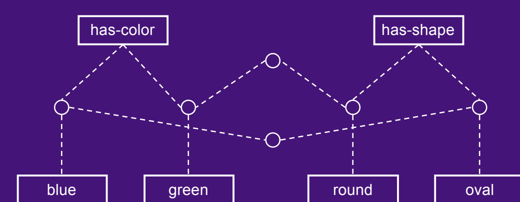
the point is, with a fan-out of 1000, if we allow 2 intermediate layers, we can almost always find a path





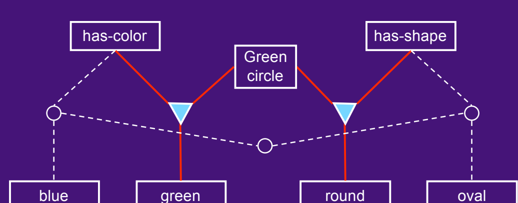
Recruiting triangle nodes

- Let's say we are trying to encode a green circle
- Activate (weak) connections between concepts (dotted lines)



Strengthen these connections

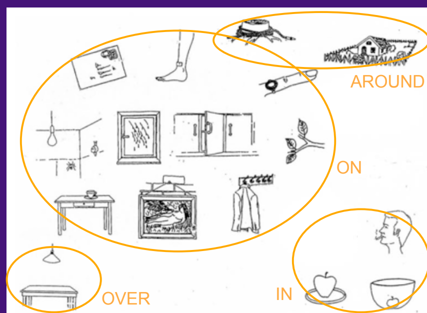
- and you end up with this picture



Session 2 outline

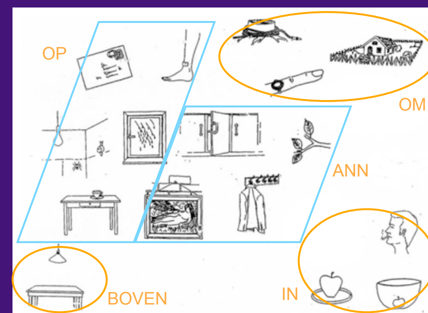
0. Models of neural computation
1. Modeling perception
   - Spatial relations and image schemas
   - Case study: spatial relations (Regier)
2. Modeling action
3. Simulative inference for language

English



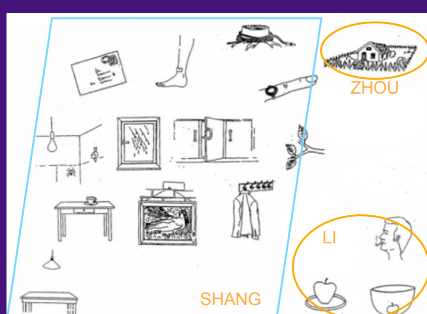
Bowerman & Pederson

Dutch



Bowerman & Pederson

Chinese



Bowerman & Pederson

Image schemas

- **Trajector / Landmark (asymmetric)**
  - The bike is near the house
  - ? The house is near the bike
- **Boundary / Bounded Region**
  - a bounded region has a closed boundary
- **Topological Relations**
  - Separation, Contact, Overlap, Inclusion, Surround
- **Orientation**
  - Vertical (up/down), Horizontal (left/right, front/back)
  - Absolute (E, S, W, N)

TR LM

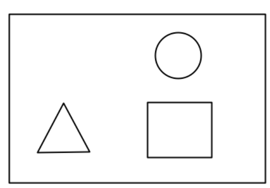
boundary
bounded region

Basis of image schemas

- **Perceptual systems**
- **Sensory-Motor routines**
- **Social Cognition**
- **Image Schema properties depend on**
  - Neural circuits
  - Interactions with the world

…all of which give rise to crosslinguistic variation!

Miniature Language Acquisition

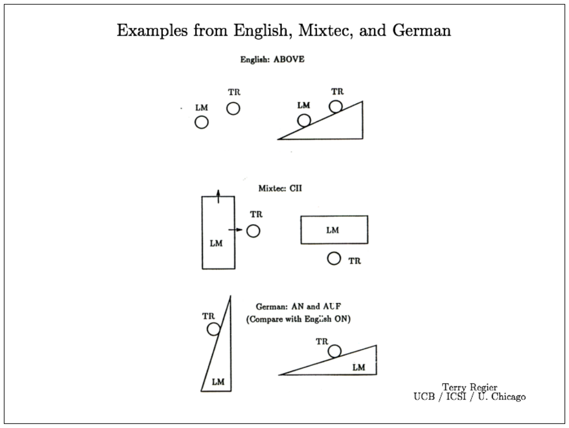


"The circle is above the square."

1. Learn to associate scenes with verbal descriptions.
2. For any new scene-description pair, tell whether the description is true of the scene.
3. Do this for any natural language

The L_0 Project at ICSI

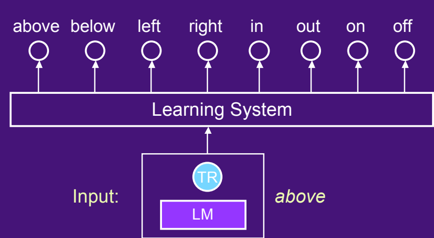
Uses structured connectionist learning methods in addressing the problem



Trajector/Landmark Schema

- **Roles:**
  - Trajector (TR) – object being located
  - Landmark (LM) – reference object
    - TR and LM may share a location (at)

Regier's Model

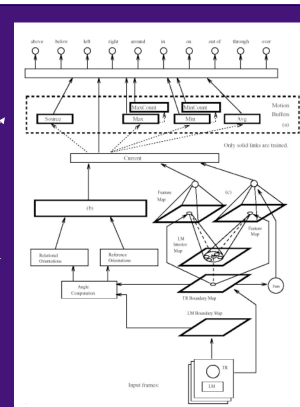


- **Training input: configuration of TR/LM and the correct spatial relation term**
- **Learned behavior: input TR/LM, output spatial relation**

Learning System

dynamic relations (e.g. into)

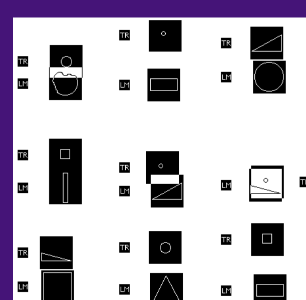
structured connectionist network (based on visual system)



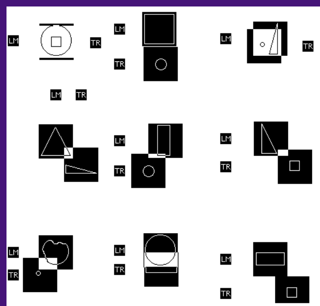
Features of the Visual System in the model

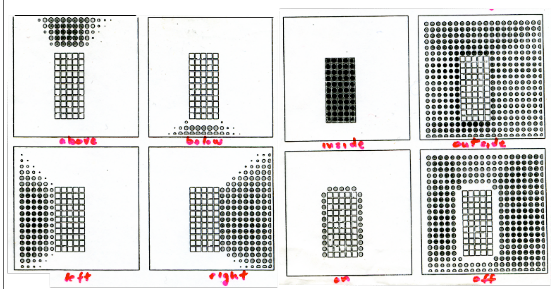
- **Orientation Sensitive cells –**
  - LGN/V1 (Hubel and Weisel)
- **Center-surround receptive fields**
  - LGN, V1 (color opponent processes) upto V4
- **Topographic Maps – All through the visual processing system.**

above – positive examples



above – negative examples





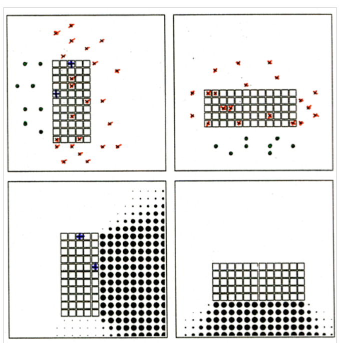


Figure 1: Some Training Data and Results for Mixtec "čii" (see text)

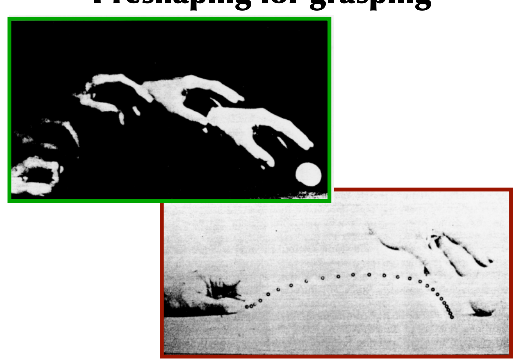
Regier model limitations

- Representational
  - **Recognition (comprehension) only**
  - **Internal representation?**
  - **inference**
- Scaling up
  - **Crosslinguistic concepts**
    - **Force dynamics, size, non-topological**
  - **Grammar**
  - **Abstract concepts**
- Uniqueness / plausibility

Session 2 outline

1. **Modeling perception**
2. **Modeling action**
   - Motor control and mirror neurons
   - Executing schemas and parameterization
   - Case study: action verbs (Bailey)
3. **Simulative inference for language**

Preshaping for grasping



Motor control: Computational requirements

(Stromberg, Latash, Kandel, Arbib, Jeannerod, Rizzolatti)

- **Hierarchical** control
  - Command signals from higher-level motor centers (motor cortex, cerebellum, basal ganglia) to muscle extensors/flexors
- **Coordination** and **concurrency**
  - distributed, parameterized coordination of cortical and sub-cortical circuits
  - Active control: action execution, dynamic interrupts

Mirror neurons

- **Neurons in monkey motor cortex fire during both execution and perception of an action** *(Gallese et al. 1996)*
- **Mirror neurons in humans** *(Porro et al. 1996)*
- **Mirror neurons activated when someone:**
  - imagines performing an action *(Wheeler et al. 2000)*
  - watches an action being performed (with and without object) *(Buccino et al. 2000)*

Monkey see, monkey do?

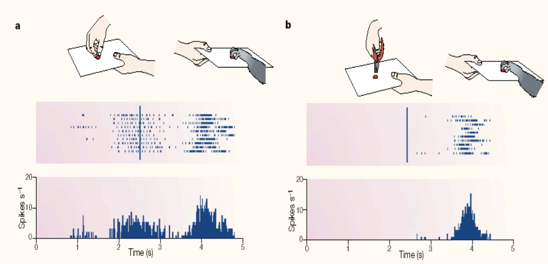
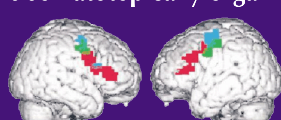


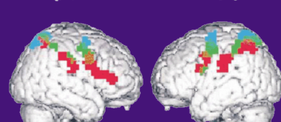
Figure 1 | **Visual and motor responses of a mirror neuron in area F5. a** | A piece of food is placed on a tray and presented to the monkey. The experimenter grasps the food, then moves the tray with the food towards the monkey. Strong activation is present in F5 during observation of the experimenter's grasping movements, and while the same action is performed by the monkey. Note that the neural discharge (lower panel) is absent when the food is presented and moved towards the monkey. **b** | A similar experimental condition, except that the experimenter grasps the food with pliers. Note the absence of a neural response when the observed action is performed with a tool. Rasters and histograms show activity before and after the point at which the experimenter touched the food (vertical bar). Adapted with permission from REF. 11 © 1996 Elsevier Science.

The Mirror System

The mirror system, like the motor system, is somatotopically organized. *Buccino et al., 2001*



humans watching videos of actions without objects



humans watching same actions with objects

Foot actions Hand actions Mouth actions

Motor/parietal circuits: summary

- **PMv (F5ab) – AIP Circuit**
  - "grasp" neurons: movements of hand prehension needed for object grasping
- **F4 (PMC) (behind arcuate) – VIP Circuit**
  - transforming peri-personal space coordinates to facilitate movement toward objects
- **PMv (F5c) – PF Circuit F5c**
  - different mirror circuits for grasping, placing or manipulating object

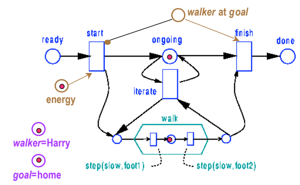
…suggest modality-independent representation of grasp action, active during both action imitation and action recognition

Modeling actions and events

- **Active representation: executing schemas (x-schemas)**
  - Extension to stochastic Petri nets
  - Fine-grained, dynamic, hierarchical control
- **X-schemas are useful for:**
  - Controlling actions
  - Monitoring actions
  - Inference

Active representations

- Executing schema (x-schema)
  - extension to stochastic Petri nets
  - Fine-grained, dynamic, parameterized control
- Useful for monitoring, control and inference



Walking:

bound to a specific *walker* with a direction or *goal*

consumes resources (e.g., energy)

may have termination condition (e.g., *walker* at *goal*)

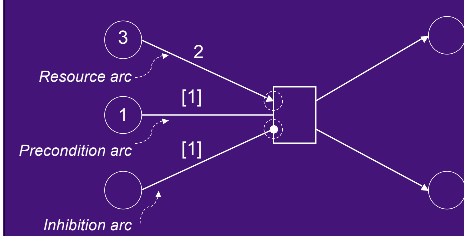
ongoing, iterative action

X-schema: Petri net extensions

- **Parameterization and dynamic binding**
  - Variable parameters
    - walk(speed=slow, destination=store1)
  - Variable objects and entities
    - grasp(cup1), push(cart)
- **Hierarchical control, durative transitions**
  - Subevents
    - walk --> step --> stance, swing phases
  - Time delay for transition firing
    - walk (duration=5 minutes)
- **Stochastic transitions, inhibition**
  - Uncertainty in world evolution and action selection

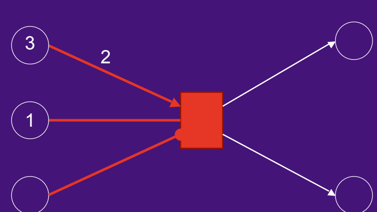
Executing schemas

Basic Mechanism



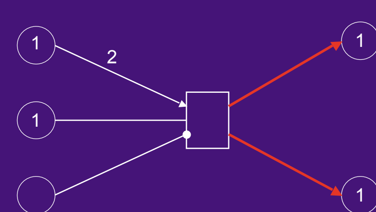
Executing schemas

Firing Semantics



Executing schemas

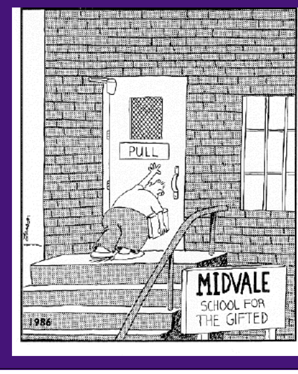
Result of Firing



Mapping: neuron to x-schema

Neuron	X-schema
Cell body	Transition
Axo-dendritic synapse	Resource arc
Axo-dendritic synapse	Inhibitory arc
Enabling condition	Precondition arc
Axon conduction	Output arc
Cell firing	Firing function

Embodied lexical semantics



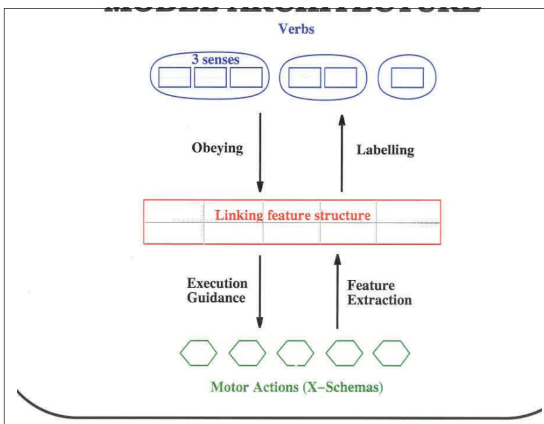
CROSSLINGUISTIC VARIATION IN HAND-ACTION VERBS

- Tamil ***thallu/ilu:*** push/pull, but implies jerkiness or suddenness (smooth motion requires a directional specifier)
- Tamil ***pudi:*** covers clutch, hold, restrain, catch; implies use of high force
- Spanish ***pulsar/presionar:*** press with the index finger *vs.* press with the palm
- Farsi ***hol-daadan/feshaar-daadan:*** two senses of pushing: move an object away from body *vs.* apply pressure to an unmoving object
- Farsi ***zadan:*** hit; strum, or play any musical instrument; object manipulation using quick motion

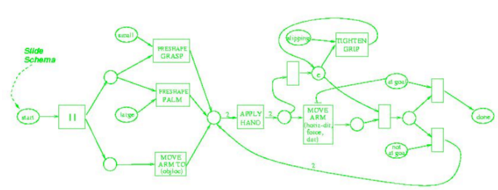
Learning verb meanings

(Bailey 1997)

A model of children learning their first verbs.
Assumes parent labels child's actions.
Child knows parameters of action, associates with word
Program learns well enough to:
1) Label novel actions correctly
2) Obey commands using new words (simulation)
System works across language
Mechanisms are neurally plausible.



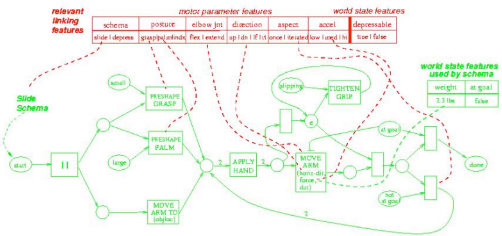
Motor Control (X-schema) for SLIDE



LINKING FEATURES

- Motivation: Extractable and linguistically adequate
- Features include:
  - Schema (which x-schema executes)
  - Hand Posture (grasp, palm, index finger, etc.)
  - Direction (toward, away, up, down, left, right)
  - Elbow Joint Motion (flex, extend, fixed)
  - Force (low, med, high)
  - Aspect (whether x-schema repeats)
  - Object Size (small, med, large)
  - Depressability (a sample object property)
- Extracted from:
  - Synergy parameters
  - Control flow and choice of synergies
  - Perceived world state

Parameters for the SLIDE X-schema



TWO SENSES OF *PUSH*

PUSH: 2 senses

sense 1

schema	posture	direction
slide 100% touch 0%	palm 60% grasp 10% index 30%	away 50% toward 5% up 15% down 30%

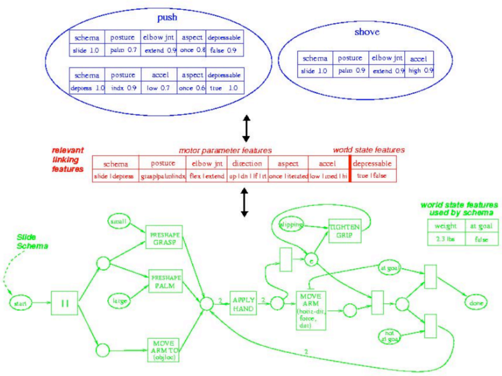
commonness 0.2

sense 2

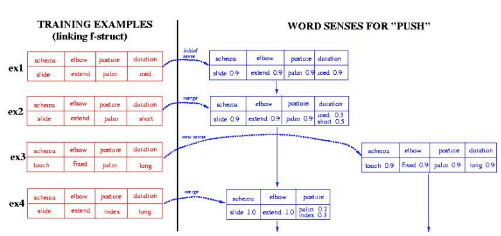
schema	posture	force
slide 0% touch 100%	palm 85% grasp 5% index 10%	low 10% med 30% high 60%

commonness 0.1

System Overview



Learning Two Senses of PUSH



Model merging based on Bayesian MDL

Verb sense learning problem

- Performance measure
  - Goal: Comprehension should improve with training
  - Criterion: need objective function to guide learning
  - "Best" model = most probable, or most compact…
- Bayesian: max posterior probability of model M given data X:
  - M is a set of word senses (or lexicon), $|M| = ws$
  - X is a set of exemplars (linking feature structures)
  - prior $P(M)$ is an exponentially decreasing function of ws

$$P(M \mid X) = \frac{P(X \mid M)P(M)}{P(X)}$$

$$P(M \mid X) = \alpha \cdot P(X \mid M)P(M)$$

$$\log P(M \mid X) = \log P(X \mid M) + \log P(M)$$

- Information-theoretic: minimum description length

$$-\log P(M \mid X) = -\log P(X \mid M) - \log P(M)$$

Results

- **English**
  - 165 training examples (18 hand action labels)
  - Evaluation
    - converges on 21 word senses
    - performance on 32 test examples : 78% recognition, 81% action
    - Mistakes are "close" fit: e.g., lift for yank
  - Learned some directional constructions (pull up)

- **Comparable performance on Farsi, Tamil**
  - identical settings, learned senses not in English
  - Tamil: 9 verb senses, 85 training, 20 test

Verb sense learning: summary

- **Model of acquisition of simple hand action verb senses**
  - Model merging: one-shot learning (fast mapping), sparse data
  - No negative evidence
  - Inductive bias = parameters over motor control schemas
  - Bidirectional: recognition and performance
  - Connectionist reduction to recruitment learning
    - triangle nodes link lexical items with motor and world-state parameters

- **Limitations**
  - No link between action and image schemas (*push through*)
  - No notion of grammar
  - No abstract senses

Session 2 outline

1. **Modeling perception**
2. **Modeling action**
3. **Simulative inference for language**
   - Event structure and aspect
   - *Frames and perspective [skipped in lecture]*

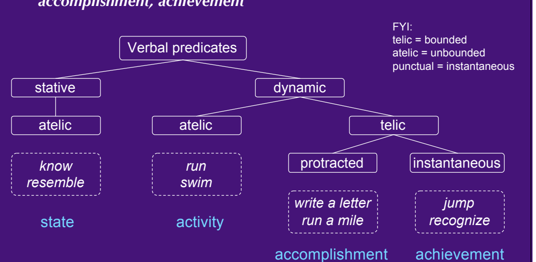
Aspect

Languages have lexical and grammatical devices for conveying information about event structure.

- Progressive: ***She was running home.***
- Perfect: ***I've had a wonderful evening.***
- Inceptive: ***She started knitting.***
- Prospective: ***She's about to leave.***
- Resumptive: ***Peace talks resume.***
- Iterative: ***They ran twice around the track.***
- Habitual: ***She runs every morning.***
- Durative: ***He played the piano for an hour.***

Aspectual classes

- **Zeno Vendler (1957)'s distinction on *state, activity, accomplishment, achievement***



Aspectual distinctions

- **Action patterns**
  - One-shot, repeated, periodic, punctual
  - Decomposition: sequential, concurrent, alternatives
- **Goal-based schema enabling/disabling**
  - Telicity, change of state
- **Generic control features**
  - Interruption, suspension, resumption
- **Resource usage**
  - Production/consumption of time, energy, objects

Richer than in traditional classes!

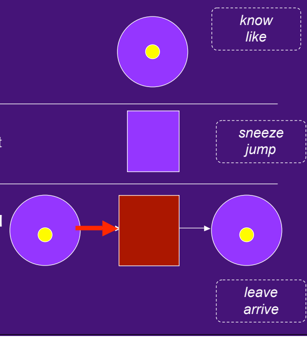
–e.g. durative/atomic, telic/atelic, stative/dynamic (VDT)

X-Schema Distinctions

State
- Obtains (*if marked*)
- Momentaneous simulation/verification

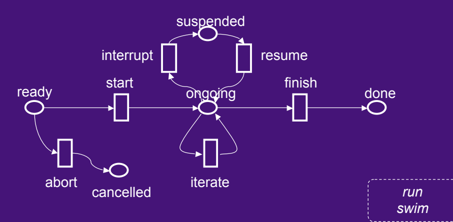
Transition
- Fires to simulate an event

Change of State
- Transition entails pre- and post-states
- Firing removes tokens from pre-state(s) and produces tokens on the post-state(s)

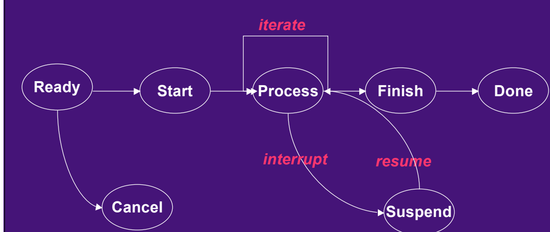


Controller X-Schema

- The **controller x-schema** captures generic event structure
- Aspectual constructions can mark (or **profile**) specific states and/or transitions



A Schema Controller

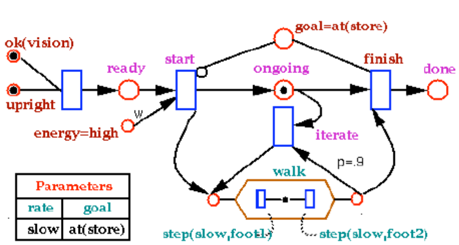


- The controller sends signals to the embedded schema.
- It transitions based on signals from the embedded schema.
- It captures higher level coordination of actions.

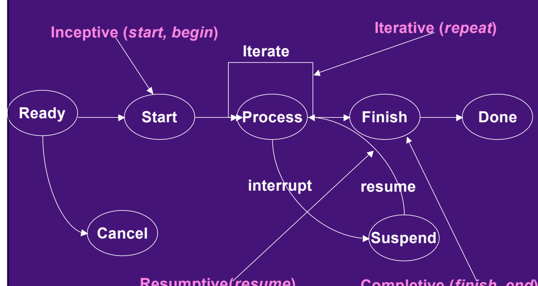
Phases, viewpoints and aspects

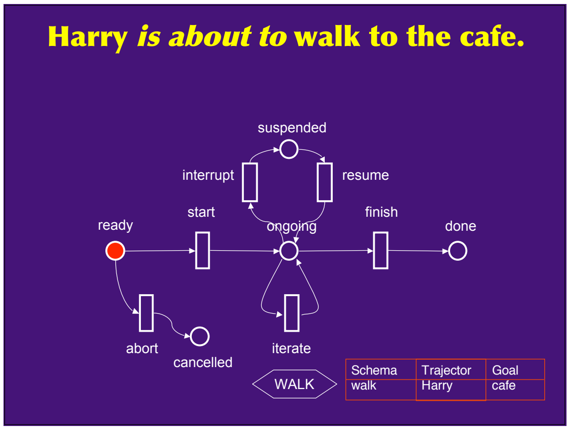
- John **is walking** to the store.
- John **is about to walk** to the store.
- John **walked** to the store.
- John **started walking** to the store.
- John **is starting** to walk to the store.
- John **has walked** to the store.
- John **has started to walk** to the store.
- John **is about to start** walking to the store.
- John **resumed walking** to the store.
- John **has been walking** to the store.
- John **has finished walking** to the store.
- John **almost** walked to the store.

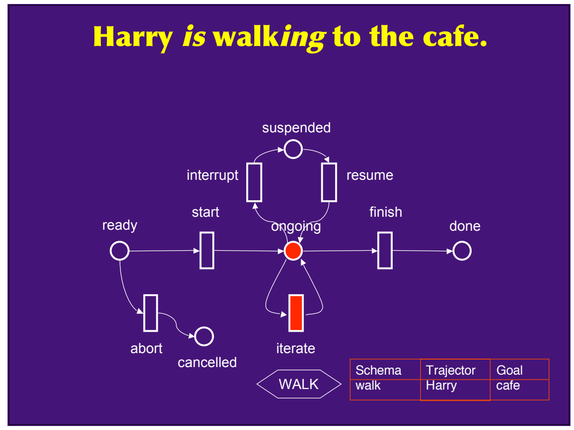
A Walk X-schema

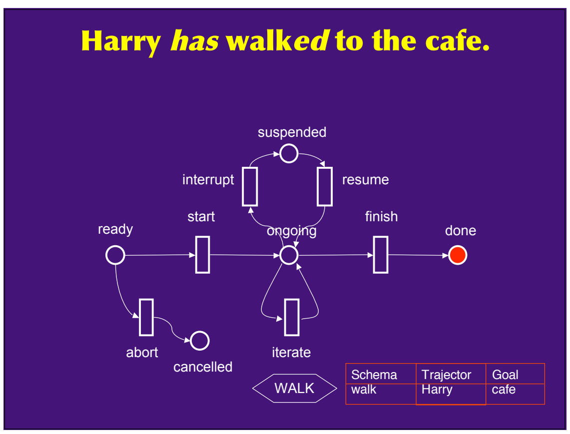


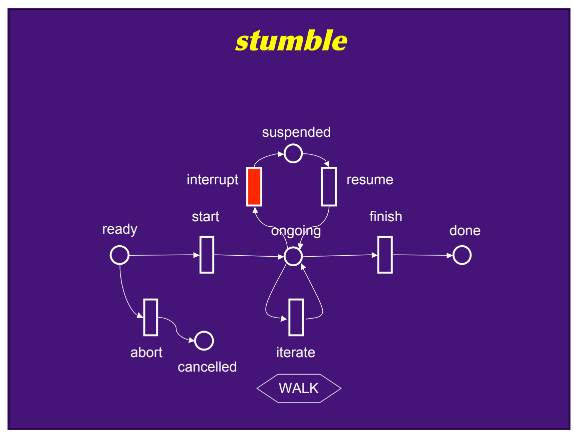
Phasal Aspects Map to Controller

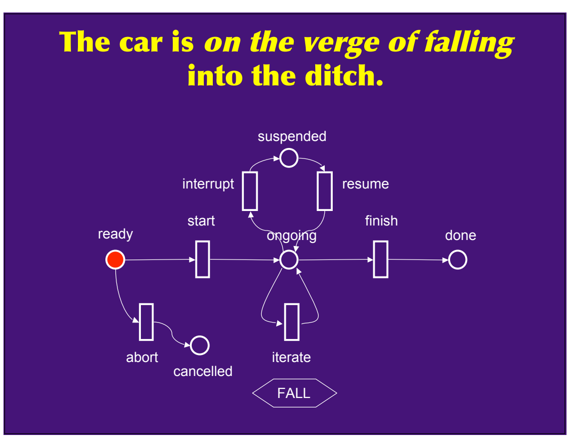


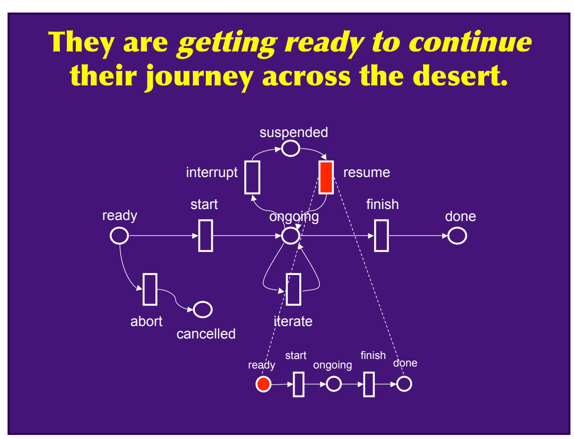

Harry is about to walk to the cafe.
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
WALK
Schema
walk
Trajector
Harry
Goal
cafe


Harry is walking to the cafe.
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
WALK
Schema
walk
Trajector
Harry
Goal
cafe

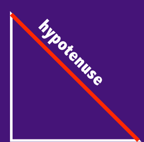

Harry has walked to the cafe.
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
WALK
Schema
walk
Trajector
Harry
Goal
cafe


stumble
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
WALK


The car is on the verge of falling
into the ditch.
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
FALL


They are getting ready to continue
their journey across the desert.
suspended
interrupt
resume
start
finish
ready
ongoing
done
abort
cancelled
iterate
ready
start
ongoing
finish
done

Frame semantics and perspective



Frames

- Frames are conceptual structures that may be culture specific
- Words evoke frames
  - The word *talk* **evokes** the Communication frame
  - The word *buy* (*sell, pay*) **evoke** the Commercial Transaction (CT) frame.
  - The words *journey, set out, schedule, reach* etc. **evoke** the Journey frame.
- Frames have roles and constraints like schemas.
  - CT has roles vendor, goods, money, customer.
- Words bind to frames by specifying binding patterns
  - Buyer binds to Customer, Vendor binds to Seller.

Buyer	Goods	Seller	Payment

She **bought** some carrots from the greengrocer for a dollar.

The greengrocer **sold** some carrots to her for a dollar.
The greengrocer **sold** her some carrots for a dollar.

She **paid** a dollar to the greengrocer for some carrots.
She **paid** the greengrocer a dollar for the carrots.

She **spent** a dollar on the carrots.

The greengrocer **charged** a dollar for a bunch of carrots .
The greengrocer **charged** her a dollar for the carrots.

A bunch of carrots **costs** a dollar.
A bunch of carrots **cost** her a dollar.

Frame-based inference

- event structure / aspectual inference
  - e.g. *buy* vs. *buying*
- perspectival inference
  - e.g. *buy* vs. *sell, buy* vs. *pay*
- resources
  - e.g. *spend, cost, worth*
- planning (goals, preconditions, effects)

How can these inferences be **unpacked**?

Simulation semantics for inference

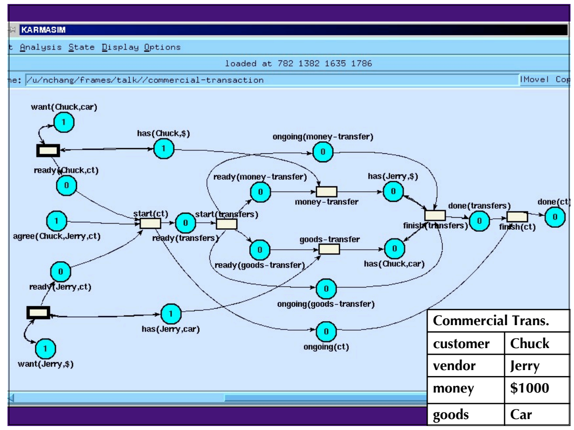
- A semantic specification (or semspec) specifies parameters for a simulation (or enactment) of the temporal and inferential structure of a frame
- Simulation engine uses x-schema (executing-schema) representation based on Petri nets *[Narayanan 1997, 1999, 2002]*

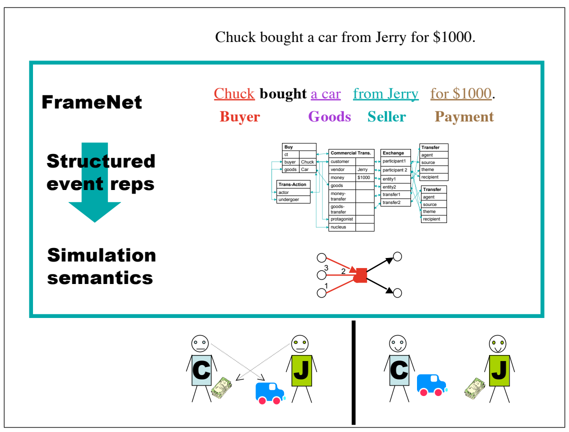
Simulation Semantics

- execution-based model of events/processes
  - tractable, distributed, concurrent, context-sensitive
- X-schemas provide natural model of
  - resource consumption/production
  - goals, preconditions, effects
  - hierarchical events (multiple granularities)

Simulation Semantics (2)

- Captures fine-grained distinctions needed for interpretation
  - aspectual inferences *[Narayanan 1997, 1999; Chang et al. 1998]*
  - metaphoric inferences *[Narayanan 1997, 1999]*
  - perspectival inferences *[Chang et al. 2002]*
  - inductive bias for language learning *[Bailey 1997, Chang 2000]*
- Captures essential features of neural computation *[Feldman & Ballard 1982, Feldman 1989, Valiant 1994]*
  - active, context-sensitive knowledge representation
  - same representational substrate for action, perception *[Boccino et al. 2001, NBL01, CNS02]*
  - natural model of concurrent and distributed computation





Next:
Embodiment and simulation-based language understanding

"What is an idea?
It is an image that paints itself in my brain."
— *Voltaire*