Building Multimodal Simulations
for Natural Language

James Pustejovsky
Nikhil Krishnaswamy
Brandeis University

EACL 2017
Valencia, Spain
April 3, 2017

Tutorial Outline

- Introduction to Multimodal Semantics (14:30-15:30)
  - Generative Lexicon Types and Habitat Theory:
  - Computational Theory of Affordances:
  - Dynamic Logic for Event Structures:
- Overview of VoxSim (15:00-16:00)
  - Module 1: Architecture and Program Flow:
  - Module 2: Object Modeling:
  - Module 3: Action and Gesture Modeling:
  - Module 4: Event Modeling: Integrating Object and Action Models
- Coffee Break (16:00-16:30)
- Creating Simulations: Modeling Novel Content (16:30-18:00)
  - Activity 1: Voxeme Modeling from 3D Geometry Library:
  - Activity 2: Behavior Attachment to a Voxeme
  - Activity 3: Adding Discriminating Attributes to Voxemes:
  - Activity 4: Creating Novel Behavior:

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Starting Assumptions

- **Language visualization and simulation generation**:
  Creating images from linguistic input; generating dynamic
  narratives in simulation environments from action-oriented
  expressions; (Chang et al., 2015; Coyne and Sproat, 2001;
  Siskind, 2001; Pustejovsky and Krishnaswamy, 2016;
  Krishnaswamy and Pustejovsky, 2016)

- **Visual Question-Answering and image content
  interpretation**: QA and querying over image datasets, based
  on the vectors associated with the image, but trained on
  caption-image pairings in the data; (Antol et al., 2015; Chao
  et al., 2015a; Chao et al., 2015b)

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Starting Assumptions

- **Gesture interpretation**: Understanding integrated spoken language with human or avatar-generated gestures; generating gesture in dialogue to supplement linguistic expressions; (Rautaray and Agrawal, 2015; Jacko, 2012; Turk, 2014; Bunt, Beun, and Borghuis, 1998)

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

The Prototype Effect 1/3

Figure: Pencil in a cup

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

The Prototype Effect 2/3

Figure: Pencil in a drawer

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

The Prototype Effect 3/3

Figure: Writing with a pencil

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

The Prototype Effect

- The pencil is in the cup. [+vertical]
- The pencil is in the drawer. [+horizontal]
- The man is using a pencil. [+diagonal]

Object Situation Disambiguation (OSD):
Disambiguate the contextual meaning of an object-denoting word to the appropriate situation.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Mental Models

- Craik (1943)

  Agents carry small-scale models of external reality in their head...

- Johnson-Laird (1983)

  A mental model represents a possibility, capturing what is common to all the different ways in which the possibility may occur (Johnson-Laird and Byrne, 2002). Used to drive inference and reasoning.

- Gentner and Stevens (1983)

  Understanding human knowledge about the world: Domain; Theoretical Approach; Methodology.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Embodiment

- Meaning centrally involves the activation of perceptual, motor, social, and affective knowledge that characterizes the content of utterances

- Understanding a piece of language is hypothesized to entail performing mental perceptual and motor simulations of its content

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Approaches to Modeling Events

- Model-Theoretic Semantics:
  Montague (1968), Davidson (1967), Kamp (1969), Partee
  (1975), Dowty (1979), Verkuyl (1972), Kim (1973), Kratzer
  (1994), Piñon (1997)

- Decompositional Semantics:
  Lakoff (1965), Fillmore (1968), Jackendoff (1972), Talmy
  (1975), Langacker (1987), Fillmore (1985), Jackendoff (1983)

- Lexical-semantic approaches:
  Higginbotham (1986), Tenny (1987), Pustejovsky (1991,
  1995), Krifka (1998), Levin and Hovav-Rappaport (1995)

- Modern Syntheses:
  Naumann (2001), Steedman (2002), Fernando (2013), van
  Lambalgen and Hamm (2005), Pustejovsky (2013)

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Cognitive Simulations of Events

- Frame Semantics
  Fillmore (1966, 1968, 1977), Jackendoff (1972, 1983), Minsky (1974), Löbner (2013)
- Mental Simulations
  Graesser et al (1994), Barselou (1999), Zwaan and Radvansky (1998), Zwaan and Pecher (2012)
- Embodiment:
  Johnson (1987), Lakoff (1987), Varela et al. (1991), Clark (1997), Lakoff and Johnson (1999), Gibbs (2005)
- Simulation Semantics
  Goldman (1989), Feldman et al (2003), Goldman (2006), Feldman (2010), Bergen (2012), Evans (2013),
- Qualitative Mental Models
  Forbus and Gentner (1997), Klenk et al (2005),

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Simulations and Grounded Cognition

- Open-class items tend to activate perceptually based simulations
  - Concrete verbs (within this class) activate motor areas
  - Abstract verbs tend not to activate these areas
- Functional items: no overt simulation. But they provide logical constraints.
- Verbs: abstracted as operational semantic procedures.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Simulations as Minimal Models

- Theorem proving (essentially type satisfaction of a verb in one class as opposed to another) provides a "negative handle" on the problem of determining consistency and informativeness for an utterance (Blackburn and Bos, 2008; Konrad, 2004)
- Model building provides a "positive handle" on whether two manner of motion processes are distinguished in the model.
- The simulation must specify *how* they are distinguished, demonstrating the informativeness of a distinction in our simulation.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Requirements for Multimodal Models of Semantics

- Internal structure of events
- Temporal anchoring and ordering of events
- Event localization: where the event takes place over time
- Rich object semantics: qualia structure, affordances, habitats
- Capturing the dynamics of the event

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Putting Space in Language

- Space as Modality: "add an operator"
  $P_\alpha(meet(john, mary))$
  (Rescher and Garson, 1968, von Wright, 1979, Bennett, 1995, etc.)

- Method of Spatial Arguments: "add an l in a relation"
  $\exists l[meet(john, mary, l) \land in(l, \text{Boston})]$
  (Whitehead, 1929, Randell et al, 1992, Cohn et al, 1997, etc.)

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

"To each their own" (Vendler, 1967)

- **Events** are temporal entities:
  modified by **temporal predicates**
- **Objects** are spatial entities:
  modified by **spatial predicates**
- **Temporal properties** of objects are derivative
- **Spatial properties** of events are derivative

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Locating Events (Davidson, 1967)

- An event is a first-order individual, e:

$$P(x_1, \ldots, x_n, e)$$

- We can identify the location of an event by a relation:

$$loc(e, l)$$

- $\exists e \exists x [smoke(j, e) \wedge in(e, x) \wedge bathroom(x)]$

a. John sang in a field.
$\exists e \exists l [sing(j, e) \wedge in(e, l) \wedge field(l)]$
b. Mary ate her lunch under a bridge.
$\exists e \exists l [eat\_lunch(m, e) \wedge under(e, l) \wedge bridge(l)]$
c. The robbery happened behind a building.
$\exists e \exists l [robbery(e) \wedge behind(e, l) \wedge building(l)]$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Locating Events (Kim, 1973, 1975) 1/2

- An event is a structured object exemplifying a property (or n-adic relation), at a time, t:

$$[(x_1, \ldots, x_n, t), P^n]$$

- We can identify the location of an object in the event:

$$loc(x, t) = r_x$$

- For purposes of event identity, we can construe an event as:

$$[(x_1, \ldots, x_n, r_{x_1}, \ldots, r_{x_n}, t), P^n]$$

$$= [([x_i], [r_{x_i}], t), P^i]$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Locating Events (Kim, 1973, 1975) 2/2

- An event is a structured object exemplifying a property (or
  n-adic relation), at a time, t:

$$[(x_1, \ldots, x_n, t), P^n]$$

- We can identify the location of an object in the event:

$$loc(x, t) = r_x$$

- For purposes of event identity, we can construe an event as:

$$[(x_1, \ldots, x_n, r_{x_1}, \ldots, r_{x_n}, t), P^n]$$

$$= [([x_i], [r_{x_i}], t), P^i]$$

- The event location, l_e, is supervenient on the object locations,
  $r_{x_1}, \ldots, r_{x_n}$.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Linguistic Approaches to Defining Paths

- Talmy (1985): Path as part of the Motion Event Frame
- Jackendoff (1983, 1990,1996): Minimal Path
- Langacker (1987): COS verbs as paths
- Goldberg (1995): way-construction introduces path
- Krifka (1998): Temporal Trace function
- Zwarts (2006): event shape: The trajectory associated with an event in space represented by a path.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Model of Motion Events

- Language encodes motion in Path and Manner constructions
- Path: change with distinguished location
- Manner: motion with no distinguished locations
- Manner and paths may compose.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Subatomic Event Structure (Pustejovsky 1991)

a. EVENT $\rightarrow$ STATE | PROCESS | TRANSITION

b. STATE: $\rightarrow$ e

c. PROCESS: $\rightarrow$ $e_1 \ldots e_n$

d. TRANSITION$_{ach}$: $\rightarrow$ STATE STATE

e. TRANSITION$_{acc}$: $\rightarrow$ PROCESS STATE

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Extensions to GL

- Qualia Structure: Can be interpreted dynamically
- Dynamic Selection: Encodes the way an argument participates in the event
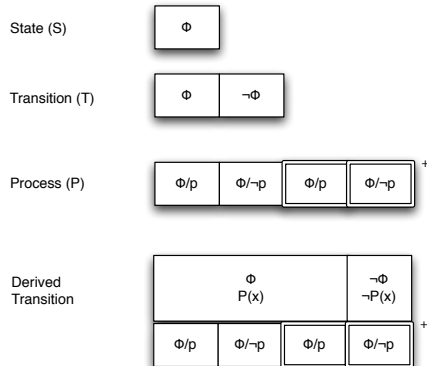- Tracking change: Models the evolution of argument attributes

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

GL Feature Structure

$$
\begin{bmatrix}
\alpha \\
\text{ARGSTR} = \begin{bmatrix} \text{ARG1} = x \\ \dots \end{bmatrix} \\
\text{EVENTSTR} = \begin{bmatrix} \text{EVENT1} = e1 \\ \text{EVENT2} = e2 \end{bmatrix} \\
\text{QUALIA} = \begin{bmatrix} \text{CONST} = \textbf{what } x \textbf{ is made of} \\ \text{FORMAL} = \textbf{what } x \textbf{ is} \\ \text{TELIC} = e_2\textbf{: function of } x \\ \text{AGENTIVE} = e_1\textbf{: how } x \textbf{ came into being} \end{bmatrix}
\end{bmatrix}
$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Inherent Dynamic Aspect of Qualia Structure

- Parameters of a verb, P, extend over sequential frames of interpretation (subevents).

- P is decomposed into different subpredicates within these events:

$$\text{Verb}(\text{Arg}_1\text{Arg}_2) \implies \lambda y \lambda x \boxed{P_1(x, y)}_A \boxed{P_2(y)}_F$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Frame-based Event Structure

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Frame-based Event Structure

Dynamic Interval Temporal Logic
(Pustejovsky and Moszkowicz, 2011)

- Formulas: ϕ propositions. Evaluated in a state, s.
- Programs: α, functions from states to states, $s \times s$. Evaluated over a pair of states, (s, s').
- Temporal Operators: $\bigcirc \phi$, $\Diamond \phi$, $\square \phi$, $\phi \, \mathcal{U} \psi$.
- Program composition:
  1. They can be ordered, $\alpha; \beta$ (α is followed by β);
  2. They can be iterated, a^* (apply a zero or more times);
  3. They can be disjoined, $\alpha \cup \beta$ (apply either α or β);
  4. They can be turned into formulas
     $[\alpha]\phi$ (after every execution of α, ϕ is true);
     $\langle\alpha\rangle\phi$ (there is an execution of α, such that ϕ is true);
  5. Formulas can become programs, ϕ? (test to see if ϕ is true, and proceed if so).

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Events are Simple and Macro programs

- The ECI programs include all identifiable basic movements within a given domain, such as: move; grasp; hold; rotate, roll

- Macro programs are complex activities, such as put, stack

$put(A, B)$
a. Given $\mathcal{C}$ being satisfied (A is clear, within reach, etc), then *grasp* A, and **while** *hold* A, *move* A until at position B.
b. $\mathcal{C}?; grasp(A); (hold(A)?; move(A))^*; on(A, B)?; ungrasp(A); \neg hold(A)?$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Labeled Transition System (LTS)

The dynamics of actions can be modeled as a Labeled Transition Systems (LTS).

An LTS consists of a 3-tuple, $\langle S, Act, \rightarrow \rangle$, where a. S is the set of states;

b. Act is a set of actions;

c. $\rightarrow$ is a total transition relation: $\rightarrow \subseteq S \times Act \times S$. An action, α provides the labeling on an arrow, making it explicit what brings about a state-to-state transition. As a shorthand for $(e_1, \alpha, e_2) \in \rightarrow$, we will also use:

$$e_1 \xrightarrow{\alpha} e_3$$

If reference to the state content (rather than state name) is required for interpretation purposes, then as shorthand for:

$(\{\phi\}_{e_1}, \alpha, \{\neg\phi\}_{e_2}) \in \rightarrow$, we use: $\boxed{\phi}_{e_1} \xrightarrow{\alpha} \boxed{\neg\phi}_{e_2}$
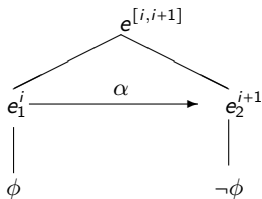
Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

State Transition

Frame-based representation:

$$\boxed{\phi}_{e_1}^{i} \quad \boxed{\neg\phi}_{e_2}^{j}$$

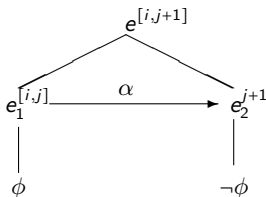$$\boxed{\phi}_{e_1}^{i} \xrightarrow{\alpha} \boxed{\neg\phi}_{e_2}^{i+1}$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Event Structure

$$e^{[i,i+1]}$$

$$e_1^i \xrightarrow{\quad \alpha \quad} e_2^{i+1}$$

$$\phi \qquad\qquad \neg\phi$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Event Structure

Mary awoke from a long sleep.
The state of being asleep has a duration, $[i,j]$, who's valuation is
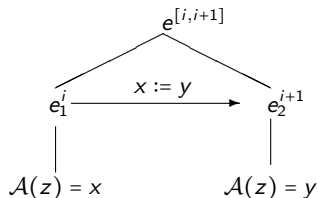gated by the waking event at the "next state", $j+1$.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Event Structure

$x := y$ (ν-transition)

"x assumes the value given to y in the next state."

$\langle \mathcal{M}, (i, i+1), (u, u[x/u(y)]) \rangle \vDash x := y$

iff $\langle \mathcal{M}, i, u \rangle \vDash s_1 \wedge \langle \mathcal{M}, i+1, u[x/u(y)] \rangle \vDash x = y$

$e^{[i, i+1]}$

$e_1^i \xrightarrow{\ x := y\ } e_2^{i+1}$

$\mathcal{A}(z) = x \qquad\qquad \mathcal{A}(z) = y$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Processes

With a ν-transition defined, a *process* can be viewed as simply an iteration of basic variable assignments and re-assignments:

$$e$$

$$e_1 \xrightarrow{\nu} e_2 \ \ldots \ \xrightarrow{\nu} e_n$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances
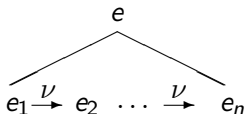
Spatial Relations in Motion Predicates

- Topological Path Expressions
  arrive, leave, exit, land, take off

- Orientation Path Expressions
  climb, descend

- Topo-metric Path Expressions
  approach, near, distance oneself

- Topo-metric orientation Expressions
  just below, just above

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Capturing Motion as Change in Spatial Relations

Dynamic Interval Temporal Logic

- **Path** verbs designate a distinguished value in the change of location, from one state to another.
  The change in value is tested.

- **Manner of motion** verbs iterate a change in location from state to state.
  The value is assigned and reassigned.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Motion Leaving a Trail

MOTION LEAVING A TRAIL:

a. Assign a value, y, to the location of the moving object, x.

$loc(x) := y$

b. Name this value b (this will be the beginning of the movement);

$b := y$

c. Initiate a path p that is a list, starting at b;

$p := (b)$

d. Then, reassign the value of y to z, where $y \neq z$

$y := z, y \neq z$

e. Add the reassigned value of y to path p;

$p := (p, z)$

e. Kleene iterate steps (d) and (e);

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
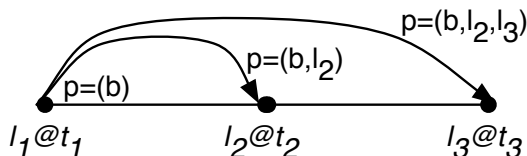Computational Theory of Affordances

Leaving a Trail

Figure: Directed Motion leaving a Trail

a. The ball rolled 20 feet.
   $\exists p \exists x [[roll(x, p) \land ball(x) \land length(p) = [20, foot]]$
b. John biked for 5 miles.
   $\exists p [[bike(j, p) \land length(p) = [5, mile]]$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Directed Motion

$$\boxed{loc(z) = x}_{e_1} \overset{x \neq y?}{\underset{}{\frown}} \xrightarrow{\nu} \boxed{loc(z) = y}_{e_2}$$
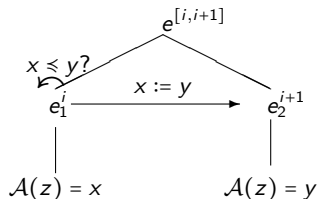
When this test references the ordinal values on a scale, $\mathcal{C}$, this becomes a *directed ν-transition* $(\vec{\nu})$, e.g., $x \lessdot y$, $x \gtrdot y$.

$$\vec{\nu} =_{df} \overset{\mathcal{C}?}{\frown} e_i \xrightarrow{\nu} e_{i+1}$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Directed Motion

$$e^{[i, i+1]}$$

$$x \lessapprox y?$$

$$e_1^i \xrightarrow{\quad x := y \quad} e_2^{i+1}$$

$$\mathcal{A}(z) = x \qquad\qquad \mathcal{A}(z) = y$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Change and Directed Motion

- Manner-of-motion verbs introduce an assignment of a location value:

  $loc(x) := y; y := z$

- Directed motion introduces a dimension that is measured against:

  $d(b, y) < d(b, z)$

- Path verbs introduce a pair of tests:

  $\neg\phi? \ldots \phi?$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Change and the Trail it Leaves

- The execution of a change in the value to an attribute $\mathcal{A}$ for an object x leaves a trail, τ.
- For motion, this trail is the created object of the path p which the mover travels on;
- For creation predicates, this trail is the created object brought about by order-preserving transformations as executed in the directed process above.
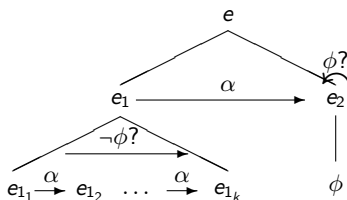
Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Accomplishments Revisited

a. John built a table.
b. Mary walked to the store.

$build(x,z,y)$		$build(x,z,y)^+$		$build(x,z,y), y = v$	
$\neg table(v)$				$table(v)$	

(i,j)

Table: Accomplishment: parallel tracks of changes

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Dynamic Event Structure for Accomplishment

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Parallel Scales define an Accomplishment

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances
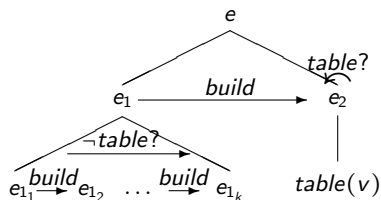
Event Locus and Spatial Aspect 1/4

- Encoding locations is generally not part of the grammatical system of a language (cf. Ritter and Wiltschko, 2005, Deal, 2008)
- Locating an event in the spatial domain is referential (except for deictic spatial morphology).
- We will distinguish between an event locus and its spatial aspect.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Event Locus and Spatial Aspect 2/4

- l_e: Event Locus: similar to Event Time in Reichenbach. it is a referential partition over the Spatial Domain, $\mathcal{D}_{\mathcal{S}}$.
  John walked.

- l_r: Spatial Aspect: a binary partitioning relative to this first partition. Similar to Reference Time.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Event Locus and Spatial Aspect 3/4

Sources of Spatial Aspect in Motion Verbs:

- ANALYTIC ASPECT: verb selects a spatial argument;
  Mary left *the room*.
  John entered *the hall*.

- SYNTHETIC ASPECT: verb is modified through PP adjunction;
  Mary swam *in the pool*.
  John walked *to the corner*.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Event Locus and Spatial Aspect 4/4

- Simple Locus: $l_e = l_r$.
  John **walked**$_{l_e, l_r}$.

- Relative Aspect: $l_e <_d l_r$.
  John **walked**$_{l_e}$ under the tree$_{l_r}$.

- Embedded Aspect: $l_e \subseteq l_r$.
  John **walked**$_{l_e}$ in the building$_{l_r}$.

- Completive Aspect: $\textbf{EC}(l_e, l_r)$, $\textbf{end}(l_r, \hat{p})$.
  John **arrived**$_{l_e}$ home$_{l_r}$.
  John **walked**$_{l_e}$ to the park$_{l_r}$

- Ingressive Aspect: $\textbf{EC}(l_r, l_e)$, $\textbf{begin}(l_r, \hat{p})$.
  John **walked**$_{l_e}$ from the park$_{l_r}$.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Event Localization

- r_{x_i}: The Kimian spatial extent of an object, x_i;
- $\hat{p}$: The path created by the motion in e;
- R_e: an embedding space (ES) for e, defined as a region containing $\hat{p}$ and r_{x_i} in a specific configuration, $\hat{p} \otimes r_{x_i}$;
- μ, the event locus: the minimum embedding space for e.
- Where μ can be defined as:
  $\forall e \forall R_e \forall \mu [[ES(R_e, e) \wedge Min(\mu, R_e)] \leftrightarrow [\mu \subseteq R_e \wedge \forall y[y \subseteq R_e \to \mu \subseteq y]]]$.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

NAME

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Habitats and Simulations Pustejovsky and Krishnaswamy (2014)

- Habitat: a representation of an object situated within a partial minimal model; Enhancements of the qualia structure.

- With multi-dimensional affordances that determine how habitats are deployed and how they modify or augment the context.

- Compositional combinations of procedural (simulation) and operational (selection, specification, refinement) knowledge.

- A habitat:
  - embeds;
  - orients;
  - positions.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Requirements on a Multimodal Semantic Simulation

1. A minimal embedding space (MES) for the simulation must be determined. This is the 3D region within which the state is configured or the event unfolds;

2. Object-based attributes for participants in a situation or event need to be specified; e.g., orientation, relative size, default position or pose, etc.;

3. An epistemic condition on the object and event rendering, imposing an implicit point of view (POV);

4. Agent-dependent embodiment; this determines the relative scaling of an agent and its event participants and their surroundings, as it engages in the environment.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Related Research on Scene and Simulation Construction

- Spatial and temporal interval logics
  - Allen Temporal Relations (Allen, 1983)
  - Region Connection Calculus (RCC8) (Randell et al., 1992)
  - RCC-3D (Albath, et al., 2010)
- Generative Lexicon, DITL (Pustejovsky, 1995; Pustejovsky and Moszkowicz, 2011)
- Static scene generation
  - WordsEye (Coyne and Sproat, 2001)
  - LEONARD (Siskind, 2001)
  - Stanford NLP Group (Chang et al., 2015)
- QSR/Game AI approaches to scenario-based simulation (Forbus et al., 2001; Dill, 2011)
- Spatial constraint mapping to animation (Bindiganavale and Badler, 1998)

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Visual Object Concept Modeling Language (VoxML)

- A modeling language for constructing 3D visualizations of concepts denoted by natural language expressions
- used as the platform for creating *multimodal semantic simulations*

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Teleotopology

- The function of space: the actions associated with a region or an object (inherently or opportunistically), i.e., Telic role values.

- The space of function: the regions defined by the Telic actions performed by an agent, or supervenient on the Telic state of an artifact, teleotopology.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Extending Qualia to Modeling Affordances

The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. It implies the complementarity of the animal and the environment. (J. J. Gibson, 1979/1986)

- Gibson (1979), Turvey (1992), Steedman (2002), Sahin et al (2007), Krippendorff (2010);
- Affordance: a correlation between an agent who acts on an object with a systematic or prototypical effect.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Semantics of Function and Purpose

There are two levels of accessibility that can be identified in a Telic role value, as illustrated below.

a. local modality: the conditions under which the activity can be performed on the object;

b. global modality: what is done with the object, and the resulting state.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Telic Qualia Role as part of Affordance Structure

Motivation for Qualia relations comes from the idea that there is a *hidden event* in the lexical representation associated with nouns denoting objects made for a particular purpose:

a. a door is for walking through

b. a window is for seeing through

c. a book is for reading

d. a beer is for drinking

e. a cake is for eating

f. a car is for driving

g. a table is for putting things on

h. a desk is for working on

i. a pen is for writing with

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Telic Values and Affordances

$\mathcal{C} \rightarrow [\pi]\mathcal{R}$

The TELIC of *sandwich*:

$$\lambda x \begin{bmatrix} \textbf{sandwich} \\ \text{AS} = \begin{bmatrix} \text{ARG1} = x : e \end{bmatrix} \\ \text{QS} = \begin{bmatrix} \text{F} = phys(x) \\ \text{T} = \lambda y \lambda e[\mathcal{C} \rightarrow [eat(e, y, x)]\mathcal{R}_{eat}(x)]] \\ \text{A} = \exists z[make(z, x)] \end{bmatrix} \end{bmatrix}$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Visual Object Concept (Voxeme)

- Object Geometry Structure:
  Formal object characteristics in R3 space
- Habitat: Embodied and embedded object:
  Orientation
  Situated context
  Scaling
- Affordance Structure:
  What can one do to it
  What can one do with it
  What does it enable
- Voxicon: library of voxemes

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Projected Voxicon: 4,000 voxemes

- 3,000 object voxemes (nouns)
- 500 program voxemes (verbs) with links to VerbNet
- 300 attribute voxemes (adjectives)
- 200 primitive relations (prepositions and stative verbs)
- Functional expressions (quantifiers, partitives, collections)
- Geometries: 512 objects complete
- Complete voxemes: 200 with affordances
- Attached behaviors: 40 distinct actions
- Animated human models: 3

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxML Elements

Entities modeled in VoxML can be:

- Objects: Physical objects (Nouns)
- Programs: Events (Verbs)
- Attributes: Properties (Adjectives)
- Functions: Quantifiers, connectives

These entities can then compose into visualizations of natural language concepts and expressions.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxML OBJECT is used for modeling nouns: 1/5

LEX	OBJECT's lexical information
TYPE	OBJECT's geometrical typing
HABITAT	OBJECT's habitat for actions
AFFORD_STR	OBJECT's affordance structure
EMBODIMENT	OBJECT's agent-relative embodiment

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Objects 2/5

- The TYPE attribute contains information to define the object geometry in terms of primitives. HEAD is a primitive 3D shape that roughly describes the object's form or the form of the object's most semantically salient subpart.

HEAD	prismatoid, pyramid, wedge, parallelepiped, cupola, frustum, cylindroid, ellipsoid, hemiellipsoid, bipyramid, rectangular_prism, toroid, sheet

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Objects 3/5

- COMPONENTS: subparts of the object
- CONCAVITY: concave, flat, or convex; refers to any concavity that deforms the HEAD shape.
- ROTATSYM (rotational symmetry) defines any of the three orthogonal axes around which the object's geometry may be rotated for an interval of less than 360 degrees and retain identical form as the unrotated geometry.
- REFLECTSYM (Reflectional symmetry): If an object may be bisected by a plane defined by two of the three orthogonal axes and then reflected across that plane to obtain the same geometric form as the original object, it is considered to have reflectional symmetry across that plane.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Objects 4/5

HABITAT defines habitats INTRINSIC to the object, regardless of
what action it participates in, such as intrinsic orientations or
surfaces, as well as EXTRINSIC habitats which must be satisfied
for particular actions to take place.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Objects 5/5

AFFORD_STR describes the set of specific actions, along with the requisite conditions, that the object may take part in. There are low-level affordances, called GIBSONIAN, which involve manipulation or maneuver-based actions (grasping, holding, lifting, touching); there are also TELIC affordances, which link directly to what goal-directed activity can be accomplished, by means of the GIBSONIAN affordances.

EMBODIMENT qualitatively describes the SCALE of the object compared to an in-world agent (typically assumed to be a human) as well as whether the object is typically MOVABLE by that agent.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Programs are used for modeling verbs

LEX	PROGRAM's lexical information
TYPE	PROGRAM's event typing
EMBEDDING_SPACE	PROGRAM's embodiment as a function of the participants and their changes over time

A PROGRAM's LEX attribute contains the subcomponents PRED, the lexeme predicate denoting the program, and TYPE, the program's type as given in a lexical semantic resource, e.g., its GL type.

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Plate

$$
\begin{bmatrix}
\textbf{plate} \\
\text{LEX} = \begin{bmatrix} \text{PRED} = \textbf{plate} \\ \text{TYPE} = \textbf{physobj} \end{bmatrix} \\[2mm]
\text{TYPE} = \begin{bmatrix} \text{HEAD} = \textbf{sheet} \\ \text{COMPONENTS} = \textbf{surface, base} \\ \text{CONCAVITY} = \textbf{concave} \\ \text{ROTATSYM} = \{Y\} \\ \text{REFLECTSYM} = \{XY, YZ\} \end{bmatrix} \\[2mm]
\text{HABITAT} = \begin{bmatrix} \text{INTR} = {}_{[1]}\begin{bmatrix} \text{UP} = align(Y, \mathcal{E}_Y) \\ \text{TOP} = top(+Y) \end{bmatrix} \\ \text{EXTR} = ... \end{bmatrix} \\[2mm]
\text{AFFORD\_STR} = \begin{bmatrix} \text{A}_1 = H[1] \rightarrow [put(x, y)]hold(y, x) \\ \text{A}_2 = \textbf{...} \\ \text{A}_3 = \textbf{...} \end{bmatrix} \\[2mm]
\text{EMBODIMENT} = \begin{bmatrix} \text{SCALE} = < \textbf{agent} \\ \text{MOVABLE} = \textbf{true} \end{bmatrix}
\end{bmatrix}
$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances
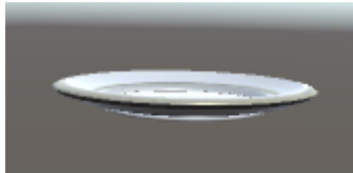
Plate

Figure: Plate voxeme instance

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxML for cup

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxML for in

$$
\begin{bmatrix}
\textbf{in} \\
\text{LEX} = \begin{bmatrix} \text{PRED} = \textbf{in} \end{bmatrix} \\
\text{TYPE} = \begin{bmatrix}
\text{CLASS} = \textbf{config} \\
\text{VALUE} = \textbf{ProperPart} \parallel \textbf{PO} \\
\text{ARGS} = \begin{bmatrix} \text{A}_1 = \textbf{x:3D} \\ \text{A}_2 = \textbf{y:3D} \end{bmatrix} \\
\text{CONSTR} = \textbf{...}
\end{bmatrix}
\end{bmatrix}
$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxML for put

$$
\begin{bmatrix}
\textbf{put} \\
\text{LEX} = \begin{bmatrix} \text{PRED} = \textbf{put} \\ \text{TYPE} = \textbf{transition\_event} \end{bmatrix} \\
\text{TYPE} = \begin{bmatrix}
\text{HEAD} = \textbf{transition} \\
\text{ARGS} = \begin{bmatrix} \text{A}_1 = \textbf{x:agent} \\ \text{A}_2 = \textbf{y:physobj} \\ \text{A}_3 = \textbf{z:location} \end{bmatrix} \\
\text{BODY} = \begin{bmatrix}
\text{E}_1 = grasp(x, y) \\
\text{E}_2 = [while(\\
\quad hold(x, y), \\
\quad move(y)] \\
\text{E}_3 = [at(y, z) \rightarrow \\
\quad ungrasp(x, y)]
\end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

VoxSim Input

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances
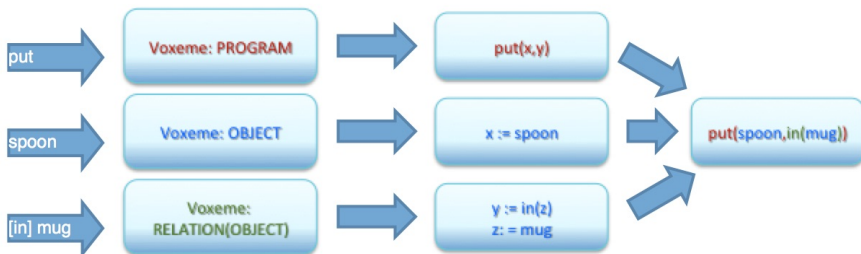
Type-driven Behavior and Constraints

Each predicate is operationalized according to its type structure

$$put(spoon, in(mug))$$

- $in(z)$: takes object, outputs location
- $put(x, y)$: path verb
- $while(\neg at(y), move(x))$

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Type-driven Behavior and Constraints

- Can test be satisfied with current object configuration?
- Can test be satisfied by reorienting objects?
- Can test be satisfied at all?

Introduction to Multimodal Semantics
Overview of VoxSim
Creating Simulations: Modeling Novel Content
References

Generative Lexicon Types and Habitat Theory
Dynamic Logic for Event Structures
Computational Theory of Affordances

Caused Motion as Rig Attachment

- Temporary parent-child relationship between joint on rig and manipulated object
- Allows agent and object to move together
- Object model + *Action model* = Event model