Do You See What I See?
Effects of POV on Spatial Relation Specifications

Nikhil Krishnaswamy and James Pustejovsky
Brandeis University

30th International Workshop on Qualitative Reasoning
Melbourne, Australia
August 21, 2017
Language users’ mental models contain a remarkable inventory of “concepts”
- Language does not directly map to thought expressed (De Saussure, 1915)
- Frame of reference and indexicality create ambiguity which is resolved through context (Kaplan, 1979)

A linguistic predicate encodes a certain level of information that can be used for reasoning

Amount and nature of that information varies between predicates

For a sentence, a set of parameters (speed, rotation, etc.) exist that make that a sentence true and a set that make it false (i.e., a different action)
Independent of their content, predicates and propositions can be expressed within a *minimal model*.

Minimal model: Universe containing set of arguments, set of predicates, interpretations of arguments, subsets defining interpretations of predicates (Gelfond and Lifschitz, 1988)

- Predicates assumed to be logic programs
- Arguments assumed to evaluate to constants

Simulation: *Minimal model* with values assigned to set of necessary and sufficient variables left underspecified in model

- Values must be defined sufficiently to show the operation of the associated model over time
- Values must be defined in a simulation or fully-specified logic program defining a predicate cannot be run
Visualization: Process linking each semantic object in the simulation to a visual object enacted in a virtual environment frame-by-frame

- Variables assigned in *simulation* are evaluated and reassigned each frame according to the program(s) currently scoping them
- Final step is rendering the complete visualization at each frame
- In a visual modality, spatial information encoded in a predicate can be revealed by simulation
- Human can see whether visualization depicts a sentence *s* or not
  - Set of values \([a]\) for parameter in *s* results in either \(\mathcal{M} \models p_s[a]\) or \(\mathcal{M} \not\models p_s[a]\).
Introduction

- Simulation allows easy storage and recovery of parameter values
  - Provides computational model of reasoning from linguistic information
- One modality of expressing a simulation is visual
  - Technology is readily available
  - Allows the creation of a shared context between multiple agents (human/human, or human/computer)
  - To gather data on information that such a simulation system provides...
    - We have to build a simulator!
Related Research

- “Simulation”: mental instantiation of an utterance, based on embodiment (Ziemke, 2003; Feldman and Narayanan, 2004; Gibbs Jr., 2005; Lakoff, 2009; Bergen, 2012; Kiela et al., 2016)
  - Argued to be ineffective in interpreting continuous or underspecified parameters (Davis and Marcus, 2016)
- Generative Lexicon, dynamic semantics (Pustejovsky, 1995; Pustejovsky and Moszkowicz, 2011; Mani and Pustejovsky, 2012)
- Orientation in QSR (Freksa, 1992; Moratz, Renz, and Wolter, 2000; Dylla and Moratz, 2004; Renz and Nebel, 2007)
- Algebraic formalisms for frames of reference (Frank, 1992; Kuipers, 2000)
Related Research

- QR as information-bearer (Joskowicz and Sacks, 1991; Kuipers, 1994)
- Cardinal directions and path knowledge (Frank, 1996; Zimmermann and Freksa, 1996)
- Object manipulation and environment navigation (Thrun et al., 2000; Rusu et al., 2008)
- QSR to improve machine learning (Falomir and Kluth, 2017)
- QSR/Game AI approaches to scenario-based simulation (Forbus, Mahoney, and Dill, 2002; Dill, 2011)
Related Research

- Spatial/temporal algebraic interval logic
  - Allen Temporal Relations (Allen, 1984)
  - Region Connection Calculus (Randell et al., 1992)
    - RCC-3D (Albath et al., 2010)

- Static scene generation
  - WordsEye (Coyne and Sproat, 2001)
  - LEONARD (Siskind, 2001)
  - Stanford NLP Group (Chang et al., 2015)

- Our approach differs by focusing on motion verbs
  (Pustejovsky, 2013; McDonald and Pustejovsky, 2014; Pustejovsky and Krishnaswamy, 2014; Pustejovsky and Krishnaswamy, 2016; Krishnaswamy and Pustejovsky, 2016a; Krishnaswamy and Pustejovsky, 2016b)
VoxML

- VoxML: Visual Object Concept Modeling Language (Pustejovsky and Krishnaswamy, 2016)
- Modeling and annotation language for “voxemes”
  - Visual instantiation of a lexeme
  - Lexemes may have many visual representation
- Scaffold for mapping from lexical information to simulated objects and operationalized behaviors
- Encodes afforded behaviors for each object
  - Gibsonian: afforded by object structure (Gibson, 1977; Gibson, 1979)
    - grasp, move, lift, etc.
  - Telic: goal-directed, purpose-driven (Pustejovsky, 1995)
    - drink from, read, etc.
VoxML

**Figure:** VoxML for a “cup”
**Figure**: VoxML for “put” and “in”
Object bounds may not contour to geometry
  - e.g., concave objects

Semantic information imposes further constraints

“in cup”: (PO | TPP | NTPP) with area denoted by cup’s interior
  - Interpenetrates bounds, but not geometry
Architectures

- Built on Unity Game Engine
- NLP may use 3rd-party tools
- Art and VoxML resources loaded locally or from web server
- Input to UI or over network

Figure: VoxSim architecture schematic
Architecture

Figure: Dependency parse for *Put the apple on the plate* and transformation to predicate-logic form.
1. Input sentence
2. Generate parse
3. Compute satisfaction conditions from voxeme composition
4. Move object to target position
5. Update relationships between objects
6. Make or break parent-child rig-attachments
7. Resolve discrepancies between Unity physics bodies and voxemes
Before executing an action, the system must determine:

1. Can test be satisfied with current object configuration?
2. Can test be satisfied by reorienting objects?
3. Can test be satisfied at all?

Figure: Object properties impose constraints on motion
"LEAN" — Theoretical formulation:
- Instruction: “Lean [[THEME]] on [[DEST]]”
- Goal: [[THEME]] is supported by [[DEST]] at an angle $\theta$
  - For this example, assume $\theta = 45^\circ$

1. Turn [[THEME]] such that major axis is $\theta$ off from $+Y$ axis
2. Move [[THEME]] so it touches a side of [[DEST]]
Modeling Events

“LEAN” — Operationalization:

- Instruction: “Lean [[THEME]] on [[DEST]]”
- Goal: [[THEME]] is supported by [[DEST]] at an angle $\theta$
  - For this example, assume $\theta = 45^\circ$
- Starting position of [[THEME]] is arbitrary
  - Not necessarily lying flat
  - Not necessarily axis-aligned
- 3D transformations take shortest path
  - Single rotation may result in unstable configuration

1. Turn [[THEME]] such that **minor axis** is $90^\circ - \theta$ off from $+Y$ axis
2. Turn [[THEME]] **about minor axis** such that major axis is $\theta$ off from $+Y$ axis
3. Move [[THEME]] so it touches a side of [[DEST]]
Modeling Events

- Three types of primitive motions
  1. TURN-1: turn(x:obj, V1:axis, EV2:axis) — turn object x so that object axis V1 is aligned with world axis V2
  2. TURN-2: turn(x:obj, V1:axis, EV2:axis, EV3:axis) — turn object x so that object axis V1 is aligned with world axis V2, constraining motion to around world axis V3
  3. PUT: put(x:obj, y:loc) — put object x at location y

\[
\begin{align*}
\text{lean} & \quad \text{LEX} = \begin{bmatrix}
\text{PRED} & \text{lean} \\
\text{TYPE} & \text{transition\_event}
\end{bmatrix} \\
\text{HEAD} & = \text{transition} \\
\text{ARGS} & = \begin{bmatrix}
A_1 & \text{x:agent} \\
A_2 & \text{y:physobj} \\
A_3 & \text{z:location}
\end{bmatrix} \\
\text{TYPE} & \quad \text{body} \\
\text{E}_1 & = \text{grasp}(x, y) \\
\text{E}_2 & = [\text{while}(\text{hold}(x, y), \text{turn}(x, y, \\
& \quad \text{align}((\text{minor}(y), \\
& \quad \text{EV} \times (90 - \theta, \text{about(EV)})))))] \\
\text{E}_3 & = [\text{while}(\text{hold}(x, y), \text{turn}(x, y, \\
& \quad \text{align}((\text{major}(y), \\
& \quad \text{EV} \times (\theta, \text{about(EV)})), \\
& \quad \text{about}((\text{minor}(y))))))] \\
\text{E}_4 & = [\text{while}(\text{hold}(x, y), \text{put}(x, y))] \\
\text{E}_5 & = [\text{at}(y, z) \rightarrow \text{ungrasp}(x, y)]
\end{align*}
\]
Demo
Underspecification

- Minimal model requires minimal parameter specification
  - “Slide the plate”
    - How fast? How far? Which direction?
  - “Put the spoon near the cup”
    - How close is “near”? 
  - “Put the block touching the plate”
    - Touching where?
- Model exists in state of non-minimal entropy
  - There exist “bits” to be set
  - Certain values result in cognitively coherent simulation
Experimental Design

- VoxSim provides method of visually testing theoretical semantic assumptions
- Unassigned parameters given values through Monte Carlo randomization
  - Unity generates random values using uniform distribution, a la standard Monte Carlo methods (Sawilowsky, 2003)
  - Values may be resampled if constraint on predicate specification is violated
- Video captured for visualizations of test sentences
  - 3 videos per input sentence
- Evaluation done through Amazon Mechanical Turk
  - Workers asked to select which of three videos best depicts the input sentence that was used to generate all three
  - Multiple answers acceptable; “None” available
  - 8 individual workers per HIT
Figure: Test environment with all objects shown. During capture of an event, all objects not mentioned in the input sentence were removed.
Evaluation

- Raw results reflect overall incidence of evaluators accepting visualization for provided utterance
- Greater probability of acceptance $\rightarrow$ parameter values better reflect utterance
  - $P(\text{acc} \mid V) \sim$ prototypicality of visualization relative to event semantics
  - Exact object coordinates and relative offsets are used to render visuals
    - Less relevant to acceptability judgment than qualitative assessment of object relations
- Discrete value set: evaluation conditioned on choice from set
- Continuous value set: evaluation conditioned on probability density over distance between objects, partitioned into subsets ($q = 5$)
### Evaluation

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Underspecified parameters</th>
<th>Possible values</th>
</tr>
</thead>
<tbody>
<tr>
<td>touching(x)</td>
<td>rel orientation</td>
<td>{left(x), right(x), behind(x), in_front(x), on(x)}</td>
</tr>
<tr>
<td>near(x)</td>
<td>transloc dir</td>
<td>(V \in {(y-x(x), y-y(x), y-z(x)) \mid d(x,y) &lt; d(edge(s(y),y)), IN(s(y)), \neg IN(y)})</td>
</tr>
</tbody>
</table>

**Table:** Predicate value assignments

- “Touching” and “Near”
  - “Touching”: discrete set
  - “Near”: continuous range
## Results

### “Touching”

<table>
<thead>
<tr>
<th>QSR (event start)</th>
<th>P(accept QSR)</th>
<th>QSR (event end)</th>
<th>P(accept QSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>behind(y)</td>
<td>0.5497</td>
<td>behind(y)</td>
<td>0.5474</td>
</tr>
<tr>
<td>in_front(y)</td>
<td>0.5692</td>
<td>in_front(y)</td>
<td>0.5816</td>
</tr>
<tr>
<td>left(y)</td>
<td>0.5753</td>
<td>left(y)</td>
<td>0.4995</td>
</tr>
<tr>
<td>right(y)</td>
<td>0.5725</td>
<td>right(y)</td>
<td>0.5560</td>
</tr>
<tr>
<td>on(y)</td>
<td>N/A</td>
<td>on(y)</td>
<td>0.6683</td>
</tr>
</tbody>
</table>
Results

“Touching”

<table>
<thead>
<tr>
<th>Movement</th>
<th>P(accept Movement)</th>
<th>Movement</th>
<th>P(accept Movement)</th>
</tr>
</thead>
<tbody>
<tr>
<td>behind→behind(y)</td>
<td>0.5347</td>
<td>left→behind(y)</td>
<td>0.5732</td>
</tr>
<tr>
<td>behind→in_front(y)</td>
<td>0.4758</td>
<td>left→in_front(y)</td>
<td>0.5853</td>
</tr>
<tr>
<td>behind→left(y)</td>
<td>0.5014</td>
<td>left→left(y)</td>
<td>0.5266</td>
</tr>
<tr>
<td>behind→right(y)</td>
<td>0.4888</td>
<td>left→right(y)</td>
<td>0.5211</td>
</tr>
<tr>
<td>behind→on(y)</td>
<td>0.7453</td>
<td>left→on(y)</td>
<td>0.6492</td>
</tr>
<tr>
<td>in_front→behind(y)</td>
<td>0.4523</td>
<td>right→behind(y)</td>
<td>0.5406</td>
</tr>
<tr>
<td>in_front→in_front(y)</td>
<td>0.6447</td>
<td>right→in_front(y)</td>
<td>0.5786</td>
</tr>
<tr>
<td>in_front→left(y)</td>
<td>0.4601</td>
<td>right→left(y)</td>
<td>0.4777</td>
</tr>
<tr>
<td>in_front→right(y)</td>
<td>0.5756</td>
<td>right→right(y)</td>
<td>0.5847</td>
</tr>
<tr>
<td>in_front→on(y)</td>
<td>0.6234</td>
<td>right→on(y)</td>
<td>0.7081</td>
</tr>
</tbody>
</table>
Results

\[ \mu_{mov} \approx 0.56236 \]
\[ \sigma_{mov} \approx 0.08108 \]

- Notable inclination against depictions where theme moves from “behind” dest to “in front,” and vice versa
  - \( P(\text{accept}|\text{behind} \rightarrow \text{in_front}(y)) \approx 0.4758 \approx \mu_{mov} - 1.07\sigma_{mov} \)
  - **Hypothesis**: POV makes it difficult to see if objects are actually touching
\[ \mu_{end} \approx 0.57256 \]
\[ \sigma_{end} \approx 0.06280 \]

- Significant inclination against depictions where theme ends to the left of dest
  - \[ P(\text{accept}|\text{left}(y)) \approx 0.4995 \approx \mu_{end} - 1.16\sigma_{end} \]
  - Apparently independent of theme’s starting location
    - More significant \( \text{in_front} \rightarrow \text{left}(y) \) and \( \text{right} \rightarrow \text{left}(y) \)
    - \[ P(\text{accept}|\text{in_front} \rightarrow \text{left}(y)) \approx 0.4601 \approx \mu_{mov} - 1.26\sigma_{mov} \]
    - \[ P(\text{accept}|\text{right} \rightarrow \text{left}(y)) \approx 0.4777 \approx \mu_{mov} - 1.04\sigma_{mov} \]
Preference for “on” specification over others

- \( P(\text{accept}|on(y)) \approx 0.6683 \approx \mu_{\text{end}} + 1.52\sigma_{\text{end}} \)
- Strongest from \( behind \rightarrow on(y) \)
- \( P(\text{accept}|behind \rightarrow on(y)) \approx 0.7453 \approx \mu_{\text{mov}} + 2.25\sigma_{\text{mov}} \)

**Hypothesis**: Occluded theme is being brought into view
Results

“Near”

| Distance quintile | P(accept|QU) |
|-------------------|-----------|
| First             | 0.7523    |
| Second            | 0.6207    |
| Third             | 0.3890    |
| Fourth            | 0.3655    |
| Fifth             | 0.1295    |
## Results

### “Near”

<table>
<thead>
<tr>
<th>Distance quintile</th>
<th>QSR (event end)</th>
<th>$P(\text{accept} \mid \text{QU,QSR})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>behind($y$)</td>
<td>0.7730</td>
</tr>
<tr>
<td>First</td>
<td>in_front($y$)</td>
<td>0.7349</td>
</tr>
<tr>
<td>First</td>
<td>left($y$)</td>
<td>0.7338</td>
</tr>
<tr>
<td>First</td>
<td>right($y$)</td>
<td>0.7712</td>
</tr>
<tr>
<td>Second</td>
<td>behind($y$)</td>
<td>0.6701</td>
</tr>
<tr>
<td>Second</td>
<td>in($y$)</td>
<td>0.5797</td>
</tr>
<tr>
<td>Second</td>
<td>left($y$)</td>
<td>0.6675</td>
</tr>
<tr>
<td>Second</td>
<td>right($y$)</td>
<td>0.5819</td>
</tr>
<tr>
<td>Third</td>
<td>behind($y$)</td>
<td>0.4151</td>
</tr>
<tr>
<td>Third</td>
<td>in_front($y$)</td>
<td>0.3644</td>
</tr>
</tbody>
</table>

References

Krishnaswamy and Pustejovsky

Do You See What I See?
### “Near”

<table>
<thead>
<tr>
<th>Distance quintile</th>
<th>QSR (event end)</th>
<th>P(accept QU,QSR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Third</td>
<td>left(y)</td>
<td>0.3945</td>
</tr>
<tr>
<td>Third</td>
<td>right(y)</td>
<td>0.3825</td>
</tr>
<tr>
<td>Fourth</td>
<td>behind(y)</td>
<td>0.1713</td>
</tr>
<tr>
<td>Fourth</td>
<td>in_front(y)</td>
<td>0.4308</td>
</tr>
<tr>
<td>Fourth</td>
<td>left(y)</td>
<td>0.2093</td>
</tr>
<tr>
<td>Fourth</td>
<td>right(y)</td>
<td>0.4699</td>
</tr>
<tr>
<td>Fifth</td>
<td>behind(y)</td>
<td>0.0972</td>
</tr>
<tr>
<td>Fifth</td>
<td>in_front(y)</td>
<td>0.1401</td>
</tr>
<tr>
<td>Fifth</td>
<td>left(y)</td>
<td>0.1250</td>
</tr>
<tr>
<td>Fifth</td>
<td>right(y)</td>
<td>0.1348</td>
</tr>
</tbody>
</table>
Results

\[ \mu_{qu} \approx 0.45140 \]
\[ \sigma_{qu} \approx 0.24192 \]

- Strong preference for ending states in close proximity (unsurprising)
  - \[ P(\text{accept}|\text{First}) \approx 0.7523 \approx \mu_{qu} + 1.24\sigma_{qu} \]
  - \[ P(\text{accept}|\text{Second}) \approx 0.6207 \approx \mu_{qu} + 0.70\sigma_{qu} \]
Results

\[ \mu_{qu,qsr} \approx \{0.75322, 0.62480, 0.38913, 0.32033, 0.12428\} \]
\[ \sigma_{qu,qsr} \approx \{0.02181, 0.05083, 0.02128, 0.15178, 0.01910\} \]

- Apparent confusion in fourth distance quintile judgments (high \(\sigma\))
  - Could be due to uncertainty of whether theme object is nearer to dest at event end than at event start
- Weak preference for “behind” relations in first 3 quintiles
  - \(P(\text{accept}|\text{First}, \text{behind}(y)) \approx 0.7730 \approx \mu_{qu=1,qsr} + 0.90\sigma_{qu=1,qsr}\)
  - \(P(\text{accept}|\text{Second}, \text{behind}(y)) \approx 0.6701 \approx \mu_{qu=2,qsr} + 0.89\sigma_{qu=2,qsr}\)
  - \(P(\text{accept}|\text{Third}, \text{behind}(y)) \approx 0.4151 \approx \mu_{qu=3,qsr} + 1.22\sigma_{qu=3,qsr}\)
Weak preference for “behind” relations in first 3 quintiles

**Hypothesis:** Foreshortening effect caused by POV causes $behind(y)$ to appear closer than it actually is
Summary

- Recorded 1,210 individual videos
- Performed 3,236 individual evaluation tasks
  - A small number of responses were rejected due to evaluators failing to answer the required question
- Provides method for generating 3D visualizations using NL interface
- Provides platform to conduct experiments on observables of motion events
- Provides intuitive way to trace spatial cues and entailments through narrative
- Used to generate data on theoretical intuitions
- Enables broader study of event and motion semantics
Future Directions

- Visualization is just one available modality to model
- As technology improves, events may be simulated aurally, haptically, or proprioceptically
- AR or VR may afford examination of human perception in immersive environments
- VoxML and simulation can be used to drive robotic agents
  - Constructing isomorphic simulation of real situation
- Interdisciplinary nature affords many extensions into other disciplines, fields, specializations
References I


References II

References III


References IV


Krishnaswamy, Nikhil and James Pustejovsky (2016a). “Multimodal Semantic Simulations of Linguistically Underspecified Motion Events”. In: *Proceedings of Spatial Cognition*. 
References VI


References VII


References


References X


[References X]