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

Introduction

- 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)

Introduction

- 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

Introduction

- 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 $\mathcal{M} \vDash p_s[a]$ or $\mathcal{M} \nvDash p_s[a]$.

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

$$
\left[\begin{array}{l}
\textbf{cup}\\
\text{LEX} = \left[\begin{array}{l}\text{PRED} = \textbf{cup}\\ \text{TYPE} = \textbf{physobj}\end{array}\right]\\
\text{TYPE} = \left[\begin{array}{l}\text{HEAD} = \textbf{cylindroid[1]}\\ \text{COMPONENTS} = \textbf{surface,interior}\\ \text{CONCAVITY} = \textbf{concave}\\ \text{ROTATSYM} = \{Y\}\\ \text{REFLECTSYM} = \{XY, YZ\}\end{array}\right]\\
\text{HABITAT} = \left[\begin{array}{l}\text{INTR} = [2]\left[\begin{array}{l}\text{UP} = align(Y, \mathcal{E}_Y)\\ \text{TOP} = top(+Y)\end{array}\right]\\ \text{EXTR} = \ldots\end{array}\right]\\
\text{AFFORD\_STR} = \left[\begin{array}{l}\text{A}_1 = H[2] \rightarrow [put(x, on([1]))]support([1], x)\\ \text{A}_2 = H[2] \rightarrow [put(x, in([1]))]contain([1], x)\\ \text{A}_3 = H[2] \rightarrow [grasp(x, [1])]\end{array}\right]\\
\text{EMBODIMENT} = \left[\begin{array}{l}\text{SCALE} = \textbf{<agent}\\ \text{MOVABLE} = \textbf{true}\end{array}\right]
\end{array}\right]
$$

Figure: VoxML for a "cup"

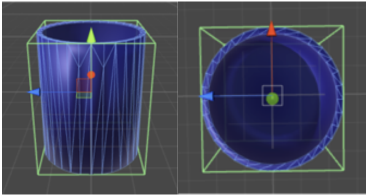
VoxML

$$\left[\begin{array}{l} \textbf{put} \\ \text{LEX} = \left[\begin{array}{l} \text{PRED} = \textbf{put} \\ \text{TYPE} = \textbf{transition\_event} \end{array}\right] \\ \text{TYPE} = \left[\begin{array}{l} \text{HEAD} = \textbf{transition} \\ \text{ARGS} = \left[\begin{array}{l} \text{A}_1 = \textbf{x:agent} \\ \text{A}_2 = \textbf{y:physobj} \\ \text{A}_3 = \textbf{z:location} \end{array}\right] \\ \text{BODY} = \left[\begin{array}{l} \text{E}_1 = grasp(x,y) \\ \text{E}_2 = [while(hold(x,y), move(y))] \\ \text{E}_3 = [at(y,z) \rightarrow ungrasp(x,y)] \end{array}\right] \end{array}\right] \end{array}\right] \quad \left[\begin{array}{l} \textbf{in} \\ \text{LEX} = \left[\begin{array}{l} \text{PRED} = \textbf{in} \end{array}\right] \\ \text{TYPE} = \left[\begin{array}{l} \text{CLASS} = \textbf{config} \\ \text{VALUE} = \textbf{ProperPart} \parallel \textbf{PO} \\ \text{ARGS} = \left[\begin{array}{l} \text{A}_1 = \textbf{x:3D} \\ \text{A}_2 = \textbf{y:3D} \end{array}\right] \\ \text{CONSTR} = \textbf{...} \end{array}\right] \end{array}\right]$$

Figure: VoxML for “put” and “in”

VoxML

- 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



VoxSim

http://www.voxicon.net/
http://www.github.com/VoxML/VoxSim

Architecture

- 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

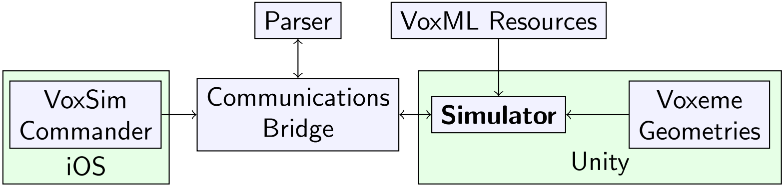
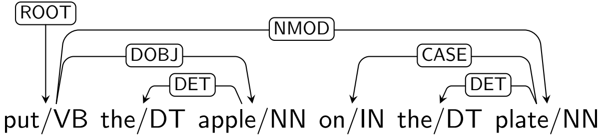


Figure: VoxSim architecture schematic

Architecture



1. $p := put(a[])$	5. $nmod := on(iobj)$
2. $dobj := the(b)$	6. $iobj := the(c)$
3. $b := (apple)$	7. $c := plate$
4. a.push$(dobj)$	8. a.push$(nmod)$
$put(the(apple),on(the(plate)))$	

Figure: Dependency parse for *Put the apple on the plate* and transformation to predicate-logic form.

Architecture

1. Input sentence
2. Generate parse
3. Compute satisfaction conditions from voxeme composition

$$
\begin{bmatrix}
\textbf{put} & \\
\textsc{type} = & \begin{bmatrix}
\textsc{head} = \textbf{process} \\
\textsc{args} = \begin{bmatrix} \textsc{A}_1 = \textbf{agent} \\ \textsc{A}_2 = \textbf{physobj} \\ \textsc{A}_3 = \textbf{location} \end{bmatrix} \\
\textsc{body} = \begin{bmatrix} \textsc{E}_1 = grasp(A_1, A_2) \\ \textsc{E}_2 = [while(hold(A_1, A_2), \\ \quad move(A_2))] \\ \textsc{E}_3 = [at(A_1, A_3) \rightarrow \\ \quad ungrasp(A_1, A_2)] \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
\quad
\begin{bmatrix}
\textbf{cup} & \\
\textsc{type} = & \begin{bmatrix}
\textsc{head} = \textbf{cylindroid[1]} \\
\textsc{components} = \textbf{surface,interior} \\
\textsc{concavity} = \textbf{concave} \\
\textsc{rotatSym} = \{Y\} \\
\textsc{reflectSym} = \{XY, YZ\}
\end{bmatrix} \\
\textsc{habitat} = & \begin{bmatrix} \textsc{Intr} = {}_{[2]}\begin{bmatrix} \textsc{up} = align(Y, \mathcal{E}_Y) \\ \textsc{top} = top(+Y) \end{bmatrix} \end{bmatrix} \\
\textsc{afford\_str} = & \begin{bmatrix}
\textsc{A}_1 = H[2] \rightarrow \\ \quad [put(x, on([1]))] \\ \quad support([1], x) \\
\textsc{A}_2 = H[2] \rightarrow \\ \quad [put(x, in([1]))] \\ \quad contain([1], x) \\
\textsc{A}_3 = H[2] \rightarrow \\ \quad [grasp(x, [1])]
\end{bmatrix}
\end{bmatrix}
$$

$$
\begin{bmatrix}
\textbf{in} & \\
\textsc{type} = & \begin{bmatrix}
\textsc{class} = \textbf{config} \\
\textsc{value} = \textbf{ProperPart} \parallel \textbf{PO} \\
\textsc{args} = \begin{bmatrix} \textsc{A}_1 = \textbf{x:3D} \\ \textsc{A}_2 = \textbf{y:3D} \end{bmatrix}
\end{bmatrix}
\end{bmatrix}
$$

Architecture

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

Semantic Processing

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?

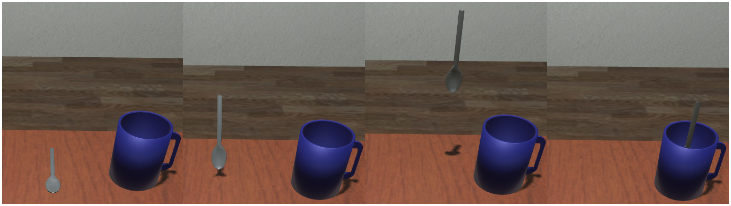


Figure: Object properties impose constraints on motion

Modeling Events

"LEAN" — Theoretical formulation:

- Instruction: "Lean [[THEME]] on [[DEST]]"
- Goal: [[THEME]] is supported by [[DEST]] at an angle θ
    - For this example, assume $\theta = 45°$

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

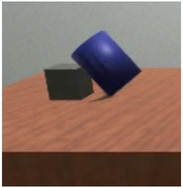


Figure: Desired goal state of "lean *x* on *y*"

Modeling Events

"LEAN" — Operationalization:

- Instruction: "Lean [[THEME]] on [[DEST]]"
- Goal: [[THEME]] is supported by [[DEST]] at an angle θ
  - For this example, assume $\theta = 45°$
- 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 $90°$-θ off from +Y axis
2. Turn [[THEME]] **about minor axis** such that major axis is θ 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**,V_1:**axis**,$\mathcal{E}_{V_2}$:**axis**) — turn object x so that object axis V_1 is aligned with world axis V_2
  2. TURN-2: turn(x:**obj**,V_1:**axis**,$\mathcal{E}_{V_2}$:**axis**,$\mathcal{E}_{V_3}$:**axis**) — turn object x so that object axis V_1 is aligned with world axis V_2, constraining motion to around world axis V_3
  3. PUT: put(x:**obj**,y:**loc**) — put object x at location y

$$\begin{bmatrix} \textbf{lean} \\ \text{LEX} = \begin{bmatrix} \text{PRED} = \textbf{lean} \\ \text{TYPE} = \textbf{transition\_event} \end{bmatrix} \\ \text{TYPE} = \begin{bmatrix} \text{HEAD} = \textbf{transition} \\ \text{ARGS} = \begin{bmatrix} \text{A}_1 = \textbf{x:agent} \\ \text{A}_2 = \textbf{y:physobj} \\ \text{A}_3 = \textbf{z:location} \end{bmatrix} \\ \text{BODY} = \begin{bmatrix} \text{E}_1 = grasp(x,y) \\ \text{E}_2 = [while(hold(x,y), turn(x,y, align(minor(y), \mathcal{E}_Y \times (90-\theta, about(\mathcal{E}_{\perp Y})))))] \\ \text{E}_3 = [while(hold(x,y), turn(x,y, align(major(y), \mathcal{E}_Y \times (\theta, about(\mathcal{E}_{\perp Y}))), about(minor(y))))] \\ \text{E}_4 = [while(hold(x,y), put(x,y))] \\ \text{E}_5 = [at(y,z) \rightarrow ungrasp(x,y)] \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

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

Experimental Design

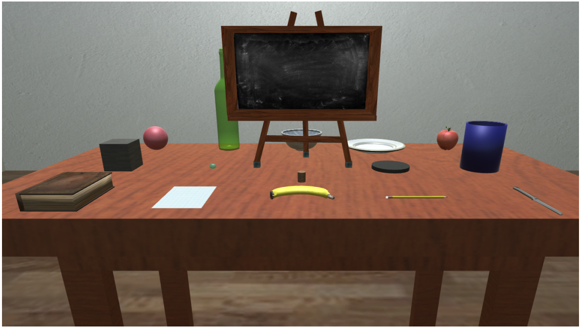


Figure: Test environment with all objects shown. During capture of an event, all objects not mentioned in the input sentence were removed.

Evaluation

- Raw results reflect overall incidence of evaluators accepting visualization for provided utterance
- Greater probability of acceptance → parameter values better reflect utterance
  - $\mathrm{P}(\text{acc} \mid V) \sim$ 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

Predicate	Underspecified parameters	Possible values
touching(x)	rel orientation	$\{left(x), right(x), behind(x), in\_front(x), on(x)\}$
near(x)	transloc dir	$V \in \{\langle y\text{-x}(x), y\text{-y}(x), y\text{-z}(x)\rangle \mid \mathrm{d}(x,y) < \mathrm{d}(\mathrm{edge}(\mathrm{s}(y),y)), IN(\mathrm{s}(y)), \neg IN(y)\}$

Table: Predicate value assignments

- "Touching" and "Near"
  - "Touching": discrete set
  - "Near": continuous range

Results

"Touching"

QSR (event start)	P(accept| QSR)	QSR (event end)	P(accept| QSR)
behind(y)	0.5497	*behind(y)*	0.5474
in_front(y)	0.5692	*in_front(y)*	0.5816
left(y)	0.5753	*left(y)*	0.4995
right(y)	0.5725	*right(y)*	0.5560
on(y)	N/A	*on(y)*	0.6683

Results

"Touching"

Movement	P(accept| Movement)	Movement	P(accept| Movement)
behind→behind(y)	0.5347	*left→behind(y)*	0.5732
behind→in_front(y)	0.4758	*left→in_front(y)*	0.5853
behind→left(y)	0.5014	*left→left(y)*	0.5266
behind→right(y)	0.4888	*left→right(y)*	0.5211
behind→on(y)	0.7453	*left→on(y)*	0.6492
in_front→behind(y)	0.4523	*right→behind(y)*	0.5406
in_front→in_front(y)	0.6447	*right→in_front(y)*	0.5786
in_front→left(y)	0.4601	*right→left(y)*	0.4777
in_front→right(y)	0.5756	*right→right(y)*	0.5847
in_front→on(y)	0.6234	*right→on(y)*	0.7081

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(accept|*behind→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

Results

$\mu_{end} \approx 0.57256$
$\sigma_{end} \approx 0.06280$

- Significant inclination against depictions where theme ends to the left of dest
  - P(accept|*left(y)*) $\approx 0.4995 \approx \mu_{end}$ - $1.16\sigma_{end}$
  - Apparently independent of theme's starting location
    - More significant *in_front→left(y)* and *right→left(y)*
    - P(accept|*in_front→left(y)*) $\approx 0.4601 \approx \mu_{mov}$ - $1.26\sigma_{mov}$
    - P(accept|*right→left(y)*) $\approx 0.4777 \approx \mu_{mov}$ - $1.04\sigma_{mov}$

Results

- Preference for "on" specification over others
  - P(accept|*on(y)*) $\approx 0.6683 \approx \mu_{end} + 1.52\sigma_{end}$
  - Strongest from *behind*→*on(y)*
  - P(accept|*behind*→*on(y)*) $\approx 0.7453 \approx \mu_{mov} + 2.25\sigma_{mov}$
  - **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"

Distance quintile	QSR (event end)	P(accept| QU,QSR)
First	*behind(y)*	0.7730
First	*in_front(y)*	0.7349
First	*left(y)*	0.7338
First	*right(y)*	0.7712
Second	*behind(y)*	0.6701
Second	*in(y)*	0.5797
Second	*left(y)*	0.6675
Second	*right(y)*	0.5819
Third	*behind(y)*	0.4151
Third	*in_front(y)*	0.3644

Results

"Near"

Distance quintile	**QSR (event end)**	**P(accept| QU,QSR)**
Third	*left(y)*	0.3945
Third	*right(y)*	0.3825
Fourth	*behind(y)*	0.1713
Fourth	*in_front(y)*	0.4308
Fourth	*left(y)*	0.2093
Fourth	*right(y)*	0.4699
Fifth	*behind(y)*	0.0972
Fifth	*in_front(y)*	0.1401
Fifth	*left(y)*	0.1250
Fifth	*right(y)*	0.1348

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 σ)
  - 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
  - $P(\text{accept}|\text{First}, \textit{behind(y)}) \approx 0.7730 \approx \mu_{qu=1,qsr} + 0.90\sigma_{qu=1,qsr}$
  - $P(\text{accept}|\text{Second}, \textit{behind(y)}) \approx 0.6701 \approx \mu_{qu=2,qsr} + 0.89\sigma_{qu=2,qsr}$
  - $P(\text{accept}|\text{Third}, \textit{behind(y)}) \approx 0.4151 \approx \mu_{qu=3,qsr} + 1.22\sigma_{qu=3,qsr}$

Results

- 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 proprioceptively
- 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

Thank You!

References I

Albath, Julia et al. (2010). "RCC-3D: Qualitative Spatial Reasoning in 3D.". In: *CAINE*, pp. 74–79.

Allen, James (1984). "Towards a general theory of action and time". In: *Arificial Intelligence* 23, pp. 123–154.

Bergen, Benjamin K. (2012). *Louder than words: The new science of how the mind makes meaning*. Basic Books.

Chang, Angel et al. (2015). "Text to 3D Scene Generation with Rich Lexical Grounding". In: *arXiv preprint arXiv:1505.06289*.

Coyne, Bob and Richard Sproat (2001). "WordsEye: an automatic text-to-scene conversion system". In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, pp. 487–496.

References II

Davis, Ernest and Gary Marcus (2016). "The scope and limits of simulation in automated reasoning". In: *Artificial Intelligence* 233, pp. 60–72.

De Saussure, Ferdinand (1915). "Course in general linguistics (1915)". In: *New York: Philosophical Library.[JL].*

Dill, Kevin (2011). "A game AI approach to autonomous control of virtual characters". In: *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).*

Dylla, Frank and Reinhard Moratz (2004). "Exploiting qualitative spatial neighborhoods in the situation calculus". In: *International Conference on Spatial Cognition*. Springer, pp. 304–322.

References III

Falomir, Zoe and Thomas Kluth (2017). "Qualitative spatial logic descriptors from 3D indoor scenes to generate explanations in natural language". In: *Cognitive Processing*, pp. 1–20.

Feldman, Jerome and Srinivas Narayanan (2004). "Embodied meaning in a neural theory of language". In: *Brain and language* 89.2, pp. 385–392.

Forbus, Kenneth D., James V. Mahoney, and Kevin Dill (2002). "How qualitative spatial reasoning can improve strategy game AIs". In: *IEEE Intelligent Systems* 17.4, pp. 25–30.

Frank, Andrew U (1992). "Qualitative spatial reasoning about distances and directions in geographic space". In: *Journal of Visual Languages & Computing* 3.4, pp. 343–371.

References IV

Frank, Andrew U (1996). “Qualitative spatial reasoning: Cardinal directions as an example”. In: *International Journal of Geographical Information Science* 10.3, pp. 269–290.

Freksa, Christian (1992). *Using orientation information for qualitative spatial reasoning*. Springer.

Gelfond, Michael and Vladimir Lifschitz (1988). “The stable model semantics for logic programming.”. In: *ICLP/SLP*. Vol. 88, pp. 1070–1080.

Gibbs Jr., Raymond W (2005). *Embodiment and cognitive science*. Cambridge University Press.

Gibson, James J. (1977). “The Theory of Affordances”. In: *Perceiving, Acting, and Knowing: Toward an ecological psychology*, pp. 67–82.

References V

- Gibson, James J. (1979). *The Ecology Approach to Visual Perception: Classic Edition*. Psychology Press.
- Joskowicz, Leo and Elisha P. Sacks (1991). "Computational kinematics". In: *Artificial Intelligence* 51.1-3, pp. 381–416.
- Kaplan, David (1979). "On the logic of demonstratives". In: *Journal of philosophical logic* 8.1, pp. 81–98.
- Kiela, Douwe et al. (2016). "Virtual Embodiment: A Scalable Long-Term Strategy for Artificial Intelligence Research". In: *arXiv preprint arXiv:1610.07432*.
- Krishnaswamy, Nikhil and James Pustejovsky (2016a). "Multimodal Semantic Simulations of Linguistically Underspecified Motion Events". In: *Proceedings of Spatial Cognition*.

References VI

Krishnaswamy, Nikhil and James Pustejovsky (2016b). "VoxSim: A Visual Platform for Modeling Motion Language". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. ACL, pp. 54–58.

Kuipers, Benjamin (1994). *Qualitative reasoning: modeling and simulation with incomplete knowledge*. MIT press.

– (2000). "The spatial semantic hierarchy". In: *Artificial Intelligence* 119.1, pp. 191–233.

Lakoff, George (2009). "The neural theory of metaphor". In: *Available at SSRN 1437794*.

Mani, Inderjeet and James Pustejovsky (2012). *Interpreting Motion: Grounded Representations for Spatial Language*. Oxford University Press.

References VII

McDonald, David and James Pustejovsky (2014). "On the Representation of Inferences and their Lexicalization". In: *Advances in Cognitive Systems*. Vol. 3.

Moratz, Reinhard, Jochen Renz, and Diedrich Wolter (2000). "Qualitative spatial reasoning about line segments". In: *Proceedings of the 14th European Conference on Artificial Intelligence*. IOS Press, pp. 234–238.

Pustejovsky, James (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

– (2013). "Dynamic Event Structure and Habitat Theory". In: *Proceedings of the 6th International Conference on Generative Approaches to the Lexicon (GL2013)*. ACL, pp. 1–10.

References VIII

Pustejovsky, James and Nikhil Krishnaswamy (2014). “Generating Simulations of Motion Events from Verbal Descriptions”. In: *Lexical and Computational Semantics (* SEM 2014)*, p. 99.

– (2016). “VoxML: A Visualization Modeling Language”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Ed. by Nicoletta Calzolari (Conference Chair) et al. Portoroz, Slovenia: European Language Resources Association (ELRA). ISBN: 978-2-9517408-9-1.

Pustejovsky, James and Jessica Moszkowicz (2011). “The qualitative spatial dynamics of motion”. In: *The Journal of Spatial Cognition and Computation*.

References IX

- Randell, D.A. et al. (1992). "A spatial logic based on regions and connection". In: *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*. Morgan Kaufmann. San Mateo, pp. 165–176.
- Renz, Jochen and Bernhard Nebel (2007). "Qualitative spatial reasoning using constraint calculi". In: *Handbook of spatial logics*, pp. 161–215.
- Rusu, Radu Bogdan et al. (2008). "Towards 3D point cloud based object maps for household environments". In: *Robotics and Autonomous Systems* 56.11, pp. 927–941.
- Sawilowsky, Shlomo S (2003). "You think you,Äôve got trivials?". In: *Journal of Modern Applied Statistical Methods* 2.1, p. 21.

References X

Siskind, Jeffrey Mark (2001). "Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic". In: *J. Artif. Intell. Res.(JAIR)* 15, pp. 31–90.

Thrun, Sebastian et al. (2000). "Probabilistic algorithms and the interactive museum tour-guide robot Minerva". In: *The International Journal of Robotics Research* 19.11, pp. 972–999.

Ziemke, Tom (2003). "What's that thing called embodiment?". In: *Proceedings of the 25th Annual meeting of the Cognitive Science Society*. Citeseer, pp. 1305–1310.

Zimmermann, Kai and Christian Freksa (1996). "Qualitative spatial reasoning using orientation, distance, and path knowledge". In: *Applied intelligence* 6.1, pp. 49–58.