Voxeme Modeling from 3D Geometry Library

- Executable available at `http://www.voxicon.net`
- Object voxemes consist of geometry + VoxML markup

Behavior Attachment to a Voxeme

- Afforded behaviors require habitat conditions to be satisfied

$H_{[2]} \rightarrow [put(x, on[1])]support([1], x)$ can be paraphrased as "In habitat 2, x can be put on component 1, resulting in component 1 supporting x"
$H_{[3]} \rightarrow [grasp(x, [1])]$ can be paraphrased as "In habitat 3, component 1 can be grasped by x"
$H_{[4]}, grasp(x, [1]) \rightarrow [lift(x, [1])]$ can be paraphrased as "In habitat 4, if x is grasping component 1, component 1 can be lifted by x"

Adding Discriminating Attributes to Voxemes

- Discriminating attributes may be nominal, such as color
    - e.g., red, blue, green, black, etc.
- or sortal, such as relative location
    - e.g., leftmost, center, rightmost

Creating Novel Behavior

- "Switch two cups"
  - Interpretation: swap the locations of two cups in scene

$$\begin{bmatrix} \textbf{switch} \\ \text{LEX} = \begin{bmatrix} \text{PRED} = \textbf{switch} \\ \text{TYPE} = \textbf{transition\_event} \end{bmatrix} \\ \text{TYPE} = \begin{bmatrix} \text{HEAD} = \textbf{transition} \\ \text{ARGS} = \begin{bmatrix} \text{A}_1 = \textbf{y[]:physobj} \end{bmatrix} \\ \text{BODY} = \begin{bmatrix} \begin{array}{lll} \text{E}_1 & = & def(w, as(loc(y[0]))), \\ & & def(v, as(loc(y[1]))) \\ \text{E}_2 & = & put(y[0], in\_front(v)) \\ \text{E}_3 & = & put(y[1], w)) \\ \text{E}_4 & = & put(y[0], v)) \end{array} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Creating Novel Behavior

- "Switch two cups"
  - Interpretation: swap the locations of two cups in scene

$$\left[\begin{array}{l} \textbf{switch} \\ \text{LEX} = \left[\begin{array}{lcl} \text{PRED} & = & \textbf{switch} \\ \text{TYPE} & = & \textbf{transition\_event} \end{array}\right] \\ \text{TYPE} = \left[\begin{array}{lcl} \text{HEAD} & = & \textbf{transition} \\ \text{ARGS} & = & \left[\, \text{A}_1 = \textbf{y[]:physobj} \,\right] \\ \text{BODY} & = & \left[\begin{array}{lcl} \text{E}_1 & = & \mathit{def}(w, \mathit{as}(\mathit{loc}(y[0]))), \\ & & \mathit{def}(v, \mathit{as}(\mathit{loc}(y[1]))) \\ \text{E}_2 & = & \mathit{slide}(y[0], \mathit{in\_front}(v)) \\ \text{E}_3 & = & \mathit{slide}(y[1], w)) \\ \text{E}_4 & = & \mathit{slide}(y[0], v)) \end{array}\right] \end{array}\right] \end{array}\right]$$

References I

- Antol, Stanislaw et al. (2015). "Vqa: Visual question answering". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2425–2433.
- Bunt, Harry, Robbert-Jan Beun, and Tijn Borghuis (1998). *Multimodal human-computer communication: systems, techniques, and experiments*. Vol. 1374. Springer Science & Business Media.
- Chang, Angel et al. (2015). "Text to 3D Scene Generation with Rich Lexical Grounding". In: *arXiv preprint arXiv:1505.06289*.
- Chao, Yu-Wei et al. (2015a). "HICO: A benchmark for recognizing human-object interactions in images". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1017–1025.

References II

Chao, Yu-Wei et al. (2015b). “Mining semantic affordances of visual object categories”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4259–4267.

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.

Jacko, Julie A (2012). *Human computer interaction handbook: Fundamentals, evolving technologies, and emerging applications*. CRC press.

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

References III

Pustejovsky, James and Nikhil Krishnaswamy (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.

Rautaray, Siddharth S and Anupam Agrawal (2015). "Vision based hand gesture recognition for human computer interaction: a survey". In: *Artificial Intelligence Review* 43.1, pp. 1–54.

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.

Turk, Matthew (2014). "Multimodal interaction: A review". In: *Pattern Recognition Letters* 36, pp. 189–195.