Visualizing Events:
Simulating Meaning in Language

James Pustejovsky (jamesp@cs.brandeis.edu)
Department of Computer Science, Brandeis University 415 South Street
Waltham, MA 02453 USA

Nikhil Krishnaswamy (nkrishna@brandeis.edu)
Department of Computer Science, Brandeis University 415 South Street
Waltham, MA 02453 USA

Keywords: event semantics; simulation semantics; habitats; affordances; qualia structure; dynamic event models

Background

In this note, we present results from a new approach to modeling the semantics of natural language, *Multimodal Semantic Simulations (MSS)*, being pursued by the first author and his students. This approach assumes both a richer formal model of events and their participants, as well as a modeling language for constructing 3D visualizations of objects and events denoted by natural language expressions. The Dynamic Event Model (DEM) encodes events as programs in a dynamic logic with an operational semantics, while the language VoxML, Visual Object Concept Modeling Language, is being used as the platform for *multimodal semantic simulations* in the context of human-computer communication.

For over two decades, Pustejovsky has been investigating the nature of events in language: the Aktionsarten and event typing associated with lexemes and phrases; how they are mapped to syntax; and how nominal semantics interacts with type coercion behaviors involving events, as modeled in Generative Lexicon. Generative Lexicon has developed a fairly sophisticated model (both data structures and mechanisms) for how the function and use of an object can be encoded as part of its lexical semantics. In addition, he has worked to develop open standards for encoding the temporal and spatial information associated with events. The resulting languages, ISO-TimeML and ISOspace, are ISO standards for their areas, and ISO-TimeML has become the most widely adopted language within the field for representing temporal information in open domains.

Multimodal Semantic Simulation Theory

According to Goldman (2006), simulation provides a process-driven theory of mind and mental attribution, differing from the theory-driven models. From the cognitive linguistics tradition, simulation semantics has come to denote the mental instantiation of an interpretation of any linguistic utterance (Feldman, 2006; Bergen, 2012). See (Markman, Klein, & Suhr, 2012) for a general review of simulation theory in psychology. While these communities do not often reference each other, it is clear from our perspective, that they are pursuing similar programs, where distinct linguistic utterances correspond to generated models that have differentiated structures and behaviors (Narayanan, 1999; Siskind, 2001; Goldman, 2006).

Prior work in visualization from natural language has largely focused on object placement and orientation in static scenes (Coyne & Sproat, 2001; Siskind, 2001; Chang, Monroe, Savva, Potts, & Manning, 2015). In previous work (Pustejovsky & Krishnaswamy, 2014; Pustejovsky, 2013a), we introduced a method for modeling natural language expressions within a 3D simulation environment, Unity. The goal of that work was to evaluate, through explicit visualizations of linguistic input, the semantic presuppositions inherent in the different lexical choices of an utterance. This work led to two additional lines of research: an explicit encoding for how an object is itself situated relative to its environment; and an operational characterization of how an object changes its location or how an agent acts on an object over time. The former has developed into a semantic notion of situational context, called a *habitat* (Pustejovsky, 2013a; McDonald & Pustejovsky, 2014), while the latter is addressed by dynamic interpretations of event structure (Pustejovsky & Moszkowicz, 2011; Pustejovsky, 2013b; Mani & Pustejovsky, 2012; Pustejovsky, 2013a). The requirements on a ”multimodal simulation semantics” include, but are not limited to, the following components:

(1) a. 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;
b. Object-based attributes for participants in a situation or event need to be specified; e.g., orientation, relative size, default position or pose, etc.;
c. An epistemic condition on the object and event rendering, imposing an implicit point of view (POV);
d. Agent-dependent embodiment; this determines the relative scaling of an agent and its event participants and their surroundings, as it engages in the environment.

In order to construct a robust simulation from linguistic input, an event and its participants must be embedded within an appropriate minimal embedding space. This must sufficiently enclose the event localization, while optionally including room enough for a frame of reference visualization of the event (the viewer's perspective).

Dynamic Event Models

Models of processes and events using updating typically make reference to the notion of a state transition (Harel, Kozen, & Tiuyn, 2000). This is done by distinguishing between formulae, ϕ, and programs, π. A formula is interpreted as a classical propositional expression, with assignment of a truth value in a specific model. We interpret models by reference to specific states, which is a set of propositions with assignments to variables at a specific index. Atomic programs are input/output relations ($[\![\pi]\!] \subseteq S \times S$), and compound programs are constructed from atomic ones following rules of dynamic logic. Events are interpreted as state sequences (sentences), created by the programs (transitions) and tests between them, which can be seen as *dependencies*. An event can be modeled as a dependency graph, where:

(2) a. Events are word sequences, $\langle e_1, e_2, \ldots, e_n \rangle$.
  b. We "label" the dependencies (programs) between words as in syntactic dependency structures.
  c. We distinguish between the *Object Model* and *Action Model*.
  d. An *Event Model* is the composition of these two.

Dependency labels are of three kinds:

(3) a. **Program** (object and action): the program moving between states.
  b. **Test**: conditions that must hold during a transition
  c. **Content**: the propositional content of a state

VoxML: Visual Object Modeling Language

We have developed a specification for a modeling language for constructing 3D visualizations of natural language expressions, used as the platform for creating multimodal semantic simulations of spatial and event semantics (Pustejovsky & Krishnaswamy, 2016). This modeling language allows for an encoding of how an object is situated relative to its environment, and an operational characterization of how an object changes its location (the object model) or how an agent acts on an object over time (the action model), thus allowing both design and representation of data structures that generate simulations of compositions of these two models into the aforementioned event model. The VoxML specification allows for dependency-parsed linguistic data to be annotated and transformed into a Dynamic Event Model, expressed within dynamic interval temporal logic (DITL), and from there into a procedurally generated visual simulation of the described event (for example, see Figures 1 and 2) .

put(obj,y)
loc(obj) := x, target(obj) := y; b := x; $(x := w; x \neq w; d(b,x) < d(b,w), d(b,y) > d(y,w)^+$

Figure 1: DEM/DITL expressions for *put(obj,y)*.

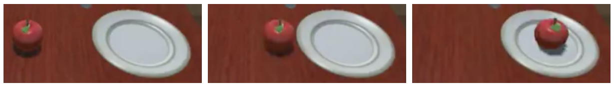

Figure 2: Program is executed

Future Directions

We are currently creating simulations of caused motion, where causation is not needed as a primitive in the logic. Rather, we employ a technique we call *rigging*, where the affected object and the actor are rigged together as one moving object. This is computationally very efficient and has interesting consequences for how to encode event causation.

Relevant Publications

Pustejovsky (2013a), Pustejovsky (2013b), Pustejovsky and Moszkowicz (2011), Pustejovsky and Krishnaswamy (2014), Pustejovsky and Krishnaswamy (2016), Mani and Pustejovsky (2012), McDonald and Pustejovsky (2014).

References

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

Chang, A., Monroe, W., Savva, M., Potts, C., & Manning, C. D. (2015). Text to 3d scene generation with rich lexical grounding. *arXiv preprint arXiv:1505.06289*.

Coyne, B., & Sproat, R. (2001). Wordseye: an automatic text-to-scene conversion system. In *Proceedings of the 28th annual conference on computer graphics and interactive techniques* (pp. 487–496).

Feldman, J. (2006). *From molecule to metaphor: A neural theory of language*. MIT press.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. OUP.

Harel, D., Kozen, D., & Tiuyn, J. (2000). *Dynamic logic* (1st ed.). The MIT Press.

Mani, I., & Pustejovsky, J. (2012). *Interpreting motion: Grounded representations for spatial language*. OUP.

Markman, K. D., Klein, W. M., & Suhr, J. A. (2012). *Handbook of imagination and mental simulation*. Psychology.

McDonald, D., & Pustejovsky, J. (2014). On the representation of inferences and their lexicalization. In *Advances in cognitive systems* (Vol. 3).

Narayanan, S. (1999). Reasoning about actions in narrative understanding. *IJCAI*, *99*, 350–357.

Pustejovsky, J. (2013a). Dynamic event structure and habitat theory. In *Proceedings of the 6th international conference on generative approaches to the lexicon* (pp. 1–10).

Pustejovsky, J. (2013b). Where things happen: On the semantics of event localization. In *Proceedings of isa-9: International workshop on semantic annotation.*

Pustejovsky, J., & Krishnaswamy, N. (2014). Generating simulations of motion events from verbal descriptions. *Lexical and Computational Semantics (* SEM 2014)*, 99.

Pustejovsky, J., & Krishnaswamy, N. (2016). Voxml: A visual object modeling language. *Proceedings of LREC*.

Pustejovsky, J., & Moszkowicz, J. (2011). The qualitative spatial dynamics of motion. *The Journal of Spatial Cognition and Computation*.

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