

Research Abstracts - 2007

Introduction

Perception & Learning

Physical, Biological & Social Systems

Systems

Theory

Reasoning by Imagining: The Neo-Bridge System

Mark A. Finlayson & Patrick H. Winston

Imagination as Common-Sense Reasoning

When asked questions such as "What is the shape of a St. Bernard's ears?" or "John and Mary kissed; did they touch?" most people report that they resort to their imagination to provide the answer. (Kosslyn 1980). In the Neo-Bridge project we seek to explore how imagination and language-understanding interact by building a system that can appeal to an internal imagined visual scene to answer questions. The system is called the *Bridge* system in acknowledgment of the idea that our intelligence relies on a "bridge" between language and vision to compute the answers to various questions. The system is the *Neo-Bridge* system because it is an updated version of the earlier, NSF-funded *Bridge* system (Bender, 2001; Bonawitz, 2003; Larson, 2003; Molnar, 2001; Shadadi, 2003). The Neo-Bridge system takes the previous *Bridge* system the next step by introducing state-of-the-art statistical natural processing in the language module, a sophisticated three-dimensional game engine for the imagination module, and new insights on the use of *Spelke* constraints (Spelke, 1990) in the question-and-explanation module. A diagram of the system architecture is shown in *Figure 1*.

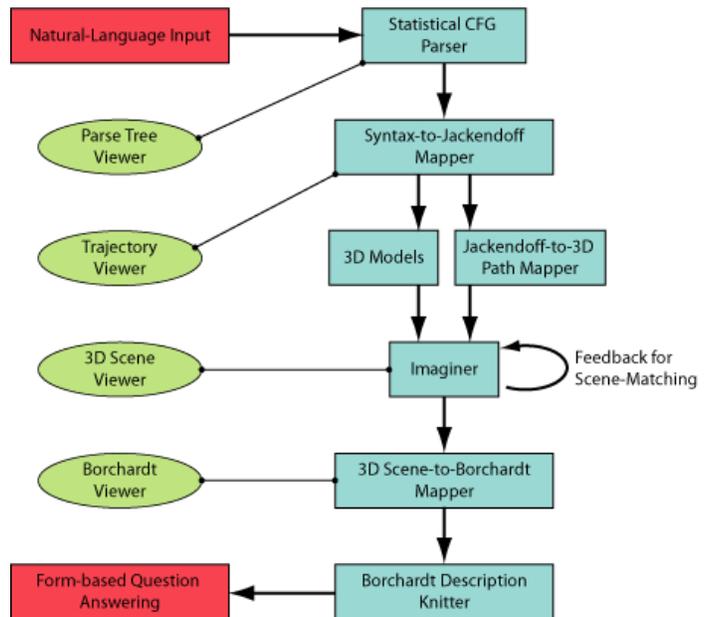


Figure 1: A Block diagram of the Neo-Bridge System. The user inputs natural language in the upper left box. This flows through the system, producing, in order, a CFG parse tree, a Jackendoff path representation, a 3D scene, a Borchardt description, and finally, a set of questions the system is able to answer about the described event.

Current State of the Neo-Bridge System

Currently the Neo-Bridge system uses the Stanford Natural Language Group's freely-available statistical parser to achieve wide language coverage (Manning, 2007). Using a set of home-grown syntax-to-semantics mappings, we construct a so-called *Jackendoff* trajectory representation (Jackendoff, 1983) that captures the movement of objects along paths (concrete or abstract) for visualization in the imaginer. *Figure 2* shows a parse of a complicated sentence by the Stanford parser, and it's translation into a Jackendoff trajectory frame describing motion along path. The imaginer uses the open-source JMonkey java 3D game engine (Powell, 2007), and a collection of freely-available 3D models to produce imagined scenes. The imagined scenes are then "unimagined" into a *Borchardt* representation (Borchardt, 1994) that contains information that was not available in the linguistic representation concerning object contact and relative motion, as well as hints as to opportunities for additional speculative reasoning by the system (violations of the *Spelke* constraints). The next step for the system is to complete the question-answering module, and introduce feedback into the system so that the imagined scene can be perturbed to test if it can be brought into a state consistent with some particular answer to a user question.

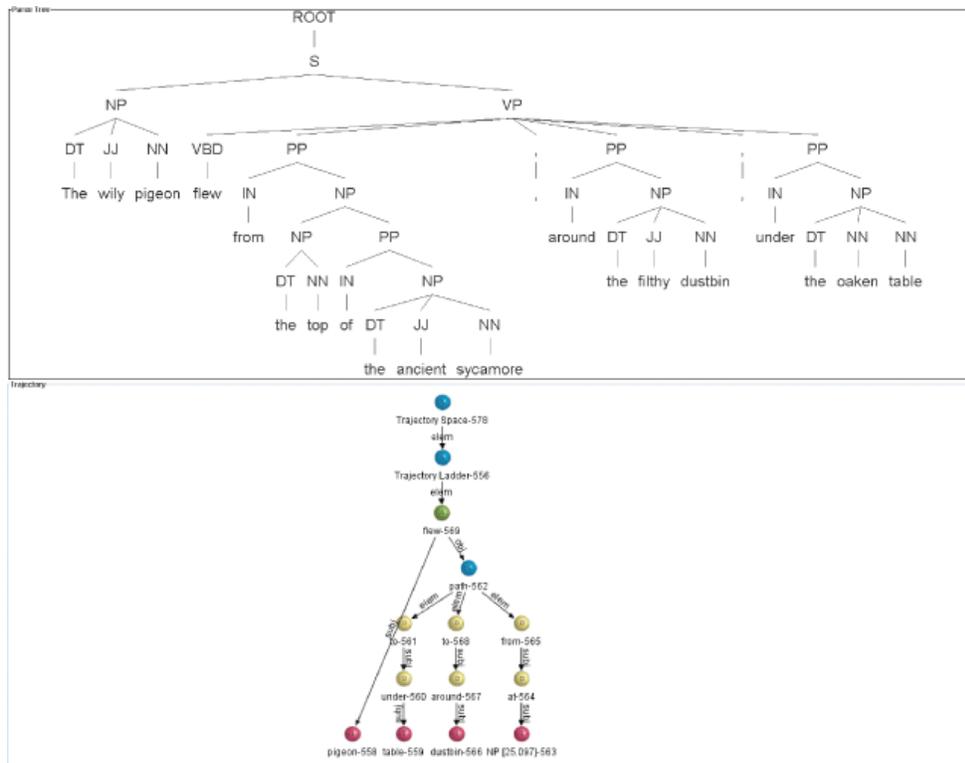


Figure 2: Shown in the top half of the figure is the probabilistic CFG parse of the sentence "The wily pigeon flew from the top of the ancient sycamore, around the filthy dustbin, under the oaken table." On the bottom shows the trajectory representation of this sentence, which has extracted the fact that the pigeon (the far lower left red node) is moving along a path from the table, by the dustbin, to the table.

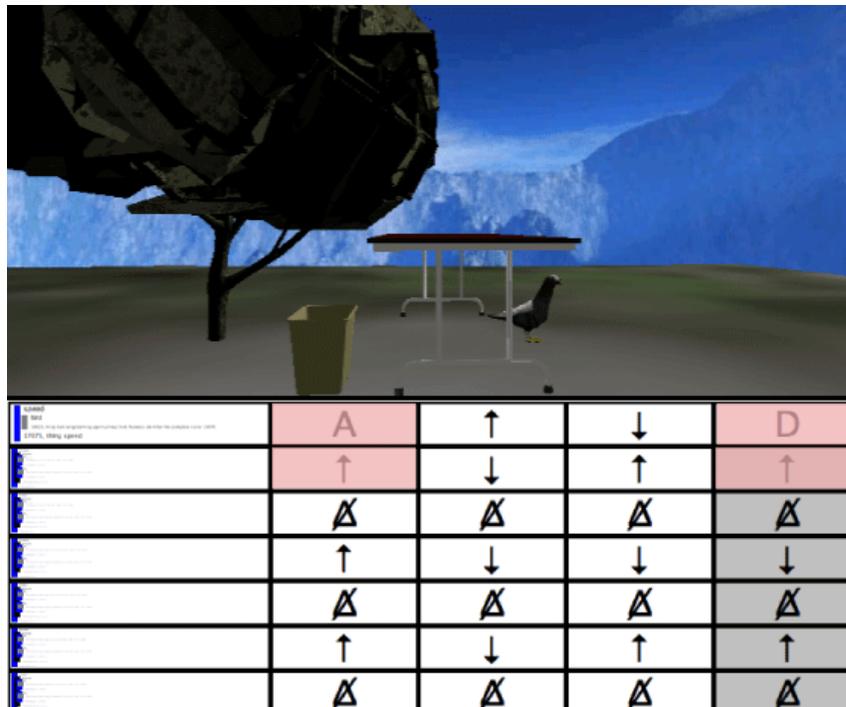


Figure 3: Shown in the top is a frame from the dynamic imagined scene from the sentence parsed in the previous figure. Below is the borchart diagram corresponding to the dynamic aspects of the scene, including the speed of the bird and the contact (or lack thereof) between objects. The leftmost column indicates the properties that are being tracked; the other columns indicate changes in those variables at times in the scene. A shaded pink cell means there was a possible violation of a Spelke principle, and these timesteps are candidates for further investigation. The symbols have the following meanings: A = "appear", D = "disappear", up = "increases", down = "decreases", slashed-triangle = "does not change"

Acknowledgements

Affiliated MIT undergraduate students are Mark Seifter, Harold Cooper, and Diana Moore. This project is funded by the NSF through grants 0211861 and IIS-0413206.

References:

Bender, J. R. (2001). *Connecting Language and Vision Using a Conceptual Semantics*. Masters of Engineering Thesis, MIT. Cambridge, MA.

Bonawitz, K. (2003). *Bidirectional Natural Language Parsing using Streams and Counterstreams*. Masters of Engineering Thesis, MIT. Cambridge, MA.

Borchardt, G. C. (1994). *Thinking between the Lines: Computers and the Comprehension of Causal Descriptions*. Cambridge, MA, MIT Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA. MIT Press.

Kosslyn, S. M. (1980). *Image and Mind*. Cambridge, MA. Harvard University Press.

Larson, S. (2003). *Intrinsic Representation: Bootstrapping Symbols from Experience*. Masters of Engineering Thesis, MIT. Cambridge, MA.

Manning, C.D. (2007). Stanford Natural Language Processing Group Parser, <http://nlp.stanford.edu/downloads/lex-parser.shtml>

Molnar, R. A. (2001). *Generalize and Sift as a Model of Inflection Acquisition*. Masters of Engineering Thesis, MIT. Cambridge, MA.

Powell, M. (2007). JMonkeyEngine, <http://www.jmonkeyengine.com>

Shadadi, A. (2003). *Barnyard Politics: A Decision Rationale Representation for the Analysis of Simple Political Situations*. Masters of Engineering Thesis, MIT. Cambridge, MA.

Spelke, E. S. (1990). Principles of Object Perception. *Cognitive Science* **14**, 29-56.