Biologically Inspired Cognitive Architectures 11, Jan. 2015, 22–28 (pre-publication version)

On the Need for Imagistic Modeling in Story Understanding

Eric Bigelow[1], Daniel Scarafoni[2], Lenhart Schubert[3]*, and Alex Wilson[4]

[1] University of Rochester
ebigelow@u.rochester.edu
[2] University of Rochester
dscarafo@u.rochester.edu
[3] University of Rochester
schubert@cs.rochester.edu
[4] University of Rochester
alexwilson@rochester.edu

Abstract
There is ample evidence that human understanding of ordinary language relies in part on a rich capacity for imagistic mental modeling. We argue that genuine language understanding in machines will similarly require an imagistic modeling capacity enabling fast construction of instances of prototypical physical situations and events, whose participants are drawn from a wide variety of entity types, including animate agents. By allowing fast evaluation of predicates such as *'can-see'*, *'under'*, and *'inside'*, these model instances support coherent text interpretation. Imagistic modeling is thus a crucial – and not very broadly appreciated – aspect of the long-standing knowledge acquisition bottleneck in AI. We will illustrate how the need for imagistic modeling arises even in the simplest first-reader stories for children, and provide an initial feasibility study to indicate what the architecture of a system combining symbolic with imagistic understanding might look like.

Keywords: imagistic modeling, natural language understanding, knowledge acquisition bottleneck, NLU architecture

1 Introduction

> *"Linguistic terms may not in the first place describe or represent meanings as such, but rather serve as triggers for activating concepts of human experience, which are far richer and more flexible than any lexical entry or formalization could possibly represent."*
> – Thora Tenbrink

*Corresponding author. The work was supported in part by ONR Award N00014-11-1-0417 and ONR STTR subcontract N00014-11-10474.

According to a long-standing line of research in cognitive science and neuroscience (with roots going back to Wilhelm Wundt and William James, or even Plato), human language understanding relies in part on the *ad hoc* creation of three-dimensional mental models, and mental images that correspond to visual projections of those models. For example, Johnson-Laird (1983) cites empirical evidence that human discourse understanding involves both symbolic representations and structural analogues of the world, where the latter become particularly prominent when the discourse provides a relatively determinate description of a configuration of objects. At the same time, even relatively determinate descriptions leave many details open, so that *"a mental model is in essence a representative sample from the set of possible models satisfying the description"* (*ibid.*, p. 165). As Johnson-Laird argues at length, the importance of such models lies in the (*nondeductive*) *inferences* they enable. While some cognitive scientists have argued against dual represention in favor of purely propositional ones (e.g., Anderson & Bower 1973), Kosslyn (1994) reviews the extensive evidence showing that *"parts of the brain used in visual perception are also used in visual mental imagery"*, and that visual cortex damage impairs not only vision but also visualization. He proceeds to propose, and marshall evidence (including PET scans) for, a general theory of how visual imagery is produced with the aid of both the visual cortex and the motor system, and the important functions it serves.

In AI, genuine language understanding is still thwarted by the knowledge acquisition (KA) bottleneck. Before we can overcome that formidable obstacle, whether by machine learning or other methods, we need at least to identify the kinds of knowledge representations required for understanding. While a great deal of AI research has addressed the question of what sorts of symbolic representations could support text comprehension, much less attention has been devoted to the potential role of mental imagery in that process, despite the insights from cognitive science noted above. This may be due in part to the frequently voiced idea that using internal three-dimensional models for comprehension would require an "inner eye" or Rylean humunculus. But this objection lacks cogency from an AI perspective, where the computational advantages of reasoning about the physical world with the aid of geometrical models and algorithms are well-established. (One of the earliest examples may be Scott Fahlman's work on reasoning about block stacking; see (Fahlman 1973).)

In the following, we illustrate how the need for imagistic representations arises in even the simplest first-reader stories. We then outline an architecture for integrated symbolic and imagistic modeling and inference, applying this to the previously introduced motivating examples. As evidence for the feasibility of such an architecture, we illustrate the functioning of two essential components on which we have been working: a broad-coverage semantic parser that produces normalized, nonindexical logical forms from English, and a preliminary imagistic modeling system (IMS) that allows construction of simple spatial scenes and evaluation of spatial predicates such as *'can-see'*, *'under'*, and *'inside'*. Much additional work will be required to bring this approach to fruition, but we contend that components of the type we have been constructing will be crucial to success in automating genuine language understanding.

Finally, we discuss related work on building imagistic models of text, and then reiterate our conclusions. The reported work does not in itself alleviate the KA bottleneck. On the contrary, we are arguing that the challenge is even larger than one might infer from most work on KA, which tends to treat the term as synonymous with acquisition of relational or rule-like knowledge. But in underscoring the need for extensive imagistic knowledge in story understanding, and outlining the possible form and use of such knowledge, we hope to be providing a better understanding of the challenges facing the natural language understanding community.

2 The Need for Imagistic Modeling

> *"The mechanisms that enable humans to tell, understand, and recombine stories separate human intelligence from that of other primates."* – Patrick Winston (2011)

The following is a simple story for beginning readers (from Lesson XXXI, Harris et al. 1889):

1. Oh, Rosy!
   Do you see that nest in the apple tree?
2. Yes, yes, Frank; I do see it.
3. Has the nest eggs in it, Frank?
4. I think it has, Rosy.
5. I will get into the tree.
6. Then I can peep into the nest.
7. Here I am, in the tree.
8. Now I can see the eggs in the nest.
9. Shall I get the nest for you, Rosy?
10. No, no, Frank! Do not get the nest.
11. Do not get it, I beg you.
12. Please let me get into the tree, too.
13. Well, Rosy, here is my hand.
14. Now! Up, up you go, into the tree.
15. Peep into the nest and see the eggs.
16. Oh, Frank! I see them!
17. The pretty, pretty little eggs!
18. Now, Frank, let us go.

One point where the need for imagistic modeling arises particularly clearly is at sentences 3 and 4. We know from the preceding two sentences that both Rosy and Frank see the nest. Yet it is clear from sentences 3 and 4 that they cannot see whether there are eggs in the nest – a fact needed to make sense of their subsequent dialogue and actions. In a purely symbolic approach, we might try to account for the visual occlusion of the eggs by means of an axiom stating that *to see the contents of a topless container, one must be near it, with one's head above it.* But there are problems with such an approach: First, the suggested axiom covers only a minor fraction of the unlimited number of ways in which occlusion can occur. To appreciate this point, consider the following (constructed) story fragments, where comprehension depends on a visual occlusion relation:

19. Rosy could not see the nest because of the thick foliage of the apple tree.
20. Jim hid from Frank behind a haystack.
21. With the curtains drawn, Ted did not notice the storm clouds gathering in the sky.
22. He finally found the "missing" glasses right under the newspaper in front of him.
23. Hoping to see the Milky Way, she stepped out of the house.
24. As he approached the mouse, it disappeared under the bed.

We noted that in order to see into a nest, the viewer should not only have a sufficiently high vantage point, but also be *near* the nest. This brings us to another problem with a purely symbolic approach: "Near" and "next to" relations are crucial to our understanding of many ordinary actions and situations, yet qualitative symbolic reasoning cannot in general tell us which of these relations hold in a specified situation. Again some simple examples suffice to indicate why understanding proximity relations is important in story understanding; the unnatural (b)-examples serve to draw attention to the proximity relations involved in the more natural (a)-examples:

25 a. Without sitting up in his bed, Tim closed and put aside the book he had been reading.
   b. #Without sitting up in his bed, Tim closed the door he had left open.

26 a. Sam heard the whir of a mountain bike behind him as he walked down the trail.
   He quickly moved aside, grumbling.
   b. Sam heard the whir of a helicopter behind him as he walked down the trail.
   #He quickly moved aside, grumbling.

27 a. Amy was walking her puppy. When it began to yelp, she quickly took it in her arms.
   b. Amy was flying her kite. #When it began to careen, she quickly took it in her arms.

28 a. Walking alongside the lioness, Joy Adamson stroked its head.
b. #Walking alongside the giraffe, Joy Adamson stroked its head.

Some of the examples also show that relative sizes and positions matter; clearly such examples can be multiplied indefintely.

It is worth noting that proximity problems also arise in cognitive robotics, where a robot may need to anticipate whether it will be next to an object (e.g., a person, food tray or book) it plans to interact with, after one or more moves or other actions. The "next-to" problem was analyzed in (Schubert 1990, 1994), and recent robotics research recognizes that cognitive robots need to construct three-dimensional spatial models of their environment (e.g., Roy et al. 2004).

As final examples, we quote two more brief stories from a similar source as our first story, in which the need for modeling spatial relations is quite apparent:

> A little girl went in search of flowers for her mother. It was early in the day, and the grass was wet. Sweet little birds were singing all around her. And what do you think she found besides flowers? A nest with young birds in it. While she was looking at them, she heard the mother bird chirp, as if she said, "Do not touch my children, little girl, for I love them dearly." (McGuffey 2005, Lesson XLII)
>
> This is a fine day. The sun shines bright. There is a good wind, and my kite flies high. I can just see it. The sun shines in my eyes; I will stand in the shade of this high fence. Why, here comes my dog! He was under the cart. Did you see him there? What a good time we have had! (McGuffey 2005, Lesson XXIX)

In the first story, note the contrast with our opening story: We infer that the girl's attention is generally turned downward, perhaps by reference to the prototypical gaze direction of someone seeking objects on the ground. So the nest she spots is likely to be situated close to or on the ground, within a field of view lower than the girl's head. This is confirmed by her ability to look at the young birds, as well as by the mother bird's perception that the girl might touch them. In the second story, the prototypical configuration of a person holding a high-flying kite at the end of a long string, and the position of the sun high in the sky, are essential to understanding why the sun shines in the kite-flyer's eyes. Further, how the shade of a high fence might shield the kite-flyer from the sun, and why a dog under a cart may or may not be noticeable, are best understood from (at least rough) physical models of these entities.

3 Combining Linguistic and Imagistic Processing

3.1 Hybrid architecture

Examples like those above indicate the need for an *imagistic modeling system* (IMS) for modeling spatial relations, supporting a general symbolic understanding and reasoning system but using techniques that exploit the very special nature of spatial relationships and interactions of complex objects. We have in mind the kind of hybridization strategy that has been successfully pursued in logic-based systems and programming languages that allow for support by taxonomic, temporal, or arithmetic constraint solvers (e.g., Frisch 1990), and more broadly in specialist-supported reasoners such as that described in (Schubert et al. 1987), or several of those in (Melis 1993).

Clearly the IMS will have to include spatial prototypes for a very large number of ordinary natural objects and artifacts, their typical poses, and their typical positioning in relation to other objects (e.g., fruits on fruit trees, trees standing on the ground, people standing, sitting, lying down, etc., birds flying, nesting, or on branch, tables and chairs in a room, etc.) Further, we need

means to assemble scenes *ad hoc* from such prototypes in accordance with verbal descriptions and symbolic knowledge, and crucially, means of "reading off" properties and relations from the assembled scenes. We consider the construction of such an IMS an important–and not very broadly appreciated–aspect of the knowledge acquisition bottleneck in AI.

Building a broad-coverage IMS will be a major challenge to the AI community, but we argue in the rest of the paper that it would indeed enable understanding of at least the simple kinds of stories we have used as examples. We do so by describing in some detail how semantic parsing, inference from background knowledge, and use of an IMS would interact in the processing of part of our opening story. The description is at least partially concrete, since we are well along in the construction a general semantic interpreter, have an inference engine that can reason with interpreted input and background knowledge, and have built a simple preliminary IMS. However, we do not yet have an integrated end-to-end story understanding system, or a significantly large set of prototypes.

3.2 Illustrating the process

We expect text understanding to proceed sentence-by-sentence, where the first step in sentence processing is semantic parsing. Our semantic parser can handle several of the simple stories we have looked at after some small adjustments. For example, We change Rosy's name to *Rosie* to prevent the Charniak parser from treating it as an adjective, and we change the question *"Has the nest eggs in it, Frank?"* to the American form *"Does the nest have eggs in it, Frank?"*. After some automatic postprocessing of the parse tree, for example to mark prepositional phrases with their type and to insert traces for dislocated constituents, our semantic interpreter successively produces (i) an initial logical form (LF) by compositional rules; (ii) an indexical LF in which quantifiers and connectives are fully scoped and intrasentential anaphors are resolved; (iii) a deindexed LF with a speech act predicate and with explicit, temporally modified episodic (event or situation) variables; and (iv) a set of canonicalized formulas derived from the previous stage by Skolemization, negation scope narrowing, equality substitutions, separation of top-level conjuncts, and other operations. The following are the parse trees and the set of *Episodic Logic* formulas (Schubert & Hwang 2000) derived automatically from sentence (1) of our lead story:

```
PARSE TREES:
‘‘‘‘‘‘‘‘‘‘‘
 (FRAG (INTJ (UH OH)) (, ,) (NP (NNP ROSIE)) (. !)),

 (SQ (AUX DO) (NP (PRP YOU))
  (VP (VB SEE) (NP (DT THAT) (NN NEST))
   (PP-IN (IN IN) (NP (DT THE)
                   (NN APPLE) (NN TREE)))) (. ?))

CANONICAL FORMULAS, WITH HEARER IDENTIFIED AS ROSIE:
‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘‘
 (SPEAKER DIRECT-OH-TOWARDS*.V ROSIE.NAME),

 (NEST8.SK NEST.N), (NEST8.SK NEW-SALIENT-ENTITY*.N),
 (TREE9.SK ((NN APPLE.N) TREE.N)),
 ((SPEAKER ASK.V ROSIE.NAME
   (QNOM (YNQ (ROSIE.NAME SEE.V NEST8.SK)))) ** E7.SK),
 ((SPEAKER ASK.V ROSIE.NAME
   (QNOM (YNQ (ROSIE.NAME SEE.V
    (THAT (NEST8.SK IN.P TREE9.SK)))))) * E7.SK)
```

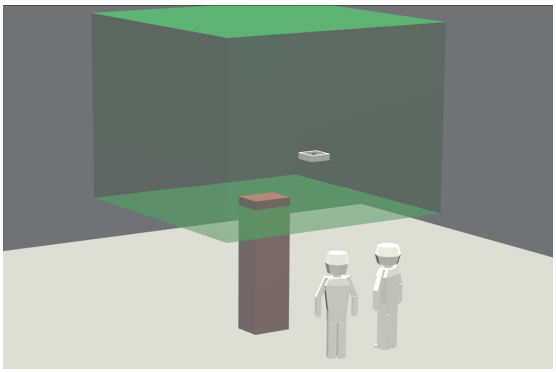

Figure 1: Scene created by IMS, allowing evaluation of query, "Can the children see eggs in the nest?"

Episodic Logic formulas, as used by the EPILOG inference engine, are in infix form, i.e., the predicate follows the first (subject) argument. Also the unnamed objects referred to have been given Skolem names, such as TREE9.SK, and the corresponding type-predications, such as (TREE9.SK ((NN APPLE.N) TREE.N)), have been separated out. "Seeing the nest in the tree" has been approximated with the conjunction of "seeing the nest" and "seeing *that* the nest is in the tree"; correspondingly the speaker's question has been broken into two parts. YNQ is a yes-no question-forming operator (with intensional semantics) and QNOM is a reification operator mapping a question intension to an abstract individual. (Together, (QNOM (YNQ ...)) are equivalent to (WHETHER ...).) The '**' and '*' operators respectively relate a sentence to a situation it characterizes, and a situation that it partially describes (see Schubert 2000).

We could now use EPILOG to infer that the speaker's question presupposes that the speaker sees the nest, and sees that it is in the tree, i.e., NEST8.SK is in fact in TREE9.SK, and hence it is (presumably) in the crown of TREE9.SK. The knowledge needed for the presupposition inference can be formally stated using the metasyntactic capabilities of EPILOG; and the inference that the nest is in the crown of the tree would be based on the stereotyped knowledge (at first in symbolic form) that a nest in a tree is normally in the crown of the tree.

Additional stereotyped knowledge is needed before imagistic modeling can begin to contribute to understanding: Trees are generally outdoors and standing upright, rooted in the ground; persons who can see a small outdoor object (such as a nest) are typically outdoors as well and near the object; further, they are usually upright on the ground, unless we have reason to believe otherwise.

At this point, we can begin to generate a model scene: Two children, one called Rosie, are outdoors on the ground, near an apple tree with a nest in its crown, and they can see the nest. (That the persons are children is a meta-inference from the genre of the book the story is taken from.) Our preliminary IMS can accept these assertions (stated as formal predications) and build a corresponding scene. The object prototypes ensure that the size relationships will be reasonably realistic.

Processing of sentences 3 and 4 leads to the inference that Rosie and Frank do not know for certain whether there are eggs in the nest. Technically, Rosie's not knowing this is an implicature of her question, (3), and Frank's not knowing this is an implicature of his phrasing *I think ...* in (4).

We assume that when a story raises the *possibility* that a certain object is in the current scene, the IMS will model the presence of that object (but flagging it as hypothetical). Thus at least one (hypothetical) egg will be placed in the nest. We also assume that the interpretive algorithm continually checks the plausibility of propositions newly inferred from the text – in this case, the plausibility of the inference that Rosie and Frank do not know for certain whether there are eggs in the nest. Using the knowledge that if a person *can see* an object at a certain place, then they *know* it is there (equivalently, if they don't know about the presence of an object, they can't see it), the algorithm would query the IMS about the visibility of the egg by the children. Evaluation of the "can-see" predicate would confirm that the children probably cannot see the (hypothetical) egg in the nest, confirming the plausibility of the inference that motivated the check. In this way, *coherence* of the current interpretation of the story is maintained, and that is an essential aspect of deep understanding.

While our implementation has not gone beyond this point, further symbolic story processing at sentence (5) would lead to the inference that Frank will end up in the crown of the apple tree if he climbs it, and by modeling this future situation we could confirm his expectation expressed by sentence (6), again using "can-see". Sentence (7) verifies the (tentative) inference that Frank carried out the intention expressed in sentence (5), and sentence (8) confirms both Frank's expectation and the model's prediction. We hope it is clear that sentences (12-16) can be tackled in a similar way.

4 Imagistic Modeling System

Our modeling environment used the Blender system (www.Blender.org), a free Gnu software system for designing and rendering complex 3D scenes. It provides a means of easily creating object models, attaching properties and constraints to those models, and performing operations in a 3D space. Blender is primarily written in python and supports native development of python scripts using its API. Python modules executed in the Blender scripting environment may access and manipulate object information in real time.

Based on input from our inference engine, the specialist constructs a 3D model for given entities and relations derived from a portion of a story describing some situation. This model may then be queried by the inference engine for spatial relations in the scene, such as 'Can-See', 'Near', 'Under', 'In', and 'Within'. Both placement and querying rely primarily on computed *acceptance areas* (AAs – in general 3D regions) in the vicinity of objects. Entities are obtained as instances of object prototypes in a database, stored in the form of 3D Blender files and XML trees. Each object is comprised of a series of part models arranged relative to an empty parent object. Default poses and configurations are also stored for objects, and these are invoked with a certain probability given that instances of all objects in a configuration are in the scene. For example, a scene with a set of chairs and a table might be automatically arranged such that the chairs encircle the table.

We use *Scene*, *Entity*, and *Predication* classes to represent information about story scenes. Each *Scene* models an individual story scene, and contains pointers to the *Entity* and *Predication* objects represented in the scene model. Entity instances contain a pointer to the Blender object representing the entity as well as methods for drawing and manipulating the object, and pointers to all predications relevant to the entity's parameters. A Predication is an instantiated spatial predicate applied to particular argument entities and includes methods imported from the predicate library. Each predicate in the predicate library has two distinct functions, one for object placement and the other for querying. Both are based on associated functions that compute AAs for parameterized 3D objects, drawing on a sublibrary of primitive spatial

operations.

For object placement, an AA is represented as a tuple of maximum and minimum values for the x, y, and z dimensions. For example, suppose that an object A whose maximum dimension is 1.0 bu (Blender units) is specified as being near a second object B whose maximum dimension is 2.0 bu, and object B is located at (2, 2, 0). A reasonable AA for placing A based on the predication [A near B] might then be: ((-1,5), (-1,5), (-3,3)). This area would be randomly sampled to determine A's modeled location.

Each query function returns a numerical value ranging from 0 to 1 depending on how well the predication is satisfied. Widely different approaches are used to query predications depending on which predicate is instantiated. 'Near', for example, performs a simple calculation depending on object sizes and distance from one another. 'Under' generates a temporary block representing its AA and returns a value depending on what proportion of the second object's volume intersects this block. A rejection technique based on predicate querying is also used to aid in object placement. After AA bounds are determined and an object is placed, predications affected by its parameters are queried. If any of these returns a value of 0, the object is placed again.

So far, we have built and tested 39 object types and 8 predicates. This is not a large number, but is enough to convince us that with many more objects (with multiple "poses" for deformable ones, and arranged in various typical configurations with other objects), and a few more predicates, we would be able to model a wide range of scenes in realistic first-reader stories.

5 Related Work

Existing research in qualitative spatial reasoning emphasizes the utility of acceptance areas and qualitative variance in predicate meanings as a function of object shape and size (Hernández 1994, Hernández et al. 1995, Clementini et al. 1997, Cohn 1997, Mukerjee et al. 2000, Cohn & Hararika 2001, Tappan 2004). Kadir et al. (2011) use Blender as their basis for modeling scenes from natural language input, utilizing similar methods to object placement as ours. WordsEye is a powerful text-to-scene conversion system capable of modeling complex visual scenes based on layers of spatial predications (Coyne & Sproat 2001, Coyne et al. 2010a, Coyne et al. 2010b). However, existing text-to-scene conversion systems such as this differ crucially from ours in that they are designed only for predicate-based scene construction and not for querying.

A Master's thesis by J. R. Bender (2001) connects simple sentences, interpreted in terms of Jackendoffian primitives, with rough 3D prism-based models; it is able to answer questions such as whether the butter on a toast is *above* the plate, where the toast has been stated to be on the plate. This work partially anticipates ours, though it did not tackle real stories, and we think that the kinds of visual predicates considered could rather easily be handled symbolically (e.g., with transitivity axioms), unlike those we implemented, such as *can see*. In related work, Finlayson and Winston (2007) showed how a game engine could be used to model a motion trajectory in a sentence about a pigeon, and to extract a part of the trajectory not explitly described. They mention that many common objects are modeled by game engines, an observation we also made at the beginning of our work; but adding the kinds of evaluable predicates to existing game engines that are needed for story understanding turned out to be far more difficult than using the Blender API to develop our own script-based modeling system.

Soar's Spatial and Visual System (SVS) (Wintermute 2009) is conceptually similar to the IMS. Predicate-specified AAs are used to create an imagistic scene representation, which can be then used to query further predicates. SVS differs from our system in that it is built to model only very simple scenes and is not designed to handle linguistic input (Wintermute 2010).

6 Conclusion

The very simple, but genuine, first-reader stories we looked at illustrated the need for spatial knowledge about humans, ordinary objects, and their typical poses and configurations in story understanding. We showed with a detailed example how such knowledge could be used to support construction of a coherent interpretation by a symbolic interpretive algorithm. The spatial models for this purpose need not be as fine-grained as those intended for creating realistic, pleasing 3D scenes. But the number of ordinary object types is very large, and their typical kinds of configurations even larger. Thus developing a large imagistic knowledge base should be considered an important challenge by the NLU community. Indeed, the more or less static modeling we have focused on needs to be supplemented eventually with subtle dynamic modeling, as is evident if one considers questions such as *If you throw a cream pie in someone's face, what happens to the pie and the face?*, or popular quips about activities that any "truly wise man" should avoid, such as playing leapfrog with a unicorn, standing barefoot behind Grandpa's rocking chair, or skipping rope under a ceiling fan.[1] Johnston & Williams' (2008) "molecular" models, which they integrate with tableaux-based reasoning, are interesting candidates for dynamic simulation, if they can be effectively used for construction of linguistically described situations, and evaluation of linguistically relevant predicates in these situations.

References

[Anderson and Bower(1973)] J. R. Anderson and G. H. Bower. 1973. Human associative memory. Winston, Washington, DC.

[Bender(2001)] J. R. Bender. 2001. Connecting Language and Vision Using a Conceptual Semantics. MS thesis, Dept. of Electrical Engineering and Computer Science, MIT, Cambridge, MA.

[Clementini et al.(1997)] E. Clementini, P. Di Felicea, and D. Hernández. 1997. Qualitative representation of positional information. *Artificial Intelligence*, 95(2):317–356.

[Cohn(1997)] A. G. Cohn. 1997. Qualitative spatial representation and reasoning techniques. *KI-97: Advances in Artificial Intelligence*, 1–30. Springer, Berlin.

[Cohn and Hazarika(2001)] A. G. Cohn and S. M. Hazarika. 2001. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1):1–29.

[Coyne and Sproat(2001)] B. Coyne and R. Sproat. 2001. WordsEye: An Automatic Text-to-Scene Conversion System. *Proc. of SIGGRAPH 2001.*

[Coyne et al.(2010a)] B. Coyne, R. Sproat, and J. Hirschberg. 2010. Spatial relations in text-to-scene conversion. *Computational Models of Spatial Language Interpretation, Workshop at Spatial Cognition.* Mt. Hood, OR.

[Coyne et al.(2010b)] B. Coyne, O. Rambow, J. Hirschberg, and R. Sproat. 2010. Frame semantics in text-to-scene generation. *Knowledge-Based and Intelligent Information and Engineering Systems*, 375–384. Springer, Berlin, Heidelberg.

[Fahlman(1973)] S. Fahlman. 1973. A Planning System for Robot Construction Tasks. MS Thesis, MIT AI Lab, June 1973, Cambridge, MA.

[Finlayson and Winston(2007)] M. A. Finlayson, P. H. Winston. 2007. Reasoning by imagining: The Neo-Bridge system. CSAIL Research Abstracts, MIT, Cambridge, MA.

[Frisch(1990)] A. Frisch. 1990. The Substitutional Framework for sorted deduction: Fundamental results on hybrid reasoning. *Artificial Intelligence*, 49: 126–136.

[1] As a more serious example, Winston (2011) mentions a friend's admonition never to wear gloves while using a table saw – imagine the glove being caught in the blade!

[Harris et al.(1889)] W. T. Harris, A. J. Rickoff, and M. Bailey. 1889. The First Reader. D. D. Merrill, St. Paul, MN.

[Hernández(1994)] D. Hernández. 1994. *Qualitative representation of spatial knowledge*, volume 804.

[Hernández et al.(1995)] D. Hernández, E. Clementini, and P. Di Felicea. 1995. *Qualitative distances*, Springer, Berlin, 45–57.

[Johnson-Laird(1983)] P. N. Johnson-Laird. 1983. *Mental models: Towards a cognitive science of language, inference, and consciousness.* Harvard Univ. Press.

[Johnston & Williams(2008)] B. Johnston and M.-A. Williams. 2008. Comirit: Commonsense reasoning by integrating simulation and logic. *Artificial General Intelligence (AGI 2008)*, 200–211.

[Kadir et al.(2011)] R. A. Kadir, A. R. Mad Hashim, R. Wirza, and A. Mustapha. 2011. 3D Visualization of simple natural language statement using semantic description. *Visual Informatics: Sustaining Research and Innovations*, 28(1):114–133. Springer, Berlin.

[Kosslyn(1994)] S. M. Kosslyn. 1994. *Image and Brain: The Resolution of the Imagery Debate.* MIT Press, Cambridge, MA.

[McGuffey(2005)] W. H. McGuffey. 2005 (original edition 1879). *McGuffey's First Eclectic Reader* (revised edition) (EBook #14640). John Wiley and Sons, New York.

[Melis(1993)] E. Melis. 1993. Working Notes of the 1993 IJCAI Workshop on Principles of Hybrid Representation and Reasoning. *Int. Joint Conf. on Artificial Intelligence.* Chambéry, France.

[Mukerjee et al.(2000)] A. Mukerjee, K. Gupta, S. Nautiyal, M. P. Singh, and N. Mishra. 2000. Conceptual description of visual scenes from linguistic models. *Image and Vision Computing*, 18(2):173-187.

[Roy et al.(2004)] D. Roy, K.-Y. Hsiao, and N. Mavridis. 2004. Mental Imagery for a Conversational Robot. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 34(3):1374–1383.

[Schubert(1990)] L. K. Schubert. 1990. Monotonic solution of the frame problem in the situation calculus: An efficient method for worlds with fully specified actions. In H. Kyburg, R. Loui and G. Carlson (eds.), *Knowledge Representation and Defeasible Reasoning*, Kluwer, Dortrecht, 23–67.

[Schubert(1994)] L. K. Schubert. 1994. Explanation closure, action closure, and the Sandewall test suite for reasoning about change. *J. of Logic and Computation* 4(5): 679–799.

[Schubert(2000)] L. K. Schubert. 2000. The situations we talk about. In J. Minker (ed.), *Logic-Based Artificial Intelligence*, Kluwer, Dortrecht, 407–439.

[Schubert & Hwang(2000)] L. K. Schubert and C. H. Hwang. 2000. Episodic Logic meets Little Red Riding Hood: A comprehensive, natural representation for language understanding. In L. Iwanska and S.C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation.* MIT/AAAI Press, Menlo Park, CA, 111–174.

[Schubert et al.(1987)] L. K. Schubert, M.-A. Papalaskaris, and J. Taugher. 1987. Accelerating deductive inference: Special methods for taxonomies, colours, and times". In N. Cercone and G. McCalla (eds.), *The Knowledge Frontier: Essays in the Representation of Knowledge*, Springer, 187–220.

[Tappan(2004)] D. Tappan. 2004. Monte Carlo simulation for plausible interpretation of natural-language spatial descriptions. *Proc. of the 17th Int. Florida Artificial Intelligence Research Society Conf.*, Miami Beach, FL.

[Winston(2011)] P. H. Winston. 2011. The Strong Story Hypothesis and the Directed Perception Hypothesis. *AAAI Fall Symposium: Advances in Cognitive Systems*, Nov. 4–6, Arlington, VA.

[Wintermute(2009)] S. Wintermute. 2009. An Overview of Spatial Processing in Soar/SVS. *Report CCA-TR-2009-01*, U. Michigan Center for Cognitive Architecture.

[Wintermute(2010)] S. Wintermute. 2010. Using imagery to simplify perceptual abstraction in reinforcement learning agents. *Proc. of AAAI-10*, July 11–15, Atlanta, GA, 1567–1573.