Knowledge-Based Constraint Satisfaction for Spatial Reasoning

Dan Tappan

New Mexico State University
Department of Computer Science
P.O. Box 30001, MSC CS
Las Cruces, NM 88003
dtappan@crl.nmsu.edu

Abstract

This linguistically motivated work addresses issues in reasoning intelligently over spatial descriptions of simple, static scenes to produce plausible graphical interpretations. It uses a combined representation that couples a semantic network for explicit knowledge with a knowledge base of frames for implicit knowledge. The knowledge base contains generalized rules for interpreting what objects are and how they should and should not be interpreted alone and in spatial interrelationships. The linguistic emphasis is on the semantics and pragmatics of underspecification, vagueness, uncertainty, and context in reasoning over spatial language and knowledge.

Introduction

Given a simple English description of a real-world scene, for instance, *a dog is in front of a cat and near a tree*, anyone can easily formulate a corresponding mental image or model. The description itself explicitly contributes only a tiny fraction of the details that such an image contains. In fact, most of the details come from an implicit, commonsense understanding of the objects in the scene and how they can and cannot be realistically depicted in three-dimensional space (among other things).

Limited spatial reasoning of this type is the goal of this work, which uses a simple representation of a description in conjunction with a relatively simple knowledge base of relevant details to define the form of a valid solution. From this form, a basic constraint satisfaction algorithm generates any number of corresponding interpretations with plausible positions and orientations for the objects. Such solutions can directly support many applications that use or could benefit from natural language like text understanding, machine translation, question-and-answer systems, query interfaces to databases, search engines, user-friendly tools for graphics and animation, and so on (Srihari 1994).

Spatial reasoning, like most intelligent processes, is a difficult computational task to emulate despite its apparent simplicity and straightforward nature. As Herskovits (1986) concludes, "[a] computational treatment … will require much greater sophistication than naive representation theory would lead us to expect." What makes the problem especially troublesome is that computers lack the vast storehouse of intricate knowledge that humans possess and the amazing abilities to reason intelligently over it. This work addresses the primary, relevant aspects of these issues in terms of what to represent, how to represent it, and when and how to use it.

Background

The knowledge representation for the explicit and implicit details of a description in this linguistically motivated work addresses shared issues in language and the spatial world. Language plays a key role in such research because it closely reflects human perception and understanding of the spatial world (Johnson-Laird 1983, Langacker 1987). In particular, four major issues are the focus. First, underspecification, or the lack of complete details in a description, requires background or so-called world knowledge to fill in the gaps in its interpretation. Second, vagueness, or the imprecise nature of descriptions, requires knowledge that defines a range of possible interpretations. Third, uncertainty, or the lack of commitment to a particular interpretation, requires knowledge of tendencies or preferences over this range. And fourth, context, or the different interpretation of objects in certain combinations with each other, requires knowledge to identify such patterns and define the differences.

These linguistic issues map to the primary spatial issue of interest: the valid and preferred spatial behaviors of the objects in a description, specifically the interpretation of their positions and orientations with respect to three contextually determined frames of reference (Herskovits 1986, Claus et al. 1988, Olivier and Tsujii 1994). The intrinsic (or object-centered) frame generally applies to objects that have a canonical front; e.g., *in front of the dog* means some position in line outward from its face. The extrinsic (or environment-centered) frame and the deictic (or viewer-centered) frame are generally the opposite case for objects without a canonical front; e.g., *in front of the tree* means in line outward from it to another position in the world that establishes a virtual front. In the extrinsic frame, this reference position is arbitrary; e.g., *in front of the tree as seen from the lake.* In the deictic frame, which is a specialized case of the extrinsic frame, it is the (usually implicit) position of the viewer; e.g., *in front of the tree (as*

seen by the viewer in the north looking south). For space reasons, this paper discusses only the implicit and deictic frames.

Table 1 shows the 25 relations for the spatial behaviors of interest. Each is of the binary form *xRy*, where *x* and *y* are objects and *R* is a relation of position, distance, or orientation. Most static, spatial prepositions in English fall into these classes (Freeman 1975, Bennet 1975, Herskovits 1986, Hill 1982, Talmy 1983, and Hawkins 1984). This research addresses through the same formalisms an additional 19 in several other classes that are beyond the scope of this paper.

Class	Relations

Position	`in-front-of` `in-front-left-of`
	`in-back-of` `in-front-right-of`
	`left-of` `in-back-left-of`
	`right-of` `in-back-right-of`
	`north-of` `northeast-of`
	`south-of` `northwest-of`
	`east-of` `southeast-of`
	`west-of` `southwest-of`

Distance	`inside`
	`outside`
	`adjacent-to`
	`near`
	`midrange-from`
	`far-from`
	`at-fringe-of`

Orientation	`facing` `facing-away-from`

Table 1: Spatial Relations

The underspecified, vague, uncertain nature of typical descriptions lacks the preciseness that a quantitative or absolute, numerical approach to spatial reasoning would require (Kuipers 1978); e.g., *the cat is 3.0 meters bearing 45.0 degrees from the dog that is located at world coordinate 20,10*. This work, like most linguistically motivated work, adopts a qualitative approach that reasons in terms of more natural, relative constraints (Mukerjee 1998); e.g., *the cat is northeast of and near the dog.* Specifically, it employs a geometric approach to intersect two-dimensional regions that are similar to Venn diagrams. The end-to-end processing of a description decomposes it into its components to determine the contextually appropriate, individual geometric constraints that, together, declaratively specify to the spatial reasoning engine the form of valid and preferred interpretations to generate. These descriptions--in fact, generally most descriptions--do not require the significantly more complex expressiveness of true three-dimensional reasoning (Xu, Stewart, and Fiume 2002).

Despite the potential of such research, very few contemporary systems exist (Wahlster 1996). CarSim (Dupuy et al. 2001) focuses on graphically rendering the results of vehicle collisions based on accident reports. WordsEye (Coyne and Sproat 2001), the closest to this work, focuses on depicting appropriate static poses for actions. Although both address text understanding and employ various degrees of knowledge representation, they focus more on producing the graphical results and less on investigating the underlying linguistic and knowledge issues. In fact, most systems that do spatial layout take a purely geometric approach and do not rely on knowledge at all (Xu, Stewart, and Fiume 2002, Yamada 1993).

Knowledge Representation

A description in this work consists of nouns, adjectives, prepositions, and various glue words like determiners, conjunctions, and the copular verb *is*. This paper does not address the adjectives, which play a related spatial role in the contextually appropriate determination of size. The nouns must refer to concrete, physical objects that are customarily present within the scenario of a zoo. Aside from the obvious visual appeal, animals and plants exhibit a variety of interesting spatial behaviors across their shape, size, capabilities, etc. The prepositions are the relations in Table 1 with determiners and conjunctions for readability and without the hyphens; e.g., *in front and left of* and *at the fringe of.*

As in most related systems (except Dupuy et al. 2001), descriptions are manually fabricated rather than acquired from existing sources to eliminate troublesome issues in parsing that are outside the scope of investigation. They must also refer to static scenes only, which is a common limitation due to the complexities of verb interpretation, movement, time dependencies, the frame problem, etc. (Adorni, Di Manzo, and Giunchiglia 1984, Sowa 1991, Srihari 1994, Coyne and Sproat 2001).

Explicit Knowledge

The representation of the explicit knowledge in a description uses a straightforward semantic network of object nodes, attribute nodes, and relation arcs, which map closely to its nouns, adjectives, and prepositions, respectively (Sowa 1991). Each object node refers to a single object in the description. Each directed arc specifies a binary relation that refers to both a constraint and a context from its source object to its target object. Figure 1 depicts the semantic network for the example to carry throughout the remainder of this paper: *the rabbit is in front of, near, and facing the giraffe.*

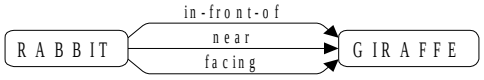

Figure 1: Semantic Network

Implicit Knowledge

The explicit knowledge in the semantic network supplies only the syntactic framework for its interpretation. Nothing in it defines the context-independent semantics of what the rabbit and the giraffe are or the context-dependent pragmatics of what it means for one to be in front of the

other, and so on. For humans, this implicit knowledge comes from an acquired understanding of the world.

The knowledge base is an inheritance hierarchy of frames that provide background details on how to interpret the object nodes and relation arcs in various contexts. This formalism is advantageous because it independently encapsulates declarative, prototypical definitions or concepts for each object that may appear in a description (Sowa 1991).

An inheritance hierarchy is analogous to a taxonomy of the plant and animal kingdom, which organizes its concepts according to similar morphology and physiology. Each concept shares the definitions of all its ancestors but extends or refines them somehow to become a more specific subconcept; e.g., a dog is a canine, which is a carnivore, which is a mammal, which is a vertebrate, and so on. As this work uses various animals for representative concepts, a natural organization for the knowledge base mirrors this real-world taxonomy.[1] It also deflects a common criticism of knowledge-based systems that ad hoc structures have no realistic connection with the real world (Mahesh 1996).

The underlying zoological hierarchy does not take into account spatial behavior, which has a significantly different and remarkably simpler organization. For example, horses, zebras, camels, and llamas all belong to different branches of this hierarchy, but they actually share the same spatial behaviors in terms of their general size, shape, presence of a canonical front, and so on. To capitalize on this observation, a separate hierarchy maps 17 spatial concepts onto the 79 animals currently in the zoological hierarchy. Figure 2 illustrates the notion of this shared hierarchy. The solid nodes define animals, and the dashed nodes define spatial behaviors.

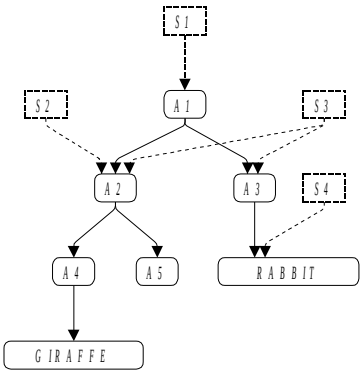

Figure 2: Dual-Hierarchy Knowledge Base

Concepts may have more than one parent from each hierarchy; e.g., *A2* inherits from *A1* as well as from *S2* and *S3*. This multiple inheritance further simplifies the organization through unrestricted sharing and combining of concepts (Mahesh 1996). Such flexibility does introduce the possibility of conflicts between contradictory or incompatible definitions, but the shallowness of the spatial hierarchy and the relatively disjoint nature of its concepts seem to mitigate this problem.

1 It omits many irrelevant distinctions, however; e.g., phylum, class, order, etc. Also, its concepts define no zoological information.

Combined Representation

The knowledge base addresses the problem of underspecification by augmenting the explicit syntactic knowledge of the semantic network with implicit semantic knowledge. The mechanism is simple: each object node in the semantic network links to its corresponding concept node in the knowledge base as Figure 3 demonstrates.

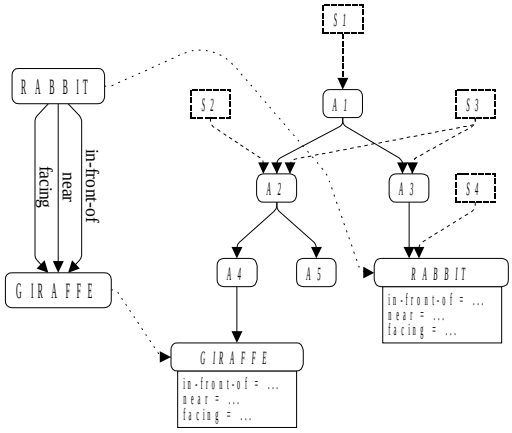

Figure 3: Semantic Network Linked to Knowledge Base

This combined representation is useful because each object node has access to the implicit spatial definitions of its concept node, which includes all the inherited definitions as well. The basis of the linked representation in this work derives from previous work on the Mikrokosmos project for knowledge-based machine translation (Mahesh 1996). It used a similar structure as a rich, interconnected text meaning representation for semantic and pragmatic analysis.

Contextual Interpretation

The knowledge base addresses the problem of context by conditionally applying definitions for default and non-default interpretations. A default interpretation occurs when an object node is either not part of a relationship or no other objects in any of its relationships affect its prototypical, spatial behavior. For example, *there is a hippo* instantiates a particular `hippo` that has no justification to differ from a standard, "generic" `hippo`. Similarly, *the hippo is in the zoo* states an inert relationship that generally imparts no different interpretation on this `hippo` than it would on any other. In other words, a default interpretation is independent of context and reflects the semantics of a concept.

A non-default interpretation is the complementary case. For example, *the hippo is in the corral* implies that its body is on the surface of world; whereas *the hippo is in the lake* implies that it is below the surface. The appropriate

vertical interpretation is critical and certainly not arbitrary or interchangeable for a `hippo`! On the other hand, either is acceptable for *the duck is in the lake.* In other words, a non-default interpretation is dependent on context and reflects the pragmatics of a concept.

This work employs two mechanisms to identify such contextual patterns for any concept. The first is by association, which triggers on specific target concepts in a relationship. The specification can be extensional by exhaustively listing all the concepts that have the same spatial effect on the source concept; e.g., `lake`, `pond`, and `pool`. It can also be intensional by indicating the branch of the hierarchy that subsumes the individual concepts; e.g., `body-of-water`. This form eliminates the need to enumerate all concepts that are equivalent in a certain respect. It also simplifies maintenance and expansion of the knowledge base because the list does not require updating if new, equivalent concepts are added to the knowledge base; e.g., `river`, `stream`.

The second mechanism is by conditional dependency, which triggers on specific properties inside the definitions of other concepts in a relationship. The frame-based formalism of definitions uses a traditional slot-filler structure to associate values with properties arbitrarily (Sowa 1991). The most common is the boolean `has-canonical-front`, which helps resolve issues with frame of reference. For example, the interpretation of *x in front of y* depends on whether *y* indicates that it has a canonical front.

Spatial Reasoning by Constraint Satisfaction

The goal of contextual interpretation over the combined representation of explicit and implicit knowledge in a description is to build a collection of spatial constraints that limit the valid, contextually appropriate positions and orientations for the graphical rendering of the objects. Satisfying these constraints thus inherently produces a valid interpretation, which is a common approach for spatial reasoning (Mukerjee 1998).

Field Constraints

The unified formalism for constraints in this work is a top-view, polar projection of two-dimensional, geometric fields that surround every object. It is similar in form to potential fields in other work (Yamada et al. 1992, Yamada 1993, Gapp 1994, Olivier and Tsujii 1994). It differs primarily in the two complementary levels of its definition and in its contextual interpretation with concepts and relations.

The first half of a field definition addresses the geometry, which constrains where others object must appear with respect to the relation that uses it. For example, Figure 4a shows a geometry for a `front` field, to which the position relation `in-front-of` and orientation relation `facing` typically bind. Similarly, Figure 4b shows a `near` field for the distance relation `near`. Although any combination of cells on the projection is available, this work finds that all 44 of the relations it currently defines bind to minor variations on wedges and rings only.

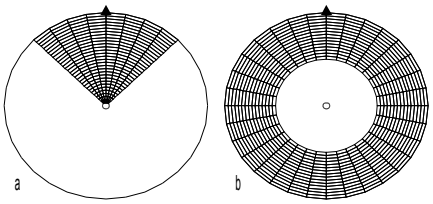

Figure 4: Geometry of Wedge and Ring Fields

The second half of a field definition addresses the topography, which overlays a probability distribution onto the geometry. Figure 5 shows which positions in Figure 4 are more consistent with an interpretation. This level reflects the "scruffy" nature of spatial relations due to vagueness and uncertainty: positions in the center of perceptual focus are more probable than those at the periphery (Mukerjee 1998, Johnson-Laird 1983). In terms of spatial inferences, the geometry of a field sanctions the positions that are legal, and the topography recommends a subset that are contextually preferred (Davis, Shrobe, and Szolovits 1993).

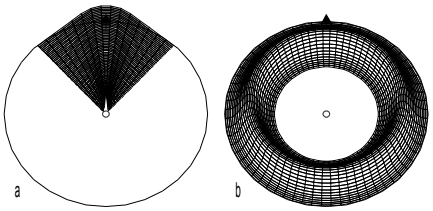

Figure 5: Topography on Wedge and Ring Fields

The relations in most descriptions interact to constrain the interpretation of objects further (Herskovits 1986). Fields easily accommodate such compositional behavior through simultaneous intersection over the geometry and topography. Figure 6 illustrates for the `front` and `near` fields the intersection that corresponds to the prepositional phrase *in front of and near.*

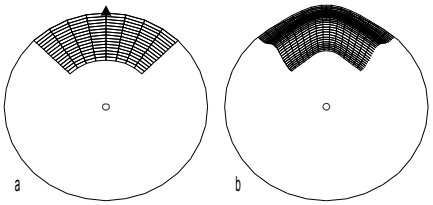

Figure 6: Intersection of `front` and `near` Fields

Two important factors play a role in the contextual application of fields. The first is frame of reference. For any concept with a canonical front, the default frame of reference is intrinsic; i.e., anything in front of it is in line with the direction it is facing. A field reflects this spatial

behavior by rotating itself so that its arrow aligns with the orientation of the object at its center (or vice versa). Thus, its `front` field aligns with this direction, and its `back`, `left`, and `right` fields respectively align 180 degrees, 90 degrees counterclockwise, and 90 degrees clockwise from it. On the other hand, for any concept without a canonical front, the default reference frame is deictic; i.e., anything in front of it is in line between itself and the viewer. In this case, the arrow aligns to the position of the viewer. Finally, all concepts support compass directions for relations like `north-of`, `south-of`, etc. In this case, the arrow always aligns to north, or top dead center.

The second factor is scale. It receives only brief mention here because it relies on the size of objects, which this paper does not address. In general, the contextual interpretation of distance depends on the size of the reference object at the center of a field (Olivier and Tsujii 1994). In other words, what is *near* for a giraffe is *far* for a rabbit. The diameter of fields reflects this behavior as Figure 7 shows.

Constraint Satisfaction

The constraint propagator finds valid solutions for the various spatial behaviors that fields define. It does so by calculating discrete values for the position and orientation of every object in the semantic network such that all their values simultaneously satisfy all their field constraints. This process uses randomization over the probability distributions in the field topographies. As such, it does not produce the same result for multiple runs over the same description. This nondeterministic behavior addresses uncertainty because there are an infinite number of valid interpretations for any description (Srihari 1994). The geographies guarantee that any solution is valid, and the topographies attempt to bias them toward more likely (or less controversial) interpretations.

The constraint propagator is a modular component of this work. Its implementation is by no means the most appropriate or efficient, and a better version could replace it easily. The justification for this rather brute-force approach is two-fold. First, this work "intelligently" addresses all its stated issues of spatial reasoning in the preceding stages; now it just mechanically fills in the blanks, so to speak. Second, most descriptions are relatively simple in their number of objects and relations because the human mind has limitations (Johnson-Laird 1983).

The objects in a description form a semantic network according to their relation arcs. This network is inherently a dependency graph that defines how the objects constrain each other. Objects that are neither directly nor indirectly interconnected form separate semantic networks that cannot interfere with each other;[2] e.g., *the dog is near the cat and the giraffe is facing the lake*. The constraint propagator can solve these constraints independently. Thus, at the top-level, it employs a divide-and-conquer strategy over one or more disjoint semantic networks.

The next level involves a greedy strategy to solve the constraints for all objects in the current network. It recursively processes every pair of objects that form a relationship. It employs the following (oversimplified) heuristics based on whether their position and orientation are set:

1. If neither object is set, solve the one with the most constraints first, then the other.
2. If one is set, solve the other.
3. If both are set and satisfy all constraints between them, then they are done.
4. If both are set and violate a constraint between them, unset them and start over at an earlier pair.

The restart mechanism uses backtracking. In Rule 4, it would re-solve the previous pair first then return to the pair that failed. If this pair failed again, it would repeat this process on it an arbitrary number of times before abandoning it for the previous pair of its previous pair, and so on.

Discussion and Summary

Figure 7 shows a representative solution for *the rabbit (R) is in front of, near, and facing the giraffe (G)*. This structure feeds directly into the graphical rendering engine that produces a corresponding three-dimensional virtual world (minus the projection details).

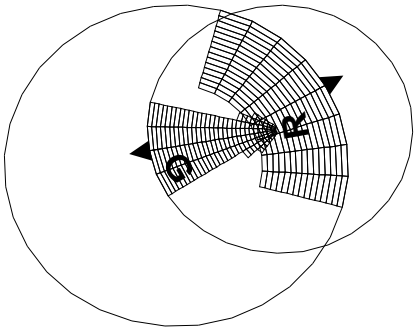

Figure 7: Rabbit and Giraffe

The geometries of the field constraints ensure that a solution is always correct, so there is no question about the effectiveness of this approach to spatial reasoning over this restricted domain of animals. The quality of the results in terms of their consistency with the real world is high as well. The true value of this work, however, is not in its results but in the mechanisms that contribute to them. This approach provides a powerful and flexible framework that successfully addresses the stated issues of interest:

- Underspecification: the knowledge base provides an expressive structure to represent varied knowledge of objects and relations.

2 Except if objects violate the global noninterpenetration constraint by invalidly embedding in each other. It is valid to embed in a container object like a corral but not in a giraffe!

- Context: the conditional manipulation of the knowledge in the knowledge base applies appropriate definitions for default and non-default interpretations.
- Vagueness: the unified formalism of fields accommodates a wide range of valid interpretations.
- Uncertainty: the nondeterministic nature of the constraint propagator uses the probabilistic definitions of fields to favor certain interpretations.

The future direction of this work will address these four issues in terms of scalability and extensibility. For the former, it will add more concepts and relations of the same or similar classes to improve the overall coverage within this domain. For the latter, it will add new concepts and relations from other domains to investigate how well this approach can solve issues in other areas of spatial reasoning.

References

Adorni, G.; Di Manzo, M.; and Giunchiglia, F. 1984. Natural Language Driven Image Generation. In Proceedings of COLING-84, 495-500. Stanford, CA.

Bennet, D. 1975. *Spatial and Temporal Uses of English Prepositions. An Essay in Stratificational Semantics*. London: Longman.

Claus, B.; Eyferth, K.; Gips, C.; Hörnig, R.; Schmid, U.; Wiebrock, S.; and Wysotzki, F. 1988. Reference Frames for Spatial Inference in Text Understanding. In Freksa, C.; Habel, C.; and Wender K., eds. Spatial Cognition--An interdisciplinary approach to representing and processing spatial knowledge 1404:214--226.

Coyne, B.; and Sproat, R. 2001. WordsEye: An Automatic Text-to-Scene Conversion System. In Proceedings of SIGGRAPH-01, 487-496. Los Angeles, CA.

Davis, R.; Shrobe, H.; and Szolovits, P. 1993. What is Knowledge Representation? *AI Magazine*, 14:17-33.

Dupuy, S.; Egges, A.; Legendre, V.; and Nugues, P. 2001. Generating a 3D Simulation of a Car Accident from a Written Description in Natural Language: the CarSim System. In Proceedings of the Workshop on Temporal and Spatial Information Processing, 1-8. Toulouse, France.

Freeman, J. 1975. The Modeling of Spatial Relations. *Computer Graphics and Image Processing* 4:156-171.

Gapp, K. 1994. Basic Meanings of Spatial Relations: Computation and Evaluation in 3D Space. In Proceedings of AAAI-94, 1393-1398. Seattle, WA.

Hawkins, B. 1984. The Semantics of English Spatial Prepositions. Ph.D. diss., University of California, San Diego.

Herskovits, A. 1986. *Language and Spatial Cognition: An interdisciplinary Study of the Prepositions in English*. Cambridge: Cambridge University Press.

Hill, C. 1982. *Up/down, front/back, left/right. A contrastive study of Hausa and English.* In Weissenborn, J. and Klein, W., eds. *Here and There. Cross-Linguistic Studies of Deixis and Demonstration.* Amsterdam: John Benjamins.

Johnson-Laird, P. 1983. *Mental Models.* Cambridge: Harvard University Press.

Kuipers, B. 1978. Modeling Spatial Knowledge. *Cognitive Science* 2:129-153.

Langacker, R. 1987. *Foundations of Cognitive Grammar, Volume 1, Theoretical Prerequisites*. Stanford: Stanford University Press.

Mahesh, K. 1996. Ontology Development for Machine Translation: Ideology and Methodology. Technical Report MCCS-96-292. Computing Research Laboratory: New Mexico State University.

Mukerjee, A. 1998. Neat vs Scruffy: A Survey of Computational Models for Spatial Expressions. In Olivier, P., and Gapp, K., eds. *Computational Representation and Processing of Spatial Expressions.*

Olivier, P.; and Tsujii, J. 1994. A computational view of the cognitive semantics of spatial prepositions. In Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94), Las Cruces, New Mexico.

Sowa, J., ed. 1991. *Principles of Semantic Networks: Explorations in the Representation of Knowledge by Computers.* New York: Academic Press.

Srihari, R. 1994. Computational Models for Integrating Linguistic and Visual Information: A Survey. *Artificial Intelligence Review* 8:349-369.

Talmy, L. 1983. *How Language Structures Space.* In Pick, H. and Acredolo, L., eds. *Spatial Orientation: Theory, Research, and Application.* New York: Plenum Press.

Wahlster, W. 1996. Text and Images. In Cole, R.; Mariana, J.; Uszkoreit, H.; Zaenen, A.; and Zue, V., eds. Survey of the State of the Art in Human Language Technology. Kluwer: Dordrecht.

Xu, K.; Stewart, J.; and Fiume, E. 2002. Constraint-Based Automatic Placement for Scene Composition. In Proceedings of the Conference on Human-Computer Interaction and Computer Graphics, 25-34. Calgary, Canada.

Yamada, A.; Yamamoto, T.; Ikeda, H.; Nishida, T.; and Doshita, S. 1992. Reconstructing Spatial Image from Natural Language Texts. In Proceedings of COLING-92, 1279-1283. Grenoble, France.

Yamada, A. 1993. Studies on Spatial Description Understanding Based on Geometric Constraints Satisfaction. Ph.D. diss., University of Kyoto.