Frame of Reference and the Representation of Commonsense Spatial Relations

Sarah A. Douglas

CIS-TR-88-11
July 12, 1988

Department of Computer and Information Science
University of Oregon

Frame of Reference and the Representation of Commonsense Spatial Relations

Sarah A. Douglas
Department of Computer and Information Sciences
University of Oregon
(503) 686-4408
CSNET: douglas@uoregon

ABSTRACT

Beginning with an analysis of existing approaches to the representation of spatial relations, this paper points out problems in interpretation caused by assuming a representation derived directly from expressions in natural language, particularly English. In particular, ambiguity exists in determining the frame-of-reference which in English can be viewer centered or reference object centered. Both viewer centered and object centered reference remain problematic. A representation which is viewer centered depends on information which is not available during interpretation. A representation which is object centered is highly typed and only available to a subset of objects. This argument coupled with numerous cross-linguistic examples suggests caution in assuming a simple universal "commonsense" description of space.

INTRODUCTION

Few researchers in artificial intelligence would question the fundamental necessity of representing and reasoning about world knowledge. An important subgoal in this process is formalizing ordinary everyday knowledge of space, best illustrated by spatial concepts such as *left*, *right*, *top* and *bottom*. (For a cogent and compelling justification of this enterprise, see Hayes, 1978, 1985a, 1985b.) Clearly commonsense spatial knowledge performs a crucial function in vision and scene analysis, robotics, natural language, spatial data bases, and qualitative simulation. However, in a recent rereading of the artificial intelligence literature, I was struck by an

apparent confusion about the nature and use of such commonsense spatial concepts and their symbolic representation.

Before plunging into the specifics of spatial representation, it is appropriate to define representation in general. Hayes (1985b) gives this succinct definition:

> Every representational formalism I know which has even the glimmerings of a clear semantics, is based on the idea of individual entities and relationshsips between them. The individuals need not be *physical* individuals; they may be abstract "things" like *the color green*....They may only exist in some imaginary world...indeed, they may actually *be* an imaginary world, or a state of it....They partake in relations with other crisp *individuals*: physical relations such as being inside of, having as color, and abstract relations such as being in 1:1 correspondence with, or bearing an analogy to, or being a physical property. (p.72)

The fundamental idea of representation is individuals and their relationships, whether one subscribes to a Tarskian model-theoretics or not, or whether one uses logic, frames, semantic nets, etc. One of the fundamental purposes of representation is to be able to identify individuals by virtue of those relationships. A spatial relation using prepositions such as "x is to the left of y" denotes a relation between a certain individual, x, and another reference individual, y. We may conceive of these relationships as more or less permanent, depending on our commitment to their logical and epistemological necessity (or our subscription to an essentialist doctrine).

Nothing I have said so far is controversial. However, I next want to present two examples, one from Patrick Winston and the other from Patrick Hayes, in which the representation of individuals having properties of spatial relations for purposes of identification becomes problematic. Winston and Hayes are not singled out for any purpose other than that they represent exemplars of the general literature.

THE PROBLEM WITH SPATIAL REPRESENTATION

Example One: Left and Right

An initial example is found in the work of Patrick Winston in his 1975 paper on "Learning structural descriptions from examples". Figure 1 shows a scene description from this paper and Figure 2 shows the knowledge representation. Winston uses the relations *Right-Of* and *Left-Of* as criterial features in determining the structure of the arch. *Right-Of* and *Left-Of* clearly relate two individuals, the B leg and the C leg, which have other properties represented: *Must-Not-Abut* (B,C); *Must-Not-Abut* (C,B); *A-Kind-Of* (*brick*,B); and *A-Kind-Of* (*brick*,C); etc.

If we take Winston's representation of an arch to be *strongly direct*, like a family tree, then if a relation is absent, it is not simply an incomplete representation, but a wrong one (c.f. Hayes, 1974). Similarly, by distinguishing between the two legs, Winston implies that the spatial distinction is crucial for identification. Two problems arise. First, examining Figure 1, we notice that leg B is left of leg C from a frame-of-reference established by the viewer. That is, the left leg, B, of the arch is the left side of the viewer of Figure 1. Second, the relations are not necessarily constant when the object is moved. We could rotate the arch 180 degrees on its horizontal axis, making the right leg the left leg and vice versa, introducing contradiction into our logical description if we believe that right and left are asymmetric. This follows from the first problem of defining the relations from a viewer centered frame-of-reference. The interpretation of these spatial relations for identification purposes is greatly limited since the frame-of-reference of the viewer is not incorporated into the representation and literally unknown and ambiguous.

Example Two: Top and Bottom

A second example comes from Hayes' paper "Naive Physics I: Ontology for liquids" (1985b). In this paper, which is an exercise in formalizing in full predicate calculus with equality a representation of commonsense knowledge about liquids, Hayes begins by formalizing the notion of container in which he eventually defines

the relations *Rwu*, "right way up", and *Top*. The expression,

Rwu(o) iff *Top(top(o),o)*

can be read as saying that an object is right way up if and only if it is a container with only one opening, *top(o)*, and that opening has a face whose surface normal has a positive vertical component, *Top*. This would be an appropriate definition for describing a "right way up " vase sitting on a table. (See Figure 3.) However, *Top*(top(o),o) is meaningless without a frame of reference with gravitational alignment.

Unlike Winston, Hayes cautions his reader with the following, "It is important to realize that *Top*, and hence *Rwu*, are not relations which are intrinsic to an object. Unlike most of what we have considered they are liable to be altered when the object is moved around in space." (p. 83). Figure 4 illustrates the same vase "upside-down". Hayes is making a very subtle distinction in representation between the *top(o)*, i.e. its unique opening, which is aligned by a frame-of-reference intrinsic to the object, and *Top(top(o),o)* as a description of an intrinsically oriented object whose alignment is coincident with gravity. This replaces the viewer centered frame-of-reference that is inherent in Winston's description with an object centered one.

Commonsense, seems to assign object intrinsic orientation (a frame-of-reference) to subclasses of objects in a fairly systematic way. But caution is necessary. Although the vase as a container in Figures 3 and 4 seems to have a "natural" top which is an opening, this is not necessarily true for all containers. If the container were a transfusion bottle, then "right-way-up", illustrated in Figure 5, and "upside-down", illustrated in Figure 6, would be the opposite of the vase. Thus, the vase and the transfusion bottle, while both containers for liquids do not share the same intrinsic orientation relative to their unique openings. Intrinsic vertical orientation is not a function of gravity, shape or containment. It is the result of human functionality. This means that a complex typing system for spatial relations exists which is much finer than a "container" object.

One last problem with Hayes' definitions is important. The relation *In(x,y)* is crucial for the enterprise of describing containers, where one object's space wholly contains another's. For example, imagine a bowl which contains an apple (Figure 7). Both

English and Chinese speakers would say "The apple is *in* the bowl." Now imagine that the bowl is turned upside down (Figure 8). English speakers would now say "The apple is *under* the bowl.", even though it is still within the containment of the bowl according to Hayes' definition. But Chinese speakers would still say "The apple is *in* the bowl.", maintaining the definition. The difficulty that arises, is that these kinds of assignments and interpretations of commonsense spatial relations are not some universal human knowing, but a highly language and culturally specific representation of space (Talmy, 1983).

Summary

In this section, I have illustrated several problems with the representation and interpretation of commonsense spatial relations:

- Spatial descriptions always have a frame-of-reference for interpretation.
- Frame-of-reference can be either viewer centered or object centered.
- Object centered reference is highly typed.
- "Commonsense" spatial descriptions may be highly idiosyncratic to particular natural languages and cultures.

In the next section, I will give a brief review of some of the complexities of the logic of spatial description in natural language with an eye to evaluating the attempt to represent their commonsense.

THE LOGIC OF COMMONSENSE SPATIAL RELATIONS

The linguistic description of space in natural languages provides the major empirical basis for understanding the representation of commonsense space. As noted above, English and most languages (but not all) provide two frames-of-reference for spatial description: one which is viewer centered and another which is object centered.

Viewer Centered Reference

Viewer centered reference is called *deixis*. Deixis comes from a Greek word meaning "pointing" or "indicating" and is now used in linguistics to relate utterances to the spatio-temporal coordinates of the act of uttering (Lyons, 1977). Since utterance means the verbal action of speaking, it is usually implicit that both speaker and hearer are physically present together and thus both viewers. Oral language predates both writing and telecommunications, and it is easy to see how language as an act could presuppose in its lexical and grammatical forms that both speaker and hearer would be physically present together.

Deictic use of space can be divided into those uses which choose proximity to the speaker as the primary element, such as the English *here* and *there*, and those uses which establish a more complex speaker centered reference system. I will not pursue the details of the first type of deixis here since they are so clearly contextual that they never appear in problems of representations of the sort I discussed earlier.

A more complex system of secondary spatial deixis exists in most languages.[1] It consists of three subsystems, each correlated with an axis in three dimensional space. These axes are based on physical parameters such as gravity and biological parameters such as the location of the face. *Above* and *Below* designate the vertical axis; *Front* and *Back* designate depth; and *Right* and *Left* designate horizontal. Whenever an orientation on a particular axis is expressed in terms of secondary spatial deixis, the spatial relation has to be interpreted as a three-place relation with the viewer's origo (basic reference point) as the third term. Thus "the arch's right leg" is the leg to my right as I am looking at it and uttering the expression:

Right-Of (B, C, *origo*(V))

where *V* is me (or you), the viewer. Since I expect that most of my readers will orient the paper in the same way as I do, I can almost guarantee an unambiguous description. But does this have any

---

[1] John Haviland, a linguist at Reed College, reports that he has encountered an Austrailian aboriginal language in which there is no secondary spatial deixis. All orientation is in reference to a global system roughly equivalent to compass points. There are no secondary spatial deixis concepts such as *left* or *right*.

meaning in a computational repesentation without the specification of the viewer centered frame-of-reference?

Deictic or viewer centered spatial relations add an epistemic quality to the description of objects that can extend far beyond simply establishing the frame-of-reference of the speaker or hearer. Languages may require the speaker to code aspects of the visual field as well. For example, Cora, a Uto-Aztecan language, requires its speakers to note whether a described object is fully accessible to the speaker's view or not (Casad & Langacker, 1985). In viewing a dog's tail from the side, part of the body may cut off the view so that while the speaker is saying "That dog has a short tail." there is an epistemic marker added which says "(Viewed from the side which might cut off my view.)". If the same dog is viewed from the back, with an unobstructed view, then the sentence has an epistemic marker added which says, "(Viewed from the back which is a full view.)". What may not be apparent from these English translations is that every description must be marked by the epistemic view of the speaker.

English may have a similar, though different type of epistemic marking. Herskovits (1985) cites the example that saying "Jim is *at* the supermarket." implies that speaker and hearer are not there also. If they were also in the supermarket, they would say, "Jim is *in* the supermarket." Thus, even apparently equivalent spatial descriptions from a logical point of view are loaded with deictic and epistemic nuances.

Object Centered Reference

Object centered use of spatial description ascribes the frame-of-reference of the axes and orientation to the reference object itself, not the speaker or hearer. However, in English only certain classes of physical objects can establish an intrinsic frame-of-reference. Persons and animals have their *Front* determined by the location of central perceptual organs, their face. Intrinsic *Front/Back* orientation for inanimate objects is derived from two sources: either their location in a functional encounter with a human (that is, front to front[1]) or their direction of travel. Thus, a CRT has a *Front* which is the surface with keyboard and screen ("Is it

[2] Some languages, notably Swahili, have a canonical encounter which is front to back.

facing us?") and a car has a *Front* which is the end that goes first down the road, well you know where I mean. From this object (not viewer) orientation, the spatial relations *Top* and *Bottom* are derivative from gravitational alignment of the object itself. Containers such as bottles which do have intrinsic orientation from function in the vertical plane, are often symmetrical in the depth and horizontal plane, and thus have no *Front/Back*, or *Right/Left*. As is evident from English, languages have a very complex system for the assignment of these object centered frame-of-references.

Using Spatial Relations for Identification

In the previous section I came to the conclusion that spatial relations can be complex expressions that are chosen to represent spatial position from several possible frames of reference: the speaker's or hearer's origo, or some reference object which in English can establish its own intrinsic origo. Thus all spatial description adds a third term, the origo, to the relation. If the complexity of spatial description in natural language seems illogical and imprecise when compared to a formal equivalent, it is important to keep in mind the purpose and context of use of spatial relations in language. The purpose of using such spatial relations is as referring expressions. If the reference is successful, the referring expression will correctly identify for the hearer the individual referent. Note that this is a different function of language from ascribing a property or asserting something. Successful reference does not depend on the truth of the description but on the hearer's ability to identify the referent from the description (Strawson, 1959). This definition of referring expressions is consistent with a Tarskian model-theoretic semantics (Lyons, 1977), since while the description may not be true in the actual world, it could be true in some possible world. But reducing the problem to one of truth-functionality prevents exploration of the perceptual and cognitive aspects, and much less, the collaborative aspects of human language.

Even reducing the problem of spatial description to succesful reference does not eliminate the problem of ambiguity. Ambiguity comes from at least two sources: determining from which origo the description is made, and determining the speaker's origo in secondary spatial deixis. How then does language cope with ambiguity? It is the speaker's choice to use either a deictic or intrinsic frame-of-reference. This choice can cause inherent

ambiguity for the hearer's interpretation which can sometimes be resolved if the hearer can problem solve the reference to a unique individual. It has been shown that in connected discourse, speakers choose secondary deixis over intrinsic when they may be encountering objects that cannot be subsumed with an intrinsic frame-of-reference (Ehrich, 1985). This strategy eliminates the confusion caused by shifting the frame-of-reference. On the other hand, single sentence descriptions are frequently given in an object centered reference system (if possible), thus confirming the speculations of Miller & Johnson-Laird (1976). As mentioned earlier secondary spatial deixis descriptions have an unambiguous interpretation in face-to-face oral discourse where the hearer can see the speaker's position.

In written text, ambiguity is more difficult to resolve since the reader does not have access to the writer's position in viewing the reference objects. Readers often have difficulty trying to resolve whether an intrinsic or deictic frame is used. This is illustrated by the familiar example of reading an automobile repair manual which locates the carburetor on the left side of the car. Is it the left of the reader standing at the "front" of the car and looking in (the deictic description), or is it the intrinsic left of the car? Frequently types of text, such as repair manuals, use a canonical intrinsic description to avoid the impossibility of determining spatial deixis. They also use pictures to help the reader orient to the referred parts.

Finally, whatever ambiguity is unresolved by these conventions or perceptual evaluations can always be resolved in oral language by collaboration and in written language by expectations that the reader can problem solve. Two recent papers (Clark & Wilkes-Gibbs, 1986; Klein, 1982) show how "talk" is structured to accommodate the fallibility and ambiguity of natural language. Language has enormous repair facilities to overcome its restricted, normally speaker-centered description.

WHAT IT WILL TAKE TO REPRESENT COMMONSENSE SPACE

Knowledge representation must begin to come to grips with these issues and the complexity of what it means to represent world knowledge and commonsense. I have briefly argued and shown by example, that commonsense spatial relations are much more

complex than the types of representations we attempt today in AI. The introduction of the third term, origo or frame-of-reference, admits context sensitivity. Researchers must be very clear what interpretation a spatial relation should have since ambiguity is inherent in a representation that does not locate the origo. Origo must be handled as a homomorphic (structure preserving) mapping from the origo object to the object of reference.

In addition to the issue of point of view, as noted above, it is the case that judgements of spatial relations such as *left*, *right*, *above*, *below* and *between*, can vary depending on visual perception of shape and size, discourse task, and contextual arrangement. Prototype positions for some situations also exist. While this is not the place to discuss all of these complexities and their implications, suffice it to say that our current research clearly supports the view of Herskovits (1985) that the inference of spatial relations is intricately complex. However, we do not yet subscribe to the view that they are computationally intractable. (See Douglas et al. 1987.) We take heart in the fact that human discourse has remarkable uniformity.

Finally, by no means does it appear that there is some universal human commonsense for spatial description if we look at the evidence of natural language. There may exist similarities between systems, but they are intersections. For example, English has a very complex typing system for determining object centered reference. Thus any description of even English commonsense spatial relations must take into account domain specifics, rather than some general (and weaker) representation.

CONCLUSIONS

This paper has avoided all the usual discussions about spatial representations and whether symbolic (descriptive) representations are sufficient or whether analogic (2-D array) representations must be introduced. Instead, I have focused on spatial relations at the knowledge level: what they require for interpretation.

I have suggested that natural language has a tremendously complex system of representation for space which varies by language and by perception and intention of the speaker. Cross-linguistic evidence should dissuade us from believing that there

exists a single "commonsense" description of space. I have also suggested that any representations of spatial relations that depends on English must contain a third term, the origo, in order for interpretation to be unambiguous. If the representation relies on a viewer centered origo, then the interpretation is inherently ambiguous unless the position of the viewer can be determined.

Finally, by adopting this position I have claimed that the function of spatial relations is not truth ascribing but for purposes of reference resolution (identifying individuals) in the world. Thus the construction of meaning in natural language becomes an inherently collaborative act between speaker and hearer constrained only by the success, and not the truth-value of the expression. This must be kept in mind when we use knowledge representations that rely on our intuitions of natural language for interpretation.

REFERENCES

Clark, H. and Wilkes-Gibbs, (1986). Referring as a collaborative act. *Cognition, 22*, 1-39.

Casad, E.H. and R.W. Langacker (1985). `Inside' and `Outside' in Cora grammar. *International Journal of American Linguistics, 53*, #3.

Douglas, S. A., Novick, D.G. and Tomlin, R.S. Consistency and variation in spatial reference. *Proceedings of the Ninth International Conference on Cognitive Science*, July 1987, Seattle, Washington.

Ehrich, Veronika (1985) The Linguistics and Psycholinguistics of Secondary Spatial Deixis. In G.A.J. Hopenbrouwers, P.A.M. Seuren, and A.J.M.M. Weijters, *Meaning and the Lexicon*. Foris Publications: Dordrecht, Holland.

Hayes, P. (1974). Some problems and non-problems in representation theory. *Proceedings AISB Summer Conference, University of Sussex*, pp. 63-79.

Hayes, P. (1978). The naive physics manifesto. In D. Michie (Ed.) *Expert Systems in the Microelectronic Age*. Edinburgh, Scotland: Edinburgh University Press.

Hayes, P. (1985a). The second naive physics manifesto. In J.R. Hobbs and R.C. Moore (Eds.), *Formal Theories of the Commonsense World.* Norwood, NJ: Ablex Publishing Corp.

Hayes, P. (1985b). Naive Physics I: Ontology for liquids. In J.R. Hobbs and R.C. Moore (Eds.), *Formal Theories of the Commonsense World.* Norwood, NJ: Ablex Publishing Corp.

Herskovits, A. (1985) Semantics and pragmatics of locative expressions. *Cognitive Science, 9,* 341-378.

Klein, W. (1982). Local deixis in route directions. In R. Jarvella & W. Klein (Eds.) *Speech, Place, and Action.* Chichester: Wiley. Lyons, J. (1977). *Semantics.* Cambridge, England: Cambridge University Press.

Miller, G. and Johnson-Laird, P. (1976) *Language and Perception,* Cambridge, MA: Harvard University Press.

Strawson, P.F. (1959). *Individuals.* London: Methuen.

Talmy, L. (1983). How language structures space. In H. Pick and L. Acredolo (Eds.) *Spatial Orientation: Theory, Research, and Application.* New York: Plenum Press.

Winston, P. (1975). Learning structural descriptions from examples. In P. Winston (Ed.) *The Psychology of Computer Vision.* New York: McGraw-Hill.

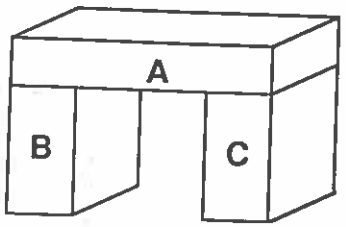

Figure 1
From P. Winston,1975)

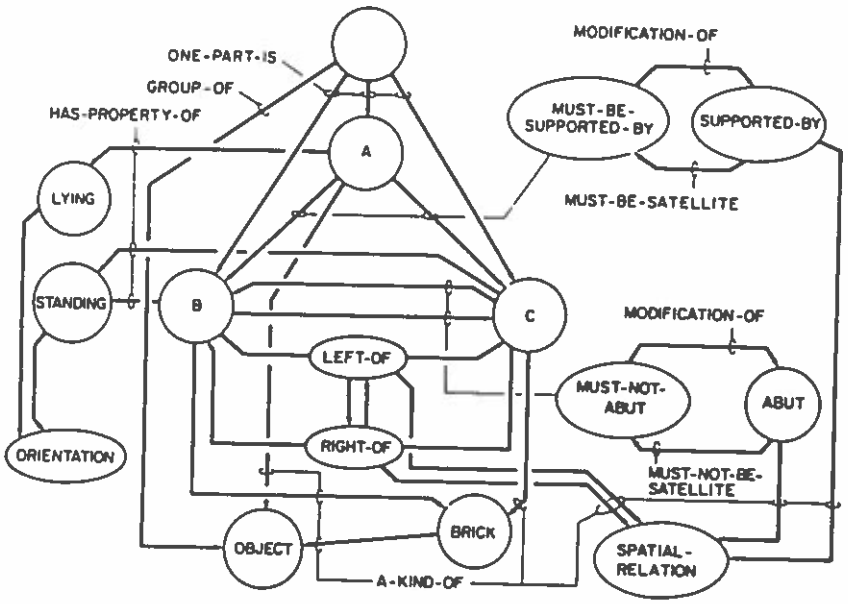

Figure 2
(From P. Winston,1975)

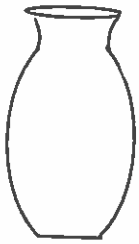

Figure 3

Figure 4

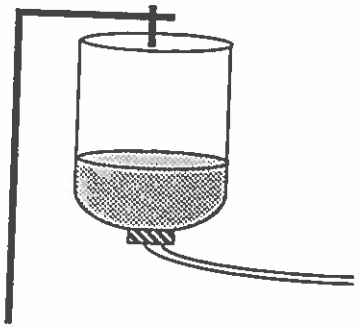

Figure 5

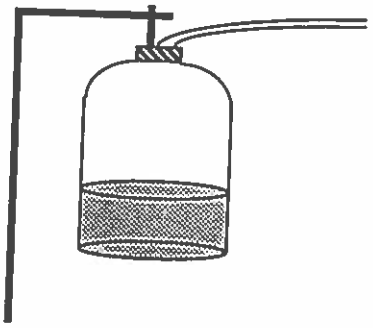

Figure 6

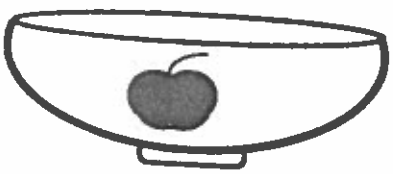

Figure 7

Figure 8