An Ontology of Information Structure

Jerry R. Hobbs
USC Information Sciences Institute
Marina del Rey, California
hobbs@isi.edu

1 Introduction

Search engines today are very good at finding information in text resources. But very often the best answer to a question is in a diagram, a map, a photograph, or a video. For example, consider the questions

> What is the Krebs cycle?
> How has the average height of adult American males varied over the last 100
>   years?
> How did the Native Americans get to America?
> What does Silvio Berlusconi look like?
> When will the various tasks on this project be completed?

The answer to the first should be a process diagram, the second a graph, the third a map with routes indicated, and the fourth a photograph. The answer to the last might best be presented in a Gantt chart.

Search engines are very much poorer at finding this kind of information, and generally they do so by looking at the associated text. One of the benefits of the Semantic Web should be that this kind of information would be encoded in a fashion that makes it more retrievable, in part by describing it better, and in part by expressing "coreference" relations among material presented in different media. For example, a diagram may bear a kind of coreference relation with a text segment in the document it is a part of. There may be a coreference relation between an object in a photograph and a noun phrase in the caption. In a news video clip I have analyzed, a woman refers to sending her "four children" to local schools, as she is standing by a wall covered with family photographs; there is also a coreference relation here.

The aim of this paper is to present an ontology[1] of the structure of information that will support a variety of statements about documents in various media, their internal

---

[1] By "ontology" I mean a logical theory, i.e., a set of predicates and functions, and a set of axioms involving the predicates and functions that constrain their possible interpretations.

structure, and how they function in the world at large, thereby providing a unified vocabulary for talking about entities that convey information. I begin by sketching an approach to anchoring symbolic systems in human cognition and discuss various levels of intentionality that occur. I then consider compositionality in different symbolic systems, and the sometimes complex coreference relations that arise from that. This theory is then applied to the specific case of diagrams as information-bearing objects, and a logical theory of Gantt charts is constructed as an illustration. Finally there is a discussion of issues raised with respect to various modalities and various manifestations of symbolic artifacts.

2 Grounding Symbols in Cognition

In this paper I will assume that we have a coherent notion of causality, as in Hobbs (2001), and change of state, as in Hobbs et al. (1987), as well as a theory of commonsense psychology at least rich enough to account for perception, planning and intentional behavior, and what I here call "cognizing", that is, taking some cognitive stance toward, such as belief, thinking of, wondering about, and so on. I will refer to the contents of thoughts and beliefs as "concepts", a general notion that subsumes propositions (Gordon and Hobbs, 2003), but also includes nonpropositional concepts like "dog" and "near", images, vague feelings of apprehension, and so on. I will assume the "ontologically promiscuous" notation of Hobbs (1985a), but for typographical convenience, I will abuse it by using predications as arguments in other predications, where a proper treatment would reify the corresponding eventualities and use those as arguments. Some of the inferences below are defeasible, and thus the underlying logic must support a treatment of defeasible inference. There are many frameworks for this, e.g., McCarthy (1980) and Hobbs et al. (1993). To minimize notational complexity, defeasibility is not made explicit in the axioms in this paper.

The basic pattern that symbols rest on is the perception of some external stimulus causing an agent to cognize a concept.

(1) $cause(perceive(a, x), cognize(a, c))$

where a is an agent, x is some entity, and c is a concept. x can be any kind of perceptible entity, including physical objects and physical properties, states, events and processes, and, as we will see later, more abstract entities as well. That is, we can perceive a ball, its roundness, and the event of someone throwing it. Among the states that can be perceived are absences. Seeing that someone's car is not in his garage can cause me to believe he is not at home. Silence, or absence of speech, can often carry very significant meaning.

The concept c may often be hard to put into words, such as the concepts triggered by music or an aesthetic design.

This pattern covers the case of a cloud reminding someone of a dog, where there is no external causal connection between the stimulus and the concept, and the case of smoke making one think of fire, where there is a causal connection, and the intermediate case of

an association that has been established by practice, as in a dinner bell making one think of food.

Some concepts are tied in such a way to the entity perceived that they can be called the "concept of" the entity. We could introduce *concept-of* as a function mapping from the entity to the concept, but since the predicate *cognize* always takes a concept as its second argument, it is simpler to build the coercion into the predicate *cognize*. Thus, if e is an entity, $cognize(a, e)$ says that agent a cognizes the concept of e. The key relation between entities and their concepts is that perceiving the entity causes the agent to cognize the concept of the entity.

(2) $cause(perceive(a, e), cognize(a, e))$

It is important to note, however, that perception can trigger many concepts and that not everything that is cognized needs to be what is perceived. Perceiving a bell can cause an agent to cognize food (as well as the bell). This makes symbols possible.

Where the concept cognized is propositional, we could talk about its truth in the world. That is, it is not only true that e is cognized; it also holds in the real world. However, this will play no role in this paper. The meanings of symbols will be strictly in terms of the concepts they invoke in the recipient.

Communication begins when another agent presents an entity causing the first agent to perceive it.

(3) $cause(present(b, x, a), perceive(a, x))$

For an agent b to present something to a is for b to cause it to be within the range of a's perception, and this causes a to perceive it.

The recipient agent a must of course be capable of cognition. A greater range of sending agents b is possible. A car that beeps when you don't fasten your seatbelt is an agent b that is presenting a signal x for the driver to cognize. It is also possible for collectives to be the sending agent, as in jointly authored documents such as the Constitution of the United States. The agents may or may not exhibit intentionality. Humans do, as do organizations of humans, whereas simple artifacts merely reflect the intentionality of their designer. Sufficiently complex artifacts may exhibit intentionality.

Causality is defeasibly transitive, so Rules (1) and (3) can be combined into the defeasible causal pattern for appropriate c's:

(4) $cause(present(b, x, a), cognize(a, c))$

That is, if b presents x to a, it will cause a to cognize the appropriate concept c. For example, a car beeps and that causes the driver to hear the beep; hearing the beep causes the driver to remember to fasten her seatbelt. So the beep reminds the driver to fasten her seat belt.

We will refer to the entity presented (x) as the symbol and to the concept evoked (c) as the meaning of the symbol.

3   Intention and Convention in Communication

Presentation by an agent can involve several levels of intentionality, and the perception can involve several levels of recognition of intentionality. First, the presentation can be entirely unintentional, as, for example, when someone conveys his nervousness by fidgeting or shaking his leg. In an abductive account of intelligent agents, an agent a interprets the environment by telling the most plausible causal story for the observables in it. Here a knows nervousness causes fidgeting and the most plausible causal story is that b's fidgeting is because b is nervous. When b says "ouch" and a infers that b feels pain, the account is exactly the same.

When the presentation is intentional, the presenter's goal is to cause the perceiver to cognize something. The presenter's intention need not be recognized. For example, I may keep the door to my office closed to lead people to believe I am not in, without wanting them to recognize my intention to communicate that.

Intention is recognized when it is part of an observer's explanation that an event occurs because the agent of the event had the goal that it occur. Defeasibly, agents do what they want to, when they can.[2]

(5) $goal(g, b) \wedge executable(g, b) \supset cause(goal(g, b), g)$

That is, if g is a goal of b's and is executable by b (or achievable by an executable action), then its being a goal will cause it to actually occur. I won't explicate *executable* here, but it means that g is (achievable by) an action of which b is the agent, and all the preconditions for the action are satisfied.

When an observer a uses this causal rule, he is recognizing the intention that lies behind the occurrence of the event.

It is most common in human communication that the intention is recognized. Agent b knows that presenting x causes a to perceive x, which causes a cognize concept c. b has the goal that a cognize c. So that causes b to present x. Agent a comes up with exactly this causal explanation of b's action of presentation, so not only does a cognize c; a also recognizes b's goal that a cognize c.

This recognition relies on agents' knowing a defeasible rule that says that

(6) $goal(g_1, b) \wedge cause(g_2, g_1) \supset goal(g_2, b)$

That is, if an agent b has a goal g_1 and g_2 tends to cause g_1, then b may have g_2 as a goal as well.

---

[2]All axioms are universally quantified on the variables in the antecedents of the highest-level implication.

In the case of communication, g_1 is $cognize(a, c)$ and g_2 is $present(b, x, a)$. The recipient observes the event of the presenting, uses axiom (5) to infer abductively that it is intentional, and uses axiom (6) together with schema (4) to recognize that b intends for a to cognize c.

We can get to full Gricean nonnatural meaning (Grice, 1989) by decomposing Rule (6) into two rules:

(7) $goal(g_1, b) \wedge cause(g_2, g_1) \supset goal(cause(g_2, g_1), b)$

(8) $goal(cause(g_2, g_1), b) \supset goal(g_2, b)$

That is, if an agent b has a goal g_1 and g_2 tends to cause g_1, then b may have as a goal that g_2 cause g_1. Moreover, if an agent b has as a goal that g_2 cause g_1, then b has the goal g_2.

When g_1 is $cognize(a, c)$ and g_2 is $present(b, x, a)$, a uses axioms (7) and (8) to explain the presentation; then a will recognize not only b's intention to have a cognize c, but also b's intention that a do so *by virtue of* the causal relation between b's presentation of x and a's cognizing c.

In order for this sort of communication to work, it must be mutually known between a and b that presenting x causes cognizing c.

Communicative conventions are causal rules of this sort that grow up in different groups. The structure of a communicative convention is

$$mb(s, cause(present(b, x, a), cognize(a, c))) \wedge member(a, s) \wedge member(b, s)$$

for a specific x and a specific c. That is, a social group s that agents a and b are members of mutually believe the causal relation between presenting x and cognizing c. For example, x might be a red flag with a white diagonal, s might be the community of boaters, and c the concept that there is a scuba diver below.

These communicative conventions can originate and take hold in a group in many different ways. The culture of a group consists in large part of a number of such rules.

Note that there is nothing particularly advanced about the arbitrariness of the symbol x. That is already there in the most primitive stage, in the connection between the bell and the food.

This completes the sketch of how the meaning of atomic symbols can be grounded in a theory of cognition: in our scheme, x is a symbol that means or represents c to a group of agents s. In an elaboration of Pease and Niles (2001) we can express this as

$$means(x, c, s)$$

I will leave out the third argument in the development of the theory of diagrams below; the community is simply the set of people expected to be able to understand the diagrams.

We next turn to how more complex symbolic objects convey more complex meanings in different modalities.

4 Composition in Symbol Systems

An atomic symbol, i.e., one that does not have interpretable parts, corresponds to some concept. Atomic symbols can be composed in various ways, depending on the type of symbol system, and the meaning of the composition is determined by meaning of the parts and the mode of composition. These composite elements can then be components in larger structures, giving us symbolic structures of arbitrary complexity.

Composition in symbol systems occurs when entities x and y, meaning c_1 and c_2 respectively, are presented with a relation R_1 between them, where R_1 conveys the relation R_2 in the target domain. Thus, we have three causal relations.

$$cause(present(b, x, a), cognize(a, c_1))$$
$$cause(present(b, y, a), cognize(a, c_2))$$
$$cause(present(b, R_1(x, y), a), cognize(a, R_2(c_1, c_2)))$$

The relation $R_1(x, y)$ can be thought of as just another entity in the symbol system, so it is subject to full Gricean interpretation just as atomic symbols are, and it can similarly be involved in the conventions of some community.

With respect to the concepts invoked, we will confine ourselves here to propositional concepts. The advantage of having a flat notation in which anything can be reified is that when composite concepts are constructed, we can view this as simply a conjunction of what is already cognized with the new relations conveyed by the composition. The recipient of the message $R_1(x, y)$ cognizes c_1, c_2, and $R_2(c_1, c_2)$.

Speech (and text as spoken) takes place in time, so the only compositional relation possible is concatenation. Within sentences, the composition of smaller units into larger units conveys primarily a predicate-argument relation between the meanings of the components. Thus, when we concatenate "men" and "work" into "men work", we are indicating that the referent of "men" is an argument or role-filler in the event denoted by "work". This view of syntax as conveying predicate-argument (and modification) relations through adjacency of constituents is elaborated in Hobbs (1998), in which an extensive grammar of English is developed in a manner similar to the Head-driven Phrase Structure Grammar of Pollard and Sag (1994).

In discourse beyond the sentence, concatenation generally conveys a coherence relation based on causality, similarity, and figure-ground (Hobbs, 1985b).

In tables, the elements in individual cells refer to some concept. The manner of composition is placement of these cells in a vertical and horizontal arrangement with other cells. Generally, the aggregate represents a set of relations: The item in a cell that is not the first in its row stands in some relation to the first element in the row. The relation is the same for all elements in that column, and is often explcitly labelled by a header at the top of the column. For example, in a table of United States presidents, we might have the year 1732 in one cell. The label on the row may be "George Washington", and the label on the column "Birth date". This spatial arrangement then conveys the relation $birthdate(GW, 1732)$.

A beep in your car can mean several things—you haven't fastened your seatbelt, the door is open, your lights are still on. When a sequence of beeps is combined with a property of that sequence, namely, a rapid frequency, the composite meaning can be more precise—your lights are still on.

A map is an interesting example of a complex visual symbolic object. There are underlying fields, indicated by colors, that represent regions of various political or geologic types. Icons are overlaid on these fields in a way that bears at least a topological relation to the reality represented, and perhaps a geometric relation. Generally there is a mixture of topological and geometric correspondences; the width of a road on a map is usually not proportional to its width in reality. The icons can have internal structure; for example, the size of a city may be represented by the number of circles around a center dot. Information is conveyed by the possible spatial relations of adjacency, distance, and orientation. For example, labels naming cities are placed near the icon representing the city.

In a process diagram, individual states may be represented by a set of icons standing in particular spatial relationships to each other. Adjacent states may be connected by arrows, the direction of the arrow indicating temporal order (Wahlster et al., 1993; Pineda and Garza, 2000).

Documents are composed of blocks of text and other symbolic elements in a particular spatial relationship. Some spatial relationships are tightly constrained. Titles need to be at the top of the document; the relation conveyed is that the title somehow indicates the content of the body of the document. Paragraphs meant to be read in sequence should be adjacent (with exceptions for page and column breaks). Paragraph breaks should also correlate with the coherence structure of the text, although often in a complex way. Other spatial relationships can be looser. A diagram, map, photograph, or sidebar only needs to be relatively near the block of text describing roughly the same thing, and the ordering of such figures should respect the order of references to them in the body of the text (Powers et al., 2003).

Web sites and PowerPoint presentations are complex symbolic objects amenable to similar analysis.

The performance of an individual in face-to-face conversation (Allwood, 2002) is also a complex symbolic object with its own compositional principles. The principal component symbolic elements are the speech stream, prosody, facial expressions, gaze direction, body posture, and gestures with the hands and arms. The primary mode of composition is temporal synchrony. The relationship conveyed by temporal synchrony could either be that the content of the two presentations are the same, as when I make a rolling gesture while describing Tweety rolling down the hill as in McNeil's experiments (McNeil, 2000), or parallel actions, as when I freeze my arm in midgesture to hold my turn while beginning a new clause.

A play is a more complex composite symbolic object, but its structure is equally amenable to a compositional account in terms of adjacency relations among the symbolic components and conjoined relations in the meaning.

In Section 6 I develop more fully a theory of diagrams, to illustrate the application of

this theory of information structure to a special case of substantial importance.

5 Coreference

In a complex symbolic object, different component presentations may convey the same concept, or more generally, closely related concepts. Thus there is a coreference relation between them. The most familiar example is between two noun phrases in speech or text. On a map, the dot with a circle around it representing a city and the label with the name of the city are in a sense coreferential, both representing the city. In a process diagram, the identity of shape or color of icons in two state depictions may indicate coreference relations; they are the same entity in different stages of the process. In a document a region of a photograph, a phrase in the caption on the photograph, and a portion of the accompanying text may all be coreferential. An iconic gesture and the accompanying description in a face-to-face conversation may also be coreferential.

It is clear that the coding of coreference relations in Web sites would be very useful for retrieval. For example, we would be able to move beyond the rather haphazard results we now get from image searches.

6 A Theory of Diagrams

6.1 Background Theories

In developing a theory of diagrams, we will need to rely on concepts from several background theories, not all of which have been described in published papers.

Set Theory: We need one relation and one function:

$member(x, s)$:	x is a member of the set s.
$card(s) = n$:	The cardinality of set s is n.

Composite Entities: A composite entity x is something that has a set of components and a set of relations among the components. We will need two relations:

$componentOf(x, s)$:	x is a component of s.
$relationOf(r, s)$:	r is a relation of s.

This depends on reifying relations (cf. Hobbs, 1985a).

Scales: One-dimensional diagram objects, intervals of time, and, by extension, events are all scales and have beginnings and ends. It will be convenient to have these concepts in both relational and functional form:

$begins(x, s)$:	x is the beginning of s.
$ends(x, s)$:	x is the end of s.
$beginningOf(s) = x$:	x is the beginning of s.
$endOf(s) = x$:	x is the end of s.

Strings: We assume there are strings of characters. They are usually symbolic objects, so they can be the first argument of the predicate *means*.

Time: In the development on Gantt charts, reference is made to concepts in OWL-Time (Hobbs and Pan, 2004). This ontology posits temporal entities (i.e., intervals and instants), the beginnings and ends of intervals, a *before* relation, Allen interval relations like $intMeets$, and temporal aggregates, which are sets of nonoverlapping, ordered temporal entities. It also handles durations and clock and calendar terms. The three predicates we need here are the following:

$TimeLine(t)$:	t is the infinite interval containing all temporal entities.
$atTime(e, t)$:	Event e occurs at instant t.
$calInt(t)$:	t is a calendar interval, i.e., a calendar day, week, month, or year.

In addition, we will need one predicate relating strings to times:

$dateStringFor(s, t)$:	s is a string describing temporal entity t.

Causality: The one predicate we need from a theory of causality or processes (Hobbs, 2001) is *enables*:

$enables(e_1, e_2)$:	Event or condition e_1 enables, or is a prerequisite for, event or condition e_2.

For Gantt charts we need a simple ontology of projects, with the following three predicates.

$Project(p)$:	p is a project.
$taskIn(z, p)$:	z is one of the tasks of project p.
$milestoneIn(m, p)$:	m is one of the milestones of project p.

A project is a composite entity among whose components are tasks and milestones, which are events. The project and its parts can have names.

$name(s, z)$:	The string s is the name of z.

Space: The actual drawing of a diagram will involve mapping the ontology of diagrams to an ontology of space. Some portion of space will have to be chosen as the ground. This will define the vertical and horizontal directions and the *above* and *rightOf* relations. In addition, the articulation between the theory of diagrams and the theory of space would have to specify the kinds of spatial regions that realize different kinds of diagram objects.

6.2   Diagram Objects

A diagram consists of various diagram objects placed against a ground, where each diagram object has a meaning. We can take the ground to be a planar surface, which thus has points. Diagram objects can have labels placed near them, and generally they indicate something about the meaning. Diagram objects, points, frameworks, meanings, and labels are discussed in turn, and then it is shown how these can be used to define Gantt charts.

Diagram objects can be classified in terms of their dimensionality. In a spatial ontology in general, we would have to specify both a dimension of the object and the dimension of the embedding space, but in this theory of diagrams, we will take our embedding space to be a two-dimensional plane. Thus, there are three types of diagram objects: $0DObject$, $1DObject$, $2DObject$.[3]

$$0DObject(x) \supset DiagramObject(x)$$

Zero-dimensional diagram objects in diagrams are the class of diagram objects that are treated as having zero dimensions in the context of the diagram. Of course, in a spatial ontology they would actually be small regions generally with some symmetries. One type of zero-dimensional diagram object is the diamond.

$$Diamond(x) \supset 0DObject(x)$$

One-dimensional diagram objects include curves ($Curve$). Three important kinds of curves are lines, rays (half-lines), and line segments.

$$LineSegment(x) \supset Curve(x)$$

A line has no beginning or end. A ray has a unique beginning but no end. A line segment has both a unique beginning and a unique end.

$$LineSegment(x) \supset (\exists! p_1, p_2)[begins(p_1, x) \land ends(p_2, x)]$$

Beginnings and ends are points, in the sense described below. It will be useful to have a term $Linear$ that covers all three types of linear diagram objects.

$$Linear(x) \equiv [Line(x) \lor Ray(x) \lor LineSegment(x)]$$

A line segment "in" a linear diagram object is one that is wholly contained in it.

---

[3]I will write only one axiom where there is a set of similar axioms. The reader should have no difficulty reconstructing the others.

$$lineSegmentIn(x, y)$$
$$\equiv LineSegment(x) \wedge Linear(y) \wedge (\forall p)[pointIn(p, x) \supset pointIn(p, y)]$$

Another kind of curve is an arrow from one point to another.

$$arrow(x, p_1, p_2) \supset Curve(x)$$

Diagrams are composite entities whose components are diagram objects.

$$Diagram(d) \wedge componentOf(x, d) \supset DiagramObject(x)$$

6.3 Points and the *at* Relation

A ground consists of *points* and any diagram object consists of points, in some loose sense of "consist of"; that is, for any ground and any diagram object there is a corresponding set of points.

$$[Ground(x) \vee DiagramObject(x)] \supset (\exists s)(\forall p)[member(p, s) \supset pointIn(p, x)]$$

A zero-dimensional object has exactly one point in it.

$$0DObject(x) \supset (\exists!p)pointIn(p, x)$$

For convenience we will say that the single point in a zero-dimensional object both begins and ends it. Points are not diagram objects.

The beginnings and ends of linear objects are points.

$$begins(p, x) \supset Linear(x) \wedge pointIn(p, x)$$

Points in the ground are partially ordered by an *above* relation and a *rightOf* relation.

$$above(p_1, p_2, g) \supset Ground(g) \wedge pointIn(p_1, g) \wedge pointIn(p_2, g)$$

A linear object is horizontal if no point in it is above any other. Similarly, vertical.

$$horizontal(x, g)$$
$$\equiv Linear(x) \wedge \neg(\exists p_1, p_2)[pointIn(p_1, x) \wedge pointIn(p_1, x) \wedge above(p_1, p_2, g)]$$

A horizontal ray all of whose points are to the right of its beginning is a rightward positive ray.

$$[ray(x) \wedge horizontal(x, g) \wedge begins(p_0, x)$$
$$\wedge (\forall p)[pointIn(p, x) \supset [p = p_0 \vee rightOf(p, p_0, g)]]$$
$$\supset rtPositive(x, g)$$

Similarly, a vertical ray all of whose points are above its beginning is a upwardly positive ray (*upPositive*). A vertical ray all of whose points are below its beginning is a downwardly positive ray (*dnPositive*).

A special kind of line segment needed for Gantt charts is a horizontal bar.

$$HBar(x) \supset (\exists g)[horizontal(x, g) \wedge LineSegment(x)]$$

When realized spatially, it will generally be thicker than other line segments.

Diagrams are constructed by placing points in diagram objects at points in the ground or in another diagram object. The *at* relation expresses this.

$$at(p_1, p_2)$$
$$\supset (\exists x_1, x_2)[pointIn(p_1, x_1) \wedge pointIn(p_2, x_2) \wedge DiagramObject(x_1)$$
$$\wedge [Ground(x_2) \vee DiagramObject(x_2)] \wedge x_1 \neq x_2]$$

The relation *at* can be extended to zero-dimensional objects in the obvious way. Typically, frameworks (see below) will be placed with respect to some points in the ground, and other diagram objects will be placed with respect to the framework or other diagram objects.

The relations of a diagram as a composite entity include its *at* relations. To say this formally we can reify the *at* relation. Thus, $at'(r, p_1, p_2)$ means that r is the *at* relation between p_1 and p_2. We can then say that r is a member of the relations of the diagram.

$$at'(r, p_1, p_2) \wedge relationOf(r, d)$$

6.4 Frameworks

Many diagrams have an underlying framework with respect to which diagram objects are then located, e.g., the lat-long framework on maps. A framework is a set of objects in a particular relationship to each other.

$$Framework(f)$$
$$\supset (\exists s)(\forall x)[member(x, s) \supset DiagramObject(x) \wedge componentOf(x, f)]$$

One very important kind of framework is a coordinate system. Here I will characterize only a rectilinear cooordinate system.

$$RCoordinateSystem(f) \supset Framework(f)$$

$$RCoordinateSystem(f) \supset (\exists g)[Ground(g) \wedge groundFor(g, f)]$$

$$RCoordinateSystem(f) \wedge groundFor(g, f)$$
$$\supset (\exists x, y)[xAxisOf(x, f) \wedge yAxisOf(y, f) \wedge rtPositive(x, g)$$
$$\wedge [upPositive(y, g) \vee dnPositive(y, g)]]$$

Two points have the same x-coordinate if there is a vertical line that contains both of them.

$$sameX(p_1, p_2, f)$$
$$\equiv (\exists l, g)[groundFor(g, f) \wedge vertical(l, g) \wedge pointIn(p_1, l) \wedge pointIn(p_2, l)]$$

Similarly for the same y-coordinate.

The x-value of a point is a point in the x-axis with the same x-coordinate.

$$x\text{-}value(p_1, p, f) \equiv (\exists x)[sameX(p_1, p, f) \wedge pointIn(p_1, x) \wedge xAxisOf(x, f)]$$

Similarly for the y-value.

6.5 Meanings

Associated with every object in a diagram is its meaning. Meaning for diagrams is thus a function mapping diagram objects into entities provided by some other ontology. Meaning is conveyed by the predication $means(x, c)$ introduced above, where x is a diagram object. There are no constraints on the second argument of $means$; it just has to be an entity in some ontology.

$$DiagramObject(x) \supset (\exists c)means(x, c)$$

The meanings of the at relations in a diagram will be specified with axioms having the following form:

$$at(x, y) \wedge p(x) \wedge q(y) \wedge means(x, a) \wedge means(y, b) \supset r(a, b)$$

That is, if a p-type diagram object x is at a q-type diagram object y in a diagram, then if x means a and y means b, then there is an r relation between a and b.

Axes in a coordinate system generally mean some set in another ontology. That set may be unordered (a set of tasks), discrete and linearly ordered (months), or continuous (time).

6.6 Labels

A label is a textual object that can be associated with objects in a diagram. The two basic facts about labels cannot be defined with precision without making reference to the cognition of the reader of the diagram.

1. A label is placed near the object it labels, in a way that allows the reader of the diagram to uniquely identify that object.

2. The content of the label as a string bears some relation to the meaning of the object that it labels, in that perceiving the string causes one to think of the meaning.

Specifying the first of these completely is a very hard technical problem (Edmondson et al., 1997). For example, often on a map one cannot correctly associate the name of a town with a dot on the map without doing the same for all nearby towns, and associating a curve on a map with the name of a road often requires abductive inferences about shortest paths and consistency of line thickness. Here I will simply say that a label can be placed at an object, and leave it to component-specific computation to determine what at means in some context.

$$label(l, x) \supset string(l) \wedge DiagramObject(x) \wedge at(l, x)$$

The second property of labels is also a difficult technical, or even artistic, problem. But a very common subcase is where the label is a name. The whole purpose of a name is to cause one to think of the object when one perceives the name, so it serves well for this property of labels.

$$label(l, x) \supset (\exists c)[means(l, c) \wedge means(x, c)]$$

6.7 Gantt Charts

A Gantt chart g for a project p is a diagram that consists of several types of components. It has a rectilinear coordinate system f where the x-axis is rightward positive and the y-axis is upward or downward positive. (The x-axis can appear at the top or the bottom of the chart.) The meaning of the x-axis is the time line or some other periodic temporal aggregate, and the meaning of the y-axis is a set of tasks and milestones in the project.

$$GanttChart(g, p)$$
$$\supset Diagram(g) \wedge Project(p)$$
$$\wedge (\exists f, x, y, t, s)[RCoordinateSystem(f) \wedge componentOf(f, g) \wedge xAxisOf(x, f)$$
$$\wedge rtPositive(x) \wedge means(x, t) \wedge TimeLine(t)$$
$$\wedge yAxisOf(y, f) \wedge [upPositive(y) \vee dnPositive(y)]$$
$$\wedge means(y, s)$$
$$\wedge (\forall z)[member(z, s) \supset [taskIn(z, p) \vee milestoneIn(z, p)]]]$$

A Gantt chart has horizontal bars representing the interval during which a task is executed.

$$GanttChart(g, p) \wedge RCoordinateSystem(f) \wedge componentOf(f, g)$$
$$\supset (\exists s)(\forall b)[member(b, s)$$
$$\supset componentOf(b, g) \wedge HBar(b)$$
$$\wedge (\exists r_1, z, p_1, t_1, q_2, t_2)[y\text{-}value(r_1, b, f) \wedge means(r_1, z)$$
$$\wedge taskIn(z, p) \wedge x\text{-}value(p_1, beginningOf(b), f)$$
$$\wedge means(p_1, t_1) \wedge begins(t_1, z) \wedge x\text{-}value(q_1, endOf(b), f)$$
$$\wedge means(q_1, t_2) \wedge ends(t_2, z)]]$$

Because a task is an event, OWL-Time allows instants as the beginnings and ends of tasks. This axiom says that a Gantt chart has a set of components which are horizontal bars representing tasks and the beginning of the bar represents the starting time of the task and the end of the bar represents the finishing time of the task.

A similar axiom says that a Gantt chart has diamonds representing milestones. We can call bars and diamonds "task icons".

$$taskIcon(x) \equiv [HBar(x) \vee Diamond(x)]$$

A Gantt chart often has arrows going from the end of one bar to the beginning of another, indicating the the first bar's task is a prerequisite for the second bar's task. A diamond can also be the source and/or target of an arrow.

$$GanttChart(g, p) \wedge RCoordinateSystem(f) \wedge componentOf(f, g)$$
$$\supset (\exists s)(\forall a)[member(a, s)$$
$$\supset componentOf(a, g)$$
$$\wedge (\exists x, z_1, p_1, y, z_2, p_2)[arrow(a, p_1, p_2) \wedge taskIcon(x)$$
$$\wedge componentOf(x, g) \wedge means(x, z_1) \wedge ends(p_1, x)$$
$$\wedge taskIcon(y) \wedge componentOf(y, g) \wedge means(y, z_2)$$
$$\wedge begins(p_2, y) \wedge enables(z_1, z_2)]]$$

A bar in a Gantt chart may have labels for the date at its beginning and end.

$$GanttChart(g, p) \wedge HBar(b) \wedge ComponentOf(b, g)$$
$$\supset [(\exists s_1)[(\forall l_1)[member(l_1, s_1)$$
$$\equiv (\exists p_1, q_1, t_1)[begins(p_1, b) \wedge label(l_1, p_1) \wedge \textit{x-value}(q_1, p_1)$$
$$\wedge means(q_1, t_1) \wedge dateStringFor(l_1, t_1)]]$$
$$\wedge card(s_1) < 2]$$
$$\wedge (\exists s_2)[(\forall l_2)[member(l_2, s_2)$$
$$\equiv (\exists p_2, q_2, t_2)[ends(p_2, b) \wedge label(l_2, p_2) \wedge \textit{x-value}(q_2, p_2)$$
$$\wedge means(q_2, t_2) \wedge dateStringFor(l_2, t_2)]]$$
$$\wedge card(s_2) < 2]]$$

The cardinality statement is a way of saying there is either zero or one label. Similarly, a diamond in a Gantt chart may have a label for a date.

Points on the y-axis of a Gantt chart can be labelled with task names.

$$[GanttChart(g, p) \wedge RCoordinateSystem(f) \wedge componentOf(f, g) \wedge yAxisOf(y, f)]$$
$$\supset (\exists s)[(\forall l)[member(l, s)$$
$$\equiv (\exists p, z)[pointIn(p, y) \wedge label(l, p) \wedge means(p, z)$$
$$\wedge l = name(z)]]$$
$$\wedge card(s) < 2]$$

Similarly, points in the x-axis of a Gantt chart can be labelled with dates or times, and line segments in the x-axis of a Gantt chart can be labelled with calendar intervals. Further elaborations are possible. The labels can have internal structure. For example, labels for subtasks may be indented. Labels for time intervals may be broken into a line for months, a line below for weeks, and so on.

7   Modalities, Media, and Manifestations

In order for communication to work, perception of the symbol must occur. Humans are able to perceive optical, acoustic, and chemical phenomena, as well as pressure and temperature. Of these modalities the optical and acoustic are by far the most important, because they offer the richest possibilities for composition. Artifact agents of course have other modalities.

Communication requires one or more devices. There must be a manner in which the presentation is carried out. Allwood (2002) categorizes these into primary, secondary, and tertiary. The primary devices or media are the ones that are human body parts and processes. The voice is used for grunts, speech and song. The hands, arms, body, face and head are used for gesture. Even at this level some encoding must be done; we need to find words for the concepts we wish to convey, and these must be mapped into sequences of articulatory gestures.

The secondary media are those involving devices external to the human body, such as marks on paper as in writing and drawings. computer terminals, telephones, videotape, and so on. These typically require multiple encodings, where the final code is known to the intended audience. The various media have different advantages and disadvantages that can be exploited for different kinds of represented content. For example, visual spatial representations can exploit more dimensions for conveying relationships than can auditory temporal representations. Hovy and Arens (1990) catalog many of these correlations.

Allwood also mentions tertiary media, including paintings, sculptures, and aesthetic designs of artifacts such as chairs. My feeling is that these are just secondary media where the content that is represented is much more difficult to capture in words.

We have a strong tendency to group together classes of symbolic entities that share the same property, especially their content, and think of them as individuals in their own right. It is probably better in an ontology of symbolic entities to view these as first-class individuals that themselves represent a particular content. Other symbolic entities may be manifestations of these individuals. The predicate *manifest* is a transitive relation whose principal property is that it preserves content.

$$manifest(x_1, x) \land means(x, c, s) \supset means(x_1, c, s)$$

(This does not take into account translations, where the s's differ.)

Thus, to use the example of Pease and Niles (2001), there is an entity called *Hamlet*. The performance of *Hamlet* manifests *Hamlet*. The performance of *Hamlet* in a particular season by a particular company manifests that, and a performance on a particular night may manifest that. A videotape of that particular performance manifests the performance, and every copy of that videotape manifests the videotape. A similar story can be told about the script of the play, a particular edition of the script, and a particular physical book with that edition of the script as its content.

The above rule should be thought of as defeasible, because variations exist, lines can be dropped, and printer's errors occur. More precisely, if some proposition occurs in the content of one symbolic entity then defeasibly it occurs in the content of symbolic entities that manifest it.

8 Summary

Information structure is one of the most basic domains in an ontology of the everyday world, along with such domains as space and time. It should be anchored in an ontology of commonsense psychology, as I have tried to sketch here, and there should be an account of how complex symbolic entities can be composed out of simpler symbolic entities in various modalities and combinations of modalities. An example has been given of how the structure and meaning of a moderately complex type of diagram can be specified. Because Web pages are an especially complex sort of symbolic entity, we can expect an ontology of information structure to be very significant in the development of the Semantic Web and other applications requiring communication of complex semantic content.

Acknowledgements

References

[1] Allwood, Jens. 2002. "Bodily Communication—Dimensions of Expression and Content", in B. Grandström, D. House, and I. Karlsson (Eds.), *Multimodality in Language and Speech Systems*, Kluwer Academic Publishers. Dordrecht, The Netherlands.

[2] Edmondson, Shawn, Jon Christensen, Joe Marks, and Stuart M. Shieber. (1997) "A General Cartographic Labeling Algorithm", *Cartographica*, Vol. 33, No. 4, pp. 13-23, Winter 1997.

[3] Gordon, Andrew, and Jerry R. Hobbs, 2003. "Coverage and Competency in Formal Theories: A Commonsense Theory of Memory", *Proceedings*, AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, Stanford, California, March 2003.

[4] Grice, Paul, 1989. "Meaning", in *Studies in the Way of Words*, Harvard University Press, Cambridge, Massachusetts, 1989.

[5] Hobbs, Jerry R. 1985a. "Ontological Promiscuity." *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 61-69. Chicago, Illinois, July 1985.

[6] Hobbs, Jerry R., 1985. "On the Coherence and Structure of Discourse", Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University.

[7] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", *Artificial Intelligence*, Vol. 63, Nos. 1-2, pp. 69-142.

[8] Hobbs, Jerry R., 1998. "The Syntax of English in an Abductive Framework", available at `http://www.isi.edu/ hobbs/disinf-chap4.pdf`

[9] Hobbs, Jerry R., 2001. "Causality", *Proceedings*, Common Sense 2001, Fifth Symposium on Logical Formalizations of Commonsense Reasoning, pp. 145-155, New York University, New York, New York, May 2001.

[10] Hobbs, Jerry R., William Croft, Todd Davies, Douglas Edwards, and Kenneth Laws, 1986. "Commonsense Metaphysics and Lexical Semantics", *Proceedings, 24th Annual Meeting of the Association for Computational Linguistics*, New York, June 1986., pp. 231-240.

[11] Hobbs, Jerry R. and Feng Pan, 2004. "An Ontology of Time for the Semantic Web", *ACM Transactions on Asian Language Information Processing*, Vol. 3, No. 1, March 2004, pp. 66-85.

[12] Hovy, Eduard H. and Yigal Arens. 1990. "When is a Picture Worth a Thousand Words?—Allocation of Modalities in Multimedia Communication", AAAI Spring Symposium on Human-Computer Interactions. Palo Alto, CA.

[13] McCarthy, John, 1980. "Circumscription: A Form of Nonmonotonic Reasoning", *Artificial Intelligence*, Vol. 13, pp. 27-39. Reprinted in M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning*, pp. 145-152, Morgan Kaufmann Publishers, Inc., Los Altos, California.

[14] McNeil, David, ed. 2000. *Language and Gesture*, Cambridge University Press, Cambridge, England.

[15] Pease, Adam, and Ian Niles. 2001. "Practical Semiotics: A Formal Theory", available at `citeseer.nj.nec.com/531589.html`.

[16] Pineda, Luis, and Gabriela Garza, 2000. "A Model for Multimodal Reference Resolution", *Computational Linguistics*, Vol. 26, No. 2, June 2000, pp. 139-194.

[17] Pollard, Carl, and Ivan A. Sag, 1994. *Head-Driven Phrase Structure Grammar*, University of Chicago Press and CSLI Publications.

[18] Power, Richard, Donia Scott, and Nadjet Bouayad-Agha, 2003. "Document Structure", *Computational Linguistics*, Vol. 29, No. 4. pp. 211 - 260.

[19] Wahlster, Wolfgang, Elisabeth André, Wolfgang Finkler, Hans-Jürgen Profitlich, and Thomas Rist, 1993. "Plan-based Integration of Natural Language and Graphics Generation", *Artificial Intelligence*, Vol. 63, Nos. 1-2, pp. 387-427.