DISCOURSE AND MEANING

PAPERS IN HONOR OF EVA HAJIČOVÁ

Edited by

BARBARA H. PARTEE

PETR SGALL

Offprint

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

This is an offprint from:

BARBARA H. PARTEE and PETR SGALL
Discourse and Meaning
John Benjamins Publishing Company
Amsterdam/Philadelphia
1996

ISBN 90 272 2146 4 (Eur.) / 1-55619-499-4 (US)

Your metaphor or mine: Belief ascription and metaphor interpretation

Yorick Wilks, John Barnden and Jin Wang

ViewGen, an algorithm and program for belief ascription, represents the beliefs of agents as explicit, partitioned proposition-sets known as environments. A way of extending *ViewGen* to the interpretation of metaphor, and in particular to the comprehension of metaphor within the belief spaces of particular agents, has been described elsewhere. The paper reports the recent implementation of this extension, as well as summarizing the argument for the claim that ordinary non-metaphorical belief ascription and the transfer of information in metaphors can both be seen as different manifestations of a single environment-amalgamation process, one in which explicitly metaphorical amalgamations are triggered by "preference breaking" in the sentence being processed. This requires a consideration of the scoping of metaphor with respect to belief contexts, analogous to the scoping of quantification and definite descriptions with respect to such contexts. The approach of assimilating metaphor to belief ascription is contrasted with Hobbs' attempt to assimilate metaphor interpretation to abduction. As a topic of ongoing and future work, the issue of mixed metaphor, of two distinct types, is briefly addressed.

1. ViewGen: The basic belief engine

A computational model of belief ascription is described in detail elsewhere (Wilks and Bień, 1979, 1983; Ballim, 1987; Wilks and Ballim, 1987; Ballim and Wilks, 1992) and is embodied in a prolog program called *ViewGen*. The basic algorithm of this model uses the notion of default reasoning to ascribe beliefs to other agents unless there is evidence to prevent the ascription.

Perrault (1987, 1990) and Cohen and Levesque (1985) have also recently explored a belief and speech act logic based on a single explicit default axiom. As our previous work has shown for some years, the default

ascription is basically correct, but the phenomena are more complex than are normally captured by an axiomatic approach.

ViewGen also avoids certain counter-intuitive assumptions, such as the *non-persistence* of ignorance about any given proposition p (see Perrault 1990). Also such systems avoid any individual-dependent criteria for ascription, such as the individual expertise notions in *ViewGen* (see below).

ViewGen's belief space is divided into a number of explicit, topic-specific partitions, called *topic environments*. *ViewGen* also generates a type of environment known as a *viewpoint*. A viewpoint consists of some person's beliefs about some topics, parcelled up into topic environments, – within *ViewGen*, – all beliefs are ultimately beliefs held by the system (e.g. the system's belief about France, what the system believes John believes about cars, etc.) and so, trivially, lie within the system's viewpoint.

The system's view of some topic (say, atoms) is pictorially represented as in Figure 1.

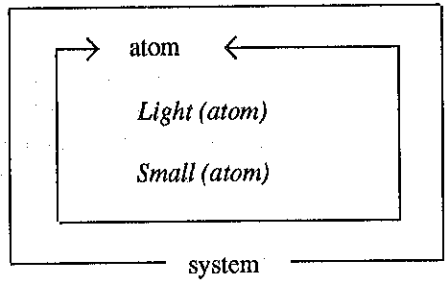

Figure 1.

The diagram contains two types of environments: First, there is the box labelled with "system" at the bottom. This is a "believe environment" or "viewpoint". Viewpoints contain topic environments, such as the box labelled with "atom" at the top. A topic environment contains a group of propositions about the topic. So, for example, the above diagram conveys that the system believes that atoms are light and small. *ViewGen*'s own "knowledge-base" is a viewpoint containing a large number of topic environments.

If the topic of a topic environment is a person, the topic may contain, in addition to the beliefs about the person, a viewpoint environment containing particular beliefs held by that person about various topics. Normally and for obvious reasons for efficiency, this is only done for those beliefs of a given person that are, as some would put it, reportable, where that will often mean beliefs that conflict with those of the system itself. For example, suppose the system had beliefs about a person called John who believes that the Earth is flat. This would be pictorially represented as in Fig.2.

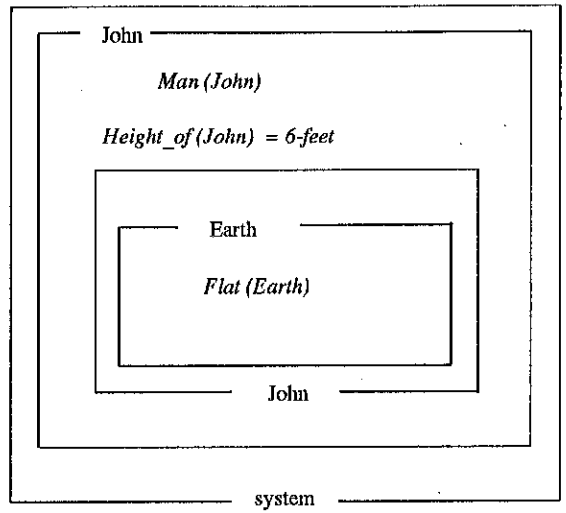

Figure 2. *The organization of beliefs about and of John*

The John viewpoint, shown as the box with "John" on the lower edge, is a *nested viewpoint*, as it is enclosed within the system viewpoint shown (through an intervening topic environment about John, shown as the box with "John" on its upper edge).

Environments are dynamically created and altered. The basic algorithm of interest in this paper is an *amalgamation* mechanism that ascribes beliefs from one viewpoint to another (or, "pushing one environment down into another"); ascribing certain beliefs, transforming some, and blocking the ascription of others. The simplest form of this algorithm, described in Wilks and Bień (1979, 1983), is that a viewpoint should be generated using a default rule for ascription of beliefs. The default ascriptional rule is to assume that another person's view of a topic is the same as one's own *except where there is explicit evidence to the contrary*. In examples of this

sort, where the topic is also the agent into whose environment an ascription is being attempted (i.e. replace "Earth" by "John" in the example), propositions in an outer topic environment E (for the topic John, in the example), are *pushed inwards* into a topic environment (for the same topic) within a believer viewpoint (John's) nested within E. Such inward pushing is central to our later observation on metaphor.

Belief ascription is a far more complex phenomenon than is shown in this brief summary and the key to our method is the delimitation and treatment of cases where the default algorithm is incorrect. We call this *atypical beliefs* and they include technical expertise, self-knowledge (itself a form of expertise), and secrets. For example, beliefs that I have about myself, such as how many fillings I have in my teeth, are beliefs that I would not normally ascribe to someone else *unless I had a reason to do so* (if, say, the person to whom I was ascribing the belief was my dentist). A representation based on lambda expressions is used in dealing with atypical beliefs, and is described elsewhere (Ballim, 1987; Ballim and Wilks, in press; Wilks and Ballim, 1987), and follows a suggestion originally made by MacCarthy and Hayes (1969). This combination of a basic default ascription rule with a mechanism for dealing with atypical belief is an original algorithm and has not, to our knowledge, been described or tested elsewhere in the literature.

2. Metaphor: Shifting the belief engine to a higher gear

Metaphor is normally explicated, formally or computationally, by a process that transfers properties by some structural mapping from one structure (the vehicle) to another (the tenor). Classic examples in AI would be the work of Falkenhainer, Forbus and Gentner (1989) and of Indurkhya (1987). All these authors are concerned, as we are, with metaphor and analogy viewed as some form of structural mapping; the difference from them of what we offer here is the linkage between that process and the process of belief ascription (and also that of "intensional identification", as explained in Ballim, Wilks and Barnden, in press).

We want to explore the possibility of applying our basic belief algorithm to metaphor, as an experiment to see if it gives insight into the phenomenon. That should not be as surprising as it may sound: metaphor has often been viewed, in traditional approaches, as "seeing one thing as

something else", a matter of viewpoints, just as we are presenting belief. We propose that propositions in the topic environment for the vehicle of a metaphor be "pushed inward" (using the standard algorithm mentioned above), into an embedded environment for the tenor, to get the tenor seen through the vehicle, or the view of the tenor-as-vehicle.

The key features here are: (1) one of the conceptual domains, namely the metaphor vehicle, is viewed as a "pseudo-believer"; (2) the pseudo-believer has a metaphorical view of a topic or domain; (3) the generation of such a view is not dissimilar from ascribing beliefs to real believers; (4) explicating this by pushing or amalgamating environments yields new intensional entities after an actual transfer of properties. So, in

Jones threatened Smith's theory by reimplementing his experiments

we would know we had a preference-breaking, and potentially metaphorical, situation from the object-feature failure on "threaten" (assuming this expects a person object), at least if we accept the argument of Wilks (1977) that metaphors could be identified, procedurally at least, with the class of preference-breaking utterances. (This includes assertions that violate class relationships, such as "An atom is a billiard ball"). The awkward cases for that broad delimitation are forms like "Connors kills McEnroe", which breaks no verb preference but is read metaphorically by some as "beat soundly at tennis". Here one might consider taking the classic Marcus-escape and use our procedural definition to rule this example out of court as a "garden path metaphor".

We could now plausibly form a metaphoric view of theory-as-person using the environment-amalgamation process sketched above. Figure 3 shows possible system environments for theory and person, and the resulting theory-as-person environment, where the arrow indicates the new environment resulting from the application of the default rule when the properties of the outer environment (for Person) are amalgamated into the inner environment (for Theory) and survive unless contradicted (as the Concrete predicate in fact is).

So, by this manoeuvre, a new and complex metaphorical property of theories is derived. It might be, of course, that this procedure of belief-overriding as a basis for metaphor would produce no different a set of plausible properties transferred than any other system (e.g. that of Falkenhainer et al.); and that would be, again, an experimental question. But its

importance or originality would lie in the fact that it was further application of an algorithm designed to explicate another phenomenon altogether (i.e., belief), and therefore yield a procedural connection between the notions, as we argued in detail in Ballim, Wilks and Barnden (1992).

There is a further interesting aspect to the connection between belief and metaphor. We have stressed a procedural connection that may seem improbable to some people. There is also the important but neglected phenomenon that the content of belief is often inherently metaphorical, and in a way that conventional theorists totally neglect by their concentration on simplistic belief examples like "John loves Mary". A far more plausible candidate might be a truth such as:

Prussia threatened France before invading it successfully in 1871.

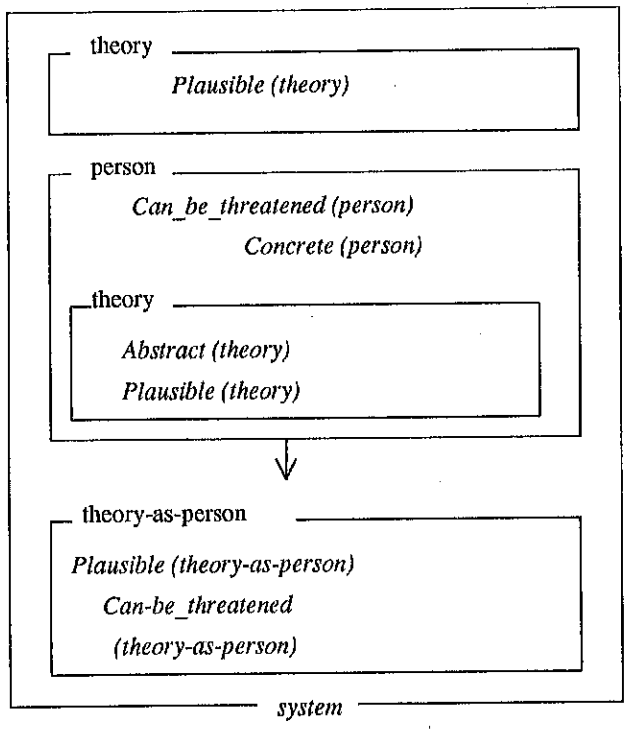

Figure 3. *Forming a theory-as-person environment*

What are we to say of this historically correct belief? What are the entities referred to by "Prussia" and "France"? Simple translation into some first-order expression like *Invade (Prussia, France, 1871)* just obscures the real problem, one for which the semantics of first order logic are no help at all.

Are the entities referred to somehow metaphorically the Prussian people, etc., or army, or a part of the army?

Following the approach described earlier, we might expect to detect breaking of linguistic preferences of the verb "threaten"; perform a trial pushdown of properties of the "People" environment (given by the conventional preferences of "threaten") into an environment for Prussia (= a land mass, the basic representation). An important safeguard, that there is no space to discuss here, would be that we examined our inventory of representations to see if we had one for "Prussia" that already expressed the (dead) metaphor of a country-name-as-a-polity (some would insist that this was a metonymy, but we do not address this alternative here).

The amalgamation of the notions of belief ascription and metaphor interpretation is described and justified in detail in Ballim, Wilks and Barnden (in press). The method as described there explicitly addresses the case of metaphor only at the top level of discourse, within the system's own notions of the component elements of the metaphor. But of course, in the general process of metaphor interpretation, we must relativize these processes to the space of any relevant believer. The system might believe the zoological truism that pigs are clean and neat, but when talking to the non-zoologist John, use the standard metaphor "He's a pig" on the assumption that John would perform the ascription against the dirty-unhygienic-pig belief that the system believes John to hold. The system must be able to model *that* metaphorical ascription, whose elements the system itself does not believe.

This relativization is included in a recent implementation of the approach within *ViewGen* by Ballim, as extended by one of us (Wang). The need for relativization underscores the benefits of our method of unifying metaphorical transference with belief ascription. We shall analyze such examples later in more detail, but we must first discuss an issue of "metaphorical scope" that is implicitly raised by the above.

3. Metaphorical scope

Consider the sentence:

John believes that a cure for terrorism is needed.

The complement of this belief report – the clause following the word "that" – can be construed as involving a terrorism-as-disease metaphor. We must first realize that there is a *metaphor scoping* issue here. The "inner scope" reading involves the idea that John himself thinks of terrorism as a disease (and we might predict that he would report his belief by means of the sentence "A cure for terrorism is needed"). The "outer scope" reading is that John believes something about terrorism that is being portrayed *by the speaker* in terms of disease-curing. John does not necessarily have a belief couched in these terms, nor would he necessarily report his belief these terms. (Perhaps John would say: "Something needs to be done to eliminate terrorism and repair the damage it has done to society".)

We now concentrate on the inner scope reading. In the process of setting up a topic environment for terrorism inside John's belief environment, the system needs to perform (i) the normal default ascription process, here moving some of its own beliefs about terrorism down to John, *as well as* performing (ii) a metaphoric transference process. This combination of tasks is simplified by our method of dressing the tasks in essentially the same algorithmic clothes.

This is especially so since it is *John's* view of disease that is important in the metaphoric transference, not the system's. After all, John might believe that diseases are caused by demonic influences, say, and this belief could affect what he would think of as reasonable ways of curing terrorism – e.g., exorcism. Thus, in task (ii), a disease topic environment must be set up within John's belief environment, and it is that topic environment that acts as the source of the metaphoric transference. The setting up of John's disease topic environment will itself, moreover, normally involve the default ascription process, moving system beliefs about disease down into John. Thus, belief ascription and metaphoric transference are tightly bound up with one another.

The case for integrating belief ascription and metaphor is yet further strengthened by bringing in the possibility that John actually believes that terrorism is *literally* a disease. (A more realistic example for the present phase of the argument could be obtained by replacing "terrorism" by "laziness", say). We could cast this as an *intensional identification* by John of the concepts of terrorism and disease. We show in Ballim, Wilks and Barnden (1992) that such identifications can themselves be handled by the same environment manipulations as are used for metaphor, although we concentrate there on the identification of descriptions of ordinary objects, such as when John takes *Jim's father* and *Frank* to be the same person whereas the system does not. Hence, we form a unity out of the trinity of belief ascription, metaphor and intensional identification. The important point for present purposes is that in interpreting "John believes that a cure for terrorism is needed" the system can abstain from making a decision about whether John *literally* or only *metaphorically* thinks of terrorism as a disease, since the same overall belief ascription process is used in either case.

One special case of the issue of a metaphor being nested in belief report complements is when the metaphor is itself about being belief or other mental states. An example of this is "John believes that one part of Mary wanted to go skiing, but another part wants to sit by the fire" where the metaphor of the mind as compartmented into sub-minds or sub-people is being used. This issue is discussed at length in Barnden (1989, 1990).

The distinction between inner and outer scope readings of metaphors within belief report complements is analogous to a classic distinction usually made in the case of object descriptions within such complements. Consider the description "Mary's uncle" in "John believes that Mary's uncle drinks gin." The description could be taken to one which is in some form directly used in John's belief (so that John might report his belief by "Mary's uncle drinks gin"). This is the inner scope (or "de dicto") reading. Alternatively, the description could merely be taken to one that is used by the speaker to define the person that John believes to drink gin; John might be thinking of the person in some completely different way. This is the outer scope (or "de re") reading.

Henceforth in this paper we consider only the inner-scope interpretation of embedded metaphor. The outer-scope case can be dealt with in a similar but simpler way.

4. Embedded metaphor ascription

In the following, the system is considering the beliefs of an agent M, with no intervening agents. Let us say that M's "peculiar" topic environment for T contains the beliefs about T that are recorded in M's space *before* an attempt to push the system's T beliefs down into M has been made. Let us

say that M's "expanded" topic environment for T contains the beliefs about T that are recorded in M's space after the attempt.

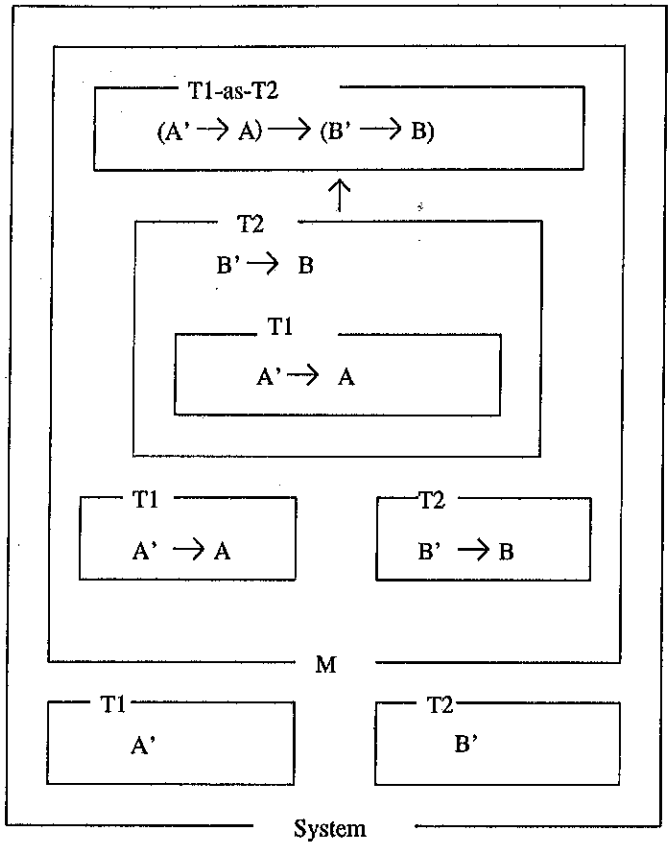

Figure 4a. Embedded metaphor ascription: (A'→ A) → (B'→ B)

Supposing T1, T2 are two topics, there are two ways the system could proceed when it takes M to see T1 as T2. The first is for the system to form M's expanded topic environments for T1 and for T2 separately, and then try to do the metaphor transference by pushing the expanded T2 beliefs down into the expanded T1 environment.

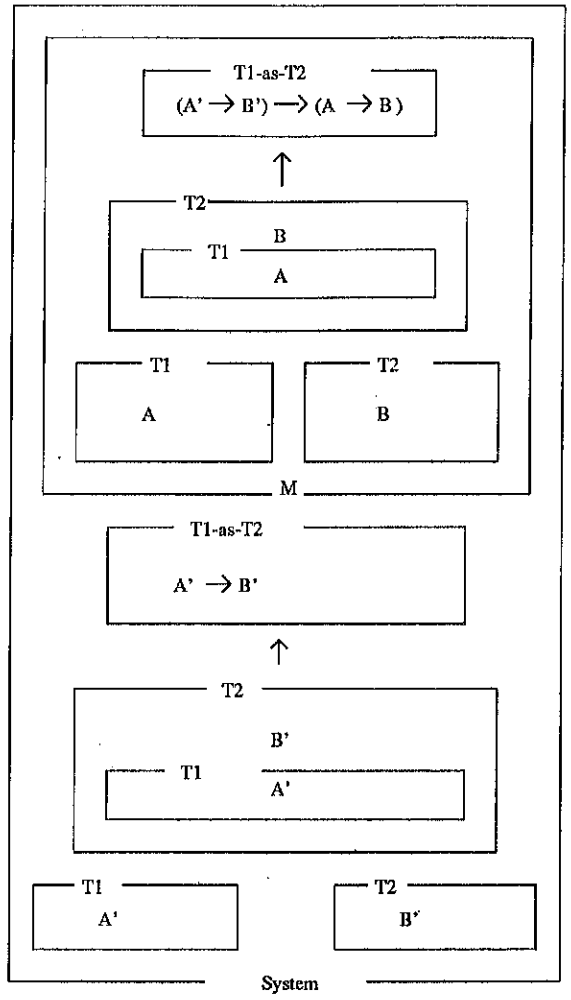

Figure 4b. Embedded metaphor ascription: (b) (A'→ B') → (A → B).

The second way is, (b) essentially, to do these steps in the other order: the system forms its own T1-as-T2 environment by one metaphorical transference, forms a *peculiar* T1-as-T2 environment within M's space by another metaphorical transference (by trying to push M's peculiar T2

beliefs into M's peculiar T1 environment), and then trying to push the system's T1-as-T2 beliefs down into M's peculiar T1-as-T2 environment.

These two possibilities can be described as $(A' \rightarrow A) \rightarrow (B' \rightarrow B)$ and $(A' \rightarrow B') \rightarrow (A \rightarrow B)$, where A', B' are the system's sets of beliefs about T2 and T1, A, B are M's peculiar sets of beliefs about T2 and T1, and $\rightarrow$ is the operation of trying to push beliefs in (the normal default ascription process). The two different processes and results are shown in Figure 4. It is clear that the two results are basically the same, but that a difference can arise from the propositions which are ascribed either from A or from B', since the earlier ascribed propositions may prevent the later ascription. Currently it seems to us that the first process, namely $(A' \rightarrow A) \rightarrow (B' \rightarrow B)$, is the correct one.

We can now consider the general process of the triggering of metaphor and the amalgamation processes that follow. Let us return to the rather tired example of Wilks (1977):

My car drinks gasoline

In that paper it was argued that this metaphor was detected or triggered by the preference failure of "drink" (not getting an animate agent) and that the metaphor was to be understood by matching into an appropriate knowledge base for Car, where some such form was found as that "Cars use gasoline". This location of "use" (rather than "leak", say) as being semantically close to "drink" resulted in a literal reading of "Cars use gasoline". The algorithm (not given in any detail here) was implemented later by Modiano. It was also argued that two metaphorical perspectives were available: this one, and one of Cars seen as Animate, from the failed preference of the verb.

If we now generalize this class of example within the belief ascription approach to metaphor of this paper, we can achieve a much more general account, one in which there is only a single perspective, although this can appear, as we noted, at any level of belief.

Suppose we again take the failed preference as triggering the metaphorical belief amalgamation, within the system's own level of belief initially. (The failure assumption is defended in section 5 below). We now access the system's belief environments for the concepts that fail to prefer each other (Car and Drink). See Figure 5 for a sketch of the propositional content. The drink environment shows that the preferred agent of the verb

is Animate, so the environment for Animate and Car are brought into focus in the system's main environment, and the metaphorical amalgamation for Car-as-animate is performed just as for the Theory-as-person case above.

We see that "Car-as-animate use gasoline" appears in the resulting intensional environment for Car-as-animate and, when that belief is present at any level of belief ascription (done here for the system's beliefs only) we see that that proposition will enter uncontradicted into the belief space of the individual who hears and is engaged in understanding "Cars drink gasoline". We assume in saying this that Cars and Cars-as-X can be identified for temporary inference purposes, and that other (so far unmentioned) ascribed beliefs about Cars-as-animate (such as they leak gasoline) will not be amalgamated into the environment of the believer of "Cars drink gasoline" since drink/leak will (hopefully) contradict each other as any such small/large, dark/light pair will be found to do. Thus the appropriate literal interpretation will enter the relevant belief environment through a single process of belief ascription.

5. Other work: Abduction is different

Hobbs (1990) has put forward a general theory of understanding and interpretation, making use of Peirce's notion of abduction: the process, purportedly a part of scientific reasoning, of explaining propositions by proving them true from assumptions you locate or construct – i.e. given b, find some a such that $a \rightarrow b$. For Hobbs, this is the basis of an *interpretation theory* for any such proposition b, so that any b is interpreted by being proved from some set of assumptions a.

One mode of general resistance to Hobbs' view starts from the observation that abduction comes from, and belongs to, a scientific and mathematical tradition of *explaining* propositions by proving them true from assumptions you locate or construct. On our view, it requires substantial argument, which Hobbs does not give, to show that this notion of abduction has anything to do with meaning, understanding or interpretation.

Consider the following silly example:

(A x). Cat (x) $\rightarrow$ Fivelegged (x)
(A x). Fivelegged(x) $\rightarrow$ Mammal (x)

from which it follows

(A x). Cat (x) → Mammal (x)

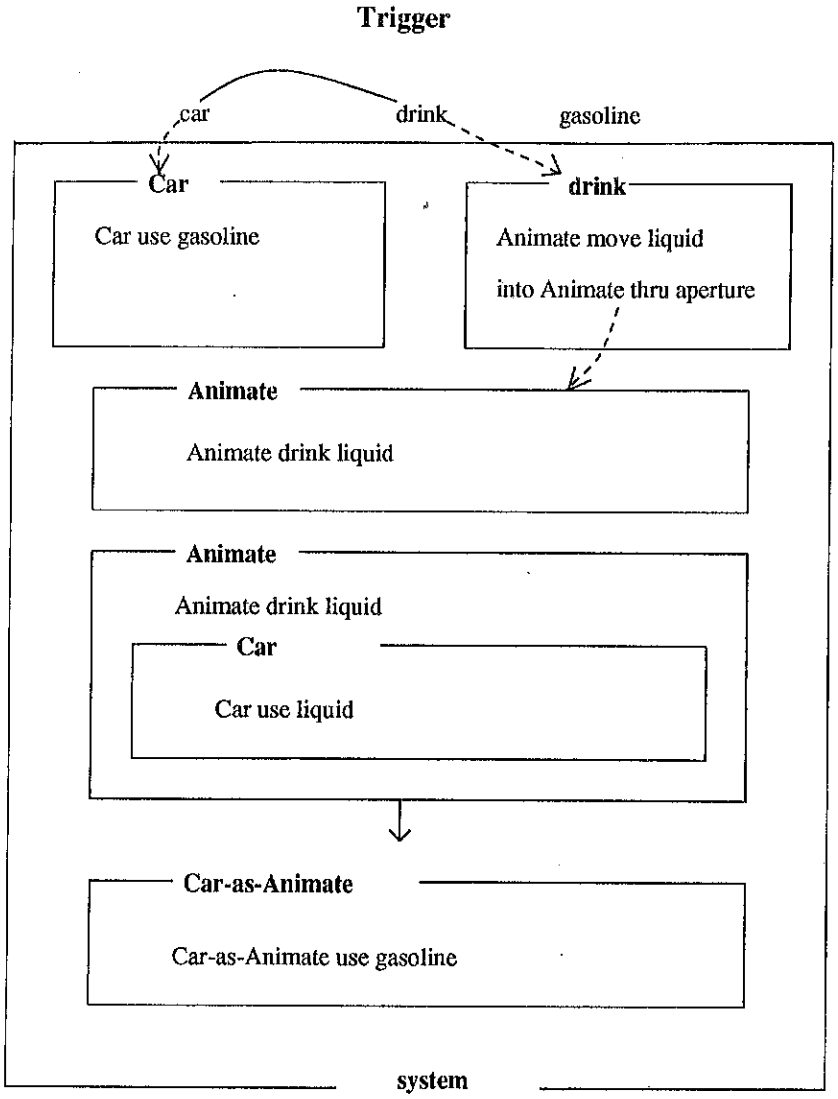

Figure 5.

The consequent is true but the first premise is wholly false. The consequent has been proved but not *interpreted* by them in any sense.

If one took a truth-conditional view of meaning, there is no way in which the (false) premises give truth-conditions for what it is to be a cat or a mammal. One might hold that they could be truth-conditions for anyone who believed held them, and for whom the above inference explained or justified "all cats are mammals". But that is not plausible, either, since, in spite of Tarski's concession of the possibility of truth "in a model or world", there is no established tradition of relativizing truth conditions to individuals as if they were beliefs. More seriously, the inference above is not a bi-conditional and hence could give no basis at all for truth-conditions in any rational sense. Hence, on any conservative view of meaning (one that one might expect to be held by an abductive logician), there is no reason to suppose abduction yields interpretation in the sense of meaning or understanding.

However, one must recognize that the tradition in which Hobbs has placed himself is wide and old, and with more implicit supporters than the parsing-as-deduction (Winograd, Colmerauer etc.) and theorem-proving traditions into which he fits: one can recall Malcolm's (1940) Wittgensteinian essay of a proof as giving the *meaning* of what was proved, again by explicit analogy to mathematical practice. Bochenski (1961) argued long ago that we understand by assuming entities and assumptions that make what we hear true. There also is, of course, Peirce himself, (though his concern was, properly, with the philosophy of science rather than general interpretation and understanding of language), but one could also add Lakoff's Generative Semantics (Lakoff 1971) which inserted propositions' logical forms into a transformational generation process so that the relation between such forms was one of inference, so that the ultimate proposition generated was "proved". The problem with all this comes when we start to take a notion of proof *seriously* in natural language understanding.

We should contrast this with what is, still, the standard view of the matter in the subject: something closer to the old notion that what something means is what it implies (as opposed to implies it). We seem to have a clear contrast here, and a paradigm of this traditional anti-abductive view would be Rieger (1975).

If we consider:

p = Castel Ivano was shelled in World War I

we can infer (provided we can construct the other necessary premises, a procedure formally indistinguishable from the backward chaining, or "hypothesis formation", phase of abduction):

→ There is a castle called Castel Ivano
→ There was a war called WWI
→ "shelled" means fired at not shucked
→ The castle has been there at least 75 years
→ Since it looks fine now, it has been restored
→ Given who fought in that war and where the castle is, it was probably shelled by the Italians or Austrians

and so on indefinitely. Some such inferences are interpretive in a clear sense of determining meaning; some offer new facts, and there is no need to try to distinguish such types clearly in making up the list above. Although what inferences can be drawn depends on the premises available (and search strategies and resources) no set of them, taken together, imply or entail the original statement p. Or rather, if they did, it would be mere chance, since, by a triviality

(a→b) ---/--- (b→a)

The inferences do show what was meant by p but do not imply it. What implies p may or may not show what it means. One can always seek evidence to explain or justify a given p, but that is a quite different matter. Or, to put that another way, hypothesis formation (the proper business of abduction) is not the same as hypothesis verification. For many people, and for the p above, that justification is

I read it in a pamphlet,

which, taken with the dubious premise, "what is in pamphlets is true" proves p, but throws no light at all on its interpretation, content or how we understand it. And that justification process is what abduction is, no more and no less.

On turning to metaphor, treated by abductive methods, Hobbs notes that "the major problem for us, then, is how to determine precisely what from the old domain *does* carry over to the new" (*ibid* p. 5). This is correct

and of course, it is the problem for every researcher in this area of AI who, almost by definition, rejects the argument of Cohen and Margalit (1972) that there can, in principle, be no algorithm selecting such a subset of features.

The important questions are whether Hobbs' method selects a different set from anyone else's system and, if it does not, whether that is compensated for by the additional intellectual coherence given by his abductive account. "But in fact", Hobbs continues, "what counts as a metaphor is determined by our theory of it" (*ibid.* p. 9).

This is probably not true, in his case or anyone else's, since, although he has a view of what counts as metaphor (see below), it is in no way connected to abduction, in that he could perfectly well have selected another metaphor-discrimination method consistently with his abductive approach. His discrimination method is negative in that he rejects the view put forward by one of us (Wilks 1977) that metaphor can be conveniently discriminated by "preference breaking", where that means either the violation of "verb preferences" as in "My car drinks gasoline" (a violated agent preference of "drink") or violated class relationship, as in "people are cattle". It is true, as Hobbs notes, that examples like the latter introduce a notion of degree that is hard to make precise, so that in:

My car is a < truck / tank

the former is taken to be true and non-metaphorical, whereas the latter is arguably metaphorical and false. He is right that this difference of degree makes metaphorical discrimination difficult, but not that it cannot be done (see further below).

His positive argument is two-fold: first, that metaphors are normal and ubiquitous and should not be processed differently from non-metaphorical utterances, and, secondly, that preference-breaking is unable to discriminate metaphor and so is falsehood, as Davidson (1978) tried to do with his phrase "metaphors are lies". As to ubiquity, one can totally accept this without also accepting that there is no need to *discriminate* metaphor from literal usage. Hobbs himself is forced to discriminate by the C predicate he introduces (see below).

Hobbs makes his second case by contrasting:

People are { 0 / not } cattle

where the first is false (and preference-breaking on our criterion) while the second is true. He argues metaphor that cannot be strongly related to phenomena like preference-breaking or truth since both the above are equally metaphorical.

But, since Hobbs wishes to explicate metaphor in terms of abduction, truth cannot be irrelevant for him: abduction is the proof of, presumably true, sentences. Perhaps the most cogent objection to Hobbs' whole thesis is that, *even if* he is right and the true statement above is as metaphorical as the false, that is very much the exception, not the standard case, and Hobbs is still burdened with a theory whose task is basically to prove falsehoods. His remedy for this, not yet fully developed, is to offer a form like:

$$(A\ x)\ (\text{Clumsy}\ (x) \rightarrow C\ (\text{Elephant})(x))$$

The intent of the formalism, at least, is clear: to give a "coercion predicate" C, such as that the "coerced elephant predicate" on the right picks out properties that are true of John, say, and help to explain the metaphor

John is an elephant.

Whatever the virtues of this device in getting round the "proof of falsehood" objection, it has other obvious drawbacks. No constructive method is given for determining how the coercion predicate works and what properties it picks out. More importantly, this abductive method could be said to disguise, not illuminate, the problem of metaphor since nothing at all in the form above tell us we are seeing a *man* as an elephant.

Hobbs' claim that some metaphors are true and unrelated to preference-breaking may be countered by the observation that, even if such cases are not rare, it may be better to think of negation as irrelevant to metaphor, rather than focusing on some metaphors being true. This irrelevance is already, to some extent, accepted in logics of *relevance*: A is as relevant to B as to not-B, if A and B are propositions. Negation does not affect relevance (see Wilks 1975), so perhaps it should not affect metaphoricity in the cattle example above. If one accepts that, Hobbs' effort to introduce truth into the discussion fails, and preference-breaking remains a handy rule of thumb for picking out metaphors. We could then rule out as metaphor what preference-breaking, in a broad sense, does not capture. Forms like

John hit the ball over the stadium (to refer to argument not baseball)

or

Connors killed McEnroe (to refer to tennis not death)

might be better seen as (true) indirect speech acts. Hobbs' best argument for an abductive approach to metaphor (apart from the pleasure of an unified approach, which we also claim) is the treatment of:

Jane is graceful but John is an elephant

where the selection of the appropriate feature for John is facilitated by the contrast with Jane. But considerable additional machinery is needed to abduce to the antonym of "graceful", so there is no sense in which his treatment of metaphor follows from a straightforward approach to logic and proof.

6. Some further work

Metaphor can involve more than the two conventional domains (tenor and vehicle). Two cases of this phenomenon in which we have some interest for the future are what we can call *parallel mixed metaphor* and *serial mixed metaphor* (or *cascaded metaphor*). In parallel mixed metaphor, the tenor domain is viewed simultaneously in terms of more than one vehicle. An example from mathematics or computer science is "the right child of this node is a leaf", which is taking a view of nodes both as parts of a biological tree and as people. In cascaded meta-phor, the tenor is viewed in terms of a vehicle which itself acts as the tenor for another metaphor (and so on). For instance, one interpretation of "A cure for terrorism is

especially urgent now that it is seeping through the whole world" has it that terrorism is being viewed, as in a previous example, as a disease, and a disease is being viewed in turn as a liquid.

Since the original belief ascription algorithm was explicitly designed to be applied recursively so as to construct any necessary level of belief ascription, it can already be applied to deal with serial mixed metaphors. The treatment is such that it makes no difference whether (i) A-as-B is seen as C or (ii) A is seen as B-as-C. This seems to accord well with our intuitions about examples we have considered, We plan to tackle parallel mixing as well in future work, especially as both types are common in the realm of metaphors of mind (which form the central concept of Barnden 1989, 1990).

Acknowledgements

This paper builds on work done in conjunction with Afzal Ballim. We are grateful also for useful suggestions from Dihong Qiu.

References:

Ballim, A. 1987. "The subjective ascription of belief to Agents". *Advances in Artificial Intelligence*, ed. by J. Hallam and C. Mellish. Chichester, England: John Wiley and Sons.

—. Y. Wilks. 1992. *Artificial Believers*. Hillsdale, N. J.: Lawrence Erlbaum Associates.

—. Y. Wilks, J. Barnden. 1992. "Belief ascription, metaphor, and intensional identification". *Cognitive Science*.

Barnden, J. A. 1989. "Belief, metaphorically speaking". *Proceedings of the 1st Intl. Conf. on Principles of Knowledge Representation and Reasoning*. San Mateo, CA: Morgan Kaufman.

—. 1990. "Naive metaphysics: A metaphor-based approach to propositional attitude representation" (Unabridged version). *Memoranda in Computer and Cognitive Science*. MCCS-90-174. Computing Research Laboratory, New Mexico State University.

Bochenski, I. 1961. *A History of Formal Logic*. Indiana: Notre Dame University Press.

Cohen, J., A. Margalit. 1972. "The role of inductive reasoning in the interpretation of metaphor". *Semantics of Natural Language* ed. by D. Davidson and G. Harman. Dordrecht: Reidel.

Cohen, P., H. Levesque. 1985. "Speech acts and rationality". *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. University of Chicago.

Davidson, D. 1978. "What metaphors mean". *Critical Inquiry*, 5, 31-48.

Falkenhainer, B., K. Forbus, D. Gentner. 1989. "The structure-mapping engine: Algorithm and examples". *Artificial Intelligence*, 41, 163.

Hobbs, J. R. 1990. "Computational aspects of metaphor". *Proceedings of NATO Workshop on Computational Theories of Communication*. Trentino, Italy.

Indurkhya, B. 1987. "Approximate semantic transference: A computational theory of metaphor". *Cognitive Science* 11, 445-480.

Lakoff, G. 1971. "On generative semantics". *Semantics*, ed. by D. Steinberg and L. Jakobovits. London and New York: Cambridge University Press.

McCarthy, J., P. Hayes. 1969. "Some philosophical problems from the startpoint of artificial intelligence". ed. by B. Meltzer and D. Mitchie. *Machine Intelligence* 4. Edinburgh University Press.

Perrault, R. 1987. Invited presentation. *Meeting of the Association for Computational Linguistics*. Stanford, Calif.

—. 1990. "An application of default logic to speech act theory". *Plans and Intentions in Communication and Discourse*, ed. by P. Cohen, J. Morgan and M. Pollack. Cambridge, Mass.: MIT Press.

Rieger, C. 1975. *Conceptional Memory and Influence*, ed. by R. Schank. Amsterdam: North Holland.

Wilks, Y. 1975. "Preference semantics". *Formal Semantics of Natural language*, ed. by E. Keenan. Cambridge, UK: Cambridge University Press

—. 1977. "Making preferences more active". *Artificial Intelligence* 8,75-97.

—. A. Ballim. 1987. "Multiple agents and the heuristic ascription of belief". *Proceedings of the 10th International Joint Conference of Artificial Intelligence*. Los Altos, CA.: Morgan Kaufmann.

—. J. Bień. 1979. "Speech acts and multiple environments". *Proceedings of the 6th International Joint Conference of Artificial Intelligence*. Tokyo, Japan.

—. 1983. "Beliefs, points of view and multiple environments". *Cognitive Science 8, 120-164*.