Non-Monotonic Formalisms for Natural Language Semantics

Richmond H. Thomason
Linguistics Department
University of Pittsburgh
Pittsburgh, PA 15260
U.S.A.

May 8, 2004

Working Draft of a Paper in Progress

This is a working draft:
version of May 8, 2004.
The material is volatile; do not quote.
Comments welcome.

Non-Monotonic Formalisms for Natural Language Semantics

1. Introduction

I will try to do three things in this paper. First, I want to situate certain problems in natural language semantics with respect to larger trends in logicism, including:

(i) Attempts by positivist philosophers earlier in this century to provide a logical basis for the physical sciences;

(ii) Attempts by linguists and logicians to develop a "natural language ontology" (and, presumably, a logical language that is related to this ontology by formally explicit rules) that would serve as a framework for natural language semantics;

(iii) Attempts in artificial intelligence to formalize common sense knowledge.

Second, I want to propose an extension of Montague's framework, and to illustrate some of its applications in the semantics of words. Third, I'll distinguish between the problem of specifying a phenomenon like natural language semantics (or providing conditions on an adequate solution), and providing an account of the related reasoning procedures. I'll try to explain briefly why I think that we are a long way from an adequate, unified account of semantics for the purposes of natural language processing.

All of this is actually a tall order. The pressure of trying to fill it given limits of space, and—even more—my present state of ignorance or confusion on many of the central issues will affect the quality of the exposition and the views themselves. But I hope that you will bear with me, since I think that a partially successful and fragmentary attempt at the larger project may convince some members of my audience that the natural language semantics community and the subgroup of the AI community interested in formalizing common sense knowledge have a great deal in common, and much to learn from one another. To a large extent, the rhetorical goal of this paper is to make this message seem plausible and exciting, and to provide semanticists with a number of relevant references from the literature in AI.

2. Logicism

The material in this and the subsequent two sections of this abstract is lifted from [37].

Let X be a topic of inquiry. X logicism is the view that X should be presented as an axiomatic theory from which the rest can be deduced by logic. *Science logicism* is expressed as an ideal in Aristotle's *Organon*. But Aristotle's logic is far too weak to serve as a means of representing Aristotelian science, and logicism was in effect remained impracticable until the 17th century, when a separation of theoretical science from common sense simplified the task of designing an underlying logic.[1]

There is a moral here about logicism. X logicism imposes a program: the project of actually presenting X in the required form. But for the project to be feasible, we have to choose a logic that is adequate to the demands of the topic. If a logic must involve explicit formal patterns of valid reasoning, the central problem for X logicism is then to articulate formal patterns that will be adequate for formalizing X.

[1] Despite the simplification, of course, a workable formalism did not begin to emerge until the 19th century.

The fact that very little progress was made for over two millennia on a problem that can be made to seem urgent to anyone who has studied Aristotle indicates the difficulty of finding the right match of topic and formal principles of reasoning. Though some philosophers (Leibniz, for one) saw the problem clearly, the first instance of a full solution is Frege's choice of mathematical analysis as the topic, and his development of the *Begriffschrift* as the logical vehicle. It is a large part of Frege's achievement to have discovered a choice that yields a logicist project that is neither impossible nor easy.

I will summarize some morals. (1) Successful logicism requires a combination of a formally presented logic and a topic that can be formalized so that its inferences become logical consequences. (2) When logicist projects fail, we may need to seek ways to develop the logic. (3) Logic development can be difficult and protracted.

3. Extensions to the empirical world

The project of extending Frege's achievement to the empirical sciences has not fared so well. Of course, the mathematical parts of sciences such as physics can be formalized in much the same way as mathematics. Though the metamathematical payoffs of formalization are most apparent in mathematics, they can occasionally be extended to other sciences.[2] But what of the empirical character of sciences like physics? One wants to relate the systems described by these sciences to observations.

Rudolph Carnap's *Aufbau*[3] was an explicit and ambitious attempt to extend mathematics logicism to science logicism, by providing a basis for formalizing the empirical sciences. The *Aufbau* begins by postulating elementary units of subjective experience, and attempts to build the physical world from these primitives in a way that is modeled on the constructions used in Frege's mathematics logicism.

Carnap believed strongly in progress in philosophy through cooperative research. In this sense, and certainly compared with Frege's achievement, the *Aufbau* was a failure. Nelson Goodman, one of the few philosophers who attempted to build on the *Aufbau*, calls it "a crystallization of much that is widely regarded as worst in 20th century philosophy."[4]

After the *Aufbau,* the philosophical development of logicism becomes somewhat fragmented. The reason for this may have been a general recognition, in the relatively small community of philosophers who saw this as a strategically important line of research, that the underlying logic stood in need of fairly drastic revisions.[5]

This fragmentation emerges in Carnap's later work, as in the research of many other logically minded philosophers. Deciding after the *Aufbau* to take a more direct, high-level approach to the physical world, in which it was unnecessary to construct it from phenomenal primitives, Carnap noticed that many observation predicates, used not only in the sciences but in common sense, are "dispositional"—they express expectations about how things will behave under certain conditions. A malleable material will deform under relatively light pressure; a flammable material will burn when heated sufficiently. It is natural to use the word 'if' in defining such predicates; but the "material conditional" of Frege's logic gives

[2]See [25].
[3][5].
[4][16], page 545.
[5]I can vouch for this as far as I am concerned.

incorrect results in formalizing such definitions. Much of [6] is devoted to presenting and examining this problem.

Rather than devising an extension of Frege's logic capable of solving this problem, Carnap suggests dropping the requirement that these predicates should be explicated by definitions. This relaxation makes it harder to carry out the logicist program, because a natural way of formalizing dispositionals is forfeited. But it also postpones a difficult logical problem, which was not, I think, solved adequately even by later conditional logics in [34] and [19]. Such theories do not capture the notion of normality that is built into dispositionals: a more accurate definition of 'flammable', for instance, is 'what will *normally* burn when heated sufficiently'. Thus, logical constructions that deal with normality offer some hope of a solution to Carnap's problem of defining dispositionals. Such constructions have only become available with the development of nonmonotonic logics.

4. Linguistic logicism

Though work in philosophical logic and its applications continues the logicist tradition to some extent, logicist projects are largely out of fashion in philosophy, and much of the work on projects of this sort is being carried on in other disciplines.

In linguistics, a clear logicist tradition emerged from the work of Richard Montague, who (building to a large extent on Carnap's work in [7]) developed a logic he presented as appropriate for *philosophy logicism.* Montague's extreme logicist position is stated most clearly in a passage in [26].

> It has for fifteen years been possible for at least one philosopher (myself) to maintain that philosophy, at least at this stage in history, has as its proper theoretical framework set theory with individuals and the possible addition of empirical predicates. ... [But] philosophy is always capable of enlarging itself; that is, by metamathematical or model-theoretical means—means available within set theory—one can "justify" a language or theory that transcends set theory, and then proceed to transact a new branch of philosophy within the new language. It is now time to take such a step and to lay the foundations of intensional languages.[6]

Montague's motivation for expanding his logical framework is the need to relate empirical predicates like 'red' to their nominalizations, like 'redness'. He argued that many such nominalizations denote properties, that terms like 'event', 'obligation', and 'pain' denote properties of properties, and that properties should be treated as functions taking possible worlds into extensions. The justification of this logical framework consists in its ability to formalize certain sentences in a way that allows their inferential relations with other sentences to be captured by the underlying logic.

Philosophers other than Montague—not only Frege, but Carnap in [7] and Church in [9]—had resorted informally to this methodology. But Montague was the first to see the task of *natural language logicism* as a formal challenge. By actually formalizing the syntax of a natural language, the relation between the natural language and the logical framework

[6] [28], pages 156–157.

could be made explicit, and systematically tested for accuracy. Montague developed such formalizations of several ambitious fragments of English syntax in several papers, of which [27] was the most influential.

The impact of this work has been more extensive in linguistics than in philosophy.[7] Formal theories of syntax were well developed in the early 70s, and linguists were used to using semantic arguments to support syntactic conclusions, but there was no theory of semantics to match the informal arguments. "Montague grammar" quickly became a paradigm for some linguists, and Montague's ideas and methodology have influenced the semantic work of all the subsequent approaches that take formal theories seriously.

As practiced by linguistic semanticists, language logicism would attempt to formalize a logical theory capable of providing translations for natural language sentences so that sentences will entail one another if and only if the translation of the entailed sentence follows logically from the translation of the entailing sentence and a set of "meaning postulates" of the semantic theory. It is usually considered appropriate to provide a model-theoretic account of the primitives that appear in the meaning postulates.

This methodology gives rise naturally to the idea of "natural language metaphysics," which tries to model the high-level knowledge that is involved in analyzing systematic relations between linguistic expressions. For instance, the pattern relating the transitive verb 'bend' to the adjective 'bendable' is a common one that is productive not only in English but in many languages. So a system for generating derived lexical meanings should include an operator ABLE that would take the meaning of 'bend' into the meaning of 'bendable'.

To provide a theory of the system of lexical operators and to explain logical interactions (for instance, to derive the relationship between 'bendable' and 'deformable' from the relationship between 'bend' and 'deform'), it is important to provide a model theory of the lexical operators. So, for instance, this approach to lexical semantics leads naturally to a model-theoretic investigation of ability,[8] a project that is also suggested by a natural train of thought in logicist AI.[9]

Theories of natural language meaning that, like Montague's, grew out of theories of mathematical language, are well suited to dealing with quantificational expressions, as in

4.1. *Every boy gave two books to some girl.*

In practice, despite the original motivation of his theory in the semantics of word formation, Montague devoted most of his attention to the problems of quantification, and its interaction with the intensional and higher-order apparatus of his logical framework.

But those who developed Montague's framework soon turned their attention to these problems, and much of the later research in Montague semantics—especially David Dowty's

[7] It is hard to explain the lack of philosophical interest in the project. Recent linguistic work in natural language metaphysics is loosely connected to earlier attempts to exploit language as a source of insights into the nature of distinctively human patterns of thought about what might be called the common sense world. I am thinking here of works like [8] and of [18]. Both of these projects grew out of a rich philosophical tradition: Cassirer's work, in particular, is firmly rooted in the European Kantian tradition. And, of course, there has been much work in the phenomenological tradition—which, however, has been much less formal.

[8] That the core concept that needs to be clarified here is ability rather than the bare conditional 'if' is suggested by cases like 'drinkable'. 'This water is drinkable' doesn't mean 'If you drink this water it will have been consumed'. (Of course, ability and the conditional are related in deep ways.) I will return briefly to the general problem of ability in Section 7.5, below.

[9] See, for example, [33].

early work in [12] and the work that derives from it—concentrates on semantic problems of word formation, which of course is an important part of lexical semantics.[10]

5. Formalizing common sense

To a certain extent, the motives of the common sense logicists overlap with Carnap's reasons for the *Aufbau.* The idea is that the theoretical component of science is only part of the overall scientific project, which involve situating science in the world of experience for purposes of testing and application; see [23] for explicit motivation of this sort. For extended projects in the formalization of common sense reasoning, see [17] and [10].

The project of developing a broadly successful logic-based account of semantic interrelationships among the lexical items of a natural language is roughly comparable in scope with the project of developing a high-level theory of common sense knowledge. Linguists are mainly interested in explanations, and computer scientists are (ultimately, at any rate) interested in implementations. But for logicist computer scientists who have followed McCarthy's advice of seeking understanding before implementing, the immediate goals of the linguistic and AI projects are not that different.

And—at the outset at least—the subject matter of the linguistic and the computational enterprise are remarkably similar. The linguistic research motivated by lexical decomposition beginning in [12] and the computational research motivated largely by problems in planning (or practical reasoning) both lead naturally to a focus on the problems of representing change, causal notions, and ability.

6. Formalizing nonmonotonic reasoning

See [15] for a good guide to the field of nonmonotonic reasoning and its early development. For subsequent developments, it is necessary to consult the primary literature. For example, see [11], [20], [24], [1], [4], [2], [14], [22], and [13]

Among the available theories of defeasible reasoning that could be applied in lexical semantics, I find circumscription the most congenial to use in attempting to apply these theories to problems of natural language semantics, for the following reasons.

(i) Circumscription is based on second-order logic, and this second-order foundation can easily be generalized to Intensional Logic.[11] Montague's apparatus does conflict with preferences that McCarthy has expressed from time to time about how to formalize intensional constructions, but this philosophical disagreement does not seem to be an obstacle to the absorption of circumscription into the framework of Intensional Logic.

[10] This emphasis on compositionality in the interpretation of lexical items is similar to the policy that Montague advocated in syntax, and it has a similar effect of shifting attention from representing the content of individual lexical items to operators on types of contents. But this research program seems to require a much deeper investigation of "natural language metaphysics" or "common sense knowledge" than the syntactic program, and one can hope that it will build bridges between the more or less pure logic with which Montague worked and a system that may be more genuinely helpful in applications that involve representation of and reasoning with linguistic meaning.

[11] See [36] for a brief description and application of the combined theory (to discourse, rather than to lexical semantics).

(ii) The more sophisticated versions of circumscription provide an explicit formalism for dealing with abnormalities.[12] I believe that such a formalism is needed in the linguistic applications.

7. Case studies

I'll illustrate the use of nonmonotonic formalisms in semantics with several case studies. In these studies, I'm merely trying to motivate the use of a nonmonotonic formalism in the semantics of words, and to suggest how it might be applied to some of the immediate problems that arise in this area. At the date of this version, I have not tried to work out the details. At this point, the abstract will become much more sketchy.

7.1. The *-able* suffix

The *-able* suffix illustrates a number of characteristics that challenge semantics.

1. There is variation in the meanings it assumes, but this variation is across a family of closely related shades of meaning. As usual in these cases, it is hard to tell whether to treat the variation by listing senses, by finding a single common meaning allowing for different uses, or by making the meaning context-dependent.

2. The meanings themselves are difficult to formalize.

3. These meanings seem to invoke references to concepts via relations of common-sense real world knowledge rather than linguistic knowledge.

4. There are exceptional patterns.

The most usual pattern of VERB+*able* works with transitive verbs V that are broadly telic. These verbs have three characteristics: they correspond to procedures that are in the normal repertoire of actions of human agents, there are normal or standard ways of initiating these actions, and there is a successful end state associated with the performance of the actions. In what I will call the paradigmatic case, the meaning of the derived adjectival form is that a thing normally will achieve the state s successfully when a test action associated with V is applied to it. The term 'successful' is deliberately used here to cover both the cases in which the state is really achieved, and in which the state is achieved without undesirable side effects. (This last condition, we can see, can shade into cases in which there are not only no undesirable side effects, but in which the state is worthy of being achieved.)

Here are some examples illustrating this paradigmatic case.

[12] See [20].

acceptable	*doable*	*observable*
adjustable	*enjoyable*	*provable*
admissible	*dispensible*	*printable*
approachable	*edible*	*readable*
applicable	*favorable*	*recognizable*
breakable	*driveable*	*recognizable*
adoptable	*expressible*	*reusable*
believable	*formalizable*	*reversible*
consumable	*expendable*	*reversible*
bearable	*flexible*	*soluble*
communicable	*learnable*	*solvable*
defensible	*fixable*	*solvable*
cleanable	*implementable*	*TeXable*
defeatable	*movable*	*trainable*
defensible	*imaginable*	*transferable*
detectable	*modifiable*	*transportable*
drinkable	*openable*	*wearable*
desirable	*loveable*	*workable*

Idea: use branching time. You perform a test action Initiate(Act). This yields a set H of normal outcome histories. You then test H for success. The test for success involves a scale of goodnesses. There are at least positions on the scale: dontcare, notbad, and good. Conditional provides an alternativeness rel ALT on histories. These are the cases where initiating action is not performed. The goal state may or may not be achieved. The scale provides a means of comparison to the alts where the goal is not achieved. R1: h dontcare h' if holds(s,h) and for all h'in H s.t. notholds(s,h') TRUE. R2: h notsomehowmuchworse wrt A if holds(s,h) and for all normal h' in ALT -A s.t. notholds(s,h') Not h'¡¡Resp h for some respect Resp R3: h overallbetter wrt A if holds(s,h) and for all normal h' in ALT -A s.t. -holds(s,h') h¿Overall h' R4: h overallmuchbetter wrt A if holds(s,h) and for all normal h' st -holds(s,h') h¿¿Overall h'

able(V,A)(R) where A is a normal procedure for bringing about V means that for all normal h in outcomes(A) h cR wrt A. able(V)(c) if for some normal procedure A for bringing about V, able(V,A)(c).

Let's see how this applies to the examples above.

1 acceptable Suppose we're talking about an offer. To say it's acceptable is not just to say that if an acceptance is initiated (say by saying "I accept") that it will be accepted. Rather, the meaning seems to be that the overall consequences are better if the offer is accepted. This is case 3.

2 adjustable The procedure here has to be a normal one. 3 admissible 4 approachable 5 applicable 6 breakable 7 adoptable 8 believable 9 consumable 10 bearable 11 communicable 12 defensible 13 cleanable 14 defeatable 15 defensible 16 detectable 17 drinkable 18 desirable

19 doable 20 enjoyable 21 dispensible 22 edible 23 favorable 24 driveable 25 expressible 26 formalizable 27 expendable 28 flexible 29 learnable 30 fixable 31 implementable 32 movable 33 imaginable 34 modifiable 35 openable 36 loveable 37 observable 38 provable 39 printable 40 readable 41 recognizable 42 recognizable 43 reusable 44 reversible 45 reversible 46 soluble 47 solvable 48 solvable 49 TeXable 50 trainable 51 transferable 52 transportable 53 wearable

54 workable

7.1.1. The ability of normal consequences of a test

The natural way to define 'x is water-soluble' is

7.2. If x were put in some water, then x would dissolve in the water.

So at first glance, it may seem that the resources for carrying out the definition that Carnap found problematic will be available in a logic with a subjunctive conditional. But suppose it happens that if one were to put this lump of salt in some water, it would be in this water—and this water is saturated with salt. The fact that the lump would not then dissolve is no reason why this salt should count as not water-soluble.

This and other such thought experiments indicate that what is wanted is not the bare subjunctive conditional, but a "conditional normality" of the sort that is used in some nonmonotonic formalisms.[13]

In a circumscriptive framework, normality is obtained by conditions on a number of abnormality predicates, which are then circumscribed, or minimized relative to certain background assumptions, in obtaining models of the nonmonotonic theory. Events are an appropriate locus for organizing these abnormality predicates not only in the case of dissolving, but in many other cases of interest for purposes of lexical decomposition.

It is convenient to think of events as classified by a system of event types, from which abnormalities and other features are inherited. In treating the dissolving example, I will make the following assumptions.

7.i. There is an event type ϕ of *put-in events*.[14] Associated with this type (and, by inheritance, with events falling under it) there is a container $container(\phi)$ and a thing moved $movee(\phi)$.

7.ii. The event type ϕ has a subtype ϕ_1, in which $container(\phi)$ is a quantity of water and $movee(\phi)$ is a quantity of salt. There is an abnormality predicate associated with ϕ_1.

7.iii. There is an event type ψ of *dissolving events*. I assume that associated with this event type (and, by inheritance, with events falling under it) there is an inception, a body, and a culmination (where the first two are events and the last is a state); also, an associated medium $medium(\psi)$ and a thing dissolved $dissolvee(e)$; also an abnormality predicate.

It will follow from general considerations about the event type ψ that if a ψ-normal event of this type occurs, its associated culmination state will also occur. (See the remarks below on telicity.)

Given this information about event types, the sort of analysis that I currently favor for dissolving amounts to this.

7.iv. Every ϕ_1-normal ϕ_1-event e_1 is also the inception of a ϕ-event e_2 such that

$$container(e_1) = medium(e_2) \text{ and } movee(e_1) = dissolvee(e_2).$$

This analysis invokes notions that have come to light in accounting for other phenomena in the analysis of word meaning. I will pass directly to these other phenomena, but will return to the problem of dispositionals briefly later, when I discuss ability.

[13] See [2], [1].

[14] This event type itself has a decompositional analysis, but we can ignore that for purposes of the example.

7.2. Telicity

I am abstracting here away from all problems having to do with time and the progressive, and concentrating on the relation between a telic event and its culminating state.[15] The most important feature of the type of telic events is that these events have three associated parts: the inception, the body, and the culmination. The inception is an initiating event. The culmination is the state that normally results. (Since the beginning, the theory of planning has concentrated on features of culminations, since these represent properties of the state that can be assumed to result if the agent performs an action.) The body is the process that normally leads to the culmination; often (as in closing a door or filling a glass), the body will consist of stages in which the goal is progressively achieved.[16] We can lay it down as a general default on telic events that the culmination of such an event will occur if its inception and body occurs. In many cases (like dissolving, or filling a glass from a tap, but not like filling a glass from a pitcher) the body will also normally occur if the inception occurs. Thus, I am likening unfulfilled telic events to Manx cats—they are objects that belong to a type that normally has a certain part, but that for some *ad hoc* reason happen to lack this part.

7.3. Agency

Some formalisms of agency in AI involve a separation of events into those that are in an agents' immediate repertoire and those that are not.[17] If such a division is adopted for linguistic purposes, we can capture agency—at least, for telic events that normally follow from their inceptions—as follows.

7.v. $Do(x, e)$ holds iff the inception e' of e is identical to an immediate action e'' that is performed by x, provided that the body of e'' occurs.[18]

For example, it follows from this account that in case someone puts a piece of salt in water and it then dissolves in the ordinary way, them this person has also performed the action of dissolving it in water, assuming that putting the salt in water is immediate an immediate action. Moreover, the action of putting the salt in water will be the inception of the dissolving event.

On this treatment, we dispense with an explicit use of any causal notions in the analysis of agency—though causal notions are certainly implicit if we believe that there is a connection between sequences of events conforming to patterns of normality and causal sequences. Since—despite the recent contributions to this subject discussed in Section 7.6—an explicit theory of common sense causality remains somewhat programmatic, I prefer such eliminative accounts. However, I'm not sure if explicit causality can be eliminated in general from the theory of agency.

7.4. Ability

'This water is drinkable' doesn't mean

[15]It should be clear, though, that I have in mind an account that would relate the truth of a progressive sentence to the occurrence of the body of an event.

[16]For ideas that are in some respects similar to these, see [35].

[17]See especially [29].

[18]I want to say that the body of a telic event occurs even if it is partial or incomplete.

7.3. *An attempt to drink this water will normally culminate in its being drunk.*

Rather, the meaning is

7.4. *Normally, one can drink this water.*

This, linguistic examples, as well as obvious motivations in AI that relate to the needs of planning, lead to a need for an account of practical ability. I don't think that such an account can be given without an extended background theory of practical reasoning. For that reason, the account that I'll sketch here may seem circular or trivial. The reason (I hope) is that the background hasn't been filled in.

Let's suppose that there is a propositional constant *practical_abnormality* that is used in practical reasoning to reject alternatives because of utility considerations. That is, if a contemplated practical alternative is shown to lead (perhaps with the aid of defaults) to this constant, the alternative has thereby been shown to one that can be ruled out of consideration. A qualitative account of practical reasoning would have to relate this constant to desires and contingent circumstances.

The definition of practical ability would then be the following, where $\Box$ represents temporal necessity.

7.vi. $can(\phi) \leftrightarrow \neg\Box[\phi \rightarrow practical\_abnormality]$

7.5. Artifacts

Many artifacts are defined in terms of their normal uses. This suggests decompositional analyses such as the following example.

7.vii. A *fastener* is an object x such that, where ϕ is the event type of using x, every ϕ-normal occurrence of an event e of type ϕ is such that *purposee* is to fasten an object to another object.

7.6. Causality

The notion of causality is usually left unanalyzed in linguistic treatments of lexical decomposition. Since it typically acts as a morphological primitive, it is easy to suppose that it is a semantic primitive. But treating it as such doesn't agree well with a logical approach to semantics. For one thing, causality would have to appear as many logical constructs of different types, and—even if these constructs are primitive—the relationships between, for instance, causality as an operator on events and causality as an operator on propositons would have to be clarified. For another, it would be difficult to account for causal reasoning or for the relationships between causality and, for instance, time and change without a model theoretic semantics for causality.

David Dowty was apparently the first linguist to take the need for a theory of this kind seriously (although some of the motivations for Dowty's project are prefigured in the earlier work of the generative semanticists). His work, especially in [12], provides an account of propositional or modal causality that integrates it with temporal and causal reasoning. As Dowty's investigation of word meaning makes clear, the crucial form of causality that is needed for this purpose is a relation between an individual and a proposition—this is the sense in which an agent causes something to be true. Unfortunately, Dowty's analysis of this sort of causality is hopelessly flawed.[19]

As far as I know, there is no way to repair this flaw, along the lines of the approach taken in [12].

From the standpoint of this paper, it would be pleasant to say that a nonmonotonic logic provides the solution to difficulties of this kind, as well as a deeper account of causality. I think it would be premature to make this claim, although it certainly seems as though interesting and important new ideas about the formalization of causality are emerging from the literature on nomonotonic logic and formalisms for reasoning about action and change. This is especially true if this literature is construed broadly and Judea Pearl's work is included, e.g. [30] and [31]. For causality in the literature on planning formalisms, see [14], [21], and [38]. And [32] is an explicit attempt to use nonmonotonic logic to account for causality.

I will not try to provide details about these ideas in this version of the paper.

I have doubts about whether the introduction of an abstract causal operator is the needed or even appropriate in analyzing the sort of causality that figures in word formation. Complex words involving causality seem to invoke structured complexes of events satisfying various constraints. Intuitively, some of these constraints are causal, but I'm not not sure that this needs to be said explicitly for linguistic purposes. This general approach is illustrated in Section 7.1, where the structure of a telic eventuality, including an inception and culmination, is used to account for dispositions without explicit reference either to times or to causality.

8. The problem of reasoning

Many researchers in the area of knowledge representation have advocated a close association between the development of systems of representation and reasoning systems. For many practitioners, this may mean only that attention must be paid to the complexity of some of the algorithms associated with the representations; for others, it may mean that the representations should be deployed in a working system whose performance has been tested in practical terms.[20] On the other hand, John McCarthy, one of the founders of AI and of the field of knowledge representation, seems to have steadfastly maintained a platonic research agenda, according to which we first clarify in logical terms the knowledge that we are trying to build into a reasoning system, without distracting ourselves with heuristics and implementation details.

Since it seems to me that we may not yet have adequate theories in which to describe our goals, and that attention to the design of successful systems is a likely source of good

[19]Dowty takes a causal operator CAUSE that takes two propositions as arguments as a primitive, and provides a model-theoretic interpretation of CAUSE. Causatives involving an agent and an outcome state are analyzed using an existential qualtifier over propositions. For instance, 'x warms y' would be formalized as

$$\exists P[\textsc{cause}(P(x), {}^\wedge\textsc{become}(\mathbf{warm'}(y))],$$

where P is a variable over functions from individuals to propositions. It follows from this that if, for instance, Bill warms the soup, then everyone warms the soup. (Suppose that Bill warms the soup; then there is a proposition p such that p causes the soup to be warm. Now, let a be an arbitrary individual. Consider a function Q that takes a into the proposition that $a = a \wedge p$. This proposition is logically equivalent to p, and therefore causal relation holds between it and the proposition that y becomes warm. So, according to the definition of agent causality, a warms the soup.

[20]See [3].

theories, and a good way of ensuring that they will ultimately have some reasonable relation to technology, I am a moderate on this issue.

However, I would not have been able to write this paper without pretending to be a platonist. Natural language semantics is such a large domain that it is very hard to think about it in the general way that linguistics seems to require while maintaining a good relation to some feasible reasoning task. The task of axiomatizing the theories that I have been contemplating in this theory is forbidding, and, if the axiomatization could be carried out, it is hardly likely that it would provide a promising path to any workable implementation.

Perhaps the best hope of closing the gap between theory and technology would be the development of compilers for higher-level planning languages that use the kinds of concepts that we have found useful for lexical decomposition. Some computer scientists are beginning to work in this direction, and compilers have even been written for simple languages of this sort.[21] But this line of research will need a lot of development before we can hope to get much help from it in figuring out how to assign procedural significance to parts of lexical semantics, and at best it is pretty slender at the moment.

[21] See [33].

Bibliography

[1] Nicholas Asher and Michael Morreau. Commonsense entailment: a modal theory of nonmonotonic reasoning. In J. Mylopoulos and R. Reiter, editors, *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pages 387–392, Los Altos, California, 1991. Morgan Kaufmann.

[2] Craig Boutilier. Conditional logics for default reasoning and belief revision. Technical Report KRR–TR–92–1, Computer Science Department, University of Toronto, Toronto, Ontario, 1992.

[3] Ronald J. Brachman. The future of knowledge representation. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, volume 2, pages 1082–1092, Menlo Park, California, 1990. AAAI Press.

[4] Gerhard Brewka. *Nonmonotonic Reasoning: Logical Foundations of Commonsense.* Cambridge University Press, Cambridge, England, 1991.

[5] Rudolph Carnap. *Der logische Aufbau der Welt.* Weltkreis-Verlag, Berlin-Schlactensee, 1928.

[6] Rudolph Carnap. Testability and meaning. *Philosophy of Science*, 3 and 4:419–471 and 1–40, 1936–1937.

[7] Rudolph Carnap. *Meaning and Necessity.* Chicago University Press, Chicago, 2 edition, 1956. (First edition published in 1947.).

[8] Ernst Cassirer. *The Philosophy of Symbolic Forms*, volume 1–3. Yale University Press, New Haven, 1955. (These works were first published (in German) in 1923–29.).

[9] Alonzo Church. The need for abstract entities in semantic analysis. *Proceedings of the American Academy of Arts and Sciences*, 80:100–112, 1951.

[10] Ernest Davis. *Common Sense Reasoning.* Morgan Kaufmann, San Francisco, 1991.

[11] James Delgrande. An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence*, 36:63–90, 1988.

[12] David R. Dowty. *Word Meaning in Montague Grammar.* D. Reidel Publishing Co., Dordrecht, Holland, 1979.

[13] Dov Gabbay, Christopher Hogger, and J.A. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 2: Nonmonotonic Reasoning.* Oxford University Press, Oxford, forthcoming.

[14] Hector Geffner. *Default Reasoning: Causal and Conditional Theories.* MIT Press, Cambridge, Massachusetts, 1992.

[15] Matthew L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann, Los Altos, California, 1987. (Out of print.).

[16] Nelson Goodman. The significance of *der logische aufbau der welt.* In Paul Schilpp, editor, *The Philosophy of Rudolph Carnap*, pages 545–558. Open Court, LaSalle, Illinois, 1963.

[17] Jerry R. Hobbs and Robert C. Moore, editors. *Formal Theories of the Commonsense World.* Ablex Publishing Corporation, Norwood, New Jersey, 1988.

[18] Otto Jespersen. *The Philosophy of Grammar.* W.W. Norton, New York, 1965. (This work was originally published in 1924.).

[19] David K. Lewis. *Counterfactuals.* Harvard University Press, Cambridge, Massachusetts, 1973.

[20] Vladimir Lifschitz. Circumscriptive theories: A logic-based framework for knowledge representation. *Journal of Philosophical Logic*, 17(3):391–441, 1988.

[21] Fangzhen Lin. Embracing causality in specifying the indirect effects of actions. In Chris Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1985–1991, San Francisco, 1995. Morgan Kaufmann.

[22] Victor Marek and Mirosław Truszczczyński. *Nonmonotonic logic: Context-Dependent Reasoning.* Springer-Verlag, Berlin, 1994.

[23] John McCarthy. Some expert systems need common sense. In H. Pagels, editor, *Computer Culture: the Scientific, Intellectual and Social Impact of the Computer*, volume 426 of *Annals of The New York Academy of Sciences*, pages 129–137. The New York Academy of Sciences, 1984.

[24] John McCarthy. Artificial intelligence, logic, and formalizing common sense. In Richmond Thomason, editor, *Philosophical Logic and Artificial Intelligence*, pages 161–190. Kluwer Publishing Co., Dordrecht, Holland, 1989.

[25] Richard Montague. Deterministic theories. In *Decisions, Values, and Groups*, volume 2, pages 325–370. Pergamon Press, Oxford, 1962. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 303–359.

[26] Richard Montague. On the nature of certain philosophical entities. *The Monist*, 53:159–194, 1969.

[27] Richard Montague. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, editor, *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, pages 221–242. D. Reidel Publishing Co., Dordrecht, Holland, 1973. Reprinted in *Formal Philosophy*, by Richard Montague, Yale University Press, New Haven, CT, 1974, pp. 247–270.

[28] Richard Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven, CT, 1974.

[29] Robert C. Moore. A formal theory of knowledge and action. In Jerry R. Hobbs and Robert C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex Publishing Corporation, Norwood, New Jersey, 1985.

[30] Judea Pearl. Embracing causality in default reasoning. *Artificial Intelligence*, 35:259–271, 1988.

[31] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, England, 2000.

[32] Yoav Shoham. *Reasoning About Change: Time and Causation From the Standpoint of Artificial Intelligence*. The MIT Press, Cambridge, Massachusetts, 1988.

[33] Yoav Shoham. Agent oriented programming. *Artificial Intelligence*, 60(1):51–92, 1993.

[34] Robert C. Stalnaker and Richmond H. Thomason. A semantic analysis of conditional logic. *Theoria*, 36:23–42, 1970.

[35] Mark Steedman and Marc Moens. Temporal ontology in natural language. In *Proceedings of the 25th Annual Conference of the Association for Computational Linguistics*, pages 1–7. Association for Computational Linguistics, 1987.

[36] Richmond Thomason. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In Philip R. Cohen, Jerry Morgan, and Martha Pollack, editors, *Intentions in Communication*, pages 326–363. MIT Press, Cambridge, Massachusetts, 1990.

[37] Richmond Thomason. Logicism, artificial intelligence, and common sense: John McCarthy's program in philosophical perspective. In Vladimir Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation*, pages 449–466. Academic Press, San Diego, 1991.

[38] Hudson Turner. A logic of universal causation. *Artificial Intelligence*, 113(1–2):87–123, 1999.