Formalizing the Semantics of Derived Words

Richmond H. Thomason Philosophy Department University of Michigan Ann Arbor, MI 48109-2110 U.S.A.

October 28, 2005

This material is taken from a long-term, but often-neglected project that goes back to 1993.

For the last five years or so, versions of this material have been posted on the web. Comments welcome.

1. Introduction

The logical approach that has been so successful in the semantic interpretation of syntactic structure has never produced a very satisfactory account of word meaning. This paper is intended to promote and illustrate an approach to that problem.

I believe that this approach leads to a wider problem that brings together elements of linguistics and philosophy in an illuminating way. But the single case study that I provide here, while it may be suggestive, does not go far enough to make a good case for the more general point. This paper is extracted from a larger collection of documents, and is intended to motivate and illustrate the ideas.

I hope that even a partially successful and fragmentary sketch of the larger project may convince some members of my audience that the natural language semantics community and the subgroup of the A.I. community interested in formalizing common sense knowledge have a great deal in common, and much to learn from one another, and that what they have to learn is useful and important for philosophy.

2. Logicism¹

I want to begin by situating 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.

Frege did a lot for logic, but I think he left us with an undeservedly narrow and unpromising version of logicism that is entirely too focused on the subject matter of mathematics and on definition as the primary method of analysis .

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 remained impracticable until the 17th century, when a separation of theoretical science from common sense simplified the task of designing an underlying logic.²

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.

¹The material in this and the subsequent section is lifted in part from [Thomason, 1991].

²Despite 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.³ But what of the empirical character of sciences like physics? Given a fomalization of mathematics, and especially of analysis, formalizing the mathematical part of a science like physics becomes a relatively routine task. But including the empirical part of such a science is a more ambitious task, calling for a formalization of the commonsense language in which experimental setups are described and observations are recorded.

Rudolph Carnap's $Aufbau^4$ 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."⁵

After the Aufbau, the philosophical development of logicism becomes somewhat fragmented: instead of "macroformalization" projects that target large-scale domains, projects tend to concentrate on the logic of certain constructions, such as tenses, modalities, and the conditional. The reason for this may have been a general recognition, in the relatively small group of philosophers who saw this as a strategically important line of research, that the underlying logic stood in need of fairly drastic revisions.⁶

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

³See [Montague, 1962].

 $^{^{4}}$ [Carnap, 1928].

 $^{{}^{5}}$ [Goodman, 1963], page 545.

⁶I can vouch for this as far as I am concerned.

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 incorrect results in formalizing such definitions.⁷ Much of [Carnap, 1936 1937] is devoted to presenting and examining this problem.

Rather than devising an extension of Frege's logic capable of solving this problem, Carnap dropped the requirement that these predicates should be explicated by definitions. Carnap wasn't happy about this relaxation, which in effect abandons dispositionals as part of a well organized logicist project, but perhaps he is right that this still leaves room for an adequate formal explication of the scientific target. From a linguistic point of view, however, Carnap's abandonment of definitions is not available—without a definition of 'T_EXable' in terms of the verb 'T_EX', we can't explain the productivity of the '-able' construction.

The analysis of dispositionals is a difficult logical problem, which was not, I think, solved adequately by later conditional logics in [Stalnaker and Thomason, 1970] and [Lewis, 1973]. 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'. Counterexamples along the lines of [Goodman, 1955] can be adapted to show that a formalization of 'flammable' using a "selection function" conditional is inadequate. The fact that a piece of paper will not burn if it is heated in the absence of oxygen doesn't show that the paper is not flammable. However, logics that deal with normality offer some hope of a solution to Carnap's problem of defining dispositionals. Such logics have only become available with the development of nonmonotonic logics.

For the last thirty-five years or so, we have seen steady steady, cumulative progress on projects that seek to provide logical theories of specific problematic constructions—the literature on the logic of conditionals is a good example of work of this kind. Much of this progress has been made not by philosophers, but by linguists and computer scientists. Works like [Dowty, 1979] and [Link, 1983] (by linguists) and [Davis, 1991] (by a computer scientist) illustrate the point that the task of providing tools for logicist projects is now undertaken by many disciplines. Unfortunately, the larger philosophical community has tended to neglect work in this area by linguists and computer scientists, despite the compelling need in philosophy for new analytical tools.

These developments provide a basis for entertaining more ambitious logicist projects, that aim at formalizing large-scale domains of linguistic and philosophical interest. The project described in this paper provides an example. It also serves to illustrate how a historically important, challenging philosophical problem can be illuminated using methods from nonmonotonic logic (a contribution from computer science) and the theory of eventuality structure (a contribution from linguistics). It relies heavily on the work of Mark Steedman, who works in both linguistics and computer science.⁸

It can also be seen as part of a linguistic project concerning the meanings of complex or

⁷It is equally natural to use the word 'when': 'A spliceable rope is one that will be spliced when you try to splice it', as well as 'A spliceable rope is one that will be spliced if you try to splice it'. The formulation with 'when' points to an important aspect of the formalization problem: time is involved.

⁸See [Steedman, 1998].

derived words.

4. Linguistic logicism

A clear logicist tradition emerged from the work of Richard Montague, a philosopher who (building to a large extent on Carnap's work in [Carnap, 1956]) developed a logic he presented as appropriate for *philosophy logicism*. This idea didn't lead to useful philosophical projects, and I myself doubt that philosophy is a promising topic for formalization, although large-scale formalizations of other domains often have a strong philosophical component. However, Montague's work did suggest and inspire logicist projects In linguistics.

Montague motivates his logical framework in [Montague, 1969] with a problem in the semantics of derived words: 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 formal ontology, and of the logical framework that goes with it, 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.

Logicians other than Montague—not only Frege, but Carnap in [Carnap, 1956] and Church in [Church, 1951]—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 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 [Montague, 1973] was the most influential.

The impact of this work has been more extensive in linguistics than in philosophy. Formal theories of syntax were well developed in the early 1970s, 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,⁹ a project that is also suggested by a natural train of thought in logicist AI.¹⁰

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 some of those who developed Montague's framework turned their attention to lexical problems, and a body of the later research in Montague semantics—especially David Dowty's early work in [Dowty, 1979] and the work that derives from it—concentrates on semantic problems of word formation, which of course is an important part of lexical semantics.¹¹

5. Formalizing common sense

Due to the influence of John McCarthy, a group of *common sense logicists* has emerged within the logically minded members of the Artificial Intelligence community. McCarthy's views have been strongly and consistently expressed in a series of papers beginning in 1959.¹² The idea is that we will not know how to build algorithms that express intelligent behavior until we have an explicit theory of the core phenomena of intelligent thought; and the term 'common sense' is merely a way of indicating the phenomena in question. In practice, the research of the AI logicists is preoccupied with much less ambitious formalization tasks having to do with specialized sorts of reasoning such as planning and temporal reasoning. But formalizing common sense remains as an important high-level goal for most of us.

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 involves situating science in the world of experience to explain the reasoning that goes into the testing and application of theories; see [McCarthy, 1984] for explicit motivation of this sort. For extended projects in the formalization of common sense reasoning, see [Hobbs and Moore, 1988] and [Davis, 1991].

⁹That 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.

 $^{^{10}}$ See, for example, [Shoham, 1993].

¹¹This 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.

 $^{^{12}}$ See the papers collected in [Lifschitz, 1990].

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 Mc-Carthy'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 [Dowty, 1979] 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 [Ginsberg, 1987] for a good early guide to the field of nonmonotonic reasoning and its early development. For subsequent developments, some good book-length treatments have become available, including [Antoniou, 1997, Brewka *et al.*, 1997, Schlechta, 1997]. Also see the relevant chapters of [Gabbay *et al.*, 1994].

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.

- Circumscription is relatively conservative from a logical point of view. For instance, its language is simply the language of classical first-order or higher-order logic, and the local semantics of expressions—their satisfaction conditions in a model—are left unchanged. This makes it relatively easy to use circumscription as a development tool.
- It is a straightforward matter to convert Montague's formalism into a circumscriptive theory.
- The more sophisticated versions of circumscription provide an explicit formalism for dealing with abnormalities.¹³ I believe that such a formalism is needed in the linguistic applications.

This version of the paper is designed to be understandable without going into technicalities. In particular, to understand the ideas behind circumscription, readers need only to know the following things.

- 1. A number of *abnormality predicates* are introduced into the language.
- 2. In defining logical consequence, attention is restricted to models in which the abnormalities are simultaneously minimized, while certain terms (the ones that are deemed independent of the abnormalities) are held constant, and certain other terms are allowed to vary.

¹³See [Lifschitz, 1988].

- 3. This has the effect of taking only certain "preferred models" into account. A theory Γ circumscriptively implies a consequence A if A is true in all the preferred models of Γ .
- 4. These preferences can be constrained by an explicit "abnormality theory" using the predicates.

7. Thesis

The following is an appropriate and illuminating logicist project.

To use a nonmonotonic version of Montague's Intensional Logic, combined with specialized domains dealing with eventuality types, plurals, and mass nouns, as the means of formalizing the logical relations between the meanings of semantically related words.

I try to make a case for this idea by illustrating it with several case studies. This version of the paper will contain only one such study. But readers familiar with lexical semantics should be able to see that the techniques can readily be generalized to other cases.

8. Case studies

The first case study (and the only one presented in this abbreviated version) has to do with words involving the suffix 'able'.

8.1. The *-able* suffix

According to [Bauer, 1983][p. 28]:

there are, however, a number of processes which are usually considered to be derivational and which do display semantic regularity: consider, for example, the formation of English adjectives in *-able* from transitive verbs. This is probably a case where there are no gaps in the derivational paradigm (any transitive verb can act as the base) and the adjectives are semantically regular, meaning 'capable of being Ved' (where V is the verb in the base).

The semantic picture is not as simple as this. In fact, 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. While it is possible to formalize separate instances in a class of semantically similar group, it can be difficult to forumulate a general template for the entire class.
- 5. There are exceptional patterns.

I have tried to address these complexities by assembling a corpus of words and looking for similarities and patterns. The following is a report on my current thinking about the semantic patterns that arise from this practice.

Typically, the VERB+*able* construction works with transitive verbs V that are broadly telic.¹⁴ Often, these verbs have three characteristics. (i) They correspond to procedures that are in the normal repertoire of actions of human agents. (ii) Often there are normal or standard ways of initiating these actions, and there is an end state associated with the performance of the actions. (iii) Often, this end state can be evaluated as to its success in aiding the execution of a plan. In the simplest, and, I think, easiest to analyze cases, all of these features are present.

8.1.1. Case 1a: the occasional conditional pattern

In the first, and simplest semantic pattern, the meaning of the derived adjectival form is that when a test action is appropriately applied to an individual, the individual will enter the outcome state associated with V. The compounds *startable* and *insurable* illustrate this. A car is startable if it will start when you go through the repertoire of actions associated with starting a car. (There is room for variation in these initiating actions; a car that is startable for someone with jumper cables might not be startable for someone without them.) And a car is insurable if applying for insurance will make it insured.

This semantic pattern is occasional—that is, it is bound to the specific occasion or situation in which it is applied. If I try to start my car, and exhaust the normal and available methods without success, the car is unstartable. Maybe the car was startable yesterday, will be startable tomorrow, and normally is startable—still, it is unstartable now. On the other hand, even in the occasional sense, it is consistent to say 'The car was startable, but I wasn't able to start it'. This would be true, for instance, if I was ignorant about the procedures for starting this particular car, or if I had lost the key, or if I couldn't find the car, or if I couldn't get into it. Startability is a capability of the car, and is separated from capabilities of the agent that are taken into account in the 'able to' construction, and from the presence or absence of enabling circumstances that are somehow independent of the starting process.

As an appendix to this paper, I have attached a corpus of about 200 *-able* compounds, along with my judgment about which semantic patterns they conform to. Out of these examples, 133 were classified as belonging to Pattern 3.

In this version of the paper, I will present example formalization instances for specific compounds of each type. The goal, of course, is to provide general formalization schemata for each semantic pattern. I'm less sure about the schemata than about the instances, so I do not pursue that goal here.

 $^{^{14}}$ There are exceptions even to the rule that construction applies only to verbs. Consider *knowledgeable*, *palatable*, and *impressionable*, for instance.

Formalization of *startable*:

(1a)
$$\forall x [Startable(x) \leftrightarrow \\ \forall e [[OCCUR(e) \land INITIATE(Start)(e) \land AFFECTED(e) = x \\ \land \neg AB(INITIATE(Start))(e)] > \\ \exists e' [TELIC(e') \land Start(x, e') \land INCEPTION(e') = e \land CULMINATES(e')]]$$

Explanation of the formalization:

Note that the formalization assumes an ontology of eventualities of the sort made popular by Davidson and others. The conditional > is the "subjunctive" conditional of [Stalnaker and Thomason, 1970]. Such a conditional is needed because, for example, a car is startable even if no attempt is made to start it. The abnormality predicate AB(X), where X is a type of initiating action, applies to those initiating actions that are appropriate, complete, and correct for the type. In the case of starting a (modern, automatic transmission) car, a nonabnormal initiating event would include (at least) inserting the ignition key and turning it fully, with the car in a neutral gear. The OCCUR predicate is true (at a time and world) of an eventuality if the eventuality actually occurs at that time and world. For each predicate P expressing an appropriate eventuality type T (such as the type of starting events), INITIATE(P)(x, e) is the eventuality type of initiating events for T. The AFFECTED function picks out the affected individual of an eventuality. The predicate TELIC applies to telic eventualities. The INCEPTION function picks out the initiating subevent of a telic eventuality. And finally, the CULMINATION function picks out the culminating state of a telic eventuality.

Putting these things together, the formalization says that a thing is startable in case if a normal inception of a starting of that thing were to occur, then this would be the inception of a telic starting event which culminates. The formalization draws on ideas about event structure which I will not explain here; see [Steedman, 1998] for background details. Restating (1a) less formally, it says that a thing is startable in case a normal attempt to start the car would succeed in starting it.

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 this example (starting) and the next one (untying), I have made the following assumptions.

- 8.i. There is a telic event type *Start* of *starting events*.
- 8.ii. There is an event type INITIATE(Start) of initiations of starting events.
- 8.iii. Associated with any event e of type *Start* is an affected object AFFECTED(e).
- 8.iv. Associated with a telic event e is a subeventuality BODY(e), the process that connects the inception of e with its culmination or nonculminating termination. Where e' is the initiation of e, BODY(e) is is *initiated process* of e': BODY(e) = INITIATED-PROCESS(e').
- 8.v. A telic event e culminates if and only if it has a culminating subeventuality CULMINATION(e).

Although the use of eventualities in (1a) belongs to an approach that is associated with Donald Davidson, the underlying ontology that is required by the approach is very different from Davidson's; (1a) invokes a quantifier over nonoccurrent (or nonexistent) events. The approach to ontology that I am assuming, and the formalization style that I prefer, is similar in many ways to the one advocated by Jerry Hobbs; see, for instance, [Hobbs, 1985].

The formalism assumes the viability of a parallel project not only in philosophical ontology, but in the formalization of common sense knowledge about actions. Whatever you may think of Gibsonian psychology, a large part of common sense knowledge concerns the *affordances* of objects in the environment, of the opportunities they provide for goal-directed action. The knowledge that is needed to supplement the account of *-able* presented here has to do with the consequences of actions. Work on planning by the A.I. community has shown that—at least in limited domains—this knowledge can be incrementally formalized, and that the formalizations will support automated goal-directed reasoning. I believe that this phenomenon is typical. Reflections on the meanings of compound words often lead directly to fundamental questions in ontology and common sense reasoning. And the inventory of word formation mechanisms (in Indo-European languages, at least) serves as a very good way of organizing and systematizing studies in these areas.

The formalization avoids two problems that attach to informal accounts of dispositionals and similar constructions, and that plague many less formal philosophical analyses. Take, for instance, the following formulation: "A car is startable in case it will start if you try to start it."

The first problem is what is meant by 'you' here. It can't mask a simple universal quantifier over agents, because the existence of agents who don't know how to start a car doesn't prevent the car from being startable. More can be said about what sort of agents are appropriate, but the process of explicating this is open-ended, and incorporating the explication in the definition of startability for a car will make it difficult to provide an account of startability in general. This is an instance of what is known (in A.I.) as *the qualification problem*. Any analytic philosopher is familiar with this problem, though as far as I know there is no general term for it in philosophy. Perhaps the primary purpose and usefulness of a nonmonotonic logic is that it provides a means of addressing this problem.

The formalization in (1a) avoids any reference to agents—this is made possible by appealing to eventualities. But in fact this doesn't solve the problem of appropriate agents, which would have to be developed in providing an abnormality theory for the INITIATE(*Start*) predicate. Nonmonotonic logic doesn't do away with the need to address such considerations, but it does provide a modular approach to axiomatization that allows them to be backgrounded and dealt with incrementally.

The second, and more difficult problem, has to do with the role of time in the explication of the *-able* suffix. The culminating state in which the car is running occurs after the initiating event. But how much of a delay is tolerable? Take a case in which a street is impassible because of roadwork. If you try to pass through the street, you will eventually get through, because the roadwork will be completed in a few days. Nevertheless, the road is impassible.

Actually, this problem is endemic. I first noticed the problem in Jonathan Bennett's analysis of teleological laws in [Bennett, 1976] (see [Thomason, 1978]), but it is likely to occur in just about any philosophical analyses that uses the conditional. Bennett's analysis is unusual in that, to some extent, it recognizes the problem and tries to deal with it explicitly. To take a simpler, hypothetical example, suppose a philosopher says that a poison is lethal in case you will die if you ingest it. Again, the 'you' in the explication is generic; this

points to the need for a theory of a normal poison-ingester, perhaps in combination with a nonmonotonic logic. But the problem of time is particularly acute in this case, since any normal poison ingester will eventually die. If 'will die' means "will die at some future time," the analysis is horribly wrong.

Bringing time explicitly into the analysis doesn't seem to help, because chronometry is precise in a way that doesn't match the meaning of these temporal conditional constructions. Poisons are a good example—some lethal poisons are slow acting, some are instantaneous. This feature of poisons is more or less independent of whether the poison is lethal. Many initiating actions, in fact, have delayed effects arising from a prolonged process that may actually need to be monitored and maintained by the agent. Consider untying a knot, or opening a window, or digging a trench. This reflection, in fact, suggests an elaboration of (1a) that is required for cases where the initiated process needs to be maintained.

Formalization of *untieable*:

(1a') $\forall x [Untieable(x) \leftrightarrow \forall e [[OCCUR(e) \land INITIATE(Untie)(e) \land AFFECTED(e) = x \land \neg AB(INITIATE(Untie))(e) \land \exists e' [INITIATED-PROCESS(e, e') \land \neg AB1(BODY(Untie))(e')]] > \exists e'' [TELIC(e'') \land Untie(x, e'') \land INCEPTION(e'') = e \land BODY(e'') = e' \land CULMINATES(e'')]]$

To paraphrase (1a'): something is untieable if initiating and maintaining an untieing process on that thing in a normal way would result in its being untied.

Formalizations (1a) and (1a') capture the intuition of "affectedness" by requiring that the initiating action (the cause) and the resultant state (the effect) be parts of a single telic eventuality. This idea is, I think, intuitive. Its success depends on an ontology of events that does not generate arbitrary, *ad hoc* telic eventualities, such as a telic eventuality whose inception is the ingestion of a poison by an agent and whose culmination is the agent's death from pneumonia 30 years later. On the other hand, we need to have enough occurring telic eventualities to underwrite the necessary causal connections. I am assuming here that an ontology meeting these requirements can be axiomatized by insisting that telic eventualities are restricted to a limited (but large) variety of types—those specified by telic verbs. Again, the problem is not dismissed out of hand, but is relegated to another portion of a general theory of common sense metaphysics. The aim is not to dismiss genuine formalization problems, but to devise a modular strategy for dealing with them.¹⁵

The formalism does not depend on a general theory of causality; in fact, it is explicitly designed to avoid the need for such a theory. I don't believe that Dowty's invocation in [Dowty, 1979] of a general causal construction was successful, even in accounting for seemingly explicit causal constructions like the *-en* suffix in verbs like *lengthen*. Despite the many advances, in recent philosophy and A.I., that have illuminated the concept of causality, it is not clear that a single, unified theory has emerged from this work that is capable of being applied to natural language semantics. I prefer approaches like the one taken here, that avoid explicit use of causality in favor of less general ontological and metaphysical notions.

¹⁵It is John McCarthy who is primarily responsible for this methodological insight. See, for instance, [McCarthy, 1986, McCarthy, 1989].

The working hypothesis, I suppose, is that natural language causality can be accounted for using event structure and a nonmonotonic logic accompanied by an abnormality theory. This approach, of course, is perfectly compatible with general accounts of causality—but I suspect that these general accounts will belong to areas of metaphysics that have little to do with language.

This theory of *-able* may provide another argument for a theory of conditionals along the lines suggested by Stalnaker, one on which conditional excluded middle is valid. Conditional excluded middle provides a simple explanation of why 'unstartable' means "incapable of starting," rather than "capable of not starting."

In some examples, the culminating state is somewhat *ad hoc* and contrived. For instance, to explicate 'bearable' along the lines of Pattern 1a, you have to say that a process of bearing (up under) a hardship culminates in a state of the hardship's being borne. It is easy to get accustomed to such contrivances, and as far as I can see they are not a serious problem.

In some cases the formalization patterns produce the right result, but do so in a way that seems unnatural. Take 'answerable', for instance. The analysis I recommend, along the lines of Pattern 1a, says that a question is answerable in case a normal attempt to answer it will result in the question's being answered. This isn't actually wrong, but it's far more natural to say that a question is answerable if there's a readily available answer. 'Soluble' (in the sense that a problem is soluble) is also like this. Applying the analysis to 'enjoyable' provides a slightly peculiar result because it's strange to say that a process that culminates, say, in the enjoyment of a movie begins with an attempt to enjoy the movie. In one extreme case, the analysis may yield the wrong result. We are taught that a theorem is provable or a number is factorable not if a (normal) attempt to prove the theorem or to factor the number would succeed, but in case there is a proof or there is a factorization. It seems that uses of *-able* that are influenced by mathematics have to be treated as special cases, maybe even as anomalous cases. I don't think that this is a serious problem.

8.1.2. Case 1b: the dispositional conditional pattern

Some compounds with *-able* take on a dispositional rather than an occasional meaning. Take 'adjustable', for instance. An adjustable table may resist adjustment, perhaps because the adjustment screws are rusted. But even though trying to adjust such a table wouldn't result in its being adjusted, the table nevertheless remains adjustable. I formalize this dispositional sense of *-able* with a more typical use of abnormality predicates.

Formalization of breakable:

(1b)
$$\forall x | Breakable(x) \leftrightarrow$$

 $\forall e[[OCCUR(e) \land INITIATE(Break)(e) \land AFFECTED(e) = x$
 $\land \neg AB(INITIATE(Break)(e)] >$
 $\exists e'[TELIC(e') \land Break(x, e') \land INCEPTION(e') = e$
 $\land [\neg AB1(BODY(Break)(e)] \rightarrow CULMINATES(e')]]]$

Explanation of the formalization:

The formalism says that something is breakable in case a normally initiated "attempt" to break it would result in its being broken, provided that the attempt initiates a process that proceeds normally. Consider, for instance, dropping a wine glass in a normal way, from a distance of three feet above a wooden floor. Suppose that 50 per cent of similar glasses dropped in just this way break, over a large number of trials. Nevertheless, the cases in which the glass does not break count as abnormal in the relevant way. I take this to mean that the semantics of AB1(BODY(Break)) reflects apprehension of risk more than mere statistics.

It makes sense to say that the dispositional semantic pattern for *-able* incorporates a generic element. This could be helpful and illuminating if we had a better understanding of the semantics of generic constructions.

It may be that the difference between Patterns 1a and 1b is due to some sort of contextual variation, or that 1b is a generic usage that is always in principle available if 1a is available. These possibilities are attractive, and are suggested in some ways by the data. But at the moment, I do not know how to formalize them properly.

8.1.3. Carnap's problem

In this section I digress briefly to state the solution that this framework provides to Carnap's problem of how to define 'water-soluble' in terms of 'water' and 'dissolve'. The solution conforms to Semantic Pattern (1b).

The formalization is as follows.

Formalization of *water-soluble*:

(1b') $\forall x [Water-Soluble(x) \leftrightarrow \\ \forall e [[OCCUR(e) \land INITIATE(Dissolve)(e) \land AFFECTED(e) = x \\ \land WATER(MEDIUM(e)) \land \neg AB(INITIATE(Dissolve))(e)] > \\ \exists e' [TELIC(e') \land Dissolve(x, e') \land INCEPTION(e') = e \\ \land [\neg AB1(BODY(Dissolve))(e)] \rightarrow CULMINATES(e')]]]$

8.1.4. Case 2: the nonharmful pattern

Consider *acceptable* or *affordable*. To say that an offer is acceptable is not to say that it will be accepted if you try to accept it. Indeed, it suggests that the offer *will* be accepted, but says that the consequences of accepting it will be harmful. To say that an item is affordable is not to say that it will be afforded if you try to afford it; it is to say that the consequences of trying to afford it will not be ruinous.

As far as I can tell, all instances of this pattern are dispositional, in the sense that they do not imply the occurrent success of the initiating action. The fact that an attempt to buy an affordable item would fail (because of a hitch in the credit transaction, or because the item is reserved for another customer) has nothing to do with whether it is affordable. To say that an offer is acceptable suggests that normally an attempt to accept it will succeed—but to say it is unacceptable also carries the same suggestion. I will assume that this is not part of the semantic content of the construction.

The formalization of this pattern invokes a HARMFUL(Accept) predicate that applies to the outcome state.

(2) $\forall x [acceptable(x) \leftrightarrow \forall e \forall e' [[OCCUR(e) \land Accept(e) \land AFFECTED(e) = x \land CULMINATION(e) = e'] > \neg HARMFUL(Accept)(e')]]$

According to (2), an offer is acceptable if, were an attempt to accept it to succeed, the result would not be harmful in the relevant way.

The last qualification is important. Even if disposing of a disposable container on some occasion has irrelevant harmful consequences (food in the disposable containers in the trash can attracts bears), this does not make the container any less disposable.

8.1.5. Case 3: the *-worthy* pattern

A third productive pattern with *-able* is used to classify the outcome not just as nonharmful but as appropriate, fitting, worthy. In a few cases, constructions with *-able* and *-worthy* are synonymous (*notable*, *noteworthy*).

Take *payable*, for instance. When a creditor says that a bill is payable, he doesn't merely mean to say that an attempt to pay it will succeed, or that it will be free of harmful effects of the relevant sort. He means that it should be paid. At least one deontic logician (Alan Anderson) analyzed *ought* as meaning that the consequences of refraining from an action would be harmful, or sanctioned. But some instances of this pattern (*honorable* and *lovable*), for instance, clearly do not have a meaning of that sort. In the interests of uniformity, I'll propose a formalization that resembles that of Pattern (2), but that says that the consequences are beneficial (in the relevant way).

(3) $\forall x [payable(x) \leftrightarrow [\forall e \forall e' [OCCUR(e) \land Pay(e) \land AFFECTED(e) = x \land CULMINATION(e) = e'] > BENEFICIAL(pay)(e')]]$

8.1.6. Resemblances among the patterns

Informally, you can state the differences between the three major patterns tidily: where V is a telic verb, V-able means that if V is attempted on an individual, the result will be successful, where 'successful' can mean (1) occurrent, (2) not harmful (in the relevant way), or (3) beneficial (in the relevant way). But first, this is not quite accurate (I have made Patterns 2 and 3 conditional on performing, not on attempting to perform the action), and second it is difficult to formalize the patterns so that the differences depend on the variation of a single simple parameter. At the moment, I think it is best to formalize different cases, in the hope that appropriate general patterns will eventually emerge.

9. Conclusion

I have said that this is part of a larger project. To get a sense of how the thesis articulated in Section 7 fares, it will be necessary to investigate a number of cases in considerable detail. I have developed partial studies of the following cases.

- 1. Some causal constructions.
- 2. Agency.
- 3. Some denominal verbs.
- 4. The *-er* of normal function, as in *fastener*.

The notion of normalcy (e.g., the normal use of an instrument to perform a task, or the normal way in which an initiated process can be expected to evolve) shows up in many of these constructions—these patterns, I believe, make a strong case for the use of a nonmonotonic logic in formalizing this domain.

It may well turn out that the thesis itself is too open-ended to be demonstrated in the form of a well demarcated formalist project, and that it will have to be confined to a core of "causal" or "action-related" constructions. The suffix *-able* may be misleading as a case study, since it is relatively productive, and logical and semantic ideas that have been developed for other purposes turn out to be relevant to its formalization. Suffixes like *-ful*, *ity*, and *-ous* may well prove to be very different in this respect; in the worst case, many word formation constructions would turn out not to have interesting, illuminating formalizations. And in many cases, the work that needs to be done seems to be more a matter of ontology than of logic.

However the thesis itself fares, I encourage the research community to pursue the formalization of word meaning. Compared to the interpretation of syntactic constructions, the study of word meaning is neglected and underdeveloped. I hope I have shown in this paper that this neglect can't be due to the absence of good linguistic data and of opportunities to deploy and test interesting, significant theories. In fact, this is a challenging and important field. Its neglect has left a large gap in semantic theories, and filling this gap offers many promising areas of investigation.

At the same time, work in this field clearly has to depart from the study of a narrowly conceived linguistic faculty that is more or less independent of other forms of knowledge and other reasoning activities. Theories of word meaning naturally and inevitably become involved with issues in ontology, knowledge representation, metaphysics, and the study of common sense reasoning. And it is only because of recent advances in the formalization of these matters that we can hope to develop theories that meet the semantic challenge. One exciting thing about this field is that work on it not only draws from linguistics, computer science, and philosophy, but that it is a very promising way of advancing each of these subjects. In particular, I believe that the systematic study of word meaning is one of the best ways of achieving new insights in metaphysics—and it can also be useful in knowledge representation.

In the end, we can hope obtain a much better understanding of the common sense world and how it is reflected in language and reasoning through cooperative work that uses the best ideas of all the relevant disciplines. I recommend this cooperative approach to anyone who is interested in projects of this kind.

Bibliography

- [Antoniou, 1997] Grigoris Antoniou. Nonmonotonic Reasoning. The MIT Press, Cambridge, Massachusetts, 1997.
- [Asher and Morreau, 1991] 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.
- [Bauer, 1983] Laurie Bauer. English Word-Formation. Cambridge University Press, Cambridge, England, 1983.
- [Bennett, 1976] Jonathan Bennett. *Linguistic Behaviour*. Cambridge University Press, Cambridge, England, 1976.
- [Boutilier, 1992] 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.
- [Brewka et al., 1997] Gerhard Brewka, Jürgen Dix, and Kurt Konolige. Nonmonotonic Reasoning: An Overview. CSLI Publications, Stanford, 1997.
- [Carnap, 1928] Rudolph Carnap. Der logische Aufbau der Welt. Weltkreis-Verlag, Berlin-Schlactensee, 1928.
- [Carnap, 1936 1937] Rudolph Carnap. Testability and meaning. *Philosophy of Science*, 3 and 4:419–471 and 1–40, 1936–1937.
- [Carnap, 1956] Rudolph Carnap. Meaning and Necessity. Chicago University Press, Chicago, 2 edition, 1956. (First edition published in 1947.).
- [Church, 1951] Alonzo Church. The need for abstract entities in semantic analysis. Proceedings of the American Academy of Arts and Sciences, 80:100–112, 1951.
- [Cross, 1986] Charles B. Cross. 'Can' and the logic of ability. *Philosophical Studies*, 50:53–64, 1986.
- [Davis, 1991] Ernest Davis. Common Sense Reasoning. Morgan Kaufmann, San Francisco, 1991.
- [Dowty, 1979] David R. Dowty. Word Meaning in Montague Grammar. D. Reidel Publishing Co., Dordrecht, Holland, 1979.

- [Gabbay et al., 1994] 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, 1994.
- [Ginsberg, 1987] Matthew L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, Los Altos, California, 1987. (Out of print.).
- [Goodman, 1947] Nelson Goodman. The problem of counterfactual conditionals. *The Journal of Philosophy*, 44:113–118, 1947.
- [Goodman, 1955] Nelson Goodman. Fact, Fiction and Forecast. Harvard University Press, 1955.
- [Goodman, 1963] 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.
- [Hobbs and Moore, 1988] Jerry R. Hobbs and Robert C. Moore, editors. *Formal Theories* of the Commonsense World. Ablex Publishing Corporation, Norwood, New Jersey, 1988.
- [Hobbs, 1985] Jerry R. Hobbs. Ontological promiscuity. In William Mann, editor, Proceedings of the Twenty-Third Meeting of the Association for Computational Linguistics, pages 61–69, Morristown, New Jersey, 1985. Association for Computational Linguistics, Association for Computational Linguistics.
- [Lewis, 1973] David K. Lewis. Counterfactuals. Harvard University Press, Cambridge, Massachusetts, 1973.
- [Lifschitz, 1988] Vladimir Lifschitz. Circumscriptive theories: A logic-based framework for knowledge representation. Journal of Philosophical Logic, 17(3):391–441, 1988.
- [Lifschitz, 1990] Vladimir Lifschitz, editor. Formalizing Common Sense: Papers by John McCarthy. Ablex Publishing Corporation, Norwood, New Jersey, 1990.
- [Link, 1983] Godehard Link. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In Rainer Bäuerle, Christoph Schwarze, and Arnim von Stechow, editors, *Meaning, Use, and Interpretation of Language*, pages 302–323. Walter de Gruyter, Berlin, 1983.
- [McCarthy, 1984] 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.
- [McCarthy, 1986] John McCarthy. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence*, 28(1), 1986.
- [McCarthy, 1989] 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.

- [Montague, 1962] 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.
- [Montague, 1969] Richard Montague. On the nature of certain philosophical entities. *The Monist*, 53:159–194, 1969.
- [Montague, 1973] 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.
- [Schlechta, 1997] Karl Schlechta. Nonmonotonic Logics. Springer-Verlag, Berlin, 1997.
- [Shoham, 1993] Yoav Shoham. Agent oriented programming. Artificial Intelligence, 60(1):51–92, 1993.
- [Stalnaker and Thomason, 1970] Robert C. Stalnaker and Richmond H. Thomason. A semantic analysis of conditional logic. *Theoria*, 36:23–42, 1970.
- [Steedman, 1998] Mark Steedman. The productions of time. Unpublished manuscript, University of Edinburgh. Available from http://www.cogsci.ed.ac.uk/~steedman/papers.html., 1998.
- [Thomason, 1978] Richmond H. Thomason. Review of Linguistic Behaviour, by Jonathan Bennett. Synthése, 39:141–154, 1978.
- [Thomason, 1991] 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.

Appendix List of Compounds in *-able*

Compound	Pattern(s)]	Compound	Pattern(s)
acceptable	2,3		curtailable	1a
accessible	1a		decipherable	1a,1b
adjustable	1b		decontaminable	1a
admissible	2		defensible	1a,1b
adoptable	2		deiceable	1a
advisable	3		desirable	3
affordable	1a		detectable	1a,1b
affectable	1a		digestible	2
amenable	IRREG		dispensable	2
answerable	1a,1b		disposable	2
approachable	2		distinguishable	1a,1b
arguable	2		doable	1a
applicable	1b		downloadable	1a,1b
available	IRREG		drinkable	2
avoidable	1a,1b		drivable	1b
bearable	1a,1b		durable	IRREG
beatable	1a		edible	1b
believable	2,3		enjoyable	3
breakable	1b		emulsifiable	1a,1b
breathable	2		excusable	3
burnable	1a,2		expressible	1b
cancelable	1a		equitable	IRREG
capable	IRREG		expendable	2b
certifiable	2,3		extractable	1a
chargeable	3		executable	1b
chewable	1a		factorizable	1a
cleanable	1a		favorable	3,IRREG
consumable	1a,2		findable	1a
constructible	1a		fixable	1a
conceivable	1a		flexible	1a
combatable	1a,2		flyable	1a
communicable	1a,2		foldable	1a
considerable	IRREG		forgettable	1a,2,3
containable	1a		forgivable	1a,2,3
comfortable	IRREG		formalizable	1b
corruptible	1a		formidable	IRREG
curable	1a,1b		foreseeable	1a
cultivatable	1a		generalizable	1a

Compound	Pattern(s)	Compound	Pattern(s)
glueable	1a	matchable	1a
graftable	1a	modifiable	1a
growable	1a	movable	1a
heatable	1a	mutable	IRREG
helpable	1a	notable	3
hireable	1a	negligible	3
honorable	3	offsetable	1a,2
imputable	3	openable	1a,1b
impressible	1a	observable	1a
improvable	1a	palatable	IRREG
inclinable	IRREG	palpable	IRREG
includable	1a,2,3	paintable	1a
instructible	1a	passable	1a,IRREG
insurable	1a	persuadable	1a
intelligible	IRREG	playable	1a,1b
interoperable	1a	pollutable	1a
interruptable	1a,2	printable	2
implantable	1b	provable	1a
implementable	1a	probable	IRREG
improvable	1a	profitable	IRREG
impressible	1a	portable	IRREG
impressionable	IRREG	publishable	2
imaginable	1a	payable	3
killable	1a	programmable	1a
knowledgeable	IRREG	preferable	3
knowable	1a,1b	payable	3
launchable	1a	potable	IRREG
learnable	1a,1b	predictable	1a
liable	IRREG	presentable	3
liftable	1a	preventable	1a,1b
likable	3	prolongable	1b
intelligible	IRREG	penetrable	1a
livable	3	programmable	1a
interoperable	1a	projectable	1a
interruptable	1a,2	protectable	1a
insurable	1a	quantifiable	1a
implantable	1b	questionable	3
lovable	3	readable	1a,1b
machinable	1a,1b	reachable	1a
manageable	1a,2	realizable	1a
manipulable	1a	reasonable	IRREG
marketable	1a	rebuttable	1a

Compound	Pattern(s)		
rechargeable	1a		
reclaimable	1a,1b		
reclosable	1a,1b		
recognisable	1a		
recollectable	1a		
reconfigurable	1b		
redeemable	1a,1b		
reducible	1a		
refinable	1a		
refutable	1a		
reliable	1a,2		
remarkable	3		
renewable	1a,1b		
rentable	1a,1b		
repairable	1a		
reputable	3		
respectable	3		
responsible	IRREG		
retractable	1a,1b		
returnable	1a,1b		
reusable	1a,1b		
reversible	1a		
revivable	1a		
revokable	1b		
scalable	1a		
searchable	1a,1b		
securable	1a		
sensible	IRREG		
sizable	IRREG		
soluble	IRREG		
solvable	1a,1b		
startable	la		
steerable	1a		
stoppable	1a		
suitable	IRREG		
supposable	1a		
suppressible	1a		
suspendable	1a		
sustainable	1a		
terminable	IRREG		
testable	la,1b		
TeXable	1a		

Compound	Pattern(s)	
tractable	IRREG	
trainable	1a	
transferable	1a	
transportable	1a	
understandable	1a	
usable	1a	
variable	IRREG	
venerable	IRREG	
valuable	3	
viewable	1a	
vulnerable	IRREG	
walkable	1a,2	
washable	1a,2	
wearable	1a	
weldable	1a	
workable	1a	