A semantic interpretation of modality in counterfactual conditionals

Lori Coulter

University of Illinois at Urbana Champaign

Abstract

This project uses a model theoretic possible worlds approach, resembling classical Formal Semantic treatments (e.g. Kratzer 1977, 1981, 1989; Lewis 1973; Veltman 2005), to interpret counterfactual conditionals with respect to a world of evaluation. The model theoretic semantics are linked with the typed feature structures in an HPSG syntax (Pollard & Sag 1994) implemented in TRALE (Penn 2004) with the Constraint Language for Lexical Resource Semantics (Penn & Richter 2004, 2005). Sets of possible worlds interact with constraints on world knowledge and constraints defining counterfactual evaluation. The truth value for a counterfactual is returned to the grammar relative to a context of evaluation.

1 Introduction

An accurate semantic interpretation of counterfactual conditionals, and modals in general, depends, to a large extent, on world knowledge. It is not possible, for instance, to interpret whether the sentence *You must run a lot*, allows the inference that the addressee ran in the past or not to be true in the actual world without world knowledge. If the speaker's knowledge is accurate, it might be possible to interpret such an inference as true in the actual world. But, on another reading, it could be a command that the addressee start running a lot for the sake of his health, in which case, he might never have run before in his life and the inference that he ran in the past would be invalid. Trying to determine whether the intended interpretation is the deontic modality of the latter interpretation or the epistemic modality of the former depends on a number of contextual factors. Some of these contextual factors, which can help circumscribe the relevant subset of world knowledge needed to make valid inferences, reside in the sentence or dialogue surrounding the modal (Coulter unpublished; Crouch 1993).

The implementation described in this paper uses propositions in a model as the framework for conducting inference. The grammar is used in conjunction with the model to determine what type of inferences to look for and which propositions are relevant. The propositions, which are first order predicate-like representations of the sentences licensed by the grammar, form sets. Sets of such sets are constrained by their conformity to various knowledge base axioms. The argument structure of verbs, for instance, in the grammar can assist in locating a background context via their encoding in the mapping from the compositional semantics of the grammar to

the propositions of the knowledge base. The use of a grammar in conjunction with knowledge base axioms and abstract propositions allows modals to have a more substantive interpretation than what would be provided by a grammar with compositional and lexical semantics alone. Although modals require such a knowledge base for accurate interpretation, they are not the only natural language phenomenon that can benefit from it: it would also facilitate natural language interpretation in what are typically considered to be intensional contexts.

The particular implementation described below represents a proof of concept and not a large-scale collaborative effort, though suggestions are given in the conclusion of how it could be incorporated into such a project. Though it supports some degree of modal interpretation in general, it focuses on achieving accuracy with counterfactual conditionals in a restricted domain. In a larger grammar, the domain restriction would limit sets of propositions in the interpretive component to those relating to the domain of the discourse under consideration. In this project only a limited domain is constructed in order to focus on the issue of modal interpretation. Modal entailment is a difficult phenomenon to characterize and remains unresolved in broad coverage entailment projects as well (MacCartney et al. 2006; Roxana Girju p.c., Mark Sammons p.c.).

The domain used in the current project is less complex than what would be required to deal with the issues and conundrums that have arisen in much of the formal semantic literature on counterfactual interpretation (e.g. Kanazawa et al. 2005, Kratzer 1989, Tichy 1976, Veltman 2005). Specifically, this project does not work with that intricate of a premise set. The propositions considered relevant for counterfactual evaluation are domain specific and somewhat general. While the implementation approach resembles relevance logic style reasoning about conditionals (e.g. Shapiro 1992), it is somewhat more intuitive and does not give the relevant world knowledge the same status as other propositions. Rather, world knowledge works as an abstract statement, similar to an axiom schema, that all plausible worlds for a given interpretation must be capable of satisfying before evaluation can precede.

The model of abstract propositions and knowledge base axioms is represented in a Prolog interpretive component. In this implementation, the Prolog interpreter works in conjunction with an HPSG syntax (Pollard & Sag 1994) implemented in TRALE (Penn 2004). The compositional semantics of the grammar is the Constraint Language for Lexical Resource Semantics (CLLRS) presented in Penn & Richter (2004, 2005). CLLRS was developed with the capability of supporting inference and entailment in typed feature structures especially with respect to semantic ambiguities involving scope and quantification. CLLRS distinguishes between lexical semantics and compositional semantics and is designed to handle the later leaving the former to standard HPSG constraints on the CONTENT value of signs (Penn & Richter 2004). But CLLRS is linked systematically to the grammar allowing some interaction between the two. This interaction is necessary for the general disambiguation of modals in that grammatical features of the sentence such as verb tense and the person value of the subject noun phrase can indicate which type of

modality is involved (Crouch 1993; Coulter unpublished). Section 3 describes the role of the grammar and various semantic components in more detail.

The modal interpretation uses a possible worlds approach that limits the worlds in which the basic meaning of a modal, for example, 'necessity' or 'possibility', is evaluated in order to achieve a representation of possibility or necessity relative to a certain context. Generally speaking, the approach resembles traditional formal semantic approaches to counterfactual modality (Kratzer 1977, 1981, 1989; Lewis 1973; Veltman 2005), in that it evaluates counterfactuals in a background that the antecedent helps to define as relevant.

Counterfactual conditionals are evaluated to be plausibly true if they are supported in a subset of the knowledge base that conforms to the appropriate world knowledge axioms. The subset of propositions in which evaluation takes place is located by the subset's compliance with the world knowledge axioms which are circumscribed primarily by the antecedent's propositional form. The interpreter uses world knowledge axioms in conditional rules as constraints in order to delimit the relevant set of possible worlds. In the disambiguation of modals in general, locating the proper contexts would require multiple sentences of the discourse, but counterfactual conditionals constitute a more tractable subcase of the problem in that the antecedent provides sufficient information to locate a context of evaluation.

The result of a query concerning the truth value of a counterfactual in the program should be intuitively plausible to a human user. In addition to getting an intuitively accurate result, the axioms that define the context of evaluation should constitute the most restrictive deviation from actual world knowledge that accommodates the antecedent. Following the basic intuitions of Lewis' (1973) account of counterfactuals, it locates the closest world to the actual world in which the antecedent is true and evaluates the counterfactual as true if the consequent is also true in that world.

For example, the sentence *If Maurice fell off of the tightrope he would've hit the ground hard* is true, generally speaking, if, in a situation nearly identical to the actual one, it follows from Maurice's falling off of the tightrope that he hits the ground hard. It is not a plausible counterfactual, for instance, if there exists a net in the actual world which would clearly catch him and prevent his collision with the floor. Similarly, the interpreter described below evaluates the truth of a counterfactual relative to the contextual background that the antecedent indicates is relevant. Presumably, additional inferences can be conducted in the same context or in a context located by a combination of the counterfactual context and additional discourse information.

2 Disambiguating Modals: The Role of World Knowledge

Kratzer (1977) observed that a modal verb, such as *must*, can be described as having a consistent core meaning of necessity, if the necessity is relative to a particular set of contextually indicated facts. An unambiguous paraphrase of a sentence with

must would include a phrase beginning with *in view of* followed by an indication of the relevant information. For instance, the sentence *Leor must leave the U.S.* could be paraphrased as *In view of the restrictions on visas, Leor must leave the U.S.* or *In view of what is known about Leor's interests abroad and long absences from work, Leor must leave the U.S.* The first paraphrase would be true if, in all possible worlds in which visa restrictions are as they are in the actual world, Leor leaves the U.S. The second would be true if in all possible worlds in which certain facts about Leor are known to be true, Leor leaves the U.S. Unfortunately, the context is rarely stated this concisely in natural language.[1] Other characterizations of modality are treated similarly in that they impose restrictions on the set of possible worlds in which a modal is evaluated or use accessibility relations to impose similar restrictions.

The difficulty posed for implementation is that such treatments, while providing deep analyses of the model theoretic semantics, assume a knowledge base. When trying to capture the intuitions computationally, questions of how to limit the set of possible worlds requires some simulation of the knowledge base. It is necessary, for instance, to get 'the set of all worlds in which visas work as they do' from sets of propositions and a set of world knowledge axioms. Trying to do this in an open domain is a daunting task, so it is an empirical question whether the Formal Semantic treatments of modals are feasible with an artificial knowledge base. The current paper constitutes an attempt to illustrate how the deeper principles of the formal treatments could work in a domain specific case.

It is important to note that the problem of modal disambiguation is far from being solved with broad coverage statistical methods. In textual entailment tasks, modals have been recognized to play a significant role and no entirely satisfactory way of handling them has been developed. In order to deal with the effects of modals, they have been characterized in relation to other modals or the absence of modals in sentences which are sufficiently similar otherwise (MacCartney et

---

[1]Even if the modal can be disambiguated between deontic and epistemic, there have been various attempts to model the context of evaluation, none of which is ideal for drawing the type of common sense inferences that broad coverage entailment projects attempt to capture. In the deontic case, the implication that the event will happen has been described as holding in all worlds in which people do as they are commanded (e.g. Heim 1982; Kratzer 1981), and the actual world is not considered to be one of these. It would be hard to define, in a realistic knowledge base, what the likelihood of actual world entailments (in a loose sense of the word) would be. Similarly, epistemic modals have plausible common sense conclusions to the degree that the speaker's world knowledge constitutes accurate premises. The problems clearly require world knowledge, the questions concern how to represent and manipulate it in order to capture the semantics of modals. The use of probabilities with conditionals has been discussed in Kaufmann (2005) and other works by the author, but the direction intended in the current work takes a different approach, primarily in that it treats world knowledge as constraints and intends to use probability for the relation between modalized propositions and the inferences that tend to be drawn about actual world propositions (e.g. For instance, to what degree does a sentence like *Sex offenders must leave their lights off on Halloween* (from Google news) corpus used in imply *Sex offenders leave their lights off on Halloween*? This type of 'inference' will never be anything but a likelihood of the event and can at best be represented as a probability based on who is enforcing the command and who is aware of it (Coulter unpublished.)

al. 2006; Girju & Roth, unpublished; Girju p.c.). For example, a modal with the core meaning of 'not possible' is predicted to entail a similar sentence without the modal, but retaining the negation (i.e. not actual). Though there has been some success with this method, it fails in a number of contexts. It is not the case, for instance, that the sentence, *There couldn't have been another shooting* entails that there was not another shooting, which is what the inferences in MacCartney et al. (2006) would predict. It can only be concluded from the sentence that, in view of what the speaker knows, it does not seem possible. The system does not take into account the fact that the conclusion is drawn from a falty premise if the speaker's world knowledge is inaccurate. A move towards implementation of a slightly deeper treatment of modality could shed light on these problems as well.

2.1 Counterfactual Conditionals as a Special Case

Counterfactual conditionals present a special case of modal interpretation in which the context of evaluation is partially identified by the antecedent. Counterfactuals form a good testing ground for locating modals in a context because the antecedent helps determine which world knowledge is necessary. The implementation described in this paper contains propositions which are generated from licit permutations of the constituents of parseable sentences from an HPSG grammar. Counterfactuals are evaluated relative to proposition-world pairs which fit certain restrictions defined based on world knowledge axioms and semantic overlap with respect to the set of actual world propositions. Given a counterfactual sentence, the program interprets it relative to the appropriate set of propositions and returns a truth value.

Counterfactual conditionals contain an antecedent clause which the speaker believes is false relative to the actual world. In order to represent the meaning of a counterfactual, it is not insightful to say it is automatically rendered true just because the antecedent is false. A counterfactual conditional with an antecedent that is false in the actual world is not considered to be true if the consequent is not true in a world like the actual world in which the antecedent is true. The counterfactual above, repeated in (1) serves as an illustration.

1. If Maurice fell off the tightrope, he would hit the ground hard.

The usual interpretation is that Maurice did not fall off the tightrope, but, imagining he had, he would have hit the ground. Part of the interpretation of counterfactuals requires that the evaluation is relative, not to the actual world, but to a similar world in which the antecedent is true. But there is the additional complexity that the world of evaluation must be similar enough to the actual world that the consequent follows fairly directly. Sentence (1) would be false, for example, if the speaker were aware of a large net spanning the floor.

In order to model this complex situation, Lewis provides a system of 'spheres'. A sphere, introduced to accommodate modal interpretation, is a set of worlds that

meet a contextually defined restriction. For example, the sphere of accessible worlds for the actual world in a sentence such as *Unsupported mass must fall* is the set of worlds which are elements of all true propositions pertaining to the laws of nature.

A system of spheres is used to define relative closeness of worlds to a given world, for instance, the actual world. The set of propositions which have the actual world as an element (and, so by definition, are true in the actual world) are true with respect to the sphere containing only the actual world. This sphere is the center of the system of spheres. A larger sphere contains those worlds that differ minimally from the actual world and a yet larger sphere contains worlds that differ minimally from those, and so on. The system is closed under union and intersection and for any two spheres, one is a subset of the other. Moving out from the singleton set in the center sphere, each sphere contains the worlds which differ minimally from the previous sphere.

The result of the system of spheres is that relative closeness to the actual world is defined with set theoretic concepts; there is no need to use world knowledge as part of the theoretical construct that indicates which worlds are closer than others, it is encoded by propositions. By this description, worlds less like the actual world are in more distant spheres. For instance, the worlds in which gravity doesn't exist are more distant from the actual world than worlds in which cats do not exist because the effects of the former are of more consequence relative to the propositions which hold in the actual world than the latter. The result is that the accessible sphere for a counterfactual conditional is the smallest sphere which contains a world in which the antecedent is true. This system supports the intuition that counterfactuals are not restricted in acceptability with respect to how distant the antecedent world is from the actual world, but from whether or not, given the antecedent, the consequent follows.

A system of spheres is difficult to implement because the task of determining contextual restrictions on accessibility spheres is re-allocated to the task of ensuring that all the correct worlds are elements of the propositions conforming to general world knowledge axioms. With respect to accessibility relations, the present implementation resembles Kratzer's (1981) representation of ambiguity in modal verbs. Kratzer's theory not only involves an ordering relation on possible worlds, but also a 'contextual background' that specifies which of the ordered worlds are relevant for the evaluation of the proposition in the scope of the modal verb. The accessibility relations in this implementation are based on a combination of world knowledge, as described by axiom schemas, and ordering of worlds fitting the schemas by overlap of the propositions true in them with those true in the actual world. This program locates the sphere of evaluation for a counterfactual in much the same way that it is located in a system of spheres, capturing the intuitive meaning of counterfactuals, but world knowledge does not need to be as comprehensibly specified.[2]

_____

[2]As an anonymous reviewer pointed out, it would be best if the implementation took into account

3 The Grammar Design

This section discusses the interpretive component of the semantics in relation to the syntax of the HPSG. While lexical semantics have standardly been located in the TFSs of the grammar and expressed in the SYNSEM value of the entry, there have been multiple approaches to incorporating a compositional semantics, as well as contextual information and model theoretic semantics (see, for instance, the summary in Copestake et al. 2006:324). This particular project will divide the semantics among the lexical semantics, the compositional semantics, and the modal logic interpreter. Other projects, such as Ginzburg & Sag (2000), have included contextual information, such as this project would allocate to the modal logic interpreter, in the TFSs of the grammar. Penn & Richter (2005) also suggest using event variables in the TFS grammar as well as including intensional types in the type signature. Possible worlds, as used for modals in this model, would then presumably involve combining propositions with world arguments in the compositional semantics. [3]

A considerable number of possible worlds are necessary to represent counterfactual interpretation. Any method of representing modals would have to consider substantial portions of hypothetical information, even if it were restricted to a discourse context. This project keeps track of the information outside of the TFSs of the grammar. A separate module with Prolog rules contains the worlds and allows logical interpretation of the first order logic like formulas in that module.

The interpretive module is ideal for allowing one to derive inferences from a disambiguated language with some reduction of the richness of structure represented in the grammar. Determining which division of labor is best for inferencing depends on what type of specification derives the most accurate inferences for a particular phenomenon and how much disambiguation or abstraction from natural language allows it to be best carried out.[4] The current implementation divides the semantics among three components, the lexical, the compositional, and the possible worlds semantics.

The HPSG in TRALE allows queries which parse the syntax of the grammar

---

the problems with Lewis' (1973) and Kratzer's (1981) account. For instance, those problems dealt with in Kratzer (1989), Tichy (1976), Veltman (2005) and others, some of which are summarized in Condoravdi & Kaufmann (2005) and Kanazawa et al (2005). Because the treatment is still rather generally applied, the nuances described in the referred to works do not affect interpretation in the current project. As the project deepens, these facts need to be accounted for.

[3]It seems that there would need to be a set of rules to build small models on the fly that had sufficient complexity to allow modal interpretation. Then the implications and world knowledge could be written as constraints. It would likely be necessary to remove at least some of this from the TFSs, which starts to look a lot like what is done here, but with world labels on propositions in the grammar. This seems like a viable modification of the current proposal. It is important to note, as well, that CLLRS supports model theoretic interpretations (Penn & Richter 2005), the component described in this project is designed to deal with possible worlds semantics.

[4]The particular division of information into the interpreter here is somewhat similar to the AKR of Bobrow et al. (2007) which adds an additional level of abstraction for various inferences beyond the compositional semantics.

as well as queries producing compositional semantic parses using CLLRS. When a user enters a modal query in the grammar, the query is sent from the grammar to the Prolog interpretive module. First the query is mapped to propositions in a first order predicate logic like form. Then it is evaluated relative to the appropriate context. The query results are then returned to the grammar along with a semantic parse of the expression.

3.1 The Grammar

The syntax of modals and conditionals in the grammar is intended to be fairly uncontroversial. This work doesn't make any bold claims about their syntactic properties. In fact, the interpretive component should support grammar designs which allow various syntactic analyses, provided that they relate straightforwardly to the compositional semantics.[5]

The common modals in counterfactual conditionals are *could have* and *would have*. The past tense modals subcategorize for a main verb which subcategorizes for its arguments. The connective 'if' has a lexical entry which combines two clauses with finite verbal heads for conditional sentences, and a lexical entry which combines a clause with a finite verbal head and a clause with a modal head to represent counterfactual conditionals. The head of conditionals is 'if' and it subcategorizes for two saturated phrases. Alternative syntactic representations of conditionals, for instance, with one of the clauses subordinate, could just as easily have been mapped to predicates in the interpretation. Modal interpretation relies primarily on the mapping between CLLRS semantic values and propositions in Prolog. In this sense, the implementation is flexible with respect to the syntax of the grammar where modals and conditionals are concerned.

The compositional semantics, CLLRS, introduces a type signature for semantic typing and the attribute LF of signs. The typing is declared in the signature of the grammar and valid compositional semantic parses satisfy standard requirements on type interaction. A portion of the type declaration for the current implementation is shown below:

```
semtype [john,location,time]:(e).
semtype [temp_phrase, loc_phrase]: (e -> t).
semtype [change_loc]: (e -> e -> e-> t).
```

The elements in square brackets to the left of the colon are the abstract arguments of the compositional semantics. Their type is declared to the right of the colon.

---

[5]The syntactic analysis here might be unduly influenced by the semantic properties I was interested in capturing. If this is the case, it would only require modifing the mapping of the syntax to CLLRS, not the mapping of the CLLRS representation to the first order predicate logic forms of the modal interpreter. Unless it were shown to be the case that the accurate syntactic form could be shown not to work with the compositional semantics given here. Then this project would require a revision of the first order predicate logic forms as well.

The TRALE implementation constitutes an HPSG with CLLRS that the modal component works with. It has lexical entries that combine using phrase structure rules as a purely theoretical HPSG would. Built into the program are queries that provide the parts of the grammar. For instance, the query `lex` will give the TFS for the word following it. The query `rec` will give a syntactic parse and `srec` a compositional semantic parse as shown below. In this way, the grammar can be queried and the licit constructions displayed. These TFS's contain the lexical semantics of the constituents as well as the compositional semantics, in the case of the semantic parse.

The LF value of Penn & Richter's (2004, 2005) semantics consists of semantically typed expressions, square brackets, parenthesis, and ˆ. The semantics is encoded in the type hierarchy as *lrs* which has the attributes of INCONT, EXCONT, and PARTS. In a given grammar, possible values for these parts are built from the typed expressions declared in the semantic type declaration of that grammar's signature. The EXCONT value is preceded by theˆsymbol and represents the maximal projection of the particular semantic expression and the INCONT, the semantic expressions in square brackets, are the semantically selected arguments of the head (see Penn & Richter (2004) for a more in-depth description of their use and interaction in HPSGs).

In the grammar implemented here, the lexical expression *in* semantically selects a locational phrase as shown in its lexical entry below.

```
in ---> (synsem: category:
            (head:preposition:temp,subcat:
            [(synsem: category: (head:case:obl,subcat:[]),
                    content:index:X),
                    lf:@sem(^P))]),
                    content:(temporal_spatial:X)),
            lf: @sem(^loc_phrase([(P)]))).
```

The LF feature has the value of *in* taking a variable as its INCONT value which is instantiated by the LF value of the noun phrase combined with it in the parse. For instance, when the grammar implementation parses a phrase such as *in Dallas*, the compositional semantics resulting is the combined semantic value of the expressions: `^in[location]`. The LF value for 'in' above combines with the LF value for 'Dallas'.

The semantic parse of a modalized sentence combines similarly when queried with the `srec` command, as shown below:[6]

```
?srec[john,would_have,arrived,in,dallas,at,three_o_clock].

^modal\\
(A:[change_loc(B:[C],D:[loc_phrase(E:[F])],G:[temp_phrase(H:[I])])])
```

---

[6]The parse is entered as a list of expression and is not intended to represent any of the structure in the HPSG. The structure is shown in the query results.

In order for the input to result in a compositional semantic parse, the combination of expressions must be compatible with the typing declared in the grammar's signature. The compositional semantics of modals and conditionals involve giving modals scope over the verbal head of the proposition. In the case of counterfactual conditionals, the modal takes scope over the consequent. The compositional semantics of *if* takes the the antecedent and consequent as semantic arguments. The CLLRS semantic forms provide a semantic parse of each sentence in the grammar and these forms can interact with scope of negation or quantification to capture semantic ambiguities. When a possible worlds semantic analysis is needed, the compositional semantic parse is mapped to a propositional representation that forms part of the sets of possible worlds. But the modal logic component is only involved when it is necessary for modal interpretation. A compositional semantic parse without an interpretation can be obtained directly from the HPSG component, but formulating the query for interpretation gives a modal logic interpretation as well as calling a compositional semantic parse in the CLLRS of the HPSG.

CLLRS provides a compositional semantics that is quite closely tied to the syntax in the grammar. By relating the compositional semantics' value for the attribute LF to the first order predicate logic forms of the modal logic interpreter, there are a series of links between the grammar and the knowledge base that can be exploited to describe the role of natural language expressions in modal disambiguation. [7]

Although modal verbs have a straightforward compositional semantics in the grammar, the lexical semantics of modals is somewhat difficult to specify since their meaning outside of a context is somewhat vague.[8] This is part of what makes disambiguation of modals a problem and why additional interpretation is helpful.

The modal intepreter is queried in the TRALE grammar and additional information about modal semantics is given. A counterfactual conditional with *could have* in the consequent is true relative to a world and a sphere if the sphere is accessible to the antecedent and the antecedent and the consequent are true in the world and there is no closer world in which the antecedent is true. The next section will describe this component in more detail.

4   The Model Theoretic Component

The model theoretic component consists of sets of propositions that represent possible worlds and world knowledge as constraints on those sets. The interpretation of counterfactual conditionals works with constraints on what constitutes a plausible world of evaluation given the antecedent. Given the set of plausible worlds of evaluation, the counterfactual with *could have* is evaluated to be true if the conse-

---

[7]There are other examples where this could be helpful. For instance, with discourse connectives. They could similarly be defined in the grammar and given constraints in the interpreter module for their meaning in a text.

[8]The WordNet lexicon, for instance, which is used for lexical disambiguation, does not contain modals since they can not be disambiguated with synsets.

quent is true in a world in which the antecedent is true, and there is no world more similar to the actual world, with respect to world knowledge axioms, in which the antecedent is true and the consequent is false. In the case of *would have*, it is true if there does not exist a world in which it is not the case that the antecedent is true and the consequent false, given the set of worlds identified by the world knowledge axioms. The rule for necessity also checks that there are no worlds more similar to the actual world in which the antecedent is true.

The possible worlds are built using proposition and world pairs in the interpretation as arguments of a predicate `is_true/2`. This works similarly to a characteristic function from propositions to $\{0,1\}$ with a world label on each function.[9] If a proposition does not hold in a world, this is represented by the absence of that proposition-world pair in the `is_true/2` predicate. A number of inferences are stated besides those relevant for counterfactual evaluation. For instance, from any world in which some event takes place, it is possible to derive that the event could take place.

A general mapping from the CLLRS compositional semantics to the propositional forms is written as a conditional rule which derives the propositional form from the semantic parse. Using this method has the additional advantage that propositions can form models built on the fly from user queries. However, in the current state, it just allows the propositions into the knowledge base. If they are not in the `is_true/2` predicate, they can not satisfy the counterfactual conditional query.

A possible world is defined as the set of propositions that are in a pair with that world.[10] The accessibility relations between worlds are defined by a number of interacting constraints. First, there are a sequence of constraints on the type of world knowledge each proposition represents. Given a domain of flight patterns, the first type of flights are Valid Flights.

Valid Flights constitute the flights which actually occur. In practical applica-

---

[9]A representative sample was permuted for the initial implementation. For certain arguments of a semantic type all possibilities were permuted to ensure a greater degree of objectivity. In other cases, the more absurd propositions were not listed for all possible arguments. So, in its current state, it is possible to get both *If John arrived in Dallas at noon then he could've departed from Chicago at noon* and *If Marry arrived in Dallas at noon then she could've departed from Chicago at noon* to fail to be possibly true counterfactuals in a query. But the latter fails because it is not in the set of true propositions for any world and the former because it is in the set of true propositions but, given the information in the antecedent, it is implausible because there are worlds more closer to reality in which John arrived in Dallas at noon and it doesn't follow in those worlds that he departed from Chicago at noon. This does not affect the practical results of the query, but could if working with larger premise sets than the antecedent. In order to get an interpretation that works equally well for any parseable sentence in the grammar, the possible worlds model needs to be implemented more efficiently. A number of methods exist for doing this, which are currently being explored in conjunction with model checking options.

[10]In order to represent this in a more traditional way, each proposition would have to correspond to the set of worlds it occurs in a proposition world pair with. This would help implement Formal Semantic treatments more literally, but I don't see any particular advantage in doing this in the current implementation.

tions, these could be built from an actual schedule in a database. The Valid Flights only include the flights in the actual world, not the individuals taking them. This arrangement allows conditionals such as *If John departed from Dallas at noon, he arrived in Chicago at 6:00* to be evaluated as true in the actual world if the event describes Valid Flights.

The Valid Flights form a subset of the Ordinary Flights and the specific subset differs based on the actual world facts represented.[11] But the set of Ordinary Flights, excepting engineering developments in increased airplane speed, does not change on a real world temporal axis. It consists of all flights which take a reasonable duration from one location to another.

The next set is not a superset of Ordinary Flights, but is disjoint from it. It is the set of Odd Flights which circle and land in the same location, but don't violate any laws of nature. They are conceivable flights in the actual world, but not the expected pattern in this domain.

Getting intuitively more distant from the actual world, there are Absurd Flights which violate basic laws of nature. For example, they allow someone to arrive and depart from the same place at the same time.

Of course, expanding this to an open domain is a large amount of work. However, it is promising that, if all possible worlds were generated from the sentences of these domain specific examples, there would be 2^{16} worlds and intuitive evaluation of counterfactuals is achieved with twelve world knowledge axioms.

4.1 Locating a Context in the Model

The accessibility relations are defined by the predicate `is_accessible` which takes as arguments, a constant which names a labeled sphere, then two world variables and a variable for a proposition.

Accessibility relations are defined in terms of the relevant world knowledge constraints. Given any two worlds, the two worlds are accessible to each other in a sphere if the proposition under evaluation conforms in those worlds to the stated constraints on world knowledge.[12] There are fourteen of these spheres defined. The first one simply states that the actual world is accessible to itself for any proposition which is true in it.[13]

```
is_accessible (sphere1,(wa,wa),Prop):-
is_true(Prop, wa).
```

For any proposition in the actual world, the fact that it is true in the knowledge base in that world is enough to derive that the actual world is accessible to itself for that proposition. This corresponds roughly to Lewis' (1973) center sphere containing only the actual world.

---

[11]A particular arbitrary set was chosen for this project.

[12]This accessibility relation is stated symmetrically, but could be stipulated not to be if it were necessary.

[13]The Prolog code is read with capital letters representing variables. The conditions occur to the right of :- with x :- y read as 'x is derivable from y.'

Moving out from the center, speaking figuratively in the system of spheres analogy, the constraints are used to allow a greater degree of accessibility. Though stated with variables for each world here, the counterfactual rule specifies the first world as the actual world. The more general statement, however, is useful for other natural language phenomena.

In sphere 6, one world is accessible to another for a proposition given that the proposition is an Ordinary Flight in each of the worlds. Since Valid Flights are Ordinary Flights, the relation is satisfied by the actual world proposition as well. This way of representing accessibilities gets some of the Lewis-style effect of having concentric spheres. It differs, however, in that the world knowledge axioms work as constraints on what worlds are accessible to each other. The particular selection of axioms limit the valid interpretations for a given sphere. In a more literal Lewis-style program, the axioms would have to be stated as propositions that, if removed, affect enough of the other propositions to constitute a significantly different set of sets of worlds. The design of the current program gives the axioms their intuitive prominence by stating them as constraints on sphere membership.

The mimicking of concentric spheres is not present in some spheres since Odd Flights are disjunct from Ordinary Flights. The intuition behind this is that other natural language expressions, like modal subordination, can locate a context as one of the Odd Flight supporting, or other non-actual spheres and reason about what would follow in such worlds, but the inferences do not hold in the actual world without the hypothetical premises. There remain in the system, however, Odd Flight containing spheres which allow accessibility to the actual world. These spheres are necessary for counterfactual evaluation.

The outermost sphere allows any proposition to be accessible to the actual world. A plausible counterfactual is not located here unless the consequent is equally absurd. This sphere captures cases like *If John were able to be in two places at the same time, and he departed from Dallas at noon, then he could've departed from Chicago at noon*.

An ordering, which is not reflexive, `is_immediately_closer/2`, is defined on the spheres as well as a transitive relation `is_closer/2`.

In order to evaluate a counterfactual, the program uses the following code, where \+ is 'not':

```
poss_true_counterfactual(Prop1, Prop2, wa, Sphere):-
(is_accessible(Sphere,(wa,W2),Prop1),
(is_true(Prop1, W2),
(true_cond(Prop1,Prop2,W2),
is_closer(wa,OtherSphere,Sphere),
\+ poss_true_counterfactual(Prop1,Prop2,wa,OtherSphere).
```

This rule derives that a counterfactual with *could have* is possibly true for the antecedent and consequent in the actual world relative to a sphere if that sphere is accessible for the world in which the antecedent is true. This condition locates an antecedent-containing sphere. The next one checks that the consequent also follows in that sphere. Last, a condition ensures that there is no sphere closer than

the one which instantiated it. Necessity for *would have* is defined similarly in terms of 'not possibly not'.

The `is_closer` line of code will satisfy the variable OtherSphere with the actual world if nothing else is in between the sphere of evaluation and the sphere containing only the actual world. This means that counterfactuals with true antecedents are evaluated to be implausible, contra Lewis'(1973) account. In order to capture the intuitions that counterfactuals which are true in the actual world reduce to material conditionals, a rule can be written which derives material conditionals from counterfactuals true in the actual world using the `true_cond/3` predicate.

The end result is that a query concerning the plausibility of a counterfactual is satisfied if the consequent holds in a world in a sphere nearest the actual world in which the antecedent holds. For example, when a user types in a query concerning the counterfactual *If John departed from Chicago at noon, he could have arrived in Dallas at 4:00*, it is satisfied as plausible. This evaluation is intuitively accurate even though there is no flight pattern on the actual itinerary under consideration in which a plane goes to the two locations at the stated times. But because it is a normal flight pattern, that is to say, nothing takes too short of a time and the claim conforms to the laws of nature, it is satisfied in the sphere of Ordinary Flights and is deemed plausible. A query such as that above is a valid counterfactual, but the non-modalized equivalent is only true if it is an actual world flight pattern. In this way, the module supports counterfactual and non-counterfactual conditional inferencing.

5 Conclusion

The model presented here constitutes a domain specific proof of concept of how traditional Formal Semantic insights can be implemented in such a way that inferencing about the plausibility of counterfactual conditionals is possible. The implementation described here invites development in either breadth or depth.

In the direction of broader coverage models, the implementation would need to be grafted into a larger grammar and made to work on broader domains. It is promising that a relatively small number of world knowledge axioms are needed when used in combination with ordering relations on propositions. It is possible that this way of handling world knowledge could have advantages in broad coverage systems. The knowledge bases used in the PASCAL RTE challenge entries, for instance, are generally built by the competitors using some degree of hard-coded world knowledge axioms. The world knowledge in the current project works as axiom schemas that propositions can satisfy. By considering only some of them to be applicable for each sphere, they limit the interpretations available in that set of worlds.

In order to get general modal interpretation, it is necessary to develop means of getting lexical semantic information to interact more intricately with the interpreter. The methods used by Bobrow et al. (2007) illustrate a promising method

to emulate if the current implementation were to develop in the broad coverage grammar direction.

As far as developments in the interpretive component are concerned, it is important to expand the temporal representations in the model. A larger knowledge base for conducting inferences can be used with model checking techniques to handle natural language entailments in larger models. Current developments involve looking into representing more complex modal and temporal relations in the Prolog interpreter. And, after implementing such developments, applying model checking with the Maude model checking module, which promises to be particularly helpful with the temporal dimension (Clavel et al. 2007).[14] Along with these developments of the interpreter, greater depth can be achieved and more of the nuances of counterfactual interpretation recognized in the Formal Semantic literature can be supported. Particularly, a more precise model theoretic characterization can be developed and some of the useful intuitions from Premise Semantics and related developments can be implemented for more complex inferences.

References

[1] Patrick Blackburn and Johan Bos. *Representation and Inference for Natural Language*. CSLI, 2005.

[2] D. G. Bobrow, C. Condoravdi, R. Crouch, V. de Paiva, L. Karttunen, T. H. King, R. Nairn, L. Price, and A. Zaenen. Precision-focused textual inference. In *ACL- PASCAL Workshop on Textual Entailment and Paraphrasing; 2007 June 28-29; Prague, Czech Republic*, 2007.

[3] Johan Bos. Exploring model building for natural language understanding. In *Proceedings of ICos-4*, 2003.

[4] Rui Pedro Chavez. Dynamic model checking of discourse representation structures with pluralities. In *Proceedings of the 7th International Workshop on Computational Semantics (IWCS7), Tilburg University, The Netherlands*, 2007.

[5] Manuel Clavel, Francisco Duran, Steven Eker, Patrick Lincoln, Narciso MartiOliet, Jose Meseguer, and Carolyn Talcott. All about maude - a high-performance logical framework: How to specify, program and verify systems in rewriting logic. In Gerhard Goos et al., editor, *Lecture Notes in Computer Science*, volume 4350. Springer, 2007.

[6] Cleo Condoravdi and Stefan Kaufmann. Modality and temporality. *Journal of Semantics*, 22:119–128, 2005.

---

[14]Other model checking applications have been recommended, including those discussed in Blackburn & Bos (2005) and Bos (2003), and Chavez's (2007) tool for dynamic model checking for first order logic.

[7] Lori Coulter. Semantic disambiguation of the modal verb 'must'. 2007 project.

[8] Richard Crouch. *The temporal properties of English conditionals and modals*. PhD thesis, Computer Laboratory, University of Cambridge, 1993.

[9] Jonathan Ginzburg and Ivan Sag. *Interrogative Investigations*. CSLI, 2000.

[10] Roxana Girju and Dan Roth. Investigating semantic entailmnet in 'v to v' constructions. 2005 project.

[11] Irene Heim. *The semantics of definite and indefinite noun phrases*. PhD thesis, University of Massachusetts, 1982.

[12] Makoto Kanazawa, Stefan Kaufmann, and Stanley Peters. On the lumping semantics of counterfactuals. *Journal of Semantics*, 22:129–151, 2005.

[13] Stefan Kaufmann. Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2):181–231, 2005.

[14] Angelika Kratzer. What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1977.

[15] Angelika Kratzer. *Words, Worlds, and Contexts: New Approaches in World Semantics*, chapter The Notional Category of Modality. Walter de Gruyter, 1981.

[16] Angelika Kratzer. An investigation of lumps of thought. *Linguistics and Philosophy*, 87(1):3–27, 1989.

[17] David Lewis. *Counterfactuals*. Harvard University Press, 1973.

[18] Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, and Christopher Manning. Learning to recognize features of valid textual entailment. In *Proceedings of the 21st International Conference of the North American Chapter of the Association for Computational Linguistics, (HLT-NAACL 2006)*, pages 41–48, 2006.

[19] Gerald Penn. Balancing clarity and efficiency in typed feature logic through delaying. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. (ACL 2004), Barcelona, July 2004*, pages 240–247, 2004.

[20] Gerald Penn and Frank Richter. Lexical resource semantics: from theory to implementation. In Stefan Müller, editor, *Proceedings of the HPSG04 Conference*. CSLI, 2004.

[21] Gerald Penn and Frank Richter. The other syntax: approaching natural language semantics through logical form composition. In H. Christiansen et al., editor, *Constraint Solving and Language Processing. Lecture Notes in Artificial Intelligence*, volume 3438, pages 48–73. Springer, 2005.

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

[23] Stuart Shapiro. Relevance logic in computer science. In Alan R. Anderson, Jr. Nuel D. Belnap, and J. Michael Dunn, editors, *Entailment*, volume Volume II. Princeton University Press, 1992.

[24] P. Tichy. A counterexample to the stalnaker-lewis analysis of counterfactuals. *Philosophical Studies*, 29:271–273, 1976.

[25] Frank Veltman. Making counterfactual assumptions. *Journal of Semantics*, 22:159–180, 2005.