

## A Semantic Interpretation of Modality in Counterfactual Conditionals.

**1. Introduction.** The logical form values in HPSG typed feature structures (TFSs) have been implemented with CLLRS (Penn & Richter 2004, 2005) in such a way that they represent semantic ambiguities and provide an interpretation that supports inference and entailment. This paper shows how an HPSG grammar implemented in TRALE can be related to a Prolog interpretive component by means of LF values to provide a truth conditional evaluation of counterfactual conditionals.

Some of the most difficult inferences to model are those involving modal verbs. The sentence *Leor must leave the U.S.* is ambiguous between the deontic reading in which some authority has commanded Leor to leave and the epistemic reading in which it is surmised that Leor leaves the country from time to time. With the deontic reading, there is no implication that Leor has ever left the U.S., but the epistemic reading implies that it is likely that he has left the U.S. at some time in the past. Inferences with respect to the future likelihood of the event vary accordingly.

Kratzer (1977) observed that a modal verb, such as *must*, has a consistent core meaning of necessity, but the necessity is relative to a particular set of contextually indicated facts. An unambiguous paraphrase would include a phrase beginning with *in view of* followed by an indication of the relevant information. The sentence above 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 in all possible worlds in which visa restrictions are as they are in the actual world and the second would be true in all possible worlds in which certain facts about Leor are known to be true. Unfortunately, the context is rarely stated this concisely in natural language.

In textual entailment tasks, modals have been recognized to play a significant role. 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or the absence of modals in sentences which are sufficiently similar otherwise (MacCartney et al 2006; Girju & Roth, unpublished.). 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. Though there has been some success with this method, it fails in certain 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, only that, in view of what the speaker knows, it does not seem possible.

Counterfactual conditionals present a special case of modal interpretation in which the context of evaluation is identified by the antecedent. Counterfactuals form a good testing ground for locating modals in a context because they do not require information beyond the sentence level. The implementation described in this paper contains propositions which are generated from 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 sphere and returns a truth value.

**2. Lewis' (1973) Theory of Counterfactual Conditionals** 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 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 introduces 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 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. 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 largest accessible sphere for a counterfactual conditional is the smallest sphere which contains a world that the antecedent is true in. 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 properly labeling propositions with respect to their elements and those of other propositions. 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 comprehensively specified.

**3.1. The Lexicon and PS Rules of the Grammar** The grammar describes single clause sentences with or without modal verbs which combine as the antecedent and consequent clauses of conditionals. The modal verbs *could have* and *would have* are sufficient to express counterfactual meaning and to represent the modal relations of possibility and necessity.

The past tense modals subcategorize only for a main verb and are ordered in the output parse by a PS rule that combines the modal with the verb and elements on the verb’s subcat list. The TRALE command ‘rec[John,could\_have,arrived,in,dallas,at,noon]’ results in the TFSs in (2) below. It is standard in semantic approaches for the modal to take scope over the finite clause, but this implementation is flexible in that it could accommodate a syntactic parse with the modal in the auxiliary position.



lf: @sem(^loc\_phrase([(P)]))).

The symbol ‘^’ is used with the semantic value of the maximal projection of the head of a phrase called the excont. In this case, the semantic value of the prepositional phrase is ‘loc\_phrase’. The square brackets indicate subterms of an expression. In (4), the loc\_phrase has a term, indicated by the variable ‘P’, as a subterm. The variable is co-indexed with the LF value of the noun phrase on the preposition’s subcat list.

When the command ‘srec [in,dallas].’ is entered, TRALE semantics combines the LF value of each lexical item as in (5).

5.        ^loc\_phrase(A:[location])

This is the semantic representation of a prepositional phrase and the variable ‘P’ is instantiated to ‘location’, which is the LF value of ‘dallas’. The variable ‘A’ indicates a sub-structure of the loc\_phrase which has ‘location’ as a subterm.

The saturated prepositional phrase can combine with the verb that subcategorizes for it. The prepositional phrase is then a subterm of the verb phrase as in (6).

6.        ?-srec[john,arrived,in,dallas,at,noon].  
          ^change\_loc(A:[john],B:[loc\_phrase(C:[location])],D:[temp\_phrase(E:[time])])

The semantics of the verb is now the excont value of the verb phrase.

Modal verbs take a sentence as their semantic argument and are treated as the semantic head as shown in (7).

7.        ?-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])])])

Propositions in the Prolog interpreter are derived from licit combinations of LF values for saturated VPs. Possible worlds are constructed from these propositions. The modal verbs and multiclausal srecs do not occur in possible worlds but are defined as modality and implication respectively on the sets of propositions. They are, however, parseable with the ‘srec’ command and have LF values in their TFSs.

**4. Counterfactual Interpretation** Counterfactual interpretation is implemented as follows. 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 consequent are true in the world and there is no closer world in which the antecedent is true as described in the Prolog rule in (8).

8.        poss\_true\_counterfactual((Prop1),(Prop2),(wa),(Sphere)):-  
          (is\_accessible((Sphere),(wa),(W2),(Prop1))),  
          (true\_cond((Prop1),(Prop2),(W2))),  
          (is\_true((Prop1),W2)),  
          (\+is\_closer((wa),(OtherSphere),(Sphere))).

The body of the rule states that the antecedent should be evaluated in an accessible sphere. Once W2 is instantiated to an accessible sphere, it must be the case that both the antecedent and the consequent of the counterfactual are true in a world in that sphere. The last clause is to check that there is no sphere closer to the actual world in which the antecedent is true. The rule for counterfactuals with ‘would\_have’ is treated similarly, but defined in terms of possibility as ‘not possibly not’.

