**The Metalinguistic Use of Vague Predicates in Conditionals**

*Stefan Hinterwimmer*, *Humboldt University Berlin*

This paper deals with the semantic and pragmatic contributions of conditionals like those in (1), which have prominent readings that can be paraphrased as given in (2).

(1)    a. If I hate anything, it's bad acting.
b. If anyone was drunk at last night's party, it was Mary.
c. If any book impressed me, it was *Finnegan's Wake*.

(2)    a. What I hate most is bad acting.
b. The drunkest person at last night's party was Mary.
c. The book that impressed me most was *Finnegan's Wake*.

Crucially, the sentences are compatible with a situation where there is more than one thing that the speaker hates, more than one person being drunk at last night's party, etc. The existence of these readings is unexpected, since standard assumptions about the meaning of conditionals in combination with ones about clefts lead us to expect readings according to which the conditions in (a) and (b) are fulfilled to be the only ones available:

(a)    The speaker is not sure whether an entity exists in the world of evaluation that satisfies the predicate introduced by the antecedent.

(b)    If this is the case, there is only one such entity, and this entity is identical to the one mentioned in the consequent.

Observe that such a reading is indeed the only one that is available to the structurally parallel examples in (3):

(3)    a. If anyone can solve this problem, it's Peter.
b. If Peter bought anything at the store, it was pizza.

Intuitively, what seems to make the difference is that the antecedent predicates in (1) are vague, and thus gradable (you can hate something to a higher or lesser degree, be drunk to a higher or lesser degree, etc.), while the ones in (3) are not (a problem is either solved or not solved etc.). Crucially, in the case of vague predicates, the standard of interpretation is left open and can be fixed by the context (cf. von Stechow 1984, Kennedy 1999 on gradable adjectives), i.e. the intensity of the negative stance a person must have towards an entity in order for the pair consisting of the person and the entity to be in the denotation of *hate* can vary from context to context.

I will show that the 'unexpected' interpretation of conditionals like those in (1) can be explained as arising from the interplay of the following factors:

(A)    Since the predicate in the antecedent is vague, the standard of interpretation can be manipulated, i.e. the positive value associated with the respective predicate can be set to the maximal degree.

(B)    The respective vague predicates can be used metalinguistically within the antecedent.

(C)    Conditionals are analysed as involving universal quantification over those worlds maximally similar to the world of evaluation where the antecedent is true (Kratzer 1986, Nolan 2003; see also Lewis 1973 on counterfactuals).

(D)    The cleft structure of the sentences enforces an interpretation according to which the entity occurring to the right of the copula is the only one that satisfies the antecedent predicate.

Ignoring the factors in (A) and (B), the combination of (B) and (C) straightforwardly accounts for the 'expected' readings of the examples in (1): first, use of the pronoun *it* in combination with the fact that the main accent most naturally falls on the NP occurring to the right of the copula indicates a cleft structure. It is well known that clefts are interpreted exhaustively:

> (4)    a. It is Paul who solved all problems. #And Mary did, too.
>
> b. Paul solved all problems. And Mary did, too.

Applying the analysis of conditionals sketched in (C) to our example (1a) thus gives us the reading shown (in simplified form) in (5):

> (5)    a. $\forall w´ [R(w, w´) \wedge \exists x [hate(sp, x) (w´)] \rightarrow \iota y.hate(sp, y) (w´) = bad\_acting]$
>
> 'In all worlds that are accessible from the world of evaluation where there is an entity *x* such that the speaker hates *x*, the unique entity that the speaker hates is *bad acting*.'

Note that this reading is dispreferred for (1a), since by using an indicative conditional the speaker implicates that she is not sure whether the antecedent is true or false in the world of evaluation. This is a strange implicature in the case of (1a), however, since persons are expected to know about their own emotions.

Turning to the 'unexpected' readings, the basic idea alluded to in (A) is that vague predicates can in principle be interpreted with respect to different standards, and that these standards can be manipulated. Technically, I (roughly) follow von Stechows's (1984) analysis of gradable adjectives, and assume that also verbs like *hate* and *impress* take an additional degree argument which can be saturated by overt degree expressions like *more* or *less*. Furthermore, I make the simplifying assumption that in the absence of an overt degree expression, the respective predicate is combined with a covert morpheme *pos*, which introduces a standard of comparison in the form of a free variable $d_s$ ranging over degrees whose value is fixed by the context:

> (6)    $[[pos]] = \lambda f_{<d,<e,<e,st>>>} . \lambda y. \lambda x. \lambda w. \exists d[f(d)(y)(x)(w) \wedge d \geq d_s]$

Concerning the examples in (1), I assume that the respective antecedent predicates can be used metalinguistically within the antecedent, i.e. the worlds quantified over do not (possibly) differ from the world of evaluation with respect to some extralinguistic state of affairs, but only with respect to the way the predicates are interpreted: the value associated with *pos* is set to the maximal (reasonable) value. This gives us (7) as the denotation of (1a):

> (7)    $\forall w´ [R(w, w´) \wedge \iota d. \forall x \forall y \forall d´[d´ \geq d \wedge hate(d´)(y)(x)(w´) \rightarrow$
>
> $[[pos]]([[hate]])(y)(x)(w´)] \wedge \forall z \forall k \forall d´´[d´´< d \wedge hate(d´´)(k)(z)(w´) \rightarrow$
>
> $\neg [[pos]]([[hate]])(k)(z)(w´)] = d_{MAX} \wedge \exists x [ [[pos]]([[hate]])(x)(sp)(w´)]$
>
> $\rightarrow \iota y. [[pos]]([[hate]])(y)(sp)(w´) = bad\_acting]$

Crucially, the sentence now says that in all worlds which only differ from the world of evaluation insofar as the standard relative to which *hate* is interpreted is set to the maximal (reasonable) value and in which there (still) is some entity such that the speaker hates this entity, the unique entity that the speaker hates is identical to *bad acting*. Now note that it would be completely superfluous to raise the standard of interpretation if there was no entity such that the pair consisting of the speaker and this entity satisfies the predicate with respect to some lower standard. This has the following consequence: The sentence is interpreted as conveying that (a) there are other entities such that the pairs consisting of the speaker and the respective entity satisfy the predicate with respect to some lower standard, which is the one that is in effect in the world of evaluation, and (b) the unique entity which survives after the standard has been raised is the one the speaker hates most in the world of evaluation.

**Selected References**

Kennedy, C. (1999): *Projecting the Adjective*. Kratzer, A. (1986): Conditionals. von Stechow, A. (1984): Comparing Semantic Theories of Comparison.