DEFEASIBILITY IN CONCEPT COMBINATION: A CRITERIAL APPROACH

BRADLEY FRANKS TERRY MYERS SCOTT McGLASHAN
CENTRE FOR COGNITIVE SCIENCE
UNIVERSITY OF EDINBURGH

CONCEPT COMBINATION AND DEFEASIBILITY

What are the semantic relations within and between lexical concepts that constrain their combination? The position advanced is that they are criterial relations. Such a view has advantages over extensional and standard intensional approaches, which fail to capture adequately the characteristics of defeasibility at play during such combinations. We will examine these questions with respect to the adjectival modification of nouns.

Approaches To Concept Combination

Treatments of concept combination divide into "extensional" and "intensional" theories. The major failings of both approaches can be traced to some shared assumptions. Extensional approaches include set-theoretic treatments of concept combination: concepts are treated as unanalyzed units, and the detailed interactions of the contents of the concepts are bypassed. "Intensional" views construe concepts according to their internal properties, and concept combination is reduced to the interactive mechanism for the inheritance of properties from the inputs. Most writers postulate some partitioning of the features into, for example, "necessary" and "characteristic" types: in combinations, the "necessary"/deductive attributes of both concepts are to be inherited by the conjunction, whilst the "characteristic" attributes might be negotiated.

The shared assumptions of the two views are as follows. Firstly, the crucial attributes are those which are deductively related to the concept: as Jackendoff (1983) has pointed out, a defining-features intensional account is formally equivalent to a meaning-postulate extensional account. Secondly, and linked to this: the semantic object is a "total object" (Landman, 1986). Total objects are crucial to classical logic and realist semantics. An object is equated with the set containing *all* of its properties: these properties are the deductively-related identity conditions. Total objects are thus predicated upon the assumption of total evidence or information, even though this may be beyond the grasp of a particular language-user. Thirdly, neither view can distinguish between the conditions of application of

a concept, and the content of that concept. Since they are (total) object-centred, the assumption is that the concept can be applied only if all of its identity conditions are fulfilled; and these are usually taken as conditions on being a particular kind of object.

Standard intensional approaches have a further quality which is founded upon the above three assumptions. The characteristic properties are construed as default properties. And an object-centred approach demands that where the characteristic features of a concept obtain, so must the necessary features.

The major difference between the extensional and intensional approaches is that the former takes the necessary attributes and operates over them as a set, in terms of formal operations only. The intensional approach operates upon the attributes themselves, allowing the further use of non-necessary attributes.

The Criterial Relation

The criterial relation (henceforth, "C-relation") is a semantic relation which may hold either between two different concepts, or between a single concept and its evidential conditions. The notion derives fairly directly from the later Wittgenstein (1953). Traditionally, criteria are said to have the following properties. Firstly, the relationship between the evidence (qua criteria) and the particular claim (qua application of a concept) is somewhere between deduction (since it confers certainty), and induction (it is defeasible). Secondly, as the criteria are necessarily good evidence for a claim, they fix the semantic content of that claim. Thirdly, to have satisfaction of the criteria for a claim is consistent with obtaining further information which overturns that claim. The relation is thus defeasible. Fourthly, criteria are generally held to be multiple. In addition to criteria a concept has symptoms, or S-relations. These are inductively related features, similar to default properties in that their defeat will not alter the identity of the particular claim. So both semantic relations are inherently defeasible.

This reflects the received criterial view. The approach adopted here extends the concepts according to an anti-realist semantics. Primarily, C-relations govern the conditions of application or use of a term, rather than conditions on identity of an object (cf. Tennant, 1987). The criteria for a concept are C-related to the mental representation of the object, which is itself S-related and C-related to its attributes, and to other concepts. C-relations are the conventionally underpinned semantic relations holding between the evidence for, and the application and evaluation of, a concept. Hence, C-relations embody partial objects based upon partial information: they concern how we may reliably "go beyond the given" evidence to infer a particular kind of conceptual entity. C-relations license defeasible

extensions of partial objects to less partial ones.

This allows us to specify the major difference between S-relations and default properties; since the properties are not tied to the identity conditions of objects, where the S-relations of a concept obtain, there is no necessity that the C-relations do so. Hence, a concept may be applied on the basis of S-relations alone, but with less warrant than a C-related application.

The final difference between the standard views and the one adopted here is that the central semantic relation is content-driven and non-transitive. This issues in a type of defeasibility which is not open to the other views.

Types of Defeasibility

Three kinds of defeasibility are pertinent: rebuttal, undercutting and default. The first two involve inter-conceptual conflict and defeat, the last one intra-conceptual. But firstly, we define a

Defeater: if P is evidence for Q, R is a defeater for this evidence iff:

R is consistent with P; and

(P & R) is not evidence for Q.

Crucially, anti-realism's denial of excluded middle denies that this is equivalent to evidence for (-Q).

Rebuttal: Denial of the Claim:

Rebutting Defeater: If P is prima facie good evidence for Q, R is a rebutting defeater for this iff:

R is a defeater; and

R is good evidence for (-Q).

Now, there are two types of rebutting defeaters:

Type I: R is a type I rebutting defeater iff:

R is a rebutting defeater; and

R is good evidence to support the claim that P would not be warrantedly assertible unless Q were so. Hence, (-P).

(The second clause is, of course, a generalization of the definition of modus tollens in classical logic: -P unless Q).

Type II: R is a type II rebutting defeater iff:

R is a rebutting defeater; and

R is good evidence to deny the claim that P would not be warrantedly assertible unless Q were so. Hence, P may be assertible.

(Here, the second clause generalizes the denial of modus tollens).

Any defeasible deduction will necessarily be a type I rebutting defeater, since modus ponens supports modus tollens. Hence, set-theoretic and standard intensional views can utilize only this type. How would this operate? A defeasible concept is applicable iff we can show that *none* of the defeating conditions obtains. So certainty is equated with necessity, and necessity is read as a deductive relation. Any weaker relation is inherently doubtful. The logical possibility of doubt in any situation is equipollent with an actual grounded doubt in the current situation.

Undercutting: Denial of the C-relation:

Defeat of C-relations occurs by an:

Undercutting Defeater: If P is prima facie good evidence for Q, R is an undercutting defeater for this iff:

> R is a defeater (not a rebutting defeater); and
>
> R is a good evidence to deny the claim that P would not be warrantedly assertible unless Q were so.

(Note, in this case, the second clause cannot be a generalization of modus tollens, since we do not have an outright denial of Q).

This type of defeater, then, attacks the *connection* between P and Q rather than Q itself. The defeater is a reason for denying that we would not have the evidence unless the conclusion were true. This does not imply either -Q or -P: it denies that a claim or generalization (Q) on the basis of P would be warrantedly assertible. But the evidence itself is still assertible, and indeed the claim could still be so.

The epistemological assumptions upon which C-relations are predicated are a great deal more naturalized than their classical realist counterparts (Baker & Hacker, 1984). The obtaining of a C-relation provides certainty in the sense that it is conclusive evidence for the claim. And it is conclusive in the sense that the evidence cannot be improved upon, even if it is multiplied: and it can nonetheless be overturned. The burden of proof is also altered: on the classical view, there is a *requirement* to check and deny the (possibly open-ended) list of potential defeaters. On an anti-realist view, there is no such requirement: if the prima facie evidence supports the claim, then, if there are no available prima facie defeaters, the claim will go through - even though there is a possibility of defeat. So the logical possibility of doubt is not equivalent to the existence of an actual grounded doubt.

Default: Denial of Inductive Properties:

This is the approach exemplified in the S-relation and utilized in frame and prototype theory. A default property is either a parameter or a value of a parameter which is linked to a concept by an inductive

relation: it is a "typical" property. Where the property is overridden, the application of the concept itself is nonetheless still justified.

ADJECTIVE-NOUN COMBINATION

Noun Phrase Constructions

One type of construction which illuminates the different kinds of defeasible semantic relations are NPs where an A modifies the N, as in sentence-type 1, below; the crucial questions concern how this modification operates, especially in cases of conflict of A and N properties, and whether we can derive sentence-types 2 and 3 from the type-1. The sentence-types can be rendered:

1: This is a (A)(N) 2: This is A 3: This is a N

It will be evident that the derivation of type-3 sentences is equivalent to the derivation of the superordinate category of the N from that of the NP, with the preservation of the sortal type. For *privative* adjective-types, the inferences to type-2 and type-3 sentences are problematic: the A functions neither predicatively nor attributively. Consider the following case:

1: This is a fake Hogbin 2: This is fake 3: This is a Hogbin

Here, we cannot infer 3 from 1, although 2 is sensical. In terms of classical logic's set-theoretic treatment, privatives are modelled by a condition which requires that, if X is a member of the set of entities denoted by N, it is a member of the relative complement of the NP set, in the particular domain. (Hoepelman, 1983).

Now, the kinds of examples with which we are concerned are ones which we term "functional privatives": ones where the inference to type-3 sentences is made problematic by the *interaction* of the semantic properties of the N and A. Thus functioning privatively is *not* an intrinsic property of the type of adjective being used - rather, it stems from the conflict of semantic features and the principles of defeasibility by which they are resolved. Examples include:

This is a stone lion

This is a plastic flower

These adjectives can function affirmatively in other contexts:

This is a stone bridge

This is a plastic chair

They each allow inferences to their appropriate sentence types 2 and 3. In the functionally privative cases, the inference to type 3 is problematic, although not completely nonsensical.

Problems With The Deductive Approach

If we map a deductive relation over total objects, the only form of defeasibilty open is type I rebuttal. So the denial of the evidence for the necessary conditions of a concept entails a denial of all of its conditions - i.e., of the concept itself. This is precisely the effect achieved by the set-theoretic relative complement method. Three problems emerge: firstly, it denies too much: with functional privatives, we would like to preserve at least some of the non-essential properties of the "lion" concept; since they support modus tollens, and require default properties to be dependent upon the obtaining of necessary conditions, standard views cannot allow this. Secondly, there is no account of the asymmetrical nature of the combination. We can argue that this will largely be at the behest of dependency relations, in which the C-relations of the A have primacy (cf. Anderson, 1986). Thirdly, the method can only tell us what the object *is not* rather than what it is: a "fake Hogbin" just is not a Hogbin, and a "stone lion" is not a lion: yet it seems that a "fake Hogbin" is somewhat more like a real Hogbin in appearance than is a Rothko. Again, a stone lion is more similar in shape to a lion than to a frog. The argument of the next section will be that the C-relation view can make good most of these flaws.

THE CRITERIAL APPROACH

We claim there are two different defeats at work here: one type II rebuttal, the other an undercutting. Ordinarily, a C-information based concept of "lion" will function as evidence for the sortal identity of an object: the conceptual evidence (P) is C-related to the identity claim (Q). But the combination of the concept "stone" (R) with P will undercut that C-relation. The combination of P with R involves type II rebuttal of the C-information lodged in P by the C-information in R. So, concept Q's use is C-related to certain kinds of information which is generated in support of the use: for example, ideas of "internal essence" of the identity of lions, derived from conventionally structured lay theories of the domain (see Murphy & Medin, 1985; Keil, 1987). If these properties were necessary/deductive conditions on a total semantic object, the overriding of them would function according to rebuttal type I; hence the term could not be applied. That this is not the case is clear from the ease with which we can understand "stone lion". The concept's use is also S-related to certain features, concerning "appearances". So we can justifiably utilize the term "lion" if we can generate appropriate C- and S-relations (i.e., P) on the basis of current context and information. But we have an undercutting of the connection between P and Q, by R: that is, the applicability of the term "lion" on the basis of P is denied. This denial of positive support for Q is not equivalent to an assertion of Q's falsity. And this also leaves P open to negotiation.

For undercutting, the combination (P & R) must be consistent; this is where the type II rebuttal of the noun's C-information, which motivates the undercutting, operates. Crucially, this type of defeat could

only be facilitated by a context of an undercutting, and not by a rebuttal of claim Q. The question is then, which has priority, P (the noun), or R (the adjective)? On the basis of linguistic dependency, the C-relations of "stone" (such as "inanimate", etc.) have the effect of overriding the C-relations of "lion" by a type II rebuttal. The full concept of "stone" is thereby rendered consistent with the undefeated S-relations of "lion". Two points are noteworthy here. First, the survival of some of the S-information is acceptable since S-relations do not require the obtaining of C-relations. Second, in this combination not all of the S-information survives: only that which is *required* by the C-relations of the "stone" concept. The other S-relations would either be directly negated (by default defeat), or rendered inconsistent (by general defeat). But the C-relations of "stone" positively require that there is some physical structure or shape: and it is these qualities of "lion" which are retained. We claim that the defeat of the C-information of "lion" by that of "stone" is type II rebuttal because it is a cancellation of the criterial *properties* of the noun; and the other defeat is by undercutting because it defeats a C-related *claim* made on the basis of such properties.

We have advocated a well-motivated intensional account of the defeasible semantic relations constraining concept combination. Such an account seems to be necessitated by the case of functional privatives. Its extrapolation to more straightforward concept combinations should be perspicuous.

REFERENCES

Anderson, J. (1986), Suprasegemental dependencies, in J. Durand (ed.), *Dependency and Non-Linear Phonology*. London: Croom Helm.

Baker, G.P., & Hacker, P.M.S. (1984), *Scepticism, Rules and Language*. Oxford: Blackwell.

Hoepelman, J. (1983), Adjectives and nouns: a new calculus, in R. Bauerle, C. Schwarze & A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*, Berlin: de Gruyter.

Jackendoff, R. (1983), *Semantics and Cognition*. Cambridge, Mass.: MIT Press.

Keil, F.C. (1987), Conceptual development and category structure, in U.Neisser (ed.), *Concepts and Conceptual Development*. Cambridge: CUP.

Landman, F. (1986), *Towards a Theory of Information*. Groningen-Amsterdam Studies in Semantics, 6, Dordrecht: Foris Publications.

Murphy, G.L., & Medin, D.L. (1985), The Role of theories in conceptual coherence, Psychological Review, 92 (3), 289-316.

Tennant, N. (1987), *Anti-Realism and Logic*. Oxford: Clarendon Press.

Wittgenstein, L. (1953), *Philosophical Investigations*. Oxford: Blackwell.