Figurative adjective-noun interpretation in a structured connectionist network

Susan Hollbach Weber
Computer Science Department, University of Rochester
hollbach@cs.rochester.edu

Abstract

Non-literal use of an adjective, whether signalled by a category error or by a value expectation violation, invokes the connotations or immediate inferences associated with that adjective in various noun contexts. Immediate inferences reflect the structure of stored knowledge, as they are available too quickly and effortlessly to involve any complex form of information retrieval. Specifically, they suggest the use of the spreading activation model of semantic memory. The relation between the inferences invoked by the adjective and salient features of the noun employed in the figurative usage are exploited by the DIFICIL connectionist inferencing system to interpret the meaning of an unfamiliar adjective-noun phrase.

The interpretation of figurative adjective-noun combinations can be modelled by exploiting the connotations that arise from the adjective's literal usages. In the phrase 'agressive diamond', for example, the high-intensity connotations associated with the adjective maps into large size in diamonds. These immediate inferences must reflect the structure of stored knowledge, as they are available too quickly and effortlessly to involve any complex form of information retrieval. Specifically, they suggest the use of the spreading activation model of semantic memory. The argument is that the patterns of immediate inferences reflect the structure of connections in the underlying spreading activation model, implemented here as a structured connectionist network.

Once the need for a figurative interpretation has been detected, the immediate inferences associated with the adjective are activated in parallel, with a view to later establishing the mapping from a source property of the adjective to a target property of the noun. Once these connotations have been catalogued and incorporated into an adjective's meaning as extended word senses, it is simply a matter of lookup. Of interest here, however, are the methods of arriving at these extended senses, assuming that each word has an unique denotation.

The interpretation of figurative adjective-noun combinations is undertaken in the context of the connectionist inferencing system known as DIFICIL, for Direct Inferences and Figurative Interpretation in a Connectionist Implementation of Language understanding [Weber, 1989b]. The cognitive model underlying the system is described in detail in [Weber, 1989a]. Briefly, categories are represented as structured collections of properties with their attendant values. Functional property values are used to establish coherent groupings of property values, known as *aspects* of the category. For example, the functional property *ripeness* generates two aspects within the *apple* category: unripe apples are green, hard and sour, while ripe apples are red, crisp and sweet. Increased activation of any member of the coalition results in increased activation for the entire aspect. Furthermore, one aspect can inhibit another if incompatible, or stimulate it if logically dependant upon it.

The organization of these aspects depends on the structural characteristics of the properties involved. Properties can permit multiple or concurrently held property values (eg. a cookie tastes at once salt and sweet) or, if the values are mutually exclusive, they can be ordered (eg. hot, warm, cool, cold) or unordered (eg. shape) [Aarts and Calbert, 1979;
Kittay, 1987]. Orthogonal to this is the fact that properties and their attendant values can be classified into three groups: perceptual, constitutive and functional. Perceptual properties are those pertaining to the senses, eg. colour, scent, taste or texture. Constitutive properties are in some sense the definitional properties of a category, often expressed in terms of genetics, compositional makeup and the like. Functional properties relate to an object's usefulness by humans, eg. tameness, edibility or state of repair. Functional properties play a special role in category representation, supplying the various perspectives from which the category can be viewed. For example, the *unripe* property provides the focus of relevance for the *sour* and *green* property values of apples.

This functional aspect model of conceptual representation has been implemented as a structured connectionist network [Feldman and Ballard, 1982] on the Rochester Connectionist Simulator [Goddard *et al.*, 1988], resulting in the DIFICIL inferencing system. The content of a knowledge base is specified with statements in a high level input language. There are six statements in the language. The *subcat* statement sets up the property inheritance or subcategorization hierarchy, and the *abstracts* statement defines the property abstraction hierarchy. The *mutex* and *invokes* statements specify the relations of mutual incompatibility and reinforcement that pertain between aspects. The *hasslot* statement establishes the properties and values belonging to a category, with syntactic variants for perceptual, constitutive and functional properties, and optional scalar positioning parameters. The *aspect* statement creates the conceptual aspects fundamental to the model. For example, the three statements

hasPslot (diamond: cut; marquise (pointy), diamond-cut (round))

hasPslot (diamond: size; large (+), medium (0), small (−))

aspect (diamond: diamond-cut [default]; small, brilliant)

would create the connectionist structures depicted in Figure 1. Categories appear as hexagons, properties as squares, values as circles, positional designators (more on these later, in the section on value mapping) as diamonds, and control nodes as rectangles. The small triangles represent the binder nodes establishing the concept–property–value triplets specified in the *hasslot* statements. The larger triangles are two/three binders [Shastri, 1985], requiring that two of their three inputs be active before firing. The pentagonal *inertial binder* labelled *hub* in the figure controls the spread of activation through the aspect; once activated, it tends to stay on even in the face of active inhibition.

In literal adjective-noun interpretation the DIFICIL system is designed to draw all available *direct inferences* pertaining to the input phrase. *Immediate inferences* are the direct inferences available at the level of the category under consideration. They are performed quickly, in a few hundred milliseconds, and without conscious thought. In DIFICIL, immediate inferences are defined to be the property values available through activation spreading from a category level aspectual hub. For example, the knowledge that diamonds are by default small and brilliant is an immediate inference. *Mediated*

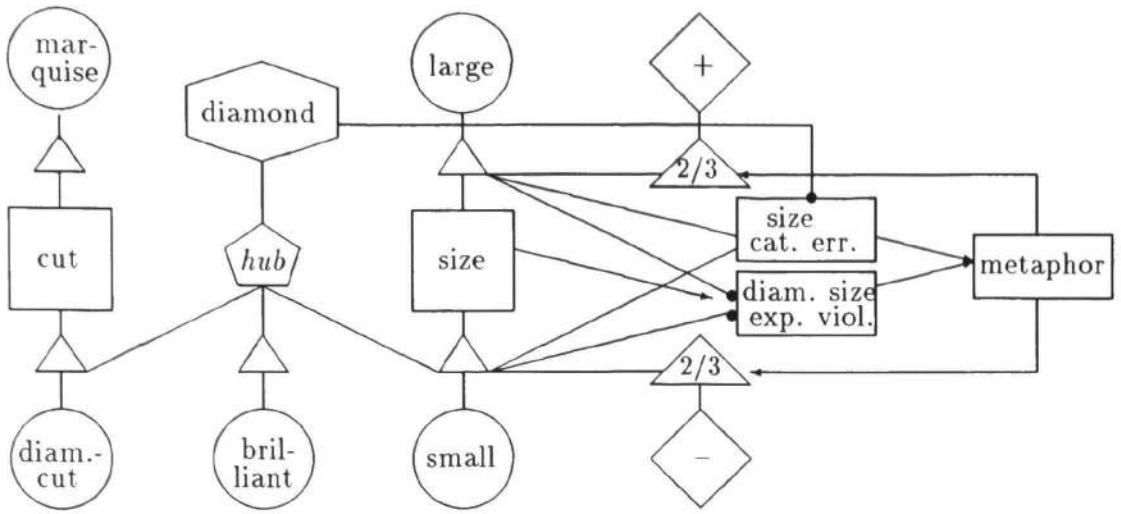

Figure 1: The *diamond-cut* aspect of diamonds, along with the semantic anomaly detection and scalar positioning mechanisms of the *size* property.

inferences are the second form of direct inference, where knowledge about a more abstract category is used to supply the information necessary to understand the discourse. Mediated inferences take somewhat longer to obtain than immediate inferences, as they require chaining up the subcategorization hierarchy created by *subcat* statements. For example, if one knew that all gemstones are expensive and that a diamond is a gemstone, the inference that diamonds are expensive is a mediated inference.

Direct inferences form the fundamental mode of operation of the system. Not only are they the primary mechanism for interpreting literal adjective-noun combinations, but the information encoded as immediate inferences is crucial to the successful interpretation of the figurative adjectival modification of a noun. Before considering how this might work, however, the prior question of distinguishing between literal and figuative usages is dealt with.

Detecting figures of speech

Figurative usage can be signalled in any number of ways, from subtle contextual cues and knowledge of conventional usage to blatant semantic anomaly. For example, the phrase 'cold woman' is ambiguous as to whether body temperature or emotional responsiveness is being referred to, an ambiguity that cannot be resolved without sentential context. On the other hand, the fact that a stare cannot literally be said to have temperature means that the phrase 'cold stare' constitutes a *category error*, a form of semantic anomaly in which a predicate is used in conjunction with a term it does not span [Keil, 1979] (recall that it is assumed throughout that each word has an unique denotation; in the case of 'cold', this would be temperature). The phrase 'cold steam' exemplifies a second form of semantic anomaly, since although steam has the property *temperature*, the only permissible value of this property is *hot*. Thus 'cold steam' constitutes an *expectation violation*, since the named property value is an unusual or impossible choice for the

property of the noun. Finally, there are idiomatic phrases, like 'cold shoulder', whose obvious literal interpretation is never considered (by native speakers of the language), since the figurative meaning is so well established.

The mechanisms for detecting the need for a figurative interpretation currently implemented in DIFICIL are limited to category error and expectation violation detection.[1] Category errors are detected on a property by property basis. A detection node, labelled *size cat. err.* in Figure 1, is created for each new property named in a *hasslot* statement. This node receives excitation from the property–value binders and inhibition from the category, so if a property value should be activated while the category is inactive, a semantic anomaly will be reported as required. Expectation violations are detected at the level of the category–property conjunction. The detection node receives inhibition from the property–value binders, and excitation from the property, so if any property value not possessed by the category is named, an expectation violation results. Both forms of semantic anomaly, when detected, transmit their activation to the global *metaphor* control node. When all possible immediate and mediated inferences have been drawn, if there is still an anomaly being reported, the *metaphor* node is activated, signalling a network-wide change of state, from literal interpretive mechanisms to figurative.

Adjective connotations

The connotations of an adjective arise from its associations within its literally allowable noun contexts. The connotations considered by DIFICIL for the purposes of figurative interpretation are the immediate inferences arising from the modification of an arbitrary noun by the given adjective. For any category known to the system, if the adjective names a property value that participates in an aspect of that category, then it will trigger a characteristic set of immediate inferences, as activation spreads from the named value to the aspectual hub and from there propagates to all related property values.

The set of all values activated in this manner form the interpretive base for understanding a figurative usage. The most straightforward interpretations arise when a property value of the target category is made available through an immediate inference associated with another category. For example, suppose it was 'known' to the system that agressive people are also large in size; then the unfamiliar figure of speech 'agressive diamond' would be interpreted as denoting a large diamond.

However, it is the indirect methods of arriving at a plausible interpretation that form the focus of this work. For example, as agressivity in people is ranked high on the intensity scale for agression, the phrase 'agressive diamond' may refer to a large diamond, since size is a salient property of diamonds and large is the high-ranking value on the intensity scale for size. On the other hand, it could be cut in a pointy shape (known as a marquise cut), since the tools of agression (weapons) tend to have pointy shapes. There are doubtless other possibilities, increasingly far-fetched, but the notion of entertaining competing interpretations in parallel is fundamental to the chosen model of spreading activation; only after the novelty of the figure has worn off, and its most likely interpretation catalogued,

[1] This of course does not preclude the possibility of eventually exploiting contextual cues or idiom recognition, if available; thanks to Jim Hendler for pointing this out.

will there be a single correct answer to the interpretation question. When considering novel figurative usages, all possibilities must be explored.

This is accomplished by two interlocking interpretive processes, one to establish all property-to-property mappings, the other to set up the value-to-value correspondences within related properties.

Property mapping

In order to establish a semantic correspondence between the property named by the adjective (eg. 'agressive' names human agression) and properties of the noun (eg, size and shape of diamonds), a property abstraction hierarchy relates all the properties in the knowledge base. As soon as the need for a figurative interpretation has been recognized, activation is permitted to spread throughout the abstraction hierarchy from the property named by the adjective. Activation will eventually spread to every property in the knowledge base, so in some sense the hierarchical arrangement is unnecessary: one could simply stimulate all properties in parallel, and achieve the same end result. But with the hierarchical spread of activation the timing delays between the various meaning hypotheses reflect their plausibility: later suggestions are increasingly implausible, as the semantic distance between the properties increases.

Value mapping

The real interpretive work is done at the level of the value-to-value mappings. There are two methods used to establish semantic correspondences: immediate inferences and scalar position conjunction.

Immediate inferences involve the notion of mutually excitatory functionally related property values. Of interest to the figurative interpretation process are the conceptual aspects participated in by the property value named by the adjective. As conceptual aspects remain inactive until enabled by the excitation of their category, it is necessary to briefly stimulate all category nodes in the network in order to establish all the connotations of the figuratively used adjective. This excitatory signal decays rapidly, however, and soon (within 10 time steps) actually inhibits all the categories previously excited. Those categories possessing an aspect in which the adjective's value participates will receive positive feedback activation sufficient to overcome this inhibitory signal for long enough to establish any further semantic mappings that may exist between properties of the aspect and properties in the noun concept.

The second mechanism for establishing semantic correspondences is *scalar position conjunction.* It turns out that most property values are scalar in nature, that is, the allowable values for a given property can be strictly ranked with respect to each other, from least to greatest [Kittay, 1987]. One example of this behaviour is the *temperature* property, whose values range from freezing through cold, cool, warm, hot and finally to boiling/burning/blistering etc. There will generally be two values that typify the positive and negative extremes (eg. hot and cold) with a third value typifying the neutral setting.

The scalar nature of a property is established with optional parameters to the *hasslot* statement, specifying the *scalar position designator* of the value. Property scales are defined implicitly by choice of position designators: for example, the *intensity* scale may have the positional designators *ipos*, short for *i*ntensity:*pos*itive (previously referred to

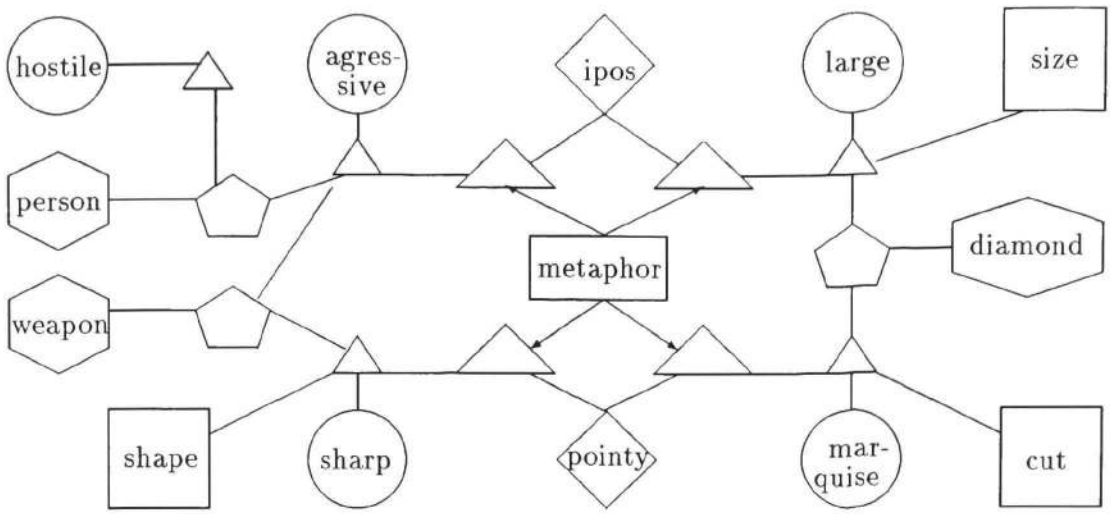

Figure 2: Schematic of the connectionist structures involved in the interpretation of the phrase 'agressive diamond'.

with the symbol '+'), *ineutral* (0), and *ineg* (−), so the relative sizes of diamonds, say, is correctly expressed by

hasPslot (diamond: size; large (ipos), medium (ineutral), small (ineg))

Unranked but mutually exclusive properties are handled in the same way, as there is no attempt to actually impose the scalar ordering implied by the choice of designators, so diamond shape, for example, could be captured with the following statement:

hasPslot (diamond: shape; marquise (pointy), pear-shape (pear-shape), ...)

Terms appearing in different semantic positions in a statement are distinguished, so the first occurrence of 'pear-shape' is taken to be a property value, while the second refers to the positional designator of the same name.

In order to interpret the phrase 'agressive diamond', it is necessary to access all the connotations of the adjective 'agressive'. This is done by briefly stimulating all the categories in the knowledge base, followed almost immediately by an increasingly inhibitory signal. Immediate inferences result for all categories with 'agressive' as a property value, as these categories receive feedback activatation sufficient to defeat the inhibitory input for long enough to establish the aspects, if any, participated in by the value. Suppose the only two such categories are weapons and people:

aspect (people: threat [non-default]; hostile, agressive);

aspect (weapons: sharp [default]; agressive).

The transmission of activation from these aspects to property values of diamonds is mediated by the scalar position designators, as depicted in Figure 2. If it has been established that sharp weapons are classified as *pointy* on the shape scale and that large size is *ipos* on the intensity scale, then it will eventually be decided that agressive diamonds are not only large but also of marquise cut.

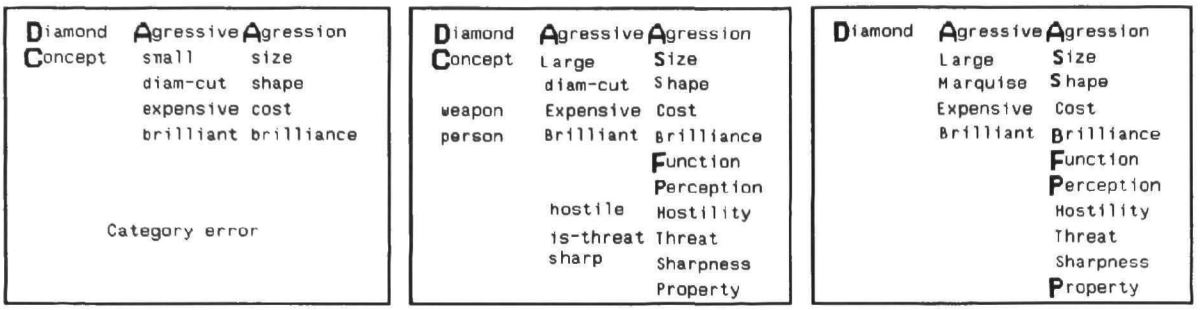

Figure 3: Time lapse shots of system output on the target phrase 'agressive diamond'.

System performance

Figure 3 shows three time-lapse pictures of the output of DIFICIL running the agressive diamond example. The graphics interface to the Rochester Connectionist Simulator permits creating icons to represent the activation levels of individual network units. The icons here take the shape of letters of the alphabet, where the first letter in the word represented by the unit is the one iconified. The icons range from a large capital letter, for significant activation, to a lower case letter, for marginal activation. Words appear in columns according to their semantics: categories on the left, property values in the middle and properties on the right. The leftmost panel shows the state of the network after about 20 simulation steps: a category error has been detected, due to the presence of the adjective 'agressive' in the absence of such nouns as 'person' or 'weapon' that would permit a literal reading. Note the marginal activation on the default property values of diamonds. Even in the face of a semantic anomaly, these default conclusions are primed for possible future reference.

The middle panel shows the state of affairs some 15 time steps later: all the categories in the knowledge base have been briefly stimulated and are now receiving gradually increasing inhibition. The two categories related to 'agressive' are able to overcome this inhibition longer than unrelated categories, due to the positive feedback from the property value. Activation is spreading rapidly through the property abstraction hierarchy, as evidenced by the long list of properties in the right hand column. The central column shows that the two aspects containing the value 'agressive' have been established: an agressive person is hostile and threatening, and an agressive weapon is sharp. Note too that the default diamond properties *expensive* and *brilliant* have been boosted in activation, while the default size *small* is easily defeated in favor of *large*, as indicated by the intensity scale conjunction of *expensive, brilliant* and *large* with *agressive*. No (non-default) conclusions have yet been drawn as to the diamond's shape.

In the leftmost panel, the final figurative interpretation is available, after a total of 40 simulation steps. All concepts but the target *diamond* have been inhibited, shutting down all value–property pairs not directly associated with the target category. Conversely, all properties in the knowledge base are active to varying degrees due to the spread of activation through the abstraction hierarchy, but as properties are incidental to their

values, this is unimportant. The relevant values appear in the central column: note that it has been established that since agressive weapons are sharp, and sharp weapons are classified as pointy on the shape scale, as are marquise-cut diamonds, the diamonds in question must be marquise-cut.

Summary

The notion that a property value occupies a scalar position on any number of classification scales associated with the property is exploited in DIFICIL to interpret figurative adjective-noun combinations. The need for a figurative interpretation can be signalled by various forms of semantic anomaly, including category error and expectation violation. When this happens, activation spreads to the immediate inferences associated with the adjective in a wide variety of noun contexts. All scalar positioning conjunctions between property values thus activated and property values of the target noun mediate the transmission of activation to the relevant properties of the noun.

Acknowledgements

Thanks to Jerry Feldman for comments on a draft. This work was supported by ONR research contract no. N00014-82-K-0193 and U.S. Army Communication-Electronics Command Grant no. DAAB 10-87-K-022.

References

[Aarts and Calbert, 1979] Jan M. G. Aarts and Joseph P. Calbert, *Metaphor and Non-Metaphor: The Semantics of Adjective-Noun Combinations*, Max Niemeyer Verlag, 1979.

[Feldman and Ballard, 1982] Jerome A. Feldman and Dana H. Ballard, "Connectionist Models and Their Properties," *Cognitive Science*, 6:205–254, 1982.

[Goddard *et al.*, 1988] Nigel Goddard, Kenton Lynne, and Toby Mintz, "The Rochester Connectionist Simulator User Manual," Technical Report 233, Computer Science Department, University of Rochester, March 1988.

[Keil, 1979] Frank C. Keil, *Semantic and Conceptual Development An Ontological Perspective*, Harvard University Press, Cambridge, Mass., 1979.

[Kittay, 1987] Eva Feder Kittay, *Metaphor: its Cognitive Force and Linguistic Structure*, Clarendon Press, 1987.

[Shastri, 1985] Lokendra Shastri, "Evidential Reasoning in Semantic Networks: A Formal Theory and its Parallel Implementation," Technical Report 166, Computer Science Deparment, University of Rochester, September 1985.

[Weber, 1989a] Susan Hollbach Weber, "Modelling semantic flexibility with a structured connectionist implementatation of functional category organization," submitted to the IEEE Conference on Neural Information Processing Systems (NIPS), May 1989.

[Weber, 1989b] Susan Hollbach Weber, "A Structured Connectionist Approach to Direct Inferences and Figurative Adjective-Noun Combinations," Technical Report 289, PhD. thesis, Department of Computer Science, University of Rochester, May 1989.