

### The double nature of negative antonymy

Antonymy is typically considered in the context of one-dimensional adjectives such as *tall-short* and *fast-slow*, where, intuitively, the positive counterpart, e.g. *tall*, references entities with much of a dimension (e.g. much height) whereas the negative, e.g. *short*, references entities *not* having much of the dimension (having little height). In accordance, on the syntactic negation theory of antonymy (Rullmann 1995; Heim 2006, 2008; Buring 2007), instead of being specified in the lexicon, antonyms are formed by a degree-negation operator, *little*, hidden in the logical form of words such as *short*.

One problem with this view is that, for some antonym pairs, our judgments as to which one is the negative are not solid; a case in point includes *simple-complex*. Moreover, in many antonym pairs our judgments counter the theory's predictions. For example, while *dangerous*, *dissimilar/ different*, *dirty* and *sick* reference existence (or much of) a property, we tend to consider them negative antonyms. In opposite, we tend to intuitively judge *safe*, *similar/ identical*, *clean* and *healthy* to be positive, despite the fact that, e.g., objects are considered *identical* (or *similar*) in, say, color iff *no* (or *not much*) *color difference* between them can be attested; individuals are considered *healthy* if they do *not* possess any serious disease; things are *clean* iff covered by *no* dirt, etc. To explain these judgments, previous work relates to negative cultural connotations, postulating that these may be dissociated from linguistic negativity (e.g., *old* is linguistically positive, despite any cultural connotations; Kennedy 1997; 2001).

This paper presents a more adequate analysis of linguistic antonymy, which takes into account the fact that many adjectives are multidimensional. This new analysis reveals that in many of the above cases negative connotations and linguistic negativity actually match. The crux is that negation may also play a role at a level higher than that of a single dimension. Logical (rather than degree-) negation can affect the operators that bind the different dimensions of a single multidimensional adjective. A corpus study yields striking support to this new view.

Let us consider some examples of the workings of multidimensional antonyms. By default, native speakers intuitively judge entities as *identical* (generally, rather than with respect to color alone) iff they are identical in *all* (or *most*) of the respects that count as relevant in the given context. By contrast, they intuitively judge entities to be *different* iff they are different in *some* (at least *one*) respect. Objects are considered *clean* iff dirt of *no* sort (dust, stains, etc.) is attested on them, while they are considered *dirty* iff dirt of *some* sort is attested. Individuals with a slight cold are *not* strictly speaking *healthy*, since they are not *all* healthy, while one type of sickness suffices to count as *sick*, etc. Considering these, among many other examples, the following rather systematic generalizations emerge. The dimensions of positive adjectives (like, e.g., *identical*, *healthy* and *clean*) are integrated by means of universal quantification. By contrast, the dimensions of the negative antonyms are integrated by means of existential quantification.

A theory of antonymy whereby negative antonyms are logical negations of their positive counterparts predicts that the force of quantifier over dimensions will systematically vary in antonym pairs, because negated universals are existential and vice versa; e.g., if (i) in order to be considered 'healthy', one has to be healthy in *every* dimension, and (ii) 'sick' is equivalent to '*not* healthy', it follows that: (iii) in order to be considered 'sick', one has to *fail* to be healthy in at least *one* dimension: For any set of predicates DIM, *healthy* is equivalent to  $\lambda x. \forall Q \in \text{DIM}, \text{healthy-wrt}(x, Q)$  iff *sick* is

equivalent to  $\lambda x. \neg \forall Q \in \text{DIM}, \text{healthy-wrt}(x, Q)$ . The latter reduces to  $\lambda x. \exists Q \in \text{DIM}, \neg \text{healthy-wrt}(x, Q)$ .

In sum, at the multidimensional level positivity appears to be linked with universal generalizations (healthy in all respects, clean of any dirt type, etc.), while negativity is linked with the existence of a counterexample (sick means not being healthy in all respects; dirty means not being free of any dirt-type, etc.)

Here is some empirical evidence for these proposals. Exception phrases are indicators of universal generalizations. The contrast between the felicity of, for example, *everyone* (or *no one*) *is happy except for Dan* and the infelicity of *#someone is happy except for Dan* illustrates this. (von Stechow 1993,4; Fox and Hackle 2006). In accordance, the above proposal predicts that speakers will generally judge exception phrases as felicitous when combined with adjectives whose dimensions combine via an implicit universal quantifier ('a *conjunctive* strategy'), but not with their negative antonyms, whose dimensions combine via an existential quantifier ('a *disjunctive* strategy'). This prediction is born out, cf. the felicity contrast in (1a,b). Furthermore, negated universals are existential and vice versa. Hence, exception phrases are predicted to be felicitous when combined with negated 'disjunctive' adjectives, but not negated 'conjunctive' adjectives. This prediction is born out too, cf. (2a,b).

- (1) a. I am **healthy** except for high blood pressure (bp) ( $\forall Q \neq \text{bp}, \text{Healthy-wrt}(I, Q)$ )  
 b. #I am **sick** except for normative blood pressure ( $\exists Q \neq \text{bp}, \neg \text{Healthy-wrt}(I, Q)$ )
- (2) a. #I am **not healthy**, except for (normal) cholesterol ("")  
 b. He is **not sick**, except for the diarrhea ( $\neg \exists Q \neq \text{diarrhea}, \neg \text{Healthy-wrt}(\text{they}, Q)$ ,  
 i.e.  $\forall Q \neq \text{diarrhea}, \text{Healthy-wrt}(\text{they}, Q)$ ).

These intuitive judgments are supported by distributional patterns in naturally occurring examples. Significantly, an analysis of 1300 naturally occurring counts of the form 'Adj. except' with 8 antonym pairs in positive vs. negative (negated or DE) contexts (as in, e.g., (1) vs. (2), respectively) supports both

- (i) the view that negative antonymy systematically affects the force of quantifier over dimension, as predicted by a logical-negation theory of antonymy, and
- (ii) the view that the nature of negative polarity in multidimensional antonyms consists of reference to counterexamples to generalizations.

An analysis of variance reveals significant differences ( $F = 10.37; P = 0.0002$ ) as follows.

- (i) In positive adjectives, conjunctivity (frequency of implicit universal dimension integration, as in, e.g., *healthy except dim* or *very similar except dim*) is higher than disjunctivity (frequency of implicit existential dimension integration, as in, e.g. *not healthy except dim* or *hardly similar except dim*).
- (ii) In negative adjectives conjunctivity is lower than disjunctivity. Moreover:
- (iii) Conjunctivity is higher in positive adjectives than in their negative antonyms, whereas disjunctivity is lower.

Does the fact that speakers regard borderline entities as e.g. *neither healthy nor sick* speak against a logical-negation analysis of *sick*? A number of studies suggest the contrary (Sauerland 2011); gaps characterize also logical negations. To conclude, the new account is economic in that dimensions need to be lexically specified only for positive adjectives (thereby explaining their cognitive prominence; Horn 1987; Giora 2006) and as a default dimensions are conjoined; negative antonyms reference negated dimension conjunctions (explaining their relative complexity and so-called negative connotations).