

### Invented Antonyms: Esperanto as a semantic lab

Esperanto's syntax and morphology were planned to be simple and transparent, while its semantic-pragmatic components (founded by Zamenhof's 1887 basic vocabulary translations) have evolved naturally through a continuous history of use, providing a special opportunity to study the interface between morphology and interpretation. We illustrate this with a study of the use of the antonymy morpheme *mal* (as in, for instance, *malalta*; 'ant.-tall'; 'short') in a 4.3 million word corpus, Tekstaro ([www.tekstaro.com](http://www.tekstaro.com); Wennergren 2003). While few non-*mal* antonyms (like *diferenca* 'different') exist, the *mal*-antonyms are significantly more productive and pervasive (as Baayen and Lieber's 1997 tests can show); thus, antonyms are easy to find.

**Linguistic correlates of negative antonyms** are commonly thought to include incompatibility with measure-phrases (cf., *two meters tall/ \*short*) and reduced likelihood of ratio-phrase modification (*twice as tall/ #short*) and nominalization (*tall* and *short* measure 'height', not *#shortness*); see Greenberg 1966, Horn 1989, Bierwisch 1989, Kennedy 2001 and references therein. These well known cross-linguistic tendencies appear to result from semantic complexity, not from overt morphological complexity or low frequency. Evidence for this is the slow and errorful acquisition of morphologically simple invented labels for negative vs. positive concepts (Klatzky et al 1973). However, these distributional tendencies admit exceptions ('non-paradigmatic cases'; see discussion below), including many positive adjectives that do not license measure phrases (*#two degrees warm/ cold*) and rarely also negative adjectives that do (*two hours late/ early*). In addition, many positive adjectives resemble negative ones in rarely licensing *twice* (*glad, wise*) and a few notorious exceptional negative adjectives (*bad, slow*) seem to license *twice* more often than their positive bases (Sassoon 2010). We use Esperanto to examine the reliability of the linguistic correlates, by asking to what extent antonym morphology (*mal*) bears features that are considered indicative of semantic negativity.

**Frequency.** The total number of adjectives in the corpus is 343,120 (not counting participles, e.g., *mi estas skribanta*, 'I am writing'), with 21,363 *mal*-antonyms. Thus, the ratio of *mal*-antonyms to all adjectives is **6.226%**; this is the first indication that *mal* marks negative antonyms, since low frequency is suggestive of negativity (Lehrer 1985).

**Measure phrases.** Of the 276 occurrences of *metro(j)* 'meter(s)' prefixed by a quantity, 41 modify adjectives (as in *dek centimetrojn longa* '10 centimeters long'), 23 nouns (*la altecon de 3-4 metroj* 'the height of 3-4 meters'), and 5 verbs (*altiĝas 12 metrojn* 'reaches-the-height of 12 meters'). Almost all the measure phrases are with positive (non-*mal*-) words such as *longa* 'long', *larĝa* 'broad', *alteco* 'height', etc. There are four occurrences of *metro(j)* with *mal* words. Two are not problematic because the *mal*-word does not directly combine with the measure phrase, which, instead, is within a PP argument of a verb/ comparative; this is also possible in English (*Kiam estas refluo kaj la akvo malleviĝas per du metroj* 'When there is low tide and the water level drops to two meters'; *pli malproksime ol unu metron* 'farther than one meter'). The two other examples do not conform to the general tendency of measure phrases not to license *mal*-antonyms; 2 out of 276 (below 1%) is a significantly lower ratio than expected (~6%). Moreover, the English translations of these exceptions license measure phrases as well (*tridek metrojn malsupren* 'thirty meters downwards' and *milojn da kilometroj malproksime* 'thousands of kilometers away'). Thus, *upward-downward* appear to be non-paradigmatic across languages (another rare case of a negative antonym licensing

measure phrases) and *far* seems to be positive, while *near* negative across languages, including Esperanto. It is *proksime* ('nearby'), not *malproksime* ('far-away'), that occurs slightly less often (560 vs. 569 tokens), with no measure phrases, and whose definition references a negative *mal*-antonym (*Apartigita per malgranda distanco* 'separated by a small distance'; [www.reta-vortaro.de](http://www.reta-vortaro.de)), illustrating that conceptual concerns may override morphological ones.

**Ratio modifiers.** 'x times as adj as' is expressed with a numeral suffixed with *oble* 'times', followed by *pli* 'more' (as in *dudekoble pli granda ol* 'twenty times as big as'). Of the 88 matches for this construction, 5 are with *mal*-words, which is not significantly different than expected (5.7% vs. ~6%). In *centoble pli malbona ol* 'a hundred times worse', we find *malbona* 'bad', which is an outlier in English as well (Sassoon 2010), supporting its status as a non-paradigmatic negative adjective. In *multoble pli malfacilan* 'many times more difficult' we find *malfacilan* 'difficult/ hard', which seems to be conceptually positive – it is also more frequent than its base form (the same regards 'hard' and 'difficult' vs. 'easy' in the American English corpus COCA, the former are both more frequent and more likely to co-occur with 'twice as'), and so is *malĝoia* 'sad', which occurs in the third example. As for the fourth example with *malfeliĉa* 'unhappy', again the distribution of *feliĉa* and *malfeliĉa* roughly corresponds to that of *oble pli felica* and *oble pli malfelica*, respectively (3 to 1, i.e. relatively similar frequencies). Thus, these adjectives are generally atypical, not just with regard to ratio modification. Moreover, *malfeliĉa* occurs in a contrastive context, which facilitates ratio modification (cf. 'twice as short' means 'half as tall'). Only discarding some of these examples do the results support the common view that ratio modifiers co-occur less often with *mal* adjectives.

**Nominalizations.** *-o* is the noun marker and *-ec* is a suffix enabling reference to intuitively abstract qualities, like English *-ness* and *-ity* (cf. *grando* 'size' vs. *grandeco* 'size'/'greatness'). A bigger sample is required to determine whether indeed *mal*-antonyms (i) have a lower ratio of *-o* nominalizations and (ii) favor the use of *ec* along with *o*, as is the case in our sample, with the notable exception of *malbono* 'the bad' (supporting its non-paradigmatic status), as well as (with regard to (i)), *malkareco* 'cheapness' and *malsameco* 'difference' (maybe due to the synonym *diferenco*). The common view that nominalization morphemes co-occur less often with *mal* adjectives is supported only discarding some non-paradigmatic cases (e.g., *malbono*).

**Discussion.** The distribution characteristic of antonyms is not due to their negativity per se (the fact that their scale is reversed in comparison with the scale of their positive bases); after all, it is characteristic only of *paradigmatic* negative antonyms. Rather, paradigmatic negative antonyms *fail to represent degree ratios*, rendering, e.g., *twice as short* and *two meters short* (= twice as short as a meter) infelicitous: Let the grammar associate positive adjectives *A* with mappings  $f_A$  of entities  $x \in D$  to degrees  $r \in \mathfrak{R}$  (Kennedy 1999) and antonyms with reversed mappings  $\lambda x \in D. r - f_A(x)$  (Sassoon 2010). Consider a context *t* in which  $x_{10}$ 's height is twice  $x_5$ 's ( $f_{tall,t}(x_{10}) = 10$ ;  $f_{tall,t}(x_5) = 5$ ) and *short*'s 'transformation value',  $r_{short,t}$ , is 1. In *t*, *tall* represents height ratios ( $2 = 10/5$ ), and so does *tall*'s non-transformed reversal,  $\lambda x \in D. -f_{tall,t}(x)$  (as  $2 = -10/-5$ ), but *short* does not ( $2 \neq (1-10)/(1-5)$ ), nor does any non-reversed transformation like  $\lambda x \in D. 1 + f_{tall,t}(x)$  ( $2 \neq (1+10)/(1+5)$ ); i.e., paradigmatic negative (e.g. *short*) and non-paradigmatic positive adjectives (*warm*) differ from paradigmatic positives (*tall*) in having a transformation value, while non-paradigmatic negatives (*bad*) are reversed but not transformed.