NOUNS, ADJECTIVES AND *MORE*, Galit W. Sassoon, 29.05.2007

Obviously, nouns and adjectives are very different types of animals. The question is precisely what this obvious difference is that forms the cause for the linguistic contrasts between them.

1 ADJECTIVES: VAGUENESS AND GRADABILITY

We can distinguish between vague and non- vague (or 'sharp') predicates:

1) **Vague predicates:** *Tall, bald, large, hot, cool*
   Have a denotation gap, $[P]^{?}$: Some entities are such that one does not know if they have that property or not. They are neither in $[P]^{+}_{c}$ nor in $[P]^{-}_{c}$
2) **Non-vague ('sharp'):** *Bird, apple*
   No denotation gap, by and large everything is in $[P]^{+}_{c}$ or in $[P]^{-}_{c}$
3) **Contrast I**: Adjectives tend to be vague and nouns tend to be sharp. But:
   Chair is a vague noun (Kamp & Partee 1995); *even (number)* is a non-vague adjective.

We can also distinguish between gradable and non-gradable predicates:

4) **Gradable predicates** (*tall, bald, large, hot, cool*) can combine with comparatives *(more P; less P*) equatives (*equally P*), and superlatives (*the most P*)
5) **Non-gradable predicates** (*Bird, apple, chair, extinct, even (number)*) cannot occur (bare) in these structures (**more P,* less P, *as P as, *the most P*).
6) **Contrast II**: By and large, vague predicates (adjectives) are gradable. Nouns are not vague and hence not gradable.

2 THE STANDARD LINGUISTIC ANALYSIS

Gradable predicate are standardly analyzed as mapping individuals to degrees (. For ex. *tall* maps its argument into a degree, $deg_{(tall,x)}$, on the dimension *height*. Thus *tall* is associated with several elements (Kennedy 1999; Rotstein and Winter 2004):

7)
   a. A set of *degrees*, S_{tall} (say – a set of number)
   b. An *ordering* on this set $\leq_{tall}$, which states for each two degrees which one represents the larger degree under *tall*
   c. A *unit of measurement* (say – centimeters)
   d. A *dimension* which these degrees measure (*height*, as opposed to say – *width*).

Sam is tall is considered true in a context c iff Sam's height, $deg_{(tall,Sam)}$, reaches the *standard* for tallness in c.

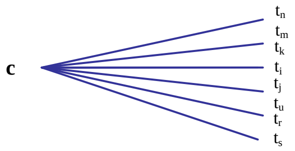
8)
   a. $[\ P\ ]^{+}_{c} = \{d \in D:\ deg_{(P,d,c)} \geq_{P,c} Standard_{P,c}\}$
   b. $[\textit{Sam is tall}\ ]_{c} = 1$ iff $deg_{(tall,[Sam]c)} \geq_{tall} Standard_{tall,c}$

In addition, given the generalization in (6), gradability is often analyzed as vagueness dependent.

Vagueness is often represented using models in which the semantic interpretation is relative to information states (or ***contexts***) c, in which ***predicate denotations are only partially known*** (van Fraassen 1969; Kamp 1975; Fine 1975; Veltman 1984; Landman 1991 etc.)

For ex.: In a partial context c the positive denotation of *tall*, $[tall]^+$, may consist of only very tall items, and the negative denotation $[tall]^-$ may consist of very short items. In such a context, we don't yet know if anything else, which is neither very tall nor very short, is tall or not. But each partial context c is extended by a set of **total contexts (supervaluations)**, t – all the possibilities seen in c to specify the complete sets of *tall* and *non-tall* things, *bald* and *non bald* entities etc.

For the analysis of gradability, it is generally, assumed (Kamp 1975; Kamp and Partee 1995) that we can use simplified vagueness models which contain but one partial context c (the ground model) and a set T_c of the total contexts t extending c. The intermediate steps between c and each t are thought to be unimportant (Figure 1). The total contexts are thought to represent different *standards* of precision (Lewis 1979). In some of them only very tall things are regarded as tall enough to be considered *tall*, in others more things are considered *tall*, etc.



Standard height in t_n:	2 meter
Standard height in t_m:	1.95
Standard height in t_i:	1.9
Standard height in t_u:	1.85
...	...

Figure 1: The context structure in a simplified vagueness model Mc

If we do not know what the standard is, we can only consider as *tall* those entities which are tall in every total context above c. van Fraassen 1969 has invented the term ***super-truth*** for truth in every total context. For example, *Sam is tall* is considered true (or super-true) in a partial context c iff Sam's degree of height, $deg_{(tall,Sam)}$, reaches the standard in every total context above c (9), that is, Sam's degree ought to exceed any degree which might yet be the standard of *tallness*.

9) Supertruth: $[Sam\ is\ tall\ ]_c = 1$ iff $\forall t \in T, t \geq c$: $deg_{(tall,[Sam]t,t)} \geq_{tall,t} Standard_{tall,t}$

Thus, it is often assumed, that a comparative statement like *Dan is taller than Sam* is true in c, that is, the degree of *Dan* exceeds that of *Sam*, iff *Dan* is tall relative to more standards, that is: *Dan is tall* is true in more total contexts above c, compared to Sam (Kamp 1975; Fine 1975). If Sam reaches a certain standard of tallness, Dan certainly reaches this standard, but not vice versa. Dan's height reaches certain standards which Sam's height does not reach.

10)

a. $[Dan\ is\ taller\ than\ Sam]_c = 1$ iff $deg_{(tall,Dan,c)} >_{tall,c} deg_{(tall,Sam,c)}$ iff:
$\{t \in T \mid [Sam\ is\ Tall]_t = 1\} \subset \{t \in T \mid [Dan\ is\ tall]_t = 1\}$

b. $[a\ is\ more\ P\ than\ b]^+_c = 1$ iff $\{t \in T \mid [P_{(a)}]_t = 1\} \subset \{t \in T \mid [P_{(b)}]_t = 1\}$

The standard analysis of comparatives in (10) has 2 problems that are addressed in parts 3-4.

3 THE ANALYSIS OF GRADABILITY

The first problem is that the standard analysis of comparatives in (10) applies to gap members only. All the entities which are already known to be *tall* in c, are *tall* in all the total contexts extending c. Hence they are wrongly predicted to be all *equally tall*. But intuitively two tall individuals can stand in the relation *taller than* to each other.

11) Wrong Prediction of (10): $\forall d_1,d_2 \in [tall]_c$: d_1 and d_2 are *equally tall*
$\forall d_1,d_2 \in [\neg tall]_c$: d_1 and d_2 are *equally tall /short*

I have shown, directly following Landman 1991's analysis, that this problem is solved by using a *standard non-simplified vagueness model*, namely one that contains many *partial contexts*, as demonstrated in Figure 2. Such a model can represent *information growth*: The order in which entities are categorized under the predicates through contexts and their extensions. We start with a zero context, c_0, where denotations are empty, and from there on, each context is followed (or extended) by contexts in which more entities are added to the denotations. In a total context t, every entity is either in the negative or in the positive denotation of each predicate. (Formally, c_1 is ***extended*** by c_2, $c_1 \leq c_2$, iff $\forall P \in PRED$: $[P]^+_{c1} \subseteq [P]^+_{c2}$ & $[P]^-_{c1} \subseteq [P]^-_{c2}$).

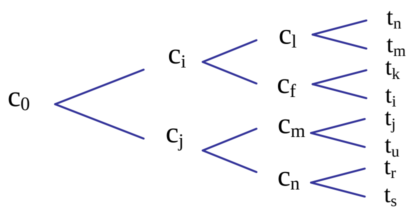


Figure 2: The contexts' structure in a standard vagueness model

In such a model, we can distinguish between the directly given denotation $[tall]^+_c$ which consists of the things which are directly known to be *tall* in c, and the ***superdenotation***, $[tall]_c$ (in (12a)), which consists also of all the things which end up being tall in all the total possibilities – all the things which, given our knowledge, must be tall:

12) Superdenotations:

a. $[tall\,]_c \;=\; \cap\{\, [tall]^+_t \mid t \in T, t \geq c\}$

b. $[\neg tall\,]_c = \;\cap\{\, [tall]^-_t \mid t \in T,, t \geq c\}$

The gradable structure of *tall* reflects the order in which entities are added to the super-denotation (the order in which entities are learnt directly or by inference to be tall).

Thus, two denotation members can stand in the relation *taller*. They do so iff one of them was added to the denotation in an earlier context. For example, (13), *Dan is taller than Sam*, is true in t iff in every c leading to t: If Sam is already considered *tall* (that is, Sam reaches the standard, be it what it may), then we can <u>infer</u> that Dan is *tall* (Dan definitely reaches the standard, given Dan's larger height). And if Dan is considered *not-tall* in c (that is, Dan does not reach the standard), then Sam is definitely *not tall* (Sam definitely does not reach the standard, given her lower height).

13) [*Dan is equally tall or taller than Sam*]$_t$ =1 iff

$\forall c \leq t$: $([Sam]_c \in [tall]_c \rightarrow [Dan]_c \in [tall]_c)$ &

$([Dan]_c \in [\neg\, tall]_c \rightarrow [Sam]_c \in [\neg tall]_c)$.

Thus, we replace (10) by (14) (*The learning constraint*): d_1 *is more P* than d_2 in a context t iff: **Either** the P-hood of d_1 is established before the P-hood of d_2 (i.e., in a context that precedes the context in which d_2 is added to the positive denotation), **Or** the non-P-hood of d_2 is established before the non-P-hood of d_1 (i.e., in a context that precedes the context in which d_1 is added to the negative denotation).

14) <u>*The learning constraint*</u>

$\forall t \in T$: $(\langle d_2, d_1 \rangle \in [\leq P]^+_t)$ In any total t, d_1 is *equally or more P* than d_2 iff:

$\forall c \leq t$: $(d_2 \in [P]_c \rightarrow d_1 \in [P]_c)$ & $(d_1 \in [\neg P]_c \rightarrow d_2 \in [\neg P]_c)$.

In any context c under t, if d_2 is P, d_1 is P, and if d_1 is $\neg$P, d_2 is $\neg$P.

In sum, P's ordering in t is the order in which entities are learnt to be P or $\neg$P (whether directly or by inference) in the contexts under t. Gradability is not connected to vagueness per se, but to the order of *vagueness removal* (learning). ⇒ It can characterize sharp predicates.

4 THE ANALYSIS OF NOUNS

The second problem of the standard analysis of vagueness and gradability is the following. The noun *chair* is vague (things like stools may be regarded as neither chairs nor non-chairs) but it is not gradable (**more chair*). Vagueness (or its removal) seems to turn adjectives, but not nouns, gradable. The nouns seem to be inherently non gradable. Hence, the standard linguistic theory does not associate nouns with a gradable structure (a set of degrees, an ordering dimension etc.)

A gradable concept structure in nouns:
The problem is that the last forty years of research in cognitive psychology have established beyond doubt that speakers consider certain entities as better examples of nouns than others (for instance, *robins* are often considered more typical *birds* than *ostriches*).
These ordering judgments are reflected in *online categorization time*: Most importantly, verification time for sentences like *a robin is a bird* is faster than for sentences like *an ostrich is a bird* (Rosch 1973; Armstrong et al 1983).

Ordering dimensions and vagueness in nouns:
In addition, speakers associate nouns with ordering dimensions (features like *feathers, flying, nesting, singing* etc.)
The classical view considered these features definitional: Necessary and sufficient conditions for membership in the denotation. But Wittgenstein 1968 (1953) and Fodor et al 1980 have shown that this idea is rarely if ever met.
For example, it is already well known that counterexamples can be found to any definitional feature you would propose for natural categories like *games* or *bachelors*.
In addition, often, speakers are uncertain about the membership status of some entities and they vary their judgments in different times or contexts, refuting the assumption that there are clear-cut criteria. For example, tomatoes fall in between the categories *fruit* and *vegetables*.
While speakers rarely (only 3% of the time in average) change their minds about the category membership of clear instances, they do so much more often (above 20% of the times in average) with regard to the membership of borderline cases, like curtains for *furniture* or avocado for *vegetables* (Murphy 2002: 20).
Crucially, the features which people link with a category like *bird* are raising the typicality of entities in the category, that is, they *are* indeed ordering dimensions, which together help to measure the typicality (and membership likelihood) of entities in the category. Thus, the standard theory in cognitive psychology associates a concept like *bird* (or the word that denotes it) with a prototype, namely:

15) The basic prototype theory:

a. A set of ***dimensions***. The feature set of *bird*:
$\mathbf{F_{(bird,c)}} = \{\textit{feathers, flying, nesting, singing, small size} \ldots\} \subseteq \text{PRED}$

b. Each dimension F has a weight W_F. For example, W_{flying} tells us how distinctive *flying* is of birds: How important *flying* is in discriminating birds from non-birds.

c. *The weighted mean hypothesis:*
For each entity d, its degree of typicality (or similarity to the prototype of P), $deg_{(d,P)}$, equals the weighted mean of d's degrees in the category features: $\mathbf{deg_{(d,P,c)} = \Sigma_{F \in F(P,c)} W_{F,c} deg_{(d,F,c)}}$

For instance, the typicality degree of a robin in *bird*, $deg_{(robin, bird)}$, is indicated by *the weighted-mean of its degrees in all the bird features:* How well it scores in *flies, sings, small* etc.
That is, the typicality of an entity in a category (e.g. *bird*) represents the extent to which it possesses the features that are distinctive of the category. The typical instances (*robins*) are more similar to the prototype, that is, they have more properties or they average better in the features, compared to the atypical instances (*ostriches*).
The weighted mean principle directly accounts for the fact that even noun features which are often thought to be *necessary* (like *unmarried* for *bachelor* and *young male* for *boy*) are no more than *important typicality features*.
First, they are not strictly necessary for membership: The degree of an entity in unimportant features, like *childish look or behavior* for *boy*, may compensate for violations in the important features. This was established for many different types of nouns, ranging from natural kinds to artifacts, through family relation categories like *uncle* and *grandmother* (Hampton 1979; 1995).
Second, nor are these features sufficient for membership: Many items which satisfy them are not clear members: The pope was never married, but is he a bachelor?

There are tight relations between the entity ordering and the denotation:
The basic cognitive theory argues that categorization is based on typicality (or similarity to the prototype). A certain degree of typicality functions as a standard, such that:

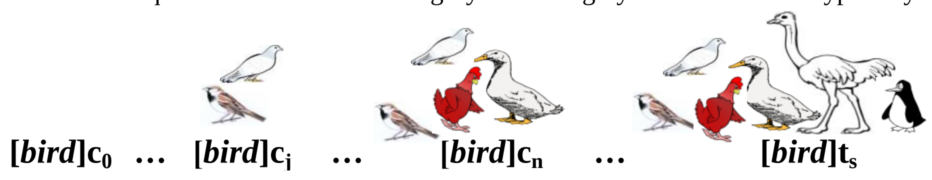
16) *The categorization criterion*: $[P] = \{d \in D \mid deg_{(d,P)} \geq standard_P\}$
An entity is classified in a category iff its typicality degree reaches the standard.

Indeed, there is abundant evidence showing that entities are positively classified iff their *average in the features* reaches criterion. Hampton 1998 analyzed the data about 500 items in 18 categories (McCloskey and Glucksberg 1978). He found a very strong coupling between their mean typicality ratings and the probability that they were categorized positively. There were also deviations, but they were highly systematic. The deviations were shown to occur due to (i) *shift of weight*, in typicality judgments compared to membership judgments, towards non-definitional perceptual criteria (which increase the typicality of non-members) (ii) *unfamiliarity* (lack of knowledge about the features of members reduces their typicality), and (iii) the existence of *competing categories*, like *kitchen utensil* and *furniture*. This reduces the likelihood of classification, but not the typicality of, say, a *refrigerator* in *furniture*.

It follows that the typicality ordering is determined by the order in which entities are added to the denotation (whether directly or by inference):
Dan is *more typical of a bird* than Sam iff Dan's bird-degree exceeds Sam's degree. That is iff: If Sam is already considered a *bird* in c (that is, Sam reaches the standard, be it what it may) then we can infer that *Dan* certainly does so, and hence it certainly counts as *a bird*. And if Dan is known *not to be a bird* (not to reach the standard), we can infer that Sam, due to its lower typicality degree, does not reach the standard and is *not a bird*.
Thus, the gradable typicality structure of *bird* reflects the order in which entities are added to the super-denotation (the order in which entities are learnt directly or by inference, to be *birds*), as expected by The learning constraint.

Some very robust findings, the *order of learning effects*, form evidence for this generalization. Most importantly, typical instances are acquired earlier than atypical ones, by children and adults (Mervis and Rosch 1981; Rosch 1973; Anglin 1977; Murphy and Smith 1982). For example, birdhood is normally determined first for *robins* and *pigeons*, later on for *chickens* and *geese*, and last for *ostriches* and *penguins*. Similarly, non-birdhood is determined earlier for *cows* than for *bats* or *butterflies*:
Second, language learners, learn faster if initial exposure is to typical category members (the crucial factor is not the amount of examples but their typicality; Mervis & Pani 1980), etc.

Figure 3: A normal acquisition order for the category *bird* is highly indicative of the typicality structure



In sum, nouns behave very much like our semantics for gradable predicates expects.

The typicality (or graded concept structure) effects in nouns, are robust and pervasive. A prominent psychologist (Murphy 2002) has written that it would be very surprising to find a cognitive task that typicality does not affect. Thus, by assuming that nouns are non gradable, in order to account for their infelicity in the comparative, linguists pay a heavy price in terms of the dissociation between the basic semantics they assume for nouns and many other things that we know about them.

Linguistic symptoms of gradability in nouns

Some phenomena which remain unexplained if nouns are assumed to be non-gradable are completely linguistic ones.

First, for example, we see in (17a) that it is sufficient to add the particle *of* to the comparative morpheme (as in *more of a bird*) and the interpretation of the noun *bird* turns gradable.

Second, it is very easy to turn a noun gradable by turning it into an adjective, either by modifying it with *typical* (17a) or simply by the adjectival morpheme 'y' (as in *birdy* or 17b).

This facts are hard to explain if nouns are merely (linguistically) non gradable.

17)

a. *A robin is more (typical) of* ***a bird*** *than an ostrich*

b. *The noun 'activity' is "nounier" / less "nouny" than the noun 'bird'*

Third, the typicality effects characterize very complex predicates. For example in (18) we see that a gradable structure pops up in a very complex noun phrase:

18) *…pretty much typical of a non-fan, non-entertainment, smart up-market British paper*

So typicality is highly productive. Within a context, we can produce typicality orderings for novel complex-concepts on the fly. It seems that *some* generative system plays a role here.

Forth, some adjectives are *multidimensional* (*healthy, intelligent, talented, good* etc.) For instance, the adjective *healthy* can be measured by dimensions such as *blood pressure*, *pulse* and *fever*. Now, the features of multidimensional (but not of one-dimensional) adjectives can be quantified over. This is demonstrated by the contrast between (19a) and (19b) (Bartsch 1986; Landman 1989). (19c) shows that quantification over dimensions is impossible in nouns.

19)

a. *Maria is healthy in every respect / generally healthy / healthy wrt blood pressure*

b. *? Maria is tall in every respect / ? Maria is generally tall / ? tall wrt height*

c. *# Tweety is a bird in every respect /# generally a bird/ # a bird wrt flying*

But like the felicity of nouns in the comparative, also quantification over the dimensions becomes possible, if the noun is slightly modified (20). Again, this fact is hard to explain if nouns are not associated with ordering dimensions.

20) *Tweety is a typical bird in every respect / generally typical of a…/ typical of…wrt flying*

Can we capture these facts while maintaining the assumption that nouns are neither vague nor gradable (in the usual sense)? This is precisely what Kamp and Partee's influential 1995 supermodel theory has attempted to do, but failed. First, it has the same problem as the standard vagueness based analysis (Kamp 1975): It fails to represent the difference in typicality levels between denotation members (or non-members). Thus, it completely fails to account for the typicality differences in sharp nouns. In addition, the supermodel theory's predictions concerning the typicality degrees in modified nouns were refuted empirically (Smith et al 1988). So the typicality effects are not accounted for by this theory.

5 NOUNS VERSUS ADJECTIVES: PROPOSAL

I propose that nouns are gradable and multi-dimensional. Crucially, even non-vague denotations are learnt gradually. This gradation forms an ordering, as *the learning constraint* predicts. The infelicity of nouns in the comparative is not due to lack of gradable meaning.

In cognitive psychology today, one-dimensional adjectives like *tall* are treated as rule based, because categorization under them does not involve averaging over dimensions. Conversely, in nouns like *bird* or *house*, categorization *is* based on averaging ('similarity to the prototype'). This distinction is important because there is evidence that rule versus similarity based categorization tasks recruit different brain systems (Ashby and Spiering 2004; Photos 2004) and their acquisition course seems to be different (perhaps due to late maturation of the rule based brain system; Keil 1979; Zelazo et al 1996; Thomason 1994).

So the distinction between one dimensional and multi-dimensional-categories seems to be a cognitively real distinction. Maybe it is this distinction that has been grammaticized into the two categories – nouns and adjectives:

21) <u>The feature constraint</u>

a. All predicates P are associated with a feature set $F_{(P,c)} \subseteq PRED$

b. In ***one-dimensional adjectives*** (like *tall*) the feature set consists of one feature. The degree of each entity equals its degree in this feature. For example:
[*Dan is tall*]$_c$ = 1 iff $deg_{(Dan,\mathit{height})} \geq standard_{\mathit{tall}}$.
$\forall d \in D$: $deg_{(d,P)} = deg_{(d,\sigma(F(P,c)))}$ & ($d \in [P]_c$ iff $deg_{(d,P)} \geq standard_{\sigma(F(P,c))}$)

c. In ***multidimensional adjectives*** (like *healthy*), in each context of use a *with respect to argument* (say – *wrt blood pressure*) selects one dimension, $wrt_{(P,c)}$, from $F_{(P,c)}$. The degree of each entity equals its degree in this dimension, and this dimension determines P's standard in that context of use. For example:
[*Dan is healthy wrt to blood pressure*]$_c$= 1 iff $deg_{(Dan,\mathit{blood\ pressure})} \geq standard_{blood\text{-}pressure}$.
$\forall d \in D$: $deg_{(d,P)} = deg_{(d,wrt(P,c))}$ & ($d \in [P]_c$ iff $deg_{(d,P)} \geq standard_{\sigma(wrt(P,c))}$)

d. In ***nouns*** the degree of an entity equals the weighted mean of its degrees in the features. The feature weights w_f and averaging method Σ may vary between uses (Smith and Minda 1998). For example:
[*Tweety is a bird*]$_c$ = 1 iff $(\Sigma_{F \in F(bird,c)} w_f deg_{(Tweety,F)}) \geq standard_{bird}$.
$\forall d \in D$: $deg_{(d,P)} = \Sigma_{F \in F(P,c)} w_f deg_{(d,F)}$ & ($d \in [P]_c$ iff $deg_{(d,P)} \geq standard_P$)

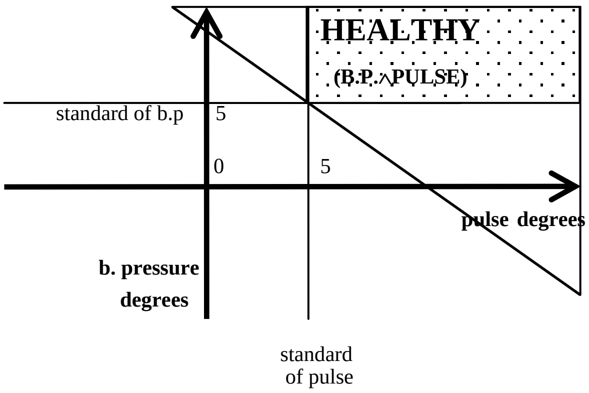


Figure 5: Multi-dimensional adjectives

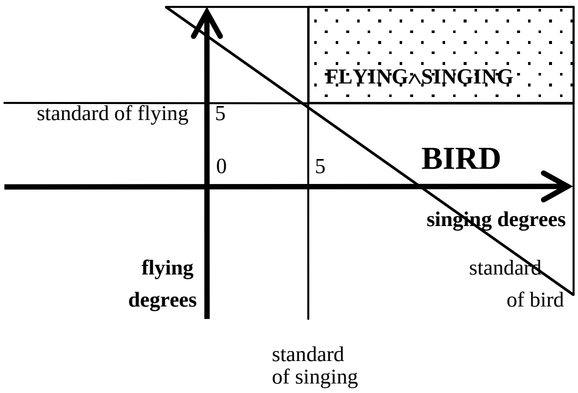


Figure 4: Nouns

1. Nouns:
The postulate in (21d) accounts for the fact that none of the features of nouns necessarily adds a categorization criterion. Even when typicality in *bird* is measured by typicality in *flying*, *singing*, and *nesting*, these features are not regarded as necessary for bird-hood (Wittgenstein 1953). Rather, entities are regarded as birds iff their *average in the features* reaches criterion (Hampton 1979; 1995).
Figure 4 demonstrates a two dimensional case, but the argument applies to n dimensional cases and all types of averaging functions. The axes stand for the set of degrees in *flying* and in *singing*. The set of birds is not given by feature intersection (that would wrongly give us the square to the right), but by averaging. An averaging formula like the one in (22) gives us the points above the straight line which represents the standard of *bird* (the triangle).

22) $[bird] = \{d \in D \mid w_{flying}deg_{(d,flying.)} + w_{singing}deg_{(d,singing)} + \ldots > standard_{bird}\}$

Given postulate (21d), we now have a reasonable story about the memory representation of predicate intensions (functions to potentially infinite sets). We only need to directly represent in memory a (small) finite sample of features and instances for each predicate. We predict (like the basic prototype theory) that knowledge of the *bird* features will trigger automatic categorization of new entities, once they become accessible, which are at least as good in the features as the known instances.

Indeed, Kiran and Thompson 2003 found that, **in aphasics and neural networks which were taught the category *features*,** training with atypical members (say, a chicken or a goose), results in spontaneous recovery of categorization of untrained more typical items (items which are better in the features, like ducks), but not of untrained less typical items (ostriches.)
Similarly, Reed and others show that in healthy subjects, **previously unavailable typical instances are frequently (falsely) assumed to be known** (Reed 1988; Mervis & Rosch 1981). Why? If less typical entities are denotation members, entities which are more typical (due to their high scores in the typicality features) should definitely be denotation members!

2. Multidimensional adjectives:
The postulate in (21c) accounts for the fact that each feature of a multidimensional adjective does add a categorization criterion. For instance, if health is measured by levels of *blood pressure*, *pulse* and *fever*, then one is *healthy* iff one is within the norm in all these respects. So the bare predicate *healthy* is interpreted as a conjunction of three one-dimensional adjectives (23a), that is, by feature intersection (23b). The set of healthy entities is the square in figure 5.

23)

a. $[\textit{Dan is healthy}]_c = 1$ iff $\forall F \in F_{(healthy,c)}$: $[\textit{Dan is healthy wrt F}]_c = 1$

b. $[\textit{healthy}]_c = [\wedge F_{(healthy,c)}] = \{d \in D \mid deg_{(d,b.p.)} > standard_{b.p}$ & $deg_{(d,pulse)} > standard_{pulse} \ldots\}$

Thus, the denotation of multidimensional adjectives like *healthy* or conjunctions like *blood pressure and pulse* cannot be indicated by averaging on the features (or conjuncts): A formula like $\mathbf{w_{b.p}deg_{(d,b.p.)} + w_{pulse}deg_{(d,pulse)} > standard}$ wrongly selects the triangle in figure 5.

3. A wrt argument and hence quantification over features makes sense only when a predicate has several features which add categorization standards (only when, indeed, an entity can fall under it in one respect and not in another respect), namely, in multi-dimensional adjectives.
A determiner which quantifies over respects states how many of the features form categorization criteria in the context of use. So we can relax the meaning of *healthy* by accommodating an existential or a vague universal quantifier over the respects:

24)

a. $[\textit{Dan is healthy in every respect}]_c = 1$ iff $F_{(healthy,c)} \subseteq [\lambda F.\textit{Dan is healthy wrt F}]_c$
(iff *Dan is healthy* wrt every *health* feature).

b. [*Dan is healthy in some respect*]$_c$=1 iff $F_{(healthy,c)} \cap [\lambda F.$*Dan is healthy wrt F*$]_c \neq \varnothing$
(iff *Dan is healthy* wrt some *health* feature).

c. [*Dan is generally healthy*]$_c$ = 1 iff, roughly, *Dan is healthy* wrt most of $F_{(healthy,c)}$.

But neither the one-dimensional adjective *tall*, nor the noun *bird*, have two different necessary criteria, so *wrt* modification and quantification over the features is infelicitous. Philosophers have argued that nouns do not have semantic necessary conditions for membership at all (Wittgenstein 1968 [1953]; Fodor et al 1980; Fodor 1998). Ordinary speakers, normally count an individual as a *bird* iff it *has bird genes*, or iff it *has a bird essence* (if they are a bit less educated). But crucially, speakers do not consider any other of the *bird* features to be a separate semantic necessary condition for birdhood (nor is this feature treated as strictly *necessary and sufficient*).
Features which merely function as domain restrictors, but do not help to distinguish *birds* from *non birds*, are irrelevant. For instance, in most contexts *animal* restricts the set of birds but also the set of non-birds: We hardly ever call prime numbers *non-birds*.
Finally, an expert on birds might in principle characterize birds by the possession of, say, 100 separate genes (which occur in every bird and only in birds). Our proposal correctly predicts that in such a context, the expert might indeed describe new species which possess only 50% or 100% of the bird genes, as *birds in this respect but not in that respect*.

25) **+/–Q(uantification over features) & wrt-argument**:

a. **+Q/wrt** predicates (e.g., *healthy*, *talented*) have <u>several</u> necessary condition.

b. **–Q/wrt** predicates (e.g., *bird* or *tall*) have at most one necessary condition.

A wrt argument may occur also in lexical entries which are derived from multi-dimensional adjectives, like their nominalizations (*health wrt b.p.*; *success wrt to profession*), or the animate nouns which are construed from them (*an Italian wrt to citizenship*). Since they are derive from adjectives, these nouns have many other adjectival features.

4. <u>The adjective *typical*</u> "turns its argument gradable" solely by adding a *wrt* role:

26)

a. [*Tweety is typical of a bird wrt flying*]$_c$ = 1 iff $deg_{(Tweety,flying)} \geq standard_{flying}$.

b. [*Tweety is typical in every respect*]$_c$ =1 iff $F_{(bird,c)} \subseteq [\lambda F.$ *Tweety is typical wrt to F*$]_c$
(iff *Tweety is typical* wrt every bird feature).

This proposal predicts the intuition that *typical of P* has more categorization criteria than *P*, although it is hard to put a finger on the exact features which add criteria, if $wrt_{(P,c)}$ isn't specified and a determiner like *generally* is accommodated.
The same analysis for nouns with a morpheme like '*y*' ('*nouny*', '*primy*') predicts the ease with which a noun can turn adjectival.

5. <u>The comparative morpheme</u> denotes an operation on the dimension of its predicative arguments. It is undefined when there is no unique dimension.
Hence, it is clearly defined in *one-dimensional* adjectives.
It is defined for *bare multi-dimensional adjectives*, because they can be interpreted wrt the conjunction of features (which then forms a unique dimension).

But it is undefined in *nouns*, where the conjunction of features cannot form a dimension. The ordering given by it is different from the noun ordering. Birds which fall under all the features but have a relatively low average are less good in *bird*, but better in the conjunction, compared to birds which violate one feature but have a higher overall average.

27) A possible semantic interpretation for *more* that would do the job:
$[d_1$ is *more P* than $d_2]_t = 1$ iff $deg_{(d1,\wedge F(P,t))} > deg_{(d2,\wedge F(P,t))}$
(d_1's degree in the conjunction of P's features exceeds d_2's degree).

In adjectives, the denotation is given by the intersection of the features ($[P]_t=[\wedge F_{(P,t)}]_t$). Hence, in adjectives (27) gives the right result: Each member ends up more P than each non member (d_1 is more P than d_2 iff the degree of d_1 exceeds the degree of d_2 in P: $deg_{(d1,P)}>deg_{(d2,P)}$).
But in nouns, *more* fails to correctly represent their ordering, so it is incompatible with them: By (27), birds which fall under all the features but have a relatively low average, end up *more birds* than birds with a higher average degree, which violate one feature. That is, the former birds end up *more birds* than the latter, though their degree is lower ($deg_{(d1,bird)}< deg_{(d2,bird)}$).

In sum, *more* selects *rule based* predicates: Predicates with one dimension which also serves as a categorization criterion. A *rule* is a predicate which may be formed by conjoining several dimensions, but not by averaging across several dimensions.

As for *typical*, this proposal correctly predicts that, say – non biological fathers, may be judged *more typical fathers* than biological fathers. None is, strictly speaking, *a typical father* (none falls under the conjunction of all the features), but among the things which are not *typical-fathers*, the real father may be much less typical.

Concerning *conjunctions of gradable adjectives*, they have at least 2 (necessary) features: The conjuncts. But conjunctions do not have a *wrt* argument. A wrt argument may only be selected for each conjunct separately, from its own feature set. If all the conjuncts select the same *wrt* argument, the use of the conjunction rather than just one conjunct becomes superfluous. And if each conjunct selects a different *wrt* argument, the conjunction remains multi-dimensional. Thus, we predict that conjunctions of gradable predicates would be non-gradable. And this prediction seems to be supported by the facts. For instance, I found (in 35 subjects) that, the preferred interpretation for *more bald and tall* is *balder and taller*. That is, *and* is not within the scope of *more*. So '*more*' does not take 2 features simultaneously.

6. The morphological complexity of the comparative versus the positive forms of predicates

In many adjectives non-members are not ordered at all. The comparative ordering is constrained to the positive denotation, as demonstrated in (28)-(29).

28) *Dan and Sam are healthy. # Dan is more / less sick than Sam*
29) *Dan and Sam are intelligent. # Dan is more/ less stupid than Sam*

Thus, the comparative forms of many adjectives presuppose that the denotation is predetermined.
But the positive forms of many adjectives in the current analysis do not presuppose that the ordering is pre-determined. The denotations of *healthy* and *sick* are determined by the dimension intersection rule, not based on the entities' degrees being above a threshold degree.
This may directly explain the fact that the morphological form of the comparative relation is cross-linguistically more complex than the positive form.
(It is also possible that in the use of *more* sometimes we do resort to averaging over the dimensions, but we employ comparisons by averaging in the positive and negative denotations ($[\wedge F(P)]_t$ and $[\neg\wedge F(P)]_t$), separately. Thus, the denotation has to be predetermined).

7. Incommensurability

According to Kennedy 1999, the infelicity of examples like (30a) is due to the fact that *tall* and *heavy* have two different ordering dimensions. This makes them incommensurable.

30)

a. *# The table is more tall / taller than it is heavy*
b. *? Dan is more hungry than tired*
c. *? Dan is happier than Sue is funny*

The main difficulty in this account is that comparisons between different dimensions freely occur in nouns (31).

31)

a. *Bling Bling says "tweet" (i'm convinced he's* ***more a bird than*** *a cat).*
b. *... giving me three bits of furniture which she didn't want anymore (a coat rack, chair, and stool thing which is really* ***more a table than*** *anything else*
c. *The "wall" was rolling backward until it come to a horizontal position, now being* ***more a table than*** *a wall*
d. *A bat is no* ***more a bird than*** *a whale is a fish* [invented]

Conceptually, it makes sense to compare the degrees of entities in two different scales only if the scales are normalized (so as to have the same range). The comparison reflects the relative position of each of the entities in the respective scale.
Comparison by normalization (relative position) is freely available in nouns because their scales are normalized in the first place (the dimensional scales are normalized for the purpose of averaging and the resulting scale has the same range). In addition, their denotation is fixed by the threshold rule, not by any other necessary condition. Thus, it makes a lot of sense to compare the relative position of things in two different nouns, so as to decide whether they should better be referred to using this or that noun. This may even affect their classification, if the nouns are seen as contrasting (*plate* versus *bowl*; *mammal* versus *bird*).
The determination of the degree function in *adjectives* (like *tall*) does not involve averaging (they are one-dimensional), so their scale need not be normalized (and it need not have a maximum or a positive range at all). Thus, interpretation by normalization is much less freely available. In addition, in the multi-dimensional adjectives, the denotation is fixed by dimensions intersection, not by a threshold rule. Thus, comparison by normalization is less likely to be informative wrt to classification; compared to the same situation in nouns (it requires a richer context).

8. Numerical degrees

In my previous talk I have claimed that numerical degrees are not intuitive primitives and that only when they are really needed, they are derived from the ordering between entities. But numerical degrees are required in order to give a semantic analysis to measure phrases like *one meter long* and *twice as happy*, in order allow for averaging across different dimensions in multi-dimensional predicates, and for other purposes.

For example, we have to assume that numerical degrees are assigned to entities by predicates like *long*, which accept numerical measure phrases like *two meters long*. How does this work? There is a convention about a certain object – *the meter* – that its *length* is 1 and the meter functions as an established conventional unit. Other objects are assigned the degree of length 1 iff they are *exactly as long as* the meter, and they are assigned the degree n iff they are *exactly as long as* the concatenation of n objects which are exactly *as long as* the meter.

Other predicates, like *happy*, can hardly have an established unit. If I decide to assign a certain mental state the degree 1, you can hardly have any evidence about the nature of this state. Thus, you will not be able to use it as an established unit for *happiness*. But the numerical degrees of entities depend on the unit which is being used. Since each one of us in each context may use another mental state of happiness as a unit, there is indeterminacy about the numerical degrees in *happy*.

Similarly, the degree function of predicates like *heavy, warm* and *cold* in the languages of the world may measure the mental states which are caused by external stimuli, rather than the stimuli themselves. Thus, they may not allow for numerical measure phrases (# *two kg heavy*).

Yet, x is *twice as long as y* is true for centimeters iff it is true for meters or inches (if you do not ignore digits after the dot). Thus, a sentence like *Dan is twice as happy as Sam* can be meaningfully used. It is true iff Dan's degree of happiness is two times Sam's degree, be the unit for *happiness* what it may (for any unit).

In sum, we have a good explanation for the indeterminacy in the numerical scale and in the mapping of entities to degrees in many predicates. At the same time, the assumption that numerical degrees are the underlying conceptual primitive allows a simple account for the use of proportional phrases like *twice as happy* (and for the typicality effects).

9. The typicality features in complex predicates

In my previous talk I have presented compositional rules, for instance, a union rule for the construction of a feature set for a conjunction or modified noun from the feature sets of the parts. My strategy was to formulate rules and then to look for explanations for the exceptions. I am now convinced that this was wrong. Why? Consider a simple case in which the two conjuncts are ordered by the same two features, F_1 and F_2. Conjunctions and modified nouns are interpreted by intersection, so their denotation can be represented by the points above the broken line in figure 6. But this means that their denotation cannot be given by averaging over the conjuncts' dimensions. This would wrongly produce a triangle as a denotation (say, the points above the dotted line in Figure 6).

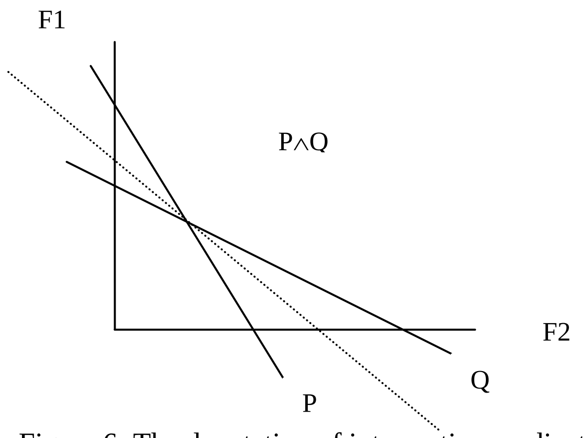


Figure 6: The denotation of intersective predicates is not given by the mean in the conjuncts' features.

This argument extends to any number of dimensions, and any type of averaging function. The crux is that categorization criteria which are based on mean functions (like: in (21d): $\Sigma F \in F(P,c)wfdeg(d,F) \geq s$) are continuous: They have a derivative (tangent) in every point. But the intersection of two such functions is a broken form, which more often than not does not have a derivative in the intersecting points. This proves that the feature set of a modified noun cannot be given by the union rule.

Thus, when we want to construct a typicality degree function and categorization algorithm for modified nouns like *male nurse, pet fish, pet bird* or *wooden spoon*, we must move to a new space of features, such that the average in them will give us the denotation. This predicts two well known phenomena: First, the ***emergence of new features*** (for example, *talks* and *lives in a cage* for *pet bird*, and *big* for *wooden spoon*) and second, the ***failure of inheritance of features from the constituents to the conjunction*** (for example, *fish* typically *live in the open oceans* and *pets* are typically *warm and affectionate*, but *pet fish* are neither).

We see that these effects are not despite, but *due*, to intersectivity!

The new features can be deduced, in the same way as they are deduced in basic lexical items.

10. Feature selection in nouns and modified nouns

The learning constraint makes the following bootstrapping mechanism or bias very plausible:

32) The learning bias: In the lack of knowledge about the category features, the early acquired members are the basis for generalization:

a. First, the early acquired members are assumed to be best in the category features.

b. Second, the category features are assumed to be precisely those properties in which these members average better than other familiar objects (which were classified later, or were classified as non-members, and hence should be worse in the features).

c. Finally, classification of new entities is inferred from their average in these features.

For example, normally, robins are classified earlier under *bird* and hence are considered more typical birds than chicken. Consequently, features in which robins average better than chicken (like *small* and *flying*) are linked to *birds*: Categorization is based on average in them.
Indeed, it was found that, **unless the features are directly taught**, acquisition is delayed if early exposure is to atypical items (say ostriches for *bird*), or even to the whole category, but not to the typical items first! (Mervis & Pani 1980). This is unexpected if features are selected based on their frequency within and outside the category, as is usually assumed. It is expected if features are selected based on the early acquired entities.
That first exposure to an atypical item slows down acquisition, supports the proposal that such situations trigger wrong inferences (32a-b). For example, if my initial exposure to birds was through ostriches, I would think that the ostrich is a representative *bird* and that its features (*running*, *ostrich size* etc.) are typical of the category. The inferences would be canceled later on, when items that fall out of the category would be discovered to be averaging better in these features than category members. But this process would slow down acquisition. Interestingly, in certain children acquisition is completely blocked within the experiment time. They refuse to abandon inferences from the learning bias. This gives further evidence for it.

To take another example, *hard* is a typicality feature of *boiled eggs*, though it is typical neither of *boiled* nor of *eggs.* The denotation is fixed by conjunct intersection and, precisely because of that, new ordering features like *hard* emerge. A feature emerges iff, all other things being equal, the higher one's degree in *boiled egg* (or the earlier one is learnt to be both an *egg* and *boiled*), the higher one's degree is in *hard* (the earlier one is learnt to be *hard*). Had there been pairs of equally good *boiled eggs,* which were identical in all except that one was *harder*, the feature *hard* would simply not have emerged!
Thus, we have a reasonable story about the acquisition of the typicality features, the typicality ordering, and with these – the intension, in simple and complex predicates.

CONCLUSIONS:

The new direction of explanation seems to be promising and fruitful, and it allows a considerable improvement in the psychological adequacy of the linguistic theory.
Empirical research is required for the purpose of establishing the facts concerning the multidimensional adjectives.

Thank you! **gala@post.tau.ac.il**