Salience and Feature Variability in Definite Descriptions with Positive-form Vague Adjectives

Raquel Fernández (raquel.fernandez@uva.nl)
Institute for Logic, Language & Computation
University of Amsterdam
P.O. Box 94242, 1090 GE Amsterdam, NL

Abstract

This paper focuses on definite descriptions involving positive-form vague adjectives such as '*tall*', '*big*' or '*heavy*'. A characteristic feature of descriptions such as '*the big triangle*' is that they can be used to refer to an object that differs from others only in the extent to which it possesses the property expressed by the adjective, even if that property would not generally be judged to be true of the object. We investigate the constraints that license the production of a definite description involving a vague adjective in the positive form and propose an approach that stresses the role of salience (modelled by means of a clustering algorithm) and the influence of feature variability (i.e. the variability of the range of values along a particular dimension such as *size*) within the conceptual categories that determine the domain of the adjective.

Keywords: Definite descriptions; relative gradable adjectives; vagueness; clustering; feature variability.

Introduction

Vague adjectives, also called "relative gradable adjectives", are expressions such as '*tall*', '*slow*' and '*expensive*', whose meaning involves reference to a dimension (e.g. height, speed, cost) along which entities for which that dimension is applicable can be ordered. A characteristic feature of these adjectives is that they can appear with different types of degree morphology, like comparative and superlative morphemes ('*taller than*', '*the heaviest*') and intensifier morphemes such as '*very*' or '*quite*'. In this paper we shall focus on definite descriptions involving the so-called *positive form* of these adjectives (e.g. '*tall*' as opposed to '*taller*' or '*tallest*'), which in most languages is the most morphosyntactically simple form a gradable adjective can take. More precisely, the question we shall address is the following: What kind of contexts license the production of a definite description involving a vague adjective in its positive form?

One of the aspects that makes the positive form particularly interesting is its striking context-dependence. To take a common example: If John's height is 185cm, he may be appropriately described as a tall man but probably not as a tall basketball player. Thus, what counts as tall (i.e. what falls in the positive extension of the adjective '*tall*') can vary from context to context, with the most relevant contextual parameter being a *comparison class* relative to which the adjective is interpreted (e.g. the set of men, the set of basketball players, etc.). In addition to being context-dependent, the positive form is also *vague* in the sense that the boundary between objects of which the adjective is definitely true and those of which it is definitely false is more a matter of gradual transitions than of sharp cut-off points. Vagueness gives rise to so-called *borderline cases* (i.e. entities for which it is not clear whether a gradable property holds or not) and is also related to the fact that often we find it inappropriate to distinguish between objects that are very close to each other along a continuum by means of a vague adjective.[1]

Consider a system that represents a domain as a set of entities and properties of those entities encoded as attribute-value pairs. As pointed out by van Deemter (2006), if gradable properties use relative values categorically, e.g. [HEIGHT = tall], then the system would never be able to produce (nor interpret) the expressions '*the tall man*' and '*the short basketball player*' as referring to the same individual in different contexts. If instead of relative values we allow more fine-grained information (such as information coming from sensors, e.g. [HEIGHT = 185cm]) as the value of gradable properties, then to produce referring expressions that do not involve measure phrases the system would need a model that specifies how this information is mapped onto descriptions containing vague adjectives such as '*tall*' in a given context. Our goal in the present paper is to explore some of the fundamental aspects needed for such a model.

We start by introducing the main aspects that characterise the felicitous use of positive-form vague adjectives in definite descriptions. We then propose to use a clustering algorithm to model how similarity within a comparison class determines the degree of salience of those points where the extensional boundary of a vague adjective can fall. After that, we show how some apparent counterexamples can be explained by taking into account different patterns of feature variability exhibited by conceptual types and refine the clustering model accordingly. We finish with a discussion of the approach and some conclusions.

Semantic Elasticity and (Dis)Similarity

Several researchers, from formal semanticists (Klein, 1980; Kyburg & Morreau, 2000; Barker, 2002) to psycholinguists (Ebeling & Gelman, 1994; Sedivy, Tenenhaus, Chambers, & Carlson, 1999; Syrett, Bradley, Kennedy, & Lidz, 2005), have pointed out the potential of positive-form vague adjectives for shifting the boundaries of their extensional meaning as a function of context. This potential is particularly apparent when the positive form is used in a definite description. Consider a pair of rods of unequal lengths. In such a context, the description '*the long rod*' can felicitously be uttered to refer to the longer of the two rods, regardless of whether any of the rods would generally be judged to be long or not.

[1] See e.g. Kamp (1975); Pinkal (1979); Williamson (1994).

Similarly to the case of '*the short basketball player*' alluded to before, the description '*the short rod*' could be used to refer to the very same object were it paired with a longer rod instead. Thus, assuming that if something is short then it is not long, whether an entity falls in the positive or negative extension of an adjective like '*long*' varies across contexts.

As mentioned above, this variation depends (at least partially) on a contextually determined *comparison class*. How to determine the right comparison class is an issue that we shall not tackle in depth in this paper. As explained in more detail in the next section, here we assume that comparison classes are computed from the salient, perceptually available context. For instance, in the situation outlined above we can consider the relevant comparison class to be restricted to the two present rods. Following Klein (1980), we define the following constraints on comparison classes:

- **(C1)** A positive-form vague adjective $\mathbf{A}$ partitions a comparison class C into two sets $\mathbf{A}^+$ and $\mathbf{A}^-$ corresponding to the positive and the negative extensions of $\mathbf{A}$, respectively.
- **(C2)** $\mathbf{A}^+$ and $\mathbf{A}^-$ must be non-empty and disjoint: $\mathbf{A}^+ \cap \mathbf{A}^- = \emptyset$, and hence $|C| \geq 2$.
- **(C3)** $\forall u, u' \in C$ if $u \in \mathbf{A}^+$ and $u' > u$ then $u' \in \mathbf{A}^+$ (and accordingly for $\mathbf{A}^-$).

By requiring that the positive and negative extensions of the adjective be non-empty within C, C2 emphasizes the importance of contrast for the interpretation of adjectives.[2] The relation $>$ in C3 defines the order needed for the right interpretation of comparative forms (if $u \in \mathbf{A}^+$ and $u' \in \mathbf{A}^-$ then u is $\mathbf{A}$-er than u').

We assume that in definite descriptions the definite article '*the*' introduces the presupposition that there is a unique salient object in the current context that satisfies the description. Given this and the constraints above, the interpretation of the description '*the long rod*' can be seen as a byproduct of (i) the fact that the salient comparison class in this context includes only two entities and (ii) the presupposition of unique existence introduced by the definite article. The idea underlying several accounts of this phenomenon[3] is that the extensional boundary of the adjective shifts as required to generate an interpretation for '*long*' in the current context that satisfies the presupposition introduced by the definite. Thus, under certain circumstances, the positive form can be used to discriminate between entities that simply differ along a particular dimension in a fashion akin to the comparative or superlative forms '*the longer/longest rod*', which do not require '*long*' to be true of the object that it is predicated of. We shall refer to this property of the positive form as *semantic elasticity*.

Semantic elasticity tells us that the denotational boundary of an adjective $\mathbf{A}$ can flexibly be shifted within a comparison class provided that C1–3 are satisfied. Given this and the presupposition of uniqueness carried by the definite, it would seem that a definite description involving the positive form can always be felicitously used to pick up the same referent that would be picked up by a superlative (e.g. '*the longest rod*'), which invariably refers to the maximal element in some ordered set. However, it turns out that this is not always the case. It has been observed that, for comparison classes that contain more than two elements, it is odd to use the positive form in a definite description when the distance between the intended referent and the closest elements in the ordered comparison class is not *salient*. Consider, for instance, a context with rods of lengths 4, 5, 9, 11, and 12. Here it seems not quite appropriate to produce the description '*the long rod*' to refer to the longest of the rods (cf. van Deemter (2004, 2006)).

If this observation is correct, then it shows that constraints C1–3 are not sufficient to account for the contextual conditions that license the presence of the positive form in a definite description. Informally, what this seems to indicate is that the boundary between $\mathbf{A}^+$ and $\mathbf{A}^-$ can fall between entities that differ from each other along the dimension associated with adjective $\mathbf{A}$ if and only if their distance along that dimension is more salient than their similarity within a particular comparison class. Borrowing Graff (2000)'s terminology, we shall refer to this restriction as the *similarity constraint* (SC). The next section presents a computational implementation of this constraint.

Computing Salience and Similarity Using Clustering

In order to provide an account of the SC a computational model needs to make precise what counts as saliently distant and conversely as similar enough. To this end, let us introduce a reflexive, symmetric, and non-transitive relation $\sim$ holding between entities $u \in C$ that satisfies the following conditions:

- $\forall u, u', u'' \in C$ such that $u > u' > u''$, if $u \sim u''$ then $u' \sim u$ and $u' \sim u''$.
- If $u' \sim u''$ then $u \in \mathbf{A}^+ \rightarrow u' \in \mathbf{A}^+$ (and accordingly for $\mathbf{A}^-$).

The effect of $\sim$ is to induce a partition of a comparison class C where subsets are non-overlapping intervals. We shall refer to subsets in this partition as *clusters*. The $\sim$ relation thus defines a set of clusters of *very similar elements*, all of them belonging to either $\mathbf{A}^+$ or $\mathbf{A}^-$. Consequently, it also defines a set of separation points or *gaps*, which we take as the points that can serve as potential denotational boundaries in C for a vague adjective $\mathbf{A}$. The idea we want to capture is that the positive form is not able to distinguish between elements that are *very similar* (i.e. related by $\sim$) given the distribution in a given comparison class. In the case of (singular) definite descriptions, this implies that the positive form will be licensed only if the target referent is not related by $\sim$ with any other element in the comparison class or, in other words, if there is a potential denotational boundary that singles out the intended referent.

[2] See Sedivy et al. (1999) and citations therein.

[3] See e.g. Kyburg and Morreau (2000) for a formal account or Syrett et al. (2005) for a psycholinguistics perspective.

Of course we do not go very far in our attempt to offer a computational cognitive model that explains the use of the positive form if we simply stipulate the existence of a relation $\sim$ without trying to model how this relation is computed. Our aim in this paper, following some ideas sketched in McNally (2007), is to explore the possibility of implementing the $\sim$ relation by means of a clustering algorithm. Other researchers who noted some of the problems associated with what we have named the SC have proposed different solutions. For instance, van Deemter (2006) uses a fixed threshold value to determine whether the distance between the intended referent of a definite description and the other elements in the comparison class allows the use of the positive form. Within the area of formal pragmatics, van Rooij (2008) has proposed to define comparison classes as semi-orders in an approach that shares some aspects with the one we adopt here. In this paper, inspired by McNally (2007) and in line with other computational work on perception and referring expressions (Thórisson, 1994; Funakoshi, Watanabe, Kuriyama, & Tokunaga, 2004; Gatt, 2006), we propose to use standard, well-understood clustering methods to model the notions of distance and similarity required. But first, a few words about the representation of the domain.

Domain Representation. We shall use records to represent the entities in the current scene or *local context*. Records are finite sets of ordered pairs $\langle \ell, v \rangle$ of *labels* and *values*, which we represent graphically as matrices with fields $[\ell = v]$. We may sometimes refer to labels as *features*. The following record represents a rod:[4]

(1) $$e_1 =_{def} \begin{bmatrix} \text{TYPE} & = & \textit{rod} \\ \text{SHAPE} & = & \text{shape23} \\ \text{LENGTH} & = & \text{12cm} \\ \text{COLOUR} & = & \text{colour67} \end{bmatrix}$$

Our representation of the local context is thus akin to the kind of knowledge base (KB) typically assumed as input in Natural Language Generation systems concerned with generating referring expressions (e.g. Dale and Reiter (1995)). As is common in such KBs, we assume a correspondence between values of label TYPE and nouns in the language, so that if an entity e is represented as a record with field [TYPE = *rod*] then it can be referred to by the noun corresponding to *rod*, i.e. '*rod*'. We also assume that fields (other than that labelled TYPE) encode properties (with labels representing particular dimensions) that can be expressible by adjectives. For simplicity's sake, we shall assume that in a definite description involving a vague predicate **A** the comparison class for **A** is determined by the head noun. That is, if the target referent is e and e is represented by a record with field [TYPE = T], then the comparison class $C = \{u_1, u_2, \ldots, u_n\}$ for **A** corresponds

[4]Feature values can be thought of as information coming from sensors. Note that the ability to perform precise measurements is not what is important here. What matters, as we shall see below, is the possibility to order (or appropriately *cluster*) the elements in a comparison set.

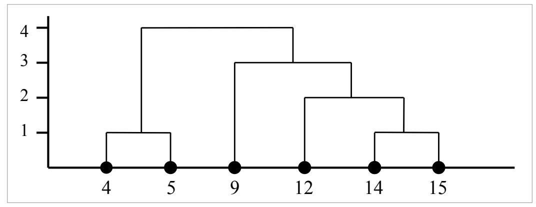

Figure 1: Dendrogram

to the set of values of the relevant dimension (e.g. LENGTH) of those entities in the local context that are represented by records with field [TYPE = T] (with T determining the head noun of the description).

Clustering. To compute where denotational boundaries can fall, and hence whether it is appropriate to place a boundary that singles out one entity (which, given constraint C3, ought to be at an end point of the ordered comparison class), we propose to use an agglomerative hierarchical clustering algorithm.[5] The algorithm produces a dendrogram with $n \in \mathbb{N}$ nested similarity levels, such that at level 0 each $u \in C$ belongs to a different cluster and at level n all $u \in C$ belong to the same cluster. Figure 1 shows a dendrogram for $C = \{4, 5, 9, 12, 14, 15\}$ with 4 similarity levels. We can now use the structure of the dendrogram to rank potential boundaries in terms of their salience. We define the degree of salience δ of a gap between u and u' in C as

(2) $$\delta(u, u' \in C) = m + 1/n$$

where m is the highest level in the dendrogram where u and u' belong to different clusters. Given this definition, the gap between 14 and 15 in Figure 1, for instance, would be assigned the score $\delta = 0.25$, while the gap between 9 and 12 would get $\delta = 0.75$, and so forth.

Since δ is defined as a ratio, it accounts for the *relative* salience of the gaps given the distribution of elements in C. The idea is that the higher the value of m with respect to the total number of levels n, the more salient the gap between u and u' in C and the less likely that $u \sim u'$ holds. We can express this kind of restrictions in terms of heuristic constraints, such as:

(3) $$u \sim u' \leftrightarrow \delta(u, u' \in C) < x$$

The form of these heuristics and the value of x may depend on several factors, such as the nature of the elements under consideration, the task at hand, or the goals of the agents. At any rate, they need to be determined empirically—a task that we have not undertaken in the current paper. Although the method needs to be empirically tuned and evaluated, we

[5]See e.g. Jain, Murty, and Flynn (1999) for an overview on data clustering. For the examples in this paper, we used the Matlab implementation of the agglomerative hierarchical clustering algorithm with euclidean pairwise distance and single-linkage distance between clusters.

see the definition of (possibly probabilistic) contraints that rely on the output of a clustering process as a promissing approach, that is significantly different from setting a fixed threshold value since the degree of salience of a particular gap entirely depends on the distribution of the elements in the contextually provided comparison class. This seems to be in accordance with the semantic elasticity exhibited by vague adjectives.

Conceptual Categories and Feature Variability

In the above model, we have used clustering to account for the relative (dis)similarity of the elements in a comparison class. This has allowed us to complement constraints C1–3 with a notion of salience according to which potential denotational boundaries can be ranked. As we shall see in this section, however, there are other factors that contribute to determining what counts as a salient gap besides the relative similarity of the objects in some perceptually available comparison class.

Crisp Judgements. The problem with the current model is more readily appreciated when we look at comparison classes that contain only two elements. In this case the notions of salience and similarity as implemented by the clustering algorithm do not add any further constraints to C1–3: as long as the two elements differ with respect to the relevant property, the prediction is that the denotational boundary of a vague adjective can fall between them (since the only available gap ought to be the most salient one), which would in turn license the use of a positive-form definite description. However, there is evidence that, in some contexts and for some types of elements, the positive form is not appropriate to make what Kennedy (2007) calls *crisp judgements*, i.e. distinctions between two entities that differ minimally with respect to the relevant property. This is not limited to definite descriptions but also holds for positive adjectives used in implicit comparison ('*compared-to*') constructions. According to Kennedy, all examples in (4) could be used to make a claim about a 100-page book as opposed to a 50-page book, but if instead of a 50-page book we have a 99-page book then only the explicit comparative forms in (4b) would be felicitous.

(4) a. The long book.
This book is long compared to that book.

b. The longer book.
This book is longer than that book.

Given our assumptions so far, accepting that $u \sim u'$ can hold for comparison classes $C = \{u, u'\}$ with only two elements violates constraint C2. This brings up a dilemma: Either we give up C2 (which is undesirable given the importance of contrast in adjective interpretation) or we allow the possibility of expanding C. Here we shall explore the latter option.

It is important to note that the crisp judgement effect does not hold across the board. For some kinds of domains, the positive form does seem to be able to make crisp judgements in the same way as the explicit comparative form does. For instance, our initial example involving two rods of unequal lengths seems to support crisp distinctions (as long as the difference in length is perceptible, the description '*the long rod*' effectively refers to the longer of the two rods). Further evidence for this comes from a set of experimental studies performed by van Deemter (2004). These experiments investigate the use of comparative, superlative and positive forms of vague adjectives in definite descriptions to discriminate between pairs of elements (such as pairs of different numbers or pairs of triangles of different size). The results obtained by van Deemter show that the positive form is significantly dispreferred when the target referent and the other element in the pair are very similar, thus confirming the crisp judgement effect. Interestingly, however, the results also show that there is variation across domains (as well as across subjects): For instance, contrary to what the crisp judgement effect would predict, when the stimuli involve triangles some subjects use the positive form in all instances regardless of the degree of similarity of the elements in each pair.[6]

Feature Variability. To account for this contrast, we propose to appeal to a distinction made in cognitive science between different patterns of feature variability within conceptual categories (see e.g. Kemp, Perfors, and Tenenbaum (2007)). The degree of homogeneity of the values of a feature allows humans—already before the age of 3—to make abstract inferences about the organization of conceptual types. For instance, we know (and can exploit this fact in forming inferences) that foods often have a common colour but not a common shape, while shape tends to be homogeneous for object types (Landau, Smith, & Jones, 1988; Booth & Waxman, 2002). Something along the same lines seems to apply to the examples we have been dealing with. For instance, while the values of the feature LENGTH for entities of type *Rod* or the feature SIZE for entities of type *Triangle* are highly heterogeneous (only knowing that something is a rod or a triangle can hardly help us to predict its length or its size, respectively), presumably people's height or books' length have a more predictable distribution of values. Our claim is that the positive form is sensitive to the degree of homogeneity of the range of values associated with the property encoded by the adjective, exhibited by the conceptual category that constitutes its domain.

Domain Representation Revisited. In order to exploit information about feature variability we need to include conceptual categories (besides individual entities) in our domain representation. To this end, we shall use Type Theory with Records (TTR) (Cooper, 2005, 2008). TTR is attractive because it offers a straightforward way of extending our representation of the local context with *record types*, which can

[6] One of the results obtained by van Deemter is that in general the superlative is the preferred form. Investigating when the superlative is preferred over the positive form in contexts that would license both is an interesting and difficult issue that falls outside the scope of this paper.

play the role of conceptual categories.[7] Example (5) shows a record type T_{rod} for the concept *rod* and an entity e_1 of that type represented as a record. Since now we have record types available, we can get rid of the field labelled TYPE in records, which we had used in (1). We add a field labelled REF for referents, which we take to be objects of the basic type *Ind(ividual)*, to both records and record types.

(5) a. $T_{rod} =_{def} \begin{bmatrix} \text{REF} & : & \textit{Ind} \\ \text{SHAPE} & : & \textit{Rod\_shape}(\text{REF}) \end{bmatrix}$

b. $e_1 =_{def} \begin{bmatrix} \text{REF} & = & \text{obj5} \\ \text{SHAPE} & = & \text{shape23} \\ \text{LENGTH} & = & \text{12cm} \\ \text{COLOUR} & = & \text{colour67} \end{bmatrix}$

Ignoring several technical details not relevant for current purposes, we say that a record e is of type T ($e : T$) iff for all fields $\langle \ell, v \rangle$ in T there is a pair $\langle \ell', v' \rangle$ in e such that $\ell = \ell'$ and $v' : v$. This means that an entity represented by a record e of type T may contain fields that are not specified by the type. This allows us to account for different patterns of feature variability. Features that are heterogeneous within a category are specified only for particular tokens (such as LENGTH and COLOUR in (5b)), while homogeneous features are part of conceptual types.

We assume that homogeneous gradable features exhibit some sort of predictable distribution. We propose to represent their value within a conceptual category in terms of the estimated parameters of the relevant probability distribution, which presumably have been inferred from exemplars of that category. For instance, if the range of values of a feature such as HEIGHT exhibited by instances of a particular conceptual category such as *Woman* is normally distributed, then we can encode this information within the record type representing that category in terms of the mean and variance of the Gaussian distribution in question: [HEIGHT : $\textit{Gaussian}(\mu, \sigma^2)$].

Refining the Clustering-based Model. We now need to refine the model so as to reflect the fact that the positive form is sensitive to information about feature variability. Sensitivity to the degree of homogeneity of a feature within a conceptual category is what will trigger the expansion of C in the appropriate contexts. Our proposal in the present paper is to use the distribution parameters of the relevant feature, whenever these are provided by the category under consideration, to generate abstract data points that complement the local comparison class. The main steps involved in the procedure are the following:

1. Let $C_L = \{u_1, u_2\}$ be the local comparison class, where u_1, u_2 are the values of the relevant feature FEAT of those entities in the local context that are of type T.
2. If T contains a field [FEAT : *Dist*] (indicating that the values of FEAT for entities of type T are to some extent homogeneous), then the parameters of the relevant probability distribution *Dist* are used to randomly generate a set of values C_R.[8]
3. Let $I \subset C_R$ be the subset of values u such that $u > u_1$ and $u < u_2$. The clustering algorithm then operates on the combined set $C = C_L \cup C_T$, where $C_T = C_R \setminus I$.[9]

This procedure allows us to expand C according to the information provided by the conceptual category determining the adjective's domain, without the need to identify C with the full set of instances of that category (e.g. the set of women or the set of books). By making the model sensitive to whether a feature is homogeneous (i.e. predictable to some degree) or heterogeneous, we are able to account for the differences between domains with respect to the crisp judgement effect. If there is no abstract information that indicates homogeneity, then the partitioning induced by the positive form proceeds solely with the elements in the local comparison class. If that information is part of the category under consideration, the positive form is sensitive to it.

The refined model offers a more general method to compute whether $\sim$ holds between two elements, even if the elements are the only members of the local comparison class. The model predicts, without the need of setting a fixed threshold, that the closer the elements are, the more likely it will be that they are related by $\sim$ and hence that the gap between them is not salient. In addition, for Gaussian distributions, the model also predicts that two elements that are both *definitely* **A** (i.e. are close to one extreme of the bell-shaped curve) will be more likely to resist crisp judgements. This seems on target given the infelicitous status of the following examples:

(6) a. [Context: two obese pigs (Kyburg & Morreau, 2000)]
$^{??}$ The fat pig is going to the fair.

b. [Context: the Sears Tower and the Hancock Building (Kennedy, 2008)]
$^{??}$ The tall one is the Sears Tower.

The aim of the current investigation has been to sketch the main components of the proposed clustering-based model. We stress once more that the model includes several parameters that need to be set empirically and that are in need of evaluation. Perhaps the most significant ones are: (i) the level of fine-graininess used to encode both the given and the randomly generated values in C; (ii) the cardinality of C_R, which presumably should be some multiple of $|C_L|$; (iii) the relationship between the degree of salience δ of a gap and $|C_R|$; and (iv) the relationship between the similarity relation $\sim$ and δ.

[7] Further details about the formal properties of TTR can be found in the above references. See also Cooper and Larsson (2009) for an approach where record types are also used to model ontological/conceptual categories.

[8] We have tested the algorithm using the Matlab function `randn` with parametrs $\mu = 166.5$ and $\sigma = 6.5$ corresponding to the distribution of height for American women aged 18–24.

[9] The substraction step is needed to preserve the actual gap between u_1 and u_2.

Conclusions

In this paper we have investigated the constraints that license the use of positive-form vague adjectives in definite descriptions. We have explored the use of clustering techniques to model the notions of salience and similarity that are required to account for the semantic elasticity of vague adjectives in these constructions. We have claimed that the positive form is sensitive to the degree of variability of the gradable property expressed by the adjective, and that this contributes to explain some differences related to the *crisp judgement* effect. The model that has emerged exploits the relative similarity of the elements in a comparison class and accounts for the influence of feature variability within conceptual categories.

Although in the current study we have focused on definite descriptions, we believe that the proposed approach may also be useful to investigate vague adjectives in predicative and indefinite contexts. Also, here we have been exclusively concerned with adjectives associated with uni-dimensional gradable properties such as LENGTH, HEIGHT or SIZE, for which the use of clustering may seem an unnecessary complication. For uni-dimensional spaces the algorithm can indeed be very simple. What matters, however, is that clustering allows us to make explicit and more precise the notions of relative similarity and salience that seem to govern the behaviour of positive-form vague adjectives. In addition, the use of clustering techniques is also attractive because it sets the stage for treating multi-dimensional gradable adjectives, such as '*safe*' or '*clever*', as well as other vague expressions such as nouns.

References

Barker, C. (2002). The dynamics of vagueness. *Linguistics & Philosophy*, *25*(1), 1–36.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, *84*(1), B11–B22.

Cooper, R. (2005). Austinian Truth, Attitudes and Type Theory. *Research on Language and Computation*, *3*, 333–362.

Cooper, R. (2008). Type Theory with Records and Unification Grammar. In F. Hamm & S. Kesper (Eds.), *Logics for linguistics structures.* Mouton de Gruyter.

Cooper, R., & Larsson, S. (2009). Compositional and ontological semantics in learning from corrective feedback and explicit definition. In *Proc. of DiaHolmia* (pp. 59–66).

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *18*, 233–266.

Ebeling, K. S., & Gelman, S. A. (1994). Children's use of context in interpreting "big" and "little". *Child Development*, *65*(4), 1178–1192.

Funakoshi, K., Watanabe, S., Kuriyama, N., & Tokunaga, T. (2004). Generating referring expressions using perceptual groups. In *Proc. of the 3rd Int'l Conf. on NLG* (pp. 51–60).

Gatt, A. (2006). Structuring knowledge for reference generation: A clustering algorithm. In *Proc. of EACL-06* (p. 321-328).

Graff, D. F. (2000). Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, *20*, 45–81.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264-323.

Kamp, H. (1975). Two theories of adjectives. In E. Keenan (Ed.), *Formal semantics of natural language* (pp. 123–155). Cambridge University Press.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental Science*, *10*(3), 307–321.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*(1), 1–45.

Kennedy, C. (2008). *Vagueness and comparison.* (Invited talk at the International Conference on Vagueness, Paris; slides available from http://semantics.uchicago.edu/kennedy/docs/paris-vagueness.pdf)

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics & Philosophy*, *4*, 1–45.

Kyburg, A., & Morreau, M. (2000). Fitting words: Vague language in context. *Linguistics & Philosophy*, *23*, 577-597.

Landau, B., Smith, L., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299-321.

McNally, L. (2007). *Classification strategies and the semantics of gradable adjectives.* (Invited talk at the 16th Amsterdam Colloquium, slides available from http://mutis.upf.es/˜mcnally/amsterdam-handout.pdf)

Pinkal, M. (1979). Semantics from different points of view. In R. Bäurle, U. Egli, & A. von Stechow (Eds.), *How to refer with vague descriptions* (pp. 32–50). Springer-Verlag.

Sedivy, J., Tenenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109-147.

Syrett, K., Bradley, E., Kennedy, C., & Lidz, J. (2005). Shifting standards: Children's understanding of gradable adjectives. In *Proc. of GALANA* (Vol. 2, pp. 353–364).

Thórisson, K. R. (1994). Simulated perceptual grouping: An application to human-computer interaction. In *Proc. of the 16th Annual Conf. of the Cognitive Science Society* (pp. 876–881).

van Deemter, K. (2004). Finetuning NLG through experiments with human subjects: the case of vague descriptions. In *Proc. of the 3rd Int'l Conf. on NLG* (pp. 31–40).

van Deemter, K. (2006). Generating referring expressions that involve gradable properties. *Computational Lingustics*, *32*(2), 195–222.

van Rooij, R. (2008). Revealed preference and satisficing behaviour. In *Proc. of the 8th Conf. on Logic and the Foundations of Game and Decision Theory (LOFT).*

Williamson, T. (1994). *Vagueness*. London: Routledge.