Context and Vagueness in REG

Vivien Mast (viv@tzi.de)
I5-[DiaSpace], SFB/TR8 Spatial Cognition, University of Bremen

Diedrich Wolter (dwolter@informatik.uni-bremen.de)
SFB/TR8 Spatial Cognition, University of Bremen

Abstract

In this paper we present a probabilistic mechanism for Referring Expression Generation (REG) that is capable of generating complex REs including combinations of vague or graded properties. We show that our mechanism covers combinations of graded properties, and propose a way to integrate feature models based on *conceptual spaces* into the system as an extension of (Mast & Wolter, 2013).

Keywords: Referring Expression Generation; Vagueness; Graded Properties; Conceptual Spaces; Color

Introduction

Referring expression generation (REG) is an essential component in language-based systems for human-computer interaction. Satisfactory results for REG are already obtained by today's approaches in scenarios that exhibit a limited number of clear-cut features. Typical examples include small sets of simple objects which are easily differentiated by their bright color or shape. A typical requirement for obtaining appropriate expressions is that object features can be grouped to crisp categorial concepts. Popular algorithms such as IA (incremental algorithm) and greedy heuristic algorithm as well as most of their extensions are based on this fundamental assumption (Krahmer & Deemter, 2012). This implies however that all occurrences of a given attribute value would be equally acceptable in a referring expression, for example every "red ball" could equally well fit this description.

In realistic scenarios we are confronted with manifold shades of red which more or less resemble one's intuition of the color red. The categorization into attributes is not obvious, as object features gradually change over from one attribute value into another. Adequacy of attributes should thus be considered as vague and gradual in REG.

A second feature of real scenarios is the role of context. If we consider attributes like "large" or "small", their interpretation is clearly dependent on context: a large hamster is likely to be smaller than a small cat and, considering a pack of small cats, even a medium-sized cat may be called large. As a consequence, in addition to the notion of graded category membership one must have a way of dealing with contextual influences on categorization.

In our work we aim at developing a computational framework for generating adequate referring expressions in human-computer interaction in realistic scenarios, i.e., we develop a mechanism for context-sensitive REG based on graded and vague category membership.

An interesting starting point for dealing with these effects has been been proposed by van Deemter (2006). In his treatment of graded properties like "tall" and "short", van Deemter relies on the perceptual context for categorization while the features are modelled as numerical values. For example, the size of mice is represented in centimetres. Based on the other animals present in the scene, the same animal could be conceived as "large mouse" or "small mouse". The system can generate an expression like "the n largest mice" for some number n, where every mouse in the target set is larger than every mouse in the distractor set. This approach requires bipolar graded attributes with two opposing values on a single numerical scale such as "tall" and "short". More complex properties such as color that require a multi-dimensional model cannot be dealt with by this approach. Moreover, as van Deemter notes, the approach does not cover combinations of vague adjectives such as "the tall fat giraffe" that are not *pareto-optimal*, i.e., where the giraffe in question is not the tallest as well as the fattest giraffe, but still the only one which is tall and fat (van Deemter, 2006).

The contribution of this work is to overcome these limitations by integrating the theory of *conceptual spaces* (Gärdenfors, 2004b, 2004a) into a probabilistic mechanism for REG. Conceptual spaces offer a cognitively motivated approach for modeling features where properties are represented as points in multi-dimensional spaces. Each dimension is represented by distinct numerical values and the similarity between two points can be calculated based on the weighted Euclidean distance. We present a probabilistic model for REG that can accommodate for the theory of conceptual spaces.

The remainder of this paper is organized as follows. First, we will discuss some fundamental terms and their relation to our work, namely *gradedness*, *vagueness*, *probability* and *context* before proceeding to introduce our probabilistic mechanism for REG. Then we will discuss in more detail some suggestions for the formalization of *conceptual spaces* and their implications for the integration of conceptual spaces into our REG mechanism. Finally, we will suggest a feature model based on similarity in conceptual spaces and discuss how the parameters for this model could determined using machine learning.

Gradedness, Vagueness, and Probability

In the literature on REG, the terms *gradedness* and *vagueness* are sometimes used interchangeably (van Deemter, 2006). However, we first briefly distinguish gradedness and vagueness before introducing a probabilistic semantics for REG that can accomodate both concepts.

In fuzzy set theory (Zadeh, 1965), each entity is a member of a given set to some degree, indicated by a membership value between 0 and 1, for example a given hue can be red to a certain degree where a bright red would have a high membership degree for the category "red", and one with a yellow tinge would have a lower membership degree for "red". This kind of *gradedness* is also a case of *vagueness*.

By contrast, Rosch's (1973) *Prototype Theory* assumes that even for two clear members of a category, one can be a *better* representative of the category than the other. For example, a bright red and a slightly more yellowish red may both be definitive members of the category "red", but the bright red may be more prototypical, i.e. more representative of the category as a whole. This type of *gradedness* is clearly not a case of *vagueness*, as it is not concerned with the boundaries of a category.

Apart from the *graded membership* interpretation of the term *vagueness*, where categories have fuzzy and overlapping boundaries, the term *vagueness* is also used for categories with clear-cut, but unknown boundaries (Lawry & Tang, 2009).

Based on this discussion, any probabilistic approach to category membership needs to clarify what the posed probabilities mean. In this paper we take an all-encompassing approach to probabilistic semantics in REG which includes vagueness and gradedness in all senses discussed above. Our central notion is $P(D|x)$, the conditional probability of a description D given an object x. We define this as the probability that a human would accept D as a good description of x with respect to the properties in question, as it could be evaluated by user studies. Thus we are involved with, for example, the probability that a person would accept "red" to be a good description for the color of a given ball. As we are only concerned with humans and their judgement, we do not take any particular position on the question whether crisp category boundaries *de facto* exist or not, or whether the differences in acceptability are due to vagueness or gradedness.

Our probability semantics must be distinguished from Frank and Goodman's (2012) probabilistic approach to REG. In their account, $P(w|x,C)$ is the probability that a speaker would *actually utter* a particular word w to refer to the object x, given the context C (Frank & Goodman, 2012). For example, that the person would say "the red one" in order to identify a particular ball. This probability is modeled by how much the label reduces uncertainty about the referent in the given context. As it relies entirely on crisp properties, it is not directly comparable to our approach.

Context

A further factor which influences acceptability or $P(D|x)$ is context. Clearly the acceptability of graded properties such as "large" or "small" depends on situational context—termed *local context* by van Deemter (2006): considering a pack of small cats, even a medium-size cat may be called large.

Further, there is also the categorial context—termed *global context* by (van Deemter, 2006). Often, for graded adjectives such as "tall" and "short", there are boundaries dependent on object category that limit acceptability (van Deemter, 2006). For example, "[if] Hans's and Fritz's heights are 210 and 205 cm, respectively, then it seems questionable to describe Fritz as the short man, even if Hans is the only other man in the local context" (van Deemter, 2006). With respect to elephants, on the other hand, 205 cm would without doubt warrant the use of the adjective "short".

According to Gärdenfors (2004b), such context effects are achieved by imposing a constraining region onto the conceptual space of a property. This region captures all the possible positions of the categorial or situational context in the feature space, and thereby influences the acceptability regions of the different terms. For example, though human skin is generally in the color range of brown–beige, humans often use color terms like "red", "yellow", or even "green" and "blue" to describe the color of a human face. The reason for this may be that the region of possible skin colors is superimposed onto color space, thereby defining the skin color closest to red as "red skin", the skin color closest to green as "green skin", etc. (Gärdenfors, 2004b).

Beyond category membership, in REG situational context also has an influence on feature selection. In situated interaction, reference is always context-dependent, as the goal of REG is to enable the listener to distinguish the target object from the distractor objects present in the scene (Olson, 1970; Hermann & Deutsch, 1976).

Classical REG approaches focus on *distinguishing descriptions*—selecting those features of an object that enable *unique identification* of the target in the given situation (Krahmer & Deemter, 2012). Here, unique identification is an absolute concept which holds if only the target object has all selected properties, but none of the distractors. Therefore, situational context may lead to a shift in property attribution: A ball which may be called "deep red" in the presence of other red balls may be called "red" if no other red balls are around.

With respect to graded properties, the concept of unique identification is not quite clear. Hermann and Deutsch (1976) show that if several properties of an object enable identification, humans choose the feature that exhibits the largest contrast between object and distractor set. For example, if trying to distinguish two objects that differ a bit in size and very much with respect to brightness, one would say "the bright one" rather than "the large one", as brightness offers a better contrast with context. But in such a case, one cannot claim that the distractors do not exhibit the given property, i.e., are not "large", as they are always large to a certain degree.

Instead of the concept of *unique identification*, we propose the concept of *discriminatory power* which is modeled in probabilistic terms by $P(x|D)$ – the probability that a human would select object x, given description D.

A Probabilistic Mechanism for REG

In this section, we present the probabilistic mechanism for computing the most appropriate referring expression with vague properties for arbitrarily complex descriptions. But what is the most appropriate description D for an object x? Clearly, one should consider the conditional probability $P(D|x)$ that D would be a good description for x as we aim to produce natural expressions. Since we also aim for effectiveness in communication we also consider conditional probability $P(x|D)$ that a human would actually select object x given description D. This leads to a two-objective optimization task of determining D that maximizes $P(D|x)$ as well as $P(x|D)$. To resolve potential optimization conflicts we introduce a parameter $\alpha \in [0,1]$ to balance acceptability $P(D|x)$ vs. discriminatory power $P(x|D)$, yielding a well-defined optimal description:

$$D_x^* := \arg\max_D (1-\alpha)P(x|D) + \alpha P(D|x) \tag{1}$$

In previous works (Mast & Wolter, 2013) we have identified that choices for α between 0.1 and 0.4 are suitable. By the Theorem of Bayes, both factors of the model are related to one another. It is thus sufficient to define a computation model for either one of them. As indicated earlier, our model is based on $P(D|x)$. Applying the Theorem of Bayes $P(x|D) = \frac{P(D|x)P(x)}{P(D)}$ involves two additional factors, the probability $P(x)$ of selecting object x and the prior probability of the description, $P(D)$. Currently, we set $P(x) = \frac{1}{N}$ where N is the total number of objects in the scene, stating equal probability for selecting any one of them. In future work, the factor might be used to capture effects of object salience. In order to determine $P(D)$ we evaluate the acceptability of D to all objects in the scene:

$$P(D) = \frac{\sum_{i=1}^{N} P(D|x_i)}{N} \tag{2}$$

Here, $x_1, \ldots, x_N$ are all objects in the scene. This equation gives us a context-sensitive distinction of the *discriminatory power* of a combination of vague attributes. Roughly speaking, $P(D)$ in Bayes Theorem dampens values for those descriptions that match many objects well. This leaves us with $P(D|x)$ as the main factor of the model.

Complex Descriptions

Our aim is to compute descriptions of arbitrary complexity, responding to the requirements of the description task at hand. We say that a complex description (or description for short) is a set of pairs of feature dimensions (color, size, location, etc.) and their respective attribute values, like for example $\{(\text{color}, \text{red}), (\text{size}, \text{small})\}$. These sets represent conjuncts in the sense that the aforementioned set refers to an object whose color is red *and* whose size is small. Sets only involving a single pair are called basic descriptions. Suppose we know the values $P(d_i|x)$ for all basic descriptions d_i involved in a description D. Then, by the laws of probability we can compute the value of $P(D|x)$—in cases where the feature attributes are independent of one another this is simply accomplished by multiplication. This stochastic independency should, according to Lawry and Tang (2009), be assumed for feature dimensions that belong to different conceptual spaces. For example, the probability of one accepting that a door is red (color feature) is assumed to be independent of the probability of accepting that the door is large (size feature). Stochastic independency of feature domains allows complex descriptions to be separated by feature domains and enables modularity. Our model is modular in the sense that it allows for different ways of modeling $P(d_i|x)$, as long as they provide, for each potential feature category, an acceptability value $P(d_i|x)$ with $0 \leq P(d_i|x) \leq 1$, indicating how well the property fits the given category. For crisp attribute categories, values 0 (not applicable) and 1 (applicable) are sufficient. It is of course desirable to use a consistent formalism for modeling feature properties as far as this is possible, in order to prevent impreciseness that might arise through inconsistent modeling, as will be discussed below.

We note that the stochastic model also handles spatial descriptions that relate one object described to another, for example “the red door to the left of the small window”; refer to (Mast & Wolter, 2013) for details.

Integration of Conceptual Spaces

The crucial question for integrating conceptual spaces with our probabilistic model is how conceptual spaces can be formalized in a way that provides a mapping to acceptance values $P(f|x)$ for each feature f, given an object x. If it is to be useful for our mechanism and for application in realistic scenarios, such a formalization needs to comply with a number of criteria. Firstly, attribution needs to be exhaustive: for every object/description pair that could be acceptable to a human in some context, the acceptability value should be larger than zero. Secondly, the empirical ordering of acceptability should be maintained. If human subjects grade one entity with a higher acceptability rating than another for a given description, the system should do so too—both within a domain, and between different domains such as color vs. size. Thirdly, the formalism should be able to deal with hierarchical concepts so that the same space may be occupied by different concepts of different levels of generality. Finally, the approach must be suitable for practical application, i.e. it should be feasible to obtain the appropriate model parameters.

Acceptability Based on Discretization

In the *conceptual spaces* approach, properties are represented by a mixed multi-dimensional parameter space. For example, color can be modeled as a combination of hue, chromaticity (or saturation) and brightness. The hue subspace is polar, i.e., it can be represented by the angle in a circle, while chromaticity and brightness are linear. Due to their interrelatedness, the latter two form a triangle, yielding a spindle as the conceptual space of color (Gärdenfors, 2004b). In these spaces, each dimension can be represented by a numerical value. The distance between two points can be calculated as the weighted

Figure 1: Dogs of different height and weight with descriptions generated by the system

Euclidean distance.

Based on research in cognitive science, Gärdenfors (2004b) suggests the following model of categorization: Categories are learnt from examples in a process of supervised learning where an abstract category prototype is formed by calculating the mean values of all category members for each dimension. The property space is then divided by *Voronoi tesselation* based on the prototypes. This method divides the multi-dimensional space into category regions in a way that each individual point is assigned to the category region of the prototype which is closest to it according to the weighted Euclidean distance (Gärdenfors, 2004b). An example in two-dimensional space is shown in Figure 2(a). For example, a given hue value would be classified as either yellow or red, depending on which color prototype it is closest to in the color space, given the weights of the dimensions.

A simple way to achieve this for an REG system would be to use available empirical information, e.g., about color names (Berlin & Kay, 1969; Sivik & Taft, 1994; Taft & Sivik, 1997; Falomir, Museros, Gonzalez-Abril, & Sanz, 2013) and identify prototypes (or prototype regions) either directly from the data where they were elicited (Berlin & Kay, 1969; Sivik & Taft, 1994) or infer them from examples by calculating the mean. In order to obtain an acceptance value, one could use Voronoi tesselation to form categories, and then transform the distance Δ_x of x from the prototype (region) into a value between 0 and 1 by relating it to the distance Δ_{max} of the most peripheral category member: $P(D|x) = 1 - \frac{\Delta_x}{\Delta_{max}}$.

Example We implemented this simple method using a scenario similar to van Deemter's (2006) "tall fat giraffe" scenario. We applied our system to a world of dogs of different height and corpulence (Figure 1), modeling the features *height* and *corpulence* as one-dimensional *conceptual spaces* and integrating them into our probabilistic framework.

Table 1 shows the exemplary creation of categorial attribute values with graded acceptability scores for height and corpulence (based on the weight/height ratio). As discussed above, one could consider either situational or categorial context for attributing labels to individuals. In this case, we use situational context only in order to better compare our approach to that of van Deemter (2006). For polar properties such as "tall" and "short", the prototype for each extreme is in principle positive infinity for "tall" and $1/$infinity for "short". Since the context imposes a limited subspace onto the conceptual space (Gärdenfors, 2004b), we consider the tallest and shortest dog present in the scene to be the prototypes for each category.

All dogs are then classified according to the principle of simple Voronoi tesselation, as suggested by Gärdenfors (2004b). Any dog closer to the "tall" prototype than to the "short" prototype is considered "tall" and vice versa. Acceptability ratings are gained by their relative closeness to the prototype such that the dog closest to the prototype has an acceptability score of 1 while one on the boundary will have a score of 0. Given this scene, our system generates the responses shown in Figure 1. Modeling "height" and "corpulence" based on normalization around the mean yields similar results (Mast & Wolter, 2013).

Discussion As figure 1 shows, our methodology enables combinations of vague attributes that are not *pareto-optimal*, thus improving upon van Deemter's (2006) approach. If a dog is both tall and fat (3), and there is another dog which may be slightly taller, but less fat (4), the system will call dog 3 "the tall fat dog" and dog 4 "the tall dog". If dog 4 is made taller (65 cm instead of 60), the system calls dog 3 "the fat dog" only, avoiding confusion with dog 4.

At the same time, our system goes beyond the *Nash arbitration plan* by considering the other dogs in the context. If there are other very fat but shorter dogs in the scene, the attribute "tall" is used for dog 3 even if dog 4 is considerably taller. For example if dog 4 is again 65 cm tall, and the weight of dog 5 is increased to 30 kg, dog 3 is called "the tall fat dog", as the attribute "tall" is needed for distinguishing it from the equally fat, but shorter dog 5.

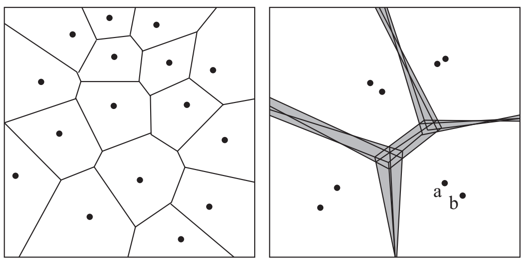
However, with respect to the criteria posited in the beginning of this section, there remain some problems. The empirical ordering is maintained within domains—shorter dogs consistently have higher acceptability ratings for "short" than taller ones and vice versa. Whether this holds between domains is an empirical question. On the other hand, the criterion of exhaustiveness is not met. This criterion states that each object/description pair that could be acceptable to a human in some context should have an acceptability value larger than zero. However, the Voronoi tesselation leads to a discretization of space where each object belongs to only one category of the given space. Namely, a dog in this situation can be either "tall" or "short" but never both. In the example given above, dog 1 is classified as "short" by the system and therefore has an acceptability value of 0 for "tall". The classification as "tall" may however be acceptable to a human in this context. Particularly, the expression "the tall skinny dog" would be useful in order to distinguish dog 1 from dog 2 which is also skinny but shorter.

Acceptability Based on Indeterminate Boundaries

In line with this problem, Douven, Decock, Dietz, and Égré (2013) criticize that Gaerdenfors' (2004b) approach of Voronoi classification has only very limited space for vagueness, namely the boundaries of the Voronoi regions. In this sense, the normalized proximity function suggested above merely represents gradedness of category membership in

	physical features			description features			
id	**height** h [cm]	**weight** w [kg]	w/h	**height**	$P(D\|x)$	**corpulence**	$P(D\|x)$
1	46	27	0.59	SHORT	0.08	SKINNY	0.87
2	35	20	0.57	SHORT	0.92	SKINNY	1.00
3	55	45	0.82	TALL	0.62	FAT	1.00
4	60	45	0.75	TALL	1.00	FAT	0.45
5	34	26	0.76	SHORT	1.00	FAT	0.57

Table 1: Exemplary operationalization of dog sizes—height and corpulence



(a) crisp categories

(b) superimposed Voronoi graphs and overlapping boundary regions (grey)

Figure 2: Voronoi graphs for characterizing category regions (Douven et al., 2013)

the sense of better vs. peripheral examples. It does not cover vagueness in the sense of unclear category boundaries. Douven et al. (2013) extend Gärdenfors' (2004b) Voronoi classification by using collated Voronoi tesselation based on prototype regions. In this case, illustrated in Fig. 2(b), a set of Voronoi tesselations is produced such that a Voronoi tesselation is performed for each combination of representatives for each region. When all Voronoi graphs are superimposed on each other, the region which is part of a given category for all graphs is considered the set of clear instances of this category while overlapping regions are considered to be boundary cases which do not clearly belong to the category. This approach is closely related to the so-called "egg-yolk" representation of spatial reasoning with indeterminate boundaries (Cohn & Gotts, 1996).

However, this approach has some counter-intuitive effects in REG: Increasing the prototypical area of a category not only increases vagueness outwards, taking in more space for potential category membership, it also increases vagueness inwards, thereby decreasing the area of clear membership.

More importantly, it remains unclear whether it yields any additional value with respect to the integration of conceptual spaces into our model. Our main goal is to assign acceptability values between 0 and 1 that integrate gradedness and vagueness, while Douven et al. (2013) solely focuses on determination of areas of vagueness. Computing acceptability values based on the areas of clear category membership and vague boundary regions could only be ad hoc without backing of a consistent theory. Particularly, for graded adjectives which are mainly context-dependent, such as the ones used in the example given above, it is not even clear how the size of the prototypical area should be determined.

Finally, any approach based on *discretization* of the conceptual space provides problems with terms that are not basic-level categories (Rosch, 1973), creating problems for our desired criterion of dealing with hierarchical organization of concepts. One may suggest applying hierarchical Voronoi tesselation for each level, subdividing the conceptual subspace of, say, red into its subconcepts, for example "crimson" and "pale red", but this is problematic as there are many subconcepts of red which may overlap strongly and which, taken together, may not cover the whole space covered by their hypernym "red".

Acceptability Based on Similarity

Therefore, rather than applying Gärdenfors' idea of learning subspaces, we suggest a methodology of learning acceptability values based directly on the concept of similarity, without the intermediary step of categorization.

As our probabilistic REG mechanism is capable of selecting one of several categories of the same conceptual space, i.e., "red" vs. "yellow" based on acceptability values and situational context, it is not necessary that the property model provide acceptability values of zero even for those elements which are clearly not members of the respective category.

For example, whether a given color should be called "red" or "yellow" or "crimson" will be determined by comparing the acceptability value of each of these properties for the target object with the average acceptability value of the same properties for other objects in the scene. In this case, the acceptability value of a given ball for "yellow" does not necessarily have to be zero even if the ball in question would not be normally classified as yellow, as long as the value is sufficiently low that it will be strongly dispreferred as compared to the other options available. Therefore, we may use *similarity* as a basis for acceptability values. Gärdenfors (2004b, 20) summarizes psychological literature, according to which the similarity s_{ij} between two objects i and j is an *exponentially decaying function* of their distance $d(i,j)$ and can be stated as

$$s(i,j) := e^{-c \cdot d(i,j)} \quad (3)$$

When we apply this similarity measure to categorization of j based on a prototype i, c determines the specificity of the category. For example, for "crimson" c will be larger than for "red", leading to a faster decline of similarity, as the category covers a smaller range of hues. This model obviously fulfills our first requirement of exhaustiveness and it can easily accomodate for hierarchical concepts by adapting parameter c individually per concept. Hierarchical organization is solved in a much more flexible way than allowed for by Voronoi tesselation. Whether the requirement of retaining empirical ordering of acceptability can be met requires empirical research. Particularly, it remains to be shown that the ordering requirement can be met for the relation of separate conceptual spaces, e.g., size vs. color.

An implementation of the dog world example discussed above using the concept of similarity to obtain acceptability values improves over the categorization-based acceptability-values in that it describes dog 1 as "the tall skinny dog" rather than just "the skinny dog" thus better differentiating it from dog 2. The descriptions for all other dogs remain the same as in the prior example. For the example, appropriate values for c were selected ad hoc, with $c_{\text{tall,short}} = 0.01$ and $c_{\text{skinny,fat}} = 100$.

The question remains of how appropriate model parameters can be determined for our model. To this end, machine learning techniques can be applied. Empirical data of acceptability ratings by human subjects such as the data collected by Sivik and Taft (1994) can be used to fit the weights for each concept parameter. For example, a user preference of "red" over "crimson" for a color in the continuum between prototypes for red and crimson that does not agree with the current model parameters leads to an update of model parameters, increasing acceptability of "red" by increasing the width of the acceptability function (lower value of c_{red}) and decreasing the width of crimson (higher value of c_{crimson}).

With respect to practicability, assigning non-zero values to all properties puts a high demand of efficiency on the REG algorithm. In this case, exhaustive search covering all properties of an object and their combinations is not an option. However, one way of dealing with this is using a variant of the greedy algorithm for REG proposed by Dale (1992) though some adaptation will be necessary to encompass our fine-grained usage of discriminatory power and the inclusion of relations.

Conclusion

In summary, we present a probabilistic approach to generating referring expressions with vague properties. Our approach goes beyond existing work in that it enables context-dependent generation covering non-pareto-optimal feature combinations and it allows for the integration of conceptually adequate property models based on the theory of conceptual spaces. We discuss the benefits and weaknesses of including categorization with Voronoi tesselation and suggest an alternate approach using a psychologically founded similarity function parametrized with machine learning.

References

Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. Berkeley: University of California Press.

Cohn, A. G., & Gotts, N. M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In P. Burrough & A. M. Frank (Eds.), *Proceedings, GISDATA specialist meeting on geographical objects with undetermined boundaries* (pp. 171–187). Taylor & Francis.

Dale, R. (1992). *Generating referring expressions: Constructing descriptions in a domain of objects and processes*. Cambridge, Mass.: MIT Press.

Douven, I., Decock, L., Dietz, R., & Égré, P. (2013). Vagueness: A conceptual spaces approach. *Journal of Philosophical Logic*, *42*(1), 137–160.

Falomir, Z., Museros, L., Gonzalez-Abril, L., & Sanz, I. (2013). A model for qualitative colour comparison using interval distances. *Displays*. (to appear)

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998.

Gärdenfors, P. (2004a). Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, *2*(2), 9–27.

Gärdenfors, P. (2004b). *Conceptual spaces: The geometry of thought*. Cambridge, Mass.: MIT Press.

Hermann, T., & Deutsch, W. (1976). *Psychologie der Objektbenennung*. Bern: Huber.

Krahmer, E., & Deemter, K. van. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173–218.

Lawry, J., & Tang, Y. (2009). Uncertainty modelling for vague concepts: A prototype theory approach. *Artificial Intelligence*, *173*(18), 1539–1558.

Mast, V., & Wolter, D. (2013). A probabilistic framework for object descriptions in indoor route instructions. In *Spatial information theory.* Berlin, Heidelberg: Springer. (to appear)

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*(4), 257–273.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Sivik, L., & Taft, C. (1994). Color naming: A mapping in the imcs of common color terms. *Scandinavian Journal of Psychology*, *35*(2), 144–164.

Taft, C., & Sivik, L. (1997). Salient color terms in four languages. *Scandinavian Journal of Psychology*, *38*(1), 29–34.

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

Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, *8*(3), 338–353.