A Label Semantics Approach to Linguistic Hedges

Martha Lewis and Jonathan Lawry
Department of Engineering Mathematics,
University of Bristol,
BS8 1TR, United Kingdom
martha.lewis@bristol.ac.uk, j.lawry@bristol.ac.uk

28th January 2014

Abstract

We introduce a model for the linguistic hedges 'very' and 'quite' within the label semantics framework, and combined with the prototype and conceptual spaces theories of concepts. The proposed model emerges naturally from the representational framework we use and as such, has a clear semantic grounding. We give generalisations of these hedge models and show that they can be composed with themselves and with other functions, going on to examine their behaviour in the limit of composition.

1 Introduction

The modelling of natural language relies on the idea that languages are compositional, i.e. that the meaning of a sentence is a function of the meanings of the words in the sentence, as proposed by [13]. Whether or not this principle tells the whole story, it is certainly important as we undoubtedly manage to create and understand novel combinations of words. Fuzzy set theory has long been considered a useful framework for the modelling of natural language expressions, as it provides a functional calculus for concept combination [30, 32].

A simple example of compositionality is hedged concepts. Hedges are words such as 'very', 'quite', 'more or less', 'extremely'. They are usually modelled as transforming the membership function of a base concept to either narrow or broaden the extent of application of that concept. So, given a concept 'short', the term 'very short' applies to fewer objects than 'short', and 'quite short' to more. Modelling a hedge as a transformation of a concept allows us to determine membership of an object in the hedged concept as a function of its membership in the base concept, rather than building the hedged concept from scratch [31].

Linguistic hedges have been widely applied, including in fuzzy classifiers [6, 7, 20, 22] and database queries [1, 3]. Using linguistic hedges in these applications allows increased accuracy in rules or queries whilst maintaining human interpretability of results [4, 23]. This motivates the need for a semantically grounded account of linguistic hedges: if hedged results are more interpretable then the hedges used must themselves be meaningful.

In the following we provide an account of linguistic hedges that is both functional, and semantically grounded. In its most basic formulation, the operation requires no additional parameters, although we also show that the formulae can be generalised if necessary. Our account of linguistic hedges uses the label semantics framework to model concepts [17]. This is a random set approach which quantifies an agent's subjective uncertainty about the extent of application of a concept. We refer to this uncertainty as *semantic uncertainty* [19] to emphasise that it concerns the definition of concepts and categories, in contrast to stochastic uncertainty which concerns the state of the world. In [19] the label semantics approach is combined with conceptual spaces [14] and prototype theory [25], to give a formalisation of concepts as based on a prototype and a threshold, located in a conceptual space. This approach is discussed in detail in section 2. An outline of the paper is then as follows: section 3 discusses different approaches to linguistic hedges

from the literature, and compares these with our model. Subsequently, in section 4, we give formulations of the hedges 'very' and 'quite'. These are formed by considering the dependence of the threshold of a hedged concept on the threshold of the original concept. We give a basic model and two generalisations, show that the models can be composed and investigate the behaviour in the limit of composition. Section 5 compares our results to those in the literature and proposes further lines of research.

2 Theoretical approach to concepts

2.1 Prototype theory and fuzzy set theory

Prototype theory views concepts as being defined in terms of prototypes, rather than by a set of necessary and sufficient conditions. Elements from an underlying metric space then have graded membership in a concept depending on their similarity to a prototype for the concept. There is some evidence that humans use natural categories in this way, as shown in experiments reported in [25]. Fuzzy set theory [30] was proposed as a calculus for combining and modifying concepts with graded membership, and extended these ideas in [32] to linguistic variables as variables taking words as values, rather than numbers. For example, 'height' can be viewed as a linguistic variable taking values 'short,' 'tall', 'very tall', etc. The variable relates to an underlying universe of discourse Ω, which for the concept 'tall' could be $\mathbb{R}^+$. Then each value L of the variable is associated with a fuzzy subset of Ω, and a function $\mu_L : \Omega \to [0, 1]$ associates with each $x \in \Omega$ the value of its membership in L. Prototype theory gives a semantic basis to fuzzy sets through the notion of similarity to a prototype, as described in [10]. In this context, concepts are represented by fuzzy sets and membership of an element in a concept is quantified by its similarity to the prototype. In this situation the fuzziness of the concept is seen as inherent to the concept. An alternative interpretation for fuzzy sets is random set theory, see [10] for an exposition. Here, the fuzziness of a set comes from uncertainty about a crisp set, i.e. semantic uncertainty, rather than fuzziness inherent in the world. This second approach is the stance taken by [19], and which we now adopt in this paper.

2.2 Conceptual Spaces

Conceptual spaces are proposed by Gärdenfors in [14] as a framework for representing information at the conceptual level. Gärdenfors contrasts his theory with both a symbolic, logical approach to concepts, and an associationist approach where concepts are represented as associations between different kinds of basic information elements. Rather, conceptual spaces are geometrical structures based on quality dimensions such as weight, height, hue, brightness, etc. It is assumed that conceptual spaces are metric spaces, with an associated distance measure. This might be Euclidean distance, or any other appropriate metric. The distance measure can be used to formulate a measure of similarity, as needed for prototype theory - similar objects are close together in the conceptual space, very different objects are far apart.

To develop the conceptual space framework, Gärdenfors also introduces the notion of integral and separable dimensions. Dimensions are integral if assignment of a value in one dimension implies assignment of a value in another, such as depth and breadth. Conversely, separable dimensions are those where there is no such implication, such as height and sweetness. A *domain* is then defined as a set of quality dimensions that are separable from all other dimensions, and a *conceptual space* is defined as a collection of one or more domains.

Gärdenfors goes on to define a *property* as a convex region of a domain in a conceptual space. A *concept* is defined as a set of such regions that are related via a set of salience weights. This casting of (at least) properties as convex regions of a domain sits very well with prototype theory, as Gärdenfors points out. If properties are convex regions of a space, then it is possible to say that an object is more or less central to that region. Because the region is convex, its centroid will lie within the region, and this centroid can be seen as the prototype of the property.

2.3 Label Semantics

The label semantics framework was proposed by [17] and related to prototype theory and conceptual spaces in [19]. In this framework, agents use a set of labels $LA = \{L_1, L_2, ..., L_n\}$ to describe an underlying conceptual space Ω which has a distance metric $d(x, y)$ between points. In fact, it is sufficient that $d(x, y)$ be a pseudo-distance. When x or y is a set, say Y, we take $d(x, Y) = \min\{d(x, y) : y \in Y\}$. In this case, the set Y is seen as an ontic set, i.e., a set where all elements are jointly prototypes, as opposed to an epistemic set describing a precise but unknown prototype, as described in [11]. Each label L_i is associated with firstly a set of prototype values $P_i \subseteq \Omega$, and secondly a threshold ε_i, about which the agents are uncertain. The thresholds ε_i are drawn from probability distributions δ_{ε_i}. Labels L_i are associated with neighbourhoods $\mathcal{N}_{L_i}^{\varepsilon_i} = \{x \in \Omega : d(x, P_i) \leq \varepsilon_i\}$. The neighbourhood can be seen as the extension of the concept L_i. The intuition here is that ε_i captures the idea of being sufficiently close to prototypes P_i. In other words, $x \in \Omega$ is sufficiently close to P_i to be appropriately labelled as L_i providing that $d(x, P_i) \leq \varepsilon_i$.

Given an element $x \in \Omega$, we can ask how appropriate a given label is to describe it. This is quantified by an appropriateness measure, denoted $\mu_{L_i}(x)$. We are intentionally using the same notation as for the membership function of a fuzzy set. This quantity is the probability that the distance from x to P_i, the prototype of L_i, is less than the threshold ε_i, as given by:

$$\mu_{L_i}(x) = P(\varepsilon_i : x \in \mathcal{N}_{L_i}^{\varepsilon_i}) = P(\varepsilon_i : d(x, P_i) \leq \varepsilon_i) = \int_{d(x,P_i)}^{\infty} \delta_{\varepsilon_i}(\varepsilon_i) \mathrm{d}\varepsilon_i$$

We also use the notation $\int_d^{\infty} \delta_{\varepsilon_i}(\varepsilon_i) \mathrm{d}\varepsilon_i = \Delta_i(d)$, according to which $\mu_{L_i}(x) = \Delta_i(d(x, P_i))$. The above formulation provides a link to the random set interpretation of fuzzy sets. Random sets are random variables taking sets as values. If we view $\mathcal{N}_{L_i}^{\varepsilon_i}$ as a random set from $\mathbb{R}^+$ into 2^{Ω}, then $\mu_{L_i}(x)$ is the single point coverage function of $\mathcal{N}_{L_i}^{\varepsilon_i}$, as defined in [18], and also commonly called a contour function [26].

Labels can often be semantically related to each other. For example, the label 'pet fish' is semantically related to the labels 'pet' and 'fish', and the label 'very tall' related to the label 'tall'. This prompts two questions: firstly, how the prototypes of each concept are related to each other, and secondly, how the thresholds of each concept are related. Two simple models for the relationships between the thresholds are given in [19]. The *consonant model* takes all thresholds as being dependent on one common underlying threshold. So, all thresholds have the same distance metric d and are related to a base threshold ε by the dependency that $\varepsilon_i = f_i(\varepsilon)$ for increasing functions f_i. In contrast, the *independence model* takes all thresholds as being independent of each other. This might hold when labels are taken from different conceptual spaces.

Between these two extremes, we model dependencies between thresholds as a Bayesian network - i.e., a directed acyclic graph whose edges encode conditional dependence between variables. The key property of this type of network is that the joint distribution of all variables can be broken into factors that depend only on each individual variable and its parents. So, for example, the network in figure 1 can be factorised as $\delta(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5) = \delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_2}(\varepsilon_2)\delta_{\varepsilon_3|\varepsilon_1,\varepsilon_2}(\varepsilon_3|\varepsilon_1, \varepsilon_2)\delta_{\varepsilon_4|\varepsilon_2}(\varepsilon_4|\varepsilon_2)\delta_{\varepsilon_5|\varepsilon_3}(\varepsilon_5|\varepsilon_3)$.

This enables calculation of the joint distribution and therefore marginal distributions in an efficient manner.

One intuitively easy example is where the dependency of one threshold ε_2 on another ε_1 is that $\varepsilon_2 \leq \varepsilon_1$. This could be taken to model the dependency of the threshold of the concept 'very tall' on the threshold of 'tall'. The label 'very tall' should be appropriate to describe fewer people than the label 'tall'. Therefore, the threshold for describing someone as 'very tall' will be narrower than the threshold for describing someone as 'tall', i.e. $\varepsilon_{\text{very tall}} \leq \varepsilon_{\text{tall}}$. This simple model will form part of the approach to modelling linguistic hedges, as outlined in the sequel.

3 Approaches to linguistic hedges

Linguistic hedges have been given varying treatments in the literature. In this section we summarise these different approaches and state the approach that we wish to take, discussing properties that hedge modifiers

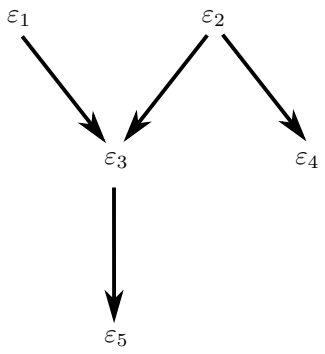

Figure 1: Example of a Bayesian network of thresholds. Dependencies between thresholds ε_i are represented by arrows.

may need. We give two specific approaches from the literature with which we will compare our results.

In [31] the idea of linguistic hedges as operators modifying fuzzy sets was introduced, so that the membership function $\mu_{hL}(x)$ of a hedged concept, hL, is a function of the membership of the base concept L, i.e. $\mu_{hL}(x) = f(\mu_L(x))$. Furthermore, truth can be considered as a linguistic variable and hence a fuzzy set [32], so that the application of a hedge can be seen as modifying the truth value of a sentence using that concept [12, 15, 32]. This second view is useful in approximate reasoning, and allows for an algebraic approach to investigating the properties of linguistic hedges, as introduced in [32], and expanded upon in [5, 12, 15]. The approach we take, however, is to view a hedge as modifying the fuzzy set associated with a concept directly, as taken by [2, 4, 9, 24]. Rather than examining the algebraic properties of hedges or their role in reasoning, we look at how hedges are semantically grounded and argue that our approach provides a particularly clear semantics.

We will propose a set of operations that may be used for both expansion and refinement of single concepts. This is in contrast to the work presented in [27] in which information coarsening is effected by taking disjunctions of labels. The idea of a hedged concept has some similarities to that of the bipolar model of concepts described in [28], since if it is appropriate to describe someone as 'very tall', it must be appropriate to describe them as 'tall', and similarly describing someone as 'quite tall' implies that it is not entirely inappropriate to describe them as 'tall'. However, we see the concepts derived by application of hedges as labels in their own right which can be used to describe data or objects.

Zadeh divides hedges into two types. A type 1 hedge can be seen as an operator acting on a single fuzzy set. Examples are 'very', 'more or less', 'quite', or 'extremely' [31]. Type 2 hedges are more complicated and include modifiers such as 'technically' or 'practically'. In [31] concepts are considered as made up of various different components, with the membership function a weighted sum of the memberships of the individual components. Type 1 hedges operate on all components equally, whereas type 2 hedges differentiate between components. For example, the hedge 'essentially' might give more weight to the most important components in a concept. Type 2 hedges are further explored in [16, 29], where components of a concept are categorised as *definitional*, *primary* or *secondary*, and the hedges 'technically', 'strictly speaking' and 'loosely speaking' are analysed in terms of these categories. Although in the following we restrict ourselves to consideration of type 1 hedges only, the treatment of concepts as having different components is mirrored by the conceptual spaces view, where each component might be seen as a dimension in the conceptual space. Further development of the framework may therefore allow a treatment of type 2 hedges.

A further distinction between types of hedge lies in the difference between powering or shifting modifiers. Powering modifiers are of the form $\mu_{hL}(x) = (\mu_L(x))^k$, where hL refers to the hedged concept and k is some real value, and shifting modifiers are of the form $\mu_{hL}(x) = (\mu_L(x - a))$. Zadeh introduces both types of modifier in his discussion of type 1 hedges [31], however his powering modifiers are most frequently cited. These are the *concentration* operator $CON(\mu_{\text{tall}}(x)) = (\mu_{\text{tall}}(x))^2$, and the *dilation* operator $DIL(\mu_{\text{tall}}(x)) =$

$(\mu_{\text{tall}}(x))^{\frac{1}{2}}$, which are often taken to implement the hedges 'very' and 'quite', (alternatively 'more or less'), respectively. [1]

The operators CON and DIL leave the core, $\{x \in \Omega : \mu_L(x) = 1\}$, and support $\{x \in \Omega : \mu_L(x) \neq 0\}$, of the fuzzy sets unchanged, which is often argued to be undesirable [3, 2, 24, 21]. In particular, [3] argue that in a fuzzy database, if a concentrating hedge is being used to refine a query that is returning too many objects, the hedge needs to reduce the number of objects returned, and hence narrow down the core. Furthermore, [22] find that classifiers using the CON and DIL operators (classical hedges) do not perform as well as those with hedges that modify the core and support of the fuzzy sets. In contrast, Zadeh himself argues that the core should not be altered. The application of a modifier 'very' to a property given by a crisp set should leave that property unchanged: 'very square' is the same as 'square'. A fuzzy set is made up of a non-fuzzy part, the core, and a fuzzy part, $\{x \in \Omega : 0 < \mu_L(x) < 1\}$. Since the core of a fuzzy set is a crisp set, it should be left unchanged. The use of classical hedges does improve performance over non-hedged fuzzy rules in expert systems [6, 7, 20], so the argument against classical hedges is a matter of degree.

The use of the CON and DIL operators to model the hedges 'very' and 'quite' is further criticised on the basis that the modifiers are arbitrary and semantically ungrounded. No justification is given for these modifiers other than that they have what seem to be intuitively the right properties [2, 8, 24]. Grounding hedges semantically is important for a theoretical account of what happens when we use terms like 'very' and also for retaining interpretability in fuzzy systems. [2, 8] both ground modifiers using a resemblance relation which takes into account how objects in the universe are similar to each other. [24] takes a horizon shifting approach.

In [24] the class of finite numbers is used as an example of the horizon shifting approach. Some numbers are certainly finite, however as numbers get larger, finiteness becomes impossible to verify. Mapping this idea onto the concept 'small', we can say that there is a class of numbers that are definitely small, say $[0, c]$. As numbers get larger than c we approach the horizon past which the concept 'small' no longer applies, expressed as $1 - \epsilon(x)(x - c)$. So:

$$\mu_{\text{small}}(x) = \begin{cases} 1 & \text{if } x \in [0, c] \\ 1 - \epsilon(x)(x - c) & \text{if } x \geq c \end{cases}$$

Now, to implement the hedge 'very', the horizon c is shifted by a factor σ and the membership function altered thus:

$$\mu_{\text{very small}}(x) = \begin{cases} 1 & \text{if } x \in [0, \sigma c] \\ 1 - \epsilon(x, \sigma)(x - \sigma c) & \text{if } x \geq \sigma c \end{cases}$$

In [24], examples of different kinds of membership functions that might be used to implement this idea are given. A linear membership function gives $\epsilon(x) = \frac{1}{a-c}$ where a is the upper limit of the membership function. To implement the hedge, the function $\epsilon(x, \sigma) = \frac{1}{\sigma(a-c)}$ is introduced, giving

$$\mu_{\text{small}}(x) = \begin{cases} 1 & \text{if } x \in [0, c] \\ 1 - \frac{x-c}{a-c} & \text{if } x \in [c, a] \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mu_{\text{very small}}(x) = \begin{cases} 1 & \text{if } x \in [0, \sigma c] \\ 1 - \frac{x-\sigma c}{\sigma(a-c)} & \text{if } x \in [\sigma c, \sigma a] \\ 0 & \text{otherwise} \end{cases}$$

[2, 8] both ground their approaches in the idea of looking at the elements near a fuzzy set in order to contract or dilate the set. The two approaches are similar, so we restrict ourselves to that of [2]. This

[1] Zadeh in fact proposes a rather more complicated hedge for 'more or less', which involves a combination of powering and shifting, however, the dilation operator is more frequently quoted in the literature.

approach introduces a fuzzy resemblance relation on the universe of discourse, and either a T-norm in the case of dilation, or a fuzzy implicator for concentration. The modifier is then implemented as follows. Consider a fuzzy set F and a proximity relation E^Z which is approximate equality, parametrised by a fuzzy set Z. As described in [2], E is modelled by $(u, v) \rightarrow E(u, v) = Z(u - v)$, where Z is a fuzzy interval centred on 0 with finite support. In terms of a trapezoidal membership function, Z can be expressed as $(-z - a, -z, z, z + a)$. Therefore, if $|u - v| \leq z$, u and v are judged to be approximately equal, i.e. $E^Z(u, v) = 1$. The set F is dilated by $E^Z(F)(s) = \sup_{r \in \Omega} T(F(r), E^Z(s, r))$, where T is any T-norm, *min* being the standard.

To understand the effect that this has on a fuzzy set F, suppose that F has a trapezoidal membership function (A, B, C, D) where $[B, C]$ is the core of F and $[A, B]$, $[C, D]$ the support, and that Z similarly is $(-z - a, -z, z, z + a)$, with the T-norm *min* used. Then $E^Z(F) = (A - z - a, B - z, C + z, D + z + a)$.

Concentration is effected in a similar way: $E_Z(F)(s) = \inf_{r \in \Omega} I(F(r), E^Z(s, r))$, where I is a fuzzy implication. If F and Z are as above with the condition that $C - B \geq 2z$, and I is the Gödel implication, then $E_Z(F) = (A + z + a, B + z, C - z, D - z - a)$.

For example, suppose we start with a set F described in trapezoidal notation as $F = (A, B, C, D) = (2, 4, 6, 8)$, and an approximate equality function parametrised by $Z = (-z-a, -z, z, z+a) = (-1, -0.5, 0.5, 1)$. The dilation of the set F using T-norm *min* is then:

$$E^Z(F) = (A - z - a, B - z, C + z, D + z + a) = (1, 3.5, 6.5, 9)$$

The concentration of the set F using the Gödel implication is:

$$E_Z(F) = (A + z + a, B + z, C - z, D - z - a) = (3, 4.5, 5.5, 7)$$

These effects are illustrated in figure 2.

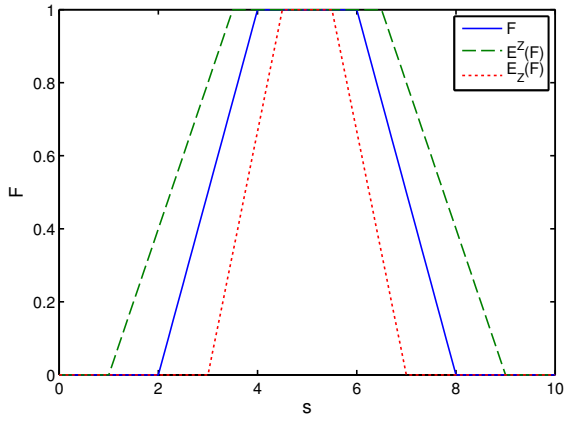

Figure 2: Illustration of the expansion and contraction modifiers proposed in [2]. The original set F can be described in trapezoidal set notation as $(2, 4, 6, 8)$. The approximate equality function is parametrised by a set $Z = (-1, -0.5, 0.5, 1)$. Dilating the set as described in [2] gives the set $E^Z(F) = (1, 3.5, 6.5, 9)$. Concentrating the set as described results in the set $E_Z(F) = (3, 4.5, 5.5, 7)$

The intuitive idea behind this approach is that if an object x_1 resembles another object x_2 that is L, then x_1 can be said to be 'quite L'. Conversely, object x_2 that is L can be said to be 'very L' only if all the objects x that resemble it can be said to be L. This formulation alters both the core and support of the fuzzy set L, which has been argued to be a desirable effect.

Following [2, 8, 24], we will propose linguistic modifiers that are semantically grounded rather than attempting to show their utility in classifiers, reasoning or to examine the algebra of modifiers. Our approach to linguistic modifiers arises very naturally from the label semantics framework, and the primary result does not require any parameters additional to the original membership function of the concept. We also show similarities between our model and the two detailed above.

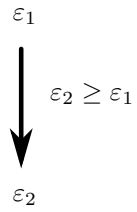

Figure 3: Directed acyclic graph representing the hedge 'quite'. The threshold ε_2 is dependent on the threshold ε_1 by $\varepsilon_2 \geq \varepsilon_1$.

4 Label semantics approach to linguistic hedges

We present three formulations of linguistic hedges with increasing levels of generality. The first assumes that prototypes are equal. Secondly, we show that an analogue holds where prototypes are not equal, and thirdly that these hold in the case where the second threshold is a function of the first. We go on to show similarities between our model and those of [2, 8, 24]. Furthermore, we show that hedges are compositional, and look at their behaviour in the limit of composition.

As described in section 2.3, LA denotes a finite set of labels $\{L_i\}$ that agents use to describe basic categories. Ω is the underlying domain of discourse, with prototypes $P_i \in \Omega$ and thresholds ε_i, drawn from a distribution δ_{ε_i}. As before, the appropriateness $\mu_{L_i}(x) = \Delta_i(d(x, P_i)) = \int_{d(x,P_i)}^{\infty} \delta_{\varepsilon_i}(\varepsilon_i)\mathrm{d}\varepsilon_i$. We use the notation $L_i =< P_i, d, \delta_{\varepsilon_i} >$.

A concept L_1 can be narrowed or broadened to a second concept L_2 using the linguistic hedges 'very' and 'quite' respectively, i.e. L_2 is defined as 'quite L_1'. The directed acyclic graph illustrating this dependency is given in figure 3. In this case, the threshold ε_2 associated with L_2 is dependent on ε_1 in that $\varepsilon_2 \geq \varepsilon_1$. In the case of 'very', we have that $\varepsilon_2 \leq \varepsilon_1$. Essentially, for 'quite', we are saying that however wide a margin of certainty we apply the label 'tall' with, the margin for 'quite tall' will be wider, and conversely for 'very'.

4.1 Hedges with unmodified prototypes

Definition 1 (Dilation and Concentration)**.** *A label $L_2 =< P_2, d, \delta_{\varepsilon_2} >$ is a* dilation *of a label $L_1 =< P_1, d, \delta_{\varepsilon_1} >$ when ε_2 is dependent on ε_1 such that $\varepsilon_2 \geq \varepsilon_1$. L_2 is a* concentration *of L_1 when ε_2 is dependent on ε_1 such that $\varepsilon_2 \leq \varepsilon_1$.*

Theorem 2 (L_2 = quite L_1)**.** *Suppose $L_2 =< P_2, d, \delta_{\varepsilon_2} >$ is a dilation of $L_1 =< P_1, d, \delta_{\varepsilon_1} >$, so that $\varepsilon_2 \geq \varepsilon_1$. Suppose also that $P_1 = P_2 = P$, and that the marginal (unconditional) distribution of ε_2, before conditioning on the knowledge that $\varepsilon_2 \geq \varepsilon_1$, is identical to δ_{ε_1}, since L_2 is a dilation of L_1. Then $\forall x \in \Omega$, $\mu_{L_2}(x) = \mu_{L_1}(x) - \mu_{L_1}(x)\ln(\mu_{L_1}(x))$.*

Proof.

$$\delta_{\varepsilon_2|\varepsilon_1}(\varepsilon_2|\varepsilon_1) = \begin{cases} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\int_{\varepsilon_1}^{\infty} \delta_{\varepsilon_1}(\varepsilon_2)d\varepsilon_2} & \text{if } \varepsilon_2 \geq \varepsilon_1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_1(\varepsilon_1)} & \text{if } \varepsilon_2 \geq \varepsilon_1 \\ 0 & \text{otherwise} \end{cases}$$

and hence,

$$\delta(\varepsilon_1, \varepsilon_2) = \delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_2|\varepsilon_1}(\varepsilon_2|\varepsilon_1) = \begin{cases} \frac{\delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_1(\varepsilon_1)} & \text{if } \varepsilon_2 \geq \varepsilon_1 \\ 0 & \text{otherwise} \end{cases}$$

Then since $\varepsilon_2 \geq \varepsilon_1$ we have that

$$\mu_{L_2}(x) = \int_0^\infty \int_{max(\varepsilon_1, d(x,P))}^\infty \delta(\varepsilon_1, \varepsilon_2) d\varepsilon_2 d\varepsilon_1 = \int_0^\infty \int_{max(\varepsilon_1, d(x,P))}^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_1(\varepsilon_1)} d\varepsilon_2 d\varepsilon_1$$
$$= \int_0^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_{d(x,P)}^\infty \delta_{\varepsilon_1}(\varepsilon_2) d\varepsilon_2 d\varepsilon_1 + \int_{d(x,P)}^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_{\varepsilon_1}^\infty \delta_{\varepsilon_1}(\varepsilon_2) d\varepsilon_2 d\varepsilon_1$$
$$= \mu_{L_1}(x) \int_0^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta(\varepsilon_1)} d\varepsilon_1 + \int_{d(x,P)}^\infty \delta_{\varepsilon_1}(\varepsilon_1) d\varepsilon_1 = \mu_{L_1}(x) - \mu_{L_1}(x) \ln(\mu_{L_1}(x))$$

□

The following example gives an illustration of the effect of applying this hedge, in comparison with the standard dilation hedge $(\mu_L(x))^{1/2}$.

Example 3. *Suppose our conceptual space $\Omega = \mathbb{R}$ with Euclidean distance and that a label L has prototype $P = 5$, and threshold $\varepsilon \sim$ Uniform$[0, 3]$. Then*

$$\mu_L(x) = \begin{cases} 1 - \frac{|x-5|}{3} & \text{if } x \in [2, 8] \\ 0 & \text{otherwise} \end{cases}$$

We can then form a new label qL with prototype $P_q = P = 5$ and threshold $\varepsilon_q \geq \varepsilon$. Then, according to theorem 2, $\mu_{qL}(x) = \mu_L(x) - \mu_L(x) \ln \mu_L(x)$. The effect of applying a dilation hedge to L can be seen in figure 4. The dilation hedge given above is contrasted with Zadeh's dilation hedge $(\mu_L(x))^{1/2}$.

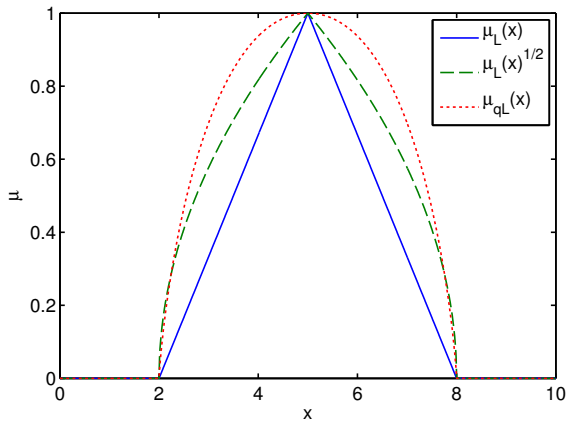

Figure 4: Plots of $\mu_L(x)$, $\mu_L(x)^{1/2}$, and $\mu_{qL}(x) = \mu_L(x) - \mu_L(x) \ln \mu_L(x)$. Values of $\mu_{qL}(x)$ near to the prototype $P = 5$ are greater than equivalent values of $\mu_L(x)^{1/2}$.

Theorem 4 ($L_2 =$ very L_1)**.** *Suppose $L_2 =< P_2, d, \delta_{\varepsilon_2} >$ is a concentration of $L_1 =< P_1, d, \delta_{\varepsilon_1} >$, so that $\varepsilon_2 \leq \varepsilon_1$. Suppose also that $P_1 = P_2 = P$, and that the marginal (unconditional) distribution of ε_2, before conditioning on the knowledge that $\varepsilon_2 \leq \varepsilon_1$, is identical to δ_{ε_1}, since L_2 is a concentration of L_1. Then $\forall x \in \Omega$, $\mu_{L_2} = \mu_{L_1}(x) + (1 - \mu_{L_1}(x)) \ln(1 - \mu_{L_1}(x))$.*

Proof.

$$\delta_{\varepsilon_2|\varepsilon_1}(\varepsilon_2|\varepsilon_1) = \begin{cases} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\int_0^{\varepsilon_1} \delta_{\varepsilon_1}(\varepsilon_2) d\varepsilon_2} & \text{if } \varepsilon_2 \leq \varepsilon_1 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{1 - \Delta_1(\varepsilon_1)} & \text{if } \varepsilon_2 \leq \varepsilon_1 \\ 0 & \text{otherwise} \end{cases}$$

and hence,

$$\delta(\varepsilon_1,\varepsilon_2)=\delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_2|\varepsilon_1}(\varepsilon_2|\varepsilon_1)=\begin{cases}\frac{\delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_1}(\varepsilon_2)}{1-\Delta_1(\varepsilon_1)} & \text{if } \varepsilon_2\leq\varepsilon_1\\ 0 & \text{otherwise}\end{cases}$$

So since $\varepsilon_2 \leq \varepsilon_1$ we have that:

$$\begin{aligned}\mu_{L_2}(x)&=\int_0^\infty\int_{min(\varepsilon_1,d(x,P))}^{\varepsilon_1}\delta(\varepsilon_1,\varepsilon_2)d\varepsilon_2 d\varepsilon_1=\int_0^\infty\int_{min(\varepsilon_1,d(x,P))}^{\varepsilon_1}\frac{\delta_{\varepsilon_1}(\varepsilon_1)\delta_{\varepsilon_1}(\varepsilon_2)}{1-\Delta_1(\varepsilon_1)}d\varepsilon_2 d\varepsilon_1\\ &=\int_{d(x,P)}^\infty\frac{\delta_{\varepsilon_1}(\varepsilon_1)}{1-\Delta_1(\varepsilon_1)}\int_{d(x,P)}^{\varepsilon_1}\delta_{\varepsilon_1}(\varepsilon_2)d\varepsilon_2 d\varepsilon_1\\ &=\int_{d(x,P)}^\infty\frac{\delta_{\varepsilon_1}(\varepsilon_1)}{1-\Delta_1(\varepsilon_1)}\left(\int_0^{\varepsilon_1}\delta_{\varepsilon_1}(\varepsilon_2)d\varepsilon_2-\int_0^{d(x,P)}\delta_{\varepsilon_1}(\varepsilon_2)d\varepsilon_2\right)d\varepsilon_1\\ &=\mu_{L_1}(x)-(1-\mu_{L_1}(x))\int_{d(x,P)}^\infty\frac{\delta_{\varepsilon_1}(\varepsilon_1)}{1-\Delta_1(\varepsilon_1)}d\varepsilon_1=\mu_{L_1}(x)+(1-\mu_{L_1}(x))\ln(1-\mu_{L_1}(x))\end{aligned}$$

□

The effect of these hedges are illustrated in the following example.

Example 5. *Suppose the label L is as described in example 3. We can form a new label vL with prototype $P_v = P = 5$ and threshold $\varepsilon_v \leq \varepsilon$. Then, according to theorem 4, $\mu_{vL}(x) = \mu_L(x) + (1-\mu_L(x))\ln(1-\mu_L(x))$ The effect of applying this contraction hedge is seen in figure 5, and again, this concentration hedge is contrasted with Zadeh's concentration hedge $(\mu_L(x))^2$.*

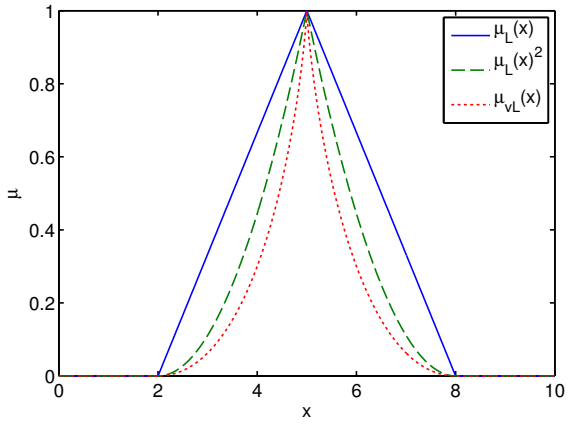

Figure 5: Plots of $\mu_L(x)$, $\mu_L(x)^2$, and $\mu_{vL}(x) = \mu_L(x) + (1-\mu_L(x))\ln(1-\mu_L(x))$. Values of $\mu_{vL}(x)$ decrease more quickly as we move from the prototype $P = 5$ than do equivalent values of $\mu_L(x)^2$.

These hedges can also be applied across multiple dimensions, demonstrated in the example below.

Example 6. *Suppose we have two labels 'tall' and 'thin'. 'Tall' has prototype $P_{tall} = 6.5ft$ and 'thin' has prototype $P_{thin} = 24in$. The appropriateness of each label is defined by:*

$$\mu_{tall}(x_1)=\begin{cases}1 & if x_1 > 6.5\\ 1-(P_{tall}-x_1) & if x_1 \in [5.5, 6.5]\\ 0 & otherwise\end{cases}$$

where the variable x_1 measures height

$$\mu_{thin}(x_2) = \begin{cases} 1 & if x_2 < 24 \\ 1 - \frac{(x_2 - P_{thin})}{4} & if x_2 \in [24, 28] \\ 0 & otherwise \end{cases}$$

where x_2 measures waist size.

Suppose further that being tall and being thin are independent of each other. The appropriateness of the label 'tall and thin' could then be defined by:

$$\mu_{tall\ and\ thin}(x_1, x_2) = \mu_{tall}(x_1)\mu_{thin}(x_2)$$

If 'tall' and 'thin' are independent, we can treat their hedges separately, so the appropriateness of a label 'very tall and quite thin' is:

$$\begin{aligned} \mu_{very\ tall\ and\ quite\ thin}(x_1, x_2) &= \mu_{very\ tall}(x_1)\mu_{quite\ thin}(x_2) \\ &= (\mu_{tall}(x_1) + (1 - \mu_{tall}(x_1)\ln(1 - \mu_{tall}(x_1)))(\mu_{thin}(x_2) - \mu_{thin}(x_2)\ln(\mu_{thin}(x_2))) \end{aligned}$$

This is illustrated in figure 6

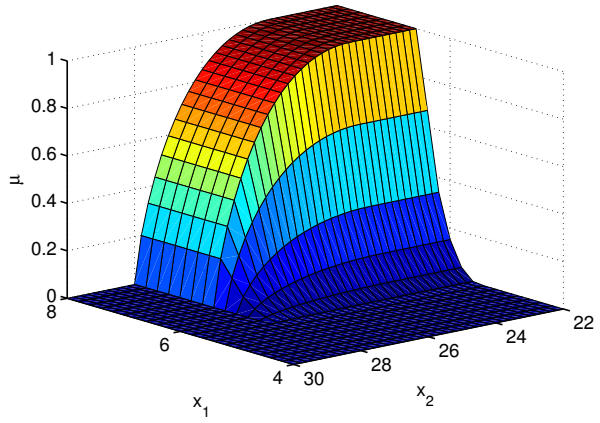

Figure 6: Plot of $\mu_{\text{very tall and quite thin}}(x_1, x_2) = \mu_{\text{very tall}}(x_1)\mu_{\text{quite thin}}(x_2)$. Notice how appropriateness drops off quickly as the variable x_1, i.e. height, decreases. In contrast, appropriateness decreases slowly as we increase the variable x_2, or waist size

4.2 Hedges with differing prototypes

As they stand, the hedges proposed leave the core and support of the fuzzy sets unchanged, which is often argued to be undesirable [2, 3, 24, 21]. A slight modification yields models of hedges in which the core, or prototype, of the concept has been changed.

Theorem 7 (Dilation)**.** *Suppose that L_2 = quite L_1, as in theorem 2, but that $P_2 \neq P_1$. Then $\mu_{L_2}(x) = \Delta_1(d(x, P_2)) - \Delta_1(d(x, P_2))\ln(\Delta_1(d(x, P_2)))$.*

Proof. Substitute $\Delta_1(d(x, P_2))$ for $\mu_{L_1}(x)$ throughout proof of theorem 2 □

Theorem 8 (Concentration)**.** *Suppose that L_2 = very L_1, as in theorem 4, but that $P_2 \neq P_1$. Then $\mu_{L_2}(x) = \Delta_1(d(x, P_2)) + (1 - \Delta_1(d(x, P_2)))\ln(1 - \Delta_1(d(x, P_2)))$.*

Proof. As above. □

Corollary 9. *If $\varepsilon_2 \geq \varepsilon_1$ and $P_2 \supseteq P_1$ then $\mu_{L_2}(x) \geq \mu_{L_1}(x) - \mu_{L_1}(x)\ln(\mu_{L_1}(x))$, and if $\varepsilon_2 \leq \varepsilon_1$ and $P_2 \subseteq P_1$, then $\mu_{L_2}(x) \leq \mu_{L_1}(x) + (1-\mu_{L_1}(x))\ln(1-\mu_{L_1}(x))$.*

Proof. $\mu_{L_2}(x) = \Delta_1(d(x,P_2)) - \Delta_1(d(x,P_2))\ln(\Delta_1(d(x,P_2)))$, but since $P_2 \supseteq P_1$, $d(x,P_2) \leq d(x,P_1)\ \forall x \in \Omega$, and so $\Delta_1(d(x,P_2)) \geq \Delta_1(d(x,P_1)) = \mu_{L_1}(x)\ \forall x \in \Omega$. Hence, $\mu_{L_2}(x) \geq \mu_{L_1}(x) - \mu_{L_1}(x)\ln(\mu_{L_1}(x))$. A similar argument shows that $\mu_{L_2}(x) \leq \mu_{L_1}(x) + (1-\mu_{L_1}(x))\ln(1-\mu_{L_1}(x))$. □

Example 10. *Suppose our conceptual space $\Omega = \mathbb{R}$ with Euclidean distance and that a label L has prototype $P = [4.5, 5.5]$, and threshold $\varepsilon \sim$ Uniform$[0,3]$. Then*

$$\mu_L(x) = \begin{cases} 1 & \text{if } x \in [4.5, 5.5] \\ 1 - \frac{|x-5|}{3} & \text{if } x \in [1.5, 4.5] \text{ or } x \in [5.5, 8.5] \\ 0 & \text{otherwise} \end{cases}$$

We form the concept qL by setting the prototype to be $P_q = [4, 6]$, and $\varepsilon_q \geq \varepsilon$. The effect of applying our dilation hedge is illustrated in figure 7.

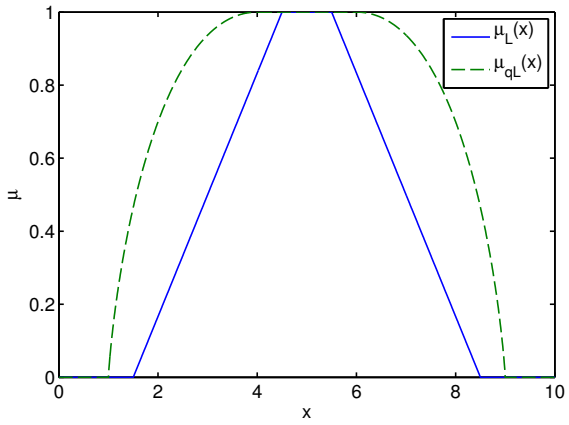

Figure 7: Plots of $\mu_L(x)$ and $\mu_{qL}(x) = \Delta(d(x,P_q)) - \Delta(d(x,P_q))\ln(\Delta(d(x,P_q)))$. The prototype of L has been expanded.

Conversely, suppose that a label L has prototype $P = [4, 6]$ and threshold $\varepsilon \sim$ Uniform$[0,3]$. Then

$$\mu_L(x) = \begin{cases} 1 & \text{if } x \in [4, 6] \\ 1 - \frac{|x-5|}{3} & \text{if } x \in [1, 4] \text{ or } x \in [6, 9] \\ 0 & \text{otherwise} \end{cases}$$

We now form the concept vL by contracting the prototype to $P_v = [4.5, 5.5]$ and setting $\varepsilon_v \leq \varepsilon$. The effect of applying the contraction hedge is illustrated in figure 8.

4.3 Functions of thresholds

It may be the case that the threshold of a given concept is greater than or less than a function of the original threshold. This could hold when a hedged concept is a very expanded or restricted version of the original concept, such as when the hedge 'loosely' or 'extremely' is used. Our formulae can also take account of this.

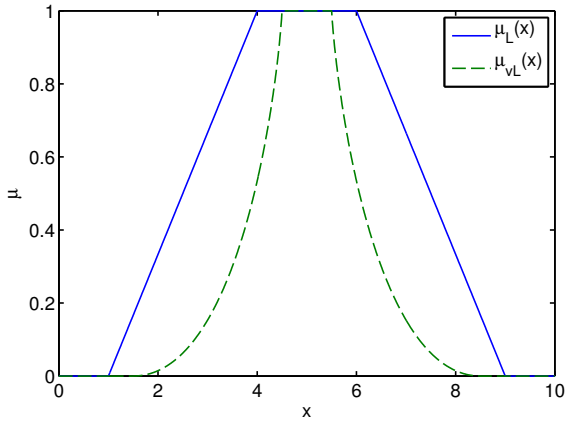

Figure 8: Plots of $\mu_L(x)$ and $\mu_{vL}(x) = \Delta(d(x, P_v) + (1 - \Delta(d(x, P_v)) \ln(1 - \Delta(d(x, P_v)))$. The prototype of L has been reduced.

Theorem 11. *Suppose* $L_2 =< P_2, d, \delta_{\varepsilon_2} >$ *is a dilation of* $L_1 =< P_1, d, \delta_{\varepsilon_1} >$ *with* $P_2 \neq P_1$ *and* $\varepsilon_2 \geq f(\varepsilon_1)$*, where* $f : \mathbb{R} \to \mathbb{R}$ *is strictly increasing or decreasing. Then*

$$\mu_{L_2}(x) = \Delta_1(f^{-1}(d(x, P_2))) - \Delta_1(f^{-1}(d(x, P_2))) \ln(\Delta_1(f^{-1}(d(x, P_2))))$$

Proof. Rewrite $\varepsilon_2 \geq f(\varepsilon_1)$ as $\varepsilon_2 \geq \varepsilon = f(\varepsilon_1)$, where $\varepsilon \sim \delta$ and is associated with a label L with prototype P. Then:

$$\mu_{L_2}(x) = \Delta(d(x, P_2)) - \Delta(d(x, P_2)) \ln(\Delta(d(x, P_2)))$$

as above

Since $f : \mathbb{R} \to \mathbb{R}$ is strictly monotone, f^{-1} exists, and $\Delta(d(x, P)) = P(d(x, P) \leq \varepsilon) = P(f^{-1}(d(x, P)) \leq \varepsilon_1) = \Delta_1(f^{-1}(d(x, P)))$.

So

$$\mu_{L_2}(x) = \Delta_1(f^{-1}(d(x, P_2))) - \Delta_1(f^{-1}(d(x, P_2))) \ln(\Delta_1(f^{-1}(d(x, P_2))))$$

as required. □

Theorem 12. *Suppose* $L_2 =< P_2, d, \delta_{\varepsilon_2} >$ *is a concentration of* $L_1 =< P_1, d, \delta_{\varepsilon_1} >$ *with* $P_2 \neq P_1$ *and* $\varepsilon_2 \leq f(\varepsilon_1)$*, where* $f : \mathbb{R} \to \mathbb{R}$ *is strictly increasing or decreasing. Then*

$$\mu_{L_2}(x) = \Delta_1(f^{-1}(d(x, P_2))) + (1 - \Delta_1(f^{-1}(d(x, P_2)))) \ln(1 - \Delta_1(f^{-1}(d(x, P_2))))$$

Proof. The proof is entirely similar to that of theorem 11 □

4.4 Links to other models of hedges

It is possible to specify the dependence of the threshold of the hedged concept on the threshold of the unhedged concept purely deterministically, i.e. by $\varepsilon_2 = f(\varepsilon_1)$, rather than $\varepsilon_2 \leq f(\varepsilon_1)$. In this case, we can show links to other models of hedges from the literature.

A simple example of a deterministic dependency is given below.

Example 13. *Suppose* $\Omega = \mathbb{R}$*, d is Euclidean distance and that* L_1 *has prototype* $P = 5$ *and* $\varepsilon \sim$ *Uniform*$[0, 3]$*. Then as before,*

$$\mu_L(x) = \begin{cases} 1 - \frac{|x-5|}{3} & \text{if } x \in [2, 8] \\ 0 & \text{otherwise} \end{cases}$$

To implement a dilation hedge, we would form a new label qL with $P_q = P = 5$ and $\varepsilon_q = k_q\varepsilon$ with $k_q > 1$. For a contraction hedge, we would form the label vL by setting $P_v = P = 5$ and $\varepsilon_v = k_v\varepsilon$ with $k_v < 1$. Then,

$$\mu_{hL}(x) = \begin{cases} 1 - \frac{|x-5|}{3k} & \text{if } x \in [5-3k, 5+3k] \\ 0 & \text{otherwise} \end{cases}$$

where $h = q$ or v, $k = k_q$ or k_v respectively.

The effect of implementing these hedges is illustrated in figure 9.

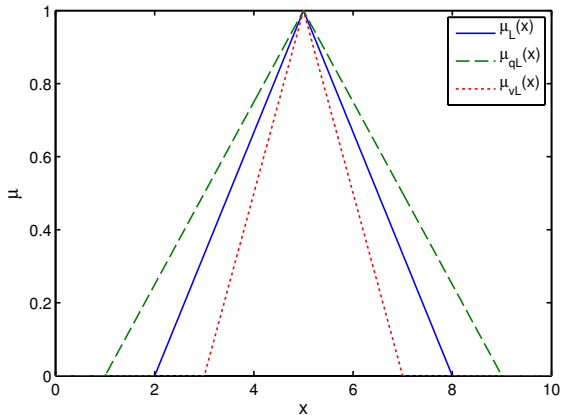

Figure 9: Plots of $\mu_L(x)$, $\mu_{qL}(x)$ and $\mu_{vL}(x)$. The prototype, a single point, remains constant on application of the hedges.

Using this approach, we can also create an effect similar to that of changing the prototype. Suppose that a label L in a conceptual space Ω has a single point P as a prototype, but that the minimum value of the threshold ε is greater than 0, for example, $\varepsilon \sim \text{Uniform}[c, a]$. Then

$$\mu_L(x) = \begin{cases} 1 & \text{if } d(x,P) < c \\ \frac{a}{a-c} - \frac{|x-P|}{a-c} & \text{if } d(x,P) \in [a,c] \\ 0 & \text{otherwise} \end{cases}$$

Suppose that a hedged concept hL is formed from L by the dependency $\varepsilon_h = k\varepsilon$ where k is a constant. Then

$$\mu_L(x) = \Delta(\frac{|x-P|}{k}) = \begin{cases} 1 & \text{if } \frac{|x-P|}{k} < c \\ \frac{a}{(a-c)} - \frac{|x-P|}{k(a-c)} & \text{if } \frac{|x-P|}{k} \in [a,c] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } |x-P| < kc \\ \frac{ka-|x-P|}{k(a-c)} & \text{if } |x-P| \in [ka,kc] \\ 0 & \text{otherwise} \end{cases}$$

This effect is illustrated in the example below.

Example 14. *Suppose that the conceptual space $\Omega = \mathbb{R}$ and that a label L has prototype $P = 5$ and threshold $\varepsilon \sim$ Uniform$[1, 2]$. Then*

$$\mu_L(x) = \begin{cases} 1 & \text{if } |x-5| < 1 \\ 2 - |x-5| & \text{if } |x-5| \in [1,2] \\ 0 & \text{otherwise} \end{cases}$$

Forming a new label qL by applying the hedge $\varepsilon_q = 2\varepsilon$ gives appropriateness measure

$$\mu_{qL}(x) = \begin{cases} 1 & \text{if } |x-5| < 2 \\ \frac{4-|x-5|}{2} & \text{if } |x-5| \in [2,4] \\ 0 & \text{otherwise} \end{cases}$$

Forming a new label vL by applying the hedge $\varepsilon_v = 0.5\varepsilon$ gives appropriateness measure

$$\mu_{vL}(x) = \begin{cases} 1 & \text{if } |x-5| < 0.5 \\ \frac{1-|x-5|}{0.5} & \text{if } |x-5| \in [0.5,1] \\ 0 & \text{otherwise} \end{cases}$$

These are illustrated in figure 10.

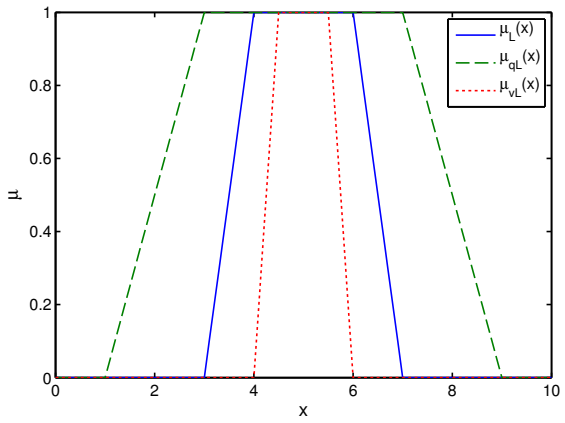

Figure 10: Plots of $\mu_L(x)$, $\mu_{qL}(x)$ and $\mu_{vL}(x)$. In this case, the set $\{x \in \Omega : \mu_{hL}(x) = 1\}$, where $h = v, q$ or nothing, expands or contracts on application of the hedges.

Notice that if we set $\Omega = \mathbb{R}^+$ and label L specified by $P = 0$, $\varepsilon \sim \text{Uniform}[c, a]$, this is identical to the linear membership model given in [24]. Specifically, we have

$$\mu_L(x) = \begin{cases} 1 & \text{if } x < c \\ \frac{a-x}{a-c} & \text{if } x \in [c,a] \\ 0 & \text{otherwise} \end{cases}$$

Forming a hedged concept hL by setting $P_{hL} = P = 0$ and $\varepsilon_{hL} = k\varepsilon$ gives

$$\mu_{hL}(x) = \begin{cases} 1 & \text{if } x < kc \\ \frac{ka-x}{ka-kc} & \text{if } x \in [kc,ka] \\ 0 & \text{otherwise} \end{cases}$$

Comparing this with the model given in section 3, we see that this is precisely the model proposed by [24] in the linear case.

Similarity between the hedging effects illustrated in figure 9 and the effects implemented in the model proposed in [2], illustrated in figure 2, can clearly be seen. To derive the model given in [2], we describe the fuzzy sets associated with labels L and hL in trapezoidal notation. Notice that $L = (P-a, P-c, P+c, P+a)$ and $hL = (P-ka, P-kc, P+kc, P+ka)$. We can render this transformation in the terms employed by [2]. Consider labels L and hL as fuzzy sets characterised by the appropriateness measure $\mu_L(x)$ and $\mu_{hL}(x)$. Then, in the case of dilation, we have:

$$qL = E^Z(L)(s) = \sup_{r\in\Omega} T(\mu_L(s), E^Z(s,r))$$

and for contraction,

$$vL = E_Z(L)(s) = \inf_{r\in\Omega} I(\mu_L(s), E^Z(s,r))$$

When T is the T-norm min, I is the Gödel implication and $Z = (-z-\alpha, -z, z, z+\alpha)$ (with the restriction that $c < z$ to ensure a well-defined set), the approach in [2] gives $qL = (P-a-z-\alpha, P-c-z, P+c+z, P+a+z+\alpha)$. If we set $z = (k-1)c$ and $\alpha = (k-1)(a-c)$, this is equal to $qL = (P-ka, P-kc, P+kc, P+ka)$.

However, we also require that $vL = (P-ka, P-kc, P+kc, P+ka)$. The approach in [2] gives $vL = (P-a+z+\alpha, P-c+z, P+c-z, P+a-z-\alpha)$, and we therefore need to set $z = (1-k)c$ and $\alpha = (1-k)(a-c)$

This formulation is not as general as given in [2], however, note that it only uses one additional parameter and no additional operators, rather than the two parameters and either a T-norm or implication used by [2].

Two more key models from the literature are the powering and shifting modifiers proposed in [31]. Recall that powering modifiers are of the form $\mu_{hL}(x) = (\mu_L(x))^k$ and shifting modifiers are of the form $\mu_{hL}(x) = (\mu_L(x-a))$. Shifting modifiers are easy to implement within our model, simply by shifting the prototype by the quantity a.

Powering modifiers can be expressed as a function of the threshold ε given a particular distribution of the threshold δ. Suppose $\Omega = \mathbb{R}$, $\varepsilon \sim U[0,c]$, giving

$$\mu_L(x) = \begin{cases} 1 - \frac{d(x,P)}{b} & \text{if } x \in [P-b, P+b] \\ 0 & otherwise \end{cases}$$

and suppose a new label hL is formed with prototype P and threshold $\varepsilon_h = f(\varepsilon)$ such that $\mu_{hL}(x) = \mu_L(x)^k$. Then $\mu_{hL}(x) = \Delta(f^{-1}(d(x,P))) = (\Delta(d(x,P)))^k$, so

$$f^{-1}(d(x,P)) = \Delta^{-1}((\Delta(d(x,P)))^k) = b - b(\frac{(b-d(x,P))^k}{b^k}) = b - \frac{(b-d(x,P))^k}{b^{k-1}}$$

and hence

$$\varepsilon_{hL} = f(\varepsilon) = b - (b^{k-1}(b-\varepsilon))^{1/k}$$

This expression seems surprisingly complicated, and there may be better ways of deriving the powering hedges that are not as a function of the threshold ε.

In this section we have shown that our general model can capture some of the many approaches found in the literature as special cases. We now go on to look at the property of compositionality that is exhibited by a number of models.

4.5 Compositionality

One of the features of hedges seen in [2, 8, 24, 31] is that they can be applied multiple times. Within the label semantics framework, this consists in expanding or reducing the threshold of a concept a number of times. The directed acyclic graph corresponding to this is shown in figure 11. We show below that expressions for 'very' and 'quite' as given in theorems 2 and 4 are compositional, and that the appropriateness of a concept after n applications of a hedge can be expressed purely in terms of the appropriateness after $n-1$ applications. We also derive expressions for the composition of deterministic hedges as described in section 4.4.

Theorem 15. *Suppose that labels $L_1, L_2, ..., L_n$ are defined by prototypes $P_1 = P_2 = ... = P_n = P$, thresholds $\varepsilon_1 \geq \varepsilon_2 \geq ... \geq \varepsilon_n$ and with a distance metric d common to all labels. Then $\mu_{L_n}(x) = \mu_{L_{n-1}}(x) + (1 - \mu_{L_{n-1}}(x)) \ln(1 - \mu_{L_{n-1}}(x))$*

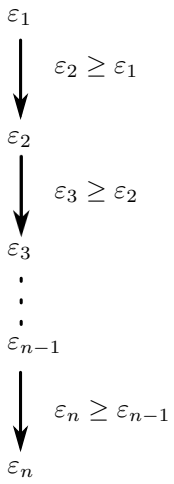

Figure 11: Threshold dependencies on multiple applications of the hedge 'quite'. Each threshold ε_i is directly dependent on the one preceding it, ε_{i-1}.

Proof. We proceed by induction on n. Theorem 2 proves this for $n = 2$. Assuming true for $n = k$, we have

$$
\begin{aligned}
\mu_{L_{k+1}}(x) &= \int_0^\infty \int_0^\infty ... \int_{max(d(x,P),\varepsilon_k)}^\infty \delta(\varepsilon_1, \varepsilon_2, ..., \varepsilon_{k+1}) \mathrm{d}\varepsilon_{k+1} ... \mathrm{d}\varepsilon_1 \\
&= \int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} ... \int_0^\infty \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} \int_{max(d(x,P),\varepsilon_k)}^\infty \delta_{\varepsilon_k}(\varepsilon_{k+1}) \mathrm{d}\varepsilon_{k+1} ... \mathrm{d}\varepsilon_1 \\
&= \int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} ... \int_{max(d(x,P),\varepsilon_{k-1})}^\infty \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} \overbrace{\int_{\varepsilon_k}^\infty \delta_{\varepsilon_k}(\varepsilon_{k+1}) \mathrm{d}\varepsilon_{k+1}}^{=\Delta_k(\varepsilon_k)} \mathrm{d}\varepsilon_k ... \mathrm{d}\varepsilon_1 \\
&\quad + \int_0^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_{\varepsilon_1}^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} ... \int_{\varepsilon_{k-1}}^{d(x,P)} \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} \overbrace{\int_{d(x,P)}^\infty \delta_{\varepsilon_k}(\varepsilon_{k+1}) \mathrm{d}\varepsilon_{k+1}}^{=\mu_{L_k}(x)} \mathrm{d}\varepsilon_k ... \mathrm{d}\varepsilon_1 \\
&= \overbrace{\int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_0^\infty \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} ... \int_{max(d(x,P),\varepsilon_k)}^\infty \delta_{\varepsilon_{k-1}}(\varepsilon_k) \mathrm{d}\varepsilon_k ... \mathrm{d}\varepsilon_1}^{=\mu_{L_k}(x) \text{ by ind. hyp.}} \\
&\quad + \mu_{L_k}(x) \int_0^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \int_{\varepsilon_1}^{d(x,P)} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} ... \int_{\varepsilon_{k-1}}^{d(x,P)} \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} \mathrm{d}\varepsilon_k ... \mathrm{d}\varepsilon_1 \\
&= \mu_{L_k}(x) + \mu_{L_k}(x) \int_0^{d(x,P)} \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} ... \int_0^{\varepsilon_3} \frac{\delta_{\varepsilon_1}(\varepsilon_2)}{\Delta_2(\varepsilon_2)} \int_0^{\varepsilon_2} \frac{\delta_{\varepsilon_1}(\varepsilon_1)}{\Delta_1(\varepsilon_1)} \mathrm{d}\varepsilon_1 ... \mathrm{d}\varepsilon_k \\
&= \mu_{L_k}(x) + \mu_{L_k}(x) \overbrace{\int_0^{d(x,P)} \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)} ... \int_0^{\varepsilon_3} \frac{-\delta_{\varepsilon_1}(\varepsilon_2) \ln(\Delta_1(\varepsilon_2))}{\Delta_2(\varepsilon_2)} \mathrm{d}\varepsilon_2 ... \mathrm{d}\varepsilon_k}^{=A}
\end{aligned}
$$

By the inductive hypothesis, $\forall i = 0...k$

$$\begin{aligned}\delta_{\varepsilon_i}(\varepsilon_i) &= -\frac{\mathrm{d}}{\mathrm{d}\varepsilon_\mathrm{i}}\Delta_i(\varepsilon_i) \\ &= -\frac{\mathrm{d}}{\mathrm{d}\varepsilon_\mathrm{i}}(\Delta_{i-1}(\varepsilon_i) - \Delta_{i-1}(\varepsilon_i)\ln(\Delta_{i-1}(\varepsilon_i))) \\ &= -\delta_{\varepsilon_{i-1}}(\varepsilon_i)\ln(\Delta_{i-1}(\varepsilon_i))\end{aligned}$$

Recursively substituting in A, we obtain

$$\begin{aligned}\mu_{L_{k+1}}(x) &= \mu_{L_k}(x) + \mu_{L_k}(x)\int_0^{d(x,P)} \frac{\delta_{\varepsilon_{k-1}}(\varepsilon_k)}{\Delta_k(\varepsilon_k)}...\int_0^{\varepsilon_3}\frac{\delta_{\varepsilon_2}(\varepsilon_2)}{\Delta_2(\varepsilon_2)}\mathrm{d}\varepsilon_2...\mathrm{d}\varepsilon_k \\ &= \mu_{L_k}(x) + \mu_{L_k}(x)\int_0^{d(x,P)}\frac{\delta_{\varepsilon_k}(\varepsilon_k)}{\Delta_k(\varepsilon_k)}\mathrm{d}\varepsilon_k \\ &= \mu_{L_k}(x) - \mu_{L_k}(x)\ln(\mu_{L_k}(x))\end{aligned}$$

□

Theorem 16. *Suppose labels $L_1, L_2, ..., L_n$ are defined by prototypes $P_1 = P_2 = ... = P_n = P$, thresholds $\varepsilon_1 \leq \varepsilon_2 \leq ... \leq \varepsilon_n$, and that distance metric d is common to all. Then $\mu_{L_n}(x) = \mu_{L_{n-1}}(x) - \mu_{L_{n-1}}(x)\ln(\mu_{L_{n-1}}(x))$*

Proof. Similar to proof of theorem 15. □

We can also derive expressions for the composition of deterministic hedges.

Theorem 17. *Suppose labels $L_1, L_2, ..., L_n$ are defined by prototypes $P_1 = P_2 = ... = P_n = P$, thresholds $\varepsilon_n = f(\varepsilon_{n-1}), \varepsilon_{n-1} = f(\varepsilon_{n-2}), ..., \varepsilon_2 = f(\varepsilon_1)$, where f is monotone increasing or decreasing, and that distance metric d is common to all. Then $\mu_{L_n}(x) = \Delta_1(f^{-(n-1)}(d(x,P))$, where f^{-k} signifies f^{-1} composed k times.*

Proof. $\mu_{L_2}(x) = \Delta_1(f^{-1}(d(x,P))$. Suppose that $\mu_{L_k}(x) = \Delta_k(d(x,P)) = \Delta_1(f^{-(k-1)}(d(x,P))$. Since $\varepsilon_{k+1} = f(\varepsilon_k)$, we have $\mu_{L_{k+1}}(x) = \Delta_k(f^{-1}(d(x,P)) = \Delta_1(f^{-k}(d(x,P)))$.

Therefore $\mu_{L_n}(x) = \Delta_1(f^{-(n-1)}(d(x,P))$ by induction. □

Since labels can be composed in this way, we can model different degrees of emphasis corresponding to the composition of multiple hedges. So, for example, we could model 'extremely L' as 'very, very L'. This is illustrated in example 18.

Example 18. *Suppose the label L is as described in example 3, i.e. L has prototype $P = 5$, and threshold $\varepsilon \sim$ Uniform$[0,3]$. We can form a new label vL with prototype $P_v = P = 5$ and threshold $\varepsilon_v \leq \varepsilon$, which has appropriateness $\mu_{vL}(x) = \mu_L(x) + (1-\mu_L(x))\ln(1-\mu_L(x))$ as shown in theorem 4. We may then form another new label vvL with prototype $P_{vv} = P_v = 5$ and threshold $\varepsilon_{vv} \leq \varepsilon_v$ with appropriateness $\mu_{vvL}(x) = \mu_{vL}(x) + (1-\mu_{vL}(x))\ln(1-\mu_{vL}(x))$ as described in theorem 15. The effect of applying this contraction hedge is seen in figure 12. We have contrasted the effect of the composed hedges with $(\mu_L(x))^4$.*

Since these hedges can be composed only an integral number of times, we cannot obtain the differences in grade that could be achieved with using various powers in a powering modifier, e.g. $\mu_L(x)^{1.73}$. However, in section 4.3 we discuss how to tune the intensity of hedges by using dependencies on functions of thresholds. We have further shown in section 4.4 how to derive powering and shifting modifiers within our framework. It would be interesting to explore how other examples of hedges can be expressed in this framework.

We have shown that when multiple hedges of the forms seen in theorems 2 and 4 are used, $\mu_{L_n}(x)$ can be expressed purely in terms of the appropriateness of the label directly preceding it. We have not been able to find a closed form solution for this recurrence, however, we can investigate the fixed points of the recurrence and examine what happens to the values of $\mu_{L_n}(x)$ as $n \to \infty$. We have also shown that deterministic hedges can be composed, and we go on to look at their behaviour in the limit of composition.

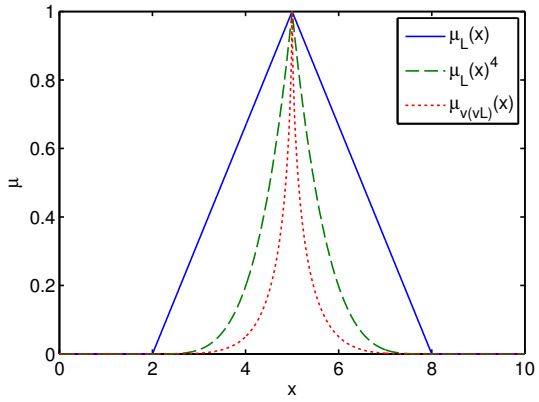

Figure 12: Plots of $\mu_L(x)$, $\mu_L(x)^4$, and $\mu_{vvL}(x)$. Values of $\mu_{vvL}(x)$ decrease more quickly as we move from the prototype $P = 5$ than do equivalent values of $\mu_L(x)^4$.

4.6 Limits of Compositions.

The following results examine the behaviour of $\mu_{L_n}(x)$ as $n \to \infty$

Theorem 19. *Suppose $L_1, ..., L_n$ are labels obtained by repeated application of the dilation operator. Then μ_{L_n} has a limit M^+ and $M^+ = 1$ $\forall x \in \Omega$ such that $\mu_{L_1}(x) \neq 0$, and $M^+ = 0$ otherwise.*

Proof. $\mu_{L_{i+1}}(x) = \mu_{L_i}(x) - \mu_{L_i}(x)\ln(\mu_{L_i}(x))$, $i = 1, .., n-1$. If $\mu_{L_1}(x) = 1$ then $\mu_{L_i}(x) = 1$ $\forall i = 1, ..., n$. Also, if $\mu_{L_1}(x) = 0$ then $\mu_{L_i}(x) = 0$ $\forall i = 1, ..., n$. Suppose $\mu_{L_i}(x) \in (0, 1)$. Then $\mu_{L_{i+1}}(x) > \mu_{L_i}(x)$, and so for $\mu_{L_1}(x) \in (0, 1)$, $\mu_{L_1}(x) < ... < \mu_{L_n}(x)$ is a strictly increasing sequence.

If a limit M^+ exists, then we will have $M^+ = M^+ - M^+\ln(M^+)$, so either $M^+ = 0$ or $\ln(M^+) = 0$. We can't have $M^+ = 0$, since we assume that $\mu_{L_1}(x) \in (0, 1)$ and the sequence is strictly increasing. Therefore, we must have $\ln(M^+) = 0$ and therefore $M^+ = 1$. So

$$\mu_{L_\infty}(x) = \begin{cases} 1 & \mu_{L_1}(x) \in (0, 1] \\ 0 & \mu_{L_1}(x) = 0 \end{cases}$$

□

Theorem 20. *Suppose $L_1, ..., L_n$ are labels obtained by repeated application of the contraction operator. Then μ_{L_n} has a limit M^- and $M^- = 0$ $\forall x \in \Omega$ such that $\mu_{L_1}(x) \neq 1$, and $M^- = 1$ otherwise.*

Proof. $\mu_{L_{i+1}}(x) = \mu_{L_i}(x) + (1 - \mu_{L_i}(x))\ln(1 - \mu_{L_i}(x))$, $i = 1, ..., n-1$. Again, if $\mu_{L_1}(x) = 1$ then $\mu_{L_i}(x) = 1$ $\forall i = 1, ..., n$. Also, if $\mu_{L_1}(x) = 0$ then $\mu_{L_i}(x) = 0$ $\forall i = 1, ..., n$, and for $\mu_{L_1}(x) \in (0, 1)$, $\mu_{L_1}(x) > ... > \mu_{L_n}(x)$ is a strictly decreasing sequence.

If a limit M^- exists, then

$$M^- = M^- + (1 - M^-)\ln(1 - M^-)$$
$$\ln(1 - M^-) = M^-\ln(1 - M^-)$$

So either $M^- = 0$ or $\ln(1 - M^-) = 0$. If $\ln(1 - M^-) = 0$ then $M^- = 1$, which is impossible since $\mu_{L_1}(x) \in (0, 1)$ and the sequence of $\mu_{L_i}(x)$ is strictly decreasing. Therefore

$$\mu_{L_\infty}(x) = \begin{cases} 0 & \mu_{L_1}(x) \in [0, 1) \\ 1 & \mu_{L_1}(x) = 1 \end{cases}$$

□

We have shown here that in the limit, the result of applying dilation or contraction modifiers multiple times is to create a crisp set. In the case of dilation, the crisp set includes the whole support of the fuzzy set associated with the original label, whereas in the case of contraction, the concept reduces to include only its prototype.

When deterministic hedges are used, i.e. $\varepsilon_2 = f(\varepsilon_1)$, the behaviour of the limit depends on the behaviour of the function f and its properties in the limit as $n \to \infty$ of f^{-n}.

Example 21. *Suppose $f(\varepsilon) = 0.5\varepsilon$. Applying this hedge multiple times will result in $\mu_{L_n}(x) = \Delta_1(2^n d(x,P))$. As $n \to \infty$, $2^n d(x,P) \to \infty$, except where $d(x,P) = 0$. Therefore,*

$$\mu_{L_\infty}(x) = \begin{cases} 0 & d(x,P) > 0 \\ 1 & d(x,P) = 0 \end{cases}$$

On the other hand, if $f(\varepsilon) = 2\varepsilon$, $\mu_{L_n}(x) = \Delta_1(2^{-n} d(x,P))$. As $n \to \infty$, $2^{-n} d(x,P) \to 0$, and hence $\mu_{L_\infty}(x) = 1 \ \forall x \in \Omega$.

The behaviour of the hedges given in example 21 is therefore different from those in theorems 19 and 20, since the concept either shrinks to a single point, in the case of contraction, or, in the case of dilation, expands to fill the entire space Ω.

5 Discussion

We have presented formulae for linguistic hedges which are both functional and semantically grounded. The modifiers presented arise naturally from the label semantics framework, in which concepts are represented by a prototype and threshold. Our hedges have an intuitive meaning: if I think that the threshold for a concept 'small' is of a certain width, then the threshold for the concept 'very small' will be narrower. On the other hand, the threshold for the concept 'quite small' will be broader. The hedges proposed are examples of 'type 1' hedges, i.e. they operate equally across all dimensions of the fuzzy set associated with a concept. The first result presented is somewhat similar to a powering modifier since the core and support of the set remain the same. In [3, 22], it is argued that this property is undesirable for hedges used in fuzzy expert systems, since if a query is returning too large a set of answers, this type of contraction hedge does not reduce this overabundance. However, although the hedges we propose do not at their simplest address the overabundance issue, we argue that they address another problem associated with powering hedges, in that they have a clear semantic grounding that the powering modifiers lack.

[2, 8] also propose modifiers that are semantically grounded, using the idea of resemblance to nearby objects. Their formulations have the properties that the core and support of the fuzzy set are both changed, thereby addressing the issue of overabundant answers [3, 22]. In our most specific case, since the prototype is not altered, the core and support of the fuzzy set representing the concept remain the same. However, our initial proposal can be generalised, as in section 4.2, to apply to the case where $P_1 \neq P_2$. Specifying a semantically meaningful way of altering the boundaries of the prototype would answer the objection that the core and support of a set should change under a linguistic hedge.

The most general result (section 4.3) shows that the formula still applies when $\varepsilon_2 \leq f(\varepsilon_1)$, or $\varepsilon_2 \geq f(\varepsilon_1)$. Combined with a distribution δ such that the lower bound of the distribution is not zero, the core and support of the fuzzy set are modified. With the condition $\varepsilon_2 = f(\varepsilon_1)$, we are able to recreate the result given in [24] for linear membership functions, and show how the model proposed by [2] has strong similarities to our own. In this case we have introduced additional parameters, so the simplicity of the original result is lost. However, the further parameters introduced are no more than those introduced by [24], and arguably fewer than those introduced by [2], who require that a resemblance relation be specified, using two additional parameters, and also, that a T-norm or fuzzy implication need to be specified. There are various choices of operator that could be used for either of these, and it is not obvious that any one is better than the others.

We have also shown that the basic case operators 'very' and 'quite' can be composed, which is not immediately obvious from the formulae (section 4.5). Further, we show that in the limit of composition the

membership of any object in the fuzzy part of L, i.e. $\{x : 0 < \mu_L(x) < 1\}$, increases to 1 in the case of 'quite' or decreases to 0 in the case of 'very' (section 4.6). This is similar to the limit of applying the powering modifiers, but differs from what would happen with the modifiers proposed by [2]. In that case, the limit of 'very' would shrink to a single point and the limit of 'quite' would expand to encompass the whole universe of discourse. This can be modelled using the deterministic hedges described in section 4.4. Although behaviour differs slightly, in fact human discourse does not apply modifiers infinitely, so the difference in behaviour is arguably not important.

Our formulation has the benefit that it can be applied in more situations than simply linguistic hedges. For example, the concept 'apple green' has a prototype different to that of just green, and the threshold for 'apple green' is likely to be smaller than the threshold for simply 'green'. Our model can take account of this.

6 Conclusions and further work

We have presented formulae for two simple linguistic hedges, 'very' and 'quite'. These formulae are functional, hence easy to compute, but also semantically grounded, in that they arise naturally from the conceptual framework of label semantics combined with prototype theory and conceptual spaces theory, and in the most specific case require no additional parameters. We have also shown that two other formulations [2, 8, 24], can be derived from this framework with equal or fewer parameters. We have shown that the hedges can be composed and have described their behaviour in the limit of composition.

Further work could look at testing the utility of these hedges in particular classifiers to compare their performance with the classical hedges and with the hedges used by e.g. [3, 22], and also to examine a trade-off between accuracy and the number of parameters used. Alternatively, investigating semantically grounded ways of expanding or reducing prototypes could have a similar impact.

The model could also be extended to the more complicated type 2 hedges such as 'essentially', or 'technically', by treating dimensions of the conceptual space heterogeneously. This requires using some type of weighting or necessity measure on the dimensions, work which is currently ongoing.

7 Acknowledgements

Martha Lewis gratefully acknowledges support from EPSRC Grant No. EP/E501214/1

References

[1] G. Bordogna and G. Pasi. A fuzzy linguistic approach generalizing boolean information retrieval: A model and its evaluation. *JASIS*, 44(2):70–82, 1993.

[2] P. Bosc, D. Dubois, A. HadjAli, O. Pivert, and H. Prade. Adjusting the core and/or the support of a fuzzy set-a new approach to fuzzy modifiers. In *Fuzzy Systems Conference, 2007. FUZZ-IEEE 2007. IEEE International*, pages 1–6. IEEE, 2007.

[3] P. Bosc, A. Hadjali, and O. Pivert. Empty versus overabundant answers to flexible relational queries. *Fuzzy Sets and Systems*, 159(12):1450–1467, 2008.

[4] B. Bouchon-Meunier. Interpretable decisions by means of similarities and modifiers. In *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on*, pages 524–529. IEEE, 2009.

[5] N. Cat Ho and W. Wechler. Hedge algebras: an algebraic approach to structure of sets of linguistic truth values. *Fuzzy Sets and Systems*, 35(3):281–293, 1990.

[6] B. Cetisli. Development of an adaptive neuro-fuzzy classifier using linguistic hedges: Part 1. *Expert Systems with Applications*, 37(8):6093–6101, 2010.

[7] A. Chatterjee and P. Siarry. Nonlinear inertia weight variation for dynamic adaptation in particle swarm optimization. *Computers & Operations Research*, 33(3):859–871, 2006.

[8] M. De Cock and E.E. Kerre. A context based approach to linguistic hedges. *International Journal of Applied Mathematics and Computer Science*, 12(3):371–382, 2002.

[9] L. Di Lascio, A. Gisolfi, and U. Cortes Garcia. Linguistic hedges and the generalized modus ponens. *International Journal of Intelligent Systems*, 14(10):981–993, 1999.

[10] D. Dubois and H. Prade. The three semantics of fuzzy sets. *Fuzzy Sets and Systems*, 90(2):141–150, 1997.

[11] Didier Dubois and Henri Prade. Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets. *Fuzzy Sets and Systems*, 192:3–24, 2012.

[12] M. El-Sayed and D. Pacholczyk. A qualitative reasoning with nuanced information. *Logics in Artificial Intelligence*, pages 283–295, 2002.

[13] G. Frege. Sense and reference. *The Philosophical Review*, 57(3):209–230, 1948.

[14] P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. The MIT Press, 2004.

[15] N.C. Ho and H.V. Nam. An algebraic approach to linguistic hedges in Zadeh's fuzzy logic. *Fuzzy Sets and Systems*, 129(2):229–254, 2002.

[16] G. Lakoff. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4):458–508, 1973.

[17] J. Lawry. A framework for linguistic modelling. *Artificial Intelligence*, 155(1-2):1–39, 2004.

[18] J. Lawry. *Modelling and reasoning with vague concepts*, volume 12. Springer, 2006.

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

[20] B.D. Liu, C.Y. Chen, and J.Y. Tsao. Design of adaptive fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 31(1):32–53, 2001.

[21] J.G. Marin-Blazquez and Q. Shen. Linguistic hedges on trapezoidal fuzzy sets: A revisit. In *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, volume 1, pages 412–415. IEEE, 2001.

[22] J.G. Marin-Blazquez and Q. Shen. Regaining comprehensibility of approximative fuzzy models via the use of linguistic hedges. *Studies in Fuzziness and Soft Computing*, 128:25–53, 2003.

[23] J.G. Marin-Blazquez, Q. Shen, and A.F. Gomez-Skarmeta. From approximative to descriptive models. In *Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on*, volume 2, pages 829–834. IEEE, 2000.

[24] V. Novák. A horizon shifting model of linguistic hedges for approximate reasoning. In *Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on*, volume 1, pages 423–427. IEEE, 1996.

[25] E. Rosch. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3):192, 1975.

[26] Glenn Shafer. *A mathematical theory of evidence*, volume 1. Princeton University Press Princeton, 1976.

[27] Yongchuan Tang and Jonathan Lawry. Linguistic modelling and information coarsening based on prototype theory and label semantics. *International Journal of Approximate Reasoning*, 50(8):1177–1198, 2009.

[28] Yongchuan Tang and Jonathan Lawry. A bipolar model of vague concepts based on random set and prototype theory. *International Journal of Approximate Reasoning*, 53(6):867–879, 2012.

[29] S. Waart van Gulik. A fuzzy logic approach to non-scalar hedges. *Towards Mathematical Philosophy*, pages 233–248, 2009.

[30] L.A. Zadeh. Fuzzy sets*. *Information and Control*, 8(3):338–353, 1965.

[31] L.A. Zadeh. A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 1972.

[32] L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences*, 8(3):199–249, 1975.