Vagueness and Learning: a Type-Theoretic Approach

Fernández, R.; Larsson, S.

Publication date
2014
Document Version
Final published version
Published in
Proceedings of the Third Joint Conference on Lexical and Computational Semantics

Citation for published version (APA):
Fernández, R., & Larsson, S. (2014). Vagueness and Learning: a Type-Theoretic Approach. In *Proceedings of the Third Joint Conference on Lexical and Computational Semantics: *SEM 2014 : Dublin, Ireland, August 23-24 2014* (pp. 151-159). ACL. http://aclweb.org/anthology/S/S14/S14-1019.pdf

Vagueness and Learning: A Type-Theoretic Approach

Raquel Fernández
Institute for Logic, Language and Computation
University of Amsterdam
raquel.fernandez@uva.nl

Staffan Larsson
Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg
sl@ling.gu.se

Abstract

We present a formal account of the meaning of vague scalar adjectives such as '*tall*' formulated in Type Theory with Records. Our approach makes precise how perceptual information can be integrated into the meaning representation of these predicates; how an agent evaluates whether an entity counts as tall; and how the proposed semantics can be learned and dynamically updated through experience.

1 Introduction

Traditional semantic theories such as those described in Partee (1989) and Blackburn and Bos (2005) offer precise accounts of the truth-conditional content of linguistic expressions, but do not deal with the connection between meaning, perception and learning. One can argue, however, that part of getting to know the meaning of linguistic expressions consists in learning to identify the individuals or the situations that the expressions can describe. For many concrete words and phrases, this identification relies on perceptual information. In this paper, we focus on characterising the meaning of vague scalar adjectives such as '*tall*', '*dark*', or '*heavy*'. We propose a formal account that brings together notions from traditional formal semanticswith perceptual information, which allows us to specify how a logic-based interpretation function is determined and modified dynamically by experience.

The need to integrate language and perception has been emphasised by researchers working on the generation and resolution of referring expressions (Kelleher et al., 2005; Reiter et al., 2005; Portet et al., 2009) and, perhaps even more strongly, on the field of robotics, where grounding language on perceptual information is critical to allow artificial agents to autonomously acquire and verify beliefs about the world (Siskind, 2001; Steels, 2003; Roy, 2005; Skocaj et al., 2010). Most of these approaches, however, do not build on theories of formal semantics for natural language. Here we choose to formalise our account in a theoretical framework known as Type Theory with Records (TTR), which has been shown to be suitable for formalising classic semantic aspects such as intensionality, quantification, and negation (Cooper, 2005a; Cooper, 2010; Cooper and Ginzburg, 2011) as well as less standard phenomena such as linguistic interaction (Ginzburg, 2012; Purver et al., 2014), perception and action (Dobnik et al., 2013), and semantic coordination and learning (Larsson, 2009). In this paper we use TTR to put forward an account of the semantics of vague scalar predicates like '*tall*' that makes precise how perceptual information can be integrated into their meaning representation; how an agent evaluates whether an entity counts as tall; and how the proposed semantics for these expressions can be learned and dynamically updated through language use.

We start by giving a brief overview of TTR and explaining how it can be used for classifying entities as being of particular types integrating perceptual information. After that, in Section 3, we describe the main properties of vague scalar predicates. Section 4 presents a probabilistic TTR formalisation of the meaning of '*tall*', which captures its context-dependence and its vague character. In Section 5, we then offer an account of how that meaning representation is acquired and updated with experience. Finally, in Section 6 we discuss related work, before concluding in Section 7.

*Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, pages 151–159, Dublin, Ireland, August 23-24 2014.

2 Meaning as Classification in TTR

In this section we give a brief and hence inevitably partial introduction to Type Theory with Records. For more comprehensive introductions, we refer the reader to Cooper (2005b) and Cooper (2012).

2.1 Type Theory with Records: Main Notions

As in any type theory, the most central notion in TTR is that of a judgement that an object a is of type T, written as $a : T$. In TTR judgements are seen as fundamentally related to perception, in the sense that perceiving inherently involves categorising what we perceive. Some common *basic types* in TTR are Ind (the type of individuals) and $\mathbb{R}^+$ (the type of positive real numbers). All basic types are members of a special type Type. Given types T_1 and T_2, we can create the *function type* $T_1 \rightarrow T2$ whose domain are objects of type T_1 and whose range are objects of type T_2. Types can also be constructed from predicates and objects $P(a_1, \ldots, a_n)$. Such types are called *ptypes* and correspond roughly to propositions in first order logic. In TTR, propositions are types of *proofs*, where proofs can be a variety of things, from situations to sensor readings (more on this below).

Next, we introduce *records* and *record types*. These are structured objects made up of pairs $\langle l, v\rangle$ of labels and values that are displayed in a matrix:

(1) a. A record type:

$$\begin{bmatrix} \ell_1 : T_1 \\ \ell_2 : T_2(\ell_1) \\ \ldots \\ \ell_n : T_n(\ell_1, \ell_2, \ldots, \ell_{n-1}) \end{bmatrix}$$

b. A record: $r = \begin{bmatrix} \ell_1 = a_1 \\ \ell_2 = a_2 \\ \ldots \\ \ell_n = a_n \\ \ldots \end{bmatrix}$

Record r in (1b) is of the record type in (1a) if and only if $a_1 : T_1$, $a_2 : T_2(a_1)$, …, and $a_n : T_n(a_1, a_2, \ldots, a_{n-1})$. Note that the record may contain more fields but would still be of type (1a) if the typing condition holds. Records and record types can be nested so that the value of a label is itself a record (or record type). We can use *paths* within a record or record type to refer to specific bits of structure: for instance, we can use $r.\ell_2$ to refer to a_2 in (1b).

As can be seen in (1a), the labels $\ell_1, \ldots \ell_n$ in a record type can be used elsewhere to refer to the values associated with them. This is a common way of constructing ptypes where the arguments of a predicate are entities that have been introduced before in the record type. A sample record and record type are shown in (2).

(2) $\begin{bmatrix} \mathrm{x} = a \\ \mathrm{c}_{man} = \mathrm{prf}(\mathrm{man}(a)) \\ \mathrm{c}_{run} = \mathrm{prf}(\mathrm{run}(a)) \end{bmatrix} : \begin{bmatrix} \mathrm{x} : \mathrm{Ind} \\ \mathrm{c}_{man} : \mathrm{man}(\mathrm{x}) \\ \mathrm{c}_{run} : \mathrm{run}(\mathrm{x}) \end{bmatrix}$

In (2), a is an entity of type individual and $\mathrm{prf}(P)$ is used as a placeholder for proofs of ptypes P. In the record type above, the ptypes $\mathrm{man(x)}$ and $\mathrm{run(x)}$ constructed from predicates are *dependent* on x (introduced earlier in the record type).

2.2 Perceptual Meaning

Larsson (2013) proposes a system formalised in TTR where some perceptual aspects of meaning are represented using *classifiers*. For example, the meaning of '*right*' (as in '*to the right of*') involves a two-input perceptron classifier $\kappa_{right}(w, t, r)$, specified by a weight vector w and a threshold t, which takes as input a context r including an object x and a position-sensor reading sr_{pos}. The sensor reading consists of a vector containing two real numbers representing the space coordinates of x. The classifier classifies x as either being to the right on a plane or not.[1]

(3) if $r : \begin{bmatrix} \mathrm{x} : \mathrm{Ind} \\ \mathrm{sr}_{pos} : \mathrm{RealVector} \end{bmatrix}$, then

$$\kappa_{right}(w, t, r) = \begin{cases} \mathrm{right}(r.\mathrm{x}) & \text{if } (r.\mathrm{sr}_{pos} \cdot w) > t \\ \neg\, \mathrm{right}(r.\mathrm{x}) & \text{otherwise} \end{cases}$$

As output we get a record type containing either a ptype right(x) or its negation, ¬ right(x). Larsson (2013) proposes that readings from sensors may count as proofs of such ptypes. A classifier can be used for judging x as being of a particular type on the grounds of perceptual information. A perceptual proof for right(x) would thus include the output from the position sensor that is directed towards x. Here, this output would be the space coordinates of x.

3 Vague Scalar Predicates

Scalar predicates such as '*tall*', '*long*' and '*expensive*', also called "relative gradable adjectives" (Kennedy, 2007), are interpreted with respect to a

[1] We are here assuming that we have a definition of dot product for TTR vectors a:RealVector$_n$ and b:RealVector$_n$ such that $a \cdot b = \Sigma_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n$. We also implicitly assume that the weight vector and the sensor reading vector have the same dimensionality.

scale, i.e., a dimension such as height, length, or cost along which entities for which the relevant dimension is applicable can be ordered. This makes scalar predicates compatible with degree morphology, like comparative and superlative morphemes ('*taller than*', '*the longest*') and intensifier morphemes such as '*very*' or '*quite*'. In this paper, our focus is on the so-called *positive form* of these adjectives (e.g. '*tall*' as opposed to '*taller*' or '*tallest*').

A property that distinguishes the positive form from the comparative and the superlative forms is its *context-dependance*. To take a common example: If Sue's height is 180cm, she may be appropriately described as a tall woman, but probably not as a tall basketball player. Thus, what counts as tall can vary from context to context, with the most relevant contextual parameter being a *comparison class* relative to which the adjective is interpreted (e.g., the set of women, the set of basketball players, etc.). In addition to being context-dependent, positive-form scalar predicates are also *vague*, in the sense that they give rise to *borderline cases*, i.e., entities for which it is unclear whether the predicate holds or not.

Vagueness is certainly a property that affects most natural language expressions, not only scalar adjectives. However, scalar adjectives have a relatively simple semantics (they are often uni-dimensional) and thus constitute a perfect case-study for investigating the properties and effects of vagueness on language use. Gradable adjectives have received a high amount of attention in the formal semantics literature. It is common to distinguish between two main approaches to their semantics: delineation-based and degree-based approaches. The delineation approach is associated with the work of Klein (1980), who proposes that gradable adjectives denote partial functions dependent on a comparison class. They partition the comparison class into three disjoint sets: a positive extension, a negative extension, and an extension gap (entities for which the predicate is neither true nor false). In contrast, degree-based approaches assume a measure function m mapping individuals x to degrees on a particular scale (degrees of height, degrees of darkness, etc.) and a *standard of comparison* or *degree threshold* θ (again, dependent on a comparison class) such that x belongs to the adjective's denotation if $m(x) > \theta$ (Kamp, 1975; Pinkal, 1979; Pinkal, 1995; Barker, 2002; Kennedy and McNally, 2005; Kennedy, 2007; Solt, 2011; Lassiter, 2011).

We build on degree approaches but adopt a perception-based perspective and take a step further to formalise how the meaning of these predicates can be learned and constantly updated through language use.

4 A Perceptual Semantics for '*Tall*'

To exemplify our approach, we will use the scalar predicate '*tall*' throughout.

4.1 Context-sensitivity

We first focus on capturing the context-dependence of relative scalar predicates. For this we define a type T_{ctxt} as follows:

(4) $\mathrm{T}_{ctxt} = \begin{bmatrix} \mathrm{c} & : & \mathrm{Type} \\ \mathrm{x} & : & \mathrm{c} \\ \mathrm{h} & : & \mathbb{R}^+ \end{bmatrix}$

The *context* (*ctxt*) of a scalar predicate like '*tall*' is a record of the type in (4), which includes: a type c (typically a subtype of Ind) representing the comparison class; an individual x within the comparison class (the argument of tall); a perceived measure on the relevant scale(s), in this case the perceived height h of x expressed as a positive real number.

The context presupposes the acquisition of sensory input from the environment. In particular, it assumes that an agent using such a representation is able to classify the entity in focus x as being of type c and is able to use some height sensor to obtain an estimate of x's height (the value of h is the sensor reading). We thus forgo the inclusion of an abstract measure function in the representation. In an artificial agent, this may be accomplished by image processing software for detecting and measuring objects in a digital image.

Besides the *ctxt*, we also assume a standard threshold of tallness θ_{tall} of the type given in (5). θ_{tall} is a function from a type specifying a comparison class to a height value, which corresponds to a tallness threshold for that comparison class. (In Section 5 we will discuss how such a threshold may be computed.)

(5) $\theta_{tall} : \mathrm{Type} \rightarrow \mathbb{R}^+$

The meaning of '*tall*' involves a *classifier* for tallness, κ_{tall}, of the following type:

(6) $\kappa_{tall} : (\mathrm{Type} \rightarrow \mathbb{R}^+, \mathrm{T}_{ctxt}) \rightarrow \mathrm{Type}$

We define this classifier as a one-input perceptron that compares the perceived height h of an individual x to the relevant threshold θ determined by a comparison class c. Thus, if θ : Type $\rightarrow \mathbb{R}^+$ and r : T_{ctxt}, then:

$$\kappa_{tall}(\theta, r) = \begin{cases} \text{tall}(r.\text{x}) & \text{if } r.\text{h} > \theta(r.\text{c}) \\ \neg\text{tall}(r.\text{x}) & \text{otherwise} \end{cases}$$

Simplifying somewhat, we can represent the meaning of '*tall*', **tall**, as a record specifying the type of context (T_{ctxt}) where an utterance of '*tall*' can be made, the parameter of the tallness classifier (the threshold θ), and a function f which is applied to the context to produce the content of '*tall*'.

(7)

$$\textbf{tall} = \begin{bmatrix} \mathrm{T}_{ctxt} = \begin{bmatrix} \text{c} & : & \text{Type} \\ \text{x} & : & \text{c} \\ \text{h} & : & \mathbb{R}^+ \end{bmatrix} \\ \theta \quad = \theta_{tall} \\ \text{f} \quad = \lambda r : \mathrm{T}_{ctxt}. \\ \quad \begin{bmatrix} \text{sit} = r \\ \text{sit-type} = \begin{bmatrix} \text{c}_{tall} : \kappa_{tall}(\theta, r) \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The output of the function f is an Austinian proposition (Cooper, 2005b): a judgement that a situation (sit, represented as a record r of type T_{ctxt}), is of a particular type (specified in sit-type). In the case of **tall**, the context of utterance (which instantiates r) is judged to be of the type where there is an individual x which is either tall or not tall, according to the output of the classifier κ_{tall}. The context of utterance in the sit field will include the height-sensor reading, which means that the sensor reading is part of the proof of the sit-type indicating that x is tall (or not, as the case may be).

Thus, to decide whether to refer to some individual x as tall or to evaluate someone else's utterance describing x as tall, an agent applies the function **tall**.f to the current situation, represented as a record $r : T_{ctxt}$. As an example, let us consider a situation that includes the context in (8), resulting from observing John Smith as being 1.88 meters tall (assuming this is our scale of tallness):

$$(8) \quad ctxt = \begin{bmatrix} \text{c} & = & \text{Human} \\ \text{x} & = & \text{john\_smith} \\ \text{h} & = & 1.88 \end{bmatrix}$$

Let us assume that given the comparison class Human, $\theta_{tall}(\text{Human}) = 1.87$. In this case, **tall**.f(*ctxt*) will compute as shown in (9). The resulting Austinian proposition corresponds to the agent's judgement that the situation in sit is one where John Smith counts as tall.

$$(9) \quad \lambda r : \mathrm{T}_{ctxt}. \begin{bmatrix} \text{sit} = r \\ \text{sit-type} = \begin{bmatrix} \text{c}_{tall} : \kappa_{tall}(\theta_{tall}, r) \end{bmatrix} \end{bmatrix}$$

$$\left(\begin{bmatrix} \text{c} & = & \text{Human} \\ \text{x} & = & \text{john\_smith} \\ \text{h} & = & 1.88 \end{bmatrix} \right) =$$

$$\begin{bmatrix} \text{sit} = \begin{bmatrix} \text{c} & = & \text{Human} \\ \text{x} & = & \text{john\_smith} \\ \text{h} & = & 1.88 \end{bmatrix} \\ \text{sit-type} = \begin{bmatrix} \text{c}_{tall} : \text{tall(john\_smith)} \end{bmatrix} \end{bmatrix}$$

4.2 Vagueness

According to the above account, '*tall*' has a precise interpretation: given a degree of height and a comparison class, the threshold sharply determines whether tall applies or not. There are several ways in which one can account for vagueness—amongst others, by introducing perceptual uncertainty (possibly inaccurate sensor readings). Here, in line with Lassiter (2011), we opt for substituting the precise threshold with a noisy, probabilistic threshold. We consider the threshold to be a normal random variable, which can be represented by the parameters of its Gaussian distribution, the mean μ and the standard deviation σ (the noise width).[2]

To incorporate this modification into our approach, we update the tallness classifier κ_{tall} we had defined in (6) so that it now takes as parameters μ_{tall} and σ_{tall}, both of them dependent on the comparison class and hence of type Type $\rightarrow \mathbb{R}^+$. The output of the classifier is now a probability rather than a ptype such as $\mathrm{tall}(x)$ or $\neg\mathrm{tall}(x)$. Before indicating how this probability is computed, we give the type of the vague version of the classifier in (10) and the vague representation of the meaning of '*tall*' in (11).

$$(10) \quad \kappa_{tall} : (\text{Type} \rightarrow \mathbb{R}^+, \text{Type} \rightarrow \mathbb{R}^+, \mathrm{T}_{ctxt}) \rightarrow [0, 1]$$

(11)

$$\textbf{tall} = \begin{bmatrix} \mathrm{T}_{ctxt} = \begin{bmatrix} \text{c} & : & \text{Type} \\ \text{x} & : & \text{c} \\ \text{h} & : & \mathbb{R}^+ \end{bmatrix} \\ \mu \quad = \mu_{tall} \\ \sigma \quad = \sigma_{tall} \\ \text{f} \quad = \lambda r : \mathrm{T}_{ctxt}. \\ \quad \begin{bmatrix} \text{sit} = r \\ \text{sit-type} = \begin{bmatrix} \text{c}_{tall} : \text{tall}(r.\text{x}) \end{bmatrix} \\ \text{prob} = \kappa_{tall}(\sigma, \mu, r) \end{bmatrix} \end{bmatrix}$$

[2]Which noise function may be the most appropriate is an empirical question we do not tackle in this paper. Our choice of Gaussian noise follows Schmidt et al. (2009)—see Section 5.1.

The output of the function **tall**.f is now a *probabilistic* Austinian proposition (Cooper et al., 2014). Like before, the proposition expresses a judgement that a situation sit is of a particular type. But here the judgement is probabilistic—it encodes the belief of an agent concerning the likelihood that sit is of a type where x counts as tall.

Since we take the noisy threshold to be a normal random variable, given a particular μ and σ, we can calculate the probability that the height r.h of individual r.x counts as tall as follows:

$$\kappa_{tall}(\mu, \sigma, r) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{r.\mathrm{h} - \mu(r.\mathrm{c})}{\sigma(r.\mathrm{c})\sqrt{2}}\right)\right]$$

Here erf is the error function, defined as[3]

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{t=0}^{x} e^{-t^2} \mathrm{d}t$$

The error function defines a sigmoid shape (see Figure 1), in line with the upward monotonicity of '*tall*'. The output of $\kappa_{tall}(\mu, \sigma, r)$ corresponds to the probability that h will exceed the normal random threshold with mean μ and deviation σ.

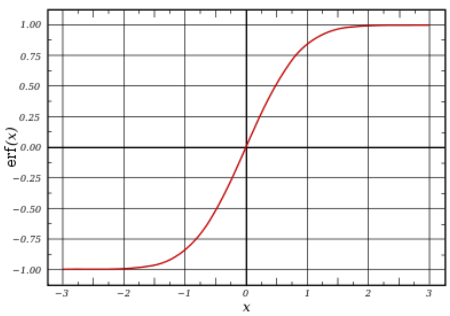

Figure 1: Plot of the error function.

Let us consider an example. Assume that we have $\mu_{tall}(\mathrm{Human}) = 1.87$ and $\sigma_{tall}(\mathrm{Human}) = 0.05$ (see Section 5.1 below for justification of the latter value). Let's also assume the same *ctxt* as above in (8). In this case, **tall**.f(*ctxt*) will compute as in (12), given that

$$\kappa_{tall}(\mu_{tall}, \sigma_{tall}, \begin{bmatrix} \mathrm{c{=}Human} \\ \mathrm{x{=}john\_smith} \\ \mathrm{h{=}1.88} \end{bmatrix}) =$$

$$\frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{1.88 - 1.87}{0.05\sqrt{2}}\right)\right] = 0.579$$

[3]For an explanation of this standard definition, see `http://en.wikipedia.org/wiki/Error_function`, which is the source of the graph in Figure 1.

(12) $\lambda r : \mathrm{T}_{ctxt}.\ \begin{bmatrix} \mathrm{sit} = r \\ \mathrm{sit\text{-}type} = \begin{bmatrix} \mathrm{c}_{tall} : \mathrm{tall}(r.\mathrm{x}) \end{bmatrix} \\ \mathrm{prob} = \kappa_{tall}(\mu_{tall}, \sigma_{tall}, r) \end{bmatrix}$

$$(\begin{bmatrix} \mathrm{c} & = & \mathrm{Human} \\ \mathrm{x} & = & \mathrm{john\_smith} \\ \mathrm{h} & = & 1.88 \end{bmatrix}) =$$

$$\begin{bmatrix} \mathrm{sit} = \begin{bmatrix} \mathrm{c} & = & \mathrm{Human} \\ \mathrm{x} & = & \mathrm{john\_smith} \\ \mathrm{h} & = & 1.88 \end{bmatrix} \\ \mathrm{sit\text{-}type} = \begin{bmatrix} \mathrm{c}_{tall} : \mathrm{tall(john\_smith)} \end{bmatrix} \\ \mathrm{prob} = 0.579 \end{bmatrix}$$

This probability can now be used in further probabilistic reasoning, to decide whether to refer to an individual x as tall, or to evaluate someone else's utterance describing x is tall. For example, an agent may map different probabilities to different adjective qualifiers of tallness to yield compositional phrases such as '*sort of tall*', '*quite tall*', '*very tall*', '*extremely tall*', etc. The meanings of these composed adjectival phrases could specify probability ranges trained independently. Compositionality for vague perceptual meanings, and the interaction between compositionality and learning, is an exciting area for future research.[4]

5 Learning from Language Use

In this section we consider possibilities for computing the noisy threshold we have introduced in the previous section and discuss how such a threshold and the probabilistic judgements it gives rise to are updated with language use.

5.1 Computing the Noisy Threshold

We assume that agents keep track of judgements made by other agents. More concretely, for a vague scalar predicate like '*tall*', we assume that an agent will have at its disposal a set of *observations* consisting of entities of a particular type T (a comparison class such as *Human*) that have been judged to be tall, together with their observed heights. Judgements of tallness may vary across individuals—indeed, such variation (both inter- and intra-individual) is a hallmark of vague predicates. We use Ω^{T}_{tall} to refer to the set of heights of those entities $x : T$ that have been considered tall by some individual. From this agent-specific set of observations, which is constantly updated as the agent is exposed to new judgements by other individuals, we want to compute a noisy threshold,

[4]See Larsson (2013) for a sketch of compositionality for perceptual meaning.

which the agent uses to make her own judgements of tallness, as specified in (11).

Different functions can be used to compute μ_{tall} and σ_{tall} from Ω^T_{tall}. What constitutes an appropriate function is an empirical matter and what the most suitable function is possibly varies across predicates (what may apply to '*tall*' may not be suitable for '*dark*' or '*expensive*', for example). Hardly any work has been done on trying to identify how the threshold is computed from experience. A notable exception, however, is the work of Schmidt et al. (2009), who collect judgements of people asked to indicate which items are tall given distributions of items of different heights. Schmidt and colleagues then propose different probabilistic models to account for the data and compare their output to the human judgements. They explore two types of models: threshold-based models and category-based or cluster models. The best performing models within these two types perform equally well and the study does not identify any advantages of one type over the other one. Since we have chosen threshold models as our case-study, we focus our attention on those here.

Each of the threshold models tested by Schmidt et al. (2009) corresponds to a possible way of computing the mean μ_{tall} of a noisy threshold from a set of observations. The best performing threshold model in their study is the *relative height by range* model, where (in our notation):

(13) relative height by range (RH-R): $\mu_{tall}(T) = \max(\Omega^T_{tall}) - k \cdot (\max(\Omega^T_{tall}) - \min(\Omega^T_{tall}))$

Here $\max(\Omega^T_{tall})$ and $\min(\Omega^T_{tall})$ stand for the maximum and the minimum height, respectively, of the items that have been judged to be tall by some individual. According to this threshold model, any item within the top $k\%$ of the range of heights that have been judged to be tall counts as tall. The model includes two parameters, k and a noise-width parameter that in our approach corresponds to σ_{tall}. Schmidt et al. (2009) report that the best fit of their data was obtained with $k = 29\%$ and $\sigma_{tall} = 0.05$.

5.2 Updating Vague Meanings

We now want to specify how the vague meaning of '*tall*' is updated as an agent is exposed to new judgements via language use. Our setting so far offers a straightforward solution to this: If a new entity $x : T$ with height h is referred to as tall, the agent adds h to its set of observations Ω^T_{tall} and recomputes μ_{tall}(Human), for instance using RH-R as defined in (13). If RH-H is used, ideally the value of k and σ_{tall} should be (re)estimated from Ω^T_{tall}. For the sake of simplicity, however, here we will assume that these two parameters take the values experimentally validated by Schmidt et al. (2009) and are kept constant. An update to μ_{tall} will take place if it is the case that $h > \max(\Omega^T_{tall})$ or $h < \min(\Omega^T_{tall})$. This in turn will trigger un update to the probability outputted by κ_{tall}.

As an example, let us assume that our initial set of observations is $\Omega^{Human}_{tall} = \{1.87, 1.92, 1, 90, 1.75, 1.80\}$ (recall this corresponds to the perceived heights of individuals that have been described as tall by some agent). This means that $\max(\Omega^{Human}_{tall}) = 1.92$ and $\min(\Omega^{Human}_{tall}) = 1.75$. Hence, given (13):

(14) $\mu_{tall}(\text{Human}) =$
$1.92 - 0.29 \cdot (1.92 - 1.75) = 1.87$

Let's assume we now make an observation where a person of height 1.72 is judged to be tall. This will mean that the set of observations is now $\Omega^{Human}_{tall} = \{1.87, 1.92, 1, 90, 1.75, 1.80, 1.72\}$ and consequently $\min(\Omega^{Human}_{tall}) = 1.72$, which yields an updated mean of the noisy threshold:

(15) $\mu_{tall}(\text{Human}) =$
$1.92 - 0.29 \cdot (1.92 - 1.72) = 1.862$

If we were to re-evaluate John Smith's tallness in light of this observation, we would get a new probability 0.64 that he is tall (in contrast to the earlier probability of 0.579 given in (12)).

5.3 Possible Extensions

The set of observations Ω^{Human}_{tall} can be derived from a set of Austinian propositions corresponding to instances where people have been judged to be tall. To update from an Austinian proposition p we simply add p.sit.h to Ω^{tall}_{Human} and recompute $\mu_{tall}(p.\text{c})$. Note that we are here treating these Austinian propositions as non-probabilistic. This seems to make sense since an addressee does not have direct access to the probability associated with the judgement of the speaker. If we were to take these probabilities into account (for instance, the use of a hedge in '*sort of tall*' may be used to make inferences about such probabilities), and if those probabilities are not always 1, we would need a different way of computing μ_{tall} than the one specified so far.

Somewhat related to the point above, note that in our approach we treat all judgements equally, i.e., we do not distinguish between possible different levels of trustworthiness amongst speakers. An agent who is told that an entity with height h is tall adds that observation to its knowledge base without questioning the reliability of the speaker. This is clearly a simplification. For instance, there is developmental evidence showing that children are more sensitive to reliable speakers than to unreliable ones during language acquisition (Scofield and Behrend, 2008).

6 Other Approaches

Within the literature in formal semantics, Lassiter (2011) has put forward a proposal that extends in interesting ways earlier work by Barker (2002) and shares some aspects with the account we have presented here. Operating in a probabilistic version of classical possible-worlds semantics, Lassiter assumes a probability distribution over a set of possible worlds and a probability distribution over a set of possible languages. Each possible language represents a precise interpretation of a predicate like '*tall*': $\mathbf{tall}_1 = \lambda x.x$'s height $\geq$ 5'6"; $\mathbf{tall}_2 = \lambda x.x$'s height $\geq$ 5'7"; and so forth. Lassiter thus treats "metalinguistic belief" (representing an agent's knowledge of the meaning of words) in terms of probability distributions over precise languages. Since each precise interpretation of '*tall*' includes a given threshold, this can be seen as defining a probability distribution over possible thresholds, similarly to the noisy threshold we have used in our account. Lassiter, however, is not concerned with learning.

Within the computational semantics literature, DeVault and Stone (2004) describe an implemented system in a drawing domain that is able to interpret and execute instructions including vague scalar predicates such as '*Make a small circle*'. Their approach makes use of degree-based semantics, but does not take into account comparison classes. This is possible in their drawing domain since the kind of geometric figures it includes (squares, rectangles, circles) do not have intrinsic expected properties (size, length, etc). Their focus is on modelling how the threshold for a predicate such as '*small*' is updated during an interaction with the system given the local discourse context. For instance, if the initial context just contains a square, the size of that square is taken to be the standard of comparison for the predicate '*small*'. The user's utterance '*Make a small circle*' is then interpreted as asking for a circle of an arbitrary size that is smaller than the square.

In our characterisation of the context-sensitivity of vague gradable adjectives in Section 4.1, we have focused on their dependence on general comparison classes corresponding to types of entities (such as Human, Woman, etc) with expected properties such as height. Thus, in contrast to DeVault and Stone (2004), who focus on the local context of discourse, we have focused on what could be called the *global* context (an agent's experience regarding types of entities and their expected properties). How these two types of context interact remains an open question, which we plan to explore in our future work (see Kyburg and Morreau (2000), Kemp et al. (2007), and Fernández (2009) for pointers in this direction).

7 Conclusions and future work

Traditional formal semantics theories postulate a fixed, abstract interpretation function that mediates between natural language expressions and the world, but fall short of specifying how this function is determined or modified dynamically by experience. In this paper we have presented a characterisation of the semantics of vague scalar predicates such as '*tall*' that clarifies how their context-dependent meaning and their vague character are connected with perceptual information, and we have also shown how this low-level perceptual information (here, real-valued readings from a height sensor) connects to high level logical semantics (ptypes) in a probabilistic framework. In addition, we have put forward a proposal for explaining how the meaning of vague scalar adjectives like '*tall*' is dynamically updated through language use.

Tallness is a function of a single value (height), and is in this sense a uni-dimensional predicate. Indeed, most linguistic approaches to vagueness focus on uni-dimensional predicates such as '*tall*'. However, many vague predicates are *multi-dimensional*, including nouns for positions ('*above*'), shapes ('*hexagonal*'), and colours ('*green*'), amongst many others. Together with compositionality (mentioned at the end of Section 4.2), generalisation of the present account to multi-dimensional vague predicates is an interesting area of future development.

Acknowledgements

The first author acknowledges the support of the Netherlands Organisation for Scientific Research (NWO) and thanks the Centre for Language Technology at the University of Gothenburg for generously funding research visits that led to the work presented in this paper. The second author acknowledges the support of Vetenskapsrådet, project 2009-1569, Semantic analysis of interaction and coordination in dialogue (SAICD); the Department of Philosophy, Linguistics, and Theory of Science; and the Centre for Language Technology at the University of Gothenburg.

References

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

Patrick Blackburn and Johan Bos. 2005. *Representation and Inference for Natural Language: A First Course in Computational Semantics*. CSLI Publications.

Robin Cooper and Jonathan Ginzburg. 2011. Negation in dialogue. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011)*, Los Angeles (USA).

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2014. A probabilistic rich type theory for semantic interpretation. In *Proceedings of the EACL Workshop on Type Theory and Natural Language Semantics (TTNLS)*.

Robin Cooper. 2005a. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3(4):333–362, December.

Robin Cooper. 2005b. Austinian truth, attitudes and type theory. *Research on Language and Computation*, 3:333–362.

Robin Cooper. 2010. Generalized quantifiers and clarification content. In Paweł Łupkowski and Matthew Purver, editors, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, Poznań. Polish Society for Cognitive Science.

Robin Cooper. 2012. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics. Elsevier BV. General editors: Dov M. Gabbay, Paul Thagard and John Woods.

David DeVault and Matthew Stone. 2004. Interpreting vague utterances in context. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING'04)*, pages 1247–1253.

Simon Dobnik, Robin Cooper, and Staffan Larsson. 2013. Modelling language, action, and perception in type theory with records. In *Constraint Solving and Language Processing*, Lecture Notes in Computer Science, pages 70–91. Springer.

Raquel Fernández. 2009. Salience and feature variability in definite descriptions with positive-form vague adjectives. In *Workshop on the Production of Referring Expressions: Bridging the gap between computational and empirical approaches to reference (CogSci'09)*.

Jonathan Ginzburg. 2012. *The Interactive Stance*. Oxford University Press.

Hans Kamp. 1975. Two theories of adjectives. In E. Keenan, editor, *Formal Semantics of Natural Language*, pages 123–155. Cambridge University Press.

John Kelleher, Fintan Costello, and Josef van Genabith. 2005. Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context. *Artificial Intelligence*, 167(1):62–102.

Charles Kemp, Amy Perfors, and Joshua B. Tenenbaum. 2007. Learning overhypotheses with hierarchical bayesian models. *Developmental Science*, 10(3):307–321.

Christopher Kennedy and Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, pages 345–381.

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

Ewan Klein. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, 4:1–45.

Alice Kyburg and Michael Morreau. 2000. Fitting words: Vague language in context. *Linguistics and Philosophy*, 23:577–597.

Staffan Larsson. 2009. Detecting and learning from lexical innovation in dialogue: a ttr account. In *Proceedings of the 5th International Conference on Generative Approaches to the Lexicon*.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*.

Dan Lassiter. 2011. Vagueness as probabilistic linguistic knowledge. In R. Nowen, R. van Rooij, U. Sauerland, and H. C. Schmitz, editors, *Vagueness in Communication*. Springer.

Barbara Partee. 1989. Possible worlds in model-theoretic semantics: A linguistic perspective. In S. Allen, editor, *Possible Worlds in Humanities, Arts and Sciences*, pages 93–123. Walter de Gruyter.

Manfred Pinkal. 1979. Semantics from different points of view. In R. Bäurle, U. Egli, and A. von Stechow, editors, *How to Refer with Vague Descriptions*, pages 32–50. Springer-Verlag.

Manfred Pinkal. 1995. *Logic and lexicon: the semantics of the indefinite*, volume 56 of *Studies in Linguistics and Philosophy*. Springer.

François Portet, Ehud Reiter, Albert Gatt, Jim Hunter, Somayajulu Sripada, Yvonne Freer, and Cindy Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7):789–816.

Matthew Purver, Julian Hough, and Eleni Gregoromichelaki. 2014. Dialogue and compound contributions. In A. Stent and S. Bangalore, editors, *Natural Language Generation in Interactive Systems*. Cambridge University Press.

Ehud Reiter, Somayajulu Sripada, Jim Hunter, Jin Yu, and Ian Davy. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, 167(1):137–169.

Deb Roy. 2005. Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, 167(1):170–205.

L.A. Schmidt, N.D. Goodman, D. Barner, and J.B. Tenenbaum. 2009. How tall is tall? compositionality, statistics, and gradable adjectives. In *Proceedings of the 31st annual conference of the cognitive science society*.

Jason Scofield and Douglas A Behrend. 2008. Learning words from reliable and unreliable speakers. *Cognitive Development*, 23(2):278–290.

Jeffrey Mark Siskind. 2001. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *Journal of Artificial Intelligence Research*, (15):31–90.

Danijel Skocaj, M Janicek, Matej Kristan, Geert-Jan M Kruijff, Aleš Leonardis, Pierre Lison, Alen Vrecko, and Michael Zillich. 2010. A basic cognitive system for interactive continuous learning of visual concepts. In *Proceeding of the Workshop on Interactive Communication for Autonomous Intelligent Robots*, pages 30–36.

Stephanie Solt. 2011. Notes on the comparison class. In *Vagueness in communication*, pages 189–206. Springer.

Luc Steels. 2003. Evolving grounded communication for robots. *Trends in cognitive sciences*, 7(7):308–312.