Automated Reasoning on Vague Concepts using Formal Ontologies, with an Application to Event Detection on Video Data*

Tommaso D'Odorico and **Brandon Bennett**
School of Computing
University of Leeds
Leeds (UK)
{sc07td, b.bennett}@leeds.ac.uk

Abstract

In this paper we lay the foundations of a formal ontology for the characterisation of vague concepts sourced from natural language, applying principles of event calculus and supervaluation semantics. We focus on a specific set of motion events and related concepts, motivated by the aim to develop an automated reasoning system able to detect occurrences of such events in video scenes. Our goal is to provide a general methodology for the formalisation of vague concepts, and to address the issue of vagueness in formal ontologies.

1 Introduction

Automated reasoning on natural language concepts is often problematic due to vagueness, an issue concerning ambiguities in the meaning of terms. In this paper, we present a formalism for the characterisation of vague concepts in ontologies, with an application to a particular reasoning task.

The motivation for this approach comes from our interest in vagueness and formal ontologies, and the involvement in Darpa Mind's Eye Challenge[1], whose goal is to automatically detect temporal occurrences of a specific set of 48 motion verbs in video scenes, and for which we are developing an automated logical reasoning system (D'Odorico and Bennett 2012). The core of this system is constituted by an ontology formally characterising the meaning of each concept relevant to our domain. The reasoning process starts with the grounding of the lower layer of the ontology with information obtained by pre-processing the video sequences, continues with the logical inference of mid- and higher-level predicates and culminates with the inference of predicates representing event occurrences in the scene.

Sec. 2 to follow provides a brief overview on vagueness and related work. The main issue our approach is addressing is the methodology with which to formulate a formal definition for the mid- and higher-level concepts in the ontology, often constituted by vague terms. For example, how would one formally define the vague concept *near*? The formalism is outlined in Sec. 3, with Sec 4 illustrating some examples of verb models and concept definitions. Finally, Sec. 5 closes with some considerations about the proposed approach.

2 Vagueness and Related Work

Vagueness is a phenomenon most ontologies struggle with when attemtping to formally define concepts sourced from natural language. Several concept classes are sources of vagueness, for example *spatial prepositions* (e.g. near, far, beside, close…); *adjectives* (e.g. tall, short, big, small, fast, slow…); *verbs* (e.g. approach, chase, exchange…) and *nouns* (e.g. group, hill, river…). Vagueness is different from uncertainty, arising from insufficient or imprecise knowledge, or generality, arising from lack of specificity.

Most of the foundational characterisations on origins and nature of vagueness can be found among the philosophical and logic communities (Fine 1975; Williamson 1994; 2003; Keefe and Smith 1997). More recently, interest has been growing within the computing and geography communities too (Varzi 2001; Bennett, Mallenby, and Third 2008; Galton 2009; Cintula et al. 2011). The ongoing debate has produced different and sometimes conflicting accounts of vagueness as intrinsic indeterminacy of things in the world (*de re* vagueness, (Tye 1990)), indeterminacy of linguistic expressions (*de dicto* vagueness, (Lewis 1986; Varzi 2001)), or indeterminacy of knowledge (*epistemic* vagueness, (Williamson 1992; 1994; Sorensen 2001)).

The work presented here is mostly concerned with the indeterminacy of linguistic expressions. In fact, most concepts defining event occurrences to be detected by an automatic reasoning system exhibit indeterminate applicability boundaries. For example, given a scene where an object a is moving towards a location b, detecting an occurrence of the event "a is arriving at b" involves the non-trivial task of establishing the applicability boundary of the concept 'arrive' to the given situation. One may formalise 'arrive' by specifying its meaning as "a is arriving at b if a is moving towards b, a is near b and a eventually stops at b", which in turn involves establishing a precise meaning for 'moving towards', 'near' and 'stop'. This simple example shows how quickly the problem can escalate even on narrow domains.

We adopt a pragmatical approach based on the *epistemic stance* (Lawry and Tang 2009), an elaboration on the epistemic characterisation of vagueness. This characterisation

*This work has been partially funded by the DARPA Mind's Eye Program (project VIGIL, W911NF-10-C-0083), whose support is gratefully acknowledged.

[1] www.darpa.mil/Our_Work/I2O/Programs/Minds_Eye.aspx

argues that there is an objectively correct set of criteria for precisely determining the applicability of a vague concept, but this set of criteria is unknown due to the uncertain and inconsistent meaning of words in natural language. The epistemic stance maintains that decision problems regarding assertions can find it useful to assume an epistemic view of vagueness and thus the existence of a clear dividing line between concept demarcations. The following types of vagueness (Bennett 2005) are the main factors affecting the individuation of such concept demarcations:

- *simple ambiguity* – a term has multiple meanings (e.g. the verb 'Pass' may mean "crossing a boundary" but also "handing an item");
- *sorites* or *threshold vagueness* – the applicability boundary of a concept is blurred and depends on the continuous variation of some observable properties of the sample to which the concept applies (e.g. the concept 'near' may be applicable to two objects a and b according to the distance separating them, for which establishing a *threshold* is non-trivial);
- *deep ambiguity* – there exist several overlapping clusters of observable properties that may establish an applicability boundary for a concept, yet it is unclear as to which ones are necessary or relevant (e.g. the fact that a is near b may be defined according to linear distance, time needed to travel from a to b or some other estimation).

This model of linguistic vagueness — assuming the existence of precise criteria determining the applicability boundary of concepts according to observable properties of objects — is suitable to the application of supervaluation semantics (Fine 1975; Keefe 2008). In this semantics, precise interpretations of vague predicates are expressed by *precisifications*. In our ontology, precisifications of vague concepts are modeled with explicit thresholds linked to observable properties relevant to the demarcation of the concept applicability boundary (see Sec. 3.4).

Despite not necessarily believing in an epistemic nature of vagueness, we do believe in the practical utility of an artificial agent behaving as if the epistemic view was correct. This underlying assumption coupled with a supervaluationist approach has guided the development of the ontology outlined in the next section.

3 Ontology

Our ontology builds upon *Event Calculus* (Kowalski and Sergot 1986; Shanahan 1999) and *Versatile Event Logic* (VEL) (Bennett and Galton 2004), formalisms designed to reason about actions and events within logic. Given an ordered set of time points $\mathcal{T} = (T, <)$, the most interesting feature of the calculus is the possibility to express that propositional expression p holds at a particular time point $t \in \mathcal{T}$, through the construct $\mathsf{HoldsAt}(p, t)$.

The purpose of our formalism is to describe real world situations, mainly objects, their properties and event occurrences. However, the task of automated event detection in which this ontology will be employed presents a few peculiar aspects bearing an influence to some of the design choices outlined in the remainder of this section.

Firstly, a computer system can only operate on a representation of the real world, and not on the real world itself. This observation led us to separate the *appearance* of things in the world, stemming from the representation, from their *evidence*, stemming from reality. This constitutes the low-level layer of the ontology, concerned with "what we know about the world" and is examined in more detail in Sec. 3.2.

Secondly, higher-level concepts that we would like to infer, and representing "what we can understand about the world", have different levels of complexity. This influenced the structuring of the ontology in a mid-level layer, concerned with objects' description, and a high-level layer, concerned with complex situations such as processes and events.

3.1 Logical Formalism

The vocabulary of our logical language can be specified by the tuple:

$$\mathcal{V} = \langle \mathcal{T}, \mathcal{O}, \mathcal{A}, \mathcal{D}, \mathcal{M}, \mathcal{Q}, \mathcal{PT}, \mathcal{F}, \mathcal{E}, \Sigma \rangle$$

where:

- $\mathcal{T}$ is the set of ordered time points (e.g. $\mathcal{T} = \{t_1, t_2 \ldots\}$);
- $\mathcal{O}$ is the set of objects;
- $\mathcal{A}$ is the set of *appearance-types* (Sec.3.2);
- $\mathcal{D}$ is the set of *evidence-types* (Sec. 3.2);
- $\mathcal{M}$ is the set of *measure-types* (Sec. 3.3);
- $\mathcal{Q}$ is the set of *quality-types* (Sec. 3.3);
- $\mathcal{PT}$ is the set of *precisification thresholds* (Sec. 3.4);
- $\mathcal{F}$ is the set of *fluents* (Sec. 3.5);
- $\mathcal{E}$ is the set of *event-types* (Sec. 3.5);
- Σ is the set of *event-tokens* (Sec. 3.5).

A formula of the type $\mathsf{HoldsAt}(p, t)$ expresses that p holds at time point t. In the following, this notation will be abbreviated to $[\![p]\!]_{[t]}$. For some propositions, time-indexing may not be relevant as they may hold regardless of a specific time point. These will be simply stated as p, equivalent to $\forall t \in \mathcal{T}\ [\![p]\!]_{[t]}$.

3.2 Appearances and Evidences

It has been pointed out that a computer system can only look at a representation of reality, which is what we call *appearance*. Appearances may result from algorithmic processing (e.g. a computer vision algorithm processing a video scene) or manual annotation by human observers of the scene. Dissimilarities in granularity and precision among different appearances can be huge; in any case they are always a partial, finite and incomplete representation of the real world and constitute all that is available for a computer system to ground the ontology. Thus the appearance represents everything the ontology knows about the world.

The set $\mathcal{A}$ of *appearance-types* represents the domain of possible appearances. Each appearance-type $(p, O_p, D_p) \in \mathcal{A}$ is a triple where p is the appearance predicate, O_p and D_p are respectively the object and description domain of p. A particular instance of such appearance-type is expressed by $[\![\mathsf{a}(p(o, d))]\!]_{[t]}$, for $o \in O_d$ and $d \in D_d$. An example of appearance-type can be the position of an object as a point on a two-dimensional discrete grid, formally $(position, \mathcal{O}, \mathbb{N} \times \mathbb{N})$. The actual appearance describing position of object o at time t on coordinate (x, y) is given by $[\![\mathsf{a}(position(o,(x,y)))]\!]_{[t]}$.

It has been said that an appearance represents everything the ontology knows about the world. However, some details about the world missing from the appearance may be logically inferred on the appearance itself, thus enriching the representation forming what we call the *evidence*. The set of *evidence-types* $\mathcal{D}$ represents the domain of such evidences, and has the same structure of $\mathcal{A}$. Given evidence-type $(p, O_d, D_d) \in \mathcal{D}$, a particular evidence is expressed by $[\![\mathsf{d}(p(o,d))]\!]_{[t]}$ for $o \in O_d$ and $d \in D_d$.

An example is given by occlusion: for instance, a person walking in space and passing behind an object. A set of appearances returned by an average tracking algorithm would represent the person's position as a bounding box at each time instant, excluding the instants in which the person is behind the object. The inferrable evidence from such an appearance is constituted by the information standing between the real world and the appearance. Even a very sophisticated ontology is unlikely to infer precise details such as the person's shape, posture and positioning. However, it is conceivable that an ontology may successfully infer the bounding boxes corresponding to the person's position whilst occluded by the wall, or infer the position of the person's hands and feet from the bounding box (and other details useful for specific event occurrences).

We call this layer of the ontology *Theory of Appearances* whose aim, as in the example above, is to augment the knowledge given by appearances. A theory may also discard 'wrong' appearances, for example spurious bounding boxes resulting from tracking algorithms.

3.3 Measures and Qualities

The theory of appearances described above is made up of low-level predicates providing precise, concrete and mostly quantitative information about objects. Measures and qualities populate the mid-level layer of the ontology, bridging the gap between quantitative/precise evidences and qualitative/vague abstractions.

A measure is a simple quantitative abstraction, generally inferred from evidences. A *measure-type* $(p, O_p, M_p) \in \mathcal{M}$ is a triple where m is the measure predicate relating objects in O_p with measure-values in M_p. Given $t \in \mathcal{T}$, $o \in O_p$ and $v \in M_p$, a particular instance of such a measure-type is expressed by $[\![\mathsf{m}(p(o,v))]\!]_{[t]}$ as in the examples below.

- $(pos, \mathcal{O}, \mathbb{R} \times \mathbb{R})$: position of an object as coordinate on a grid assuming that, for each object, its position is either given as evidence-type $(pos, \mathcal{O}, \mathbb{N} \times \mathbb{N}) \in \mathcal{D}$ or can be calculated as centroid of a bounding box (evidence-type $(bbox, \mathcal{O}, \mathbb{N}^4) \in \mathcal{D}$

$$
\begin{aligned}
&[\![\mathsf{m}(pos(o,(x,y)))]\!]_{[t]} \equiv [\![\mathsf{d}(pos(o,(x,y)))]\!]_{[t]} \vee \\
&\vee \Big[[\![\mathsf{d}(bbox(o,(b_x,b_y,b_w,b_h)))]\!]_{[t]} \wedge \\
&\wedge x = \frac{b_x + b_w}{2} \wedge y = \frac{b_y + b_h}{2} \Big]
\end{aligned}
\tag{1}
$$

- $(dist, \mathcal{O} \times \mathcal{O}, \mathbb{R})$: distance between objects $o_1, o_2 \in \mathcal{O}$

$$
\begin{aligned}
&[\![\mathsf{m}(dist((o_1,o_2),d))]\!]_{[t]} \equiv [\![\mathsf{m}(pos(o_1,(x_1,y_1)))]\!]_{[t]} \wedge \\
&\wedge [\![\mathsf{m}(pos(o_2,(x_2,y_2)))]\!]_{[t]} \wedge d = edist((x_1,y_1),(x_2,y_2))
\end{aligned}
\tag{2}
$$

where $edist : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is a function calculating the euclidean distance between two points.

A quality is a richer, qualitative and often vague abstraction inferred from evidences, measures and other qualities. A *quality-type* $(p, O_p, Q_p) \in \mathcal{Q}$ is a triple where p is the quality predicate relating objects in O_p with quality-values in Q_p. Given $t \in \mathcal{T}$, $o \in O_p$ and $v \in Q_p$, a particulare instance of such quality-type is expressed by $[\![\mathsf{q}(p(o,v))]\!]_{[t]}$. Below there are some examples of quality-types that are relevant to our domain:

- Type of an object: $(type, \mathcal{O}, \{\text{person}, \text{car}, \text{box}, \text{other}\})$;
- Size of an object: $(size, \mathcal{O}, \{\text{pocketSize}, \text{carrySize}, \text{manSize}, \text{largeSize}\})$;
- Speed of an object: $(speed, \mathcal{O}, \{\text{slow}, \text{walkSpeed}, \text{runningSpeed}, \text{fast}\})$;
- Relative positioning of one object with respect to another: $(relPosition, \mathcal{O} \times \mathcal{O}, \{\text{left}, \text{right}, \text{over}, \text{under}\})$;
- Topological relation between two objects (RCC-8 relation set (Randell, Cui, and Cohn 1992)): $(rccRel, \mathcal{O} \times \mathcal{O}, \{\text{DC}, \text{EC}, \text{PO}, \text{EQ}, \text{TPP}, \text{NTPP}, \text{TPP}^{-1}, \text{NTPP}^{-1}\})$
- Proximity of two objects: $(proximity, \mathcal{O} \times \mathcal{O}, \{\text{near}, \text{far}, \text{veryFar}\})$.

Contrary to most measures in $\mathcal{M}$, the ambiguous nature of qualities makes their formalisation difficult. This problem is addressed by precisifications, introduced below.

3.4 Precisifications

A formal method to establish whether a vague concept holds can be obtained through the ideas in Supervaluation Semantics (Fine 1975; Keefe 2008). In this theory, a formula may admit multiple models, each obtainable via an assignment of referents to terms and truth-values to predicates. Such an assignment is called a *precisification*, and allows to obtain a precise interpretation of a vague term. This approach preserves classical logic inference rules, hence it is preferred over multi-valued logics such as Fuzzy Logic (Zadeh 1975) for our task of ontology reasoning.

Supervaluation Semantics and the epistemic stance lead us to Standpoint semantics (Bennett 2011), an elaboration of supervaluation semantics where the precisification is explicitly embedded in the language syntax. Specifically, the precise criteria governing the extension of a concept's applicability boundary are modeled in terms of *applicability thresholds* for one or more observable properties.

The set of precisification thresholds $\mathcal{PT}$ is constituted by couples (t, V_t), where V_t is the range of admissible values for threshold t. A precisification P is an assignments of values to precisification thresholds, i.e. $P \subseteq \{(t, v_t) \mid (t, V_t) \in \mathcal{PT} \wedge v_t \in V_t\}$ (assuming $\nexists\, (t, v_1), (t, v_2) \in P \wedge v_1 \neq v_2$).

Precisifications are explicited in the language with the syntax demonstrated in the following example. The vague concept o_1 *is near* o_2 can be made precise by specifying a threshold on an observable property, such as the linear distance between a and b. We can define the proximity quality with value 'near' by introducing precisification threshold $(minNear, v) \in \mathcal{PT}$ and adding precisification P as parameter to the definition:

$$\begin{aligned}&[\![\mathsf{q}[P](proximity((o_1, o_2), \text{near}))]\!]_{[t]} \equiv \\ &\quad \exists(minNear, \delta) \in P \wedge [\![\mathsf{m}(dist((o_1, o_2), d))]\!]_{[t]} \wedge d < \delta\end{aligned} \tag{3}$$

The simple definition above states that the *proximity* quality near holds between objects o_1 and o_2 if their distance measure is smaller than the *minNear* threshold specified by precisification P.

3.5 Fluents, Processes and Events

We distinguish between two types of time-dependent formal expressions: propositional expressions whose validity can be stated over time (*fluents*) and expressions referring to temporal entities occurring over some interval (*events*).

A fluent's truth-value may be established at single time points. Fluents describe either a state that may hold or not hold, or a process that may be active or inactive at each time point. Given $[\![f]\!]_{[t]}$, we define $\mathsf{HoldsOver}(f, [t_1, t_2])$ and $\mathsf{HoldsOn}(f, [t_1, t_2])$ to express that f holds over the interval $[t_1, t_2]$:

$$\mathsf{HoldsOver}(f, [t_1, t_2]) \equiv \forall\, t\big[(t_1 \leq t \leq t_2) \rightarrow [\![f]\!]_{[t]}\big] \tag{4}$$

$$\begin{aligned}&\mathsf{HoldsOn}(f, [t_1, t_2]) \equiv \mathsf{HoldsOver}(f, [t_1, t_2]) \;\wedge \\ &\quad \wedge\, \exists\, t'_1\Big[t'_1 < t_1 \wedge \forall\, t'\big[t'_1 < t' < t_1 \rightarrow \neg[\![f]\!]_{[t']}\big]\Big] \wedge \\ &\quad \wedge\, \exists\, t'_2\Big[t_2 < t'_2 \wedge \forall\, t'\big[t_2 < t' < t'_2 \rightarrow \neg[\![f]\!]_{[t']}\big]\Big]\end{aligned} \tag{5}$$

If $\mathsf{HoldsOver}(f, [t_1, t_2])$ is true for some t_1, t_2, from (4) it follows that $\mathsf{HoldsOver}(f, [t_i, t_j])$ is also true, for every $t_1 < t_i < t_j < t_2$. Conversely, $\mathsf{HoldsOn}(f, [t_1, t_2])$ holds only on the greatest continuous temporal interval over which f is true, i.e. there are no t'_1, t'_2 such that $t'_1 < t_1, t_2 < t'_2$ and $\mathsf{HoldsOn}(f, [t'_1, t'_2])$. In the formulae to follow, the notation $\mathsf{HoldsOn}(f, [t_1, t_2])$ will be shortened in $[\![f]\!]_{[t_1, t_2]}$.

An event represents a complex action and we distinguish between *event-types* and *event-tokens* (Bennett and Galton 2004). An event-type $e \in \mathcal{E}$ is associated with a set of episodes of a particular event, for example: 'John approaches Mary', formalised as $\mathsf{Approach}(\mathrm{John}, \mathrm{Mary})$. An event-token $\sigma \in \Sigma$ constitutes an occurrence of a particular event-type over a temporal interval. To express occurrence of event type $e \in \mathcal{E}$ over time interval $i \in \mathcal{I}$ we introduce the construct $\mathsf{Occurs}(e, i)$. For clarity of notation in the formulae to follow, event-types are capitalised as in Approach.

4 Verb Models

This section illustrates some sample verb models applying the principles outlined in the previous section. Definitions in this ontology are not intended as exhaustive semantic characterisations of concepts, such as the ones that would be produced following a sistematic linguistic analysis. Our formalisation needs to strike a compromise between the complexity of meaning, the practical task of detecting occurrences of such concepts on a coarse and imprecise representation of the real world and the implementability of these definitions in Prolog, the logic-programming language of our automated reasoning system. For this reason, we cannot formulate very complex definitions, as they would either be too difficult to disambiguate, or impractical to break down and ground on the Theory of Appearances.

The formalisations to follow are intended as an illustration on how to apply such methodology. Most definitions are parameterised with precisification thresholds; a recurring one is the *detection window* T_w. Defining whether $[\![\mathsf{f}]\!]_{[t]}$ holds often involves examining whether other predicates hold at instants preceding and/or following t. Threshold T_w allows to quantify and discretise such span.

4.1 Verbs of Proximity

A few verbs in our set seem to involve the definition of the notion of proximity, notably Approach and Arrive/Leave.

Given two objects $o_1, o_2 \in \mathcal{O}$, the fact that o_1 is approaching o_2 can be defined by specifying that o_1 has to be both getting closer and moving towards o_2. The fluent getCloserTo holds true at time t if, given precisification thresholds T_w and T_d, the distance between objects o_1 and o_2 decreases of at least T_d over detection window of length T_w. The definition below uses the measure $(dist, \mathcal{O} \times \mathcal{O}, \mathbb{R})$ defined in (2):

$$\begin{aligned}&[\![\mathsf{getCloserTo}[P](o_1, o_2)]\!]_{[t]} \equiv \exists(T_w, w), (T_d, d) \in P \\ &\quad \exists t_s, t_e, d_s, d_e[t - t_s = t_e - t = w \wedge d_s - d_e > d \wedge \\ &\quad \wedge [\![\mathsf{m}(dist((o_1, o_2), d_s))]\!]_{[t_s]} \wedge [\![\mathsf{m}(dist((o_1, o_2), d_e))]\!]_{[t_e]}\end{aligned} \tag{6}$$

The fluent moveTowards considers whether object o_1 is heading in the direction of o_2 irrespective of o_2's movements (measure $(pos, \mathcal{O}, \mathbb{R}^2)$ in the definition below defined in (1)):

$$\begin{aligned}&[\![\mathsf{moveTowards}[P](o_1, o_2)]\!]_{[t]} \equiv \exists(T_w, w), (T_d, d) \in P \\ &\quad \exists t_s, t_e, p_{1s}, p_{2s}, p_{1e}[t - t_s = t_e - t = w \wedge \\ &\quad \wedge [\![\mathsf{m}(pos(o_1, p_{1s}))]\!]_{[t_s]} \wedge [\![\mathsf{m}(pos(o_2, p_{2s}))]\!]_{[t_s]} \wedge \\ &\quad \wedge [\![\mathsf{m}(pos(o_1, p_{1e}))]\!]_{[t_e]} \wedge \\ &\quad \wedge\, edist(p_{1s}, p_{2s}) - edist(p_{1e}, p_{2s}) > d\end{aligned} \tag{7}$$

It is now possible to define the event Approach:

$$\begin{aligned}&\mathsf{Occurs}(\mathsf{Approach}[P](o_1, o_2), [t_1, t_2]) \equiv \\ &\quad [\![\mathsf{getCloserTo}[P][T_w, T_d](o_1, o_2)]\!]_{[t_1, t_2]} \wedge \\ &\qquad \wedge [\![\mathsf{moveTowards}[P][T_w, T_d](o_1, o_2)]\!]_{[t_1, t_2]}]\end{aligned} \tag{8}$$

The above definitions refer to the rather simple concepts of distance and position. Depending on the data grounding the ontology, one could employ a finer characterisation taking into account, for example, the type of objects involved, the terrain surrounding them, the different paths one object could take, the presence of constraints blocking a particular path, the effort required for each path etc.

The verb Arrive and its opposite Leave appear harder to formalise. Events described by verbs of this kind are punctual or near-punctual events, i.e. their occurrences generally span a small temporal interval whose duration and precise individuation is questionable. Also, their meaning suggests a directional motion and the existence of a destination (Arrive) or a location the object is moving away from

(Leave); the manner of motion is not specified but is often dependent on the type of objects involved (Levin 1993). A simple example is a train arriving at a station. Most formalisations would agree on the fact that the interval for such an event occurrence would terminate on the instant the train stops at the platform, but disagree over the starting instant. A possibility is to establish a boundary around the destination the verb refers to (the spatial object 'station'), and formalise Arrive as starting on the instant the train crosses the boundary, such as in the following definition (o_2 is a generic object, representing an actual spatial object or a spatial area):

$$\begin{aligned}&\mathsf{Occurs}(\mathsf{Arrive}[P](o_1,o_2),[t_s,t_e]) \equiv \exists t_1,t_2,t_3,B\\ &\quad\exists (T_w,w),(T_d,d),(T_b,b)\in P\ [t_1<t_s<t_2<t_e \wedge\\ &\quad\wedge [\![\mathsf{m}[P](bdry(o_2,B))]\!]_{[t_1,t_e]} \wedge\\ &\quad\wedge [\![\mathsf{moveTowards}[P](o_1,B)]\!]_{[t_1,t_s]} \wedge\\ &\quad\wedge \mathsf{Occurs}(\mathsf{Enter}(o_1,B),[t_s,t_2]) \wedge \mathsf{Occurs}(\mathsf{Stop}(o_1),[t_3,t_e])\end{aligned} \tag{9}$$

The above definition introduces a number of ambiguous sub-concepts to be formalise. Definition of measure $[\![\mathsf{m}(bdry(o,B)]\!]_{[t]}$, which holds if B is an area representing the boundary of object o at time t, is debatable. In general, concrete/atomic objects show precise boundaries (e.g. people, vehicles, small items), whilst abstract/complex objects show vague boundaries (e.g. cities, stations), whose ambiguity depends on the object's nature and also on the situation and event considered. A suitable boundary given the task of establishing the starting point of the event "train arriving at station", could be, for instance, the line past which a train approaching the station can be seen from the platform (but there may be many platforms, with different fields of view, etc.). However, if a different instance of Arrive is considered, e.g. a person arriving at a station, it is likely that another set of criteria is more appropriate for individuating the station boundary, e.g. a particular distance from the station entrance. Most practical, usable definitions will establish b as a crisp boundary according to precisification threshold T_b.

The verb Enter is vague too; a possible interpretation may define an occurrence of Enter as starting when the front of the train is touching the boundary, and finishing when the entire train has crossed the boundary, but other interpretations may be acceptable. The verb Stop is another vague punctual or near-punctual event with very limited duration, spanning the interval immediately preceding the instant in which motion ceases. Its ambiguity is mostly due to the difficulty of determining such interval. Formalising Stop would require establishing some individuation criteria for the interval probably precisified by a threshold (for example monotonic deceleration).

4.2 Topology

The meaning of verbs Enter and its opposite Exit can be modeled in terms of the changes in the topological relations holding between the event participants. Given the quality $rccRel$ (Sec. 3.3) expressing the topological relations of the RCC-8 calculus (Randell, Cui, and Cohn 1992), a formalisation of the occurrence of event $\mathsf{Enter}(o_1,o_2)$ over interval $[t_s,t_e]$ is given as:

$$\begin{aligned}&\mathsf{Occurs}(\mathsf{Enter}[P](o_1,o_2),[t_s,t_e]) \equiv\\ &\quad\exists t_1,t_2,t_3[t_1<t_s<t_2<t_e<t_3 \wedge\\ &\quad\wedge [\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{DC}))]\!]_{[t_1,t_s]} \wedge\\ &\quad\wedge [\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{EC}))]\!]_{[t_s,t_2]} \wedge\\ &\quad\wedge [\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{PO}))]\!]_{[t_2,t_e]} \wedge\\ &\quad\wedge [\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{TPP}))]\!]_{[t_e,t_3]}]\end{aligned} \tag{10}$$

The quality $\mathsf{q}(rccRel((o_1,o_2),\mathrm{DC}))$ holds if o_1 and o_2 are topologically disconnected. The following formalisation slightly modifies the original RCC-8 formulation by introducing threshold T_{DC} representing the minimum amount of separation between two boundaries to be considered disconnected:

$$\begin{aligned}&[\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{DC}))]\!]_{[t]} \equiv \exists B_1,B_2,I_1,I_2,d\\ &\quad\exists (T_{DC},\delta),(T_b,b)\in P\ [\ [\![\mathsf{m}[P](bdry(o_1,B_1))]\!]_{[t]} \wedge\\ &\quad\wedge [\![\mathsf{m}[P](bdry(o_2,B_2))]\!]_{[t]} \wedge [\![\mathsf{m}[P](int(o_1,I_1))]\!]_{[t]} \wedge\\ &\quad\wedge [\![\mathsf{m}[P](int(o_2,I_2))]\!]_{[t]} \wedge I_1\cap I_2=\emptyset \wedge B_1\cap B_2=\emptyset \wedge\\ &\quad\wedge [\![\mathsf{m}(dist((B_1,B_2),d))]\!]_{[t]} \wedge d>\delta]\end{aligned} \tag{11}$$

The measures $\mathsf{m}[P](bdry(o,B))$ and $\mathsf{m}[P](int(o,I))$ hold true if B and I are sets of points representing respectively the boundary and the interior of object o, with threshold (T_b,b) precisifying vague boundaries/interiors. The measure $\mathsf{m}(dist((B_1,B_2),d))$ holds true if d is the distance between the two boundaries B_1 and B_2. There may be several possible formalisations of this measure; one appropriate to the disconnection relation would hold for d being the minimum distance between any pair of points belonging to the two boundaries.

The quality $\mathsf{q}(rccRel((o_1,o_2),\mathrm{EC}))$ holds if o_1 and o_2 are externally connected. RCC-8 specifies that two spatial regions are in such a topological relation only if they share a common boundary point and no interior points. The following definition allows for a small degree of separation or overlap between the regions (thresholds T_{DC} and T_{EC}):

$$\begin{aligned}&[\![\mathsf{q}[P](rccRel((o_1,o_2),\mathrm{EC}))]\!]_{[t]} \equiv \exists\, B_1,B_2,I_1,I_2,d\ \exists (T_{DC},\delta),\\ &\quad(T_{EC},\delta'),(T_b,b)\in P\ [\ [\![\mathsf{m}[P](bdry(o_1,B_1))]\!]_{[t]} \wedge\\ &\quad\wedge [\![\mathsf{m}[P](bdry(o_2,B_2))]\!]_{[t]} \wedge [\![\mathsf{m}[P](int(o_1,I_1))]\!]_{[t]} \wedge\\ &\quad\wedge [\![\mathsf{m}[P](int(o_2,I_2))]\!]_{[t]} \wedge\\ &\quad\wedge [\ \ [I_1\cap I_2=\emptyset \wedge B_1\cap B_2=\emptyset \wedge [\![\mathsf{m}(dist((B_1,B_2),d))]\!]_{[t]}\\ &\qquad\quad \wedge d<\delta] \vee [I_1\cap I_2=\emptyset \wedge B_1\cap B_2\neq\emptyset] \vee\\ &\qquad\quad \vee [I_1\cap I_2=I\neq\emptyset \wedge B_1\cap B_2\neq\emptyset \wedge |I|<\delta']\quad]]\end{aligned} \tag{12}$$

The above definition distinguishes three possible configurations of external connection by means of a disjuntion, respectively: 1) o_1 and o_2 are actually disconnected but extremely close to each other; 2) o_1 and o_2 are truly externally connected; 3) o_1 and o_2 are actually partially overlapping, but the amount $|I|$ of points in the intersection of their interior parts is smaller than threshold T_{EC}.

A similar methodology can be followed to formalise the remaining RCC relations. The above formalisation is so far incomplete, as it requires the definition of sub-concepts boundary, interior and intersection. It is also open to further specifications, for example the minimum distance for disconnection (threshold T_{DC}) and the maximum overlap for external connection (threshold T_{EC}) may be made proportional to the size of the two objects.

4.3 Contact

This category comprises generic verbs such as Touch, Hold and specific characterisations of contact motions such as Hit, PickUp, Carry.

The temporal nature of Touch is disputable: one interpretation regards it as a state, i.e. the event-type Touch(o_1, o_2) occurs on any interval where o_1 and o_2 are in contact, another interpretation regards it as a punctual or near-punctual event, such that the above event-type holds on a small temporal interval starting immediately before o_1 comes into contact with o_2, and terminates once movement or contact ceases (with several interpretations demarcating such interval differently). A simple definition based on the topological relations in (11) and (12) is the following:

$$
\begin{aligned}
&\mathsf{Occurs}(\mathsf{Touch}[P](o_1, o_2), [t_s, t_e]) \equiv \exists\, t_1, t_2, t_3 \\
&\quad [t_1 < t_s < t_2 < t_e \wedge t_s < t_3 < t_e \wedge \\
&\quad \wedge [\![\mathsf{q}[P](\mathit{rccRel}((o_1, o_2), \mathrm{DC})]\!]_{[t_1]} \wedge \\
&\quad \wedge [\![\mathsf{q}[P](\mathit{rccRel}((o_1, o_2), \mathrm{EC}))]\!]_{[t_2]} \wedge \\
&\quad \wedge [\mathsf{Occurs}(\mathsf{Stop}(o_1), [t_3, t_e]) \vee \\
&\qquad \vee [\![\mathsf{q}[P](\mathit{rccRel}((o_1, o_2), \mathrm{DC}))]\!]_{[t_e]}] \quad]
\end{aligned}
\tag{13}
$$

The verb Hit is generally a specialised occurrence of Touch with a movement that could be described as fast, sudden, forceful and possibly involving specific contact parts. Whilst it is conceivable to extend definition in Eq 13 to specify speed of objects or contact parts, the characterisation of force constitutes a greater challenge.

The verb PickUp is very specific, as it generally refers to the event starting with person o_1 reaching for object o_2 of small-to-medium size with a particular body part and ending with the person holding the object. Formalising PickUp hence requires the formalisation of Hold, Reach and generally requires the identification of the position of the hands attached to a person. It is often the case that this position has to be inferred in the Theory of Appearances (Sec. 3.2), as it is a too fine-grained detail for most object representations.

4.4 Structured events

This latter category comprises most events that can be characterised as an extended sequence of more specific events. For example, given objects o_1, o_2 of type person and a generic object o_3, the event Give(o_1, o_2, o_3) occurs on the temporal interval corresponding to an occurrence of the sequence in which o_1 has or holds o_3, o_1 moves o_3 towards o_2, o_2 reaches for o_3 and eventually o_2 has or holds o_3.

Another example is the verb Exchange, which also exhibits multiple meanings. One possible characterisation of Exchange is given by a sequence in which two objects of type person exchange position. Another characterisation is given by the sequence in which two objects of type person exchange one or two objects, likely to result in an interleaving of Give and Receive occurrences.

5 Conclusion

The sample verb models in the previous section show the complexity of our main task of defining complex vague concepts from natural language in a formal ontology. The attempt of defining even one motion verb often unfolds a variety of sub-concepts, interpretations and ambiguities. This makes testing the validity of our approach in a practical implementation rather difficult.

We have managed to test part of the methodology and formalism exposed so far in ProVision, our Prolog-based automated reasoning system for event detection. ProVision grounds the ontology on the representation of a video scene (*annotation*) and, via logical inferences, produces a list of predicates representing event occurrences in the scene. In our tests, annotations are text files produced by human observers containing the position of two-dimensional rectangular bounding boxes for each object detected in the scene and each video frame. Tests have been carried out on a set of 1302 vignettes for the definitions of verbs Approach and Hold with encouraging results (see (D'Odorico and Bennett 2012) for implementation and result details).

Looking at the development of the ontology and at the experimental results, we can identify several issues affecting the accomplishment of our event detection task. Firstly, motion verbs concept definitions exhibit a complexity that can escalate quickly. Secondly, the applicability thresholds introduced to disambiguate vague concepts lead to the issue of designing an effective mechanism for establishing appropriate threshold values. Ideally, we would imagine to automatically infer threshold values given type of objects and contextual information.

This leads us to another prominent issue, which consists in the limitations of the data available to us so far. The representation of objects as simple rectangular bounding boxes cannot provide enough evidence for the detection of verbs with very specific meanings. The formalisation of concepts within the ontology has to watch out for the risk of specifying too many detailed semantic characteristics of concepts, manifestations of which may be too challenging to detect on the data available to us. This problem may be alleviated by further development of the theory of appearances. For example, we could imagine to extract three-dimensional coordinates, detect occlusion or attempt to correct error and noise.

There are a number of other approaches to event detection tasks through Machine Learning and Inductive Logic Programming (Dubba, Cohn, and Hogg 2010; Dubba 2012; Sridhar, Cohn, and Hogg 2010). Given the very particular nature of the events to be recognised in this task, we believe our approach's main strength is to potentially provide for a greater specification of each verb's semantic characteristics, which may not be completely understood by learning techniques. In fact, the characterisations within our ontology can be easily augmented by integrating further or different concept definitions suited to a particular task. This approach and methodology also have the potential to be generalised to other domains and automated reasoning tasks on qualitative, vague concepts.

References

Bennett, B., and Galton, A. P. 2004. A Unifying Semantics for Time and Events. *Artificial Intelligence* 153(1-2):13–48.

Bennett, B.; Mallenby, D.; and Third, A. 2008. An Ontology for Grounding Vague Geographic Terms. In Eschenbach, C., and Grüninger, M., eds., *FOIS*, volume 183 of *Frontiers in Artificial Intelligence and Applications*, 280–293. IOS Press.

Bennett, B. 2005. Modes of Concept Definition and Varieties of Vagueness. *Applied Ontology* 1(1):17–26.

Bennett, B. 2011. Possible Worlds and Possible Meanings: a Semantics for the Interpretation of Vague Languages. In *Commonsense 2011: Tenth International Symposium on Logical Formalizations of Commonsense Reasoning*, AAAI Spring Symposium. Stanford University: AAAI.

Cintula, P.; Fermuller, C.; Godo, L.; and Hajek, P., eds. 2011. *Understanding Vagueness. Logical, Philosophical and Linguistic Perspectives*, volume 36 of *Studies in Logic*. College Publications.

D'Odorico, T., and Bennett, B. 2012. Detecting events in video data using a formal ontology of motion verbs. In Stachniss, C.; Schill, K.; and Uttal, D., eds., *Proceedings of Spatial Cognition 2012*, volume 7463 of *Lecture Notes in Artificial Intelligence*, 338–357. Berling Heidelberg: Springer-Verlag.

Dubba, K. S. R.; Cohn, A. G.; and Hogg, D. C. 2010. Event Model Learning from Complex Videos using ILP. In *proceedings of ECAI 2010 - 19th European Conference on Artificial Intelligence, Lisbon, Portugal, August 16-20*, volume 215 of *Frontiers in Artificial Intelligence and Applications*, 93–98. IOS Press.

Dubba, K. S. R. 2012. *Learning Relational Event Models from Video*. Ph.D. Dissertation.

Fine, K. 1975. Vagueness, truth and logic. *Synthese* 30:263–300.

Galton, A. 2009. Spatial and Temporal Knowledge Representation. *Earth Science Informatics* 3(3):169–187.

Keefe, R., and Smith, P. 1997. *Vagueness: a reader*. MIT press.

Keefe, R. 2008. Vagueness: supervaluationism. *Philosophy Compass* 3(2):315–324.

Kowalski, R. A., and Sergot, M. J. 1986. A Logic-based Calculus of Events. *New Generation Computing* 4(1):67–95.

Lawry, J., and Tang, Y. 2009. Uncertainty modelling for vague concepts: a prototype theory approach. *Artificial Intelligence* 173:1539–1558.

Levin, B. 1993. *English Verb Classes and Alternations - A Preliminary Investigation*. The University of Chicago Press.

Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Randell, D. A.; Cui, Z.; and Cohn, A. G. 1992. A Spatial Logic Based on Regions and Connection. In Nebel, B.; Rich, C.; and Swartout, W. R., eds., *Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92). Cambridge, MA, October 25-29, 1992*, 165–176. Morgan Kaufmann.

Shanahan, M. 1999. The Event Calculus Explained. In *Artificial Intelligence Today: Recent Trends and Developments*, volume 1600 of *Lecture Notes in Computer Science*. Springer Berling / Heidelberg. 409–430.

Sorensen, R. A. 2001. *Vagueness and Contradiction*. Oxford University Press.

Sridhar, M.; Cohn, A. G.; and Hogg, D. C. 2010. Unsupervised Learning of Event Classes from Video. In *proceedings of the 24th AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15*, AAAI 2010, 1631–1638.

Tye, M. 1990. Vague objects. *Mind* XCIX(396):535–557.

Varzi, A. C. 2001. Vagueness in georgraphy. *Philosophy and Geography* 4:49–65.

Williamson, T. 1992. Vagueness and Ignorance. *Proceedings of the Aristotelian Society, Supplementary Volumes* 66:145–177.

Williamson, T. 1994. *Vagueness*. The Problems of Philosophy. Routledge, 1998 edition.

Williamson, T. 2003. Vagueness in Reality. In Loux, M. J., and Zimmerman, D. W., eds., *The Oxford Handbook of Metaphysics*. Oxford University Press.

Zadeh, L. A. 1975. Fuzzy Logic and Approximate Reasoning. *Synthese* 30(3-4):407–428.