Spatial Descriptions in Type Theory with Records

Simon Dobnik and Robin Cooper

Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Box 200, 405 30 Göteborg, Sweden
{simon.dobnik@,robin.cooper@ling}.gu.se
http://www.flov.gu.se

Abstract. We present how TTR (Type Theory with Records) can model both geometric perception and conceptual (world) knowledge relating to the meaning of spatial descriptions for a robotic agent.

1 Introduction

TTR [2,3] is a type theory with records which leads to a view of meaning which is tightly linked to perception and classification. An agent makes *judgements* that an object a (an individual or a situation) is of type T (written as $a : T$). The notion of truth is related to such judgements. A type T is "true" just in case there is something a such that $a : T$. However, types are independent of their extensions (also known as *proof objects* or *witnesses*), for example, an agent may know a type but not its extension or two agents may disagree about the extension of a type. An agent learns judgements through his interaction with its environment and other agents. The type systems that agents develop converge to a common standard through constant refinements.

Types are either basic or complex (that is constructed from components). Examples of basic types in this paper are *Ind* and *Real* whose witnesses are individuals and real numbers respectively. Examples of complex types are types constructed from predicates and arguments (*p-types*) such as left(a,b) (intuitively the type of situation where a is to the left of b) and *record types* such as $\begin{bmatrix}\text{a}:\mathit{Ind}\\ \text{b}:\mathit{Ind}\\ \text{c}_{\text{left}}:\text{left(a,b)}\end{bmatrix}$. Record types are sets of fields, pairs which consist of a label (represented to the left of the ':') and a type (to the right). There may not be two fields with the same label. A witness for this record type will be a record with fields with the same labels and witnesses of the corresponding type (and possibly also additional fields with other unique labels). Labels with a 'c' are used here where the type is a p-type (intuitively a constraint on the objects in the record). Such types are often *dependent* in that they are constructed from objects in other fields of the record being judged.

Type theory is attractive as a theory for relating perception to higher level conceptual reasoning because it is based on the notion of judging objects to be of types which can be regarded as an abstract theory of perception. Thus it provides us with a theory that encompasses both low-level perception and high-level semantic reasoning in a way that is not usual in standard linguistic approaches to formal semantics. Thus it offers the possibility of connecting the kind of work in implementations of perception by robots to high level semantics. It is frequently not trivial to connect models of robot perception to natural language semantics in a systematic way (for an approach see [13]). Furthermore, by keeping linguistic and perceptual meaning representations in separate modules their interaction can be hard to explore. We are attempting to bridge this gap.

We attempt to illustrate this approach by sketching how this type theory might model how spatial relations can be generated by a mobile agent's perception of its environment. We show how types representing geometric knowledge required for the meaning representations of spatial descriptions are built from sensory observations (perceptual knowledge). We also demonstrate how perceptual types interact with linguistic type representations (conceptual knowledge). Our aim here is not to say anything new about either sensory perception or about the semantic analysis of the semantics of spatial expression but to give some idea of how both could be comprehended in a single theory.

2 Representing robot states and updates

We can do little more here than indicate some of the types involved and how they are related to each other. See [6] for a more detailed proposal concerning how robot learning might be modelled. We will use types to model the partial information that the robot has about the state that it is in. The type of the initial state of the robot may be:

$$\mathit{InitRobotState} = \left[\begin{array}{lcl} \text{self} & : & \left[\begin{array}{lcl} \text{a} & : & \mathit{Ind} \\ \text{pnt} = \left[\begin{array}{lcl} \text{x} & = & 0 \\ \text{y} & = & 0 \end{array}\right] & : & \mathit{Point} \\ \text{orient}{=}0 & : & \mathit{Real} \\ \text{c}_{\text{point}} & : & \text{observed\_point(self.pnt)} \\ \text{c}_{\text{loc}} & : & \text{located(self.a,self.pnt)} \end{array}\right] \\ \text{c}_{\text{robot}} & : & \text{robot(self.a)} \\ \text{pm} = [\text{self.pnt}] & : & \mathit{PointMap} \\ \text{objects}{=}[\text{self}] & : & [\mathit{Object}(\text{pm})] \\ \text{c}_{\text{object\_map}} & : & \text{obj\_map(objects, pm)} \\ \text{beliefs}{=}[] & : & \mathit{RecType} \\ \text{time}{=}0 & : & \mathit{Time} \end{array}\right]$$

The 'self'-field requires a record corresponding to a located individual, a point in (two-dimensional) space represented as a record with fields for the x- and y-coordinates (initially set to 0), an orientation represented as a real number, initially 0, and two constraints which require that the robot has observed the point at which it is located.

The notation *label=value:Type* used in the 'pnt' and 'orient' fields here is known as a *manifest field* and is used to represent a field $label{:}Type_{value}$, where the $Type_{value}$ is a restriction of *Type* so that its unique witness is *value*. The 'pm'-field is for a point map modelled as a list of points, initially the singleton list containing the location of 'self'. The point map is a list of individuated point landmarks as built with a SLAM procedure [5] . The 'objects'-field is for a list of objects assembled from the point map (that is, an object map based on 'pm'), initially the singleton list containing 'self'. This is an object map. As the robot moves around it discovers new landmarks which are added to the point map and their estimate of global location (relative to the robot's origin) is continuously improved. Since at this point no point landmarks have been discovered yet, the list of objects built from these landmarks, and the list of beliefs about these objects is also empty. At the time $t + 1$ the agent may transition to a new state by moving and making new observations. It may also hear an utterance made by its conversational partner.

SLAM gives us a geometric representation of the environment containing abstracted point landmarks in a global coordinate frame from which angles and distances required for geometric representation of spatial descriptions can be determined [10,12,14]. The robot's list of objects represented in the 'objects'-field of the state are located at points and regions within this point map. A geometric representation of a region or a volume consists of a group of 2-dimensional points from the point map that can be hulled with a convex hull.

The types *PointObject* and *RegionObject* are relative to a point map, and this is represented by functions returning a type (dependent types):

$$\mathit{PointObject} = \lambda p{:}\mathit{PointMap}(\begin{bmatrix} \mathrm{a} & : & \mathit{Ind} \\ \mathrm{pnt} & : & \mathit{Point} \\ \mathrm{orient} & : & \mathit{Real} \\ \mathrm{c_{point}} & : & \mathrm{observed\_point}(\mathrm{pnt},p) \\ \mathrm{c_{loc}} & : & \mathrm{located}(\mathrm{a},\mathrm{pnt}) \end{bmatrix})$$

$$\mathit{RegionObject} = \lambda p{:}\mathit{PointMap}(\begin{bmatrix} \mathrm{a} & : & \mathit{Ind} \\ \mathrm{reg} & : & \mathit{PointMap} \\ \mathrm{orient} & : & \mathit{Real} \\ \mathrm{c_{region}} & : & \mathrm{region}(\mathrm{reg},p) \\ \mathrm{c_{loc}} & : & \mathrm{located}(\mathrm{a},\mathrm{reg}) \end{bmatrix})$$

$$\mathit{Object} = \lambda p{:}\mathit{PointMap}(\mathit{PointObject}(p) \vee \mathit{RegionObject}(p))$$

(See [6, p.8] for a characterization of the predicates 'observed_point' and 'region'.) Once the robot has identified located objects in this way it can compute spatial relations between these objects by comparing their 'pnt' (location point) or 'reg' (location region) fields. Beliefs about such spatial relations, coded by p-types) will be added to the 'beliefs'-field in the robot state.

3 Representing spatial relations

Geometrically, the spatial relation 'to the left of'[1] holds between three individuals conceptualised as objects of type *RegionObject*: the located object, the reference object and the viewpoint which determines the orientation of the reference frame [7,9]. If o_1, o_2, o_3:*RegionObject* and f_{relation} is a spatial relation classifier[2] of type $\mathit{Region} \rightarrow \mathit{Region} \rightarrow \mathit{Orientation} \rightarrow \mathit{Type}$ then

$$\begin{array}{l} e\text{:left}(o_1.\text{a},o_2.\text{a},o_3.\text{a}) \text{ iff } e : f_{\text{relation}}(o_1.\text{reg}, o_2.\text{reg}, o_3.\text{orient}) \\ \text{and } f_{\text{relation}}(o_1.\text{reg}, o_2.\text{reg}, o_3.\text{orient}) = \text{left}_{geom}(o_1.\text{reg}, o_2.\text{reg}, o_3.\text{orient}). \end{array}$$

Two relativisations or transformations of region locations must be performed before the classification can take place (both of which can be expressed in our formalism): (i) the (global) coordinate frame must be rotated to correspond to the orientation of o_3; and (ii) the origin of the global coordinate frame must be transposed so that it is identical to the centre point of the region of location of o_2 (cf. [11]). Since o_1's region of location has been relativised we only need to learn one classifier function regardless of the viewpoint. The TTR representation allows us to combine perceptual classification with qualitative spatial representation [1].

The new belief [e:left(o_1.a,o_2.a,o_3.a)] is merged with the robot's beliefs in the 'beliefs'-field of the robot state and can be used, for example to answer a question about the location of o_1.[3]

The influence of world knowledge on the semantics of the spatial descriptions goes beyond conceptualisation of objects. For example, [4] describe experiments involving pictures of a man holding an umbrella at various angles and with various degrees of exposure to rain presented to human observers and conclude that for the spatial relation 'over' the satisfaction of the constraint 'umbrella provides protection from rain' is more than 'the umbrella is within the geometric spatial template for 'over'. A predicate representing 'over' would obey something like the following conditional (not biconditional):

$$e\text{:over}(\text{o}_1.\text{a},\text{o}_2.\text{a},\text{o}_3.\text{a}) \text{ if } e\text{:}\begin{bmatrix} \text{c}_{\text{rain}} & : & \text{rain}(\text{o}_3.\text{a}) \\ \text{c}_{\text{umbrella}} & : & \text{umbrella}(\text{o}_2.\text{a}) \\ \text{c}_{\text{over}_{\text{geom}}} & : & \text{over}_{geom}(\text{o}_2.\text{vol},\text{o}_1.\text{vol}) \\ \text{c}_{\text{protects}} & : & \text{protects}(\text{o}_3.\text{a},\text{o}_1.\text{a},\text{o}_2.\text{a}) \end{bmatrix}$$

where o_1, o_2 and o_3 are of type *VolumeObject* similar to *RegionObject* except that three dimensional volumes are used rather than two dimensional regions.

Geometrically ($\text{c}_{\text{over}_{\text{geom}}}$), the umbrella must be in a particular spatial configuration with the man which can be trained as a classifier . 'over_{geom}' is typically not susceptible to perspective shifts as the viewpoint is fixed by the gravity and hence the third object that would determine the viewpoint is not necessary. Hence, before the classification

[1] We are considering the relative notion here, not that which is based on the intrinsic orientation of some object which has a front and a back.

[2] See [8] for a TTR account of classifier learning from human interaction.

[3] Objects o_2 and o_3 would have to be selected separately beforehand. The reference object o_2 should be some contextually salient object. The viewpoint object o_3 should be the agreed viewpoint in the discourse.

takes place only the origin of the global coordinate frame must be transposed to the centre point of the volume of location of o_2.

The constraint c_{protects} represents a conceptual constraint on witnesses of the ptype over(o_1.a, o_2.a, o_3.a) where the ptype protects(o_3, o_1, o_2) may in its turn also rely on a perceptual classifier. What is important here is that this constraint can have been learned by the agent not through perceptual observation but through linguistic communication, for example by being explicitly told that protection from the rain is required. Alternatively it could have been learned by hypothesising this fact after observing situations of humans, umbrellas and rain. Through reasoning humans are able to create increasingly more abstract types which are ultimately grounded in perception[4]. In our view there is no clear cut-off point between low level perceptual knowledge and high-level conceptual knowledge as traditionally assumed.

Since we assume that the geometric meaning constraint $c_{\text{over}_{\text{geom}}}$ is determined by a probabilistic classifier, the acceptable deviations of the umbrella from the prototypical vertical upright position and their gradience are accounted for. The representation predicts that a situation where a man holds an umbrella in the upright position and therefore the $c_{\text{over}_{\text{geom}}}$ constraint is defined with high probability but the umbrella does not provide protection from the rain cannot have the denotation of the ptype over(o_3,o_1,o_2) since the constraint c_{protects} is not satisfied. Since ptypes such as over(o_3,o_1,o_2) may be characterised by probabilistic knowledge as well, we could regard all constraints as expressing a degree of belief that particular situations are of particular types (see [6, p.17–18] for more details and also probabilistic TTR [in prep.]).

4 Conclusion and further work

We have presented a brief sketch how TTR can be used to represent different meaning components of spatial descriptions. Its strengths are that it considers meaning representations to be based on perception and that it can represent different meaning modalities in a unified way. It thus bridges the gap between models of natural language and models of perception. In such a model it becomes transparent that there are many similarities in the way an agent learns and applies the meanings of linguistic and non linguistic representations. Being a formal computational model it is well suited for modelling language and perception in artificial agents which will be the focus of our work in the future.

Acknowledgements We thank 3 anonymous reviewers of the CoSLI 3 workshop for their valuable comments. We would also like to thank Staffan Larsson for important discussion in connection with this work.

References

1. Cohn, A.G., Renz, J.: Qualitative spatial representation and reasoning. In: Frank van Harmelen, V.L., Porter, B. (eds.) Handbook of Knowledge Representation, Foundations of Artificial Intelligence, vol. 3, chap. 13, pp. 551–596. Elsevier (2008)

[4] Although we do not, of course, claim that all types are grounded in physical perception

2. Cooper, R.: Austinian truth, attitudes and type theory. Research on Language and Computation 3(2), 333–362 (2005)
3. Cooper, R.: Type theory and semantics in flux. In: Kempson, R., Asher, N., Fernando, T. (eds.) Handbook of the Philosophy of Science, General editors: Dov M Gabbay, Paul Thagard and John Woods, vol. 14. Elsevier BV (2012)
4. Coventry, K.R., Prat-Sala, M., Richards, L.: The interplay between geometry and function in the apprehension of Over, Under, Above and Below. Journal of memory and language 44(3), 376–398 (2001)
5. Dissanayake, M.W.M.G., Newman, P.M., Durrant-Whyte, H.F., Clark, S., Csorba, M.: A solution to the simultaneous localization and map building (SLAM) problem. IEEE Transactions on Robotic and Automation 17(3), 229–241 (2001)
6. Dobnik, S., Cooper, R., Larsson, S.: Modelling language, action, and perception in type theory with records (January 2013), `http://sites.google.com/site/typetheorywithrecords/drafts/perceptual-ttr-post-proceedings.pdf`, manuscript
7. Herskovits, A.: Language and spatial cognition: an interdisciplinary study of the prepositions in English. Cambridge University Press, Cambridge (1986)
8. Larsson, S.: The TTR perceptron: Dynamic perceptual meanings and semantic coordination. In: Artstein, R., Core, M., DeVault, D., Georgila, K., Kaiser, E., Stent, A. (eds.) SemDial 2011 (Los Angelogue): Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue. pp. 140–148. Los Angeles, California (September 21—23 2011)
9. Levinson, S.C.: Space in language and cognition: explorations in cognitive diversity. Cambridge University Press, Cambridge (2003)
10. Logan, G.D., Sadler, D.D.: A computational analysis of the apprehension of spatial relations. In: Bloom, P., Peterson, M.A., Nadel, L., Garrett, M.F. (eds.) Language and Space, pp. 493–530. MIT Press, Cambridge, MA (1996)
11. Maillat, D.: The semantics and pragmatics of directionals: a case study in English and French. Ph.D. thesis, Committee for Comparative Philology and General Linguistics, University of Oxford, Oxford, UK (May 2003)
12. Regier, T., Carlson, L.A.: Grounding spatial language in perception: an empirical and computational investigation. Journal of Experimental Psychology: General 130(2), 273–298 (2001)
13. Roy, D.: Semiotic schemas: a framework for grounding language in action and perception. Artificial Intelligence 167(1-2), 170–205 (Sep 2005)
14. Zwarts, J., Winter, Y.: Vector space semantics: A model-theoretic analysis of locative prepositions. Journal of Logic, Language and Information 9, 169–211 (2000)