Type Theory with Records: a General Framework for Modelling Spatial Language

Simon Dobnik, Robin Cooper and Staffan Larsson

University of Gothenburg, Centre for Language Technology,
Dept. of Philosophy, Linguistics & Theory of Science, Gothenburg, Sweden
{simon.dobnik@, robin.cooper@ling, staffan.larsson@ling}.gu.se

1. Introduction

Cross-disciplinary research has shown that spatial language is dependent on several contextual factors, for example geometrical arrangement of the scene (Regier and Carlson, 2001), the type of objects referred to and their interaction (Coventry et al., 2001) , and alignment in dialogue (Watson et al., 2004) among others. Although the contribution of these contextual factors has been well-studied, several questions relating to the modelling and representation in the domain of computational models for situated conversational agents still remain.

Some basic requirements for such a representational system are: (i) formal accuracy; (ii) the ability to represent information from different modalities; (iii) bridging perceptual and conceptual domains; (iii) adaptability and learnability of representations; (iv) formal ability to capture the meaning relations typically found in human reasoning and language.

In building situated conversational agents, several systems have been proposed but none of them capture all of these requirements. For example, *semiotic schemas* (Roy, 2005) account for the meaning of words that refer to entities and actions it is not straightforwardly evident how they relate to other linguistic representations. (Krujiff et al., 2007) adopt a layered model with distinct representations at each layer. For example, there is a feature map corresponding to features from sensory observations, a navigation graph containing way-points, topological map of areas, and a conceptual map of an ontology of objects. Although there exist mechanisms by which these representational levels interact, the kinds of representations at each level are quite distinct from each other and are shaped by different operations. The question we would like to address is whether such representational levels and operations can be generalised by taking inspiration from the way humans assign, learn and reason with meaning.

2. Type Theory with Records

Type Theory with Records (TTR) (Cooper et al., 2014) provides a theory of natural language semantics which views meaning as tightly linked to perception and classification. It is based on the notion of an agent *judging* situations/invariances in the world to be of types (written as $a : T$) which can be regarded as an abstract theory of perception (Larsson, 2013). This 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 as well as it offers robotics the possibility of connecting the implementations of perception to high level semantics. The type system in TTR is rich in comparison to that found in traditional formal semantics (entities and truth values). Types in TTR are represented as matrices or *record types* containing label-value pairs where labels are constants and values can be either basic (Ind, Real) or record types. The corresponding proof-objects of record types are records.

TTR also incorporates a theory of interaction as it takes the view that agent learns judgements through their interaction with its environment and other agents.

2.1 Classification of objects and spatial relations

Assuming we have a pointmap (a two or three dimensional matrix of real numbers) representing an agent's sensory input of its environment, we can try to detect objects in that environment and classify them as being of various sorts such as chairs, tables etc. We introduce a type *Pointmap* whose witnesses are such matrices. We assume an *object detection function* which is of type

$$(\textit{Pointmap} \rightarrow \text{set}(\begin{bmatrix} \text{reg} & : & \textit{Pointmap} \\ \text{pfun} & : & (\textit{Ind}\rightarrow\textit{Type}) \end{bmatrix}))$$

that is, it maps a pointmap and to set of records specifying (1) the region (reg) occupied by an object (which should be required to be a subpointmap of the input pointmap), and (2) a property which is to be associated with the object. This is a function, labelled 'pfun' which maps an individual a ptype, a type constructed from a predicate and an argument.[1] An example of such a property would be the function $\lambda x{:}\textit{Ind}$. chair(x). Such a function thus associates regions in a pointmap with properties of individuals. Given such a pairing of a pointmap and a property (modelled in TTR as a record with two fields) it is straightforward to construct the type of situations which contain an individual located at the region and having the property. The following function would return such a type:

$$\lambda r{:}\begin{bmatrix} \text{reg:}\textit{Pointmap} \\ \text{pfun:}(\textit{Ind}\rightarrow\textit{Type}) \end{bmatrix} . \begin{bmatrix} \text{a} & : & \textit{Ind} \\ \text{loc} & : & \text{location(a, } r\text{.reg)} \\ \text{c} & : & r\text{.pfun(a)} \end{bmatrix}$$

Let us call such a function an *individuation function*. By combining the use of an object detection function and an individuation function we can map between the perceptual

[1] Types constructed from predicates (ptypes) are to be thought of intuitively as types of situations which show the predicate to hold of the argument. Within type theory, types play the role of propositions, rather than, for example, sets of possible worlds.

domain of point clouds and the “logical” domain of individuals with various properties using the same type theoretical apparatus for both domains.

2.2 Object function and interaction

Spatial descriptions, e.g. “in”, are not only sensitive to geometric arrangements of scenes modelled by spatial templates but also to the type of and interaction between objects related. These spatial relations can be automatically generalised over in terms of hypernym classes (Dobnik and Kelleher, 2014).

Ontological knowledge, e.g. that a chair is furniture, can be modelled by similar functions which map records to types:

$$\lambda r{:}\left[\begin{array}{l}\text{a:}\mathit{Ind}\\ \text{c:chair(a)}\end{array}\right].\ \text{furniture}(r.\text{a})$$

We can use such a function to generate a superordinate type from a type generated by an individuation function.

A preposition like “in” is associated with several distinct types of spatial situations as has been confirmed by empirical work. Each type of situation representing a spatial template involves a different interplay of geometric and conceptual knowledge spanning the domain of point clouds and “logical” individuals. An example of a function sketching such an inference is given below:

$$\lambda r{:}\left[\begin{array}{lcl}\text{o}_1 & : & \left[\begin{array}{lcl}\text{a} & : & \mathit{Ind}\\ \text{reg} & : & \mathit{Pointmap}\\ \text{c} & : & \text{person(a)}\end{array}\right]\\ \text{o}_2 & : & \left[\begin{array}{lcl}\text{a} & : & \mathit{Ind}\\ \text{reg} & : & \mathit{Pointmap}\\ \text{c} & : & \text{furniture(a)}\end{array}\right]\\ \text{st} & : & \text{spatial-template}_{in5}(o_1\text{.reg},o_2\text{.reg})\end{array}\right].\ \text{in(r.}o_1\text{.a,r.}o_2\text{.a)}$$

This function takes account of conceptual properties of objects that could be obtained by computing relevant hypernyms such as ‘person’ and ‘furniture’ and an associated spatial template that relates the point clouds associated with them. It then generates a type of situation which involves a conceptual spatial relation between individuals.

2.3 Accommodating frame of reference (FoR)

Agents in conversation align to the primed FoR and continue to use it. Speakers initiating conversation tend to be egocentric: they generate description from their point of view. In the examples below s_0^{Alex} an s_1^{Sam} represent dialogue information states of the agents Alex and Sam respectively. These are modelled as records of the types indicated. Thus, for example, the origin for Alex’s frame of reference in s_0^{Alex} is s_0^{Alex}.private.for-origin. Hearers assume that speakers take this strategy as represented by s_0^{Sam}.shared.for-origin.

Alex: The chair is to the left of the table.

$$s_0^{Alex}:\left[\begin{array}{lcl}\text{priv} & : & \left[\begin{array}{l}\text{objs:}\left[\begin{array}{l}\text{o}_0{:}\Sigma_0\\ \text{o}_1{:}\Sigma_1\\ \text{o}_2{:}\Sigma_2\\ \text{o}_3{:}\Sigma_3\end{array}\right]\\ \text{bel:}\left[\begin{array}{l}\text{c}_{me}\text{=}\left[\text{c:me}(\Uparrow^2\text{objs.o}_0\text{.a})\right]{:}\mathit{RecType}\\ \text{c}_{left}\text{=}\left[\text{c:left}(\Uparrow^2\text{objs.o}_2\text{.a, }\Uparrow^2\text{objs.o}_3\text{.a})\right]{:}\mathit{RecType}\end{array}\right]\\ \text{for-origin=objs.o}_0\text{.a:}\mathit{Ind}\\ \text{agenda= [}\left[\begin{array}{l}\text{move:}\mathit{Assertion}\\ \text{cont=}\Uparrow^2\text{bel.c}_{left}{:}\mathit{RecType}\end{array}\right]\text{]:list(}\mathit{DMove}\text{)}\end{array}\right]\\ \text{shared:} & & \left[\text{c}_{in-focus}{:}\Uparrow\text{priv.objs.o}_2\text{.a}\right]\end{array}\right]$$

Sam: Aha.

$$s_1^{Sam}:\left[\begin{array}{lcl}\text{priv} & : & \left[\begin{array}{l}\text{objs:}\left[\begin{array}{l}\text{o}_0{:}\Sigma_0\\ \text{o}_1{:}\Sigma_1\\ \text{o}_2{:}\Sigma_2\\ \text{o}_3{:}\Sigma_3\end{array}\right]\\ \text{bel }:\left[\text{c}_{me}\text{=}\left[\text{c:me}(\Uparrow^2\text{objs.o}_0\text{.a})\right]{:}\mathit{RecType}\ \right]\end{array}\right]\\ \text{shared:} & & \left[\begin{array}{l}\text{speaker=}\Uparrow\text{priv.objs.}_0\text{.a:}\mathit{Ind}\\ \text{c}_{in-focus}{:}\Uparrow\text{priv.objs.o}_2\\ \text{latest-move: }\left[\begin{array}{l}\text{speaker=}\Uparrow^2\text{priv.objs.o}_0\text{.a:}\mathit{Ind}\\ \text{cont=}\left[\text{c:left}(\Uparrow^3\text{priv.objs.o}_2\text{.a, }\Uparrow^3\text{priv.objs.o}_3\text{.a})\right]{:}\mathit{RecType}\end{array}\right]\\ \text{for-origin=speaker:}\mathit{Ind}\end{array}\right]\end{array}\right]$$

The field labelled ‘priv.objs’ is for a record of objects which the agent has established where each of the Σ_i is a type returned by an individuation function on the basis of the pointmap that the agent has constructed on the basis of perceptual experience.

Here we assume for brevity that Sam and Alex use identical labels for the objects and that ‘o_0’ labels Alex, o_1 labels Sam, and o_2 and o_3 label the objects they are talking about. This simplification is not necessary. Neither is it necessary that the two agents have identified the same objects and part of the future plans in this project is to show how differences in perception of the physical domain can be resolved by communication between the agents using dialogue.

3. Conclusion

We propose that TTR is a very suitable candidate for representing and reasoning with the meaning of spatial descriptions in conversational agents and sketch some of its strengths with examples. Our future work will involve implementing this framework in a computational application.

References

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. 2014. A probabilistic rich type theory for semantic interpretation. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 72–79, Gothenburg, Sweden, 27 April.

Kenny R. Coventry, Mercè Prat-Sala, and Lynn Richards. 2001. The interplay between geometry and function in the apprehension of Over, Under, Above and Below. *Journal of Memory and Language*, 44(3):376–398.

Simon Dobnik and John Kelleher. 2014. Exploration of functional semantics of prepositions from corpora of descriptions of visual scenes. In *Proceedings of the Third Workshop on Vision and Language*, pages 33–37, Coling, Dublin, Ireland.

Geert-Jan M. Krujiff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: what, where... and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138. Special issue on human and robot interactive communication.

Staffan Larsson. 2013. Formal semantics for perceptual classification. *Journal of Logic and Computation*, online, Dec. 18.

Terry Regier and Laura A. Carlson. 2001. Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General*, 130(2):273–298.

Deb Roy. 2005. Semiotic schemas: a framework for grounding language in action and perception. *Artificial Intelligence*, 167(1-2):170–205, September.

Matthew E. Watson, Martin J. Pickering, and Holly P. Branigan. 2004. Alignment of reference frames in dialogue. In *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, Chicago, USA.