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 arising from the interaction of the agent with the environment through perception and other agents through situated conversation, for example geometrical arrangement of the scene (Regier and Carlson, 2001), the type of objects referred to and their interaction (Coventry et al., 2001), visual and discourse salience of objects (Kelleher et al., 2005), alignment in dialogue (Watson et al., 2004), and gesture (Tutton, 2013) among others. Although the contribution of these contextual factors has been well-studied, several questions relating to the modelling, representation and information fusion of these sources in the domain of computational models for situated conversational agents still remain. In particular, (i) how is an agent able to determine the meaning and reference of such descriptions and how do different contextual factors represented as features combine with each other to form bundles that represent the meaning of spatial descriptions; (ii) how can be mechanisms by which different sources of contextual knowledge interact with each other to form the meaning of a spatial description modelled in a formal and precise way; (iii) how do the meaning representations of spatial descriptions combine with meaning representations of other spatial or non-spatial words in a language in order to form the meanings of sentences; and (iv) how can an agent reason about linguistic and non-linguistic representations to plan its actions.

Some basic requirements for such a representational system are: (i) formal accuracy; (ii) the ability to represent information from different modalities; (iii) classification to bridge perceptual and conceptual domains (c.f. grounding (Harnad, 1990)); (iii) adaptability and learnability of representations (situated agents may enter new physical and conversational contexts to which they must adapt); (iv) sufficient formal strength of representations which captures the meaning relations typically found in human reasoning and consequently language (Blackburn and Bos, 2005).

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) focus on meanings being functions from perceptual events and actions of a robot. Although these routines 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. Relying on a general framework like this allows us to formulate representations that can be more easily related to each other and also to create computational models that more closely correspond to human cognition.

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. The example below shows a judgement that a record (a matrix with = as a delimiter) containing a sensory reading is of a type (with : as a delimiter). The traditional distinction between symbolic and sub-symbolic knowledge is blurred in this framework as both can be assigned appropriate types. The framework is attractive for modelling spatial descriptions, which being itself symbolic, also require reference to the perceptual/geometric properties of the scene.

$$\begin{bmatrix} \text{a} & = & \text{ind}_{26} \\ \text{sr} & = & \text{[[34,24],[56,78]...]} \\ \text{loc} & = & \text{[45,78,0.34]} \end{bmatrix} : \begin{bmatrix} \text{a} & : & \text{Ind} \\ \text{sr} & : & \text{list(list(Real))} \\ \text{loc} & : & \text{list(Real)} \end{bmatrix}$$

There are several relations of meaning components that the type system allows us to capture: classification by *functional application* of types (Section 2.1), generalisation and specialisation by the notion of *sub-typing* (Section 2.2) and

type-merging operations such as *asymmetric-merge*, meaning constituency by *dependent types* (Section 2.2), temporal sequencing of type judgements as a string of events, etc. 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. The type systems that agents individually develop are dynamic, probabilistic and converge to a common standard through constant interactive refinements (Section 2.3).

2.1 Classification of objects and spatial relations

Classification is modelled as a functional mapping of information of one type to another. The function takes a record of sensory readings of the type on the left and returns a type of the object on the right which in this case is a predicate type, e.g. chair(a). The record associated with the o_1 variable is a *manifest field*, a way of fully specifying type, which stores a proof that the object is of this type. The function $f_{pointmap}$ returns a region of the absolute point map occupied by the object.

$$\lambda r{:}\begin{bmatrix} \text{a} & : & \text{Ind} \\ \text{sr} & : & \text{list(list(Real))} \\ \text{loc} & : & \text{list(Real)} \end{bmatrix} \left(\begin{bmatrix} \text{o}_1{=} \begin{bmatrix} \text{a} & {=}r.\text{a} \\ \text{sr} & {=}r.\text{sr} \\ \text{loc} & {=}r.\text{loc} \\ \text{reg} & {=}f_{pointmap}(r) \end{bmatrix} & : & f_{objclass}(r) \end{bmatrix} \right)$$

such that $f_{objclass}(r)$ = ClassPred(r.a) where ClassPred is one of chair, box, alex,...

2.2 Object function and interaction

If r : chair(a) then

$$\text{r} : \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{sr} & : & \text{list(Real)} \\ \text{loc} & : & \text{list(Real)} \\ \text{reg} & : & \text{f}_{pointmap}\text{(r)} \\ \text{c}_{\text{hyp}} & : & \text{furniture(x)} \end{bmatrix}$$

If r : box(a) then

$$\text{r} : \begin{bmatrix} \text{x} & : & \text{Ind} \\ \text{sr} & : & \text{list(Real)} \\ \text{loc} & : & \text{list(Real)} \\ \text{reg} & : & \text{f}_{pointmap}\text{(r)} \\ \text{c}_{\text{hyp}} & : & \text{phys-obj(x)} \end{bmatrix}$$

$$\lambda\text{r}: \begin{bmatrix} \text{o}_1 & : & \begin{bmatrix} \text{a} & : & \text{Ind} \\ \cdots & & \\ \text{reg} & : & \text{f}_{pointmap}\text{(r)} \\ \text{c}_{\text{hyp}} & : & \text{person(a)} \end{bmatrix} \\ \text{o}_2 & : & \begin{bmatrix} \text{a} & : & \text{Ind} \\ \cdots & & \\ \text{reg} & : & \text{f}_{pointmap}\text{(r)} \\ \text{c}_{\text{hyp}} & : & \text{furniture(a)} \end{bmatrix} \\ \text{st} & : & \text{spatial-template}_{in5}(o_1\text{.reg},o_2\text{.reg}) \end{bmatrix} (\text{in}(\text{r}.o_1,\text{r}.o_2))$$

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 which can be automatically generalised over as hypernym classes (Dobnik and Kelleher, 2014). Hyponym/hypernym relations can be expressed with *subtyping* (e.g. chair $\sqsubseteq$ furniture $\sqsubseteq$ artefact $\sqsubseteq$ physical object $\sqsubseteq$ entity). The types with the labels st, o_1 and o_2 are *dependent types* on their containing record type. We can also represent that "in" is associated with several distinct types of spatial situations which is confirmed empirically and that each type of situations involves a different interplay of geometric and conceptual knowledge.

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 (private.for-origin=objects[0] : *Object*). Hearers assume that speakers take this strategy (private-for-origin=last-move.c_s.speaker/2 : *Object*).

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

$$\begin{bmatrix} \text{private} & : & \begin{bmatrix} \text{t=1} & : & \textit{Time} \\ \text{agenda= [} \begin{bmatrix} \text{m:}\textit{Assertion} \\ \text{cnt=[beliefs[0]]:list(}\textit{RecType}\text{)} \end{bmatrix} \ldots \text{]} & : & \text{list(}\textit{DMove}\text{)} \\ \text{beliefs= [[}s_1\text{:left(objects[2],objects[3])], [}s_1\text{:me(objects[0])]]} & : & \text{list(}\textit{RecType}\text{)} \\ \text{objects= [}o_0\text{, }o_1\text{, }o_2\text{, }o_3\text{]} & : & \text{list(}\textit{Object}\text{)} \\ \text{for-origin=objects[0]} & : & \textit{Object} \end{bmatrix} \\ \text{shared} & : & \begin{bmatrix} \text{in-focus=private.objects[2]} & : & \textit{Object} \end{bmatrix} \end{bmatrix}$$

Sam: Aha.

$$\begin{bmatrix} \text{private} & : & \begin{bmatrix} \text{t=2} & : & \textit{Time} \\ \text{agenda= []} & : & \text{list(}\textit{DMove}\text{)} \\ \text{beliefs= [[}s_1\text{:me(objects[1])]]} & : & \text{list(}\textit{RecType}\text{)} \\ \text{objects= [}o_4\text{, }o_5\text{, }o_6\text{, }o_7\text{]} & : & \text{list(}\textit{Object}\text{)} \end{bmatrix} \\ \text{shared} & : & \begin{bmatrix} \text{in-focus=private.objects[2]:}\textit{Object} \\ \text{last-move= [} \begin{bmatrix} \text{m:}\textit{Assertion} \\ c_s\text{=speaker(m,private.objects[0])} & : & \textit{Object} \\ c_h\text{=hearer(m,private.objects[1])} & : & \textit{Object} \\ \text{cnt=[[}p_1\text{:left(private.objects[2],...[3])]]} & : & \text{list(}\textit{RecType}\text{)} \end{bmatrix} \text{]} & : & \text{list(}\textit{DMove}\text{)} \\ \text{beliefs=[[}s_1\text{:last-move.cnt]]} & & & : & \text{list(}\textit{RecType}\text{)} \\ \text{for-origin= last-move.}c_s\text{.speaker/2} & & & : & \text{Object} \end{bmatrix} \end{bmatrix}$$

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

Patrick Blackburn and Johan Bos. 2005. *Representation and inference for natural language. A first course in computational semantics*. CSLI Publications.

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. Association for Computational Linguistics.

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, Dublin, Ireland, August. Dublin City University and the Association for Computational Linguistics.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D*, 42(1–3):335–346, June.

J.D. Kelleher, F. Costello, and J. van Genabith. 2005. Dynamically structuring updating and interrelating representations of visual and linguistic discourse. *Artificial Intelligence*, 167:62–102.

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, December 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.

Mark Tutton. 2013. A new approach to analysing static locative expressions. *Language and Cognition*, 5:25–60, 3.

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.