REPRESENTING DIFFERENT KINDS OF REFERENCE WITH TYPE THEORY WITH RECORDS

Simon Dobnik, Robin Cooper & Staffan Larsson (University of Gothenburg)
{simon.dobnik@, robin.cooper@ling, staffan.larsson@ling}.gu.se

Type Theory with Records (TTR) (Cooper, 2012) provides a theory of natural language semantics which views meaning as tightly linked to perception and classification. An agent has access to its local world in which it can make *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 their interaction with its environment and other agents. The type systems that agents develop are dynamic and converge to a common standard through constant refinements (Cooper and Larsson, 2009; Larsson, 2013).

TTR 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 (Cooper et al., 2014). Thus it offers the possibility of connecting the kind of work in implementations of perception by robots to high level semantics (Dobnik et al., 2013). It is not trivial to connect models of robot perception to natural language semantics in a systematic way (for an approach see Roy (2005)). 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.

Traditionally, reference is considered as a relation between a linguistic expression and some properties of a physical world. For example, spatial descriptions can be modelled as *spatial templates* or *potential fields* which identify regions of applicability of a particular spatial relation between a landmark and a target object (Regier and Carlson, 2001). However, there are at least two other kinds of referential relations that the semantics of these descriptions contextually depend on. Experimental research by Coventry and Garrod (2004) shows that spatial prepositions are sensitive to what objects are involved in the relation and how they interact with each other: such knowledge (also known as *functional*) is part of an agent's knowledge about the world and its ability to predict the outcome of situations. Spatial descriptions are also referential in respect to the current dialogue in which they are used. For example, perspective of spatial relations such as "in front of the chair" is frequently left out in conversation as it is can be agreed upon or aligned between conversational partners (Dobnik et al., 2014). Spatial descriptions are referential in all these cases in the sense that their semantics picks out some invariances of the current context in which a description is used.

We argue that TTR is ideally suited for modelling semantics of spatial descriptions because (i) meaning is evaluated in a form of judgements which are made as an agent explores the world, taking into account changing contexts which give rise to dynamic meaning representations required for capturing the referential nature of spatial descriptions; (ii) as a powerful knowledge representation system it allows us to represent meaning distinctions ranging from basic perceptual concepts which make up lexical semantics of words to entire discourses or dialogue game boards; (iii) the sub-typing allows us to capture underspecification and semantics relatedness of descriptions and situations; (iv) functional updates to existing types allow us to account for how new types are learned from observing new information. In this presentation we demonstrate how these mechanisms can capture the referential aspects of spatial descriptions discussed in the previous paragraph. The benefit of applying TTR in this domain is that it provides a new unified model of the meaning representation of spatial descriptions that can be related to other work in natural language semantics in this theory, for example the study of dialogue. It is also fully computationally implementable. Our ongoing goal is to use it as a knowledge representation in situated conversational agents.

References

Robin Cooper and Staffan Larsson. 2009. Compositional and ontological semantics in learning from corrective feedback and explicit definition. In *Proceedings of the 13 the Workshop on the Semantics and Pragmatics of Dialogue (SemDial: DiaHolmia)*, pages 59–66.

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.

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. Elsevier BV.

Kenny R Coventry and Simon C Garrod. 2004. *Saying, seeing, and acting: the psychological semantics of spatial prepositions*. Psychology Press, Hove, East Sussex.

Simon Dobnik, Robin Cooper, and Staffan Larsson. 2013. Modelling language, action, and perception in Type Theory with Records. In Denys Duchier and Yannick Parmentier, editors, *Constraint Solving and Language Processing: 7th International Workshop, CSLP 2012, Orléans, France, September 13–14, 2012, Revised Selected Papers*, volume 8114 of *LNCS*, pages 70–91. Springer.

Simon Dobnik, John D. Kelleher, and Christos Koniaris. 2014. Priming and alignment of frame of reference in situated conversation. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (DialWatt)*, Edinburgh, 1–3 September.

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.

Representing different kinds of reference with Type Theory with Records

Simon Dobnik, Robin Cooper & Staffan Larsson
{simon.dobnik@, robin.cooper@ling, staffan.larsson@ling}.gu.se
Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden

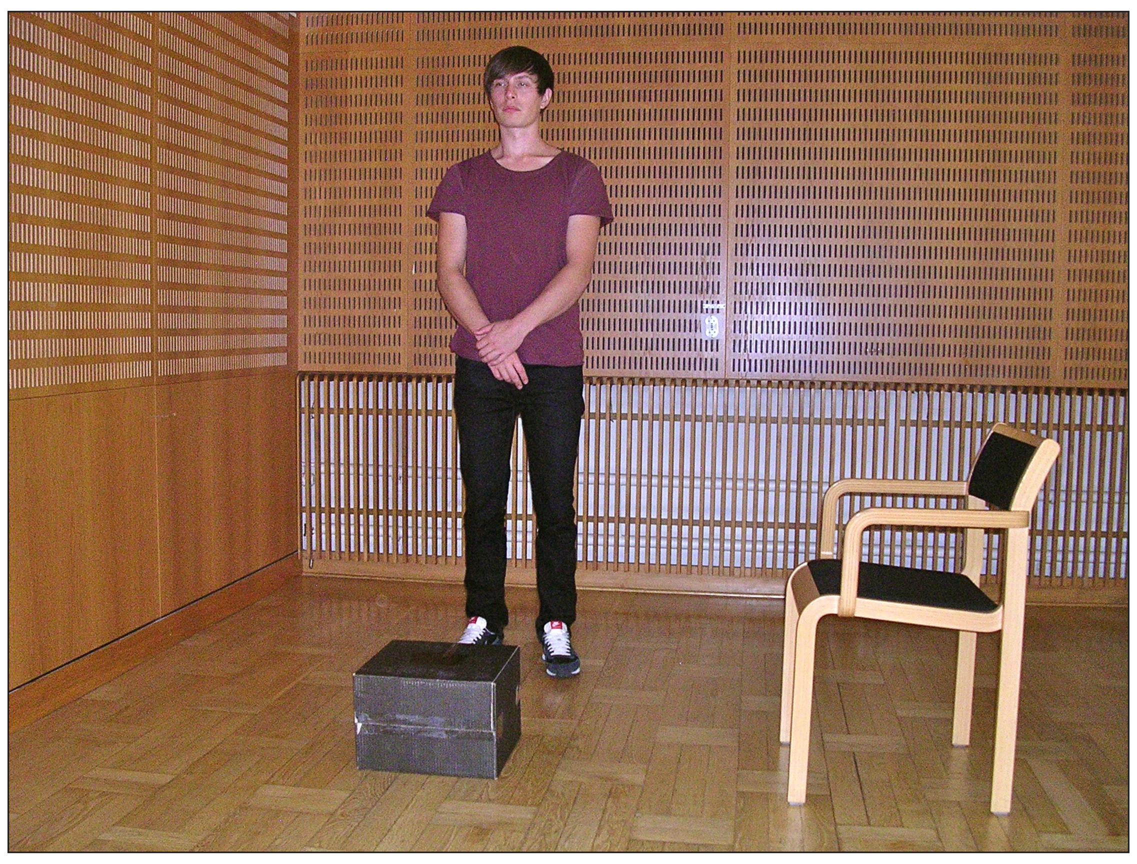
Aims

- How does a situated agent moving through its environment and making successive linguistic and perceptual observations assign meaning and reference?
- Explore meaning and reference of spatial descriptions in Type Theory with Records (TTR) (Cooper, 2012) which accounts for:
  - how linguistic and perceptual meaning and reference interact,
  - how reference is regulated,
  - how meaning is updated (learned) in light of new experience.

The referential nature of spatial descriptions

Spatial descriptions such as "the box to the left of the chair" establish (at least) three kinds of reference with their linguistic and physical context:

- reference to geometric properties of the scene;
- reference to function and typical use of objects;
- reference to dialogue history.



Reference is regulated by (probabilistic) judgements that situations and objects are of a particular type (in symbols, $a : T$) (Cooper et al., 2014).

The rich type system of TTR allows us to design representations from perceptual sensory readings to discourse.

From sensory readings to concepts

Proofs objects of record types are records which include sensor readings (verification).

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

Functions are applied to records of the required types or types.

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

where

$$f_{objclass} : \begin{bmatrix} \text{a} & : \text{Ind} \\ \text{sr} & : \text{list(list(Real))} \\ \text{loc} & : \text{list(Real)} \\ \text{reg} & : f_{pointmap}(r) \end{bmatrix} \rightarrow \text{Type}$$

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

Each agent builds their own inventory of objects but as they share the environment (and there are no classifier errors) their types are identical.

The interplay of geometry and typical use of objects/function

The objects of prepositions add world-knowledge or functional constraints to spatial relations.

Such constraints can be captured by automatic clustering the types of objects occuring with a particular preposition into hypernym classes (see (Dobnik and Kelleher, 2014)).

If r : alex(a) ∨ sam(a) then

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

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

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

Hyponym/hypernym relations can be conveniently expressed in TTR with sub-typing (e.g. ocean ⊑ body of water ⊑ thing ⊑ physical entity ⊑ entity) allows us to relate different representations.

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

- A spatial description such as "in" may be associated with several distinct types of situations.
- Each type of situations involves complex gemetric and conceptual knowledge.
- Each class of situations requires a unique geometric representation.

Accommodating frame of reference (FoR)/Perspective

Perspective can be assigned by any contextually salient object, including the speaker and the hearer.

Agents in conversation align to the primed FoR and continue to use it in the conversation (Dobnik, Kelleher, and Koniaris, 2014).

Speakers initiating conversation tend to be egocentric: they generate description from their point of view (private.for-origin=objects[0] : *Object*).

Hearers assume this strategy (private-for-origin=last-move.c_s.speaker/2 : *Object*).

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

$$\begin{bmatrix} \text{private} : \begin{bmatrix} \text{t=1} & : \textit{Time} \\ \text{agenda=} \left[\begin{bmatrix} \text{m:}\textit{Assertion} \\ \text{cnt=[beliefs[0]]:list(}\textit{RecType}\text{)} \end{bmatrix} \ldots \right] & : \text{list(}\textit{DMove}\text{)} \\ \text{beliefs= [[s}_1\text{:left(objects[2],objects[3])], [s}_1\text{:me(objects[0])]]} & : \text{list(}\textit{RecType}\text{)} \\ \text{objects= [o}_0\text{, o}_1\text{, o}_2\text{, o}_3\text{]} & : \text{list(}\textit{Object}\text{)} \\ \text{for-origin=objects[0]} & : \textit{Object} \end{bmatrix} \\ \text{shared} : \begin{bmatrix} \text{in-focus=private.objects[2]} : \textit{Object} \end{bmatrix} \end{bmatrix}$$

Sam: Aha.

$$\begin{bmatrix} \text{private} : \begin{bmatrix} \text{t=2} & : \textit{Time} \\ \text{agenda= []} & : \text{list(}\textit{DMove}\text{)} \\ \text{beliefs= [[s}_1\text{:me(objects[1])]]} & : \text{list(}\textit{RecType}\text{)} \\ \text{objects= [o}_4\text{, o}_5\text{, o}_6\text{, o}_7\text{]} & : \text{list(}\textit{Object}\text{)} \end{bmatrix} \\ \text{shared} : \begin{bmatrix} \text{in-focus=private.objects[2]:}\textit{Object} & \\ \text{last-move=} \left[\begin{bmatrix} \text{m:}\textit{Assertion} & \\ \text{c}_\text{s}\text{=speaker(m,private.objects[0])} & : \textit{Object} \\ \text{c}_\text{h}\text{=hearer(m,private.objects[1])} & : \textit{Object} \\ \text{cnt=[[p}_1\text{:left(private.objects[2],... [3])]]} & : \text{list(}\textit{RecType}\text{)} \end{bmatrix} \right] & : \text{list(}\textit{DMove}\text{)} \\ \text{beliefs=[[s}_1\text{:last-move.cnt]]} & : \text{list(}\textit{RecType}\text{)} \\ \text{for-origin= last-move.c}_\text{s}\text{.speaker/2} & : \text{Object} \end{bmatrix} \end{bmatrix}$$

References

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

Cooper, Robin, 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.

Dobnik, Simon 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.

Dobnik, Simon, John D. Kelleher, and Christos Koniaris. 2014. Priming and alignment of frame of reference in situated conversation. In *Proceedings of the 18th Semdial Workshop (DialWatt)*, Edinburgh, 1–3 September.