Toward perceptually grounded formal semantics

Simon Dobnik, Staffan Larsson, Robin Cooper

Department of Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden

GÖTEBORGS UNIVERSITET

Aims

Computational models of dynamic meaning in dialogue

Computational models of symbolic and perceptual meaning

Traditional formal semantics

Model-theoretic semantics does not deal with perception and dynamic meaning.

#1: What is the norm for set membership?

This is a green ball.

True if in a given model the referent a_{37} is a member of a set containing green objects $F(\text{green}) = \{\ldots, a_{37}, \ldots\}$.

#2: Norm affected by perception (geometry)

The chair is to the left of the table.

True if in a given model the referents a_{56} and b_{61} can be found in the set of pairs defined by $F(\text{left}) = \{\ldots, \langle a_{56}, b_{61}\rangle, \ldots\}$.

#3: Competition of norms: geometry *vs.* function

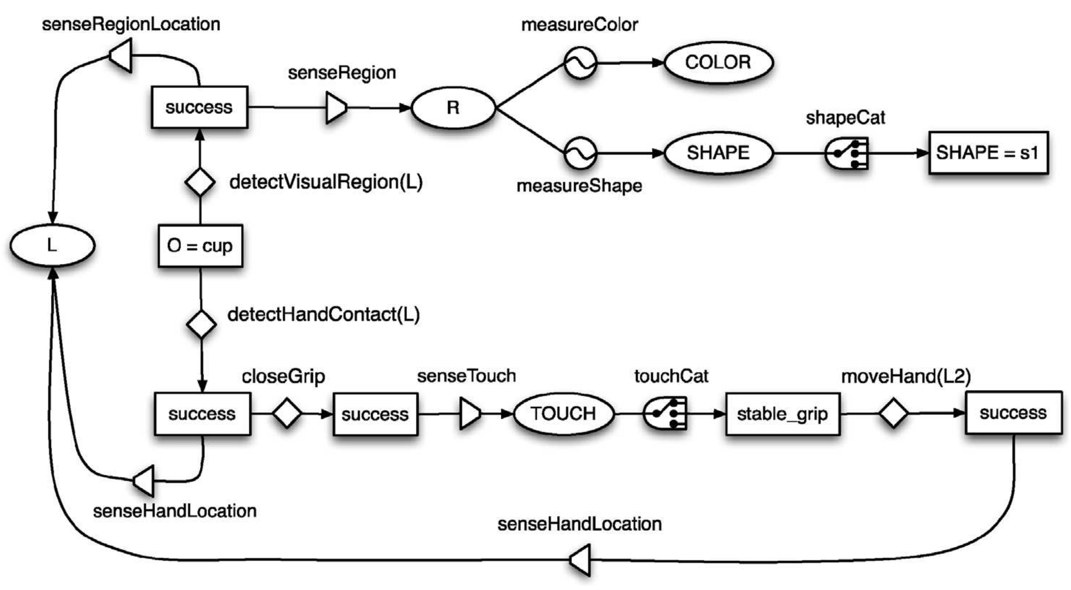
The umbrella is over a man.

#4: Dynamic norm

A: *I like bears.*
...
A: *That's a nice bear.*
B: *Yes, it's a nice* panda.
A: *Panda.*

Grounding language in vision

Conceptual categories (*a cup*) are defined by action and perception (Roy 2005, p.190).



No account of how distributional and categorical meaning is:

- composed;
- reasoned about;
- compared;
- modified: refined or generalised.

Type Theory with Records (TTR)

Types are intensional categories.
Perception is assignment to types.
Agents may have different type systems.

$$\begin{array}{ll} a & : \text{Apple} \\ \text{Proof objects} & : \text{Type} \end{array}$$

Types may have a more complex structure: **record types**.

$$\left[\begin{array}{lcl} \text{x} & : & \text{Ind} \\ \text{c}_{a-shape} & : & \text{apple-shape(x)} \\ \text{c}_{a-colour} & : & \text{apple-colour(x)} \\ \text{c}_{a-taste} & : & \text{apple-taste(x)} \\ \text{c}_{a-smell} & : & \text{apple-smell(x)} \end{array}\right]$$

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

$$\left[\begin{array}{lcl} \text{a} & = & \text{ind}_{26} \\ \text{sr}_{img} & = & [[34,24],[56,78]\ldots] \\ \text{c}_{sf} & = & [[45,78],[63,12]] \end{array}\right] : \left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{sr}_{img} & : & \text{RVector}^n \\ \text{c}_{sf} & : & \text{prf(sfocus(sr}_{img}\text{,a))} \end{array}\right]$$

Functions are applied to records of the required types:

$$\lambda r{:}\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{sr}_{img} & : & \text{RVector}^n \\ \text{c}_{sf} & : & \text{prf(sfocus(sr}_{img}\text{,a))} \\ \pi & : & \text{ClassifierKnowledge} \end{array}\right] \ldots$$

$$\left(\left[\text{c}_{a-shape} = \left[\begin{array}{lcl} \text{a} & = & \text{r.a} \\ \text{sr}_{img} & = & \text{r.sr}_{img} \\ \text{c}_{sf} & = & \text{r.c}_{sf} \\ \pi & = & \text{r.}\pi \end{array}\right] : f(\text{r.sr}_{img}, \text{r.}\pi, \text{r.a})\right]\right)$$

Classification is a mapping from sensory readings to types:

$$f(\text{r.sr}_{img}, \text{r.}\pi, \text{r.a}) = \begin{cases} \text{apple-shape(r.a)} \vee \\ \neg\ \text{apple-shape(r.a)} \end{cases}$$

If something is apple-shaped, it might be an apple (cf. enthymemes).

$$\lambda r{:}\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{c}_{a-shape} & : & \text{apple-shape(a)} \end{array}\right] \ldots \left(\left[\text{c}_{apple} \;:\; \text{apple(r.a)}\right]\right)$$

The more constraints can be verified/grounded, the higher the certainty that an individual is an apple.

TTR and dynamic meaning

The meaning in TTR can be updated as agents interact in dialogue.
Each agent has its own **take on a situation**.
Agents coordinate meaning.

Coordinating perceptual knowledge

R: *The chair is to the left of the table.*
H: *The chair is behind the table.*
R: *OK.*

Initially, the robot classifies every relation as "to the left of".

$$\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{b} & : & \text{Ind} \\ \text{l}_a & : & \text{loc(a)} \\ \text{l}_b & : & \text{loc(b)} \\ \pi & : & \text{ClassifierKnowledge} \\ \text{c}_{rel} & : & f(\text{l}_a,\text{l}_b,\pi) = \left\{\ \text{left(a,b)}\right. \end{array}\right]$$

The corrective feedback from a human is used to update the relation type:

$$\pi' = \text{retrain\_classifier}(\pi, \text{l}_a, \text{l}_b, \text{behind})$$

$$f(\text{l}_a,\text{l}_b,\pi) = \begin{cases} \text{left(a,b)} \vee \\ \text{behind(a,b)} \end{cases}$$

Coordinating symbolic knowledge

The robot does not know about tables yet and the perceptional knowledge alone may not be enough to distinguish between tables and chairs.

$$\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{c}_{ch} & : & \text{chair(a)} \\ \text{c}_{ot} & : & \text{class(a,furniture)} \\ \text{c}_{osh} & : & \text{chair-shape(a)} \end{array}\right]$$

R: *This is a chair.*
H: *No, it's a table.*
R: *A table.*
H: *One sits on a chair but one keeps food on a table.*
R: *Aha.*

The object shape classifier is updated so that it also evaluates to table-shape(a) (π' = retrain_classifier(π, RVector^n,table)) and a new record type for tables is created. New categorical type constraints are also added.

$$\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{b} & : & \text{Ind} \\ \text{c}_{ch} & : & \text{chair(a)} \\ \text{c}_{ot} & : & \text{class(a,furniture)} \\ \text{c}_{ot} & : & \text{class(b,human)} \\ \text{c}_{s} & : & \text{provides\_support(a,b)} \\ \text{c}_{osh} & : & \text{chair-shape(a)} \end{array}\right]$$

$$\left[\begin{array}{lcl} \text{a} & : & \text{Ind} \\ \text{b} & : & \text{Ind} \\ \text{c}_{tb} & : & \text{table(a)} \\ \text{c}_{ot} & : & \text{class(a,furniture)} \\ \text{c}_{ot} & : & \text{class(b,food)} \\ \text{c}_{s} & : & \text{provides\_support(a,b)} \\ \text{c}_{osh} & : & \text{table-shape(a)} \end{array}\right]$$

Categorical world knowledge is useful for directing visual search and modelling object affordances.

References

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

Cooper, Robin. 2005. Austinian truth, attitudes and type theory. *Research on Language and Computation* 3:333–362.

Larsson, Staffan. 2011. The TTR perceptron: Dynamic perceptual meanings and semantic coordination. *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011 - Los Angelogue).* September 21–23, 2011 Los Angeles, California. 140–148.