Universität Stuttgart

How to combine speech and action

Tillmann Pross

Overview

This poster presents a conservative extension of Discourse Representation Theory with anchors for temporal and non-temporal discourse referents in which the semantics of predicates is specified with the help of a branching-time account of action [Pross 2010]. The semantic processing of verbal and non-verbal content is embedded into a real-time architecture for the agent-based control of action and thought (a so-called Belief-Desire-Intention (BDI) Interpreter, [D'Inverno et. al. 2004]) and instantiated on a robot platform.

The basic problem: Combining speech and action

- Consider Searle's classical account of speech acts:
- The pragmatics of sentences in imperative mood (where R stands for the subject and P for the predicate of the imperative) [Searle 1969]:
  "If the sentence is imperative, its imperative illocutionary force indicating device (F term) determines that the object referred to by the R term is to do the act specified by the P term, and so on through other examples" (p. 122)
- **But**: How does a predicate P specify an act?
  "to know the meaning of a (...) predicate expression is to know under what conditions it is true or false of a given object." (ibid., p. 125)
- **But**: This type of truth-conditional account of the meaning of predicates fails to specify how a predicate specifies an act.
- Can we define a truth-conditional semantics for predicates such that the meaning of predicates specifies actions?

The Framework of Grounded Discourse Representation Theory

Consider the following example setup. Fred and robot Clara are situated at a table. On the table is a cube, a slat and a screw. Fred utters (1) to Clara.

(1) Fred: Give me the cube!

- (1) is a request for Clara to choose from her future possibilities for action that course of action which renders true that Fred has been given a cube.
- The framework adopted for analysis of the example makes use of the following machinery to relate the Discourse Representation Structure for (1) with actions:
  - Temporal Anchors: The semantics of events in DRT is stated in terms of anchors that specify branching temporal structures (corresponding to plans of agents) for particular events.
  - BDI-based control of the interpretation of DRSs and sensorimotor action (i.e. the interpretation of a DRS is a complex action that an agent performs)
  - Dynamic models: The model structures that the agent uses to evaluate DRSs are the same information structures that the BDI-based control algorithm manipulates and employs in order to deal with the dynamic evolution of the agent's internal BDI-configuration and the agent's environment. That is, models can change in response to e.g. the interpretation of DRSs, the adoption of new intentions or the perception of new configurations of the environment.
- The interpretation process of DRSs is based on the resolution of 'variable' anchor sources for discourse referents, i.e. a speech act is considered as the transmission of a DRS with unresolved anchor sources (i.e. unspecified reference) in order to move the interpretation task of the DRS (i.e. the resolution of the anchors of the DRS) from the speaker to the heaer.
- Two possible options exist for the interpretation of DRSs (roughly corresponding to the distinction between conceptual and procedural meaning drawn in [Sperber and Wilson 1993]):
  - Plain interpretation: the anchors of the DRS can be resolved with respect to the current configuration of the model of the interpreting agent.
  - Reactive interpretation: the anchors of the DRS can not be resolved to the current model. Consequently, the interpreting agent is supposed to transform the model according to the actions specified by the temporal anchors of the DRS in a way such that the anchors of the DRS can be resolved.
- The distinction between plain and reactive interpretation assings DRSs not only a meaning in terms of their context change potential but also in terms of their 'model change potential'.

Action-based predicate semantics

In the following, **Things** is a set of individuals, **S** assigns scenarios, **T** branching temporal structures and **Attitude** attitudes to agents x at a time t, **Prop** is a function from properties to individuals and SEM is a set of lexicalized annotations of particular individuals and events with properties resp. temporal structures.

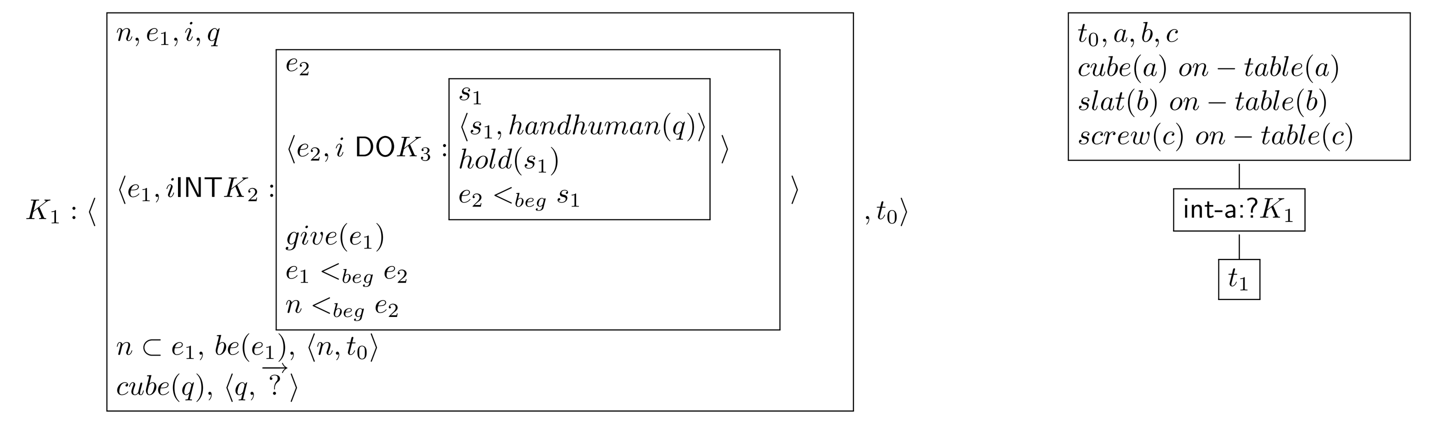
- $\langle s, R(x_1, \ldots, x_n)\rangle \vDash_{M,t} handle(s)$
  plain: iff $\exists G = \{\langle x_1, a_1\rangle, \ldots, \langle x_n, a_n\rangle\}$ sth.
  $SEM_{handle}(a_1, \ldots, a_n) \in$ **Prop**$(a_1, \ldots, a_n)(t)$;
  where $SEM_{handle}(a_1, \ldots, a_n)$ is an annotation of a particular individual with properties
  reactive: belief-add$(x, t, SEM_{handle}(a_1, \ldots, a_n))$
- $\langle e, x\text{INT}K\rangle \vDash_{M,S,P,t} handle(e)$
  plain: iff $[K]_{M,t} \in$ **Attitude**(Int, x, t);
  reactive: intention-add$(x, SEM_{handle}(e))$
- $(n \subseteq e)\ \langle e, x\text{DOK}\rangle \vDash_{M,S,P,t} handle(e)$
  plain: iff $\exists[S; t_0, n] \in$ **S**$(x)(t)$ and $\exists[P; n, t_1] \in$ **T**$(x)(n)$ sth.
  $(S \cup P) \in SEM_{handle}(e)$ and $\vDash_{M,t_1} K$;
  where $SEM_{handle}(e)$ is an annotation of a particular event with a temporal structure
  reactive: goal-add$(x, SEM_{handle}(e))$

Analysis of "Give me the cube!"

"Lexicalized" anchors and annotations used in the example analysis:

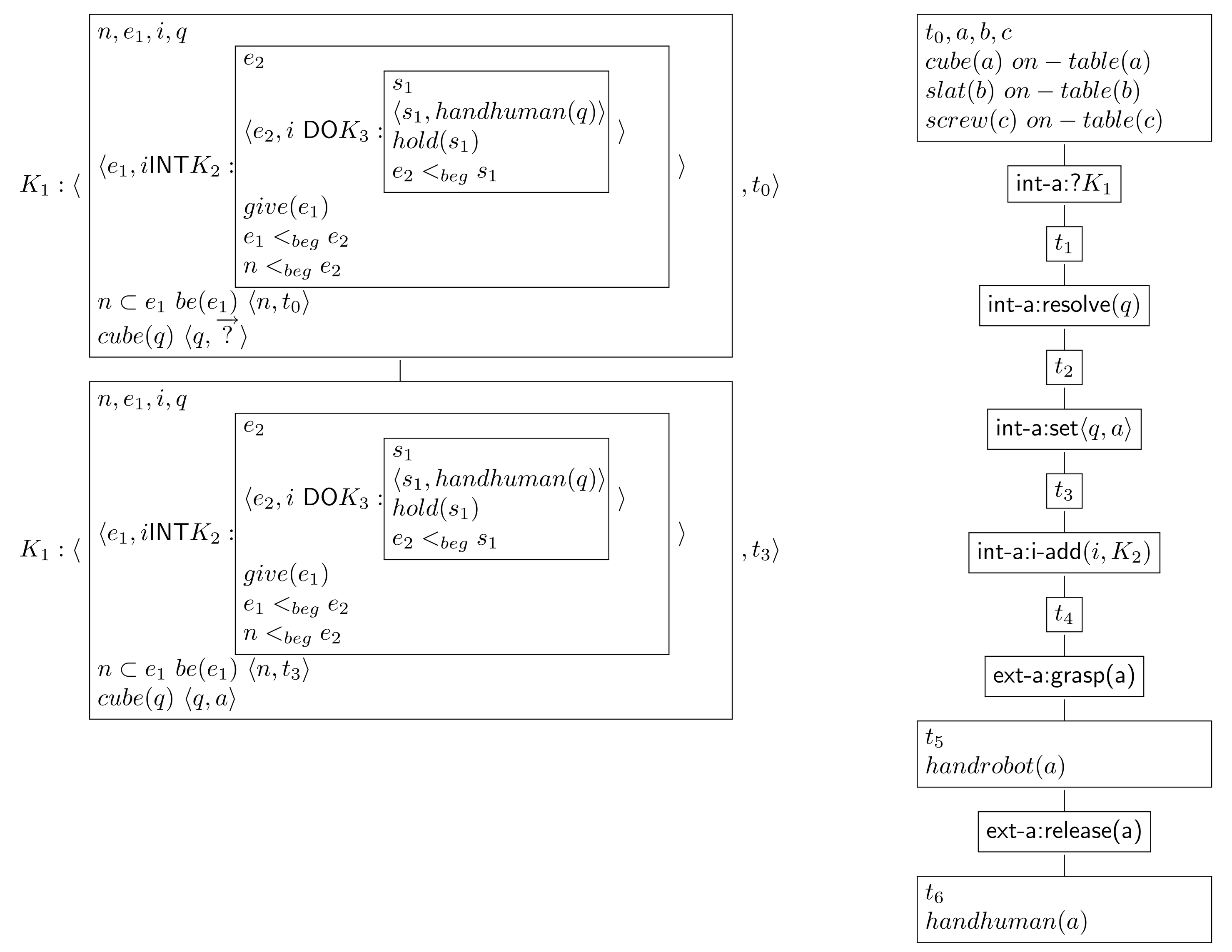
handle	DRS	SEM
'cube'	[x | $cube(x)$, $\langle x, g\rangle$]	[g | $cube(g)$]
'give'	[x, e | $give(e)$, $\langle e, x$DO [y, s | $\langle s, handhuman(y)\rangle$, $\langle y, d\rangle$] $\rangle$]	t_0 - ext-a:grasp(d) - t_1 - ext-a:release(d) - t_2 : $handhuman(d)$

Example: Initial Configuration



The cognitive state of Clara at t_1, $CS(Clara)(t_1)$. Left: Representation K_1 of utterance (1) "Give me the cube". q is a discourse referent for 'cube', i is a discourse referent for Clara, e_1 and e_2 are discourse referents for events - e_2 stands for 'give', e_1 represents that e_2 is the content of an intention. s_1 represents the goal state of 'give'. n represents the current now of Clara at the time at which K_1 was added to $CS(Clara)(t)$. Right: Clara's Model at t_1, $MODEL(Clara)(t_1)$, the information structure against which Clara evaluates K_1 (represented as an atomic action int-a?K_1).

Example: Final Configuration



Modelling of Clara's first-person perspective on the interpretation of example (1) ("Give me the cube"). The figure shows Clara's cognitive structure at t_6, $CS(Clara)(t_6)$. Earlier stages of processing are recorded in $CS(Clara)(t_6)$, representing a discourse history. Clara's interpretation attempt int-a:?K_1 at t_1 invokes a plan that pushes K_1 to the list of DRSs to be interpreted. Next, it is checked whether K_1 contains variable anchor sources. The variable anchor source '?' for q in K_1 triggers a plan int-a:resolve(q) for the resolution of this source. The anchor source for q is resolved at t_3 (under consideration of the definiteness constraint on q) to $a \in MODEL(Clara)(t_2)$. Once the variable anchor source for q has been resolved to a, K_1 is passed over to the main interpretation process. As a plain interpretation of K_1 fails at t_3 - up to t_3 there is no temporal structure in which the SEM of 'give' could be embedded - a reactive interpretation of K_1 is executed. This results in an extension of Clara's model at t_3 with the temporal annotation SEM for 'give'. As the temporal structure $t_4 - t_5 - t_6$ is added to Clara's intentions by the command i-add, K_1 can be anchored in $MODEL(Clara)(t_4)$ and Clara realizes a successful interpretation of (1). In turn, Clara's BDI-interpreter executes the adopted intention such that Fred has the cube in his hands at t_6.

References

[D'Inverno et. al. 2004] D'Inverno, M. D.; Luck, M.; Georgeff, M. P.; Kinny, D. and Wooldridge, M. J.: The dMARS Architecture: A Specification of the Distributed Multi-Agent Reasoning System. Autonomous Agents and Multi-Agent Systems 9:5-53.

[Pross 2010] Pross, T.: Grounded Discourse Representation Theory. Towards a semantics-pragmatics interface for human-machine collaboration. PhD Thesis, Institute for Natural Language Processing, University of Stuttgart, 2010.

[Searle 1969] Searle, J. R.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, 1969.

[Sperber and Wilson 1993] Sperber, D. and Wilson, D.: Linguistic Form and Relevance. Lingua 90(2):1-25.

Institute for Natural Language Processing