Grounded event semantics for robots

Tillmann Pross

University of Stuttgart, Institute for Natural Language Processing

July 15

Outline

Plan

1. Motivation
   - Sample Setup
   - Events in Formal Semantics
   - Starting Point
2. Representing the robot's internal configuration
   - Sensorimotor Control
   - Adding a future
   - Branching Time Structure
3. Event Segmentation in Cognitive Science
4. Event Segmentation in DRT with Anchors
   - Branching-Time Semantics for DRT
   - DRS Interpretation: Direction of Fit
   - Example: Give me the cube!
5. Beyond Robots: Prospects and Problems

Clara

European research project 'JAST', TU Munich

Joint Action Task

Construct objects from a Baufix construction kit in collaboration of a human and a robot.

Example

Human Fred and robot Clara are situated at a table. On the table is a cube and a slat.
What is the 'meaning' that Clara should assign to (1) in order to response correctly?

(1) Fred: Give me the cube!

Event semantics

In formal semantics, events are atomic theoretical terms of a logical form formalism:

(2) a. $\lambda x \lambda y \lambda z \lambda e. give(e) \wedge Agent(e,x) \wedge Theme(e,y) \wedge Goal(e,z)$

b. $g \vDash_M e : R(x_1, \ldots, x_n)$ iff $\langle g(e), g(x_1), \ldots, g(x_n) \rangle \in I(R)$

Events described by occurrences of 'give me the cube' are events that stand in some 'agent'-relation to the one who is doing the giving, some 'theme'-relation to the thing that is given and some 'goal'-relation to the one who is given the thing.

Starting Point

- What is an event?
- Humans have an intuitive understanding of what events in general and the event described by “give the cube” are (but this does not imply agreement in theorizing about events, see e.g. Casati and Varzi [1996])
- Robots have no such intuitive understanding of events and explanation of events such as “give the cube” in terms of further symbols won't help out ('symbol grounding problem' Searle [1980])

Why is this interesting at all?

“From a theoretical perspective, one of the major attractions of robotic implementations – in contrast to purely theoretical discussions of potential mechanisms in human – is that the entire grounding mechanism needs to be spelled out clearly.” Thill et al. [2014]

(This issue occupied my mind for quite some time.)

Motivation for studying Computational Linguistics, Studienarbeit, Diplomarbeit, Dissertation and some papers

(Pross [2005, 2006, 2010a,b, 2011, 2014])

Objectives

→ Extend the sensomotoric and planning capabilities of a robot with the ability to construct, maintain and interpret complex symbolic representations of events.

← Ground complex symbolic representations of events in the sensomotoric control and planning of a robot.

Plan

1. Motivation
    - Sample Setup
    - Events in Formal Semantics
    - Starting Point
2. Representing the robot's internal configuration
    - Sensorimotor Control
    - Adding a future
    - Branching Time Structure
3. Event Segmentation in Cognitive Science
4. Event Segmentation in DRT with Anchors
    - Branching-Time Semantics for DRT
    - DRS Interpretation: Direction of Fit
    - Example: Give me the cube!
5. Beyond Robots: Prospects and Problems

Interface to Sensorimotor Control

- Starting point: Interface specification of object recognition and motor control provided by the engineers.
- No need to bother about the granularity of motor actions and issues of object recognition
- Additional assumption here: NLP pipeline deals with construction and generation of semantic representations

Sensorimotor Structure (SMS)

(3)

s_1

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SLAT, L - TABLE\}$

A-GRASP(β)

s_2

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SLAT, L - HANDROBOT\}$

Incorporating Rational Agency

"The fundamental purpose of brains is to produce future."
[Dennett, 1991, p. 177]

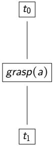
- Sensorimotor control deals with the perceptions and the motor actions of the robot.
- What the sensorimotor control does not provide is a concept of the future.
- A concept of the future is necessary for the robot in order to interact with its environment in a rational way (e.g. to achieve goals).

Intentions as choice with commitment

- Planning theory of intention (Bratman [1987], Cohen and Levesque [1991])
- Deliberate about partially specified plans and commit to the execution of one of the underspecified plans.
- Computer Science/Distributed Artificial Intelligence:
  Belief-Desire-Intention (BDI) Model of Rational Agency (Singh et al. [1999], Inverno et al. [2004])

Example Plan: grasp an object

(4)

Type:	grasp
Invocation:	an object is perceived
Context:	the perceived object can be reached
Feedback:	handrobot(a)
Body:	

Control Algorithm

```
do
   options:=option-generator(perceptions,B,G,I);
   selected-options:=deliberate(options,B,G,I);
   update-intentions(selected-options,I);
   execute(I);
   get-new-perceptions();
   drop-successful-attitudes(B,G,I);
   drop-impossible-attitudes(B,G,I);
until quit.
```

Rational Control Structure

The BDI-configuration of the robot is represented in a Rational Control Structure (RCS).

- A Rational Control Structure has a tree-like structure of annotated times:
    - The trunk of the tree is defined by translating the sensorimotor information into times annotated with the perceived state of affairs
    - The branches of the tree are defined by the options of future action generated by the Belief-Desire-Intention Interpreter
    - Division of past and future by the 'actuality' of the SMS
    - Future times and actions are indicated by primes

Acquisition of Grounding

- Times: $s_i \rightarrow \langle t_n, s_i \rangle$
- Time Annotations:

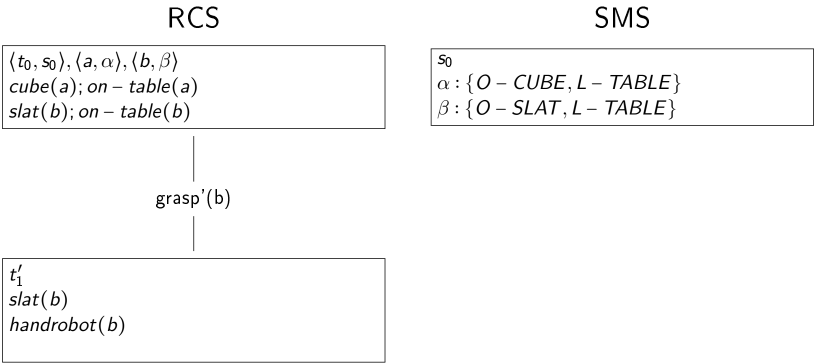
$$\alpha : \{O-CUBE, L-TABLE\} \rightarrow \boxed{\begin{array}{l} \langle a, \alpha \rangle \\ cube(a) \\ on-table(a) \end{array}}$$

- Transitions between Times: $A-GRASP(\alpha) \leftrightarrow grasp(a)$

Example: Grasp a slat

Sensing an object invokes the plan for grasping it. Deliberation adds grasping the slat as a new intention. The respective planning structure is added to Control Structure as the 'future' of the robot.

(5)



RCS

$\langle t_0, s_0\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a); on-table(a)$
$slat(b); on-table(b)$

grasp'(b)

t_1'
$slat(b)$
$handrobot(b)$

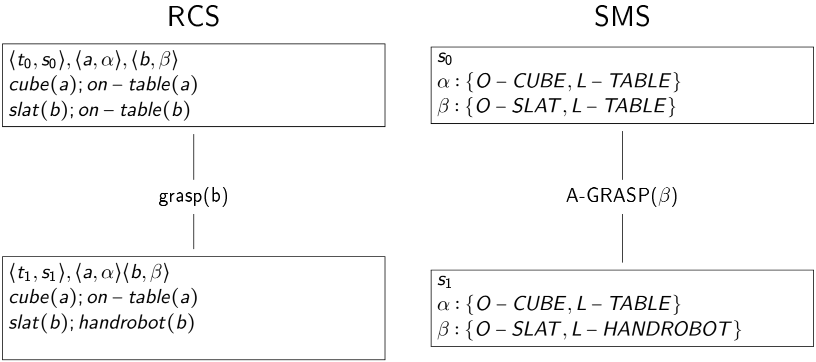
SMS

s_0
$\alpha : \{O-CUBE, L-TABLE\}$
$\beta : \{O-SLAT, L-TABLE\}$

Grasp a slat

Intention to grasp the slat has been realized by execution of the plan for grasping.

(6)



RCS

$\langle t_0, s_0\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a); on - table(a)$
$slat(b); on - table(b)$

grasp(b)

$\langle t_1, s_1\rangle, \langle a, \alpha\rangle\langle b, \beta\rangle$
$cube(a); on - table(a)$
$slat(b); handrobot(b)$

SMS

s_0
$\alpha : \{O - CUBE, L - TABLE\}$
$\beta : \{O - SLAT, L - TABLE\}$

A-GRASP(β)

s_1
$\alpha : \{O - CUBE, L - TABLE\}$
$\beta : \{O - SLAT, L - HANDROBOT\}$

Formal Definition of a Rational Control Structure

A Rational Control Structure defines a grounded branching structure of annotated times $E = \{\mathbf{T}, I, Actions\}$ of an agent x at time t, where

- $\mathbf{T} = \langle <, \mathbf{Times} \rangle$ is a branching structure of times of an agent x at time t.
- I associates times $t \in \mathbf{Times}$ with annotations, i.e. I is a function from times to time annotations.
- *Actions* is a set of atomic actions, where each element of *Actions* constitutes a transition between two times.

Underlying formalism: Computational Tree Logic Emerson [1990] applied to Multi-Agent-Systems Singh [1994]

Plan

Identifying events

- How do events relate to the branching structure of times?
- More precisely: How can a robot identify a certain portion of times *as* an event?
- How does a person come to perceive that one meaningful event has ended and another has begun?

Event segmentation

- Experiments in Cognitive Science show that ongoing activities are automatically and spontaneously segmented into hierarchically organized parts (e.g. Zacks and Iyer [2001], Zacks and Swallow [2007]).
- These hierarchically organized segments of 'temporal variation' correspond to discrete events.
- What are the hierarchical structures used to segment events?

Event segmentation

- Events are segmented with the help of "bottom-up processing of sensory features such as movement and [. . .] top-down processing of conceptual features such as actors' goals." [Zacks and Swallow, 2007, p. 80]
- Top-down processing is preferred by the subjects of the experiments (just because language encodes the top-down structuring of events with goals).

Bottom-Up event segmentation

How do people structure their everyday activities? Behaviour episodes [Barker and Wright, 1954, p.236]:

1. Change in the “sphere” of the behavior from verbal to physical to social to intellectual, or from any one of these to another.
2. Change in the part of the body predominantly involved in a physical action [. . .]
3. Change in the physical direction of the behavior. [. . .]
4. Change in the behavior object “commerced with” [. . .]
5. Change in the present behavior setting. [. . .]
6. Change in the tempo of activity [. . .]

Top-Down event segmentation

Top-Down structuring of events by hierarchical decomposition of goals
High complexity of activity requires high complexity of structuring

- Plans/Intentions
- Behaviour/Desires
- Causes/Effects

⇒ An agent uses for the same hierarchical structure of actions to plan his own actions and to segment events.

Plan

1. Motivation
    - Sample Setup
    - Events in Formal Semantics
    - Starting Point
2. Representing the robot's internal configuration
    - Sensorimotor Control
    - Adding a future
    - Branching Time Structure
3. Event Segmentation in Cognitive Science
4. **Event Segmentation in DRT with Anchors**
    - Branching-Time Semantics for DRT
    - DRS Interpretation: Direction of Fit
    - Example: Give me the cube!
5. Beyond Robots: Prospects and Problems

Anchors in DRT

- Implementation of event segmentation in DRT with anchors
- Kamp [1984]: anchors represent an agent's relation of acquaintance with an object in the outside world.
- Here, anchors are pairs of discourse referents (a 'floater') and RCS 'things' (a 'source'): $\langle x, a \rangle$

Anchors

Anchors are part of the lexical semantics of nouns such as 'cube', they define how the anchored discourse referent can be 'identified' in the Rational Control Structure.

SEM	$\langle x, a \rangle$ $cube(x)$
ID	$cube(a)$

Anchor Types

An anchoring relation $\langle \mathit{floater}, \mathit{source} \rangle$ is called

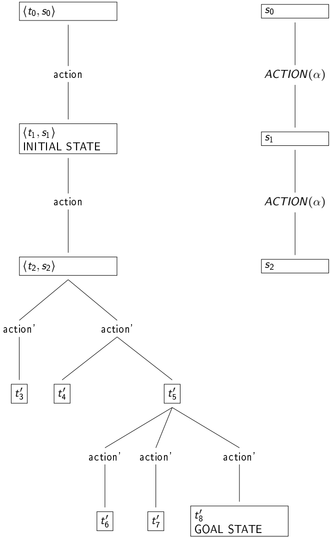
- internal iff its floater is a DRS-reference marker and its source a RCS-reference marker
- external iff its floater is a DRS-reference marker and its source is a RCS-reference marker anchored in a SMS object

Representing segmentation with anchors

- The anchor source of an event specifies the conditions under which the event can be segmented/'identified' in a RCS structure.

DRS/RCS/SMS

(7) $\langle$ $\boxed{\begin{array}{l}\langle e, \mathit{SEGMENTATION}\rangle \\ \text{Bottom-Up: Differences between } t_0\text{-}t_1 \text{ and } t_0\text{-}t_8 \\ \text{Top-Down: Goal } t_8 \text{ and Subgoals } t_2,\ t'_5\end{array}}$ $, t_1\rangle$



Temporal Anchors

Temporal anchors are part of the lexical semantics of verbs

SEM $\boxed{\begin{array}{l} \langle e,\ x\ \mathrm{DO}\ K_{goal} \rangle \\ name(e) \end{array}}$,

ID specifies the segmentation conditions for e in terms of

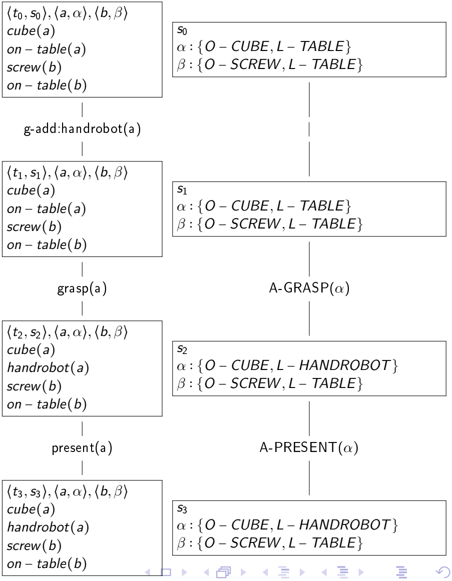
  - a tree-like structure of annotated times with leaves rendering the goal state described by K true.

Give an object

Type:	give
Invocation:	g-add($\boxed{handhuman(x)}$)
Context:	$\langle x, a \rangle$
Feedback:	$\boxed{handhuman(a)}$
Body:	$\boxed{t_0}$ $\boxed{g - add(handrobot(a))}$ $\boxed{t_1}$ $\boxed{present(a)}$ $\boxed{t_2}$ $\boxed{release(a)}$ $\boxed{t_3}$

“I am giving a cube”

(8)
$$\left\langle \boxed{\begin{array}{l} \langle e_1, \mathrm{iINT} \boxed{\begin{array}{l} \langle e_2, \mathrm{iDO} \boxed{handhuman(x)} \rangle \\ e_2 \subset e_1 \\ give(e_2) \end{array}} \rangle \\ cube(x) \\ \langle x, a \rangle \\ e_1 \subseteq n \\ \langle n, [t_0, t'_4] \rangle \end{array}}, t_3 \right\rangle$$



$\langle t_0, s_0\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a)$
$on-table(a)$
$screw(b)$
$on-table(b)$

s_0
$\alpha: \{O-CUBE, L-TABLE\}$
$\beta: \{O-SCREW, L-TABLE\}$

g-add:handrobot(a)

$\langle t_1, s_1\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a)$
$on-table(a)$
$screw(b)$
$on-table(b)$

s_1
$\alpha: \{O-CUBE, L-TABLE\}$
$\beta: \{O-SCREW, L-TABLE\}$

grasp(a)

A-GRASP(α)

$\langle t_2, s_2\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a)$
$handrobot(a)$
$screw(b)$
$on-table(b)$

s_2
$\alpha: \{O-CUBE, L-HANDROBOT\}$
$\beta: \{O-SCREW, L-TABLE\}$

present(a)

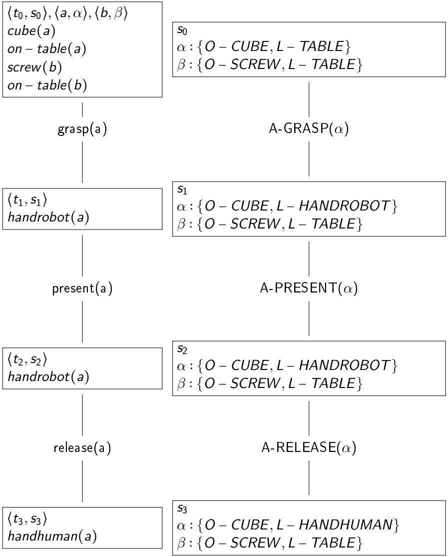
A-PRESENT(α)

$\langle t_3, s_3\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a)$
$handrobot(a)$
$screw(b)$
$on-table(b)$

s_3
$\alpha: \{O-CUBE, L-HANDROBOT\}$
$\beta: \{O-SCREW, L-TABLE\}$

“I gave a cube”

$$(9)\qquad \left\langle \boxed{\begin{array}{l} \langle e_2, i\text{CAUSE}\ \boxed{handhuman(x)}\ \rangle \\ \langle x, a\rangle \\ cube(x) \\ give(e_2) \\ e_2 < n \\ \langle n, t_3\rangle \end{array}}, t_3 \right\rangle$$



$$\boxed{\begin{array}{l} \langle t_0, s_0\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle \\ cube(a) \\ on-table(a) \\ screw(b) \\ on-table(b) \end{array}} \qquad \boxed{\begin{array}{l} s_0 \\ \alpha : \{O-CUBE, L-TABLE\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

grasp(a) A-GRASP(α)

$$\boxed{\begin{array}{l} \langle t_1, s_1\rangle \\ handrobot(a) \end{array}} \qquad \boxed{\begin{array}{l} s_1 \\ \alpha : \{O-CUBE, L-HANDROBOT\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

present(a) A-PRESENT(α)

$$\boxed{\begin{array}{l} \langle t_2, s_2\rangle \\ handrobot(a) \end{array}} \qquad \boxed{\begin{array}{l} s_2 \\ \alpha : \{O-CUBE, L-HANDROBOT\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

release(a) A-RELEASE(α)

$$\boxed{\begin{array}{l} \langle t_3, s_3\rangle \\ handhuman(a) \end{array}} \qquad \boxed{\begin{array}{l} s_3 \\ \alpha : \{O-CUBE, L-HANDHUMAN\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

Using the Rational Control Structure as a Model Theory

A given Rational Control Structure $E = \{\mathbf{T}, I, Actions\}$ of an agent x at time t defines

- The set of RCS times **Times**
- The set of RCS things **Things**
- The set of RCS properties **Properties**

Assignment Functions

- A function **Time** that assigns RCS structures to an agent τ at t, i.e. the time structure of an agent at t:
  $\mathbf{T}(\tau)(t)$
- A function *Plan* that assigns Plans (Structures) to an agent τ at t:
  $P(\tau)(t)$
- A function *Scenario* that assigns Scenarios (Branches) to an agent τ at t:
  $S(\tau)(t)$
- A function **Property** that assigns **Properties** to (tuples of) RCS-things $\langle a_1, \ldots, a_n \rangle$ at t:
  $\mathbf{P}(a_1, \ldots, a_n)(t)$

Assignment Functions

For a given BDI-interpreter instance for an agent x, we can define the following function:

- A function **Attitudes** that assigns attitudes of a certain type ϕ (Desires or Intentions) to an agent x at t:
  Attitudes$(\phi)(x)(t)$

Models for Grounded DRSs

- A model M at a time t of an agent x is a tuple
  - $M(x)(t) = \langle Plans, Scenarios, \mathbf{Time}, \mathbf{Property}, \mathbf{ID}, \mathbf{Things}, \mathbf{Attitudes}, \rangle$
- The notion of an 'embedding' or 'interpretation' of a DRS in M is replaced by a 'successful anchoring' of a DRS in M.
- Identifying a successful anchoring of a DRS is itself a complex action of the interpreting agent.

Identification of DRS-reference markers for physical objects

- $\langle x, a \rangle \vDash_{M,t} name(x)$ iff $ID_{name}(x) \in \mathbf{Property}(a)(t)$

Direction of Fit: Anchoring of DRSs

- Passive DRS interpretation
  - Anchor a DRS in a given Model
  - Word-to-world fit Austin [1962]
  - Conceptual meaning Sperber and Wilson [1993]
  - Declarative Semantics Gabbay [1987]
- Active DRS interpretation
  - Anchor a DRS by changing a given Model according to the ID-specification of involved time-individuals
  - World-to-word fit
  - Procedural Meaning
  - Imperative Semantics

Identification of DRS-reference markers for temporal objects: DO-segmentation

Present: $\langle e, xDOK\rangle \vDash_{M,S,P,t} name(e)$

- passive: iff $\exists[S;t,n] \in \mathbf{S}(x)(t)$ and $\exists[P;n,t_1] \in \mathbf{T}(x)(n)$ sth.
  $(S \cup P) \in ID_{name}(e)$ and
  $\vDash_{M,t_1} K$ and
  $[K]_{M,t} \in \mathbf{Attitude}(Do, x, t)$
- active: g-add$(x, ID_{name}(e))$

Future: $\langle e, xDOK\rangle \vDash_{M,S,P,t} name(e)$

- passive: iff $\exists[P;n,t_1] \in \mathbf{T}(x)(n)$ sth. $P \in ID_{name}(e)$ and
  $[K]_{M,t} \in \mathbf{Attitude}(Do, x, t)$
- active: g-add$(x, ID_{name}(e))$

Identification of DRS-reference markers for temporal objects: INT-segmentation

$\langle e, x\mathrm{INT}K \rangle \vDash_{M,S,P,t} name(e)$

- passive: similar to DO but $[K]_{M,t} \in \textbf{Attitude}(Int, x, t)$
- active: i-add$(x, ID_{name}(e))$

Variable anchor sources

DRSs constructed from utterances have variable 'unresolved' anchor sources:

- Specific reference
  - $\langle \mathit{floater}, \overrightarrow{\mathit{source}} \rangle$
- Unspecific reference
  - $\langle \mathit{floater}, ! \rangle$ External anchor
  - $\langle \mathit{floater}, ? \rangle$ Internal anchor
  - $\langle \mathit{floater}, ?a \rangle$ Anaphoric

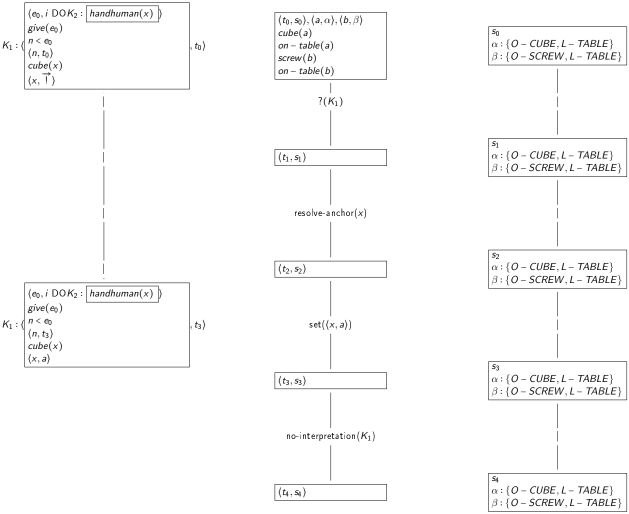
“Give me the cube!”

(10)

$$\left\langle \boxed{\begin{array}{l} \langle e_0, i \text{ DO } \boxed{handhuman(x)} \rangle \\ give(e_0) \\ n < e_0 \\ \langle n, t_0 \rangle \\ cube(x) \\ \color{red}{\langle x, \overrightarrow{!} \rangle} \end{array}}, t_0 \right\rangle \qquad \boxed{\begin{array}{l} \langle t_0, s_0 \rangle, \langle a, \alpha \rangle, \langle b, \beta \rangle \\ cube(a) \\ on - table(a) \\ screw(b) \\ on - table(b) \end{array}} \qquad \boxed{\begin{array}{l} s_0 \\ \alpha : \{O - CUBE, L - TABLE\} \\ \beta : \{O - SCREW, L - TABLE\} \end{array}}$$

$$(11)\qquad K_1 : \left\langle \boxed{\begin{array}{l} \langle e_0, i \text{ DO } \boxed{handhuman(x)} \rangle \\ give(e_0) \\ n < e_0 \\ \langle n, t_0 \rangle \\ cube(x) \\ \langle x, \overrightarrow{!} \rangle \end{array}}, t_0 \right\rangle \qquad \begin{array}{c} \boxed{\begin{array}{l} \langle t_0, s_0 \rangle, \langle a, \alpha \rangle, \langle b, \beta \rangle \\ cube(a) \\ on - table(a) \\ screw(b) \\ on - table(b) \end{array}} \\ | \\ ?'(K_1) \end{array} \qquad \boxed{\begin{array}{l} s_0 \\ \alpha : \{O - CUBE, L - TABLE\} \\ \beta : \{O - SCREW, L - TABLE\} \end{array}}$$

(12)



$K_1 : \langle$

$\langle e_0, i\ \mathrm{DO}\, K_2 : \boxed{handhuman(x)} \rangle$

$give(e_0)$

$n < e_0$

$\langle n, t_0 \rangle$

$cube(x)$

$\langle x, \overrightarrow{!} \rangle$

$, t_0 \rangle$

$K_1 : \langle$

$\langle e_0, i\ \mathrm{DO}\, K_2 : \boxed{handhuman(x)} \rangle$

$give(e_0)$

$n < e_0$

$\langle n, t_3 \rangle$

$cube(x)$

$\langle x, a \rangle$

$, t_3 \rangle$

$\langle t_0, s_0 \rangle, \langle a, \alpha \rangle, \langle b, \beta \rangle$

$cube(a)$

$on - table(a)$

$screw(b)$

$on - table(b)$

$?(K_1)$

$\langle t_1, s_1 \rangle$

resolve-anchor(x)

$\langle t_2, s_2 \rangle$

set$(\langle x, a \rangle)$

$\langle t_3, s_3 \rangle$

no-interpretation(K_1)

$\langle t_4, s_4 \rangle$

s_0

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SCREW, L - TABLE\}$

s_1

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SCREW, L - TABLE\}$

s_2

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SCREW, L - TABLE\}$

s_3

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SCREW, L - TABLE\}$

s_4

$\alpha : \{O - CUBE, L - TABLE\}$

$\beta : \{O - SCREW, L - TABLE\}$

(13)

$$\left\langle \boxed{\begin{array}{l} \langle e_1, i\mathrm{INT} \boxed{\begin{array}{l} \langle e_2, i\mathrm{DO} \boxed{handhuman(x)} \\ e_2 \subset e_1 \\ give(e_2) \end{array}} \rangle \\ cube(x) \\ \langle x, a \rangle \\ e_1 \subseteq n \\ \langle n, [t_6, t'_{10}] \rangle \end{array}}, t_6 \right\rangle$$

$$\boxed{\begin{array}{l} \langle t_5, s_5 \rangle, \langle a, \alpha \rangle, \langle b, \beta \rangle \\ cube(a) \\ on-table(a) \\ screw(b) \\ on-table(b) \end{array}}$$

g-add(i, K_2)

$$\boxed{\langle t_6, s_6 \rangle}$$

$$\boxed{confirm-has-interpretation(K_1)}$$

$$\boxed{t'_7}$$

$$\boxed{g-add'(handrobot(a))}$$

$$\boxed{t'_8}$$

$$\boxed{present'(a)}$$

$$\boxed{t'_9}$$

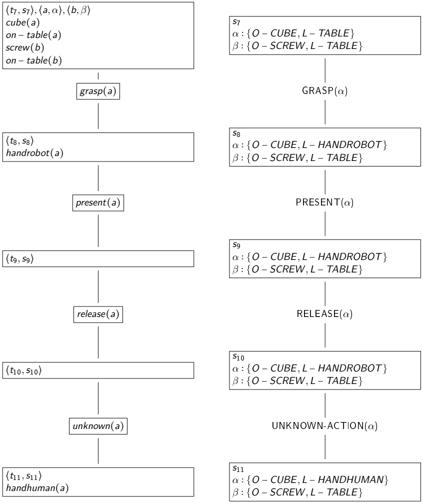
$$\boxed{release'(a)}$$

$$\boxed{\begin{array}{l} t'_{10} \\ handhuman(a) \end{array}}$$

$$\boxed{\begin{array}{l} s_5 \\ \alpha : \{O-CUBE, L-TABLE\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

$$\boxed{\begin{array}{l} s_6 \\ \alpha : \{O-CUBE, L-TABLE\} \\ \beta : \{O-SCREW, L-TABLE\} \end{array}}$$

(14)



$\langle t_7, s_7\rangle, \langle a, \alpha\rangle, \langle b, \beta\rangle$
$cube(a)$
$on-table(a)$
$screw(b)$
$on-table(b)$

$\boxed{grasp(a)}$

$\langle t_8, s_8\rangle$
$handrobot(a)$

$\boxed{present(a)}$

$\langle t_9, s_9\rangle$

$\boxed{release(a)}$

$\langle t_{10}, s_{10}\rangle$

$\boxed{unknown(a)}$

$\langle t_{11}, s_{11}\rangle$
$handhuman(a)$

s_7
$\alpha : \{O-CUBE, L-TABLE\}$
$\beta : \{O-SCREW, L-TABLE\}$

GRASP(α)

s_8
$\alpha : \{O-CUBE, L-HANDROBOT\}$
$\beta : \{O-SCREW, L-TABLE\}$

PRESENT(α)

s_9
$\alpha : \{O-CUBE, L-HANDROBOT\}$
$\beta : \{O-SCREW, L-TABLE\}$

RELEASE(α)

s_{10}
$\alpha : \{O-CUBE, L-HANDROBOT\}$
$\beta : \{O-SCREW, L-TABLE\}$

UNKNOWN-ACTION(α)

s_{11}
$\alpha : \{O-CUBE, L-HANDHUMAN\}$
$\beta : \{O-SCREW, L-TABLE\}$

Plan

1. Motivation
   - Sample Setup
   - Events in Formal Semantics
   - Starting Point
2. Representing the robot's internal configuration
   - Sensorimotor Control
   - Adding a future
   - Branching Time Structure
3. Event Segmentation in Cognitive Science
4. Event Segmentation in DRT with Anchors
   - Branching-Time Semantics for DRT
   - DRS Interpretation: Direction of Fit
   - Example: Give me the cube!
5. **Beyond Robots: Prospects and Problems**

Applications of grounding events in hierarchical action structures

Apart from Robotics, the grounding of events in hierarchical action structures provides valuable insights in various subfields of semantics and pragmatics

- Vendler Classes, Tense (e.g. Steedman [2002], Van Lambalgen and Hamm [2004])
- Verbs at the Syntax-Semantics Interface (e.g. Force Dynamics (Copley and Harley [2014])
- Intentional Verbs like "fetch" (Kamp [2007])
- Speech Acts (e.g. Cohen and Perrault [1986])
- Discourse Structure (e.g. Grosz and Sidner [1986])
- Frames (e.g. Fillmore [1982], Minsky [1972])
- Narration (e.g. Rumelhart [1975], Schank and Abelson [1977])
- ...

Acquisition of Plans

The acquisition, granularity, size and grounding of a plan/script/frame library is an unsolved problem since the early days of Artificial Intelligence and Linguistics.

Learning 'script' knowledge

Growing interest in learning event descriptions from e.g.

- web-experiments (Regneri et al. [2010])
- motor/brain data (Guerra-Filhoa and Yiannis [2012])
- perceptions (Matuszek et al. [2012])
- videos (Regneri et al. [2013])

Goal: statistical correlation between event *descriptions*/script knowledge (i.e. full sentences) and **bottom-up segmentation** of *events*.
But: event *descriptions* specify hierarchical **top-down segmentation** conditions.

Decomposition of actions and language

- Event descriptions (i.e. linguistic expressions) can be decomposed into their linguistic constituents: they have a compositional denotational semantics
- If we want the semantics of event descriptions to be defined in terms of hierarchical goal structures, we need to relate the composition of event descriptions with a compositional account of the structure of actions and goals.
- How can we decompose actions into hierarchical goal structures that mirror the compositional structure of event descriptions in order to capture top-down segmentation?

Outlook: Learning goals and actions with affordances

- Maybe we should turn things upside down and focus on which actions things afford instead of what actions do to things in a theory of affordances Gibson [1982]
- E.g. a door affords egress and ingress, a knife affords cutting and scraping, . . .
- Affordances offer a way in which perceptual learning can be linked to goals and actions (which makes it particularly interesting for robots, e.g. Rome et al. [2008]).
- Steedman [2002]: affordance-based planning and combinatory categorial grammar are based on the same two combinatory operations – functional composition and type-raising – → there is an immediate mapping between language and action.

Are verbs the locus of events?

- Affordances are indexed with objects rather than events or result states.
- Consequently, the affordance-based perspective on event semantics challenges the predominant approach of decomposing eventsà la Dowty [1979], Levin [1999] in the lexicon with DO/CAUSE/BECOME and result states in particular and the grammatical functions of "Agent" and "Patient" Dowty [1991] in general.

References I

J. L. Austin. *How to do Things with Words*. Oxford University Press, New York, 1962.

R. G. Barker and H. F. Wright. *Midwest and its children: The psychological ecology of an American town.* Row, Peterson and Company, Evanston, 1954.

Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, 1987.

Roberto Casati and Achille Varzi, editors. *Events*. Dartmouth, 1996.

P. R. Cohen and J. H. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1991.

P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. pages 423–440, 1986.

Bridget Copley and Heidi Harley. A force-theoretic framework for event structure. Unpublished Manuscript,, 2014.

D. Dennett. *Consciousness explained.* Little, Brown and Co., Boston, New York, 1991.

D. R. Dowty. *Word Meaning und Montague Grammar*. Springer, New York, 1979.

David Dowty. Thematic proto-roles and argument selection. *Language*, 67(3):547 – 619, 1991.

References II

E. A. Emerson. Temporal and modal mogic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Vol. B*, Amsterdam, 1990. North-Holland Publishing Company.

Charles J. Fillmore. Frame semantics. In The Linguistic Society of Korea, editor, *Linguistics in the Morning Calm. Selected Papers from SICOL-1981*, pages 111–137. Hanshin, Seoul, 1982.

D. M. Gabbay. The declarative past and imperative future: Executable temporal logic for interactive systems. In *Temporal Logic in Specification*, pages 409–448, London, 1987. Springer.

James J. Gibson. *Wahrnehmung und Umwelt*. Urban & Schwarzenberg, 1982. Up 482.

B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

Gutemberg Guerra-Filhoa and Aloimonosb Yiannis. The syntax of human actions and interactions. *Journal of Neurolinguistics*, 25(5):500–514, 2012.

M. D' Inverno, M. Luck, M. P. Georgeff, D. Kinny, and M. J. Wooldridge. The dMARS architecture: A specification of the distributed multi-agent reasoning system. *Autonomous Agents and Multi-Agent Systems*, 9:5–53, 2004.

References III

Hans Kamp. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo M. V. Janssen, and Martin Stokhof, editors, *Truth, Interpretation and Information: Selected Papers from the Third Amsterdam Colloquium*, pages 1–41. Foris Publications, Dordrecht, 1984.

Hans Kamp. Intentions, plans and their execution: Turning objects of thought into entities of the external world., 2007. Unpublished Manuscript, University of Stuttgart.

Beth Levin. Objecthood. an event structure perspective. In *Proceedings of CLS 35*, pages 223–47. Chicago Linguistic Society, 1999.

C. Matuszek, N. FitzGerald, L. Zettlemoyer, L. Bo, and D. Fox. A Joint Model of Language and Perception for Grounded Attribute Learning. *ArXiv e-prints*, June 2012.

M. Minsky. A framework for representing knowledge. In P. H. Winston, editor, *The psychology of computer vision*. McGraw-Hill, New York, 1972.

Tillmann Pross. Argumentationslinien der formalen pragmatik. Studienarbeit, Institute for Natural Language Processing, University of Stuttgart, 2005.

Tillmann Pross. The pragmatics of aspect. Diplomarbeit, Institute for Natural Language Processing, University of Stuttgart, 2006.

Tillmann Pross. *Grounded Discourse Representation Theory*. PhD thesis, Institute for Natural Language Processing, Institute for Natural Language Processing, University of Stuttgart, Germany, 2010a. URL `d-nb.info/1000641007/34`.

References IV

Tillmann Pross. Metalanguage dynamics. In *Proceedings of the Workshop on Theories of Information Dynamics and Interaction and their Application to Dialogue at ESSLI 2010*, pages 66–78, Copenhagen, DK, 2010b.

Tillmann Pross. How to combine speech and action. Poster at Theoretical Pragmatics 2011, Berlin, 2011.

Tillmann Pross. An action-theory based treatment of temporal individuals. In Ruth Hagengruber and Uwe Riss, editors, *Philosophy, Computing and Information Science*, volume 3 of *History and Philosophy of Technoscience: 3*, pages 179–188. Pickering & Chatto, London, GB, 2014.

Michaela Regneri, Alexander Koller, and Manfred Pinkal. Learning script knowledge with web experiments. In *Proceedings of ACL 2010*, Uppsala, Sweden, July 2010. Association for Computational Linguistics.

Michaela Regneri, Marcus Rohrbach, Dominikus Wetzel, Stefan Thater, Bernt Schiele, and Manfred Pinkal. Grounding action descriptions in videos. *Transactions of the Association for Computational Linguistics (TACL)*, 1:25–36, 2013.

Erich Rome, Joachim Hertzberg, and Georg Dorffner, editors. *Towards Affordance-Based Robot Control*, volume 4760 of *Lecture Notes in Computer Science*. Springer, 2008.

References V

D. E. Rumelhart. Notes on a schema for stories. In D. G. Bobrow and A. Collins, editors, *Representation and Understanding: Studies in Cognitive Science.*, pages 211 – 236. Academic Press, London, 1975.

R. C. Schank and R. P. Abelson. *Scripts, plans, goals and understanding: an inquiry into human knowledge structures.* L. Erlbaum Associates, Hillsdale, 1977.

J. R. Searle. Minds, brains, and programs. *Behavioral and Brain Sciences*, 3:417 – 424, 1980.

M. P. Singh. *Multiagent Systems. A theoretical framework for Intentions, Know-How and Communications*. Springer, New York, 1994.

M. P. Singh, A. S. Rao, and M. P. Georgeff. Formal methods in DAI: Logic-based representation and reasoning. In G. Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, chapter 8, pages 331 – 376. MIT Press, Cambridge, MA, 1999.

D. Sperber and D. Wilson. Linguistic form and relevance. *Lingua*, 90(2):1 – 25, 1993.

M. Steedman. Plans, affordances, and combinatory grammar. *Linguistics and Philosophy*, 25(5):723 – 753, 2002.

S. Thill, S. Pado, and T. Ziemke. On the importance of a rich embodiment in the grounding of concepts: Perspectives from embodied cognitive science and computational linguistics. *Topics in Cognitive Science*, pages 1–14, 2014.

References VI

Michiel Van Lambalgen and Fritz Hamm. *The proper treatment of events*. Wiley Blackwell, 2004.

J. M. Zacks and K. M. Swallow. Event segmentation. *Current Directions in Psychological Science*, 16(2):80–84, 2007.

Tversky B. Zacks, J. M. and G. Iyer. Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General*, 130:29 – 58, 2001.

Steedman: Affordances

(15) $push(x) = \lambda x.\{shut(x) \multimap open(x); open(x) \multimap shut(x)\}$

(16) $affordances(door) = \{push\}$

Affordance conception of door as a function mapping doors into (second-order) functions from their affordances like pushing to their results (e.g. open).

(17) $door' = \lambda x_{door} \lambda p_{affordances(door)}.px$