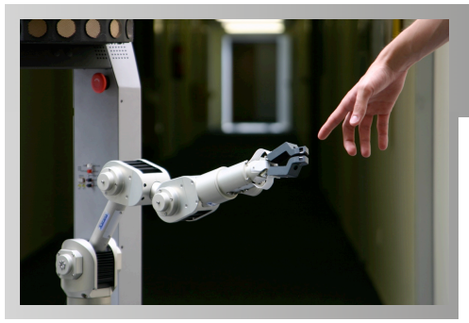
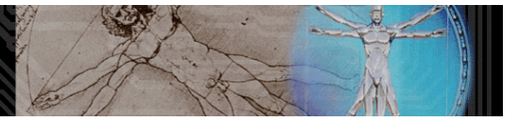




"How to make talking robots"

Pierre Lison

Cognitive Systems @ Language Technology Lab

German Research Center for Artificial Intelligence (DFKI GmbH)

Saarbrücken, Germany

http://www.dfki.de/cosy/

These slides are based on:

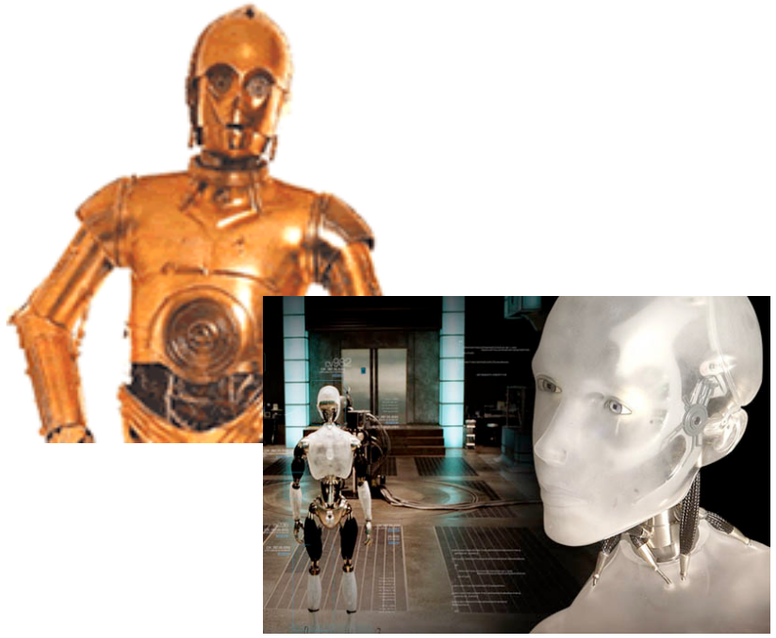
"How to make talking robots",
Geert-Jan M. Kruijff, IEEE RO-MAN'07 Tutorial,
Jeju Island (Korea), August 26 2007

"What makes a cognitive system understand situated dialogue?"
Geert-Jan M. Kruijff, invited talk at the Department of Linguistics,
University of Texas @ Austin, March 20 2006

Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

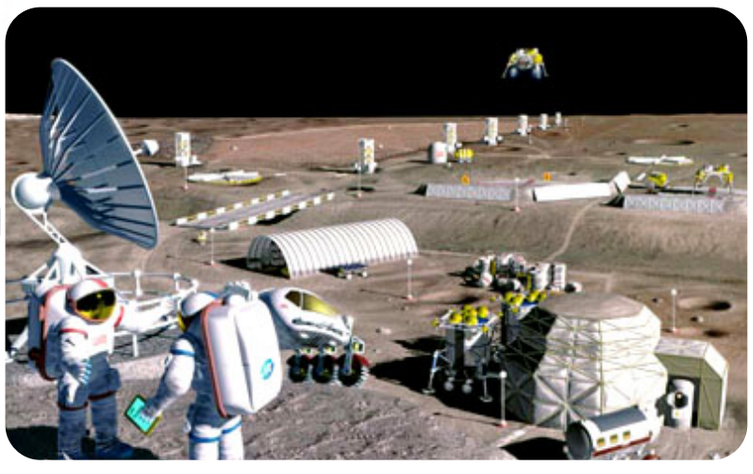
Talking robots?

Most of you probably know these guys ...



and we are already able to do a few things ... (though it ain't no Hollywood)

What we'd like to do is to get robots to help us out, at home ...



... and in the office (wherever the job takes us).

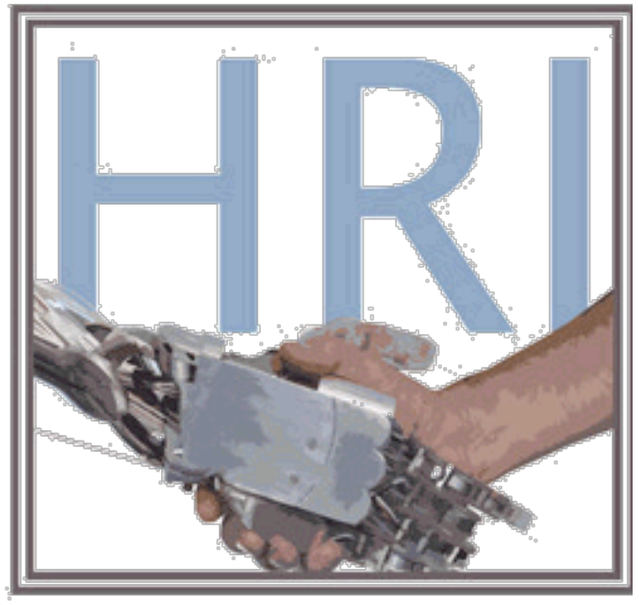
Goal and overview

- Goal of this talk:
  - Give a broad overview of what "embodied interaction" is about
  - Fundamental question: *"What makes an embodied, situated cognitive system understand?"*
- Overview:

1. Generalities on Human-Robot Interaction
2. What is a cognitive system?
3. Incremental processing of situated dialogue
   - What *is* spoken dialogue processing? What are we trying to do?
   - How "should" we process spoken dialogue, linguistically?
   - How could situation awareness aid spoken dialogue processing?

Human-robot interaction

- How do we make robots that actually *understand* what we say? And that understand *why, when* and *how* they should say something?



- Situated dialogue understanding
  - Understanding and producing language, relative to a current or imaginable situation in which the agents are situated
  - Why? Language often refers to reality, discusses actions and plans that affect that reality
- **Situated dialogue understanding requires situated understanding**

The Cosy project

- Goal: *Foundations of integrated, continuously developing cognitive architectures for embodied interactive agents*
- Scenario-based validation of approaches
- (EU FP6 IP; 6M€; September 2004 - August 2008)

KTH Stockholm:
Vision, exploration

ALU Freiburg:
Collaborative planning

U.Birmingham:
Architecture, planning

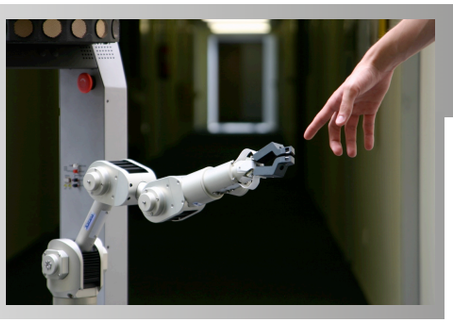
LPE-CNRS Paris:
Sense of space

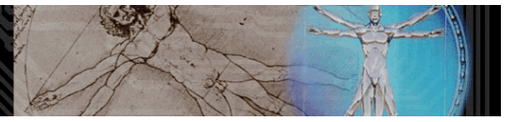
DFKI, Saarbrücken:
Communication

TU Darmstadt:
Vision

U. Ljubljana:
Vision, mapping

EU Cognitive Systems Unit





What is a cognitive system?

Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

Cognition = perception + intelligence

- What is a cognitive *system*?
  - Cognition is more than intelligence: it is intelligence set in reality
  - "Cognition = perception + intelligence"
- What about embodiment and experience?
  - Embodiment modulates how a system sees, experiences, reality
  - "Cognition = embodiment[perception+intelligence]"
  - A system encompasses the embodied experience and the intelligence
  - "Cognitive system = embodiment [architecture + experience]", with "architecture = intelligence + perception"

Varieties of embodiment

- *What counts as a body?*

embodiment variation

physical realization

physical embodiment

organismoid embodiment

organismal embodiment

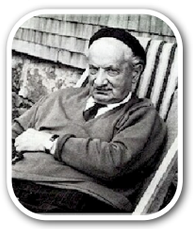
(Chrisley & Ziemke 2005)

Philosophy and embodiment

- From dualism …
  - *Cogito ergo sum*: the separation of mind and body
  - *How do the mental and physical realms interact?*
- to behaviorism …
  - Investigating (reactive) behavior linking mental and physical realms
  - *What are the processes that underly cognition?*
- to contextualized phenomenology
  - Human activity is not context-free manipulation of representations, but contextualized experience of the body-environment system
- The question of development …
  - Language is an inherently socially situated activity, and thus language use and development require embodiment
  - The body plays an essential role in development (interaction)
- and the phenomenology of mind
  - The way we perceive an object is determined by possibilities for bodily interaction

R. Descartes
(1596-1650)

B.F. Skinner
(1904-1990)



M. Heidegger
(1889-1976)

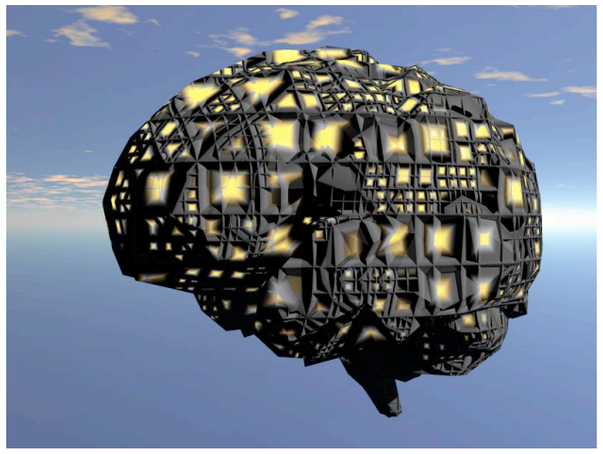
L.S. Vygotsky
(1896-1934)

J. Piaget
(1896-1980)

M. Merleau-Ponty
(1908-1961)

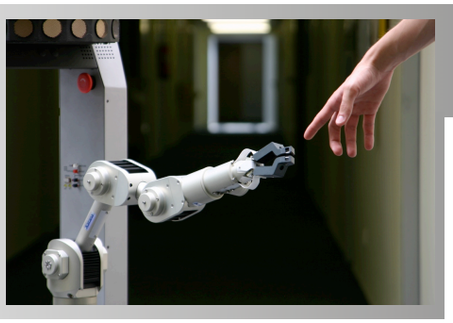
Cognitive architectures

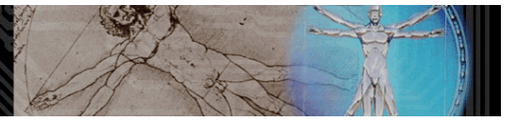
- A specification of the underlying system infrastructure
- What are the underlying issues?
  - What representational assumptions are made?
  - How are things stored - what are the characteristics of memories?
  - What processes operate on these memories?
  - How are perception and motor control integrated?



Properties of how knowledge is processed as a window on an architecture's capabilities

- An orthogonal perspective: properties
  - Properties of cognitive architectures from which capabilities arise
  - Properties of knowledge and its *representation (what is being represented, and how?), organization (what knowledge can be accessed, and how?),* and *use* (*how can knowledge be used?*)
  - The *mechanisms for acquiring and revising* knowledge
  -





Incremental processing of situated dialogue

Deutsches Forschungszentrum für Künstliche Intelligenz
German Research Center for Artificial Intelligence

What is situated dialogue?

- What is *dialogue*?
  - Spoken ("verbal") and, possibly, non-verbal interaction between two or more participants
  - The verbal and non-verbal interaction express meaning, which needs to be understood by all participants ("grounding") for the interaction to be successful
- What is *situated* dialogue?
  - A form of dialogue in which the expressed meaning may (spatio-temporally) refer to a physical environment
  - Grounding situated dialogue thus involves being able to resolve the linguistic references to the (own) situation awareness, to yield a common ground in how to understand the environment

What is spoken dialogue processing

- Different levels of processing (comprehension)
  - Auditory: speech recognition, (speaker localization & tracking)
  - Grammatical: syntactic structure, semantic structure
    "A grammar specifies the relation between how well-formed syntactic structures express an underlying (linguistic) meaning"
  - Discourse: contextual reference resolution (anaphora, ellipsis), rhetorical relation resolution, (clarification triggers)
    "Discourse interprets utterance meaning relative to the established context, establishing how it contributes to furthering the discourse"
- Challenges for spoken dialogue processing
  - Robustness in speech recognition: noise, speaker-independence, out-of-vocabulary; (intonation, emotion)
  - Robustness to ill-formed utterances: partial, ungrammatical utterances
  - Ambiguity, uncertainty in contextual interpretation of utterances

Spoken language analysis

Dialogue context meaning

structural dialogue analysis
[SDRT+paradigmatic choice]

SF Networks

Linguistic meaning

natural language grammar analysis
[OpenCCG]

CCG Grammar

String

(off-the-shelf) speech recognition
[Sphinx4, Nuance]

Speech Grammar

Audio signal

```
(@d0:dvp(body ^
 <AType>informative.attributive.endurant.perspective.spatial ^
 <TurnT>continuous ^
 <Mood>ind ^
  <ContentBody>(b3:state ^ be ^
                <Mood>ind ^
                <Restr>(w3:person ^ we) ^
                <Scope>(....)))))
```

```
@b1:state(be ^
      <Mood>ind ^
      <Restr>(w1:person ^ we) ^
      <Scope>(i1:region ^ in ^
            <Plane>horizontal ^
            <Positioning>static ^
            <Dir:Anchor>(o1:location ^ office ^
                  <Delimitation>unique ^
                  <Quantification>specific_singular)))
```

"We are in the office"

Grammatical analysis

- Why do a proper grammatical analysis?
  - Proper analysis is better than keyword spotting!
  - Analysis reveals *structure*, shows how content is built up
  - Analysis does not need to be "brittle": usefulness of partial analyses, possibilities for probabilistic inference, out-of-vocab

- What is the goal of such analysis?
  - Grammar relates form to meaning; the goal is to build a representation of the meaning(s) an utterance expresses
  - Conceive of linguistic meaning is ontologically richly sorted, relational structures to provide a basis for further interpretation

Parsing using CCG

- Bottom-up chart parsing, based on the CKY algorithm
- Combinatory Categorial Grammar (Steedman, 2000):
  - Mildly context-sensitive grammar formalism
  - Fully lexicalized
  - Building structures using combinatory rules
  - Applicability of combinatory rules is controlled through modalized slashes
- Computational aspects of CCG:
  - Polynomial parseability; statistical optimization in chart-based generation
  - Open-source development platform: OpenCCG [`http://openccg.sf.net`]
  - Statistical, broad-coverage parsing models for CCG

Incremental processing

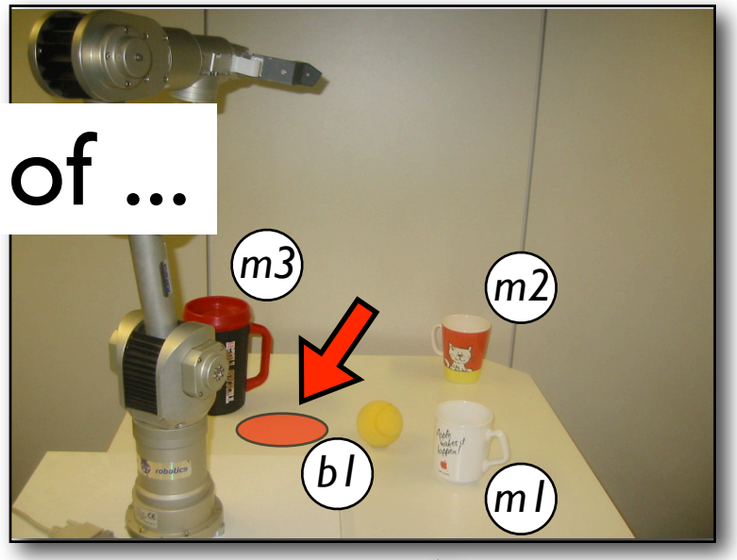
- What is incremental processing?
  - Processing an utterance in a "left-to-right" order, i.e. in the order as the words are heard
  - After each step, we have a set of analyses covering the left ("already heard") part of the utterance
  - Also, the categories for these analyses indicate what type(s) of categories (and possibly, content) we expect to the (immediate) right
- What is the point of incremental processing?
  - After each step, we can prune uninteresting interpretations from the parser's search space
  - After each step, we can prime selectional attention in other processes (even modalities), based both on built content and on *expected* content

Incremental processing: example

put

the mug

left of ...



dialogue

@p1:action(**put** ^
<Actor>(r1:hearer ^ **robot**) ^
<Patient>(x1:thing) ^
<Dir:WhereTo>(l1:location))

@p1:action(**put** ^
<Actor>(r1:hearer ^ **robot**) ^
<Patient>(x1:thing ^ **mug**) ^
<Dir:WhereTo>(l1:location))

potential syntactic ambiguity

catsys

put:
Actor: Causative-Agent (r1)
Patient: Graspable-Object (x1)
Goal: Location(l1)

put:
Actor: Causative-Agent (r1)
Patient: Graspable-Object (x1) ^ mug(x1)
Goal: Location(l1)

put:
Actor: Causative-Agent (r1)
Patient: Graspable-Object (x1) ^ mug(x1)
Goal: Location(l1) ^ left-of(l1,?b1)

visual bnd

thing(x1):
mug(m1), mug(m2),
mug(m3), ball(b1)

thing(x1):
mug(m1),
ball(b1)

thing(x1):
mug(m1)

thing(x1):
mug(m1)

potential visual referent ambiguity

spatial wm

topological:
near(m1, b1)
near(m1, arm), near(b1, arm),
near(m3, arm)

projective:
right-of(m1, b1)
left-of(free-region(f1), b1)

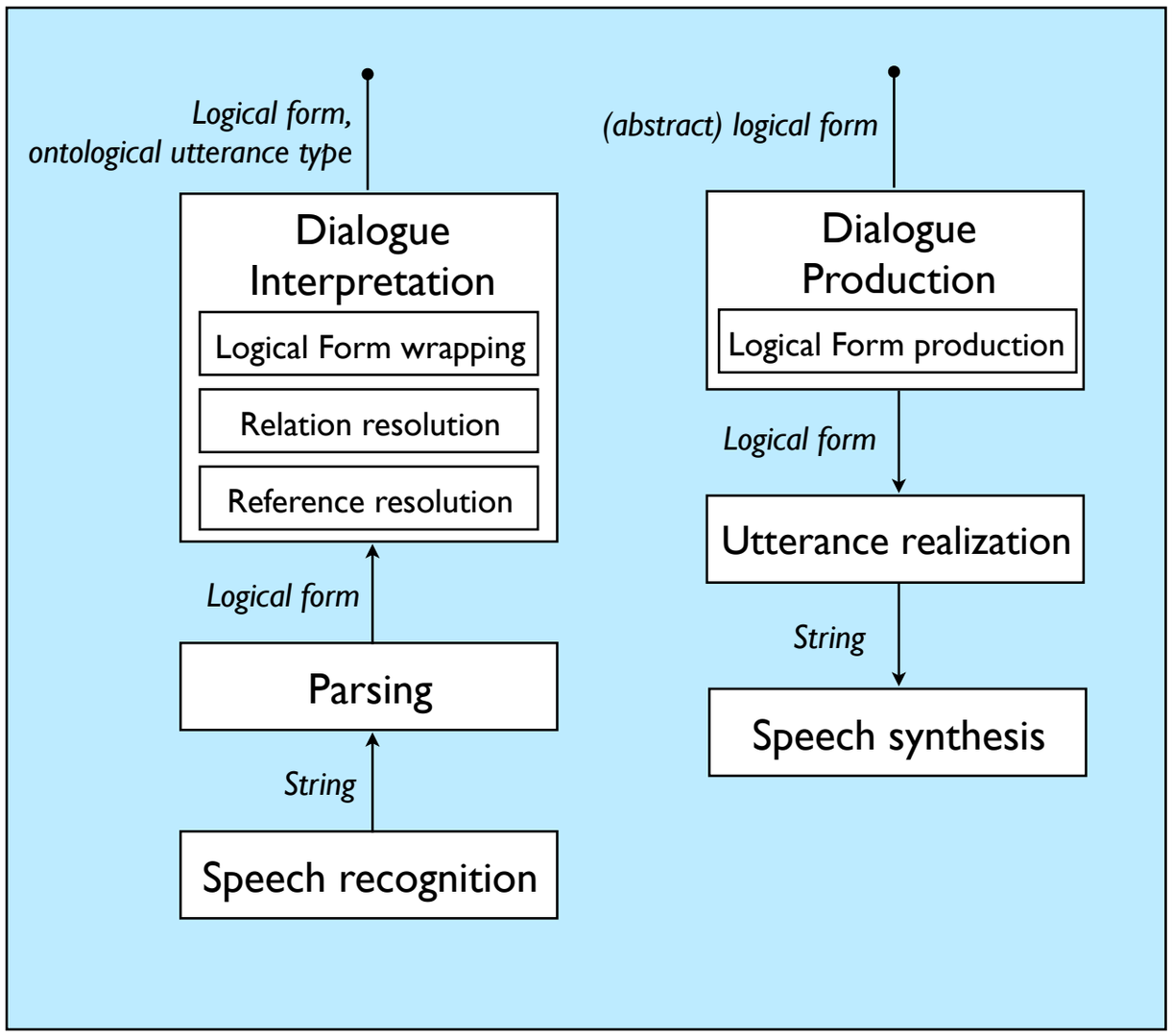
manipulation

topokinetic:
graspable(m1, arm)
graspable(b1, arm)

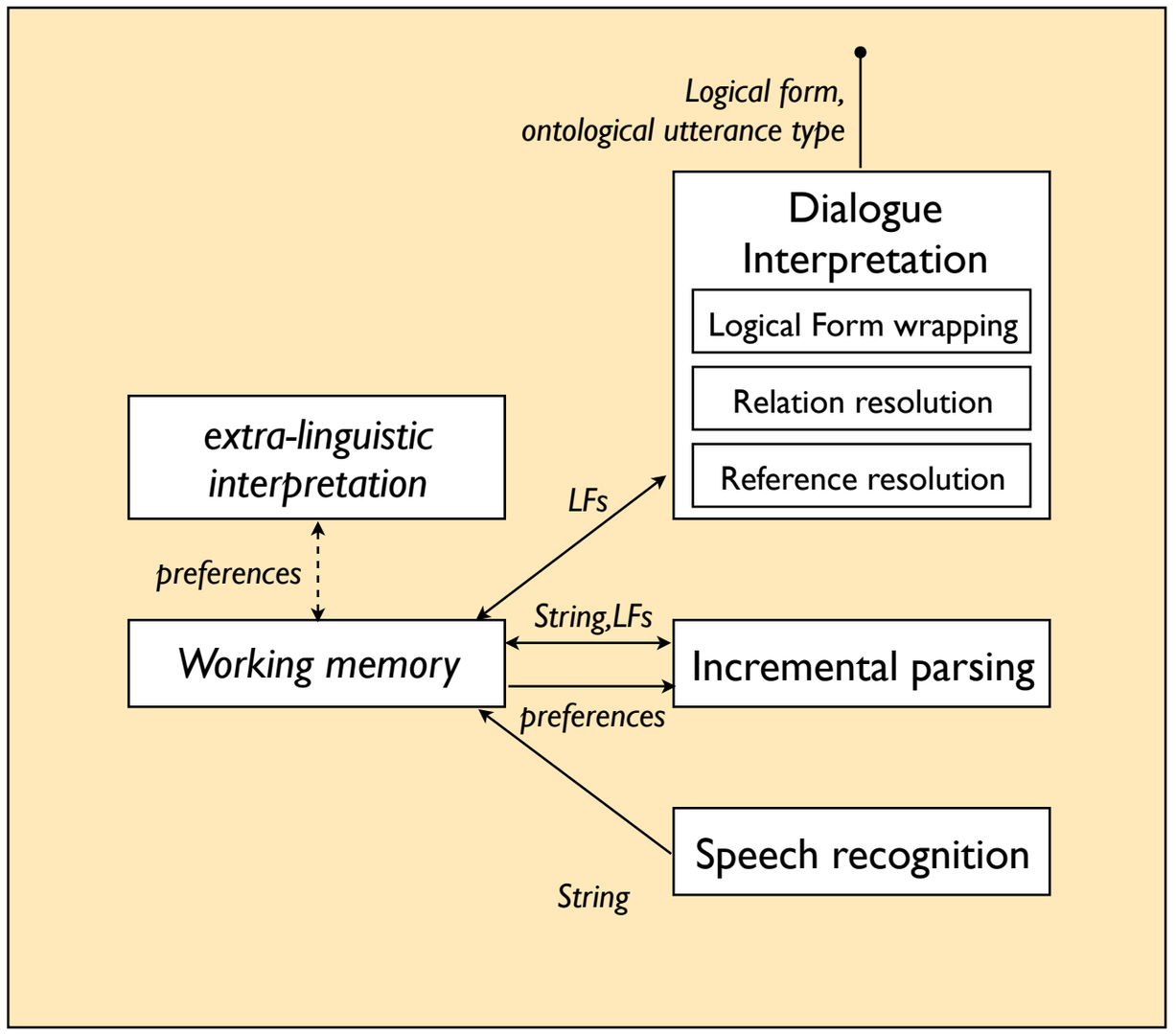
topokinetic:
graspable(m1, arm),
plan: reach(m1)

topokinetic:
graspable(m1, arm),
plan: reach(m1)
reachable(free-region(f1))

Incremental processing model



sequential model

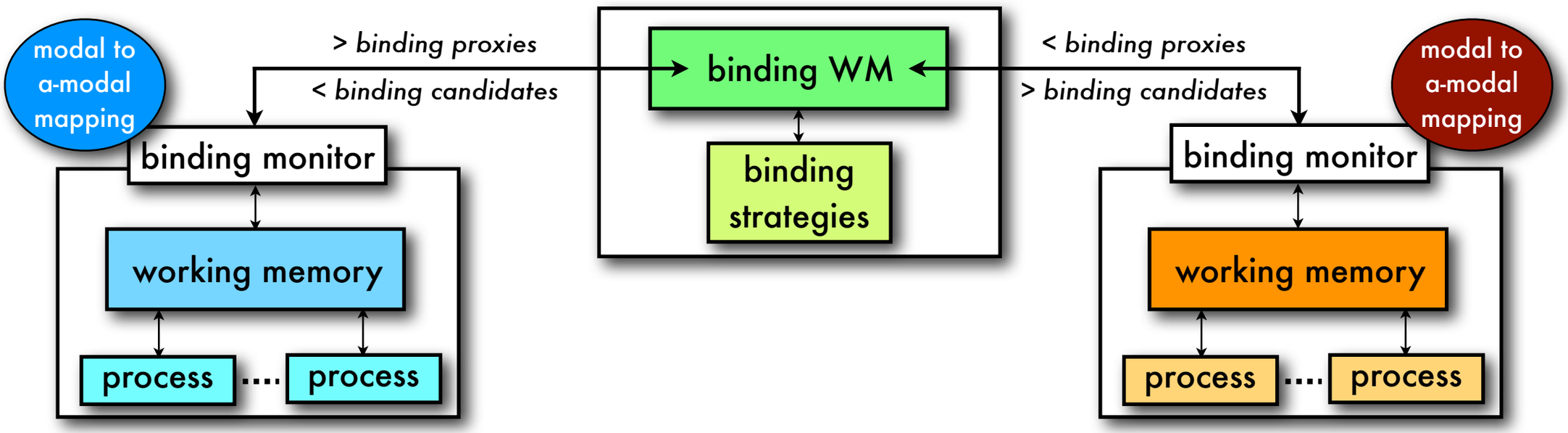


incremental model (analysis side)

- From a sequential to an incremental processing model
  - Information exchange needed to enable pruning during processing
  - Use a local working memory to enable this exchange

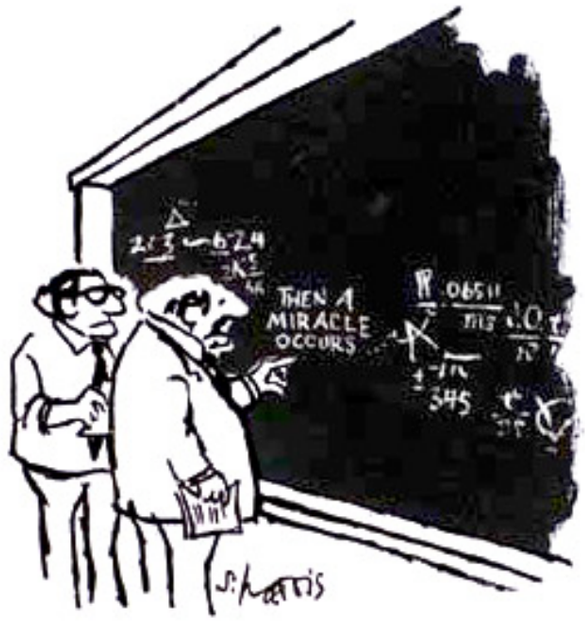
Integrating the dialogue context

- Context sensitivity in situated dialogue
  - Sensitivity to both discourse context and situated context
  - Content from both needs to be integrated to prime processing
  - Priming effects at multiple levels: content-preferences (lexical, syn-semantic), references, and ellipsis
- Connecting linguistic content with other modalities



Cross-modal content association

Language



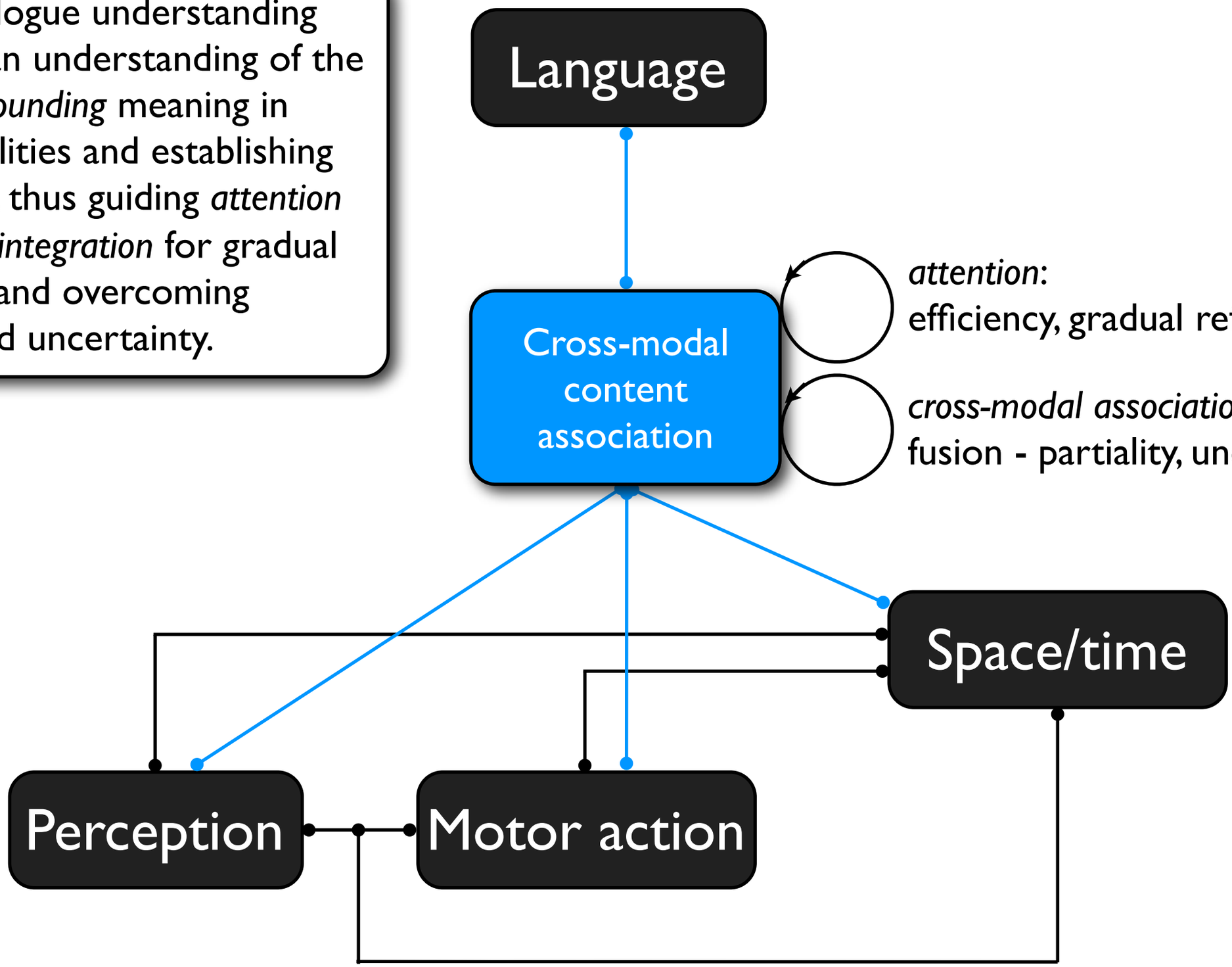
Space/time

Perception

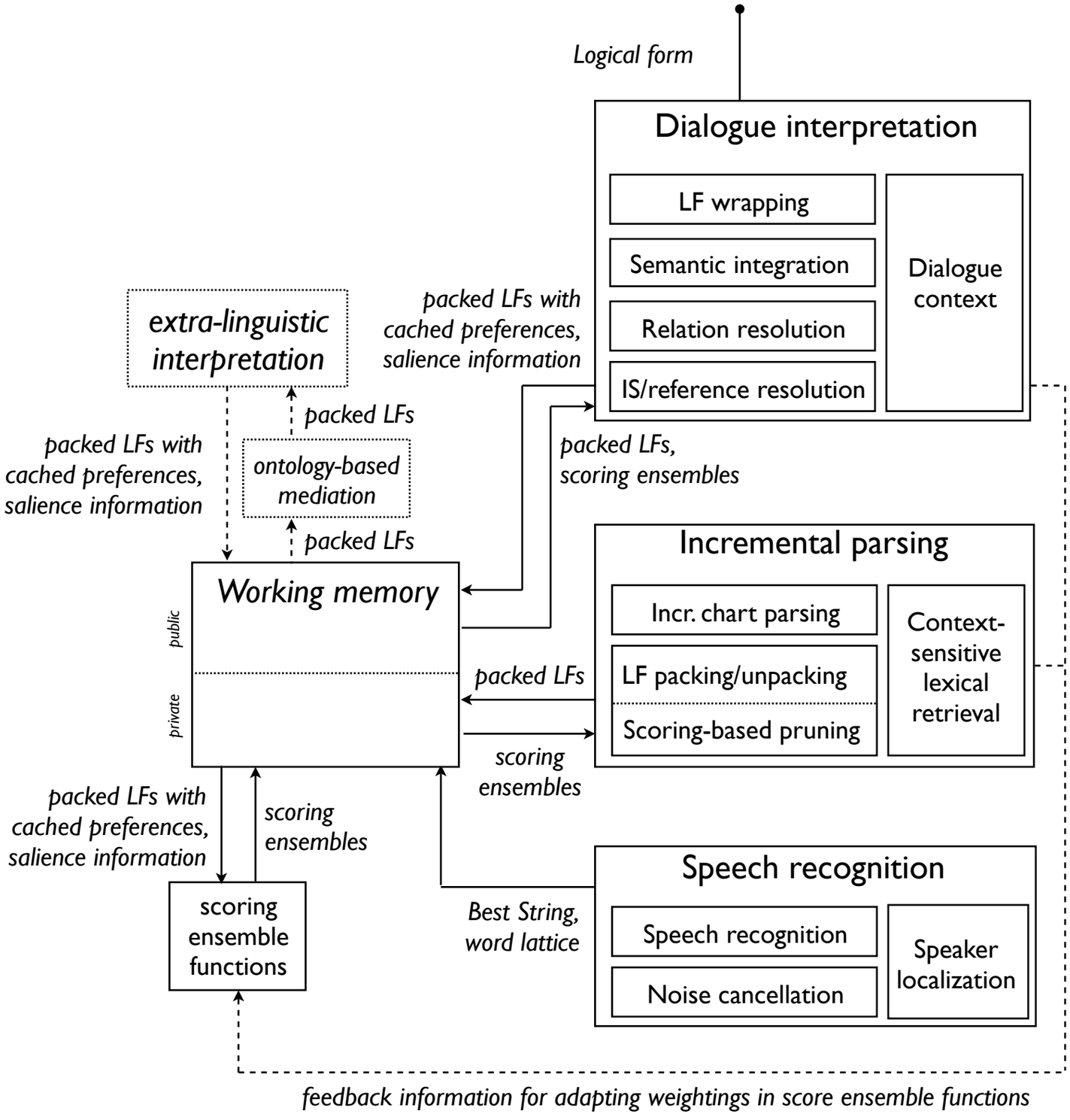
Motor action

Cross-modal content association

Situated dialogue understanding is based in an understanding of the situation, *grounding* meaning in other modalities and establishing *expectations*, thus guiding *attention* and *content integration* for gradual refinement and overcoming partiality and uncertainty.



Comprehension architecture

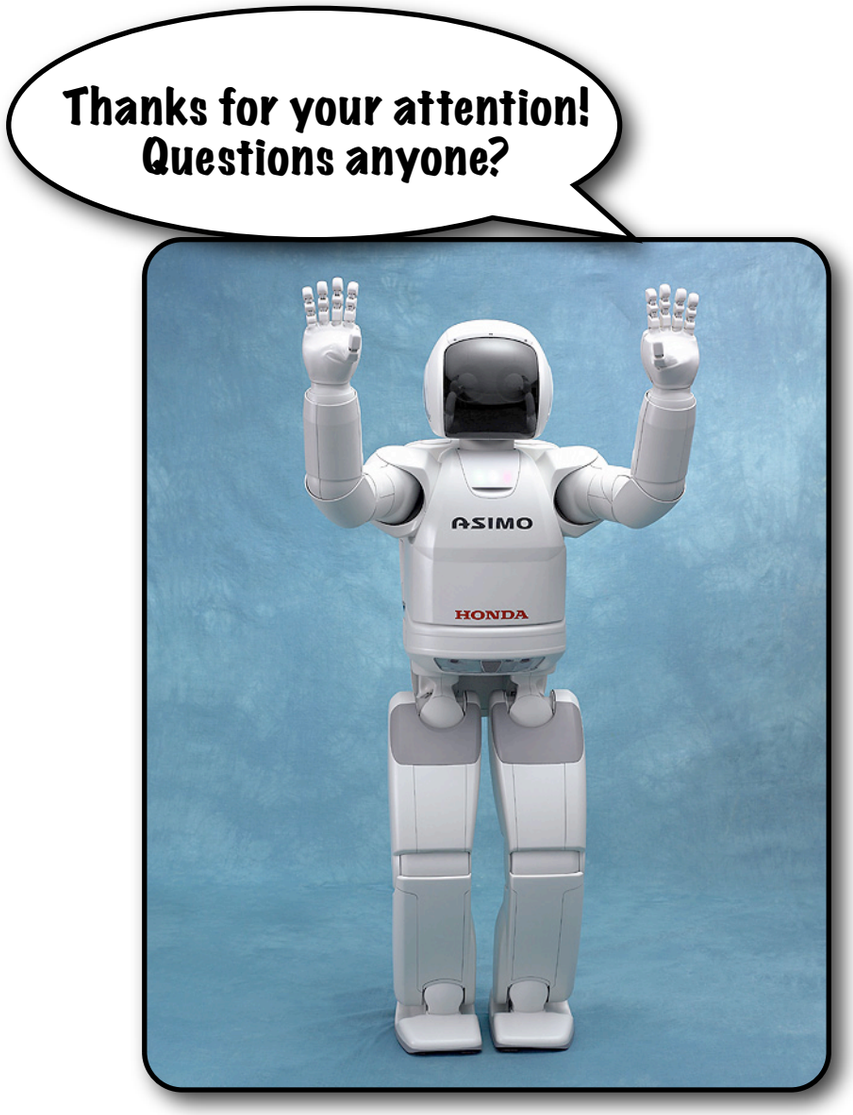


- Context pruning
  - Lexical meanings *(What meanings are active?)*
  - Syntactic attachment *(What relations are supported?)*
  - Semantic, referential indexicality *(What meanings, references can be resolved to the context?)*

Summary

- Human Robot Interaction:
  - How do we make robots that actually *understand* what we say, and understand *why, when* and *how* they should say something?
  - Situated dialogue understanding requires situated understanding
- A cognitive perspective on HRI:
  - Cognition is intelligence set in reality
  - Embodiment modulates how a system sees, experiences, reality
- Incremental processing of situated dialogue
  - Different levels of processing: auditory, grammatical, discourse
  - The interpretation of an utterance is built incrementally
  - Interpretation combines information from multiple levels within the dialogue system, as well as from other modalities

Questions



For publications, talks, videos:

http://www.dfki.de/cosy/

http://www.cognitivesystems.org