Dialogue model specification and interpretation for interacting with service robots

Luis Pineda
and
The Golem group
luis@leibniz.iimas.unam.mx
http://leibniz.iimas.unam.mx/~luis
IIMAS, UNAM, México
September, 2009

The Golem project team (current)

- Luis A. Pineda
- Héctor Avilés
- Paty Pérez Pavón
- Ivan Meza
- Wendy Aguilar
- Montserrat Alvarado
- External collaborators
- Students

Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and future work

Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and future work

Main assumptions…

- The interaction cycle must be tight!
  - The agent needs to keep in touch with the world!
  - Only reactive behavior?
- Human-level communication (e.g. Coordinated language and vision) needs to rely on symbolic representations
- The objects of representation are interpretations (not the world directly!)
- Perception is guided by intentions and memory
- Interpretation depends on the context

The Context

- Every interpretation act is performed in context:
  - A spatial and time situation (indexical)
  - A set of agents (I, you…also indices)
  - A discourse or interaction history (anaphoric)
  - A conceptual domain
  - A set of potential intention and action types that can be expressed by the agents during the interaction
- The context is a big thing!
- Can it be modelled explicitly in a simple way?

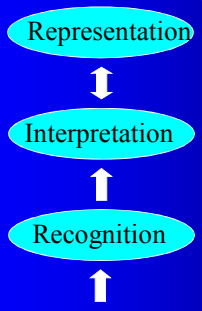
Agent's expectations

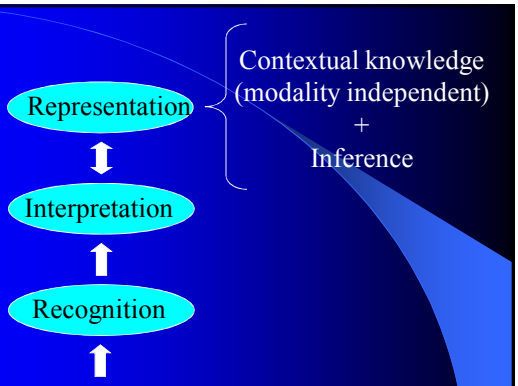
- Linguistic expectations (communicative)
  - What are the speech acts that can be expressed by interlocutors in specific conversational situations?
- Environmental expectations (without intent)
  - What events are expected to appear in the situation that are perceived by the different senses?
- How to deal with the unexpected?

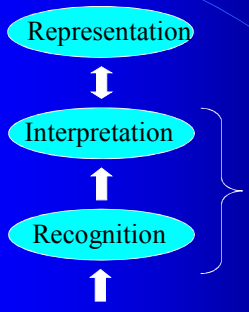
Content

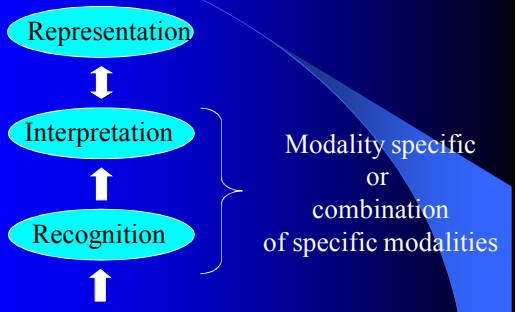
- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Future work

A three layers architecture

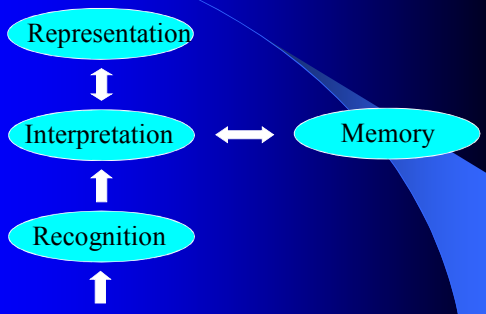




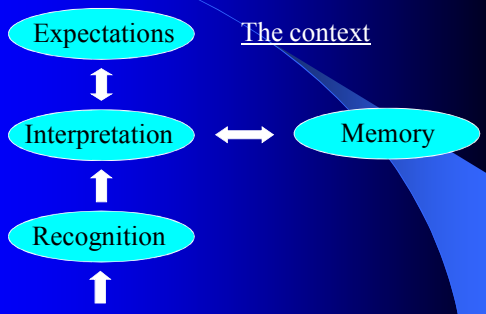




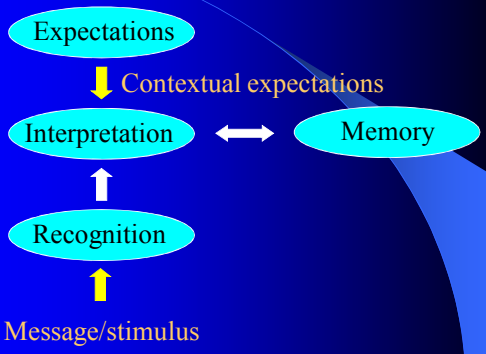
Modality specific images…



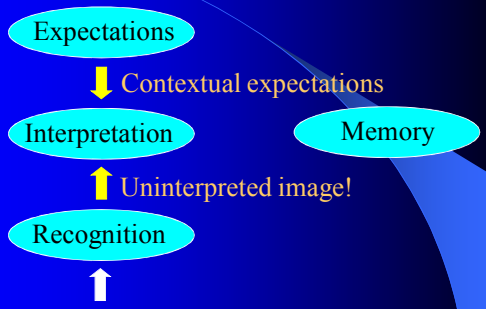
Expectations driven interpretation



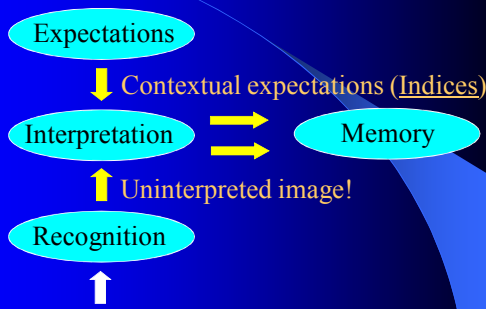
An interpretation act…



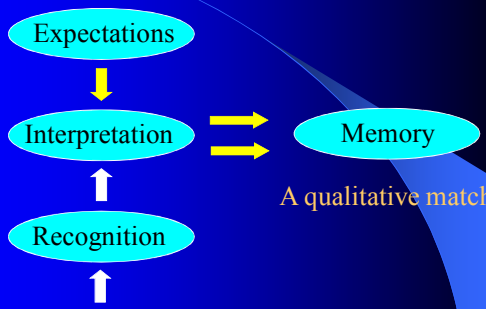
An interpretation act…



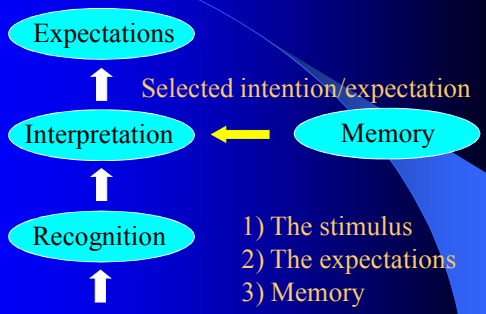
Memory is indexed by expectations!



Memory retrieval

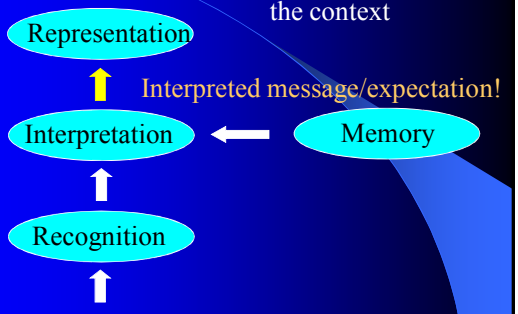


The elements of interpretation…

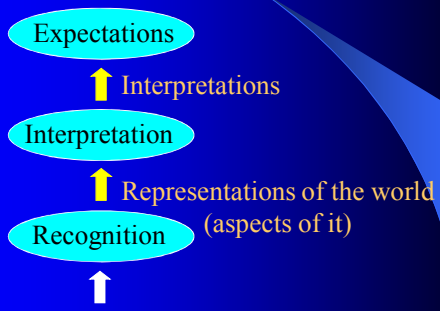


1) The stimulus
2) The expectations
3) Memory

Interpretation relative to the context



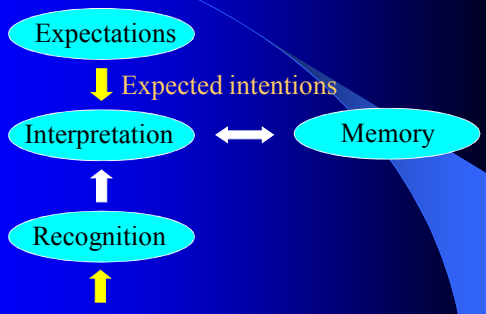
The objects of representation are interpretations!



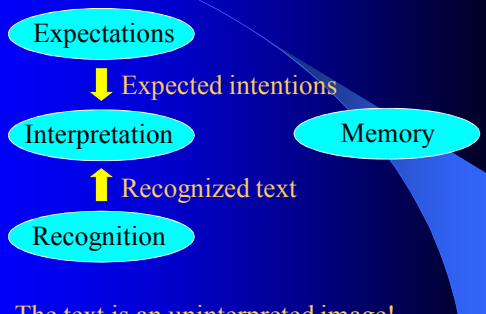
Interpretation Examples:

Linguistic & visual

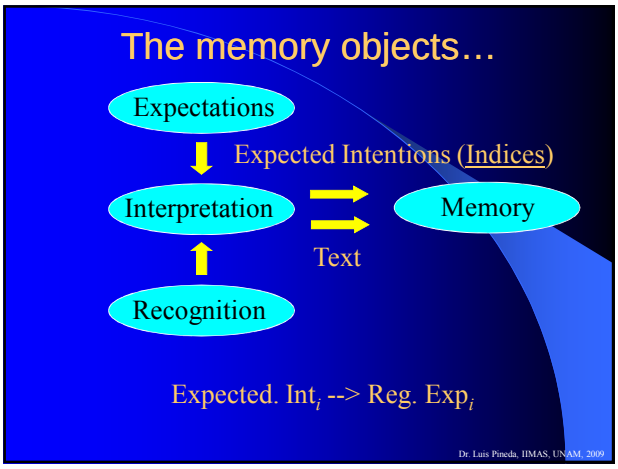
Linguistic Interpretation

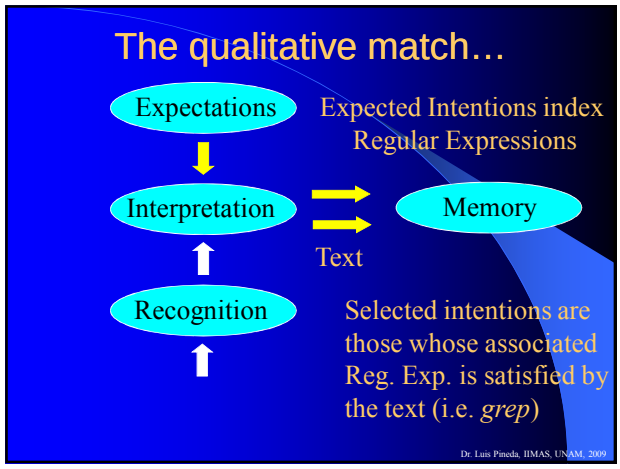


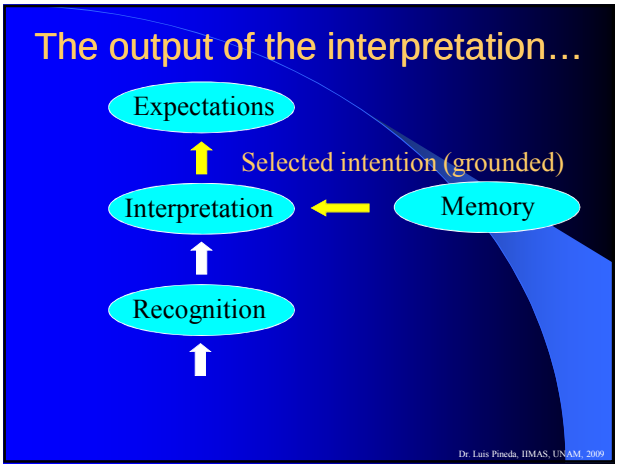
Spoken recognition…

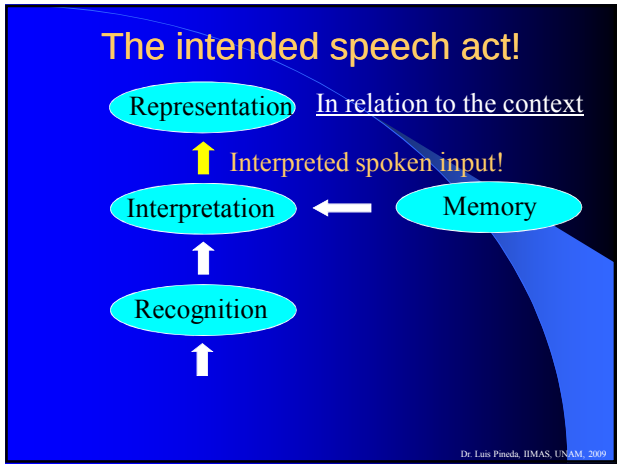


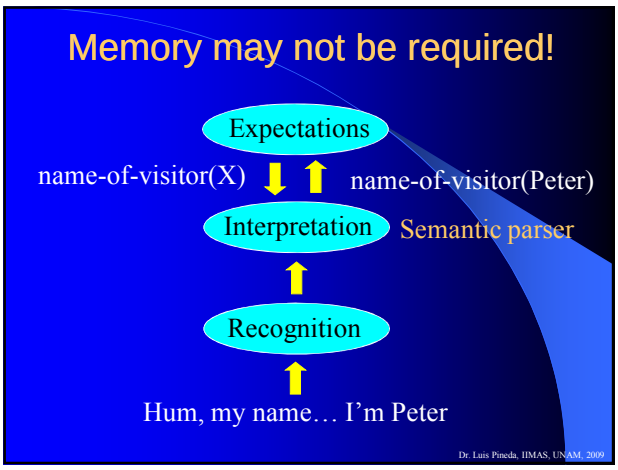
The text is an uninterpreted image!

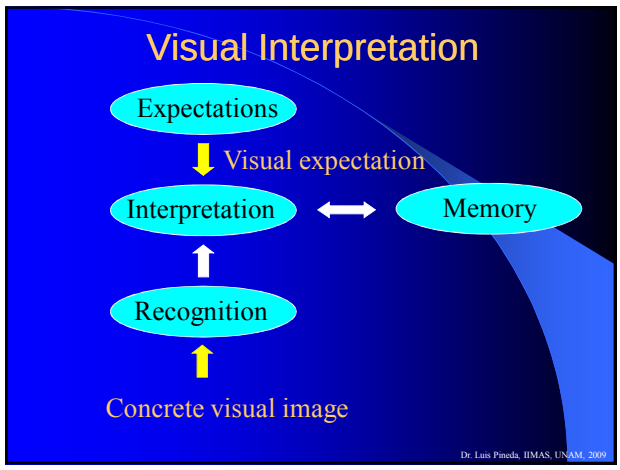

The memory objects…
Expectations
Expected Intentions (Indices)
Interpretation
Memory
Text
Recognition
Expected. Int$_i$ --> Reg. Exp$_i$
Dr. Luis Pineda, IIMAS, UNAM, 2009


The qualitative match…
Expectations
Expected Intentions index Regular Expressions
Interpretation
Memory
Text
Recognition
Selected intentions are those whose associated Reg. Exp. is satisfied by the text (i.e. *grep*)
Dr. Luis Pineda, IIMAS, UNAM, 2009

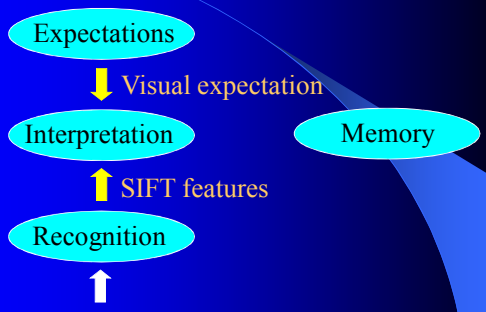

The output of the interpretation…
Expectations
Selected intention (grounded)
Interpretation
Memory
Recognition
Dr. Luis Pineda, IIMAS, UNAM, 2009


The intended speech act!
Representation
In relation to the context
Interpreted spoken input!
Interpretation
Memory
Recognition
Dr. Luis Pineda, IIMAS, UNAM, 2009

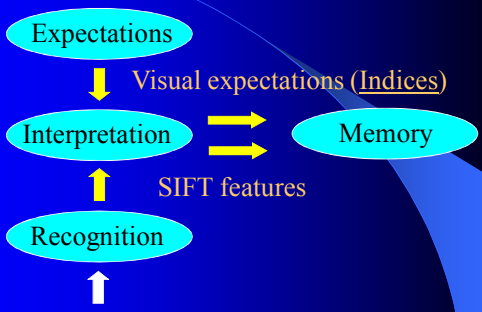

Memory may not be required!
Expectations
name-of-visitor(X)
name-of-visitor(Peter)
Interpretation
Semantic parser
Recognition
Hum, my name… I'm Peter
Dr. Luis Pineda, IIMAS, UNAM, 2009


Visual Interpretation
Expectations
Visual expectation
Interpretation
Memory
Recognition
Concrete visual image
Dr. Luis Pineda, IIMAS, UNAM, 2009

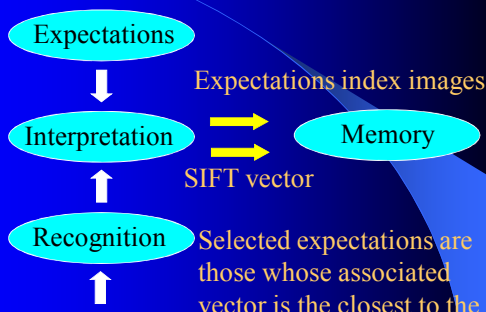
Spoken recognition…



The memory objects…

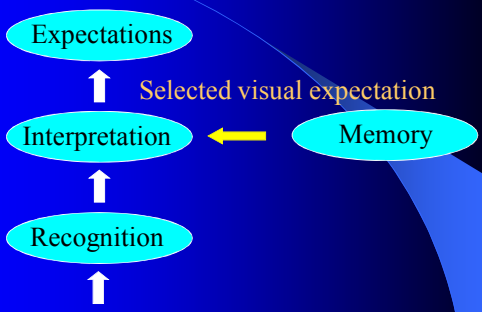


The qualitative match…

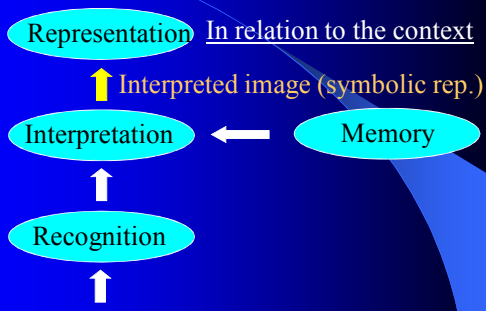


Selected expectations are those whose associated vector is the closest to the vector of the external image

The qualitative match



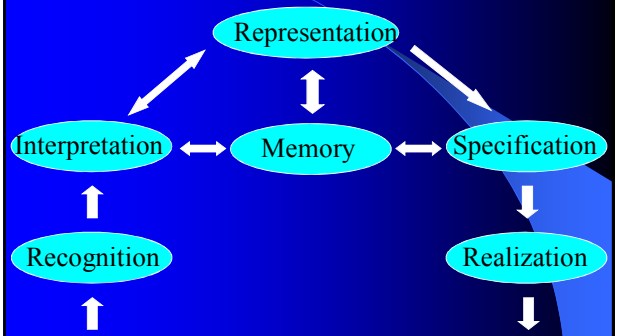
Interpretation of the visual act



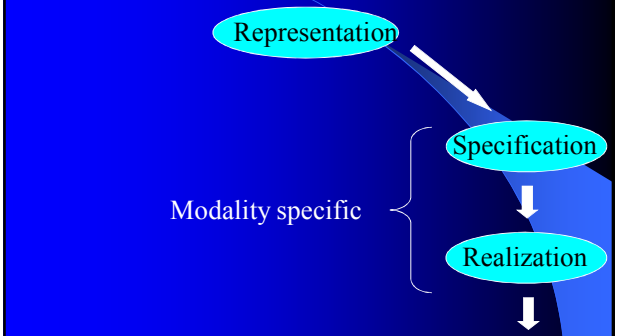
The full architecture:

Interpretation and Action

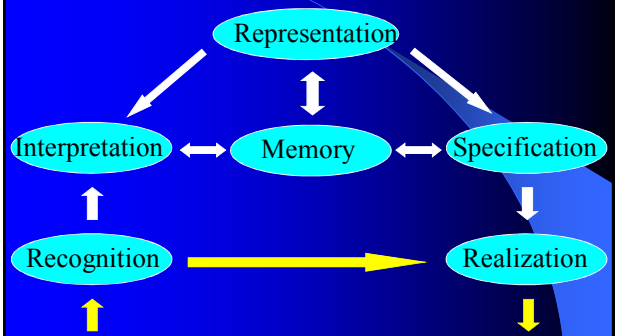
The full architecture



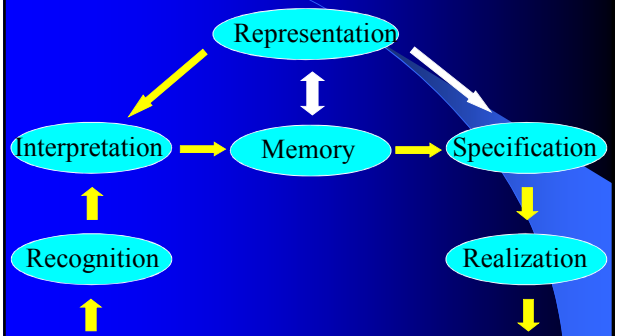
Output actions



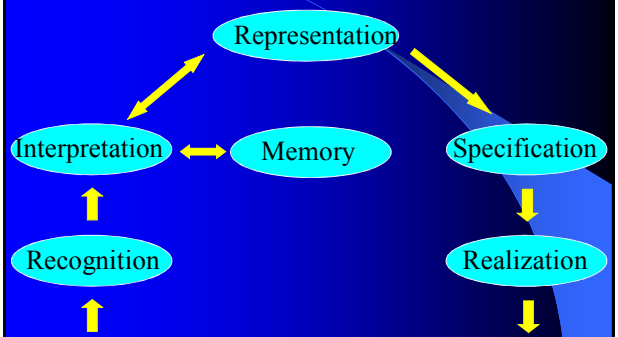
Reactive behavior



Schematic behavior…



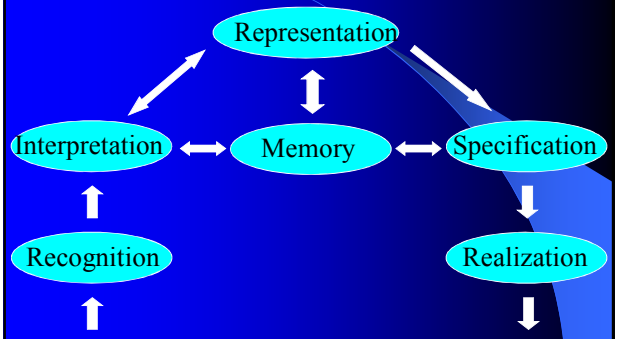
Our current working setting…



Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and future work

Dialogue Models



Dialogue Models

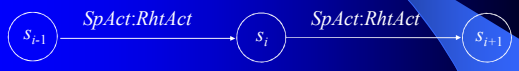
The formalism:

Functional Recursive Transition Networks (F-RTN)

Interactive situation

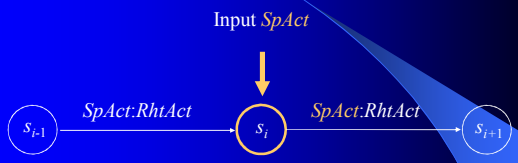
- The conversational situation
  - Information state
  - Modality oriented
  - Speech acts and expectations
  - Rhetorical Acts: actions performed by the agent (spoken, motor, etc.)
- Relative to a context!

Conversational Situation

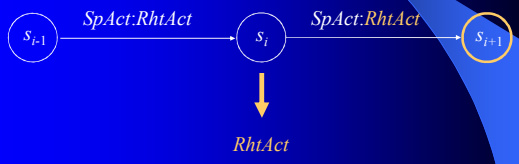


Only a specific set of intentions is meaningful

Conversational Situation



Conversational Situation



Intention types

- Listening
- Visual
- Telling
- Recursive
- Error
- Final

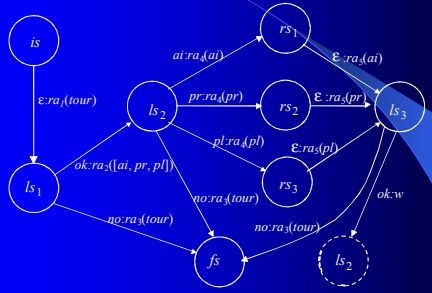
Rhetorical acts (Actions)

- Structure:
  - Lists of basic acts
  - Basic Acts: Modality specific actions
- Function:
  - Direct external behavior
  - Can direct internal behavior (embedded in the interpretation state): Razoning, planning, theorem-proving, problem-solving

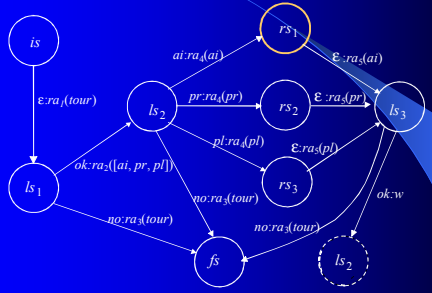
The Model

- Conversational model:
  - Set of dialogue models
- Dialogue model:
  - Set of situations
  - Set of rhetorical acts
- Declarative specification from a task analysis

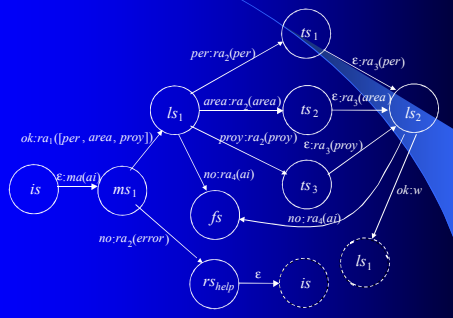
A Dialogue Model



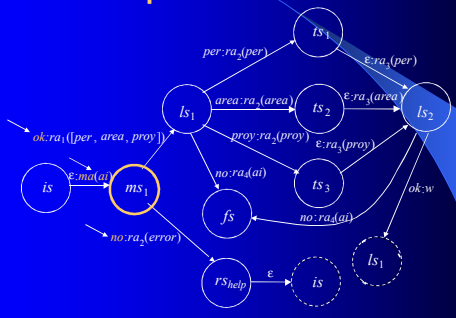
A Recursive Situation



A subordinated DM



Modality independent specification

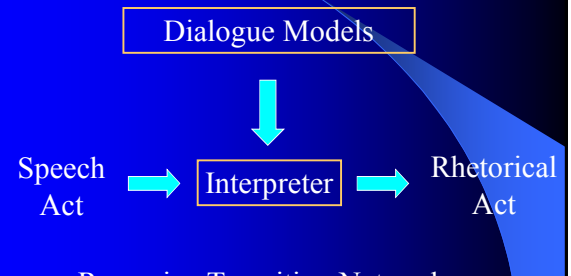


Basic expressive power: Recursive Transition Networks

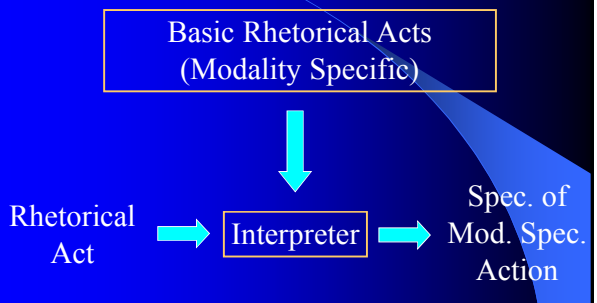
Declarative specification of situations

```
[id ==> ls_2,
        type ==> listening,
        in_pairs ==> [ls_1 => ok:ra_2([ai,pr,pl]),
                      ls_3 => ok:ra_2([ai,pr]),
                      ls_3 => ok:ra_2([ai,pl]),
                              ls_3 => ok:ra_2([pr,pl])],
        out_pairs ==> [ai:ra_4(ai) => rs_1,
                       pr:ra_4(pr) => rs_2,
                       pl:ra_4(pl) => rs_3,
                       no:ra_3(tour) => fs],
        expected_intentions ==>
                       [ai,pr, pl, no]
      ],
```

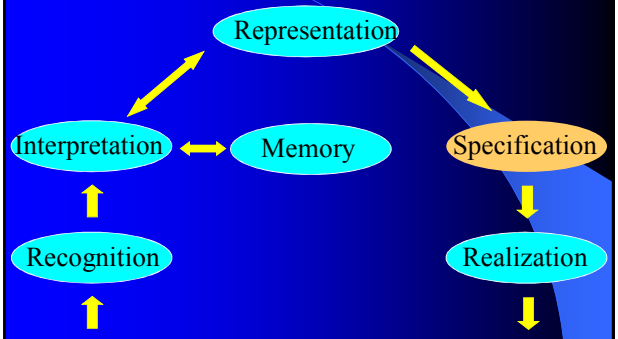

Dialogue Manager = Interpreter program



Generation (NL & motor action)



Specification of actions



Representational objects (Speech acts, Rhetorical acts and Situations)

- Constants
- Grounded predicates
- Variables
- Predicates with free variables
- Functions

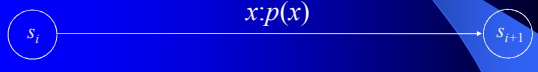
Constants

Definite intentions and actions

Grounded Predicates

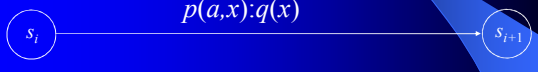
Definite intentions and actions

Variables



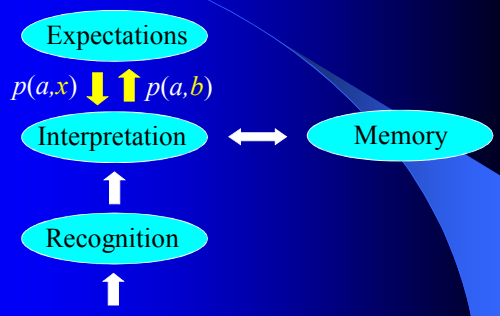
Abstract intentions and actions

Predicates with free-variables (The indexical context)

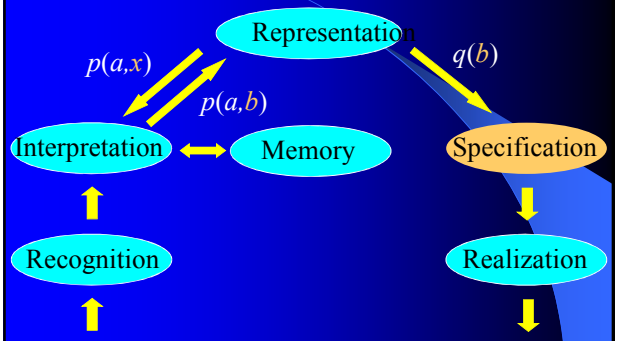


Abstract intentions and actions

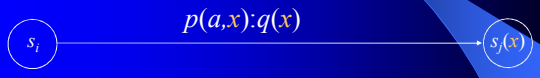
Collecting arguments from perception



Passing values from input to output

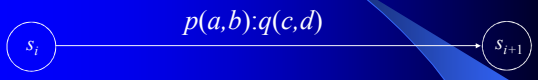


Situations can have parameters



Abstract intentions, actions and situations

The interaction history (anaphoric context)



The full path is recorded once an arc has been traveled (i.e. grounded), for all arcs traveled:

dm_1: s_i =>$p(a,b)$:$q(c,d)$ => s_j

The interaction history is a list of these terms

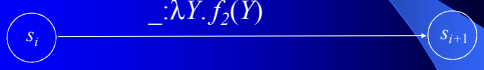
Functions

Functional definition of intentions, actions and next situations: X, Y and Z depend on the interaction history, the current DM and the current situation.

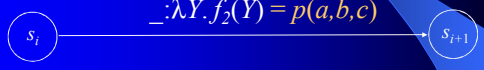
Functional Actions

Suppose a scenario in which the robot is giving a tour to a visitor and there is a situation where it has to make an offer, but taking into account the offers it has made before…

Functional Expression of Actions



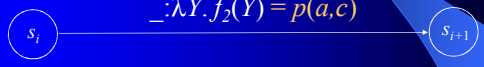
Evaluation…



Where p = *do you want me to do …*

Robot: *Do you want me to do a or b or c?*
User: *Please, do b*

But if we come again to this situation…



Where p = *do you want me to do …*

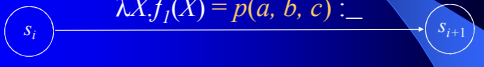
Robot: *Do you want me to do a or c?*
User: Please, do *a*

Functional Definition of Expectations

Suppose the robot is expecting to see one among a number of posters at a specific situation during the interaction (at the same position)

Functional Definition of Expectations

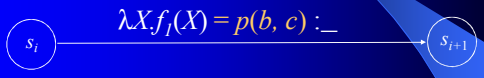
Evaluation…



Robot expects to see a poster *a or b or c* at the current situation

Robot recognizes poster *a*

But if we come to this situation again…



Robot focuses in recognizing *b* or *c*

The unexpected…

- Every situation can be followed by a recursive situation embedding a recovery dialogue model
- The information gathered is passed as a parameter to the recovery model
- The recovery model may use the interaction history…
  - *You already saw this poster… do you want so see it again?*

Functional Definition of Next Situations

Suppose the robot responds up to four questions from the user for each poster… then he or she might get bored!

Functional Definition of Next Situations

Functional definition of next situations

Providing information…

Where $s_i = s_{i+1}$

Where the system offers to explain a topic of the current poster (not explained before!)

But after the fourth time…

Where s_j is a "continuation dialogue":

Do you want to see another poster?

Declarative specification

```
[id ==> ls_2,
        type ==> listening,
        in_pairs ==>  [ls_1 => λx.f(x): λx.g(y),
                       ls_3 => ok:λx.h(x),
        out_pairs ==> [λx.f(a,x):λx.h(y)
                                => λz.rs(z)),
                       no:ra_3(tour) => fs],
          ],
```


Possible value's types of functions…

- Constants
- Grounded predicates
- Variables
- Predicates with free variables
- Functions

The full expressive power:

Functional Recursive Transition Networks (F-RTN)

Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and future work

Implementation

- Agents environment:
  - Open Agent Architecture (linux)
- Dialogue models spec. & interpretation:
  - Prolog
- Speech recognition:
  - Sphinx
  - Acoustic models: The Corpus DIMEx100
- Speech synthesis:
  - Mbrola & Festival

Implementation

- Robot navigation:
  - Player/Stage
- Vision:
  - OpenCV
  - SIFT
- Robot: Magellan Pro (RWI)
- Fixed Applications with PCs

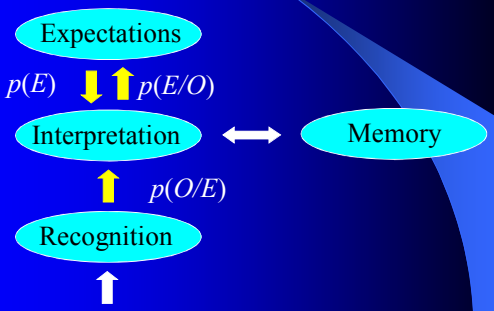
Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and Future work

Content

- Introduction
- A multimodal architecture for service robots
- Dialogue models
- Implementation
- Applications:
  - Tell me about the magazine!
  - Guess the card!
  - A guided tour with pointing acts
- Conclusion and future work

A Bayesian Architecture (with symbolic knowledge)



Future work

- Speech and language processing:
  - More roboust speech recognition (Language Models)
  - More general NLP capabilities (e.g. Semantic parsing)
  - Richer linguistic explanations

Future work

- Dialogue Models
  - A new DM interpreter (more general)
  - Introduce multimodal situations: listening and seeing
  - Extend the range of speech acts that can be understood (grounding strategies, turn taken strategies)
  - More general recovery strategies in the DM
  - Introduce a stochastic component in the F-RTN
  - Introduce internal actions (i.e. “thinking”)

Future work

- Vision and navigation
  - More intelligent navigation (e.g. Metric maps, qualitative plans and interaction cycles of move, see, ask)
  - More varied recognition and action devices
  - More control input devices handled from DM directly (contact, laser, ultrasonic, infrared, etc.)

Future work

- Architecture
  - Integration of reactive behavior
  - Intergration of schematic behavior
  - Use a multimodal data-base to implement the memory component

Construct diverse applicatios with the same platform!

Many thanks!