Human Robot Interaction

Emanuele Bastianelli, Daniele Nardi

`bastianelli@dis.uniroma1.it`

Department of Computer, Control, and Management Engineering

Sapienza University of Rome, Italy

SAPIENZA
UNIVERSITÀ DI ROMA

Introduction

- Robots are going to be even **more present** in everyday life
- Different Purposes:
  - Healthcare, Safety & Rescue, Entertainment, …
- **Non-expert-user** interaction capabilities needed

What is Human Robot Interaction?

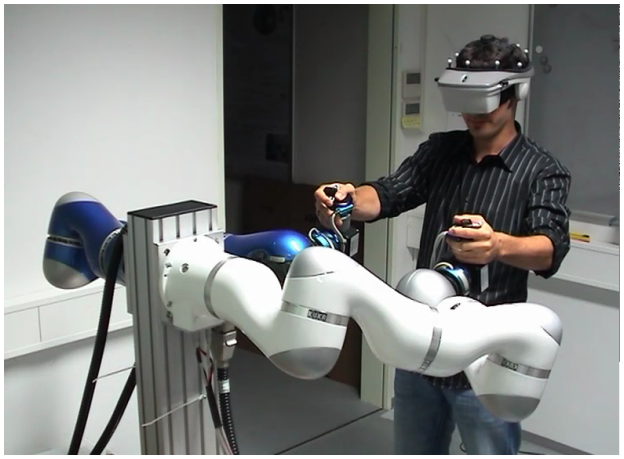
"*Human-robot interaction is the field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans*" (Goodrich)

HRI vs Human Computer Interaction

- Human-robot interaction is bidirectional (robots are not passive entities like computers!)
- Human-robot interaction is asymmetric (robots have not the same cognitive skills of humans)
- HCI techniques and metrics might not be applicable to HRI
- Robots are perceived as living entities

Human Robot Interaction

- *When human actions have effect on robots (and vice versa)*
- How many ways to interact with a robot?
- Tele-operation and other forms
  - operated using some physical interface
  - shared autonomy
- Safe Physical HRI
- Symbiotic Robotics
- **Social HRI**

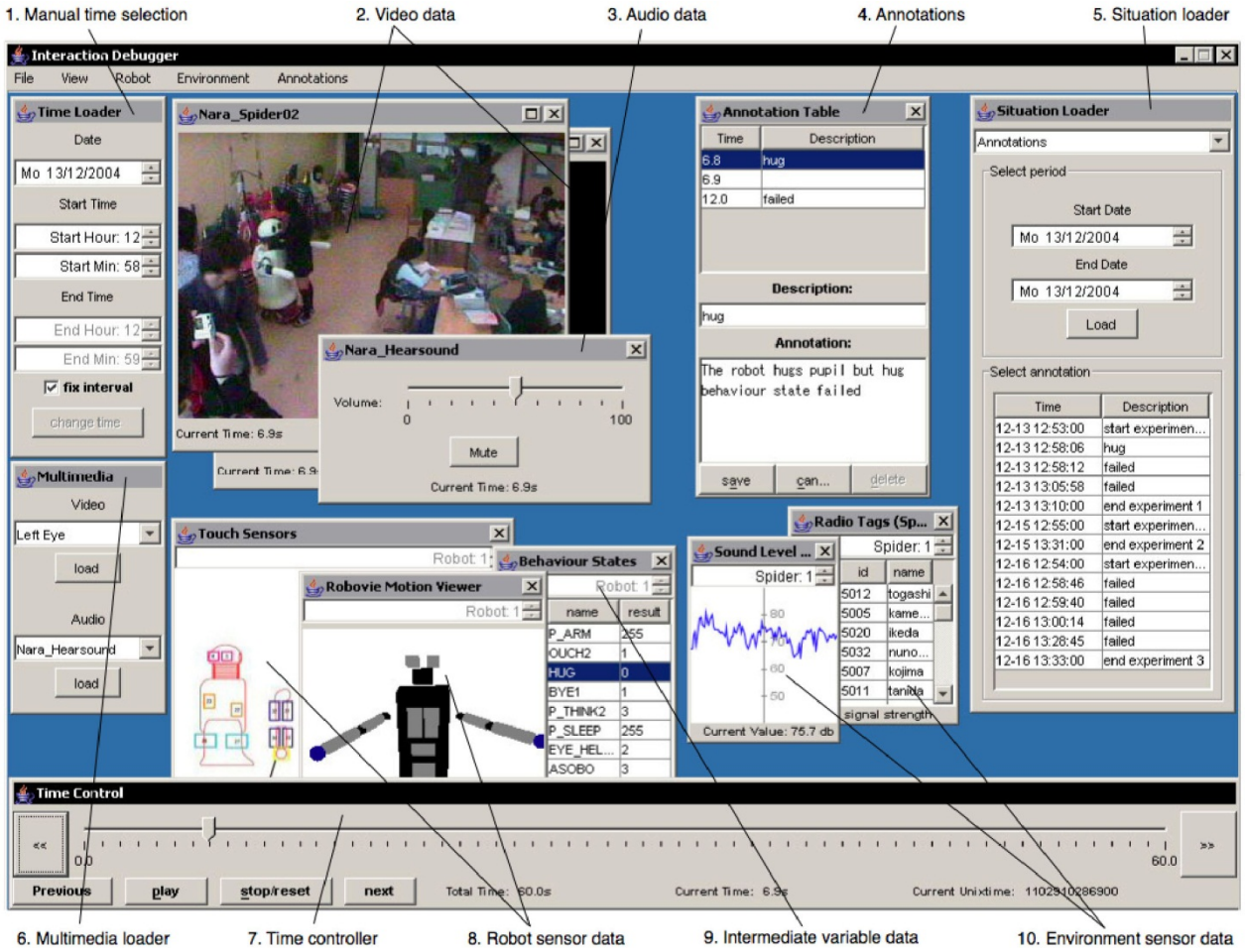


Human Robot Interaction

- **According to** [Scheutz 2011] **a robot with human-like interaction capabilities must be:**
  - *real time*
  - *parallel*
  - *spoken*
  - *embodied*
  - *situated*
  - *dialogue based*

Scheutz, M., Cantrell, R., Schemerhorn, P.: *Toward humanlike task-based dialogue processing for human robot interaction.* AI Magazine 34(4), 64–76 (2011)

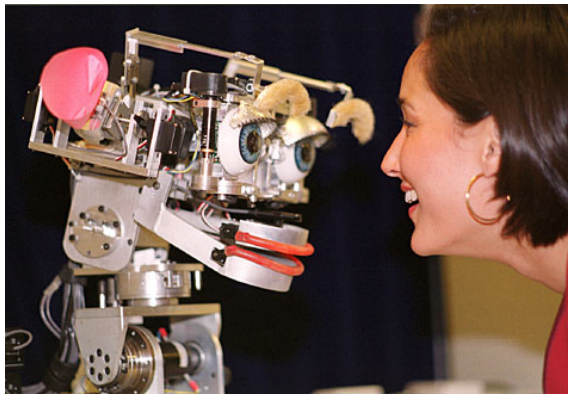
Not a social interface



Social Human Robot Interaction

- *Modeling interaction between humans and robots as the "natural" interaction between humans*

- HRI studies a variety of interaction modalities
  - **Natural Language**
  - Gestures
  - Facial Expressions
  - Non Verbal Interactions
  - Empathy
  - …



Possible Input (for the Robot)

- Hand-held devices
- Speech
- Touch
- Temperature
- Olfaction

People

- Positions and velocities
- Gestures
- Race? Gender?

Head

- Gaze
- Facial Expressions

Possible output (of the Robot)

Body

- Position
- Speed

Head

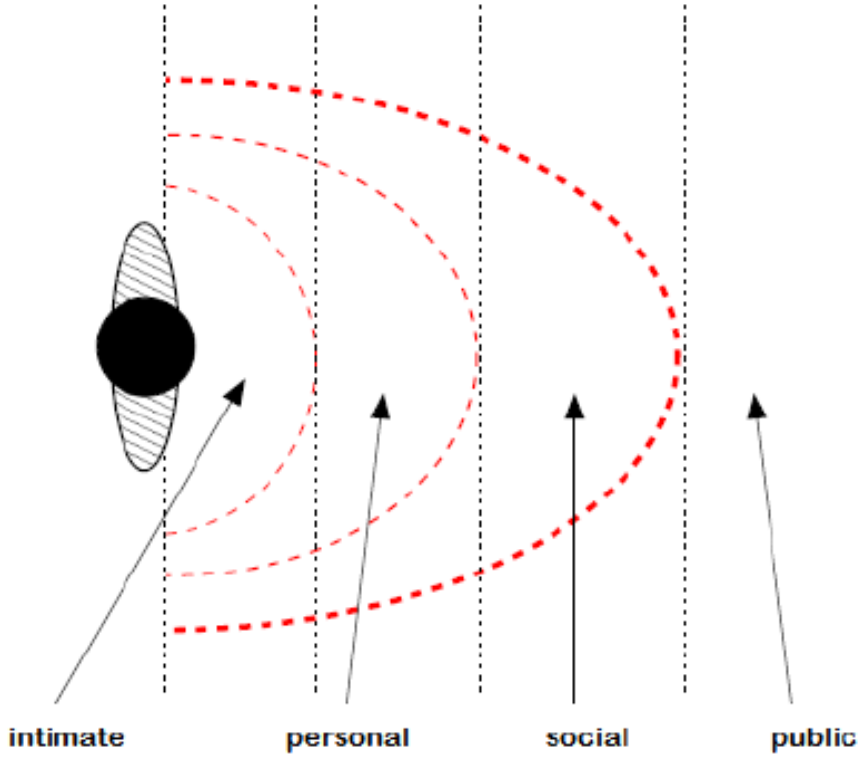
- Turning
- Eye motion
- Facial expressions

Arms

- Grab objects (shake hands)
- Speech
- Sound
- Lighting

Proxemics

- Posture
- Facing angle
- Distance
- Touch
- Eye Contact
- Thermal Heat
- Smell
- Vocal loudness



Human Robot Interaction in Natural Language

- Natural Language is an expressive, flexible and intuitive interface

- **Aims of Natural Language HRI**: *providing robots with the ability of interacting in a natural way with humans, using NL* (aka **Natural Language Understanding**)

- Imply complex processing: robots need to understand and reason on what is being said
  - Speech Recognition
  - Natural Language Processing and Understanding
  - Grounding

Grounding

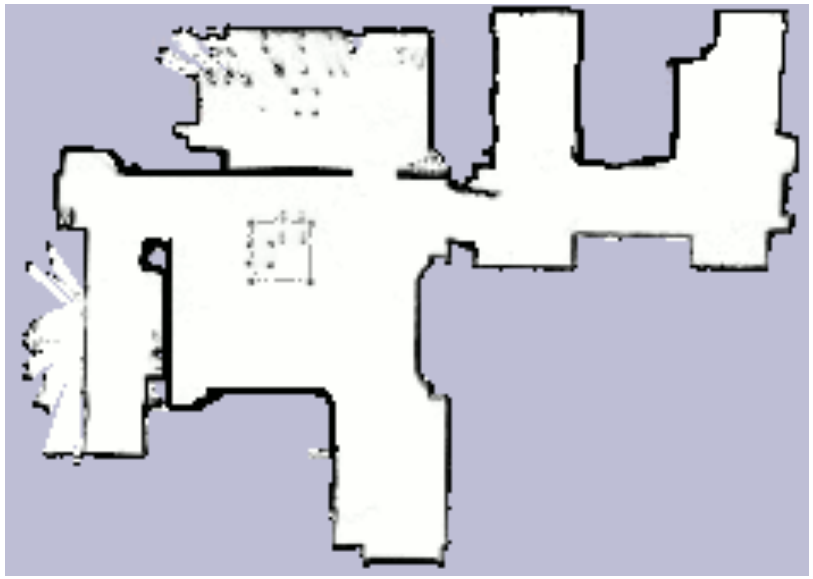
- *"ground a symbol meaning in something other than just more meaningless symbols"* [Harnad,1990]

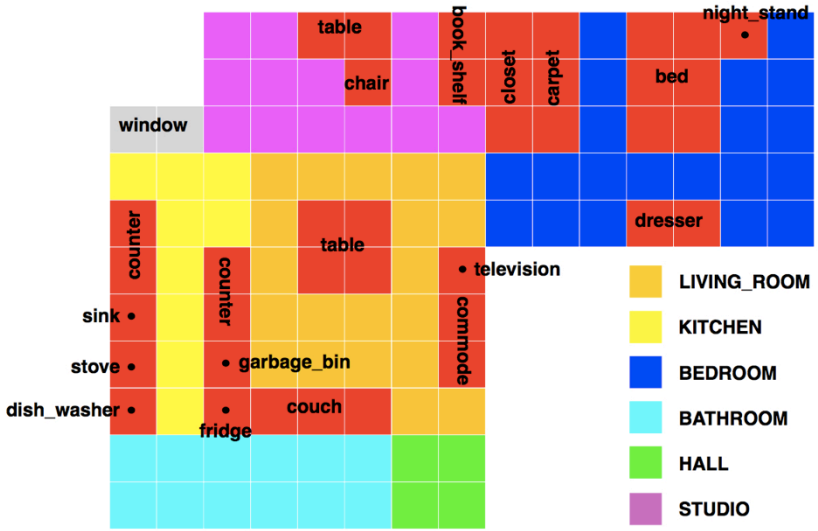
- or Anchoring: *"the process of creating and maintaining the correspondence between symbols and sensor data that refer to the same physical objects"* [Coradeschi&Saffiotti,2003]

" bring the can in the trash bin"

Semantic Maps

"A semantic map … is a map that contains, in addition to spatial information about the environment, assignments of mapped features to entities of known classes. Further knowledge about these entities, independent of the map contents, is available for reasoning in some knowledge base with an associated reasoning engine." [Nüchter&Hertzberg,2008]





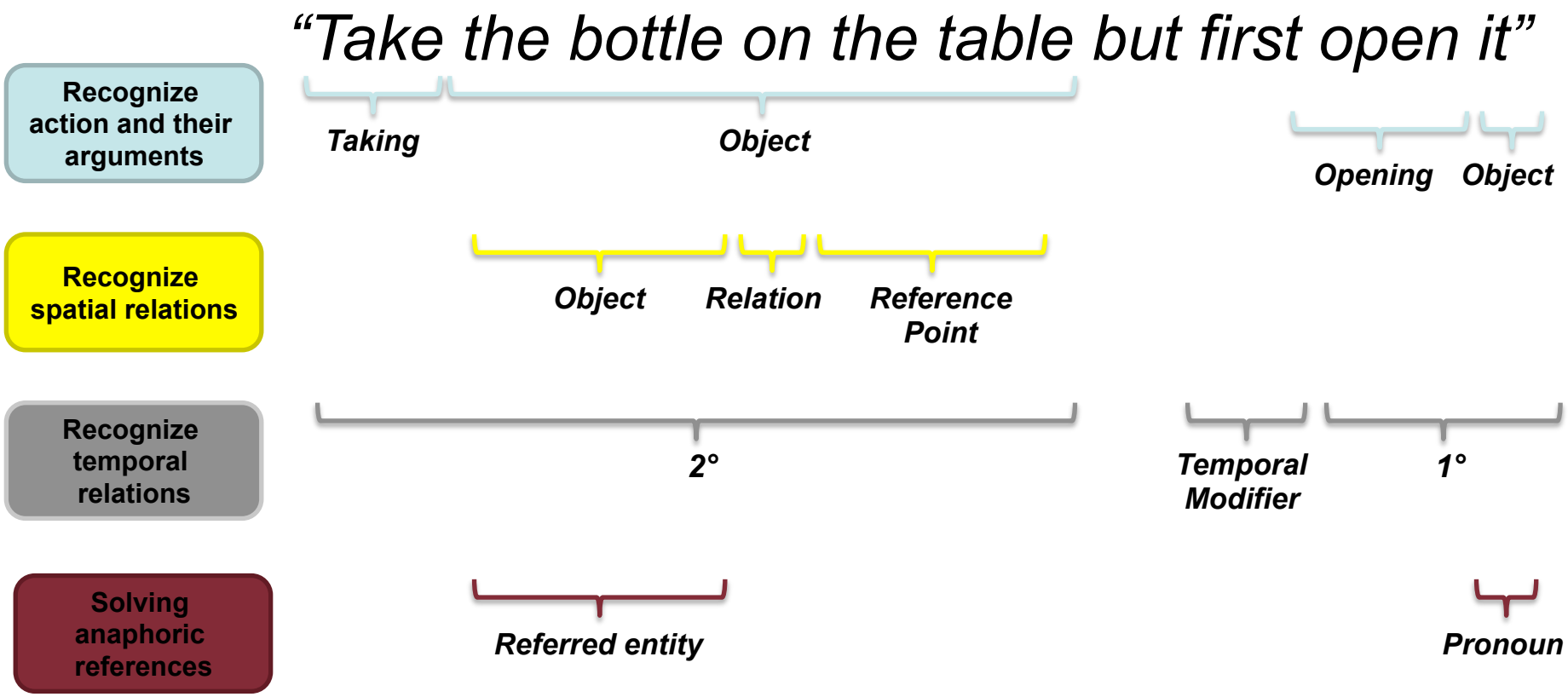
Semantic Maps

- **Grounding**: semantic maps are needed to close the loop with perception
- *Semantic Mapping* is the process of building semantic maps
- Fully automatic semantic mapping
- Human Augmented Mapping
  - Involves the interaction with the user

Human Augmented Mapping

Dipartimento di Ingegneria Informatica
Automatica e Gestionale Antonio Ruberti
Sapienza
Università di Roma

Natural Language Processing

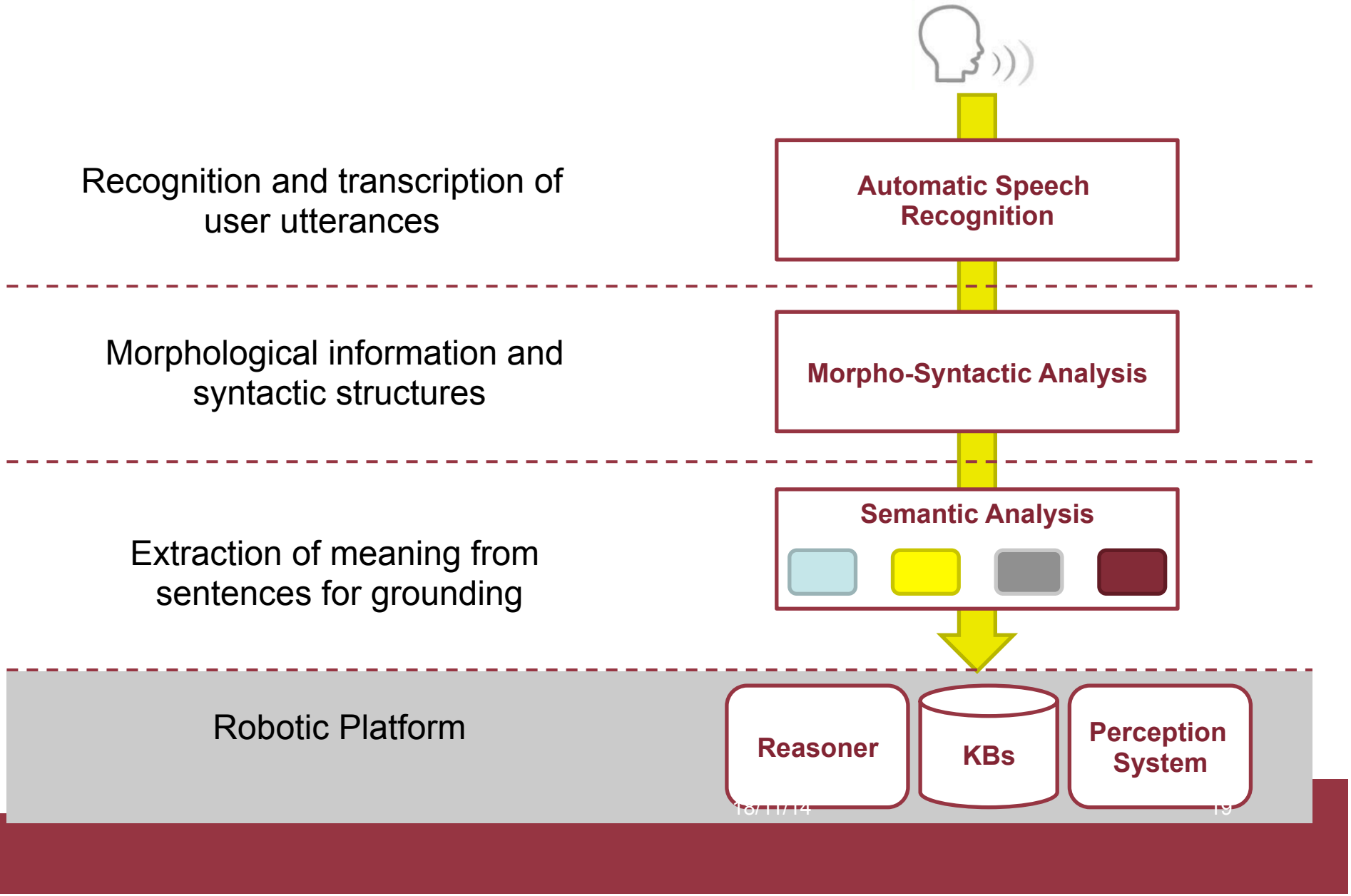


. . .

Natural Language Processing

- How can we provide all this information to the robot?
  - Natural Language Processing: **Semantic Analysis**
- Different Semantic Theories vs. One Single Theory
- Complex and challenging task
  - Many different approaches proposed

Processing Chain



Automatic Speech Recognition

- *Translation of spoken words into text*

- Command and control
  - Grammar Based
  - High Performance
  - Controlled Language

```
S      -> Verb Object
Verb   -> “take” | “grab” | …
Object -> “bottle” | “glass” | …
…
```

- Morpho-syntactic and Semantic processing can be embedded in the recognition process
  - Semantic Attachments

Automatic Speech Recognition

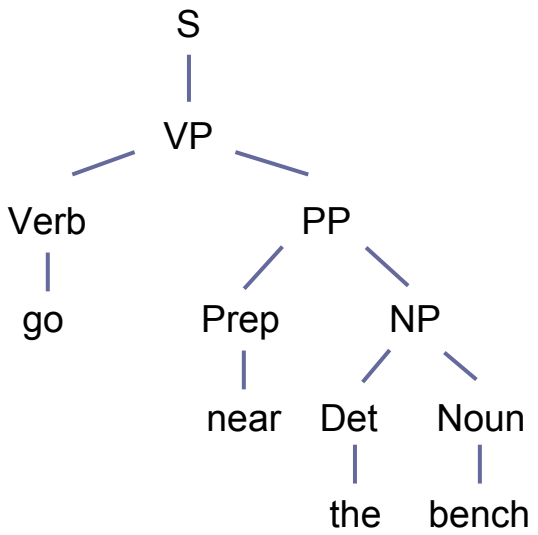
- *Translation of spoken words into text*

- Free form speech
  - Based on huge models acquired by learning
  - High computation capacity needed
  - Open Language

- Needs some subsequent processing to interpret the recognized utterances

Morpho-Syntactic Analysis

- *Morphological* and *Syntactic Analysis* produces information about grammatical nature of words and assigns syntactic structure to sentences
  - features used in the semantic parsing processing



- Performed as preprocessing step for Semantic Analysis

Semantic Analysis

- *Semantic parsing* (analysis) used to give a structure to the meaning of a sentence

Extraction of all semantic aspects needed for grounding

- One single module or cooperation of dedicated processors
  - Semantics of Actions
  - Spatial Semantics
  - Temporal Semantics
  - …

Semantic Analysis

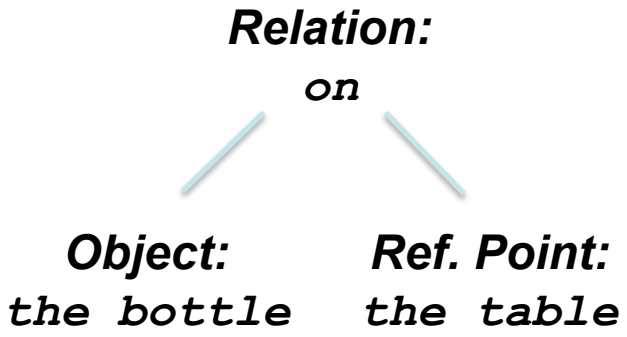
- *Semantic parsing* (analysis) used to give a structure to the meaning of a sentence
- Example:

"take the bottle on the table"

Actions

Taking
Verb: *take* Theme: *the bottle* Source: *on the table*

Spatial Relations



Homework 1/4

Giving command to MARRtino in Natural Language (implementing a simple NL processing chain)

1. Recognizing “open loop” motion commands…
   - “go forward”
   - “go backward”
   - “turn right”
   - “turn left”
   - “stop”
2. …and of *grounded* motion commands
   - e.g. “go to the kitchen”, “move near the closet”, …
   - Semantic Map needed

Homework 2/4 – Processing steps

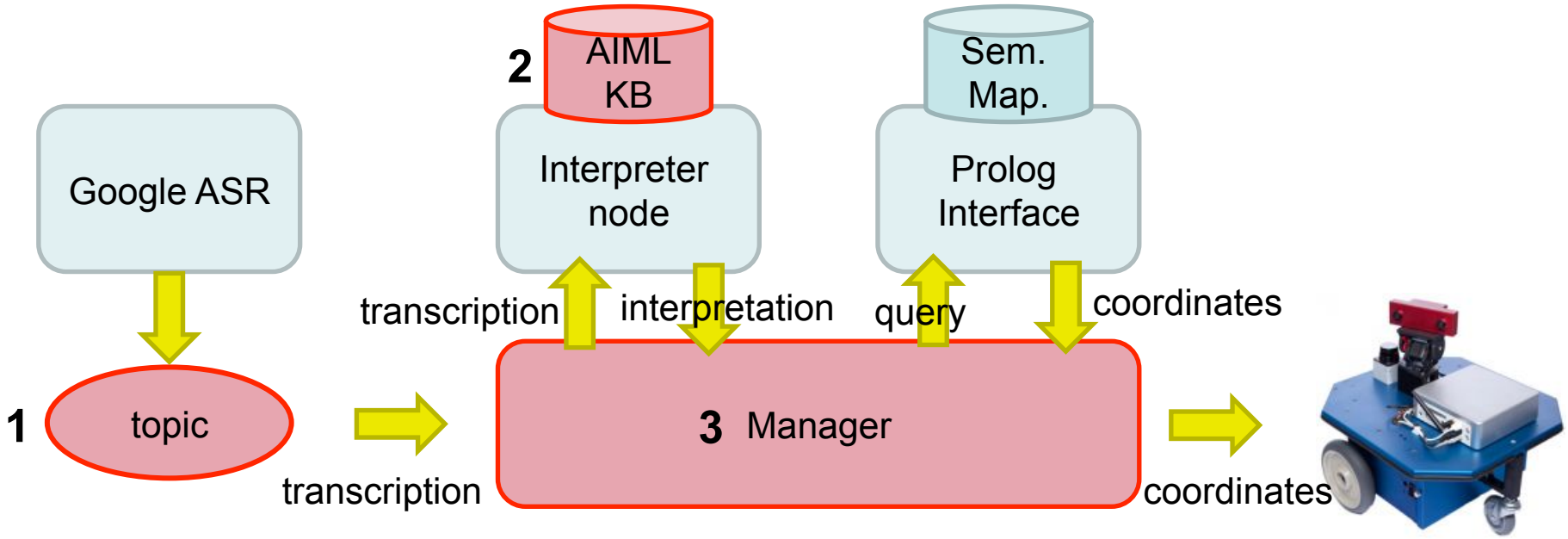
- Google Speech to Text for Speech Recognition
- Semantic interpretation of transcriptions through Artificial Intelligence Markup Language (AIML)
- Grounding through a Semantic Map, queried using Prolog

Homework 3/4 - Implementation

- **What will be provided**:
  - Interface to Google ASR (Python)
  - ROS node embedding an AIML interpreter (Python)
  - ROS node embedding an interface to Prolog (C++)
  - A Metric Map of an environment (.ppm)
  - A Semantic Map built on the Metric Map of the same environment (Prolog)

Homework 4/4

- What you need to implement:
  1. A ROS node embedding the Google ASR
  2. A simple AIML Knowledge Base to parse motion commands
  3. A ROS node that manages the interaction between the modules and execute the final command



Google

- You need an API key.

1. Go to this link: https://cloud.google.com/console and create your own project.
2. Join this group here: https://groups.google.com/a/chromium.org/forum/?fromgroups#!forum/chromium-dev
3. In your project go to APIs & auth > APIs , and activate Speech API (only 50 requests for each key).
4. Go to Credentials and make your client.
5. Generate a Browser key.

- For more information: http://www.chromium.org/developers/how-tos/api-keys

AIML 1/3

- Provides a method to interpret Natural Language
- *"Stimulus/Response"* (S/R) pattern (used in common chatbots)
- **Stimulus** represents what the user may say, and is the input of the Interpreter
- **Response** represents what the user expects as answer, given the corresponding simulus. It is the output, that can be:
  - A string
  - A system call

AIML 2/3

- The language to be interpreted is defined using AIML, an extension of XML language

- Implements the S/R pattern
  - stimulus coded as <pattern> tag
  - response coded as <template> tag

- e.g.:

```
<category>
    <pattern>go to the kitchen</pattern>
    <template>action:GOTO_dest:kitchen</template>
</category>
```

AIML 3/3

- AIML supports the use of regular expressions inside the <pattern> tag

```
<?xml version="1.0" encoding="UTF-8"?>

<aiml version="1.0.1" xmlns="http://alicebot.org/2001/
AIML-1.0.1"
    xmlns:html="http://www.w3.org/1999/xhtml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://alicebot.org/2001/
AIML-1.0.1 http://aitools.org/aiml/schema/AIML.xsd">

<category>
      <pattern>GO TO THE *</pattern>
      <template>action:GOTO_dest:<star/></template>
</category>
```

PyAIML

- The AIML ROS node use PyAIML as interpreter
- Can be downloaded from http://pyaiml.sourceforge.net/
- Very easy to install
  - Just run the `setup.py` script
- Import the library with `import aiml`

Prolog Interface 1/2

- ROS interface to Prolog
- It consults Prolog file defined in the launch file

```
<launch>
<node pkg="PrologInterface" type="prologInterface"
name="prologInterface" output="screen">

<param name="prolog_binary" value="path_to_prolog_binary" type="str"/>
<param name="prolog_path" value="path_to_prolog_file_dir" type="str"/>
<param name="prolog_file_1" value="file_name_1" type="str"/>
<param name="prolog_file_2" value="file_name_2" type="str"/>

</node>
</launch>
```

Prolog Interface 2/2

- Qurerying service to the consulted Prolog KB
  - service name: `prolog_query`

- `prologSrv.srv` structure

```
string predicate
string[] arg
---
solution[] ris
```

- `solution.msg` structure

```
string[] atoms
```

Semantic Map

- Basic version of a semantic map
- Reports only coordinates about rooms and objects in a Prolog KB
  - X, Y, Theta coordinates

```
locationOf(kitchen,20,40,180).
locationOf(dining_room,10,60,0).
locationOf(sofa,40,20,90).
…
```