Challenges for Dialog in Human-Robot Interaction

Dialogs on Dialogs Meeting

October 7th 2005

Hartwig Holzapfel

About me

- Studied Computer Science in Karlsruhe (Germany)
- Minor field of study Computational Linguistics Stuttgart (Germany)
- Diploma Thesis on Emotion-Sensitive Dialogue at ISL, Prof. Waibel
- Scientific employee/PhD student at Karlsruhe/Prof. Waibel since 2003
- Recent Projects
  - FAME: EU Project: Facilitating Agents for Multicultural Exchange presented at Barcelona Forum/ACL 2004
  - SFB 588: collaborative research effort at Karlsruhe on Humanoid Robots
- Research (within above projects):
  - Multimodal (speech+pointing synchronous and fleximodal)
  - Multilingual Aspects
  - ASR in dialogue context
  - Current: Cognitive Architecture for Robots and Learning

Outline

- Robots

- SFB588: The humanoid-Robots project
- The Robot „Armar“
- Interaction scenarios

- Multimodal Interaction
- Multilingual Speech Processing
- Cognitive Architectures

- Open Tasks

Humanoid Robots

- Why humanoid:
  - Humanoid body facilitates acting in a world designed for humans
  - Use Tools designed for humans
  - Interaction with humans
  - Intuitive multimodal communication
  - Other aspecs like understand human intelligence

- Kind of Humanoid Robots
  - Service Robots
  - Assistants
  - Space
  - Help for elderly persons

Humanoid Robots (some examples)

- Cog
- ASIMO
- QRIO
- GuRoo
- Kismet
- Nursebot
- PINO Open Plattform
- HOAP 2
- Sarcos Robot
- Robonaut
- ARMAR

SFB588 - the humanoid Robot Project

- Started 2001
- 2nd phase started 2004 targeting for an integrated system
- Current robot-platform ARMAR
- New platform in development
- Goals: Household and Kitchen scenarios


Selected Interaction Scenarios

- Loading and unloading the dishwasher
- Proactive behaviour: coffee service
- „Bring me something“

Bring me something

- Interaction:

- Detect persons
  - Detect person visually
  - Respond to person
- Initiate Interaction (what can I do for you?)
- Recognize speech (distant?) and gestures (bring me this cup)
- Locate objects, update environment model
- Find, go to, grasp, and bring object to person
- Recover from error states

Challenges

- Multimodal communication
- Multilingual (our Robot lives in Germany)
- Uncertain information about environment
- Distant speech
- New words, new objects and new actions
  - Semantic description
  - Attributes
  - Visual features
  - Task description
- Introducing new persons
  - Name, Hobbies, ..
  - Visual ID
  - Voice ID
- Floating domain-boundaries

Multimodal Interaction

Multimodal Interaction with

A humanoid robot

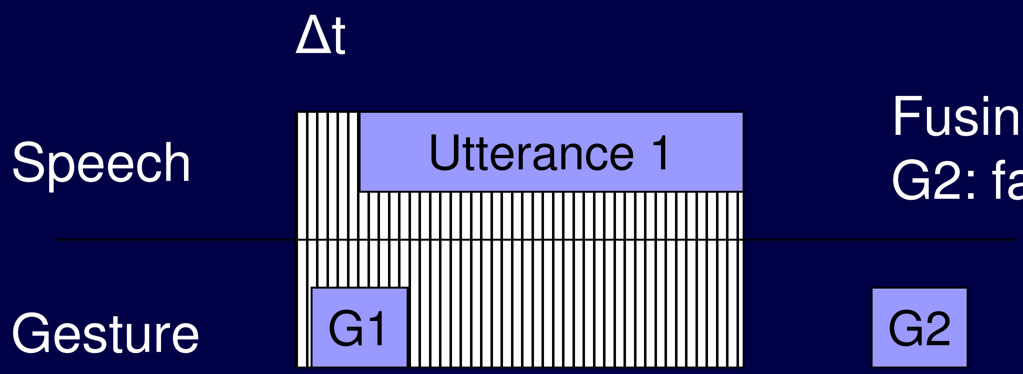
- **Visual Perception of the user**
  - Person Tracking
  - Gaze / Head orientation
  - Gesture Recognition
- **Speech Recognition**
  - Distant microphones
  - Spontaneous speech
- **Dialog Manager**
  - Multimodal Parsing

Take the cup!

"Which cup do you want me to take?"

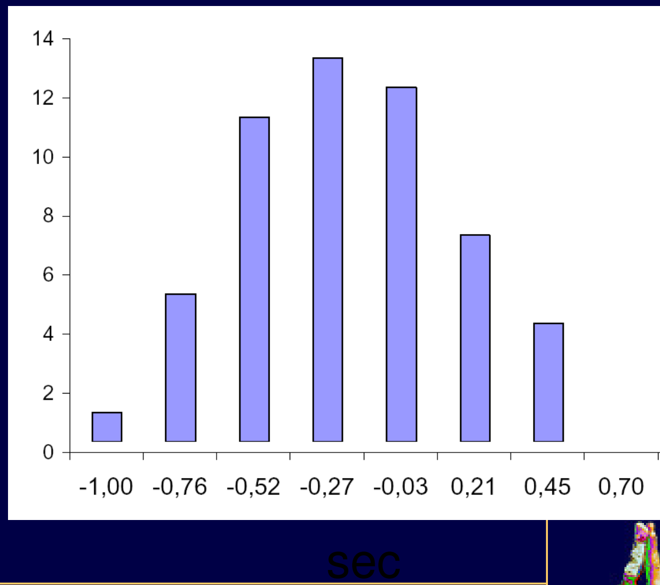
This one!

Multimodal Fusion

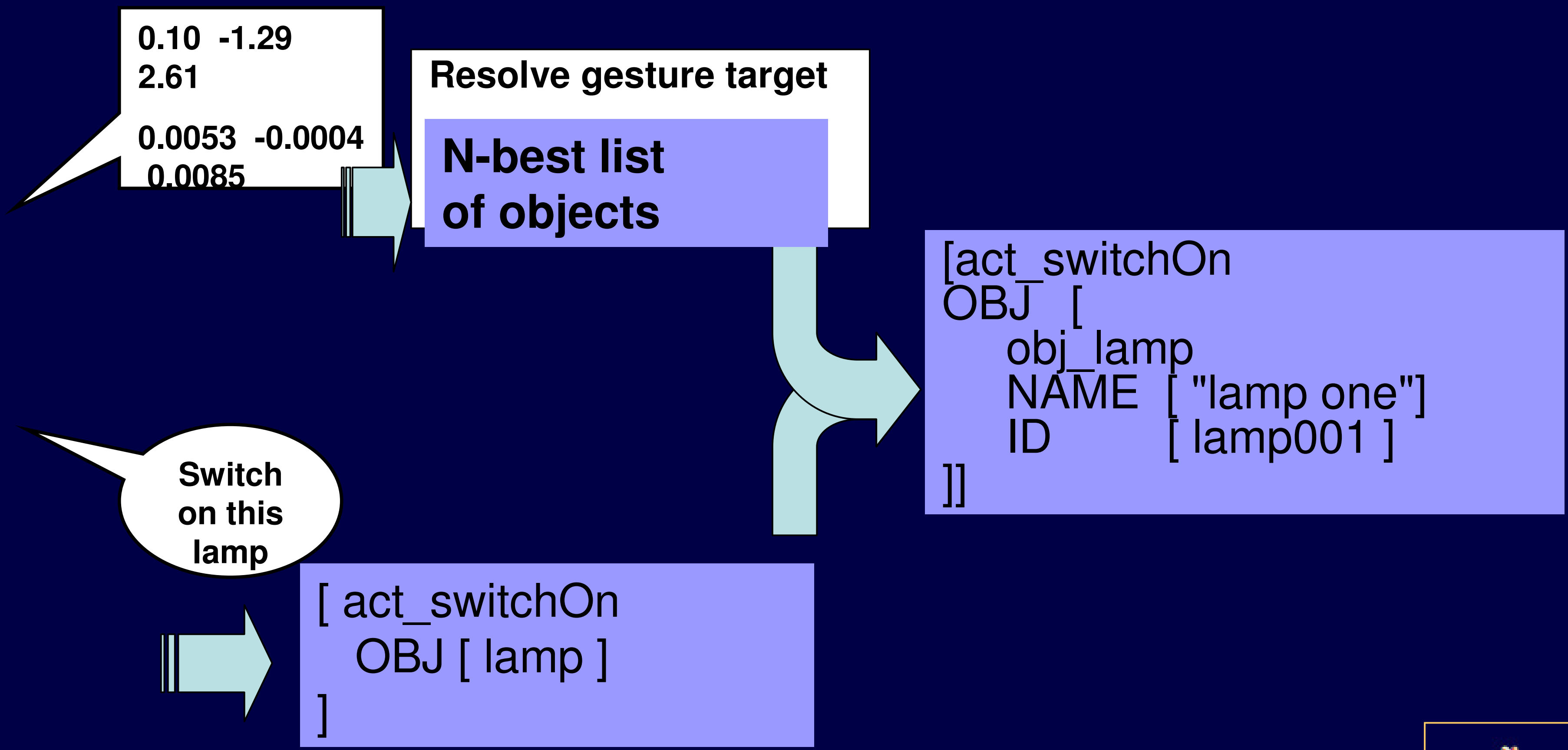


Fusing utterance 1 and G1
G2: false positive

Temporal correlation between
Speech and pointing gesture

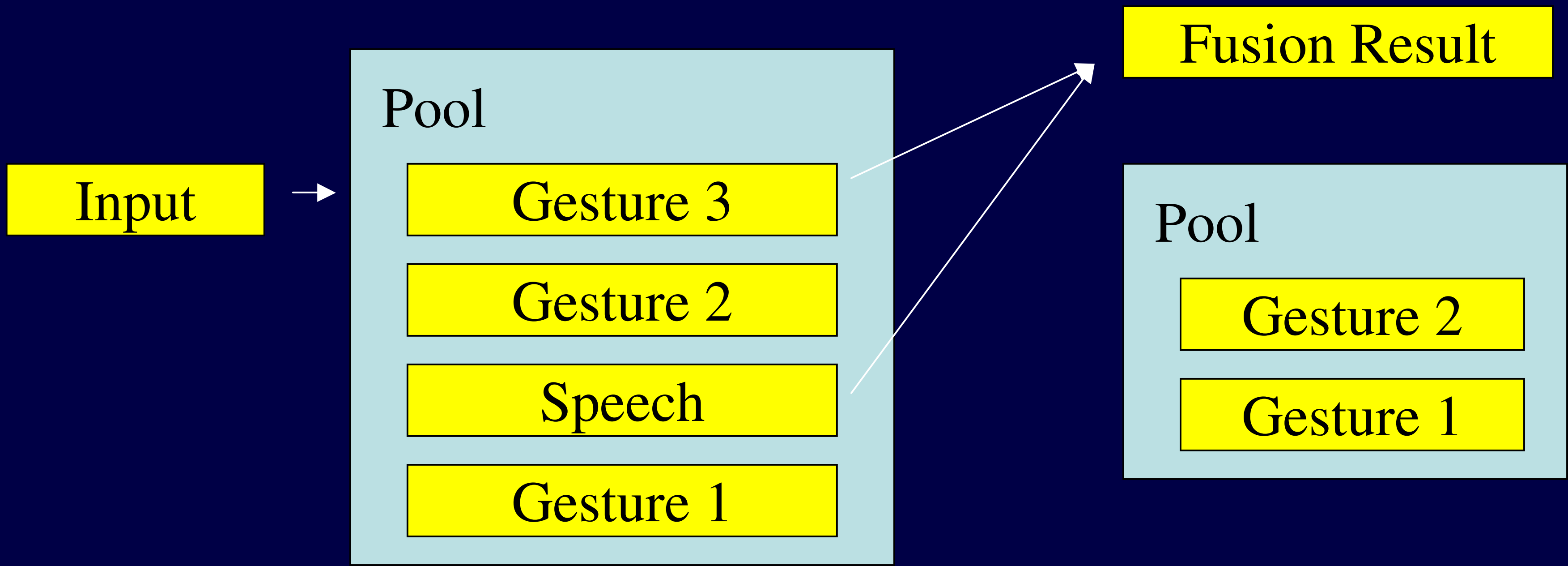


Fusing Speech and Pointing Gestures



Multimodal Parsing

- Pool of semantic tokens
- Parsing rules for fusion of tokens



Experiments and Evaluation

- Fusion for n-best Lists

Gesture Detection (rt)	
Recall	87%
Precision	47%

Gesture Recognizer (rt, relative errors)	
First Hypo	44%
N-best	94%

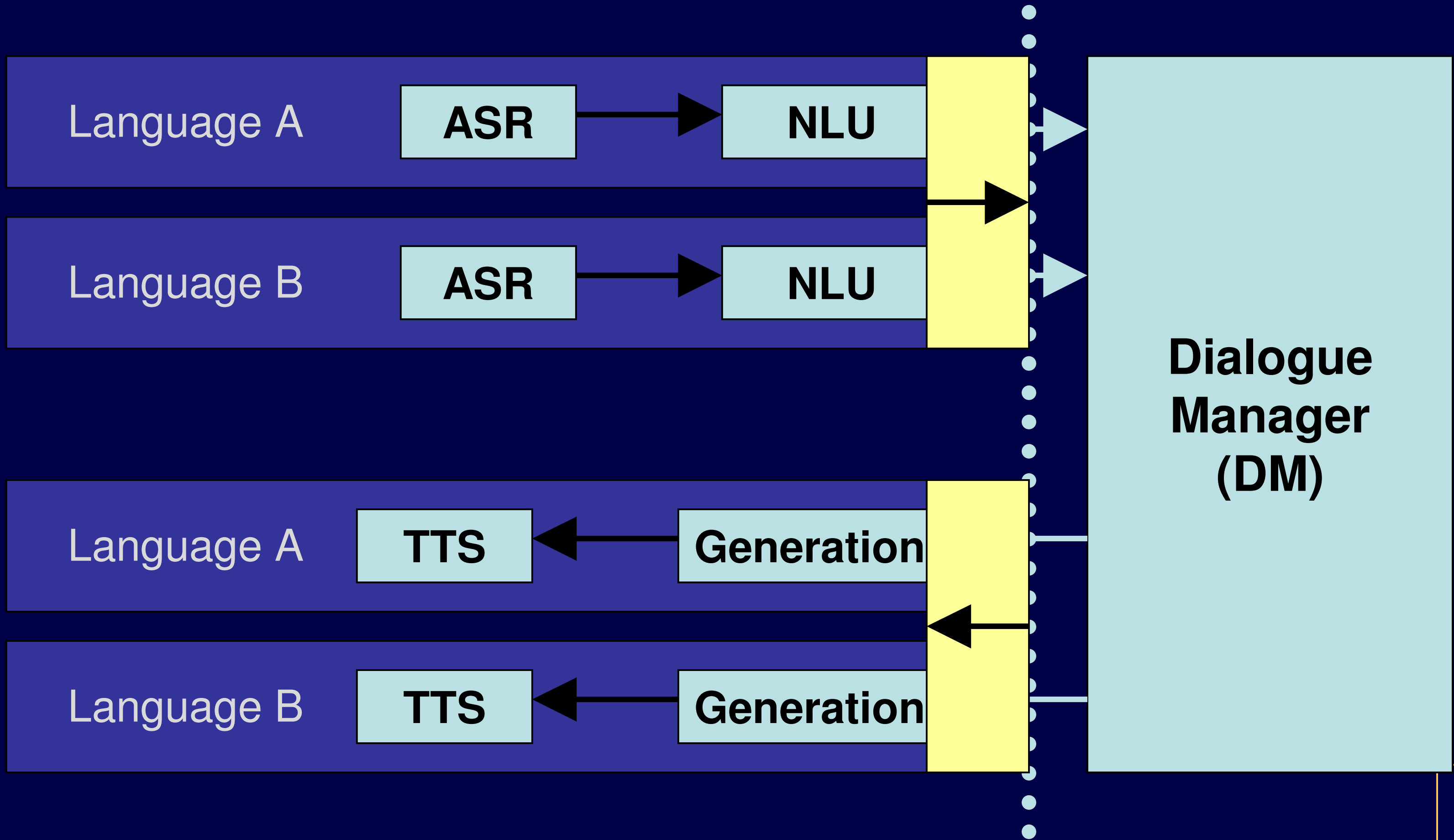
Speech Recognition (0,8*rt)	
WER	24%
SER	33%

Fusion	
Speech + Nbest G	74%
Nbest S + nbest G	76%

Multilingual Speech Processing

- Why?
  - German lab,
  - To get native speakers we need to build a German system
  - However, best ASR system is English
  - International Visitors

Designing a multilingual system

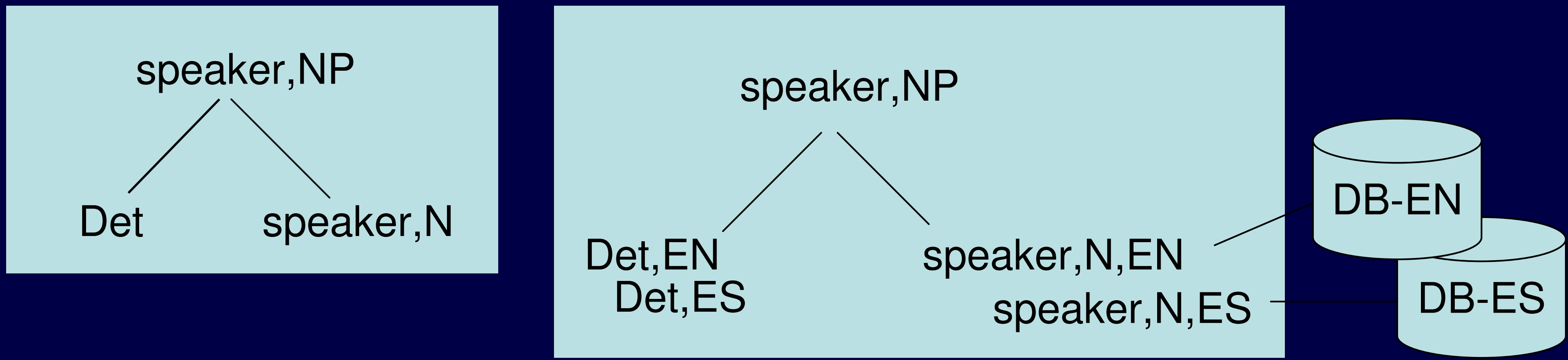


Input Grammar - Rule Interfaces

- Software engineering offers principles for programming languages
- Usage of Interfaces for common functionality
- Rule interfaces define
  - Common semantic information
  - Abstract grammar nodes

Multilingual information from databases – semantic grammars

- Proper nouns are read from databases
  - Syntactic phrase structure
  - Imported nouns form construct rules



Speaker,N,EN -> ‚name1‘ : ‚name2‘ : ‚name3‘;

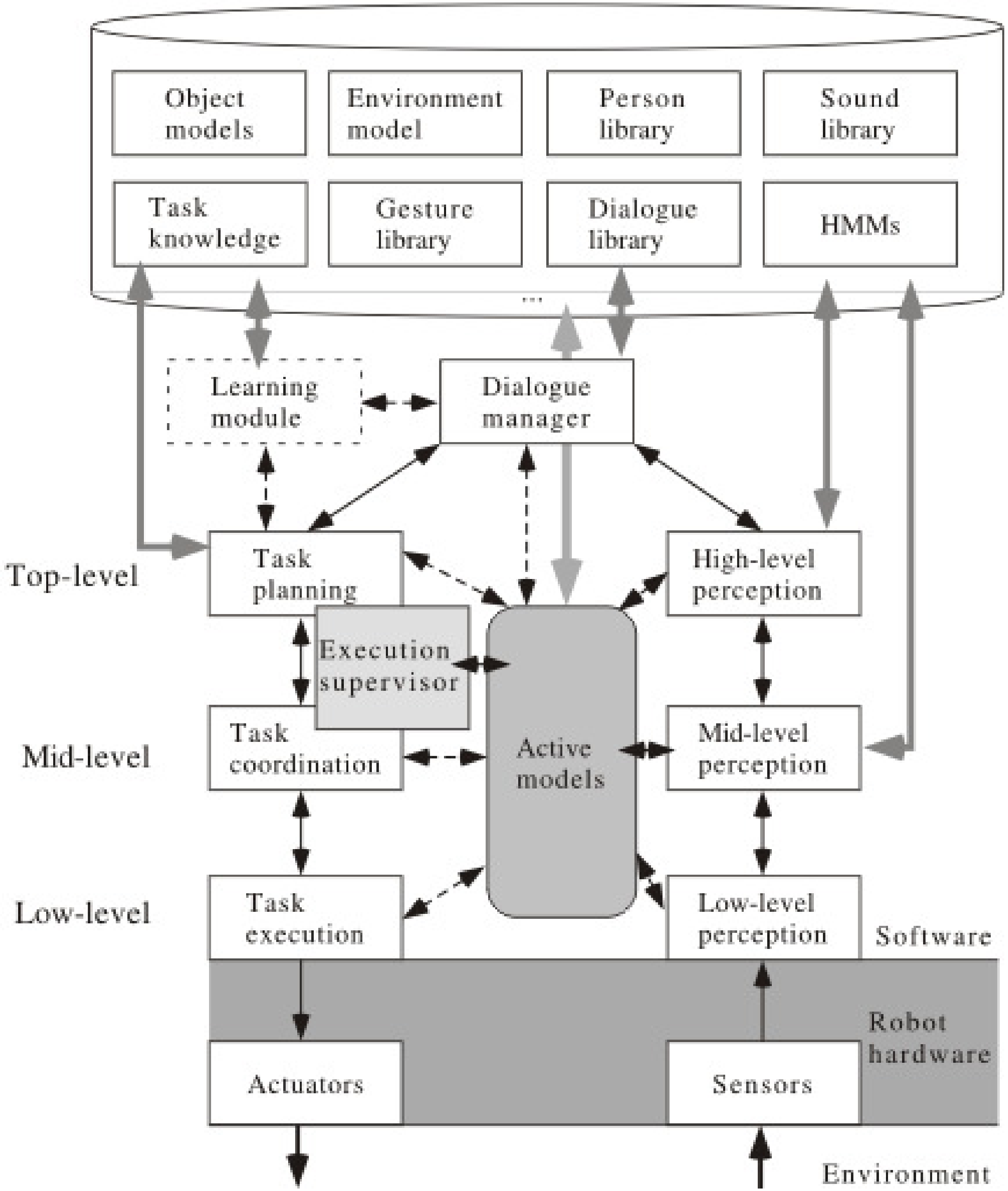
Experiences of using these concepts

- FAME demonstrator (http://isl.ira.uka.de/fame)
  - 5 persons working on grammars: 2 English, 2 Spanish, 1 German, only English as output
  - English and Spanish developed in parallel roughly same amount of time, German developed afterwards by using rule interfaces and grammar porting
- SFB humanoid robots (German research effort http://sfb588.uni-karlsruhe.de)
  - 3 persons working on grammars and generation:
    2 English (experts) - developing, 1 German (student) - translating
  - German application works reliably (grammars and generation)

Cognitive Architecture

- Integrate dialogue into complete system architecture
- Distribution of cognitive abilities:
  - Simple dialogue manager with intelligent controller architecture
  - vs. Cognitive abilities in dialogue control
- Both approaches already exist
  - Dialog centered systems with control of background application
  - Vs. Intelligent architecture and adding speech commands
- Our current approach tries to model the complete architecture for a robot, dialogue only as a component
  - Competing Model of input by the user and current robot tasks
  - Conflicting resource access

Cognitive Architecture



Communication Models

- Interpret and forward User commands to the platform
  - Test if actions are possible
- Receive information by the platform to resolve information => query user
  - Request new information
  - Recover from errors
- Maintain user‘s goal model, update according to system and task state
  - Request information from system model
  - (When is the goal fulfilled)
  - Challenge: Interpret input by the user in the right context
- Request output channels (speech/multimodal)
- Request resources to receive input by the user

Open Tasks

- Initiate Interaction: detect persons, obtain attention and start dialog
- Attention modelling
- Learn new objects
  - Detect unkown words referencing objects
  - Introduce words, semantic meaning
  - Get visual “understanding” of these objects
- Learn about persons
  - ID: voice and vision
  - Names: new words
  - Social relations: what is this person doing here?
- Learn new actions
  - New sentence constructions
  - Relate semantics to robot actions