We present a cognitive architecture for humanoid robots interacting with objects and caregivers. The architecture is foundational to the MACSi Project: it is designed to support experiments to make a humanoid robot gradually enlarge its repertoire of known objects and skills combining autonomous learning, social guidance and intrinsic motivation. This knowledge may lead the robot to autonomous learning of affordances.

**COGNITION**
- Persistent knowledge, learning processes, reasoning, affordances, ...
- Autonomous behavior
- Decision making
- Social guidance

**EGOSPHERE**
- Proprioceptive data
- Episodic knowledge
- Contextual information (people, objects...)

**PERCEPTION**
- Cameras
- Proprioception
- Sound sensor
- Rgbd-sensor

**ACTION**
- Motor commands

**OBJECTS RECOGNITION**
- Objects in the scene are extracted from depth maps through segmentation; their visual appearance is characterized by complementary features such as SURF and SUPERPIXELS' properties.

**MULTIMODAL PEOPLE TRACKING**
- Human caregivers interacting with the robot are identified through a multimodal approach, combining embodied information (cameras) with external sensors (rgbd-sensors and microphone array).

**USING SOCIAL GUIDANCE**
- The human caregiver asks the robot to recognize an object, take it and put it in a box: on success, a positive reward is given to the robot.

**TOWARDS LEARNING AFFORDANCES**
- The robot pushes an object to detect its motion and learn the effect of its action (e.g. the blue car rolls).