Motivation

- Problem of connecting language to the world
- Current NLP systems construct representations of meaning that use symbolic descriptions (first order predicate calculus) of the world as conceived by human designers.
- ‘Circular definitions’
- Need Natural language systems that can interpret the speakers meaning in everyday scenarios.
- Words make sense when they refer to the particulars of a particular situation.
- The listener cannot do the ‘right’ thing unless it knows something about the physical situation to which the words refer.
Possible avenues to explore..

• Use example from robotics and avoid knowledge representations which rest on symbolic, human generated descriptions of the world.
• Systems must connect internal representations to external physical entities.
• Proposes a theory of Language grounding

Progress..
Meaning

• Words refer to entities and situations in the world – ‘fly’ and ‘soup’
• Language use is situated. ‘this’ and ‘is’ connect the speech act to a region of space and time
• Agents use language to pursue goals : “get me another soup” or “I will report to the manager”.

Grounding

• Grounding is a causal predictive cycle by which an agent maintains beliefs about the world.
• Belief is an explicit information structure that exists in the mind of the agent.
• Primary belief - concrete, physical world of objects, spatial relations.
• Abstract Belief – primary beliefs connected by analogy and metaphor – ‘Ideas are food’.
Desiderata for a Theory of Language Grounding

1. Unification of representational primitives – objects, properties, situations should be constructed from the same set of primitives.
2. Cross-modal Translatability – represent information derived from perception and language in a common form.
3. Integrated space of actions – Motor and speech acts should be expressed in a single integrated space of actions.

Theory of signs, projects and schemas

- A sign is a physical pattern that which only exists as a sign relative to an interpreter.
- Signifies an object
- Three classes:
  - Natural – shaped by nomic physical laws. The agent is attuned to specific sensory input and can only detect signs that appear within those channels.
  - Indexical
  - Intentional – situate beliefs relative to the frame of reference. Necessary as parameters for control.
Beliefs about signs

- a-signs: “a distribution over all possible observations within an input channels domain”
- form elements of schemas that enable an agent to encode signs and make context-dependent predictions about the observation of new signs

d-signs: belief about the output of a discrete categorization process which maps a continues domain a-sign to a discrete domain.

Ripley – ‘a conversational helping hand’

- Ripley, to investigate connections between natural language semantics, perception, and action
- Ripley is a custom constructed manipulator robot designed for grounded language experiments
- Features – 7 DOF, hardcoded to look up and “make eye” contact with people.
- A visual subsystem that tracks regions and maintains correspondence between regions.
- Detection and tracking of a region – “an object that is instantiated in Ripley’s mental model”
- It understands a limited set of spoken requests.
- There is a spatial language interpreter which maps requests into goals with respect to objects in Ripley’s mental model.
Ripleys representation of underlying worlds

- Schema to represent active perception required for touching to gauge compliance, that provides a basis for grounding words as ‘soft’ and ‘hard’.
- Close grip – action primitive, runs a motor controller connected to the grip motor.
- A sensor projection, senseGripResistance is connected to D and projects an a-sigh with ‘compliance’.
• Ripley perceives indexical signs of objects such as cups to represent beliefs about spatial location.
• When an object is touched an encoding of the objects location is formed.
• Transformed into a two dimensional space corresponding to positions on the surface of the robots workspace.
• Complex actions can be represented by constructing structured schemas by using a combination of natural a-signs, natural d-signs, indexical a-signs etc.
• For example – lifting the heaviest object amongst the given ones
Intentional signs

- Used for goal driven communication.
- “what’s this fly doing in my soup” – issuing a directive to get a new bowl of soup.
- Intentional signs are physical directives that stand for something to someone
- Speech acts are assembled using lexical units
- pairs of a-signs, d-signs connected to defining schemas through projections
- A speech categorizer is implemented using Hidden Markov Models.