Multi-Modal Question-Answering: Questions without Keyboards

**Goal:** Allow players in a 3D virtual world to take virtual photographs of objects and then use the resulting photos to interact with NPCs (non-player characters).

**Interaction:** Player shows photo to NPC and clicks on an item to bring up a menu of dynamically generated queries.

Selecting a query from this menu, e.g., “What is this?” selects the set of nodes in the KB that correspond to that object.

Treelets containing the selected nodes are combined to create: “That is Mr. Jones’ house.” From which we generate the final response.

NPC KB
NPC knowledge and info required for deictic references.

Photo KB
Generic description of objects for use when the NPC’s KB is lacking specific knowledge.

Combined KB
Treelets containing the selected nodes are combined to create:

Image
Object Locator Map

Object Descriptors
Generic description for each object stored as LF:

Tobj house
Tsub this
Possr Mr._Jones
AppInfo 102

Tobj house
Tsub that
Possr Mr._Jones
AppInfo 102

Tobj name
Tsub Mr._Jones
Possr I
AppInfo 23

(Top) Scene Graph (3D Virtual World) +
Virtual Photo

Virtual Photo

Object Descriptors

Generic description for each object stored as LF:

Tobj house
Tsub this
Possr blue
AppInfo 102

Tobj house
Tsub this
Possr Mr._Jones
AppInfo 102

Tobj name
Tsub Mr._Jones
Possr I
AppInfo 23

(And many more...)