Language Understanding for Collaboration in the Real World

Professor Donna Byron
CSE 885
October 18, 2005

Who Am I?

- MBA in MIS in 1987, worked for 9 years on Financial software
- Completed PhD in Computer Science summer 2002
- Joined OSU Sept 2002
- Research area is Artificial Intelligence, specifically Natural Language Understanding
- Teaching the AI curriculum
  - 630 (AI Intro), 732 (Computational Ling)

What does a grad student in this area need?

- Interest in learning about:
  - Language and communication processes
  - Multi-modal interfaces (graphics, voice, keyboard)
  - AI design principles
  - Machine Learning
  - Data representation and compression
  - Knowledge Representation
  - Logic and Reasoning

Research Area: Natural Language Understanding

Build artifacts with linguistic competence

1. Spoken dialog systems
   - Book a flight
   - Schedule your medication for the day
   - Correct your alignment for physical therapy exercise
   - Task-specific language is easier to process
2. Computer processing of language data
   - Document translation
   - Searching for documents about a topic
   - Searching for photos based on captions
   - Use a text collection to answer questions

Why use a spoken interface:

- Hands-free interaction:
  - Impaired dexterity
  - Typophobic client base
  - Device too small for keyboard
  - Hands are busy
    - Surgeons
    - Driving
    - Daycare center
- Mobile computing
- Not tied to the desk
- Ubiquitous computing (Smart Rooms)

What is a spoken dialog system?

- Interacts with natural language
- More efficient than command-and-control language
- Not natural:
What is a spoken dialog system?

- More natural:
  - "There's another one at 10:59.
  - Is there anything earlier?"
  - "There's another one at 10:59.
  - Is there anything earlier?"

What is a spoken dialog system?

- Understands language usage conventions

  - Airlines reservation system:
    - "What can I do for you?"
    - "I need to fly to Boston on Monday"
    - (Find me a flight to Boston on Monday)
  
  - Appointment manager:
    - "What can I do for you?"
    - "I need to fly to Boston on Monday"
    - (Please help me rearrange my appointments)

What is an intelligent device

- Brings additional skills to the interaction
- Understands the task domain
- Monitor a sequence of actions to infer the user's intent
- Be proactively helpful, offer assistance as appropriate

Basic components of a spoken dialog system

- Speech in: "Could we use truck1?"
- ASR
- Word sequence
- Language Understanding
- Intelligent Agent
- Act: modify world
- Ask: isAvailable(truck1)
- Language Generation
  - Speak: Communicate with user
  - Synth: Speech
- Interpret: Language
  - Convert the observed signal into the correct meaning
  - Requires more than just hearing correctly
  - Example: "Could my Claritin cause headaches?"
  - Speech recognition: Kitabu kiko wapi?
  - You still have to know:
    - The vocabulary
    - How words in a sequence are related
    - Idioms, politeness, and other usage customs
Phases of Interpretation

1. Syntactic Analysis

2. Semantic Analysis

3. Intention Rec

YN-Question: \[ S = NP + VP \]
[Note: The image contains a diagram with a tree structure representing syntactic analysis and semantic analysis with an example of a question and its analysis steps.]

Referring Expression Forms:

Same item in the world, variety of expressions

“Karen Holbrook, president of the Ohio State University, ...”
“President Holbrook”
“Aunt Karen”
“the lady in the white suit”
“a university president” “the president” “she”

What constitutes an appropriate response from Mr. briefcase?

What constitutes an appropriate response from Mr. groceries?

Referring Expression Meanings:

Same expression, variety of referents

“Karen Holbrook, president of the Ohio State University, ...”
“the president”
“Karen”
“she”

Sense vs. Reference

- To process REs, you must answer the questions WHAT and WHICH
- What kind of thing is it? (a newspaper, a building, an amount)
  - This is understanding the sense
- Which one is it? (Koffolt, Dreese, Larkins)
  - This is understanding the reference
- To work on a task with you, the system must understand the reference
Referring Expressions

- Descriptive noun phrases, names, and pronouns
- A large source of ambiguity: “I left the paper”
  - a term paper
  - an instruction sheet
  - an individual copy of the local newspaper
  - the newspaper company
    - “I left the paper today.”
  - The building housing the newspaper company
    - “I left the paper right at 5 but the traffic was heavy”

Forms of REs

- Headed phrase (sometimes ambiguous):
  - Definite Description: “the cat with the red collar”
  - Proximate Desc: “this button”
  - Distal Desc: “that chair in the corner”
- Headless (always ambiguous):
  - Pronouns:
    - Demonstrative “this” “that” “these” “those”
    - Personal “his” “it” “them” “her”
    - Locative “here” “there”
  - One Anaphora:
    - “the one in the other room” “another one”

Contextual ambiguity

- Words + Context = meaning
- Context might include:
  - Intonation of utterance (prosody)
  - Previous part of the conversation
    - “I’m sorry Dave, I’m afraid I can’t do that”
  - Objects visually available in the setting
    - Get the door
  - Location of the conversation
    - “the president”

RE Forms and context

- Different RE forms are differentially sensitive to context:
  - Personal pronouns want a referent ‘in focus’
    - “I found the ammo clip” “Can you pick it up?”
  - Demonstrative pronouns want a referent that is less focused, but ‘in short term memory’
    - “I found the ammo clip” “That’s good”
  - Deictics are sensitive to the time/place/speaker
    - “I found the ammo clip” “Good, that goes here”

Steps to understand a reference

1. Are these words talking about:
   - a) something NEW or
   - b) something I already know about
2. If a) construct a representation
3. If b) find the thing in memory
   - Search in a collection of entities
   - Filter the collection based on properties
     - “The dog” just look for dogs
   - Choose among remaining entities based on ?

Embodied Conversational Agent

1. User and System getting information through multiple modalities
2. Multiple communication channels open between interlocutors (Language, Posture, Gesture, Gaze)
3. Speak from a specific location in 3D space
4. Agent is not omniscient regarding objects in the world
Situated Language Example

**Designer:** Where would you like the refrigerator.
**Client:** How about on that wall behind you.
**Designer:** (Turning around to look at the wall) Do you mean to the left of the window?
**Client:** No I was thinking more over here.
(looks, gestures, and takes a step toward desired spot).
**Designer:** Oh okay. Would you like a prep area underneath the overhead light?
**Client:** Yes. Can we swap the light to match that pretty one in the hallway?
(session7: 4:12)

New RE forms in situated language

New kinds of referring expressions:
- ‘the table to your left’
- ‘turn on the lights’
- ‘What’s that doing there?’
- ‘Move the table/that table from the hall in here’

Quake task setup

(Example: session7 3:18 – 3:55)
- Two human players in a quake level
- The task is a treasure hunt: one player has an instruction sheet with objectives:

Steps to understand a reference

1. Are these words talking about:
   a) something NEW or
   b) something I already know about
2. If a) construct a representation
3. If b) find the thing in memory
   - Search in a collection of entities
   - Filter the collection based on properties
     “The dog” just look for dogs
   - Choose among remaining entities based on ?

Using words to form D

“Send the engine at Avon along with a boxcar to pick up the cargo.”

Well understood: D={the engine, Avon, a boxcar, the cargo}
Using objects instead of words

“Send the engine at Avon along with a boxcar to pick up the cargo.”

Interpreted as Command:
- Event: move
- What: ENGINE1
- Dest: AVON
- With: BOXCAR?
- Why: Load: ORANGES1

Less often: D={ENGINE1, AVON, BOXCAR7, ORANGES1}

Referring to high-order referents

“Send the engine at Avon along with a boxcar to pick up the cargo.”

Interpreted as Command:
- Event: move
- What: ENGINE1
- Dest: AVON
- With: BOXCAR7
- Why: Load: ORANGES1

“That will take 3 hours.” (the loading action)
“But they’re all full.” (all boxcars in the domain)
“But they don’t go in boxcars.” (oranges in general)

Referring to high-order referents

“Send the engine at Avon along with a boxcar to pick up the cargo.”

Interpreted as Command:
- Event: move
- What: ENGINE1
- Dest: AVON
- With: BOXCAR7
- Why: Load: ORANGES1

New technology:
D={ENGINE1, AVON, BOXCAR7, ORANGES1, moving, loading, engines, cities, boxcars, cargoes, }

Mixing language and visual info

1) Door to a room opens
2) “This is a room we need to be in”.

Interpreted as Command:
- Event: move
- What: ENGINE1
- Dest: AVON
- With: BOXCAR7
- Why: Load: ORANGES1

Research Goals

- Making richer models of the Domain D
  - The visual environment
  - The larger setting (like President Holbrook)
  - Building D correctly is an open research problem
- Incorporating new information sources as filters
  - Which RE form was used?
  - Where is the speaker looking?
  - What is he pointing at?
- Making reference processing modular

Moving into realistic environments

- Understand language about the user’s surroundings
- Combine discourse and environment to form referential domain
  - What parts of visual history to keep around?
  - For how long?
  - What things in a scene are salient/topical?
- Incorporating new information sources into referent selection
  - Gesture
  - Gaze
**Working from many angles**

- Data collection: recording human conversations, building test data from text collections
- Analysis: Human experiments to understand how language works
- Design: Developing and testing new algorithms
- Implementations: Building robust, pluggable components for well-understood processing

**Human Behavior Experiments**

Setup: “Put the cup on the saucer.”
Target: “Now put it/that next to the butterfly.”

- vs -

Setup: “Put the cup next to the saucer.”
Target: “Now put it/that next to the butterfly.”

**Natural Language Processing Projects**

- Testbed for situated language understanding
  - Use Quake data to design context management
  - Quakebot hooked to language understanding software
  - Human does a task in the quake world with bot
- Building an API for reference understanding
  - Component for other researchers to use
  - Initially built in LISP, porting to Python
  - Language-independent: works for English or Korean
  - API definition will make it easier for teams to implement good reference software

**Language reduction in maps**

(w/ D. Byron & students)

- Task: automated travel assistant
- Context-aware
  - Time
  - Place
  - Usual routes
  - User preferences
- Can suggest alternative routes when traffic reports come in
  - “315 is backed up, you might want to use Fishinger instead”

**Want to learn more?**

- Join the AI mail group and come to AI presentations
- Join the gradcog mail group and go to CogSci talks
- Join the Comp Ling discussion group in Ling and attend meetings (LING 795Y)
- Take AI courses:
  - 630/730/779/694=Vision/794=Speech Rec
  - 788/888 offered by any AI faculty
  - CSE 732
- Come by and talk about your interests

**Important Courses to Take: Natural Language Track**

- Artificial Intelligence (CSE630)
- Linguistics 601, 684.01, 684.02
- Computational Linguistics (CSE732)