LUI 2.0: Next Steps in the Language User Interface
Deborah A. Dahl, Ph.D.
Conversational Technologies
Mobile Voice 2016
San Jose
Intelligent Personal Assistants

• Since 2010
• Allow users to interact with a system using spoken natural language
• Can answer many everyday questions
Examples

• Personal assistants
  – Siri, Google Now, Cortana, Amazon Echo and others
  – get general information and accomplish everyday tasks

• Enterprise Assistants
  – Nuance Nina, Openstream EVA
  – get enterprise information and accomplish enterprise tasks
With natural language, users’ questions are expressed in their terms.
How are we doing?

- Current technology is very good, but far from human
- Plenty of room for improvement
- Where to start?
Steps in Spoken Dialog Processing

- Recognize speech
- Understand language
- Manage dialog
- Do something
- Respond to the user

"I found three sushi restaurants nearby, here they are on the map."
Today’s Natural Language Technology

• Most assistants can handle simple questions –
  – What time is it?
  – Where is the nearest pizza restaurant
• What about more complex questions?
  – Let me know after 5 minutes has passed
  – Which has more calories, broccoli or carrots?
An Exploratory Natural Language Understanding Test

- 6 commercial personal assistants
- 150 English utterances
- All results are averages
Test Protocol

• Spoke each test utterance to the 6 assistants and recorded the results

• Scoring
  – Not testing speech recognition — only used correct speech recognition results
  – Just doing a web search doesn’t show understanding
  – Not testing domain knowledge — if an assistant couldn’t do something but admitted it, that was ok
# Some Test Utterances

<table>
<thead>
<tr>
<th>Easier</th>
<th>Harder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convert 5 pounds to kilograms</td>
<td>Count to 10 by twos</td>
</tr>
<tr>
<td>How many square millimeters are there in one acre?</td>
<td>Does every planet have moons?</td>
</tr>
<tr>
<td>Will it rain tomorrow?</td>
<td>Will the temperature have gone down to 50 degrees by midnight</td>
</tr>
<tr>
<td>Will I need an umbrella tomorrow?</td>
<td>It’s time for Japanese food.</td>
</tr>
<tr>
<td>Are there any Thai or Vietnamese restaurants near here?</td>
<td>Set a 1-minute timer and a 2-minute timer</td>
</tr>
<tr>
<td>What is the nearest star?</td>
<td>Is Jupiter larger than Saturn?</td>
</tr>
<tr>
<td>...and what about tomorrow?</td>
<td>What is the largest planet in the solar system and what's its diameter?</td>
</tr>
<tr>
<td>It won’t rain tomorrow, will it?</td>
<td>Aside from Pluto, what dwarf planets are there?</td>
</tr>
</tbody>
</table>
Test 1: Performance on a variety of domains

- Time and date
- Everyday knowledge
- Weather
- Local business search
- Reminders
- Arithmetic
- Timers
- Out of domain
- Some domains not tested – sports, news, movies, music, traffic
Results for Domains

overall percent correct by domain

- weather
- time and date
- arithmetic
- local business search
- everyday knowledge
- reminders
- timers and alarms
Test 2: Basic and complex utterances

Two kinds of utterances

• basic
  “what time is it”
  “is this Friday”

• complex
  – “Which has more calories, broccoli or an apple?”
  – “Let me know after 5 minutes has passed”
What makes the language of an utterance “complex”?

(https://xkcd.com)
Some aspects of complexity

Omitted information:
Pronouns (anaphora) and ellipsis

Comparisons

Indirectness:
Implicature

Beyond the here and now:
Time expressions, possibility, probability, contingency, negatives

Out of domain

Talking about more than one thing:
Conjunction and plurals, recurring time
Results for complexity

percent correctly answered with one try (avg. of 6 systems)
Overall

• 61% correct for “basic”
• 29% correct for “complex”
Lots of room for improvement – one example

Negation

• Aside from Pluto, what dwarf planets are there?
• Which planets don't have moons?
• Are there any non Japanese Asian restaurants near here?
Other observations

- Speech recognition was usually correct
- No assistant engaged in real dialog, usually just one follow-up question possible
- Need dialog processing to handle pronouns
- Assistants often don’t know what they don’t know
Do complex constructions matter?
Is covering the “basic” constructions good enough?

- People don’t mind repeating and rephrasing questions.
- But you’re telling people that they can say anything they want.
- People will lose patience with repeating questions.
How can systems improve their natural language processing?
Some strategies

• Look at the utterances *before* the successful ones

• Look for conjunctions, comparatives, and negative words

• Develop better context representations for handling pronouns and ellipsis

• Look at techniques beyond intent recognition and slot filling
Conclusions

• Natural language coverage is very good across different virtual personal assistants
• There is still room for improvement!