Speech and Language Processing

Chapter 24: part 1
Dialog(ue) and Conversational Agents

Outline
- Introduction and History
- The Linguistics of Conversation
Conversational Agents

- **AKA:**
  - (Spoken) Dialogue Systems
  - SDS
  - Intelligent Personal Assistants
  - Chatbots

Applications

- Information providing
  - Amtrak’s Julie, CMU Let’s Go and Skylar
- Customer Care
  - T-Mobile’s Call Center, AT&T Call Routing
- Training
  - Research platforms: e.g. Pitt ItSpoke
- Siri, Cortana
- Amazon Alexa, Google Home
- Communicating with robots
- Anything with limited screen/keyboard (mobile apps)
- Fun and games....

- Goal: Emulate Human-Human Behavior?
A travel dialog: Communicator

S1: Hello. You've reached the [Communicator]. Tell me your full name
U2: Hi I'd like to fly to Seattle Tuesday morning
S3: Traveling to Seattle on Tuesday, August eleventh in the morning. Your full name?
U4: John Doe
S5: Hello, John, I'm glad to hear from you again. You can now ask me for help.
S6: OK. Let's see I have a United flight departing Pittsburgh at ten forty-five a.m.
   arriving into Seattle at one fifty-seven p.m. Is that OK?
U7: Yes
S8: Will you return to Pittsburgh from Seattle?
U9: No, I will go to Los Angeles.
S10: Traveling to Los Angeles. On what date
U11: Sorry, can we go back to talk about the previous leg again.
S12: For the first leg, you are on a United flight 1446 departing Pittsburgh at ten forty-five a.m.
   tomorrow, arriving in Seattle at one fifty-seven p.m. Do you want to take that?
U13: Actually I want to leave on Wednesday.
S14: A flight on Wednesday.
S15: I have a flight departing Pittsburgh at ten forty-five a.m., arrives Seattle at one fifty-seven p.m. Is that OK?
U16: Yes.

Figure 19.1 The travel domain: a fragment from a successful conversation between a user (U) and the Communicator system (S) of Xu and Rudnicky (2000).

Call routing: ATT HMIHY

S: How may I help you?
U: Can you tell me how much it is to Tokyo?
S: You want to know the cost of a call?
U: Yes, that's right.
S: Please hold on for rate information.

S: How may I help you?
U: What's the area code for Chicago?
S: You want an area code.
U: (silence)
S: The area code for Chicago is 312.

Figure 19.2 The call routing/telephone information domain: two actual conversations between a user (U) and the AT&T How May I Help You (HMIHY) system (S) (Gorin et al., 1997).
A tutorial dialogue: ITSPoke

S: I didn’t understand. What force acts on all objects within close proximity to earth?
U: *the force of gravity*
S: Fine. Besides the Earth’s gravitational force, are there any other forces acting on the person, keys, and elevator after the release?
U: *no*
S: Yep. Thus, keep in mind that after the release, the only force on the person, keys, and elevator is the force of gravity. If gravity is the only force acting on an object, will it be moving or staying still?
U: *it will be moving*
S: Fine. Not only are the person, keys, and elevator moving, they have only gravitational forces on them. When an object is falling and has only gravitational force on it, it is said to be in what?
U: *freefall*

Figure 19.3 The tutoring domain: part of an actual conversation between a student user (U) and the ITSPoke system (S) of (Litman and Stillman, 2004), based on the Why2-Atlas text-based tutoring system (?).

Dialogue Systems: A Brief History

<table>
<thead>
<tr>
<th>Computer Prototypes</th>
<th>Interactive Telephone</th>
<th>Devices (e.g. smartphones)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>1970</td>
<td>1980</td>
</tr>
<tr>
<td>1990</td>
<td>2000</td>
<td>2010</td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ELIZA
(Chatbots)

*Men are all alike.*
*IN WHAT WAY*
*They’re always bugging us about something or other.*
*CAN YOU THINK OF A SPECIFIC EXAMPLE*

[Weizenbaum, 1966]

---

SHRDLU
(Artificial Intelligence)

*Pick up a big red block.*
*OK*
*Grasp the pyramid.*
*I DON’T UNDERSTAND WHICH PYRAMID YOU MEAN*

[Winograd, 1971]
Dialogue Systems: A Brief History

ELIZA
(Chatbots)

SHRDLU
(Artificial Intelligence)

Allen, Cohen, & Perrault
(Plan-based)

How many hotels are there in Cambridge.
I KNOW OF SIX. HOTELS IN CAMBRIDGE
[Glass et al., 1995]
### Dialogue Systems: A Brief History

**ELIZA**  
(Chabotns)

**SHRDLU**  
(Artificial Intelligence)

**SIRI**  
(hybrid approach)

**VODIS, VOYAGER**  
(Speech)

### Recent System Landscape

[a] Cortana: Microsoft’s new personal assistant for Windows Phone 8.1.

[b]udi: In-car multimedia and navigation system from Audi.

[c] Ada and Grace, the virtual guides at the Museum of Science in Boston, developed by the USC Institute for Creative Technologies [7].

[d] Hildie, an interactive scheduling assistant [3].

[e] Nao, the interactive robot created by Aldebaran Robotics.

[Lison and Meena, 2014]
Are we done yet?

The Future Directions of Dialogue-Based Intelligent Personal Assistants (SIGdial 2016 panel)

- Yes (one of the panelists)
  - no bottleneck but privacy!

- No (audience members - especially senior NLPers)
  - pragmatics, social dimensions, ...

Linguistics of Human Conversation

- Turn-taking
- Speech Acts
- Grounding
- Conversational Structure
- Implicature
Turn-taking

- Dialogue is characterized by turn-taking.
  - A:
  - B:
  - A:
  - B:
  - ...
- Resource allocation problem:
- How do speakers know when to take the floor?
  - Total amount of overlap relatively small (5% - Levinson 1983)
  - Don’t pause either
  - Must be a way to know who should talk and when.

Turn-taking rules

- At each transition-relevance place of each turn:
  - a. If during this turn the current speaker has selected B as the next speaker then B must speak next.
  - b. If the current speaker does not select the next speaker, any other speaker may take the next turn.
  - c. If no one else takes the next turn, the current speaker may take the next turn.
Implications of subrule a

- For some utterances the current speaker selects the next speaker
  - Adjacency pairs
    - Question/answer
    - Greeting/greeting
    - Compliment/downplayer
    - Request/grant
  - Silence between 2 parts of adjacency pair is different than silence after
    - A: Is there something bothering you or not?
    - (1.0)
    - A: Yes or no?
    - (1.5)
    - A: Eh
    - B: No.

Turn-taking Behaviors
Important for SDS

- System understanding:
  - Is the user backchanneling or is she taking the turn (does ‘ok’ mean ‘I agree’ or ‘I’m listening’)?
  - Is this a good place for a system backchannel?

- System generation:
  - How to signal to the user that the system system’s turn is over?
  - How to signal to the user that a backchannel might be appropriate?
Types of Behavior

- **Smooth Switch**: S1 is speaking and S2 speaks and takes and holds the floor
- **Hold**: S1 is speaking, pausing, and continuing to speak
- **Backchannel**: S1 is speaking and S2 speaks -- to indicate continued attention -- not to take the floor (e.g. *mhmm, ok, yeah*)
- How do people coordinate these behaviors with their interlocutor?
- Acoustic-prosodic and lexical cues....

Speech Acts

- **Austin (1962)**: An utterance is a kind of action
- **Clear case**: *performatives*
  - I name this ship the Titanic
  - I second that motion
  - I bet you five dollars it will snow tomorrow
- **Performative verbs** *(name, second)*
- Austin’s idea: not just these verbs
Each utterance is 3 acts

- Locutionary act: the utterance of a sentence with a particular meaning
- Illocutionary act: the act of asking, answering, promising, etc., in uttering a sentence.
- Perlocutionary act: the (often intentional) production of certain effects upon the thoughts, feelings, or actions of addressee in uttering a sentence.

Locutionary and illocutionary

- “You can’t do that!”
- Illocutionary force:
  - Protesting
- Perlocutionary force:
  - Intent to annoy addressee
  - Intent to stop addressee from doing something
## The 3 levels of act revisited

<table>
<thead>
<tr>
<th></th>
<th>Locutionary Force</th>
<th>Illocutionary Force</th>
<th>Perlocutionary Force</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can I have the rest</td>
<td>Question</td>
<td>Request</td>
<td>Intent: You give me</td>
</tr>
<tr>
<td>of your sandwich?</td>
<td></td>
<td></td>
<td>sandwich</td>
</tr>
<tr>
<td>I want the rest of</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>your sandwich</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Give me your</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sandwich!</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### The 3 levels of act revisited

<table>
<thead>
<tr>
<th>Locutionary Force</th>
<th>Illocutionary Force</th>
<th>Perlocutionary Force</th>
</tr>
</thead>
<tbody>
<tr>
<td>Can I have the rest of your sandwich?</td>
<td>Question</td>
<td>Request</td>
</tr>
<tr>
<td>I want the rest of your sandwich</td>
<td>Declarative</td>
<td>Request</td>
</tr>
<tr>
<td>Give me your sandwich!</td>
<td>Imperative</td>
<td>Request</td>
</tr>
</tbody>
</table>

### 5 classes of speech acts: Searle (1975)

- **Assertives**: committing the speaker to something’s being the case
  - suggesting, putting forward, swearing, boasting, concluding
- **Directives**: attempts by the speaker to get the addressee to do something
  - asking, ordering, requesting, inviting, advising, begging
- **Commissives**: committing the speaker to some future course of action
  - promising, planning, vowing, betting, opposing
- **Expressives**: expressing the psychological state of the speaker about a state of affairs
  - thanking, apologizing, welcoming, deploring
- **Declarations**: bringing about a different state of the world via the utterance
  - I resign; You’re fired)
Dialogue Acts in SDS

- Roughly correspond to Illocutionary acts
  - Motivation: Improving Spoken Dialogue Systems
  - Many coding schemes (e.g. DAMSL)
  - Many-to-many mapping between DAs and words
    - Agreement DA can be realized by Okay, Um, Right, Yeah, ...
    - But each of these can express multiple DAs, e.g.

  S: You should take the 10pm flight.
  U: Okay
      ...
      ...that sounds perfect.
      ...
      ...but I’d prefer an earlier flight.
      ...(I’m listening)

- DA recognition important for
  - Turn recognition (which grammar to use when)
  - Turn disambiguation, e.g.

  S: What city do you want to go to?
  U1: Boston. (reply)
  U2: Boston? (request for information)
  S: Do you want to go to Boston?
  U1: Boston. (confirmation)
  U2: Boston? (question)
Automatic DA Detection

- Can we detect DAs automatically?
  - Lexicons are domain-dependent
  - ASR output is errorful
- Typical methods
  - Supervised learning
  - Sequence learning / tagging

Back to 1980: Analyzing Intention in Utterances (Allen & Perrault)

- Cooperative behavior in dialogue
  - Indirect speech acts, helpful responses, ...

- A knowledge-based approach
  - Plan construction & execution (agent A)
  - Plan inference & obstacle detection (Agent B)
  - Plan construction & execution (agent B)
What about Siri?

Indirect Speech Acts (beyond literal meaning)

Me: I don’t know Megan’s address
Siri: Don’t worry about it, Diane

Me: Are you able to tell me Megan’s address
Siri: We were talking about you, not me

Me: Can you tell me Megan’s address
(Do you know Megan’s address)
Siri: Here’s the address for Megan ...

-----------------------------
Me: Tell me Megan’s address
Siri: Here’s the address for Megan ...
Helpful Responses
(more information than requested)

Me: Is Steve Jobs the chairman of Apple
Siri: Thanks for asking about Steve. Here’s Apple’s web page about him

Me: What day of the week is election day
Siri: Election Day is on Tuesday, November 8 2016

Me: When can I watch the World Series
Siri: The Indians face the Cubs in game 3 of the World Series tomorrow at 8:08 PM

Indirect + (Mixed) Helpful

Me: I’m hungry
Siri: I don’t want you feeling all peckish. Here’s what I found.

15 results displayed
Filtered to be nearby
... but some of the 15 were closed
Back to the Future? From Siri to Viv


Soapbox

- While cooperative dialogue research based on speech acts was largely abandoned for more practical approaches (my own research included!)
  - "There is much to be gained from recognizing not just what was said, but why; from identifying conclusions people naturally draw from both what has been said and hasn't" (2016 Workshop CFP)

- Maybe not such an upbill battle anymore?
Grounding

- Dialogue is a collective act performed by speaker and hearer
- Common ground: set of things mutually believed by both speaker and hearer
- Need to achieve common ground, so hearer must ground or acknowledge speakers utterance.
- Clark (1996):
  - Principle of closure. Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it
  - Need to know whether an action succeeded or failed

Clark and Schaefer: Grounding

- Continued attention: B continues attending to A
- Relevant next contribution: B starts in on next relevant contribution
- Acknowledgement: B nods or says continuer like *uh-huh, yeah, assessment (great!)*
- Demonstration: B demonstrates understanding A by paraphrasing or reformulating A’s contribution, or by collaboratively completing A’s utterance
- Display: B displays verbatim all or part of A’s presentation
Grounding Strategies from Weak to Strong

I need to get your homework by tonight.

- Continued attention
  ...
- Next contribution
  I should be finished after class.
- Acknowledgment
  Mhmm...
- Demonstration
  You need this soon.
- Display
  You need to get my homework by tonight.

A human-human conversation (find display, acknowledgement)

| C1 | ...I need to travel in May. |
| A1 | And, what day in May did you want to travel? |
| C2 | OK uh I need to be there for a meeting that’s from the 12th to the 15th. |
| A2 | And you’re flying into what city? |
| C3 | Seattle. |
| A3 | And what time would you like to leave Pittsburgh? |
| C4 | Uh hmm I don’t think there’s many options for non-stop. |
| A4 | Right. There’s three non-stops today. |
| C5 | What are they? |
| A5 | The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm. |
| C6 | OK I’ll take the 5ish flight on the night before on the 11th. |
| C7 | OK. |
Grounding examples

- Display:
  - C: I need to travel in May
  - A: And, what day in May did you want to travel?

- Acknowledgement
  - C: He wants to fly from Boston
  - A: mm-hmm
  - C: to Baltimore Washington International
  - [Mm-hmm (often transcribed “uh-huh”) is a backchannel, continuer, or acknowledgement token]

Grounding Examples (2)

- Acknowledgement + next relevant contribution
  - And, what day in May did you want to travel?
  - And you’re flying into what city?
  - And what time would you like to leave?

- The and indicates to the client that agent has successfully understood answer to the last question.
Grounding and Dialogue Systems

- Grounding is not just a tidbit about humans
- Is key to design of conversational agent
- Why?
  - HCI researchers find users of speech-based interfaces are confused when system doesn’t give them an explicit acknowledgement signal
Discourse Structure

- Dialogue variants of discourse chapter

Pragmatics

- Context-dependent meaning, invited inference, intended meaning – vs. “propositional content”
- Indirect Speech Acts
- Implicature
Conversational Implicature

- A: And, what day in May did you want to travel?
- C: OK, uh, I need to be there for a meeting that's from the 12th to the 15th.

- Note that client did not answer question.
- Meaning of client's sentence:
  - Meeting
    - Start-of-meeting: 12th
    - End-of-meeting: 15th
  - Doesn't say anything about flying!!!!!
- What is it that licenses agent to infer that client is mentioning this meeting so as to inform the agent of the travel dates?

Conversational Implicature (2)

- A: ... there's 3 non-stops today.
- This would still be true if 7 non-stops today.
- But no, the agent means: 3 and only 3.
- How can client infer that agent means:
  - only 3
Grice: conversational implicature

- Implicature means a particular class of licensed inferences.
- Grice (1975) proposed that what enables hearers to draw correct inferences is:
- Cooperative Principle
  - This is a tacit agreement by speakers and listeners to cooperate in communication

4 Gricean Maxims

- Relevance: Be relevant
- Quantity: Do not make your contribution more or less informative than required
- Quality: try to make your contribution one that is true (don’t say things that are false or for which you lack adequate evidence)
- Manner: Avoid ambiguity and obscurity; be brief and orderly
Relevance

- A: Is Regina here?
- B: Her car is outside.
- Implication: yes
  - Hearer thinks: why would he mention the car? It must be relevant. How could it be relevant? It could since if her car is here she is probably here.
- Client: I need to be there for a meeting that's from the 12th to the 15th
  - Hearer thinks: Speaker is following maxims, would only have mentioned meeting if it was relevant. How could meeting be relevant? If client meant me to understand that he had to depart in time for the mtg.

Quantity

- A: How much money do you have on you?
- B: I have 5 dollars
  - Implication: not 6 dollars
- Similarly, 3 non stops can't mean 7 non-stops as hearer thinks:
  - if speaker meant 7 non-stops she would have said 7 non-stops
- A: Did you do the reading for today's class?
- B: I intended to
  - Implication: No
  - B's answer would be true if B intended to do the reading AND did the reading, but would then violate maxim