Introduction to Dialogue

Caroline Sporleder

Universität des Saarlandes

Sommersemester 2009

15.07.2009
S1: Hello. You’ve reached the [...] Tell me your full name.
U2: hi I’d like to fly to seattle tuesday morning
S3: Travelling 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
S: I didn’t understand that. 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
Dialogues consist of **turns**. Speakers are able to negotiate turns quite efficiently (few overlaps). Turn changes happen at (para-)linguistically signalled **transition relevance places**. ⇒ people can figure out quickly who should talk next and when they should start.

**Turn Taking Rules** (Sacks et al. 1974)

1. If during this turn the current speaker has selected A as the next speaker then A must speak next.
2. If the current speaker does not select the next speaker, any other speaker may take the next turn.
3. If no one else takes the next turn, the current speaker may take it.
Dialogue often consists of adjacency pairs

- question-answer
- greeting-greeting
- request-grant
- compliment-downplayer

Silence between the two parts of an adjacency pair is significant ⇒ interpreted as a refusal to respond or a dispreferred response (e.g. failure to grant a request)

Example

A: Is there something bothering you or not?  
   (1.0)  
A: Yes or no?  
   (1.5)  
A: Eh?  
B: No.
Speaker and hearer need to establish **common ground**. Hearer must signal to speaker that they have understood; hearer has to **ground** speaker’s utterance.

**Principle of closure** (Clark 1996)  
Agents performing an action require evidence, sufficient for current purposes, that they have succeeded in performing it.

**Closure can be achieved by:**

- continued attention
- next contribution
- acknowledgment (e.g., nodding, *uh-huh*)
- demonstration (e.g., paraphrasing)
- display (e.g., verbatim repetition)
Example (J&M, p. 853)

C: . . . I need to travel in May.
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.
A: And you’re flying into what city?

Example (J&M, p. 853)

C: He wants to fly from Boston to Baltimore.
A: Uh-huh

Utterances like *uh-huh* are call backchannel.
Example (J&M, p. 853)

C: . . . I need to travel in May.
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.
A: And you’re flying into what city?

Example (J&M, p. 853)

C: He wants to fly from Boston to Baltimore.
A: Uh-huh

Utterances like *uh-huh* are call backchannel.
Example (J&M, p. 853)
C: . . . I need to travel in May.
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.
A: And you’re flying into what city?

Example (J&M, p. 853)
C: He wants to fly from Boston to Baltimore.
A: Uh-huh

Utterances like *uh-huh* are call backchannel.
Example (J&M, p. 853)
C: ... I need to travel in May.
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.
A: And you’re flying into what city?

Example (J&M, p. 853)
C: He wants to fly from Boston to Baltimore.
A: Uh-huh

Utterances like *uh-huh* are call backchannel.
Dialogue structure is often partly conventionalised, i.e. there is a certain conversational structure to which the participants adhere. Deviation from this structure can lead to confusion.
Example: Opening Telephone Conversations

Stage 1: enter a conversation, with summons-response adjacency pair
Stage 2: identify speakers
Stage 3: establish joint willingness to converse
Stage 4: raise the first topic (usually done by the caller)
Example: Opening Telephone Conversations

**Stage 1:** enter a conversation, with summons-response adjacency pair

**Stage 2:** identify speakers

**Stage 3:** establish joint willingness to converse

**Stage 4:** raise the first topic (usually done by the caller)

---

**Example (J&M 2009, p. 855, from Clark 1994)**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Speaker and Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A1: (rings B’s telephone)</td>
</tr>
<tr>
<td>1,2</td>
<td>B2: Benjamin Holloway</td>
</tr>
<tr>
<td>2</td>
<td>A3: This is Prof. Dwight’s secretary.</td>
</tr>
<tr>
<td>2,3</td>
<td>B4: ooh, yes ——</td>
</tr>
<tr>
<td>4</td>
<td>A5: uh:m, about the lexicology seminar</td>
</tr>
<tr>
<td>4</td>
<td>B6: yes</td>
</tr>
</tbody>
</table>
Example: Opening Telephone Conversations

Stage 1: enter a conversation, with summons-response adjacency pair
Stage 2: identify speakers
Stage 3: establish joint willingness to converse
Stage 4: raise the first topic (usually done by the caller)

Example (J&M 2009, p. 855, from Clark 1994)

C1: (rings)
O2: Directory Enquiries, for which town please?
C3: Could you give me the phone number of Mrs. Smithson?
O4: Yes, which town is this at please?
C5: Huddleston.
O6: Yes, and the name again?
C7: Mrs. Smithson.
**Stage 1:** enter a conversation, with summons-response adjacency pair

**Stage 2:** identify speakers

**Stage 3:** establish joint willingness to converse

**Stage 4:** raise the first topic (usually done by the caller)

---

**Example (J&M 2009, p. 855, from Clark 1994)**

C1: (rings)

O2: Directory Enquiries, **for which town please?**

C3: Could you give me the phone number of Mrs. Smithson?

O4: Yes, which town is this at please?

C5: Huddleston.

O6: Yes, and the name again?

C7: Mrs. Smithson.
Stage 1: enter a conversation, with summons-response adjacency pair
Stage 2: identify speakers
Stage 3: establish joint willingness to converse
Stage 4: raise the first topic (usually done by the caller)

Example (J&M 2009, p. 855, from Clark 1994)

C1: (rings)
O2: Directory Enquiries, for which town please?
C3: Could you give me the phone number of Mrs. Smithson?
O4: Yes, which town is this at please?
C5: Huddleston.
O6: Yes, and the name again?
C7: Mrs. Smithson.
Example: Opening Telephone Conversations

Stage 1: enter a conversation, with summons-response adjacency pair
Stage 2: identify speakers
Stage 3: establish joint willingness to converse
Stage 4: raise the first topic (usually done by the caller)

Example (J&M 2009, p. 855, from Clark 1994)

C1: (rings)
O2: Directory Enquiries, for which town please?
C3: Could you give me the phone number of Mrs. Smithson?
O4: Yes, which town is this at please?
C5: Huddleston.
O6: Yes, and the name again?
C7: Mrs. Smithson.
**Stage 1:** enter a conversation, with summons-response adjacency pair

**Stage 2:** identify speakers

**Stage 3:** establish joint willingness to converse

**Stage 4:** raise the first topic (usually done by the caller)

---

Example (J&M 2009, p. 855, from Clark 1994)

C1: (rings)
O2: Directory Enquiries, for which town please?
C3: Could you give me the phone number of Mrs. Smithson?
O4: Yes, which town is this at please?
C5: Huddleston.
O6: Yes, and the name again?
C7: Mrs. Smithson.

⇒ phone-based dialogue systems often use **open** rather than directive **prompts** (e.g., “How may I help you?” vs. “How may I direct your call?”)
Architecture of Spoken Dialogue Systems (J&M 2009)

Speech Recognition -> Natural Language Understanding

Dialogue Manager

Text-to-Speech Synthesis -> Natural Language Generation

Task Manager
Architecture of Spoken Dialogue Systems (J&M 2009)

Speech Recognition → Natural Language Understanding → Dialogue Manager → Task Manager

Text-to-Speech Synthesis → Natural Language Generation
Architecture of Spoken Dialogue Systems (J&M 2009)

- Speech Recognition
- Natural Language Understanding
- Dialogue Manager
- Text-to-Speech Synthesis
- Natural Language Generation
- Task Manager

Caroline Sporleder csporled@coli.uni-sb.de

Computational Models of Discourse
Architecture of Spoken Dialogue Systems (J&M 2009)

Speech Recognition → Natural Language Understanding → Dialogue Manager → Task Manager

Text-to-Speech Synthesis → Natural Language Generation → Dialogue Manager
Architecture of Spoken Dialogue Systems (J&M 2009)

- Speech Recognition
- Natural Language Understanding
- Dialogue Manager
- Task Manager
- Natural Language Generation
- Text-to-Speech Synthesis
Architecture of Spoken Dialogue Systems (J&M 2009)

- Speech Recognition
- Natural Language Understanding
- Dialogue Manager
- Text-to-Speech Synthesis
- Natural Language Generation
- Task Manager

Caroline Sporleder csporled@coli.uni-sb.de

Computational Models of Discourse
Example: Frame and Slot Semantics

Show me morning flights from Boston to San Francisco on Tuesday.

SHOW:
FLIGHTS:
  ORIGIN:
    CITY: Boston
  DATE:
    DAY-OF-WEEK: Tuesday
  TIME:
    PART-OF-DAY: morning
DESTINATION:
  CITY: San Francisco
Example: Frame and Slot Semantics

Show me *morning* flights from *Boston* to *San Francisco* on *Tuesday*.

SHOW:
FLIGHTS:
  ORIGIN:
    CITY: Boston
    DATE:
      DAY-OF-WEEK: Tuesday
    TIME:
      PART-OF-DAY: morning
  DESTINATION:
    CITY: San Francisco
Filling the frame slots:
manually built context free semantic grammars

<table>
<thead>
<tr>
<th>Example Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOW</td>
</tr>
<tr>
<td>→ show me</td>
</tr>
<tr>
<td>DEPART–TIME–RANGE</td>
</tr>
<tr>
<td>→ (after</td>
</tr>
<tr>
<td>HOUR</td>
</tr>
<tr>
<td>→ one</td>
</tr>
<tr>
<td>FLIGHTS</td>
</tr>
<tr>
<td>→ (a) flight</td>
</tr>
<tr>
<td>AMPM</td>
</tr>
<tr>
<td>→ am</td>
</tr>
<tr>
<td>ORIGIN</td>
</tr>
<tr>
<td>→ from CITY</td>
</tr>
<tr>
<td>DESTINATION</td>
</tr>
<tr>
<td>→ to CITY</td>
</tr>
<tr>
<td>CITY</td>
</tr>
<tr>
<td>→ Boston</td>
</tr>
</tbody>
</table>
Filling the frame slots:

stochastic grammars (e.g., Hidden Markov Models)
Architecture of Spoken Dialogue Systems (J&M 2009)

- Speech Recognition
- Natural Language Understanding
- Text-to-Speech Synthesis
- Natural Language Generation
- Dialogue Manager
- Task Manager

Caroline Sporleder csporled@coli.uni-sb.de
Computational Models of Discourse
content planning (what to say)
generation (how to say it)
1. content planning (what to say) → dialogue manager
2. generation (how to say it)
1. content planning (what to say) → dialogue manager
2. generation (how to say it)
   - template-based generation (templates with canned text)
   - deep generation
Template-based Generation

- pre-specified, manually-written text templates for each stage of the dialogue (e.g., “What time do you want to leave ORIGIN?”)

Deep Generation

- sentence planner
- surface ralizer
- prosody assigner
Design issues

- text should sound coherent and natural (e.g. use of cue phrases)
Design issues

- text should sound coherent and natural (e.g. use of cue phrases)

Example: Unnatural Text

*Please tell me where you want to fly.*

*...*

*Please tell me on what date you want to fly.*

*...*

*Please tell me your preferred departure time.*
Design issues

- text should sound coherent and natural (e.g. use of cue phrases)

Example: Better Text

*First, tell me where you want to fly.*

...  

*Next, I’ll need to know on what date you would like to fly.*

...  

*Thanks. And what is your preferred departure time?*
Design issues

- text should sound coherent and natural (e.g. use of cue phrases)
- repeated prompts should get shorter

Example: Shortening of prompts

S: Now, what’s the first company to add to your watch list?
U: Cisco
S: What’s the next company?
U: IBM
S: Next one?
U: Intel
...
Design issues

- text should sound coherent and natural (e.g. use of cue phrases)
- repeated prompts should get shorter
- take into account memory load of user

Example: Reduce memory load

U: I want to fly from Boston to San Francisco next Thursday morning.
S: There are 10 flights from Boston to San Francisco next Thursday. The earliest is with United and leaves at 6:55 am and arrives at 2:17 pm their time. Would you like information about the later flights?
Architecture of Spoken Dialogue Systems (J&M 2009)

- Speech Recognition
- Natural Language Understanding
- Dialogue Manager
- Task Manager
- Text-to-Speech Synthesis
- Natural Language Generation
Single Initiative

- system dictates the flow of the conversation
- modelled as manually built finite state automaton
- advantage: system knows what to expect at each stage, which makes speech recognition and language understanding easier
- disadvantage: lack of flexibility
Example: Single Initiative System

- What city are you leaving from?
- Where are you going?
- What date do you want to leave?
- Is it a one-way trip?
- Do you want to go from <FROM> to <TO> on <DATE>?
- What date do you want to return?
- Do you want to go from <FROM> to <TO> on <DATE> returning on <RETURN>?
- Book the flight

Yes or No?
Mixed Initiative

- initiative shifts between user and system
- dialogue aim is to fill the slots in the frame (e.g. destination, departure date etc.) but system leaves more control to user by employing open prompts (e.g., “How may I help you?”)
Mixed Initiative

- initiative shifts between user and system
- dialogue aim is to fill the slots in the frame (e.g. destination, departure date etc.) but system leaves more control to user by employing open prompts (e.g., “How may I help you?”)

Note: systems can also be designed to switch between single and mixed initiative