Dialogue Models and Dialogue Systems

Prof. Marilyn Walker
www.dcs.shef.ac.uk/~walker
Joe Polifroni
j.polifroni@dcs.shef.ac.uk
Francois Mairesse
f.mairesse@sheffield.ac.uk

Class Organization

- Essay 90% (topics available in two weeks)
- In class 'thought pieces' (on each topic) 10%
- Webpage:

http://www.dcs.shef.ac.uk/~francois/dialog-art/

Topics to be covered

- Architecture of spoken dialogue systems
- Evaluation of spoken dialogue systems
- User modelling
- Spoken Language Generation
- Learning

Today's topics

- Types of Spoken Dialogue Systems
- Architecture of SDS
- Components of SDS
- DM in context of SDS

Spoken Dialogue Systems

- Intelligent agent interacting with humans by voice to complete a variety of tasks
- Many deployed systems
- Can understand what people say
- Sounds human when responds
- Can pass the Turing test

Listen again …

- Open-ended prompt
- Multiple requests in one utterance
- Confirmation subdialogue
- Reprompting
- Remembering user goal across confirmation subdialogue
- Rapid speech
- Slightly odd synthesis
- Implicit, then explicit confirmation
- Multiple responses
- Politeness behavior

Types of dialogue systems

- Chatbots
  - Seek to emulate human-human behavior
  - Aim to pass the Turing Test
- Tutorial
  - Goal: instruct a user
  - Topics:
    - Language learning
    - Car repair
    - Algebra

Types of dialogue systems (cont'd.)

- Task-oriented
  - Process based
    - Transfer money in bank accounts
    - Pay bill with service provider
  - Information based
    - Book a flight
    - Find a restaurant
    - Find directions

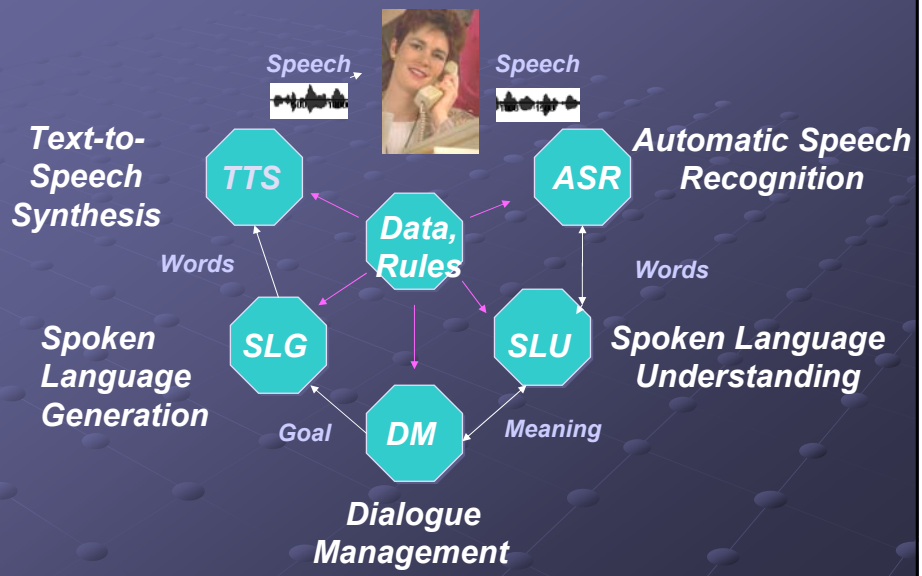
What each type of system is trying to model

- Chat
  - Common sense/human knowledge/politeness behavior
- Tutorial
  - Underlying process/step-by-step requirements/pedagogical theory
- Task-oriented
  - Task requirements
    - Steps required to achieve goal
    - Data needed to achieve goal

Output considerations

- Chat:
  - Formal/informal language
  - Friendliness
  - Human-like speech (including hesitations/false starts?)
- Tutorial
  - Clarity
  - Step-wise presentation of concepts
- Task-oriented
  - Clarity of questions
  - Verbosity

Spoken Dialogue Systems



Audio server

- Purpose: data capture
- Input: speech; Output: digitized version of speech
- Considerations:
  - Availability
  - Bandwidth
  - Drop-out
  - Barge-in

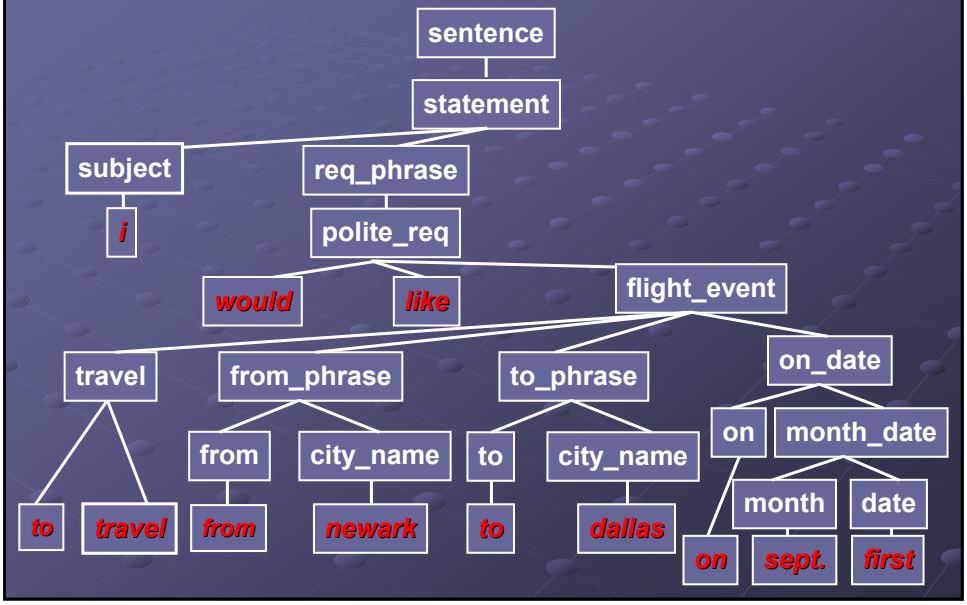
Automatic Speech Recognition

- Purpose: transcribe the speech
- Input: digital speech
- Output: String/N-best list representing hypothesized words
- Considerations:
  - Vocabulary size
  - Grammar type
  - Speech type
    - Isolated word/continuous speech
    - Spontaneous speech/read speech
    - Accented speech

Natural Language Understanding

- Purpose: produce meaning representation from ASR output
- Input: String/N-best list
- Output: Meaning representation
- Considerations:
  - Type of grammar
    - Finite-state
    - Full parse
    - Word-spotting

SLU: Example Full Parse



SLU: Word-Spotting Output

- *I would like to fly from Newark to Dallas on September first*
  - I would like to fly from <city_name> *Newark* </city_name> to <city_name> *Dallas* </city_name> on September first
  - I would like to fly <from_place> *from* <city_name> Newark </city_name> </from_place> <to_place> *to* <city_name> Dallas </city_name> </to_place> on September first
  - I would like to fly <from_place> from <city_name> Newark </city_name> </from_place> <to_place> to <city_name> Dallas </city_name> </to_place> on <date> *September first* </date>
  - <req_flight> *I would like to fly* </req_flight> <from_place> from <city_name> Newark </city_name> </from_place> <to_place> to <city_name> Dallas </city_name> </to_place> on <date> September first </date>

SLU: Example Output

Request

Flight_event

From: *EWR*

To: *DFW*

Date: *01Mar2006*

Flight_event

From: *EWR*

To: *DFW*

Date: *01Mar2006*

Dialogue Management

- Purpose: decide what system's next action should be.
- Input: a meaning representation from SLU

Request_phrase
Flight_event
From: *EWR*
To: *DFW*
Date: *01Sep2005*

- Output: High-level communicative goal(s)
  - *Confirm-info-sofar, Get-next-info*

Natural Language Generation

- Purpose: produce an output string to be shown/spoken to the user
- Input: Representation from DM
- Output: String for TTS (possibly marked for prosody, etc.)
- Considerations:
  - Verbosity
  - Level of formalism
  - "Elegance"

Text-to-Speech Synthesis

- Purpose: speak string to user
- Considerations:
  - Human-like
  - Flexibility
- Formant-based synthesis (Stephen Hawking)
  - Human vocal tract modelled and speech truly "synthesized"
- Concatenative Synthesis
  - Bits of human speech glued together (concatenated)
  - More natural but inherently limited: can only produce what you have units for in the database)

How are all these pieces put together?

- One possibility: Dialogue Manager
- Reason: Dialogue Manager is in the "middle" in time and functionality
  - DM comes after ASR and NLU
  - DM precedes NLG and TTS
  - DM is often only component that has access to database (and reasoning about database)

Dialogue Management and Audio Server

- If barge-in is enabled, how is truncated input interpreted:

User: I'm interested in Thai restaurants in North London.
System: *I know of 8 Thai rest-*
User: Wait, that's not what I wanted.

User: I'm interested in Thai restaurants in North London.
System: *I know of 8 Thai restaurants in North London. There's Banh Mi, Thai Palace, Gold-*
User: Wait, that's the one I wanted.

Dialogue Management and ASR

- How can DM constraint recognition choice?

User: I'm going to Dallas on May eighteenth.
System: *Okay, where are you leaving from?*
User: Dulles.

User: I want to return on May twentieth.
System hears:
i want to return on may twelfth
i want to return on may twentieth
After dialogue management:
System: *So that's returning on May twelfth.*

Dialogue Management and NLP

- Words/phrases are interpreted *in context*

User: I need to book a flight.
System: *Okay, where are you leaving from?*
User: Dulles.

Dialogue management and NLG

- Tailor response to fit user model/current history

User: I'm interested in Thai restaurants in North London.
System: *I know of 8 Thai restaurants in North London. Two of them have very high food quality: Banh Mi and Golden Siam.*

Dialogue Management and TTS

- Emphasize new/pertinent information

User: I'm interested in Thai restaurants in North London.
System: *I know of 8 Thai restaurants in North London. Two of them have very high food quality: Banh Mi and Golden Siam.*
User: Actually, what about Chinese restaurants.
System: Okay, Chinese restaurants in North London.

Conclusions

- For systems to become more human-like, more intelligence needs to be added to all system components
- All components of a spoken dialogue system interact with each other
- Dialogue manager can be the logical module for mediating the interaction