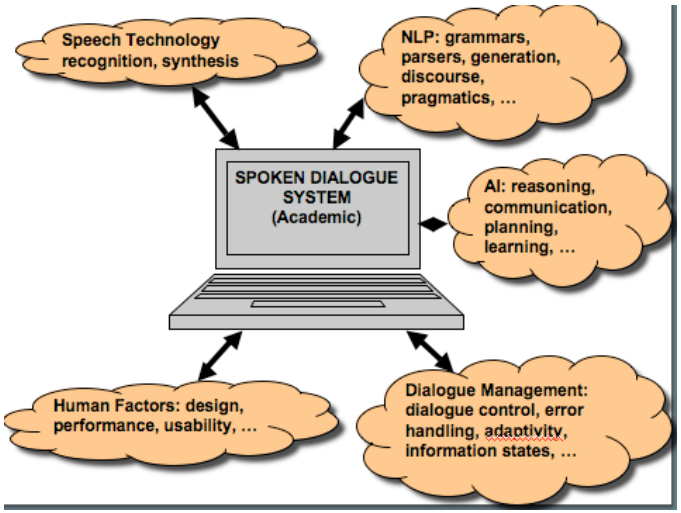
Advanced Dialogue Modeling for Practical Applications: Introduction

Ivana Kruijff-Korbayova

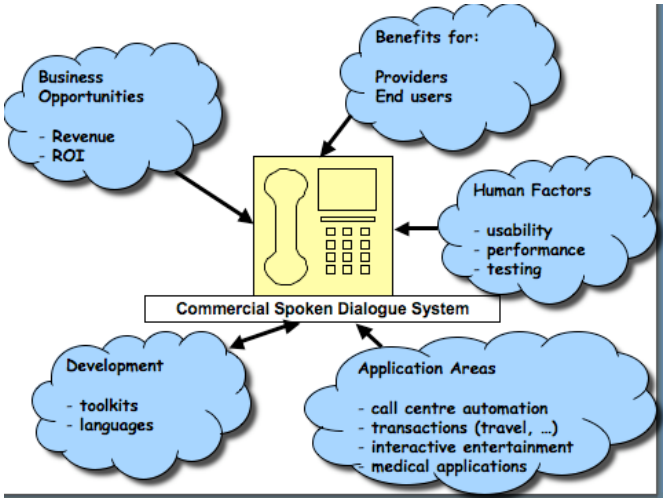
Dialogue Systems

- Emerging interdisciplinary area since the early 1990s
  - integration of speech technology, natural language processing, AI, dialogue / communication theory, human factors, …
  - scientific / academic – based research
  - commercially driven R&D
  - achievements and challenges

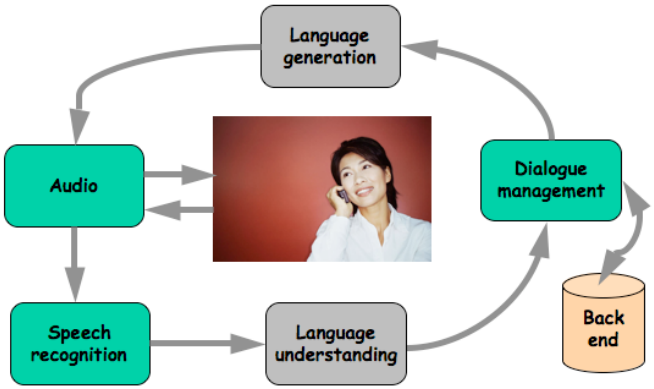
Dialogues System Research



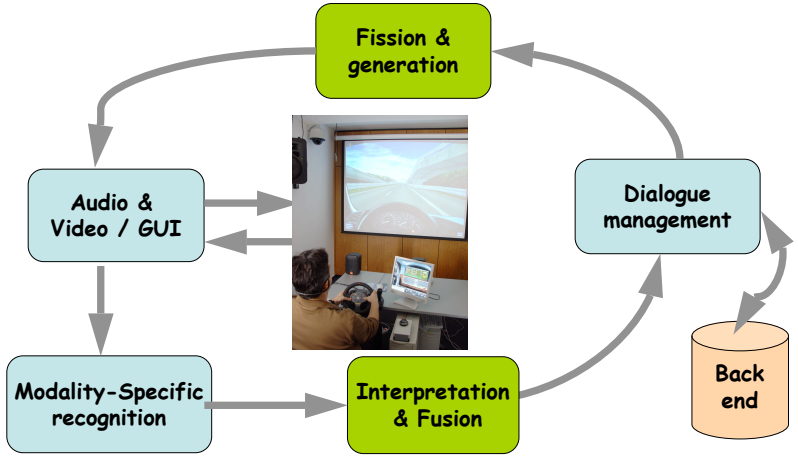
Dialogue System Industry



Typical Pipeline Architecture

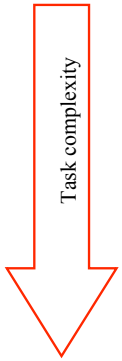


Typical Pipeline Architecture (Multimodal)



Dialogue Control

- Finite state systems
  - Sequence of predefined steps (dialogue script)
- Frame-based systems (form-filling)
  - Task represented as a set of slots to fill (frame, template)
- Agent-based systems
  - Joint problem solving by collaborating agents

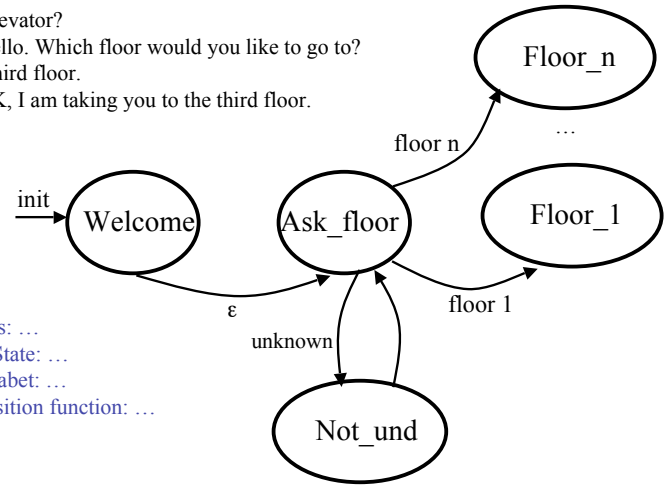


Finite State Systems

- Sequence of predefined steps (dialogue script)
- Set of states and transitions
  - State determines system utterance
    - Carefully designed prompts at each state
  - User utterance determines transition to next state (deterministic)
    - Speech recognition tailored to expected responses given the state
- No recursion! (=no nested subdialogues)
- System-driven interaction (system initiative)
- Rigid dialogue flow

FSM: Example 1

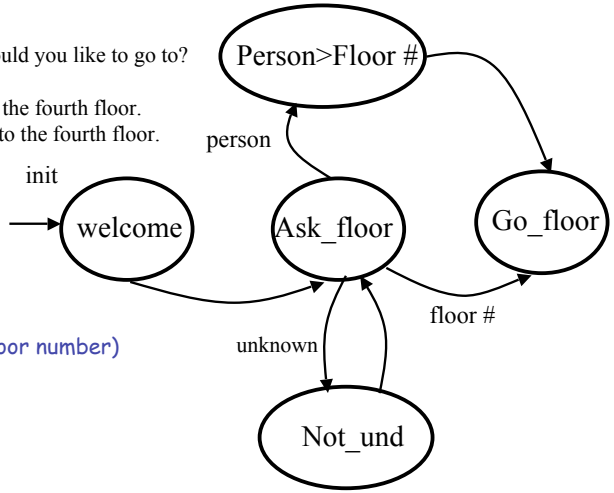
U: Elevator?
S: Hello. Which floor would you like to go to?
U: Third floor.
S: OK, I am taking you to the third floor.



States: …
Init-State: …
Alphabet: …
Transition function: …

FSM: Example 1 (extended)

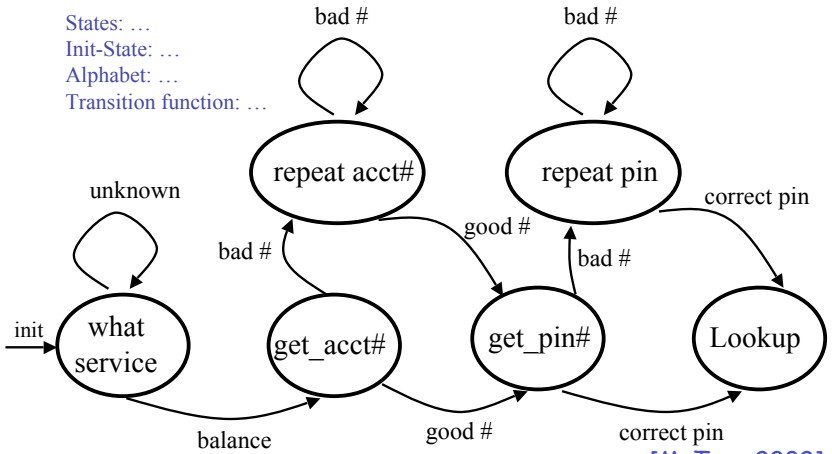
U: Elevator?
S: Hello. Where would you like to go to?
U: Prof. Barry.
S: Prof. Barry is on the fourth floor.
I am taking you to the fourth floor.



(variable for floor number)

FSM: Example 2

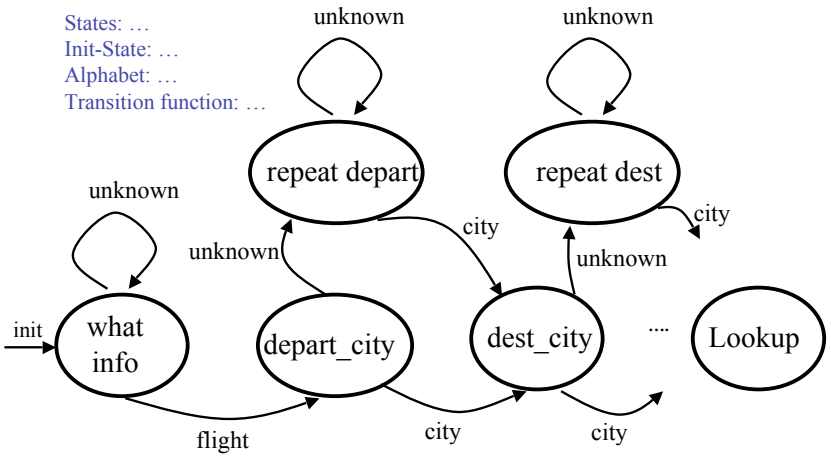
States: …
Init-State: …
Alphabet: …
Transition function: …



[McTear 2002]

FSM: Example 3

States: …
Init-State: …
Alphabet: …
Transition function: …



Finite State Systems: Sum Up

- Advantages
  - Fixed prompts can be pre-recorded
  - Speech recognition and input interpretation can be tuned for each state
- Disadvantages
  - Very rigid dialogue flow
  - Inhibiting user initiative
  - In principle can make more flexible, but it quickly gets very complex
  - Only suitable for simple fixed tasks

Frame-Based Models

- Task represented as a set of slots to fill
  - Frame (form): what info should be supplied by user
  - Dialogue states: which slots are filled
- Strategies for selecting the next system action (given dialogue state)
- Enables mixed initiative ("over-answering")
- Somewhat more flexibility

departure_city	?
departure_date	?
destination_city	?
return_date	?
…	

Frame-Based Models

S: What can I do for you?
U: I want to fly to Paris

departure_city	?
departure_date	?
destination_city	Paris
return_date	?
...	

S: Where will you fly from?
U: From Berlin on August 1st.

departure_city	Berlin
departure_date	1/8/05
destination_city	Paris
return_date	?
...	

Frame-Based Models

- Deciding what to do next
  - Next unfilled slot (similar to FS-based)
  - Slot-combination weighting
  - Ontology-based coherence
- Database lookup
  - Delayed (typically; after certain slots filled)
  - Immediate (can be "expensive", but helpful)

Ontology-based coherence

S: What is the patient's sex?
U: Female with severe nipple discharge
S: What is the patient's age?
U: Fifty five
S: Is the discharge bilateral?
U: No

S: What is the patient's sex?
U: Female with severe nipple discharge
S: Is the discharge bilateral?
U: No
S: What is the patient's age?
U: Fifty five

[Milward&Beveridge 2003]

Delayed vs. Immediate Lookup

S: What can I do for you?

U: I want to fly from Saarbruecken to Luxembourg.

S: When would you like to fly?

U: April 1st.

S: At what time would like to fly?

U: In the morning.

S: Sorry. There is no flight from Saarbruecken to Luxembourg on April 1st in the morning.

S: What can I do for you?

U: I want to fly from Saarbruecken to Luxembourg.

S: Sorry. There is no flight from Saarbruecken to Luxembourg.

VoiceXML

- VoiceXML is a web-based markup language for representing spoken dialogs
  - Form: basic unit of functionality
  - Field: prompts for and accepts user input
  - Prompt: sequence of audio elements or TTS messages
  - Audio: audio file or TTS message to play
  - Filled: processes input, can pass control to other forms
- Form Interpretation Algorithm
  - Defines how fields in a form are filled in , and how the fill ordering can be modified
- Global event handlers for (e.g., error handling, help)
  - Define behavior when predefined global conditions occur
- VoiceXML application collects and processes info, and plays back info
- VoiceXML assumes a voice browser
  - Info conveyed to user by audio output (synthesized and/or recorded)
  - Info received from user as audio input (voice and/or telephone keypad tones)
  - (analogical to HTML)

Frame-Based DM: Sum Up

- Advantages
  - Enables some user initiative
  - More flexible
- Disadvantages
  - Speech recognition more difficult, because user input less restricted
  - Not every task can be represented by a frame

Agent-Based DM

- (Collaborative) problem solving by (rational) agents
- Dialogue involves recognition, construction and execution of plans
- Agents have knowledge about solving tasks
  - deciding on goals (objectives): adopt, select, defer, abandon, release
  - forming plans to achieve goals (recipes)
  - executing those plans (acting)
  - revising decisions (re-planning, abandoning goals, etc.)
- Agents communicate to establish common ground
- Agents reason about beliefs and actions

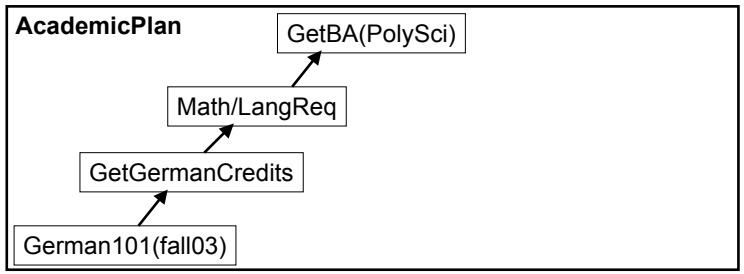
Collaboration

- Communication is a joint activity: agents collaborate to establish and achieve their goals
- Neither agent can accomplish the task alone
  --> mixed initiative
- Need joint goals and mutual understanding
- --> cooperation
- --> grounding

Intention Recognition

Given: plan for getting a BA

U: I'll take German 101 fall semester.



Interleaved Planning&Acting

User: Send ambulance one to Parma right away.
System: OK. [sends ambulance]
System: Where should we take the victim once we pick them u
User: Rochester General Hospital.
System: OK.

[Blaylock et al. 2003]

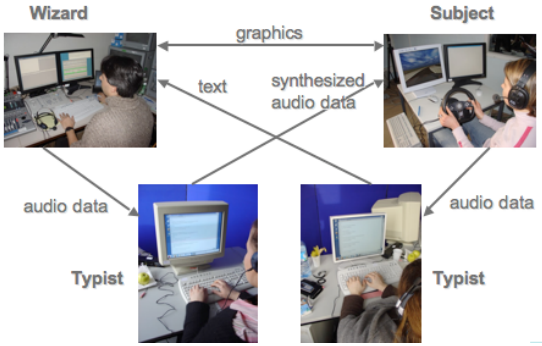
Agent-based: Sum Up

- Advantages
  - Enables flexible and adaptive dialogue modeling
  - Any task can be modeled
- Disadvantages
  - Intention recognition is difficult
  - Lots of reasoning (see QUD-based DM for "shortcuts")

  --> "expensive"

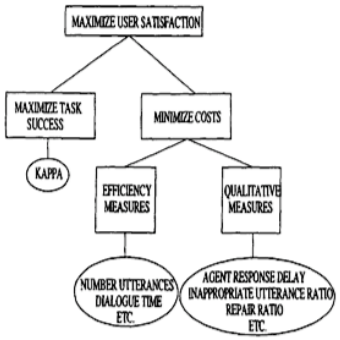
Development Methodologies

- Requirement Specification
  - Analysis of human-human dialogues
  - Wizard-of-Oz experiments (simulations) to gather user behavior samples and test design ideas in early stages of development
    - e.g., the TALK project WOZ experiment setup:



Development Methodologies

- Usability Evaluation
  - PARADISE framework [Walker et al. 1997]:
    - Maximize user satisfaction through maximizing task success while minimizing dialogue costs
    - User satisfaction (surveys)
    - Objective measures:
      - Task success (in terms of filling a set of slots)
      - Dialogue costs:
        - » Efficiency, e.g., no. of turns and time
        - » qualitative phenomena, e.g., no. of inappropriate utterances or repairs
    - Performance function: relative contribution of objective factors to user satisfaction



Deployment Platforms

- PC
  - GoDIS
  - Circuit-Fix-It Shop, TRIPS/TRAINS
  - Autotutor, Why-Atlas, BE&E, PACO …
- Telephone
  - Philips Train Timetable System, Deutsche Bahn info, …
  - It-Spoke weather
- Embedded voice systems
  - HAL (Home Automated Living), D'Homme project
- In-car voice or multimodal systems
  - BMW navigation, TALK project: MP3 player
- PDA, tablet PCs, next generation phones
  - MATCH, SmartKom
- Embodied agents and robots
  - REA, SAM, MRE, …
  - WITAS
  - MEL, BIRON, COSY system, Companions

DEMOS

Key Issues for the Future

- Pervasive systems
  - distributed dialogues: shifts between dialogue situations
  - concurrent dialogues: multitasking (co-ordination, synchronisation, redundancy)
  - interaction model needs to be predominantly event-based (external events, opportunistic)
- Adaptivity:
  - Systems need to be dynamically adaptive in a number of different ways: to the environments in which they are used (modality), to their user's preferences and needs (personalisation), to changes in task and context, to interaction progress.
- Ability to learn:
  - Systems need to be able to learn from interactions with users in order to provide an optimally usable interface that matches the current environment and user.
- Standardization:
  - There is a need for a common set of standards to support re-usability for developers and to support usability for the users of spoken dialogue systems, e.g. constraining vs. open-ended prompts, explicit vs. implicit verification.

References

- D. Jurafsky and J. Martin (2000): Speech and Language Processing, Chapter 19.
- McTear (2002): Spoken Dialogue technology. In ACM Surveys. pp. 1-80
- VoiceXML Forum: http://www.voicexml.org/
- Walker, Marilyn A., Litman, Diane J., Kamm, Candace A., Abella, Alicia (2000): PARADISE: A Framework For Evaluating Spoken Dialogue Agents. In Proc. Of the ACL Conference.