Language Technology II
Language-Based Interaction

Manfred Pinkal
Ivana Kruijff-Korbayová

Course website:
www.coli.uni-saarland.de/courses/late2

The Speaking Elevator (1)

- User: Elevator
- System: Which floor do you want to go?
- User: Four /Fourth floor
- System: <Command execution>

Advantages of a speech-based interface

- Allows device operation for disabled persons: elevator, wheel-chair, all kinds of in-home applications, train-ticket automata, … ("Barrier-free Environment")
- Helps in cases where hands and eyes are needed for another task: Driving, surgery, … ("hands-on"/"eyes-on" applications)
- Supports automation in cases where only the acoustic channel is available: Telephony applications (travel information, call centers, …)

The Speaking Elevator (2)

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: I'm taking you to the Fourth floor.
  <Command execution>

Speech-based interfaces: Problems (1)

- Automatic speech recognition (ASR) is not 100% reliable. Reliability varies with a number of factors (speaker, physical environment, background noise, ...).

Speech-based interfaces: Problems (1)

- Automatic speech recognition (ASR) is not 100% reliable. Reliability varies with a number of factors (speaker, physical environment, background noise, ...).
- But:
  - Dialogue modelling offers techniques that support robustness and effectiveness:
    - Grounding techniques (German: "Verständigungssicherung")
    - Clarification questions (by system)
    - Correction (by user)
  - Also: Quality and reliability of ASR are improving; confidence information supports sensible clarification strategies.

The Speaking Elevator (3)

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: I'm sorry I didn't understand you. Please specify a floor number.
- User: Four.
- System: I'm taking you to the fourth floor.

<Command execution>

The Speaking Elevator (4)

- User: Elevator
- System: Which floor do you want to go?
- User: Fourth floor
- System: I'm taking you to the fifth floor.
- User: No, fourth floor
- System: I'm taking you to the fourth floor.

<Command execution>

Speech-based interfaces: Problems (2)

"...If you'd like to hear all of your options again, press 49. If you've forgotten why you called in the first place, press 50."

Speech-based interfaces: Problems (2)

- For more complex applications, users have difficulties to keep track of the interaction: Limited memory, non-transparent dialogue structure (as compared to menu-based interaction).

Disadvantages of speech-based interfaces

- For more complex applications, users have difficulties to keep track of the interaction: Limited memory, non-transparent dialogue structure (as compared to menu-based interaction).
- But:
  - There are linguistic techniques to structure the dialogue (e.g., resuming previous dialogue; decomposing a choice between many options into a sequence of decisions in a manageable alternative set)
  - There are options to complement/support NL dialogue by other modalities (written text, graphics)

First Generation Dialogue Systems

- "Command and Control" pradigm
- Restricted vocabulary, keywords, system-initiated and controlled rigid dialogue structure
- Optimising for robustness and effectiveness (to make good for poor speech recognition quality)
- Since late Eighties, early Nineties

Second Generation Dialogue Systems

- Flexible, adaptive, cooperative dialogue systems
- Optimising for efficiency and user saticfaction
- Research systems since mid-nineties, first commercial systems recently

The Speaking Elevator (5)

- User: Elevator
- System: Which floor do you want to go?
- User: To professor Barry's office.
- System: I'm taking you to the fourth floor.

<Command execution>

Features of second-generation "natural" dialogue systems (1)

- Content orientation: Users say *what* they want
- Flexibility: Users say it *the way* they want, with no need to learn the specific commands a device "expects": Lexical choice, reference identification, syntax
- Collaborative clarification of ambiguous or underspecified user input.

The Speaking Elevator (6)

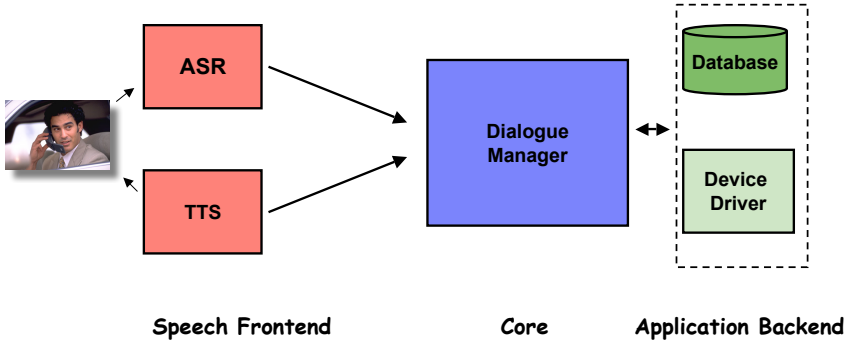
- User: Elevator
- System: Which floor do you want to go?
- User: To the phonetics department.
- System: Fourth floor or fifth floor?
- User: Professor Barry's office.
- System: I'm taking you to the fourth floor.

<Command execution>

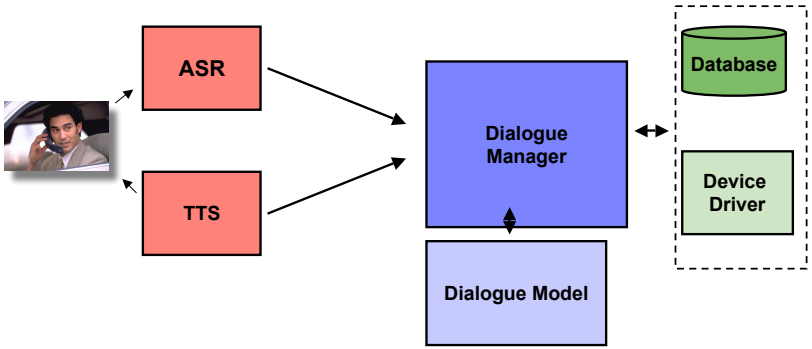
Features of second-generation "natural" dialogue systems (2)

- Adaptivity:
  - System adapts to preferences, expertise, ability of user, and to context
  - Alignment in utterance legth and speed, syntax, lexicon, semantic interpretation
- Interactive, context-sensitive help
- Mixed-Initiative: Flexible dialogue structure, jointly controlled and driven forward by user and system.
- ...

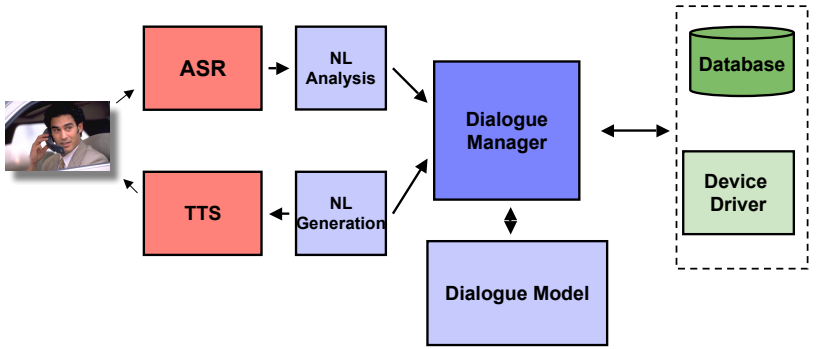
Basic Architecture of a Dialogue System



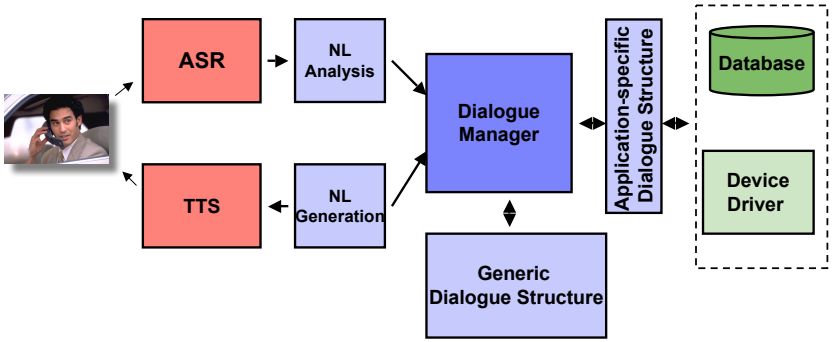
Basic Architecture of a Dialogue System



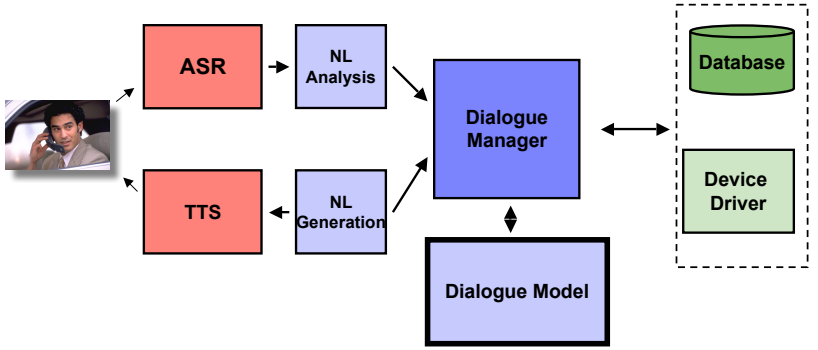
Basic Architecture of a Dialogue System



Basic Architecture of a Dialogue System



Basic Architecture of a Dialogue System



Dialogue Modelling

- Finite Automata /Transition Graphs

Dialogue Modelling

- Finite Automata /Transition Graphs
- ... augmented with subgraphs and global variables

Dialogue Modelling

- Finite Automata /Transition Graphs
- ... augmented with subgraphs and global variables
- Template-based /Form filling dialogue

Template-based Dialogue Modelling

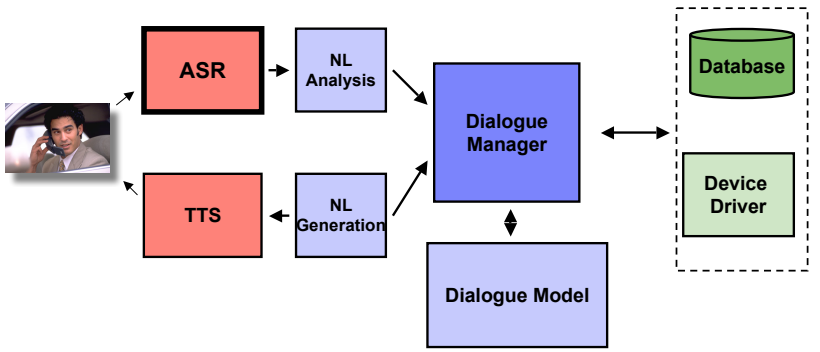
From	
To	
Date	
Departure time	

S: Where do you want to go?
U: To Frankfurt on June 15.
S: Which time do you want to go?
U: I want to leave from Saarbrücken at 9 a.m.

Dialogue Modelling

- Finite Automata /Transition Graphs
- ... augmented with subgraphs and global variables
- Template /Form Filling
- Information State Update

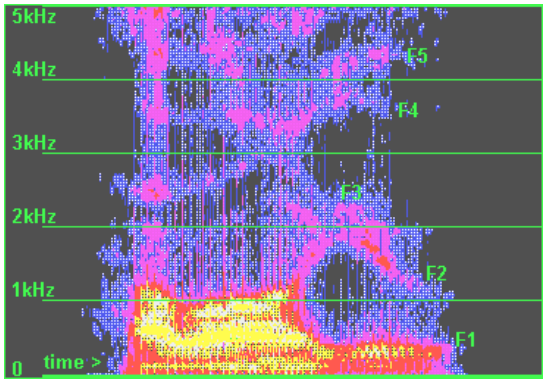
Basic Architecture of a Dialogue System



The Speech Front-end: ASR

- Input: An acoustic signal

A Spectrogram



How are you?

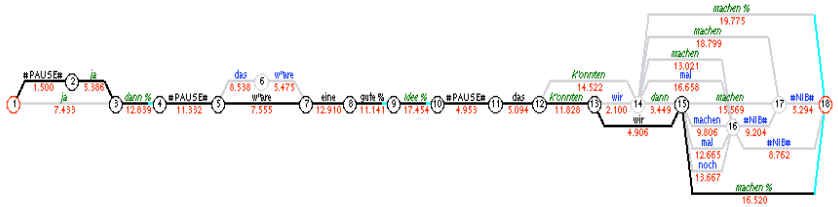
The Speech Front-end: ASR

- Input: An acoustic signal
- Processing:
  - Feature vector extraction
  - Statistical Phone, Word, Language Models – HMMs (Hidden Markov Models)

The Speech Front-end: ASR

- Input: An acoustic signal
- Processing:
  - Feature vector extraction
  - Statistical Phone, Word, Language Models – HMMs (Hidden Markov Models)
- Output (alternatively):
  - Best hypothesis / n-best
  - Word Lattice

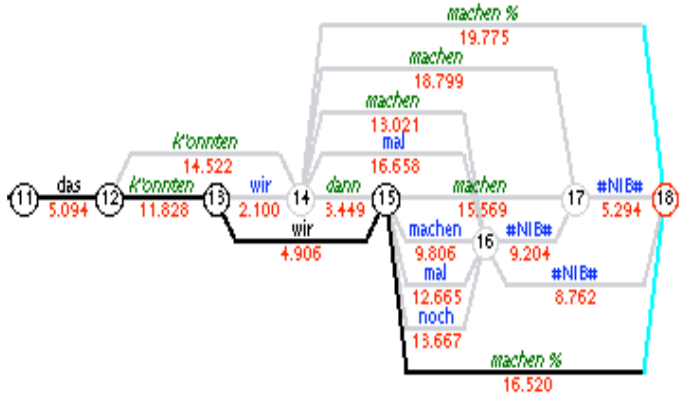
A Word Lattice



Ja, das wäre eine gute Idee. Das könnten wir dann machen.

Yes, that would be a good idea. We could do it.

A Word Lattice



Performance of ASR Systems

- Standard performance measure is "Word Error Rate" (WER):
  - The Minimum Edit Distance between best hypothesis and correct string: Insertions+Substitutions+Deletions
  - divided by total number of words in correct string.
- Example:
  - *Ja, das wäre eine gute Idee. Das könnten wir dann machen.*
  - Ja, dann wäre eine gute Idee. Das könnten wir machen.

Performance of ASR Systems

- Performance is dependent on an number of factors:
  - Acoustic quality of input
    - background noise
    - room acoustics
    - recording and transmission quality (headset, free microphone, standard phone, mobile phone)
  - Phonetic quality of input (voice, accent)
  - Linguistic quality of input (single words - continuous speech in reading quality - spontaneous speech)

Performance of ASR Systems

- Available processing resources
- Size of Lexicon (20 – 200.000)
- Perplexity (something like the weighted average number of choices for a random variable): If the recognizer expects 100 possible word forms with equal probability, on the avarage, the perplexity of its language model is 100.

Dictation Systems

- Dictation Systems (e.g. ViaVoice, Dragon Dictate):
  - Large lexicon (> 50.000)
  - High perplexity
  - Speaker-dependent ASR, personal training necessary

Domain-restricted dialogue systems

- Dialogue Systems:
  - Small domain- and application-dependent lexicon
  - Low perplexity
  - Context-specific language models/ Recognition grammars
  - Only content words matter
  - Grounding and clarification strategies available.
  - Reasonable results with speaker-independent ASR

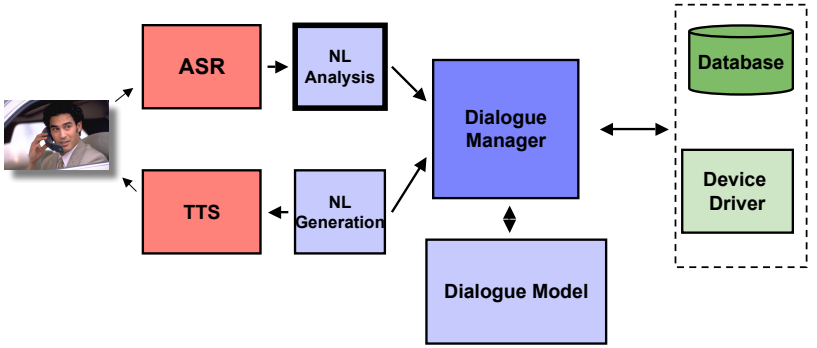
"Expectation-driven" Speech Recognition

- Context-specific language models (usually encoded in terms of BNF recognition grammars)
  - Fourth floor or fifth floor?
- Reweighting Recognition Results by Plausibility Considerations

Advanced Features of ASR

- Confidence Values (word and turn level)
- Out-of-Vocabulary words (OOV)
- Concept Recognition

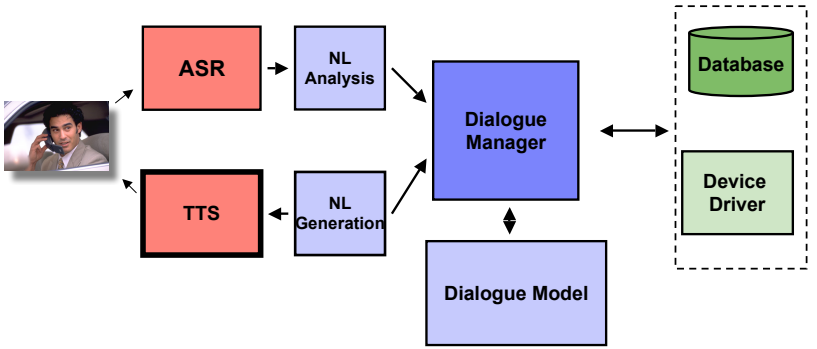
Basic Architecture of a Dialogue System



NL Analysis

- Deep grammatical analysis still difficult (lack of coverage, robustness), for most applications not necessary.
- Flat and partial analysis based on key-words or key-phrases ("semantic parsing"), ignoring function words and other irrelevant lexical material.
  - Will the flight from London scheduled for four thirty p.m. arrive in time?
- Recognition grammar and grammar for NL analysis often coincide.

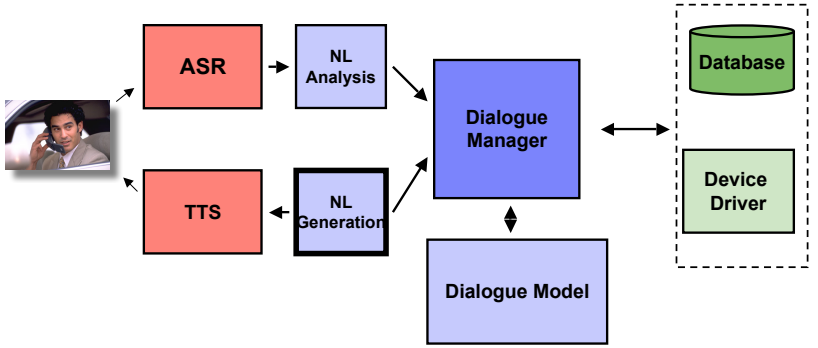
Basic Architecture of a Dialogue System



Text-to-Speech

- Diphone synthesis
  - better, but still rather poor quality
- Word concatenation
- Pre-recorded speech
  - high quality, but expensive and inflexible
- Unit selection
- http://www.naturalvoices.att.com/demos/index.html

Basic Architecture of a Dialogue System



NL Generation

- Usually on the basis of patterns and slot-filling.
  - The next flight to London Heathrow will leave as scheduled at 4:30 p.m.