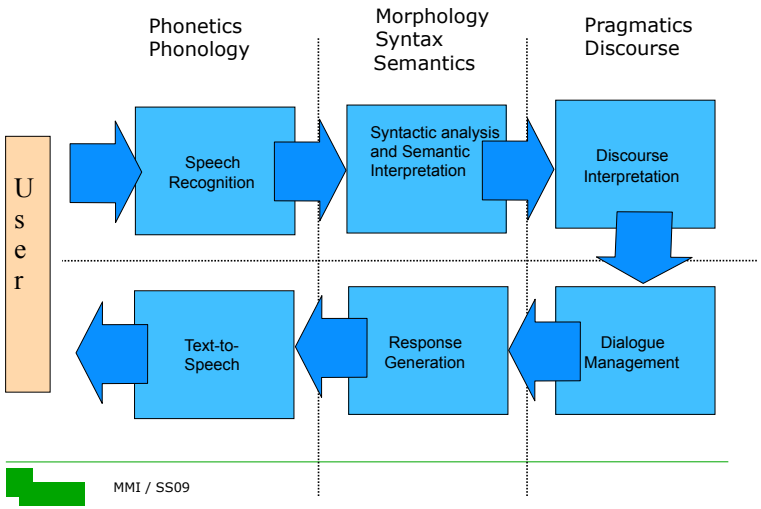
Human-Computer Interaction

Session 10
Natural Dialog Interaction (continued)

Classical SLDS structure



Voice Command

Current automotive speech technology at BMW

- Artikel auf *Spiegel Online* vom 25.6.2009

Voice Command

Automotive voice command (BMW)

Spoken Language Dialogue Systems (SLDS)

A system that allows a user to speak his queries in natural language and receive useful spoken responses from it

Provides an interface between the user and a computer-based application that permits spoken interaction in a "relatively natural manner"

What is a dialogue?

- multiple participants exchange information
- all participants pursue (ideally) the same goal
- discourse develops over the dialogue
- some conventions and protocols exist

- general structure
  - Dialogue = [episodes]+ (topic changes)
  - Episodes = [turn]+ (speaker changes)
  - Turn = [utterance]+ (function changes)

A lot to be handled...

- in both monologue and dialogue
  - information status: what is given, what is new?
  - coherence: how do the utterances fit together?
  - references: what is being referred to?
  - speech acts: what is the intention of the speaker?
  - implicature: what can be inferred from it?

- +only in dialogue
  - turn-taking: who has the the right to speak?
  - initiative: who is seizing control of the dialogue?
  - grounding: what info is settled between the speakers?
  - repair: how to detect and repair misunderstandings?

Resolve references

- Ellipsis
  - People often utter partial phrases to avoid repetition
    A: At what time is "Titanic" playing?
    B: 8pm
    A: And "The 5th Element"?
  - Necessary to keep track of the conversation to complete such phrases
- Some words are only interpretable in conext
  - Anaphora: "I'll take it", he said.
  - Temporal/spatial: "The man behind me will be dead tomorrow."

Handle information structure

Distinguish two parts of one utterance

- Theme:
  Part of a proposition that repeats known information to create cohesive connection to previous propositions („discourse cohesion")
- Rheme:
  Part of a proposition that contributes new information

Example: Who is he? He is (Theme) a student. (Rheme)

- There can be purely rhematic/thematic utterances

(Bolinger; Halliday, 1960's)

Understand speech acts

- Every utterance is an action performed by the speaker in a real speech situation
- Obvious in *performative* sentences: „I name this ship titanic.", „I bet you 5 bugs."
- Any sentence in a speech situation constitutes three kinds of acts:
  - Locutionary act: the utterance of the sentence „I'm cold."
  - Illocutionary act: the action in uttering it (asking, answering, commanding, …) → informing that I'm cold.
  - Perlocutionary act: the production of effects upon the addressee and ultimately the world → get window closed
- speech act explicates the illocutionary act

Austin (1962), Searle (1975)

Understand indirect meaning

S: „What day in May do you want to travel?"
U: „Uh, I need to be there for a meeting that's from the 12th the 15th."

- U does not answer the question, expects hearer to draw certain inferences
- Theory of conversational implicature: hearer can draw inferences because they assume the conversants follows four maxims (Grice, 1975):
  - Maxim of Quantity: Be exactly as informative as required
  - Maxim of Quality: Make your contribution one that is true
  - Maxim of Relevance: Be relevant.
  - Maxim of Manner: Be understandable, unambiguous, brief, and orderly

→ Maxim of Relevance allows S to know that U wants to travel by the 12th.

Understand grounding

Allwood, 1976;
Clark & Shaefer, 1989

- Interlocutors are trying to establish common ground, a set of mutual beliefs
- Listener must ground a speaker's contribution by acknowledging it, signaling understanding or agreement
- Various ways to do this:

S: „I can upgrade you to an SUV at that rate."

- Continued attention/permission to proceed - U gazes appreciatively at S
- Relevant next contribution - U: „Do you have an Explorer available?"
- Acknowledgement, "backchanneling" - U: „Ok/Mhm/Great!"
- Display/repetition - U: „You can upgrade me to an SUV at the same rate?"
- Request for repair- U: „Huh?"

Manage initiative

Control - the ability/license to bring up new topics, to start tasks, to pose questions, etc.

- System-initiative:
  system always has control, user only responds to system questions

- User-initiative:
  user always has control, system passively answers user questions
- Mixed-initiative:
  control switches between system and user either using fixed rules or dynamically based on participant roles, dialogue history, etc.

Initiative strategies

- System initiative (spoken "form filling")
  S: Please give me your arrival city name.
  U: Baltimore.
  S: Please give me your departure city name
  U: Boston
  S:…

Rigid, restricted vocabulary, rigid, NLP easy and more accurat,

- User initiative
  U: When do flights to Boston leave?
  S: At 8:30 AM and 3:45 PM.
  U: How much are they?
  S:…

requires good NLP, users must be aware of possible words

- Mixed initiative
  S: Where are you traveling to?
  U: I want to go to Boston.
  S: At time do you want to fly?
  U: Are there any cheap flights?

natural, open, unpredictable, hard to model, requires NLP and complex dialogue manag.

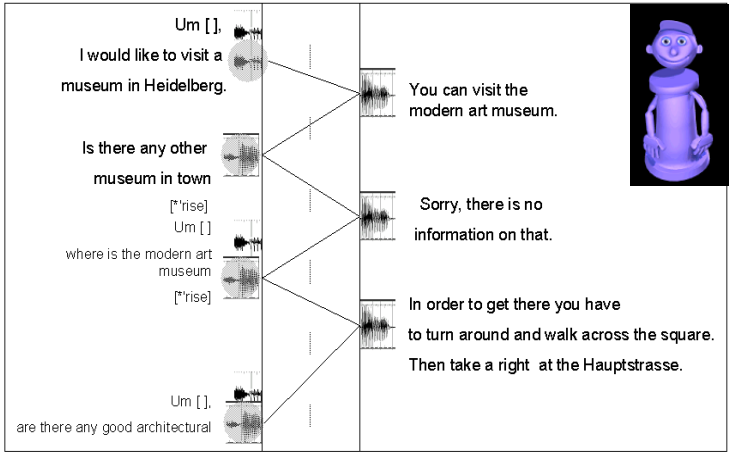
Understand turns and utterances

- Turn = [utterance]+
- But what is an utterance?
  - Not a syntactic sentence (may span several turns)
    A: We've got you on USAir flight 99
    B: Yep
    A: leaving on December 1.
  - Not a turn (multiple utterances may occur in one turn)
    A: We've got you on USAir flight 99 leaving on December. Do you need a rental car?
- Dialogue is characterized by *turn-taking*
  - Who should talk next?
  - When should they talk?

Manage turn-taking

- People know well when they can take the turn
  - Only little speaker overlap (~5% in English)
  - But little silence between turns either, a few of 1/10 s
    - Less than needed to plan motor routines for speaking
    - Speakers usually start motor planning before previous speaker has finished talking !!
- How do we know?
  - Schegloff (1968): *Adjacency pairs* set up speaker expectations and give rise to discourse obligations
    - QUESTION → ANSWER, REQUEST → GRANT, ...
    - Silence inbetween is dispreferred → pauses disturb users!
  - Sacks et al. (1974): *transition-relevance places* and rules that govern turn-taking, e.g.
    - If current speaker does not select next speaker, any other speaker may take next turn

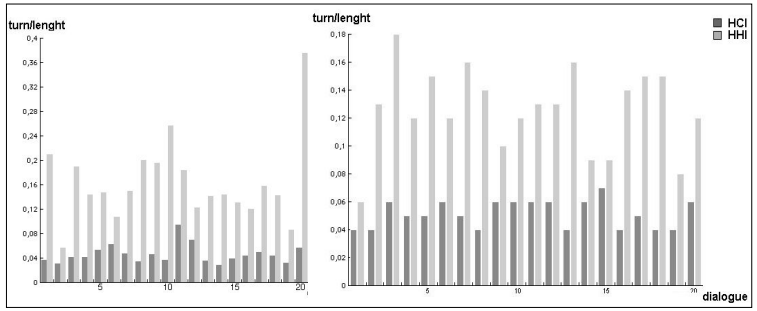
Usual structure of HCI dialogues



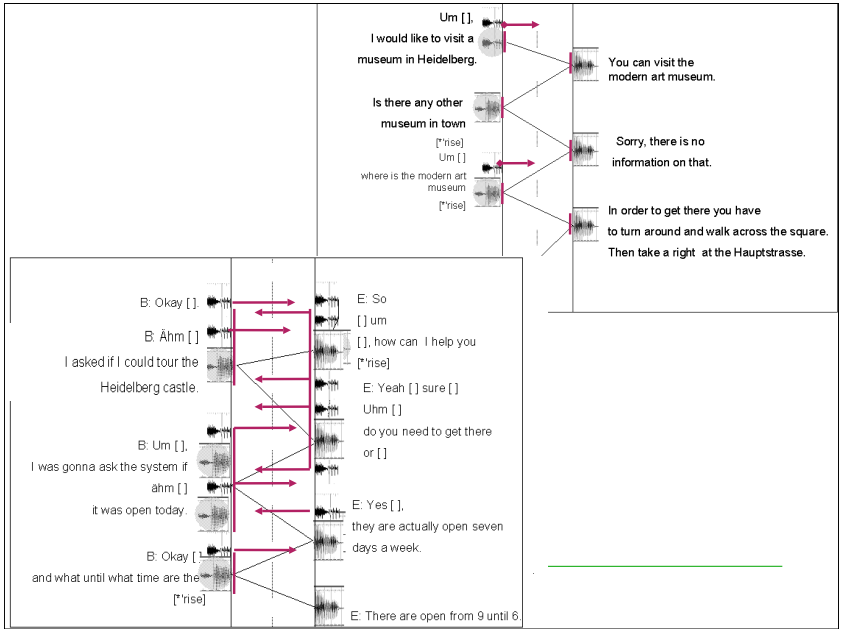
Measuring dialogue efficiency

Highly significant loss of dialogical efficiency in HCI vs. HHI using the PARADISE metric: Walker et al (2001) - dialogue turns / dialogue length

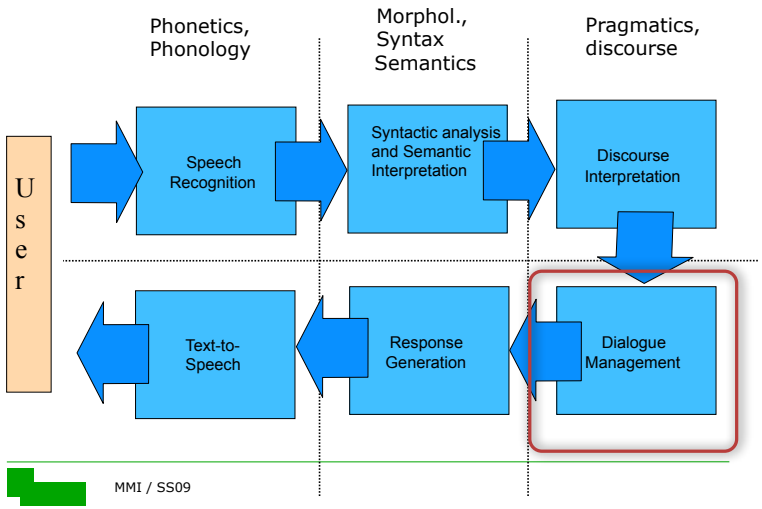
Robert Porzel, Uni Bremen



English Data German Data



Classical SLDS structure



DIALOGUE_MANAGER (DM)

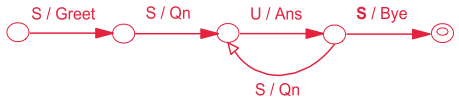
```
while conversation is not finished
  if user has completed a turn
  then interpret user's utterance
  if system has obligations
  then address obligations
  else if system has turn
  then if system has intended conversation acts
     then call generator to produce NL utterances
     else if some material is ungrounded
     then address grounding situation
     else if high-level goals are unsatisfied
     then address goals
     else release turn or attempt to end conversation
  else if no one has turn
  then take turn
  else if long pause
  then take turn
```

The heart of the system...

Jurafsky & Martin, 2000

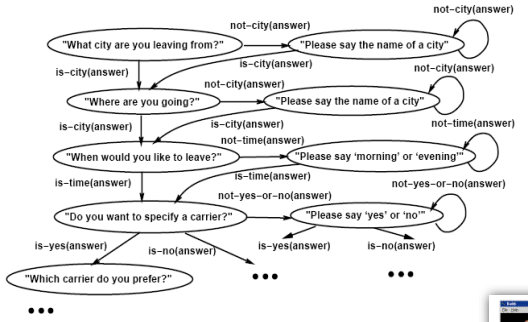
Finite state machine DM

Finite State Dialogue Grammar

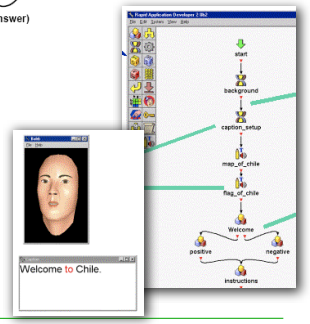


- Graph specifies all legal dialogues (dialogue grammar)
  - Nodes: system's questions
  - Transitions: possible paths through the network
  - Each state represents a stage in the dialogue ("now"), rarely with complete dialogue history
- System has initiative
- Context is fixed by the question being asked
- Used widely in commercial applications

Finite state machine DM



(Jurafsky & Martin, 2000)



Do-it-yourself example: CSLU Toolkit
http://cslu.cse.ogi.edu/toolkit/

Frame-based DM

Prompt: Where and when do you want to travel?
Grammar: <input of departure and arrival city, date and time>
Help: Please specify the departure and arrival city, date and time

FROM
Prompt: From which city are you leaving?
Grammar: <input of a city>
Help: Tell me the name of the city you want to leave from

TO
Prompt: To which city do you want to travel?
Grammar: <input of a city>
Help: Tell me the name of the city you want to travel to

WHEN
Prompt: When do you want to travel?
Grammar: <input of date and time>
Help: Please specify date and time of your journey

Filled: SELECT * FROM connections WHERE departure like 'FROM' AND destination like 'TO' *AND* time like *'WHEN'*

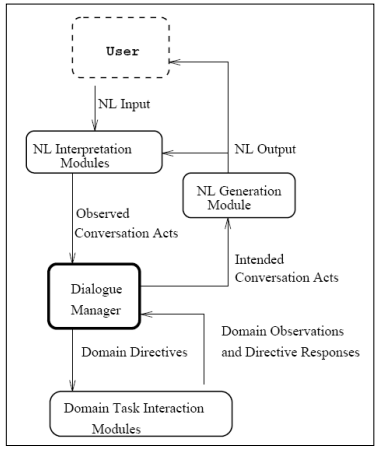
Frame-based DM

- template (frame) containing slots to be filled
  - destination: London, date: unknown, time of departure: 9
- questions to fill slots, conditions under which they can be asked
  - condition: unknown(origin) & unknown(destination)
    question: "Which route do you want to travel?"
  - condition: unknown(destination)
    question: "Where do you want to travel to?"
- system loops and decides next question based on what information has been elicited and what not yet
- system has initiative, dialogue more flexible, develops based on the current state of the system
- Commercially used, parts of standards: VoiceXML, SALT
  - bad for negotiation, planning, mixed-initiative

Intention-/plan-based DM

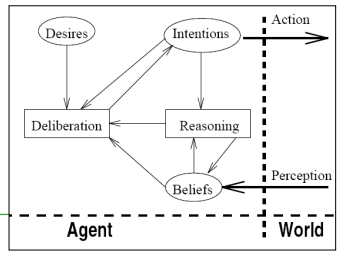
- **Idea**: dialogue arises from the collaboration of two or more agents in solving a task
  - there are goals to be reached
  - plans are made to reach those goals
  - the goals and plans of the other participants must be inferred or predicted
  - goals may involve changing the beliefs of others
  - models of the mental state of participants are used
- draws on methods from Artificial Intelligence
- permits more complex interaction between user, system, and underlying application
- allows for mixed-initiative dialogue

Example: TRAINS (Traum, Allen, 1996)



- Design system as agent with own mental states (Bratman, 1987)
  - Beliefs: world model
  - Desires: goals
  - Intentions: plans to pursue

Reasoning: derive new beliefs
Deliberation: decide actions



„Conversational Agency" (David Traum)

- Extending BDI to social attitudes that link one agent to others in dialogue
  - about the conversational partner, including mutual beliefs about the other's mental state

REQUEST(speaker,hearer,act)
body: MB(hearer, speaker, speaker WANT hearer DO act)
effect: hearer WANT hearer DO act

→ hearer thinks that speaker wants him to do an act

  - about what the agent *should* do, but not necessarily wants to: discourse obligations that inform deliberation

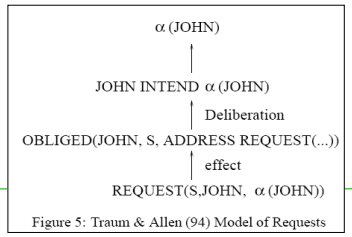


Figure 5: Traum & Allen (94) Model of Requests

Early example: TRAINS
(Traum, Allen, 1996)

TRAINS dialogue manager

- Explicit representation of **conversational state**
  - private and mutual beliefs, beliefs about user beliefs
  - proposals (to represent insincere or tentative acts)
  - domain plans (goals+actions+objects+constraints), either private, proposed or shared
  - discourse goals, represented as scripts specifying goals in different phases of conversation
  - obligations implied by received dialogue acts
  - intended acts to be generated
  - local initiative (who is expected to speak next)
  - stack of accessible discourse units
  - discourse structure information

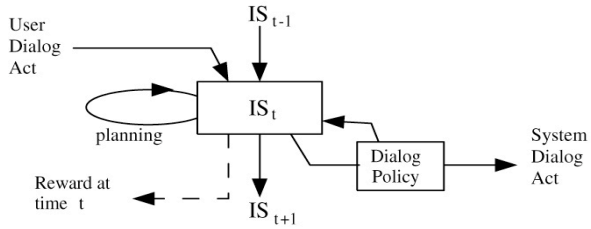
TRAINS dialogue manager

- Reactive: system will deliberate as little as possible until it can act, running in cycles
- No long-range plans, one step at a time

- Prioritized list of sources for deliberations
  1. Discourse obligations
  2. Weak obligation: don't interrupt user's turn
  3. Intended speech act (→ NLG + state update)
  4. Weak obligation: grounding (acknowledge, repair)
  5. Discourse goals: proposal negotiation
  6. High-level discourse goals (domain reasoning)

Information State approach

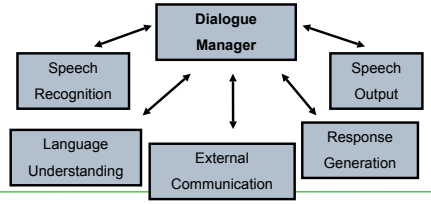
- Central data structure(s) to define conversational state
  - employed in deciding on next actions
  - updated in effect of dialogue acts by either speaker
- operational semantics of plans stated as update rules
- dialogue manager = definition of the contents of the IS + description of update processes

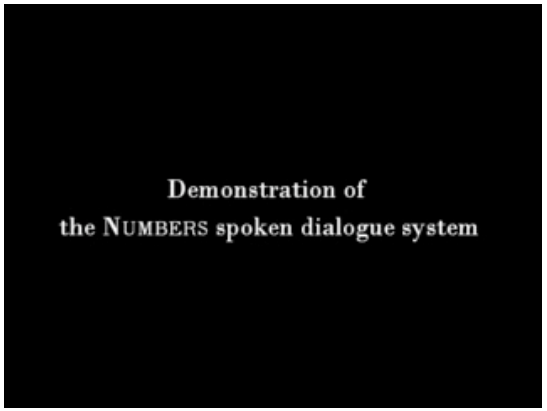


(Traum & Larsson, 2003)

SLDS architectures

- Pipeline structure with message passing
  - + Incrementality (D. Schlangen, Uni Potsdam)
- Blackboard
  - System = distributed, collaborating agents
  - Dialogue manager hosts central data structures (IS)
  - Rationale: Importance of context/discourse for all stages





- Incremental dialogue processing
  - G. Skantze (KTH Stockholm), D. Schlangen (Uni Potsdam)

Summary

Features/ dialogue control	State-based	Frame-based	Intention-based
Input	Single words or phrases	NL with concept spotting	Unrestricted NL
Verification	Explicit confirmation of each turn or at end	Explicit & implicit confirmation	Grounding
Dialogue Context	Implicitly in dialogue states	Explicitly represented Control represented with algorithm	Model of System's BDI + dialogue history
User Model	Simple model of user characteristics / preferences	Simple model of user characteristics / preferences	Model of User's BDI

Next session: multimodal interaction

speech recognition
& lip reading