Spoken Language Interaction

Dialogue systems –
systems that can converse

Konversationsprogramme

Eliza (J. Weizenbaum, 1969)

- Programm handelt/spricht wie ein Psychotherapeut
- Arbeitet ausschließlich mit Schlüsselworten und Kontextmustern
- Wissen ist abgelegt in Regeln/Skripten

„Eliza-Effekt“

- Eliza wirkt wie ein verstehender Gesprächspartner
- Gesprächspartner interpretieren Bedeutungen in die Aussagen des Gegenüber hinein

Bei schon einfachen Kniffen überfordert:

```
Benutzer: "Ich möchte auf keinen Fall über's Wetter reden!"
Bot: "Gern! Hier in Bielefeld regnet es mal wieder."
```

ELIZA - Skripte

```
ELIZA SCRIPT FILE:
Sxxxxxxxxx - DEFINE SIGNON MESSAGE
Txxxxx
Txxxxx - DEFINE PAIR OF WORDS TO TRANSPOSE
Nxxxxx - RESPONSE FOR NULL ENTRY
Mxxxxx - DEFINE RESPONSE FOR LATER USE OF "MY„
Xxxxxxx - DEFINE RESPONSE FOR NO KEYWORD FOUND
Kxxxxxx - DEFINE KEYWORD
Rxxxxxx - DEFINE RESPONSE FOR PREVIOUS KEYWORD
...
```

```
TRANSPOSITION TABLE
T ARE „
T AM „
T AM „
T ARE „
T WERE „
...
```

```
SIGNON MESSAGE
SHI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT?
SSO HOW ARE YOU DOING TODAY?
...
```

```
NO KEYWORD FOUND
XPLEASE GO ON.
XWHAT DOES THAT SUGGEST TO YOU?
...
```

```
NULL RESPONSES
Nyou're gonna wear your enter key out if you keep
    that up.
...
```

```
DEFINE KEYWORDS AND RESPONSES
KCAN I „
RPERHAPS YOU DON'T WANT TO*.
RDO YOU WANT TO BE ABLE TO*?
KFRIEND „
RWHY DO YOU BRING UP THE TOPIC OF FRIENDS?
RDO YOUR FRIENDS WORRY YOU?
RDO YOUR FRIENDS PICK ON YOU?
...
```

Backgrounds

What is dialogue?

What distinguishes it from monologue?

What is a dialogue?

- multiple participants exchange information
- all participants pursue (ideally) the same goal
- discourse develops with the dialogue
- some conventions and protocols exist
- general structure
  - Dialogue = [episodes]+ (topic changes)
  - Episodes = [turn]+ (speaker changes)
  - Turn = [utterance]+ (function changes)

There is a lot to handle…

- 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?

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 | a student.

Theme | Rheme

- There can be purely rhematic/thematic utterances

(Bolinger; Halliday, 1960‘s)

Coherence

„John hid Bill's car keys. He likes spinach."

- Hearers try to find out how utterances connect
  - Hearer will either question the coherence, or construct an explanation that makes it coherent.

- Informational approach: *coherence relations* between the information that the parts convey
  - Hobbs (1979), for example:
    - Result: „John bought an Acura. His father went ballistic."
    - Explanation: „John hid Bill's car keys. He was drunk."
    - Elaboration: „John bought an Acura. A big new SUV."
  - Rethorical Structure Theory (Mann & Thompson, 1987)
  - *discourse connectives*: explicit words that constrain - but not create! - possible relations, e.g. „because" or „and"

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
  - Illocutionary act: the action in uttering it (act of asking, answering, promising, commanding, …)
  - Perlocutionary act: the production of effects upon the feelings, thoughts, actions of the addressee
- speech act describes the illocutionary act

Austin (1962), Searle (1975)

Coherence (revisited)

- Intentional approach: adopting speech act theory, hearer is to infer plan-based intention of the speaker
- Recognize discourse structure
  - Based on cue words/phrases or prosody
  - Based on mental model of interpretation, e.g. beliefs, desires & intentions (BDI) (Grosz & Sidner, 86)
    - The discourse has a purpose (DP) and…
    - …each segment has a purpose (DSP) that plays a role in achieving the DP („subdialogues“)
    - Two relations between DPS‘s: Dominance and Satisfaction-precedence
  - Said to be „AI-complete“, i.e. a full human-like AI needed

Reference

- 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.”

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 conversation follows 4 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.

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?
- Appears fluid but not obvious, no computational model exists

Turn-taking

- People know well when they can take the turn
  - 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 when a speaker is…
  - giving up or taking a turn?
  - holding the floor?
  - interruptable?
- How do we know when…
  - its our turn obligatorily? or optionally?

Simple Turn-Taking Rules

- Sacks et al. (1974): Rules apply at each *transition-relevance place* of each turn:
  - If current speaker has selected A as next speaker, then A must speak next
  - If current speaker does not select next speaker, any other speaker may take next turn
  - If no one else takes next turn, the current speaker may take next turn
- TRPs are where the structure of the language allows speaker shifts to occur
- TRPs tend to occur at utterance boundaries (not sentence boundaries)

Turn-taking

- Schegloff (1968): *Adjacency pairs* set up next speaker expectations, gives also rise to discourse obligations
  - QUESTION → ANSWER
  - GREETING → GREETING
  - COMPLIMENT → DOWNPLAYER
  - REQUEST → GRANT
- Silence after the first part of a pair is *significant*
  - Significant silence is *dispreferred* (like a "No")

  A: Is there something bothering you or not? (1.0s)

  A: Yes or no? (1.5s)

  A: Eh?

  B: No.

→ Pauses in dialogue systems with slow speech recognizers disturb users (Yankelovich et al., 1995)

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

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

requires good NLP, users must be aware of possible words

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

Grounding

- There's a purpose to conversations
- Participants are trying to come to a meeting of minds, they're trying to establish common ground (a set of mutual beliefs)
- Hearers must ground a speakers utterances by making it clear whether or not understanding has occurred
- Various ways to do this…

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

User: ????

Stalnaker, 1978

Establishing grounding

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

U:

- Continued attention/permission to proceed

  (U gazes appreciatively at S)
- Relevant next contribution

  U: Do you have an Explorer available?
- Acknowledgement, “backchanneling”

  U: Ok/Mhmmm/Great!
- Display/repetition

  U: You can upgrade me to an SUV at the same rate?
- Request for repair

  U: Huh?

Clark & Shaefer, 1989

Dialogue systems

Dialogue managers

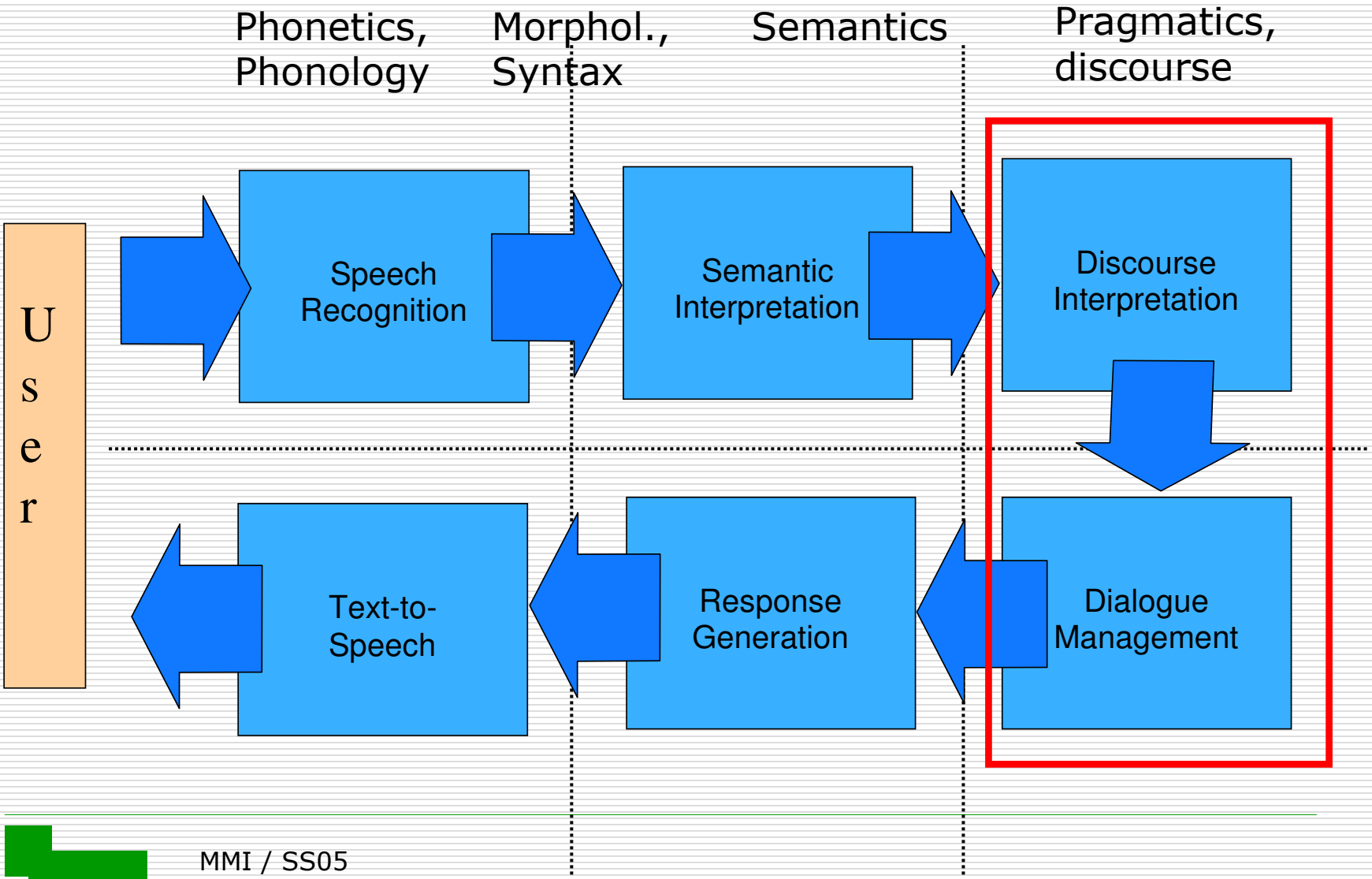
Finite-state based

Frame-based

Agent-based

Information states

Spoken Dialogue System



Dialogue Manager

Central tasks

- Interpretation of input
- Maintenance of discourse context
- Determine if information suffices to identify the task, handle repairs
- Determine what is expected or reasonable given context, planning of system responses
- Communicate with external applications (database, etc.)
- Manage communication flow

DIALOGUE_MANAGER

```
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
```

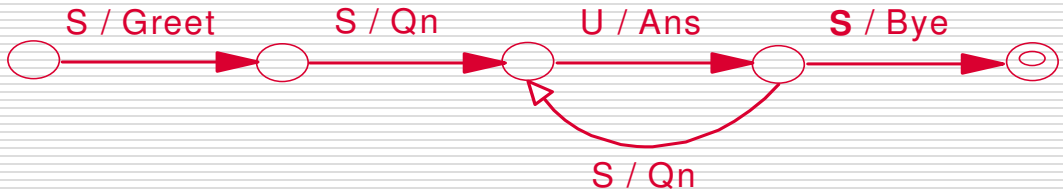
Jurafsky & Martin, 2000

Dialogue managers

- Continuum of approaches
  - Finite-state automaton/pattern-matching
  - Frames/production rules
  - Agent/plan-based

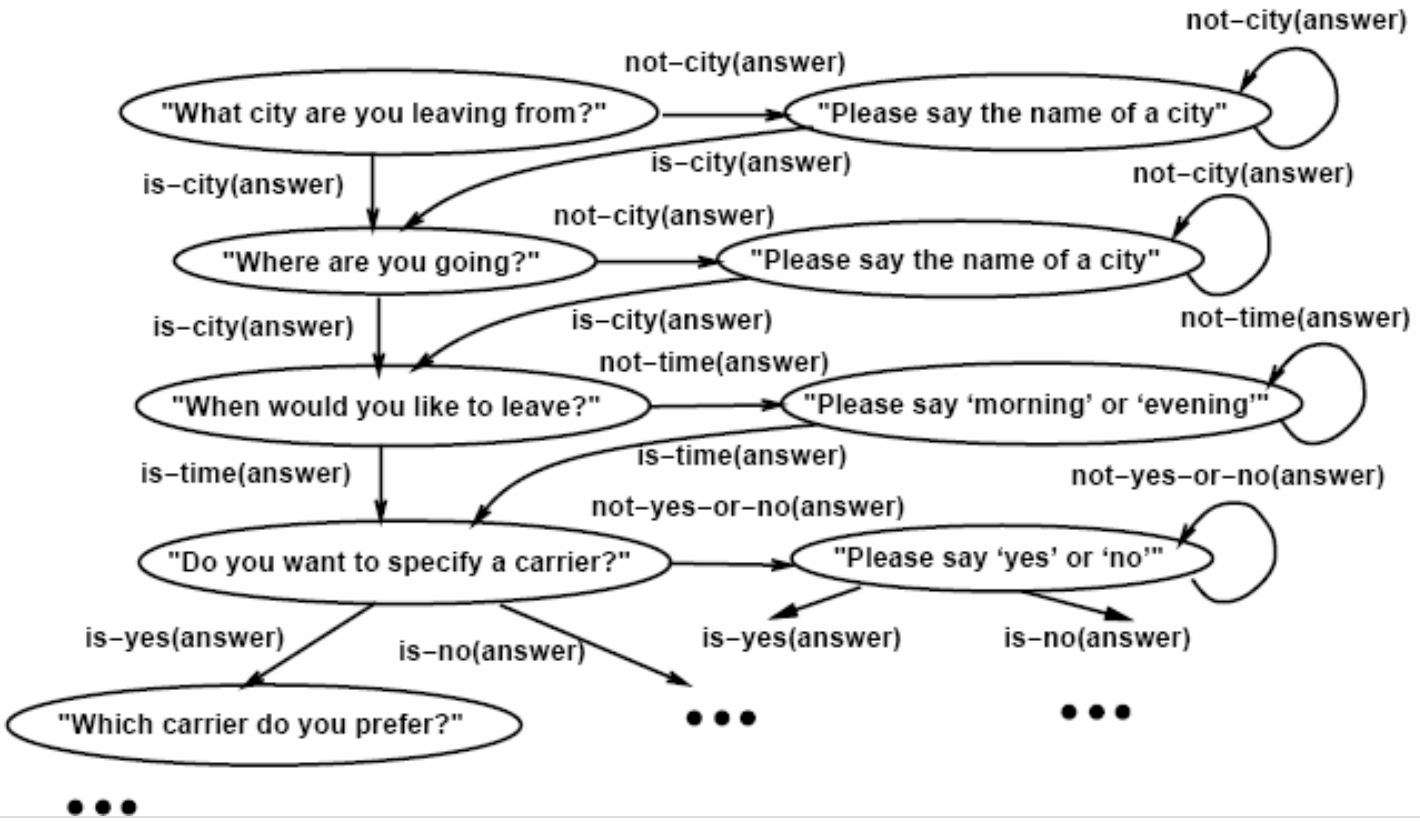
FSA dialogue manager

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

FSA dialogue manager



(Jurafsky & Martin, 2000)

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

FSA dialogue control

- Advantages
  - suitable for well-structured tasks with pre-determined sequence of questions
  - dialogue can be modelled graphically
  - can include sub-dialogues for sub-tasks e.g. getting a date
  - some empirical evidence that users prefer a predictable control flow
- Disadvantages
  - very little “strict” structure exists in real dialogues…
  - problem when dialogues deviate from predetermined path
  - difficult for user to make corrections
  - difficult for user to introduce unexpected information
  - not suitable for more complex tasks (combinat. explosion)

Frame-based dialogue manger

- template (frame) containing slots to be filled
  - destination: London, date: unknown, time of departure: 9
- questions to fill slots, with 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 to be asked based on what information has been elicited and what not yet (based on question conditions or priorities)
- system has initiative
- dialogue more flexible, develops based on the current state of the system
- e.g. VoiceXML, SALT

Frame-based dialogue manager

- Problems:
  - user can provide more information than was asked for in the system prompt e.g.

    S: where are you travelling to?
    U: London on Friday
  - user's answer could include various permutations of the required information e.g.

    Destination

    Destination + Date

    Destination + Time

    Destination + Date + Time

    Destination + Time + Date
  - other problems accounting for user initiative, grounding, etc.

Problems: complex Tasks

- users with wide range of different levels of knowledge would require wide range of system responses and NLP capabilities
- the state of the world may change dynamically during the course of the dialogue - not possible to specify all possible configurations in advance;
- dialogues involving
  - negotiation of some task to be achieved
  - planning and other types of collaborative interaction
  - mixed-initiative

Agent-based/plan-based control

- dialogue arises from the collaboration of two or more agents, as they cooperate to solve 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 of an agent-based system

U: I'm looking for a job in the Calais area. Are there any servers?

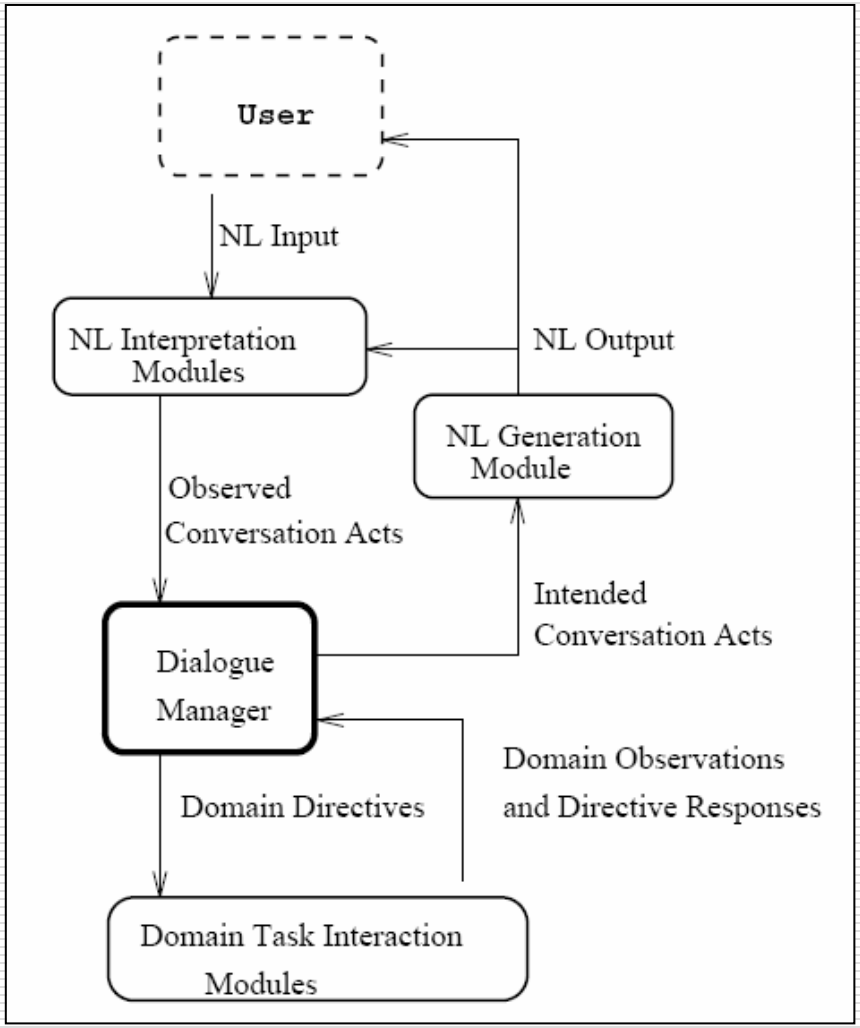
S: No, there aren't any employment servers for Calais. However, there is an employment server for Pas-de Calais and an employment sever for Lille. Are you interested in one of these?

- system recognizes user's needs and attempts to provide a more co-operative response

Agent-based modeling

- different approaches depending on what is modeled
  - Planning and plan recognition
  - Beliefs, desires, and intentions logics (BDI)
  - Rational agency

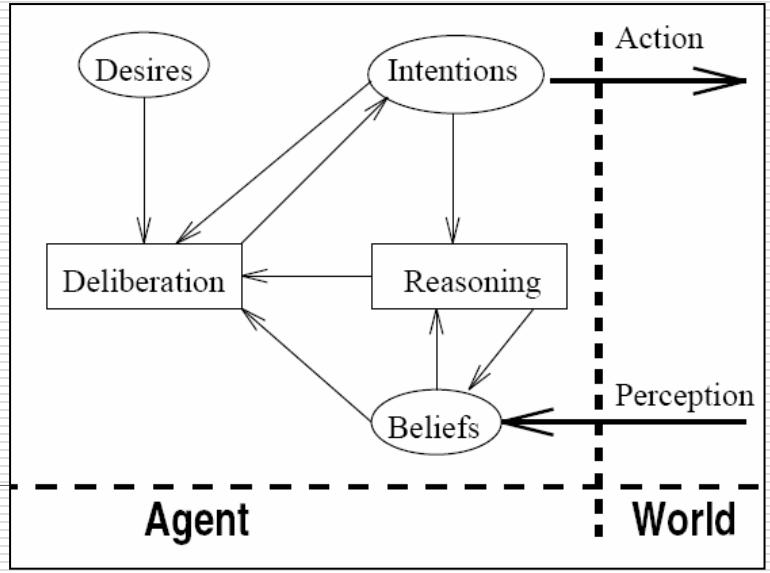
Example: TRAINS (Traum, Allen, 1996)



- 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 (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

source of obligation	obliged action
S_1 Accept or Promise A	S_1 achieve A
S_1 Request A	S_2 address Request: accept **or** reject A
S_1 YNQ whether P	S_2 Answer-if P
S_1 WHQ P(x)	S_2 Inform-ref x
utterance not understood or incorrect	repair utterance
S_1 Initiate DU	S_2 acknowledge DU
Request Repair of P	Repair P
Request Acknowledgement of P	acknowledge P

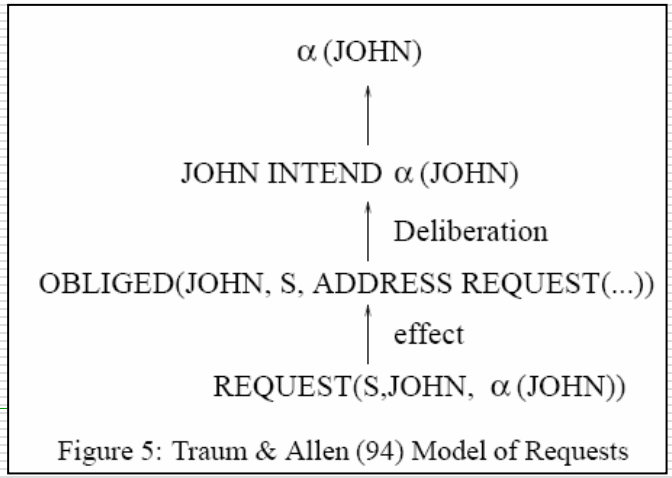


Figure 5: Traum & Allen (94) Model of Requests

Conversational agency (cont.)

- Dialogue structured in discourse units (DUs), built up by single-utterance grouding acts
- Extent speech acts to multi-level conversation act theory

Level	Act Type	Sample Acts
<UU	**Turn-taking**	take-turn keep-turn
UU	**Grounding**	Initiate Repair Ack Continue
DU	**Core Speech Acts**	Inform YNQ Accept Request
>DU	**Argumentation**	Elaborate Q&A

Table 2: Conversation Act Types

TRAINS-93 dialogue manager

- Context representation: 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
  - intended acts to be generated
  - local initiative (who is expected to speak next)
  - stack of accessible discourse units
  - discourse structure information

Example

U: “I want to go to Pittsburgh in May.” (INFORM1)

- Conversational state
  - Discourse obligations: NONE
  - Turn holder: system
  - Intended acts: NONE
  - Unacknowledged speech acts: INFORM1
  - Discourse goals: get-travel-goal, create-travel-plan
- Conversational state update
  - Intended acts: REQUEST1, ACKNOWLEDGE-INFORM1

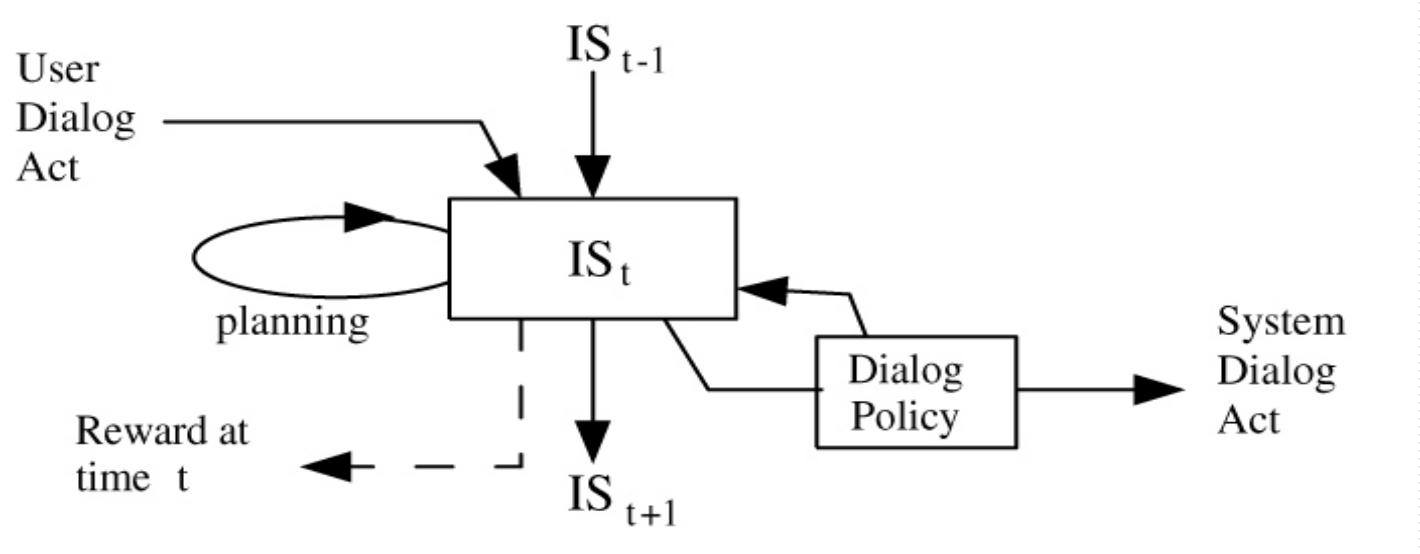
S: “And, what day in May did you want to travel?”

TRAINS-93 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 today

- 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 determined by update rules
- dialogue manager = definition of the contents of the IS + description of update processes



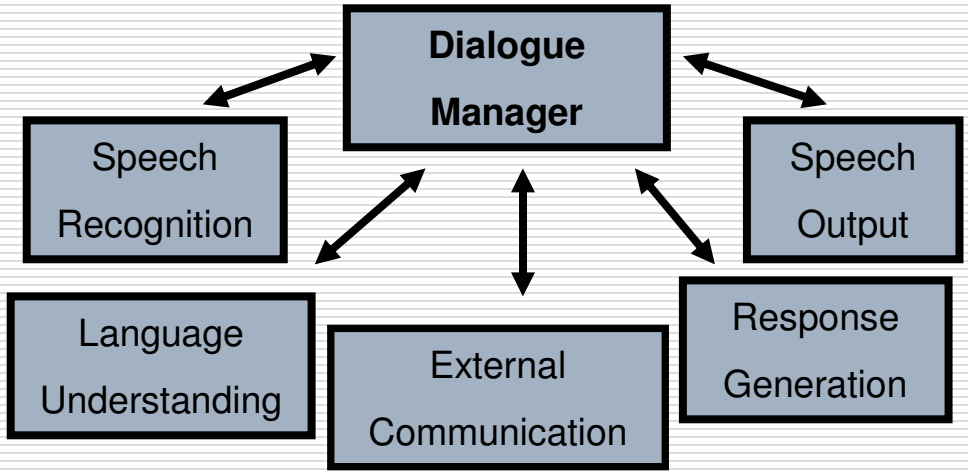
(Traum & Larsson, 2003)

Example: Information State

```
output:
  < hello, welcome to the edinburgh informatics flight
    booking system.what is your departure city?  >
lastspeaker: user
recogninput: < edinburgh >
input: < edinburgh >
lastmoves:
  < [edinburgh],u,
    [([greet],s),([ask_user_start_city],s)] >
filledslotsvalues:
  < [([ask_user_start_city],s)],[[edinburgh]] >
turn: system
oplansteps:
  ( [ask_user_destination_city] , [release_turn] )
nextmoves: < [ask_user_destination_city],s >
int: < [release_turn] >
. . .
```

Agent-based architectures

- Pipeline
- Blackboard
  - System = distributed, collaborating agents
  - Dialogue manager hosts central data structures (IS)
  - Rationale: Importance of context/discourse for all stages



Summary

Features/ dialogue control	State-based	Frame-based	Agent-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

Some dialogues systems

- Commercial Sytems:
  - small vocabulary (~100 words)
  - closed domain
  - system initiative
- Research systems:
  - Larger (but still limited) vocabulary (~1000 words)
  - closed domain
  - (limited) mixed initiative
- Some important applied research systems
  - TRAINS [1991-2000]: flight/train timetables
  - Verbmobil [1993-2000]: translation, scheduling
  - TRINDI, GODIS [1998-2000]: SmartHome
  - …

Recent Trends

- Complete models
- Stochastic and hybrid methods
- Multimodality
- Efficiency and robustness
- Domain-dependent instantiations
- Focus on measurable improvements for specific subproblems
  - reduction of speech recognizer word errors
  - improved quality of translation