Human-Computer Interaction

Termin 10:
Natural Dialog Interaction
### The evolution of user interfaces

<table>
<thead>
<tr>
<th>Year</th>
<th>Paradigm</th>
<th>Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950s</td>
<td>None</td>
<td>Switches, punched cards</td>
</tr>
<tr>
<td>1970s</td>
<td>Typewriter</td>
<td>Command-line interface</td>
</tr>
<tr>
<td>1980s</td>
<td>Desktop</td>
<td>Graphical UI (GUI), direct manipulation</td>
</tr>
<tr>
<td>1980s+</td>
<td>Spoken Natural Language</td>
<td>Speech recognition/synthesis, Natural language processing, dialogue systems</td>
</tr>
<tr>
<td>1990s+</td>
<td>Natural interaction</td>
<td>Perceptual, multimodal, interactive, conversational, tangible, adaptive</td>
</tr>
<tr>
<td>2000s+</td>
<td>Social interaction</td>
<td>Agent-based, anthropomorphic, social, emotional, affective, collaborative</td>
</tr>
</tbody>
</table>
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* with the application in a “relatively natural manner”.
Levels of sophistication

Controlled language
- limited vocabulary, simple grammar (e.g. command language)

Natural language
- Single utterances (e.g. question & answer), huge vocabulary, complex grammar, grammatical variation, etc.

Natural dialogue
- Longer stretch of speech interaction of two interlocutors

User

Phonetics, Phonology

Speech Recognition

Morphol., Syntax, Semantics

Syntactic analysis and Semantic Interpretation

Natural language understanding

Response Generation

Natural language generation

Text-to-Speech

Dialogue Management

Discourse Interpretation

Pragmatics, discourse
Natural language understanding

Tree classical steps:

1. Syntax analysis/parsing:
   - Determine sentence structure from sequence of words

2. Semantic interpretation/understanding:
   - Determine word meanings and the overall meaning of their composition in the sentence

3. Discourse interpretation/pragmatic analysis:
   - Use context information to complete and disambiguate sentence meaning
   - Determine intention behind the sentence

Input: Tarzan kissed Jane.

Syntax

Parsing

parse tree:

sentence

Noun_phrase

Verb_phrase

Noun

Verb

Noun_phrase

Tarzan

kissed

Jane

Semantic Interpretation

Internal representation:

person: tarzan

agent

kiss

object

instrument

lips

Semantics

Conceptual/world knowledge interpretation

expanded representation:

possess

pet: cheetah

experiencer

love

object

person: tarzan

agent

kiss

object

person: jane

instrument

lips

location

jungle

location

To:

question answerer, database query handler, translator, etc.

MMI / SS05
Ziel: Baumartige Zerlegung des sprachlichen Ausdrucks in seine Komponenten gemäß einer Grammatik

PARSE ("the dog is dead", G):
[S: [NP: [Article: the][Noun: dog]]
[VP: [Verb: is][Adjective: dead]]

- Grammatik: Formale, endliche Beschreibung der *Struktur* aller Elemente einer (oft unendlichen) Sprache
- Parsing = Suchen nach einer möglichen Ableitung eines Satzes in einer Grammatik → Ableitungsbaum
- Beispiel für „Tim aß den Löwen“
Sprachverstehen - Interpretation

☐ Aufgabe: Bedeutungsrekonstruktion
  ■ Was ist die Bedeutung von „Er beginnt um zwei im Raum V2-122.“?

☐ Unterscheide:
  ■ Semantisches Potential: Linguistisch bestimmte Bedeutung, lässt sich allein mit linguistischem Wissen ermitteln

\[ \text{Begin}(e,t,l) \land \text{Event}(e) \land \text{Time}(t) \land \text{Location}(l) \]
\[ \land \text{Equal}(t,2) \land \text{Room}(l,V2-122,?b) \]

■ Aktueller semantischer Wert: Volle Interpretation unter Anwendung nicht-linguistischens Wissens (Kontext, Domäne, Welt):

\[ \text{Begin}(e,t,l) \land \text{Event}(e) \land \text{Time}(t) \land \text{Location}(l) \]
\[ \land \text{Equal}(t,2) \land \text{Room}(l,V2-122,?b) \]
\[ \land \text{Talk}(e,s,l) \land \text{Proffessor}(s,\text{Cambridge}) \]
\[ \land \text{Name}(s,\text{Steven – Hawking}) \land \text{Building}(b,\text{Uni – Bielefeld}) \land \ldots \]
Interpretation

linguistisches Wissen

Semantik  Phonologie  Morphologie  Syntax

Lexikon  Grammatik (Regelwerk)

nicht-linguistisches Wissen  linguistisch bestimmte Bedeutung  Situation/Kontext

domänenspezifisches Wissen, Allgemeinwissen  Physisch, örtlich, zeitlich, Textzusammenhang, Intension

Interpretation
Semantic interpretation/analysis

Ziel: Bestimmung des semantischen Potenzials

- Umformung des Parse-Baumes in eine interne Repräsentation (z.B. Prädikatenlogik, Frames, ...)

- Zwei wesentliche Schritte:
  1. Lexikalische Semantik: Bestimmung der Bedeutung einzelner Wörter
     - Probleme: Homonymie, Polysemie (bank/bank), Synonyme (big/large), Antonyme (boy/girl, hot/cold)
     - Resourcen, z.B. WordNet (http://wordnet.princeton.edu/)
  2. Satzsemantik: Konstruktion der Gesamtbedeutung aus den Einzelbedeutungen (kompositionelle Semantik),
     - häufig anhand des Parse-Baums, erweitert mit sem. Kategorien (Nname, Aktionsbeschreibung, etc.) syntaktisch-semantisches Parsing
Discourse interpretation

Ziel: Von Satzsemantik zu Text-/Diskurssemantik

□ Nötige Wissensquellen (über ling. Wissen hinaus):
  ■ Domänenwissen (banking transaction)
  ■ Diskurswissen (satzübergreifend)
  ■ Weltwissen (*Common-sense knowledge*, Situationswissen)

□ Beispiel:
U: I would like to open a fixed deposit account.
S: For what amount?
U: Make it for 8000 Rupees.
S: For what duration?
U: What is the interest rate for 3 months?
S: Six percent.
U: Oh good then make it for that duration.
Zentrale Fragen der Diskursinterpretation.

- **Referenzauflösung**: Worauf wird Bezug genommen?
  - Ellipsen: ausgelassene Wörtern oder Phrasen
  - Anaphern: „John likes that blue car. He buys it."

- **Intentionserkennung**: Was will der Sprecher?
  - „Do you have the time?“ → will die Zeit wissen
  - „When is the last train to London?“ → will nach London

- **Informationsstruktur**: Was ist bekannt und was neu?

- **Rhetorische und narrative Struktur**: Wie ist der Bezug zum vorher Gesagten?

Benötigt Diskurskontext; siehe unten
Geht's auch ohne Syntax und Semantik?

Z.B. mit "keyword-spotting":

- durchsuchen der Benutzereingabe nach bestimmten Schlüsselworten, z.B. "Wetter", und generieren einer Antwort, die zum Schlüsselwort passt
- Einfach, aber skalierbar (grosse Zahl an Regeln)
- Grundlage vieler Chatterbots
  - Eliza (Weizenbaum, 1969), die Mutter aller chatterbots
  - Jabberwacky.com
  - Anna ([www.ikea.de](http://www.ikea.de))
- bereits bei einfachen syntaktischen Kniffen überfordert

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

- **Goal:**
  - produces understandable and appropriate output in natural language, possibly along with prosodic information

- **Input:**
  - some underlying non-linguistic representation of information

- **Output:**
  - text to speak, prosodic information

- **Knowledge sources required:**
  - linguistic knowledge (of language)
  - domain and world knowledge

Language Generation

- Simplest generation method is using templates, mapping representation straight to text template (with variables/slots to fill in).
  - loves(X, Y) → X “loves” Y
  - gives(X, Y, Z) → X “gives the” Y “to” Z

- Templates are very rigid, much more to NLG in general..
  - Better: “John eats the cheese and apple, then sneezes. He then laughs.”

- Getting good style involves working out how to map many facts to one sentence, when to use pronouns, when to use connectives like “then” etc.
Tasks in NLG

- **Content Planning**
  - Content Determination
  - Discourse planning (Document Structuring)

- **Micro-planning**
  - Aggregation
  - Lexicalisation
  - Referring Expression
  - Generation

- **Surface Realisation**
  - Linguistic Realisation
  - Structure Realisation

- **What to say, how to order and structure it**
- **How to break it up into sentences and words, how to refer to objects**
- **How to express things in terms of grammatically correct sentences**
1. Content Planning

Goals:
- to determine what information to communicate (content)
- to determine how to structure this information to make a coherent text/discourse

Results: *messages*, predefined data structures that...
- correspond to informational elements (units)
- collect underlying data in ways convenient for ling. expression

- Essentially, a domain-dependent expert-systems-like task
- Common approaches:
  1. based on observations about common utterance structures
  2. based on reasoning about discourse coherence and the purpose of the utterance
Content plan (aka. document plan)

- Tree structure with messages at its leaf nodes
- *Rhetorical Structure* (RST): distinction between *nucleus*, the central segment, and the *satellite*, the more peripheral one, and relations between them (e.g. elaboration, contrast, ...)
- Example from *WeatherReporter* system (Reiter et al.):

```
DOCUMENTPLAN

SATELLITE-01 [SEQUENCE]
  |  NUCLEUS
  |  cooler than average

SATELLITE-02 [SEQUENCE]
  |  NUCLEUS
  |  drier than average

SATELLITE-01 [ELABORATION]
  |  NUCLEUS
  |  average # raindays

SATELLITE-02 [ELABORATION]
  |  NUCLEUS
  |  rain spell

SATELLITE-01 [CONTRAST]
  |  NUCLEUS
  |  rain so far

SATELLITE-01 [CONTRAST]
  |  NUCLEUS
  |  rain amounts
```
2. Microplanning

Goal:
- convert a content plan into a sequence of sentence or phrase specifications

Tasks:
- Aggregation via *conjunction, ellipsis, or embedding*
  - Heavy rain fell on the 27th and [] on the 28th.
- Lexicalisation (choosing words)
- Reference: how to refer to entities
  - initially: full name, relate to salient object, specify location
  - subsequently: Pronouns, definite NPs, proper names, possibly abbreviated
3. Surface realisation

Goal:

to convert text specifications into actual text

Purpose:

to hide the peculiarities of English (or whatever the target language is) from the rest of the NLG system

Tasks:

- Structure realisation
  - Choose markup to convey document structure

- Linguistic realisation using specialized grammars
  - Insert function words
  - Choose correct inflection of content words
  - Order words within a sentence
  - Apply orthographic rules
Conducting natural dialogue
Hi, how are you today?

I'm fine. How can I help you?

I'm looking for a cookie recipe that my mom used to make.

Do you want to look it up in the net, or in your personal notes?

Let's check the net first.

Alright, here is a site about baking.

User

Speech Recognition

Text-to-Speech

Syntactic analysis and Semantic Interpretation

Response Generation

Dialogue Management

Phonetics, Phonology

Morphol., Syntax Semantics

Discourse Interpretation

Pragmatics, discourse

Us r

Phonetics,
Phonology

Morphol.,
Syntax
Semantics

Pragmatics,
discourse

MMI / SS06
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 for a *perfect* SLDS

- 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 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?
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 describes the illocutionary act

Austin (1962), Searle (1975)
Coherence over speech acts

“John hid Bill’s car keys. He likes spinach.”

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

- Informational approach: assume coherence relations
  - Hobbs (1979), for example:
    - Explanation: “John hid Bill’s car keys. He was drunk.”
    - Elaboration: “John bought an Acura. A big new SUV.”
  - constrained by discourse connectives („because“, ...)

- Intentional approach: infer plan-based intention of speaker
  - from discourse structure, based on cue words/phrases, prosody, and mental model of interpretation
  - 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 context**
  - 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 the conversation 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.
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 appropriate 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 to speak?
  - 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
  - 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!
Grounding

- Interlocutors are trying to establish **common ground**, a set of **mutual beliefs**
- Hearers must **ground** a speaker’s contribution by signaling understanding or agreement thru **feedback**
- 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?“

Allwood, 1976; Clark & Shaefer, 1989
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 accurate.

requires good NLP, users must be aware of possible words

natural, open, unpredictable, hard to model, requires NLP and complex dialogue management.
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
HCI Spotting: Not a Hard Task

I would like to visit a museum in Heidelberg.

Is there any other museum in town?

where is the modern art museum?

Um [], are there any good architectural

You can visit the modern art museum.

Sorry, there is no information on that.

In order to get there you have to turn around and walk across the square. Then take a right at the Hauptstrasse.
Something Missing Perhaps

E: So
[] um
[] how can I help you
["rise"]
E: Yeah [] sure []
Uhm []
do you need to get there
or []

B: Okay [],
I was gonna ask the system if
ähm []
it was open today.

E: Yes [],
they are actually open seven
days a week.

B: Okay [],
and what until what time are the
["rise"]

E: There are open from 9 until 6.
Relay

Efficient Hand-Over

- turn-taking
- turn-giving
- overlaps

You can visit the modern art museum.

Is there any other museum in town?

Um []
I would like to visit a museum in Heidelberg.

Um []
where is the modern art museum?

[‘rise] 
Sorry, there is no information on that.

[‘rise] 
In order to get there you have to turn around and walk across the square. Then take a right at the Hauptstrasse.

B: Um []
I was gonna ask the system if it was open today.

Ahm []
I asked if I could tour the Heidelberg castle.

B: Okay []
and what until what time are the

E: Um []
do you need to get there or []

E: So []
[‘um], how can I help you

[‘rise] 
E: Yeah [] sure []

[‘rise] 
Um []
do you need to get there or []

E: Yes [], they are actually open seven days a week.

B: Okay []
and what until what time are the

[‘rise] 
E: There are open from 9 until 6.
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

Jurafsky & Martin, 2000
Some dialogues systems

- **Commercial Systems:**
  - 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
  - ...

MMI / SS05
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

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

(Jurafsky & Martin, 2000)
Frame-based dialogue manager

- 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

- e.g. VoiceXML, SALT

- not good for dialogues involving negotiation, planning, mixed-initiative
Agent-based/plan-based control

- 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 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 server for Lille. Are you interested in one of these?

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

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

<table>
<thead>
<tr>
<th>source of obligation</th>
<th>obliged action</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁ Accept or Promise A</td>
<td>S₁ achieve A</td>
</tr>
<tr>
<td>S₁ Request A</td>
<td>S₂ address Request: accept or reject A</td>
</tr>
<tr>
<td>S₁ YNQ whether P</td>
<td>S₂ Answer-if P</td>
</tr>
<tr>
<td>S₁ WHQ P(x)</td>
<td>S₂ Inform-ref x</td>
</tr>
<tr>
<td>utterance not understood or incorrect</td>
<td>repair utterance</td>
</tr>
<tr>
<td>S₁ Initiate DU</td>
<td>S₂ acknowledge DU</td>
</tr>
<tr>
<td>Request Repair of P</td>
<td>Repair P</td>
</tr>
<tr>
<td>Request Acknowledgement of P</td>
<td>acknowledge P</td>
</tr>
</tbody>
</table>

→ hearer thinks that speaker wants him to do an act

\[ \alpha (JOHN) \]

JOHN INTEND \( \alpha (JOHN) \)

Deliberation

OBLIGED(JOHN, S, ADDRESS REQUEST(...))

effect

REQUEST(S,JOHN, \( \alpha (JOHN) \))

Figure 5: Traum & Allen (94) Model of Requests
Conversational agency (cont.)

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

<table>
<thead>
<tr>
<th>Level</th>
<th>Act Type</th>
<th>Sample Acts</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;UU</td>
<td>Turn-taking</td>
<td>take-turn</td>
</tr>
<tr>
<td></td>
<td></td>
<td>keep-turn</td>
</tr>
<tr>
<td>UU</td>
<td>Grounding</td>
<td>Initiate Repair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ack Continue</td>
</tr>
<tr>
<td>DU</td>
<td>Core Speech Acts</td>
<td>Inform YNQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accept Request</td>
</tr>
<tr>
<td>&gt;DU</td>
<td>Argumentation</td>
<td>Elaborate Q&amp;A</td>
</tr>
</tbody>
</table>

Table 2: Conversation Act Types

MMI / SS05
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)
Today: 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 determined by update rules
- dialogue manager = definition of the contents of the IS + description of update processes

(Traum & Larsson, 2003)
## Summary

<table>
<thead>
<tr>
<th>Features/dialogue control</th>
<th>State-based</th>
<th>Frame-based</th>
<th>Agent-based</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td>Single words or phrases</td>
<td>NL with concept spotting</td>
<td>Unrestricted NL</td>
</tr>
<tr>
<td><strong>Verification</strong></td>
<td>Explicit confirmation of each turn or at end</td>
<td>Explicit &amp; implicit confirmation</td>
<td>Grounding</td>
</tr>
<tr>
<td><strong>Dialogue Context</strong></td>
<td>Implicitly in dialogue states</td>
<td>Explicitly represented Control represented with algorithm</td>
<td>Model of System’s BDI + dialogue history</td>
</tr>
<tr>
<td><strong>User Model</strong></td>
<td>Simple model of user characteristics / preferences</td>
<td>Simple model of user characteristics / preferences</td>
<td>Model of User’s BDI</td>
</tr>
</tbody>
</table>
Recent Trends

- More comprehensive models
- Probabilistic and hybrid methods
- Multimodality
- Efficiency and robustness

... getting "modest"
- Domain-dependent instantiations
- Focus on measurable improvements for specific subproblems, e.g.
  - reduction of speech recognizer word errors
  - improved quality of translation