Language Technology II: Natural Language Dialogue

Dialogue Phenomena (1)

Ivana Kruijff-Korbayová
ivana.kruijff@dfki.de

Introduction

- A dialogue system engages in interaction with a human *as a participant/agent*

- So, it needs to have *a model* of what such interaction(s) looks like
  - → What needs to be modeled?
  - → How?

- Easy and pleasant interaction is an essential design aspect
  - → What characterizes easy and pleasant interaction?

Introduction

- **How do we know what conversations look like?**
  - Study human-human conversations
    - Ultimate benchmark for "naturalness"
    - BUT, dialogue systems have specific requirements
  - Study human-computer conversations:
    data collected with actual systems
    - Realistic, but confined to implemented functionality
  - Study simulated human-computer conversations
    data collected in Wizard-of-Oz studies, where a human
    simulates (part of) the system (given an algorithm)

Characteristics of Conversation

- Human-human conversation

- Human-computer interaction
  - Humans change their language use
  - Nevertheless, humans tend to treat computers as rational social agents and so (the "better" the interaction, the more) the essential characteristics remain

Characteristics of Dialogue

- Linguistic properties:
  - Cohesive devices:
    anaphora (pronouns, etc.), lexical cohesion, ellipses, fragments
  - Structure manifested in the participants' contributions

- Dialogue-specific phenomena
  - Turn-taking
  - Grounding: achieving mutual understanding
  - Error recovery (identifying and resolving misunderstandings)
  - Dialogue acts / speech acts; indirectness
  - Sequences of dialogue acts
  - Mixed initiative (either participant can be in control); collaboration

- Spontaneous speech characteristics

Cohesion & Dialogue Economy

- For reasons of economy, parts of structure are often "abbreviated" or omitted
  $\Rightarrow$ anaphoric reference, ellipsis and fragments

- The missing structure can normally be recovered from the previous utterances and from the context

- Keeping track of the context is essential to coherent dialogue

- Without modeling these phenomena, dialogue can appear unnatural or even go wrong

Cohesion & Dialogue Economy

U: Do any samples contain bismuth and ruthenium?
S: Yes.
U: Give me their overall analyses.

U: Do any samples contain bismuth and ruthenium?
S: No.
U: Then what do they contain?

A: What time is Twelfth Night playing tonight?
B: It starts at 8:10 p.m.
A: And Hamlet?

G: where are you in relation to the top of the page just now?
F: uh, about four inches
G: four inches?
F: yeah
G: where are you from the left-hand side?
F: about two.

Characteristics of Dialogue

- Linguistic properties:
  - Cohesive devices:
    anaphora (pronouns, etc.), lexical cohesion, ellipses, fragments
  - Structure manifested in the participants' contributions

- Dialogue-specific phenomena
  - Turn-taking
  - Grounding: achieving mutual understanding
  - Error recovery (identifying and resolving misunderstandings)
  - Dialogue acts / speech acts; indirectness
  - Sequences of dialogue acts
  - Mixed initiative (either participant can be in control); collaboration

- Spontaneous speech characteristics

Characteristics of Dialogue

- Spontaneous speech-related phenomena:
  - pauses and fillers („uh", „um", „..., like, you know,...")
  - prosody, articulation
  - disfluencies
  - overlapping speech

- Spontaneous conversation vs. practical dialogs: open-ended, topic drifts vs. goal/task-orientedness → joint activity

Today's Lecture

- Turn-taking

- Initiative and Collaboration

- Grounding

Turn Taking

Turn Taking

- Dialogue participants take turns (like in a game): A, B, A, B

- Dialogue turn = a continuous "contribution" to the dialogue from one speaker

- Though it is generally not obvious when a turn in natural dialog is finished, turn-taking appears fluid in normal conversation:
  - Minimal pauses between speakers (few hundred ms)
  - Less than 5% speech overlap

- How does it work?

Turn Taking Rules

- Conversation analysis (Sacks et al. 1974)

- When can one take a turn:
  - Transition-relevance place (TRP) --- places where the dialog/ utterance structure allows speaker shift to occur (typically at utterance boundaries, but also smaller units)
  - TRP signals include syntax (phrase boundaries), intonation, gaze, gesture; cultural conventions apply

- Who speaks next
  - At each TRP (current speaker A):
    - If A selected B as next speaker, B should speak
    - If A did not select the next speaker, then anyone may take a turn
    - If no-one else takes a turn, then A may (continue)
  - To get a turn if not selected, a speaker must "jump in" at a TRP

Turn Taking Rules

- Exercise:

- When do we get pauses or lapses?

- When do we get overlaps?

Turn Taking in Dialogue Systems

- Rigid:
  - System speaks until it completes it's turn, then waits
    - Problems: long turns; too long or too short waiting
  - System lets User to finish turn, then starts
    - Problem: wrong determination of end of user's turn

- With barge-in:
  - User barge-in: system allows an interruption
    - Open-mic: system listening all-the-time
      - Problem: talk directed at system vs. noise (vs. other talk); backchannel vs. taking the turn
    - Push-to-talk: user pushes button to take the turn
  - System barge-in:
    - When appropriate at all?

Initiative
&
Collaboration

Initiative

- Who is in control of the dialogue progression?
  - Being the one who's talking does not necessarily mean being in control, e.g., just answering a question

- How to decide whether to take initiative (move forward)

- Dialogue initiative vs. task initiative

- Human-human conversation: varied initiative patterns
  - Generally, mixed initiative: either participant can assume initiative, depending on knowledge, skills, situation, etc.

Initiative in Dialogue Systems

- Fixed initiative model (one participant in control)
  - System-initiative: system drives dialogue by prompting user; if done well, very efficient; otherwise may be unnatural and inconvenient for user
  - User initiative: user can do/say what wants when wants (if knows what); may be difficult for system, if too many possibilities; may work well in constrained domains

- Partial mixed initiative model
  - Allowing some constrained mixed initiative

Collaboration

- Conversation (and communication in general) is **a joint activity**
  - has a purpose (agreed on by the participants)
  - involves collaboration/cooperation
  - requires coordination of actions among agents
  - requires common ground

- Collaborating (being cooperative): helping each other to accomplish goals by, e.g.,
  - Cooperative interpretation beyond literal meaning (inference)
  - Cooperative answering
    - Complying with requests or directives when possible
    - Providing more information than requested (when it is relevant or useful), also correcting false presuppositions or misconceptions
    - Intensional answers and generalizations
  - Taking initiative when this helps to accomplish the joint activity

Collaboration & Inference

- Discourse participants "read between lines"
  - Conversational implicatures due to Gricean maxims
  - Informativity principle: try strongest interpretation

- and cooperatively accommodate implicit assumptions, if possible
  - Accommodation of presuppositions: adjusting common ground unless conflicting information
  - Resolving reference to entities in common ground

- If not possible, repair problems, if relevant

Maxims of Conversation

- ## Cooperative Principle (Grice 1975)

    "Make your contribution such as it is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged."

- ## Maxims of cooperative conversation

    – Maxim of quality

    – Maxim of quantity

    – Maxim of relevance

    – Maxim of manner

- ## Conversational implicatures arise based on

    – Assumed adherence to or blatant violation of maxims

Collaboration & Common Ground

- Entering a conversation, dialogue participants presuppose that there exists certain shared knowledge
  - → **common ground**
    - introduced by Stalnaker (1978) based on older family of notions: common knowledge (Lewis, 1969), mutual knowledge or belief (Schiffler, 1972)
    - stock of knowledge taken for granted, i.e. assumed to be known both by the Speaker and the Hearer
      sum of their mutual, common or joint knowledge, beliefs, and suppositions

- sources of common ground:
  - evidence about social, cultural communities people belong to, academic backgrounds, etc. (*communal common ground*)
  - direct personal experiences (*personal common ground*)

Grounding

Common Ground

- What does it mean to mutually know that p?

Common Ground

- The Byzantine generals problem:
    - actual mutual knowledge cannot be achieved in situations in which communication is fallible.

- Individual agents act on their individual beliefs or assumptions about what their CG is

- Stalnaker (speaker presupposition)
    - Discrepancies may lead to failures in communication
    - A context is close enough to being nondefective if the divergences do not affect the conversation issues

Language Technology II: Natural Language Dialogue
Ivana Kruijff-Korbayová

Grounding

(Clark 1996)

- **Principle of (joint) closure**: Agents performing a (joint) action require (CG) evidence, sufficient for current purposes, that they have succeeded in performing it

- The optimal evidence isn't usually the strongest, most economical and most timely evidence possible, because that may be too costly.

Grounding in Joint Activities

- Grounding = process of augmenting CG (by engaging in conversation)

- CG status is not all-or-nothing:
  - Graded evidence, Feeling of knowing
  - Also false consensus

- Grounding means establishing CG well enough for current purposes, at all levels of interpretation

Clark's Joint Action Ladder
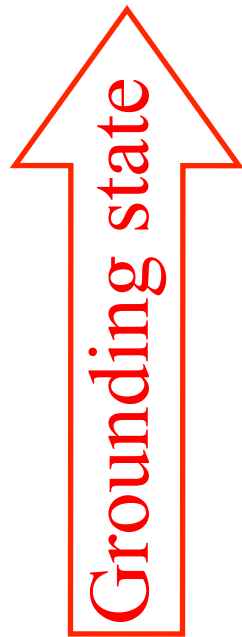
	Interpretation level	Speaker's actions	Hearer's actions
4	Intention	S proposes project w	H considers project w
3	Proposition	S signals that p	H recognizes that p
2	Signal	S presents signal s	H identifies signal s
1	Channel	S executes behavior t	H attends to behavior t

Grounding state

- Downward evidence; upward completion

Grounding Feedback

1. Continued attention

2. Relevant next contribution (presupposing understanding, e.g., answer to question, doing action)

3. Acknowledgement: asserting understanding with
    1. A backgrounded acknowledgement token (= continuer / backchannel), e.g., "Yeah", "mmm-mm", "I see", "uh-huh", or nodding;
    2. An assessment, e.g., that's great
    3. Unison completion

4. Demonstration (by paraphrasing, reformulating or cooperatively completing)

5. Display (verbatim repetition)

weaker

Grounding Problems

- H can be in one of the following states:
  - H did not notice that S's uttered U
  - H noticed, but did not hear it correctly
  - H heared it correctly, but did not understand it

- Grounding problems are due to
  - Lack of perception or understanding
  - Ambiguity
  - Conflicts (cannot link to CG)
  - Misunderstanding (discrepancies in CG)

- Clarification and repair strategies, e.g., ask for repetition, rephrase, clarify

Grounding & Contributions

- Contribution (to conversation):

  - **Presentation phase**: A presents a signal for B to understand. He assumes that, if B gives evidence e or stronger, he can believe that B understands what he means by it.

  - **Acceptance phase**: B accepts A's signal by giving evidence e' that she believes she understands what A means by it. She assumes that, once A registers e', he too will believe she understands.

$$C_i: \quad \textcircled{1} \xrightarrow{\textit{Presentation}} \textcircled{2} \xrightarrow{\textit{Acceptance}} \textcircled{3}$$
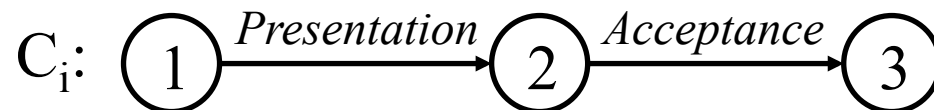
Expanded Contributions

- Expanded acceptance phase:
  - often when B has trouble understanding A's presentation ⇒ grounding subdialog, e.g., request for verification, clarification, repair

- Expanded presentation phase ("installments"):
  - often when A anticipates B may have trouble understanding (or when A unsure) ⇒ dividing up and (possibly) requesting feedback through grounding subdialog, e.g., request for confirmation

Expanded Contributions

- Packaging of Installments:
  What is the optimal size of a contribution?

  – The smaller the chunks with grounding feedback, the more certainty, but the communication takes longer

  – The larger the chunks, the more danger of snowball effect of a misunderstanding at some point

  – Working memory constraints

$\Rightarrow$ Variable size, depending on skills and purposes.

Verification Strategies in Systems

- **Pessimistic**: Immediate explicit feedback (and verification request)
  - S: Where do you want to go?
  - U: Hamburg.
  - S: Traveling to Hamburg. (OK?)
  - U: Yes.
  - S: When do you want to go?

- **Optimistic**: Delayed explicit feedback by summarizing at task end
  - …
  - S: So. Traveling from Saarbrücken to Hamburg on Monday June 6
  - …

- **Carefully optimistic**: Immediate "implicit" feedback by incorporating material to be grounded in the next system turn
  - S: Where do you want to go?
  - U: Hamburg.
  - S: And when do you want to go to Hamburg?

Grounding Strategies in Systems

- **Pessimistic**: Immediate explicit feedback (and verification request)
  - S: Where do you want to go?
  - U: Hamburg.
  - S: Traveling to Hamburg. (OK?)
  - U: Yes.
  - S: When do you want to go?

  Easy immediate repair
  Tedious

- **Optimistic**: Delayed explicit feedback by summarizing at task end
  - …
  - S: So. Traveling from Saarbrücken
  - …

  Efficient when correct
  Delayed repair more difficult

- **Carefully optimistic**: Immediate "implicit" feedback by incorporating material to be grounded in the next system turn
  - S: Where do you want to go?
  - U: Hamburg.
  - S: And when do you want to go to Hamburg?

Grounding Strategies in Systems

- Factors
  - ASR confidence(s) below/above a threshold
  - pragmatic plausibility (Gabsdil & Lemon 2004)
    - combining ASR confidence with task interpretation confidence (plausible actions in context)
  - context-adaptive strategies
    - dialogue progress so far

    → reinforcement learning: learn optimal strategies from data based on rewards for „good" dialogue and user satisfaction (Lemon et al. 2006)

To be continued …

Conversation Structure

Conversation Structure

- Global: the overall structure of an entire conversation

- Local: relations between pairs or sequences of turns

Global Structure

- Generic structure of a conversation:

  – Opening: "initialization" (establish contact, greetings, pleasantries)

  – Body: exchange about the subject matter(s) (accomplishing task(s), discussing topic(s)), sometimes a task is ended by a summary

  – Closing: winding down, farewell, breaking contact

- Conventions apply in all sections

Global Structure

Opening

↓

Task

↓

More? — *yes*

no

↓

Closing
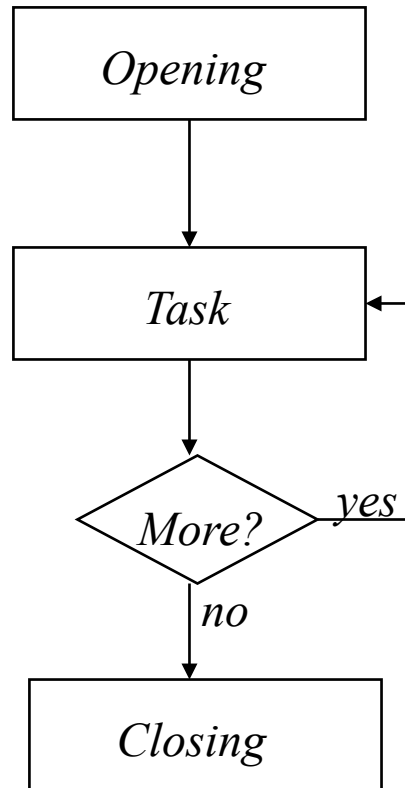
+ *task info*
+ *control options*

Global Structure
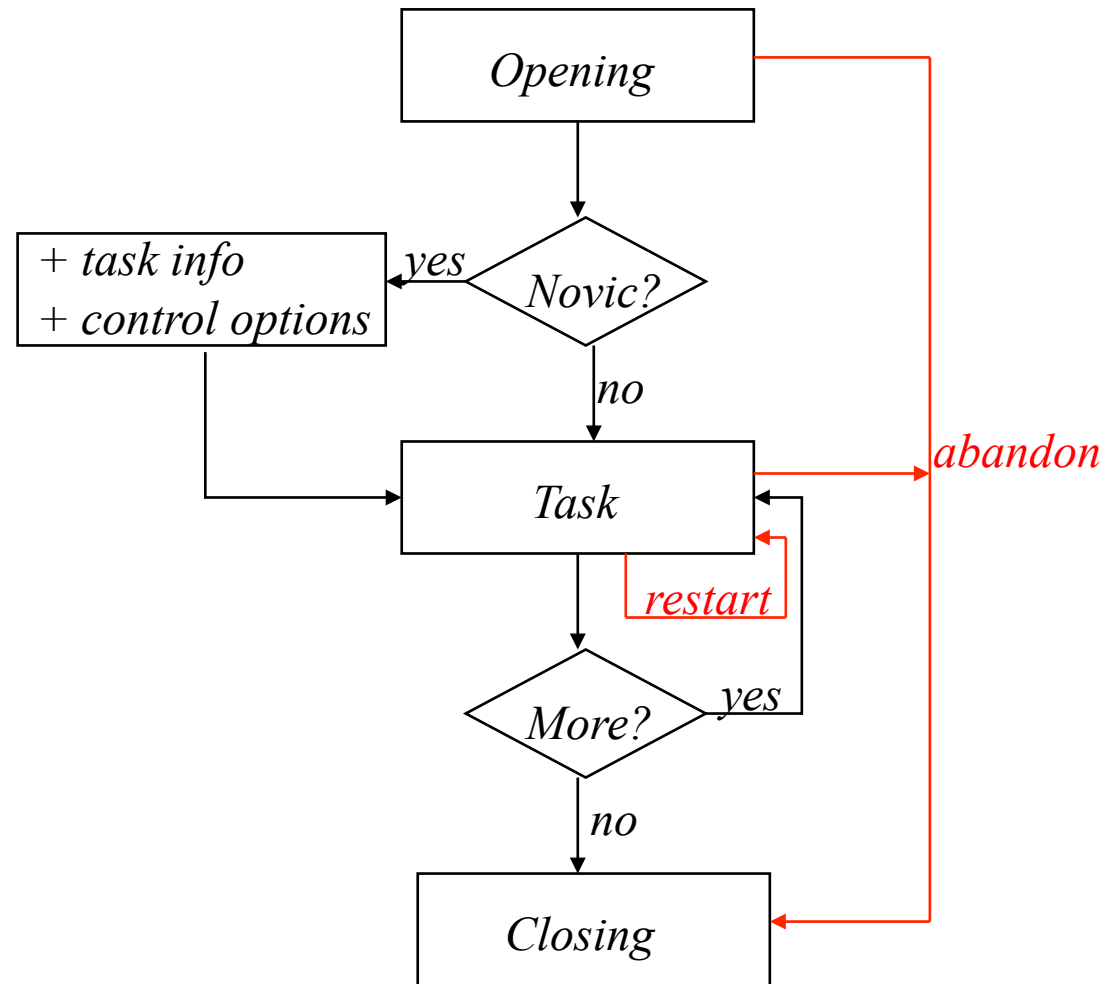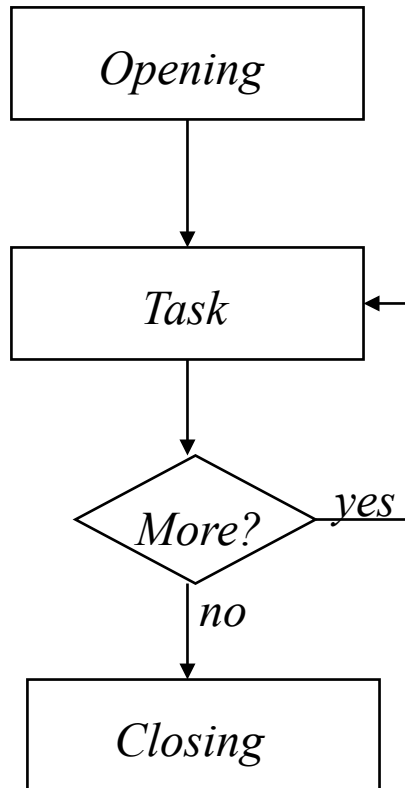
Local Structure

- ## Adjacency pairs (Schegloff 1968):

  – Adjacent turns

  – Produced by different speakers

  – Ordered: First^Second

  – Typed: particular First requires a particular Second

    - Greet-greet, ask-answer, request-grant, offer-accept, compliment-downplay, etc. $\Rightarrow$ preferences, expectations

- ## Insertion sequences: APs can be embedded

  – E.g., "sub-dialogue", misapprehension-correction, clarification

Local Structure: Preferences

- ## Significant silence (option 1 at a TRP)

  - ### If A selects B to speak next, but B doesn't

    - Then (assuming B has heard & understood)
      B's silence can be interpreted as a hesitation to give a dispreferred Second,
      e.g., B does not know the answer to A's question, B's response to A's offer or request is negative, etc.

- ## Other cases (silence at options 2 or 3 at a TRP) are just insignificant delays (pauses or lapses)

Local Structure: Insertions

- "Sub-dialogue":

  A: Where are you going?

  B: Why do you want to know?

  A: I thought I'd come with you.

  B: I'm going to the supermarket.

- Clarification:

  A: I'd like three sausages.

  B: Which ones? Merquez or Lyoner?

  A: Merquez.

  B: Here you go.

- Misapprehension-Correction:

  A: I'd like three sausages.

  B: Three pairs.

  A: No, three single pieces.

  B: OK.

A: When is the next train from SB to Hamburg?
B: The next train to Homburg Hauptbahnhof is at 1 p.m.
A: Hamburg, not Homburg.
B: Ah, Hamburg?
A: Yes.
B: OK, the next connection to Hamburg is at 3 p.m.

Dialogue Structure and Coherence

- Grosz and Sidner (1985)
  - Linguistic structure: discourse segments signaled by cues, e.g., discourse markers, prosody, etc.
  - Intentional structure: discourse segment purposes and relations between them (dominance, satisfaction-precedence); subdialogues vs. true interruptions
  - Attentional structure: entities in focus spaces corresponding to discourse segments; antecedents for anaphoric links; stack-model of focusing

Dialogue Economy

Dialogue Acts
(Conversation Acts)
(Dialogue Moves)

Speech Acts

- Dialogue acts evolved from speech acts

- Speech act theory: do things with words (Austin, Searle)

- Utterances are acts that change context
  - Locutionary act: the act of uttering the words with their semantic content
  - Illocutionary act: the communicative act the speaker performs in saying the words --> speech acts
  - Perlocutionary act: the act that occurs as a result of the utterance (e.g., making someone laugh, scared…)

Speech Acts

Assertive	S commits to sth being the case	Comment, suggest, swear, boast, conclude
Directive	S attempts to get H do sth	Ask, order, request, beg, invite, advise
Commissive	S commits to future course of action	Promise, plan, vow, bet, oppose
Expressive	S expresses psychological state	Thank, apologize, welcome, deplore
Declarations	S changes world	Resign, name, fire

Dialogue Acts/Moves

- Generalization of speech acts to conversational functions of utterances at various levels

- Various taxonomies, typically tuned for a specific task or domain

- Attempts at reusable schemes:
  - Conversation acts (Traum and Hinkelman 1992, Traum 1994)
  - DAMSL (1997)
  - DATE (2001)

DAMSL

- DAMSL: hierarchical general DA classification scheme for task-oriented dialogue
  - Forward looking function (like speech act)
  - Backward-looking function: relationship to previous utterance(s) by other speaker (including grounding)
  - Information level
    - Task: doing the task
    - Task management: talking about the task
    - Communication management: managing communication
    - Other
  - Communicative status: intelligibility, interpretability, self-talk...

DAMSL:Forward Looking Functions

- Statement   a claim

- Information request   a question
  - Check   a question confirming info

- Influence-on-addressee (= Searle's directives)
  - Open-option   a weak suggestion or list of options
  - Action-directive   command or instruction

- Influence-on-speaker   (= Searle's commissives)
  - Offer   offer to do something (subject to confirmation)
  - Commit   commitment to do something

- Conventional
  - Opening   greetings
  - Closing   farewell
  - Thanking   thanking and responding to thanks

Backward Looking Functions

- Agreement (speaker's attitude toward an action, plan, object, etc.)
  - Accept
  - Accept part
  - Maybe
  - Reject
  - Reject part
  - Hold

- Answer (answer to question)

- Understanding (whether speaker understood previous turn)
  - Signal-non-understanding
  - Signal-understanding
    - Acknowledgement (demonstrated by a continuer or assessment)
    - Repeat-paraphrase (demonstrated by a repetition or rephrase)
    - Completion (demonstrated by collaborative completion)

Dialogue Games/Sequences

- Some sequences of dialogue acts occur regularly, are even conventionalized; cf. adjacency pairs
    - Greeting-greeting
    - Question-answer
    - Compliment-downplayer
    - Accusation-denial
    - Offer-acceptance
    - Request-grant
    - …

- Obligation to respond

- Preferred responses

Dialogue Act Recognition

- How do we decide what DA a user input is, e.g., statement vs. info-request

- At first glance, this looks simple

  - Different syntax:
    - Yes-no-questions have subj-verb inversion
    - Statements have declarative syntax
    - Commands have imperative syntax

- However, the mapping between surface form and illocutionary act is not one-to-one

Dialogue Act Recognition

- For example, what looks like a yes/no question

    Can you give me a list of the flights from A to B

  Can be a polite form of directive or request

    Please give me a list of flights from A to B

- What looks like a statement

    And you said you wanted to travel next week

  Can actually be a question, used to verify sth. (but, intonation!)

Dialogue Act Recognition

- Another example of "indirectness":

  A: That's the telephone.

  B: I'm in the bath.

  A: OK.

- Can be paraphrased as follows:

  A requests B to perform action (answer phone)

  B states reason why he cannot comply (in bath)

  A undertakes to perform action (answer phone)

Dialogue Act Recognition

- Idiom-based model:
  - Literal meaning (direct speech act)
  - Idiomatic meaning (indirect speech act)

  the grammar would list idiomatic meanings for each construction, e.g., Can you X? would have request as one possible meaning

- Inferential model: indirect speech acts arrived at by inference

Automatic DA Recognition

- This is on of the tasks of the dialogue management module (see next lecture)

- Plan-based interpretation
  - Essentially the inference model, differences lie in amount and depth of actual reasoning
  - Symbolic
  - Requires hand-coding and domain-knowledge

- Cue-based recognition
  - Essentially derived from the idiom model
  - Using a combination of utterance features and context features (supervised machine learning methods)
  - Requires hand-annotated data

Conclusions

- Characteristics of human-human dialogue that also (should) hold for human-computer dialogue:
  - Turn-taking
  - Initiative and Collaboration
  - Global and local structure
  - Dialogue economy
  - Dialogue acts and indirectness
  - Grounding

- but they present challenges for modeling

Reading

- D. Jurafsky and J. H. Martin. Speech and Language Processing. Chapter 19. Prentice Hall. 2000.

- H. Clark. Using Language. Chapters 4 and 8. Cambridge University Press. 1996.

- DAMSL annotation manual

  http://www.cs.rochester.edu/research/cisd/resources/ damsl/RevisedManual/

Exercise

- Purpose of dial-a-dialogue assignment:
  - See for yourselves how and to what extent the presented dialogue phenomena are handled in a sample system
  - What would you want the/a system to be able to do
  - Think what it takes to achieve that behavior

- See course website:

  http://www.coli.uni-saarland.de/courses/late2/

  --> Schedule --> Exercise June 9 2005