Foundations of Speech and Language Technology:

Dialogue

Ivana Kruijff-Korbayová
korbay@coli.uni-sb.de
www.coli.uni-sb.de/~korbay/ → Teaching

Introduction

- Monologue vs. dialogue/conversation
- Written vs. spoken language
- Dialogue systems: NLI, ECA, HRI
- Application domains: database info, education/training, entertainment, assistance, collaboration
- What properties do conversations have?

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 the system, given an algorithm)

Conversation is a Joint Activity

- Collaboration
- Turns and turn-taking
- Speech acts
- Grounding
- Local and global structure
- Conversational Implicatures

Collaboration

- Conversation (and communication in general) is a joint activity:
  - It has a purpose (agreed on by the participants)
  - It involves cooperation: neither agent can accomplish the task alone
- Collaborating (being cooperative): helping each other to establish and accomplish goals by, e.g.,
  - Cooperative interpretation beyond literal meaning (inference)
  - Cooperative responses
    - 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

Conversation is joint action: Turns and Turn-Taking

Turn Taking

- Dialogue turn = a continuous contribution from one speaker
- Dialogue participants take turns (like in a game): A, B, A, B
- Though it is generally not obvious when a turn in natural dialog is finished, turn-taking appears fluid in normal conversation:
  - Speech overlap < 5%
  - Minimal pauses between speakers (< 100 ms)
- How does it work?

Turn Taking Rules

- Conversational analysis (Sacks et al. 1974)
- Transition-relevance places
  = places where the utterance structure allows speaker shift to occur (typically at utterance boundaries, but also smaller units)
  - TRPs are signaled by syntax (phrase boundaries), intonation, gaze, gesture; also cultural conventions apply
- 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 take it
  (To get a turn if not selected, a speaker must "cut in" at a TRP)

Turn-Taking Rules: Implications

- Turn switching:
  - There are explicit signals
  - Some turns require a response from another speaker:
    - question - answer
      A: What time is it?
      B: It's 5 p.m. / B: I don't have a watch.
    - greeting - greeting, compliment - downplayer, request - grant
  - Such pairs are called adjacency pairs (Schegloff, 1968)
  - Preferred vs. dispreferred second parts

Turn-Taking Rules: Implications

- Significance of silence: the absence of a second part of an adjacency pair can carry meaning, e.g.,

  Mom: What color is the shirt?

  Child: --

  Mom: Will you please put the book back?

  Child: --

  A (clearly addressing B): Hi, how are you?

  B (ostensively continuing their activity): --

Turn Taking Models
in Human-Computer Dialog

- Rigid: no interruptions
  - System speaks until it completes it's turn, then waits
    - How long to wait? (Is the user still there? Did s/he hear?)
  - System lets User to finish turn, then starts
    - How to determine end of user's turn? (Is the user finished?)
  - Problems: user must wait for end of system turn; misrecognition if user speaks too early; a pause may be mistaken for end of turn
- Flexible: with barge-in:
  - User barge-in: system speaks until it completes it's turn or until it detects an interruption
    - Open-mic: system listening all-the-time
      - Problems: talk vs. noise; also system's own talk is "noise"
    - Push-to-talk: user pushes button to take a turn
      - Problems: button pushed or released too early
  - System barge-in: When is a TRP? When appropriate at all?

Conversation is Action:
Speech Acts

Speech Acts

Speech act theory (Wittgenstein 1953, Austin 1962, Searle 1975)

- Each utterance is an action performed by the speaker

  I declare the Olympic Games open.

  I object!

  I bet 5 Euro it will snow.

- Every utterance constitutes three kinds of acts on context:
  - Locutionary act: the act of uttering the words with their semantic content
  - Illocutionary act (= speech act): the communicative act the speaker performs by saying the words, e.g., asking, answering, promising, etc.
  - Perlocutionary act: the effect of the utterance on the hearer's feelings, thoughts or actions, intended or unintended (e.g., making someone laugh, scaring someone, making someone stop, etc.)

Speech Act Types

| 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

Conversation is Joint Action: Grounding

Grounding

- Conversation is a collective act performed by the speaker and the hearer
- Speaker and hearer need to establish mutual knowledge / common ground (Stalnaker 1978)
- Principle of (joint) closure:

  Agents performing a (joint) action require evidence, sufficient for current purposes, that they have succeeded in performing it (Clark 1996)

Grounding

Contribution = joint linguistic act:

$$S \xrightarrow[\text{S: Presentation}]{Init(I, U_k, DU_i)} 1 \xrightarrow[\text{H: Acceptance}]{Ack(R, U_k, DU_i)} F$$
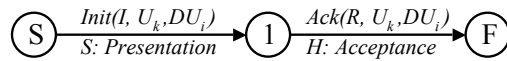
1:A=I: Move the boxcar to Corning
2:A=I: and load it with oranges
3:B=R: OK

Positive Grounding Feedback

- Continued attention
- Relevant next contribution
- Acknowledgement (nod or "continuer", e.g., uh-huh, yeah; or assessment, e.g., that's great)
- Demonstration (by paraphrasing, reformulating or cooperatively completing)
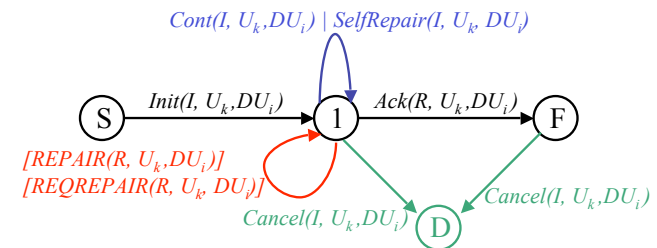- Display (verbatim repetition)

weaker

Grounding Problems

- Grounding problems are due to
  - Lack of perception or understanding
  - Ambiguity
  - Conflicts (differences in beliefs)
  - Misunderstanding (misinterpretation)
- Clarification and repair strategies, e.g., repair, ask for clarification, repetition, rephrase
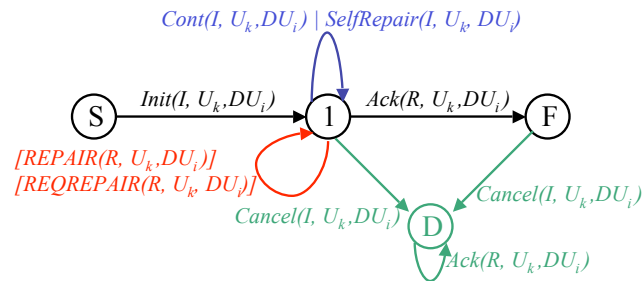
Grounding Acts

Traum (1999):

$Cont(I, U_k, DU_i)$ | $SelfRepair(I, U_k, DU_i)$

$S \xrightarrow{Init(I, U_k, DU_i)} 1 \xrightarrow{Ack(R, U_k, DU_i)} F$

[REPAIR(R, U_k, DU_i)]
[REQREPAIR(R, U_k, DU_i)]

$Cancel(I, U_k, DU_i)$ $Cancel(I, U_k, DU_i)$ D

1:A: Move the boxcar to Corning   Init(A,1,DU1)
2:A: and load it with oranges   Cont(A,2,DU1)
3:B: OK   Ack(B,3,DU1)

Grounding Acts

Traum (1999):

$Cont(I, U_k, DU_i) \mid SelfRepair(I, U_k, DU_i)$

$Init(I, U_k, DU_i)$ $Ack(R, U_k, DU_i)$

S → 1 → F

$[REPAIR(R, U_k, DU_i)]$
$[REQREPAIR(R, U_k, DU_i)]$

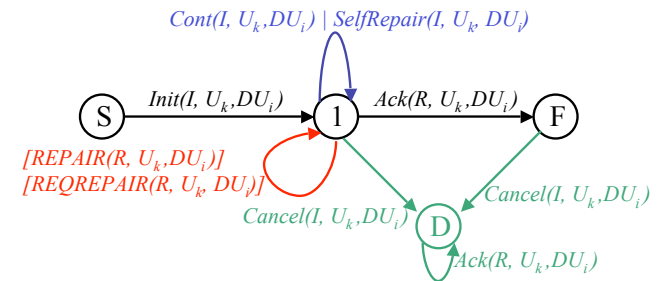$Cancel(I, U_k, DU_i)$ $Cancel(I, U_k, DU_i)$

D

$Ack(R, U_k, DU_i)$

(5) 1:A: Move the boxcar to Bath
   2:B: Bath?
   3:A: Oh, Corning.
   4:B: OK

Init(A,1,DU1)
ReqRepr(B,2,DU1)≈Init(B,2,DU2)
Ack(A,3,DU2) ≈ Repair(A,3,DU1)
Ack (R,4,DU1)

Grounding Acts

Traum (1999):

$Cont(I, U_k, DU_i) \mid SelfRepair(I, U_k, DU_i)$

$Init(I, U_k, DU_i)$ $Ack(R, U_k, DU_i)$

S → 1 → F

$[REPAIR(R, U_k, DU_i)]$
$[REQREPAIR(R, U_k, DU_i)]$

$Cancel(I, U_k, DU_i)$ $Cancel(I, U_k, DU_i)$

D

$Ack(R, U_k, DU_i)$
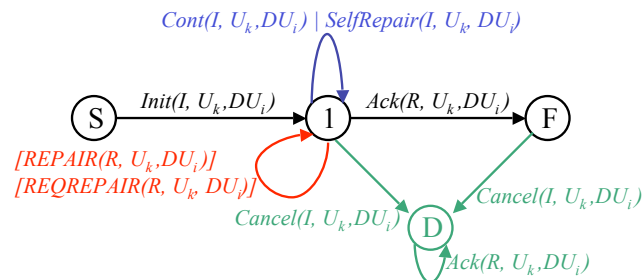
(4) 1:A: Move the boxcar to Bath
   2:B: To Corning
   3:A: Oh, sure.

Init(A,1,DU1)
Repair(B,2,DU1)≈Init(B,2,DU2)
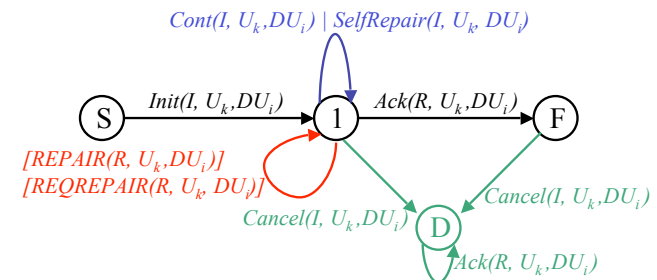Ack (A,3,DU2)

Grounding Acts

Traum (1999):

$Cont(I, U_k, DU_i) \mid SelfRepair(I, U_k, DU_i)$

$Init(I, U_k, DU_i)$ $Ack(R, U_k, DU_i)$

S → 1 → F

$[REPAIR(R, U_k, DU_i)]$
$[REQREPAIR(R, U_k, DU_i)]$

$Cancel(I, U_k, DU_i)$ $Cancel(I, U_k, DU_i)$

D

$Ack(R, U_k, DU_i)$

1:A: Move the boxcar to Bath
2:A: I mean Corning
3:B: OK

Init(A,1,DU1)
SelfRep(A,2,DU1)
Ack(B,3,DU1)

Grounding Acts

Traum (1999):

$Cont(I, U_k, DU_i) \mid SelfRepair(I, U_k, DU_i)$

$Init(I, U_k, DU_i)$ $Ack(R, U_k, DU_i)$

S → 1 → F

$[REPAIR(R, U_k, DU_i)]$
$[REQREPAIR(R, U_k, DU_i)]$

$Cancel(I, U_k, DU_i)$ $Cancel(I, U_k, DU_i)$

D

$Ack(R, U_k, DU_i)$

1:A: Move the boxcar to Bath
2:A: oh, forget that
3:B: OK

Init(A,1,DU1)
Cancel(A,2,DU1)
Ack(B,3,DU1)

Verification Strategies

- 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?
  - = very safe, but tedious!
- 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?
  - = User may still correct misunderstanding right away
- Delayed explicit feedback by summarizing at task end
  - S: So. Traveling from Saarbrücken to Hamburg on Monday June 6
  - = may be completely mislead, hard to recover from a misunderstanding

Handling Non-Understanding

When a system cannot recognize or interpret a user's input to a prompt:

- "Sorry I did not understand"
- Ask for repetition
- Repeat prompt as was
- Reformulate prompt
  - S1: When do you want to go?
  - S2: Say the date you want to depart.
- Provide suggestions/help
  - S3: You can say today or a date.
- A combination of these strategies is useful

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, requesting confirmation
  - Dialogue initiative vs. task initiative
- Basically, two models:
  - Fixed initiative model (one participant in control)
    - System-initiative: can drive dialogue as desired by prompting user, but may be unnatural and inconvenient for user (directive prompts)
    - User initiative: can do what & when wants (if knows what to do), but difficult for system, because it doesn't know what is coming
  - Mixed initiative model (either participant can assume initiative, depending on knowledge, skills, situation, etc.)
    - Typical in human-human conversation
   - How to decide whether to take initiative?

Conversation Structure

Conversation Structure

- Local: adjacency pairs = speech act sequences that occur regularly, are even conventionalized (obligation to respond, expected responses)
  - Greeting-greeting
  - Question-answer
  - Compliment-downplayer
  - Accusation-denial
  - Offer-acceptance
  - Request-grant
  - …
- Global: Common overall organization
  - e.g., telephone call(s):
    1. Enter a conversation with a summons-response adjacency pair
    2. Identify speakers
    3. Establish joint willingness to converse
    4. Raise the first topic (usually done by the caller)
  - Certain speech acts are expected depending on the place in conversation

Conversation Structure

Conversational Implicature
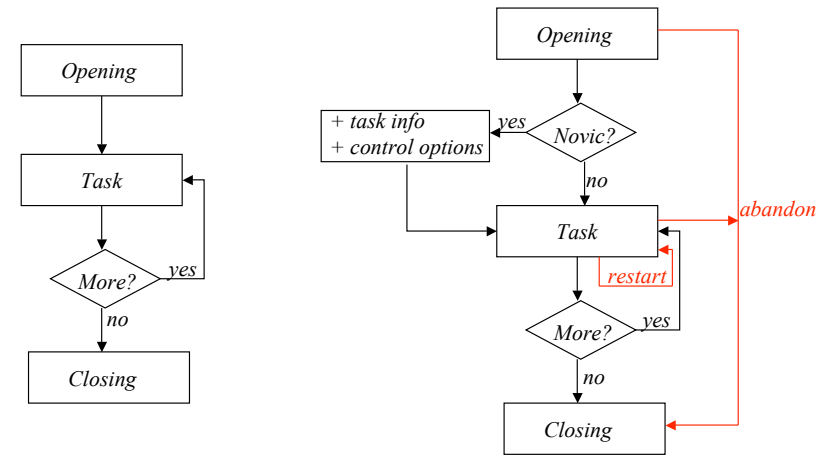
Maxims of Conversation

- Cooperative Principle
  - Maxim of quality
  - Maxim of quantity
  - Maxim of relevance
  - Maxim of manner
- Conversational implicatures arise based on
  - Assumed adherence to maxims
  - Blatant violation of maxims

Dialogue Economy

- Maxim of quantity, maxim of manner --> Dialogue utterances are often "reduced": they contain anaphora or ellipsis, or are fragmentary
- The full intended meaning can normally be recovered w.r.t. the previous utterances and from the situational context
- Keeping track of the context is therefore essential to coherent dialogue
- Without modeling these phenomena, dialogue can appear unnatural or even go wrong

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: about four inches
G: four inches?
F: yeah

Conclusions

- Characteristics of human-human dialogue that also (should) hold for human-computer dialogue:
  - Collaboration: purpose, cooperation, initiative
  - Turn-taking
  - Dialogue acts
  - Grounding
  - Global and local structure
  - Conversational implicature
- but they present challenges for modeling

Reading

Basic:
[1] D. Jurafsky and J. Martin (2009): Speech and Language Processing. Second Edition. Chapter 24.1, pp. 847-855.

Further:
[2] D. Jurafsky and J. Martin (2009): Speech and Language Processing. Second Edition. Chapter 24.2-8, pp. 847-894.

[3] B. Grosz, M. Pollack and C. Sidner (1989): Discourse. In Foundations of Cognitive Science. M. Posner (ed.). MIT Press. pp. 437-468.

[4] D. Traum (1999): Computational Models of Grounding in Collaborative Systems. AAAI Fall Symposium on Psychological Models of Communication. Pp. 124-131. http://people.ict.usc.edu/~traum/Papers/psych.ps