Natural Language Processing

Lecture 20

Today's Outline

New Links (thanks Steffi!)

Help Design a Conversational Agent!

- www.cs.cmu.edu/~sshriver/exp/kwds2/newkwds1.htm

Meet My Conversational Agents, and Friends

Review

- discourse structure
- why is dialogue different?

Finish Dialogue and Conversational Agents

- representing and interpreting dialog acts
- dialogue structure and coherence
- conversational agents

Some of My Conversational Agents

TOOT

RLDS (later known as NJFUN)

CobotDS

(Also analyzed data from Annie, ELVIS, HMIHY, W99)

CobotDS Transcript

C1S. **Hello, who is this?** *(CobotDS prompt)*
C1U. *Buster. (phone user response)*
C2S. **Secret password?**
C2U. *Eight three two two.* [C(a), successful login]
C3S. **Hi, Buster. What do you want to do?**
C3U. *Who.*
C4S. **You who. I am with Cocobot, Betwixt, and HFh.**
C4U. *Wave.* [C(b)]
C5S. **You wave.**
C5U. *Where.*
C6S. **You where. I am in The Cobotatory.**
HFh says How was the movie?, to you. [L(a)]
C6U. *Smile Betwixt.* [C(c)]
C7S. **You smile to Betwixt.**
C7U. *Say HFh.*
C8S. **What message would you like to pass?**
C8U. *Fantastic.* [C(d), smalltalk ASR grammar]
C9S. **You say Fantastic, to HFh. Betwixt waves to you.** [L(b)]
C9U. *Say.*
C10S. **What message would you like to pass?**
C10U. *What's going on in the MOO?.* [C(e), smalltalk ASR grammar]
C11S. **You say What's going on in the MOO, to everyone.**
C11U. *Summarize.*
C12S. **You summarize. Recently, HFh and Betwixt chatted.**
Betwixt and GreenBean bickered.
HFh and Natto had a friendly conversation.
I added to the noise. HFh made the most noise.
C12U. *Grammar (ASR rejects the utterance, and returns nothing).*
C13S. **Next. HFh says I am providing hoops wisdom to Betwixt, to you,**
and then, Betwixt says, HFh is pontificating on the Knicks again. , to you. [L(c),L(d)]

LambdaMoo Transcript

L1. HFh [to Betwixt]: I was able to see the game when I got home.
L2. Betwixt [to HFh]: I missed the first half because I was busy shoveling.
L3. Cobot turns to pick up the phone. [C(a)]
L4. Cobot begins talking to Buster! [C(a)]
L5. Cobot holds up a sign: Buster passes on a wave from the phone. [C(b)]
L6. HFh pats Cobot on the head.
L7. HFh [to Cobot]: phone: How was the movie? [L(a)]
L8. Betwixt [to HFh]: Overall, it's one of their better stretches this year.
L9. Cobot smiles at Betwixt. [C(c)]
L10. Cobot [to Betwixt]: That was from Buster. [C(c)]
L11. Betwixt [to Cobot]: phone: wave [L(b)]
L12. Cobot [to HFh]: Fantastic [C(d)]
L13. Cobot [to HFh]: That was from Buster. [C(d)]
L14. HFh nods to Betwixt.
L15. HFh [to Betwixt]: But it can't last. The soap opera always takes over at some point.
L16. Cobot holds up a sign: Buster says, 'What's going on in the MOO' from the phone. [C(e)]
L17. HFh [to Cobot]: phone: I am providing hoops wisdom to Betwixt. [L(c)]
L18. Betwixt [to Cobot]: phone: HFh is pontificating on the Knicks again. [L(d)]

Try Other Conversational Agents!

Academic

- CMU Movieline: 412 268 1185
- CMU Communicator (travel): 412 268-1084 or -5144
- MIT Jupiter (weather): 1-888-573-8255

Commercial

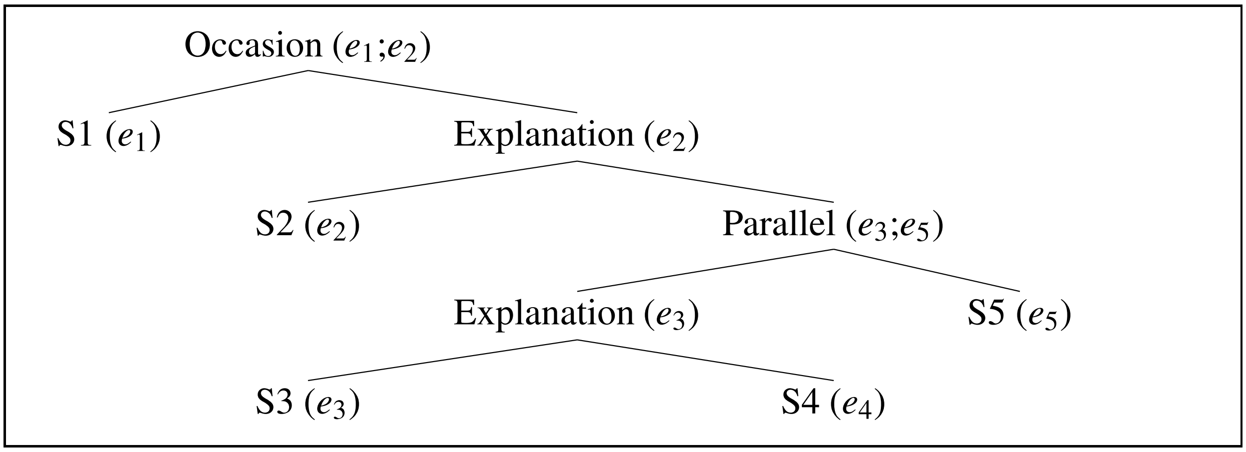
- Nuance: 1-888-NUANCE-8 (demo)
- Speechworks: United Flight Info 1.800.824.6200
- Tellme: 1 800 555 tell

Review: Discourse Structure

A (tree) structure of discourse segments (locally coherent utterances)

Coherence Relation-Based Example

- John went to the bank to deposit his paycheck. (S1)
- He then took a train to Bill's car dealership. (S2)
- He needed to buy a car. (S3)
- The company he works for now isn't near any public transportation. (S4)
- He also wanted to talk to Bill about their softball league. (S5)



Review: Dialogue is Different

Turns and utterances

- speakers know how to take turns (*who* should talk next, and *when* they should talk)
- conversation analysis provides an approach to turn-taking
- transition-relevance places are typically at utterance boundaries, but utterance segmentation is a difficult problem

Grounding

- a hearer must ground a speaker's utterances (by making it clear that (believed) understanding has occurred), or else indicate that a grounding problem occurred

Implicature

- conversational implicature means a particular class of licensed inferences (that the speaker expects the hearer to draw)
- Grice's maxims for conversation explain how hearers draw such inferences

Review: Dialogue Acts

Austin (1962) observed that dialogue utterances are a kind of speaker *action*, or <u>speech act</u>.

The utterance of any sentence in a real situation constitutes three kinds of act: locutionary, illocutionary, and perlocutionary.

Searle uses the term speech acts to classify illocutionary acts (1975).

DAMSL

A recent computational, expanded, hierarchical dialogue act tagging scheme (Dialogue Act Markup in Several Layers)

Forward looking level (draws from Searle/Austin speech acts)

- Statement: a claim made by the speaker
- Info-Request: a question by the speaker
  - Check: a question for confirming information
- Influence-on-addressee: Searle's directives
  - Open-option: a weak suggestion or listing of options
  - Action-directive: an actual command
- Influence-on-speaker: Austin's commissives
  - Offer: speaker offers to do something (subject to confirmation)
  - Commit: speaker is committed to doing something
- Conventional: other
  - Opening: greetings
  - Closing: farewells
  - Thanking: thanking and responding to thanks

DAMSL (continued)

Backward looking level (draws from grounding, adjacency pairs, . . .)

- Agreement: speaker's response to previous proposal
  - Accept
  - Accept-part
  - Maybe
  - Reject-part
  - Reject
  - Hold
- Answer: answering a question
- Understanding: whether speaker understood previous
  - Signal-non-understanding
  - Signal-understanding
    - Ack: continuer or assessment
    - Repeat-rephrase: repetition or reformulation
    - Completion: collaborative completion

Biased towards task-oriented dialogue, so also Switchboard (shallow) DAMSL

A Tagged Dialogue using DAMSL

[assert]	C_1:	...I need to travel in May.
[info-req,ack]	A_1:	And, what day in May did you want to travel?
[assert, answer]	C_2:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
[info-req,ack]	A_2:	And you're flying into what city?
[assert,answer]	C_3:	Seattle.
[info-req,ack]	A_3:	And what time would you like to leave Pittsburgh?
[check,hold]	C_4:	Uh hmm I don't think there's many options for non-stop.
[accept,ack]	A_4:	Right.
[assert]		There's three non-stops today.
[info-req]	C_5:	What are they?
[assert, open-option]	A_5:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
[accept,ack]	C_6:	OK I'll take the 5ish flight on the night before on the 11th.
[check,ack]	A_6:	On the 11th?
[assert,ack]		OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
[ack]	C_7:	OK.

Dialogue Act Tagging Algorithms

Sometimes there are obvious mappings from surface forms to dialogue acts

- STATEMENT *I don't care about lunch.*
- ACTION-DIRECTIVE *Show me the flights from Pittsburgh.*

But there are also many violations, or Indirect Speech Acts

- ACTION-DIRECTIVE *Can you show me the flights from Pittsburgh?*
- ACTION-DIRECTIVE *It's hot in here.*

A continuum of solutions

- plan inference model (derive only one of literal or indirect meaning)
- cue or idiom model (both literal and indirect meanings)

The Inferential Approach: Searle

Can you give me a list of the flights from Atlanta?

1. X has asked me a question about whether I have the ability to give a list of flights.
2. I assume that X is being cooperative in the conversation (in the Gricean sense) and that his utterance therefore has some aim.
3. X knows I have the ability to give such a list, and there is no alternative reason why X should have a purely theoretical interest in my list-giving ability.
4. Therefore X's utterance probably has some ulterior illocutionary point.
5. A preparatory condition for a directive is that the hearer have the ability to perform the directed action.
6. Therefore, X has asked me a question about my preparedness for the action of giving X a list of flights.
7. Furthermore, X and I are in a conversational situation in which giving lists of flights is common and expected.
8. Thus, in the absence of any other plausible illocutionary act, X is probably requesting me to give him a list of flights.

Plan Inference / Recognition

Making the inferential approach computational

- an AI planning (STRIPS) inspired model
- Allen, Cohen, Perrault in the 70's, and others since

Domain Acts

- BOOK-FLIGHT(A,C,F)

Speech Acts

- INFORM(S,H,P)
- INFORMIF(S,H,P)
- REQUEST(S,H,ACT)

Surface Acts

- SURFACE-REQUEST(S,H,ACT)

Plan Inference (continued)

Plan Inference Heuristics

- Action-Effect Rule
- Precondition-Action Rule
- Body-Action Rule
- Know-Desire Rule
- Extended Inference Rule

See page 737 to trace indirect speech act interpretation of *Can you give me a list of flights from Atlanta*, e.g.

- S.REQUEST(S,H,InformIf(H,S,CanDo(H,Give(H,S,LIST))))
- output: REQUEST(S,H,Give(H,S,LIST))

Plan Inference Since

Multi-level analyses (beyond question-answering)

Formalization

Automatic acquisition

Probabilities rather than heuristics

Collaborative and multi-agent approaches

Efficiency

See "Techniques for Plan Recognition", *User Modeling and User-Adapted Interaction*, 10-year Anniversary Issue

Cue-Based Interpretation

Less sophisticated, data-driven, more efficient alternative to plan inference.

Multiple sources of knowledge provide dialogue act cues

- words and collocations
- prosody
- conversational structure
- combinations of the above

Words and Collocations

- *please* usually signals REQUEST
- word n-grams for each dialogue act (e.g., *so you*, *sounds like* are common REFORMULATION bigrams)

Prosody

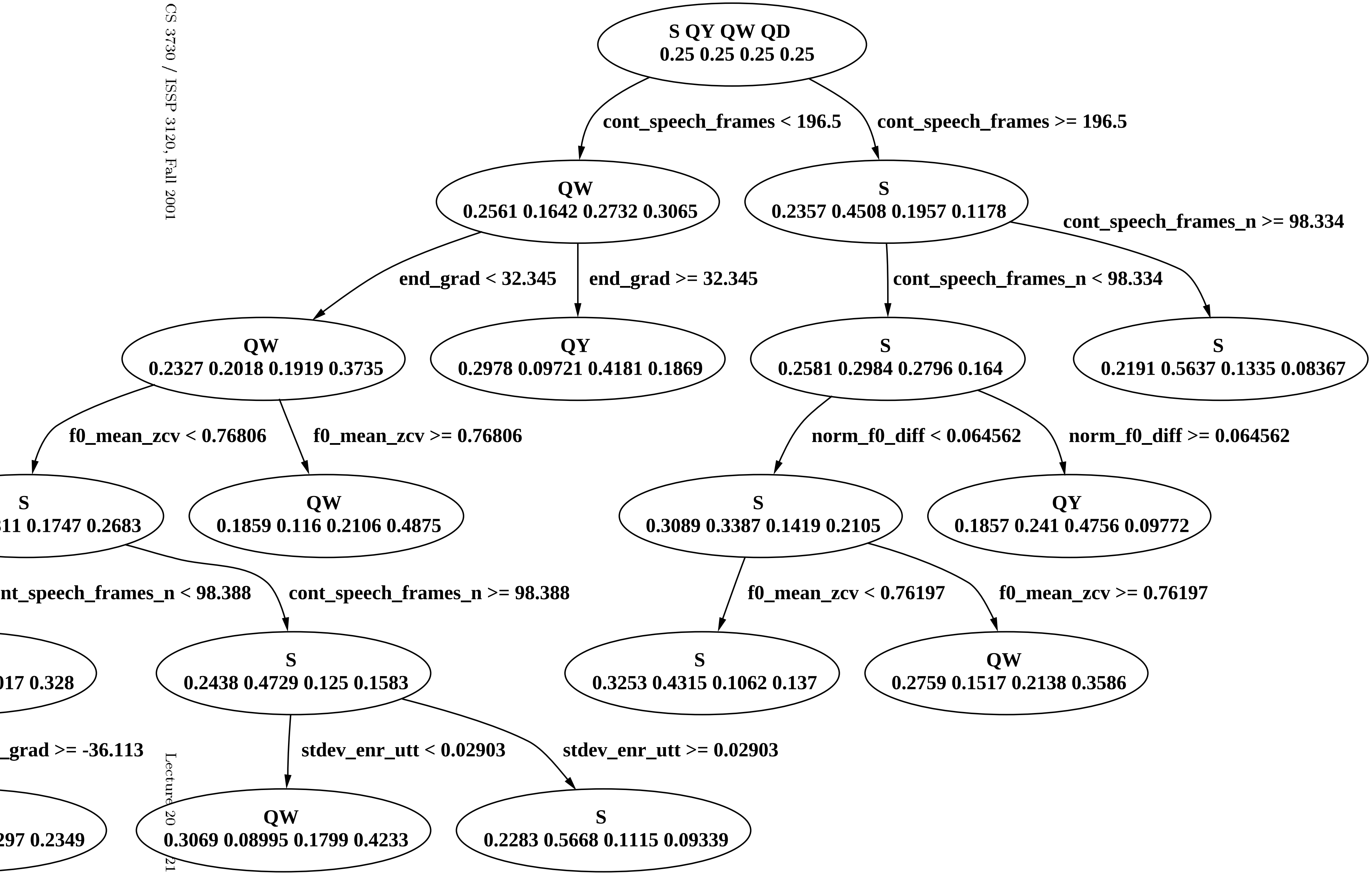
Decision trees for using prosody to classify speech acts

Speech Acts (classes)

- STATEMENT (S)
- YES-NO QUESTIONS (QY)
- WH-QUESTIONS (QW)
- DECLARATIVE-QUESTIONS (QD)

Prosody (features)

- pitch or fundamental frequency (F0) contour
- energy or loudness
- temporal duration


S QY QW QD
0.25 0.25 0.25 0.25
cont_speech_frames < 196.5
cont_speech_frames >= 196.5
QW
0.2561 0.1642 0.2732 0.3065
S
0.2357 0.4508 0.1957 0.1178
end_grad < 32.345
end_grad >= 32.345
cont_speech_frames_n < 98.334
cont_speech_frames_n >= 98.334
QW
0.2327 0.2018 0.1919 0.3735
QY
0.2978 0.09721 0.4181 0.1869
S
0.2581 0.2984 0.2796 0.164
S
0.2191 0.5637 0.1335 0.08367
f0_mean_zcv < 0.76806
f0_mean_zcv >= 0.76806
norm_f0_diff < 0.064562
norm_f0_diff >= 0.064562
S
QW
0.1859 0.116 0.2106 0.4875
S
0.3089 0.3387 0.1419 0.2105
QY
0.1857 0.241 0.4756 0.09772
cont_speech_frames_n >= 98.388
f0_mean_zcv < 0.76197
f0_mean_zcv >= 0.76197
S
0.2438 0.4729 0.125 0.1583
S
0.3253 0.4315 0.1062 0.137
QW
0.2759 0.1517 0.2138 0.3586
stdev_enr_utt < 0.02903
stdev_enr_utt >= 0.02903
QW
0.3069 0.08995 0.1799 0.4233
S
0.2283 0.5668 0.1115 0.09339

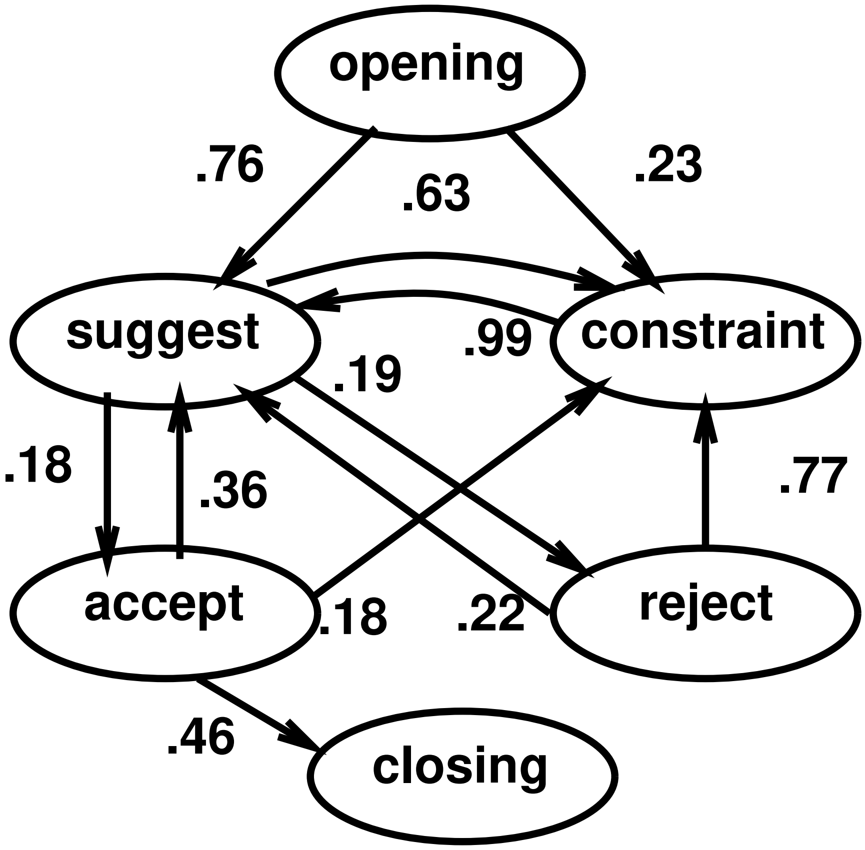
Conversational Structure

Capture observations such as *yeah* is typically an AGREEMENT after a PROPOSAL but a BACKCHANNEL after an INFORM

N-grams for dialogue act sequences (generalization of adjacency pairs)

Previous tagging algorithms (e.g. TBL)

A dialogue act HMM:



Structure / Coherence in Dialogue

<u>Intentional</u> versus previous <u>informational</u> approach.

Discourse structure a la Grosz and Sidner (1986)

- linguistic structure
- intentional structure
- attentional state

Intentional structure

- discourse purpose
- discourse segment purpose (DSP)

Two "coherence" relationships

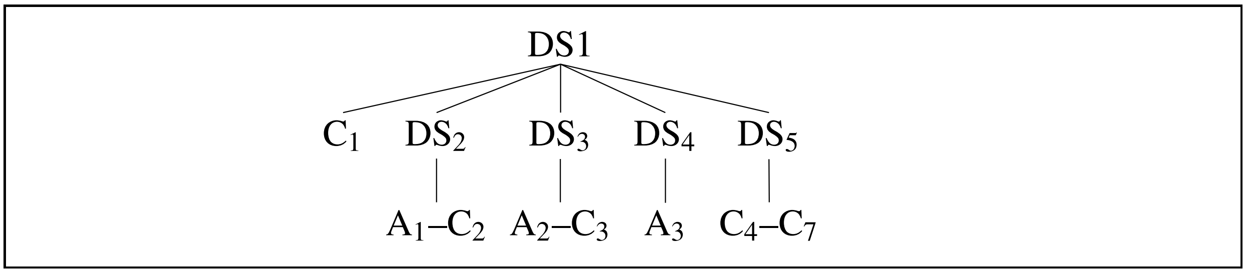
- <u>dominance</u>: DSP1 dominates DSP2 if satisfying DSP2 is intended to provide part of the satisfaction of DSP1
- <u>satisfaction-precedence</u>: DSP1 satisfaction-precedes DSP2 if DSP1 must be satisfied before DSP2

Example

C_1:	. . . I need to travel in May.
A_1:	And, what day in May did you want to travel?
C_2:	OK uh I need to be there for a meeting that's from the 12th to the 15th.
A_2:	And you're flying into what city?
C_3:	Seattle.
A_3:	And what time would you like to leave Pittsburgh?
C_4:	Uh hmm I don't think there's many options for non-stop.
A_4:	Right. There's three non-stops today.
C_5:	What are they?
A_5:	The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.
C_6:	OK I'll take the 5ish flight on the night before on the 11th.
A_6:	On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.
C_7:	OK.

- I1: (Intend C (Intend A (A find a flight for C)))
- I2 (A1-C2): (Intend A (Intend C (Tell C A departure date)))
- I3 (A2-C3): (Intend A (Intend C (Tell C A destination city)))
- I4 (A3): (Intend A (Intend C (Tell C A departure time)))
- I5 (C4-C7): (Intend C (Intend A (A find a nonstop flight for C)))

Example (continued)



- I1: (Intend C (Intend A (A find a flight for C)))
- I2: (Intend A (Intend C (Tell C A departure date)))
- I3: (Intend A (Intend C (Tell C A destination city)))
- I4: (Intend A (Intend C (Tell C A departure time)))
- I5: (Intend C (Intend A (A find a nonstop flight for C)))

I1 dominates I2, I3, I4, I5

I2 satisfaction-precedes I5

I3 satisfaction-precedes I5

Intentions

Can be implemented using AI-planning formalisms

Recognition algorithms range from inferential to cue-based

Integration of Informational/Intentional coherence

- Moore and Pollack (1992)

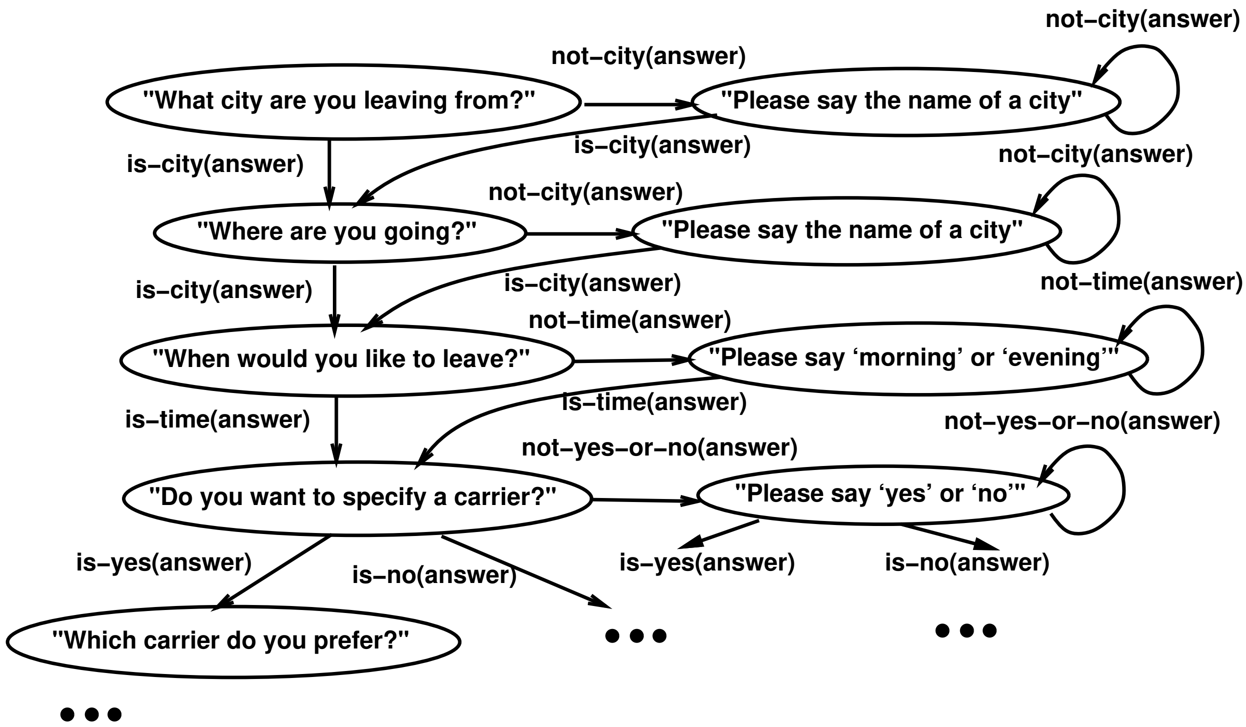
Dialogue Managers

Needed to build conversational agents

Continuum of approaches

- pattern-matching
- finite-state automaton
- MDP
- frames
- production rules
- plan-based

FSA Dialogue Manager



TOOT used this approach.

Compact for system initiative.

Frame-Based Dialogue Manager

Slot: Question

- DepartureLocation: “From what city are you leaving?”
- DepartureTime: “When would you like to leave?”
- Airline: “”
- . . .

Still has problems accounting for user initiative, grounding, etc.

Conversation-Act Dialogue Manager

Dialog arises from the interaction of two or more agents, as they cooperate to achieve a set of goals

- interaction is viewed in terms of "conversation acts"
- action choice is based on goals and rules of behavior

Example based on Rochester's TRAINS System

```
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
```

Example

Client: *I want to go to Pittsburgh in May.* (INFORM1)

System state

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

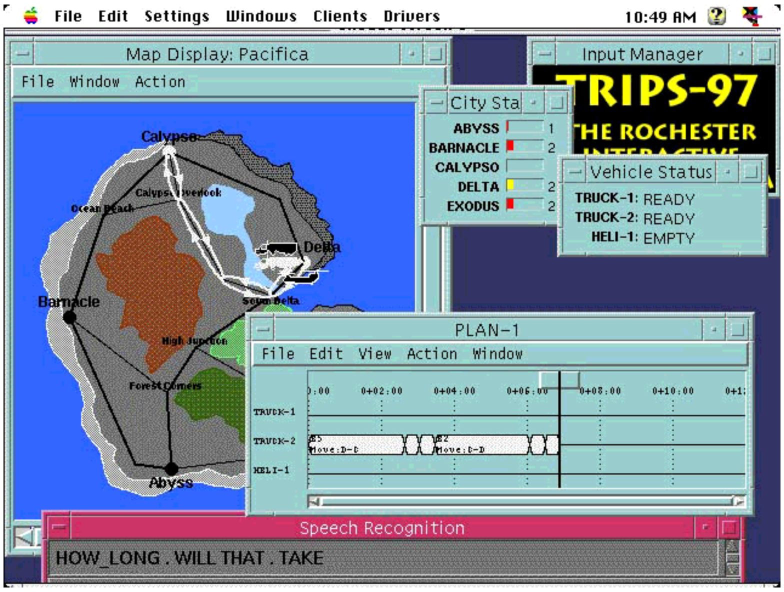
System state update (to address grounding, goals)

- Intended speech acts: REQUEST-INFORM1, ACKNOWLEDGE1

System: *And, what day in May did you want to travel?*

TRIPS Screen Dump

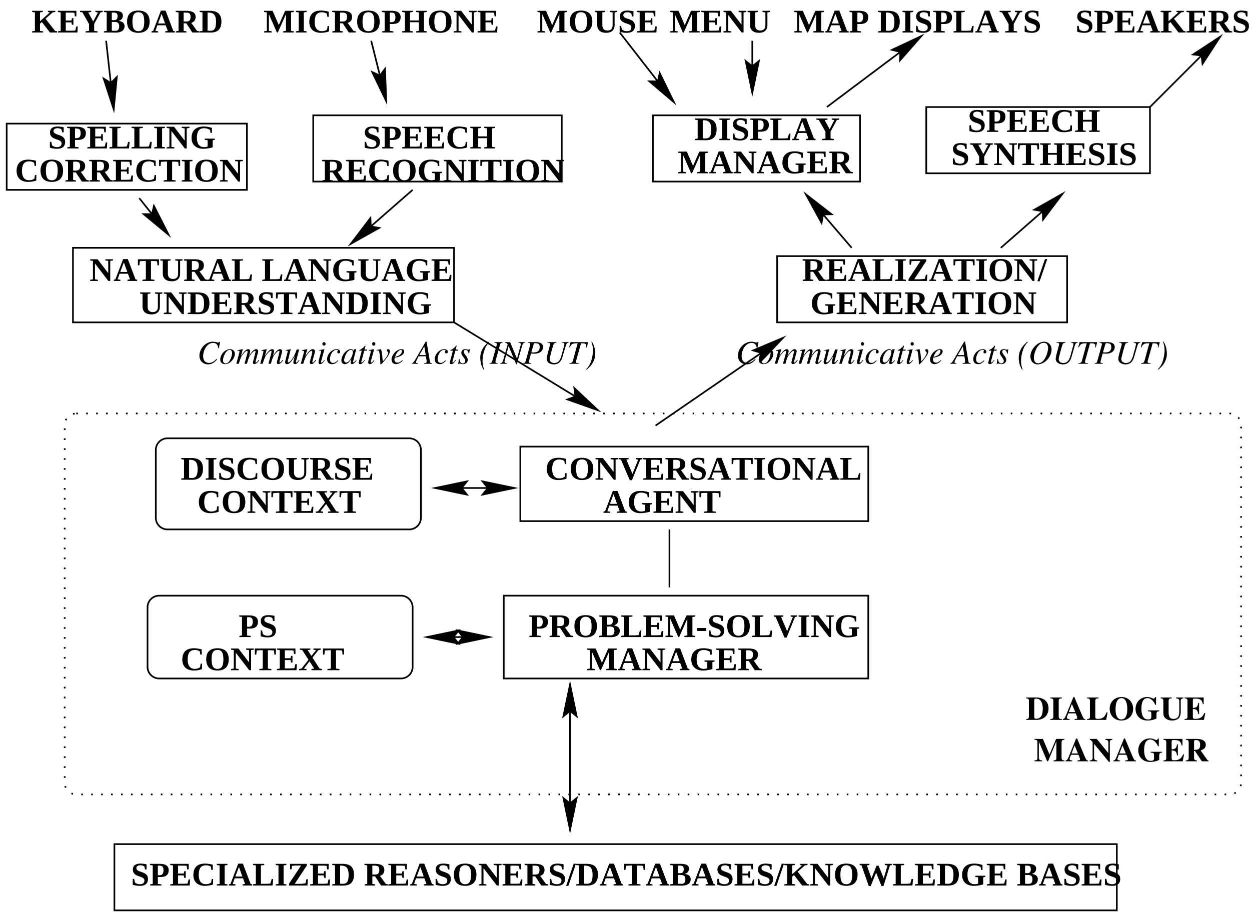
TRAINS became TRIPS



TRIPS Multimodal Dialog

	Speech	Screen
U:	Hello.	
T:	Hi there.	
U:	Show me a map of Pacifica.	
T:	Ok.	*Display map*
U:	Where are the people?	
T:	There are 2 groups of people at E, 2 at C, 2 at B, and 1 group of people at A.	*Display people chart*
U:	And where are the transports?	
T:	There are 2 trucks and 1 helicopter at D.	*Display vehicle chart*
U:	Use a truck to get the people from C to D.	
T:	Your wish is my command.	*Display plan* *Update other displays*
U:	How long will that take?	
T:	It will take 6 hours and 42 minutes.	
U:	What if we went along the coast instead?	
T:	That option would take 10 hours 27 minutes.	*Display 2nd plan window*
U:	Forget it.	
T:	Ok.	*Delete 2nd plan window*
U:	Use the other truck to get people from E to D.	
T:	Ok.	*Enhance displays*
U:	Let's use the helicopter instead.	
T:	Ok.	*Change displays*
...		

TRIPS Architecture



Communicative Act Analysis

U:	Hello.	*Greeting*
T:	Hi there.	*Greeting*
U:	Show me a map of Pacifica.	*Direct Request*
T:	Ok.	*Acceptance*
U:	Where are the people?	*Question*
T:	There are 2 groups of people at E, 2 at C, 2 at B, and 1 group of people at A.	*Answer*
U:	And where are the transports?	*Question*
T:	There are 2 trucks and 1 helicopter at D.	*Answer*
U:	Use a truck to get the people from C to D.	*Direct Request*
T:	Your wish is my command.	*Acceptance*
U:	How long will that take?	*Question*
T:	It will take 6 hours and 42 minutes.	*Answer*
U:	What if we went along the coast instead?	*Question*
T:	That option would take 10 hours 27 minutes.	*Answer*
U:	Forget it.	*Direct Request*
T:	Ok.	*Acceptance*
U:	Use the other truck to get people from E to D.	*Direct Request*
T:	Ok.	*Acceptance*
U:	Let's use the helicopter instead.	*Suggestion*
T:	Ok.	*Acceptance*

...

Circuit Fix-It Shop

1. S: This is the circuit-fixit shop. How may I help you?
2. U: RS111 is not working.
3. S: I am familiar with that circuit.The LED is supposed to be displaying alternately flashing one and seven.
4. U: *There is no wire on one zero four.*
5. S: *There is supposed to be a wire between connector one zero four and connector one zero two.*

6: U: The circuit is working.

7. S: Glad to have been of assistance. Good-bye.

Dialog occurs when theorem proving fails

- utterance 2: initiates goal inference
- utterance 3: user model guides response
- utterances 5, 7: unexpected subdialog switches tracked

Analysis of the Plan-Based Approach

Advantages

- generality
  - communicative dimension
  - task dimension
- unifying framework
  - language and reasoning
  - speech and other modalities

Disadvantages

- computational complexity
- (traditionally) knowledge-based
- restricted management of uncertainty

Recent Trends

Efficiency and robustness

Stochastic and hybrid methods

Domain-dependent instantiations

Focus on measurable improvements for specific subproblems

- reduction of ASR word errors
- improved quality of translation

Examples

VERBMOBIL (speech to speech translation)

- addition of domain-dependent communicative acts
- communicative act prediction using language modeling methods
  - 74% of next acts are in top 3 predictions
- statistical predictions narrow down knowledge-based predictions
- recognized communicative acts help resolve ambiguity

Similar projects at CMU

Designing Dialogue Systems

Human-human transcripts

WOZ

Iterative design

Learning

Evaluating Dialogue Systems

PARADISE: PARAdigm for DIalogue System Evaluation (Walker, Litman, Kamm, Abella)

Collect experimental data

- Kappa or some other success measure
- Cost measures that can be measured reliably
- User satisfaction ratings per dialogue interaction

Use multiple linear regression to model User Satisfaction from kappa and costs

$$\text{UserSatisfaction} = (\alpha * \mathcal{N}(\kappa)) - \sum_{i=1}^{n} w_i * \mathcal{N}(c_i)$$

Test Fit (Amount of Variance Explained)

Determine Which Predictor Variables Contribute

Obtain Weights for Predicting US, given Predictor Values

Sample PARADISE Results

TOOT Estimated Performance Function:

$$\mathrm{P} = .45\mathcal{N}(Succ.) + .35\mathcal{N}(ASRAcc.) - .42\mathcal{N}(BargeIns)$$

Generalized TOOT and ELVIS Performance Function:

$$\mathrm{P} = .23\mathcal{N}(Succ.) + .43\mathcal{N}(ASRAcc.) - .21\mathcal{N}(Time)$$

Also

See slides 10, 11, 18, 19, 20, 24 from Allen's Dialogue Modeling for Spoken Language Systems Tutorial

For Next Time

Chapter 20