Introduction to Natural Language Processing

Chapter 19

Outline

Review

- reference resolution
- text coherence
- discourse structure

Dialogue and Conversational Agents

- why dialog is different
- representing and interpreting dialog acts
- dialogue structure and coherence
- conversational agents

Review: Text Coherence

Coherent vs. incoherent texts

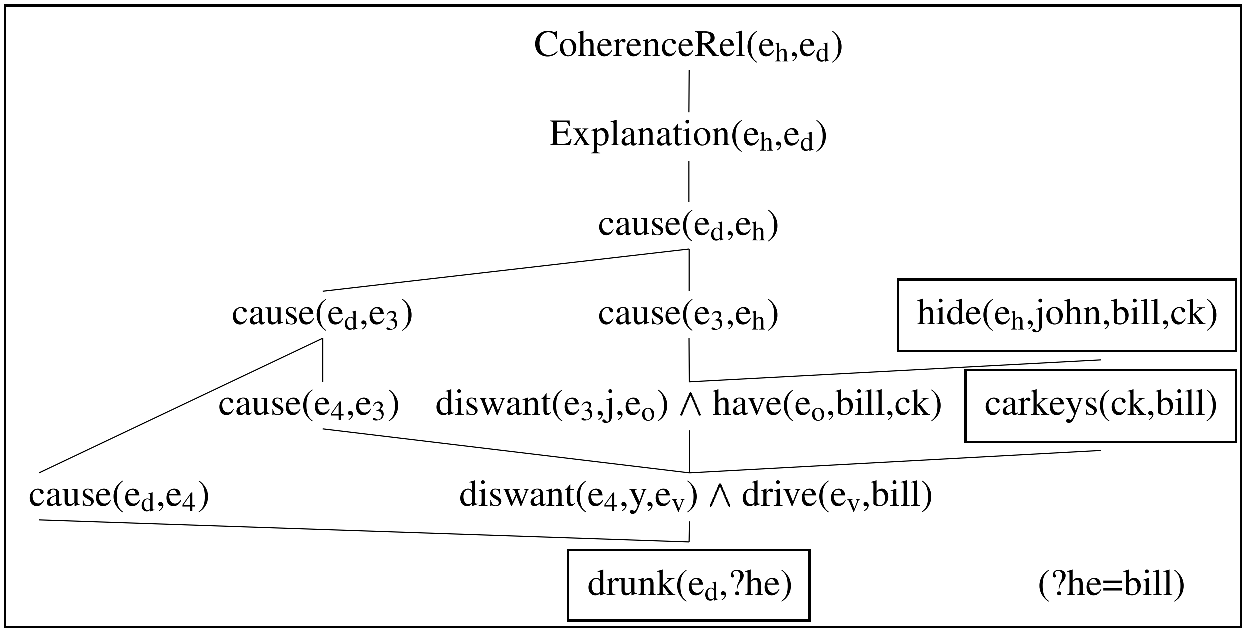
- John hid Bill's car keys. He was drunk
- ? John hid Bill's car keys. He likes spinach.

Example Coherence Relations from Hobbs

- Result
  - Infer that the state or event asserted by the meaning of the first sentence (S_0) causes or could cause the state or event asserted by the meaning of the second sentence (S_1).
  - John bought an Acura. His father went ballistic.
  - *and*
- Explanation
  - Infer that the state or event asserted by S_1 causes or could cause the state or event asserted by S_0.
  - John hid Bill's car keys. He was drunk.
  - *because*

Review: An Abductive Algorithm

John hid Bill's car keys. He was drunk.



The coherence derivation causes the inference of implicit information, reference resolution.

Dialogue and Conversational Agents

"**Conversation** or **dialogue** is the most fundamental and specially privileged arena of language."

What makes dialogue different?

- turn-taking
- grounding
- implicature

Speech act representation and interpretation.

New approaches to coherence and structure.

Example Conversational Agents

Why2 Physics Tutor

Phone systems you called at the beginning of the semester

NPR show

- http://www.npr.org/ramfiles/me/20020214.me.05.ram

Turns and Utterances

Dialogue is characterized by turn-taking.

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.

Speakers know how to take turns (*who* should talk next, and *when* they should talk)

- little overlap (around 5% in English - although depends on the domain!)
- not much silence between turns either

Turns

Conversation Analysis provides a socio-linguistic approach to turn-taking (e.g., Sacks et al.).

Turn-Taking Rule (simplified):

At each transition-relevance place of each turn:

- If during this turn current speaker has selected A as the next speaker, then A must speak next.
- If current speaker does not select the next speaker, any other speaker may take the next turn.
- If no one else takes the next turn, the current speaker may take the next turn.

Transition-relevance places are where the structure of the language allows speaker shifts to occur.

Conversation Analysis (cont.)

Adjacency pairs (set up next speaker expectations)

- GREETING GREETING
- QUESTION ANSWER
- COMPLIMENT DOWNPLAYER
- REQUEST GRANT

Significant silence (follows first part of an adjacency pair)

- A: Is there something bothering you or not?
- (1.0)
- A: Yes or no?
- (1.5)
- A: Eh?
- B: No.

Implications for spoken dialogue systems

Utterances

Transition-relevance places are typically at <u>utterance</u> boundaries.

Recall that (spoken) utterances are typically shorter, contain more pronouns, have repairs . . . , compared to written sentences.

Many theories take the utterance as the primitive unit, but utterances are difficult to segment

- a single utterance may occur across several turns
  - A: We've got you on USAir flight 99
  - B: Yep
  - A: leaving on December 1.
- multiple utterances may occur in a single turn
  - A: We've got you on USAir flight 99 leaving on December 1. Do you need a rental car?
- linguistic boundary cues include cue words, ngrams, prosody

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 <u>utterance</u> boundaries, but utterance segmentation is a difficult problem

Grounding

Conversational participants must constantly establish common ground (or mutual belief).

Thus, 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.

Acknowledgement

- continuer / backchannel / acknowledgement token (also nods if vision available)
- Example
  - A: ...returning on U.S. flight one.
  - C: *Mm hmm*
- grounds A's utterance, and also returns turn

Display (stronger method)

- display all or part of utterance to be grounded verbatim
- Example
  - C: OK I'll take the 5ish flight on the 11th.
  - A: *On the 11th?*

Request for repair

- indicate lack of grounding
- Example
  - C: OK I'll take the 5ish flight on the 11th.
  - A: *Huh?*
  - C: I'll take the 5ish flight on the 11th.

Conversational Implicature

Pragmatics: the study of how language is used to accomplish goals (Chapter 1); beyond literal meaning.

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.

Example

- A: What day in May did you want to travel?
- C: I need to be there for a meeting that's from the 12th to the 15th.
- A: OK. There are 3 non-stops on the 11th.

Implicature-licensed inferences

- the meeting information answers the request for travel dates
- there are not 4 non-stops

Grice's Maxims

Maxim of Quantity: Be exactly as informative as is required:

- Make your contribution as informative as is required (for the current purposes of the exchange).
- Do not make your contribution more informative than is required.

Maxim of Quality: Try to make your contribution one that is true:

- Do not say what you believe to be false.
- Do not say that for which you lack adequate evidence.

Maxim of Relevance: Be relevant:

Maxim of Manner: Be perspicuous:

- Avoid obscurity of expression.
- Avoid ambiguity.
- Be brief (avoid unnecessary prolixity).
- Be orderly.

Which maxims license previous inferences?

Dialogue Acts

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

Example: performative sentences

- I name this ship the Titanic.
- I second this motion.
- I bet you five dollars it will snow tomorrow.

Speech Acts

The utterance of any sentence in a real situation constitutes three kinds of act.

<u>Locutionary acts</u>

- the utterance of a sentence with a particular meaning

<u>Illocutionary acts</u>

- the act of asking, answering, promising, etc. in uttering a sentence

<u>Perlocutionary acts</u>

- the (often intentional) production of certain effects upon the feelings, thoughts, or actions of the addressee in uttering a sentence

You can't do that.

- locutionary: utterance
- illocutionary force: protesting
- perlocutionary effect: stopping or annoying the hearer

Speech Acts

Searle uses this term to classify illocutionary acts (1975).

- Assertives: committing the speaker to something's being the case (*suggesting, putting forward, swearing boasting, concluding*)
- Directives: attempts by the speaker to get the addressee to do something (*asking, ordering requesting, inviting advising, begging*)
- Commissives: committing the speaker to some future course of action (*promising, planning, vowing, betting, opposing*)
- Expressives: expressing the psychological state of the speaker about a state of affairs (*thanking, apologizing, welcoming, deploring*)
- Declarations: bringing about a different state of the world via the utterance (including many of the performative acts above: *I resign, you're fired*)

Review II: Dialogue is Different

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 III: Dialogue Acts

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

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

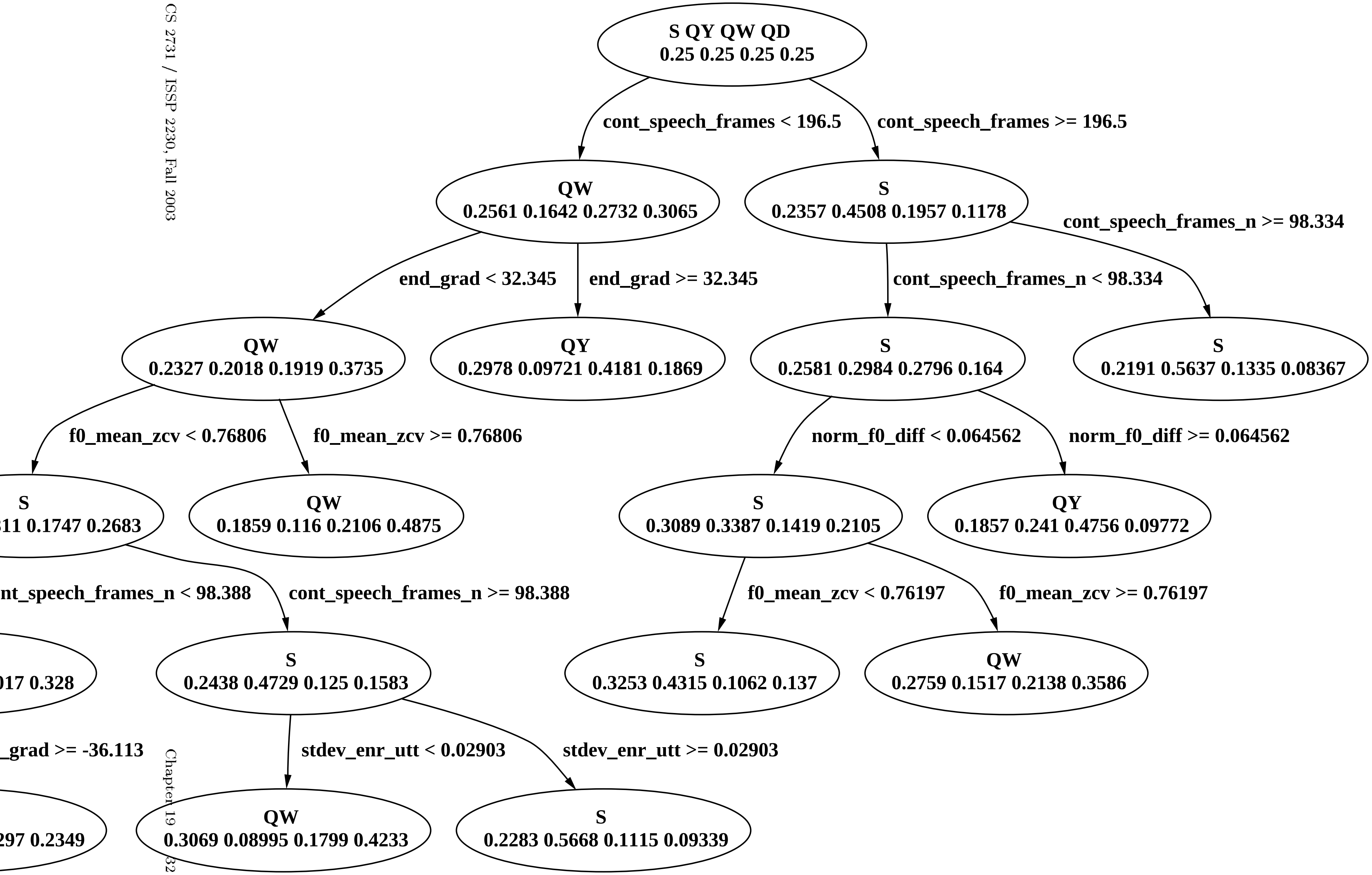
Decision trees for using prosody to classify speech acts

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
311 0.1747 0.2683
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
nt_speech_frames_n < 98.388
cont_speech_frames_n >= 98.388
f0_mean_zcv < 0.76197
f0_mean_zcv >= 0.76197
017 0.328
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
_grad >= -36.113
stdev_enr_utt < 0.02903
stdev_enr_utt >= 0.02903
297 0.2349
QW
0.3069 0.08995 0.1799 0.4233
S
0.2283 0.5668 0.1115 0.09339

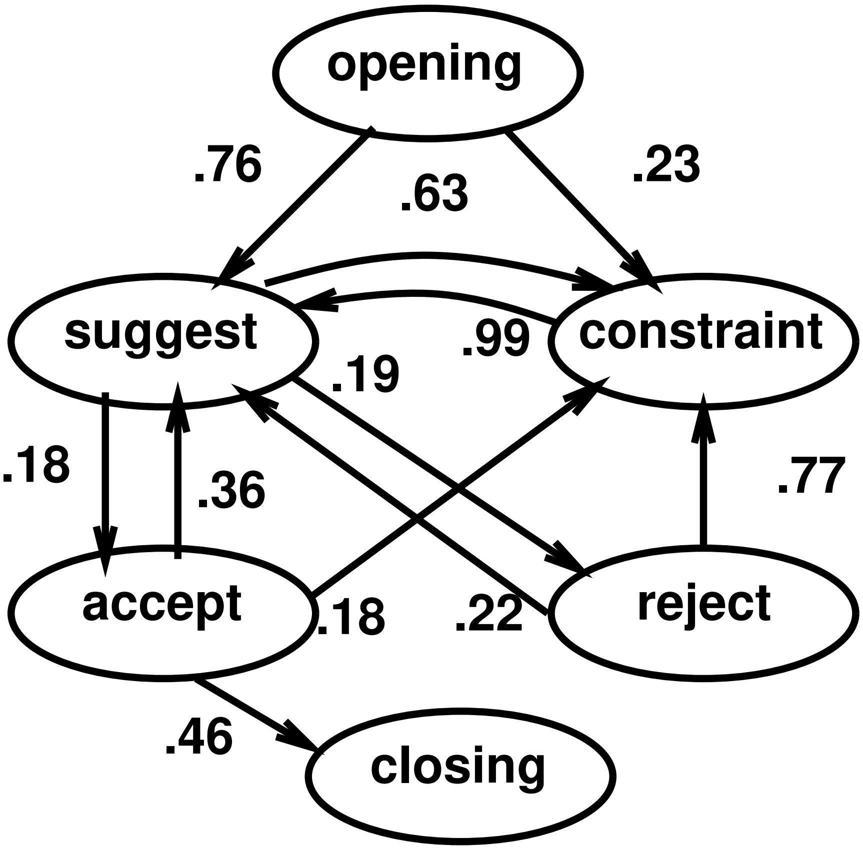
Conversational Structure

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

Intentional versus previous informational 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

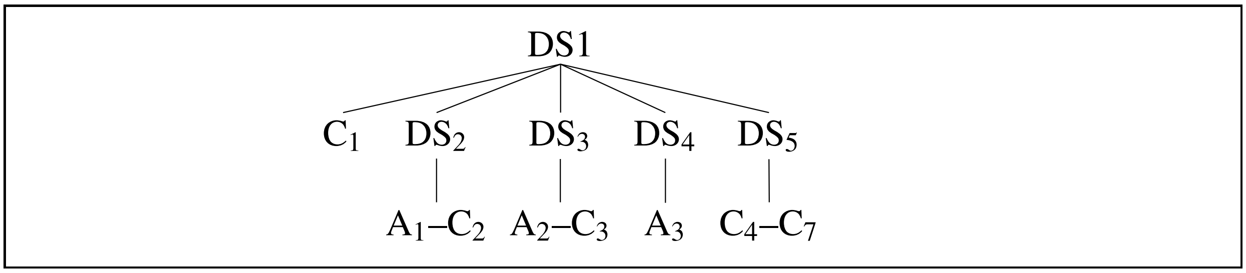
- dominance: DSP1 dominates DSP2 if satisfying DSP2 is intended to provide part of the satisfaction of DSP1
- satisfaction-precedence: DSP1 satisfaction-precedes DSP2 if DSP1 must be satisfied before DSP2

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)

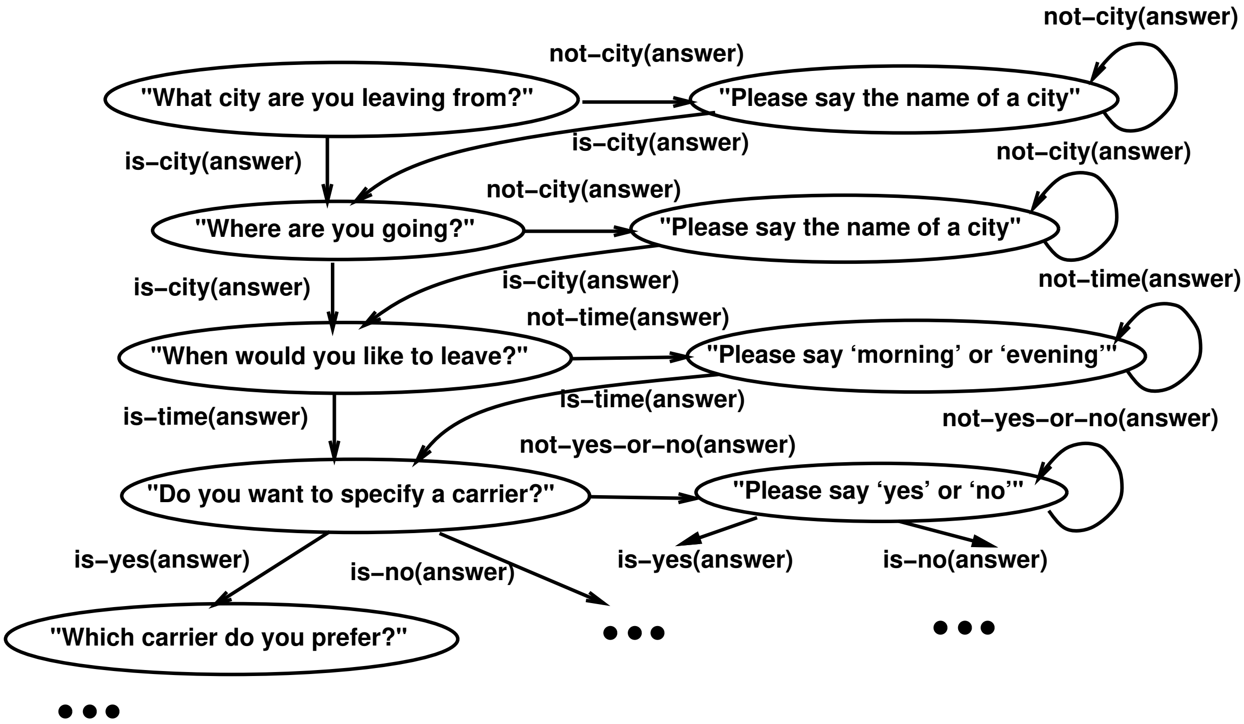
Dialogue Managers

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)

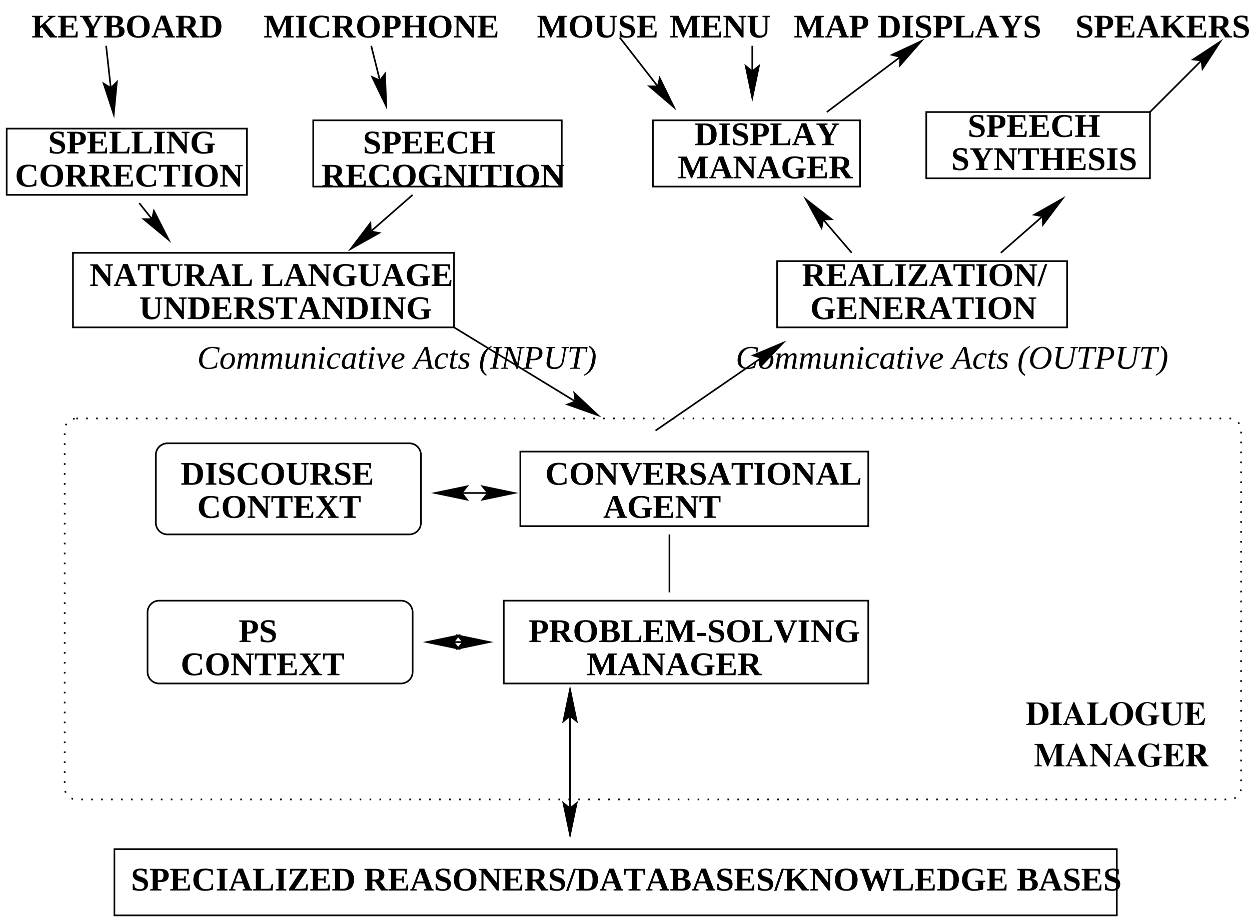
- Intended speech acts: REQUEST-INFORM1, ACKNOWLEDGE1

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

TRIPS 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

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

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:

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