# The upper reaches of NLP: Discourse and Dialog

(in one fast-moving blur)

**Christopher Manning - CS224N**

**(Largely recycling slides by Jim Martin, Julia Hirschberg, and Jennifer Chu-Carroll)**

1

## What Makes a Discourse Coherent?

The reason is that these utterances, when juxtaposed, will not exhibit coherence. Almost certainly not. Do you have a discourse? Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book.

2

## Better?

Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence.

3

## What makes a text coherent?

- **Appropriate use of coherence relations between subparts of the discourse -- rhetorical structure**
- **Appropriate sequencing of subparts of the discourse -- discourse/topic structure**
- **Appropriate use of referring expressions**

4

## Rhetorical Structure Theory

(Mann, Matthiessen, and Thompson '89)

- **One theory of discourse structure, based on identifying relations between segments of the text**
  - Nucleus/satellite notion encodes asymmetry
  - Some rhetorical relations:
    - Elaboration (set/member, class/instance, whole/part…)
    - Contrast: multinuclear
    - Condition: Sat presents precondition for N
    - Purpose: Sat presents goal of the activity in N
  - How many rhetorical relations are there? MMT say 23

5

## Relations

- **A sample definition**
  - Relation: evidence
  - Constraints on N: H might not believe N as much as S thinks s/he should
  - Constraints on Sat: H *already believes or will believe* Sat
- **An example:**

  George Bush favors big business.

  He is sure to veto House Bill 1711.

6

## Automatic Rhetorical Structure Labeling

- **Same old story by now…**
  - Get a group of annotators to assign a set of RST relations to a text
  - Extract a set of surface features from the text that might signal the presence of the rhetorical relations in that text
  - Train a supervised ML sequence model based on the training set

7

## Classifier Features

- Explicit markers: *because, however, therefore, then, etc.*
  - But often there is *no* explicit marker
- Tendency of certain syntactic structures to signal certain relations: Infinitives are often used to signal purpose relations:
  - Use rm *to delete files*.
- Ordering
- Tense/aspect
- Intonation
- Lexical Chains

8

## Reference Resolution

U: Where is A Bug's Life playing in Summit?
S: It is playing at the Summit theater.
U: When is it playing there?
S: It's playing at 2pm, 5pm, and 8pm.
U: I'd like 1 adult and 2 children for the first show. How much would that cost?

- **Knowledge sources:**
  - Domain knowledge
  - Discourse knowledge
  - World knowledge

9

## Referring Expressions: Definition

- **Referring expressions provide an additional kind of glue that makes texts cohere.**
- **Referring expressions are words or phrases, the *semantic interpretation* of which is *a discourse entity* (also called referent)**
  - Discourse entities are *semantic objects* and they can have multiple *syntactic realizations* within a text

10

## Discourse sounds bad without them

- U: Where is A Bug's Life playing in Summit?
- S: A Bug's Life is playing at the Summit theater.
- U: When is A Bug's Life playing at the Summit theater?
- S: A Bug's Life's playing at 2pm, 5pm, and 8pm.
- U: I'd like 1 adult and 2 children for the first show. How much would 1 adult and 2 children for the first show cost?

11

## Reference Resolution: In Theory

- Focus stacks:
  - Maintain recent objects in stack
  - Select objects that satisfy semantic/pragmatic constraints starting from top of stack
  - May take into account discourse structure
- Centering (Grosz 1995):
  - Backward-looking center (Cb): object connecting the current sentence with the previous sentence
  - Forward-looking centers (Cf): potential Cb of the next sentence
  - Rule-based filtering & ranking of objects for pronoun resolution

12

## Centering theory. Motivation

- **(Grosz 1995) examines interactions between local coherence and the choice of referring expressions**
  - Pronouns and definite descriptions are not equivalent with respect to their effect on coherence
  - They make different inference demands on the hearer or reader.

13

## Task: Anaphora resolution

- **Finding in a text all the referring expressions that have one and the same denotation**
  - Pronominal anaphora resolution
  - Anaphora resolution between named entities
  - Full noun phrase anaphora resolution
  - Zero anaphors (fairly rare in English; everywhere in Chinese/Japanese/…)

14

## Pronominal anaphora resolution

- **Rule-based vs statistical**
  - (Ken 1996), (Lap 1994) vs (Ge 1998)
- **Performed on full syntactic parse vs on shallow syntactic parse**
  - (Lap 1994), (Ge 1998) vs (Ken 1996)
- **Type of text used for the evaluation**
  - (Lap 1994) computer manual texts (86% accuracy)
  - (Ge 1998) WSJ articles (83% accuracy)
  - (Ken 1996) different genres (75% accuracy)

15

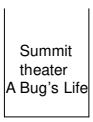
## Reference Resolution in Dialog Systems in Practice

- Non-existent: does not allow the use of anaphoric references
- Allows only simple references:
  - utilizes the focus stack reference resolution mechanism
  - does not take into account discourse structure information
- Example:

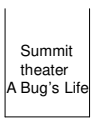
U: Where is A Bug's Life playing in Summit?

16

S: A Bug's Life is playing at the Summit theater.



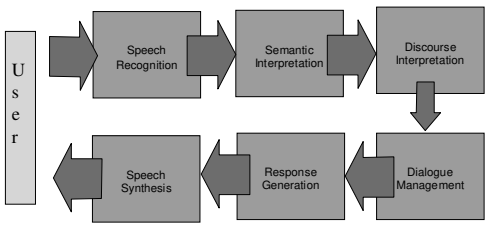

17




U: When is **it** playing **there**?

18

## Spoken Dialogue System




19

## Dialogue vs. Monologue

- Monologue and dialogue both involve interpreting
  - Information status (given and new info)
  - Coherence issues
  - Reference resolution
  - Speech acts, implicature, intentionality
- Dialogue involves managing
  - Turn-taking
  - Grounding
  - Detecting and repairing misunderstandings
  - Initiative and confirmation strategies

20

## Segmenting Speech into Utterances

- What is an `utterance'?
  - Why is end of utterance detection harder than end of sentence?
  - Single syntactic sentence may span several turns
    - A: We've got you on USAir flight 99
    - B: Yep
    - A: leaving on December 1.
  - Multiple syntactic sentences may occur in single turn
    - A: We've got you on USAir flight 99 leaving on December. Do you need a rental car?
  - Intonational definitions: intonational phrase, breath group, intonation unit

21

## Turns and Utterances

- Dialogue is characterized by turn-taking:
  - Who should talk next
  - When they should talk
- Turns in recorded speech:
  - Little speaker overlap (around 5% in English)
  - But little silence between turns either
- How do we know when a speaker is giving up or taking a turn? Holding the floor? How do we know when a speaker is interruptable?

22

## Talking to Computers

- Spoken dialogue systems make it possible to accomplish real tasks *without talking to a real person*
  - A big development in the last 10 years!
- Keys to success
  - Sticking to goal-directed interactions in a limited domain
  - Priming users to adopt a vocabulary you can recognize
  - Segmenting the task into manageable stages
  - Judicious use of system vs. mixed initiative

23

## Overall System Strategies

- Touch-tone replacement:
  - S: For checking information, press or say one.
  - U: One.
- System Initiative (Control freak)
  - S: Please give me your arrival city name.
  - U: Baltimore.
  - S: Please give me your departure city name
  - U: Boston
  - S:…
- Rigid, unnatural, difficult with chatty users

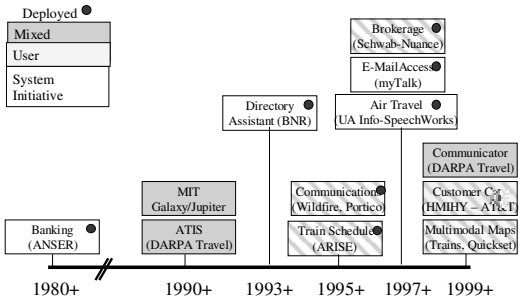
24

## Overall System Strategies

- **Mixed initiative**
  - S: How may I help you?
  - U: I want to go to Boston.
  - S: What day do you want to go to Boston?
- **User Initiative**
  - S: How may I help you?
  - U: I want to go from Boston to Baltimore on November 8.

25

## *Some Representative Spoken Dialogue Systems*




26

## Dialogue Management Example

- **Dialogue 1:**

S: Would you like movie showtime or theater playlist information?
U: Movie showtime.
S: What movie do you want showtime information about?
U: Saving Private Ryan.
S: At what theater do you want to see Saving Private Ryan?
U: Paramount theater.
S: Saving Private Ryan is not playing at the Paramount theater.

27

## Dialogue Management Example

- **Dialogue 2:**

S: How may I help you?
U: When is Saving Private Ryan playing?
S: For what theater?
U: The Paramount theater.
S: Saving Private Ryan is not playing at the Paramount theater, but it's playing at the Madison theater at 3:00, 5:30, 8:00, and 10:30.

28

## Comparison of Sample Dialogues

- **Dialogue 1:**
  - System-initiative
  - Implicit confirmation
  - Merely informs user of failed query
  - Mechanical
  - Least efficient
- **Dialogue 2:**
  - Mixed-initiative
  - No confirmation
  - Suggests alternative when query fails
  - More natural
  - Most efficient

29

## Intention Recognition

B: I have to wash my hair.

30

A: Would you like to go to the hairdresser?

- **B's utterance should be interpreted as an** acceptance **of A's proposal**.

31

A: What's that smell around here?

- **B's utterance should be interpreted as an** answer **to A's question**.

32

A: Would you be interested in going out to dinner tonight?

- **B's utterance should be interpreted as a** rejection **of A's proposal**.

33

## Intention Recognition (Cont'd)

- **Goal: to recognize the intent of each user utterance as one (or more) of a set of dialogue acts based on context**
- **Sample dialogue actions:**
  - Switchboard DAMSL
    - Conventional-closing
    - Statement-(non-)opinion
    - Agree/Accept
    - Acknowledgment
    - Yes-No-Question/Yes-Answer
    - Non-verbal
    - Abandoned
  - Verbmobil
    - Greet/Thank/Bye
    - Suggest
    - Accept/Reject
    - Confirm
    - Clarify-Query/Answer
    - Give-Reason
    - Deliberate

34

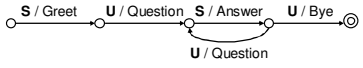
## Intention Recognition: In Theory

- **Knowledge sources:**
  - Overall dialogue goals
  - Orthographic features, e.g.:
    - punctuation
    - cue words/phrases: "but", "furthermore", "so"
    - transcribed words: "would you please", "I want to"
  - Dialogue history, i.e., previous dialogue act types
  - Dialogue structure, e.g.:
    - subdialogue boundaries
    - dialogue topic changes
  - Prosodic features of utterance: duration, pause, F0, speaking rate

35

## Intention Recognition: In Theory (Cont'd)

- Finite-state dialogue grammar:
  - e.g.




- Plan-based discourse understanding:
  - Recipes: templates for performing actions
  - Inference rules: to construct plausible plans
- Empirical methods:
  - Probabilistic dialogue act classifiers: HMMs
  - Rule-based dialogue act recognition: CART, Transformation-based learning

36

## Intention Recognition: In Practice

- Makes assumptions about (high-level) task-specific intentions: e.g.,
  - Call routing: *giving destination information*
  - ATIS: *requesting flight information*
  - Movie information system: *movie showtime or theater playlist information*
- Does not allow user-initiated complex dialogue acts, e.g., clarification, or indirect responses

S1: What's your account number?
U1: Is that the number on my ATM card?

S2: Would you like to transfer $1,500 from savings to checking?
U2: If I have enough in savings.



## Intention Recognition: In Practice (Cont'd)

- User utterances can play one of two roles:
  - Identify one of a set of possible task intentions
  - Provide necessary information for performing a task
- Based on either keywords in an utterance or its syntactic/semantic representation
- Maps keywords or representations to intentions using:
  - Template matching
  - Probabilistic model
  - Vector-based similarity measures



## Grounding (Clark & Shaefer '89)

- Conversational participants don't just take turns speaking…. they try to establish common ground (or mutual belief)
- H must ground a S's utterances by making it clear whether or not understanding has occurred
- How do hearers do this?



S: I can upgrade you to an SUV at that rate.

- Continued attention
  (U gazes appreciatively at S)
- Relevant next contribution
  U: Do you have a RAV4 available?
- Acknowledgement/backchannel
  U: Ok/Mhmmm/Great!
- Demonstration/paraphrase
  U: An SUV.
- Display/repetition
  U: You can upgrade me to an SUV at the same rate?
- Request for repair
  U: I beg your pardon?



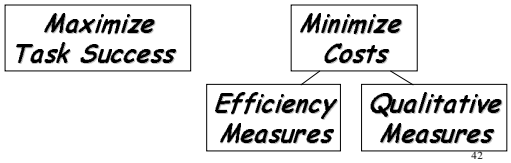
## Detecting Grounding Behavior

- Evidence of system misconceptions reflected in user responses (Krahmer et al '99, '00)
  - Responses to incorrect verifications
    - contain more words (or are empty)
    - show marked word order (especially after implicit verifications)
    - contain more disconfirmations, more repeated/corrected info
  - 'No' after incorrect verifications vs. other ynq's
    - has higher boundary tone
    - wider pitch range
    - longer duration
    - longer pauses before and after
    - more additional words after it



## Evaluation

- Performance of a dialogue system is affected both by *what* gets accomplished by the user and the dialogue agent and *how* it gets accomplished



## Metrics

- **Efficiency of the Interaction:User Turns, System Turns, Elapsed Time**
- **Quality of the Interaction: ASR rejections, Time Out Prompts, Help Requests, Barge-Ins, Mean Recognition Score (concept accuracy), Cancellation Requests**
- **User Satisfaction**
- **Task Success: perceived completion, information extracted**

43

## User Satisfaction Metrics

- TTS Performance
  - Was system easy to understand in this conversation?
- ASR Performance
  - In this conversation, did system understand what you said?
- Task Ease
  - In this conversation, was it easy to do what you wanted?
- Interaction Pace
  - Was the pace of interaction appropriate in this conversation?
- User Orientation
  - In this conversation, did you know what you could say at each point of the dialog?
- System Response
  - How often was the system sluggish and slow to reply to you in this conversation?
- Expected Behavior
  - Did system work the way you expected it to in this conversation?

44

## Identifying Misrecognitions and User Corrections Automatically

(Hirschberg, Litman & Swerts)

- Identifying when conversation has gone astray and recovering, mainly by examining user's response
- Collect corpus from interactive voice response system
- Identify speaker 'turns'
  - that are incorrectly recognized
  - where speakers first aware of error
  - that correct misrecognitions
- Identify prosodic features of turns in each category and compare to other turns
- Use ML to train a classifier to do this

45

## Results

- **Reduced error in predicting misrecognized turns to 8.64%**
- **Error in predicting 'awares' (12%)**
- **Error in predicting corrections (18-21%)**

46

## Turn Types

TOOT: Hi. This is AT&T Amtrak Schedule System. This is TOOT. How may I help you?

User: Hello. I would like trains from Philadelphia to New York leaving on Sunday at ten thirty in the evening.

TOOT: Which city do you want to go to? Misrecognition site

User: New York. Correction occurs

Aware site

47

## Conclusions

- **Spoken dialogue systems present new possibilities but also new problems**
  - Recognizing speech introduces a new source of errors
  - Additional information provided in the speech stream offers new information about users' intended meanings, emotional state (grounding of information, speech acts, reaction to system errors)
- **Spoken dialogue systems rather than web-based interfaces?**

48