Dialogue Acts and Information State

Julia Hirschberg

CS 4706

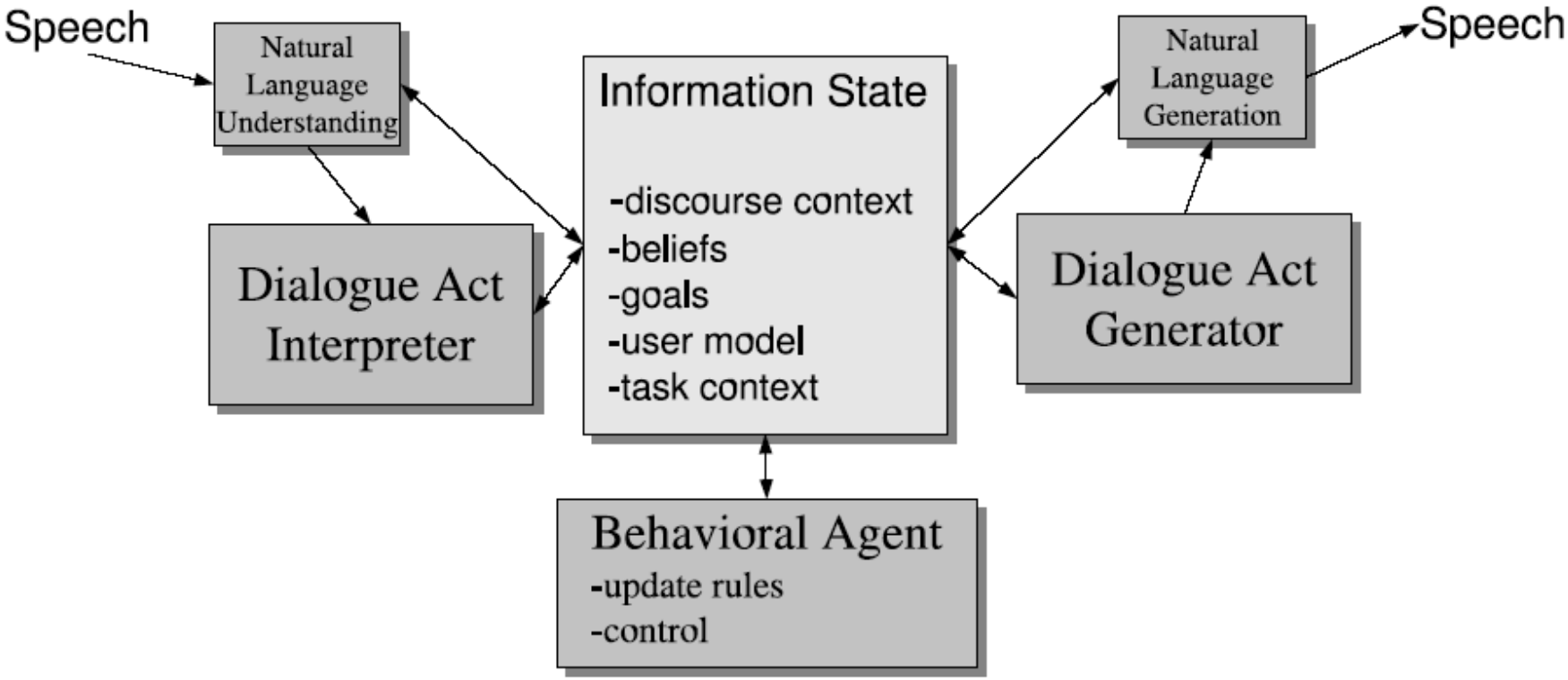
Information-State and Dialogue Acts

- If we want a dialogue system to be more than just form-filling, it
- Needs to:
  - Decide when user has asked a question, made a proposal, rejected a suggestion
  - **Ground** user's utterance, ask clarification questions, suggestion plans
- Good conversational agents need sophisticated models of interpretation and generation – beyond slot filling

Information-State Architecture

- Information state representation
- Dialogue act interpreter
- Dialogue act generator
- Set of update rules
  - Update dialogue state as acts are interpreted
  - Generate dialogue acts
- Control structure to select which update rules to apply

Information-state



Dialogue acts

- AKA **conversational moves**
- Actions with (internal) structure related specifically to their dialogue function
- Incorporates ideas of grounding with other dialogue and conversational functions not mentioned in classic Speech Act Theory

Speech Act Theory: Reminder

- John Searle *Speech Acts* '69
  - Locutionary acts: semantic meaning/surface form
  - Illocutionary acts: request, promise, statement, threat, question
  - Perlocutionary acts: Effect intended to be produced on Hearer: regret, fear, hope

What Kind of Speech Acts do we need for a Real Task: Verbmobil

- Two-party scheduling dialogues
- Speakers were asked to plan a meeting at some future date
- Data used to design conversational agents which would help with this task
- Issues:
  - Cross-language
  - Machine translation
  - Scheduling assistant

Verbmobil Dialogue Acts

THANK	thanks
GREET	Hello Dan
INTRODUCE	It's me again
BYE	Allright, bye
REQUEST-COMMENT	How does that look?
SUGGEST	June 13th through 17th
REJECT	No, Friday I'm booked all day
ACCEPT	Saturday sounds fine
REQUEST-SUGGEST	What is a good day of the week for you?
INIT	I wanted to make an appointment with you
GIVE_REASON	Because I have meetings all afternoon
FEEDBACK	Okay
DELIBERATE	Let me check my calendar here
CONFIRM	Okay, that would be wonderful
CLARIFY	Okay, do you mean Tuesday the 23rd?

Automatic Interpretation of Dialogue Acts

- How do we automatically identify dialogue acts?
- Given an utterance:
  - Decide whether it is a QUESTION, STATEMENT, SUGGEST, or ACKNOWLEDGMENT
- Recognizing illocutionary force will be crucial to building a dialogue agent
- Perhaps we can just look at the form of the utterance to decide?

Can we just use the surface syntactic form?

- YES-NO-Qs have auxiliary-before-subject syntax:
  - Will breakfast be served on USAir 1557?
- STATEMENTs have declarative syntax:
  - I don't care about lunch
- COMMANDs have imperative syntax:
  - Show me flights from Milwaukee to Orlando on Thursday night

Surface Form != Speech Act Type

	Locutionary Force	Illocutionary Force
Can I have the rest of your sandwich?	Question	Request
I want the rest of your sandwich	Declarative	Request
Give me your sandwich!	Imperative	Request

Dialogue act disambiguation is hard! *Who's on First*?

Abbott: Well, Costello, I'm going to New York with you. Bucky Harris the Yankee's manager gave me a job as coach for as long as you're on the team.
Costello: Look Abbott, if you're the coach, you must know all the players.
Abbott: I certainly do.
Costello: Well you know I've never met the guys. So you'll have to tell me their names, and then I'll know who's playing on the team.
Abbott: Oh, I'll tell you their names, but you know it seems to me they give these ball players now-a-days very peculiar names.
Costello: You mean funny names?
Abbott: Strange names, pet names...like Dizzy Dean...
Costello: His brother Daffy
Abbott: Daffy Dean...
Costello: And their French cousin.
Abbott: French?
Costello: Goofe'
Abbott: Goofe' Dean. Well, let's see, we have on the bags, Who's on first, What's on second, I Don't Know is on third...
Costello: That's what I want to find out.
Abbott: I say Who's on first, What's on second, I Don't Know's on third….

Dialogue act ambiguity

- Who's on first
  - INFO-REQUEST
  - or
  - STATEMENT

Dialogue Act ambiguity

- Can you give me a list of the flights from Atlanta to Boston?
  - Looks like an INFO-REQUEST.
  - If so, answer is:
    - YES.
  - But really it's a DIRECTIVE or REQUEST, a polite form of:
  - Please give me a list of the flights…
- What looks like a QUESTION can be a REQUEST

Dialogue Act Ambiguity

- What looks like a STATEMENT can be a QUESTION:

Us	OPEN-OPTION	**I was wanting to make some arrangements for a trip that I'm going to be taking uh to LA uh beginning of the week after next**
Ag	HOLD	**OK uh let me pull up your profile and I'll be right with you here. [pause]**
Ag	CHECK	And you said you wanted to travel next week?
Us	ACCEPT	**Uh yes.**

Indirect Speech Acts

- Utterances which use a surface statement to ask a question
  - And you want to….
- Utterances which use a surface question to issue a request
  - Can you get me…

DA Interpretation as Statistical Classification

- Lots of clues in each sentence that can tell us which DA it is:
  - Words and Collocations:
    - Please or would you: good cue for REQUEST
    - Are you: good cue for INFO-REQUEST
  - Prosody:
    - Rising pitch is a good cue for INFO-REQUEST
    - Loudness/stress can help distinguish yeah/AGREEMENT from yeah/BACKCHANNEL
  - Conversational Structure
    - Yeah following a proposal is probably AGREEMENT; yeah following an INFORM probably a BACKCHANNEL

Disambiguating Ambiguous DAs Intonationally

- Nickerson & Chu-Carroll ’99: Can info-requests be disambiguated reliably from action-requests?
- Modal (Can/would/would..willing) questions
  - Can you move the piano?
  - Would you move the piano?
  - Would you be willing to move the piano?

Experiments

- Production studies:
  - Subjects read ambiguous questions in disambiguating contexts
  - Control for given/new and contrastiveness
  - Polite/neutral/impolite
- Problems:
  - Cells imbalanced
  - No pretesting
  - No distractors
  - Same speaker reads both contexts

Results

- Indirect requests (e.g. for action)
  - If L%, more likely (73%) to be indirect
  - If H%,46% were indirect: differences in height of boundary tone?
  - Politeness: can differs in impolite (higher rise) vs. neutral
  - Speaker variability

Statistical Classifier Model of DA Interpretation

- Goal: decide for each sentence what DA it is
- C**lassification task**: 1-of-N classification decision for each sentence
  - With **N** classes (= number of dialog acts).
  - Three probabilistic models corresponding to the 3 kinds of cues from the input sentence.
    - Conversational Structure: Probability of one dialogue act following another P(Answer|Question)
    - Words and Syntax: Probability of a sequence of words given a dialogue act: P("do you" | Question)
    - Prosody: probability of prosodic features given a dialogue act : P("rise at end of sentence" | Question)

Corpus Studies: Jurafsky et al ‘98

- Lexical, acoustic/prosodic/syntactic differentiators for yeah, ok, uhuh, mhmm, um…
- Labeling
  - Continuers: Mhmm (not taking floor)
  - Assessments: Mhmm (tasty)
  - Agreements: Mhmm (I agree)
  - Yes answers: Mhmm (That’s right)
  - Incipient speakership: Mhmm (taking floor)

Corpus

- Switchboard telephone conversation corpus
  - Hand segmented and labeled with DA information (initially from text)
  - Relabeled for this study
  - Analyzed for
    - Lexical realization
    - F0 and rms features
    - Syntactic patterns

Results: Lexical Differences

- Agreements
  - yeah (36%), right (11%),...
- Continuer
  - uhuh (45%), yeah (27%),…
- Incipient speaker
  - yeah (59%), uhuh (17%), right (7%),…
- Yes-answer
  - yeah (56%), yes (17%), uhuh (14%),...

Results: Prosodic and Syntactic Cues

- Relabeling from speech produces only 2% changed labels over all (114/5757)
  - 43/987 continuers --> agreements
  - Why?
    - Shorter duration, lower F0, lower energy, longer preceding pause
- Over all DA's, duration best differentiator but…
  - Highly correlated with DA length in words
- Assessments: That's X (good, great, fine,…)

Generating Dialogue Acts

- Confirmation
- Rejection

Confirmation

- Another reason for grounding
  - ASR Errors: Speech is a very errorful channel
  - Even for humans in noisey conditions
  - Humans use grounding to confirm that they've heard correctly
  - ASR is much worse than humans!
- Conclusion: SDS need to do even more grounding and confirmation than humans

Explicit confirmation

- S: Which city do you want to leave from?
- U: Baltimore
- S: Do you want to leave from Baltimore?
- U: Yes

Explicit confirmation

- U: *I'd like to fly from Denver Colorado to New York City on September 21st in the morning on United Airlines*
- S: Let's see then. I have you going from Denver Colorado to New York on September 21st. Is that correct?
- U: *Yes*

Implicit confirmation: display

- U: *I'd like to travel to Berlin*
- S: When do you want to travel to Berlin?

- U: *Hi I'd like to fly to Seattle Tuesday morning*
- S: Traveling to Seattle on Tuesday, August eleventh in the morning. Your name?

Implicit vs. Explicit

- Complementary strengths
- Explicit: Easier for users to correct system's mistakes (Can just say "no")
- But explicit is cumbersome and long
- Implicit: Much more natural, quicker, simpler (if system guesses right).

Implicit and Explicit

- Early systems: all-implicit or all-explicit
- Modern systems: adaptive
- How to decide?
  - ASR system can provide confidence metric.
    - Expresses how convinced system is of its transcription of the speech
  - If high confidence, use implicit confirmation
  - If low confidence, use explicit confirmation

Computing Confidence

- Simplest: Use acoustic log-likelihood of user's utterance
- More features might help
  - Prosodic: utterances with longer pauses, F0 excursions, longer durations
  - Backoff: did we have to backoff in the LM?
  - Cost of an error: Explicit confirmation before moving money or booking flights

Rejection

- e.g., VoiceXML “nomatch”
- “I’m sorry, I didn’t understand that.”
- Reject when:
  - ASR confidence is low
  - Best interpretation is semantically ill-formed
- Option: 4-tiered level of confidence:
  - Below confidence threshhold, reject
  - Above threshold, explicit confirmation
  - If even higher, implicit confirmation
  - Even higher, no confirmation

DA Detection Example: Correction Detection

- Despite clever confirmation/rejection strategies, dialogue systems still make mistakes
- If system misrecognizes an utterance, and either
  - Rejects
  - Via confirmation, displays its misunderstanding
- Then user has a chance to make a correction
  - Repeat themselves
  - Rephrasing
  - Saying “no” to the confirmation question.

Learning from Human Behavior (Krahmer et al ’01)

- Learning from human behavior
  - ‘go on’ and ‘go back’ signals in grounding situations (implicit/explicit verification)
  - Positive: short turns, unmarked word order, confirmation, answers, no corrections or repetitions, new info
  - Negative: long turns, marked word order, disconfirmation, no answer, corrections, repetitions, no new info 

- Hypotheses supported but…
  - Can these cues be identified automatically?
  - How might they affect the design of SDS?

Corrections

- Unfortunately, corrections are harder to recognize than normal sentences
  - Swerts et al (2000): Corrections misrecognized twice as often (in terms of WER) as non-corrections
  - Why?
    - Prosody seems to be largest factor: hyperarticulation
    - Example from Liz Shriberg
      - "NO, I am DE-PAR-TING from Jacksonville) 
    - Hyperarticulation

A Labeled dialogue (Swerts et al)

	Turn	Corr Type
S	On which day of the week do you want to leave?	
S	Sorry, I didn't hear anything from you. Please tell me your departure date	
U	*Um tomorrow*	NA
S	You can say the departure date. For example, you can say 'today' or 'tomorrow'. Or you can say 'Monday', 'Tuesday', 'Wednesday'...	
U	*Tomorrow*	REP
S	You can say the departure date. For example, you can say 'today' or 'tomorrow'. Or you can say 'Monday', 'Tuesday', 'Wednesday'...	
U	*Saturday*	PAR
S	Do you want to leave on Saturday?	
U	Yes	
S	What time do you want to leave?	
U	Eight o'clock A M	
S	Do you want to leave around ten oclock p-m	
U	*Eight o'clock*	OMIT
S	Do you want to leave around ten oclock p-m	
U	*No*	OMIT
S	What time do you want to leave?	
U	*Eight o'clock A M*	ADD

Distribution of Correction Types

	Add	Add/Omit	Omit	Par	Rep
All	8%	2%	32%	19%	39%
After Misrec	7%	3%	40%	18%	32%
After Rej	6%	0%	7%	28%	59%

Machine Learning to Detect User Corrections

- Build classifiers using features like
  - Lexical information (words “no”, “correct”, “I don’t”, swear words)
  - Prosodic features (various increases in F0 range, pause duration, and word duration that correlation with hyperarticulation)
  - Length
  - ASR confidence
  - LM probability
  - Dialogue features (e.g., repetitions)

But….

- What to do when you recognize a user is trying to correct the system?

Summary

- Dialogue Acts and Information State
- Dialogue Acts
  - Ambiguities and disambiguation
- Dialogue Acts: Recognition
  - ML approaches to DA classification
- Dialogue Acts: Generation
  - Confirmation Strategies
  - Rejections
- Dialogue Acts: Detecting Corrections

Next

- Evaluating Spoken Dialogue Systems