Discourse & Dialogue: Introduction

Ling 575 A
Topics in NLP
March 30, 2011

Roadmap

- Definition(s) of Discourse
- Different Types of Discourse
  - Goals, Modalities
  - Topics, Tasks in Discourse & Dialogue
- Course structure
- Overview of Theoretical Approaches
  - Points of Agreement
  - Points of Variance
- Dialogue Models and Challenges
- Issues and Examples in Practice
  - Spoken dialogue systems

What is a Discourse?

- Discourse is:
  - Extended span of text

What is a Discourse?

- Discourse is:
  - Extended span of text
  - Spoken or Written

What is a Discourse?

- Discourse is:
  - Extended span of text
  - Spoken or Written
  - One or more participants

What is a Discourse?

- Discourse is:
  - Extended span of text
  - Spoken or Written
  - One or more participants
  - Language in Use

What is a Discourse?

- Discourse is:
  - Extended span of text
  - Spoken or Written
  - One or more participants
  - Language in Use
  - Expresses goals of participants
    - Processes to produce and interpret

Why Discourse?

- Understanding depends on context
  - Referring expressions: it, that, the screen
  - Word sense: plant
  - Intention: Do you have the time?

Why Discourse?

- Understanding depends on context
  - Referring expressions: it, that, the screen
  - Word sense: plant
  - Intention: Do you have the time?

- Applications: Discourse in NLP
  - Question-Answering
  - Information Retrieval
  - Summarization
  - Spoken Dialogue
  - Automatic Essay Grading

Different Parameters of Discourse

- Number of participants
  - Multiple participants -> Dialogue

Different Parameters of Discourse

- Number of participants
  - Multiple participants -> Dialogue
- Modality
  - Spoken vs Written

Different Parameters of Discourse

- Number of participants
  - Multiple participants -> Dialogue
- Modality
  - Spoken vs Written
- Goals
  - Transactional (message passing)
  - Interactional (relations, attitudes)
  - Task-oriented

Major Topics & Tasks

- Reference:
  - Resolution, Generation, Information Structure

Major Topics & Tasks

- Reference:
  - Resolution, Generation, Information Structure
- Intention Recognition

Major Topics & Tasks

- Reference:
  - Resolution, Generation, Information Structure
- Intention Recognition
- Discourse Structure
  - Segmentation, Relations

Major Topics & Tasks

- Reference:
  - Resolution, Generation, Information Structure
- Intention Recognition
- Discourse Structure
  - Segmentation, Relations
- Fundamental components:
  - How do they interact with dimensions of discourse?
    - # Participants, Spoken vs Written, ..

Dialogue

- Systems
  - Components
  - Dialogue Management
  - Evaluation
- Turn-taking
- Politeness
- Stylistics

Course Structure

- Discussion-oriented course:

Course Structure

- Discussion-oriented course:
  - Class participation

Course Structure

- Discussion-oriented course:
  - Class participation
  - Presentations
    - Topic survey

Course Structure

- Discussion-oriented course:
  - Class participation
  - Presentations
    - Topic survey
  - Project:
    - Proposal
    - Progress
    - Final report

Course Perspectives

- Foundational:
  - Linguistic view:
    - Understanding basic discourse phenomena
    - Analyzing language use in context

Course Perspectives

- Foundational:
  - Linguistic view:
    - Understanding basic discourse phenomena
    - Analyzing language use in context
- Practical/Implementational:
  - Computational view:
    - Developing systems and algorithms for discourse tasks

Course Projects

- Reflect linguistic and/or computational perspectives

Course Projects

- Reflect linguistic and/or computational perspectives
- Option 1: Analytic (Required for Ling elective credit)
  - In-depth analysis of linguistic discourse phenomena
    - Reflect understanding of literature
    - Analyze real data
    - ~15 page term paper

Course Projects

- Reflect linguistic and/or computational perspectives
- Option 1: Analytic (Required for Ling elective credit)
  - In-depth analysis of linguistic discourse phenomena
    - Reflect understanding of literature
    - Analyze real data
    - ~15 page term paper
- Option 2: Implementational
  - Implement, extend algorithms for discourse/dialogue tasks
  - Shorter write-up of approach, evaluation

Reference & Knowledge

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

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Reference & Knowledge

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

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Reference & Knowledge

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

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Reference &Knowledge

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

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Intention Recognition

U: What time is A Bug's Life playing at the Summit theater?

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Intention Recognition

U: What time is A Bug's Life playing at the Summit theater?

- Using keyword extraction and vector-based similarity measures:

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Intention Recognition

U: What time is A Bug's Life playing at the Summit theater?

- Using keyword extraction and vector-based similarity measures:
  - Intention: Ask-Reference: _time

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Intention Recognition

U: What time is A Bug's Life playing at the Summit theater?

- Using keyword extraction and vector-based similarity measures:
  - Intention: Ask-Reference: _time
  - Movie: A Bug's Life

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Intention Recognition

U: What time is A Bug's Life playing at the Summit theater?

- Using keyword extraction and vector-based similarity measures:
  - Intention: Ask-Reference: _time
  - Movie: A Bug's Life
  - Theater: the Summit quadplex

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

Computational Models of Discourse

- 1) Hobbs (1985): Discourse coherence based on small number of recursively applied relations

Computational Models of Discourse

- 1) Hobbs (1985): Discourse coherence based on small number of recursively applied relations

- 2) Grosz & Sidner (1986): Attention (Focus), Intention (Goals), and Structure (Linguistic) of Discourse

Computational Models of Discourse

- 1) Hobbs (1985): Discourse coherence based on small number of recursively applied relations
- 2) Grosz & Sidner (1986): Attention (Focus), Intention (Goals), and Structure (Linguistic) of Discourse
- 3) Mann & Thompson (1987): Rhetorical Structure Theory: Hierarchical organization of text spans (nucleus/satellite) based on small set of rhetorical relations

Computational Models of Discourse

- 1) Hobbs (1985): Discourse coherence based on small number of recursively applied relations
- 2) Grosz & Sidner (1986): Attention (Focus), Intention (Goals), and Structure (Linguistic) of Discourse
- 3) Mann & Thompson (1987): Rhetorical Structure Theory: Hierarchical organization of text spans (nucleus/satellite) based on small set of rhetorical relations
- 4) McKeown (1985): Hierarchical organization of schemata

Computational Models of Discourse

- 1) Hobbs (1985): Discourse coherence based on small number of recursively applied relations
- 2) Grosz & Sidner (1986): Attention (Focus), Intention (Goals), and Structure (Linguistic) of Discourse
- 3) Mann & Thompson (1987): Rhetorical Structure Theory: Hierarchical organization of text spans (nucleus/satellite) based on small set of rhetorical relations
- 4) McKeown (1985): Hierarchical organization of schemata

Discourse Models: Common Features

- Hierarchical, Sequential structure applied to subunits
  - Discourse “segments”
  - Need to detect, interpret

Discourse Models: Common Features

- Hierarchical, Sequential structure applied to subunits
  - Discourse “segments”
  - Need to detect, interpret
- Referring expressions provide coherence
  - Explain and link

Discourse Models: Common Features

- Hierarchical, Sequential structure applied to subunits
  - Discourse “segments”
  - Need to detect, interpret
- Referring expressions provide coherence
  - Explain and link
- Meaning of discourse more than that of component utterances

Discourse Models: Common Features

- Hierarchical, Sequential structure applied to subunits
  - Discourse "segments"
  - Need to detect, interpret
- Referring expressions provide coherence
  - Explain and link
- Meaning of discourse more than that of component utterances
- Meaning of units depends on context

Theoretical Differences

- Informational (Hobbs/RST)
  - Meaning and coherence/reference based on inference/abduction
  - Versus

Theoretical Differences

- Informational (Hobbs/RST)
  - Meaning and coherence/reference based on inference/abduction
  - Versus
- Intentional (G&S)
  - Meaning based on (collaborative) planning and goal recognition, coherence based on focus of attention

Theoretical Differences

- Informational (Hobbs/RST)
  - Meaning and coherence/reference based on inference/abduction
  - Versus
- Intentional (G&S)
  - Meaning based on (collaborative) planning and goal recognition, coherence based on focus of attention
- "Syntax" of dialog act sequences
  - versus

Theoretical Differences

- Informational (Hobbs/RST)
  - Meaning and coherence/reference based on inference/ abduction
  - Versus
- Intentional (G&S)
  - Meaning based on (collaborative) planning and goal recognition, coherence based on focus of attention
- "Syntax" of dialog act sequences
  - versus
- Rational, plan-based interaction

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?
- Are discourse segments psychologically real or just useful?
  - How can they de recognized/generated automatically?

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?
- Are discourse segments psychologically real or just useful?
  - How can they de recognized/generated automatically?
- How do you define and represent “context”?
  - How does representation interact with ambiguity resolution (sense/reference)

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?
- Are discourse segments psychologically real or just useful?
  - How can they de recognized/generated automatically?
- How do you define and represent “context”?
  - How does representation interact with ambiguity resolution (sense/reference)
- How do you identify topic, reference, and focus?

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?
- Are discourse segments psychologically real or just useful?
  - How can they de recognized/generated automatically?
- How do you define and represent “context”?
  - How does representation interact with ambiguity resolution (sense/reference)
- How do you identify topic, reference, and focus?
- Identifying relations without cues?
- Discourse and domain structures

Challenges

- Relations:
  - What type: Text, Rhetorical, Informational, Intention, Speech Act?
  - How many? What level of abstraction?
- Are discourse segments psychologically real or just useful?
  - How can they de recognized/generated automatically?
- How do you define and represent “context”?
  - How does representation interact with ambiguity resolution (sense/reference)
- How do you identify topic, reference, and focus?
- Identifying relations without cues?
- Computational complexity of planning/plan recognition
- Discourse and domain structures

Dialogue Modeling

- Two or more participants – spoken or text
  - Often focus on task-oriented collaborative dialogue

Dialogue Modeling

- Two or more participants – spoken or text
  - Often focus on task-oriented collaborative dialogue

- Models:
  - Dialogue Grammars: Sequential, hierarchical constraints on dialogue states with speech acts as terminals
    - Small finite set of dialogue acts, often "adjacency pairs"
      - Question/response, check/confirm

Dialogue Modeling

- Two or more participants – spoken or text
  - Often focus on task-oriented collaborative dialogue

- Models:
  - Dialogue Grammars: Sequential, hierarchical constraints on dialogue states with speech acts as terminals
    - Small finite set of dialogue acts, often "adjacency pairs"
      - Question/response, check/confirm
  - Plan-based Models: Dialogue as special case of rational interaction, model partner goals, plans, actions to extend

Dialogue Modeling

- Two or more participants – spoken or text
  - Often focus on task-oriented collaborative dialogue

- Models:
  - Dialogue Grammars: Sequential, hierarchical constraints on dialogue states with speech acts as terminals
    - Small finite set of dialogue acts, often "adjacency pairs"
      - Question/response, check/confirm
  - Plan-based Models: Dialogue as special case of rational interaction, model partner goals, plans, actions to extend
  - Multi-layer Models: Incorporate high-level domain plan, discourse plan, adjacency pairs

Dialogue Modeling Challenges

- How rigidly do speakers adhere to dialogue grammars?
  - How many acts? Which ones?
  - How can we recognize these acts? Pairs? Larger structures?

Dialogue Modeling Challenges

- How rigidly do speakers adhere to dialogue grammars?
  - How many acts? Which ones?
  - How can we recognize these acts? Pairs? Larger structures?
- Mental models
  - How do we model the beliefs and knowledge state of speakers?

Dialogue Modeling Challenges

- How rigidly do speakers adhere to dialogue grammars?
  - How many acts? Which ones?
  - How can we recognize these acts? Pairs? Larger structures?
- Mental models
  - How do we model the beliefs and knowledge state of speakers?
- Discourse and domain structures

Practical Considerations

- Full reference resolution, interpretation:

Practical Considerations

- Full reference resolution, planning:
  - Worst case NP-complete, AI-complete

Practical Considerations

- Full reference resolution, planning:
  - Worst case NP-complete, AI-complete
- Systems must be (close to) real-time

Practical Considerations

- Full reference resolution, planning: Worst case NP-complete, AI-complete
- Systems must be (close to) real-time
  - Complex models of reference -> Interaction history
    - Often stack-based recency of mention

Practical Considerations

- Full reference resolution, planning:
  - Worst case NP-complete, AI-complete
- Systems must be (close to) real-time
  - Complex models of reference -> Interaction history
    - Often stack-based recency of mention
  - Planning/Inference -> state-based interaction model

Practical Considerations

- Full reference resolution, planning:
  - Worst case NP-complete, AI-complete
- Systems must be (close to) real-time
  - Complex models of reference -> Interaction history
    - Often stack-based recency of mention
  - Planning/Inference -> state-based interaction model
  - Questions: Initiative (system/user driven?)
    - Corpus collection
    - Evaluation

Spoken Dialogue Modeling

- Building interactive spoken language systems
  - Based on speech recognition and (often) synthesis

Spoken Dialogue Modeling

- Building interactive spoken language systems
  - Based on speech recognition and (often) synthesis
- Dominated by practical considerations
  - Limitations of: speech recognizer accuracy, speed, coverage; speech synthesizer speed, fluency, naturalness; plan/intention recognition and reasoning speech and effectiveness

Spoken Dialogue Modeling

- Building interactive spoken language systems
  - Based on speech recognition and (often) synthesis
- Dominated by practical considerations
  - Limitations of: speech recognizer accuracy, speed, coverage; speech synthesizer speed, fluency, naturalness; plan/intention recognition and reasoning speech and effectiveness
- Often simplistic but implementable models

Spoken Dialogue Modeling

- Building interactive spoken language systems
  - Based on speech recognition and (often) synthesis
- Dominated by practical considerations
  - Limitations of: speech recognizer accuracy, speed, coverage; speech synthesizer speed, fluency, naturalness; plan/intention recognition and reasoning speech and effectiveness
- Often simplistic but implementable models
- Design and evaluation challenges
  - What is the best dialogue? Fastest? Fewest errors? Most “natural”?

Dialogue Management: Motivating Examples

- 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.

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

DM: Motivating Examples (Cont'd)

- 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.

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

DM: Motivating Examples (Cont'd)

- Dialogue 3:

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 playing at the Fairmont theater at 6:00 and 8:30.
U: I wanted to know about the Paramount theater, not the Fairmont 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.

From Carpenter and Chu-Carroll, Tutorial on Spoken Dialogue Systems, ACL '99

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
- Dialogue 3:
  - Mixed-initiative
  - No confirmation
  - Suggests alternative when query fails
  - More natural
  - Moderately efficient

Comparison of Sample Dialogues

- Dialogue 1:
  - System-initiative
- Dialogue 2:
  - Mixed-initiative
- Dialogue 3:
  - Mixed-initiative

Comparison of Sample Dialogues

- Dialogue 1:
  - System-initiative
  - Implicit confirmation
- Dialogue 2:
  - Mixed-initiative
  - No confirmation
- Dialogue 3:
  - Mixed-initiative
  - No confirmation

Comparison of Sample Dialogues

- Dialogue 1:
  - System-initiative
  - Implicit confirmation
  - Merely informs user of failed query
- Dialogue 2:
  - Mixed-initiative
  - No confirmation
  - Suggests alternative when query fails
- Dialogue 3:
  - Mixed-initiative
  - No confirmation
  - Suggests alternative when query fails

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
- Dialogue 3:
  - Mixed-initiative
  - No confirmation
  - Suggests alternative when query fails
  - More natural
  - Moderately efficient

Dialogue Evaluation

- System-initiative, explicit confirmation
  - better task success rate
  - lower WER
  - longer dialogues
  - fwer recovery subdialogues
  - less natural
- Mixed-initiative, no confirmation
  - lower task success rate
  - higher WER
  - shorter dialogues
  - more recovery subdialogues
  - more natural

Candidate measures from Chu-Carroll and Carpenter

Dialogue System Evaluation

- Black box:
  - Task accuracy wrt solution key
  - Simple, but glosses over many features of interaction

Dialogue System Evaluation

- Black box:
  - Task accuracy wrt solution key
  - Simple, but glosses over many features of interaction

- Glass box:
  - Component-level evaluation:
    - E.g. Word/Concept Accuracy, Task success, Turns-to-complete
  - More comprehensive, but Independence? Generalization?

Dialogue System Evaluation

- Black box:
  - Task accuracy wrt solution key
  - Simple, but glosses over many features of interaction
- Glass box:
  - Component-level evaluation:
    - E.g. Word/Concept Accuracy, Task success, Turns-to-complete
  - More comprehensive, but Independence? Generalization?
- Performance function:
  - PARADISE[Walker et al]:
    - Incorporates user satisfaction surveys, glass box metrics
    - Linear regression: relate user satisfaction, completion costs

Dialogue Management

- Controls flow of dialogue
  - Openings, Closings, Politeness, Clarification,Initiative
  - Link interface to backend systems
- Mechanisms: increasing flexibility, complexity
  - Finite-state
  - Template-based
  - Learning-based
- Acquisition
  - Hand-coding, probabilistic dialogue grammars, automata, HMMs

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/Multiparty
- How do we integrate different information sources?
  - Task plans and discourse plans
  - Multi-modal cues: Multi-scale
    - syntax, semantics, cue words, intonation, gaze, gesture
- How can we learn?
  - Cues to discourse structure
  - Dialogue strategies, models

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/Multiparty
- How do we integrate different information sources?
  - Task plans and discourse plans
  - Multi-modal cues: Multi-scale
    - syntax, semantics, cue words, intonation, gaze, gesture
- How can we learn?
  - Cues to discourse structure
  - Dialogue strategies, models

Broad Challenges

- How should we represent discourse?

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/ Multiparty

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/Multiparty
- How do we integrate different information sources?
  - Task plans and discourse plans
  - Multi-modal cues: Multi-scale
    - syntax, semantics, cue words, intonation, gaze, gesture
- How can we learn?
  - Cues to discourse structure
  - Dialogue strategies, models

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/ Multiparty
- How do we integrate different information sources?

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/ Multiparty
- How do we integrate different information sources?
  - Task plans and discourse plans
  - Multi-modal cues: Multi-scale
    - syntax, semantics, cue words, intonation, gaze, gesture
- How can we learn?

Broad Challenges

- How should we represent discourse?
  - One general model?
  - Fundamentally different? Text/Speech; Monologue/Multiparty
- How do we integrate different information sources?
  - Task plans and discourse plans
  - Multi-modal cues: Multi-scale
    - syntax, semantics, cue words, intonation, gaze, gesture
- How can we learn?
  - Cues to discourse structure
  - Dialogue strategies, models