Lecture 3

Dialogue Management

2001/2002

Lecture Plan

Attention, Intentions and the Structure of Discourse

- Three components of discourse structure: linguistic, intentional, attentional
- Examples
- Linguistic signals of discourse segmentation: cue phrases, intonation

Collagen: A Collaboration Manager for Software Interface Agents

- Basic Description
- The Theory Applied
- The Collagen Architecture

Basic reading: (Grosz and Sidner 1986)
(Rich and Sidner 1998)

Discourse Structure Theory

(Grosz and Sidner 1986)

Three inter-related discourse structuring principles:

Linguistic: discourse segments and their relations (e.g., embedding) are signaled in the linguistic form of expressions

Attentional: at every point in the discourse, a set of entities is salient (i.e., in the center of attention); there are transitions between attentional states

Intentional: each discourse segment has a unique purpose (DSP); there are relations between DSPs (satisfaction-precedence vs. dominance)

These together supply the information needed by discourse participants to determine *how an individual utterance fits with the rest*, i.e., why it was said in that context and what it means. Also, certain *expectations* about what is to come are thus formed.

(What about things that are said against our intentions, or choosing between contradictory intentions?)

Linguistic Structure

- basic elements are *utterances*, they get aggregated into *discourse segments*
- *embedding relationship* can hold between segments
- discourse segmentation has been observed across a wide range of discourse types: task-oriented dialogues, descriptions of apartments, Watergate transcripts, informal debates, explanations, therapeutic discourse, narratives
- two-way interaction between discourse segment structure and utterances
  - utterances can convey information about structure: cue phrases, intonation, etc.
  - structure can constrain interpretation of utterances: referring expressions

Example 1

B: I ordered some paint a week ago.

A: Yes

B: and I wanted to order some more

A: Yes how many tubes?

B: What's the price?

A: I'll work it out for you.

B: Thanks

A: 3 pounds

B: 3 pounds?

A: Yes

B: That's for the large tube?

A: Yes

B: I'll ring back. I wasn't sure about the price you see

A: OK

Example 2

Q. Can you please describe your house?

A. . . . then in the kitchen . . .
there's a large window which faces the backyard with two smaller windows directly flanking it and . . .
if we're facing . . . towards the backyard
now on the righthand side is . . . a sliding glass door and a small window . . .
on the left is a stove and a refrigerator . . .

Intentional Structure

- *discourse segment purpose* (DSP): the intention that leads to the initiation of a new discourse segment (other intentions are not part of Intentional Structure)
  E.g. int. that some agent: intends to perform some task; believes some prop.; intends to identify an object; knows some property of an object.

  - All intentions apart from DSPs belong to the Attentional Structure. DSPs belong to the ...?

- DSPs are intentions that are meant to be recognized, i.e. recognition of the DSP is essential to its achieving its intended effect (cf. "meaning-nn" in (Grice 1969))

- structural relations among DSPs:

  - DSP_1 dominates DSP_2 if the satisfaction of DSP_1 comes (in part) from the satisfaction of DSP_2
  - DSP_1 satisfaction-precedes DSP_2 if DSP_1 must be satisfied before DSP_2

Intentional Structure

- DSPs form a tree-structure of sub-intentions eventually grounded in communicative actions.
  - Because DSPs are part of the Attentional Structure, they lead a tree structure made up of the subintentions in the Attentional Structure.
- A discourse segment can only serve a single DSP, though a later DSP can take advantage of what has been achieved by an earlier one.
- Isomorphism between explicit realization of DSPs and embedding of discourse segments.
  - Isomorphism in the sense that whenever a different DSP is realised, we have a different, and hence embedded, discourse segment.
- Discourse understanding relies on recognizing DSPs and the structural relations between them.

Attentional State

- = an abstraction of the participants' focus/center of attention as their discourse unfolds. It serves to summarise information from form previous utterances needed in subsequent interpretations/processing.
- Modeled by a set of *focus spaces*, arranged in a *stack*.
- A focus space is associated with each discourse segment. It contains those entities that are *salient* (explicitly mentioned or implicitly involved). The lower the segment in the embedding, the higher the position of the corresponding space in the stack.
- Dynamics: i.e., it evolves with the discourse. Transition rules specify conditions for adding (=PUSH) and deleting (=POP) spaces.
- The relationships among DSPs, i.e., intentional structure, determine pushing and popping of focus spaces.

Attentional State

- While the intentional structure provides a complete record of the DSPs, the attentional state only contains information relevant to purposes in a portion of the intentional structure.
- Normally, the attentional state is empty at the conclusion of a discourse. *(really???)*
- It is the attentional state that can directly constrain the interpretation of referring expressions.

Application of the Theory: "Interruptions"

Because processing an utterance requires ascertaining how it fits with previous discourse, it is crucial to decide which parts of the previous discourse are relevant to it and which cannot be.

- **true interruption**: different unrelated purposes, different entities, i.e., separate focus spaces.
- **flashback and filling in missing places**: DSP satisfaction-precedes the DSP of the interrupted segment and is dominated by another segment's DSP
- **digression**: separate DSP, but overlapping focus spaces; Two participants, two different DSPs for the same segment.
- **semantic returns** (noninterruptions): explicit reintroduction of entities and/or DSP

Application of the Theory: Cue Words

Discourse segment boundaries and/or transitions in intentional and/or attentional structure can be signaled by linguistic means. For example:

- segment boundary:
  (i) opening: "now", "well";
  (ii) closing: "anyway", "OK"

- new dominance: "for example"

- new satisfaction-precedence: "first", "second", "further", ... , "finally", etc.

COLLAGEN

Main reading: (Rich and Sidner 1998),
Also: (Rich, Sidner, and Lesh 2000)

Basic Description

- It is a Dialogue Management system using collaboration, i.e., discourse participants coordinate their actions and divide the labour towards the achievement of shared goals.
- The particular system it is integrated in also comprises a "software agent" added to an existing GUI. (Closer look at architecture in a minute.)
- The implementation they used for testing Collagen is in the domain of travel application.
- Communication, observation and interaction channels both between the agent and the user symmetrically, as well as between them and an application window (GUI).

Basic Description

- The agent can initiate the conversation and help out with suggestions on which the next focus in planning should be. It mimics the human collaborate.
- If the user gets stuck, the agent can access the programming application interface (API) and suggest a solution specific to the users problem. (e.g. Flying with a different airline that does give itineraries for the preferred route).
- Although the final suggestion comes from an application specific recipe, the strategy to be followed itself is application independent.
  - It is part of the collaboration manager. It consists of proposing the next executable step in the current recipe.
  - The application specific recipes are part of the travel agent. Recipes come in two forms:
    - A library of steps for realising each goal
    - Rules for arbitrary pattern-action (Which step to be chosen based both on application state and on discourse)

Basic Description

- Generic methodology for each step:
  - Recipe identification for current goal (e.g. work out itinerary backwards or forwards)
  - It can be identified by asking the user.(e.g. to chose between options).
  - A goal may be achieved when all its parameters are known and all its predecessors have been achieved.
  - A goal may be performed by either CP, unless otherwise specified.

The Theory applied

- SharedPlans, i.e., the participants must have common beliefs about goals, actions, capabilities, intentions and commitments.
- At the beginning of a discourse session participants only have partial SharedPlans.
  - The collaboration is completed when they have a totally SharedPlan.
- SharedPlans have embedded SharedPlans within them; substeps in the total process.
- No framework provided for interleaving planning (deciding how the goal should be achieved) and execution (performing actions toward the decided goal).
- Attentional state: each focus state corresponds to a SharedPlan and is associated with a DSP. The repertoire of DSPs is held in the intentional state and the stack is manipulated via it (Lochbaum 1994).

The Theory Applied: Discourse State Representation

- Plan trees represent partial SharedPlans
  - They consist of alternating act and recipe nodes.
  - Every node has bindings that have to be realised before it can be realised itself.
  - Bindings have inter-constraints in the ways they can be realised, defined in their recipe library definitions.
  - Both bindings and their propagation in the tree use a truth maintenance system. Therefore, non-monotonic changes in the discourse state are enabled.
- Question: Plan trees application-independent or not?

The Theory Applied: Discourse State Representation

- Focus stack and history list interact. Items are popped off the focus stack and pushed onto the History list.

  - Only when a focus space (partial SharedPlan) does not have a parent one (that satisfaction precedes it) is it pushed onto the history list (Why?).
  - If it does have a parent partial SharedPlan, it is kept in the focus stack and treated as closed.

Processing Algorithms

- Discourse interpretation: it must work out how the current action contributes towards the realisation of the DSP.
- Five main cases. The current action:
  - directly achieves the DSP,
  - is one of the steps in one of the recipes for the current DSP (What do you think happens when the step is not in the recipe?),
  - identifies the recipe to be used,
  - identifies who should perform the step or the DSP,
  - identifies an unspecified parameter of either a step or a DSP.
- If the act each time achieves one of these cases, the act is added to the partial SharedPlan representation. If it completes the DSP, the focus stack is popped.
- If none hold, there is an interpretation and a new segment is pushed onto the stack, its first element being the act. The DSP may not be known yet.

The Collagen Architecture

- The agent is a "black box". It can be a rule based expert system, a neural net or an ad hoc collection of code. The choice is up to the implementor of every specific application.
- The current system does not provide a decision making device. (What more would we need to add to the system to enable decision making?)
- It only chooses to perform the highest action in the agenda.

The Collagen Architecture: The Basic Execution Cycle

1. The Interpretation Module updates the discourse state.
2. That causes a new agenda of possible acts towards the DSP completion to be pushed onto the Discourse Generation Module.
3. The user can select an act from the menu, which is fed by the agenda. In recent applications speech recognition and processing has been implemented, as well. (It is not clear when the agent is supposed to take the turn. Any ideas?)
4. With the selection of one of the objects in the menu, the cycle closes and a new one begins that will handle the new act.

The Collagen Architecture: Segmented Interaction History

- It provides a structural guide to the problem solving for the user.
- It serves as a menu for History-Base Transformations.
  (More about that in a sec.)
- It reflects the Linguistic Segment Structure. (Wie bitte?)

History-Based Transformations

- They transform either the application state, or the discourse state, or both.
- Only elements of open spaces can be transformed, but some transformations can be augmented, for instance, a closed one (More in a sec).
- They are applied on segments.
  - Transformation for the current segment can be chosen from the menu as well as the proposed next steps.
  - For previous segments the desired segment to be changed has to be selected first from the interaction history window.
- A copy of the focus stack (All segments to be addressed e.t.c) is kept both at the start and the end of every segment. (Why?)

History-Based Transformations: Examples

- *Stopping*: It is the simplest one and it is used by some of the more complex ones.
  - The current segment is popped off the focus stack and the application state remains the same.
  - If the segment has a parent one that dominates it, any bindings done towards the realisation of the parent segment that were based on the stopped one must be undone.
- *Returning*: Both the application and the discourse state are reset to an earlier point in the problem solving, e.g. the original airline.
- Three forms of Return: retry, revisit, undo. They follow Return.
  - In retry and revisit there are two segments after the Return; one for the transformation taking place and one for the acts being performed within it.
    * *Retry*: Used for going back to a previous DSP and approaching it a different way. The segment can be closed or open. The original recipe is abandoned. (Why?)
    * *Revisit*: Used to pick up where a previous DSP was left off. The original recipe used is preserved. Applied to closed ones. (Why?)
- Stop is done at the appropriate point in the discourse state before retry or revisit.

Task Modelling

- Artificial Discourse Language: Internal representation of dialogue acts uses the one proposed by (Sidner 1994). E.g. Propose For Accept:

  PFA(*(t, $participant_1$, belief, $participant_2$)*)

  - The way to form beliefs uses application-independent operators. Everything else is application-specific.

  PFA(37, agent, SHOULD(add-airline(*t*, agent, ua), user))
  Meaning: "Propose I add United specification."

- Recipe Library:
  - Contains recipes indexed by objectives.
  - Partially ordered sequence of acts (steps) with constraints between them comprise one recipe.

Initiative

- *Global Level*, i.e., having something relevant to say at each point. *Local Level*, i.e., When and how to say it.
- *Global Level*:
  - Claim; No need for discourse plan operator on moves that compel the agent to perform a certain act. It is build in the agenda. (True or not? Clue: There is "an explicit choice of things to say". How does the system chose between them?)
  - Negotiation: Resolving differences in belief, fundamental to collaboration (G&S, Grice). A first step done in (Sidner 1994) with the Artificial Discourse Language.
- *Local Level*:
  - Only ad hoc mechanisms.

Different Applications Mentioned

- Setting up and programming a VCR.
- Symbol Editor: Guides the user and also automatically performs many of the tedious subtasks.
- Teaching the student how to operate a gas turbine engine and generator configuration. It just describes the next step to be taken. (Claim: "Teaching and assisting are better thought of as different points of a spectrum." What is missing from Collagen for teaching to be appropriate? How easily could it be adopted?)
- Programming a home thermostat.

Questions to Take Home.

- How is G&S computationally beneficial?
- They try to reduce intentions to their interconnections. Maybe good enough for NLU, but what about NLG?
- Ideas on turn-taking: When is the Agent supposed to take the turn and how?
- What happens when a step is performed by the user and cannot be traced in the agenda?
- What would need to be added to the system to enable decisions.
- How familiar must the user be with the system? Why?

References

Grice, H. P. (1969). Utterer's meaning and intentions. *Philosophical Review 68*(2), 147–177.

Grosz, B. and C. L. Sidner (1986). Attention, intentions and the structure of discourse. *Computational linguistics 12*, 175–204.

Lochbaum, K. E. (1994). Using collaborative plans to model the intentional structure of discourse. Ph.D. thesis TR-25-94, Harvard University, Cambridge, MA.

Rich, C. and C. L. Sidner (1998, March). Collagen:a collaboration manager for software interface agents. Technical Report TR-97-21a, Mitsubishi Electric Research Laboratories.

Rich, C., C. L. Sidner, and N. Lesh (2000). Collagen: Applying collaborative discourse theory to human-computer interaction. Technical Report TR-2000-38, Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, Massachusetts 02139.

Sidner, C. L. (1994). An arificial discourse language for collaborative negotiation. In B. Hayes-Roth and R. Korf (Eds.), *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Menlo Park, California, pp. 814–819. American Association for Artificial Intelligence Press.

Also more Collaboration at http://www.cit.dk/coconut