A Method for Generating Narrative Discourse to Prompt Inferences

James Niehaus
Charles River Analytics Inc.
625 Mt. Auburn St.
Cambridge, MA, USA
jniehaus@cra.com

R. Michael Young
Department Of Computer Science
North Carolina State University
Raleigh, NC, USA
young@csc.ncsu.edu

ABSTRACT

Narratives that prompt inferences can be more interesting in that they provide the reader with the opportunity to reason about the narrative world, participating in its construction. These narratives can also be more concise and direct, as details can be filled in by the reader. On the other hand, narratives that leave out important information without the opportunity to infer this information may be incoherent. To generate narratives that prompt inferences a system must 1) employ a theory of how inferences are prompted and 2) provide a capacity for creating narratives that satisfy inference goals. This paper presents is a novel algorithm for generating discourse plans that prompt inferences according to a theory of online inferencing in narrative discourse. Though other approaches have generated narrative and discourse structures to influence the reader's perception of the narrative, this is the first approach to present an empirically based cognitive model of online inference generation. The algorithm is a partial-order planning approach to discourse generation, selecting events to tell the reader from an input story plan.

Categories and Subject Descriptors

I.2.m [**Computing Methodologies**]: Artificial Intelligence—*Miscellaneous*

General Terms

Algorithms theory

Keywords

Narrative generation, Discourse generation, Inferencing, Cognitive models of discourse comprehension

1. INTRODUCTION

Inferences in narrative comprehension are pieces of information that readers add to their understanding of the narrative to improve comprehension [7]. Narratives that prompt inferences can be more interesting in that they provide the reader with the opportunity to reason about the narrative world, participating in its construction. These narratives can also be more concise and direct, as details can be filled in by the reader. On the other hand, narratives that leave out important information without the opportunity to infer this information may be incoherent. Previous narrative generation approaches have not addressed this form of online narrative inferencing.

To generate narratives that prompt inferences a system must 1) employ a theory of how inferences are prompted and 2) provide a capacity for creating narratives that satisfy inference criteria (such as search within a space of narratives). If the generation process does not know how an inference is prompted it cannot select solutions that meet inference criteria, and generation must be able to represent and explore the space of possible narratives to find a solution. In this paper, we concentrate on plot-level online inferences that are sets of events that the reader infers. Online inferences are those that occur during active reading, not while the reader is reflecting upon what has been read. Hence the output of our algorithm is discourse structure, defining which elements are told when. We do not generate plot structure, that defines what happens in the story world, nor medium representations, e.g. how to construct well-formed natural language descriptions.

To employ a theory of how inferences are prompted, we use a computational model of the empirically supported Constructivist theory of narrative discourse comprehension [15] [7]. This theory predicts that inferences are most likely when they are necessitated and enabled. Necessitated inferences are those that achieve one of the reader's goals, such as maintaining causal coherence or predicting the actions of characters. Enabled inferences are those which are coherent within the reader's understanding of the story structure and may be quickly discovered.

To provide a capacity for creating narratives that satisfy inference constraints, we employ partial-order planning. Story plans represent the plot, or *fabula*, of the narrative and discourse plans represent constraints on the telling, or *szjuet*, of the narrative. A planning library contains templates of possible events within the narrative world. This representation is similar to that of [17] and [15]. The main contribution of this paper is a novel algorithm for generating discourse plans that prompt inferences according to a theory of online inferencing in narrative discourse.

2. RELATED WORK

INT3 2010 June 18, Monterey CA, USA

Previous work has presented an approach to *predicting* causal and intentional inferences from a given discourse structure [15]. This paper employs a similar representation of inferencing for the purposes of *generation*. This work also employs the IPOCL planner: a partial order narrative planner that attempts to satisfy rules of character intentionality [17]. IPOCL maintains frames of commitment to track the relationship between character actions and character intentions in order to increase character believability. We employ the IPOCL plan structure for our notions of character intentions.

Other related work occurs in narrative understanding systems, narrative generation systems with user models or reader models, and cognitive models of narrative discourse comprehension. Some narrative generation systems have recognized the need for a model of the user or reader to improve the quality of the static narrative or of the interaction in an interactive narrative. Finally, cognitive models provide an empirically rigorous study of how humans process narrative text. However, these models often lack the formality of a computational implementation and cannot be executed "out of the box", and few have been employed for generational purposes. The following paragraphs review relevant material in these categories.

Other narrative generation systems have incorporated implicit or explicit reader models, of which inferencing may be considered an instance. Mott's U-Director [13] employed a decision theoretic planner to choose the next best event according to criteria of narrative progression through a plot graph, story world state, and the system's model of the user's state. The user's state is defined by a model of the user's goals, beliefs, and experiential state, and more detailed user models of have been explored in subsequent research [12] [9] [11]. This user modeling is distinct from reader modeling in that it attempts to understand the actions of the user in a virtual world, not the reader's narrative comprehension. Cheong and Young's Suspenser [3] generated narrative discourses from a plot representation to fulfill a model of suspense. In Suspenser, suspense was rated by the inverse of the number of solutions available to the problems of the protagonist. If the protagonist had few or no solutions to his problems, suspense was rated as high, if he had many solutions to his problems, suspense was rated as low. Bae and Young's Prevoyant [1] use flashback and foreshadowing to reader surprise. Surprise is measured by a computation of unexpectedness, whether the reader can detect missing events, and postdictability, whether the story makes sense afterwards. Though both Suspenser and Prevoyant model some aspects of narrative comprehension for dramatic effect, neither define a model criteria for prompting online inferences of the type discussed here.

Early work on situation models (a.k.a. mental models) in narrative comprehension [2] has expanded into an active field of research that examines which information readers maintain, generate, and forget while reading. This specification of the creation and maintenance of situation models enables the prediction of reading processes that may aid narrative generation. Included in this research are examinations of reader inferencing. Graesser, Millis, and Zwaan ([8]) identify many types of inferences which may occur: "goals and plans that motivate characters' actions, character traits, characters' knowledge and beliefs, character emotions, causes of events, the consequences of events and actions, properties of objects, spatial contexts, spatial relationships among entities, the global theme or point of the text, the referents of nouns and pronouns, the attitudes of the writer, and the appropriate emotional reaction of the reader." The work in this paper is concerned with generating inferences about possible sequences of events either to satisfy the world state or to predict character actions.

Graesser, Singer, and Trabasso ([7]) present a Constructivist theory of discourse processing; the primary principle of which is that the reader attempts a *search (or effort) after meaning*. This principle is broken into three critical assumptions. *The reader goal assumption* states that "the reader constructs a meaning representation that addresses the reader's goals." *The coherence assumption* states that "The reader attempts to construct a meaning representation that is coherent at both the local and global levels". *The explanation assumption* states that "The reader attempts to explain why actions, events, and states are mentioned in the text." [7] Graesser, Singer, and Trabasso identify a number of classes of inferences that may be constructed by the reader, and provide support for each. The work presented in the present paper enacts these assumptions in the reader's search after meaning.

2.1 Two Types of Inferences

Our representation and treatment of inferences follows [15], here we provide a brief overview. The class of inferences treated here consist of single steps or sequences of steps that will be predicted or assumed to have happened by the reader. Myers et al. suggest that inferences are more often made when inferencing is *necessitated* to understand the story and *enabled*, so that the inference is not difficult to construct [14] [18]. Graesser and Clark [5] note that readers rarely make unconstrained predictive inferences and that when they are asked to, the inferences often prove to be untrue. For the restricted view of online inferencing in this work, causal and intentional inferences are constructed when they are both necessitated and enabled.

Causal inferences are steps inferred to maintain causality. Causal inferences are necessitated when the world state changes without explanation, and the reader infers that something must have happened to change the world state. Causal inferences are enabled when the preconditions of the steps in the inference are presented in the discourse and when the search for the inference is small and well constrained. Figure 1 shows a narrative with a causal inference necessitated; the reader may infer that Nathan dropped the melon.

Intentional inferences are steps inferred to fulfill a character's stated intention. Intentional inferences are necessitated when characters have unfulfilled intentions, and the reader infers that a character will act to achieve his goals. As with causal inferences, intentional inferences are enabled when the preconditions of the steps in the inference are presented in the discourse and when the search for the inference is small and well constrained. Figure 2 shows a narrative with a intentional inference necessitated; the reader may infer that Brian will order the cheeseburger.

2.2 Definitions

2.2.1 Input

The input to the generation algorithm is a representation of the story, a representation of the possible story world ac-

1. Nathan's hands were still wet from the rain.
2. He grabbed at the large melon from the market shelf.
3. As he left the market, Nathan angrily wiped melon chunks off his pants.

Figure 1: Example narrative with causal necessity.

1. Brian had been famished since he started his new, healthy diet.
2. The waiter handed him a menu.
3. At the top of the menu was a bacon cheeseburger.

Figure 2: Example narrative with intentional necessity.

tions, and a set of inferences to prompt. The story is represented as an IPOCL plan, as defined in [17], which contains the events, objects, facts, and characters as well as causal, ordering, and intentional relations between them. The possible story world actions is an IPOCL planning domain, a set of templates for creating new steps in the plan. The set of inferences is a set of plan steps and a specification of whether they are to be causal or intentional inferences.

Definition, (Causal Inference Criterion) *A causal inference criterion is a tuple* $< s, b >$*, where* s *is an IPOCL step and* b *is a set of bindings over the free variables in the step.*

Definition, (Intentional Inference Criterion) *An intentional inference criterion is a tuple* $< s, b >$*, where* s *is an IPOCL step and* b *is a set of bindings over that step.*

The causal inference criteria define which steps are to be inferred by the reader due to the causal structure of the narrative. The intentional inference criteria define steps to be inferred by the reader due to the intentional structure of the narrative. The bindings over the free variables given in the criterion may range from no bindings to completely bound.

2.2.2 *Output*

The output of the algorithm is a sequence of discourse content, a total ordering over a subset of the elements in the story. The sequence of discourse content represents a proposition level telling of a narrative which may be translated to a medium such as text or film. The sequence may be constructed from an IPOCL plan input by choosing a subset of the elements from the plan and then choosing a total order over these elements.

Definition, (Discourse Plan) *A discourse plan is a tuple,* $< S, B, O, L, C, D, I >$*, where* S *is a set of steps,* B *is a set of binding constraints on the free variables in* S*,* O *is the set of ordering constraints on steps in* S*,* L *is a set of causal links between steps in* S*,* C *is a set of frames of commitment,* D *is the subset of* S *which are causally prompted inferences, and* I *is the subset of* S *which are intentionally prompted inferences.*

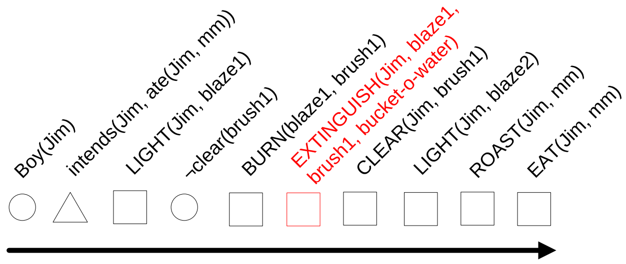

Figure 3: A Sequence of Discourse Content. Events are squares, intentions are triangles, and propositions are circles.

Definition (Sequence of Discourse Content) *A sequence of discourse content is a tuple,* $< R, S, B, O, L, C, T >$*, where* S *is a set of steps,* R *is a set of propositions from the preconditions and effects of the steps in* S*,* B *is a set of bindings on the free variables in* S*,* O *is a set of ordering constraints on steps in* S*,* L *is a set of causal links between steps in* S*,* C *is a set of intentional frames of commitment, and* T *is a total ordering over* $S \bigcup R \bigcup C$*.*

A discourse plan is an intermediate representation used by the algorithm before creating the final sequence of discourse content. Discourse plans are IPOCL plans with denotations for inferred steps which are to be removed from the final discourse. Before output, a linearization of the discourse plan (minus inferred steps) is chosen to serve as the sequence of discourse content that is the final output. Any linearization of a discourse plan is a satisfying sequence of discourse content.

The sequence of discourse content contains three types of elements: steps, propositions, and frames of commitment. The steps represent events in the narrative such as "Jim kicked the ball". The propositions represent single facts about the world such as "The vase rested on the table" and may also serve to introduce and identify objects in the world: "There was a girl named Sally". The frames of commitment identify intentions of the characters as in "Bob wanted to eat some ice cream.". The total ordering over these elements asserts the order in which they should be presented to the reader. Thus, if a ordering over step s proposition p and frame of commitment f specifies $s < p < f$, then the step s is presented, followed by the proposition p, followed by the frame of commitment f. Figure 3 shows a visualization of an example sequence of discourse content.

The sequence of discourse content in figure 3 can be used to create a narrative in a medium such as text or film by realizing each discourse element within the medium. For instance, a simple template based system might be used to translate each element into a sentence for a text, or a visual discourse generator may select a sequence of shots and actions to convey the discourse. Figure 4 shows a manual translation from the sequence of discourse content in Figure 3 to text.

3. GENERATION ALGORITHM

This section presents a partial-order planning algorithm to generate a discourse plan to prompt inferences by selecting and ordering content from a story plan. Generating a

1. There was a boy named Jim.
2. Jim wanted to eat a marshmallow.
3. He lit a fire,
4. but the surrounding brush was not properly cleared!
5. The brush caught fire.
6. Jim quickly put out the fire.
7. Then he cleared the brush out of the way.
8. He relit the fire,
9. and roasted his marshmallow.
10. He ate his marshmallow.

Figure 4: The sequence of discourse content from Figure 3 translated to text, performed manually.

discourse plan is accomplished by a search in the space of partial plans. Each plan maintains a list of flaws, and children in the search space are generated by choosing a flaw and applying the appropriate resolution method. Causal and intentional inference criteria are treated as flaws requiring that specific steps be inferred. These flaws are solved by including elements from the event log to satisfy necessity and enablement criteria.

The necessitation and enablement criteria differentiate the inference process from simple precondition and effect chaining of planning operators. The planner must create a suitable "hole" within the causal and intentional structure of the IPOCL plan in which the reader can add the appropriate inference. Thus, the planner must motivate the inference, through necessitation, and lead the reader close to the inference through enablement. Informally, the planner necessitates inferences by either presenting an unfulfilled intention or an unfulfilled action precondition to the reader. The planner enables the inference by providing enough information that the possibilities for the inference are few. The reader may then may the small leap that constructs the inference.

The algorithm operates as follows. First the inference criteria are addressed. The criteria for causal inferences specify a single step in an inference. To fulfill one of these criterion, the step must be part of a inference that is both enabled and causally necessitated. Likewise, the criteria for intentional inferences specify a single step. To fulfill and intentional inference criterion, the step must be part of an inference that is both enabled and intentionally necessitated.

The starting plan is an empty discourse plan, and refinement operators are used to generate children to satisfy flaws in the parent plan. Flaws in the discourse plan represent comprehension criteria that have not yet been satisfied or potential consistency problems. The user specifies the initial causal inferencing, intentional inferencing, and focus criteria, and each criteria is converted to a flaw in the discourse plan.

Discourse plans are generated in partial-order, then linearized to form a sequence of discourse content. A discourse plan is an IPOCL plan that contains steps, bindings, causal

Let P be a Discourse Plan, R be an IPOCL Plan, F be a list of flaws, and Δ be an operator library.
Algorithm Refine-Discourse-Plan(P, R, F, Δ)

1. **Termination Check** If F is empty, return P.
2. **Flaw Selection and Refinement** Let f be either 1) a causal inference flaw $f = \{fs, fb\}$, 2) an intentional inference flaw $f = \{fs, fb\}$, 3) a link to intention flaw $f = \{fs, fc\}$, 4) an open precondition flaw $f = \{fs, fp\}$, or 5) an open effect flaw $f = \{fs, fe\}$. Switch on the flaw type:
   (a) **Prompt Causal Inference** Let s be a new step of type fs with bindings fb, and e be a chosen effect of s. Add s to P with bindings fb. Create a new open effect flaw in F for e. Create new open precondition flaws in F for every precondition of s.
   (b) **Prompt Intentional Inference** Let s be a new step of type fs with bindings fb, e be a chosen effect of s, and c be a frame of commitment from R. Add s to P with bindings fb, and add c to P if it does not exist in P. Create a new link intention flaw in F for e and c. Create new open precondition flaws in F for every precondition of s.
   (c) **Link Open Effect** Either 1) let s be a step with precondition p that unifies with an effect of fs. Add a protected causal link from fs to s in P. Or 2) let s_{add} be a new step with precondition p_{add} that unifies with fs. Add a protected causal link from fs to p_{add} and add a new link open effect flaw in F for s_add. If s_{add} does not exist in R add new open precondition flaws in F for s_{add}.
   (d) **Link to Intention** Let g be the goal of fc or a precondition of fc. Either 1) let e be an effect of fs that unifies with g. Add a protected causal link from e to g in P. Or 2) let s_{add} be a new step with effect e_{add} that unifies with g. Add a protected causal link from e_{add} to g and add a new link intention flaw in F for fe and s_add.
   (e) **Establish Open Precondition** Resolve in the standard manner (see [16]) by adding a causal link l from either an existing step or a new step. If l is from a new step s_{add} that does not exist in R create new open precondition flaws for s_{add} (but do not otherwise).
3. **Consistency Check** If P does not exist or is inconsistent, backtrack.
4. **Recursive Invocation** Call Refine-Discourse-Plan(P, R, F)

Figure 5: Algorithm for refining a discourse plan to remove all flaws.

links, orderings, and frames of commitment. A discourse plan may also include inferred steps, which are not included in the final sequence of discourse content. These inferred steps are used as place holders to ensure that inferences are enabled and necessitated. The final discourse plan must be constructed such that any linearization is a satisfying sequence of discourse content.

Figure 5 is the planning algorithm for generation. The algorithm selects a flaw and calls the appropriate routine to refine the plan to remove the flaw. The refinement process may create new flaws, such as the open precondition or open effect flaws that are never specified by the user. The search proceeds until a solution is found, a search limit is reached, or the search space is fully exhausted.

Causal Inference flaws denote steps that are to be causally inferred. Intentional Inference flaws denote steps that are to be intentionally inferred. Link to intention flaws denote steps that are to be causally linked to an intention (for intentional inferencing necessitation). Open precondition flaws denote preconditions of inferred steps that are to be established in the plan (for inferencing enablement). Open effect flaws denote effects of inferred steps that are to be causally linked to non-inferred steps (for causal necessitation). Lastly, focus flaws denote elements which are to have a minimum focus.

Discourse plans with Causal Inference flaws are refined to include the inferred step as a place holder for further refinement. After the step is inserted, open precondition flaws for each precondition are added and a single open effect flaw is created to ensure the inference is necessitated. If a solution is found further down in the tree, it will have satisfied all of the preconditions of the inferred step and used one of its effects to establish a precondition of a non-inferred step, either directly or through a causal chain.

Discourse plans with Intentional Inference flaws are also refined to include the inferenced step as a place holder for further refinement. After the step is inserted, open precondition flaws for each precondition are added and a link to intention is created to ensure the inference is necessitated. If a solution is found further down in the tree, it will have satisfied all of the preconditions of the inferred step and used one of its effects to establish a goal of a frame of commitment, either directly or through a causal chain.

Discourse plans with Link to Intention flaws are refined to create a causal chain between a step and a frame of commitment. The step may be directly linked to the frame of commitment in a single refinement or the refinement may add an inferred step to the causal chain reaching from the frame to the step. If a new step is added, then 1) new Open Precondition flaws are created for each of its preconditions and 2) a new Link to Intention flaw is created to link the original step to the new step. If a solution is found further in the tree, it will have satisfied all of the preconditions of the inferred steps and it will have causally linked the original step to the frame of commitment.

Discourse plans with Open Precondition flaws are refined, similar to the standard manner [16], to add a causal link to the precondition. The link may be created between existing steps in the plan, or a new step may be added at the head of the link. If the new step is taken from the event log, then every ordering with both steps in the discourse plan is added to the discourse plan to maintain consistency with the event log. If the new step is inferred, then new Open Precondition flaws are created for each of its preconditions. Hence, chains of inferred steps must eventually have their preconditions satisfied by steps from the event log.

Discourse plans with Open Effect flaws are refined to add a causal link from the effect. The link may be created between existing steps in the plan, or a new step may be added at the tail of the link. As in Open Precondition refinements, If the new step is taken from the event log, then every ordering with both steps in the discourse plan is added to the discourse plan to maintain consistency with the event log. If the new step is inferred, then 1) new Open Precondition flaws are created for each of its preconditions other than the one being linked and 2) a new Open Effect flaw is created for the new step. Thus, chains of inferred steps must eventually have an effect satisfy a precondition of a non-inferred step.

- ...
- Intends (Sally (has Sally Bluedress))
- WITHDRAW-MONEY (Sally Bank)
- PICKPOCKET (Robbie Sally Money1 Mainstreet Darkalley)
- REPORT-STOLEN (Sally Sheriff Money)
- BUY-DRINKS-FOR (Robbie Barney Money1)
- LAY-TO-DRUNKEN-SLEEP (Robbie Barney Darkalley)
- TAKE-THING-OFF-SLEEPER (Robbie Barney Sixshooter Darkalley)
- ...

Figure 6: A selection of some of the events in the story. Events leading up to a bank robbery in a western town. Each event represents an IPOCL plan step or frame of commitment.

4. EXAMPLE

This section presents an example to illustrate the operation of the algorithm. Before the algorithm is executed, the domain author must define the inputs. The domain author creates an IPOCL plan (either by manual encoding or using a tool to generate the story) representing the happenings of a wild west town leading up to a bank robbery; Figure 6 lists some of these events. The domain author provides a planning library for this domain defining templates for events. Finally, the domain author must choose the inference criteria that the planner will attempt to satisfy. Examining the story, the domain author decides upon the following inference goals, from which inference flaws are defined.

1. The reader should causally infer that Robbie pickpocketed Sally.
2. The reader should intentionally infer that Robbie will hold up the bank.

These flaws, the story, and the planning library are input into the algorithm, and the search for a discourse plan proceeds as follows. Beginning with an empty discourse plan,

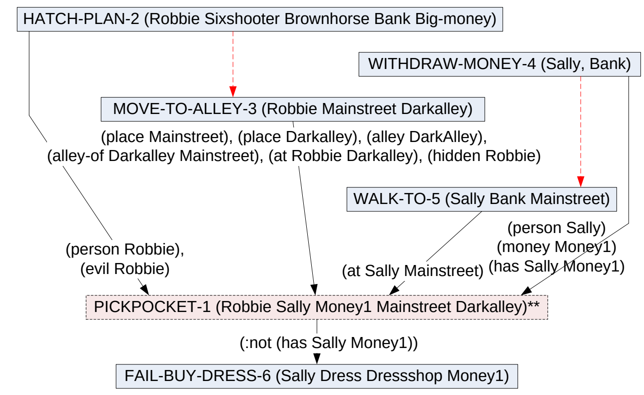

Figure 7: The Discourse Plan after resolving the causal inference flaw.

the algorithm chooses the causal inference flaw that indicates an inference of Robbie pickpocketing Sally. The pickpocket step is inserted into the discourse plan, and new flaws are created for the open preconditions of the pickpocket step and to link one of the effects of the step. One of the effects of pickpocket is that Sally no longer has any money, and the step FAIL-BUY-DRESS-6 is found in the story with a matching precondition. By including FAIL-BUY-DRESS-6, the reader is alerted that Sally does not have any money. This step is added to the discourse plan, but because it is present in the original story plan and the reader is not required to understand how it can occur, no new open precondition flaws are created.

Next, the open preconditions of the pickpocket step are addressed. To enable the pickpocket inference, the reader must be aware that each precondition is satisfied (e.g. Sally has some money, she is on Main Street, Robbie is hidden in the dark alley, the dark alley is next to Main Street, Robbie is a disreputable character). Steps from the story plan are found to fulfill each of these preconditions. No new open precondition flaws are created for these steps because they occur in the story plan. As described above, orderings between steps are preserved to maintain the coherence of the discourse plan. For instance, because Sally withdraws money before she walks to Mainstreet in the story plan, an ordering constraint is placed between these events in the discourse plan. Figure 7 shows the resulting intermediate plan; the preconditions of the pickpocket step are linked, and the causal inference is now both necessitated and enabled.

After the causal inference flaw and ensuing flaws are addressed, the algorithm attempts the intentional inference flaw that indicates an inference of Robbie holding up the bank. The hold-up step is inserted into the discourse plan, and new flaws are created for the open preconditions and to link the step to a frame of commitment (denoting a character's intention). First, steps are included or reused to satisfy the open preconditions of the hold-up step. Next, the algorithm addresses the link to intention flaw by attempting to find an effect of the hold-up step that unifies with a goal for one of the story's frames of commitment. However, there is no such match, and the inference must be extended to a second inferred step. The algorithm inserts a new inferred step that unifies with Robbie's frame of commitment to have a lot of money: Robbie collects money from the bank heist. Flaws for the open preconditions of the collect-money step

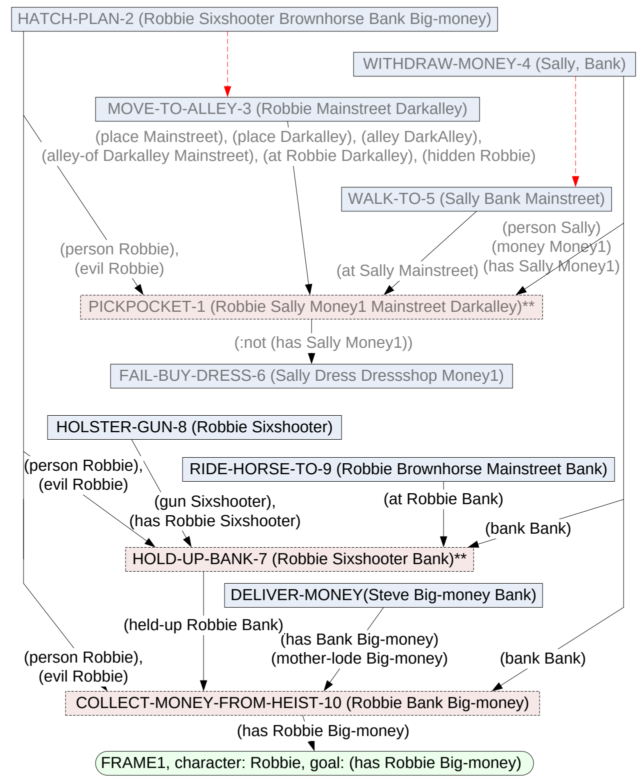

Figure 8: The Discourse Plan after resolving the intentional inference flaw.

are created as well as a flaw to link the hold-up step to the collect-money step. A unifying effect the hold-up step is found to satisfy a precondition of the collect-money step (that the bank must be held up before a robber can collect money), and the other open preconditions are satisfied by adding and reusing steps from the story plan.

Figure 8 shows the resulting final discourse plan with no flaws; the preconditions of the hold-up and collect-money steps are linked, and the intentional inference is now both necessitated and enabled. A sequence of discourse content satisfying the original inference criteria can be created by selecting a linearization of this plan and removing all of the inferred steps. An example output, using text templates for steps and frames of commitment, is given below.

"Sally withdrew some money from the Town Bank. She walked out of the Bank onto Main Street. Robbie hatched an evil plan. Robbie hid in the dark alley next to Main Street. Sally reached in her purse to pay for her new dress, but her money was gone! Robbie holstered the six shooter. Robbie wanted gold, lots of gold. Steve delivered the gold bars to the Town Bank. Robbie rode his horse down Main Street to the Bank."

5. DISCUSSION

To make the workings of the inference generation clear, the examples in this paper have largely been simplistic, toy examples. However, causal and intentional inferences are not restricted to such simple texts. Consider this excerpt from Sir Arthur Conan Doyle's "The Adventures of Sherlock Holmes", in which Watson (narrating) is passing Holmes's abode:

"I was seized with a keen desire to see Holmes again, and to know how he was employing his extraordinary powers. His rooms were brilliantly lit, and, even as I looked up, I saw his tall, spare figure pass twice in a dark silhouette against the blind. He was pacing the room swiftly, eagerly, with his head sunk upon his chest and his hands clasped behind him. To me, who knew his every mood and habit, his attitude and manner told their own story. He was at work again. He had risen out of his drug-created dreams and was hot upon the scent of some new problem. I rang the bell and was shown up to the chamber which had formerly been in part my own." [4]

This skillful introduction prompts an intentional inference concerning Watson. The reader may infer, correctly, that Watson will meet Holmes fondly and discuss old times. This passage also enables further causal and intentional inferences concerning Holmes, himself. The narration has identified Holmes's current task and referenced possible intentions (solving a new crime). The forthcoming interaction between Holmes and Watson is placed in this context, and new causal inferences that fill in intervening actions are enabled (e.g. Holmes has acquired clues from a crime scene after beginning an investigation). The astute reader will not be surprised when the characters quickly begin discussing crime scene clues after a brief reminiscence of past encounters.

Though an inference may be predicted successfully for a large number of readers, some readers may not make even the most strongly predicted inference. Causal and intentional inferences, as presented in this paper, require the reader to understand the context, understand the character intentions, agree with the author about how the narrative world operates, and agree with the author about the most plausible inference. Numerous misunderstandings can occur during reading. The reader may not have the appropriate background to make predictions about the narrative world, or the operation of the narrative world may be underdefined. Inferences can occur, however, when the author establishes common ground with the reader, perhaps through a shared view of the world or through explanation and examples within the narrative. The inferences prompted by the technique in this paper assume that this common ground has been established, and that the reader would describe the operation of the narrative world using similar operators as given to the discourse planner.

6. CONCLUSIONS AND FUTURE WORK

This paper presents a method for generating sequences of discourse content to satisfy two types of online inference criteria. Though other approaches have generated narrative and discourse structures to influence the reader's perception of the narrative [3] [1] [13] [12] [9] [11], this is the first approach to present an empirically based cognitive model of online inference generation. The algorithm employs a partial-order planning approach to discourse generation, selecting events to tell the reader from an input story plan.

The approach in this paper might be extended to encompass other aspects of narrative discourse comprehension including focus, coherence, and cohesion. Focus measures the salience of objects, facts, and events in the reader's mind [18]. Elements which have more focus throughout the story are often better comprehended and remembered. Coherence is a measure of the ability for the reader to make sense of the narrative [6]. The approach in this paper maintains local coherence between individual events in the discourse plan, improving the reader's ability to make and understand the relevant inferences, but does not provide a structure for global coherence, explaining why the events are included. Cohesion is a measure of the local interrelatedness of the elements of the story [6]. Cohesive narratives do not jump wildly from topic to topic, but center on a few elements for extended sections.

While this paper presents a solution based on the Constructivist Theory of narrative discourse comprehension, other competing and complimentary cognitive models of discourse comprehension may be the subject of future work in narrative generation. Memory Resonance models like that of [10] predict inferences and focus based on local coherence and the activation of elements due to relational links. These models may include restrictions imposed by the constraints of short term memory and by other limited cognitive capabilities. Embodied Cognition models like that of [19] predict the activation of perception and sensory properties of objects and events as if the reader were in the narrative world, experiencing the events with his or her own senses. Models such as these may inform future narrative generation systems, providing an empirical basis to support and evaluate implementations.

7. REFERENCES

[1] B.-C. Bae and R. M. Young. A use of flashback and foreshadowing for surprise arousal in narrative using a plan-based approach. In *Proceedings of the International Conference on Interactive Digital Storytelling (ICIDS 08)*, Erfurt, Germany, November 2008.

[2] G. Bower and D. Morrow. Mental models in narrative comprehension. *Science*, 247(4938):44, 1990.

[3] Y.-G. Cheong and R. M. Young. Narrative generation for suspense: Modeling and evaluation. In *Proceedings of the International Conference on Interactive Digital Storytelling (ICIDS 08)*, Erfurt, Germany, November 2008.

[4] A. Doyle. *The Adventures of Sherlock Holmes.* Wordsworth Editions Ltd, 1992.

[5] A. Graesser and L. Clark. *Structures and procedures of implicit knowledge.* Ablex, Norwood, NJ, 1985.

[6] A. Graesser, D. McNamara, M. Louwerse, and Z. Cai. Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36:193–202, 2004.

[7] A. Graesser, M. Singer, and T. Trabasso. Constructing inferences during narrative text comprehension. *Psychological Review*, 101(3):371–395, 1994.

[8] A. Graesser, M. Singer, and T. Trabasso. Discourse comprehension. *Psychological Review*, 48:163–189, 1997.

[9] S. Lee, S. McQuiggan, and J. Lester. Inducing user affect recognition models for task-oriented environments. In *Proceedings of the Eleventh International Conference on User Modeling*, pages 380–384, Corfu, Greece, 2007.

[10] G. McKoon and R. Ratcliff. Inference during reading. *Psychological Review*, 99(3):440–466, 1992.

[11] S. McQuiggan and J. Lester. Modeling and evaluating

empathy in embodied companion agents. *International Journal of Human-Computer Studies*, 65(4):348–360, 2007.

[12] B. Mott, S. Lee, and J. Lester. Probablistic goal recognition in interactive narrative environments. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 187–192, Boston, Massachusetts, 2006.

[13] B. Mott and J. Lester. U-director: A decision-theoretic narrative planning architecture for storytelling environments. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multiagent Systems*, pages 977–984, Hakodate, Japan, 2006.

[14] J. L. Myers, M. Shinjo, and S. A. Duffy. Degrees of causal relatedness and memory. *Journal of Memory and Language*, 26:453–465, 1987.

[15] J. Niehaus and R. M. Young. A computational model of inferencing in narrative. In *Proccedings of the AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II*, Stanford, CA, 2009.

[16] J. Penberthy and D. Weld. Ucpop: a sound, complete, partial order planner for adl. In *Proceedings Third International Conference on Principles of Knowledge Representation and Reasoning (KR-92)*, pages 103–114, 1992.

[17] M. O. Riedl and R. M. Young. An intent-driven planner for multi-agent story generation. *Proceedings of the 3rd International Joint Conference on Autonomous Agents and Multi Agent Systems, New York*, 2004.

[18] R. A. Zwaan and G. A. Radvansky. Situation models in language comprehension and memory. *Psychological Bulletin*, 1998.

[19] R. A. Zwaan, R. A. Stanfield, and R. H. Yaxley. Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13(2):168–171, 2002.