STORY ANALYSIS

Story analysis is the subfield of discourse understanding (qv) having to do with understanding stories—coherent narratives in written form. A story analysis program may demonstrate its analysis in two ways. First, it may produce a paraphrase or summary of the story that includes information not explicitly mentioned in the text. Second, it may answer questions about the story. For example, given the simple story:

> The princess was asleep in the castle when a dragon came and took her away. The knight searched far and wide for the princess. Finally he found her in the dragon's den. After a mighty fight, he slew the dragon. The princess's father, the king, rewarded the knight with gold pieces and his daughter's hand.

A story analysis program might produce a summary, and answer questions as follows:

> Summary:
> A dragon kidnapped a princess.
> A prince rescued her, and they were married.
> Q1: What happened after the knight killed the dragon?
> A: He returned her to the castle.
> Q2: Why did the prince kill the dragon?
> A: Because he wanted to save the princess.

For both questions, the answer is not explicitly stated in the text, but it can be inferred with a high degree of confidance. There are other possible summaries and answers. Question Q2 might have been answered "Because dragons are evil, and they must be punished." This reveals a level of understanding that is sufficient for the story at hand, but would not yield a complete analysis of a story in which dragons are protagonists, or in which evil-doers go unpunished.

Story Analysis programs are interesting because they can provide insight into the structure of stories, the notion of coherence, and the interaction between events, goals, and plans. Constructing a story analysis program requires complete, explicit theories for all three of these components, and the theories can be easily tested by observing the programs responses to sample stories and questions. Improper responses can yield clues to improved theories, and proper responses serve as evidence supporting the theory.

Unfortunately, there is a vast amount of knowledge necessary to understand even a simple story. Thus, the failure of a program to properly understand any individual text might be due to a minor piece of missing information, rather than a major flaw in the theory. It is difficult to distinguish these cases. Similarly, given that a program performs well on one text, it is difficult to predict how it will perform on other texts. These difficulties place constraints on the usefulness of story analysis programs as theoretical tools.

It should be mentioned that there is a companion line of research, story generation, that makes use of much of the same theoretical tools. A story generation program takes as input a description of a situation and some characters, and produces as output a story about the characters in that situation, introducing new characters and plot twists as necessary. This line of work will not be covered in this entry, but representative work includes Meehan [**?**], Dehn [**?**] and Lebowitz [**?**].

Although there are few practical successes at this point, there is hope that discourse understanding technology may lead to practical programs which scan newswire stories, encyclopedia entries, scientific article abstracts, and other archival sources to retrieve useful information based on a thorough understanding of the texts. This could lead to a vast improvement over current techniques based on keyword search.

Story Analysis: Early History

The psychological and computational study of story analysis can be said to have begun with Bartlett's *Remembering* [**?**], a far-seeing study of how appropriate schemata are used in the recall of stories. Since then we have seen a number of disparate approaches to the problem, including a series of computer models which convincingly mimic human understanding capabilities on certain examples—but fail completely on others.

Another landmark analysis was by Propp [?], who proposed certain formal properties shared by all folk-tales. This prompted the study of story grammars, a theory which treated stories as "big sentences"—units of text with a structure that could be parsed, and hence could be grammatical or ungrammatical. The units of the grammar included settings, events, goals, attempts, and so on. The story grammar approach was introduced by Lakoff [?] and Rummelhart [?], and was adopted whole-heartedly by a number of psychologists. A good book-length summary is Mandler [?].

In experiments by Mandler [?] and Thorndyke [?], subjects were presented with a story, and later asked to recall it. The subjects remembered goal-directed actions better than non-goal directed actions; they remembered high-level actions better than low-level ones; and they tended to make recall errors for actions that appeared in a non-canonical position, moving the actions back in place. This was taken as evidence for a plan-based story grammar theory. The idea was that subjects formed a mental representation of the story consisting of a parse according to the story-grammatical rules.

However, a subsequent experiment by Lichtenstein and Brewer [?] provided another explanation. They showed subjects videotapes of goal-directed activities, such as setting up a slide projector. Recall of these videotapes showed substantially the same effects as recall for stories. This suggests that the recall of stories is due at least in part to a general capacity to understand purposeful action, rather than a story-specific grammar. Besides this capacity to understand actions, story analysis also depends on general language skills, and on story-specific conventions. For example, it is a convention of fables that they often start with "Once upon a time." Many fables involve groups of three: the king has three children, and the third succeeds where the first two fail. It is not clear if this is just a story-telling convention, or if it is routed in some more fundamental property of human cognition (perhaps three is just the right number for the memory we have).

Computational vs. Non-Computational Approaches

There is a major split between story analysis researchers with different backgrounds. They disagree on the questions of methodology, theoretical assumptions, and goals. In psychology and text linguistics, the methodology is: given a story and an interpretation of that story, describe the processes that could have lead to that interpretation. Many of these researchers adopt the story grammar approach. Some psychologists, following Bartlett, concentrate on the processes. For others, the goal is to describe and characterize stories, so the issue of "storyness" vs. non-storyness is often an important concern. In AI, the methodology is: given a text, specify the knowledge and processes that could lead to a reasonable interpretation of that text. Many AI researchers adopt a schema-interpretation approach, while others work within a more general abductive reasoning paradigm. The goal is to characterize the interpretation process, and therefore often ignores the question of storyness, assuming that the program will only be presented with valid stories. Instead, the key issues are semantic and pragmatic interpretation, and the recognition of plans and other known concepts.

An example of the two camps arguing at cross-purposes is [?]. In this article, Wilensky attacks story grammars, and suggests a theory of reasoning and interpretation in its place. Various respondents defend their position assuming their own research paradigm, and thus each side claims victory in the end.

In this entry we concentrate on the computational approach, while recognizing that non-computational work such as that by Grimes, [?], Flammer [?], van Dijk [?] and Kintsch [?, ?] and has been influential to the computationalists, in addition to being worthy work in its own right.

Script-Based Analysis

The first outline of a computational theory of story analysis was Charniak's thesis [?]. Charniak pointed out many problems, but never implemented anything beyond the prototype stage. Thus Cullingford's SAM (Script Applier Mechanism) program [?, ?, ?] was the first real story analyzer. SAM's basic algorithm was

to identify an appropriate schema, or script, (qv) and then to interpret all subsequent information as "filling in the blanks" in that script. The following is an example of SAM'S capabilities:
Input Text:

John went to a restaurant. He ordered a hot dog. the waiter said they didn't have any. He asked for a hamburger. when the hamburger came, it was burnt. He left the restaurant.

Question-Answering Output:

Q: Did John sit down in the restaurant?
A: `Probably.`

Q: Did John order a hot dog?
A: `Yes.`

Q: What did the waiter serve John?
A: `The waiter served John a hamburger.`

Q: Did John pay the check?
A: `No.  John was angry because the hamburger was overdone and so he left the restaurant.`

The question-answering program used to demonstrate SAM's analysis was designed by Lehnert [**?**].

On the surface, this seems quite impressive. First, SAM has recognized the appropriateness of the restaurant script. Second, the roles of waiter, customer, and meal have apparently been integrated into this schema. Third, steps of the script that were not explicitly mentioned in the text—sitting down, leaving without paying—were inferred by SAM. A number of subtleties are required to make this possible; for example, SAM has to match the input "the hamburger came" to the question "what did the waiter serve" to come up with the answer "the waiter served John a hamburger." This is possible because the restaurant script includes a step where the order is brought to the table, and the agent of that step is, by default, the waiter.

Unfortunately, SAM, like most AI programs, was not made available for peer review, and there was no formal experimental tests of its abilities. It performed well on pre-selected stories, but could not handle stories that deviated from known scripts, that did not make the relevant script obvious, or that involved more than one script. SAM also had an extremely limited syntactic repertoire. Despite its limited abilities, SAM served as an early example of a working computerized story analyzer, and prompted further research. Psychologists were interested in how scripts were represented and used by humans [**?**]. Other script-based programs were constructed. For example, FRUMP [**?**] (qv) skimmed newspaper articles from the UPI news wire, such as the following:
Input Text:

MOUNT VERNON,IL (UPI) A SMALL EARTHQUAKE SHOOK SEVERAL SOUTHERN ILLINOIS COUNTIES MONDAY NIGHT, THE NATIONAL EARTHQUAKE INFORMATION SERVICE IN GOLDON, CO, REPORTED.

SPOKESMAN DIN FINLEY SAID THE QUAKE MEASURED 3.2 ON THE RICHTER SCALE, "PROBABLY NOT ENOUGH TO DO ANY DAMAGE OR CAUSE ANY INJURIES." THE QUAKE OCCURRED ABOUT 7:48 PM CST AND WAS CENTERED ABOUT 30 MILES EAST OF MOUNT VERNON, FINLEY SAID. IT WAS FELT IN RICHLAND, CLAY, JASPER, EFFINGTON AND MARION COUNTIES.

SMALL EARTHQUAKES ARE COMMON IN THE AREA, FINLEY SAID.
Output:

```
SELECTED SKETCHY SCRIPT $EARTHQUAKE

ENGLISH SUMMARY:

THERE WAS AN EARTHQUAKE IN ILLINOIS WITH 3.2 RICHTER SCALE READ-
  ING.
```

FRUMP is discussed further in the entry for scripts.

Goal-Based Analysis

If SAM was a response to Bartlett's schemata, then the next program to come out of the Yale school can be seen as an answer to Propp's characterization of goal-based stories. Wilensky's PAM (Plan Applier Mechanism) [**?**, **?**] processed non-stereotypical stories by tracking character's goals, plans, and actions. For example:

Input Text:

John wanted money. He got a gun and walked into a liquor store. He told the owner he wanted some money. The owner gave John the money and John left.

Question-Answering Output:

Q: Why did John get a gun?
A: `Because he wanted to rob a liquor store.`

Q: Why did the shopkeeper give John the money
A: `Because the shopkeeper didn't want to get hurt.`

Q: Why did John leave?
A: `Because John didn't want to get caught.`

PAM has no script for 'robbing a liquor store,' but it does know that people often want money, that stores have money, that a gun can be used to threaten bodily harm, that people want to avoid bodily harm, and so on. PAM attempts to explain each input either as a goal for some character, as a step in a known plan, or as a sub-step of some intermediate plan for a known goal. This can involve a complex explanation of arbitrary length. Consider the input:

Willa was hungry. She picked up the Michelin guide.

For the first sentence, PAM supposes that Willa will have the goal of satisfying her hunger. This is a known goal, and so needs no further explanation. For the second sentence, PAM builds a chain of seven inferences: (1) Picking up the guide is a plan for possessing it. (2) Possessing can be a sub-goal of having, so PAM hypothesizes that Willa wants to have the guide. (3) Possessing an object with a known function can be a sub-goal of performing the function, in this case, reading the guide. (4) Reading is a plan for knowing new information. The Michelin guide is a source of information about restaurants (among other things), so PAM hypothesizes that Willa wants to know the location of a restaurant (and makes other hypotheses as well). (5) Knowing a restaurant's location is instrumental to going there. (6) Being at a restaurant is instrumental to eating there. (7) Eating would satisfy Willa's known hunger goal. Thus, PAM has arrived at an explanation.

PAM goes beyond SAM in being able to process non-stereotypical stories, but it suffers from brittleness in many of the same ways. Perhaps more important than the PAM program itself was the theory of goal interaction it was based upon [**?**]. Wilensky discusses a number of relations that goals can have to one another. For example, there can be conflict (or concord) between two goals held by the same person (or by different persons). Each of these four situations gives rise to various strategies. Consider the example:

> John wanted to watch television but Mary wanted to go out to a movie. John told Mary their car should not be driven before replacing the shocks.

To understand this story, we must first understand that a limited resource—time—is in contention. We assume there is not time to both watch television and go out to a movie. Thus, John and Mary are in a goal competition situation. John has surmised that Mary has planned to drive the car, and he has constructed a counter plan to convince her not to. If Mary accepts his argument, it means that she believes he is being truthful (even though she knows that lying is a possible means of convincing) and that her goal of seeing a movie is less important than her goal of preserving the health of the car and its occupants.

The entry on frames discusses Wilensky's work from another perspective.

Integrated Analysis

The two programs we have reviewed championed different knowledge structures: scripts for SAM, and plans and goals for PAM. Both programs were successful in pointing out how important this knowledge is, but neither program was able to handle stories that strayed outside their area of expertise. SAM could not handle the Liquor Store or Michelin Guide story because it lacked the appropriate scripts. PAM, could handle a simple story like "Pat hit Kim. She cried." because it had knowledge of character's goals and reactions. But it could not a handle a superficially similar story like "The wind blew the vase off the table. It broke." because the story involves no animate characters, and hence no goals.

A more general and robust program would have to be able to handle a number of different knowledge structures. This capability was provided by Dyer's BORIS program [**?**]. BORIS was *integrated* in that it combined rules for dealing with different knowledge structures at the same time. Where SAM looked for scripts and PAM looked for plans and goals, BORIS searched for an explanation at each of four levels: scriptal, goal/plan, thematic, and role. For example, in the main example Dyer uses, two old friends meet for lunch to discuss some business. The meeting is understood by BORIS three ways: in terms of the restaurant script, the friendship theme, and the businessman role. While this approach may be a good heuristic in many cases, it will fail in others. An event could just as easily have multiple explanations at the same level. For example, in "John was hungry when it started to rain. He ducked into a restaurant", a single action can be interpreted as a plan towards two goals: staying dry and having food.

BORIS was a triumph in its coverage of disparate knowledge structures. In fact, it can be seen as primarily a theory of memory organization and access, and only secondarily as a story analyzer. One of the interesting facts about BORIS is that its memory of a story can be affected by the questions it is asked. In effect, it fills in part of its representation of the story as a by-product of answering questions. This is in accord with psychological experiments by Bartlett and by Loftus.

BORIS incorporates and distinguishes seventeen classes of knowledge: object primitives, scripts, settings, goals, plans, affects, themes, interpersonal relationships, physical states, events, social acts, memory organization packets, thematic abstraction units, scenes, scenarios, reasoning, and beliefs. These structures interact only in certain predefined ways. For example, inferences can be made to connect events to goals (as in PAM), but emotional affect can never be directly related to settings. Thus, if given the passage "At the restaurant Bill punched John in the mouth. John got mad." and the question "How did John feel at the restaurant?", BORIS would not be able to directly recall John's affect. Instead, it would have to first recall the restaurant setting, then the restaurant script. From that BORIS would recall that the punching event occurred as an unexpected event in the restaurant script and infer some goal of John's, and finally infer his affect.

Despite—or perhaps because of—this vast and varied amount of knowledge, BORIS was in some ways even more brittle than its predecessors. Each knowledge type—indeed, each allowable combination of knowledge types—had a set of rules associated with it. Some sample rules are shown here:

```
IF an act at the destination setting of a transition scenario
   is enabled by that setting
THEN build an enables link between the goal achieved
   and the change of proximity goal in the scenario

IF the word just read is a pronoun
   and a human is found with matching gender and case
THEN bind the concept to that human

IF the word just read is 'glass' THEN
IF it is followed by a liquid
   THEN glass is used as a measure (e.g. a glass of coke)
   ELSE glass is used as a material (e.g. a glass plate)
```

```
IF the story refers to a meal
  and a human is found modified by 'with'
THEN that human is an eater in the meal
```

The rules were too *ad hoc*, and the interactions too complicated, to allow the rule base to scale up gracefully. Whereas SAM and PAM handled dozens of stories (admittedly short and contrived ones), BORIS was shown to answer questions on only two stories, both involving the same subject matter. BORIS's capabilities were deep, but not broad. Clearly, we need either a sophisticated learning mechanism to acquire the knowledge BORIS needs, or else an entirely new approach.

To appreciate the ramifications of continuing or abandoning the BORIS approach, one has to be aware of the tension that exists throughout AI between *strong* and *weak* methods. Strong methods are heuristics aimed at specific problems, such as the script-application algorithm that embodied SAM. Weak methods are general algorithms such as exhaustive search or resolution theorem proving that can be applied to a wide variety of problems. Strong methods tend to fail when the problem at hand varies from the original target domain, or when it is necessary to combine evidence from separate parts of the problem to arrive at the whole solution. Weak methods tend to fail by considering too many possible combinations, and never getting to an answer. Most weak methods are provably intractable—in other words, there will always be problems for which they will take too much time to solve.

BORIS was the epitome of a strong method. Every fact was encoded as a rule; every new piece of information made the algorithm more complicated. In the next two sections, we will examine approaches based on weak methods. In the last section, we show the approach that seems dominant in the 1990's. It is also based on weak methods.

Unified Analysis

BORIS can be seen as an attempt to deal with all the types of knowledge structures that were proposed by Schank and Abelson in [**?**], as well as several that were introduced since then. There is general agreement that having this knowledge is very helpful, perhaps absolutely necessary, but the knowledge would be easier to use if the *ad hoc* rules could be replaced with a simpler algorithm based on a unified weak method.

Norvig [**?**, **?**] uses the weak method of marker passing, a spreading activation mechanism which passes markers through a network of knowledge structures, looks for relevant marker collisions, and makes inferences based on those collisions. The model stems from the TLC program of Quillian [**?**, **?**], which had the ability to make some interesting inferences, but suffered from a lack of maturity in both the knowledge representation and the linguistic tools available at the time. Small [**?**] and Waltz [**?**] also make use of the idea of spreading activation to do parsing and semantic interpretation, and Hendler [**?**] uses marker passing to do problem solving.

Norvig's FAUSTUS program features a representation language capable of encoding a variety of knowledge structures in a uniform format. The algorithm is to pass markers from each concept in the parsed representation of the input to neighboring concepts. These markers keep track of where they came from, and so when a marker collision occurs, the paths of the two markers can be consulted to see what inference, if any, to suggest. In general, there may be contradictory suggestions, and the final interpretation consists of a coherent subset of the possible inferences. There is a measure of how coherent a set of inferences is, and the program chooses a set of inferences that tends to maximize this measure of coherence. In practice, there are too many possibilities to consider them all, so the program uses an approximation technique that looks at a small number of possibilities at a time.

The important point is that coming up with an interpretation of a text is just a matter of looking for connections among the concepts mentioned. Of course, the exact connections to consider is a critical problem.

FAUSTUS looks for specific types of marker paths that suggest specific inferences. An 'elaboration' path suggests a relation between two objects, and a 'double elaboration' path suggests a new object related by two separate relations to two given objects. A 'reference' path suggests that one object is co-referential with another, and a 'concretion' path suggests that an object should be categorized as some more specific type. These are the only paths that matter; all other marker collisions are ignored. The paths themselves are specified in terms of the primitive links that are traversed. Primitive links cover epistemological notions like taxonomic hierarchy (ISA), identity and non-identity, and the existence of a relation which maps between two categories. Primitive links do not mention any domain concepts, such as person, buying, selling, and so on. An example follows:

Input Text:

A cobbler sold a pair of boots to a hiker.

Inferences:

```
The selling is a shoe-store-transaction.
    (This is a CONCRETION inference.)
There is a walking such that it is the purpose of the boot and the
   hiker is the object-moved of it.
    (This is a DOUBLE-ELABORATION inference.)
```

Rather than answer questions, FAUSTUS outputs inferences as it makes them. The program's knowledge base includes the equivalent of a script for buying and selling in a store, and it also has a more specific version of this concept, called shoe-store-transaction: a store-transaction where the merchandise is shoes and the merchant is a shoe salesman. Markers from the concepts 'boot' and 'cobbler' intersect at shoe-store-transaction and suggest the concretion inference, that is, they suggest that the selling took place in the cobbler's shoe store. The inference is eventually accepted, since there is no evidence to contradict it. The second inference is suggested by a collision at 'walking' originating at 'boots' and 'hiker.' The markers take different paths to arrive at this collision, so a different inference is suggested, and again the inference is accepted.

FAUSTUS is able to process a variety of texts with a simple control structure coupled with a complex knowledge base. This means it can be extended by adding knowledge rather than changing the algorithm. Recall that BORIS was *integrated*: it used different knowledge sources in one program, while FAUSTUS is truly *unified*: all knowledge is in the same format. However, FAUSTUS has a difficult time deciding on a proper interpretation in the face of conflicting suggestions. There are heuristics for favoring one suggestion over another, and for deciding which suggestions to consider first, but there is no provision for weighing a set of suggestions against an alternate interpretation stemming from another set. The knowledge is unified, but not *commensurable.* FAUSTUS can also be seen as an abductive reasoning system; these will be discussed further below.

As an aside, we note that this view of understanding as finding connections can be traced back to Bartlett [**?**], who says (Ch. XII, sec. 1) "all the cognitive processes which have been considered, from perceiving to thinking, are ways in which some fundamental 'effort after meaning' seeks expression. Speaking very broadly, such effort is simply the attempt to connect something that is given with something other than itself."

Coherence-Based Analysis

Another story analyzer based on weak methods was Alterman's NEXUS program [**?**]. NEXUS searched for a solution using breadth-first search, while FAUSTUS used a marker-passing algorithm, but the result is much the same. The difference is that Alterman specified in advance the connections he was looking for. Thus, his program was limited in the inferences it could make, and could not be readily extended to make different ones. However, the connections chosen constitute a theory of coherence; of what makes one line of a story go with the previous lines. In particular, it is a theory of event coherence. The idea is if each event can be hooked up to another, then the various objects will also match up appropriately. Alterman looks for taxonomic, partonomic, and temporal relations between events. The taxonomic relation says, for example, that driving is a kind of moving. Partonomic relations hold between complex events and their subparts. There are four temporal relations, representing events that necessarily (or probably) occur before (or after) another. In the following example, the event coherence structure built by NEXUS is passed to SUM [?], which generates a summary.

Input Text:

The pig trotted towards the stream carrying a bundle of clothes. The animal expertly soaked and scoured the laundry. The pig hung the clothes in the sun to dry. The pig gathered her laundry and trotted home.

Summarization:

`The pig cleaned clothes at the stream.`

NEXUS misses many inferences that could be made by BORIS or FAUSTUS. However, NEXUS does capture something interesting about the notion of "storyness," or at least of coherence, in a highly compact theory involving only seven relations.

Notice that there is a difference between a paraphrase and a summarization of a story. A paraphrase uses different words to indicate that the program is capable of understanding rather than just mimicing. A summary must pick out the important points in a story. Thus, generating a summary indicates a succesful understanding of the events in a story, and it also indicates that the relations between those events have been understood. Summarization returns us to the idea of storyness.

A more ambitious and complex summarization theory is Lehnert's plot unit theory. Here all events are characterized as having a positive, negative or neutral mental state. These are connected with links denoting a motivation, actualization, termination or equivalence. The entire text is described in theses terms. The interesting part is that small groups of nodes and links group together into recognizable sub-graphs. The smallest of these are the 15 primitive plot units: problem, failure, perseverance, success, etc. These fit together to form complex plot units such as nested subgoals, fortuitous problem resolution, and double-cross. Given this analysis, summarization to Lehnert is a matter of recognizing complex plot units, and noticing how they fit together. The PLUGG program [?] was used to generate the following summary of Arnold Toynbee's synopsis of the New Testament:

> Jesus makes an appeal to the masses. The government wants to maintain authority over the masses. Jesus causes a scandal. Jesus takes the law into his own hands to avenge God. The authorities arrest Jesus. Jesus is crucified. Jesus' death is a triumph. Jesus is worshipped.

This is a satisfactory summary, but it should be noted that it was produced from a hand-translation from the original synopsis into the plot unit formalism. No current program is robust enough to analyze the original text.

Both Lehnert and Alterman's models of summarization suffer from the same criticism as story grammars: they characterize events and goals solely by their connections to their effects, and not by their intrinsic interest or importance. Thus, the following two texts would be assigned identical structures, while intuitively only John's plight seems like the basis of a real story:

> John had been happily married for several years.
> Then his wife contracted cancer.
> Fred had been happily writing for several minutes.
> Then his pen started to run out of ink.

Wilensky's theory of story points[?] was designed to address these problems, but is still open to some of the same criticisms.

Hobbs [?] presents another attempt at defining coherence. Hobbs poses a very important question: why should a discourse (or story) be longer than one sentence? He gives four answers. (1) It may be that more than one thing happened, and the speaker wants to convey this. However, to be a coherent discourse, the events have to be related, but they need not be related by necessary causation. Hobbs calls this the 'occasion' relation; it is similar to Alterman's 'probably happens after' relation. (2) The speaker may want to place an evaluative judgement on the discourse, as when a discourse is introduced with "the most remarkable thing happened..." (3) The speaker needs to link what is new with what he thinks the listener knows. Thus, settings can be introduced by the 'background' relation, and unusual events can be mediated by an explanation: "He was in a foul mood. He hadn't slept well that night." (4) The speaker can make his exact meaning clear by giving examples at varying levels of generality. The relations of 'parallel,' 'contrast,' 'generalization' and 'exemplification' are used for this.

Hobbs presents a program, DIANA-2, which recognizes instances of his coherence relations. For example, given the set of instructions:
Leave the building. Turn. Go to the corner.
DIANA-2 recognizes that leaving the building is the occasion that allows turning to be appropriate, and turning then occasions going to the corner. Similar work addressing coherence in discourse is by Mann and Thompson [**?**]. Sperber and Wilson [**?**] present a model that is based on relevance rather than coherence, but is also similar. See discourse understanding for more on this topic.

Commensurable Analysis

In 1990, story analysis is in a period of retrenchment. We have seen some early successes that didn't scale up well. We have seen tentative models of summarization in-the-small, and of discourse coherence. However, we have yet to see one model that comes with a broad knowledge base, makes local coherence inferences, arrives at an overall understanding of the text's important points, themes, and moral, and distinguishes stories from non-stories. The most disheartening fact is that partial successes have not been amenable to fine-tuning and improvement. Instead, they have been discarded in favor of brand new, theoretically incompatible efforts.

In hind sight, it is easy to see why this has come about. Partly it has to do with the fact that the subject matter is complex, and there is no one established tradition for important problems like knowledge representation and linguistic formalisms. But more importantly, past research efforts were just not suitable for re-use. Single-purpose systems, like SAM and PAM, had underlying algorithms tuned to one particular knowledge structure. Thus, they could not be extended to handle other types of knowledge. Integrated systems, like BORIS, proved too complex to maintain. Unified systems, like FAUSTUS, offer hope, but are still unable to make fine distinctions when there are competing interpretations. We need a methodology where evidence for competing interpretations is commensurable as well as unified. In this section we will review two systems with just that property. Both claim to be a model of semantic interpretation rather than a model of story analysis *per se*, but semantic interpretation is a large part of the battle, and the architecture of these systems should make it possible to add knowledge of story conventions.

The TACITUS project [**?**] adopts a model called abductive inference, which is inference to the best explanation. They concentrate on applying knowledge to problems in local pragmatics, that is, the pragmatics of a single sentence at a time. Consider the sentence:

Disengaged compressor after lube-oil alarm.

Here the problems include: finding a referent for the compressor and alarm, determining the relation between lube-oil and alarm, deciding if the prepositional phrase modifies the compressor or the disengaging, and determining that "after" refers to the time of the sounding of the alarm, not to the alarm itself. The approach is to build a logical form representation of the sentence, with certain aspects left unspecified. The next step is to specify all the variable parts of the logical expression, either by identifying variables with known objects and relations, or by just accepting new variables. To accept a new variable is to add new information, and the trick is not to add more new information than necessary; in other words, to make the sentence cohere by grounding it in known referents. New information is more likely to appear in certain linguistic constructs than in others. Hobbs assigns a "dollar" cost to each new assumption, with the costs weighted by these linguistic factors. Definite noun phrases are normally used to refer to known objects, so the cost of assuming one is high—$10 for the purposes of this example. Metonymic connections and noun-compound interpretations have an even higher cost, $20, since to assume them is to fail to find a connection all together. At the other end of the scale, non-nominal propositions are often new information, so they get a low cost, $3. The logic form for the example sentence, with costs attached, is as follows:

$$\exists(e,x,c,j,k,y,a,o) Past(e)^{\$3} \wedge disengage'(e,x,c)^{\$3}$$
$$\wedge\, compressor(c)^{\$5} \wedge after(j,k)^{\$3}$$
$$\wedge\, event(j)^{\$10} \wedge rel(j,y)^{\$20} \wedge (y = c \vee y = e)$$
$$\wedge\, event(k)^{\$10} \wedge rel(k,a)^{\$20} \wedge alarm(a)^{\$5}$$
$$\wedge\, noun - compound(o,a)^{\$20} \wedge lube - oil(o)^{\$5}$$

Interpretation is now a matter of finding or assuming values for *e,x,c,j,k,y,a* and *o* that minimize the total cost.

Charniak and Goldman [**?**, **?**] propose a similar model. Instead of using costs of assumptions, Charniak and Goldman's program uses probabilities as the universal currency. Different possible interpretations are represented as logical disjunctions. The most probable interpretation is chosen, according to Bayesian probability theory. For example, given the text:

John went to the supermarket. He found the milk on the shelf. He paid for it.

the program will select as the most probable interpretation a model that has paying as a step of the supermarket-shopping event, and has 'it' referring to the milk rather than the shelf or the supermarket. Milk is preferred because there is additional evidence for it, stemming from the fact that in a shopping event one pays for the selected goods. Thus, the program has duplicated an inference that SAM would make using scripts. What this model has that SAM lacked, though, is the ability to rule out a script-based inference in the face of competing evidence. For example, suppose the passage above was read in the context "John needed some shelving. He heard that the local supermarket was selling off some milk-damaged shelves." This context could be used to build another model, where 'it' refers to the shelf. The decision between the two models would depend on the initial probabilities and on the combination rules, but the important point is that all evidence can be considered.

Schubert and Hwang [**?**] present another system based on an abductive probabalistic logic. They outline how the system could be used to understand part of the Little Red Riding Hood story. However, the full system is not yet implemented.

Ng and Mooney [**?**] return to a coherence-based analysis, but they provide a metric for measuring coherence, so that their system can search for the interpretation that maximizes coherence. Norvig and Wilensky [**?**] critically evaluate these commensurable abduction systems. They point out that maximizing a single probability-like measure is theoretically suspect, because there is a fundamental trade-off between probability and informativeness. The more assumptions one makes, the greater the rewards of inferring something, and the greater the risk of making the mistake. If there is a single measure that can be maximized, it must be a measure of utility to the listener, not probability.

Compositional Analysis

These last two systems have a more direct connection to formal logic than the other systems we have reviewed, so it is important to contrast them with the best-known logical formalism for semantic interpretation, Montague semantics. In Montague semantics, there is a one-to-one correspondence between syntactic and semantic forms. That is, every rule of the grammar builds a corresponding logical expression, and these are combined compositionally to form the logical interpretation of the whole sentence. Montague semantics has enjoyed a strong following in linguistics, but has generally been seen as unable to cope with the pragmatic complexities necessary for story analysis. A Montague grammar interpretation of a sentence cannot be sensitive to the context, since it must be built compositionally from its components, without any provision for these components communicating with each other. Because of such problems, many AI researchers have adopted a "pragmatics first" approach, effectively ignoring syntax. While this has lead to a better understanding of semantics and pragmatics, it has forced researchers to work with syntactically simple, artificial texts. The abductive approach is a major advance because it allows syntactic, semantic and pragmatic factors to be considered on an equal footing.

There have been attempts to extend Montague's compositional approach, as in Kamp's [**?**] discourse representation theory, or Hirst's [**?**] work on disambiguation. A more recent attempt is the Candide system [**?**]. Their solution is to use a *least commitment grammar,* which leaves undecided a number of problems, such as prepositional phrase attachment, noun-noun compound interpretation, and quantifier scopings. In effect, the syntax produces a vague parse, which corresponds to a logical form encompassing several disjunctions. These disjunctions can then be disambiguated in the presence of any context information. The interpretations must still operate within the framework of compositionality, so they are not as free to consider interactions as are Charniak and Hobbs, but they claim that the guidance offered by compositionality will make interpretation more efficient. Charniak counters by using his probability measures not only to decide the best final interpretation, but also to limit search while building the interpretation: the least likely partial interpretations are discarded before they are fully expanded.

Conclusions

As the most recent research has shown, we are a long ways from a good model of the interpretation of a single sentence, let alone a complete story. There are basic questions to be answered in a number of fields.

Despite advances in linguistics, grammatical coverage is often inadequate to handle even simple children's stories. More complete grammars and lexicons are needed.

It has been clearly shown that knowledge of typical plans, goals, and schemata are essential for story analysis. Cataloging these is an even larger undertaking than doing a grammar or lexicon, in the same way that producing an encyclopedia is more difficult than a dictionary. Representing even a tiny fraction of common-sense knowledge will require both painstaking work and theoretical breakthroughs. Advances in machine learning (qv) may help.

There is still room for work on the basic question of what it means to interpret a text. Most authors assume that their task is to come up with a single interpretation of each sentence, the best possible model. But many stories are intentionally ambiguous, and have interpretations at multiple levels. Norvig [**?**] discusses multiple interpretations. Hobbs [**?**] attempts to summarize what AI and Cognitive Science can learn from the field of literary criticism, and vice versa. In conclusion, stories are complex entities. They rely on linguistic and literary conventions. They refer to events, characters plans and goals. They have morals and points, and they are conveyed by story tellers who have motives of their own. All these must be recognized before full appreciation of a text can be claimed. This makes story analysis perplexingly difficult, but it also makes it a fruitful field of research.

References

[1] Jim Meehan. Using planning structures to generate stories. *American Journal of Computational Linguistics*, (Microfiche 33):77–93, 1975.

[2] Natalie Dehn. Story generation after TALE-SPIN. *IJCAI-81*, 1:16–18, 1981.

[3] Michael Lebowitz. Creating a story-telling universe. pages 63–65, Karlsruhe, West Germany, 1983.

[4] F. C. Bartlett. *Remembering: A Study in Experimental Social Psychology*. Cambridge University Press, Cambridge, 1932.

[5] V. Propp. Morphology of the folktale. Baltimore, 1958. Translation by Laurence Scott. Edited by Svatava Pirkova-Jakobson.

[6] George P. Lakoff. Structural complexity in fairy tales. *The Study of Man*, 1:128–150, 1972.

[7] David Rumelhart. Notes on a schema for stories. In Daniel G. Bobrow and Allan Collins, editors, *Representation and Understanding*, pages 211–236. Academic Press, New York, 1975.

[8] Jean M. Mandler. *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Lawrence Erlbaum, Hillsdale, NJ, 1984.

[9] Jean M. Mandler and Nancy S. Johnson. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, (9):111–151, 1977.

[10] Perry W. Thorndyke. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, (9), 1977.

[11] E.H. Lichtenstein and W.F. Brewer. Memory for goal-directed events. *Cognitive Psychology*, 12:412–445, 1980.

[12] Robert Wilensky. Story grammars versus story points. *J. of The Behavioral and Brain Sciences*, 6(4), 1983.

[13] J. Grimes. *The Thread of Discourse*. Mouton and Co., The Hague, 1975.

[14] August Flammer and Walter Kintsch. *Discourse Processing*. North-Holland, Amsterdam, 1982.

[15] Teun A. van Dijk and Walter Kintsch. *Strategies of Discourse Comprehension*. Academic Press, New York, 1983.

[16] Walter Kintsch. Aspects of text comprehension. In Jean-Francois Le Ny and Walter Kintsch, editors, *Language and Comprehension*. North-Holland, Amsterdam, 1982.

[17] Walter Kintsch. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95:163–182, 1988.

[18] Eugene Charniak. Toward a model of children's story comprehension. Technical Report AI-TR-266, Cambridge, MA, 1972.

[19] Roger Schank. SAM a story understander. Technical Report Yale University Computer Science Research Report #43, 1975.

[20] R. E. Cullingford. Script application: Computer understanding of newspaper stories. Technical Report Research Report #116, 1978.

[21] Richard E. Cullingford. SAM. In Karen Sparck-Jones Barbara J. Grosz and Bonnie Lynn Webber, editors, *Readings in Natural Language Processing*, pages 627–649. Morgan Kaufmann, Los Altos, 1986 (1981).

[22] Wendy Lehnert. *The Process of Question Answering*. Erlbaum, Hillsdale, NJ, 1978.

[23] G. H. Bower, J. B. Black, and T. Turner. Scripts in memory for text. *Cognitive Psychology*, 11:177, 1979.

[24] G. DeJong. Prediction and substantiation: A new approach to language processing. *Cognitive Science*, 3(3):251–273, 1979.

[25] Robert W. Wilensky. *Understanding Goal-based Stories*. Yale University Computer Science Research Report, New Haven, CT, 1978.

[26] Robert Wilensky. *Planning and Understanding*. Addison-Wesley, Reading, MA, 1983.

[27] Michael G. Dyer. In-depth understanding. Technical Report Research Report #219, 1982.

[28] Roger C. Schank and Robert P. Abelson. *Scripts, Plans, Goals and Understanding*. Erlbaum, Hillsdale, N.J., 1977.

[29] Peter Norvig. A unified theory of inference for text understanding. Technical Report Ph.D Thesis, 1986.

[30] Peter Norvig. Marker passing as a weak method for text inferencing. *Cognitive Science*, 13(4):569–620, 1989.

[31] M. Ross Quillian. Semantic memory. In Marvin Minsky, editor, *Semantic Information Processing*. Massachusetts Institute of Technology Press, Cambridge, 1968.

[32] M. Ross Quillian. The teachable language comprehender: A simulation program and theory of language. *Communications of the ACM*, (12):459–476, 1969.

[33] Steven Small and Chuck Rieger. Parsing and comprehension with word experts. In Wendy G. Lehnert and Martin H. Ringle, editors, *Strategies for Natural Language Processing*, pages 89–148. Lawrence Erlbaum Associates, Hillsdale, NJ, 1982.

[34] David L. Waltz and Jordan B. Pollack. Massively parallel parsing: a strongly interactive model of natural language interpretation. *Cognitive Science*, 9(1):51–74, 1984.

[35] James A. Hendler. Integrating marker-passing and problem-solving. Technical Report Technical Report No. CS-85-08, Providence, RI, 1985.

[36] Richard Alterman. A dictionary based on concept coherence. *AI Journal*, 25(2):153–186, 1985.

[37] Richard Alterman. Summarization in the small. In Noel E. Sharkey, editor, *Advances in Cognitive Science 1*, pages 72–93. Ellis Horwood/Wiley, Chichester/New York, 1986.

[38] W. Lehnert and C. Loiselle. Plot unit recognition for narratives. Technical Report No. 83-39, Univ. of MA, Amherst, MA, 1983.

[39] Jerry R. Hobbs. Towards an understanding of coherence in discourse. In Wendy G. Lehnert and Martin H. Ringle, editors, *Strategies for Natural Language Processing*, pages 223–243. Erlbaum, Hillsdale, 1982.

[40] W. Mann and S. Thompson. Relational propositions in discourse. *Discourse Processes*, 9(1):57–90, 1986.

[41] D. Sperber and D. Wilson. *Relevance*. Harvard University Press, Cambridge, MA, 1986.

[42] Jerry R. Hobbs, Mark Stickel, Paul Martin, and Douglas Edwards. Interpretation as abduction. pages 95–103, Buffalo, NY, 1988.

[43] Eugene Charniak and Robert Goldman. A logic for semantic interpretation. pages 87–94, Buffalo, NYU, 1988.

[44] Eugene Charniak and Robert Goldman. A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. Detroit, MI, 1989.

[45] Lenhart K. Schubert and Chung Hee Hwang. An episodic knowledge representaion for narrative texts. pages 444–458, 1989.

[46] H. T. Ng and R. J. Mooney. The role of coherence in constructing and evaluating abductive explanations. Stanford, 1990.

[47] Peter Norvig and Robert Wilensky. A critical evaluation of commensurable abduction models for semantic interpretation. Helsinki, 1990.

[48] H. Kamp. In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language*, pages 227–322. Mathematisch Centrum, Amsterdam, 1981.

[49] Graeme Hirst. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge University Press, Cambridge, 1986.

[50] Martha E. Pollack and Fernando C. N. Pereira. An integrated framework for semantic and pragmatic interpretation. pages 75–86, Buffalo, NY, 1988.

[51] Peter Norvig. Multiple simultaneous interpretations of ambiguous sentences. *Proceedings of the 10th Annual Conference of the Cognitive Science Society*, 1988.

[52] Jerry R. Hobbs. *Literature and Cognition*. CSLI, Stanford, CA, 1990.

PETER NORVIG
University of California, Berkeley