COGNITIVE SCIENCE 4 (293–331) (1981)

Plot Units and Narrative Summarization*

WENDY G. LEHNERT
Department of Computer Science
Yale University, New Haven, Connecticut 06520

In order to summarize a story, it is necessary to access a high level analysis of the story that highlights its central concepts. A technique of memory representation based on plot units appears to provide a rich foundation for such an analysis. Plot units are conceptual structures that overlap with each other when a narrative is cohesive. When overlapping intersections between plot units are interpreted as arcs in a graph of plot units, the resulting graph encodes the plot of the story. Structural features of the graph then reveal which concepts are central to the story, and which concepts are peripheral. Plot unit analysis is currently being investigated as a processing strategy for narrative summarization by both computer simulation and psychological experiments.

1. INTRODUCTION

When a person reads a narrative story, an internal representation for that story is constructed in memory. We examine the contents of this memory representation by asking the reader simple questions about the story. Typical question-answering behavior will reveal evidence for numerous inferences, causal chain constructions, and the predictive integration of information into instantiated knowledge structures (Lehnert, 1978; Graesser, 1981). While question answering provides us with a method for examining the contents of a memory representation, the task of question answering does not readily yield a more global picture of the memory representation as a whole. We can only guess at how the various pieces fit together within a single structure.

If we are interested in the structure of narrative memory representations, the summarization task is a rich (and largely untapped) source of

*The author is indebted to Roger Schank, Chris Riesbeck, Jim Meehan, Art Graesser, Haj Ross, and Marty Ringle for their extensive and helpful comments on various drafts of this paper. This work was supported in part by the Advanced Research Projects Agency under contract N00014-75-C-1111 and in part by the National Science Foundation under contract IST7918463.

enlightening phenomena. When a reader is asked to summarize a story, vast amounts of information within the memory representation are selectively ignored, in order to produce a distilled version of the original narrative. This process of simplification relies on a global structuring of memory that allows search procedures to concentrate on central elements of the story while ignoring peripheral details. We intuitively expect that some global or "macro" structure is holding memory together, but a precise formulation of this structure is much more elusive.

Any process model for summarization that attempts to utilize high-level narrative structures must confront a number of difficult questions. How is the hierarchical ordering of a memory representation constructed at the time of understanding? Exactly what elements of the memory representation are critical in building this structure? What search processes are used to examine memory during summarization? How are summaries produced after memory has been accessed? In this paper, we will propose a method for narrative analysis and summarization that addresses each of these issues.

In the next four sections, we develop a representational system of plot units for high-level structural analysis. Sections six through eight examine the conceptual content of narrative summaries in terms of plot units, and present a rigorous framework for describing narrative cohesion. Given these conventions, we are then in a position to propose a simple process-oriented description of summary generation in section nine. Section ten rounds out the process model by outlining some bottom-up recognition techniques needed to produce plot unit analyses, and the conclusion presents a brief comparison between the proposed system and story grammars, closing with some comments on other applications for plot unit structural analysis.

2. AFFECT STATE PATTERNS

In the system about to be proposed, emotional reactions and states of affect are central to the notion of a plot or story structure. The structure of a narrative text is a configuration of plot units, and each plot unit is itself a configuration of smaller entites called "affect states." Affect states do not attempt to describe complex emotional reactions or states of desire in the detail that inference mechanisms would require; they merely mark gross distinctions between "positive" events, "negative" events, and mental events of null or neutral emotionality. A number of researchers have tackled the problem of affect far more seriously than this (deRivera, 1977; Roseman, 1979); and our notion of a plot unit should naturally dovetail with these more sophisticated representational systems. But right now, it is useful to see how far these very gross differentiations can carry us in our quest for summarization algorithms.

The notation we'll use for plot units involves abbreviations for our three affect states:

+ (Positive Event) Events that please
– (Negative Event) Events that displease
M (Mental State) Mental states (w/ neutral affect)

Each of these affect states occur with respect to a single character, and events involving multiple characters require multiple affect states. For example, if John marries Mary, the event is presumably positive for John and Mary, while Mary's father (who can't stand John) experiences a negative event.

As a story progresses, a linear map of chronologically-ordered affect states is created for each character in the story. Upon examining a number of affect state maps from a variety of stories, general patterns begin to emerge. These patterns lead us to the notion of a plot unit. For example, consider the affect map for John in the following story:

> When John tried to start his car this morning, it wouldn't turn over. He asked his neighbor Paul for help. Paul did something to the carburetor and got it going. John thanked Paul and drove to work.

The affect analysis for John consists of three affect states:

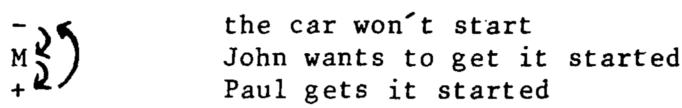


Figure 1.

To make causality explicit, we will connect appropriate pairs of affect states with pairwise causal links. John's three affect states above are connected by three different link types. An aversive event *motivated* John to get his car started; John *actualized* this desire by getting Paul to start the car, and Paul's assistance *terminated* the original difficulty. This pattern represents an affect configuration (or plot unit) that is extremely pervasive in narrative texts: *resolution of a problem by intentional means.*

Now suppose we extend our analysis to include the affect states of Paul:

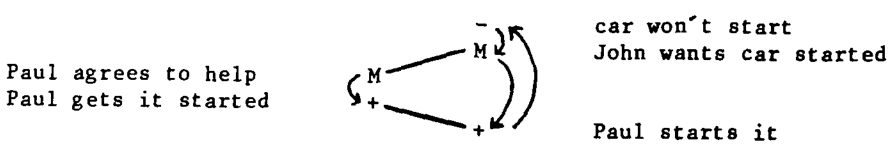


Figure 2.

The diagonal links signify causalities of affect across characters. When Paul agrees to help, he is assuming John's state of desire; now Paul wants to get the car started too. This configuration of embedded achievement across characters signals an *honored request.* When this configuration is followed by its symmetrical counterpart, we have an instance of an *exchange.* Favors (voluntary services) are often exchanged, and the idea of a loan is a special case of exchanged services:

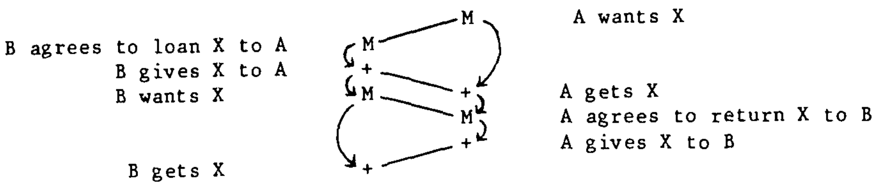


Figure 3.

If we were missing the last two positive events in this structure, we would understand that A is obligated to B; it would be up to A to complete the symmetry of the configuration. A *trade* is also a special case of an exchange, where the two transactions occur simultaneously:

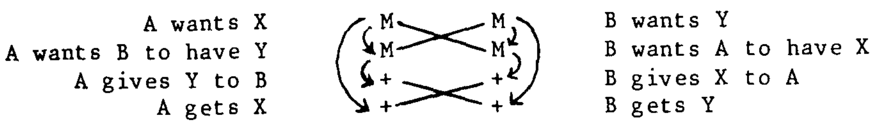


Figure 4.

A number of standard affect configurations arise in this manner which allow us to recognize narrative structures and build plot structures from affect states. But, before we can identify standard configurations, we must present a system of causal links that will be used to join pairs of affect states.

3. CAUSAL LINKS

A link which runs from a negative event to a mental state describes motivation, while a link running from a mental state to a positive event describes actualization. To make these and other distinctions explicit, we will use a system of four link types: MOTIVATION (m), ACTUALIZATION (a), TERMINATION (t), and EQUIVALENCE (e). Motivation links describe causalities behind mental states, and actualization links describe intentionalities behind events. The termination link is not used when an event per se is terminated, but when the affective impact of that event is surplanted or displaced. For example, a second marriage many "terminate" a prior divorce in this sense, if it nullifies the emotional reactions to that divorce. Equivalent events and states are linked when multiple perspectives of a single affect state can be separated. The use of these links will hopefully be clarified by some examples given in the next section. But first, we will describe the syntax of these four links.

Each link describes an oriented arc between two affect states: m-links and a-links point forward in time, while t-links and e-links point backward in time. With three affect states and four link types, there are 36 possible pairwise configurations, if we consider all the possible combinations. But in fact, only 15 of these will occur when we observe some syntactic constraints on the use of causal links. To summarize these constraints, the following table illustrates which combinations occur by marking legal configurations with an "*".

	M M	M +	M –	+ M	– M	+ –	– +	+ +	– –
m	*			*	*				
a		*	*						
t	*					*	*	*	*
e	*					*	*	*	*

The constraints can be described as follows:

m-links must point to a mental state.
a-links must point from a mental state to an event.
t-links and e-links must point:
a) from a mental state to a mental state, or
b) from an event to an event.

Figure 5.

These links have been given an orientation for intuitive convenience, rather than notational necessity. For m-links and a-links, the pointer moves from a temporal antecedent to its consequent. With t-links, the pointer goes from a subsequent event to the prior event it terminates. E-links are used to identify redundant state descriptions which appear at different times. The backward orientation of e-links is therefore arbitrary.

4. PRIMITIVE PLOT UNITS

Our 15 legal pairwise configurations will act as the building blocks for more complex configurations. We will refer to them as the "primitive plot units," and each will be specified by name.

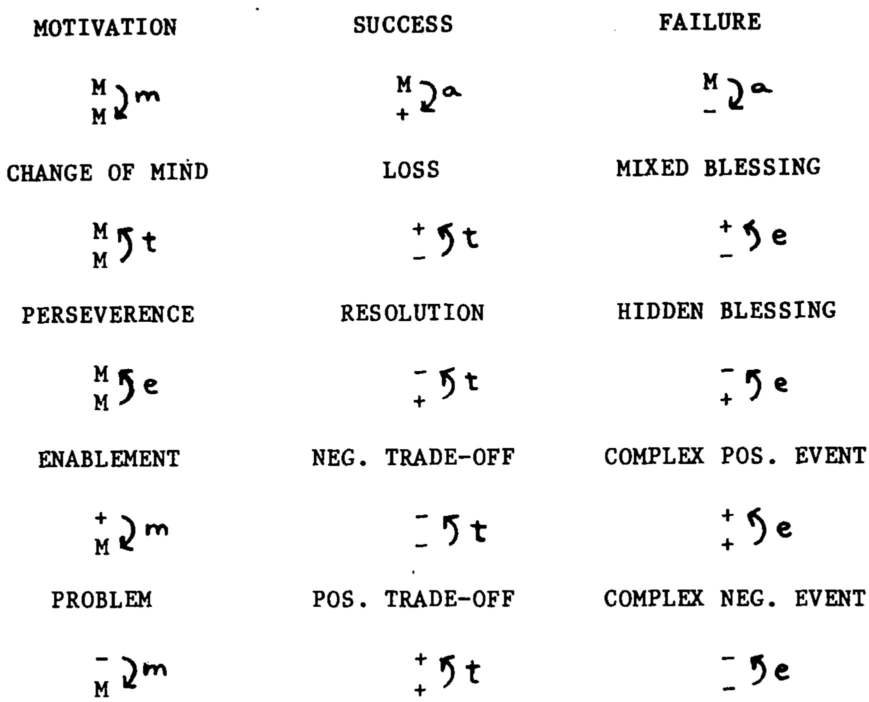


Figure 6.

Sometimes, a primitive plot unit will appear without other interceding affect states. This occurs most commonly with the units "problem," "enablement," and "motivation." Other primitive plot units tend to be broken up by interceding affect states. For example, it may take months (with lots of interceding emotional reactions) to find out that a job promotion is now leading to an ulcer. This would be an example of a mixed blessing, or a good thing turned sour.

EXAMPLES OF PRIMITIVE PLOT UNITS

PROBLEM:	You get fired and need a job. You bounce a check and need to deposit funds. Your dog dies and you long for companionship.
SUCCESS:	You ask for a raise and you get it. You fix a flat tire. You need a car so you steal one.
FAILURE:	Your proposal of marriage is declined. You can't find your wallet. You can't get a bank loan.

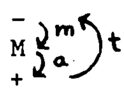
RESOLUTION:	Your broken radio starts working again. They catch the thief who has your wallet. You fix a flat tire after a blow out.
LOSS:	Your big income tax refund is a mistake. The woman you love leaves you. The car you just bought is totaled.
POS. TRADE-OFF:	You buy a new Toyota and then inherit a Porsche. You take a day off and then realize it's a holiday. You get a raise and then win the Irish Sweepstakes.
NEG. TRADE-OFF:	You get fired so you don't have to take a lousy job assignment. Your car blows up so you don't have to make the next insurance payment. You lose the election so you don't have to placate demanding voters.
PERSEVERENCE:	You want to get married (again). You reapply to Yale after being rejected. You want to ski again after a bad skiing accident.
HIDDEN BLESSING:	You get audited and they owe you. You sprain an ankle and win damages. Your mother dies and you inherit a million.
MIXED BLESSING:	You buy a car and it turns out to be a lemon. You fall in love and become insanely jealous. Your book is reviewed but they hate it.
CHANGE OF MIND:	You apply to Harvard and then to Yale. You want to buy a car but decide against it. You want to see a movie until a friend pans it.
MOTIVATION:	You need advice so you decide to ask a friend. You want to buy a car so you apply for a loan. You want to reach a client so you call him.
ENABLEMENT:	You decide to celebrate after a raise. You receive a book and decide to read it. You get a loan and have to pay it back.
COMPLEX POS:	A gift is indicative of close friendship. Your raise signifies recognition. You win respect by getting a rolls royce.
COMPLEX NEG:	You lose $100 when your wallet is stolen. You break an arm in a car accident. Your house burns down and you aren't covered.

These primitive plot units will serve as building blocks for more complicated plot configurations. They do not, by themselves, provide us with all of the recognition abilities we need. We will now expand beyond our set of primitive units, in order to describe more complicated situations.

5. COMPLEX PLOT UNITS

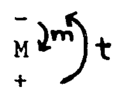
Using the 15 primitive plot units, we can build larger plot units to represent general plot configurations. For example, the string (– M +) of three affect states is used by three different plot units that are distinguished only by the causal links involved:

INTENTIONAL PROBLEM RESOLUTION



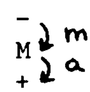
= problem
& success
& resolution

FORTUITOUS PROBLEM RESOLUTION



= problem
& resolution

SUCCESS BORN OF ADVERSITY

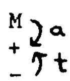


= problem
& success

Figure 7.

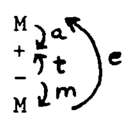
These are examples of complex plot units that are commonly found in narrative texts. Other closely related plot units include:

FLEETING SUCCESS



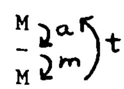
= success
& loss

STARTING OVER



= success
& loss
& problem
& perseverence

GIVING UP



= failure
& problem
& change of mind

Figure 8.

Many complex plot units can be transformed into different units by way of a minor variation:

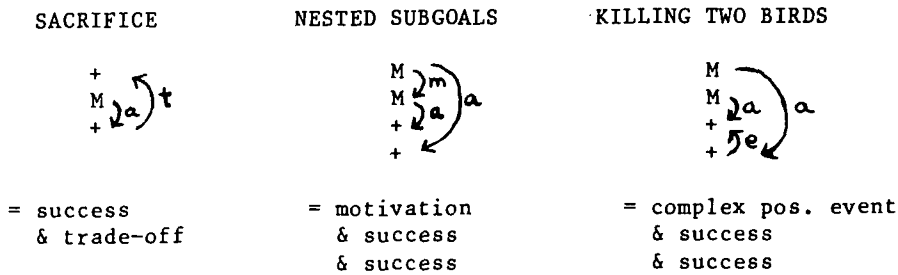


Figure 9.

Thus far, we have concentrated on plot units that describe configurations within a single character. A large number of complex plot units involve multiple characters. Plot units with more than one character require cross-character causal links. These will be represented by diagonal segments between affect states, where the higher affect state precedes the lower affect state in time. While we found it useful to distinguish four types of intra-character links in building the primitive plot units, we will not need to distinguish cross-character links. Cross-character links can occur between any pair of affect states, and their interpretation will rely on the following conventions:

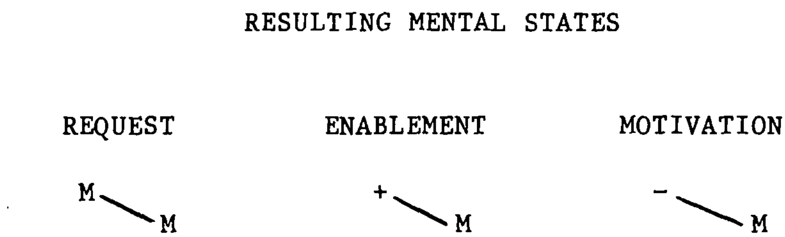


Figure 10.

These configurations describe the initiation of a goal state as a direct response to another character's situation. All of the resulting mental states are initiated by free choice. In the case of "M/M", the resulting mental state occurs in response to a request. This resulting mental state may assume the desires of the initiator, or it may oppose them. The request configuration does not commit us to any assumptions about the contents of the two mental states, or how their contents are related. In the cases of " + /M" and " – /M", we have mental states enabled or motivated by vicarious events. For example, a desire to celebrate is normally enabled by a positive event, while a desire to help out is typically motivated by a negative event.

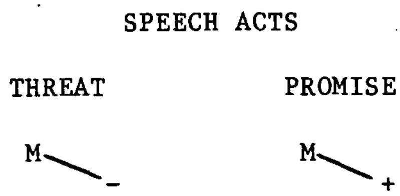


Figure 11.

These two configurations describe communications which result in positive and negative affect states. The antecedent in either case is a mental state describing the intentions of that character. These two configurations often appear in tandem when an agreement is achieved by coercion, i.e., a promise is motivated by a threat.

Figure 12.

Shared events are shared in the sense that two characters are affected by them in a similar manner. The same event is experienced by both people as either a positive or negative event. Mixed events are just the opposite. Here, the same event is experienced differently by both people: one is affected positively, and one negatively.

These nine cross-character configurations can now be used to build complex plot units involving two characters. Some of the most common configurations involving two characters are those that describe cooperative agreements and behavior. In the simplest case, a request is made and the respondent behaves either cooperatively or not:

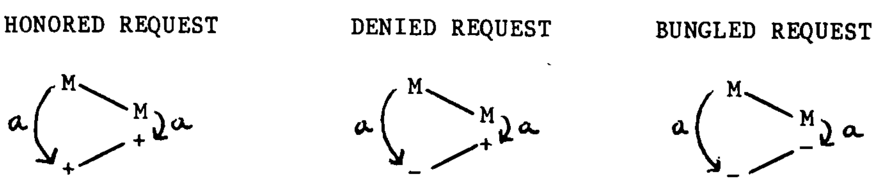


Figure 13.

In our story about John's car, we had an instance of an honored request. John asked Paul for help, and Paul got the car started for John. In this situation, the second character assumes the mental goal state of the first character. Paul wanted to get the car started too. When a request is denied, we should assume different mental states. If Paul tells John that he's too busy, we should not assume that Paul wanted to get John's car started.

A slightly different situation arises when the speech act of a threat is invoked, instead of a request:

EFFECTIVE COERCION INEFFECTIVE COERCION BUNGLED COERCION

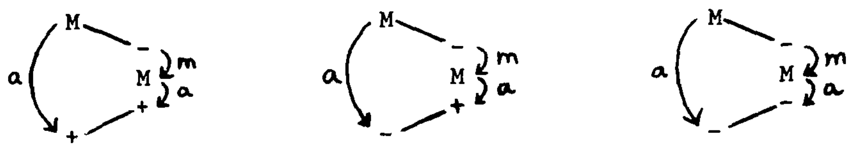


Figure 14.

In these situations the respondent is confronted with a problem situation that can be resolved with either cooperative behavior or a challenging denial. These situations are very common, and in some cases it is appropriate to represent them in greater detail. For example, what if Paul agrees to help John get his car started, but then fails to do so? In some stories, the extraction of an agreement receives enough attention to warrant its own affect analysis:

PROMISED REQUEST HONORED PROMISED REQUEST BUNGLED

Figure 15.

When John asks Paul for help, he has set up a subgoal for getting help. If Paul agrees to help, Paul satisfies John's subgoal by making a promise. If Paul then succeeds in helping John, the top level goal is achieved as well. But if Paul fails, his actions amount to nothing more than good intentions that were bungled. The plot units for an honored request and a promised request that is honored are very similar. When a request is honored, we have a request and shared success. When a request is promised and then honored, we have nested subgoals, a request, a promise, and shared success. These are identical except for details about the agreement as an interaction that is separate from the service performed. This more detailed level of description is necessary when we try to represent "good intentions" that fail in response to a request or threat.

If we examine the notion of a threat at this level of detail, we can see the difference between a threat that is agreed to and successful, versus a threat that is sincerely agreed to but unsuccessful anyway. (These are elaborations on effective coercion and bungled coercion).

COERCED AGREEMENT HONORED   COERCED AGREEMENT BUNGLED

Figure 16.

In both of these cases the respondent intends to go along with the threat. When the threat succeeds, it is because the respondent succeeds. When the threat fails, it is because the respondent fails. In both cases, the respondent promises to cooperate. A slightly different situation arises when the respondent promises to cooperate, but intentionally fails to come through:

DOUBLE-CROSS

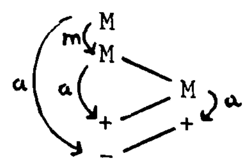


Figure 17.

In a double cross, the respondent deceptively agrees to go along, and then intentionally does something to foil the other's goal. This unit contains subgoals, a request, a promise, and a mixed event of success and failure. We could also represent a double cross in response to coercion, if the request were replaced with a threat:

COERCED DOUBLE-CROSS

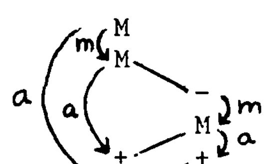


Figure 18.

This coerced version of a double cross is somewhat more self-contained than the plain double cross since it is motivated by a coercive act. We can see symmetry in the negative consequences to both characters. In addition to cooperative and uncooperative responses, people often interact in unsolicited ways:

UNSOLICITED HELP

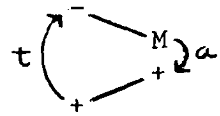


Figure 19.

In this case, the problem state is completely assumed by an intervening character who is motivated by the initial problem state to initiate his own assistance. If Paul had noticed John's problem and volunteered his services, we would have a case of unsolicited help.

Any of the preceding plot units for cooperative behavior can be embedded in a problem resolution. For example, a problem resolution via a successful threat would look like:

PROBLEM RESOLUTION BY EFFECTIVE COERCION

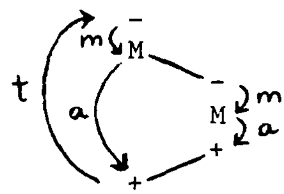


Figure 20.

In addition to the various ways that one character can react to another's desires, there are also a number of standard plot configurations that describe situations of reciprocation. When cooperative behavior is reciprocated, we arrive at plot units for obligation, exchange, and trades:

OBLIGATION SERIAL EXCHANGE SIMULTANEOUS EXCHANGE

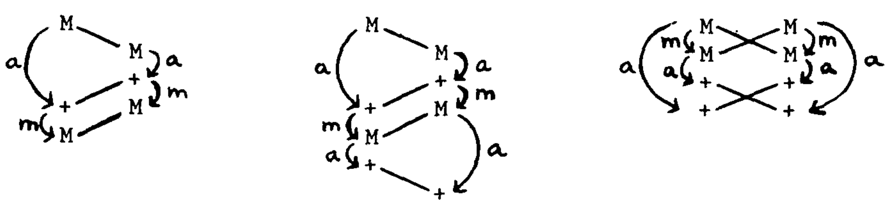


Figure 21.

Notice that the unit for a serial exchange (of requests) is very similar to the unit for a simultaneous exchange (or trade). In a serial exchange, the requests are satisfied one after another, while in a simultaneous exchange, requests are handled in parallel. The same affect states and link configurations occur in both units; only the temporal sequencing of the affect states is different.

Another variation on exchanged requests occurs when the respondent agrees to honor the initial request, pending a conditional request of his own. Paul could have agreed to fix John's car, if John would first give him a beer. Then we would have two requests with one being conditional on the completion of the other:

REQUEST HONORED WITH CONDITIONAL REQUEST

REQUEST HONORED WITH CONDITIONAL PROMISE

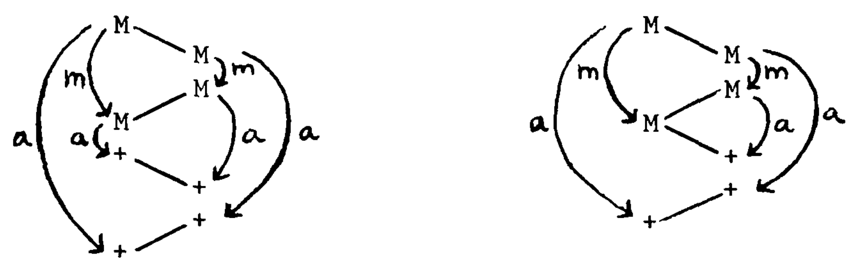


Figure 22.

Of course John may only promise to give Paul a beer. In this case, the request is met with a conditional promise. While we expect John to honor his promise, he may not. If he doesn't, we will find the pattern for a double cross.

Many plot units are recognized by predictive processing off of primitive plot units. For example, whenever a promise unit is encountered, we must activate expectations for success resulting in a positive shared event, or success resulting in a positive mixed event. That is, a primitive unit for a promise always sets up expectations for the complex units describing an honored or reneged promise.

HONORED PROMISE

RENEGED PROMISE

Figure 23.

Promises and cooperative behavior are not the only plot units that rely on shared and mixed events. Other complex plot units include:

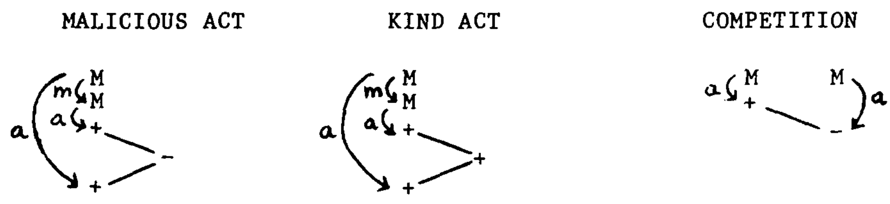


Figure 24.

Finally, a number of plot units involve variable affect states:

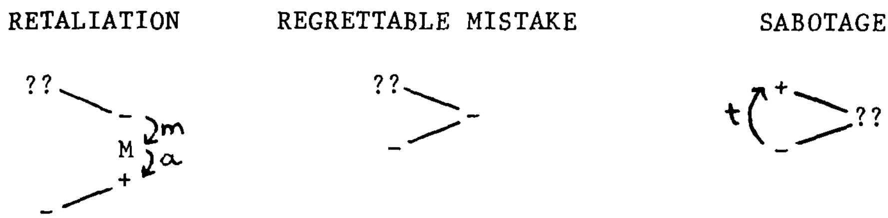


Figure 25.

The unspecified affect state here signifies a "wild card" for the purposes of pattern recognition. Any affect state will match an unspecified state.

This section has attempted to show how complex units can be constructed to provide infinite variations of plot structures. For example, a kind act with a resulting trade-off will amount to an act of self-sacrifice, while a fortuitous problem resolution with a trade-off will merely signify an undesirable side effect. It would be pointless to try to enumerate at this time all of the possible combinations that are useful for plot recognition, although the question of what constitutes a valid unit is very interesting in its own right. From a psychological viewpoint, we might expect different people to operate with different sets of plot units. This could account for at least some of the individual differences that appear in summarization data, and might further provide an interesting basis for developmental theories of reading comprehension.

6. SUMMARIZATION

By recognizing plot units, we can achieve a high-level analysis of activities and interactions within a narrative. We should expect to find evidence for this "chunking" of information in paraphase and summarization behavior. To see how this works, consider the following narrative:

> John was thrilled when Mary accepted his engagement ring. But when he found out about her father's illegal mail-order business, he felt torn

> between his love for Mary and his responsibility as a policeman. When John finally arrested the old man, Mary called off the engagement

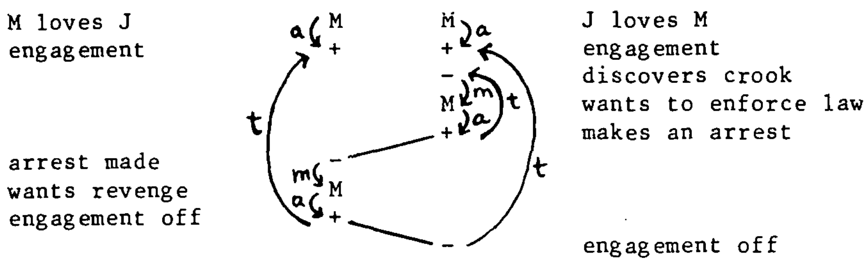


Figure 26.

The affect analysis for John and Mary reveal configurations of a trade-off from retaliation on the part of Mary, and a problem resolution leading to loss for John.[1] Ideally, one might expect good summaries to convey each of these four plot units (trade-off, retaliation, problem resolution, and loss). A stronger claim about summaries would argue that any summary which does not convey all four plot units is an unacceptable summary:

"When John arrested Mary's father, she interfered with his wedding."
(no trade-off for Mary)

"When John arrested an old crook, Mary called off their engagement."
(no retaliation for Mary)

"When Mary's father was arrested, she called off her engagement."
(no problem resolution for John)

"When John arrested Mary's father, she called off her engagement."
(no loss for John)

But a summary that includes all four plot units provides an accurate description of the story:

"When John arrested Mary's father, she called off their engagement."
(all units present)

Of course "inclusion" here means inclusion by inference, as well as by explicit mention. We must infer that there is a causality between John's act and Mary's act in order to understand retaliation, but this inference had to be made with the original narrative as well.

Ultimately, an affect analysis in terms of plot units should allow us to predict the sorts of summaries that human subjects will produce. But initially, we must study actual summary behavior, in order to develop a process model that converts plot unit configurations into narrative summaries. Consider the following story:

[1] We will ignore the initial success units for reasons that will be explained in section 9.

John and Bill were competing for the same job promotion at IBM. John got the promotion and Bill decided to leave IBM to start his own consulting firm, COMSYS. Within three years COMSYS was flourishing. By that time John had become dissatisfied with IBM so he asked Bill for a job. Bill spitefully turned him down.

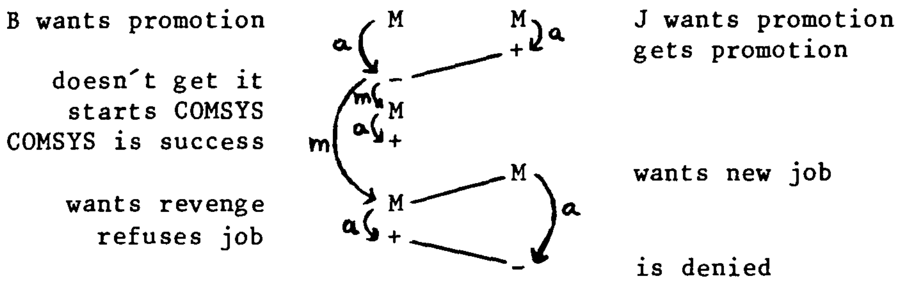


Figure 27.

Here we have a competitive situation between John and Bill in which John wins. Bill's failure turns into success out of adversity, and then he retaliates against John for his initial failure. John sets the stage for Bill's retaliation by asking Bill for a job. John consequently experiences a failure when Bill uses this opportunity to get revenge by denying John's request. The plot units here are (1) competition which subsumes (2) John's success and (3) Bill's failure, (4) success born of failure, (5) retaliation, and (6) a denied request which subsumes (7) John's request and (8) Bill's denial. To see how these units are integrated into summaries of the story, we will look at 10 summaries provided by experimental subjects. The subjects were asked to read the story, and were then instructed to summarize the story in one sentence. The summaries which appear below are verbatim responses, except for some name corrections (subjects frequently reversed Bill and John).

SUMMARIZATION BEHAVIOR

(1) John, Bill compete for a job which John wins, causing Bill to quit the company and start his own firm (COMSYS), which leads to Bill's spiteful rejection of John's request for a job some years later at Bill's successful company.

(2) Bill was spiteful when John asked him for a job, because they had once competed for the same job at IBM.

(3) John and Bill were both competing for a job at IBM which John got so Bill started his own business and later had the opportunity to turn John down when John wanted a job.

(4) John got promoted by IBM, so Bill his friend, started his own business which soon flourished and when John came asking for a job, Bill spitefully turned him down.

(5) Bill and John worked for IBM, and were friends until three years later Bill turned John down when he asked for a job.

(6) Bill turned John down for a job because John had beat him out of a promotion when they both worked at IBM.

(7) Bill started his own business COMSYS after losing out to John for a job at IBM and later out of spite refused to give John a job when John was dissatisfied with his old one.

(8) Bill, who lost a job promotion to his competitor John, establishes a lucrative consulting firm of his own, and rejects John's request for a job later on.

(9) John beat out Bill for a promotion at IBM whereupon Bill decided to leave and form his own company, COMSYS, which was flourishing within three years, and which John turned to for a job when he was fed up at IBM which he did not get due to Bill's spite.

(10) John was promoted at IBM instead of Bill, so when Bill left IBM to start his own firm and the business flourished, he turned John down when the latter, dissatisfied at IBM, applied to Bill for a job.

In analyzing these summaries for the presence of plot units, we find that a number of units are present only in an implicit manner. For example, (1,2,4,7,9) explicitly refer to "spite" and therefore make explicit reference to the retaliation unit. Summaries (3,6,8, and 10) are constructed with suggestive causalities, and the presence of retaliation is only implicitly present. These implicit cases can be contrasted with (5) where there is no basis for a retaliation unit whatsoever. When retaliation is implicit, it is conveyed by the causal constructions of clause formation. Other plot units may be implicitly present by conceptual entailment. For example, in (3, 6, and 7), the request (for a job) is implicit from the verb phrase "to turn down," since this expression describes a denied request. In all of the other summaries, John's request is explicit. In the chart on the next page, we have marked with an "IMP" those plot units which are implicit in the text.

Other plot units are implicitly present by processes of inference. For example, in (2), the explicit presence of retaliation and competition force us to infer that John won and Bill lost. The patterns of competition and retaliation wouldn't overlap at a negative event, if Bill got the job. Without this overlap, we would say that it just doesn't make sense for Bill to get the job and then feel spiteful about it. This inference is a "role-binding inference," driven by the retaliation unit.

"X is spiteful toward Y" sets us up for:

1. a causal antecedant: Y causes a [–] for X, and
2. a causal consequent: X causes a [–] for Y.

Since competitive resolution entails the configuration needed by (1),

Figure 28.

we can establish who is the winner and who is the loser by a role-binding inference (If X is spiteful toward Y, X is the loser, and Y is the winner). So we can say that Bill's failure and John's success are present by implicit inference in summary (2). The presence of denial is also supplied by the retaliation unit where the structure for a negative mixed event is encoded.

If we analyze these summaries for the presence of our eight plot units, we get the following distribution:

- competition (COMP)
- Bill's success (BS)
- Bill's failure (BF)
- John's failure (JF)

retaliation (RET)
John's success (JS)
denied request (DR)
John's request (JR)

	COMP	BS	RET	DR	BF	JS	JR	JF
1	X	X	X	X	X	X	X	X
2	X		X	IMP	IMP	IMP	X	IMP
3	X	X	IMP	X	X	X	IMP	X
4		X	X	X		X	X	X
5				X			X	X
6	X		IMP	X	X	X	IMP	X
7	X	X	X	X	X	X	IMP	X
8	X	X	IMP	X	X	X	X	X
9	X	X	X	IMP	X	X	X	X
10	X	X	IMP	X	X	X	X	X

Figure 29.

We could postulate a rough qualitative ranking of the summaries, based on the number of plot units present. It is the case that most (6) of the summaries reference all eight plot units, while the summary containing the least (3) plot units is arguably the worst summary.

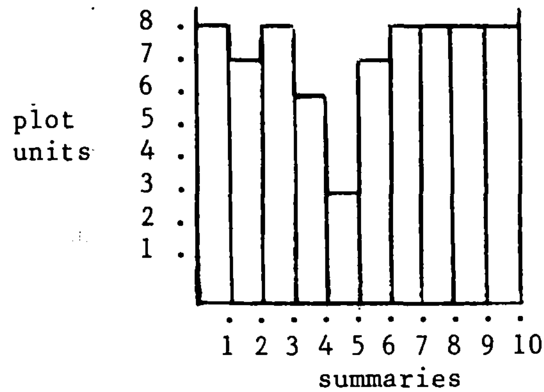


Figure 30.

This distribution suggests that the plot unit analysis is central to paraphrase production. Summary (5) with three plot units does seem to be the poorest summary, while the others (albeit stylistically different) are more on a par in terms of their content.

By analyzing the nature of plot unit occurrences in terms of their explicit expression or implicit presence, we begin to see that some units are "pivotal" in driving inferences about other units. The identification of pivotal units will be very important in the actual process of summarization. We will return to this idea in section 9 when we outline the process model for summarization.

To see how summaries are built from plot configurations, we can look at these 10 summaries in terms of their clause constructions. In the following abstractions, we have abbreviated all clauses that describe plot units and identified them accordingly. What remains is a structural backbone for the sentences generated:

1. [COMP] which [JS,BF] causing [BS] which leads to [RET,DR,JF] of [JR].
2. [RET] when [JR,] because [COMP]. (infer JS,BF,DR,JF)
3. [COMP] which [JS,BF] so [BS] and later had the opportunity to [DR,JF]. (implicit JR, RET)
4. [JS] so [BS] and when [JR], [DR,RET,JF].
5. [DR,JF] when [JR].
6. [DR,JF] because [COMP,JS,BF]. (implicit JR,RET)
7. [BS] after [COMP,JS,BF] and later [DR,RET,JF]. (implicit JR)
8. Bill, who [COMP,JS,BF], [BS] and [DR,JF,JR] later on. (implicit RET)

9. [COMP,BF,JS] whereupon [BS] and which [JR] which [JF] due to [RET]. (infer DR)
10. [COMP,BF,JS] so when [BS], [DR,JF] when [JR]. (implicit RET)

These skeletons reveal natural "clumps" of information. For example, Bill's failure and John's success are naturally tied to their competition. This follows from the fact that competition entails units for success and failure. If Bill's success is mentioned at all, it occurs in isolation of other units, and always follows the COMP-JS-BF clump (When their order is inverted in (7) the connective makes their relationship explicit). John's failure and his denied request tend to appear together, and can be easily combined with retaliation, when retaliation is made explicit. The choice of specific connectors appears to be determined by retaliation, since the causality connecting other plot units serves to convey retaliation implicitly. The interplay between global factors (like retaliation), and more local entities (like John's job request) can be handled in a variety of ways. Some constructions are stylistically more pleasing than others, and the use of implicit and inferential information seems central to the more successful strategies.

7. NARRATIVE COHESION

It is possible to assess the cohesiveness of a narrative by analyzing its connectivity across plot units. For example, in the COMSYS story, we have a totally coherent text: John's success causes Bill's failure and this motivates Bill to become successful on his own. Bill then exploits an opportunity to retaliate against John by causing John to fail in his job hunting. The causal chain is not quite linear, but it is completely connected:

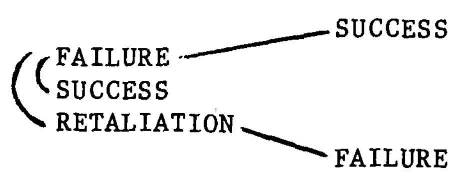


Figure 31.

Suppose the last sentence of the COMSYS story was:

"Bill gave John a key position in his company."

Then, we would have a slightly different set of plot units:

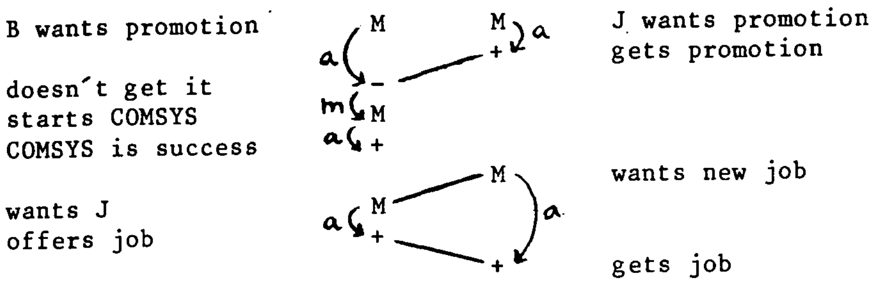


Figure 32.

Now we no longer have retaliation. Instead, we have Bill honoring a request by John which yields John another success. The story is no less plausible, but its cohesiveness is lessened; now there is no connectivity between the first three and the last two plot units. Bill's success enables him to help John, but there is no affect-oriented connection to unify the story.

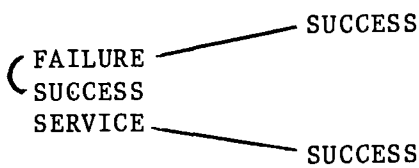


Figure 33.

We would tend to say that Bill gave John a good job *in spite* of the fact that John won the promotion Bill wanted. We are more surprised to see Bill act magnanimously; a retaliation seems more likely. But this exceptation is not founded on any knowledge of Bill's personality or attitude toward John. We have no such information to help us predict his behavior. It is instead a general expectation for narrative unity. We have a preference for cohesive narratives, and retaliation allows us to tie everything together. If Bill offers John the job, we cannot establish total connectivity across all plot units.

This type of expectation derives from our knowledge about narratives rather than our knowledge about the world in general. It is a weak expectation in the sense that it can be easily overridden by specific knowledge. For example, if we knew that Bill was in the habit of "turning the other cheek," then we would not expect retaliation.

It could be argued that any expectation for retaliation in the COMSYS story is really an expectation about an eye for an eye rather than narrative unity. Bill was John's rival and Bill will want to get even. This level of expectation relates to the symmetry of a story. A story is weakly cohesive, if it has a symmetry in its cross-character affect causalities. Retaliation is a plot unit that completes the symmetry of aversive causalities. When Bill refuses

John the job, we have both strong cohesion (total connectivity across plot units), and weak cohesion (symmetry in the cross-character affect links). But when Bill offers John a good job, we have neither strong nor weak cohesion.

When a narrative embodies total symmetry, we detect this immediately and remember it as a salient feature of the story. For example, consider *The Gift of the Magi* by O. Henry. This is a story about a young couple who want to buy each other Christmas presents. They are both very poor. Della has long beautiful hair, and Jim has a prized pocket watch. To get money for the presents, Della sells her hair and Jim sells his pocket watch. Then, she buys him a gold chain for his watch, and he buys her an expensive ornament for her hair. When they find out what they've done, the are consoled by the love behind each other's sacrifices.

The story exhibits an extreme symmetry:

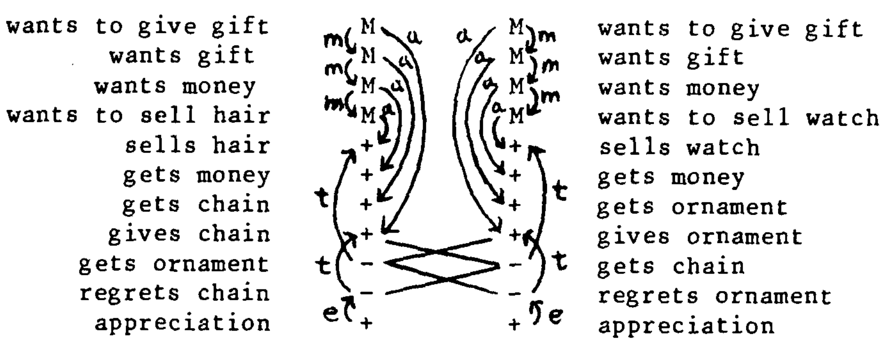


Figure 34.

This configuration involves (1) nested subgoals and (2) achievement (in getting and giving the gifts), (3) loss (in no longer having the things they sold), (4) another loss (in no longer having pleasure from the act of giving) (5) regrettable mistakes (the bad gifts), and (6) hidden blessings (in realizing what the gifts signify). Not only is there complete symmetry across both characters, but there are ironic causalities across the plot units. For example, the sense of loss does not occur until the top-level goals are achieved (when the gifts are exchanged). At the same time, this loss is also the basis for a hidden blessing at the end of the story, when they realize how the gifts signify their unselfish love for each other.

Symmetries of this sort can be heavily exploited in long-term memory representations. For example, a long-term recall of this story might include the fact that (1) Della sold her hair to buy Jim a gift, and (2) Jim bought Della an ornament for her hair. If these facts are augmented by knowledge of symmetry, a subject might then remember that (3) Jim sold X in order to buy the ornament, and (4) Della's gift to Jim was no longer appropriate after he sold X. If (3) and (4) were remembered by symmetric reconstruction, the actual identity of X and Della's gift might be forgotten.

Narrative cohesion will be an important factor for effective memory retention: cohesive texts (as defined by connectivity across plot units) should be remembered with greater accuracy than non-cohesive texts. While this claim is not central to the problem of text summarization, we can expect the two problems to be strongly related. But before we can proceed with either problem, we must become a bit more rigorous about the notion of connectivity across plot units.

8. CONNECTIVITY DEFINED

This section will develop the terminology necessary for a precise statement of our process model. As humans, we can look at graphic affect representations for narratives, and perceive rough degrees of connectivity within those representations. But a computational model that relies on connectivity will have to manipulate a precise formulation of connectivity. So we must now resort to a few dry definitions. Once we have a suitably precise terminology, the actual process model will follow with relative ease.

In all that follows, let A and B be plot units.

DEFINITION: A is *related* to B if and only if A and B share a common affect state. (for convenience, we assume A ≠ B)

DEFINITION: A is *connected* to B if and only if one of the following conditions hold:
a) A = B
b) A is related to B
c) there is a sequence of intervening plot units Ul, . . . , Un such that A is related to Ul, Ui is related to Ui + l, and Un is related to B.

DEFINITION: A *family* around A is the set of plot units that are related to A. The family around A will be designated as F(A).

DEFINITION: A *cluster* around A is the set of plot units that are connected to A.

In all that follows, let F be a family and K be a cluster.

DEFINITION: A *entails* B if and only if all affect states contained in B are also contained in A. (we may say that A entails B or that B is entailed by A).

DEFINITION: Let A be a plot unit contained in K. A is a *top level plot unit in K* if and only if A is not entailed by any other plot units contained in K.

DEFINITION: The *size of K* is the number of top level plot units contained in K. The size of K will be designated as o(K).

DEFINITION: F is a *maximal family* in K if and only if F is a family contained in K and $o(F) \geq o(G)$ for all families G contained in K.

DEFINITION: A is a *pivotal unit* in K if and only if the family around A is a maximal family in K.

DEFINITION: K is a *simple cluster* if and only if K has one pivotal unit.

DEFINITION: We will define a distance metric on K as follows:

(i) $d(A,B) = 0$ if $A = B$

(ii) $d(A,B) = 1$ if A is related to B

(iii) $d(A,B) = k$ if $U_1, \ldots, U_{k-1}$ is the shortest sequence of plot units connecting A and B.

DEFINITION: The *span* of K is defined as the $\max\{d(A,B) | \text{A and B are units in K}\}$

DEFINITION: Let K be a simple cluster. The *depth* of K is defined as the $\max\{d(A,B) | \text{A is the pivotal unit, B is a unit in K}\}$.

These definitions describe simple graph structures that can be readily recognized in pictorial representations. We will look at three examples of plot unit graphs, but first, a few observations:

1. Maximal families may contain more than one pivotal unit. It is therefore possible to have a cluster with a unique maximal family that is not a simple cluster.
2. The definition for relatedness describes the simplest condition possible. We may later need to refine this to distinguish units that share n affect states ($n = 1,2,3$, etc.) and units whose shared affect states have certain properties of connectivity in terms of the affect links between them.
3. The notion of a top-level unit is relative to the specification of some set of plot units. This allows us to examine the effect that different set specifications have on summarization behavior. For example, if we didn't include a unit for a denied request, the top-level units for that configuration would drop down to the request, success, and positive mixed event. Various set specifications might be a key to individual differences in summary behavior.
4. More entailment between plot units results in simpler graph structures.
5. Larger units (in terms of affect states) are likely to result in greater connectivity, as well as simpler graph structures.

The best way to get a sense of all this is to play with some concrete examples of the definitions in action.

Consider the story of John's broken engagement:

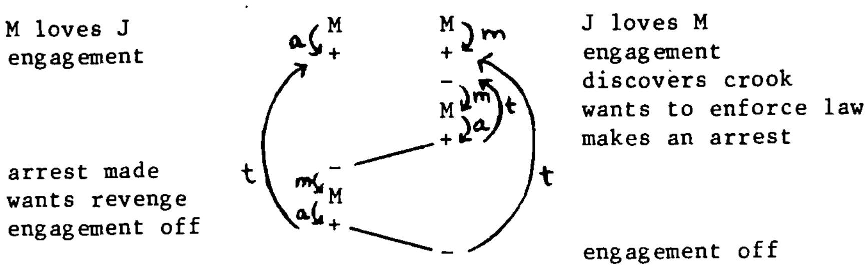


Figure 35.

There are six top-level plot units:

- Mary's success [MS]
- John's success [JS]
- Mary's trade-off [TO]
- Mary's retaliation [RET]
- John's resolution [RES]
- John's loss [JL]

The families for these units can be represented with a connectivity graph:

```
F(MS) = {TO}
F(JS) = {JL}
F(RES) = {RET}
F{JL) = {RET, JS}
F(TO) = {RET, MS}
F(RET) = {TO, RES, JL}
```

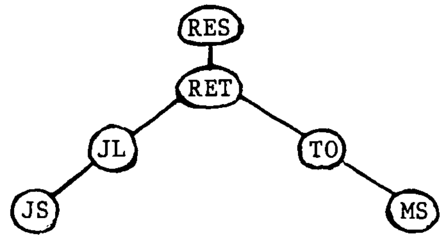


Figure 36.

There is only one maximal family and one pivotal unit (RET). This makes the cluster of six units a simple cluster. It has a depth of 2 and a span of 4.

****

Now consider the COMSYS story.

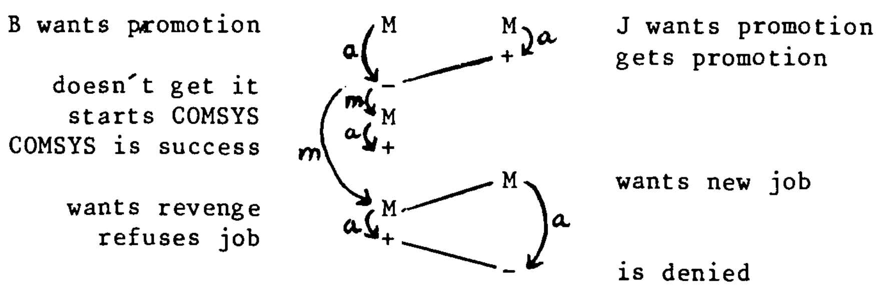


Figure 37.

There are four top-level plot units:

- success born of failure [SBF]
- competition [COM]
- retaliation [RET]
- request denied [RD]

The connectivity graph for these families is:

```
F{RD) = {RET}
F(COM) = {RET, SBF}
F(SBF) = {COM, RET}
F(RET) = {COM, RD, SBF}
```

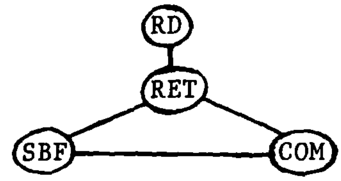


Figure 38.

There is only one maximal family and one pivotal unit (RET). This yields a simple cluster with a depth of 1 and a span of 2.

****

Both of the previous stories result in fairly simple affect connectivity. For our last example, we will look at *The Gift of the Maji.*

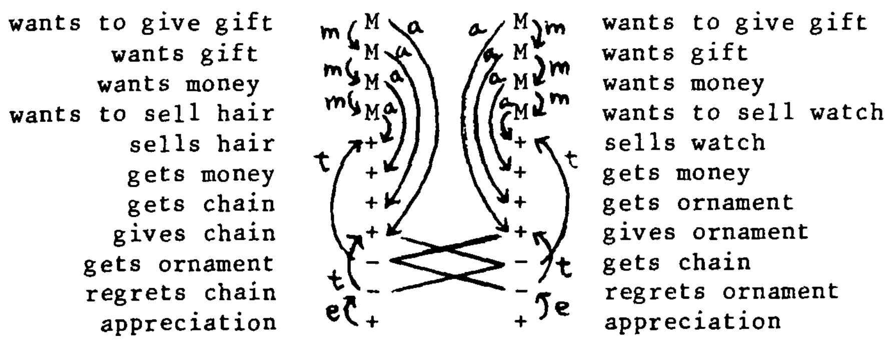


Figure 39.

Now we have 10 top-level plot units. Because of the symmetry of the story, we see the same units appearing for both the husband and the wife. These will be prefaced with "H" and "W" to signify which is which.

- [HN] nested subgoals [WN]
- [HM] regrettable mistake [WM]
- [HL1] loss of object [WL1]
- [HL2] loss of achievement [WL2]
- [HB] hidden blessing [WB]

The connectivity graph for these looks like:

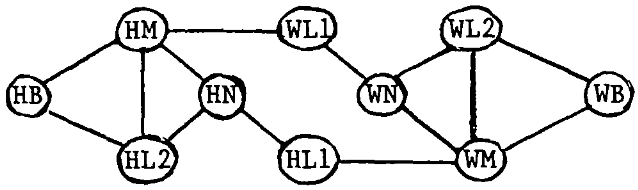


Figure 40.

There are two maximal families, F(HM) and F(WM), with pivotal units HM and WM. Because this is not a simple cluster, the depth of the cluster is not defined. The span is 5.

9. SUMMARY GENERATION

Once the connectivity of the plot units has been established, we can drive a process of summary generation based on affect analysis. The generation of summaries for arbitrary narratives will require an extensive process model that can handle various classes of plot unit configurations. For example, the summary process for a simple cluster will have to be different from the summary process for a cluster with multiple pivotal units.

We will not attempt to present the complete solution here. But, we can discuss the solution for a simple case in order to illustrate the techniques needed for the general case. We will therefore outline the process model for summaries of simple clusters, and then discuss methods for extending this solution to arbitrary clusters.

9.1 Summarization of Simple Clusters

The algorithm for generating summaries of simple clusters is a five-step process:

- STEP 1: Find all top-level plot units in K.
- STEP 2: Derive the plot unit graph structure.
- STEP 3: Identify the pivotal unit, P.
- STEP 4: Generate a base-line summary (S) from a frame for P.
- STEP 5: Augment S with information from plot units related to P.

Steps 1, 2, and 3 are simple manipulations based on the definitions of section 6.[2] Steps 4 and 5 require some explanation.

[2](Step 3 involves summing the rows of an adjacency matrix)

9.1.1 Plot Unit Generational Frames. All plot units are associated with generational frames which designate how the plot unit can be expressed in natural language. For example, a frame for the "competition unit" might look like this:

X and Y both [M], but Y [+].

where X, Y, M, and + are slots in the affect configuration:

Figure 41.

This frame would give us summaries like, "Fred and Hank both loved Mary, but Hank married her." Or, "Bill and John both wanted the same job at IBM, but John got it." While this is a general frame that can be applied to any situation of competition, this very general frame may be overridden by knowledge-specific frames which are dependent on the specific instantiations of the affect states. For example, when the competition is for a job promotion, we can say "Y was promoted over X at IBM," and this will convey the entire competition unit as well. This knowledge-specific frame can be invoked whenever the concept of a promotion appears in the parallel mental states by a scheme of double indices on competition and promotion. The specification and selection of knowledge-specific generational frames will be a major problem for the production of smooth summaries.

9.1.2. Integrating Related Units. Once a generational frame is chosen for the pivotal unit, we will transform the resulting base-line summary into a final summary by integrating any additional information from plot units related to the pivotal unit. By delimiting our integration to those units which are directly related to the pivotal unit, we essentially delete from our summary any information that is more peripheral to the heart of the cluster. The effectiveness of this cut-off heuristic is open to further investigation. Perhaps the cut-off boundary should be a function of cluster depth and/or span.

The actual integration of new information into the base-line summary can be handled in roughly two ways: (1) the addition of a new clause, or (2) the further refinement of existing references in the base-line summary. To see how these two techniques work, we will consider the "COMSYS" and "Broken Engagement" stories.

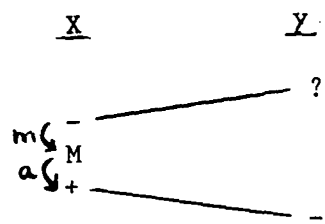


Figure 42.

We will use a general frame for retaliation in our base-line summaries:

"Because Y's [?] caused a [–] for X, X (later) [+]ed to cause a [–] for Y."

This frame allows us to build base-line summaries for both stories:

> "Because John prevented Bill from getting a job at IMB, Bill later prevented John from getting a job."

> "Because John did something bad to Mary's father, she prevented his engagement."

Notice that both of these summaries already convey the plot units for John's failure, Bill's failure, and John's loss. This is because the specification of a negative event that is part of a failure or loss unit will automatically communicate the notion of that failure or loss.

The base-line summary for the "COMSYS" story must now be augmented by the units for competition, success born of failure, and a denied request. Competition and the denied request will be integrated by a further specification of existing references. Success born of failure will require a new clause:

S: "Because John prevented Bill from getting a job at IBM, Bill later prevented John from getting a job."

S + competition:

> "Because John was promoted over Bill at IBM, Bill later prevented John from getting a job."

S + competition + success born of failure:

> "Because John was promoted over Bill at IBM, Bill started his own company, and later prevented John from getting a job."

S + competition + success born of failure + denied request:

> "Because John was promoted over Bill at IBM, Bill started his own company, and later refused to give John a job when he asked for one."

The base-line summary for the "Broken Engagement" story must be augmented with the units for the problem resolution and the trade-off. Both of these units will be integrated by further specification of existing references.

S: "Because John did something bad to Mary's father, she prevented his engagement."

S + problem resolution:

"Because John arrested Mary's father, she prevented his engagement."

S + problem resolution + trade-off:

"Because John arrested Mary's father, she called off their engagement."

While this algorithm specifies the general structure of the summarization process, there are a number of problem areas which require extensive work:

1. Plot unit generational frames must be specified for both simple and complex plot units.
2. Knowledge-specific generational frames must be designed for those concepts which lend themselves to special verbs or constructions.
3. A selection process for the best generational frame must be designed.
4. The integration of additional information into the base-line summary must be described in detail for both further specifications and clause additions.
5. The interplay between (3) and (4) must be studied.

In addition to these fundamental problems within the process model, we must also examine the problems of summarization for clusters of more than one pivotal unit, and stories of more than one cluster.

9.2 Summarization of Arbitrary Clusters

It may be the case that stories involving more than one cluster cannot be easily reduced to a one-sentence summary. In fact, when a story has a single cluster, but that cluster has more than one pivotal unit, it may be difficult to derive a one-sentence summary. These hypotheses can be tested by examining a number of stories. For now, we will look at *The Gift of the Magi* to get a sense of what the difficulties are.

Recall that there were 10 top-level plot units for this story:

- [HN] nested subgoals [WN]
- [HM] regrettable mistake [WM]
- [HL1] loss of object [WL1]
- [HL2] loss of achievement [WL2]
- [HB] hidden blessing [WB]

"Loss of object" refers to the sense of loss experienced in no longer having his watch (or her hair). "Loss of achievement" refers to the sense of loss experienced in giving a gift that turns out to be a mistake (they can no longer feel good about the gifts they gave). The first sense of loss occurs as soon as the gifts are exchanged: they each realize that the gift they received is inappropriate. The second sense of loss comes with the regrettable mistake: they each realize that the gift they gave is inappropriate. The connectivity graph reveals that there are two pivotal units:

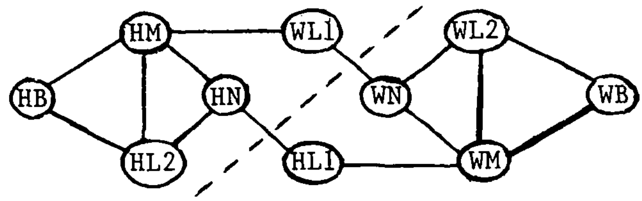


Figure 43.

HM and WM yield maximal families; and the units for nested goals (WN and HN) and loss of objects (WL1 and HL1) are "boundary units" in the sense that they are related to units from both maximal families. Before we discuss possible algorithms, let us look at a sample summary:

> A woman sold her long locks of hair so she could buy her husband a watch chain for Christmas. But when she gave him the chain, she found out that he had sold his watch so he could buy her a comb for her hair. Initially they regretted their expensive gifts, but then they realized how much love was signified in the sacrifices made.

This summary assumes the wife's point of view for the first two sentences. Because of the symmetry in this story, it is natural to infer plot units concerning the husband as information becomes available. Let us take another look at this summary, this time identifying plot units as they are conveyed. Plot units in parentheses are inferred by shifting the perspective:

> A woman sold her long locks of hair so she could buy her husband a watch chain for Christmas [WN]. But when she gave him the chain, she found out that he had sold his watch [WL2,(HL1)] so he could buy her a comb for her hair [HN,WL1,(HL2)]. Initally they regretted their expensive gifts [WM,HM], but then they realized how much love was signified in the sacrifices made [WB,HB].

The boundary units seem to be conveyed first, while the two units with the largest span occur at the very end. If this story is representative, it suggests that summaries should start with units that bound maximal families, and then proceed to those units which are more isolated later. In fact, we would have an acceptable summary, if we deleted the last sentence altogether.

While pivotal units are central for stories with simple clusters, clusters with more than one maximal family are organized around the boundaries between those families. This summary started off with WN, a boundary unit from F[WM]. The other boundary unit from F[WM] is HL1. Can we build a summary starting with HL1? Try this one:

> When Jim received a watch chain from Della for Christmas, he explained that he had sold his watch [HL1,(WL2)]. He sold his watch so he could buy Della a comb for her hair [HN], but he didn't know that she sold her hair [HL2,(WL1)] in order to buy him the watch chain [WN].

In the first summary HL1 follows from WL2 by inference. In the second summary WL2 follows from HL1 by inference. Similarly, in the first summary, HL2 follows from WL1 by inference, and in the second summary, WL1 follows HL2. Since these pairs of plot units are inferentially dependent on each other, we should not allow them to distract us from the actual flow of control that is at work here. To see the pattern emerge, let us identify the HL1-WL2 pair as "X" and the HL2-WL1 pair as "Y". The order of presentation for plot units in the first summary is:

$$\mathrm{WN} - \mathrm{X} - \mathrm{HN} - \mathrm{Y}$$

The order of presentation in the second summary is:

$$\mathrm{X} - \mathrm{HN} - \mathrm{Y} - \mathrm{WN}$$

These presentations differ by a simple rotation of one unit. We could change the perspectives on these two summaries to get:

$$\mathrm{HN} - \mathrm{Y} - \mathrm{WN} - \mathrm{X}$$

$$\mathrm{Y} - \mathrm{WN} - \mathrm{X} - \mathrm{HN}$$

which gives us all four rotations. There are 20 more possible arrangements, and it is possible to generate summaries that correspond to all the permutations. Therefore, any random ordering of the four boundary units will provide a good summary. Can we get a decent summary out of anything less than these four units? Consider:

> Della had sold her long locks of hair to buy her husband a watch chain [WN], and he sold his watch to buy her a comb for her hair [HN].

This summary is based on two boundary plot units of maximal connectivity.

$$o(F[\mathrm{WN}]) = o(F[\mathrm{HN}]) = 3 > 2 = o(F[\mathrm{WL1}]) = o(F[\mathrm{HL1}]).$$

Perhaps minimal summaries can always be derived from maximally connected boundary units. To answer this question and others like it, we have to study the affect analysis for a number of narratives. How often do clusters arise with two or more maximal families? Can stories with multiple clusters be reduced to single sentence summaries? Does the algorithm outlined in section 9.1 work for all simple clusters? How does symmetry affect the process of summarization? These questions can only be resolved by testing proposed algorithms on a variety of narratives.

10. RECOGNIZING PLOT UNITS

Thus far, we have explained how plot units can be used to generate summaries, but we haven't explained where the plot units come from in the first place. While the intuitive notion of plot units may be attractive, this is a useless notion for a process model, unless we can specify the processes which will analyze text and produce plot units as output.

The first step in the derivation of plot units is the derivation of affect states. We cannot possibly identify a plot unit unless its component states are available to us first. The identification of an affect state is actually a fairly straightforward process, if we can assume the computional power of a predictive knowledge-based story understander. Using the knowledge structures outlined in (Schank & Abelson, 1977), we can recognize affect states in terms of fairly fundamental taxonomies:

MENTAL STATE (M)

1. initiating an A, D, E or I – goal
2. missing an enabling condition
3. needing a goal subsumption state
4. suspension of absence of a positive interpersonal theme
5. a plan is intended

POSITIVE EVENT (+)

1. achieving an A, D, E, or I – goal
2. obtaining a necessary enabling condition
3. achieving a goal subsumption state
4. initiating or resuming a positive interpersonal theme
5. intended plan succeeds
6. getting news about (+:1-5) for some person you care about
7. getting news about (–:1-7) for some person you dislike or hate
8. getting news about (M:1-5) for some person you dislike or hate

NEGATIVE EVENT (–)

1. A, D, E, or I – goal is thwarted
2. P or C – goal is initiated
3. script interference is encountered
4. initiation or intensification of a negative interpersonal theme
5. intended plan fails

6. termination of a positive interpersonal theme
7. losing a necessary enabling condition
8. getting news about (– :1–7) for some person you care about
9. getting news about (M:1–5) for some person you care about
10. getting news about (+ :1–5) for some person you dislike or hate

A discussion of this terminology would take us far afield from our central concerns, but the interested reader can find ample discussion of these references in (Schank & Abelson, 1977). Recognition for these entities has been implemented in a number of knowledge based systems, including SAM (Cullingford, 1978), PAM (Wilensky, 1978), and BORIS (Dyer & Lehnert, 1980, Lehnert et al., 1981).

Once the three primary affect states are recognized, we can implement a predictive system of demons to build specific plot units. For example, the appearance of a mental state should construct a demon that can be activated by

1. another mental state (a possible m, t, or e-link)
2. a positive event (a possible a-link)
3. a negative event (a possible a-link)

If another mental state is encountered, the demon should check to see if there is a subgoal relationship (m-link), mutual exclusion (t-link), or equivalence (e-link) at work. If one of these can be verified, we have identified a primitive plot unit.

As soon as a primitive plot unit is identified, demons for complex plot units are constructed. For example, a subgoal unit should predict the possible occurrence of such nested subgoals, as a request, a threat, or a kind or malicious act. In this way, a hierarchical structure of predictions can be implemented which looks for successively complicated plot units, as information appears to support those possibilities.

The verification of specific affect links will rely on specific instantiations behind each affect state. There is no way of knowing whether or not two mental states should be joined by a motivational link, unless we can establish a subgoal relationship between them. This subgoal relationship is naturally dependent on the specific content of each mental state. But these checks do not ask for anything that a standard inference mechanism would not need to know anyway. The information needed for these verifications should already be present in the system for inference purposes that are independent of plot unit recognition. Therefore, we are essentially constructing these units as a side effect of existing processes. We do not need to propose additional knowledge structures to build plot units. They will fall out quite naturally from other memory manipulations with a minimal amount of extra processing.

11. CONCLUSIONS

It appears that a high-level analysis for narratives can be derived from configurations of three primary affect states. These configurations consist of primitive and complex plot units whose overlapping structures allow us to measure the connectivity and symmetry of character interactions. Once the plot units in a story have been properly identified, they provide us with a framework for text summarization.

Relatively little progress has been made on the problem of text summarization. Thus far, the only program to attempt it has been the FRUMP system [DeJong, 1979]. FRUMP analyzes UPI stories in about 50 domains, and provides summaries based on a top-down extraction of relevant information in those domains. Summaries driven by a single knowledge domain do not exhibit much variation, since the summaries are all based on an a priori set of expectations about that domain. For example, an earthquake story will be summarized in terms of (1) where it occurred, (2) what the Richter scale registered, and (3) how many people were killed or injured. All earthquake summaries will describe those three components when they are available. This style of summarization is completely top-down and driven by specific expectations. FRUMP cannot deal with unexpected information, and its summaries will reflect total ignorance of anything unexpected, even if the unexpected information is critically important.

Other top-down strategies have been proposed (although not implemented) which rely on a story grammar approach to text analysis (Rumelhart, 1975, 1977; Thorndyke, 1977). This approach is best typified by Rumelhart, 1975). Rumelhart points out that a number of short narratives fall into what he calls the "EPISODE" schema. The EPISODE schema about protagonist "P" consists of:

1. EVENT "E" CAUSES "P" TO DESIRE GOAL "G"
2. "P" TRIES TO GET "G" UNTIL OUTCOME "O" OCCURS

Each of the relational terms in this schema (CAUSE, DESIRE, & TRY) refer in turn to other schema which will likewise be instantiated by particular variables within a given story. The EPISODE schema provides a root node for a hierarchical tree structure that will expand to arbitrary depth as the schemata on each level are instantiated and expanded in a recursive manner.

Rumelhart uses summarization data to illustrate how various levels of detail coincide with expansions to a particular level within the tree structure. A level 0 summary is based on the root node alone. A level 1 summary is based on the first level of expansion from the root, a level 2 summary is based on the second level of expansion, and so forth. While his data supports story grammars very generally, a closer analysis of summarization data comparing specific story grammar predictions to plot unit predictions appears to verify plot unit analyses over story grammars (Lehnert, Black, & Reiser, 1981).

Rumelhart does not discuss the recognition problem for his top-down analysis of stories. Story grammars have since been criticized for being computationally naive [Black & Wilensky, 1979], and therefore of dubious value in a process model of narrative text comprehension. But apart from these processing issues, there is the overwhelming limitation of all top-down processors: a story grammar cannot characterize input that does not conform to its expectations. Just as FRUMP cannot deal with input outside of its knowledge domains, a story grammar would be of no help when confronted with a story whose plot was not a priori anticipated by the grammar. To what extent does *The Gift of the Magi* conform to the EPISODE schema? A hierarchical story grammar simply cannot be general enough to capture large variations in plot structures. Even when knowledge of narratives does seem to be operating in a predictive manner, many of these predictions can be incorporated in terms of plot unit and affect state predictions as well (consider the symmetry arguments for strong and weak cohesion in section 7).

There is infinite variation in the number of plots that are possible, and people can understand a story with a new plot line whether they've seen a similar plot before or not. This suggests that plot recognition must be based on bottom-up processing, rather than a top-down analysis. We can attain the flexible recognition capabilities of a bottom-up analysis scheme by constructing configurations of primitive and complex plot units. Most of the information needed to recognize affect states has already been incorporated in predictive knowledge-based systems (Schank & Riesbeck, 1981) to drive various inference mechanisms; and the extra processing needed to link these affect states together into plot units is not difficult [Reiser, 1981].

Because affect states are based on information about plans, goals and themes, affect analysis will not be applicable to stories which do not contain information along these lines. Using this approach, we will not be able to handle descriptions of sunsets, burnt steaks, or waking up in the morning. Or course, it is not clear that people can comfortably summarize stories that center on perceptual descriptions either, so this limitation is not a cause for concern. It would not be difficult to generate a summary that said, "This is a story describing a sunset." That is probably what a human would be reduced to as well.

On a related note, we should point out that plot unit analyses are also inappropriate for expository texts. There is some evidence that expository texts are structured by hierarchical trees (Kintsch, 1974, Kintsch & vanDijk, 1975); and the applicablity of plot units (or lack of it) to a given text might provide us with a criterion for distinguishing an expository text from narratives. For example, the "Circle Island" story (Thorndyke, 1977) cannot be analyzed with plot units, and would therefore be classified as expository text rather than narrative text.

In this paper, we have stressed the relevance of affect analysis for the task of narrative summarization. There are also other applications to explore with plot units. For example, a high-level analysis of a story is probably used as an index into long-term memory. Such an index would determine when the story can be remembered, and under what conditions information from the story can be accessed (Schank, 1979). The ability to recognize similarities across stories should also be affected by a high level of narrative analysis. In a recent experiment, 36 subjects were instructed to group 36 stories according to their plot similarities. A cluster analysis of the data shows that plot unit structures predict the resulting groups very effectively (Reiser, Lehnert and Black, 1981). Plot units may also have a place in generational tasks. While processes of generation and understanding differ by much more than a procedural inverse (Meehan, 1976), it might be the case that stories are generated from initial affect configurations as the starting point.

While all of these other areas are potentially relevant to plot units, the task of narrative summarization has the advantage of presenting a relatively clean I/0 problem. We can give stories (input) to human subjects, and ask them to produce summaries (output). This data is initially valuable in the design of a summarization process, and it will eventually allow us to test resulting programs for their psychological ability. If the process model becomes sufficiently sophisticated, it would be appropriate to study individual differences across subjects. Perhaps it would even be possible to analyze a subject's summarization behavior on a few key texts, and then predict subsequent summarization behavior on completely different texts. For the moment, however, it will suffice to specify a system that can generate reasonable summaries for a variety of narrative texts. This paper has attempted to show how such a goal may be realized in the near future.

REFERENCES

Black, J. B., & Wilensky, R. "An Evaluation of Story Grammars." *Cognitive Science, Vol. 3,* No. 3, 1979, 213-230.

Cullingford, R. E. "Script Application: Computer Understanding of Newspaper Stories." Department of Computer Science Research Report #116, Yale University, New Haven, Conn., 1978.

DeJong, G. F. "Skimming Stories in Real Time: An Experiment in Integrated Understanding." Department of Computer Science Research Report #158, Yale University, New Haven, Conn., 1979.

deRivera, J. "A Structural Theory of the Emotions." *Psychological Issues,* Monograph No. 40. New York: International Universities Press, 1977.

Dyer, M., & Lehnert, W. "Memory Organization and Search Processes for Narratives." Department of Computer Science Research Report #175, Yale University, New Haven, Conn., 1980.

Graesser, A. *Prose Comprehension Beyond the Word.* New York: Springer-Verlag, 1981.

Kintsch, W. *The Representation of Meaning in Memory.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1974.

Kintsch, W., & vanDijk, T. "Recalling and Summarizing Stories." *Language, Vol. 40,* 1975, 98–116.

Lehnert, W. G. *The Process of Question Answering.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1978.

Lehnert, W., Black, J., & Reiser, B. "Summarizing Narratives" *Proceedings of the 7th International Joint Conference on Artificial Intelligence.* Vancouver, B.C., 1981.

Lehnert, W. G., Dyer, M. G., Johnson, P. N., Yang, C. J. "S. BORIS—An experiment in in-depth understanding of narratives." Accepted for publication by the *Journal of Artificial Intelligence.*

Meehan, J. R. "The Metanovel: Writing Stories by Computer," Department of Computer Science Research Report 74, Yale University, New Haven, Conn., 1976.

Reiser, B. R. "Story structure and summarization" (unpublished manuscript) Department of Computer Science, Yale Univeristy, New Haven, Conn., 1981.

Reiser, B. R., Lehnert, W. G., & Black, J.B. "Recognizing Thematic Units in Narratives." *Proceedings of the Third Annual Conference of the Cognitive Science Society,* Berkeley, Calif., 1981.

Roseman, I. "Cognitive Aspects of Emotion and Emotional Behavior," 87th Annual Convention of the American Psychological Association, New York, N.Y., 1979.

Rumelhart, D. E. "Notes on a Schema for Stories." In D. G. Bobrow, & A. M. Collins (Eds.), *Representation and Understanding.* New York: Academic Press, 1975.

Rumelhart, D. E. "Understanding and Summarizing Brief Stories." In D. Laberge, & S. Samuels (Eds.), *Basic Processing in Reading, Perception, and Comprehension.* Lawrence Erlbaum. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schank, R. C. *Conceptual Information Processing.* North Holland, Amsterdam, 1975.

Schank, R. C. "Reminding and Memory Organization: An Introduction to Mops," Department of Computer Science Research Report 170, Yale University, New Haven, Conn., 1979.

Schank, R. C., & Abelson, R. P. *Scripts, Plans, Goals, and Understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Schank, R., & Riesbeck, C. *Inside Computer Understanding.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1981.

Thorndyke, Perry W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology,* 1977, *9,* 77–110.

Wilensky, R. "Understanding Goal-based Stories." Department of Computer Science Research Report 140, Yale University, New Haven, Conn., 1978.