

WHAT MAKES SAM RUN?
SCRIPT BASED TECHNIQUES
FOR QUESTION ANSWERING

Wendy Lehnert
Yale University

I. INTRODUCTION

The A.I. group at Yale has developed a program SAM (Script Applier Mechanism) which models processes in story understanding. The basic theoretical construct underlying SAM is the notion of a script (Schank & Abelson 75). Script based knowledge is mundane information which tends to lie in the periphery of consciousness. Going to a restaurant, watching a football game, taking a bus, and paying bills are examples of script activities. Scripts describe situations which are frequently encountered with negligible or predictable variations. People in situational scripts are seldom conscious of script related actions. The acts which define a script are things which people automatically do or expect to occur. Since a person who has not seen a football game will not fully comprehend stories about football games, we assume script based knowledge is used in story understanding.

SAM utilizes a restaurant script in order to paraphrase, summarize, and answer questions concerning stories about eating out in a restaurant. The question answering techniques described in this paper were devised as SAM evolved.

The following story is an input story which SAM has processed:

John went to a restaurant. The hostess seated John. The hostess gave John a menu. The waiter came to the table. John ordered lobster. John was served quickly. John left a large tip. John left the restaurant.

SAM responds in a sample question answering session:

- Q. Why did John go to a restaurant?
A. So John could eat.
- Q. Did the waiter give John a menu?
A. No, the hostess gave John a menu.
- Q. Why did the hostess give John a menu?
A. So John could order.
- Q. What happened when the hostess gave John a menu?
A. John read the menu. The waiter saw that John was at the table. The waiter went to the table.
- Q. What did John eat?
A. Lobster.
- Q. Who brought John the lobster?
A. The waiter brought John the lobster.
- Q. What happened when John ate the lobster?

A. John indicated to the waiter that he would like a check. The waiter gave John the check. John remembered that the waiter had served the lobster quickly. John left the waiter a large tip.

Q. Did John enjoy the lobster?
A. Probably.

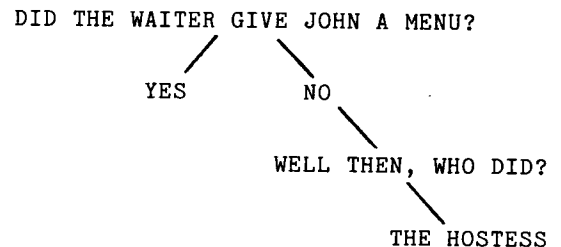
Q. Who paid the bill?
A. John.

Q. Why did John pay the bill?
A. So he could leave.

II. INTERPRETING QUESTIONS: ESTABLISHING FOCUS

A recurring issue in natural language processing which is particularly evident in question answering is the problem of focus. What is interesting about a statement? Where is attention directed? What aspects of a statement are significant? Many questions have more than one acceptable answer and to a large extent people manage to agree on which answers seem most natural. The appropriateness of a response is a function of focus. Questions usually have a focus (or emphasis) which renders one response more appropriate than another.

Consider the sample input story above and the question "Did the waiter give John a menu?" The most natural answer to this is "No, the hostess gave John a menu." To arrive at this response we must go beyond the original yes or no question and answer a second question "Who gave John a menu?"



The interesting problem here is how we picked up this second question. By going on to ask who gave John the menu we have interpreted the original question to focus on the actor who executes the transfer of the menu. How did this emphasis arise? We could have gone on to ask "Well then, what did the waiter give John?" or even "Well then, what did the waiter do?" Emphasis in this direction would elicit answers like:

No, the waiter gave John a check.
No, the waiter brought John his meal.
No, the waiter took John's order.

While each of these is an acceptable answer, they are less natural than:

No, the hostess gave John a menu.

So to arrive at the best answer we have to focus on the actor being the most important or interesting component of the question. How do we do this?

CONJECTURE:

WHEN GIVEN A CHOICE OF FOCUS,
TAKE VARIATION OVER EXPECTATION.

This conjecture is based on the premise that variables are more interesting than constants, i.e. the unexpected is more worthy of attention than the expected.

In general, implementing such a rule may be hard, but within the context of a script, it's easy. Every script is characterized by a set or sequence of actions specific to that script. In a restaurant the patron expects to receive a menu, sit down at a table, order, eat, pay, etc. Expected acts such as these are constants within the script. We are surprised to hear things like:

John went to a restaurant but he didn't eat.
John went to a restaurant and didn't pay the check.
When John went to the restaurant he sat on the floor.

In these cases our expectations have been violated because the script constants of eating, paying, and sitting at a table have been contradicted or over-ruled. When given a question, we examine the question statement in order to establish which components comprise a script constant. Once we know which script constant matches our question statement, we take the object of focus to be that element of the question statement which is not a part of the script constant (if one exists). Since this extraneous element should be a script variable (being non-constant) we have established the appropriate focus.

In our example, the act of transferring a menu to the patron is a script constant. We expect an ATRANS of the menu to John. Who gives him the menu is a script variable since we would not be surprised to hear it was the hostess or a waiter. Perhaps even the cook gave him the menu or he got it himself. A similar situation occurs when John gets the check. We expect him to get a check, but the actor of the transfer is variable. Of course these variables assume default bindings in the absence of explicit data; unless I hear otherwise, I assume the waiter brings the check.

Whenever the answer to a did-question is "No", it is natural to augment the negative response with a correction or explanation of some sort. There are two classifiable situations when the initial response is negative. In one case a focus exists and can be determined by our rule (as in the waiter giving John the menu). In the other case no focus is found in the question. No focus is found in "Did John

sit down?" or "Did John pay the check?" because the actions in question are full script constants with no possibility of variation within the expectations of the script. In such instances where no focus exists the expectations of the script have been violated. (John should have sat down and he should have paid the check.) Whenever expectations are violated, the natural question to be asked is "How come?" This is equivalent to "Why didn't John sit down?" or "Why didn't John pay the check?" Answers to these will either be wierd-oriented answers or interference-oriented answers (see part IV). When the original question statement does have a focus, the answer is found by matching the constant part of the statement against the script acts. Once a script act is matched, we instantiate the variable bindings and return the resulting conceptualization as the best augmentative answer.

So in applying the rule of variation over expectation (V/E) to the question "Did the waiter give John a menu" we identify giving John a menu as a script constant and the actor binding as a script variable. Therefore the focus of attention falls on the actor, and we augment the minimally correct response "No" with the most natural addition, "the hostess gave John a menu."

Script variables also occur in instrumentality, manner, mode, and time fillers, e.g. "Why did John drive to the restaurant?" If we're in the restaurant script, we expect John to get to the restaurant. How he gets there is variable. Applying V/E to the question, we establish the focus to be on driving. "Did John eat his meal in 10 minutes?" We expect John to eat his meal. How long it takes him is variable. Applying V/E we determine the focus to be on the time it took John to eat.

If more than one variable occurs in a question, some hierarchy must be invoked to establish the focus. In answering "Why did John drive to the restaurant at 4:00A.M.?" we presumably find that going at 4:00A.M. is more interesting than driving. People need to be able to resolve focus in order to understand what a question is driving at. "Why did John roller skate to the restaurant at 4:00A.M.?" tends to have the effect of two different questions: "Why did he roller skate?" and "Why did he go at 4:00A.M.?" The ambiguity in this question results from the ambiguity of focus. We have trouble deciding which is more interesting, the mode of transportation, or the hour. When focus is not resolved a question seems confused or ill-defined.

III. ANSWERING WHAT-HAPPENED-WHEN-QUESTIONS

Part of SAM's internal representation for the input story is a causal chain of conceptual dependency diagrams (Schank 75). A causal chain is an alternating sequence of states and actions in which each state enables the following action and each action results in the following state. Understanding the relationship between two

conceptualizations is at least in part reflected by the ability to construct a causal chain between them.

If I hear that John had a bad car accident, I am not surprised to hear that his car was wrecked, or that he was injured, or that Mary doesn't want to ride with him any more. These all relate to the accident as causal consequences. Even when expectations are violated as in "John was in a car accident. He was overjoyed." We try to make sense out of it by constructing the most feasible causal chain we can: Maybe the car was worthless anyway and John was not too badly hurt but he was insured and they're giving him a big settlement and he can really use the money for some reason.

When SAM processes an input story, causal chains are established between consecutive input conceptualizations. Generating causal chains in a situational script is easy because the script contains all the expected actions which will fill in a chain between any two acts of the script. A major part of the script data base consists of various causal paths throughout portions of the restaurant script. When SAM receives a what-happened-when question, it matches the act in question against its corresponding script counterpart and simply returns that portion of the causal chain representation of the story which begins with the act in question and ends at the next conceptualization mentioned in the input story.

IV. ANSWERING WHY-QUESTIONS

Once we have interpreted a question by establishing its focus, we still have to answer the question. The most interesting class of questions in this respect seem to be why-questions. There appear to be roughly four types of answers to why-questions. Two are script based and two require data outside of scripts. The script based answers have implementable heuristics (currently incorporated in SAM).

(1) WIERD-ORIENTED ANSWERS (non-script based)

In any script context we may get an unexpected occurrence which is relevant to the script. Answers dependent on the wierd occurrence may relate back to it in a number of ways. Consider the following examples:

Ex.1: John went to a restaurant and broke his wrist when the chair he was sitting on collapsed. John sued the restaurant.

Q. Why did John sue the restaurant?
A. His chair collapsed and he was injured.

Ex.2: John went to a restaurant and found out that everyone got a free drink of their choice. John ordered the cheapest drink they had.

Q. Why did John order a cheap drink?
A. I have no idea.

Ex.3: John went to a restaurant and ordered a hamburger. When the waiter was

carrying it from the kitchen he dropped it on the floor. John ate the hamburger and left.

Q. Why did John eat the hamburger?

A. He must not have known it was dropped.

In Ex.1, a causal chain can be constructed between the wierd occurrence and the act in question. The act in question is consistent with our expectations after the wierd occurrence; the chair collapsing and resulting injury are the causal antecedents of John suing the restaurant.

In Ex.2, no causal chain can be constructed between the wierd occurrence and the act in question, so we are at a loss to answer the question.

In Ex.3, our expectations are violated as in Ex.2, but here we can account for the discrepancy, and we use the explanation as our answer. We expect a causal chain which includes John refusing the hamburger. Since this construction is contradicted when we hear that John ate the hamburger, we reconstruct the causal chain and account for the validity of the new construction in our answer.

The difficulties in arriving at answers of this type are apparent:

- 1) Since scripts normally run in the background of a story line and are rarely in the foreground, we need to be able to identify wierd occurrences as distinguished from commonplace occurrences which are irrelevant to the script. For example, how do we know that smoke coming from a wall is wierd and smoke coming from an open barbecue pit in a steak house is OK. Similarly, if John stands up and starts making a toast, this is not wierd unless perhaps there is no one else at his table. Some very strong inference mechanisms or higher level structures must come into play in the problem of recognizing wierdness.
- 2) We need to know if the act in question is consistent with the wierd occurrence (as in Ex.1) or if it violates expectations (as in Ex.2). This is equivalent to knowing when a causal chain can be constructed between two conceptualizations and when no such chain exists.
- 3) If our expectations have been violated, we need to be able to construct feasible explanations whenever possible (as in Ex.3). Constructing a feasible explanation is equivalent to constructing a believable causal chain. In Ex.3, the causal chain behind our explanation is arrived at by suppressing the inference that John knew about the waiter dropping his hamburger. Since this is the key to a valid causal construction, we zero in on it for our answer.

In general the problems of recognizing an unusual occurrence or constructing a causal chain are major issues which are far from resolved.

(2) EXTERNALLY-ORIENTED ANSWERS
(non-script based)

Questions like "Why did John walk to the restaurant?" or "Why did John order a hamburger?" require data from outside of the script. Little can be said about these general script exits until we have developed some data structures outside of scripts.

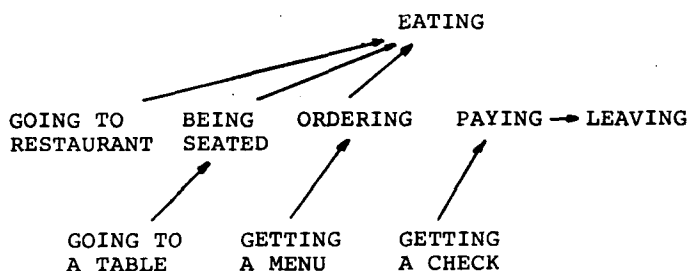
(3). GOAL-ORIENTED ANSWERS (script based)

These occur in one of two ways:

- 1) The focus of the question (as determined by V/E) is a variable whose default binding is a character in the script.
- 2) The question has no focus (via V/E).

- Q1: Why did John go to the restaurant?
(focus:nil)
Q2: Why did John go to a table?
(focus:nil)
Q3: Why did the hostess give John a menu?
(focus:hostess)
Q4: Why did the waiter give John a check?
(focus:waiter)
Q5: Why did John pay the check?
(focus:nil)

Each script has a static goal structure which consists of scriptgoals and a set of subgoals. The subgoals may exist on different levels of detail. The hierarchy of the restaurant script has only one level of subgoals and one scriptgoal. The goal structure for the restaurant script looks like:



The top level of this structure contains the scriptgoal of eating. The second level represents the subgoals of the restaurant script and the third level contains other acts found in the script (not all shown).

Goal-oriented answers are derived by the following rules:

- a) If the act in question is a subgoal, go to the next goal in the next level up. If no such goal exists, go to the next goal in the same level.

b) If the act in question is not a subgoal, go to the next goal in the lowest level of subgoals.

c) If the act in question is a scriptgoal, there is no goal-oriented answer. It probably has an externally-oriented answer.

Using the goal algorithm we can answer questions Q1-Q5.

- A1: So he could eat.
A2: So he could sit down.
A3: So he could order.
A4: So he could pay.
A5: So he could leave.

Notice that this goal structure is oriented with respect to the central character of the script, in this case the restaurant patron. If we were in a restaurant script with respect to the waiter we would answer Q4 with something like "Because it's his job" or "Because John was done eating". Intrinsic to all scripts is a point of view.

(4) INTERFERENCE-ORIENTED ANSWERS
(script based)

These are similar to wierd-oriented answers but are distinguished by being more commonplace. The restaurant script contains alternative paths which contain occurrences of goal interference. For example, if no tables are available, we have interference with the goal of being seated.

Ex.1: John went to a restaurant and ordered a hotdog. The waiter said they didn't have any. So John ordered a hamburger.

Ex.2: John went to a restaurant and was told he'd have to wait an hour for a table. John left.

Ex.3: John went to a restaurant. He read the menu, became very angry and left.

A goal interference predicts an action which will be either a resolution or consequence of the interference. Therefore any question which points to such a resolution or consequence is explained by the interfering occurrence.

Q1: Why did John order a hamburger?
A1: The waiter said they didn't have any hotdogs.

Q2: Why did John leave?
A2: He was told he'd have to wait an hour for a table.

Q3: Why did John leave?
A3: He became very angry.

V. THE THEORETICAL SIDE OF SAM

The problems of interpreting a question or finding the best answer to a why-question are both characterized by the necessity of knowing what is interesting about the question. Interpretation is facilitated by establishing focus. Answering a why-question may entail examination and construction of causal chains or knowledge of goal hierarchies. In any case, the solution to what is interesting lies within some structural representation of the story. When we are within the confines of a script, the problem is relatively trivial since the structures we need are predetermined and static. Outside of a script we need dynamic processes which can generate the needed representation as we go along. To date, systems based on uncontrolled inferencing and propositional reasoning have failed to be effective precisely because no higher level structures were invoked to give the processing direction. The difference between a blind inferencing mechanism and a clever one is this crucial ability to determine what is deserving of attention. By studying the structures implicit in scripts, we may gain some insight concerning what types of guidance mechanisms exist and how analogous structures may be generated in contexts beyond scripts.

The proposed heuristic of variation over expectation is theoretically significant insofar as it suggests an alternative to what might be called a propositional approach to memory retrieval. Suppose we know that the host gave John a menu, and we need to answer "Did the waiter give John the menu?" How are we to answer this question without recourse to scripts or the idea of focus? Suppose we approached the problem propositionally. One possible line of analysis might entail the following reasoning:

- (i) The act of transferring a menu to a restaurant patron usually occurs once in the course of a dinner out.
- (ii) The act of transferring a menu to a restaurant patron is executed by one actor only.
- (iii) The host and the waiter are two different actors.

Given these three suppositions and some deductive reasoning capacity, we are in a position to conclude that the answer to the question is "No".

There are a number of problems with an approach of this type. In the first place, it is probably impossible to implement. We need some very clever inferencing to pull (i) out of the blue. Then deduction and inferencing must combine in some mysterious way to extract (ii) from (i). All in all, the whole argument smells like theorem proving, a technique which has proved ineffective and is certainly not the way people work. But ignoring all these objections, even if you could implement it, the fact remains that this has simply not done a very good job of answering the question. It yields only a minimally

correct response and has no indication of the point of the question; there is no way of knowing how to augment the initial response "No".

By examining non-script based approaches to this question, it seems clear that the best possible answer can be derived only from a data base which enables us to establish the focus of the question. There is no way that the natural answer to this question can be found without some sense of what is interesting about the question.

VI. CONCLUSIONS

In the area of memory organization, there is much controversy over categories of world knowledge and corresponding models of memory. At present, there is an ongoing debate concerning episodic vs. semantic memory (Tulving 72). Episodic memory emphasizes experiential knowledge of the world, while semantic memory accommodates abstractions derived from experience. It is generally conceded that people must have both episodic and semantic knowledge. Contention arises when retrieval mechanisms are described which bias one data structure over another (Schank 74; Ortony 75). The problem of course is which types of knowledge are used for what purposes and how.

Analysis of memory retrieval mechanisms usually proceeds along one of two routes. On one hand, there is speculation about memory retrieval in general, without reference to things people actually do. On the other hand, there are psychological experiments which study very specific tasks that people never encounter outside of a psychological test. Neither approach has taught us much about the nature of human memory. The development of computer models has the distinct advantage of forcing us to identify and account for memory processes which people really have and use all the time.

Trying to answer whether or not the waiter gave John a menu led to the concept of focus and a heuristic for determining focus. Question answering using focus works because it is founded on recognizing what people find interesting. As people live from day to day, they experience various activities and situations. Some of these activities are more engaging than others, and some situations are more interesting than others. If we can discover a metric which assesses the relative interest-appeal of assorted human experiences, then we can use this metric to establish general focus in story understanding. Whatever metric we design will have to examine experiential data bases since the phenomenon of being interested in something is inherent in experience and cannot be derived.

A system relying on purely semantic data will never know where to focus because the experiential element of what is interesting has been distilled out of its data base. It might be argued that perhaps a function exists which would operate on a

semantic network of propositions and evaluate the focus of a statement or story. Suppose this could be done. Then what is the point of abstracting experiential data in the first place? Why develop a purely semantic conceptual representation if we're just going to turn around and recreate the experiential data that's been thrown away?

No one is denying that people have the ability to abstract principles from experience and acquire knowledge which is not episodic in nature. We all know that most swans are white and Ancient Greece was polytheistic. The issue is a question of exactly where and how semantic knowledge is used in natural language processing. SAM has demonstrated the power of episodic memory organization in the task of story understanding and question answering. While it is certainly not true that episodic memory is going to account for the memory organization underlying all thought processes, we are constructing models which illustrate a theory of episodic memory in language processing.

REFERENCES

- Ortony, A., How Episodic is Semantic Memory? In Proceedings of Theoretical Issues in Natural Language Processing, Cambridge MA, 1975.
- Schank, R.C., and Abelson, R.P., Scripts, Plans, and Knowledge. Presented at the 4th International Joint Conference on Artificial Intelligence, Tbilisi, USSR. August, 1975.
- Schank, R.C., Is there a Semantic Memory? Castagnola, Switzerland: Istituto per gli Studi Semantici e Cognitivi, 1974 (mimeo).
- Schank, R.C., The Structure of Episodes in Memory. In D.G. Bobrow and A.M. Collins (eds.), Representation and Understanding. New York: Academic Press, 1975.
- Tulving, E., Episodic and Semantic Memory. In E. Tulving and W. Donaldson (eds.), Organization of Memory. New York: Academic Press, 1972.