SCRIPT-BASED GENERATION AND EVALUATION OF EXPECTATIONS IN TRAFFIC SCENES

Gudula Retz-Schmidt

IFI-HH-M-136/85

October 1985

This paper has been presented at the GWAI-85, 9.Fachtagung über Künstliche Intelligenz, 23.9. - 28.9.1985, in Dassel/Solling, West Germany. Proceedings as GWAI-85 (H. Stoyan, ed.), Informatik-Fachberichte, Springer-Verlag Berlin/Heidelberg/New York/Tokyo (to be published).

SKRIPT-BASIERTE GENERIERUNG UND EVALUIERUNG VON ERWARTUNGEN IN STRASSENVERKEHRSSZENEN

Zusammenfassung

Die Erkennung und Verbalisierung von Ereignissen in Realwelt-Bildfolgen hängt in bestimmten Fällen (u.a. Futur-, Präsens-, Negativaussagen und Verwendung bestimmter Verben) von Erwartungen ab. Diese können sich u.a. auf Alltagswissen stützen. Diese Arbeit handelt von Skripten als einer Art der Repräsentation von Alltagswissen über stereotype Ereignisfolgen. Sie beschreibt die Verwendung von Skripten als ein Mittel zur Generierung von Erwartungen über Ereignisse in der Domäne der Straßenverkehrsszenen. Die Verwendung dieser Erwartungen und ihrer Überprüfung zum Zweck der Verbalisierung werden vorgestellt. Schließlich werden einige Grenzen dieses Ansatzes diskutiert.

Die in dieser Mitteilung beschriebene Arbeit wurde an der Universität Hamburg als Teil des Projektes NAOS durchgeführt.

SCRIPT-BASED GENERATION AND EVALUATION OF EXPECTATIONS IN TRAFFIC SCENES

Gudula Retz-Schmidt

Fachbereich Informatik, Universität Hamburg,
Schlüterstraße 70, 2000 Hamburg 13

Current address: SFB 314, FB-10 Informatik IV,
Universität des Saarlandes, 6600 Saarbrücken 11

Abstract

The recognition and verbalization of events in real-world image sequences in certain cases (including future-tense, present-tense, negative statements, and the use of certain verbs) depends on expectations. These can be based on common-sense knowledge. This paper deals with scripts as a way of representing common-sense knowledge about stereotypical sequences of events. It describes the use of scripts as a means of generating expectations about events in the domain of traffic scenes. The use of these expectations and their evaluation for the purpose of verbalization are presented. Finally, some of the limitations of this approach are discussed.
The work described in this paper was performed at the University of Hamburg as part of the project NAOS.

1. Introduction

This work is based on the system NAOS, which simulates a human speaker who describes a traffic scene to a hearer who cannot himself see the scene. As described elsewhere [1, 2, 3, 4] the system NAOS uses event models for the recognition of events in traffic scenes. Event models are a representation of classes of events in a relational notation, organized around verbs of change, in NAOS restricted to a subset of the verbs of locomotion, and certain other concepts which are important for a simple scenario (e.g. "stehen" = "stand") [3]. The event recognition process tries to match the event models against the "geometrical scene description" (GSD), the output of the scene analysis component (which is currently still supported by human interaction [4]), stored in an associative database. The recognized events (i.e. instantiated event models) are added to the GSD and can then be verbalized, i.e. used to generate a scene description in natural language [4].

But there are still many sensible statements describing a scene which the process described above cannot yield. For instance in certain situations it might be sensible to include the negative statement

"Der Bus hielt nicht an der Bushaltestelle an." ("The bus didn't stop at the bus stop.") in the description of a scene. On the other hand a lot of other negative statements which are also true in the same situation, like for instance "The lorry didn't take off and fly away.", are not sensible. What makes the difference between the former and the latter? The reason why some negative statements are sensible in a given situation, whereas others are not, is that they express the fact that an expectation of the speaker has been contradicted (or the speaker's model of the hearer's expectations; this aspect is not dealt with in NAOS except in the case of answering explicit questions of the hearer).

The same phenomenon is involved in the German verbs "weitergehen", "weiterfahren", and "stehenbleiben", which mean "not stop walking", "not stop driving", and "not resume moving" respectively and thus are one-word paraphrases of negated verbs.

In order to be able to make negative statements we need a mechanism for generating expectations and comparing them with reality.

We encounter a similar task when we want to make statements in present tense about non-durative, composite events (e.g. "überholen" = "overtake"), i.e. verbalize composite events while they are actually happening and not yet completed. We will only make such a statement if we expect the event to be completed.

Even more extreme in this respect are future-tense statements. They are based on expectations about events that haven't even begun at the time of utterance.

What is needed in all three cases are expectations. We will now look at a model of common-sense knowledge and its use for generating expectations.

2. Scripts

When we are watching a traffic scene, we usually don't know the intentions and goals of the participating agents. Nevertheless in many cases we are able to expect what will happen next. For instance, if we see red traffic lights, we expect cars approaching the traffic lights to slow down and stop in front of them. This is possible because we have knowledge about what events usually happen in certain situations or in connection with other events or, to put it another way, knowledge about typical sequences of events.

Scripts have been introduced as a model of this kind of knowledge. "A script is a structure that describes appropriate sequences of events in a particular context." [5, p.41]. "In actual use, scripts represent a knowledge structure composed of stereotyped sequences of events that define some common everyday occurence, ..." [6, p.6]. "A SCRIPT is a stereotypic event sequence in a specific situational context." [7, p.4].

Scripts were originally developed as a means of supporting story understanding [5, 6, 8]. More generally, they were claimed to be a model of human memory structures that facilitate the understanding of situations and events in daily life [5]. Whether or not one accepts this claim, scripts seem to be a knowledge-representation structure sufficiently general and thus also appropriate for the purpose of understanding scenes.

The next sections will be dealing with the use of scripts for generating expectations in NAOS and with the evaluation of these expectations for the purpose of generating natural-language statements.

3. An Example

To get an idea of the representation of the scripts used in NAOS let us look at a simple example.

A stereotypic event sequence in a situation where an agent is near some traffic lights could be represented as shown in figure 1.

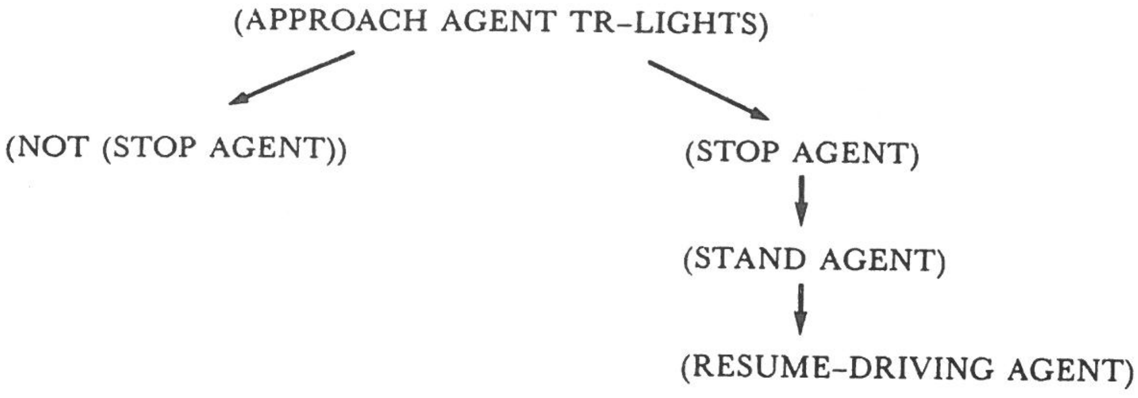

Figure 1: Skeletal TRAFFIC-LIGHTS script

The script in figure 1 has a tree structure. Each node consists of a tuple which specifies an event in a relational notation. AGENT and TR-LIGHTS are variables.

This kind of representation might be sufficient for story understanding, where the task is to fill gaps, i.e. fill in details that have been left out by the storyteller, because expectations derived from it are only used for implicit reasoning. In the case of explicit verbalization, however, more certainty and thus more detailed and exact expectations (including the approximate place and time of occurrence of the events) are needed. Hence the representation scheme for the description of scenes needs to be more powerful.

Figure 2 shows an example of the representation used in NAOS.

In figure 2 each node consists of two parts: an event component (E), that refers to an action of the agent (these components essentially contain the same information as the skeletal script shown in figure 1) and a background component (B), that refers to important additional characteristics of the situation. Each component consists of a tuple or a conjunction of tuples. The notation of the tuples corresponds to the notation of the event models and the GSD in NAOS. T1 to T6 are time variables. They always occur in pairs denoting interval boundaries. RED and GREEN are constants.

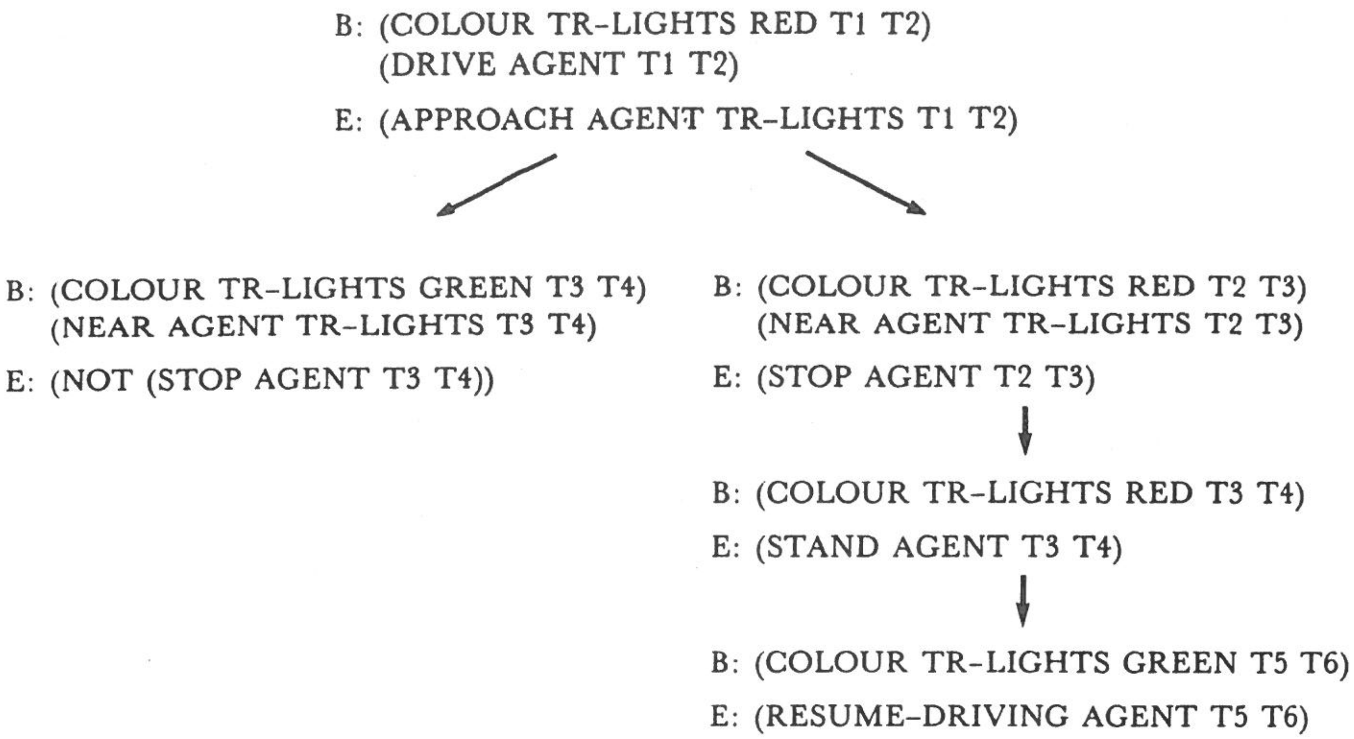
B: (COLOUR TR-LIGHTS RED T1 T2)
(DRIVE AGENT T1 T2)
E: (APPROACH AGENT TR-LIGHTS T1 T2)
B: (COLOUR TR-LIGHTS GREEN T3 T4)
(NEAR AGENT TR-LIGHTS T3 T4)
E: (NOT (STOP AGENT T3 T4))
B: (COLOUR TR-LIGHTS RED T2 T3)
(NEAR AGENT TR-LIGHTS T2 T3)
E: (STOP AGENT T2 T3)
B: (COLOUR TR-LIGHTS RED T3 T4)
E: (STAND AGENT T3 T4)
B: (COLOUR TR-LIGHTS GREEN T5 T6)
E: (RESUME-DRIVING AGENT T5 T6)

Figure 2: TRAFFIC-LIGHTS script used in NAOS

4. Expectations

In the process described below scripts are used for the generation of expectations, and the geometrical scene description (GSD) is used for the evaluation of these expectations, i.e. for determining whether an expectation is satisfied or not.

The process of generation and evaluation of expectations inspects the content of the GSD at consecutive instances of time. At each instance of time the following actions can be performed: New scripts can be activated, new expectations can be generated using active scripts, and existing expectations can be evaluated.

In order for a script to be activated, its header (i.e. its root) must be instantiated. In the case of our example (figure 2) the header is:

B: (COLOUR TR-LIGHTS RED T1 T2)
(DRIVE AGENT T1 T2)
E: (APPROACH AGENT TR-LIGHTS T1 T2)

The process thus matches the header against the GSD and tries to instantiate all variables of the header. Whenever a header can be completely instantiated, the corresponding script is activated. Several scripts can be active at the same time.

Expectations are generated in the following way: For an instance of time the current position of each active script can be determined by instantiating the tree (or rather the appropriate branch of the tree) down to the maximally possible depth at that instance of time. The deepest node that can be instantiated is the current position. The successors of the current position (if it has successors) are then taken as the current expectations. If the current position of a script has changed, compared to the last instance of time, the process of the generation of expectations will yield new current expectations as well.

5. Evaluation and Verbalization

If both the B and E components of an expectation can be instantiated, the expected event (i.e. the E component of the expectation) together with the label SATISFIED and the evaluation time is stored in the database of expectations (an associative database similar to the one of the GSD). This entry in the expectation database can then be used to generate a past tense statement comprising the expected event. This kind of statement, however, can be generated in NAOS without the aid of expectations as well.

If the B component can be instantiated and the E component is a non-durative, composite event and can be partly (i.e. at the beginning) instantiated, the expected event together with the label BEGUN and the evaluation time is stored in the expectation database. This entry can then be used to generate a present tense statement comprising the expected event. In the case of durative events present tense statements can be generated without the aid of expectations as well because, if a durative event is happening during some time interval, the same applies to all subintervals thereof.

Future tense statements can be generated using the current expectations. However, it seems reasonable to impose the additional condition on this kind of verbalization, that only those expectations that are the only successors of a current node should be verbalized. If there were alternative expectations, the certainty of each one would be too low and it would be difficult to decide which one to verbalize. In the case of only one expectation the expected event together with the label EXPECTED and the generation time is stored in the expectation database and can be used to generate a future tense statement. The same holds if the process of the evaluation of expectations yields, that the current position of a script hasn't changed compared to the last instance of time. In this case, if there was an entry in the expectation database comprising the expected event, the entry is updated, i.e. the generation time or the last evaluation time is replaced by the new evaluation time. The new entry can then be used to generate a future tense statement.

If the B component can be instantiated, but the E component can't, the expected event together with the label CONTRADICTED and the evaluation time is stored in the expectation database. This entry can then be used to generate a negative past tense statement.

The processes descibed above are implemented in UCI-LISP and FUZZY on a DEC-10 system.

6. Conclusions

In this paper we have introduced scripts as a model for the representation of knowledge about stereotypic sequences of events and as a means for generating expectations in the domain of traffic scenes. Making use of the generation and evaluation of expectations we can extend the set of possible statements about traffic scenes in the system NAOS.

Of course we are limited to those statements that can be derived from scripts. Without a script we can't generate any expectations. For instance, if we only have the script shown in figure 2, we can't generate a statement like "Die Ampel wurde nicht grün." ("The traffic lights didn't turn green."). This would only be possible if we had a script in which the traffic lights are the agent and in which the knowledge is represented that traffic lights turn red and green every now and then. This points out another limitation which lies in the tree structure of the scripts used in this approach. In certain cases, like for instance the behaviour of traffic lights, a net structure that permits cycles seems to be more adequate to represent the knowledge.

Another problem lies in the exactness of the scripts. In order to cope with the problems that are presented by real-life traffic scenes a lot more details have to be represented in the nodes. For instance the TRAFFIC-LIGHTS script would have to represent the knowledge that cars usually stop directly in front of red traffic lights only if there is no other car in front of them and that otherwise they usually stop directly behind the last car of the queue in front of the traffic lights.

A more severe limitation is the fact that, using scipts, we can only deal with frequently occurring sequences of events, whereas human speakers have ways of expecting events in more unusual situations as well. One way to overcome this limitation would be to incorporate models of the kind of plans, goals, and themes (cf. [5]) and to make use of reasoning mechanisms about intentions of agents.

Finally, there are still types of statements that can't be generated by NAOS. One such kind are causal statements. In order to be able to generate sentences containing words like "weil" ("because"), "da" ("since"), "denn" ("for"), "deshalb" ("therefore") etc., one would need to represent knowledge about causal (and not only temporal and spatial) relations between events. If this had to be incorporated into NAOS, it would also rise the question, how deep and how exact the knowledge should be, i.e. whether only qualitative, common-sense knowledge or quantitative, scientific knowledge too (e.g. physical laws, cf. [9]) should be represented.

Acknowledgements

I would like to thank Bernd Neumann and Hans-Joachim Novak for their valuable comments and suggestions.

References

[1] B. Neumann: Towards Natural Language Description of Real-World Image Sequences. GI 12. Jahrestagung, Informatik-Fachberrichte 57, Springer, 1982, 349-358

[2] H.-J. Novak: On Verbalizing Real-World Events: An Interface of Natural Language and Vision. In: B. Neumann (ed.): GWAI-83, 7th German Workshop on Artificial Intelligence, Dassel/Solling, September 1983, Springer, 1983

[3] B. Neumann and H.-J. Novak: Event Models for Recognition and Natural Language Description of Events in Real-World Image Sequences. IJCAI-83, 1983, 724-726

[4] B. Neumann: Natural Language Description of Time-Varying Scenes. Report 105, Fachbereich Informatik, Universität Hamburg, August 1984

[5] R.C. Schank and R.P. Abelson: Scripts, Plans, Goals and Understanding. Hillsdale/New Jersey: Erlbaum, 1977

[6] R.C. Schank and C.R. Riesbeck (eds.): Inside Computer Understanding: Five Programs Plus Miniatures. Hillsdale/New Jersey: Erlbaum, 1981

[7] W.G. Lehnert: Text Processing Effects and Recall Memory. Research Report 157, Yale University, Department of Computer Science, 1979

[8] R.E. Cullingford: Script Application: Computer Understanding of Newspaper Stories. Research Report 116, Yale University, Department of Computer Science, 1978

[9] K.D. Forbus: Spatial and Qualitative Aspects of Reasoning about Motion. AAAI-80, 1980, 170-173