In: IJCAI-83, pp. 643-646.

OVER-ANSWERING YES-NO QUESTIONS:

Extended Responses in a NL Interface to a Vision System

Wolfgang Wahlster[1], Heinz Marburger[1], Anthony Jameson[2], Stephan Busemann[1]

[1] Research Unit for Information Science
and Artificial Intelligence,
University of Hamburg,
Mittelweg 179,
D-2000 Hamburg 13,
F.R. Germany

[2] Psychologisch Laboratorium,
Faculteit der Sociale Wetenschappen,
Katholieke Universiteit,
Montessorilaan 3,
NL-6500 HE Nijmegen,
The Netherlands

ABSTRACT

This paper addresses the problem of over-answering yes-no questions, i.e. of generating extended responses that provide additional information to yes-no questions that pragmatically must be interpreted as wh-questions. Although the general notion of extended responses has already been explored, our paper reports on the first attempt to build a NL system able to elaborate on a response as a result of anticipating obvious follow-up questions, in particular by providing additional case role fillers, by using more specific quantifiers and by generating partial answers to both parts of questions containing coordinating conjunctions. As a further innovation, the system explicitly deals with the informativeness-simplicity tradeoff when generating extended responses. We describe both an efficient implementation of the proposed methods, which use message passing as realized by the FLAVOR mechanism and the extensive linguistic knowledge incorporated in the verbalization component. The structure of the implemented NL generation component is illustrated using a detailed example of the system's performance as an interface to an image understanding system.

1. INTRODUCTION

Typical of a human dialog partner is the ability to recognize when it is appropriate to provide more than a mere literal, direct answer to a question and to decide what that extended response should be. Hence natural language (NL) dialog systems that are only capable of simple direct responses to users' questions will necessarily be regarded as uncooperative. For this reason much of the current work on cooperative NL dialog systems includes efforts to go beyond question answering [2] by generating extended responses.

Several types of extended responses have been investigated:

(E1) pointing out incorrect presuppositions [7],
(E2) generating unsolicited justifications [14],
(E3) pointing out discontinuities in the domain of discourse which suggest that a small change in the original question would result in a much more helpful response [13],
(E4) offering to ‘monitor’ for information requested by the user as the system learns of it [10].

In the present paper we discuss some of the semantic and pragmatic issues related to (E5) - another important class of extended responses - and describe the evaluation and generation modules of our dialog system HAM-ANS[1] [4], which are the first such components that realize this type of response in an AI system:

---

[1] Research on HAM-ANS is currently being supported by the German Ministry of Research and Technology (BMFT) under contract 08IT1503 8.

(E5) elaborating on a response to a yes-no question in order to anticipate obvious follow-up questions of the user; in particular

(a) filling optional deep case slots in the case frame associated with a verb used in the request (cf. Figs. 1 and 2),
(b) using more specific quantifiers in the answer than in the question,
(c) generating explicit partial answers to both parts of questions containing coordinating conjunctions.

In contrast to previous AI work on extended responses of type (E1) - (E4), we will describe the results of our research not in the computational context of NL DB query systems, but rather in the context of a NL interface to a vision system for understanding image sequences [11].

In one of the applications of our German language dialog system HAM-ANS, we presuppose the following situational context: the system is observing a street intersection and supplies on the phone the user, who is familiar with the scene but cannot see it from his remote location, with information about the traffic at that intersection (cf. Fig. 1). Although HAM-ANS is a full-fledged, operational dialog system with a powerful parsing component, we will restrict our presentation here to those modules of HAM-ANS's evaluation and generation components which are responsible for over-answering yes-no questions and necessary for verbalizing the resulting extended responses.

2. OVER-ANSWERING YES-NO QUESTIONS

Extended responses to yes-no questions are answers that contain more than a plain 'yes' or 'no'. We call the process of generating extended responses that provide more specific or additional information *'over-answering'*. To make our discussion more concrete, let us consider the following interaction with HAM-ANS (cf. also Fig. 1 for an additional example):

(1) User: *Has a yellow car gone by?*
(2) HAM-ANS: *Yes, one yellow one on Hartungstreet.*

2.1. Providing Additional Case Role Fillers

By answering as in (2), the system has in effect interpreted (1) as a sort of indirect wh-question (cf. [8]) which could be paraphrased:

(3) *Has a yellow car gone by? If so, where?*

For (1), the fact that an extended response is generated is closely related to the cognitive process underlying the determination of the truth value of the proposition in question, namely the verification of the presence of a certain type of locomotion in the analyzed image sequence (cf. also Fig. 1). Note that it is impossible to verify that 'a yellow car has gone by' without determining the spatial location of this event. This means that the LOCATIVE slot in the case frame for 'to go by' is filled as a *side effect* of the search for an answer.

HAM-ANS's policy of restricting extended responses to information which would have had to be obtained even for a minimal direct answer is a simple, general strategy for keeping the cost of over-answering within reasonable limits, in terms of both computation and the complexity of the resulting utterances. Like the other heuristics to be presented below, it may be complemented with more expensive and domain-specific approaches which take the user's assumed interests and expectations into account in the selection of unsolicitated information for verbalization (see e.g., [5]).

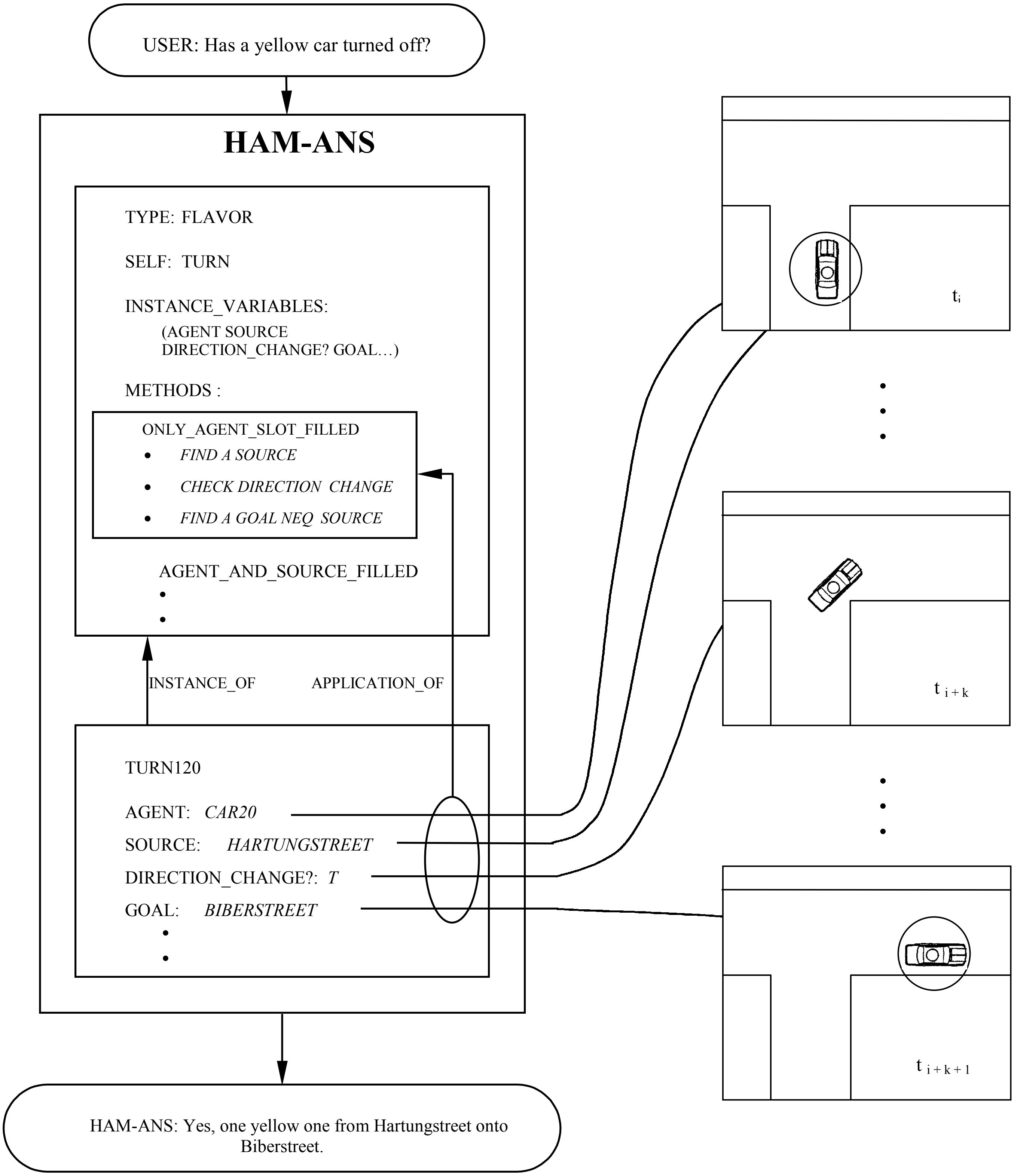

Fig.1: Case slot filling as side effect of visual search

Let us now consider the example given in Fig.1 in more detail. Motion concepts are represented in a specialization hierarchy using the object-oriented FLAVOR formalism [3, 15]. The root flavor is the motion concept 'to move'. Descendants in the tree, e.g. 'to go by', 'to turn', inherit the declarative and procedural information, instance variables and methods, respectively, contained in their parents. The instances of a flavor, e.g. TURN120, denote a specific event in the domain of discourse.

The task of checking the truth value of the proposition in the question is accomplished through message passing. These messages include: creating instances of motion concepts, instantiating deep case slots specified in questions, and activating appropriate methods, e.g. ONLY-AGENT_SLOT_FILLED. These methods then activate further methods which are either owned by the flavor, (e.g. 'CHECK DIRECTION CHANGE') or inherited from parents (e.g. 'FIND A SOURCE', -'FIND A GOAL NEQ SOURCE'). In Fig.1 the latter two methods must check, among other things, whether or not the object that the agent is located on satisfies the selection restrictions of the deep case or not. This check produces an object name, which is used to fill the corresponding deep case slot. These slots are then used in the formation of the extended response.

In order to verbalize extended responses as informative but concise utterances like (2), HAM-ANS must elide some parts of the original question. This task is performed by the ellipsis generation component (cf. Fig. 2), which eliminates those parts of the semantic representation of complete extended answers that are identical to the stored representation of the question [6].

2.2 The Use of More Specific Quantifiers

As a final remark on the interaction (1)-(2), let us point out that the indefinite article 'a' in (1) was replaced in (2) by the more precise quantifier word 'one'. The over-answering rule responsible for this result is illustrated more clearly in the interaction (4)-(5), in which the vague 'several' is replaced by 'two':

(4) User: *Have several vehicles stopped on Biberstreet?*
(5) HAM-ANS: *Yes, two.*

When an existential or numerical quantifier appears in a question, the corresponding quantifier word in the answer denotes a specific number. This use of precise quantifier words in answers will usually be pragmatically appropriate, since it increases the *informativeness of answers* without substantially increasing their complexity.

In general, of course, increased informativeness can often be achieved only at the price of greater complexity. The complexity introduced into the answer to (6) by the mention of the two streets involved is not so great as to be distracting to

(6) User: *Has a station wagon driven by?*
(7) HAM-ANS: *Yes, three on Biberstreet and two on Hartungstreet.*

the user, but if the five station wagons happened to have driven on five different streets, a response which mentioned each of them would be irritatingly long and tedious:

(8) System: *Yes, one on Hartungstreet, one on Biberstreet, one on ...*

2.3 The Informativeness-Simplicity Tradeoff

In the absence of specific knowledge concerning the nature and amount of information desired by the user, some general heuristics are required to guide the system's decision when informativeness-simplicity tradeoffs arise. A variety of different sets of rules for this purpose are conceivable; the following set of heuristics, which are possible instantiations of the *Gricean Maxims of Quantity*, has proved useful within HAM-ANS:

(H1) Avoid superfluous complexity, i.e. complexity not justified by any increase in informativeness.

(H2) Do not allow a certain maximum degree of complexity to be exceeded.
(H3) Within these limits, maximize the amount of information presented.

As an example, we sketch here the criteria used in cases like (1), (4) and (6), i.e. when there is some nonempty set of relevant objects which satisfy the predicate specified by the question's verb phrase. In general, the evaluation component will have accumulated more specific information concerning each of these objects than that specified in the original predicate. To determine whether this additional detail should be mentioned in the answer, the system partitions the objects into *equivalence classes* according to whether the corresponding more specific assertions are identical. If only one or two equivalence classes result (as in examples (1) and (6), respectively), the additional information is mentioned in the answer. If there are more than two equivalence classes, they are not mentioned separately, and the original predicate of the question is applied in the answer to all of the objects collectively.

The principle (H1) is reflected here in the formation of equivalence classes, since this sort of grouping can lead to a reduction in the complexity of the answer without any loss of informativeness. The maximum degree of complexity allowed (cf. rule H2) corresponds here to a two-part coordination such as the answer to (6) . Finally, the policy of rule H3 is reflected here in the reporting of the additional detail.

The specific criteria just sketched are, of course, very narrow in scope. In HAM-ANS analogous criteria are applied, e.g., when the question contains a coordinating conjunction (as in (9) below, cf. E5 (c) above) and when it has nested quantifiers.

(9) User: *Have a Station wagon and a truck driven away along Biberstreet?*
(10) HAM-ANS: *No, one station wagon, but no trucks.*

3. THE GENERATION OF EXTENDED RESPONSES

In this section the capabilities of our implemented system are briefly illustrated by some comments on a trace of the process of generation of the answer (2). The trace, shown in Fig. 2, starts after a successful parse which produces a SURF representation of the input question. SURF is a linguistically motivated declarative representation language which is used as a target and a source language by the parser and the NL generator, respectively. The SURF representation is transformed into a DEEP representation, which is better suited for the evaluation processes, which draw inferences on the basis of the user's utterances and derive answers to questions. The main task of this transformation process is to

```
HAS A YELLOW CAR GONE BY?

;;SURF representation of input sentence

  (af-d:   EVENT
   (t-s:   (q-d: THE (r: 1 1))   (lambda:  x0  (af-a:  ACT x0   GO_BY)))
   (d-e:   role-list:
   (rl-s : agent:
        (lambda:  x0
          (af-a:  AGENT  x0
         (t-s:  (q-qt :  A)
             (d-o : AND     (lambda:    x0(af-a:  ISA  x0   CAR))
                            (lambda:  x0   (af-a:  REF x0    YELLOW)))))))

 mod:
 (d-m:  tense:  (lambda:  x0(af-a:   TENSEx0PERF))
        voice:  (lambda:    x0   (af-a:   VOICE x0ACT)))))

**  Normalisation: transformation into DEEP representation

;;DEEP representation of input sentence

(f-d:  (t-q: (for:   (q-d:   THE    (r: 1    1))   x57)  (af-a:  ACT  x57  GO_BY))
```

```
 (f-d :  (t-q:  (for:  (q-qt:   A)     x58)
           (f-o:    AND  (af-a  ISA    x58   CAR) (af-a:  REF x58   YELLOW)))
(f-e:   role-list:
 (rl-d:
  agent:  (af-a:    AGENT    x57   x58))
mod:
 (f-m: tense:       (af-a:    TENSE     x57    PERF)    voice:   (af-a:    VOICE x57    ACT)))))
```

** Evaluation

;; Generation of reference objects fitting the agent description

```
            (CAR20)
```

;; Testing of a partially instantiated case frame

```
(f-e:   role-list:
 (rl-d:
  agent:   (af-a:   AGENT    GO_BY    CAR20))
mod:
 (f-m:   tense:  (af-a:    TENSE GO_BY    PERF)  voice:   (af-a:    VOICE GO_BY    ACT)))
```

;;Application of flavor method ONLY_AGENT_SLOT_FILLED

;;DEEP representation of answer

```
(f-d:   (t-q:     (for:     (q-d:     THE     (r: 1 1))     x57) (af-a:     ACT    x57   GO_BY))
 (f-d:   (t-q:     (for:     (q-qt:    (r: 1   1))   x58)
        (f-o:     AND     (af-a:    ISA   x58   CAR)   (af-a:    REF     x58    YELLOW)))
   (f-e:    role-list:
    (rl-d:
     agent:   (af-a:    AGENT    x57   x58)
     locative:(af-a:    LOCATIVE    x 57  *ON HARTUNGSTREET))
mod:
 (f-m:  tense:  (af-a:    TENSE x57  PERF)   voice:  (af-a:    VOICE  x57  ACT)))))
```

** Inverse normalisation: transforming into SURF representation

** Ellipsis generation

;;Elliptified SURF representation of answer

```
(rl-s:   agent:
   (lambda:  x0
      (af-a: AGENT x0
      (t-s: (q-qt:  (r:1 1))
             (d-o: AND
                (lambda: x0 (af-a: ISA x0 CAR))
                 (lambda: x0 (af-a: REF x0 YELLOW))))))
locative:
(lambda: x0   (af-a:    LOCATIVE   x0     *ON HARTUNGSTREET)))
```

;; Verbalised structure of answer

```
          (SENTENCE (AGENT   (NP  (N:    SG)      ONE     YELLOW   (ELLIPSIS    CAR)))
                              (LOCATIVE  (PP  *ON     (NP   (N:    SG)      HARTUNGSTREET))))
```

**Surface transformations

YES, ONE YELLOW ONE ON HARTUNGSTREET.

Fig. 2: Example trace (translated from German) of the generation of an extended response

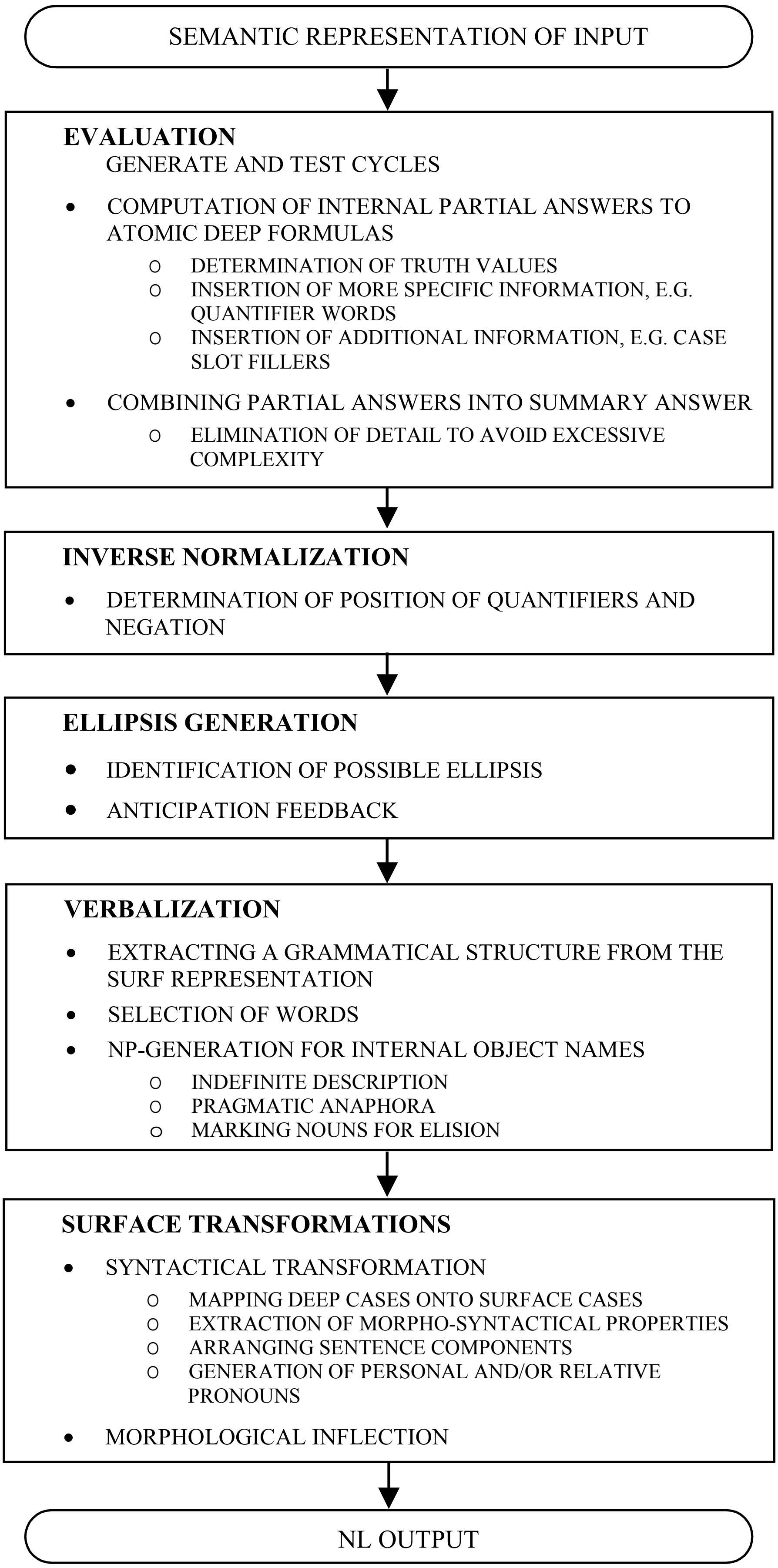

Fig. 3: The structure of HAM-ANS's evaluation and generation components

determine the scope of quantifiers (cf. Figs. 2 and 3). The pattern-directed evaluation component of HAM-ANS includes all of the processes for deriving the semantic structure of an extended response, as discussed in the previous section. After the DEEP representation of the resulting answer has been subjected to the inverse transformation the ellipsis generation component tries to reduce the structure which represents the complete response. The verbalization process uses translations rules attached to the various categories of SURF expressions; it is thus guided directly by the message to be expressed rather than by the hierarchical structure of a grammar. The possibly elliptical preterminal structure of the answer is mapped onto a well-formed German utterance by the surface transformation component, which uses extensive linguistic knowledge (cf. Figs. 2 and 3). In this component, the terminal word forms are generated using morpho-syntactic properties (e.g. for number (marked in Fig. 2 with N:), surface case, and gender) and morphological classification schemes for the different inflectional paradigms of German nouns and verbs. The properties extracted from the preterminal structure and the word lexicon are assigned to noun phrases according to rules of agreement [1].

The structures following the deep cases in Fig. 2 are treated as sentence components, e.g. AGENT becomes the subject of the sentence. The noun marked for elision can not be elided until its gender has been extracted. The preterminal preposition *ON causes the generation of ‘on’ within a LOCATIVE phrase, but ‘onto’ within a GOAL phrase. The analogous problem in German is to provide the inflection processes with the appropriate surface case, i.e. dative and accusative, respectively.

4. LIMITATIONS AND FUTURE WORK

Our current research efforts are not being concentrated on interactions like (11)-(12) and (13)-(14), where in the system's reply the motion concept used in the questions is modified by an adverb or replaced with a more specific one. The

(11) User: *Has the car stopped?*
(12) System: *Yes, very abruptly.*

(13) User: *Has the car gone by?*
(14) System: *Yes, it sped away.*

generation of answers of this type would require a more elaborate evaluation strategy and the use of more knowledge sources, e.g. knowledge about common sense standards [11, 12].

On the other hand the methods described can easily be extended to handle *temporal adverbs*. In interaction (15)-(16), the extended response indicates when the event took place. Because the time

(15) User: *Did the bus stop?*
(16) System: *Yes, it just did.*

of the beginning and the end of an event is recorded during the evaluation phase, we plan to provide for the generation of answers like (16) using methods similar to those described in [9] for the analysis of temporal descriptions.

REFERENCES

(1) BUSEMANN, S.: Problems involving the automatic generation of utterances in German. Memo ANS-8, Research Unit for Information Science and AI, Univ. of Hamburg, April 1982.

(2) COHEN, P. R., PERRAULT, C. R.. ALLEN, J. F.: Beyond question answering. In: Lehnert, W. G., Ringle, M. H. (eds.): Strategies for natural language processing. Hillsdale, N.J.: Erlbaum, 1982, pp. 245-274.

(3) DI PRIMIO, F., CHRISTALLER, T.: A poor man's flavor system. Technical report, ISSCO, Geneva, 1983.

(4) HOEPPNER, W., CHRISTALLER, TH., MARBURGER, H., MORIK, K., NEBEL, B., O'LEARY, M., WAHLSTER, W.: Beyond domain independence: Experience with the development of a German language access system to highly diverse background systems. In: Proc. 8th IJCAI, Karlsruhe, 1983.

(5) JAMESON, A.: Impression monitoring in evaluation-oriented dialog: The role of the listener's assumed expectations and values in the generation of informative statements. In: Proc. 8th IJCAI, Karlsruhe 1983.

(6) JAMESON, A., WAHLSTER, W.: User modelling in anaphora generation: Ellipsis and definite description. In: Proc. ECAI-82, Orsay 1982, pp. 222-227.

(7) KAPLAN, S. J.: Cooperative responses from a portable natural language data base query system. Ph.D. Thesis, Computer Science Department, Univ. of Pennsylvania, 1979.

(8) KIEFER. F.: Yes-no questions as wh-questions. In: Searle, D. R., Kiefer, F., Bierwisch, M. (eds.): Speech act theory and pragmatics. Dordrecht: Reidel, 1980, pp. 97-119.

(9) MARBURGER, H., NEUMANN, B., NOVAK, H.-J.: Natural language dialogue about moving objects in an automatically analyzed traffic scene. In: Proc. 7th IJCAI, Vancouver 1981, pp. 49-51.

(10) MAYS, E., JOSHI, A. K., WEBBER, B. L.: Taking the initiative in natural language data base interactions: Monitoring as response. In: Proc. ECAI-82, Orsay 1982, pp. 255-256.

(11) NEUMANN, B.: Towards natural language description of real-world image sequences. In: Nehmer, J. (ed.): GI-12. Jahrestagung. Berlin: Springer, 1982, pp. 349-358.

(12) NOVAK, H.-J.: On the selection of verbs for natural language description of traffic scenes. In: Wahlster, W. (ed.): Proc. 6th German workshop on AI, Berlin: Springer 1982, pp. 22-31.

(13) SIKLOSSY, L.: Impertinent question-answering systems: Justification and theory. In: Proc. ACM National Conference, 1978, pp. 39-44.

(14) WEBBER, B., JOSHI, A.: Taking the initiative in natural language data base interactions: Justifying why. In: Proc. 9th COLING, Prague 1982, pp. 413-418.

(15) WEINREB, D.. MOON, D.: Lisp Machine Manual (4th ed.). MIT, 1981.