A COOPERATIVE YES-NO QUERY SYSTEM FEATURING DISCOURSE PARTICLES

Kjell Johan SÆBØ

Seminar für natürlich-sprachliche Systeme
*Tübingen**

and
Norges allmennvitenskapelige forskningsråd

Abstract

Cooperative dialog systems will offer extended answers to questions, that is, they will volunteer information not explicitly asked for. A complete response will be complex and the member sentences will evince an extensive parallel, the indirect answer substituting an alternative for a focus in the question. Research on discourse particles has shown that they are necessary to ensure coherence between adjacent sentences evincing an extensive parallel, that is, that they reflect discourse relations as given in complex answers, so that such answers emerge as core contexts. Thus the proper mode of representation for discourse particles in a system coincides with the framework of cooperative question-answering. The PASSAT system centers on the rôle of particles in characterizing and reflecting relations such as underlie complex response.

Discourse Particle Semantics

Discourse particles are meta expressions in a natural language: They express discourse relations, which are necessary properties of (complex) discourses, that is, they refer to things in the language and not to things in the world. It follows that they do not influence truth conditions. Instead, they affect coherence: When occurring in a complex discourse such that the discourse relation "contradicts" its meaning, a discourse particle may cause incoherence.

?Edison invented the telegraph.
Marconi also invented the phonograph.

Conversely, discourse particles can cause coherence too: When one does not occur in a complex discourse such that the discourse relation "entails" its meaning, an incoherence comes about which the occurrence would prevent. What causes that incoherence is the occurrence of the *empty*, or *zero*, particle.
So nonempty particles are sometimes necessary to reflect discourse relations, in other words, they substitute positive for negative presuppositions.

?Edison invented the telegraph.
He invented the phonograph.

And, those presuppositions do not refer to the world but to the environment of discourse. On a classical truth presupposition, enriched by a sensitivity to context, the empty particle in the example presupposes that Edison did not invent the telegraph. But then, he in fact did not, so the incoherence is not explained. Similarly, on classical terms the particle in the former example presupposes that Marconi invented the telegraph, but then, he in fact did.

So discourse particles, empty or not, react not to what is or is not the case but to what is or is not supposed to be the case: To the context. The proper context category for the discourse particle category is a complex sentence: A sentence pair. Any two sentences in sequence occasion the empty or some nonempty discourse particle in front of or within the second sentence. Thus a language in a model theory of discourse particles will consist in a pair:

$$L = < SP , DP >$$

SP is a set of sentence pairs sp, the individual constants, and DP is a set of discourse particles dp including the empty particle *dp0*, the predicates; and for any dp and sp, dp(sp), the application of dp on sp, is to represent the occurrence of dp in sp. An interpretation rule is to state a necessary condition for the coherence of any dp(sp) in terms of the meaning of sp and the meaning of dp. Thus the model in the theory will consist in a triple:

$$M = < DR , \mathcal{f} , Þ >$$

DR is a set of discourse relations, $\mathcal{f}$ is an assignment mapping constants, i.e. sentence pairs, onto members of DR, and Þ is an interpretation mapping predicates, i.e. discourse particles, onto subsets of DR. The denotation of any element of DP dp is defined as the set of discourse relations Þ(dp). The interpretation rule states that for any sp, dp, dp(sp) is only coherent if $\mathcal{f}$(sp) is not an element of Þ(dp') for any dp' different from dp.

It is assumed that the denotations of the discourse particles - the elements of the picture of DP under Þ - are all disjunct, i.e. that for any dp1, dp2 in DP, the intersection of Þ(dp1) and Þ(dp2) is empty. Thus for any sp and dp such that $\mathcal{f}$(sp) is in Þ(dp), dp'(sp) is incoherent for any dp' different from dp. This means that a sentence pair instantiating a discourse relation that belongs to the denotation of a particle is only coherent if occasioning the particle, in other words, that a particle is necessary with respect to every discourse relation in its denotation. As DP contains the empty particle *dp0*, this principle corresponds to a negative presupposition.

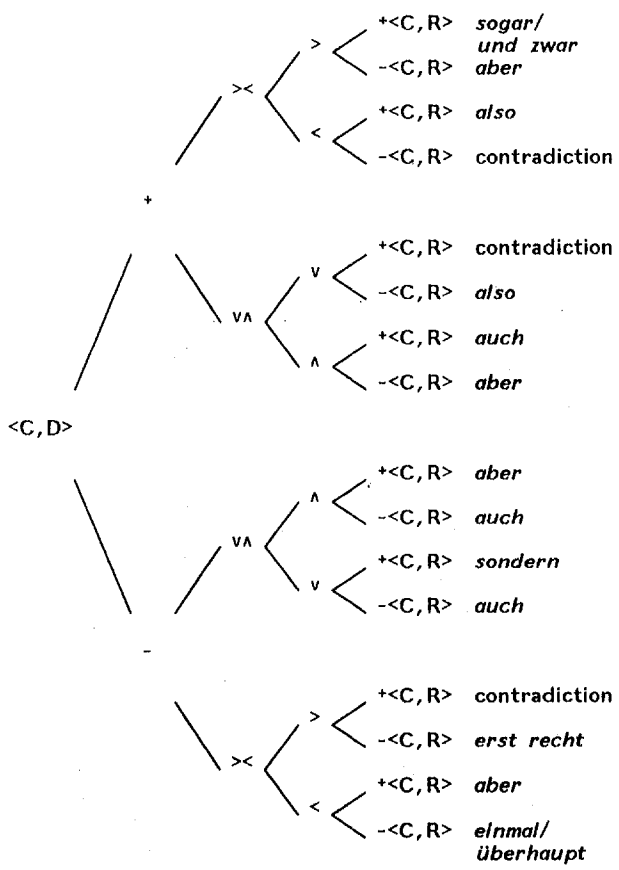
So there are discourses necessitating this particle or other in virtue of the relations they instantiate. A discourse relation is an abstraction on the way the two members of the sentence pair compare to each other. Any two sentences can be segmented into a *copart*, a *depart*, and a *repart*, meaning the portion common to both sentences, the distinct portion of the first sentence, and the distinct portion of the second sentence. The common denominator of discourse relations in the denotation of any nonempty particle is that the de- and the repart are minimal: That the two sentences differ in only one description. Discourses necessitating nonempty particles are characterized by an extensive parallel. Importantly, polarity

* *The paper is based on research done in the project LILOG, financed and supervised by IBM Germany.*

change does not count as a de- or a repart but serves as an extra parameter to differentiate specific discourse relation classes.

Marconi invented the telegraph,
?(<u>but</u>) he didn't invent the phonograph.

Discourse relations encode information on two other parameters: Whether the de- and the repart are in a semantic relation (scalarity or hyponymy) such that one sentence entails the other (negation exempt), and, if so, whether entailment is left-to-right or reversely, and if not, whether they are in a semantic relation (exclusion) such that one sentence entails the negation of the other, or the two sentences are compatible. These four features - polarity distribution accounts for two - give rise to sixteen cases, thirteen providing slots for German discourse particles:



C = copart, D = depart, R = repart;

+/- = positive/negative polarity;

>< = depart and repart are ordered:
> = R>D (+<C,R> entails +<C,D>),
< = R<D (+<C,D> entails +<C,R>);

∨∧ = depart and repart are not ordered:
∨ = D and R exclude each other (+<C,D> entails -<C,R>),
∧ = D and R do not exclude each other (+<C,D> and +<C,R> are compatible).

Complex Response: The Basic DP Situation

It is desirable to equip a computer system to execute the semantic theory sketched above. There are several possible approaches to this, but one embodies decisive advantages. That is a particular *generative* approach.

It is a cornerstone of the theory that discourse particles not only introduce requirements on contexts but also have a communicative necessity; that contexts require *them* in that the nonoccurrence can be as damaging to coherence as can the occurrence. In representing the model in an automatic process it is especially desirable to capture this aspect.

One way to go is to have a program test surface-language inputs and give notices of error whenever incoherence occurs. On this course, every piece of information is produced by the user, who must be acquainted with the theory in order to obtain an interesting reaction. Or one can have a program generate particles in accordance with the syntax and semantics of the contexts it generates. Again, there are more than one way to go: The input can be of a general nature, e.g. a description of a discourse relation, or it can be specific, e.g. a pair of discourse representation structures. Either way, the user has to specify the context unit to be generated, a complex discourse with a parallel, as long as she addresses it directly at some level.

However, once such contexts are embedded in a communicative setting to motivate them independently, there is no need to prescribe anything. There is another, *indirect* generative approach which promises spontaneous and systematic generation of proper contexts, and an interesting application: Dialog systems capable of cooperative question answering.

This is an independently motivated field of research in Artificial Intelligence and in Computational Linguistics, seeking to simulate that crucial feature of human dialog behavior that answers are far from always formed in strict accordance with the semantic structure of the question. Frequently in actual conversation, answerers are expected to elaborate, in particular on a *yes* or a *no*. A realistic yes-no query system will be prepared to offer additional information in the form of extended answers, and several systems in this spirit have recently been devised.

As it happens, complex responses are key contexts for discourse particles. A complete response consists in a sentence sequence. This means that the proper context category for particles is given a priori in this framework. Moreover, the context unit as such, a complex discourse with an extensive parallel, is given as well because a simple response is inadequate just in case corresponding information on a *relevant alternative* to a focus in the query is available to the responder. As far as wh- questions are concerned, the focus coincides with the wh- position. As regards yes-no questions, it may be any item suitable as a depart in a particle scheme. It is adequate, then, to supplement the simple response by the corresponding information. The sentence frame of the question carries over from the direct to the indirect answer as the focus (and possibly polarity too) is exchanged. (Occasionally, the focus (depart) is empty so that the alternative (repart) adds a piece of information, typically an adjunct.) Thus complex response creates discourse relations such as necessitate discourse particles in a systematic way and on independent grounds. It may be considered the basic situation of utterance for discourse particles.

There is consensus that an extension to a response is appropriate if and only if the information it conveys is relevant. The challenge consists in defining what constitutes relevance in each single case. It is a prime desideratum to develop general guidelines for selecting alternatives. Relevant means relevant to the goal of the dialog, and any sensible approch will take Grice's Maxim of Quantity as a point of departure: Make your contribution as informative as is required for the current purpose of the exchange. There are various ways of exploiting this principle. One is to relate information to the assumed practical purpose of the query so as to suggest surrogate courses of action in case the direct answer is negative. Such a strategy is pursued by KAPLAN (1983). Alternatives will be identified on the basis of a functional synonymy:

- *Do you have a match?*
- *No, but I have a lighter.*

Another approach, adopted by JOSHI et al. (1984), is to focus on world knowledge so as to correct false default inferences licensed by the direct answer by stating exceptions to normal courses of events. Alternatives will be identified by way of stereotypes:

- *Is Tweety a bird?*
- *Yes, but he cannot fly.*

In the theory of *scalar conversational implicature*, applied to question-answering by HIRSCHBERG (1985), the Maxim of Quantity is revised to refer to the *strength* of an utterance: Make the strongest relevant claim justifiable by evidence. The concept of relevance remains, but it is anchored to linguistic knowledge by a semantic relation: Strength as surfacing in scalarity. Such a strategy embodies two clear advantages: A stronger version of a question, whether positive or negative, cannot rationally be known to the questioner in advance; and, the search for a stronger version can be guided by rules which must be represented in a reasonably intelligent system anyhow, namely, lexical relations and meaning postulates.

Semantic scales are defined by tuples of lexical items linearly ordered by entailment. Consider as an example the pair *possible* and *probable* and a query *Is it possible* or *Is it probable* for some proposition *It*, and assume the adjective to be the focus. The answer *no* to the former question will answer the other one too, as will the answer *yes* to the latter question. The answer *yes* to the former question will not, nor will the answer *no* to the latter, yet a responder is required to make the strongest relevant claim, and provided the other item counts as relevant, there is a straightforward way to do so:

- *Yes, it is even probable.*
- *No, it is not even possible.*

In fact, if the maxim is revised to require the responder to *assess* the strongest relevant proposition, two more responses emerge as adequate, again on the condition that the other item counts as relevant:

- *Yes, but it is not probable.*
- *No, but it is possible.*

In a wider sense, semantic scales are defined by tuples of lexical items arranged by entailment in a hierarchy of set inclusion and exclusion. Consider as an example the quadruple *Scandinavian*, *Danish*, *Norwegian*, and *Swedish*, and queries *Is it Scandinavian* etc. for some referent *It*, and assume the adjective to be the focus. The answer *no* to the first question will answer all the other questions too, as will the answer *yes* to any subsequent question. The converse is not the case, yet a responder is required to make the strongest relevant claim, and provided the other items count as relevant, there is a straightforward way:

- *Yes, (and in fact) Danish/Norwegian/Swedish.*
- *No, (but) Norwegian/Swedish// Danish/Swedish//Danish/Norwegian.*
- *No, not Scandinavian at all.*

The proviso was made that the other items count as relevant, as the responses were given on the maxim "Assert/assess the strongest proposition relevant". Note, however, that a certain measure of relevance is secured by the circumstance that that proposition is not the strongest proposition as such, corresponding to a contradiction, or just any strong proposition, but one among a limited number stronger than another, in fact, exactly one as polarities go, corresponding to a (true) sentence entailing the question supplied with a sign and obtained by exchanging one item. So a link is established between the direct and the indirect answer prior to relevance considerations. Relevant alternative candidates are selected on the basis of independently accessible linguistic knowledge. The relevance question is reduced from *What is relevant?* to *Is this relevant?*; the discovery procedure is transformed into a decision procedure, and this process is low-level and domain-independent.

Items that are interconnected by a semantic relation such as scalarity and hyponymy seem to tend to be mutually relevant so that irrelevance cases can be considered exceptions to the rule. It is not impossible that the assessment of a higher value is irrelevant once a value is confirmed or that the assessment of a lower value is uninteresting once a value is denied, but it is as improbable as it is that items arrived at on more pragmatic considerations are irrelevant. Likewise, one cannot exclude that the confirmation of, say, a subkind once a superkind is confirmed or the confirmation of a sisterkind or the denial of a superkind once a natural kind is denied is uninteresting, but one can think it equally improbable. So regularly - *by default* - when there is a scale or a hierarchy around the item in question, all items in that scale or hierarchy will enter into the set of propositions at issue, then on meaning postulates, defining the interrelationships in terms of entailment, one alternative can be identified as the informative in dependence on the distribution of polarities in that set.

With reference to the parametric discourse relations and particles paradigm presented above, the sketched cases of complex response cover five relation classes. Each of these is strongly motivated in the framework of cooperative response insofar as any complex answer patterned on it is appropriate in principle. Given a query <C,D> where D is the focus, any complex response +/-,+/-<C,R> where R is an alternative such that the parameters are chosen accordingly - and truthfully - is basically adequate.

When moving upward on a scale, in case the stronger statement is verified, the particle *sogar* applies; otherwise, if it is falsified, the particle *aber*. When moving downward on a scale, in case the weaker statement is verified, the particle *aber* applies again; otherwise, if it is falsified, the particle *einmal*. When moving downward in a hierarchy and the statement is verified, the specification particle *und zwar* applies. When moving upward and the statement is falsified, the particle *überhaupt* applies. Finally, moving sideways to verify, *sondern* is the particle.

The PASSAT System

The tiny database query system PASSAT, consisting in one PROLOG II program comprising approximately 100 clauses, is designed to demonstrate a register of rules regulating choices, of alternatives to lexical items and of particles of discourse, in accordance with semantic relations and in terms of polarity. It is devised to imitate a natural performance in three respects:

- *Quality*, the search for the true response;
- *Quantity*, the quest for the exhaustive response - information on a relevant alternative;
- *Coherence*, the search for the discourse particle;

and lexical entailments underlie all three aspects. So while the system is primarily intended as an illustration of a facet of centerpiece functions of German discourse particles, it is at the same time a small but systematic model of complex response principled on independently available knowledge.

PASSAT exploits a sortal hierarchy of natural kinds and a scalar structure of ranked items to arrive at relevant alternative data and to select appropriate discourse particles to bridge the gaps, borrowing its terminology and database facts from shipping. Such computations rely on a variety of modules:

- *Lexicon*. Here, semantic relations between and among lexical items, such as sortal sameness, "antonymy" (Disjunctivity: Difference and sortal sameness), hyponymy, intersectivity (cross concepts, uniting different-sort items), "synonymy" (comparability in a strict sense), and scalarity, are designated and defined in their own terms.

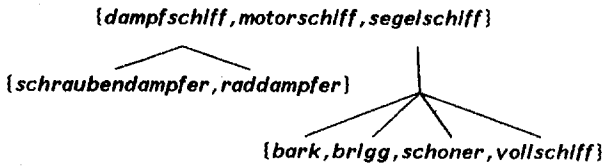
- *Meaning postulates*, where semantic relations introduced in the lexicon are defined and interpreted by entailment, that is, in terms of (necessary) polarity in view of the sentence context.

- *Alternative relations*, stating conditions on which one item constitutes an alternative to another in terms of lexical relations and (simple) polarity in view of the sentence context.

- *Particle relations*, stating sufficient conditions for the output of a certain particle in terms of alternative relations and (simple) polarity in view of the sentence context.

- *Response rules*, evaluating original queries and perceived and received substitute queries (alternatives in the sentence context) in terms of database facts or meaning postulates.

- *Knowledge base*, containing the minimal amount of primitive fact (no facts that are deducible from other facts on meaning postulates), representing (predominantly positive) polarity.



This is one of the two sortal hierarchies PASSAT is acquainted with. Questions to be understood by PASSAT are of the form

- Ist x P?

where *x* is an individual name (that of a ship) and *P* is a predicate, e.g. a common noun, so throughout it is a question of a constant's membership in a set. First answers (*Ja* or *Nein*) are by a large measure calculated by way of meaning postulates defined on lexical relations like hyponymy or antonymy, and these same relations go in turn to compute second answers. Once a first response is found, PASSAT goes on to seek alternatives: Provided that answer was yes, it seeks to enhance the specificity of the predicate, e.g. to restrict the set denoted by the noun by moving downward in the sortal hierarchy:

- Ist fortuna segelschiff?
- Ja, und zwar bark.

In case polarity is negative in the first run, the system seeks to increase informativity by e.g. searching for the set to which the individual does belong (moving sideways in the sortal hierarchy):

- Ist preciosa bark?
- Nein, sondern brigg.

These two basic principles are enriched and extended by a *recursive* mechanism: As soon as an alternative to the subject of interrogation has been determined, the search goes on for an alternative to that alternative, entering into the rôle of the subject of interrogation, and so on:

- Ist concordia dampfschiff?
- Nein, sondern segelschiff, und zwar schoner.

- Ist prudentia schoner?
- Nein,
überhaupt nicht segelschiff,
sondern dampfschiff,
und zwar raddampfer.

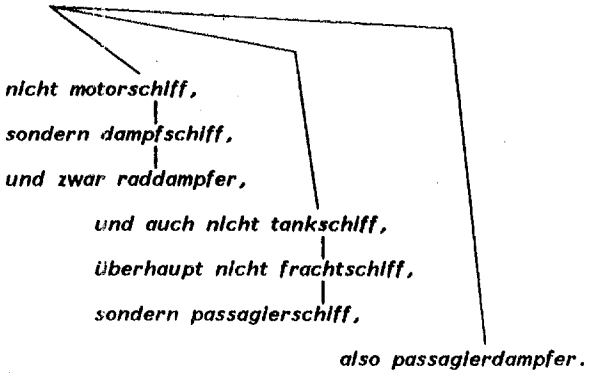
On the other hand, PASSAT is acquainted with another hierarchy too. The concept *ship* is subdivided on two equivalent points of view, the locomotion and the function:

And composites are introduced which combine these two hierarchies: For a constant to be a member of such a set, it must belong to both sets denoted by the two components:

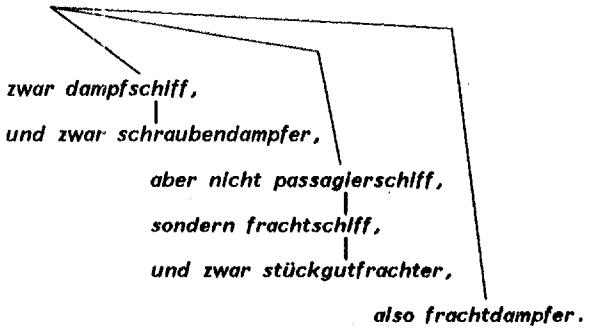
{*frachtdampfer,passaglerdampfer,motortanker*}

And here, a context sensitivity inside the complex answer has been installed (by means of an extra variable position in the alternative relations) to permit a second and a third alternative to the first answer to be stalled until the sequences of "lower-level" alternatives to the second and third answers (first and second alternatives) are exhausted, to be readdressed with *backtracking*:

- *Ist prudentia motortanker?*
- *nein,*



- *Ist poseidon passaglerdampfer?*
- *nein,*



The conditions under which a cross-concept like *frachtdampfer* is an (ultimate) alternative to another are rather complex insofar as it requires numerous steps to come to a conclusion as to whether to draw a conclusion by use of *also* (approximating English *so*). It depends on the arrangement of both of the two intersected kinds, in casu, *dampfschiff* and *frachtschiff*, in relation to the other pair, say, *motorschiff* and *tankschiff* uniting to *motortanker*.

- *Ist poseidon motortanker?*
- *nein,*
  nicht motorschiff,
  sondern dampfschiff,
  und zwar schraubendampfer,
  und auch nicht tankschiff,
  sondern stückgutfrachter.

Thus *frachtdampfer* is no alternative to *motortanker* because the two corresponding component kinds *tankschiff* and *frachtschiff* are downward specific (the former is more specific than the latter), whereas the converse is not the case - *motortanker* does form an alternative to *frachtdampfer* as *frachtschiff* and *tankschiff* are upward specific; the latter is more specific than the former.

[*gestrandet,gescheitert*]

This is one of two scales known to PASSAT. Again, comprehensible queries are of the form *Ist x P?* where *x* is an individual name and *P* a predicate, but this time the predicate is not a common noun but a (perfect participle) intransitive verb. Once a primary answer is given to a query, a search starts for an alternative answer once over, and given a positive primary response, the system seeks, again, to enhance the specificity of *P*, only now not by seeking to restrict the set by downward movement with respect to a hierarchy but by upward movement on the scale, to assess the next value irrespectively of whether it is valid or not:

- *Ist preciosa gestrandet?*
- *Ja, sogar gescheitert.*

- *Ist fortuna gestrandet?*
- *Ja, aber nicht gescheitert.*

Given a primary answer with polarity negative, however, as before, PASSAT tries to increase the information value nevertheless through strengthening the statement, but not by searching for confirmation sideways or a more comprehensive denial upward in a hierarchy,

- *Ist prudentia schoner?*
- *nein, überhaupt nicht segelschiff, sondern dampfschiff...*

but by assessing the next value in the downward direction on the scale induced by the verb:

- *Ist fortuna gescheitert?*
- *nein, aber gestrandet.*

- *Ist felicia gescheitert?*
- *nein, nicht einmal gestrandet.*

Now there is another scale known to the system:

[*gekentert,gesunken*]

And the two scales are associated with one another in a structure which presents a pragmatic case of alternativity (the only one in the system). More precisely, the items *gekentert* and *gestrandet* are in a symmetric relation termed *syn* as a pseudo-synonymy case, with the consequence (and purpose) that in case the - primary or secondary - answer to either one of the lower values - as a query or alternative - is negative - in the first or the second instance, and the answer to the other lower value is positive, then that other lower value is treated as an alternative, on the consideration that in view of the higher goal of the query, it will be of interest:

- *Ist concordia gestrandet?*
- *nein, wohl aber gekentert, sogar gesunken.*

- *Ist concordia gescheitert?*
- *nein, nicht einmal gestrandet,*
  wohl aber gekentert, sogar gesunken.

Limitations

The test of any natural-language system, whether generative or interpretative, is in its measure of generality or flexibility, in its aptness for expansion and extension in various dimensions. As far as the present program is concerned, these dimensions can be identified with a range of linguistic modules: Morphology (1), lexicon (2), syntax (3), semantics (4), and pragmatics (5).

(1) Deliberately, no morphology has been built into the system. Relevant items would have been (a) the indefinite article (*eln*), (b) gender variants (*eln/-e*) and (c) coherent forms (*keln/-e*). These refinements are omitted in order not to pay undue complexity to such inessentials, though the implementation would be feasible.

(2) The lexical items and relations are not casuistic in the sense that they are unrepresentative of hierarchies and scales in German. Parallel structures can be added or substituted without difficulty. Only, real hierarchies and scales do not exist in isolation but in integration in a taxonomic superstructure. One problem is that a concept (e.g. *schiff*) may be open to specification in sequence (e.g. *ja, und zwar segel-schiff, und zwar vollschiff*), another is that a concept (e.g. *schiff*) may be open to specification in two directions (e.g. *ja, und zwar passaglerschiff und motorschiff*). Before the concept *schiff* enters into PASSAT in the obvious way, a method must be developed to determine how far and which way relevance is to reach in each case in view of the user's interest.

(3) Deliberately, only a minimal syntax has been built into the system. This is, again, to accentuate the central principles, but more to not create the impression that interesting syntax problems have been solved. Thus the rules of ellipsis have not been explored. PASSAT uses total ellipsis throughout - though not on deliberation, but by necessity. It could instead use partial ellipsis discriminately to put out answers like *neln, zwar ist sie eln x, und zwar eln y, aber sie ist keln z, sie ist überhaupt keln u, sondern eln v* - yet it would still do so not by first generating and then reducing complete structures but by producing those strings blindly.

(4) The system suffers a serious shortcoming in not assessing the lexical relations in meaning postulates and alternative relations in terms of the sentence context semantic structure. By accident, predicates (common nouns or verbs) occur in predicative position with the copula throughout, so the semantic relation invariably comes to the surface. As soon as contexts are introduced where the noun e.g. serves to restrict quantification over a binary relation, or, as a simplification, it is an argument of a binary relation, as in *hat fortunatus eine brigg* (compatible with *fortunatus hat eine bark*), the relation ceases to carry semantic relevance, and the system must consider some semantic representation to judge whether meaning postulates and alternative relations apply.

(5) The pragmatic open problem lies in where the focus lies, more exactly, to which component of the question any alternative can be considered relevant in particle terms, what part is the plausible depart in the first instance. At the current state of the system, the focus is located once and for all in the one-place predicate *P*, yet it is a commonplace that yes-no questions are systematically ambiguous insofar as their topic-focus structure has consequences for what constitutes a proper substitute, namely, a sentence where the topic stays the same and the focus changes. Thus it could be that a question like *Ist concordia eine bark* is intended to, in the event of a negative response, elicit not a continuation like *sondern eln schoner* but an extension like *aber fortuna ist eine* - for example, in case the higher goal of the query is to ascertain that there is a bark available for a higher purpose still. In principle there are means in a language to posit focus, and those means may be syntactical, like word order or clefting. Such are not, however, available to PASSAT, being a German-interfaced system, so without the phonology Germans use it is necessary to exploit a memory of past exchange, so as to address the higher goal of the query directly.

References

HIRSCHBERG, J. (1985) *A Theory of Scalar Implicature*. Univ. of Pennsylvania diss.

JOSHI, A./B.WEBBER/R.WEISCHEDEL. (1984) "Preventing False Inferences", in *Proceedings of COLING '84*.

KAPLAN, J. (1983) "Cooperative Responses from a Portable Natural Language Database Query System", in Brady/Berwick (eds.): *Computational Models of Discourse*, Cambridge, Mass.: MIT Press, 167-208.

SÆBØ, K.J. (1988) *A Model for Discourse Particles*. LILOG Report n, Stuttgart: IBM.