INFERENCING ON LINGUISTICALLY BASED SEMANTIC STRUCTURES

Eva Hajičová, Milena Hnátková

Department of Applied Mathematics
Faculty of Mathematics and Physics
Charles University
Malostranské n. 25
118 00 Praha 1, Czechoslovakia

ABSTRACT

The paper characterizes natural language inferencing in the TIBAQ method of question-answering, focussing on three aspects: (i) specification of the structures on which the inference rules operate, (ii) classification of the rules that have been formulated and implemented up to now, according to the kind of modification of the input structure the rules invoke, and (iii) discussion of some points in which a properly designed inference procedure may help the search of the answer, and vice versa.

I SPECIFICATION OF THE INPUT STRUCTURES FOR INFERENCING

A. Outline of the TIBAQ Method

When the TIBAQ (text-and-inference based answering of questions) project was designed, main emphasis was laid on the automatic build-up of the stock of knowledge from the (non-pre-edited) input text. The experimental system based on this method converses automatically the natural language input (both the questions and new pieces of information, i.e. Czech sentences in their usual form) into the representations of meaning (tectogrammatical representations, TR´s); these TR´s serve as input structures for the inference procedure that enriches the set of TR´s selected by the system itself as possibly relevant for an answer to the input question. In this enriched set suitable TR´s for direct and indirect answers to the given question are retrieved, and then transfered by a synthesis procedure into the output (surface) form if sentences (for an outline of the method as such, see Hajičová, 1976; Hajičová and Sgall, 1981; Sgall, 1982).

B. What Kind of Structure Inferences Should Be Based on

To decide what kind of structures the inference procedure should operate, one has to take into account several criteria, some of which seemingly contradict each other: the structures should be as simple and transparent as possible, so that inferencing can be performed in a well-defined way, and at the same time, these structures should be as "expressive" as the natural language sentences are, not to lose any piece of information captured by the text.

Natural language has a major drawback in its ambiguity: when a listener is told that the criticism of the Polish delegate was fully justified, one does not know (unless indicated by the context or situation) whether s/he should infer that someone criticized the Polish delegate, or whether the Polish delegate criticized someone/something. On the other hand, there are means in natural language that are not preserved by most languages that logicians have used for drawing consequences, but that are critical for the latter to be drawn correctly: when a listener is told that Russian is spoken in SIBERIA, s/he draws conclusions partly different from those when s/he is told that in Siberia, RUSSIAN is spoken (capitals denoting the intonation center); or, to borrow one of the widely discussed examples in linguistic writings, if one hears that John called Mary a REPUBLICAN and that then she insulted HIM, one should infer that the speaker considers "being a Republican" an insult; this is not the case, if the speaker said that then she INSULTED him.

These and similar considerations have led the authors of TIBAQ to a strong conviction that the structures representing knowledge and serving as the base for inferencing in a question-answering system with a natural language interface should be linguistically based: they should be deprived of all ambiguities of natural language and at the same time they should preserve all the information relevant for drawing conclusions that the natural language sentences encompass. The experimental system based on TIBAQ, which was carried out by the group of formal linguistics at Charles University, Prague (implemented on EC 1040 computer, compatible with IBM 360) works with representations of meaning (tectogrammatical representations, TR´s) worked out in the framework of functional generative description, or FGD (for the linguistic background of this approach we refer to Sgall, 1964; Sgall et al., 1969;

Hajičová and Sgall, 1980).

C. Tectogrammatical Representations

One of the basic tenets of FGD is the articulation of the semantic relation, i.e. the relation between sound and meaning, into a hierarchy of levels, connected with the relativization of the relation of 'form' and 'function' as known from the writings of Prague School scholars. This relativization makes it possible to distinguish two levels of sentence structure: the level of surface syntax and that of the underlying or tectogrammatical structure of sentences.

As for a formal specification of the complex unit of this level, that is the TR, the present version (see Plátek, Sgall and Sgall, in press) works with the notion of basic dependency structure (BDS) which is defined as a structure over the alphabet A (corresponding to the labels of nodes) and the set of symbols C (corresponding to the labels of edges). The set of BDS's is the set of the tectogrammatical representations of sentences containing no coordinated structures. The BDS's are generated by the grammar $G = (V_N, V_T, S, R)$, where $V_T = A \cup C$, $A = \{(a^q, GR)\}$, a is interpreted as a lexical unit, q is a variable standing for t and f (contextually bound and non-bound, respectively) and GR is interpreted as a set of grammatemes belonging to a; C is a set of complementations ($c \in C$, where c is an integer denoting a certain type of complementation, called a functor), C' denotes the set $\{<, >, <_c, >_c\}$ for every $c \in C$.

To represent coordination, the formal apparatus for sentence generation is to be complemented by another alphabet Q, where $q \in Q$ is interpreted as types of coordination (conjunctive, disjunctive, adversative, ..., apposition) , and by a new kind of brackets denoting the boundary of coordinated structures; $Q' = \{[_q,]_q\}$ for every $q \in Q$. The structures generated by the grammar are then called complex dependency structures (CDS).

Coming back to the notions of elementary and complex units of the tectogrammatical level, we can say that the complex unit of the TR is the complex dependency structure as briefly characterized above, while the elementary units are the symbols of the shapes a, q, c, q, the elements of GR, and the parentheses. The lexical units a are conceived of as elementary rather than complex, since for the time being we do not work with any kind of lexical decomposition. Every lexical unit is assigned the feature 'contextually bound' or 'non-bound' . The set of grammatemes GR covers a wide range of phenomena; they can be classified into two groups. Grammatemes representing morphological meaning in the narrow sense are specific for different (semantic) word classes: for nouns, we distinguish grammatemes of number and of delimitation (indefinite, definite, specifying):for adjectives and adverbs, grammatemes of degree, for verbs, we work with grammatemes of aspect (processual, complex, resultative), iterativeness (iterative, non-iterative), tense (simultaneous, anterior, posterior), immediateness (immediate, non-immediate), predicate modality (indicative, possibilitive, necessitive, voluntative), assertive modality (affirmative, negative), and sentential modality (declarative, interrogative, imperative). The other group of grammatemes is not - with some exceptions - word-class specific and similarly as the set of the types of complementations is closely connected with the kinds of the dependency relations between the governor and the dependent node; thus the Locative is accompanied by one member of the set {in, on, under, between, ...}.

The dependency relations are very rich and varied, and it is no wonder that there were many efforts to classify them. In FGD, a clear boundary is being made between participants (deep cases) and(free) modifications: participants are those complementations that can occur with the same verb token only once and that have to be specified for each verb (and similarly for each noun, adjective, etc.), while free modifications are those complementations that may appear more than once with the same verb token and that can be listed for all the verbs once for all; for a more detailed discussion and the use of operational criteria for this classification, see Panevová 1974; 1980; Hajičová and Panevová, in press; Hajičová, 1979; 1983. Both participants and modifications can be (semantically) optional or obligatory; both optional and obligatory participants are to be stated in the case frames of verbs, while modifications belong there only with such verbs with which they are obligatory.

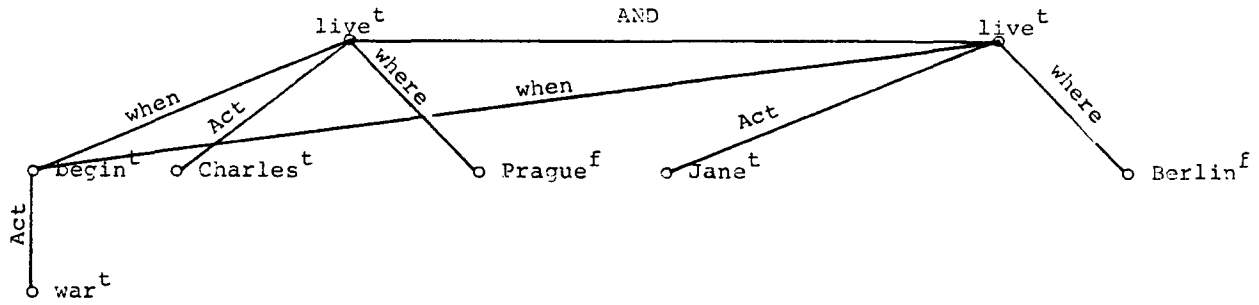
In the present version of FGD, the following five participants are distinguished: actor/bearer, patient (objective), addressee, origin, and effect. The list of modifications is by far richer and more differentiated; a good starting point for this differentiation can be found in Czech grammars (esp. Šmilauer, 1947). Thus one can arrive at the following groupings:

(a) local: where, direction, which way,
(b) temporal: when, since when, till when, how long, for how long, during,
(c) causal: cause, condition real and unreal, aim, concession, consequence,
(d) manner: manner, regard, extent, norm (criterion), substitution, accompaniment, means (instrument), difference,

benefit, comparison.

In our discussion on types of complementations we have up to now concentrated on complementations of verbs; with the FGD framework, however, all word classes have their frames. Specific to nouns (cf. Pitha, 1980), there is the partitive participant (a glass of water) and the free modifications of appurtenance (a leg of the table), of general relationship (nice weather), of identity (the city of Prague) and of a descriptive attribute (golden Prague).

To illustrate the structure of the representation on the tectogrammatical level of FGD, we present in Fig. 1 a complex dependency structure of one of the readings of of the sentence "Before the war began, Charles lived in PRAGUE and Jane in BERLIN" (which it has in common with "Before the beginning of the war, Charles lived in PRAGUE and Jane lived in BERLIN");to make the graph easier to survey, we omit there the values of the grammatemes.



the linearized form:

$$\langle\langle(war^t, \{sing, def\}) >_{Act} (begin^t, \{anter, compl, noniter, nonimmed, indic, affirm, before\}) >_{when} [\langle(Charles^t, \{sing, det\}) >_{Act} (live^t, \{anter, compl, noniter, nonimmed, declar, indic, affirm\}) \leq_{where} (Prague^f, \{sing, def, in\}) > \langle (Jane^t, \{sing, def\}) >_{Act} (live^t, \{anter, compl, noniter, nonimmed, declar, indic, affirm\}) \leq_{where} (Berlin^f, \{sing, def, in\}) >]_{AND}$$

Fig. 1

II INFERENCE TYPES

A. Means of Implementation

The inference rules are programmed in Q-language (Colmerauer, 1982), which provides rules that carry out transformations of oriented graphs. Since the structures accepted by the rules must not contain complex labels, every complex symbol labelling a node in TR´s has the form of a whole subtree in the Q-language notation (in a Q-tree).

The set of TR´s constitutes a semantic network, in which the individual TR´s are connected into a complex whole by means of pointers between the occurrences of lexical units and the corresponding entries in the lexicon. (Questions of different objects of the same kind referred to in different TR´s will be handled only in the future experiments.)

The following procedures operate on TR´s:

(i) the extraction of (possibly) relevant pieces of information from the stock of knowledge;
(ii) the application of inference rules on the relevant pieces of information,
(iii) the retrieval of the answer(s).

The extraction of the so-called relevant pieces of information is based on matching the TR of the input question with the lexicon and extracting those TR´s that intersect with the TR of the given question in at least one specific lexical value (i.e. other than the general actor, e.g. one, the copula, etc.); the rest of the trees (supposed to be irrelevant for the given question) are then deleted.

The set of relevant TR´s is operated upon by the rules of inference. If a rule of inference has been applied, both the

source TR as well as the derived TR constitute a part of the stock of knowledge and can serve as source TR's for further processing. In order to avoid infinite cycles, the whole procedure of inferencing is divided into several Q-systems (notice that rules within a single Q-system are applied as long as the conditions for their application are fulfilled, i.e. there is no ordering of the rules).

B. Types of Inference Rules

1. Rules operating on a single TR:

(i) the structure of the tree is preserved; the transformation concerns only (a) part(s) of the complex symbol of some node of the CDS (i.e. label(s) of some node(s) in the Q-tree of the TR):

(a) change of a grammateme:

$V_{perform}$-Possib (N_{device}-Act) (X-Pat) ... ==

$V_{perform}$-Indic (N_{device}-Act) (X-Pat) ...

Note: In our highly simplified and schematic shapes of the rules we quote only those labels of the nodes that are relevant for the rule in question; the sign == stands for "rewrite as"; N_{device} stands for any noun with the semantic feature of "device", $V_{perform}$ for a verb with the semantic feature of action verbs, Possib and Indic denote the grammatemes of predicate modality.

Ex.: An amplifier can activate a passive network to form an active analogue. == An amplifier activates a passive network to form an active analogue.

(b) change of a functor (type of complementation):

V-use (N_i-Pat) (N_j-Accomp) ... ==
V-use (N_i-Regard) (N_j-Pat) ...

Ex.: Operational amplifier is used with negative feedback. == With operational amplifier negative feedback is used.

$V_{perform}$ (N_i-Act) (N_j-Pat) ... ==
$V_{perform}$ (D_{gen}-Act) (N_i-Instr) (N_j-Pat)..

Ex.: Operational amplifiers perform mathematical operations == Mathematical operations are performed by means of operational amplifiers.

Note: Act, Pat, Instr, Accomp, Regard stand for the functors of Actor, Patient, Instrument, Accompaniment and Regard, respectively; D_{gen} denotes a general participant.

(c) change of the lexical part of the complex symbol accompanied by a change of some grammateme or functor:

V_i-Possib ((few)N_i) (V-use(N_k-Accompneg) ...)... ==V_i-Necess ((most)N_i) (V-use (N_k-Accompposit)...)...

Ex.: With few high-performance operational amplifiers it is possible to maintain a linear relationship between input and output without employing negative feedback.== With most ... it is necessary to maintain ... employing negative feedback.

(ii) a whole subtree is replaced by another subtree:

Ex.: a negative feedback == a negative feedback circuit

(iii) extraction of a subtree to create an independent TR:

- relative clause in the topic part of the TR

V_i (V_j-Gener-L(...))... ==
V_j-Gener-L (...)

Ex.: An operational amplifier, which activates a passive network to form an active analogue, is an unusually versatile device. == An operational amplifier activates a passive network to form an active analogue.

Note: L stands for the grammateme "contextually bound", R for "non-bound", Gener for the functor of general relationship.

- causal clause in TR's with affirmative modality

V_i-Affirm (V_j-Cause (...))... ==
V_j (...)

Ex.: Since an operational amplifier is designed to perform mathematical operations, such basic operations as ... are performed readily. == An operational amplifier is designed to perform mathematical operations.

- deletion of an attribute in the focus part of a TR

V_i (N_j-R (X-Gener-R)) ... ==
V_i (N_j-R) ...

Ex.: Operational amplifiers are used as regulators ... to minimize loading of reference diodes permitting full exploitation of the diode´s precision temperature stability. == Operational amplifiers are used as regulators ... to minimize loading of reference diodes.

(iv) the transformation gives rise to two TR´s

distributivity of conjunction and disjunction (under certain conditions: e.g. for the distributivity of disjunction to hold, the grammateme of Indic with the main verb is replaced by the grammateme of Possib)

Ex.: Operational amplifiers are used in active filter networks to provide gain and frequency selectivity. == Operatinal amplifiers are used in active filter networks to provide gain. Operational amplifiers are used in active networks to provide frequency selectivity.

2. Rules operating (simultaneously) on two TR´s
(the left-hand side of the rule refers to two TR´s)

- conjoining of TR´s with the same Actor

Ex.: An operational amplifier activates a passive network to form an active analogue. An operational amplifier performs mathematical operations. == An operational amplifier activates .... and performs ....

- use of definitions: the rule is triggered by the presence of an assertion of the form "X is called Y" and substitutes all occurrences of the lexical labels X in all TR´s by the lexical label Y

III EFFECTIVE LINKS BETWEEN INFERENCING AND ANSWER RETRIEVAL

A. The Retrieval Procedure

The retrieval of an answer in the enriched set of assertions (TR´s) is performed in the following steps:

(a) first it is checked whether the lexical value of the root of the TR is identical with that of the TR of the question; if the question has the form "What is performed (done, carried out) by X?", then the TR from the enriched set must include an action verb as a label of its root;

(b) the path leading from the root to the wh-word is checked (yes-no questions are excluded from the first stage of our experiments); the rightmost path in the relevant TR must coincide with the wh-path in its lexical labels, contextual boundness, grammatemes and functors (with some possible deviations determined by conditions of substitutability: Singular - Plural, Manner - Accompaniment, etc.); the wh-word in the question must be matched by a lexical unit of the potential answer, where the latter may be further expanded;

(c) if also the rest of the two compared TR´s meet the conditions of identity or substitutibility, the relevant TR is marked as a full answer to the given question; if this is not the case but at least one of the nodes depending on a node included in the wh-path meets these conditions, then the relevant TR is marked as an indirect (partial) answer.

B. Towards an Effective Application of Inference Rules

In the course of the experiments it soon became clear that even with a very limited number of inference rules the memory space was rapidly exceeded. It was then necessary to find a way how to achieve an effective application of the inference rules and at the same time not to restrict the choice of relevant answers. Among other things, the following issues should be taken into consideration:

The rules substituting subtrees for subtrees are used rather frequently, as well as those substituting only a label of one node (in the Q-tree, i.e. one element of the complex symbol in the CDS), preserving the overall structure of the tree untouched. These rules operate in both directions, so that it appears as useful to use in such cases a similar strategy as with synonymous expressions, i.e. to decide on a single representation both in the TR of the question and that included in the stock of knowledge; this would lead to an important decrease of the number of TR´s that undergo further inference transformations.

Only those TR´s are selected for the final steps of the retrieval of the answer (see point (a) in III.A) that coincide with the TR of the question in the lexical label of the root, i.e. the main verb. If the inference rules are ordered in such a way that the rules changing an element of the label of the root are applied before the rest of the rules, then the first step of the retrieval procedure can be made before the application of other inference rules. This again leads to a con-

siderable reduction of the number of TR´s on which the rest of the inference rules are applied; only such TR´s are left in the stock of relevant TR´s

(i) that agree with the TR of the question in the label of the root (its lexical label may belong to superordinated or subordinated lexical values: device - amplifier, etc.),

(ii) that include the lexical label of the root of the question in some other place than at the root of the relevant TR,

(iii) if the question has the form "Which N ...", (i.e. the wh-node depends on its head in the relation of general relationship), then also those TR´s are preserved that contain an identical N node (noun) on any level of the tree.

The use of Q-language brings about one difficulty, namely that the rules have to be formulated for each level for the tree separately. It is possible to avoid this complication by a simple temporary rearrangement of the Q-tree, which results in a tree in which all nodes with lexical labels are on the same level; the rules for a substitution of the lexical labels can be then applied in one step, after which the tree is ´returned´ into its original shape.

These and similar considerations have led us to the following ordering of the individual steps of the inference and retrieval procedure:

1. application of rules transforming the input structure to such an extent that the lexical label of the root of the tree is not preserved in the tree of a potential answer;

2. a partial retrieval of the answer according to the root of the tree;

3. application of rules substituting other labels pertinent to the root of the tree;

4. partial retrieval of the answer according to the root of the tree;

5. application of inference rules operating on a single tree;

6. application of inference rules operating on two trees;

7. the steps (b) and (c) from the retrieval of the answer (see III.A above).

REFERENCES

Colmerauer A., 1982, Les systemes Q ou un formalisme pour analyser et synthétiser des phrases sur ordinateur, mimeo; Germ.transl. in: Prague Bull. of Mathematical Linguistics 38, 1982, 45-74.

Hajičová E., 1976, Question and Answer in Linguistics and in Man-Machine Communication, SMIL,No.1,36-46.

Hajičová E., 1979, Agentive or Actor/Bearer, Theoretical Linguistics 6, 173-190.

Hajičová E., 1983, Remarks on the Meaning of Cases, in Prague Studies in Mathematical Linguistics 8, 149-157.

Hajičová E. and J. Panevová, in press, Valency (Case) Frames of Verbs, in Sgall, in press.

Hajičová E. and P. Sgall, 1980, Linguistic Meaning and Knowledge Representation in Automatic Understanding of Natural Language, in COLING 80 - Proceedings, Tokio, 67-75; reprinted in Prague Bulletin of Mathematical Linguistics 34, 5-21.

Hajičová E. and P. Sgall, 1981, Towards Automatic Understanding of Technical Texts, Prague Bulletin of Mathematical Linguistics 36, 5-23.

Panevová J., 1974, On Verbal Frames in Functional Generative Description, Part I, Prague Bulletin of Mathematical Linguistics 22, 3-40; Part II, PBML 23, 1975, 17-52.

Panevová J., 1980, Formy a funkce ve stavbě české věty /Forms and Functions in the Structure of Czech Sentence/, Prague

Pitha P., 1980, Case Frames for Nouns, in Linguistic Studies Offered to B. Siertsema, ed. by D.J.v.Alkemade, Amsterdam, 91-99

Plátek M., Sgall J. and P. Sgall, in press, A Dependency Base for a Linguistic Description, to appear in Sgall, in press.

Sgall P., 1964, Zur Frage der Ebenen in Sprachsystem, Travaux linguistiques de Prague I, 95-106.

Sgall P., 1982, Natural Language Understanding and the Perspectives of Question Answering, in COLING 82, ed. by J. Horecký, 357-364.

Sgall P., ed., in press, Contributions to Functional Syntax, Semantics and Language Comprehension, to appear in Amsterdam and Prague.

Sgall P., Nebeský L., Goralčíková A. and E. Hajičová, 1969, A Functional Approach to Syntax, New York.

Šmilauer V., 1947, Novočeská skladba /A Present-Day Czech Syntax/, Prague.