PERSPECTIVE PAPER: LINGUISTICS

Petr Sgall
Charles University

Introduction

The book by Sparck Jones and Kay (1973), which analyzes the importance of linguistics for various domains of information science on the basis of an exhaustive examination of the trends and achievements of information retrieval, as well as the two subsequent surveys of the work in the field by Walker (1973) and Damerau (1976) make it possible to reconsider the role of linguistics in connection with the use of computers, taking into account the development of research in the last two decades, and the conditions of this development, as determined by internal and external factors. We would like to divide our remarks into four sections: (1) preliminary remarks on the relationships between linguistics and document retrieval, fact retrieval, and communication of man with a robot; (2) questions of linguistic theory and its new developments, as relevant for information science; (3) linguistic aspects of fact retrieval; (4) connections between linguistics and programming or computer science in a narrow sense.

Linguistics and Information Science

Document retrieval, which was the main object of the book by Sparck Jones and Kay, requires a detailed linguistic elaboration of many domains. Among them are the lexicon and its structure (word formation, semantic relationships between lexical units); morphemics (the word as a set of word forms); word classes; classes of phrases (especially, the structure of the noun phrase as the typical shape of a term consisting in more than one word, from tooth-brush and consonant cluster through man-of-war and man in the street to bookkeeping by double entry, etc.); types of contexts (for statistical and other studies of cooccurrence of words); and, of course, many other questions. But still, Sparck Jones and Kay are undoubtedly correct when they write (pp. 61, 105, 111, 151, 158) that document retrieval has been connected, first of all, with "the use in a more or less *ad hoc* manner of linguistic information", with "quasilinguistic techniques" (for instance, the occurrence of prepositions is often used to identify noun phrases which might be relevant). On the other hand, the application of full-fledged linguistic models (transformational grammar, predictive analysis) can be found in connection with fact retrieval, and is only exceptional in document retrieval systems. (Let us note that, for inflectional, and perhaps also agglutinating, languages, an analysis of the endings of successive words may be perhaps even more useful than the analysis of prepositions is for English texts; see Kirschner and Buranova (in press).

Only a relatively small part (Chapter 7) of the book by Sparck Jones and Kay was devoted to fact retrieval, but the authors stated that "work in this area shows more of the shape of

things to come than anything else. . ."(p. 174). Walker and Damerau have confirmed this view explicitly, and their comprehensive surveys of the development of information retrieval since 1971 show convincingly that the expansion resulting from experimenting with fact retrieval (question answering) systems is most important for the future of the domain.

Thus, there are good reasons for a reexamination of the utility of linguistics for information science to be focused on fact retrieval rather than on automatic indexing and other methods of document retrieval. First, it is in the former domain that linguistic methods and linguistic systems may contribute to the basis of the study. With document retrieval it would suffice, probably, to use linguistic techniques based on a traditional approach, the discussions concerning methodological questions and linguistic theories in the strict sense being rather irrelevant here for the time being. And second, the development of research in information retrieval has been determined, since the beginning of the 1970's, by question answering systems rather than by investigations in document retrieval.

Therefore, we want to devote our contribution to the relationship between linguistics and fact retrieval. In the next section some issues of the theory of language will be examined from the viewpoint of the question -- formulated by Sparck Jones and Kay -- whether a substantial improvement of linguistic theories, "which are still far from adequate" (p. 198; cf. also pp. 80, 187), will contribute effectively to the "interpenetration" of linguistics and information science as two "natural bedfellows". Before starting that section, it might be useful to characterize briefly some more recent aspects of this question of the utility of linguistics for information science and artificial intelligence, especially from the viewpoint of what new insights and requirements Winograd's (1972) SHRDLU has brought into the discussion.

It is certainly possible to argue against the specific methods of grammatical and semantic description chosen by Winograd, claiming, e.g., that he has no large knowledge of linguistic writings, that he simplifies matters so that it will be necessary to redefine many of his linguistic notions if his system is to work with thousands of words instead of two hundred, or if his robot is to manipulate more differentiated objects, etc. Also, it is possible to maintain that his approach to questions of language has not directly brought a new paradigm of linguistic theory. But still, his personal knowledge of specific linguistic questions (in the "traditional" sense of the term linguistics, which in his terms includes transformational grammar, too) might not be decisive in this respect. (Let us remember that, in his time, Franz Bopp was said to have too weak a knowledge of Latin to be considered a good linguist, but it was he who introduced a programme of Indo-European comparative linguistics, which was carried out in the development of linguistics through the next hundred years).

The main new factors Winograd's approach has introduced into linguistics consist neither in his choice of linguistic theory, nor in his opinions on this or that type of phrase. They concern primarily the relationship between a description of linguistic competence and other fields, particularly, a theory of performance, pragmatics, a theory of reference. To illustrate this point, let us first examine briefly some linguistic aspects of Winograd's study.

It appears at first that Winograd has turned upside down many of the values considered basic in algebraic as well as in structural linguistics. While analyzing the text, he does not proceed from one level of its representation to another, he does not work with the term sentence (not using any unit between clause and text), and he applies a grammar that was not equipped with an explicit framework before, i.e., Halliday's systemic grammar. Winograd states that the generative power of his recognition routine exceeds that of

context-free grammars (p. 82) and can be compared with that of unrestricted rewriting systems (cf. pp. 43ff regarding Wood's transition networks, which could be translated almost directly into PLANNER programs). He points out that man, in distinction to deductive systems based on predicate logic, has "a large set of heuristics and procedures for solving problems at different levels of generality" (p. 39) and uses them also as a hearer in the process of communication, when he resolves the ambiguity of words and phrases, applying not only his linguistic competence, but also his knowledge about the world to exclude unacceptable parsing variants (p. 33 and *passim*).

After a more detailed examination of Winograd's approach, one may infer that the distinctions between his standpoints and those having been more or less common in algebraic linguistics, as far as the points mentioned are concerned, need not be regarded as crucial. The division of linguistic description into levels is present in Winograd's study as well as with Chomsky or, e.g., in the writings of the classical Prague school of structural linguistics. Each of the levels is described by a specific device in his system: The so-called morphographemic representation is achieved (or, morphemic analysis is performed) by his input program, combined with the dictionary look-up (pp. 5, 73ff, 93ff). Syntax is, of course, held apart from semantics, and the level of linguistic semantics is treated by a component of his system other than the (deductive) analysis of the robot's knowledge of the world (the former being handled by the component called Semantics, pp. 126ff*, the latter by Blocks, pp. 117ff). Certainly, the interaction between these components is typical for Winograd's system; at many points of the parsing, Semantics can be called to check whether this or that parsing variant is to be pursued, and Semantics may call Blocks to see whether this or that noun-phrase, for instance, would be nonsensical. Thus it is meaningful to argue whether a NP *the street in a car* (in *I rode down...*) should be excluded by the semantic or rather by the cognitive component. (Winograd speaks here, p. 23, about the semantic analyzer, without discussing the borderline between linguistic competence and knowledge of the world from this viewpoint.) In any case, the interaction between components is possible only if these components differ from each other (we will return to this point below).

Winograd's attitude to the term sentence is not so deeply distinct from classical linguistic viewpoints as it might seem. He just speaks (pp. 17ff) about a major clause ("which could stand alone as a sentence"), more or less avoiding the traditional term so as to be able to treat coordinated major clauses in a way similar to sequences of utterances in a text. But this last point -- the absence of a strict distinction between sentence boundary and coordination -- has been well known for decades in linguistics (cf. Daneš, 1951; Pitha, 1967; Allerton, 1969; Katz and Fodor's, 1963, maxim concerning the synonymy of coordinate conjunction with sentence boundary is discussed in Sgall, 1973).

In discussing the form of Winograd's grammar and the remaining features of his approach mentioned above (generative power, use of heuristic procedures, and of the knowledge of the world during the process of understanding a text), we must bear in mind that his aim is not to describe a language system, or linguistic competence, but to describe linguistic performance (or the functioning of language in the process of communication), to present a model of language use, as he himself puts it (p. 3). Thus, if he has been successful in showing that, for his aim, this or that (or none) of the approaches known from algebraic linguistics is as suitable as his own theoretical basis, this does not prove anything about the

----

*We assume that, e.g., the terms "semantic subject" and "semantic (first, second) object" (p. 132) belong to the level of output representations of this component.

Petr Sgall

appropriateness of linguistic theories as such. It is another goal to describe the understanding of language -- where shortcuts using extralinguistic knowledge are necessary for disambiguation and also perhaps involve the necessity of a greater generative power -- than to describe language as such, with every detail of its intricate structure (but excluding non-linguistic items). One can discuss whether this or that goal is more appropriate, but, first of all, it is necessary to distinguish between the two of them, which might be made more difficult with such formulations as the following: "Language is a process of communication between people, and is inextricably enmeshed in the knowledge that those people have about the world"(p. 26). This process, that of communication or of linguistic performance, is distinct from language itself; this process is identical not with language, but with its use or functioning. Different languages can be used in such a process, even if exactly the same knowledge about the world is involved. A theoretical description of language itself definitely cannot be identical with a description of the process of communication. Some applications of linguistics require a model of language use, but they are still limited to experiments with a very restricted universe and also with a very restricted portion of natural language. If, in Winograd's experiment, only 200 lexical units, only a dozen extra-linguistic objects, and only one pair of interlocutors are involved, then, certainly, it would be out of place to use a very complicated grammar of English, with a detailed description of English syntax. But if linguistically oriented systems of artificial intelligence are to be applied practically to a wider domain of tasks, will it not be useful to state explicitly that the system of language (langue, linguistic competence) is complicated enough to deserve to be described in itself? Such a description appears necessary, if the interaction with other subsystems of human knowledge should be handled (e.g., at the relevant points of a syntactic description, cognitive procedures are then called to intervene). If I am not mistaken, Winograd's system acts in this way, in principle, and it remains only to state it (and to check whether his type of grammar is capable of being complemented as to account for most different phenomena of the syntax of natural language).

One more argument might be useful to make the idea of a relatively independent description of language more plausible. It is needed not only by pure theory, but also by different types of applications, where we are not concerned with a given and restricted subset of the universe, but where it is necessary to be prepared to handle most divergent sets of non-linguistic objects. Leaving out language learning (which could appear as too remote from the domain of artificial intelligence, although in the epoch of programmed learning this remoteness might not be quite certain), let us recall machine translation. Although Winograd refers to machine translation only as "having failed" (p. 42), he remarks about some of the systems built with the aim of an automatic syntactic analysis (which could be used for MT as well as for information retrieval or in the framework of man-machine communication in artificial intelligence), namely, those of Petrick and, especially, of Woods. He does not mention the approaches of Kay, of Mel'čuk, of the French group (Vauquois), of Kittredge (TAUM). I do not want to discuss the perspectives and possibilities of these groups now. I would like only to mention that the "failure" as yet to find procedures to handle the whole domain of grammar of natural language may well be due to the impossibility of concentrating the work of the rather small individual groups* and of equipping them with a thorough knowledge of different linguistic theories. But it is necessary, in the given context, to point out that the parsers designed for translation

----
*Walker's remarks concerning the possible effect of a concentration of the resources expended on question answering systems (p. 81) also may be well applied to the scattered research on machine translation. As Damerau notes (pp. 27-29), several groups continue to develop operationally-oriented (even if not purely commercial) systems of translation, others being interested in systems of translation that could be used inside other systems (e.g., as a vehicle for their evaluation)

cannot use much factual information, so that they have to rely on linguistic analysis proper to an extent greater than Winograd's approach has. More precisely, they also have to work with "general knowledge" (and preference or priority for this or that parsing variant), as well as with the knowledge of a given branch, and they should be able to use factual information gained in translating the previous portion of a text (not only the information from the local, but also that from the overall discourse context). For this, Winograd's way of handling the interaction of the components of his system might be of highest importance for the formulation of recognition routines. But still, MT systems must be so formulated as to provide for an interaction of the linguistic system proper (and that of general knowledge) with several (interchangeable) stocks of non-linguistic knowledge -- and also for an interaction of these stocks of knowledge with several (interchangeable) descriptions of natural languages. Thus, a relative independence of linguistic descriptions should be acknowledged.

Even if we could proceed to discuss various minor inadequacies and misformulations of Winograd's, it will be more useful to state that, although his combination of a linguistic description with that of a part of the world (in the form he declares) does not constitute a major contribution to the development of linguistics, there are at least two aspects in which he (as well as other authors who have achieved positive results in man-machine communication, cf. the writings of Hays, Kay, Klein, Schank, Wilks and others) really contributes to a new understanding of the structure of language and of the tasks of linguistics:

(a) The use of linguistic descriptions in artificial intelligence provides, as Winograd puts it, "a rigorous test for linguistic theories" (p. 2).

Although this, of course, is not a direct operational testing of (explicit and complete) theories, and although such testing has not yet proceeded far enough to permit it to be evaluated properly, it might be useful to add here several remarks concerning this point.

In accordance with what has been presumed already by many linguists, Winograd's approach (see especially p. 16) corroborates the view that such applications as those in the domain of artificial intelligence (as well as MT) are better served by those types of linguistic description that connect meaning and sounds explicitly. The fact that Winograd has not chosen any of the more or less formally elaborated systems of stratificational or functional description might have been caused by biographical factors. In any case, the underlying linguistic approach he has chosen, Halliday's systemic grammar, has much in common with the types of description elaborated by Mel'čuk, Lamb, Vauquois, Hays, by our Prague group, etc., especially in what concerns the relationship between semantics and syntax.* There are several levels, each of them having its proper syntax, and they are treated as ordered between the level of sound and that of meaning, each of these two bordering on non-linguistic domains. Inside transformational grammar, this would correspond more to the approach of generative semantics. Like most of the authors quoted above, and like Lakoff, McCawley, and Postal, Winograd also rejects the necessity of a specific level of deep syntax, formulating his programs as operations connecting each pair of the adjacent levels in the sequence of graphemics, morphemics, syntax, and semantics (as for the relationship of the last level to the non-linguistic domain, see below).

----
*Also other aspects are in common (which is not surprising, since Halliday as well as most of the other quoted authors use the classical European structural linguistics as their point of departure), e.g., the use of complex symbols comprising one lexical unit and the indices or grammatemes accompanying it, instead of handling grammatical units as if they were independent lexical items.

Another point relevant in this respect corroborates some views that are opposite to Chomskyan transformational grammar. The parsers of Winograd, Woods, and others do not use ordered rules (such as the ordered transformations in some versions of TG), but rather proceed from text to grammar. That is, while passing through the representation of a sentence, they try to find a solution for the given word or word group, instead of checking a list of rules and passing through the sentence representation with each rule separately, to see whether, at the given point of the analysis, it can be applied. With Winograd, this attitude is pronounced to such a degree that he does not proceed from the representation of the sentence (his major clause) on one level to that on another. Rather, he goes up to the highest levels (including that of the non-linguistic knowledge) already with the first NP that is being parsed, to eliminate the "nonsensical" parsing variants as soon as possible, so as not to burden the further portions of parsing by having to consider them. It is to be noted that with some systems (such as the functional generative description) this text-to-grammar orientation has been present in the generative as well as in the recognition routines.

> (b) The other, and probably most important point in which the impact of Winograd's approach influences the development of linguistics in a decisive way may be seen in his "imperative form" of representing knowledge and semantics (see esp. p. 116).

In his system, every semantic definition of a word has the form of a program for checking whether the mentioned object (class, relation, etc.) is present in the machine's memory. Also every declarative major clause interpreted by the machine is considered an instruction to deduce consequences from the given statement that may be useful in the further search for answers to the input questions. In his way it is possible to design a system more effective than one based entirely on deductive logic (the formulae of which give us no "strategy information", p. 39) and more similar to natural language (in which, as Winograd states there "is no sharp line dividing atomic concepts from nonatomic ones"; "the meaning of any concept depends on the entire knowledge of the speaker"; etc., pp. 26ff).

Certainly, it has to be shown whether this "imperative" and "program" form of semantic representations and definitions is not connected to a large extent with the specific properties of a dialogue between man and robot. Other purposes (e. g., fact retrieval, an automatical encyclopedia, or a theoretical description of language) might require some modifications or restrictions of this principle. But, as far as we can judge now, it is a sound principle which might be basically useful not only for technical applications, but also for the pure theory of linguistics. Incidentally, it is possible to quote at least three other sources of a similar principally new approach, using instructions, operations, etc., instead of semantic definitions or representations. The first of them is the well known approach of Lewis, of Montague grammars, etc., which use an operation (function, mapping) as the correlate of the meaning of word or of a word group (e.g., the intensional sense of an NP being represented by a function or prescription telling us what conditions an individual or a class of individuals must meet to belong to the extensional meaning of that NP). Another source of such a treatment can be seen in the handling of definitions as instruction in some approaches to modelling of cognition (as seen in the writings of D. G. Hays and his collaborators).

The third source of such an "imperative" understanding of statements is that connected with considerations of topic and focus (theme and rheme, functional sentence perspective, etc.). Although Halliday (1967) describes the "theme" and "focus" structures of English in detail (also using some of the results of Czech linguists, such as Mathesius, Daneš, and Firbas), in the book of Winograd these two aspects of the "imperative approach" are not

connected directly. Only in his paragraph on questions (pp. 137–141), does he makes a more or less systematic use of the concept of focus. We shall see later what the importance of these notions is inside the new understanding of semantics.

New Developments in Linguistic Theory

One cause of the fact that the interpenetration of linguistics and information science is not yet as intrinsic or intensive as it could be consists in the long development of·linguistics as oriented to "human" (i.e., more or less pedagogical, educational) applications. For centuries, linguistics has been applied almost exclusively to such practical purposes as the teaching and learning of foreign languages, or the care and culture of standard languages. Linguistic theory has developed under a steady influence of the requirements of these applications, which have some important features in common; these features are determined first of all by the fact that the user of the results of linguistic research always is human (teacher, pupil, reader). And a human being is, as a native speaker of one of the natural languages, accustomed to use a language.

This means that the characteristics of language use that are more or less common to different languages (and partly do not change even if a dialect is promoted to standard language) were by far not studied with such an intensity as were those domains of linguistics connected with features characteristic for individual languages, or for groups of them (grammar, lexicon, phonology). Also the causes of the relative lack of systematic semantic studies in traditional and structural linguistics can be found partly in the fact that semantic phenomena, even if they are not fully identical in different languages, appear as a connecting link between different languages rather than as points of difference between them.

Under the influence of the educational applications, linguistics was understood as studying first of all the system of language (the rules linking the sounds of a language with a poorly understood repertoire of semantic units), not its use. The problems of language use with their semantic, psychological, sociological, and other implications have not been altogether ignored, and they have been studied quite intensively, in some linguistic schools, under the headings of stylistics, psycholinguistics, sociolinguistics, etc. But most of these disciplines (or their predecessors) have not been urgently needed for practical applications, and their objects have not been so appropriate for structural modelling, so that most theoretical linguistic analyses concentrated on the structure of language, on grammar.

When Chomsky's first writings made it possible to handle grammar as a formal system, it was not all of linguistics that gained a level of formalization comparable to that of physics or chemistry, but only the then supposed core of linguistics. And one of the main findings achieved between 1965 and 1975, due to the experiments in fact retrieval, language uncerstanding, etc., as well as to theoretical research in semantics, consists in the evidence that formalization alone does not suffice to allow linguistics to be as useful in technical applications (in man-machine communication) as physics or chemistry are for industry. The requirement emerged for studying the questions of the use of language with the same intensity as those of grammar.

A pupil, as a user of a natural language, knows how to use the items of another language he is taught or reads about. He knows from his mother tongue what sentences may follow each other, how to identify the antecedent of an anaphorical pronoun, how to switch between *I* and *you* in a dialogue, and, even though he would not be able to use the proper terms, he is able to distinguish rather safely, without explicit training (by mere analogy to some clear

examples) an actor from a goal, a topic from a focus, etc. In these and many other points he just handles the items of the foreign language (or of a "higher" form of his mother tongue) in the same way that he has been accustomed to handle their counterparts in his own language. He must be instructed how to express the subject, the passive, what is the shape of the intonation centre, etc., in the language described (of course in many cases he can do without explicit instruction). But the functioning of such items, many of their possible combinations in a sentence or in longer contexts, etc., are already familiar to him. When he achieves the linguistic competence of his "new" language, he can perform (some secondary points aside), using the ability he had gained for another language (if the cultural backgrounds of the two languages do not differ too much).

For a machine it is not enough to be instructed in linguistic competence. Many aspects that had been neglected in linguistics (especially in the U.S.) before the 1960's are now beginning to be studied systematically, and the influence of technical applications is very intensive for some of them. The study of the structure of text, rapidly developing first of all in Western and Eastern Germany, is connected not only with the requirements of the theory of literature, but also with those of information retrieval. A sound basis for the classification of texts, a description of the coherence of texts including such areas as anaphorical relations, etc., are equally necessary for both of them. This applies also to the analysis of dialogue, of questions and answers, of the semantic patterns of personal pronouns. Other aspects of pragmatics, especially those connected with reference, are important for the automatic treatment of language in artificial intelligence and fact retrieval as well as in modern psychological research concerning the structure of human memory, knowledge and interaction with the world. Linguistic competence in Chomsky's sense constitutes only a small portion of the mechanism of communication, and if we want -- as Martin Kay formulated the task at the conference on computational linguistics in Varna, May 1975 -- to understand how it is possible that a string of symbols has the effect of changing the inner state of the mind of a human being who perceives the string, then we must regard computational linguistics not only as one of the applications of theoretical linguistics, but also as one of its main advanced posts, responsible for the future theoretical development itself.

We must realize, of course, that up to now not many theoretical linguists have devoted much attention to computers and information science. Only exceptionally (see, e.g., the writings of Mel'čuk, Lamb, Garvin, Kay, as well as the efforts of the Prague group of algebraic linguistics) has the work in theoretical linguistics been connected with a more or less direct concern with computing. This lack of understanding of the importance of contemporary "applications" certainly is connected with the fact that questions of performance, pragmatics, and of the variability of language use (as Walker, p. 88, puts it) have not yet been studied to an extent comparable with that of grammar. But the trend is positive; the younger generation of linguists (perhaps more in Western Europe than elsewhere, but Lakoff's recent writings also belong here) is beginning to stress the necessity for studying the main aspects of linguistic performance. Up to now, perhaps no fully reliable theoretical point of departure has been found in this domain, but new light has been shed on many empirical issues in connection with text linguistics and the theory of style. The study of pragmatical questions has found a more advantageous position in that some of the issues of pragmatics (especially the status of pragmatical indices or points of reference) have been elucidated from the point of view of semantics (in the writings of Montague, David Lewis and others; Schnelle speaks about a "semanticization" of a portion of pragmatics).

The question raised by Sparck Jones and Kay that we referred to earlier, about whether improvement of linguistic theories will help the interpenetration of linguistics and information science, can be answered only by "we still can hope", as far as performance and

pragmatics are concerned. But we want to show that as regards semantics, especially the semantics of the sentence, the answer already may be more positive.

We have already mentioned that the study of linguistic semantics also was largely neglected, although in this case the relationship between the two main reasons mentioned above may be other than in the case of performance and pragmatics (the difficult accessibility of semantic phenomena being stronger than the fact that their description was not necessary for traditional applications of linguistics). The study of semantics (not only in the lexicon, but also in the structure of the sentence) has never been fully absent in some linguistic approaches. Thus, in the Prague school, semantic issues have been paid attention since its very beginnings in the 1920's. Nevertheless, fact retrieval and language understanding have been studied without much connection with linguistic semantics, although semantics clearly belongs to the proper object of these disciplines. The lack of influence of linguistic semantics on information science has not only external causes. It was not easy to apply the results of the European schools of linguistics, which have not used an explicit language and have not been well known in America, where most of the computer models were developed. But we must admit that the lack of influence was mainly conditioned by the relatively low level of understanding of the nature of semantic phenomena in linguistics. One of the main characteristics of the new situation (after the development between 1965 and 1975, which has been described by Sparck Jones and Kay, Walker, and Damerau) consists in the progress made in the semantics of natural language.

It is well known that the new attention devoted to semantics in transformational theory led to the division of this school into the Chomskyan wing and the wing of generative semantics. Furthermore, the development of these two new theories has confirmed both the claim that semantics belongs to the core of linguistics (the relationship between sounds and meanings is the main object of the study of language systems) and the claim that a systematic attempt to include semantics in the transformational description would lead to the necessity of changing the structure of the description (cf. Sgall, 1964, p. 95). Moreover, one can see that questions of semantics still belong to the most bothersome issues in both versions of transformational description. Neither the research based on semantic markers in the sense of Katz, nor that of lexical decomposition with Lakoff has brought the possibility of coherent descriptions, the adequacy of which for some not quite restricted class of phenomena really could be checked. The attempts at a formulation of a variant of predicate logic which could form the basis for an adequate description of the semantics of natural language are still far from being successful. It seems that these attempts would have to repeat, in a sense, the development of modal logic, logic of tense, intensional logic, if they actually were to reach their aim, namely adequacy for the full range of a natural language. Natural language is a universal system (in which everything can be formulated that man is capable to formulate, even though not without ambiguity and vagueness), and one of the aims of the development of modern logic and model theory has been precisely to achieve the possibility of accounting adequately for the semantics of linguistic phenomena.

Certainly, since the period analyzed by Spark Jones and Kay, much has changed with respect to semantics. While their chapter on semantics is almost entirely devoted to questions of lexicon (so that they could state -- p. 139 and elsewhere -- that semantics was not important for information retrieval, at least if questions of term classification are disregarded, see p. 159 and 171), the development of semantics since 1970 concerns first of all the relationships between semantics and syntax. If transformational theory, which is, in a sense, a continuation of the anti-semantic descriptive linguistics (neo-Bloomfieldian structuralism), although in many respects a polemic continuation, meets serious difficulties where attempting an integration of semantics into linguistic description, then it may be useful to look for a better understanding of semantics in other trends of linguistics, and in logic.

First of all, there is the Montague trend, which, as we have already noted, has contributed to a sound treatment of such pragmatical indices (or points of reference) as *I, you, here, now*. We can state briefly that the semantics of these four elements may be accounted for by variables which are free in the semantic representation of a sentence (or in the hyper-sentence inside this representation), while the specific values of these variables -- corresponding to their reference -- are determined directly by every occurrence of the given sentence in a given text and situation (since the speaker and the addressee, as well as the place and the time point of the utterance are determined by the utterance token itself, in the primary form of communication). Other pragmatical indices, such as in *my, we, yesterday*, and probably also *the country, this building*, may be analyzed semantically as based on the four elementary ones. This trend (represented, with respect to linguistic semantics, first of all in the writings of Schnelle and Dahl) also underlies various logically oriented writings (the most detailed analysis of natural language semantics from this point of view has been given by Cresswell). Howver, we shall see later in this section that the neglect of linguistic structuring, which has always been connected with the analyses of natural language semantics by logicians, interferes even here in the understanding of the relationship between the semantic analysis based on truth values and the (linguistically important) study of sense.

Those trends in modern linguistics that base their semantic analysis on semantic insights and attitudes of structural linguistics in Europe for the most part have not yet elaborated correct formal systems. This applies to Lamb's stratificational or cognitive linguistics as well as Mel'cuk's system meaning <=> text and many other approaches. They often work without testable criteria and their semantic classification might be suspect in that it is difficult to check them for completeness and adequacy, to exclude personal attitudes in completing them, etc. This observation applies, of course, also to other semantic and cognitive systems, none of which has been used up to now, as far as I know, by more than a small group of authors. Nevertheless, although these approaches are not mentioned by Sparck Jones and Kay (including Kay's, 1970, own contribution), it follows from their book (pp. 35ff, 71) that transformational description is connected with various drawbacks which other systems do not always share. Having worked on questions of the framework of generative description for almost two decades, I am convinced that at least in two respects it is worth while to look for a sound basis of semantic analysis in these stratificational or functional approaches:

(a) The tradition, based in structural linguistics, of working with diagnostic contexts in searching for semantic equivalence (synonymy) and semantic difference (ambiguity. homonymy) of surface units has led to elaboration of such criteria as the dialogue test (for distinguishing between obligatory "cases" and free adverbials, see Panevová, 1974, Panevová and Sgall, in prep.), and the question test (for topic and comment, see Sgall, Hajičová, and Benešová, 1973, Section 3.2, and also the commentation test of Posner, 1972), and to a better understanding of the test of negation (in distinguishing not only meaning proper and pre- supposition, but also allegation, see Hajičová, 1974; Sgall, Hajičová, and Benešová, 1973, Section 4.2).

These and similar criteria are testable and make it possible to determine the units needed on different levels of language structure. The search for such criteria will spread necessarily with the further research in linguistic semantics. The discussions on semantics in transformational linguistics should be checked from this viewpoint in connection with the distinction between linguistic meaning and factual knowledge. We would like to add a remark on this distinction here.

Not only must the description of syntax be held apart from describing some domain of non-linguistic data, as its long history has shown with overwhelming evidence, but also inside semantics we have to distinguish between linguistic meaning (i.e., the meaning of

linguistic units -- elementary and complex -- as well as rules that allow constructing the meaning of a complex whole from that of its constituent parts) and non-linguistic (cognitive, ontological, factual) content. The latter cannot be described by purely linguistic means. Only the former can be handled in a general way, for a given language in its universality, while the latter must be treated in close connection with the specific task, with the domain of the application. Knowledge of the world must concern either toy blocks on a table, or driving a car, etc., but it cannot cover all of them at once in a single model. Certainly, an area of "general knowledge" can be identified, the content of which may be useful for many or perhaps all types of use of natural language in man-machine communication, but for the time being this area, in some sense intermediate between meaning and content, has not yet been characterized to a sufficient extent.

The issue concerning the distinction between linguistic meaning and cognitive content has been discussed for decades. From Hjelmslev to Dokulil and Coseriu, European structural linguists have maintained that it is necessary not to deal directly with factual knowledge. On the other side, most adherents of transformational description maintain that it is the task of anyone who wants to "insert a level of linguistic meaning" between that of cognitive content and syntax to bring arguments strong enough to justify the complication of the system. But are cognition and meaning really two levels in this sense? Is not meaning just the single form our "model of the world" can have? We shall see in the next paragraph that there are arguments for a "European" answer to these questions.

(b) The systems working with a linear ordering of levels (from semantics to sounds or graphemes) are more directly connected with (and thus, applicable for) the procedures of synthesis and analysis, which are not only immediately needed for the practical purposes of man machine communication, but which also may be understood (with some restrictions) as hypothetical formal models of human linguistic mechanisms.

In some cases it was even possible to show that such a system is connected with a generative power weak enough to characterize natural language as a specific system (see Plátek, 1974, for the Prague functional generative description). It appears most important for the status of linguistic semantics that the sequence of levels from phonetics to semantic representation be seen as the proper domain of linguistics, while the relationship between semantic representation and logical languages or cognitive systems is the domain of cooperation of linguists with logicians, psychologists, and specialists in models of cognitive structures (including semantic networks). It has been shown that Carnap's distinction between intension and intensional structure (directly relevant for the handling of non-extensional contexts, especially of belief sentences), or, in other words, the distinction between truth-functional semantics and linguistic meaning (or sense, in logical terms) can be accounted for in a systematic way, for a natural language, by a description of the functional type, if the semantic representations of sentences are viewed as describing the linguistic meaning (or sense, or intensional structure) of the sentences, while procedures translating these semantic representations into logical and/or cognitive structures specify the truth-functional values of the sentences. (See Sgall et al., in prep., for the relationship between linguistic meaning and Carnap's intensional structure; the translation into a logical language is examined in Sgall, Hajičová, and Benešová, 1973, Section 7.6; Hajičová, Křížek, and Sgall, 1975, and the Appendix here).

This treatment of the semantics of the sentence includes, for instance, such distinctions as that between the two sentences in (1), which appear as fully synonymous (their difference being described by an optional transformation or in some equivalent way), and those in (2) to (4), which differ in their linguistic meaning similarly as those in (5) to (7), although they

do not differ cognitively. The intensions of the two sentences in (2) to (4) are identical, but the difference in the intensions of the sentences in (5) to (7) shows that this is not given by the linguistic structure of those sentences, but only by specific conditions involving, in addition, the lexical or morphological cast or "setting", i.e., the meaning postulates of individual morphemes.

(1) (a) They persuaded Jim not to return into that company.
(b) They persuaded Jim that he should not return into that company.

(2) (a) The old table is yellow.
(b) The yellow table is old.

(3) (a) Tom sold Jack a car.
(b) Jack bought a car from Tom.

(4) (a) John talked about Jack to JANE.*
(b) John talked to Jane about JACK.

(5) (a) Any old table is yellow.
(b) Any yellow table is old.

(6) (a) Tom sells cars to the citizens of Glasgow.
(b) The citizens of Glasgow buy cars from Tom.

(7) (a) John talked about many problems to few GIRLS.
(b) John talked to few girls about many PROBLEMS.

Thus it appears, of course, that linguistic semantics proper is connected with a rather narrow concept of synonymy; many so called paraphrases will turn out not to be fully synonymous, i.e., to take different truth values for different possible worlds or states of affairs. It may be considered superfluous to work with such detailed classification of the meanings of sentences, since in many cases they overlap semantically to such an extent that the distinction between them seems not to be relevant. It can be said, however, that this concept of strict synonymy is inherent to natural language, so that, if "the way in which semantics and syntax are combined" (Walker, p. 82) is considered as one of the main points, it is just this strict synonymy which can be accounted for by linguistic methods. The fact that with a specific lexical cast such differences in linguistic meaning as that between (2)(a) and (b) are not relevant, for the cognitive content, can be accounted for by rules translating semantic representations into cognitive networks (if a single object has two properties, it is not important which of them is structured as the main predicate of the assertion). Similarly, the difference between (3) and (6) will be treated by describing the converse predicates by means of different meaning postulates (or other devices of lexical semantics) which would coincide in case the transaction concerns two people only and is formulated as accomplished.

It may be worth while to devote more place to a similar account of the difference between the pairs (4) and (7), since here the questions of topic, focus, and communicative dynamism (or functional sentence perspective) are involved. These questions have not yet been studied systematically enough in most linguistic frameworks.

If the relationship between "given" and "new" knowledge is well understood, it becomes clear that, if the speaker formulates a declarative sentence and communicates it to a hearer as a statement, in the general case, he instructs the hearer to pick up certain "established items" of information (i.e., items that have been already contained in his memory, primarily in a part of it that has been already activated by the given discourse

----
*Capitals denote the intonation centre here.

112

and its situation) and to connect them with items presented as new (or not directly recoverable), i.e., to modify them somehow, place them in new relationships with other items or with each other, etc.*

As linguistic considerations have shown, these two kinds of items involved in a sentence are divided just in accordance with the dichotomy of what Chomsky now calls presupposition and focus (only his formulations must be made more precise in some points, see Hajicova and Sgall, 1975). Thus, in the situation of Winograd's dialogue with his robot, if, after point 13, p. 11, the machine is told "I own blocks which are not red", the words *I* and *own* are in the topic here (since the sentence "The blue pyramid is mine" immediately precedes), they are used as referring to items of information the hearer (machine) shares with the speaker (man) in the activated part of memory. They are used as recoverable, as "established" items *about which* something is communicated, i.e., which are to be modified somehow. This somehow is specified by the rest of the sentence, i.e., by its focus: the NP *blocks which are not red* refers to information not exactly "new" (the machine "knows" that there are such blocks in its "universe"), but not directly recoverable (in Halliday's terms), or, more exactly, to information which is to be added to the topic in the hearer's memory. It is defined here what objects "are owned by the speaker", and it can be checked that this is decisive for the action performed by the machine while "understanding" this sentence. This checking should include the fact that the focus includes an "exhaustive list" (see Kuno, 1972) of items standing in the stated relationship to the topic (exhaustive down to a certain threshold, in the general case). But this holds only for cases in which the verb belongs to the topic; in the primary case the verb belongs to the focus (to what is stated about the subject, etc.), so that there is no such exhaustiveness. If, e.g., at the given point the machine were told "the red pyramid is supported by a green block", then the first NP would constitute the topic, the rest being the focus of this sentence. That is, the machine would be instructed to put the established item *the red pyramid* in the relationship *be supported by* to the item *a green block* (or, more exactly, to check, whether this relationship really is new -- which it is not, in the given case, or whether it is compatible with the knowledge the machine has, and only if after these steps it proves useful, to add the new information to the given stock of knowledge).

In the functional generative description of language, this aspect of the declarative sentence is accounted for by its semantic representation having the form of a dependency tree exemplified in Figure 1. There the superscript b with the main verb denotes that the verb belongs to the topic, while the expanding elements (participants as well as free adverbials) stand to the left of their governor if they belong to the topic, and to its right, if they belong to the focus. With embedded clauses, the concept of topic is thus relativized in a straightforward way. A procedure for generating such trees has been formulated, and it has been shown that no excessive generative power is needed. It is not always easy to find the boundary between topic and focus when parsing a sentence, but the basis of such an analysis may be formulated according to the transformational rules for focus, as given in Hajicova and Sgall (1975). Such rules reflect not only the dichotomy, but also the whole hierarchy of communicative dynamism (or "deep word order").

It is well known today that the topic/focus dichotomy as well as the scale of communicative dynamism are semantically relevant, in the general case; the examples (4) and (7), taken over from Lakoff, illustrate this fact. It is possible, in the context of Winograd's study, to point out that the sentence (8)(a) should be interpreted in another way than (8)(b).

----

*For a more detailed discussion see Sgall, Hajičová, and Benešová (1973), especially pp. 10ff., 70-73, 158ff, 251ff.
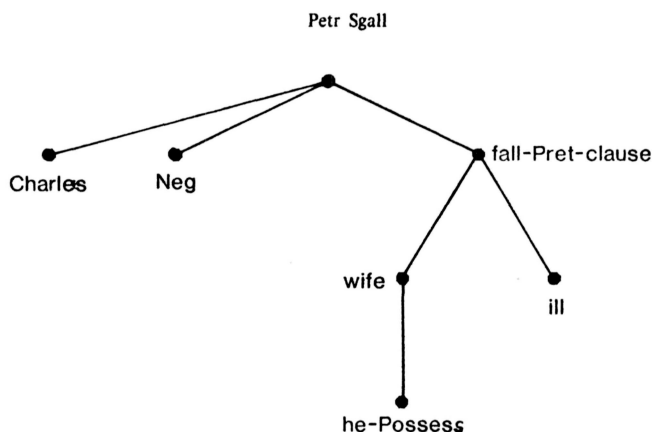
113

Petr Sgall

Fig.1 -- **Charles didn't come since his wife fell ill**

(8) (a) Now cubes are mostly in the BOX.
(b) Now in the box there are mostly CUBES.

With our approach, these distinctions are accounted for by the scale of communicative dynamism, i.e., by the order of the elements of the corresponding semantic representations, which is correlated with the order of quantifiers in a logical formula. (For the translation procedure yielding the mentioned correlation, see the Appendix and the references quoted there.)

The dichotomy of topic and focus is also connected with the organization of human memory, or of the stock of knowledge (and other kinds of psychological phenomena). Human memory is a vast domain, structured in various ways, and if an act of communication is to be effective, the understanding of the message should not require more than a minimal effort on the part of the hearer. Only some elements of his memory are foregrounded by the situation of the discourse, and the required effort is smaller if some of these elements are chosen as the established items by the speaker and if the lexical units referring to them are marked as such, being (primarily) placed at the beginning of the message. Point after point, the message can be expanded; or, in other terms, the communicative act consists in a structure of messages linking with each other. Thus, in the unmarked case, the established items are referred to prior to the specification of their desired modification.

If the sentence of a natural language is considered the systemic form of an elementary communicative act, its structure may be expected to reflect the basic conditions of communication. This standpoint then allows us to understand why the sentence includes, besides the syntactic patterning (consisting in the hierarchy of verbs and their

participants, with their inner structure, the finite verb of the main clause representing the central point of this hierarchy), also a communicative patterning, in which the parts referring to the established items (in the above sense) are distinguished from the parts concerning the modification the speaker has in mind (the added information, in the ideal case). Thus, when formulating a sentence, the speaker has to choose the main predicate and the topic (the established items).

In the elementary case, the two hierarchies coincide (in such sentences as *Jack SLEPT* and *Mary LIVES* the coincidence is most complete), and they were also put together in the Aristotelian formulations of linguistic and logical structures. During the development of logic and linguistics they were not only held apart, but the second of them had been neglected almost completely for long centuries. Only in the last decades has it been studied in a more systematic way, and attempts been made to investigate systematically the interplay of the two hierarchies in cases that are not so elementary as the above examples. In these more complicated cases, the established items are not always identical with the subject of the sentence (they may include the subject, but also other parts of the sentence, or may even lie outside the subject), the verb need not specify the modification wanted by the speaker completely (it may even refer to some of the established items, if states, activities, etc., are referred to as already known, etc.), the verb may have more participants (with a free choice of those that refer to the established items), some of them may contain other verbs with their participants, etc.

Thus it is advisable to distinguish the part of the sentence referring to the established items also terminologically (as topic, or theme) from the other part (specifying the desired modification in the above sense, and called comment, rheme, or focus). In other terms, the topic may be called the contextually bound part of the sentence (with embedded sentences, of course, we come to a whole hierarchy, so that it may be advantageous to hold the terms apart for the sake of a more detailed classification). Contextual boundness does not mean co-textual here, since not only are items known from the previous portion of the given text included, but also those given by the situation of the discourse. Thus, it is necessary to use, in the description of the structure of sentences, some pragmatic data concerning the stock of knowledge shared by the speaker and by the hearer(s). The elements of the stock of shared knowledge (more exactly, the stock the speaker himself has and supposes to be shared by the hearer(s), too) should be classified according to the degree to which they are foregrounded (activated) in the situation of the given discourse. This means that besides a (generative or other) description of language itself, a description of the functioning of language in communication (a description of linguistic performance) must contain also another mechanism, describing the stock of knowledge. Inside this stock, the elements (or at least some of them) are partially ordered in such a way that the ordering relation may be interpreted as a scale of foregrounding.* Some of the elements of the stock of shared knowledge are, so to say, permanently foregrounded, and thus the speaker can use them as contextually bound, and also as presupposed, at any stage of a discourse. It seems that, first of all, the indexical elements *I, you, here, now* belong there, but also other notions closely related to them (*my mother, my wife, my children, my country, my town, this year, this month, today,* etc.), but perhaps not all nouns of unique reference in the

----

*The parallel between the whole stock of shared knowledge, as distinct from its (most) foregrounded elements, and permanent memory, as distinct from temporary memory, is evident and calls for experimental checking (cf. also Chafe, 1973, and the literature quoted there.) We may note only quite briefly that it is this stock of shared knowledge, which should probably be used instead of such unclear notions as "universe of discourse" or "actual world", if semantic research is to be combined with a constructive approach.

universe of discourse (cf. Kuno, 1973, p. 39). Other nouns, the reference of which usually must be specified afresh for every discourse (or even for a certain part of the discourse), can be foregrounded by this very specification (or, if their referents attract the attention of the participants of the discourse, also by deixis).

This mechanism would account for the possible use of *My wife has read, in a German weekly, that...* as a beginning of a discourse even in a situation where the speaker's wife is not present, and not known to the hearer. As we have said, it would be necessary to work with a hierarchy of the elements of the stock of knowledge, using at least a partial ordering, since, of course, *my aunt, your mother, my third grandson, the Old Town of our capital, last century, Paris*, etc., are noun phrases very suitable for the use inside the topic of a discourse opening (being connected with at most trivial presuppositions), but, under certain circumstances, *my teacher's niece, Aconcagua*, or *the age of Michelangelo* also would do.

The foregrounded elements can be mentioned, in a discourse, in two different ways: (a) as contextually bound, and (b) in the focus, along with elements that have not yet been foregrounded. Compare *Charles saw HIM* with the object foregrounded (and therefore pronominalizable) but included in the focus (and therefore stressed) against *Charles SAW him*, with the object not only foregrounded (known) but also contextually bound (therefore unstressed). Therefore, the topic cannot be identified with the known, given (or foregrounded) elements.

If an element of the stock of shared knowledge occupies a relatively low position in the scale of foregrounding, it can be introduced in a discourse as contextually non-bound, i.e., in the focus of a sentence. Such a mentioning gives the element a higher degree of foregrounding; in some respects it is possible to conceive the element mentioned last to be more foregrounded than the elements mentioned before, the foregrounding of which already shades away step by step (if they do not belong to the permanent part of the foregrounded elements).

The hierarchical organization of the stock of shared knowledge is relevant also for an investigation of the structure of text. It may be assumed that any break in the fluent line of discourse or text is connected with a more or less considerable change in the set of (most) activated elements of the stock of shared knowledge. The extent of this change depends on whether we are concerned with a simple pause between two paragraphs (after which an issue may be re-evoked that was not mentioned immediately before), with a deeper break (e.g., between chapters), or even with a pronounced discontinuity of the text, evoked by an outside interference into the discourse. In the first two cases, the change concerned might be characterized as a switch of a common theme or "hypertheme". These notions, which must be understood as relativized or stretched along some scale, could be analyzed much more explicitly than up to now, if the interplay of topics (themes) of utterances, as identified, e.g., by the question test, were understood as reflecting the changes in the degree of activation of the corresponding elements in the stock of shared knowledge. Typically, an element is activated by its first mentioning in a rhematic position. By this very fact it becomes available as a possible thematic element in the following part of the text, and, if the speaker switches to another theme afterwards, the activation of the given element is reduced. After the newly chosen theme has been "exhausted", or saturated, it is possible to return to the former theme again, but this possibility is restricted. If it was the theme of the first utterance of paragraph A, it can well emerge again at the beginning of paragraph B, but if another theme is chosen here, the degree of activation of the original theme is again reduced, etc.

It seems, from this point of view, as we have already remarked, that often the element mentioned last (i.e., the focus of the last sentence of the preceding portion of the text) can be conceived as more foregrounded than the elements mentioned before, the foregrounding of which already shades away. This would point to the possibility that the part of human memory corresponding to the foreground of the stock of shared knowledge could be described by a device to some extent similar to a pushdown automaton (but, certainly, restricted to a finite storage). Some examples corroborate the view that an item that has been mentioned later carries a smaller degree of communicative dynamism than another element that was mentioned at some earlier point of the discourse, if in the present, i.e., repeated occurrence they both are contextually bound.

In almost every text, of course, the situation is complicated by various factors. In a text having the qualities of a work of art, there are deviations of different types possible. In a technical text, the topic is, as a rule, rather complex, consisting of a relatively large number of items activated partly by the text belonging to a certain domain of knowledge, and partly by the relationship between technical terms known in this domain.

The relevance of the phenomena now described as topic and focus, connected with the semantics of word order, quantification, and negation (see also the remarks by Sparck Jones and Kay, pp. 125 and 185) was one of the main linguistic topics of the 1973 Conference of Formal Semantics in Cambridge, U.K.* Also in some of the contributions at the International Joint Conference on Artificial Intelligence in Tbilisi, 1975, considerable attention was devoted to these questions.** An approach using the concept of stock of shared knowledge with its structure as briefly outlined above might be useful not only in that it allows for handling meaning in terms of procedures (cf. the discussion of Winograd's work above and also Walker, p. 82), but, most of all in serving as a basis for referential semantics in a sense similar to that of Winograd (pp. 134; 168ff). A noun phrase connected with the label ("grammateme") "definite" or "specific" (which should be present in the semantic representations of sentences) is understood as referring to an object that is represented in the stock of shared knowledge. If this noun phrase has been used in the topic of the utterance, its referent must be looked for among the foregrounded elements of the stock of knowledge. The degree of communicative dynamism carried by the noun phrase shows in any case whether the object referred to is more or less foregrounded. These and other aspects connected with the proposed account of topic and focus lead to a language sharing such advantages of some logical languages as the existence of general inference procedures, but lacking the drawback of these languages in that their assertions do not indicate how they should be used; see the Appendix.

Linguistic Aspects of Fact Retrieval

According to Damerau (1976, p. 1), linguistics has not had much influence on language processing in fact retrieval systems. However, the (case) frames Damerau so often quotes did come from linguistics (the linguist Fillmore formulated his case theory following some insights of European structural linguistics). Furthermore, the question Damerau raises on p. 6 -- concerning the possibility of using frames not only in implementations

----

*See for instance the paper by Vennemann in Keenan (1975); a very important aspect of the semantics of discourse is analyzed by Isard in the same volume.

**See especially the contributions by Mylopoulos et al. (1975) and Vayncveyg (1975).

exploring toy worlds, but also in those corresponding to realistic situations -- concerns basic linguistic issues. If implementation should concern a system the empirical adequacy of which was, to a certain extent, checked before actual programming, then such a system, corresponding to realistic situations, hardly can be classed as not being linguistic. Schank and Winograd have derived their systems from quite specific linguistic theories. Probably Stevens and Rumelhart (quoted by Damerau, p. 9, as having presented a study "of syntax rather than semantics") also have been inspired by linguistic studies in syntax (and semantics). In other words, the impact of linguistics in fact retrieval cannot be ignored, and how could it, when the *content* of a document, the *meaning* of utterances, the *understanding* of texts are involved.

What matters, however, is the fact that theoretical linguistics is not often used directly (or adequately) in fact retrieval, which appears, up to now, as a rather isolated, although rapidly growing branch of linguistic engineering. We have already discussed some causes of this lack of contact, especially those connected with the lack of due understanding of the importance of computational "applications" on the side of theoretical linguists. But, on the other side, as Sparck Jones and Kay have noted (p. 198), at least in one aspect the onus is on the documentalist, who should provide a proper specification of what he needs to satisfy his retrieval objectives. Up to now it does not seem that an adequate (and sufficiently general) specification of these objectives -- either for fact or for document retrieval -- has been achieved. And, it is the specialist in fact retrieval, who should be able to use (at least, by means of a team cooperation) recent findings of linguistic theory, even if they have not yet found their way to the academic curriculum. If Damerau (p. 25) draws an analogy between some of the artificial intelligence systems and the dreams of alchemists, he may be right, but with these systems, it is linguistic semantics that is lacking, and some of its issues are by far more accessible today than those of chemistry were for the alchemists some three or four centuries ago. In any case, the value of any solution of a general problem in fact retrieval will depend, first of all, on the universality of the solution. If the given approach is adequate and can be used without specific restrictions (i.e., if the conditions for its effective use have been stated already), then its detailed elaboration is more or less a mere matter of routine. In the other case, if we do not yet know whether the approach can be transferred from the restricted domain of a given experiment to a universe of realistic situations, i.e., if we do not know where it is possible to apply our approach, there always is the danger that sooner or later it will be necessary to resort to ad hoc solutions which might not be free of the risk that the approach as a whole would be invalidated by them. And, furthermore, in the domain of meaning, content, and understanding natural language, it is precisely linguistic semantics that should take the responsibility of the universality of the solutions. We attempted to show in the last section that there are good reasons to state that linguistic semantics is prepared to bear this responsibility.

The universal solution is certainly not the cheapest, in the general case, and it might be prudent not to use all of its force where this is not needed. To apply this way of reasoning, however, one must know first the universal solution, and second, what parts or aspects of it are not necessary for the given task. Therefore, it would be more advisable to experiment with more general procedures and systems, which should be reduced then to yield effective and economical models for commercial use. Unfortunately, the world is not organized wisely enough, and one of the main goals of almost each of the scattered experimenting groups is to show in a relatively very short time that it is possible, using a very limited amount of preliminary work, to construct a practically useful system.

One of the main problems of fact retrieval seems to consist in what has been characterized by Sparck Jones and Kay (p. 44) as decoupling. When higher types of information systems are concerned, the task is not only to separate the language used to communicate with the core of the system from the form of the data store and from the logical operations needed to manipulate it (since languages may change, while the data should remain unchanged). In addition, the task is to find for each of them its proper

characteristics, including features that would make the given language as effective as possible. One of the main conclusions of the developments of linguistic semantics, as summarized in the last section, consists in the advantages of having a specific language -- the set of semantic representations -- between the natural language (serving as the input and output language of the whole system) and the language of the stored information. The latter language may well have the form of cognitive or "semantic" networks, but it appears difficult to get rid of such drawbacks as those quoted, in connection with these networks, by Damerau (p. 4) from Woods (the lack of distinction between extensional and intensional, as well as asserted and hypothetical nodes, etc.), if the corresponding items in surface sentences, which may be used as cues to such distinctions, are left without counterparts in the inner language of the system. It appears that the use of a more sophisticated linguistic framework, and also of a more reliable psychological basis (as is present in the cognitive networks of Hays, 1975) could be useful in these respects. Also the distinction between the bulk of the data and its foregrounded part (in the sense of the previous section) should be provided for. The language of logical operations (or, in more modern terms, of the brain of the system) might be connected with the existence of general inference procedures and other advantages of logical languages. Here we are in a better situation than in the domain of linguistic semantics, since logical systems have been studied much more intensively. Thus, even if "special natural deduction systems" are proposed for information systems (cf. Damerau, quoting Reiter, p. 7), it can be hoped that the condition we have already mentioned will be satisfied. In using a special system for a restricted purpose, it can be checked relatively easily what has been left out from a universal system that is adequate without the special restrictions.

We can give here (in the Appendix) only a quite short characterization of the main features of a system of fact retrieval using the level of semantic representations and their translations into a logical metalanguage which may be used by the brain.

Linguistics and the Future of Programming

The overall aim of language processing is to make the computer understand human language as such. It is not necessary to describe formally all the variation of language according to style, local dialects, affected speech, etc., since most of this variation is not likely to be used when computers are addressed.

To put it quite simply, man-machine communication presupposes that one of its participants has learned the other's language. This burden, hitherto carried by (computer) people, should be passed over to the computer alone.

If programming numerical tasks as well as other tasks not requiring a regular communicative interaction between man and machine are left aside for the moment, we might state that the direct use of computers by specialists in other branches than computer science is not possible without such a development. What is needed, is a compiler for English and other natural languages. In any case, the significance of linguistics for the future of artificial intelligence is connected with making programming languages more and more human, i.e., making them include reasonable and expandable subsets of natural languages.

Not only is programming as such involved, since if a free conversation with the computer were possible to a certain extent, many different degrees of man-machine cooperation would emerge, some of them being more akin to the formulation of a program, others to its use. As for the necessary and possible freedom in the use of natural language, two extreme cases could be left out of consideration. On the one hand, the standard tasks that are typical for certain areas of the current use of computers, like numerical tasks or document retrieval, do not require a free use of natural language. On the other hand

it may be assumed that the "informational explosion" and the automation of its handling will be connected with some requirements, concerning a standardization of their linguistic means, upon the authors of texts in this or that branch of technology, so that a "full" formalization of natural language can still remain beyond our dreams.

Thus it is possible, as for the domain of fact retrieval, to imagine a situation where the authors writing about, say, electronics, or chemistry, etc., will be advised not to use sentences longer than twenty words, the pronoun *which* without a preposition, the conjunction *as* in positions where *because* can be used instead, sequences of more than two noun (or prepositional) phrases without an intervening verbal form, etc.* Otherwise their texts will not be well understood by the system designed to construct and update the automatic encyclopaedia of the given branch of technology, and customers using this system as a source of their information probably will not find and use the author's results.

Certainly, it will not be necessary for the authors of texts themselves to take care of respecting the standardizing prescriptions. They will be aided by a more or less drastic preediting since, in the first period of the foreseen epoch, the standardization rules may be somewhat complicated and some of them may require specific linguistic knowledge. But the functioning of such systems itslef will yield more and more rough material and impulese of various kinds (frequent types of errors at different levels, contextual environments permitting criteria for their solution to be found, etc.) so that it will be possible to complement the parsers or compilers to such an extent that preediting will eventually become superfluous. Some restrictions on the authors may even then be necessary, but these will be comparable to the requirements of standardization in other fields. Thanks to their usefulness, they probably will not represent such a large burden for the authors as the orthographic conventions of our times do.

It is already possible now in formulating research programs to reckon with the fact that the main attention should be concentrated on the syntactic and semantic analysis of a properly chosen subset of the natural language, which soon might be able to occupy the place of programming languages. Several contributions illustrating this development are being prepared for the International Conference on Computational Linguistics in Ottawa, July 1976.** These tendencies certainly will be supported by such factors as the increasing accessibility of man machine interaction, in the course of which much of the ambiguity and vagueness of natural language can be treated (thus yielding new material for further developments in parsers).

Another aspect of the relationship between programming and linguistics consists in such approaches to automatic programming as Heidorn's (quoted by Damerau, p. 22), where natural language is used in the input to describe a problem (a solution of which is to be programmed), and also in one of the outputs, in which this problem is rephrased by the computer. As Damerau remarks, if such a system is tested with a real class of users, its uti ity may be evaluated; and, we may add, such testing could be very important for linguistic research. Every system using natural language, if used on a larger scale, will bring new possibilities for the refinement of linguistic frameworks.

As Walker (p. 74) states, the most significant effect on linguistics in the long term will result not from computer use oriented directly toward linguistic aims, but from systems of question answering or computer understanding. The interplay of factual knowledge and of

----

*Or, to put it in a more practical way, such a sequence would always be analyzed as in "the drawer of the table in the room of my secretary", rather than "book of adventures by Stevenson in my library".

**See especially the contributions by Burton and Woods on a compiling system for augmented transition networks, by Schlesinger, and by Landsbergen.

linguistic competence, the different aspects of language use as changing the information stored in memory, and also the ways of acquisition of language can be investigated by means of the study of such systems. Experiments with question answering systems and with programming in natural languages will be of major importance for theoretical linguistics. In this sense, the future of linguistics is closely connected with computer science and information systems; but also the future of the latter disciplines cannot be imagined without linguistics.

Appendix

Let us describe, as an example, the programme for man-machine communication being prepared, step by step, by the group of algebraic linguistics of Charles University in Prague. The algorithms being prepared can be divided into the following groups:

(i) Grammatical analysis of input texts (i.e., of Czech and English texts on electronics), yielding unambiguous semantic representations of subsequent sentences (in the form of linearized dependency graphs).

(ii) Algorithms translating from the semantic representations (which still have a linguistic character) into a logical language, well adapted to the purposes of the theory of inference, etc.; in this form the information gained from the natural language input is to be stored in the computer memory (and confronted with further such information, gained later from other texts).

(iii) Algorithms looking for information needed according to questions formulated by the users in Czech and English and analyzed according to the algorithms of (i).

(iv) Algorithms for the synthesis of answers in Czech and English.

We want to characterize the solution of two problems that appeared crucial for (i) and (ii), as well as for the form of the semantic representations (SR's) of sentences (the language serving as the output of (i) and input of (ii), among other purposes).

The chosen linguistic approach is connected with the following form of SR's (where every A stands for a lexical unit, accompanied, possibly, by markers of "morphological meanings" (grammatemes), such as Plural, Preterite, etc., the subscript denoting the function of the given item in the sentence: Actor, Objective, Dative, Place, Direction, etc.):

(1) $V (A_1^b, A_2^b, \ldots, A_j^b, A_{j+1}, \ldots, A_n)$

Each of the participants, denoted here as A_x, can of course itself consist in a g roup headed by the verb, so that a structure of the shape (1) may be embedded, under certain conditions (specified by a generative grammar using dependency syntax) into another such structure. The superscript b is attached to those elements that are contextually bound, i. e., included in the topic; the others constitute the comment or the focus of the sentence.*

The topic/focus dichotomy is treated here along the lines we have already characterized above for the empirical background; this also concerns the position of the negation operator. If one takes into account also the delimiting features (such as Definite,

----

*This tectogrammatical level, the set of SR's of this shape, could serve also as an interlingua for automatic translation; certainly, in many details of the lexicon and also in some features concerning the grammatemes and other items of the syntax of this level, languages differ. It is necessary to treat such differences in a way similar to that in which idioms are treated in binary translation. For a relatively detailed characterization of the tectogrammatical level, cf. Sgall et al. (1969), Klein and Stechow (1974), Sgall et al. (1975).

Indefinite, Specifying) the shape of an SR as given in (1) should be complemented in the following way:

$$(2) \quad V_{m_0} ((P_1 \; a_{m_1})^b_{k_1}, (P_2 \; a_{m_2})^b_{k_2}, \ldots, (P_j \; a_{m_j})^b_{k_j}, (P_{j+1} \; a_{m_{j+1}})_{k_{j+1}} \ldots, (P_n \; a_{m_n})_{k_n})$$

where the indices of the form m_i specify lexical units, the indices of the form k_i specify the "cases" (participants, syntactic functions) and P_i stands for one of the operators SPEC, DEF, INDEF, EV, FEW, MANY, i. e., for delimiting features (understood here in the sense of Bierwisch; for a discussion cf. Krizek, 1973; for the purpose of our present discussion, we also class here *every*, *many*, and *few* as delimiting features). The formula (2) represents an SR of a sentence with a contextually non-bound verb; if the verb is contextually bound, then V is attached a superscript b.

To formulate a procedure translating (2) to the predicate calculus, we have introduced a new operator $St \; x$ (read as: "for x such that ..."). St is an operator with a free number of arguments; in its nature, it is close to the epsilon operator but it is generalized for a greater (arbitrary, but finite) number of name arguments having a set character. We write $St(x) \; (F(x))(G(x))$, to be read "for any x for which $F(x)$ holds, $G(x)$ holds", which can be compared with Russell's $G(x(F(x)))$, and x stands for a sequence of set name variables (where no ambiguity can arise, we do not distinguish the name of an element from that of the set having only this one element).

The first step of this translation procedure, determined by Rule T, depends on whether the verb occurring in the given SR is or is not contextually bound.

> Rule T: If the SR has the shape (2), it is rewritten as (3); if the SR differs from (2) only in that V has the superscript b, the SR is rewritten as (4):
>
> $$(3) \quad St \; (x_{k_1}, x_{k_2}, \ldots, x_{k_j}) \; (a_{m_1} \; (x_{k_1}^{P_1}) \; \& \; \ldots \; \& \; a_{m_j} \; (x_{k_j}^{P_j})) \; (St \; (R, x_{k_{j+1}}, \ldots, x_{k_n})$$
> $$(V_{m_0} \; (R) \; \& \; a_{m_{j+1}}(x_{k_{j+1}}^{P_{j+1}}) \; \& \; \ldots \& \; a_{m_n} \; (x_{k_n}^{P_n}))(R(x_1, \ldots, x_n)))$$
>
> $$(4) \quad St \; (x_{k_1}, x_{k_2}, \ldots, x_{k_j}, R) \; (a_{m_1} \; (x_{k_1}^{P_1}) \; \& \; a_{m_2} \; (x_{k_2}^{P_2}) \; \& \; \ldots \; \& \; a_{m_j} \; (x_{k_j}^{P_j}) \; \& \; V(R))$$
> $$(St \; (x_{k_{j+1}}, \ldots, x_{k_n}) \; (R(x_1, \ldots, x_n)) \; (a_{m_{j+1}} \; (x_{k_{j+1}}^{P_{j+1}}) \; \& \; \ldots \; \& \; a_{m_n} \; (x_{k_n}^{P_n})))$$

Note: It would be more accurate if the superscript denoting the delimiting operator that binds the variable x_{k_i} ($i = 1, 2, \ldots, n$) were included also in the first parenthesis following the symbols St or R; but no misunderstanding can arise here.

Afterwards we eliminate, step by step, the occurrences of the symbol St (for a technical presentation of the whole procedure, cf. Sgall, Hajicova and Benesova, 1973, pp. 199-203) in such a way that it is replaced by one of the usual quantifiers (determined by the given delimiting feature). After other modifications stated in an algorithmic form, a formula of the second order predicate calculus results (see the examples in the quoted book; also Hajicova et al., 1975).

If an SR contains the symbol Neg, Rule T is applied first, disregarding Neg; afterwards,

(a) If Neg has not been assigned the superscript b, then the symbol NON (for logical negation) is written to the left of the translation of the string $G(w)$ in the formula representing the result of the application of T to the given SR.

(b) If Neg^b is present in the translated SR (with V^b), then NON V instead of V is written in the resulting formula.

The elimination of the operator St leads to the common form of formulae of predicate logic. If this operator (defined for instance by means of an axiom) is retained, one of the drawbacks of predicate logic, well known in the domain of artificial intelligence (as Winograd also quoted in his study), can be removed. The formula in its common form does not point out how it should be used, while the corresponding representation using the operator St denotes explicitly the distinction (and the boundary) between topic and focus. (The importance of this from the viewpoint of his "imperative form" of semantic representations, Winograd clearly shows in the case of input questions, as we have already seen.) Thus we are informed directly what parts of the information concerned should be only identified (in the activated part of the robot's storage), and what parts of the SR should be used to modify that "given" information.

To be even more specific, let us quote an example from a logical discussion, where this drawback of the formulae can be illustrated (this example is far from belonging to those in which the quoted drawback could lead to real difficulties). If, e.g., Suppes characterizes P, the set of rules of a context-sensitive grammar, by P being included in V_N^* x V^+ (where $V^+ = V^* - \{0\}$), and the corresponding set connected with a context-free grammar by P being included in V_N x V^+, then the reader must compare all the symbols of both formulae to find the difference. If, instead, the latter formula is written in a shape corresponding not just to "P is a part of the Cartesian product of V_N and V^+", but to "(here) it is V_N, the Cartesian product of which with V^+ includes P", the symbol V_N is unambiguously characterized as the (only part of) focus of the given assertion, and, in this way it is pointed out that the relevant distinction concerns this symbol and its counterpart in the former formula. The identification of this counterpart is much easier than a general search for the relevant distinction, which would be necessary if the common shape of the formulae were used. Clearly, with more complicated (sets of) assertions, differences of this kind would be much more pronounced.

Thus, our example shows how it is possible to translate sentences into representations which are close enough to predicate logic for its laws of inference, etc., to be applicable almost directly, but which lack some of its drawbacks.

References

Allerton, D. J. "The Sentence as a Linguistic Unit." *Lingua*, 1969, 22, 27-46.

Chafe, W. L. "Language and Memory." *Language*, 1973, 49, 261-281.

Cresswell, M. J. *Logics and Languages.* London, Methuen, 1973.

Dahl, Ö. On Points of Reference. *Logical Grammar Reports*, 1972, 1.

Daneš, F. "K vymezení syntaxe" (On the Specification of Syntax). *Jazykovedne Studie*, 1951, 4, 41-45.

Damerau, F. J. "Automatic Language Processing." In Williams, M.E., ed., *Annual Review of Information Science and Technology*, Volume 11. Washington, D.C., American Society for Information Science, 1976.

Hajičová, E. "Meaning, Presupposition, and Allegation." *Philologica Pragensia*, 1974, 17, 18-25.

Hajičová, E., Křížek P., and Sgall, P. "A Generative Approach To Semantics." *Linguistica Silesiana*, 1975, 1, 17-31.

Petr Sgall

Hajicova, E, and Sgall, P. "Topic and Focus in Transformational Grammar" *Papers in Linguistics*, 1975, 8.

Hays, D. G. *Cognitive Structures.* State University of New York at Buffalo, 1975.

Isard, S. "Changing the Context." In Keenan, E.L., ed., *Formal Semantics of Natural Language.* London, Cambridge University Press, 1975. Pp. 287-296.

Katz, J. J., and Fodor, J. A. "The Structure of a Semantic Theory." *Language*, 1963, 39, 170-210.

Kay, M. "From Semantics to Syntax." In Bierwisch M., and Heidolph, K. eds., *Progress in Linguistics.* The Hague, Mouton, 1970. Pp. 114-126.

Kay, M. "The MIND System." In Rustin, R., ed., *Natural Language Processing.* New York, Algorithmics Press, 1973, 155-188.

Keenan, E. L., ed., *Formal Semantics of Natural Language.* London, Cambridge University Press, 1975.

Kirschner, Z., and Buráňová, E. "A System of Automatic Indexing." (Paper presented at the Prague Conference on Language Processing, 1975.) *Nauchno-Texnicheskaya Informaciya*, in press.

Klein, W., and von Stechow, A., eds. *Functional Generative Grammar in Prague.* Kronberg/Taunus, 1974.

Křížek, P. "Towards a Formal Account of the Semantics of Noun Phrases." *Prague Bulletin of Mathematical Linguistics*, 1973, 20, 43-58. [Also in Klein, W., and von Stechow, A., eds., *A Functional Generative Grammar in Prague.* Kronberg/Taunus, 1974. Pp. 105-124.]

Kuno, S. "Functional Sentence Perspective." *Linguistic Inquiry*, 1972, 3, 269-320.

Kuno, S. *The Structure of the Japahese Language.* Cambridge, Massachusetts, 1973.

Lakoff, G. "On Generative Semantics." In Steinberg, D.D., and Jakobovits, L.A., eds., *Semantics.* Cambridge, Cambridge University Press, 1971. Pp. 232-296.

Lamb, S. M. Outline of Stratificational Grammar. New York, 1966.

Lewis, D. "Semantics." In Davidson, D., and Harman, G., eds., *Semantics of Natural Language.* Dordrecht, Reidel, 1972. Pp. 169-218.

Mel'čuk, I. A. *Opyt teorii lingvisticheskix modeley 'smysl <=> tekst'.* Moscow, 1974.

Montague, R. *Formal Philosophy: Selected Papers of Richard Montague.* Edited by R.H. Thomason. New Haven, Yale University Press, 1974.

Mylopoulos, J., Cohen, P., Borgida, A., and Sugar, L. "Semantic Networks and the Generation of Context." In *Advance Papers of the Fourth International Joint Conference on Artificial Intelligence*, Volume 1. Tbilisi, 1975. Pp. 134-142.

Panevová, J. "Verbal Frames in Functional Generative Description." *Prague Bulletin of Mathematical Linguistics*, 1974, 22, 3-40; 1976, 24, 17-52.

Panevová, J., and Sgall, P. "Operational Criteria for Distinguishing Cases From Free

Adverbials." *International Review of Slavic Linguistics*, in press.

Petrick, S. R. "Semantic Interpretation in the REQUEST System." In Zampolli, A., ed., *Computational and Mathematical Linguistics*. Proceedings of the International Conference on Computational Linguistics, Pisa, 27 August - 1 September, 1973, Volume 2. Firenze, Olschki, in press.

Pitha, P. "K vymezovani rozsahu gramatik" (On the Specification of the Scope of Grammars). Slovo a Slovesnost, 1967, 28, 1-6.

Plátek, M. "On One System of Sets of Languages Close to Context-Free Languages." *Teorie a Metoda*, 1974, 6, 103-120. [Also in Klein, W., and von Stechow, A., ed., *A Functional Generative Grammar in Prague*. Kronberg/Taunus, 1974. Pp. 105-124.]

Plath, W. J. "Transformational Grammar and Transformational Parsing in the REQUEST System." In Zampolli, A., ed., *Computational and Mathematical Linguistics*. Proceedings of the International Conference on Computational Linguistics, Pisa, 27 August - 1 September, 1973, Volume 1. Firenze, Olschki, in press.

Posner, R. *Theorie des Kommentierens*. Frankfurt/Main, 1972.

Schank, R. C. "The Conceptual Analysis of Natural Language." In Rustin, R., ed., *Natural Language Processing*. New York, Algorithmic Press, 1973, Pp. 291-309.

Schnelle, H. *Sprachphilosophie und Linguistik*. Reinbek bei Hamburg, 1973.

Sgall, P. "Zur Frage der Ebenen im Sprachsystem." *Travaux Linguistiques de Prague*, 1964, 1, 95-106.

Sgall, P. "K Programu Lingvistiky Textu" (On the Programme of Linguistics of Text). *Slovo a Slovesnost*, 1973, 34, 39-43. [Also in Klein, W., and von Stechow, A., ed., *A Functional Generative Grammar in Prague*. Kronberg/Taunus, 1974. Pp. 369-381.]

Sgall, P., et al. "Xarakteristika Semanticheskoy Zapisi Predlozheniya." *Problemy Sozdaniya i Razvitiya Mezhdunarodnoy Sistemy NTI*, 1975, 2, 33-49.

Sgall, P., Hajičová, E., and Benešová, L. *Topic, Focus and Generative Semantics*. Kronberg/Taunus, 1973.

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

Sgall, P., Procházka, O., and Hajičová, E. "On the Autonomy of Linguistic Semantics." *Theoretical Linguistics*, in preparation.

Sparck Jones, K., and Kay, M. *Linguistics and Information Science*. New York, Academic Press, 1973.

TAUM. *Rapport du Mois d'Aout, Projet de Traduction Automatique de l'Universite de Montreal*, 1973.

Vayncveyg, M. "Model associyativnoy pamyati ispolzuyushchaya ponyatiyie 'vazhnosti'." *Advance Papers of the Fourth International Joint Conference on Artificial Intelligence*, Tbilisi, Volume 1. Pp. 165-168.

Vauquois, B. "La Traduction Automatique à Grenoble." *Documents de Linguistique Quantitative*, 1975, 24.

Vennemann, T. "Topics, Sentence Accent, Ellipsis: A Proposal for Their Formal Treatment." In Keenan, E.L., ed., *Formal Semantics of Natural Language*. London, Cambridge University Press, 1975. Pp. 313-328.

Walker, D. E. "Automated Language Processing." In Cuadra, C. A., and Luke, A. W., eds., *Annual Review of Information Science and Technology*. Volume 8. Washington, D. C., American Society for Information Science, 1973. pp. 69-119.

Wilks, Y. *Grammar, Meaning, and the Machine Analysis of Language*. London, Routledge and Kegan Paul, 1972.

Winograd, T. *Understanding Natural Language*. Edinburgh, Edinburgh University Press, 1972.

Woods, W. A., and Makhoul, J. "Mechanical Inference Problems in Continuous Speech Understanding." *Artificial Intelligence*, 1974, 5, 73-91.