AFCRL-67-0269

AD652644

SURVEY OF AUTOMATED LANGUAGE PROCESSING 1966

D. G. Bobrow
J. B. Fraser
M. R. Quillian

Bolt Beranek and Newman Inc
50 Moulton Street
Cambridge, Massachusetts

Contract No. AF19(628)-5065

Project No. 8668

Scientific Report No. 7

D D C
RECEIVED
JUN 6 1967
C

This research was sponsored by the Advanced Research Projects Agency under ARPA Order No. 627, Amendment No. 2

April 1967

Stanley R. Petrick

Distribution of this document is unlimited

Prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

ARCHIVE COPY

AFCRL-67-0269

SURVEY OF AUTOMATED LANGUAGE PROCESSING 1966

D. G. Bobrow
J. B. Fraser
M. R. Quillian

Bolt Beranek and Newman Inc
50 Moulton Street
Cambridge, Massachusetts

Contract No. AF19(628)-5065

Project No. 8668

Scientific Report No. 7

This research was sponsored by the Advanced Research Projects Agency under ARPA Order No. 627, Amendment No. 2

April 1967

Stanley R. Petrick

Distribution of this document is unlimited

Prepared for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

ABSTRACT

This report is a survey of Automated Language Processing done in 1966. It is limited in scope to analytical processing of natural language, excluding work in programming languages, speech recognition, and statistical processing of text. It focuses on work aimed at generating and analyzing sentences of a natural language. This survey has four major sections:

The first, on syntactic theory, contains a summary of the principal assumptions underlying work in generative grammar and a report of the most significant developments in theoretical and descriptive work in syntax. The second section, on semantic theory, attempts to provide some dimensions along which we can judge various theories that have been proposed and developed in the literature in 1966. A number of empirical studies related to semantics and psycholinguistics are reported in a third section. Finally, a fourth section discusses various computer systems for manipulation of natural language. These range from systems that support linguistic studies to systems that are attempting to utilize natural language as a communication medium.

TABLE OF CONTENTS

ABSTRACT iii

SECTION I
INTRODUCTION 1

SECTION II
SYNTACTIC THEORY 3

SECTION III
SEMANTIC THEORY 9

SECTION IV
PSYCHOLINGUISTIC STUDIES RELATED TO SEMANTICS 17

SECTION V
COMPUTER SYSTEMS 21

CONCLUSION 31

REFERENCES 33

BLANK PAGE

SECTION I

INTRODUCTION

The area of automated language processing covers, in its broadest sense, any use of computers to process any type of language. This, obviously, is too broad a field to cover in this chapter, so we begin by delimiting our scope. First, only the processing of natural language will be considered, thus excluding work on artificial languages such as programming languages and command and control languages. Furthermore, we are interested here only with analytical (i.e., non statistical) processing of natural languages, thereby excluding most work in automatic indexing, abstracting, content analysis and stylistic analysis. Finally, we restrict our examination to analytic processing of natural language that is represented in its input or output form in normal orthography, thus excluding work in speech analysis and synthesis, and optical pattern recognition. Accordingly, we will focus on the work being performed to generate and analyze sentences of a natural language (usually English) in terms of some grammar, or data base, or both, and then doing something with this analysis - perhaps only printing it out, or at the other extreme, answering questions contained in the input.

This survey has four major sections. The first, on syntactic theory, contains a summary of the principal assumptions underlying work in generative grammar and a report of the most significant developments in theoretical and descriptive work in syntax. The second section, on semantic theory, attempts to

provide some dimensions along which we can judge various theories that have been proposed and developed in the literature in 1966. A number of empirical studies related to semantics and psycholinguistics are reported in a third section. Finally, we conclude with a discussion of various computer systems for manipulation of natural language. These range from systems that support linguistic studies to systems that are attempting to utilize natural language as a communication medium. We have, in general, considered literature from 1966 only, since Simmons survey (106) of last year covered material through 1965.

SECTION II

SYNTACTIC THEORY

The linguistic framework within which almost all of the current wor. in language processing (as delimited earlier) is carried out involves the theory of language developed by Chomsky and others that was introduced in Syntactic Structure (17), and elaborated on in Aspects of the Theory of Syntax (14). A very lucid presentation of this material can be found in "Computer Analysis of Natural Languages" by Susumu Kuno (68). However, let us review quickly the assumptions underlying work in generative grammars of this type. First, the goal of linguists in writing grammars is to state clearly and concisely those facts that a native speaker of the language knows about his language. This knowledge has been called the competence of the speaker of a language; thus a grammar may be thought of as a model for the ideal speaker's competence. Such knowledge of the language is, to be sure, not always obvious. For example, a speaker of English can provide the correct plural ending on any noun, even if it is a nonsense word; few speakers recognize, however, that the different endings they systematically evoke are directly correlated with the final sound of the word, a regularity statable in three simple phonological rules. Furthermore, many of the regularities of a language can be captured only by abstract representations of what are eventually the surface facts.

But to state such regularities of a language is, by itself, not the sole end. A more far-reaching goal consists of finding universal features of all languages, from Chinese to English to Sanscrit, and of determining a maximally simple specification of those universal facts, along with the idiosyncratic features of particular languages. Underlying this approach is, obviously, the assumption that all natural languages share a large common area of similarity, albeit at a rather abstract level. Furthermore, although there is some disagreement on this point, it is generally assumed that there does not exist at present any discovery procedure for determining the most insightful linguistic analysis for a particular part of some given language in terms of a particular linguistic theory. Nor, it appears, is there any reason to assume one will be forthcoming in the near future.

A transformational grammar is intended to be neither a psychological model for the user of a language nor a pedagogical tool. The speaker's competence in a language, those facts about the language that a grammar is intended to capture, is not always reflected in his actual performance, for instance, while he is being affected by noise, interruptions, indigestion, and the like. A theory of performance is an important area of study, but it is not currently being considered by most linguists. They feel that the restrictions that are necessarily placed on the existing language processing systems are completely ad hoc. Consequently, although most grammars that have been used in language processing are generative in nature, beginning with the syntactic portion of a sentence first and then accounting for the semantic and phonological portions, this should be not held as a model for actually producing speech nor for its interpretation by the hearer. Nor should any rules in any grammar to date be thought of as analogous to the actual neurophysiological manipulations that occur.

Hopefully, the relationships that the rules of a grammar depict have some psychological relevance, but at present this is only speculation.

A transformational grammar consists of three major rule components: a syntactic component, a semantic component, and a phonological component. The syntactic component is the only creative component of the transformational grammar. It contains a base subcomponent (a context-free phrase structure grammar), a transformational subcomponent, and a lexicon. The structures produced by the base component, with lexical substitutions, are designed to represent all of the semantic relationships of the sentence being generated. The semantic component, applied to these abstract constructs generated by rules of the base subcomponent, produces a semantic interpretation for a sentence. This implies that many paraphrases having disparate syntactic surface structures will have similar underlying base structures. The rules of the transformational subcomponent, acting on a base tree and on trees resulting from the application of earlier transformations, produce the surface tree of the sentence. The structure of the surface (derived) tree is intended to reflect not only the actually occurring word order of the sentence, but also psychologically significant groupings of words. The phonological rules apply to this surface structure to produce a representation of spoken language. Thus, the semantic and phonological sets of rules are purely interpretive on, respectively, the base and derived structures of a sentence.

There were two outstanding contributions in the area of syntax last year, the first by Lakoff (70), the second by Fillmore (26). Lakoff investigated certain cases of syntactic irregularity and suggested that lexical items should have rule-exception features

in addition to the other syntactic features already proposed in Aspects (16). This innovation makes possible the exclusion from (or inclusion in) the domain of a transformation some specific lexical items that are not otherwise distinguished as a group. For example, the passive transformation applies to all sentences having a verb followed by a direct object noun phrase (John hit the ball; the ball was hit by John) with the exception of sentences containing a few verbs like "marry" and "befit." By positing the existence of a rule feature [± passive] to be associated with each verb, the passive transformation is now redefined to be applicable to only those sentences containing verbs specified as [± passive]. The implication of this suggestion is to make the base structure of the grammar far more abstract than previously. For example, just as the verb "act" will be specified as [± nominalization] in order to permit a nominalization transformation to derive the noun "action" from the verbal form "act", so will there be posited a verb "idea" similarly specified such that the nominalization transformation obligatorily derives the noun "idea." We also point out here that this same piece of work by Lakoff contains some new and certainly suggestive analyses of parts of English syntax, independent of the suggestions of rule features, such as treating prepositions, adjectives, and verbs in a relatively uniform manner.

The second paper, by Fillmore, is an investigation of the basic form of the base structure of a transformational grammar. He suggests that Chomsky's emphasis on underlying subject and object is an unnecessary one and will not play a significant role in the semantic rules that interpret a base structure. He argues that the relationship between the lexical items "door" and "close" in the sentence "The door closed" and "John closed the door" is exactly the same and that there is no way of expressing this fact

in the base structure formalism presented in Aspects. Accordingly, he suggests a revision of the base structures of sentences, such that every sentence has an analysis as: possibly certain adverbials, an auxiliary, and a proposition. The proposition contains the verb and various types of phrases such as the ergative, agentive, instrumental, etc. One of these phrases, dominated by the proposition, eventually becomes the grammatical subject of the sentence, subject to very rigorous constraints; which are only sketched out in his paper. Fillmore's paper is certainly provocative but much further development of the ideas is needed before their linguistic merit can be evaluated.

Various other papers appeared in which small parts of English syntax were considered within the transformational framework, among them papers by Fraser (31), Peters (50h), Lakoff (70), Ross (50d), Rosenbaum & Lochak (95), Postal (90), Schane (101), and Langendoen (72). All of these were written from the linguist's point of view. Most of them presuppose at least some knowledge of linguistic theory and were not written with a view towards computer implementation of the rules and regularities that they propose.

There were two notable efforts reported in the literature during the last year for which the goal was to compile a grammar of English that accounts for a sizable subset of the language. The first, by Chapin (6), and Geiss (116), represents the current version of the grammatical rules developed for the MITRE Analysis Procedure for Transformational Grammars and presents a set of sample sentences processed by that analysis procedure. These rules represent a revised version of the grammar to be found in The English Preprocessor Manual (84) and consist of corrections, refinements, and the introduction of a few new constructions.

The second paper is that by Rosenbaum & Lochak (95), which contains both an introduction to some of the linguistic theory underlying its grammar, and the conventions utilized, and a set of rules for a part of English syntax with an emphasis on the predicate complement construction. This grammar, which does not make use of syntactic features, will shortly be replaced by a more sophisticated version that uses features and also reflects a number of innovations in linguistic theory.

A third effort, currently underway at the University of California at Los Angeles under the direction of Prof. Robert Stockwell, is attempting to construct a consistent transformational grammar of English syntac that reflects all work performed to date. This project is about half way through its two and one-half year term and no publications can be expected until at least the Fall of 1967.

SECTION III

SEMANTIC THEORY

Last year, Simmons' review of semantic theory suggested that there now exist several "major fragments" of a comprehensive theory of semantics. He mentioned specifically the work of Katz, Fodor and Postal, of the Cambridge Language Research Unit, and of Quillian, and implied that these separate fragments might eventually come together to make some harmonious semantic theory. Perhaps this is true, but no such coalescence has appeared this year, and, in fact, a good percentage of the year's publications in this field seem to have amounted to attacks, open or implicit, by one semantic theorist on another, especially attacks on and by the Katz group. Berlyne (3), Osgood (86), and Bolinger (8) have all published papers written specifically as critical attacks on the Katz position. Two other statements of approaches to semantics that appeared during the year (Weinreich (117), and Quillian (91),) both went to considerable lengths to show why they thought Katz's position was wrong, if not, indeed, absurd. Katz (60), meanwhile, has replied vociferously to at least some of his detractors.

We will not recount here the multifarious disagreements involved in all this, much less take sides. However, we would like to propose one sort of "dimension," along which it seems many semantic theories may be located. Essentially, this dimension is the degree of complexity of the material that is assumed to constitute semantic information. For a performance model, this information

is what would have to be stored in the memory of the device that produced and/or comprehended language. A position lying near the "complexity" end of this dimension would assume that semantic information consists of complex configurations, forming overall a network of nodes, interrelated to one another by different kinds of links. A position closer to the middle of this continuum might assume that semantic information is structured into trees, but still trees using several different kinds of labeled linkages. Further along on this continuum one might find a theory that assumed a tree structure for semantic information but now one that used only a single kind of linkage between nodes of the tree. Finally, a position near the "simplicity" end of our continuum might assume that semantic information consists simply of unordered aggregates of semantic features. Semantic theories of today seem to be spread all along this continuum.

The semantic theory of Katz (61,60) and his co-authors, Fodor (29) and Postal (hereafter KFP) lies very close to the simplicity end of this continuum, although, as Weinreich points out, the KFP theory really seems to occupy two different positions at once. Both of these positions are near the simplicity end of our continuum. First, KFP assumes that the semantic information to be prestored in the lexicon is all to be in the form of very simple trees with unlabeled links, with essentially two kinds of labeled nodes (markers and distinguishers). Second, however, KFP says that the final result of the semantic component of a language is to be an even simpler sort of information, viz, a single bundle of unordered properties for each sentence.

Weinreich feels it would be desirable to be consistent about the complexity of semantic structures (p. 419):

> Another way of saying it is that definitions of words have semantic structures of the same general form as sentences of a language.... This principle explains the possibility of freely introducing new words into a language by stipulating their meanings through expressions couched in words of language L. Of course, a particular complex expression, if it is not tautologicous, contains more features in its semantic structure than any one of its constituents; but the form of the structures, as represented in the theory, is the same for simplexes and complexes.

Weinreich points out that while a few constructions, e.g., simple adjective-noun constructions can indeed be viewed as forming simple, unordered sets of semantic features (a process he calls "linking"), a great many others cannot, and he proposes other sorts of combination procedures --"nesting," "delimiting," "modalization" -- to produce more complex configurations. Thus, Weinreich's own position on our continuum lies somewhat further out toward the complexity end than does KFP, since Weinreich would allow for semantic information of somewhat more complex form. However, Weinreich would still like to treat as many syntactic constructions as possible as forming simple unordered sets of features. Weinreich also discusses a number of other issues in his paper, and argues that his view of semantics is much more compatible with transformational grammars than is KFP. However, on the issue of how semantic information is to be structured, Weinreich only proposes notations to represent nonlinking configurations when this notation can be taken readily from symbolic logic or transformational theory. As a consequence, he still thinks of semantic information as arranged into structures that are restricted to trees, although he suggests at several points that this may be inadequate.

Another paper with assumptions that place it close to the simplicity end of our continuum is that of Kiefer (65), who says (p.228):

> ...It seems obvious that the semantic relations involve a hierarchic structure of semantic categories, therefore the semantic relations are defined with such a system of semantic categories taken for granted.

Kiefer goes on to explore the implications that would follow if there were, for any one language, a single hierarchy into which all semantic categories could be placed. He finds that one could then define different kinds and degrees of similarity between word meanings, depending upon the number and remoteness of what amounts to the superset memberships they share. On the other hand, Kiefer does say (p. 228-9)

> ...The deeper we penetrate into the semantics of natural language the more structured the set of semantic categories seems to be. To put it differently, the more facts about language we want to describe by means of semantic categories, the more we have to impose on K_s (in the semantic memory store) a complicated structure. ...So far, we do not know how complicated the structure of K_s really is.

The position of the C.L.R.U. group (110) on our spectrum seems to lie further toward the complexity end than any of those yet considered, since the semantic structures they assume are nets of relations, without a tree structure necessarily imposed, or any limitation that semantic elements all be single binary features, such as +_animate. However, the C.L.R.U. semantic network is still simple in that it does not label the <u>kind</u> of link connecting semantic categories. The C.L.R.U.'s effort has been devoted to the question of how, on the one hand, one might <u>derive</u> such links from empirical investigation, or, on the other hand, to how one might cluster words or concepts together into categories. The C.L.R.U.'s clustering techniques rely on <u>counting</u> the number

of similarities found between categories, and never on any qualitative difference between kinds of link. It should be noted that these research aims are thus in contrast to the other work reported here, which is concerned with how semantic information could be used if one did have it, rather than how to obtain that information.

At about the same level of complexity as the C.L.R.U. concepts, one finds the networks proposed as models of memory in certain psychological simulation models, for example, by Frijda (35) and also in the "EPAM" models. These later models, which use discrimination nets, are relatively quite well understood and have properties which many people working in language, especially linguists seem not to be aware of. Wynn (119) has given a particularly thorough and insightful investigation of EPAM-type models. It seems to us that a study of such graphs could add a great deal to the linguist's knowledge of what sorts of structures for semantic information are available to be used in conjunction with their syntactic models. In this regard, the works of Reich (92), although incomplete and somewhat chaotic, are at least a first attempt to forge a merger between linguistic problems and an appreciation of what can be accomplished with a sophisticated network model.

An interesting exploration of general properties of graphs was reported during the year by Mann & Jensen (79). These authors describe techniques allowing rapid, guided searching of fairly complex networks, although their techniques are very much limited by being applicable only to the finding of a path between two given nodes that, although perhaps long, still always has all of its links running in one direction. Two-way paths, i. e., ones

made up of two separate subpaths leading from each of the given nodes to some third common node. The importance of allowing this kind of path through a common ingredient has been demonstrated by Quillian's program (91).

Finally, at probably the farthest extreme of the continuum that we have proposed lies Quillian's model. This represents semantic information by general, recursive graph structures (which unlike tree structures also allow loops within the data network) and by complex configurations built up with qualitatively differing kinds of links. Quillian argues that the most important point about such a network is that it can be used by a computer to answer a great many more semantic questions than were anticipated at the time the information was stored, because such information is rich enough to allow its use in reorganized form without becoming nonsense.

Simmons (106, also, is now working with a network model similar to that advocated by Quillian and is addressing himself in particular to the question of how inferences within data stored in such a network may be made. Simmons' work is discussed further in Section V. The inference problem in such networks reduces to the question of when one path connecting two nodes in the data store can be rewritten and/or collapsed into another shorter path. If a connective is rewritten, some paraphrase of the original relationship is generated. If a long path is collapsed into a shorter one, an inference has, in effect, been drawn. Two works relevant to these processes have appeared recently. The first is a dissertation by Elliot (23) in which he attempts to describe prepositions by properties relevant to such inference making. To describe these prepositions, Elliot includes standard

properties such as transitivity, reflexivity, etc., as well as some properties less often considered in this context, such as "one-leader." If prepositions or something like them are to appear as connectives in a network of semantic data, then such properties are essential to drawing inferences from that data.

The second interesting work in regard to inference making is the paper by Fillmore, mentioned earlier in the section on syntactic theory. Fillmore is specifically concerned about how best to formulate parts of a transformational grammar, but his work is relevant to the type of connectives that should be used in a store of semantic data. His point, essentially, is that to specify that some noun is, say, the subject of some verb, is an inadequate specification of their relationship. This inadequacy shows up in a network model when the verb has to be rewritten to derive or recognize a paraphrase. Thus, the verb "to swarm" can be rewritten as "to cluster," if its subject is, say, "bees." However, if the subject of swam is, say, "garden," as in "the garden swarms with bees," then rewriting the verb as "clusters" produces an erroneous paraphrase, unless the verb's subject is also changed. This example shows that, within a store of semantic information, a link labeled "subject of" is inadequately differentiated. See Quillian (91) for more discussion of Fillmore's work from this point of view.

In summary, it would appear that there are now two principal techniques available for representing the structures of semantic information that is to be associated with language. One of these techniques is to view this information as generated by some set of recursive rules, such as a grammar or a computer program. The other is to view this information as some kind of network of

associated links. Either of these techniques may be simple, and consequently bear low information, or complex, and consequently rich in information. Skinner's view of language (108) was that a very simple associative net, and what amounted to trivially simple rules, could adequately explain language. Chomsky (16) argued tellingly against this view, proposing that much more complex rule systems were required. Some of Chomsky's followers have proposed semantic theories to be based on syntactic rule systems and what amounts to a store of information structured so minimally that it contains relatively little information. This may, in fact, be acceptable. On the other hand, this may put an impossible load on the system of semantic rules.

SECTION IV

PSYCHOLINGUISTIC STUDIES RELATED TO SEMANTICS

The number of studies more or less relevant to semantics has continued to increase during the year. A good number of these have been psychologically oriented and connected in one way or another with the Cognitive Studies Center at Harvard. Early work, such as that by Miller (82), appeared to support the hypothesis that people really operate with something functionally similar to a transformational grammar in their heads. Some of the strongest apparently supportive data for this hypothesis is to be found in a study by Savin (49), which appears to show that the transformational complexity of a sentence directly predicts the amount of short-term memory space a subject has left, after reading the sentence, to use for storing a list of nonsense syllables. Other supportive evidence was reported by Fodor (28), who found that subject tended to perceive a click superimposed on an aurally presented sentence, not at the place where the click was superimposed, but instead slightly displaced forward or backward in the sentence; moreover this displacement could be predicted by the syntactic constituent structure of the sentence. However, followup studies often seem to have been discouraging for those who might like to think that transformations have any simple or direct psychological validity. The conflicting evidence on both sides is too tentative and diverse to attempt to synthesize here, but many of the psychological studies involved are summarized in the Harvard Cognitive Center reports (48) and in the survey by Ervin-Tripp & Slobin (24).

Another major source of information on such studies is the Journal of Verbal Learning and Verbal Behavior. The interested reader will also want to look into the survey by Creelmann (19), which is a very thorough examination of "semantic generalization" studies, stemming from Razran's original work. Another recent study, done by Kohlers (66), would appear to be very damaging to the hypothesis that a single grammar must appear as a "component of any reasonable performance model" (unless, of course, one denies that a human is a reasonable performance model!). Kohlers presented bilingual subjects with sentences primarily of one language, but with a few words of those sentences translated into the second language that the subject knew. He found that these subjects could read such texts silently at almost the same speed they could read single-language text, but if asked to read aloud, their time increased very significantly. Kohlers feels this result cannot be wholly explained simply by increased motor difficulty involved in switching the position of the tongue, throat, etc., when speaking different languages, so that, it would seem, some really different information processing must occur in production of sentences as distinct from comprehension of sentences.

Another psychological study relevant to semantics is reported by Paige & Simon (87), who in part were attempting to find out whether or not people solving algebra word problems think in a manner similar to the way that Bobrow's computer program, STUDENT (4), worked. On this question, the answer seems to be a mixed yes and no, depending upon the particular person, on the instructions given him, and so on. One interesting finding resulted from giving the subjects problems that, in fact, violated physical reality (such as requiring for solution that a board be of negative length). Some subjects simply misperceived such problems so

that they did make sense, thus illustrating the pervasive influence of stored knowledge of the task on sentence comprehension.

Two studies dealing with the disambiguation of polysemantic words occurred during the year, one primarily an analytic study. by Rubinstein (97), the other, a computer simulation by Quillian (91). In the latter study it was found, somewhat to the investigator's surprise, that a computer program, working with almost no syntactic information whatsoever, was able to correctly disambiguate 12 out of 19 ambiguous words in 5 sentences of running text. The program utilized semantic information stored in network form.

Another interesting empirical study is reported by MacKay (78). In his study, subjects were asked to complete ambiguous sentences. Judging the difficulty of the task by the time necessary for the subject to complete the sentences, MacKay found that he could distinguish between ambiguities at three levels: lexical, derived (surface) structure, and deep structure. This supports the psychological significance of Chomsky's formulation of a grammar. In addition, the fact that very few subjects even noted the ambiguities indicates perhaps that analysis and understanding are usually concerned with mapping input onto some one consistent, previously viable, interpretation and not with a search for all possible meanings. MacKay concludes with the observation (p.426):

> An attempt to fit these results to several theories of the processing of ambiguous sentences led us to the conclusion that ambiguity interferes with our understanding of a single meaning of a sentence, and that the degree of interference varies with the linguistic level at which the ambiguity occurs.

Another type of study pertinent to semantic and syntactic performance models is that which studies the acquisition of language in children. Earlier work in this area is summarized in Brown (11) and in Ervin-Trip & Slobin (24).

Another study that promises to be of great interest to semantic theory is being undertaken by John Olney* at System Development Corp. Olney is keypunching the entire Webster's Seventh Collegiate Dictionary and will try to answer basic questions in lexicography, such as which words and which kinds of relationships are used most often in the definitions of other words. This is, therefore, an empirical investigation of the semantic network inherent in an actual dictionary, which should prove of considerable interest.

*Personal communication.

SECTION V

COMPUTER SYSTEMS

Computer systems lie on a continuum between two extremes. At one end we have systems which manipulate language for its own sake. providing linguists a tool to better study the English language (and other languages). On the other extreme we have the attempts to make useful systems (or at least demonstration systems) which accept natural language questions as input and then utilize some data base to answer the questions. Of course in some of these question-answering systems part of the aim is to develop a performance theory of language; and similarly some of the theoretical tools are also thought of as being steps toward systems which will ultimately be able to communicate in natural language.

A large project at IBM (80) is concerned with the specification and utilization of a transformational grammar. This project is investigating the applicability of transformational theory as a theoretical support for linguistics aspects of language data processing. As mentioned earlier, this IBM group has been developing a transformational grammar of English which includes a large lexicon. In addition they have been developing computer programs to aid in the grammatical and lexicographic work. They have implemented a LISP program to test the generative power of a transformational grammar. Given the deep structure of a sentence (in a notation developed by the group) a sequence of transformation names, and a transformational grammar, it will

run through the proposed derivation of the sentence, checking for the applicability of the requested transformations and the applicability of other obligatory transformations not requested, but possible. Thus this batch-processing system can help a linguist to determine the consistency of a grammar and the language it generates. Systems similar in purpose to this IBM system are currently being designed for on-line use at Stanford, by Joyce Friedman, at System Development Corporation, by Londe, and at Bolt Beranek and Newman, by Bobrow, Fraser, and Teitelman.

The program at IBM for support of lexicon development involves removing much of the clerical work from the developer. It allows additions, deletions, and modifications to the lexicon, partial or complete printout in several convenient formats, and scans of various sorts to assist the lexicographer. This program is written in SNOBOL for the IBM 7094.

The IBM system described above is useful in helping a linguist test a generative grammar. For analysis of English language input to a computer, however, syntactic analysis programs are necessary. At the MITRE Corp., there has been continued development of their transformational grammar parsing system. This system was described in Simmons survey last year and in Kuno's recent survey (68). Differences between this system and a similar transformation parsing system developed by Petrick (89) were also described. We point out here that only Petrick's system represents an algorithm which handles formally defined class of transformational grammar. The current version is running in LISP on the UNIVAC M-460, and IBM's M44, 7094, 7044, and Q-32. In this past year a lexical lookup system has been programmed for the MITRE system, as reported by Walker &

Bartlett (115). In addition, an efficient context-free parsing algorithm for determining surface structure has been programmed. This algorithm allows a compact representation of multiple trees ("super-trees") in the same structure. In application of the reversal of the generative transformation rules in finding a transformational (deep structure) parsing, a technique has been developed to apply these transformations to these "super-trees," which represent simultaneously a number of phrase structure trees.

Improvement in the MITRE grammar continued to be made and several versions of the MITRE grammar have been issued. The most important change that has been made since the most recent version appeared is a new treatment of coordinate conjunction by Schane (111). Other work in progress deals with lexical analysis and the incorporation into the grammar of syntactic features.

Some interesting investigations of the properties of the MITRE syntactic analysis procedure have been carried out by Friedman. In particular, Friedman (33) shows that an algorithm exists for constructing the surface grammar in the form of the grammar if a weak condition on surface structure is met, namely, that no symbol has two nonadjacent occurrences in a branch without an intervening S node, where S is the only recursive symbol in the grammar.

Elsewhere, Friedman (32) proves that there are transformational grammars for which no surface grammar exists unless suitable restrictions are imposed on the grammar. If this is the case, then algorithms for the construction of surface grammars do exist but there is no algorithm for a minimal surface grammar. A third paper by Friedman (34) discusses the validity and generality of the MITRE syntactic analysis procedure. The problem of

the validity concerns the fact that although the current MITRE procedure is thought to be correct for the specific grammars for which it was developed, there is no proof that some cases have not been overlooked. The generality problem, on the other hand, concerns the difficulties of how to automatically generate the additional components required to make the current procedure applicable to any one of a class of grammars. Possible approaches to the solution of both these problems are presented.

Work on predictive analysis of English has been continued at Harvard under the leadership of Susumo Kuno. This system was described in last year's survey. It has been put in a broader context in Kuno's brilliant survey "Computer Analysis of Natural Languages" (68), in which he presents a coherent overview of the entire field. This paper provides an excellent introduction for the novice and a good review for the sophisticated researcher in the art of language processing. Recently Kuno (67) reported on an algorithm that will accept any context-free grammar and convert it into a standard-form grammar for the predictive analyzer. In addition, the standard-form grammar produced by the analyzer is augmented in such a way that the trees produced by the predictive parser provide information about derivations that would have been found using the original context-free grammar. This augmented predictive analyzer, as a parsing algorithm for arbitrary context-free languages, is compared with two other parsing algorithms: a selective top-to-bottom algorithm similar to Irons' (55) "error-correcting parse algorithm, and an immediate constituent analyzer that is an extension of the Sakai-Cocke algorithm (99) for normal grammars. The comparison is based on several criteria of efficiency, covering core storage requirements, complexity of the programs, and processing time. Then comparison concluded that the augmented predictive analyzer is a parsing

algorithm comparable, to if not superior, to the other parsing algorithms in its overall efficiency. The selection of one algorithm over the others depends on individual application and on one's own taste.

Some interesting work in predictive analysis has been coming from the English Language Research Unit at the University of Edinburgh. In their paper, "A Model for the Perception of Syntactic Structure," Thorne et al (113) propose a predictive analysis system that is designed to be a first stage in a transformational parsing system.

In contrast to the Harvard System, their program is considered to be a model for syntactic perception, rather than an automatic parsing device. The principal implication of this goal is that it is not essential to have, at the first stage, a routine that looks up every word in a form-class dictionary, as is done in the Kuno system. Part of their reason for rejecting such a dictionary lookup (which would operate on every word in a sentence) is that, in most cases, the information obtained in this way increases the complexity of the initial analysis procedure rather than simplifying it. For example, in English virtually every word that can be a noun can also be a verb; given that, a decision must be made as to which of the two roles a word is playing in a sentence under consideration. It therefore seems more economical, if nothing else, to work out the part of speech in terms of fulfillment of predictions based upon nonambiguous items in the same sentence. They propose to have a closed-class dictionary of approximately 2000 items, and some affix rules to give them other information concerning words in the sentence. The result of their predictive analysis is a rough bracketing of the sentence, which is all they claim is needed for a second-stage transformational analysis. This bracketting will contain far fewer ambiguities

than might be obtained if the substructure of these rough groupings were elaborated. This first-stage predictive analyzer is said to be working, and a full transformational parsing system should be working sometime in 1967.

Ernest von Glasersfeld (40) and his group, who have just moved to Athens, Georgia from Milan, Italy, have a different and interesting approach to the analysis of natural language. They have been exploring the applicability to English of a correlational grammar, as implemented in a multi-store parsing procedure, similar to the Sakai-Cocke algorithm mentioned earlier. In their recent report, they describe advances in several areas: classification of words in terms of correlation indices, i.e., by means of code numbers representing the word possibilities of forming syntactic combinations with other words; reclassification, i.e., assigning correlation indices to word combinations; and grammatical and semantic factor analysis, i.e., analysis of grammatical functions and semantic content in terms of constant factors across several languages. The most outstanding example of this latter work, by Brian Dutton of this group, is his analysis (22) of factors underlying the use of pronouns in languages.

String analysis of language derived from work of Harris at the University of Pennsylvania (114) has been continued by Sager at the Linguistic String Project at New York University. This group has a number of working parsing programs that produce an interesting form of sentence analysis. Their system is basically unchanged from that described in the survey by Bobrow (67). The output of analysis is a center string, which represents the skeleton of a sentence, and various substrings, which are adjuncts to portions of the sentence and modify various elements in the

sentence. Recently, they have described a procedure for mapping their analysis program, in which the grammar is embedded in the program itself, into a standard Backus-normal-form grammar.

Bohnert & Backer (7) have been working on a still different approach to the problem of representing the deep structure of natural language. Their thesis, as with Darlington earlier, is that the notation of symbolic logic offers the most promising "deep structure" notation for natural language analysis, and that machine translation to such notation is the best proving ground for this analysis. Machine translation to logic, they claim, has a practical promise in that the result can be acted upon by logical deduction programs. Therefore, if results are stored in this logical notation, information implicit in the store can be deduced by standard theorem-proving techniques. Their approach has been to begin with a simple logic-like language, which they claim is subset of English, and then work step-wise toward a more fully English-like syntax. For each step so far they have written programs for the translation of the language into logic. This is basically an outgrowth of the Carnap school of semantics. The most suggestive part of their work is in exploration of certain parallels between mathematical logic and English; an example is factoring, which leads to their own interesting theory of conjunction.

A number of papers on question-answering systems have appeared this year. William Woods (50 1) has written a very interesting paper on this subject. His paper proposes a uniform method for performing semantic interpretations of English questions for a question-answering system that has a well-structured data base. In particular, the Official Airline Guide is considered as a data base for a proposed interpreter that takes the output

from a hypothetical parsing algorithm of the Chomsky transformational variety and translates it to an expression in a formal query language. The semantic meaning of the input question is represented in a form similar to ordinary quantificational logic, with an additional question and set-definitional capability. The semantic theory underlying the interpretation of a question is not the KFP theory, but one similar to that developed by Bobrow (4). Relations asserted by a sentence are determined by the main verb and its modifiers, and properties of the subject and objects. Arguments of relations are specified by noun phrases, which may denote specific objects, quantified variables (with range perhaps restricted by relative clauses) question words, and functionally defined objects. Woods presents a set of semantic rules in a "pattern→action" format to specify a correspondence between usage restrictions and corresponding semantic meanings of elements for the question-answering system for the Official Airline Guide. The major drawback of this paper is that it reflects only careful thought, but no computer implementation.

Although the material presented is clear, and apparently consistent, there are still a number of fuzzy areas (such as the feasibility of retrieval given a specification of the type proposed), which would only be cleared up by an implementation of such a system. Despite this, the paper is well worth reading, because of its clarity and the number of new ideas it presents about treatment of language objects. It generalizes on the approach taken by Thompson (112) and Craig et al (18) on the DEACON system, which was described last year, and in Simmons' survey of question-answering systems.

Simmons (106) has continued work on question-answering. In his latest approach, questions and statements are syntactically

analyzed to permit derivation of a normalized kernel of the form subject-verb-nominal. The approach to kernelization depends on cues offered by the surface structure that indicate a kernel string should be derived. When kernels have been derived for a question and for a string that may be an answer, the two sets of kernels are compared for an identical match. If this is not found, the pair are checked for a match after certain meaning-preserving equivalence operators have been applied to the kernels. The approach is interesting but seems limited by the ad hoc method for the determination of the kernel strings. The system also faces problems in storing kernels in a fashion that makes relevant ones more immediately accessible than irrelevant in the large data base.

Weizenbaum (118), in a new version of ELIZA, a program for conversing in English, has demonstrated the strong effect of context on understanding of English input. He presents cogent arguments to demonstrate that this is not an artifact of machines but is true for humans as well. His program, as it stands, provides for switching of context to allow different interpretations of the same input as a function of history and frame of reference. In two distinctly different contexts, one simulating a psychiatrist of Rogerian persuasion, and another simulating a mathematical assistant, he shows how the same remarks can be interpreted in two very different ways and lead to two very different types of conversation. His program is designed to deal, not with complete text, but rather with fragments of natural language that occur in real conversations. Therefore, it cannot rely on texts being grammatically complete or correct. (No theory that depends on parsing of presumably well-formed sentences can be of much help in this task.) Since incomplete utterances are characteristic of communication between men, Weizenbaum's approach is most suggestive.

Quillian and Bobrow (5) have two programs under development whose merger, it is hoped, will be able to accept a very wide class of English input and report how the assertions made are related to a prior data store. The first program is a predictive parser with only a very limited dictionary of closed class words. Its output will be a rudimentary "chunking" or bracketing of the elements of the input sentence, rather than a full characterization of its syntactic structure. The second program is a semantic comparison program, which, for each chunk in the sentence, will report that the meaning of the chunk is:

1) Equivalent in meaning to a specified part of the current data store.
2) Contradictory to the current data store.
3) Incomprehensible, given the current data store.
4) Not equivalent, but sufficiently similar to information in the store to allow a tentative interpretation that can be added to the store.

An on-line "teacher" will provide the input sentences and monitor the responses and interpretation of the program. This, hopefully, will allow graceful growth both of the semantic store and of the capabilities of the program, thus increasing the breadth of "understanding" of the system, as a function of context and the history of interaction.

CONCLUSION

No great stride forward has been made this year toward a program that can really "understand" a large body of text. In fact, there is no indication that a major breakthrough is just around the corner. Much more basic research seems necessary before one can think of a general language processing system.

The emphasis on the science rather than on the technology of computational linguistics is particularly evident in the report (85) of the NRC Automatic Language Processing Advisory Committee (ALPAC). The committee undertook a two-year study of the use of computers in the translation of foreign languages. It found little justification at present for massive support of machine translation per se, finding it to be, overall, slower, less accurate, and more costly than that provided by human translators. The committee recommended that computers be used to gain a deeper understanding of the nature of language and for machine-aided translation rather than pure machine translation. "Computational linguistics--study of parsing, sentence generation, structure, semantics, statistics and quantative linguistic matters, including experiment in translation, with machine aids or without, should be supported as a science and not be judged by any immediate or forseeable contributions to practical translation." Judging by this recommendation, it seems that the direction of Federal support (and hence much work) will swing more toward the

theoretical and basic research end of the spectrum than toward the engineering and practical approach at this time.

REFERENCES

1. ALTMANN, BERTHOLD. A Natural Language Storage and Retrieval (ABC) Method: Its Rationale, Operation, and Further Development Program. J. Chem. Doc., 6 (Aug. 1966) 154-157.

2. ANTAL, LASZLO. Content, Meaning, and Understanding. Mouton, The Hague, The Netherlands, 1964, 63 p. (Janua Linguarum, Series Minor, Nr. 31) Reviewed by I. I. Revzin in: Vop. Yazykoz, 15 (Jan. - Feb. 1966) 115-120.

3. BERLYNE, DANIEL E. Mediating Responses: A Note on Fodor's Criticisms. J. Verb. Learn. Verb Behav., 5 (Aug. 1966) 408-411.

4. BOBROW, DANIEL G. A Question-Answering System for High School Algebra Word Problems. In: American Federation of Information Processing Societies. AFIPS Conference Proceedings, vol. 26; 1964 Fall Joint Computer Conference. Spartan, Baltimore, Md., 1964, p. 591-614.

5. BOBROW, DANIEL G. Problems in Natural Language Communication with Computers. Bolt Beranek and Newman, Inc., Cambridge, Mass., Aug. 1966, 19 p. (Report no. Scientific-5, BBN-1439) AFCRL 66-620 (AD-639 323) Also published in: IEEE Trans. Hum. Factors Electron., 4FE-8 (Mar. 1967)

6. BOBROW, DANIEL G. Syntactic Analysis of English by Computer: A Survey. In: American Federation of Information Processing Societies. AFIPS Conference Proceedings, vol. 24; 1963 Fall Joint Computer Conference. Spartan, Baltimore, Md., 1963, p. 365-387.

7. BOHNERT, HERBERT G. & PAUL O. BACKER. Automatic English-To-Logic Translation in a Simplified Model. A Study in the Logic of Grammar. Final report, 1961-1966. IBM, Watson Research Center, Yorktown Heights, N. Y., Mar. 1966, 117 p. (AFOSR 66-1727)

8. BOLINGER, DWIGHT. The Atomization of Meaning. Language, 41 (Oct.-Dec.1965) 555-573.

9. BOWLES, EDMUND A. The Humanities and the Computer: Some Current Research Problems. Comput. Automat., 15 (Apr. 1966) 24-27.

10. BREWER, J. English Article Insertion: Further Studies. Bunker-Ramo Corp., Canoga Park, Calif., Sept. 1965.

11. BROWN. ROGER. Social Psychology. The Free Press, New York, 1965.

12. BURGER, JOHN F., ROBERT E. LONG, & ROBERT F. SIMMONS. An Interactive System for Computing Dependencies, Phrase Structures and Kernels. System Development Corp., Santa Monica, Calif., 28 Apr. 1966, 27 p. (SP-2454)

13. CARMODY, BASIL T. & PAUL E. JONES, JR. Automatic Derivation of Microsentences. Commun. ACM, 9 (June 1966) 443-449.

14. CHOMSKY, NOAM. Aspects of the Theory of Syntax. M. I. T. Press, Cambridge, Mass., 1965, 251 p. (Massachusetts Institute of Technology. Research Laboratory of Electronics. Special Technical report no. 11) (AD-616 323)

15. CHOMSKY, NOAM. Current Issues in Linguistic Theory. Mouton, The Hague, The Netherlands, 1964, 119 p. (Janua Linguarum, Series Minor, nr. 38)
Revised and expanded version of a report presented to the session: "The Logical Basis of Linguistic Theory," Ninth International Congress of Linguists, Cambridge, Mass., 1962.
Also published by: Humanities, New York, 1966, 119 p.

16. CHOMSKY, NOAM. Review of R. B. Skinner's "Verbal Behaviour" (Appleton-Century-Crofts, New York, 1957) Language, 35:1 (1959) 26-58.
Reprinted in: Fodor, Jerry A. & Jerrold J. Katz. The Structure of Language; Readings in the Philosophy of Language. Prentice-Hall, Englewood Cliffs, N. J., 1964, p. 547-578.

17. CHOMSKY, NOAM. Syntactic Structures. 's-Gravenhage: Mouton, The Hague, The Netherlands, 1957, 116 p.

18. CRAIG, JAMES A., SUSAN C. BERENZNER, HOMER C. CARNEY, & CHRISTOPHER R. LONGYEAR. DEACON: Direct English Access and Control. In American Federation of Information Processing Societies. AFIPS Conference Proceedings, vol. 29; 1966 Fall Joint Computer Conference, Nov. 7-10, San Francisco, Calif. Spartan Books, Washington, D. C., 1966, p. 365-380.

19. CREELMAN, MARJORIE B. The Experimental Investigation of Meaning. Springer, New York, 1966.

20. CUADRA, CARLOS A. (ed.) Annual Review of Information Science and Technology, (American Documentation Institute. Annual Review series, vol. 1). Interscience Publishers, New York, 1966, 351 p.

21. DEUTSCH, K. W. On Theories, Taxonomies, and Models as Communication Codes for Organizing Information. Behav. Sci., 11 (Jan. 1966) 1-17.

22. DUTTON, B. Analysis and Definition of Linguistic Closed Systems. Beit. Ling. Inf., 9 (1966)

23. ELLIOTT, ROGER W. A Model for a Fact Retrieval System. Ph. D. Thesis. University of Texas Computation Center, Austin, May 1965, 152 p. (TNN-42)

24. ERVIN-TRIPP, S. M. & D. I. SLOBIN. Psycholinguistics. In: Annual Review of Psychology. Annual Reviews, Inc., Palo Alto, Calif., 1966, vol. 17, p. 435-474.

25. FEDER, JEROME. The Linguistic Approach to Pattern Analysis: A Literature Survey. Lab. for Electroscience Research, New York Univ., N. Y., Feb. 1966, 48 p. (Tech. Report no. TR-400-133) AFOSR 66-1802 (AD-637 497)

26. FILLMORE, CHARLES J. A Proposal Concerning English Prepositions. In: Proceedings of the Seventeenth Annual Round Meeting. Georgetown Univ. Press, Washington, D. C. 1966.

27. FILLMORE, CHARLES J. Toward a Modern Theory of Case. Ohio State University Research Foundation, Columbus, Aug. 1966, p. 1-24. (Project on Linguistic Analysis, Report No. 13)

28. FODOR, JERRY A. More about Mediators: A Reply to Berlyne and Osgood. J. Verb. Learn. Verb. Behav., 5 (Aug. 1966) 412-415.

29. FODOR, JERRY A. & JERROLD KATZ. The Structure of a Semantic Theory. Language, 39 (Oct.-Dec. 1963) 170-210.
Also published in: Jerry A. Fodor & Jerrold J. Katz (eds.) The Structure of Language; Readings in the Philosophy of Language. Prentice-Hall, Englewood Cliffs, N. J., 1964, p. 479-518.

30. FOX, L. (ed.) Advances in Programming and Non-Numerical Computation. Symposium Publications Div., Pergamon Press, Oxford, New York, etc., 1966, 218 p. Contents:
a. GILL, S. Introduction to Automatic Computing, p. 3-27.

b. WOODWARD, P. M. List Programming, p. 29-48.
c. BARRON, D. W. & C. STRACHEY. Programming, p. 49-82.
d. FOSTER, J. M. Artificial Languages, p. 83-96.
e. LANDIN, P. J. A λ-Calculus Approach, p. 97-141.
f. GILL, S. A Survey of Non-Numerical Applications, p. 145-154.
g. COOPER, D. C. Theorem-Proving in Computers, p. 155-182.
h. MICHIE, D. Game-Playing and Game-Learning Automata, p. 183-196;
Appendix: Rules of SOMAC, by John Maynard Smith, p. 196-200.
i. NEEDHAM, R. M. Information Retrieval and Some Cognate Computing Problems, p. 201-218.

31. FRASER, B. Some Remarks on the Verb Particle Combination in English. In: Proceedings of the Seventeenth Annual Round Meeting. Georgetown Univ. Press, Washington, D. C., 1966.

32. FRIEDMAN, JOYCE. An Investigation of Surface Grammars. MITRE Corp., Bedford, Mass., 1965.

33. FRIEDMAN, JOYCE. A Surface Grammar Algorithm. MITRE Corp., Bedford, Mass., 1965.

34. FRIEDMAN, JOYCE. Validity and Generality of the MITRE Syntactic Analysis Procedure: A Problem Statement. MITRE Corp., Bedford, Mass., 1966.

35. FRIJDA, NICO H. & A. D. de GROOT. Towards a Model of Human Memory. Amsterdam Univ., Amsterdam, Holland.

36. GARVIN, PAUL L. Fulcrum Technique for Chinese-English Machine Translation. Final Report, 1 July 1964-30 June 1965. Bunker-Ramo Corp., Canoga Park, Calif., Jan. 1966, 177 p. (RADC TR-65-340) (AD-627 763)

37. GARVIN, PAUL L., JOCELYN BREWER, & MADELINE MATHIOT. Prediction-Typing--A Pilot Study in Semantic Analysis. Bunker-Ramo Corp., Canoga Park, Calif., Jan. 1966, 88 p.

38. GLASERSFELD, ERNST VON. An Approach to the Semantics of Prepositions. In: Josselman, H. (ed.) Proceedings of the Conference on Computer-Related Semantic Analysis, Las Vegas, December 3-5, 1965. Wayne State Univ. Press, Detroit, 1966.

39. [GLASERSFELD, ERNST VON, et al.] Automatic English Sentence Analysis. Final Scientific report for 1 Mar. 1964 - 30 June 1965. Language Research Section, Istituto Documentazione della Associazione Meccanica Italiana (IDAMI) Milan, Italy, 30 June 1965, 113 p. (Report No. ILRS-T-11, 65063) (AFOSR 65-1908) (AD-622 570)

40. GLASERSFELD, ERNST VON & BRIAN DUTTON. Supralinguistic Classification and Correlations. Istituto Documentazione della Associazione Meccanica Italiana (IDAMI) Milan, Italy, Mar. 1966. (Informal Report No. ILRS-T-13, 660320).

41. GLICKERT, PETER. A Codification of English Words. 1325 E. Street, N.W., Washington, D. C., 1966.

42. GOLD, E. MARK. Usages of Natural Language. Part I and II, Basic Principles. Institute for Mathematical Studies in the Social Sciences, Stanford Univ., Calif., Apr. 1966.

43. GRAVES, PATRICIA A., DAVID G. HAYS, MARTIN KAY, & THEODORE W. ZIEHE. Computer Routines to Read Natural Text with Complex Formats. RAND Corp., Santa Monica, Calif., Aug. 1966, 141 p. (Report No. RM-4920-PR) (AD-637 303)

44. GRUBER, JEFFREY S. Studies in Lexical Relations. Massachusetts Institute of Technology, Cambridge, Mass., Sept. 1965.

45. GRUNWALD, ARNOLD P. A Braille-Reading Machine. Science, 154 (7 Oct. 1966) 144-146.

46. HAMP, ERIC, FRED W. HOUSEHOLDER, & ROBERT AUSTERLITZ (eds.) Readings in Linguistics II. University of Chicago Press, Chicago, 1966, 395 p.

47. HARPER, KENNETH E. Studies in Inter-Sentence Connection. RAND Corp., Santa Monica, Calif., Dec. 1965, 21 p. (Report no. RM-4828-PR) (AD-626 572)

48. HARVARD UNIVERSITY. CENTER FOR COGNITIVE STUDIES. Fifth Annual Report, 1964-1965. William James Center for the Behavioral Sciences, Cambridge, Mass., 1965, 72 p.

49. HARVARD UNIVERSITY. CENTER FOR COGNITIVE STUDIES. Sixth Annual Report, 1965-1966. The Center for Cognitive Studies, Center for the Behavioral Sciences, William James Hall, Cambridge, Mass., 1966, 61 p.

50. HARVARD UNIVERSITY. COMPUTATION LABORATORY. Mathematical Linguistics and Automatic Translation. Anthony G. Oettinger, Principal Investigator. Cambridge, Mass., Aug. 1966, 1 vol. (various pagings) (Report No. NSF-17 to the National Science Foundation) Contents:

a. LAKOFF, GEORGE. Stative Adjective and Verbs in English, p. I-1--I-16.
b. LAKOFF, GEORGE & JOHN ROBERT ROSS. A Criterion for Verb Phrase Constituency, p. II-1--II-11.
c. LAKOFF, GEORGE. A Note on Negation, p. III-1--I-16.
d. ROSS, JOHN ROBERT. A Proposed Rule of Tree-Pruning, p. IV-1--IV-18.
e. ROSS, JOHN ROBERT. Relativization in Extraposed Clauses (A Problem which Evidence is Presented that Help is Needed to Solve) p. V-1--V-3.
f. LAKOFF, GEORGE & STANLEY PETERS. Phrasal Conjunction and Symmetric Predicates, p. VI-1--VI-49.
g. HOFFMAN, T. RONALD. Past Tense Replacement and the Modal System, p. VII-1--VII-21.

h. PETERS, STANLEY. A Note on Ordered Phrase Structure Grammars, p. VIII-1--VIII-19.

i. LEWIS, DAVID J. Compositions of Right and Left Sequential Machines, p. IX-1--IX-16.

j. WOODS, WILLIAM A. & SUSUMU KUNO. A Recognition Procedure for Arbitrary Context-Free Grammars and the Upperbound of Recognition Time, p. X-1-- X-26.

k. HERRINGER, JAMES, MARGARET WEILER, & ELEANOR HURD. The Immediate Constituent Analyzer, p. XI-1--XI-47.

l. WOODS, WILLIAM A. Semantic Interpretation of English Questions on a Structured Data Base, p. XII-1--XII-44.

51. HAYS, DAVID G. An Annotated Bibliography of Publications on Dependency Theory. RAND Corp., Santa Monica, Calif., Mar. 1965, 22 p. (Report no. RM-4479-PR) (AD=613 469).

52. HAYS, DAVID G. Processing Natural Language Text. RAND Corp., Santa Monica, Calif., Oct. 1966, 8 p. (P-3461) (AD-640 658)

53. HAYS, DAVID G. (ed.) Readings in Automatic Language Processing. American Elsevier Pub. Co., New York, 1966, 202 p. Contents:

a. HAYS, DAVID G. Introduction, p. 1-7.

b. GORN, SAUL. Specification Languages for Mechanical Languages and Their Processors-- A Baker's Dozen, p. 9-32.

c. KAY, MARTIN & THEODORE W. ZIEHE. Natural Language in Computer Form, p. 33-49.

d. LAMB, SYDNEY M. & WILLIAM H. JACOBSEN, JR. A High-Speed Large-Capacity Dictionary System, p. 51-72.

e. HAYS, DAVID G. Parsing, p. 73-82.

f. KUNO, SUSUMU. The Predictive Analyzer and a Path Elimination Technique, p. 83-106.

g. HAYS, DAVID G. Connectability Calculations, Syntactic Functions, and Russian Syntax, p. 107-125.

h. DINNEEN, DAVID A. The Grammar of Specifiers, p. 127-136.

i. EDMUNDSON, H. P. & DAVID G. HAYS. Research Methodology for Machine Translation, p. 137-147.

j. LAMB, SYDNEY M. On the Mechanization of Syntactic Analysis, p. 149-157.

k. LUHN, H. P. Keyword-in-Context Index for Technical Literature, (KWIC Index) p. 159-167.

l. SALTON, GERARD, Automatic Phrase Matching, p. 169-188.

m. YNGVE, VICTOR H. A Framework for Syntactic translation, p. 189-198.

54. HENDERSON, MADELINE M., JOHN S. MOATS, MARY ELIZABETH STEVENS, & SIMON M. NEWMAN. Cooperation, Convertibility Among Information Systems: A Literature Review. U. S. Dept. of Commerce, National Bureau of Standards, Washington, D. C., 15 June 1966, 140 p. (National Bureau of Standards Miscellaneous Publications 276)

55. IRONS, R. An Error-Correcting Parse Algorithm. Commun. ACM, 6 (Nov. 1963) 669-673.

56. JOHNSON, NEAL F. The Influence of Associations between Elements of Structured Verbal Responses. J. Verb. Learn. Verb. Behav., 5 (Aug. 1966) 369-374.

57. JOHNSON, NEAL F. On the Re tionship between Sentence Structure and the Latency in Generating the Sentence. J. Verb. Learn. Verb. Behav., 5 (Aug. 1966) 375-380.

58. JOINT PUBLICATIONS RESEARCH SERVICE, Washington, D. C. Foreign Developments in Machine Translation and Information Processing.

No. 181 Scientific-Technical Information, no. 11, 1965. 15 Apr. 1966, 171 p. (JPRS-35061) Order from CFSTI or ETC: as TT-66-31499. Trans. of Nauch.-Tekh- Inform. (USSR) no. 11 (1965) 1-60.

No. 182 GIRSHBERG, YU V., A. M. DUBITSKAYA, N. S. KOLCHINSKAYA. 15 Apr. 1966, 14 p. (JPRS-35063) Order from CFSTI or ETC: as TT-66-31501. Trans. of Nauch.-Tekh. Inform. (USSR) no. 12 (1965) 45-48.

59. KASHER, ASA. Data-Retrieval by Computer; a Critical Survey. Hebrew Univ., Jerusalem, Israel, Jan. 1966, 72 p. (Technical Report No. 22 to Office of Naval Research, Information Systems Branch). (AD-631 748)

60. KATZ, JERROLD J. Mr. Pfeifer on Questions of Reference. Found. Lang., 2 (Aug. 1966) 241-244.

61. KATZ, JERROLD J. PAUL M. POSTAL. An Integrated Theory of Linguistic Descriptions. Massachusetts Institute of Technology, Cambridge, 1964. (Research Monograph No. 26)

62. KAY, MARTIN. The Tabular Parser: A Parsing Program for Phrase Structure and Dependency. RAND Corp., Santa Monica, Calif., July 1966, 49 p. (RM-4933-PR) (AD-635 934)

63. KAY, MARTIN, FREDERICK VALADEZ, & THEODORE ZIEHE. The Catalog Input/Output System. RAND Corp., Santa Monica, Calif., Mar. 1966, 71 p. (Report No. 4540-PR) (AD-629 606).

64. KELLOGG, CHARLES H. An Approach to the On-Line Interrogation of Structured Files of Facts Using Natural Language. System Development Corp., Santa Monica, Calif., 29 Apr. 1966, 86 p. (SP-2431/000/00)

65. KIEFER, FERENC. Some Semantic Relations in Natural Language. Found. Lang., 2 (Aug. 1966) 228-240.

66. KOLERS, PAUL A. Reading and Talking Bilingually. Amer. J. Psychol., 79 (Sept. 1966) 357-376.

67. KUNO, SUSUMU. The Augmented Predictive Analyzer for Context-Free Languages -- Its Relative Efficiency. Commun. ACM, 9 (Nov. 1966) 810-823.

68. KUNO, SUSUMU. Computer Analysis of Natural Languages. Presented at Symposium on Mathematical Aspects of Computer Science; American Mathematical Society, New York, April 5-7, 1966.

69. KURKI-SUONIO, REINO. On Top-to-Bottom Recognition and Left Recursion. Commun. ACM, 9 (July 1966) 527-528.

70. LAKOFF, GEORGE. On the Nature of Syntactic Irregularity. [Doctoral Dissertation, Indiana Univ.] In: Harvard University. Computation Laboratory. Mathematical Linguistics and Automatic Translation. Anthony G. Oettinger, Principal Investigator. Cambridge, Mass., Dec. 1965, 1 vol. (Report No. NSF-16 to the National Science Foundation.)

71. LANGENDOEN, TERENCE. Some Problems Concerning the English Expletive 'It'. Ohio State University Research Foundation, Columbus, Aug. 1966, p. 104-134.

72. LANGENDOEN, TERENCE. The Syntax of the English Expletive 'It'. In: Proceedings of the Seventeenth Annual Round Meeting. Georgetown Univ. Press, Washington, D. C., 1966.

73. LEHMANN, W. P. & E. D. PENDERGRAFT. Machine Language Translation Study. Quarterly Progress Report no. 11, 1 Nov. 1965 - 31 Jan. 1966. Linguistics Research Center, Univ. of Texas, Austin, May 1966, 121 p. (ECOM 02162-11) (AD-635 505)

74. LIEBERMAN, D., P. ROSENBAUM, D. LOCHAK, & F. BLAIR. Specification and Utilization of a Transformational Grammar. IBM, Thomas J. Watson Research Center, Yorktown Heights, N. Y., Mar. 1966, 305 p. (Report No. Scientific-1) (AFCRL 66-270) (AD-635 520).

75. LINDSAY, ROBERT K. A Heuristic Parsing Procedure for a Language Learning Program. University of Texas, Austin, 28 May 1964. (Information Processing Report Number 12)

76. LINDSAY, ROBERT K. & JANE M. LINDSAY. Reaction Rime and Serial Versus Parallel Information Processing. J. Exp. Psychol., 71 (Feb. 1966) 294-303.

77. LOCKHEED MISSILES & SPACE COMPANY. ELECTRONIC SCIENCES LABORATORY. Automatic Indexing and Abstracting, Part 1. Annual Progress Report. B. D. Rudin, Principal Investigator. Palo Alto, Calif., Mar. 1966, 1 vol. (various pagings) (Report no. M-21-66-1-Pt.1) (AD-631 241)

78. MacKAY, D. G. To End Ambiguous Sentences. Percept. Psychophys., Vol. 1 (1966)

79. MANN, WILLIAM C. & PAUL A. JENSEN. A Data Structure for Directed Graphs in Man-Machine Processing. Computer Command and Control Co., Washington, D. C., 20 June 1966, 76 p. (Report no. 77-106-1) (AD-636 251)

80. MARCKWARDT, ALBERT H. Linguistics and the Teaching of English. Indiana University Press, Bloomington, 1966, 160 p.

81. MATTHEWS, G. HUBERT. Analysis of Synthesis of Sentences of Natural Languages. In: 1961 International Conference on Machine Translation of Languages and Applied Language Analysis. (Proceedings of the Symposium held at the National Physical Lab., Teddington, England, Sept. 1961) Her Majesty's Stationery Office, London, 1962, p. 531-542.

82. MILLER, G. A. Some Preliminaries to Psycholinguistics. Amer. Psychol., 20 (1965) 15-20.

83. MINSKY, MARVIN L. Artificial Intelligence. Sci. Amer., 215 (Sept. 1966) 246-252, 257-258, 260.

84. MITRE CORPORATION. English Preprocessor Manual. Language Processing Techniques Sub-Department, Information Sciences Dept., MITRE Corp., Bedford, Mass., Dec. 1964. (Report no. SR-132)

85. NATIONAL ACADEMY OF SCIENCES--NATIONAL RESEARCH COUNCIL. AUTOMATIC LANGUAGE PROCESSING ADVISORY COMMITTEE. Language and Machines; Computers in Translation and Linguistics. National Academy of Sciences--National Research Council, Washington, D. C., 1966, 124 p. (Publication 1416)

86. OSGOOD, CHARLES E. Meaning Cannot Be r_m? [Theoretical Note] J. Verb. Learn. Verb. Behav., 5 (Aug. 1966) 402-407.

87. PAIGE, JEFFRY M. & HERBERT A. SIMON. Cognitive Processes in Solving Algebra Word Problems. Carnegie Institute of Technology, Pittsburgh, Pa., 18 Mar. 1965. (Complex Information Processing, Paper # 80)

88. PAPP, FERENC. Mathematical Linguistics in the Soviet Union. Mouton, The Hague, The Netherlands, 1966.

89. PETRICK, S. R. & J. Keyser. Syntactic Analysis. Physical Sciences Research Report #324, May 1967. Bedford, Mass. AFCRL 67-0305

90. POSTAL, PAUL M. On So-Called "Pronouns" in English. In: Proceedings of the Seventeenth Annual Round Meeting. Georgetown Univ. Press, Washington, D. C. 1966.

91. QUILLIAN, M. ROSS. Semantic Memory. Ph. D. Thesis. Carnegie Institute of Technology, Pittsburgh, Pa., Feb. Also published by Bolt Beranek and Newman Inc. Cambridge, Mass., Oct. 1966. (Scientific Report No. 5)

92. REICH, PETER A. An Algorithm Which Executes a Network of Finite Automata Used by Stratificational Grammar. Dept. of Psychology and Mental Health Research Institute, Univ. of Michigan, Ann Arbor, 11 June 1966.

93. REITMAN, WALTER R. Cognition and Thought; An Information Processing Approach. Wiley, New York, 1965, 312 p.

94. RITCHIE, R. W. & P. S. PETERS, JR. Generative Power of Transformational Grammar. (To be published)

95. ROSENBAUM, P. & D. LOCHAK. The IBM Core Grammar of English. In: Lieberman, D. [et al.] Specification and Utilization of a Transformational Grammar. International Business Machines Corp., Thomas J. Watson Research Center, Yorktown Heights, N. Y., Mar. 1966, Part 1. (Scientific report no. 1, Contract no. AF19 (628)-5127) (AFCRL 66-270) (AD-635 520)

96. ROSENBERG, SHELDON. Recall of Sentences as a Function of Syntactic and Associative Habit. J. Verb. Learn. Verb. Behav., 5 (Aug. 1966) 392-396.

97. RUBENSTEIN, HERBERT. Problems in Automatic Word Disambiguation. In: Josselman, H. (ed.) Proceedings of the Conference on Computer-Related Semantic Analysis, Las Vegas, December 3-5, 1965. Wayne State Univ. Press, Detroit, 1966.

98. RUBIN, J. Optimal Classification into Groups: An Approach for Solving the Taxonomy Problem. Unpublished IBM report. Mar. 1966.

99. SAKAI, ITIRO. Syntax in Universal Translation. In: 1961 International Conference on Machine Translation of Languages and Applied Language Analysis (held in Teddington, England, Sept. 1961) Her Majesty's Stationery Office, London, England, 1962, vol. 2, p. 593-609.

100 SAMMET, JEAN E. The Use of English as a Programming Language. Commun. ACM, 9 (Mar. 1966) 228-230.

101. SCHANE, S. A. A Schema for Sentence Coordination. MITRE Corp., Bedford, Mass., Aug. 1966, 64 p. (MTP-10) (AF. ESD. TR-66-305) (AD-640 570)

102. SEBEOK, THOMAS A. (ed.) Current Trends in Linguistics. Vol. 3: Theoretical Foundations. Mouton, The Hague, The Netherlands, 1966.

103. SHAPIRO, PAUL A. ACORN--An Automated Coder of Report Narrative. Biostatistics Dept., Roswell Park Memorial Institute, Sept. 1966.

104. SHILLAN, D. A Method and a Reason for Tune Analysis of Language. Cambridge Language Research Unit, Cambridge, England, Nov. 1965. (M.L.179).

105. SHOUP, JUNE E. Research on Grammatical Analysis of Spoker Language. Final Report 2 Feb. 1964 - 1 Feb. 1966; Communication Sciences Lab., Michigan Univ., Ann Arbor, 1 Feb. 1966, 14 p. (AFOSR 66-1566) (AD-637 566)

106. SIMMONS, ROBERT F. Automated Language Processing. In: Cuadra, Carlos (ed.) Annual Review of Information Science and Technology, (American Documentation Institute. Annual Review series, vol. 1). Interscience Publishers, New York, 1966, p. 137-169.

107. SIMON, HERBERT A. & ALLEN NEWELL. Heuristic Problem Solving by Computer. In: Sass, Margo A. & William D. Wilkinson (eds.) Computer Augmentation of Human Reasoning. Spartan Books, Washington, D. C., 1965, p. 25-35.

108. SKINNER, B. F. Verbal Behaviour. Appleton-Century-Crofts, New York, 1957.

109. SMOKLER, H. Informational Content: A Problem of Definition. J. Phil., 63 (14 Apr. 1966) 201-211.

110. SPARCK JONES, K. Semantic Markers. Cambridge Language Research Unit, Cambridge, England, Nov. 1965. (M.L. 181)

111. TERZI, PAOLO. An Hypothetical Mechanism of the Origin of Ideas as an Instrument for the Automatic Analysis of Language [Un ipotetico meccanismo delle origini delle idee quale strumento per l'analisi automatica del linguaggio]. Istituto Lombardo, Accademia di Scienze e Lettere, Milan, 17 Feb. 1966.

112. THOMPSON, FREDERICK B. English for the Computer. In: American Federation of Information Processing Societies. AFIPS Conference Proceedings, vol. 29; 1966 Fall Joint Computer Conference, November 7-10, San Francisco, Calif. Spartan Books, Washington, D. C., 1966, p. 349-356.

113. THORNE, J. P., H. M. DEWAR, H. WHITFIELD, & P. BRATLEY. A Model for the Perception of Syntactic Structure. English Language Research Unit, Univ. of Edinburgh, Scotland, 1966.

114. UNIVERSITY OF PENNSYLVANIA. DEPARTMENT OF LINGUISTICS. NSF TRANSFORMATIONS PROJECT. Report on the String Analysis Programs. Introductory Volume. Philadelphia, Pa., Mar. 1966, 182 p. Contents:

a. SAGER, NAOMI. A Syntactic Analyzer for Natural Language.
b. MORRIS, JAMES. The IPL String Analysis Program.
c. SAGER, NAOMI & MORRIS SALKOFF. Outputs from Scientific Articles.
d. RAZE, CAROL. The FAP String Analysis Program.

115. WALKER, D. E. & J. M. BARTLETT. The Structure of Languages for Man and Computer: Problems in Formalization. MITRE Corp., Bedford, Mass. (Report no. SS-10) Also published in: First Congress on the Information System Sciences, Hot Springs, Va., 19-21 November 1962. Session 10, p. 45-105. (AD-435 102)

116. WALKER, D. E., P. G. CHAPIN, M. L. GOIS, & L. N. GROSS. Recent Developments in the MITRE Syntactic Analysis Procedure. MITRE Corp., Bedford, Mass., Sept. 1966, 40 p. (AF. ESD. TR-66-306)

117. WEINREICH, U. Explorations in Semantic Theory. In: Sebeok, Thomas A. (ed.) Current Trends in Linguistics. Mouton, The Hague, The Netherlands, 1965, vol. 3.

118. WEIZENBAUM, JOSEPH. ELIZA--A Computer Program for the Study of Natural Language Communication between Man and Machine. Commun. ACM, 9 (Jan. 1966) 36-45. (AD-623 745)

119. WYNN, W. H. An Information-Processing Model of Certain Aspects of Paired-Associate Learning. Unpublished Ph. D. Dissertation. Univ. of Texas, Austin, Jan. 1966.

120. ZIFF, P. More on Understanding Utterances. System Development Corp., Santa Monica, Calif., 8 June 1966, 16 p. (SP-2504)

121. ZWICKY, ARNOLD M., JOYCE FRIEDMAN, BARBARA C. HALL, & DONALD E. WALKER. The MITRE Syntactic Analysis Procedure for Transformational Grammars. In: American Federation of Information Processing Societies. AFIPS Conference Proceedings, vol. 27, Part 1: Fall Joint Computer Conference. Spartan Books, Washington, D. C., 1965, p. 317-326. (AD-629 455)

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Bolt Beranek and Newman Inc
50 Moulton Street
Cambridge, Massachusetts 02138

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

SURVEY OF AUTOMATED LANGUAGE PROCESSING 1966

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)

Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Bobrow, D. G. Quillian, M. R.
Fraser, J. B.

6. REPORT DATE

April 1967

7a. TOTAL NO. OF PAGES

54

7b. NO. OF REFS

121

8a. CONTRACT OR GRANT NO.

AF19(628)-5065

b. PROJECT NO.

8668

ARPA Order No. 627, Amendment No. 2

DOD element
6154501R

d. DOD subelement n/a

9a. ORIGINATOR'S REPORT NUMBER(S)

Scientific Report #7

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

AFCRL-67-0269

10. DISTRIBUTION STATEMENT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11. SUPPLEMENTARY NOTES

PREPARED FOR
Hq. AFCRL, OAR (CRB)
United States Air Force
L.G.Hanscom Field, Bedford, Mass. 01730

12. SPONSORING MILITARY ACTIVITY

Advanced Research Projects Agency

13. ABSTRACT

This report is a survey of Automated Language Processing done in 1966. It is limited in scope to analytical processing of natural language, excluding work in programming languages, speech recognition, and statistical processing of text. It focuses on work aimed at generating and analyzing sentences of a natural language. This survey has four major sections: The first, on syntactic theory, contains a summary of the principal assumptions underlying work in generative grammar and a report of the most significant developments in theoretical and descriptive work in syntax. The second section, on semantic theory, attempts to provide some dimensions along which we can judge various theories that have been proposed and developed in the literature in 1966. A number of empirical studies related to semantics and psycholinguistics are reported in a third section. Finally, a fourth section discusses various computer systems for manipulation of natural language. These range from systems that support linguistic studies to systems that are attempting to utilize natural language as a communication medium.

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

Unclassified
Security Classification

A-31408

Unclassified
Security Classification

14. KEY WORDS	LINK A ROLE	LINK A WT	LINK B ROLE	LINK B WT	LINK C ROLE	LINK C WT
Natural Language Processing						
Syntactic Analysis						
Semantic Theory						
Psycholinguistics						
Generative Grammars						
Computational Linguistics						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

Unclassified
Security Classification

A-31409