# MULTIPLE-PATH SYNTACTIC ANALYZER *

S. KUNO AND A. G. OETTINGER

Harvard Computation Laboratory, Cambridge, Mass., USA

## 1. INTRODUCTION

The method of predictive syntactic analysis [1-4]) aims at obtaining a single most probable description of the structure of an input sentence in a single left-to-right scan through the sentence. The computer program uses a storage area called the *prediction pool*. At any intermediate point in the analysis of a sentence, the prediction pool contains a single set of *predictions*, generated by the processing of the preceding words, that may be fulfilled by the remaining words. The prediction pool is similar to a pushdown store in that the prediction fulfilled and discarded is usually among the topmost in the pool, and in that the newly generated predictions are placed above the remaining predictions.

Experiments on Russian and English texts have demonstrated the capability of predictive analysis for handling complex sentence structures including many levels of subordination or coordination, but the results have been disappointing for the following reasons:

1. There are many syntactically ambiguous sentences in natural texts. Provision for determining all legitimate alternative syntactic structures is therefore essential from both the theoretical and the practical points of view. Neither estimates of the reliability of syntactic analysis, nor significant attacks on the problems of choosing the semantically correct structures, are possible without such provisions. A storage area termed *hindsight* has been provided in predictive analysis programs in the hope of enabling at least local alternative parsings, but practical use of this facility now appears inordinately difficult.

2. When a single-path analysis comes to a dead end, determining which of the previous branch points was the cause of the failure poses serious problems.

3. Owing to the lack of an effective method for distinguishing paths which have already been followed from those which have not, it has not been possible to try different paths in a systematic loop-free sequence.

A new method has been developed for extending the predictive approach by including effective and economical provisions for multiple analyses of syntactically ambiguous sentences. The prediction pool for this method is of variable size, consisting of one or more *subpools*, each of which contains a set of predictions

corresponding to a path that may lead to an acceptable structure for the complete sentence. Each subpool is a pushdown store in the strict sense; that is, only the topmost prediction in each subpool is tested against the next word of the sentence.

After the $(k - 1)$st word in a sentence has been processed, the prediction pool contains a subpool for each sentence structure compatible with the first $(k - 1)$ words. The topmost prediction of each subpool is then tested against all the homographs of the $k$-th word. By a simple process of grammar table look-up, each allowable combination of a prediction and a homograph is associated with new predictions which replace the topmost prediction of the appropriate subpool. Subpools for which no allowable combination exists are discarded. The subpools resulting from this process are used in turn for the processing of the $(k + 1)$st word. After the processing of the last word of a sentence, only those subpools which have no predictions remaining are retained in the prediction pool. By tracing back the paths that have yielded these subpools, the alternative acceptable syntactic structures of the sentence are obtained.

## 2. DICTIONARY AND SYNTACTIC WORD CLASSES

Each word of an input sentence is looked up in a dictionary and is coded for membership in all the syntactic word classes to which it belongs. For example, the input English sentence THEY ARE FLYING PLANES. will be coded as shown in table 1. Punctuation marks are treated like words in the ordinary sense.

## 3. GRAMMAR TABLE

A grammar table is a rectangular array defining the *grammatical matching function* G of a language whose syntax is described in terms of a set of predictions P, a set of syntactic word classes S, and a set of syntactic role indicators R. Each prediction stands for a certain syntactic structure recognized in the language. To each ordered argument pair $(P_i, S_j)$, G assigns a set, possibly empty, of ordered pairs

$$G(P_i, S_j) = \{[(p_1{}^1, p_1{}^2, \ldots, p_1{}^{m_1}), (r_1)],$$

$$[(p_2{}^1, p_2{}^2, \ldots, p_2{}^{m_2}), (r_2)], \ldots, [(p_q{}^1, p_q{}^2, \ldots, p_q{}^{m_q}), (r_q)]\},$$

where $p_k{}^l \in P$ and $r_k \in R$ $(k = 1, 2, \ldots, q; l = 1, 2, \ldots, m)$.

Each element of $G(P_i, S_j)$ corresponds to a set of structures that may follow when the syntactic structure represented by the given prediction $P_i$ is initiated by

a word belonging to class $S_j$. Whenever $P_i$ and $S_j$ are grammatically incompatible, $G(P_i, S_j) = \Phi$, the empty

Table 1. Sample from an English Dictionary.

| English Word | Class Code | Comments (not stored in machine) |
|---|---|---|
| THEY | PRN | personal pronoun in the nominative case |
| ARE | BE1 | finite complete intransitive verb, as in "They *are* in the sky." (A prepositional phrase, according to the present grammar, is considered to be adverbial, and cannot fulfill the role of a complement or object of a verb.) |
| | BE2 | finite copula, as in "They *are* students." and "They *are* good." |
| | BE3 | finite auxiliary verb for the progressive form, passive voice, and be-to form, as in "They *are* coming.", "They *are* seen.", and "They *are* to come here." |
| FLYING | RI1 | present participle of complete intransitive verb, as in "They are *flying* to Boston" and "It is a *flying* plane." |
| | RT1 | present participle of single-object transitive verb, as in "He is *flying* a plane." |
| | GI1 | gerund of complete intransitive verb, as in "*Flying* is pleasant." |
| | GT1 | gerund of single-object transitive verb, as in "*Flying* a plane is pleasant." |
| PLANES | NOU | noun, as in "They are *planes*." |
| | VI1 | finite complete intransitive verb, as in "The glider *planes*." |
| | VT1 | finite single-object transitive verb, as in "He *planes* the surface of the board." |
| | PRD | period as end of sentence punctuation |

set. Each couple $[(P_i, S_j), G(P_i, S_j)]$ in $G$ is a *rule* of the grammar. A rule subsumes as many *subrules* as there are members of $G(P_i, S_j)$, each subrule being

defined as a couple $[(P_i, S_j), g_k(P_i, S_j)]$, where

$$g_k(P_i, S_j) \in G(P_i, S_j).*$$

In the present English grammar, the rule for $(P_i, S_j) = $ = (SENTENCE, PRN) (PRN = personal pronoun in the nominative case) consists of the subrules shown in table 2. The vertical lines in $g_k(P_i, S_j)$ denote the boundaries of predictions; the slashes in the English examples denote the boundaries of the corresponding structures. The prediction which comes last in a subrule is the first to be tested, hence the subrules should be read from right to left. The syntactic role indicator $r$ is SUBJECT OF PREDICATE VERB OF DECLARATIVE SENTENCE for the first five subrules, and SUBJECT OF PARTICIPIAL PHRASE for the last three subrules.

In practice, rules are stored sequentially in machine memory in the alphabetic order of their argument pairs, those pairs for which $G(P_i, S_j) = \Phi$ being omitted. The present experimental grammar has approximately 3400 subrules.

## 4. ANALYSIS OF A SENTENCE

The procedure for analysing a sentence will be explained in this section using THEY ARE FLYING PLANES. as an example. At the beginning of the analysis of the sentence, the prediction of SENTENCE is stored in the prediction pool. Next, this prediction is paired with the syntactic word class (PRN) of the first word (THEY) to form an argument which, when looked up in the grammar table, yields the eight sets of predictions and syntactic role indicators shown in table 2; these new predictions then replace the initial prediction of SENTENCE. The prediction pool now contains eight subpools each of which corresponds to a different way, recognized by the present grammar, of terminating the sentence initiated by PRN.

The analysis proceeds to the second word, and the three syntactic word classes (BE1, BE2, BE3) assigned to ARE are coupled with the topmost prediction of each of the eight subpools in the prediction pool. The resulting arguments are

---

* This characterization of the grammatical matching function is due to Warren Plath.

Table 2. The Subrules of $G$ (Sentence, PRN).

$g_1$(SENTENCE, PRN)  = | PERIOD | PREDICATE (as in " / They / go / .")

$g_2$  = | PERIOD | PREDICATE | ADJECTIVE CLAUSE (as in " / We / who are ready to die / salute you / .")

$g_3$  = | PERIOD | PREDICATE | COMMA | ADJECTIVE PHRASE | COMMA (as in "/ They / , / knowing the truth / , / came to the right conclusion / .")

$g_4$  = | PERIOD | PREDICATE | COMMA | SUBJECT PHRASE | COMMA (as in " / We/ , / the people of the United States / , / love peace / .")

$g_5$  = | PERIOD | PREDICATE | SUBJECT PHRASE | AND-OR (as in " / They / and / John / came / .")

$g_6$  = | SENTENCE | COMMA | PARTICIPLE (as in " / They / having done the right thing / , / we can trust them.")

$g_7$  = | SENTENCE | COMMA | PARTICIPLE | SUBJECT PHRASE | AND-OR (as in " / They / and / John / having done the right thing / , / we can trust them.")

$g_8$  = | SENTENCE | COMMA | PARTICIPLE | COMMA | SUBJECT PHRASE | COMMA (as in " / They / , / the Russians / , / having said no / , / we took a decisive step.")

(PREDICATE, BE1), (PREDICATE, BE2),
★ (PREDICATE, BE3); (ADJECTIVE CLAUSE,
BE1),...; (COMMA, BE1)...; (COMMA,
BE1),...; (AND-OR, BE1),...; (PARTICIPLE,
BE1),...; (AND-OR, BE1),...; (COMMA,
BE1),....

Each of these 24 arguments is looked up in the grammar table, but only (PREDICATE, BE1), (PREDICATE, BE2) and (PREDICATE, BE3) yield nonempty $G(P_i, S_j)$. All the subpools stored in the prediction pool—except that with PREDICATE as the topmost prediction—are discarded, since the predictions in them cannot be fulfilled by the sentence in question.

The subrules for (PREDICATE, BE1), (PREDICATE, BE2) and (PREDICATE, BE3) are shown in table 3. All of these subrules have PREDICATE VERB as their syntactic role indicator.

Now, the new predictions given by these grammar subrules replace the topmost prediction (PREDICATE) of the subpool which orginally contained | PERIOD | | PREDICATE. Twelve new subpools, all of which have the prediction of PERIOD as the bottom prediction, are generated and stored in the prediction pool. These subpools in turn are used for the processing of the next word, FLYING, and so on.

The current grammar yields three analyses for THEY ARE FLYING PLANES. The first in table 4 shows the syntactic structure of the sentence applicable when THEY refers to planes. The second shows the syntactic structure of the sentence acceptable when THEY refers to people. Analysis No. 03 is semantically absurd, but it reflects the structure of a sentence such as "The facts are smoking kills.", which is not semantically absurd. The same three analyses would be obtained for "The facts are smoking kills.", but only one would be semantically correct for this sentence.

The analyses obtained for THEY ARE FLYING PLANES. can be limited to the two syntactically and

semantically acceptable ones simply by deleting the subrule pertaining to

$g_4$(PREDICATE, BE2) = | DECLARATIVE CLAUSE.

However, the semantically correct analysis of "The facts are smoking kills." would thereby be lost, leaving only two unacceptable analyses. The easy way out of this particular dilemma would be to rule out "The facts are smoking kills." as ill-formed and accept only "The facts are: smoking kills.". The problem is, however, a more general one to which the solution must be sought, not within the presently defined precincts of syntax but in the shadowy realm of semantics. A set of multiple analyses provides, for the first time, a firm base from which to start such an exploration.

The same grammar produced two analyses for IT HAS ALREADY BEEN MENTIONED THAT A RESPONSE MAY BE LEARNED BY THE MACHINE IF ENCOURAGED BY THE EXPERIMENTER. Analysis No. 01 (table 5) shows the structure of the sentence in which IT is a *temporary* subject without normal pronominal reference and in which THAT introduces the *real* subject noun clause. This analysis corresponds to the way in which the sentence is generally understood. Analysis No. 02 (table 6) shows the structure of the sentence in which THAT introduces an adverbial clause, with IT referring to something mentioned before. The syntactic structure reflected in the analysis corresponds to the structure of such semantically similar but more normal expressions as "It has already been mentioned so that a response may be learned by the machine..." or "It has already been mentioned lest a response may inadvertently be learned by the machine..."

The second analysis is due to subrules pertaining to adverbial "that"-clauses as in "We eat that we may live." or "It has been kept polished that it may glitter forever." One solution is to keep the syntactic classifi-

Table 3. The Subrules of G (Predicate, BE*i*).

$g_1$(PREDICATE, BE1) = | ADVERBIAL PHRASE (as in "They are / *in the sky* / .")

$g_2$      = | PREDICATE | AND-OR-COMMA | ADVERBIAL PHRASE (as in "They are / *in the sky* / *and* / *are over us* / .")

$g_1$(PREDICATE, BE2) = | NOUN COMPLEMENT (as in "They are / *students* / .")

$g_2$      = | ADJECTIVE COMPLEMENT (as in "They are / *good* / .")

$g_3$      = | NOUN CLAUSE (as in "The fact is / *that it smells* / .")

$g_4$      = | DECLARATIVE CLAUSE (as in "The fact is / *it smells* / .")

$g_5$      = | PREDICATE | AND-OR-COMMA | NOUN COMPLEMENT (as in "You are / *a student* / *and* / *are a scholar* / .")

$g_6$      = | PREDICATE | AND-OR-COMMA | ADJECTIVE COMPLEMENT (as in "They are / *diligent* / *and* / *are working every day* / .")

$g_1$(PREDICATE, BE3) = | PARTICIPLE (as in "They are / *coming* / ." and "They are / *seen* / ."

$g_2$      = | INFINITIVE (as in "They are / *to come* / .")

$g_3$      = | PREDICATE | AND-OR-COMMA | PARTICIPLE (as in "They are / *coming* / *and* / *are staying here* / ." and "They are / *seen* / *and* / *are heard* / .")

$g_4$      = | PREDICATE | AND-OR-COMMA | INFINITIVE (as in "They are / *to come* / *and* / *are to stay here* / .")

**ANALYSIS NO. 01**

| ENGLISH | SWC | SYNTACTIC ROLE | CLAUSE LEVEL | PREDICTION | | PARTIAL INTERPRETATION OF PREDICTION CODES |
|---|---|---|---|---|---|---|
| THEY | PERSONAL PRN NOM | SUBJECT OF PREDICATE VERB | DECLARATIVE SENT | SEX | | SENTENCE |
| ARE | BE-COPULA | PREDICATE VERB | DECLARATIVE SENT | PD | VBX | PREDICATE |
| FLYING | PRESENT P OF VII | COMPLEMENT OF PREDICATE VERB | DECLARATIVE SENT | PD | N3A | NOUN COMPLEMENT (A) |
| PLANES | NOUN I | COMPLEMENT OF PREDICATE VERB | DECLARATIVE SENT | PD | N6A | NOUN COMPLEMENT (B) |
| . | PERIOD | END OF SENTENCE | END OF SENTENCE | PD | | PERIOD |

**ANALYSIS NO. 02**

| ENGLISH | SWC | SYNTACTIC ROLE | CLAUSE LEVEL | PREDICTION | | PARTIAL INTERPRETATION OF PREDICTION CODES |
|---|---|---|---|---|---|---|
| THEY | PERSONAL PRN NOM | SUBJECT OF PREDICATE VERB | DECLARATIVE SENT | SEX | | SENTENCE |
| ARE | BE2-AUXILIARY | PREDICATE VERB | DECLARATIVE SENT | PD | VBX | PREDICATE |
| FLYING | PRESENT P OF VTI | PREDICATE VERB | DECLARATIVE SENT | PD | PAA | PARTICIPLE |
| PLANES | NOUN I | OBJECT OF PREDICATE VERB | DECLARATIVE SENT | PD | N2A | OBJECT PHRASE (A) |
| . | PERIOD | END OF SENTENCE | END OF SENTENCE | PD | | PERIOD |

**ANALYSIS NO. 03**

| ENGLISH | SWC | SYNTACTIC ROLE | CLAUSE LEVEL | PREDICTION | | PARTIAL INTERPRETATION OF PREDICTION CODES |
|---|---|---|---|---|---|---|
| THEY | PERSONAL PRN NOM | SUBJECT OF PREDICATE VERB | DECLARATIVE SENT | SEX | | SENTENCE |
| ARE | BE2-COPULA | PREDICATE VERB | DECLARATIVE SENT | PD | VBX | PREDICATE |
| FLYING | GERUND OF VII | GERUND-SUBJECT | COMPLEMENT CLAUSE | PD | SGE | DECLARATIVE CLAUSE |
| PLANES | COMPLETE VI | PREDICATE VERB | COMPLEMENT CLAUSE | PD | VBE | PREDICATE |
| . | PERIOD | END OF SENTENCE | END OF SENTENCE | PD | | PERIOD |

Table 5. Analyzed Sentence

**ANALYSIS NO. 01**

| ENGLISH | SWC | SYNTACTIC ROLE | CLAUSE LEVEL | PREDICTION | | PARTIAL INTERPRETATION OF PREDICTION CODES |
|---|---|---|---|---|---|---|
| IT | TEMPORARY SUBJECT | TEMPORARY SUBJECT | DECLARATIVE SENT | SEX | | SENTENCE |
| HAS | HAVE3-TENSE AUX | PREDICATE VERB | DECLARATIVE SENT | PD NCC | VBX | PERIOD |
| ALREADY | ADVERB 1 | ADVERB | DECLARATIVE SENT | PD NCC | PFA | PERIOD / PREDICATE |
| BEEN | PAST P OF BE3 | PREDICATE VERB | DECLARATIVE SENT | PD NCC | PFA | PERIOD / PAST PART. PERFECT |
| MENTIONED | PAST P OF VTI | PREDICATE VERB | DECLARATIVE SENT | PD NCC | PPA | PERIOD / PAST PART. PERFECT |
| THAT | NOUN CONJUNCTION | CONJUNCTION | SUBJECT CLAUSE | PD | | PERIOD / PAST PART. PASSIVE |
| A | ADJECTIVE 3 | SUBJECT OF PREDICATE VERB | SUBJECT CLAUSE | PD SGC | N4A | PERIOD / NOUN CLAUSE |
| RESPONSE | NOUN I | SUBJECT OF PREDICATE VERB | SUBJECT CLAUSE | PD VBC | | PERIOD / NOUN CLAUSE / DECLARATIVE CLAUSE / SUBJECT PHRASE (B) |
| MAY | AUXILIARY VERB | PREDICATE VERB | SUBJECT CLAUSE | PD VBC | | PERIOD / PREDICATE |
| BE | INFINITE BE3 | PREDICATE VERB | SUBJECT CLAUSE | PD BVA | | PERIOD / PREDICATE |
| LEARNED | PAST P OF VTI | PREDICATE VERB | SUBJECT CLAUSE | PD PAA | | PERIOD / NON-FINITE VERB |
| BY | PREPOSITION | PREPOSITION | SUBJECT CLAUSE | PD | | PERIOD / PARTICIPLE |
| THE | ADJECTIVE 3 | OBJECT OF PREPOSITION | SUBJECT CLAUSE | PD N2G | | PERIOD |
| MACHINE | NOUN I | OBJECT OF PREPOSITION | SUBJECT CLAUSE | PD N5G | | PERIOD / OBJECT PHRASE (A) |
| IF | ADVERB CONJUNCT | CONJUNCTION | ADVERBIAL CLAUSE | PD | | PERIOD / OBJECT PHRASE (B) |
| ENCOURAGED | PAST P OF VTI | PARTICIPIAL VERB | ADVERBIAL CLAUSE | PD PAC | | PERIOD / PARTICIPLE |
| BY | PREPOSITION | PREPOSITION | ADVERBIAL CLAUSE | PD | | PERIOD |
| THE | ADJECTIVE 3 | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N2G | | PERIOD / OBJECT PHRASE (A) |
| EXPERIMENTER | NOUN I | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N5G | | PERIOD / OBJECT PHRASE (B) |
| . | PERIOD | END OF SENTENCE | END OF SENTENCE | PD | | PERIOD |

Table 6. Analyzed Sentence (continued)

**ANALYSIS NO. 02**

| ENGLISH | SWC | SYNTACTIC ROLE | CLAUSE LEVEL | PREDICTION | | PARTIAL INTERPRETATION OF PREDICTION CODES |
|---|---|---|---|---|---|---|
| IT | INDEFINITE PRN | SUBJECT OF PREDICATE VERB | DECLARATIVE SENT | SEX | | SENTENCE |
| HAS | HAVE3-TENSE AUX | PREDICATE VERB | DECLARATIVE SENT | PD | VBX | PERIOD / PREDICATE |
| ALREADY | ADVERB 1 | ADVERB | DECLARATIVE SENT | PD | PFA | PERIOD / PAST PART. PERFECT |
| BEEN | PAST P OF BE3 | PREDICATE VERB | DECLARATIVE SENT | PD | PFA | PERIOD / PAST PART. PERFECT |
| MENTIONED | PAST P OF VTI | PREDICATE VERB | DECLARATIVE SENT | PD | PPA | PERIOD / PAST PART. PASSIVE |
| THAT | ADVERB CONJUNCT | CONJUNCTION | ADVERBIAL CLAUSE | PD | | PERIOD |
| A | ADJECTIVE 3 | SUBJECT OF PREDICATE VERB | ADVERBIAL CLAUSE | PD SGG | N4A | PERIOD / DECLARATIVE CLAUSE / SUBJECT PHRASE (B) |
| RESPONSE | NOUN I | SUBJECT OF PREDICATE VERB | ADVERBIAL CLAUSE | PD VBG | | PERIOD / PREDICATE |
| MAY | AUXILIARY VERB | PREDICATE VERB | ADVERBIAL CLAUSE | PD VBG | | PERIOD / PREDICATE |
| BE | INFINITE BE3 | PREDICATE VERB | ADVERBIAL CLAUSE | PD BVA | | PERIOD / NON-FINITE VERB |
| LEARNED | PAST P OF VTI | PREDICATE VERB | ADVERBIAL CLAUSE | PD PAA | | PERIOD / PARTICIPLE |
| BY | PREPOSITION | PREPOSITION | ADVERBIAL CLAUSE | PD | | PERIOD |
| THE | ADJECTIVE 3 | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N2G | | PERIOD / OBJECT PHRASE (A) |
| MACHINE | NOUN I | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N5G | | PERIOD / OBJECT PHRASE (B) |
| IF | ADVERB CONJUNCT | CONJUNCTION | ADVERBIAL CLAUSE | PD | | PERIOD |
| ENCOURAGED | PAST P OF VTI | PARTICIPIAL VERB | ADVERBIAL CLAUSE | PD PAC | | PERIOD / PARTICIPLE |
| BY | PREPOSITION | PREPOSITION | ADVERBIAL CLAUSE | PD | | PERIOD |
| THE | ADJECTIVE 3 | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N2G | | PERIOD / OBJECT PHRASE (A) |
| EXPERIMENTER | NOUN I | OBJECT OF PREPOSITION | ADVERBIAL CLAUSE | PD N5G | | PERIOD / OBJECT PHRASE (B) |
| . | PERIOD | END OF SENTENCE | END OF SENTENCE | PD | | PERIOD |

cation of words simple and the grammar rules general and to tolerate getting Analysis No. 02 for "It has already been mentioned that a response may be learned . . ." and Analysis No. 01 for "It has been kept polished that it may glitter forever." Another approach is to make the syntactic classification of words more detailed, and the grammar rules more complicated, so that a *real* subject noun clause introduced by "that" is admitted only when it is preceded by "it" plus special verbs and adjectives ("*It is known* that the number is greater than zero."; "*It is clear* that there is no solution to the question.") and so that an adverbial clause introduced by "that" is admitted only in other cases. Further theoretical and empirical studies are necessary to determine which choice to make for any given purpose.

The column headed PREDICTION in tables 5 and 6 shows the status of a prediction subpool before the processing of the word in the same row. Whenever PD (PERIOD prediction) appears alone in that column, the word in the preceding row is the last of a well-formed substring of the sentence. Thus the sentence could have been terminated with a period in any position marked by a slash: IT HAS ALREADY BEEN MENTIONED / THAT A RESPONSE MAY BE LEARNED / BY THE MACHINE / IF ENCOURAGED / BY THE EXPERIMENTER.

## 5. PROGRAM

The analysis program for an IBM 7090 was written to follow only one path at a time, so that all data transfers and table references required in the course of analysis might be effected entirely within the core memory (32 000 words).

A path is determined in part by the choice of a single homograph $S_{\beta_k}$ for each word position $k$ ($k = 1, 2,...,n$), where $n$ is the number of words in a sentence. If the $k$th position has $\alpha_k$ homographs $S_{\beta_k}(\beta_k = 1, 2, . . ., \alpha_k)$, then the total number of distinct selections is

$$N = \prod_{k=1}^{n} \alpha_k.$$

These $N$ selections are effectively enumerated by means of a variable radix representation in which the $k$th digit $\beta_k$ is initially set to 1 for all $k$; then $\beta_n$ is incremented by unity until $\beta_n = \alpha_n + 1$, following which $\beta_n$ is reset to 1 and $\beta_{n-1}$ is increased by a unit carry, and so on in the usual way until $\beta_k = \alpha_k$ for all $k$.

Let $p_k$ be a subpool in the prediction pool following the analysis of the $k$th word, and $P_{ik}$ the topmost prediction in $p_k$. The number of paths from word $k$ to the homograph $S_{\beta_{k+1}}$ due to $p_k$ is then equal to the number $\gamma_k$ of subrules

$$[(P_{ik}, S_{\beta_{k+1}}),\ g_{i_k}(P_{ik}, S_{\beta_{k+1}})]\quad (l_k = 1, . . ., \gamma_k).$$

A single path from $k$ to $k + 1$ is thus determined by fixing $\beta_{k+1}$ and $l_k$. When

$$g_{i_k}(P_{ik}, S_{\beta_{k+1}})$$

replaces $P_{ik}$ at the top of $p_k$, a new subpool $p_{k+1}$ and a

corresponding $P_{i,k+1}$ are obtained, and the single path may, if possible, be extended to $k + 2$.

The extension will not be possible if either

$$k + 2 = n + 1 \text{ or } G(P_{i,k+1}, S_{\beta_{k+2}}) = \Phi.$$

In the former case, a path has been found through the sentence which, if $p_{k+1}$ is empty, corresponds to an acceptable analysis. In the latter case, the path has no continuation to and through $S_{\beta_{k+2}}$; hence $\beta_{k+2}$ is incremented. If $\beta_{k+2} + 1 = \alpha_{k+2} + 1$, extension of the path from $k$ to $k + 1$ and on to $k + 2$ is ruled out completely and hence a new path from $k$ to $k + 1$, determined by $l_k + 1$, is tried; $\beta_{k+1}$ is reset to 1. Similarly, if $k + 2 = n + 1$, the path from $k$ to $k + 1$ determined by $l_k + 1$ is checked. In either case, when $l_k + 1 = \gamma_k + 1$, the paths from $k$ to $S_{\beta_{k+2}}$ have been exhausted, $\beta_{k+1}$ is incremented, a new $\gamma_k$ is provided, and $l_k$ is reset to 1. If $\beta_{k+1} + 1 = \alpha_{k+1} + 1$, then $\beta_{k+1}$ is set to 1 and a new path from $k - 1$ to $k$ is tried, providing a new $P_{ik}$. The process terminates when $l_k = \gamma_k$, and $\beta_k = \alpha_k$ for all $k$.

Thus, branchings caused by homography (membership of a given word form in more than one syntactic word class) and by multiple functions of a given word class (more than one subrule in a grammar rule) are followed in a systematic loop-free sequence in which any given partial path is never followed more than once. The amount of core-storage required in the course of analysing a sentence is proportional to $n$.

## 6. MINIMIZING THE NUMBER OF PATHS TO BE FOLLOWED

It was originally feared that the number of different paths to be taken, and hence the processing time, would grow exponentially with $n$, and make the method impractical. The programming technique of section 5 has, however, proved to be a very effective means of discarding irrelevant paths: if no path is open to $S_{\beta_k}$ because $G(P_{i,k-1}, S_{\beta_k}) = \Phi$, then at least $\prod_{k+1}^{n} \alpha_i$ path continuations are eliminated at one stroke.

Several other techniques have been developed that eliminate additional irrelevant paths without destroying any paths which may yield acceptable analyses.

## 7. RUNNING TIME

The analysis of THEY ARE FLYING PLANES. on an IBM 7090 took less than one second. The analysis of IT HAS ALREADY BEEN MENTIONED THAT A RESPONSE MAY BE LEARNED BY THE MACHINE IF ENCOURAGED BY THE EXPERIMENTER. took approximately one minute. Twelve minutes were needed for the analysis of the 35-word sentence: A SHEAR STRESS APPLIED DURING THE RECOVERY HAD NO EFFECT ON THE AMOUNT OF RECOVERY, IF THE STRESS WAS LESS THAN THE INSTANTANEOUS YIELD POINT, IRRESPECTIVE OF THE DIRECTION OF THE STRESS.

The time necessary for the analysis of a sentence is not directly proportional to the length of the sentence, since it strongly depends on the nature of the sequence of homograph sets that are assigned to the words in the sentence.

## 8. CONCLUSION

A program for multiple-path syntactic analysis of English has been written for both the Univac I and the IBM 7090 and tested on a variety of sentences. Experiments so far have yielded satisfactory results, and have given hints as to what should be done toward improving the definitions of word classes and of grammar rules and toward further reducing running time.

The application of the system to the analysis of Russian is now being studied, and it is expected that the basic principles of the method offer a convenient framework for the development of more powerful syntactic analyzers for both English and Russian. Since arbitrary sets of homographs can be assigned to one or more word positions, the system is also an experimental tool for the study of distributional and generative grammars.

## 9. ACKNOWLEDGEMENTS

## 10. REFERENCES

1) Alt, F. L. and I. Rhodes: *Recognition of Clauses and Phrases in Machine Translation of Languages*. Proc. of the First International Conference on Machine Translation of Languages and Applied Language Analysis, NPL 1961. (to appear).

2) Oettinger, A. G.: *Automatic Syntactic Analysis and the Pushdown Store*. Proc. of Symposia in App. Math., Amer. Math. Soc. 12 (1961).

3) Rhodes, I.: *A New Approach to the Mechanical Translation of Russian*. NBS Report No. 6295 (1959).

4) Sherry, M.: *Comprehensive Report on Predictive Syntactic Analysis*. Mathematical Linguistics and Automatic Translation, Report No. NSF-7, Section I, Harvard Computation Laboratory (1961).