SYNTACTIC STRUCTURE AND AMBIGUITY OF ENGLISH*

Susumu Kuno and Anthony G. Oettinger
Computation Laboratory of Harvard University
Cambridge, Massachusetts

1. INTRODUCTION

This paper is in two parts. The first (Section 2) gives an evaluation of the performance of the multiple-path syntactic analyzer to date, with emphasis on the nature and the consequences of syntactic ambiguities in English sentences and suggestions for the refinement of the grammar. The remainder of the paper is concerned with certain concrete implications of the theoretical description of multiple-path predictive analysis provided by recent work of Evey[4, 5] and Greibach[6, 8]. A modification of the form of the current grammar is proposed which should yield a new grammar with additional intuitive appeal, a simplified version of the present analysis program, and sentence structure descriptions in the form of a generalized parenthesis-free notation readily interpretable as a tree.

The basic technique of multiple-path predictive analysis has been described previously (Kuno and Oettinger[12, 13]). The grammar and other details of the operating system are given in full in two recent reports (Kuno[10, 11]).

The grammar is essentially a set of directed productions as defined by Greibach[6, 8]. A directed production is written as $(P, c) \rightarrow c\, P_1 \ldots P_k$ where c is a terminal symbol (syntactic word class) and the P's are intermediate symbols (predictions). Each prediction stands for a syntactic structure ascribed by the grammar to a string of the language, such as "S" (sentence), "VP" (predicate), "SP" (subject phrase), "PD" (period), etc. A syntactic role indicator is adjoined to each production to describe the role played by the word class c when fulfilling the prediction P. For example, (S, prn) → prn VP PD, (*SV*) indicates that a sentence may be initiated ("S" is an initial symbol) by a prn (personal pronoun in the nominative case) serving as subject of a predicate verb (*SV*), and that the pronoun should be followed by a predicate ("VP") and a period ("PD").

For any given English sentence the analyzer, now in operation on Harvard's IBM 7090, produces explicitly all parsings of the sentence implicit in the current version of the grammar, which has been designed to accept as well-formed most sentences that appear or may appear in scientific papers.

The analyzer, based on a predictive technique originally proposed by Rhodes[17], is abstractly characterized as a directed production analyzer or dpa (Greibach[6]). Every dpa is the inverse of a context-free phase structure generator (psg) in a standard form with productions $P \rightarrow c\, P_1 \ldots P_k$. It is an inverse in the sense that the dpa will accept as well-formed precisely those strings generated by the psg. Since Greibach has shown that for every psg (in the sense of Chomsky) there is a psg in standard form which generates precisely the same set of strings, every psg has a dpa as an inverse, and the intuitively evolved multiple-path predictive analyzer therefore turns out to have even

* This work has been supported in part by the National Science Foundation under Grant G-24833.

greater generality and esthetic appeal than was originally hoped for.

The mechanism of analysis may be characterized as a non-deterministic pushdown store transducer. According to results of Chomsky[2] and Evey[4], the set of all languages that can be either accepted or generated by this class of machines is precisely the set of all context-free phrase structure languages. Earlier conjectures of Oettinger[15] regarding the role of pushdown stores in syntactic analysis are thus confirmed and, although other mechanisms have been suggested (Matthews[14], Sakai[19]) or implemented (Robinson[18]) there is now good reason for regarding the pushdown store transducer as a "natural" device and not merely as a convenient programming trick.

Conceptually, the analyzer operates as follows. The topmost prediction P (intermediate symbol) in a prediction pool (pushdown store) is used to form a couple (P, c) with the word class c of the word being scanned. If there is no production in the grammar with couple (P, c), the pool is abandoned. Otherwise, the symbol "P" is deleted from the pool, as many copies of the pool are made as there are productions with couple (P, c), the elements $P_1 \ldots P_k$ of each production are loaded into the corresponding pool, the process moves to the next word and continues with each of the new pools in turn. The process is initiated with a single pool containing only the initial symbol "S"; it yields an acceptable structure for a sentence whenever a period (or equivalent) is reached and the pool is empty after removal of the prediction of the period; it terminates when all pools have been abandoned or have led to acceptable structures. Since a given word may belong to more than one syntactic word class, means for cycling through the possible word class combinations must be superimposed on this basic non-deterministic pushdown store machine, but this adds no essential features or complications.

Each distinct sentence structure is displayed both as a list of couples (P, c) consistent with the characterization of the system as a directed production analyzer and in a more conventional tree form related to its characterization as the inverse of a phrase structure generator.

2. THE OUTPUT OF THE ANALYZER

2.1 The application of the analyzer to English text has, on the whole, yielded results that are encouraging in the sense that intuitively satisfactory and semantically acceptable structures are produced for a wide range of sentences. Where a sentence is commonly regarded as inherently ambiguous (e.g., "They are flying planes."), the analyzer produces several structures each reflecting one of the distinct interpretations.

There has been, to date, no difficulty in extending the grammar to yield acceptable analyses for sentences rejected by earlier versions, and no major difficulties are anticipated on this score in the future. Catastrophic increase in the size of the grammar seems unlikely; in fact, the current grammar of 2100 rules is descended from an earlier version with 3500 rules with some increase in power on the way.

To be sure, certain common "idiomatic" structures are still maltreated owing to the absence of idiom tables. These have been deliberately omitted to resist the temptation toward excessive *ad hoc* use of such tables to handle apparently difficult constructions that, after some thought, turn out to be amenable to clear-cut systematic treatment within the frame-work of a dpa. Certain rare types of linked structures (e.g., such strings as abcd . . . abcd . . .) known to be beyond the scope of context-free phrase structure grammars must eventually be accounted for either by introducing the equivalent of less restrictive productions (thereby significantly deviating from pushdown store techniques) or by some *ad hoc* truncating technique (thereby sacrificing some conceptual elegance for the sake of a sound engineering solution). These and other sins of omission are not, however, of prime concern to us today.

The most serious problem for the immediate future is the matter of ambiguity. A sentence is ambiguous relative to a given dpa (psg) if that sentence is analyzed (generated) by the dpa (psg) in more than one way. Dealing with ambiguity is hard for both formal and psychological reasons.

Formally, there is a class of unpleasant theoretical results that tell us that the ambiguity problem is recursively unsolvable for context-free languages even of greatly restricted generality (Chomsky and Schützenberger[3], Greibach[7]), i.e., no general algorithm can be found for determining whether or not a given dpa (psg) will analyze (generate) some sentence in more than one way. The outlook for practically interesting decidable subsets is dim, and so experimental search for special solutions in special cases is our only recourse.

In a grammar that purports to describe a natural language, the question is not so much the existence of ambiguity but, worse yet, matching the ambiguity of the grammar to that observed in the language. From this point of view, there are three types of ambiguities: those that should be in the grammar because they are seen in the language, those that should not but are readily eliminated, and the rest. Obviously, the first two types cause no trouble. The elimination of the second type usually corresponds to an enlargement of the precincts of syntax at the expense of what otherwise would be regarded as semantics.

It is, however, a major problem to classify an ambiguity. Is it there because the grammar is at fault? Or are we unhappy with it merely because our mind is fixed on one plausible interpretation to the exclusion of others? At this stage one's disciplined inclination is to answer yes to the first question. Consider, however, the following sentence: "People who apply for marriage licenses wearing shorts or pedal pushers will be denied licenses."† Silly but clear, isn't it? But have you thought that "*People* who apply . . . *or pedal pushers* . . ." could be denied licenses? Dope pushers would be! Or perhaps it is "*People* who *apply* for . . . *or* (who) *pedal* pushers . . ." ? People do pedal bicycles. Are they wearing shorts, or are they applying for shorts that happen to be wearing marriage licenses? Will they *be denied* licenses? Or will they be *denied licenses*? There are more which the current grammar relentlessly exhibits.

Less frivolous cases will now be considered. Space permits only a sampling of both good and bad. Details may be found in Kuno[11] or run your own; grammar, dictionary and program are available to responsible investigators.

† For this and several other valuable test sentences we are indebted to Professor F. W. Harwood of the University of Tasmania who challenged our ability to deal with them.

2.2 The first example to be considered will be a clear-cut one. It will serve primarily to illustrate various features of the analyzer and its output and to demonstrate that there are well-behaved English sentences that are properly treated by the analyzer. Two additional examples will then be used to exhibit ambiguities of the second and third type.

Figure 1 is a fragment of the grammar table. The argument pairs are couples (P, c). The new predictions (NEW PREDS) are right-hand sides $P_1 \ldots P_k$ of directed productions (P, c) $\rightarrow$ c $P_1 \ldots P_k$. Thus the rule entry of 7X, MMM-3 corresponds to a directed production (7X, mmm) $\rightarrow$ mmm XD MC. As mentioned earlier, the syntactic role indicator (SR) partly specifies the role c plays when fulfilling the prediction P. The role is completely specified by the syntactic role indicator in conjunction with indices (e.g., "A" of "XD-A" in 7X, MMM-3) associated with predictions. The agreement test indicator (AGREE TEST) introduces an apparent deviation from a strict pushdown transducer, but Greibach (Section 2.3)[8] has shown that it functions purely as an abbreviation technique without altering the fundamental nature of the grammar and analyzer. The structural and shift codes (STRUCT, SHIFT CD) are used by an editing program to turn the output of the dpa into a tree representation.

Definitions of a few of the 133 word classes (terminal symbols) presently used in the grammar are given in Fig. 2. A list of all 82 current predictions (intermediate symbols) is given in Fig. 3.

Sentence 1 is "The increase in flow stress was attributed to vacancies, which have appreciable mobility at — 72". Figure 4 shows the word class codes associated by the English dictionary with each word in this sentence. "S", "P", "C" and "Y" as the fourth character denote singular, plural, common, and subjunctive, respectively.

The unique analysis produced for this sentence is shown in Fig. 5. In any analysis, a single word class (SWC) together with a mne-

ARGUMENT PAIR	SR	AGREE TEST	NEW PREDS	MNEMONIC DESCRIPTIONS OF PREDICTIONS	STRUCT, SHIFT CC	ENGLISH EXAMPLES
					***	LANGUAGE
7X,GT1-0	YY	00100		SUBJECT MASTER	SA	PROCESSING
			7X-X	SUBJECT MASTER	0 S	MECHANISMS
						WILL BE NEEDED
7X,MMM-0	YY	10010		SUBJECT MASTER	S	TRANSLATION
						WILL BE NEEDED
7X,MMM-1	YY	10010		SUBJECT MASTER	S	TRANSLATION
			AP-	POST-POSITIONAL ADJ	1 SPM	PERFORMED
						(AUTOMATICALLY)
						WILL BE NEEDED
7X,MMM-2	YY	10010		SUBJECT MASTER	S	TRANSLATION
			AC-	ADJECTIVE CLAUSE	1 S7S	WHICH
					(S7V)	IS PERFORMED
						(AUTOMATICALLY)
						WILL BE NEEDED
7X,MMM-3	YY	00001		SUBJECT MASTER	S	ANALYSIS
			XD-A	(A) AND (B)	0 +	AND
			MC-X	NOUN SUBJECT	0 S	SYNTHESIS
						WILL BE NEEDED
7X,MMM-4	YY	00001		SUBJECT MASTER	S	ANALYZERS
			CN-A	COMMA	0 ,	,
			MC-X	NOUN SUBJECT	0 S	TRANSFORMERS
			XC-A	(A,B,) AND (C)	0 +	AND
			MC-X	NOUN SUBJECT	0 S	SYNTHESIZERS
						WILL BE NEEDED
7X,MMM-5	YY	10010		SUBJECT MASTER	S	ANALYZERS
			CN-A	COMMA	0 ,	,
						(AUTOMATIC)
			1C-X	SUBJECT	0 S	ANALYZERS
			CN-A	COMMA	0 ,	,
						WILL BE NEEDED
7X,NOU-0	YY	00100		SUBJECT MASTER	SA	TRANSLATION
			7X-X	SUBJECT MASTER	0 S	PROGRAM
						WILL BE NEEDED
					***	SPACE
7X,NUM-0	YY	00100		SUBJECT MASTER	SA	COMMUNICATIONS*
					(SA)	GREAT
			4X-X	MODIFIED SUBJECT	0 S	DIFFICULTIES
						ARE TO BE
						CONSIDERED

Figure 1. Fragment of Grammar Table.

AAA	common features of ADJ, ADK, ADM, ADN, ADO, and ART
AAB	common features of ADK and ADN
ADJ	(a) adjectives which can be modified by "very" such as "beautiful", "red", etc. "Many, much, few, little" are excluded from this class. (b) adjectives in the superlative degree, excluding "most" and "least". Ex. "prettiest, best, worst".
ADK	adjectives in the comparative degree: "older, better".
ADL	"very, only, same" as adjectives.
ADM	"most" and "least".
ADN	"more" and "less".
ADO	"many, much, few, little".
ADP	"such".
ART	ART is for noun-phrase introducers such as definite and indefinite articles ("the, a"), demonstrative adjectives ("this, that, these, those"), possessive pronouns ("my, your"), pro-adjectives ("another, any") and titles ("Dr.").
AUX	auxiliary verbs: "will, shall, can, may, do, does, etc.".
AV1	usual adverbs: "quickly, fast".
AV2	adverbs homographic with prepositions: "He came <u>in</u>.".
BE1	finite forms of "be" as a complete intransitive verb: "He <u>is</u> well.", "He moved that the problem <u>be</u> up for discussion.".
BE2	finite forms of "be" as a copula which has to be followed by a noun complement or by an adjective complement: "He <u>is</u> tall.", "I am a student.".
BE3	finite forms of "be" as an auxiliary verb for the progressive form, passive voice, or be-to-form: "He <u>is</u> running.", "He <u>was</u> killed.", "He <u>is</u> to come here.".
BI1	the basic form of BE1: "be".
BI2	the basic form of BE2: "be".
BI3	the basic form of BE3: "be".
CCO	non-subjunctive adverbial clause introducers: "before, after, since".
CMA	comma, semicolon, colon, dash, and parenthesis.
CO1	noun clause introducing conjunctions "that", "if" and "whether".
II1	the basic form of VI1.
II3	the basic form of VI3.
IT1	the basic form of VT1.
IT2	the basic form of VT2.
IT6	the basic form of VT6.
IT7	the basic form of VT7.

Figure 2. Fragment of Class Definitions.

Prediction, Mnemonic Description			
1X	SUBJECT	IO	INTERROG PRN ACC
33	AS-CLAUSE	IQ	INTERROG PRN COMPL
4X	MODIFIED SUBJECT	IX	COMPLETE VI
7X	SUBJECT MASTER	LB	RELATIVE PRONOUN ACC
88	THAN-CLAUSE	MX	NOUN SUBJECT
A1	ATTRIBUTIVE ADJ	N2	OBJECT
A2	DISCONTINUOUS ADJ	N3	NOUN COMPLEMENT
AC	ADJECTIVE CLAUSE	N5	MODIFIED OBJECT
AI	ADJECTIVE	N6	MODIFIED COMPLEMENT
AP	POST-POSITIONAL ADJ	N8	OBJECT MASTER
AR	ARTICLE	N9	COMPLEMENT MASTER
B1	INFINITE VT1	NC	NOUN CLAUSE
BV	INFINITE VERB	ND	NOUN CL WITH NO OBJ
BW	INF VERB WITH NO OBJ	NE	CONDITIONAL NOUN CLAUSE
BX	INF COMPLETE VI	NQ	NOUN OBJECT
BY	INFINITE COPULA	PA	PARTICIPLE
C2	ADVERB CLAUSE CONJ	PB	PART WITH NO OBJ
C3	AS (OF COMPARISON)	PD	PERIOD
C8	THAN (OF COMPARISON)	PF	PERFECT PARTICIPLE
CM	COMMA, AND, OR	PG	PERF PART WITH NO OBJ
CN	COMMA	PH	PERF PARTICIPLE VI
CX	COPULA	PI	PERF PART COPULA
DA	ADVERB	PJ	PERF PART BE1
DB	ADVERB AFTER BE1	Q1	PERF PARTICIPLE VT1
DC	THERE, HERE	QU	QUESTION MARK
DM	DUMMY PREDICTION	R1	PARTICIPLE VT1
DN	ADVERBIAL NOUN PHR	RR	PARTICIPLE VI
DP	PREPOSITIONAL PHR	RS	PRES PART COPULA
DQ	PREPOSITION	SE	SENTENCE
EX	BE2 (COPULA)	SF	DECLAR CL WITH NO OBJ
FX	BE3 (AUXILIARY)	SG	DECLARATIVE CLAUSE
G1	GERUND OF VT1	SH	CONDITIONAL DECLAR CL
GR	GERUND	TX	SIMPLE OBJ VT
HX	HAV3 (TENSE AUX)	UX	AUXILIARY VERB
I1	TO-INFIN VT1	VX	PREDICATE
ID	INTERROG ADVERB	WX	PREDICATE WITH NO OBJ
IF	TO-INFINITIVE	XC	(A,B,) AND (C)
IG	TO-INFIN WITH NO OBJ	XD	(A) AND (B)
IH	TO-INFIN COMPLETE VI	ZC	(A,B,) AND (C) (DROP)
II	TO-INFIN COPULA	ZD	(A) AND (B) (DROP)
IN	INTERROG PRN SUBJECT	ZM	COMMA, AND, OR (DROP)

Figure 3. List of Predictions.

monic interpretation (SWC CODE) is selected among those originally given as in Fig. 4. Classes mmm or nnn and aaa or aab account for features common to several noun and adjective classes respectively, and have been introduced explicitly to achieve certain practical economies. However, only the parent class appears in this column of the analysis output. For example, although the rule (4X, mmm) accounts for "increase" as "nou", it is "nou" which appears as SWC in Fig. 5.

The data in the "SYNTACTIC ROLE" column of Fig. 5 give a rough idea of the role of each word in the sentence. The syntactically and semantically acceptable sentence structure produced by the analyzer is exhibited in more explicit detail by the tree in Fig. 6. This tree is based in an obvious way on the data in the "SENTENCE STRUCTURE" column of Fig. 5; the latter format, which is easier to lay out on a standard printer than a tree, is produced by an editing program from the dpa output which will be described shortly. The structure symbols used in both representations are defined in Fig. 7. The tree representation, which is both intuitively appealing and useful in certain applications, has features of both phrase structure and dependency trees (Hays[9]); its nature is examined more closely by Greibach (Section 3)[8], and in Section 3 of this paper.

The heart of the output, corresponding to the output of a dpa, is given in the columns "RL NUM" (Rule Number) and "PREDICTION POOL" of Fig. 5. Before the processing of "flow", the pushdown store holds | PD- | VS-A | NQ-G . The couple (NQ, nou) specifies the rule that accepted "flow" as nou used attributively. The right-hand element of the corresponding production (NQ, nou) →

```
THE FOLLOWING ANALYSES WERE MADE USING ENGLISH GRAMMAR     000068   PREDICTION POOL SIZE =    100   NESTER MAXIMUM =     5
ANALYSES OF SENTENCE NUMBER   000001

WORD                   HOMOGRAPHS
-----                  ------------
THE                    AAA   ART
INCREASE               NNNS  MMMS  NOUS  VT1P  IT1   VI1P  II1
IN                     PRE   AV2
FLOW                   NNNS  MMMS  NOUS  VI1P  II1   VT1P  IT1
STRESS                 NNNS  MMMS  NOUS  VT1P  IT1
WAS                    BE1S  BE2S  BE3S
ATTRIBUTED             VT1C  PT1
TO                     TCIS  PRF
VACANCIES              NNNP  MMMP  NOUP
,                      CMA
WHICH                  RL1   RL2   IPN   IPO   IAD
HAVE                   HV1   HAVP  VT1P  IT1   VT3P  IT3   VT4P  IT4   VT5P  IT5
APPRECIABLE            AAA   ADJ
MOBILITY               NNNS  MMMS  NOUS
AT                     PRF
-72                    NNNC  MMMC  NUMC
                       PRD
```

Figure 4. Coding of Sentence 1.

***** ANALYSIS NUMBER OF SENTENCE NUMBER 000001

ENGLISH	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
						SE
THE	1SA	ART	PRO-ADJECTIVE	SUBJECT OF PREDICATE VERB	SEAAA0	
						PD VZA4ZA
INCREASE	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	4XMMM0	
						PD VSA
IN	1SPR	PRE	PREPOSITION	PREPOSITION	VXPRE0	
						PD VSANQG
FLOW	1SPOA	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNOU0	
						PD VSAN8G
STRESS	1SPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N8MMM0	
						PD VSA
WAS	1VX	BE3S	BE3-AUXILIARY	PREDICATE VERB	VXBE30	
						PD PAA
ATTRIBUTED	1V	PT1	PAST P OF VT1	PREDICATE VERB	PAPT10	
						PD
TO	1VPR	PRE	PREPOSITION	PREPOSITION	PDPRE0	
						PD NQG
VACANCIES	1VPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNN2	
						PD AC
,	1VPO,	CMA	COMMA	INSERTION	ACCMA0	
						PD AC
WHICH	1VPO7S	RL1	RELATIVE PRN NOM	SUBJECT OF PREDICATE VERB	ACRL10	
						PD VCF
HAVE	1VPO7V	VT1P	SINGLE OBJECT VT	PREDICATE VERB	VXVT11	
						PD N2A
APPRECIABLE	1VPO7CA	ADJ	ADJECTIVE 1	OBJECT OF PREDICATE VERB	N2AAA0	
						PD N5A
MOBILITY	1VPO7C	NOUS	NOUN 1	OBJECT OF PREDICATE VERB	N5MMM0	
						PD
AT	1VPO7CPR	PRE	PREPOSITION	PREPOSITION	PDPRE0	
						PD NQG
-72	1VPO7CPO	NUMC	NOUN 2	OBJECT OF PREPOSITION	NQNNN0	
						PD
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRD0	

PCOL OVERFLOWS= 0 NUMBER TEST FAILURES= 14 SHAPER OVERFLOWS= 550 NESTER OVERFLOWS= 155 TIME= 0.0 MINUTES

Figure 5. Analysis of Sentence 1.

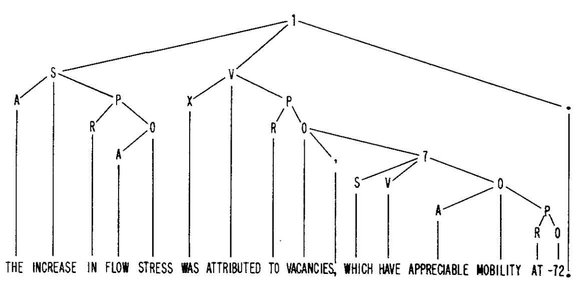


Figure 6. Tree for Sentence 1.

1	declarative	S	subject
2	interrogative	V	verb
3	imperative	O	object
4	subject clause	C	complement
5	object clause	D	adverb
6	complement clause	P	phrase
7	adjective clause	A	attributive
8	adverbial clause	M	participle
		G	gerund
		X	auxiliary verb
		R	phrase or clause introducer (preposition or conjunction)
		E	adverbial noun phrase
		.	period
		,	comma
		+	and/or/but
		=	question mark

Figure 7. Structure Symbols.

nou N8 replaces NQ-G in the pushdown store yielding | PD- | VS-A | N8-G | as a new state for the subsequent processing of "stress".

The dpa itself treats certain adverbs and prepositional phrases as "floating" structures, since little is understood as yet about reliable ways of relating them to the structures they modify. This is reflected, for example, by the fact that, although the VS prediction accepts "in" as a (floating) preposition, it is restored to the pushdown store by the production (VX, pre) → pre NQ VX‡. In the editing process, however, certain experimental assumptions have been made. Thus, for example, "in flow stress" and "to vacancies" are provisionally connected to "increase" and to "attributed" respectively. The matter turned out all right in this case, but later examples will show that this success, regrettably, is not universal, and many interesting open questions remain. The under-

‡ This one generic production serves to handle not only VS ("S" for singular) but also VP ("P" for plural) and similar variants denoted by "X".

lying dpa output readily lends itself to experimentation with a variety of potentially useful or elegant representations.

The actual analyzer is not in fact rigorously a dpa. For one thing, it is truncated in a way that reduces it to what any operating machine is for all practical purposes, namely, a finite state machine. Moreover, certain departures are made for the sake of operating economy from the straightforward specification of productions and from strict pushdown store operation. As mentioned regarding the agreement test, Greibach has shown that these departures have no theoretical significance.

So-called "droppable" predictions (Greibach[8], Section 2.3) are another case in point. Their introduction to condense certain pairs of productions into one led to the elimination of 900 productions from the grammar table and a speed-up of machine operation by a factor of 2.5.

The final line of Fig. 5 tells about various tests made during the analysis. The program is written so that only a certain maximum number of predictions (100 in the current version) can be stored in the prediction pool at one time. The maximum must be large enough to allow the pool to accommodate a great many pairs of predictions which are droppable. It was sufficient for the analysis of Sentence 1, as shown by "POOL OVERFLOWS = 0".

The number of otherwise successful lookups in the grammar table which were discarded because they failed the agreement test between the fulfilled prediction (e.g., of a 3rd person singular verb) and the processed syntactic word class (e.g., VT1*P*) is given as "Number Test Failures". Fourteen paths were discontinued with the help of the agreement test, as is shown by "NUMBER TEST FAILURES = 14". These paths probably pertained to the interpretation of "flow" or "stress" as predicate verb of subject "increase" ("The increase in flow stresses . . .).

"Shaper Overflows" indicates the number of paths that were discontinued because of the shaper test, which eliminates pools such that the number of words remaining to be processed is less than the minimum number needed to fulfill the remaining predictions in the pool. There were 550 such instances in the analysis of Sentence 1.

"Nester Overflows" indicates the number of paths that were discontinued because of the nesting test, a comparison of the number of non-droppable predictions in the prediction pool against an allowable maximum each time a new pool is formed. This test effectively truncates the dpa.

Since the number of non-droppable predictions in an active pool corresponds roughly to the depth of nesting of the next word class to be processed on the assumption that all the droppable predictions will eventually be dropped, and in line with the hypothesis of Yngve[20] that English sentences usually do not have a depth of nesting greater than about seven, it is expected that a small finite maximum number of predictions will suffice for the processing of well-formed sentences from natural habitats. At any stage of the analysis of a sentence, therefore, any prediction pool containing more than the maximum number of predictions can be discarded on the assumption that it predicts a depth of nesting never reached by well-formed sentences.

The maximum number was originally set at 12 in order to gain confidence that legitimate paths would be discontinued on this basis only very rarely, if at all. It turned out that no legitimate analysis had a prediction pool which contained more than *six* non-dropped predictions at any stage of the analysis. It is therefore reasonably improbable that a legitimate analysis will be lost because of the nesting test. In case of serious doubt, the maximum can be readily raised albeit at an unpleasant price in machine time. The version of the program with which Sentence 1 was processed had the maximum depth of nesting set to 8. 155 paths were discontinued due to the nesting test. Experiments to test the effect of limiting the maximum extent of self-embedding are also under way.

The analysis of the sentence took less than 0.1 minutes.

2.3 Figure 8 shows the word class coding of Sentence 2, which reads "Economic studies show that it could be a billion-dollar-a-year

ANALYSES OF SENTENCE NUMBER 000002

WORD	HOMOGRAPHS
ECONOMIC	AAA ADJ
STUDIES	VT1S VI1S NNNP MMMP NOUP
SHOW	VT1P IT1 VI1P II1 VT2P IT2 VT6P IT6 VT7P IT7 NNNS MMMS NOUS
THAT	CC1 NNNS PRZS AAA ART CCO RL1 RL2 RL6
IT	TITS PRNS PRO
COULD	AUXC
BE	BE1Y BE2Y BE3Y BI1 BI2 BI3
A	AAA ART
BILLION-DOLLAR-A-YEAR	AAA ADJ
BUSINESS	NNNS MMMS NOUS
BY	PRE AV2
THE	AAA ART
1970*S	NNNC MMMC NUMC
	PRD

Figure 8. Coding of Sentence 2.

business by the 1970's.". Four distinct analyses were obtained for this sentence, mainly due to the interpretations of "show" and of "that". It turns out in this case that all but one can be eliminated by appropriate modifications of the grammar.

Analysis No. 1 (Fig. 9) treats "that" as a conjunction (CCO) which introduces an adverbial clause with the meaning of "in order that" or "because of the fact that", and "show" as a complete intransitive verb (VI1). This analysis can be made semantically acceptable in a marginal way by replacing the original word forms by others syntactically equivalent in the sense that they are either classified alike in the present grammar, or belong to distinct classes that produce the same new predictions when fulfilling a given prediction (e.g., PRZ and NOU both fulfill a prediction P for which there are rules (P, nnn)). The substitute forms were manually inserted in the column "ENGLISH SUBSTITUTE".

Although the interpretation of "that" as CCO is somewhat far-fetched in this particular sentence, the coding of "that" as CCO is needed for such sentences as "It has been kept polished *that* it may glitter forever.", "I am happy *that* you have succeeded." and "I am surprised *that* he did not pass.". Therefore, the possibility of eliminating this interpertation on general grounds is ruled out.

Analysis No. 2 (Fig. 10) treats "show" as a double object transitive verb (VT2), "that" as an indirect object of "show", and "the 1970's" as a direct object of the verb. In this analysis, "that" is modified by the adjective clause ("7") "it could be a billion-dollar-a-year business by". The indicated substitutions make this structure quite plausible so that it too cannot be eliminated on general grounds.

A minor but confusing flaw of Fig. 10 should be pointed out. Although the prediction of an indirect object is represented by "NQ" and that

***** ANALYSIS NUMBER OF SENTENCE NUMBER 000002

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SMC	SMC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
ECONOMIC	DILIGENT	1SA	ADJ	ADJECTIVE 1	SUBJECT OF PREDICATE VERB	SEAAA0	SE
STUDIES	UNDERGRADUATES	1S	NOUP	NOUN 1	SUBJECT OF PREDICATE VERB	4XMMM0	PD VZA4ZA
SHOW	WORKED	1V	VI1P	COMPLETE VI	PREDICATE VERB	VXVI10	PD VPA
THAT	(SO) THAT	.1VBR	CCO	ADVERB CONJ 2	CONJUNCTION	PDCCO1	PD
IT	THEY	1VBS	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	SGPRN0	PD SGG
COULD	COULD	1VBVX	AUXC	AUXILIARY VERB	PREDICATE VERB	VXAUX0	PD VSG
BE	BE	1VBV	BI2	INFINITE BE2	PREDICATE VERB	BVBI21	PD BVA
A	HONORS	1VBCA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAA0	PD N3A
BILLION-DOLLAR-A-YEAR		1VBCA	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	N6ADJ0	PD N6A
BUSINESS	CANDIDATES	1VBC	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMM0	PD N6A
BY	IN	1VBCPR	PRE	PREPOSITION	PREPOSITION	PDPRE0	PD
THE	TWO	1VBCPCA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAA0	PD NQG
1970*S	YEARS	1VBCPC	NUMC	NOUN 2	OBJECT OF PREPOSITION	N5MMM0	PD N5G
.	.	1.	PRD	PERIOD	END OF SENTENCE	PDPRD0	PD

Figure 9. Analysis No. 1 of Sentence 2.

***** ANALYSIS NUMBER 2 OF SENTENCE NUMBER 000002

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
ECCNOMIC	AUTOMATIC	1SA	ADJ	ADJECTIVE 1	SUBJECT OF PREDICATE VERB	SEAAAO	SE
STUDIES	TRANSLATION	1S	NOUP	NOUN 1	SUBJECT OF PREDICATE VERB	4XMMMO	PD VZA4ZA
SHOW	SHOWS	1V	VT2P	DOUBLE OBJECT VT	PREDICATE VERB	VXVT2O	PD VPA
THAT	THOSE	1C	PRZS	INDEFINITE PRN	OBJECT OF PREDICATE VERB	NONNN2	PD N2ANQA
IT	[(WHOM) IT	1C7S	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	ACPRNO	PD N2AAC
COULD	HAS	1C7VX	AUXC	AUXILIARY VERB	PREDICATE VERB	WXAUXO	PD N2AWSF
BE	BEEN	1C7V	BI2	INFINITE BE2	PREDICATE VERB	BWBI21	PD N2ABWA
A	DEPENDENT	1C7CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD N2ADQ N3A
BILLICN-DOLLAR-A-YEAR		1C7CA	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	N6ADJO	PD N2ADQ N6A
BUSINESS		1C7C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD N2ADQ N6A
BY	UPON]	1C7CPR	PRE	PREPOSITION	PREP PHRASE DISCONTINUOUS	DQPREO	PD N2ADQ
THE	THE	1CA	ART	PRO-ADJECTIVE	OBJECT OF PREDICATE VERB	N2AAAO	PD N2A
1970*S	DIFFICULTIES OF...	1C	NUMC	NOUN 2	OBJECT OF PREDICATE VERB	N5MMMO	PD N5A
.	.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 10. Analysis No. 2 of Sentence 2.

of a direct object by "N2" in the subrule (VX, VT2)-O, the two structures are not distinguished by the current diagramming routine: both are represented by the same structure symbol "O". Therefore, in the structure diagram of Fig. 10, it looks as if the basic pattern of the sentence were "S" (subject) -"V" (verb) -"O" (object) -"." (period), although the presence of two "1O's", one for "that" and the other for "1970's", indicates that the sentence has two distinct object heads. The boundaries of the indirect and direct objects are not explicit in the diagram, but have to be identified with the aid of the pushdown history in the column "PREDICTION POOL". The distinction between structure symbols for an indirect and direct object has to be embodied in the diagramming routine.

An Analysis No. 3 (Fig. 11), "that" is regarded as a noun conjunction (CO1) which introduces a nominal object clause ("5") of "show" as a noun clause transitive verb (VT6). This is the analysis which is semantically acceptable except for the dependency of the floating prepositional phrase "by the 1970's" which was not handled as well here as similar structures in Sentence 1, for reasons mentioned in Section 2.2.

In Analysis No. 4 (Fig. 12), "that" is regarded as an indirect object of the VT7 "show", with "it could be a billion-dollar-a-year business by the 1970's", without an introductory conjunction, interpreted as the object clause of "show". Here again plausible substitutions exist, so that elimination on general grounds is not indicated.

One way of eliminating Analysis No. 1 of Sentence 2 is to preclude the use of "that" by itself as a CCO except when preceded by certain adjectives and past participles of "emotion" as in "I am *happy* (*glad*, *sorry*, etc.) that you have succeeded" and "I am *surprised* (*disappointed*, *delighted*, etc.) that you have passed". Indeed,

***** ANALYSIS NUMBER 3 OF SENTENCE NUMBER 000002

ENGLISH	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
ECCNOMIC	1SA	ADJ	ADJECTIVE 1	SUBJECT OF PREDICATE VERB	SEAAAO	SE
STUDIES	1S	NOUP	NOUN 1	SUBJECT OF PREDICATE VERB	4XMMMO	PD VZA4ZA
SHOW	1V	VT6P	NOUN CLAUSE VT	PREDICATE VERB	VXVT6O	PD VPA
THAT	15R	CO1	NOUN CONJUNCTION	CONJUNCTION	NCCO1O	PD NCD
IT	15S	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	SGPRNO	PD SGD
COULD	15VX	AUXC	AUXILIARY VERB	PREDICATE VERB	VXAUXO	PD VSD
BE	15V	BI2	INFINITE BE2	PREDICATE VERB	BVBI21	PD BVA
A	15CA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAAO	PD N3A
BILLICN-DOLLAR-A-YEAR	15CA	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	N6ADJO	PD N6A
BUSINESS	15C	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMMO	PD N6A
BY	15CPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
THE	15CPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD NQG
1970*S	15CPO	NUMC	NOUN 2	OBJECT OF PREPOSITION	N5MMMO	PD N5G
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 11. Analysis No. 3 of Sentence 2.

***** ANALYSIS NUMBER 4 OF SENTENCE NUMBER 000002

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
							SE
ECONOMIC	ECONOMIC	1SA	ADJ	ADJECTIVE 1	SUBJECT OF PREDICATE VERB	SEAAA0	
							PD VZA4ZA
STUDIES	STUDIES	1S	NOUP	NOUN 1	SUBJECT OF PREDICATE VERB	4XMMM0	
							PD VPA
SHOW	SHOW	1V	VT7P	OBJ-NOUN CL VT	PREDICATE VERB	VXVT70	
							PD SGDNQA
THAT	THEM	1C	PRZS	INDEFINITE PRN	OBJECT OF PREDICATE VERB	NQNNN0	
							PD SGD
IT	(THAT) IT	1SS	PRNS	PERSONAL PRN NOM	SUBJECT OF PREDICATE VERB	SGPRN0	
							PD VSD
COULD	COULD	1SVX	AUXC	AUXILIARY VERB	PREDICATE VERB	VXAUX0	
							PD BVA
BE	BE	1SV	BI2	INFINITE BE2	PREDICATE VERB	BVBI21	
							PD N3A
A	"	1SCA	ART	PRO-ADJECTIVE	COMPLEMENT OF PREDICATE V	N3AAA0	
							PD N6A
BILLION-DOLLAR-A-YEAR		1SCA	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	N6ADJ0	
							PD N6A
BUSINESS		1SC	NOUS	NOUN 1	COMPLEMENT OF PREDICATE V	N6MMM0	
							PD
BY		1SCPR	PRE	PREPOSITION	PREPOSITION	PDPRE0	
							PD NQG
THE		1SCPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAA0	
							PD N5G
1970*S		1SCPO	NUMC	NOUN 2	OBJECT OF PREPOSITION	N5MMM0	
							PD
.		1.	PRD	PERIOD	END OF SENTENCE	PDPRD0	

PCOL OVERFLOWS= 0 NUMBER TEST FAILURES= 0 SHAPER OVERFLOWS= 1264 NESTER OVERFLOWS= 62 TIME= 0.1 MINUTES

Figure 12. Analysis No. 4 of Sentence 2.

the probability of the occurrences of such sentences as "It has been kept polished *that* it may glitter forever." will be fairly low since one would more often say "so that" or "in order that" on such occasions.

The emegence of Analysis No. 1 may also be attributed to the coding of "show" as an intransitive verb (VI1) for sentences such as "They *show up* every morning at eight." or "The tuberculosis tests often *show up* positive.". Since "show" as an intransitive verb seems always to require a special adverb to follow it, establishing such a subclass of VI1 and making a prediction of such an adverb will make it possible to discard Analysis No. 1 of Sentence 2.

In Analysis No. 2 "that" is interpreted as the indirect object of a double object transitive verb (VT2). It is common to use "that" as an indirect object of a VT2, as in "He gave *that* serious consideration." which means "He gave serious consideration to *that* (matter).". However, the occurrence of "that" as PRZ modified by an adjective clause which is not itself introduced by the relative pronoun "which", either in the nominative case (RL1) or in the accusative case (RL2), has some unusual features. It is awkward but admissible to say "He does that which pleases most of his constituents.". ("He does what pleases . . . " is smoother), but it is poetic, perhaps normally ungrammatical, to omit "which" and say "He has that all people desire."; the spoken version requires intonational gymnastics to be understood, and like the written version, seems more at home in a sermon than in scientific prose. Analysis No. 2 could be deleted by prohibiting "that" as PRZ from being modified by an adjective clause not introduced by "which".

As for the fourth analysis, on the assumption that "that" as PRZ is used as the object of a VT7 only in sentences one might address to children such as "We tell that when and where it should stop.", with "that" meaning "(to) that toy", and that such sentences most likely never appear in scientific papers (Would Piaget accept this?), one could eliminate Analysis No. 4 by prohibiting the acceptance of "that" by the indirect object prediction ("NQ") for verbs of category VT7. At such junctures one must be prepared to make explicit decisions about what will be regarded as grammatical and what will not, and assess the consequences of these decisions!

2.4 Sentence 3, "Slime formation is dependent on size of particles formed by mechanical means, amount of metal in the amalgam, and purity of solutions." was coded as shown in Fig. 13. The five analyses obtained for this sentence are shown in Figs. 14 through 18.

The selected syntactic word classes are the same in the first two (Figs. 14 and 15). It therefore is not homographs, but multiple functions of word classes that give rise to these two analyses. The differences between the two can best be appreciated by looking at the structure symbols which the "," and "+" symbols connect together in the sentence structure diagrams. In Analysis No. 1, two ","s and a "+" appear at the same level, showing that "means", "amount" and "purity" are all objects of the preposition "(formed) by" (i.e., "formed by

ANALYSES OF SENTENCE NUMBER 000003

WORD	HOMOGRAPHS				
-SLIME	NNNS	MMMS	NOUS	VT1P	IT1
FORMATION	NNNS	MMMS	NOUS		
IS	BE1S	BE2S	BE3S		
DEPENDENT	AAA	ADJ	NNNC	MMMC	NOVC
ON	PRE	AV2			
SIZE	NNNS	MMMS	NOUS	VT1P	IT1
OF	PRE				
PARTICLES	NNNP	MMMP	NOUP		
FORMED	VT1C	PT1	VI1C	PI1	
BY	PRE	AV2			
MECHANICAL	AAA	ADJ			
MEANS	NNNS	MMMS	NOUS	VT1S	VT6S
,	CMA				
AMOUNT	NNNC	MMMC	NOUC	VI3P	II3
OF	PRE				
METAL	NNNS	MMMS	NOUS		
IN	PRE	AV2			
THE	AAA	ART			
AMALGAM	NNNS	MMMS	NOUS		
,	CMA				
AND	XCO				
PURITY	NNNS	MMMS	NOUS		
OF	PRE				
SOLUTIONS	NNNP	MMMP	NOUP		
.	PRD				

Figure 13. Coding of Sentence 3.

***** ANALYSIS NUMBER 1 OF SENTENCE NUMBER 000003

ENGLISH	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
SLIME	1SA	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENOUO	SE
FORMATION	1S.	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	7XMMMO	PD VZA7ZA
IS	1V	BE2S	BE2-COPULA	PREDICATE VERB	VXBE20	PD VSA
DEPENDENT	1C	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	AIADJO	PD AIA
ON	1CPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SIZE	1CPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQG
OF	1CPOPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
PARTICLES	1CPOPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NONNN1	PD NQG
FORMED	1CPOPOPM	PT1	PAST P OF VT1	POST-POSITIONAL PART-ADJ	APPT10	PD AP
BY	1CPOPOPMPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
MECHANICAL	1CPOPOPMPOA	ADJ	ADJECTIVE 1	OBJECT OF PREPOSITION	NQAAAO	PD NQG
MEANS	1CPOPOPMP$\boxed{0}$	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMM3	PD NSG
	1CPOPOPMP$\boxed{,}$	CMA	COMMA	COMPOUND OBJECT	CNCMAO	PD NQGXCBNQGCNB
AMOUNT	1CPOPOPMP$\boxed{0}$	NOUC	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQGXCBNQG
OF	1CPOPOPMPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
METAL	1CPOPOPMPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQGXCBNQG
IN	1CPOPOPMPOPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
THE	1CPOPOPMPOPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD NQGXCBNQG
AMALGAM	1CPOPOPMPOPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD NQGXCBNSG
,	1CPOPOPMP$\boxed{,}$	CMA	COMMA	COMPOUND OBJECT	XCCMAO	PD NQGXCB
AND	1CPOPOPMP$\boxed{+}$	XCO	COORDINATE CONJ1	COMPOUND OBJECT	XCXCOO	PD NQGXCB
PURITY	1CPOPOPMP$\boxed{0}$	NOUS	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQG
OF	1CPOPOPMPOPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SOLUTIONS	1CPOPOPMPOPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQG
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 14. Analysis No. 1 of Sentence 3.

means . . . , amount, . . . and purity . . . "). In Analysis No. 2, on the other hand, "size", "amount" and "purity" appear at the same level, forming a three-member object of "(dependent) on", (i.e., "dependent on size . . . , amount, . . . and purity . . . "). Although it is the second analysis that is semantically acceptable for this particular sentence, the first analysis is syntactically as legitimate as the second one. (See Type 2 ambiguity, Section 4.1 of Greibach[8]). Rejecting it in this case requires much deeper insight into semantics than

***** ANALYSIS NUMBER 2 OF SENTENCE NUMBER 000003

ENGLISH	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
SLIME	1SA	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENOUO	SE
FORMATION	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	7XMMMO	PD VZA7ZA
IS	1V	BE2S	BE2-COPULA	PREDICATE VERB	VXBE20	PD VSA
DEPENDENT	1C	ADJ	ADJECTIVE 1	COMPLEMENT OF PREDICATE V	AIADJO	PD AIA
ON	1CPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SIZE	1CP[C]	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNN3	PD NQG
OF	1CPOPR	PRE	PREPOSITION	PREPOSITION	CNPREO	PD NQGXCBNQGCNB
PARTICLES	1CPOPC	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNN1	PD NQGXCBNQGCNBNQG
FORMED	1CPCPOPM	PT1	PAST P OF VT1	POST-POSITIONAL PART-ADJ	APPT10	PD NQGXCBNQGCNBAP
BY	1CPOPOPMPR	PRE	PREPOSITION	PREPOSITION	CNPREO	PD NQGXCBNQGCNB
MECHANICAL	1CPOPOPMPOA	ADJ	ADJECTIVE 1	OBJECT OF PREPOSITION	NQAAAO	PD NQGXCBNQGCNBNQG
MEANS	1CPOPOPMPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD NQGXCBNQGCNBN5G
,	1CP[,]	CMA	COMMA	COMPOUND OBJECT	CNCMAO	PD NQGXCBNQGCNB
AMOUNT	1CP[C]	NOUC	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQGXCBNQG
OF	1CPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
METAL	1CPOPC	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQGXCBNQG
IN	1CPOPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
THE	1CPOPCPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD NQGXCBNQG
AMALGAM	1CPOPCPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD NQGXCBN5G
,	1CP[,]	CMA	COMMA	COMPOUND OBJECT	XCCMAO	PD NQGXCB
AND	1CP[+]	XCO	COORDINATE CONJ1	COMPOUND OBJECT	XCXCOO	PD NQGXCB
PURITY	1CP[C]	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQG
OF	1CPCPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SOLUTIONS	1CPCPC	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQG
.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 15. Analysis No. 2 of Sentence 3.

***** ANALYSIS NUMBER 3 OF SENTENCE NUMBER 000003

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
SLIME	THE	1SA	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENOUO	SE
FORMATION	FACT	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	7XMMMO	PD VZA7ZA
IS	IS	1V	BE2S	BE2-COPULA	PREDICATE VERB	VXBE23	PD VSA
DEPENDENT	(THAT) PEOPLE	16S	NOVC	NOUN 3	SUBJECT OF PREDICATE VERB	SGNNNO	PD SGE
ON	AT	16SPR	PRE	PREPOSITION	PREPOSITION	VXPREO	PD VCE
SIZE	THE	16SPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD VCENQG
OF	TIME	16SPOPR	PRE	PREPOSITION	PREPOSITION	VXPREO	PD VCE
PARTICLES		16SPOPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD VCENQG
FORMED	BELIEVED	16V	VI1C	COMPLETE VI	PREDICATE VERB	VXVI10	PD VCE
BY	IN	16VPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
MECHANICAL	HUMAN	16VPOA	ADJ	ADJECTIVE 1	OBJECT OF PREPOSITION	NQAAAO	PD NQG
MEANS	RIGHTS	16VPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMM3	PD N5G
,	,	16VP,	CMA	COMMA	COMPOUND OBJECT	CNCMAO	PD NQGXCBNQGCNB
AMOUNT	(THE) IMPORTANCE	16VPO	NOUC	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQGXCBNQG
OF	OF	16VPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
METAL	DIGNITY	16VPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQGXCBNQG
IN	OF	16VPOPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQGXCB
THE	THE	16VPOPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD NQGXCBNQG
AMALGAM	INDIVIDUAL	16VPOPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD NQGXCBN5G
,	,	16VP,	CMA	COMMA	COMPOUND OBJECT	XCCMAO	PD NQGXCB
AND	AND	16VP+	XCO	COORDINATE CONJ1	COMPOUND OBJECT	XCXCOO	PD NQGXCB
PURITY	FREEDOM	16VPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQG
OF	OF	16VPOPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SOLUTIONS	SPEECH	16VPOPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQG
.	.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 16. Analysis No. 3 of Sentence 3.

is now available. Cases such as these underline the great importance of retaining human links in any chain for natural language data processing and the danger of relying on any method of syntactic analysis that does not properly account for ambiguities.

The emergence of a compound object "amalgam, and purity" or "metal, and purity" has been precluded since the current grammar regards as ill-formed the use of a comma for a two-member compound noun phrase. This structure has been excluded from the grammar

***** ANALYSIS NUMBER 4 OF SENTENCE NUMBER 000003

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
SLIPE	THE	1SA	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENOUO	SE
FORMATION	FACT	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	7XMMMO	PD VZATZA
IS	IS	1V	BE2S	BE2-COPULA	PREDICATE VERB	VXBE23	PD VSA
DEPENDENT	(THAT) PEOPLE	16S	NOVC	NOUN 3	SUBJECT OF PREDICATE VERB	SGNNNO	PD SGE
ON	UP (TILL NOW)	16D	AV2	ADVERB 2	ADVERB	VXAV20	PD VCE
SIZE	SUPPORTED	16V	VT1P	SINGLE OBJECT VT	PREDICATE VERB	VXVT11	PD VCE
OF	[EXCEPT FOR	16VPR	PRE	PREPOSITION	PREPOSITION	N2PREO	PD N2A
PARTICLES	THOSE	16VPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NONNN1	PD N2ANOG
FORMED	REFERRED	16VPOPM	P11	PAST P OF V11	POST-POSITIONAL PART-ADJ	APPI10	PD N2AAP
BY	TO (BEFORE)]	16VPOPMPR	PRE	PREPOSITION	PREP PHRASE DISCONTINUOUS	DQPREO	PD N2ADQ
MECHANICAL	PURE	16OA	ADJ	ADJECTIVE 1	OBJECT OF PREDICATE VERB	N2AAAO	PD N2A
MEANS	DEMOCRACY	16O	NOUS	NOUN 1	OBJECT OF PREDICATE VERB	N5MMM3	PD N5A
,	,	16,	CMA	COMMA	COMPOUND OBJECT	CNCMAO	PD NQAXCBNQACNB
AMOUNT	PRINCIPLE	16O	NOUC	NOUN 1	OBJECT OF PREDICATE VERB	NQNNNO	PD NQAXCBNQA
OF	OF	16OPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQAXCB
METAL	INVIOLABILITY	16OPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD NQAXCBNQG
IN	OF	16OPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD NQAXCB
THE	INDIVIDUAL	16OPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD NQAXCBNQG
AMALGAM	RIGHTS	16OPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD NQAXCBN5G
,	,	16,	CMA	COMMA	COMPOUND OBJECT	XCCMAO	PD NQAXCB
AND	AND	16+	XCO	COORDINATE CONJ1	COMPOUND OBJECT	XCXCOO	PD NQAXCB
PURITY	FREEDOM	16O	NOUS	NOUN 1	OBJECT OF PREDICATE VERB	NQNNNO	PD NQA
OF	OF	16OPR	PRE	PREPOSITION	PREPOSITION	PDPREO	PD
SOLUTIONS	SPEECH	16OPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD NQG
.	.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

Figure 17. Analysis No. 4 of Sentence 3.

***** ANALYSIS NUMBER 5 OF SENTENCE NUMBER 000003

ENGLISH	ENGLISH SUBSTITUTE	SENTENCE STRUCTURE	SWC	SWC CODE	SYNTACTIC ROLE	RL NUM	PREDICTION POOL
SLIPE	THE	1SA	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	SENOUO	SE
FORMATION	FACT	1S	NOUS	NOUN 1	SUBJECT OF PREDICATE VERB	TXMMMO	PD VZATZA
IS	IS	1V	BE2S	BE2-COPULA	PREDICATE VERB	VXBE23	PD VSA
DEPENDENT	(THAT) THE SYSTEM	16S	NOVC	NOUN 3	SUBJECT OF PREDICATE VERB	SGNNNO	PD SGE
ON	UP (TILL NOW)	16D	AV2	ADVERB 2	ADVERB	VXAV20	PD VCE
SIZE	IGNORED	16V	VT1P	SINGLE OBJECT VT	PREDICATE VERB	VXVT11	PD VCE
OF	[ON	16VPR	PRE	PREPOSITION	PREPOSITION	N2PREO	PD N2A
PARTICLES	THE PRINCIPLE	16VPO	NOUP	NOUN 1	OBJECT OF PREPOSITION	NONNN1	PD N2ANOG
FORMED	AGREED	16VPOPM	P11	PAST P OF V11	POST-POSITIONAL PART-ADJ	APPI10	PD N2AAP
BY		16VPOPMPR	PRE	PREPOSITION	PREP PHRASE DISCONTINUOUS	DQPRE1	PD N2ADQ
MECHANICAL		16VPOPMPOA	ADJ	ADJECTIVE 1	OBJECT OF PREPOSITION	NQAAAO	PD N2ADQ NQG
MEANS		16VPOPMPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMM3	PD N2ADQ N5G
,		16VPOPMP,	CMA	COMMA	COMPOUND OBJECT	CNCMAO	PD N2ADQ NQGXCBNQGCNB
AMOUNT		16VPOPMPO	NOUC	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD N2ADQ NQGXCBNQG
OF		16VPOPMPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD N2ADQ NQGXCB
METAL	*	16VPOPMPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NONNNO	PD N2ADQ NQGXCBNQG
IN		16VPOPMPOPOPR	PRE	PREPOSITION	PREPOSITION	XCPREO	PD N2ADQ NQGXCB
THE		16VPOPMPOPOPOA	ART	PRO-ADJECTIVE	OBJECT OF PREPOSITION	NQAAAO	PD N2ADQ NQGXCBNQG
AMALGAM		16VPOPMPOPOPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	N5MMMO	PD N2ADQ NQGXCBN5G
,		16VPOPMP,	CMA	COMMA	COMPOUND OBJECT	XCCMAO	PD N2ADQ NQGXCB
AND		16VPOPMP+	XCO	COORDINATE CONJ1	COMPOUND OBJECT	XCXCOO	PD N2ADQ NQGXCB
PURITY		16VPOPMPO	NOUS	NOUN 1	OBJECT OF PREPOSITION	NQNNNO	PD N2ADQ NQG
OF	UPON]	16VPOPMPR	PRE	PREPOSITION	PREP PHRASE DISCONTINUOUS	DQPREO	PD N2ADQ
SOLUTIONS	IDIOMATIC EXPRESSIONS	16O	NOUP	NOUN 1	OBJECT OF PREDICATE VERB	N2NNNO	PD N2A
.	.	1.	PRD	PERIOD	END OF SENTENCE	PDPRDO	PD

POOL OVERFLOWS= 0 NUMBER TEST FAILURES= 3 SHAPER OVERFLOWS= 13437 NESTER OVERFLOWS= 20645 TIME= 1.3 MINUTES

Figure 18. Analysis No. 5 of Sentence 3.

not because it would be difficult to recognize it as well-formed, but rather because its inclusion at this time would cause an excessive increase in the number of semantically unacceptable analyses for common sentence types which do not have such a structure among their normal semantically acceptable analyses. For example, a sentence such as "Time passes, and the world changes." would give two semantically unacceptable analyses if a compound noun phrase "amalgam, and purity" were allowed as well-formed. In one analysis, "time (NOU) passes

(NOU), and the world" would be regarded as a compound subject of "changes (VI1)", while in the second analysis, "passes (NOU), and the world (NOU) changes (NOU)" would be regarded as a compound object of the imperative verb "time (IT1)".

The emergence of a three-member noun phrase "particles, amount, and purity" has also been precluded since the current grammar does not accept a post-positional adjective, participle or clause which modifies the first member of a compound noun phrase. Rules can readily be added to the grammar to enable the analyzer to accept these structures. Such rules, if embodied in the grammar, would yield a semantically and syntactically acceptable analysis for sentences such as "I like wine *imported from France*, beer from Germany and sake from Japan.", although they would have the unpleasant effect of producing a semantically unacceptable analysis "particles, amount, and purity" in cases such as Sentence 3.

The remaining three analyses of Sentence 3 represent a structure similar to that of "The fact is smoking kills." in which "smoking kills" constitutes a complement clause of "is". The problem here is that too many means of eliminating these three analyses suggest themselves and that the consequences of any alternative are difficult to predict in detail *a priori*.

One obvious technique would be to treat "The fact is smoking kills." as ill-formed insisting instead on "The fact is: smoking kills.". In the absence of enforceable normative techniques, and that is the usual practical situation, this choice is less attractive than might appear at first thought.

A more promising approach might be based on a refinement of word classes. This leaves open the acceptability of the three analyses under appropriate substitution. Since "formation" does not seem to belong to the category of nouns such as "fact", "plan" and "idea" which, as the subject of a copula "be", can introduce a complement clause, all three analyses could be discarded by refining the nominal class definitions. Again in all three analyses, the word form "dependent" is interpreted as NOVC with the meaning of "one who depends on or looks to another for support", as in "I have one *dependent*." or "The *dependent* and the underprivileged need greater educational opportunities.". The analyses could therefore be deleted also by refining the specification of noun classes in order to group "dependent" with other nouns which cannot form the head of a noun phrase without being preceded by one of such noun phrase introducers as "one", "a", "the", or "my".

In Analysis No. 3 (Fig. 16), the complement clause is composed of "dependent" as subject and "formed" as predicate verb. "Formed" is a complete intransitive verb (VI1C) as in "Ice *formed* under the wings.". The analysis can be made semantically acceptable by replacing the original word forms by those manually inserted in the column "ENGLISH SUBSTITUTE".

In Analysis No. 4 (Fig. 17) the complement clause is composed of "dependent" as plural subject, "size" as predicate verb, and "mechanical means, amount . . . , and purity . . . " as object of the predicate verb. Much remains to be studied about the behavior of adverbs of the class AV2 ("on") which are accepted as floating structures in the current grammar.

The occurrence of a prepositional phrase ("of particles . . . ") between a predicate verb and an object is not uncommon, as in "The author sketches *in the first chapter* an outline of historical and descriptive linguistics.". The interpretation of "formed by" as a post-positional modifier of "particles"—with "formed" as PI1—raises important problems. The current grammar accepts any PI1 as a post-positional modifier if it is followed by PRE. This provision is for structures such as "This is the boy *run over* by a car.", "This is a topic *come across* in various places.". It seems, however, that certain members of PI1 cannot be used as post-positional modifier and that each member of PI1 that can be used as post-positional modifier can be followed only by a limited class of prepositions peculiar to itself. This suggests the necessity for some refinement of verb classification.

Analysis No. 5 (Fig. 18) has a complement clause whose predicate verb "size" governs the object "solutions" which is widely separated from the verb by a long prepositional phrase "of particles . . . purity of". The prepositional phrase has a structure similar to that of "on the principle agreed [intervening prepositional phrase] upon" (see English substitutes in Fig. 18). The asterisk in the column "ENGLISH SUBSTITUTE" indicates that the interpretation given for the corresponding word forms cannot be made semantically acceptable by any English substitutes.

Although Analysis No. 5 can be discarded by any of the techniques proposed in the preceding three paragraphs, its emergence also suggests the necessity for more careful study of floating structures. First, it is possible to establish a prediction of a preposition which cannot accept a floating structure. Such a prediction could be generated after the processing of verbs such as "agreed", "run", "come", and "talked" (all PI1) of "This is a principle *agreed upon* by the people.", "This is the boy *run over* by the car.", "This is a topic *come across* in various places." and "This is a book *talked about* in various circles." respectively, since a floating structure seldom appears between "agreed" and "upon", "run" and "over", and so forth.

The provision would, however, also rule out less frequent structures on the borderline of grammaticality such as "This is the principle *agreed* finally *upon* by the people.", "This is the boy *run* completely *over* by the car.", "This is a topic *come* constantly *across* in various places.", "This is a book *talked* constantly *about* in various circles.". This may or may not be desirable. In any case, although some would agree that the above four sentences with inserted adverbs "finally", "completely", "constantly", and "constantly" are well-formed, even if colloquial and awkward, most would agree that the replacement of each of these adverbs by a longer adverbial phrase would turn the sentences into ill-formed sentences. It would be most unlikely to have sentences such as "This is the principle agreed *finally and unanimously* upon by the people.", or "This is the principle agreed *with no opposition* upon by the people.". The problem here is that the intervening phrases are *too long*.

The criterion of whether an inserted structure is *too long* or *not too long* is quite subjective at this moment. It does not always depend upon the number of words in such a structure, but upon the relative length of the structure in connection with those structures which precede and/or succeed it. Contrast " . . . thereby insuring *against all enemies* the peace and security of . . . " with " . . . thereby insuring *against interference from noise due to excessive crowding of channels* radio astronomy." and with " . . . thereby insuring *against interference from noise due to excessive crowding of channels* not only radio astronomy but also other scientific and communication enterprises that require freedom from interference.".

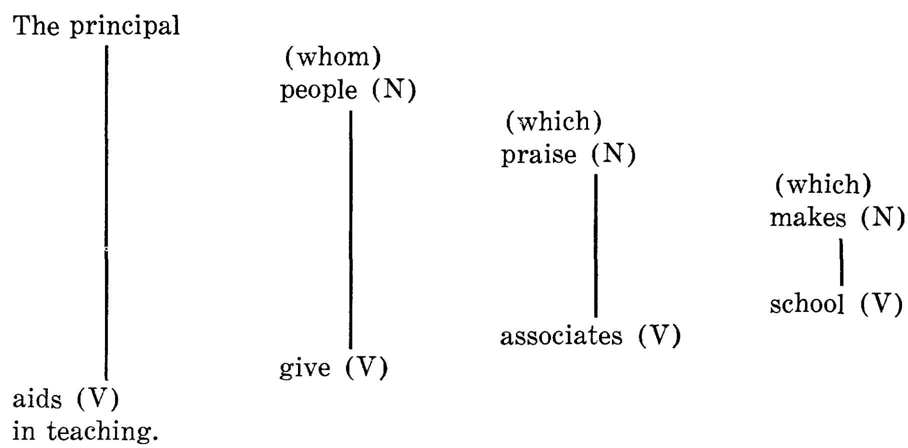
If such a criterion (more or less pertaining to style) could be successfully formalized, the automatic syntactic analysis of languages could be greatly improved.

2.5 The version of the English analyzer (referred to as 1963-FJCC version) used for Sentences 1, 2 and 3 of this section differs from the version (refered to as 1962-IFIP version), described in our previous paper[12, 13], in the following two points: (1) the system has been entirely reprogrammed to attain higher efficiency in program performance, resulting in a speed-up of processing time by a factor of 5 over the 1962-IFIP version; (2) the feature of "droppable" predictions mentioned in Section 2.2 of this paper has been added with an increase in speed by a factor of 2.5. Hence the new version is an order of magnitude faster than the old.

Several other techniques, now being planned for incorporation in the 1963-FJCC version, will eliminate irrelevant paths in syntactic analysis without destroying any paths which may yield acceptable analyses.

(a) Generalized Shaper: At each stage of analysis, the program compares the number of (non-droppable) "comma" predictions and "and" predictions respectively with the number of commas and ands remaining to be processed in the sentence. If the former is greater than the latter, the path is discarded. A similar comparison is to be made between participial predictions and participial word classes. This technique, originated by Plath for Russian,[16] has been experimentally programmed for the 1962-IFIP version where it reduced processing time by a factor of 5. It is expected that this technique, when incorporated in the 1963-FJCC version, will increase the speed by a factor of at least 3. It is yet to be determined where the break-even lies between the time required for making such tests and the time saved by the elimination of irrelevant paths due to such tests.

(b) Self-embedding Test: Independent of the "Nester" test described in Section 2.2, the program checks how many self-embedded structures a given prediction pool contains at each stage of the analysis of a sentence. For initial experiments, any pool which contains more than 3 predicate and clause predictions will be discarded on the assumption that it predicts a structure too deeply self-embedded ever to occur in natural well-formed sentences. This test is expected to be effective especially when a given sentence has a series of contiguous nouns because it would reject the possibility of the first noun being modified by an adjective clause initiated by the second noun (as in "The boy *people* (N) *praise* (V) is . . . ") and the second noun in turn being modified by another adjective clause initiated by the third noun, and so on. For example, this test would accept the syntactically and semantically acceptable interpretation of "The principal people praise makes school associates give aids in teaching." as "The principal (whom) people (N) praise (V) makes (V) school (N) associates (N) give (V) aids (N) in teaching.", while rejecting the interpretation of the same sentence as containing three self-embedded adjective clauses:



The expected processing time of sample sentences by the projected program incorporating these additional features is shown in Fig. 19, together with the actual processing time of the same sentences by the 1962-IFIP version and the 1963-FJCC version.

In addition to these two techniques which are now being programmed, another technique of Plath's for avoiding repetitive local parsings is now being studied for the English analyzer. This technique, already programmed for the Russian analyzer, has proved to be effective for longer sentences by sharply bounding the exponential dependence of processing time on sentence length toward the limiting case of log exp or linear dependence.

3. DIRECTED PRODUCTION ANALYZER AND PHRASE STRUCTURE GENERATORS

3.1 The primary output of a dpa is the sequence of couples (P, c) and of prediction pools. This output specifies the structure of a sentence in a definite and useful way but, standing alone, lacks the intuitive immediacy of the more familiar immediate constituent or dependency tree structural representations. The mapping from this form of output to a more natural tree form, effected by an editing program using the syntactic role indicators, prediction indices, and shifting codes as described and displayed in Section 2, is only one of many possible ones, of which several might well be both more appealing and more useful.

Sentence no.	Sentence Length	No. of Analyses	Sentence	1962-IFIP	1963-FJCC	Expected
				mins	mins	mins
1	17	1	The increase in flow stress was attributed to vacancies, which have appreciable mobility at -72.	0.4	0.0	0.0
2	14	4	Economic studies show that it could be a billion-dollar-a-year business by the 1970's.	0.9	0.0	0.0
3	25	5	Slime formation is dependent on size of particles formed by mechanical means, amount of metal in the amalgam, and purity of solutions.	11.2	1.3	0.3
4	18	1	Single strain reversals at -72 not only produced the N effect but also increased the flow stress.	0.7	0.1	0.0
5	35	12	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.	120.0?	10.3	2.0
6	16	40	People who apply for marriage licenses wearing shorts or pedal pushers will be denied licenses.	2.0	0.1	0.0
7	17	4	Nearly all authorities agree that this will be the first practical, large-scale use of space.	0.9	0.0	0.0
8	16	3	A clutch of major companies has been pressing to get such a system into being.	13.5?	0.1	0.0
9	23	7	The U.S. has reached a momentous point of decision in a project that only a few years ago would have seemed improbable.	2.5?	0.2	0.0
10	23	25	Technologically speaking, there are three basic contending schemes, with a number of variations, for orbiting a communication satellite.	22.5	1.5	0.3

Figure 19. Processing Time.

Greibach[8] (Section 3.1) has made it clear that a strictly bi-unique correspondence between an arbitrary psg and an inverse dpa is too much to hope for: every dpa is the inverse of infinitely many psg's and, furthermore, given a psg-dpa pair, it is undecidable in general whether or not the latter is the inverse of the former. She has shown, however (Figure 6 of Greibach[8]; Section 6.2 of Greibach[6]), that the passage from psg to dpa can be restricted so as to proceed in a unique "natural" way.

It follows that, given a psg, a dpa can be constructed which will, together with a mapping of remarkable conceptual simplicity that can be effected with less cumbersome apparatus than that of Section 2, display the structure of a sentence in conventional phrase structure form. There is also some empirical evidence to the effect that, given a dpa with shifting codes, etc., a psg can be constructed from it which, when converted to standard form in the "natural" way, yields something close to the original dpa. Hence there is some hope that, although the mapping from dpa to psg is not abstractly single valued, the loop of Fig. 6 of Greibach[8] might at least be closed in a unique self-consistent way.

Consider the directed productions

$$(\text{S, art}) \rightarrow \text{art SP}' \text{ VP PD } (SV) \tag{1}$$
$$(\text{SP}', \text{nnn}) \rightarrow \text{nnn } (SV) \tag{2}$$
$$(\text{VP, vi}) \rightarrow \text{vi } (PV) \tag{3}$$
$$(\text{PD, prd}) \rightarrow \text{prd } (ES) \tag{4}$$

where "art" stands for an article, "SP′" for a subject phrase modified by an adjective, and "VP" and "PD" for a predicate and a period, respectively. "*SV*" is a role indicator for the subject of a verb, "PV" for a predicate verb, and "*ES*" for an end-of-sentence mark. These productions are sufficient for the obvious analysis of the sentence "The summer came.".

The psg productions

$$\text{S} \rightarrow \text{SP VP PD} \tag{5}$$
$$\text{SP} \rightarrow \text{T SP}' \tag{6}$$
$$\text{T} \rightarrow \text{art} \tag{7}$$
$$\text{SP}' \rightarrow \text{nnn} \tag{8}$$
$$\text{VP} \rightarrow \text{vi} \tag{9}$$
$$\text{PD} \rightarrow \text{prd} \tag{10}$$

are adequate to generate the same sentence with tree structure as in Fig. 20.

Productions (5)-(7) of the psg correspond to production (1) of the dpa, in the manner de-

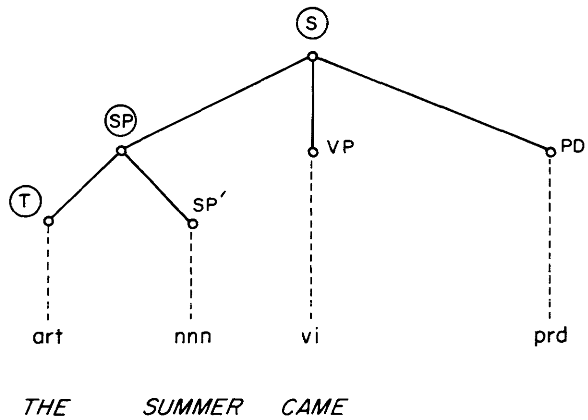


Figure 20. Tree Structure for "The summer came."

scribed by Greibach[8] (Figs. 1, 2 and 3). It is the absence from (1) of the circled symbols of Fig. 20, which we shall call *virtual* predictions, which in a sense differentiate the dpa from the psg. When "the" as "art" is accepted by the rule (1), the predictions S, SP and T are virtually fulfilled in whole ("T") or in part ("S", "SP"). The fact that T is a constituent of SP is essentially what is denoted by the role indicator (*SV*) in (1). Since SP′ is also a constituent of SP it is at a lower level than "VP" and "PD", although it appears undistinguished from the latter in (1). It is this disparity which is corrected by the shifting codes associated with dpa productions of the actual English grammar. With the fresh insight yielded by Greibach's theoretical results it appears possible, if desirable, to dispense with the *ad hoc* tree mapping apparatus built into the editing program in favor of more natural and elegant techniques.

These techniques are based on a new extension of parenthesis-free or Polish prefix notation in which predictions are treated as functors which, unlike conventional functors, do not have a fixed degree, but instead are explicitly labelled with a degree determined by the actual or virtual production by which they are expanded.

3.2 Consider the following augmented directed production as a replacement for production (1).

$$(\mathrm{S}, \mathrm{art}) \rightarrow \mathrm{S}_3\ \mathrm{SP}_2\ \mathrm{T}_1\ \mathrm{art}\ \mathrm{SP}'\ \mathrm{VP}\ \mathrm{PD}. \qquad (11)$$

In (11), subscripted expressions are interpreted as functors of degree specified by their subscripts. All other expressions are interpreted as variables (functors of degree 0). It is an immediate consequence of this interpretation that the right-hand side of any production written in this form is itself a well-formed string in parenthesis-free notation. Hence a grammar of this type would lend itself to mechanical checks for the well-formation of its rules, a property of considerable practical importance, say, in verifying the key-punching of a large grammar table.

If the subscripted expressions in (11) are ignored, (11) corresponds directly to (1). The subscripted expressions may, however, also be identified with the virtual predictions of Fig. 20. "T", as a functor of degree 1, has argument "art"; "SP", of degree 2, has as arguments the well-formed formulas "T_1 art" and "SP′", and the three arguments of S_3 are the well-formed formulas "SP_2 T_1 art SP′", "VP", and "PD". Any functor whose scope includes only subscripted expressions and terminal symbols corresponds to a wholly fulfilled virtual prediction (e.g., "T_1"), otherwise to a partially fulfilled one (e.g., "SP_2").

The production (11) ascribes degree 3 to "S". Other productions need not ascribe the same degree. Thus, in

$$(\mathrm{S}, \mathrm{ii}) \rightarrow \mathrm{S}_2\ \mathrm{VP}_1\ \mathrm{ii}\ \mathrm{PD} \qquad (12)$$

"S" is ascribed degree 2. The terminal symbol "ii" stands for the infinite form of an intransitive verb, and (12) accounts for structures such as "Go.".

3.3 It is obvious how to get augmented directed productions of the form (11) or (12) from the phrase structure tree of any sentence, since that tree is always finite. However, the productions of an arbitrary psg may provide for infinite left-branching structures (e.g., X → XY), hence more subtle difficulties arise when mapping the psg into a psg in standard form because the application of *every* production of such a psg must yield a terminal symbol.

Greibach's normal form theorem not only shows that such provisions can be made effectively for an arbitrary psg but it also implicitly converts left-branching structures into right-branching ones by eliminating such productions as X → XY while creating or retaining others of forms such as X → aXZ (Fig. 5 of Greibach[8]). As a consequence, and so far as phrase structure grammars are concerned, the direction of branching is shown to be not so much an intrinsic property of a language as a property of a grammar describing the language although the freedom of self-embedding is preserved. In fact, even a language so inherently "left-to-right" in appearance as parenthesis-free notation itself can be generated, hence analyzed, entirely in a right-to-left mode! In view, however, of the fact that English is written and read from left-to-right, of the desirability of generating (analyzing) a terminal symbol each time a production is applied, and of Yngve's[20] arguments about the desirability of limited left-branching, the psg in standard form and the corresponding dpa suggest themselves as potential mechanisms for speakers and hearers respectively, and hence as worthy objects of further study by psychologists and linguists. The authors are deeply impressed with the simplicity and elegance of the corresponding machine realization of such grammars but this, of course, is in itself no argument at all in favor of their adoption as explanatory models for human synthesis and analysis of sentences without some careful experimentation. It should go without saying that if transformations in the sense of Chomsky[1] are to be applied to any given sentence, the phrase markers for the sentence must be at hand. The realization of a dpa is therefore an essential prerequisite for the effective application of transformational grammars to sentence analysis.

As pointed out in Section 3.1, going from a dpa to a psg (other than the obvious but non-intuitive standard form inverse) is not a simple matter, and considerable theoretical and experimental work remains to be done. The available structure symbols and shift codes do appear to lead readily to the conversion of the current dpa grammar to one whose rules are augmented directed productions. Whether the resulting psg can, using Greibach's normal form theorem, be reconverted to the current dpa, thereby closing the loop of Fig. 6 (Greibach[8]), remains to be seen, but there is some ground for optimism at present.

3.4 Quite fortunately, the analysis program for a system based on augmented directed productions can be precisely that for the present one except that the former requires two prediction pools instead of one. The first pool is used for storing fulfilled (subscripted) predictions and terminal symbols[§], the second for storing unfulfilled (and therefore active) predictions. Each time the topmost prediction in the active pool is processed against a word class of the next word, the subscripted predictions and the terminal symbol of the subrule are stored in the fulfilled pool in the same order as they appear in the formula. Remaining active (non-subscripted) predictions are stored in the active pool. The performance of these two pools is illustrated below using "The man saw the boy." as an example.

For the sake of simplicity of explanation, only the path which leads to the acceptable analysis of this sentence is followed here. Augmented directed productions which are needed for this path are:

Note

vt: transitive verb

V: verb prediction

OP: object phrase prediction

OP′: modified object phrase prediction

prd: period

$$(\mathrm{S, art}) \rightarrow \mathrm{S_3\ SP_2\ T_1\ art\ SP'\ VP\ PD} \quad (13)$$

$$(\mathrm{SP', nnn}) \rightarrow \mathrm{SP'_1\ nnn} \quad (14)$$

$$(\mathrm{VP, vt}) \rightarrow \mathrm{VP_2\ V_1\ vt\ OP} \quad (15)$$

$$(\mathrm{OP, art}) \rightarrow \mathrm{OP_2\ T_1\ art\ OP'} \quad (16)$$

$$(\mathrm{OP', nnn}) \rightarrow \mathrm{OP'_1\ nnn} \quad (17)$$

$$(\mathrm{PD, prd}) \rightarrow \mathrm{PD_1\ prd} \quad (18)$$

Figure 21 shows the status of the two pools after the processing of each word of the sentence. Initially, the fulfilled pool is empty, and the active pool contains the initial symbol S. The second line shows that after the processing of "the" as art, "S_3 SP_2 T_1 art" of (13) have been stored in the fulfilled pool, and that "SP′

§ Formally, the first pool is simply an output tape with writing-head only and not a pushdown store, since nothing is read from it in the course of further analysis.

Fulfilled Pool bottom ⟷ top	Active Pool top ⟷ bottom
	S
S_3 SP_2 T_1 art	SP' VP PD
S_3 SP_2 T_1 art SP'_1 nnn	VP PD
S_3 SP_2 T_1 art SP'_1 nnn VP_2 V_1 vt	OP PD
S_3 SP_2 T_1 art SP'_1 nnn VP_2 V_1 vt OP_2 T_1 art	OP' PD
S_3 SP_2 T_1 art SP'_1 nnn VP_2 V_1 vt OP_2 T_1 art OP'_1 nnn	PD
S_3 SP_2 T_1 art SP'_1 nnn VP_2 V_1 vt OP_2 T_1 art OP'_1 nnn PD_1 prd	

Figure 21. Analysis of "The man saw the girl."

VP PD" have been stored in the active pool, replacing the previous topmost prediction S. The series of symbols contained in the fulfilled pool in the last line of Fig. 21 is the output of the analysis of this sentence.

In this proposed system, discontinuous structures such as "*It* is true *that he is right.*" can probably be treated in the same way as in the current analyzer; a production for (S, it) will be

$$(S, it) \rightarrow S_4\, SP_1\, it\ VP\ NC\ PD \qquad (19)$$

where "it" stands for a temporary subject, and NC for a noun clause which is the true subject of the sentence. If it is desired that the connection between SP and NC be explicitly identified, it is possible to assign special marks to these predictions showing that they constitute a single (discontinuous) structure.

Adverbs, prepositional phrases, etc. will be accepted as floating structures and their symbols ignored when checking for well-formation of a production as a whole. For example,

$$(VP, adv) \rightarrow ADV_1^*\ adv\ VP \qquad (20)$$

indicates that, although an adv is joined to the structure ADV of degree 1, ADV itself is outside the structure of VP, with its dependency undetermined.

3.5 Since the analyzer output is in parenthesis-free form, it can be interpreted as a phrase-structure tree without further formal ado, although, if a more graphic tree form is desired, additional editing is obviously possible.

The output of such an analyzer has an abvious kinship to that of Yngve's[2] model of random sentence production. There are, however, significant differences beyond the obvious one that it is easier to describe the structure of a sentence being synthesized than that of one being analyzed.

In Yngve's model, no degrees are assigned to non-terminal symbols such as "S", "NP", "VP", etc. The difference is non-trivial. There is the bonus of a mechanical check for well-formation of rules, not a negligible factor in major clerical enterprises. More important, however, there is the fact that Yngve's output formulas become ambiguous whenever rules are included in the grammar of which the left-hand symbol is the same but the number of right-hand symbols is different. In order to make Yngve's output unambiguous, SP of SP → T SP′ has to be distinguished from SP of SP → nnn. This would entail duplication of rules such as S → SP VP PD: one rule for SP with degree 1, the second rule for SP with degree 2. The proposed technique for specifying degrees for fulfilled predictions resolves this dilemma.

REFERENCES

1. CHOMSKY, N., "Three Models for the Description of Language," *IRE Trans. on Information Theory,* Vol. IT-2, 1956.
2. CHOMSKY, N., "Formal Properties of Grammars," in R. R. Bush, E. H. Galanter, and R. D. Luce (Eds.), *Handbook of Mathematical Psychology,* Vol. 2, New York, Wiley, 1962.
3. CHOMSKY, N., and SCHÜTZENBERGER, M. P., "The Algebraic Theory of Context Free Languages," in P. Braffort and D. Hirschberg (Eds.), *Computer Programming and Formal System,* North Holland Publishing Co., Amsterdam, 1963.
4. EVEY, J., "The Theory and Application of Pushdown Store Machines" (Doctoral Thesis), *Math. Ling. and Automatic Translation,* Report No. NSF-10, Harvard Computation Laboratory, 1963.
5. EVEY, J., "Application of Pushdown Store Machines," *AFIPS Proceedings of 1963 Fall Joint Computer Conference,* Las Vegas, Nevada, November 12-14, 1963.
6. GREIBACH, S., "Inverses of Phrase Structure Generators" (Doctoral Thesis), *Math. Ling. and Automatic Translation,* Report

No. NSF-11, Harvard Computation Laboratory, 1963.

7. GREIBACH, S., "The Undecidability of the Ambiguity Problem for Minimal Linear Grammars," *Information and Control,* Vol. 6, No. 2, 1963.
8. GREIBACH, S., "Formal Parsing Systems" (Draft report), submitted for publication, 1963.
9. HAYS, D. G., "On the Value of Dependency Connection," *Proc. of the 1961 International Conference on Machine Translation of Languages and Applied Language Analysis,* her majesty's Stationery Office, London, 1962.
10. KUNO, S., "The Current Grammar for the Multiple-path English Analyzer," *Math. Ling. and Automatic Translation,* Report No. NSF-8, Harvard Computation Laboratory, 1963.
11. KUNO, S., "The Multiple-path Syntactic Analyzer for English," and "Study of Analysis Output of the Multiple-path English Analyzer," *Math. Ling. and Automatic Translation,* Report No. NSF-9, Harvard Computation Laboratory, 1963.
12. KUNO, S., and OETTINGER, A. G., "Multiple-path Syntactic Analyzer," *Information Processing 62,* North-Holland Publishing Co., Amsterdam, 1963.
13. KUNO, S., and OETTINGER, A. G., "Multiple-path Syntactic Analyzer" (updated version), *Math. Ling. and Automatic Translation,* Report No. NSF-8, Harvard Computation Laboratory, 1963.
14. MATTHEWS, G. H., "Analysis by Synthesis of Natural Languages," *Proc. of 1961 International Conference on Machine Translation of Languages and Applied Language Analysis,* her majesty's Stationery Office, London, 1962.
15. OETTINGER, A. G., "Automatic Syntactic Analysis and the Pushdown Store," *Proc. of Symposia in Applied Math.,* Vol. XII, American Mathematical Society, Providence, Rhode Island, 1961.
16. PLATH, W., "Multiple-path Syntactic Analyzer of Russian," *Math. Ling. and Automatic Translation,* Report No. NSF-12, Harvard Computation Laboratory, 1963.
17. RHODES, I., "A New Approach to the Mechanical Translation of Russian," *National Bureau of Standards Report,* 1959, No. 6295, also in *Mechanical Translation,* Vol. 6, 1961.
18. ROBINSON, J. J., "Preliminary Codes and Rules for the Automatic Parsing of English," *Memorandum RM-339-PR,* The RAND Corporation.
19. SAKAI, I., "Syntax in Universal Translation," *Proc. of the 1961 International Conference on Machine Translation of Languages and Applied Language Analysis,* her majesty's Stationary Office, London, 1962.
20. YNGVE, V., "A Model and an Hypothesis for Language Structure," *Proc. of the American Philosophical Society,* Philadelphia, 1960.