[*International Conference on Machine Translation of Languages and Applied Language Analysis*, National Physical Laboratory, Teddington, UK, 5-8 September 1961]

AN APPROACH TO THE SEGMENTATION PROBLEM IN SPEECH ANALYSIS AND LANGUAGE TRANSLATION

by

GERARD SALTON and R.W. THORPE

INTRODUCTION

THE generation of proper word boundaries is an important part of several problems in information processing. Specifically, the speech recognition problem is often described as the production of a phonemic transcript, followed by the assembly of phonemes into complete words.[1,2,3,4] The automatic translation of certain natural or artificial languages, such as, for example, Chinese and Japanese to English,[5,6,7] or English to Braille[8] also requires the generation of words in the output language which may correspond either to several items of input, or to only part of an input item.

The segmentation problem is often complicated by the fact that each item of input may be associated with several possible output correspondents, only one of which is acceptable in any given context. Frequently, the reduction of each set of multiple correspondents is at least partly dependent upon the proper recognition of word boundaries. The English phoneme sequence/aban/ might, for example, correspond to the indefinite article "a" followed by the noun "ban", or it might form a verb or noun prefix as in "abandon", or "abandonment". Similarly, the Chinese character □ (dzi), which may be translated as "self" when standing alone, may in combination with other characters be translated variously as "freedom", "self-defence", "ego", "originality", "naturally", "freely", "liberalism", and so on.

The generation of syntactically well-formed sentences in the output language is a common requirement for the set of problems under consideration. Since the material being processed does not, however, consist of complete syntactic units, it is first necessary to generate the appropriate structural information before any method based on syntax can be used.

Two principal techniques are therefore proposed for the recognition of word boundaries and the elimination of multiple correspondents. The first consists of a *particle analysis* designed to attach to each output correspondent, whether consisting of a complete word or not, a set of grammatical indicators giving information about the possible role of the given correspondent in a sentence. The second is a type of syntactic *analysis* which compares certain predicted structural features in each sentence with the syntactic Information actually attached to each output correspondent. The acceptable correspondents are those for which the grammatical indicators match the predictions. Methods are given to select a single acceptable correspondent from each set of multiple correspondents and to assemble particles which do not constitute complete words into complete output items.

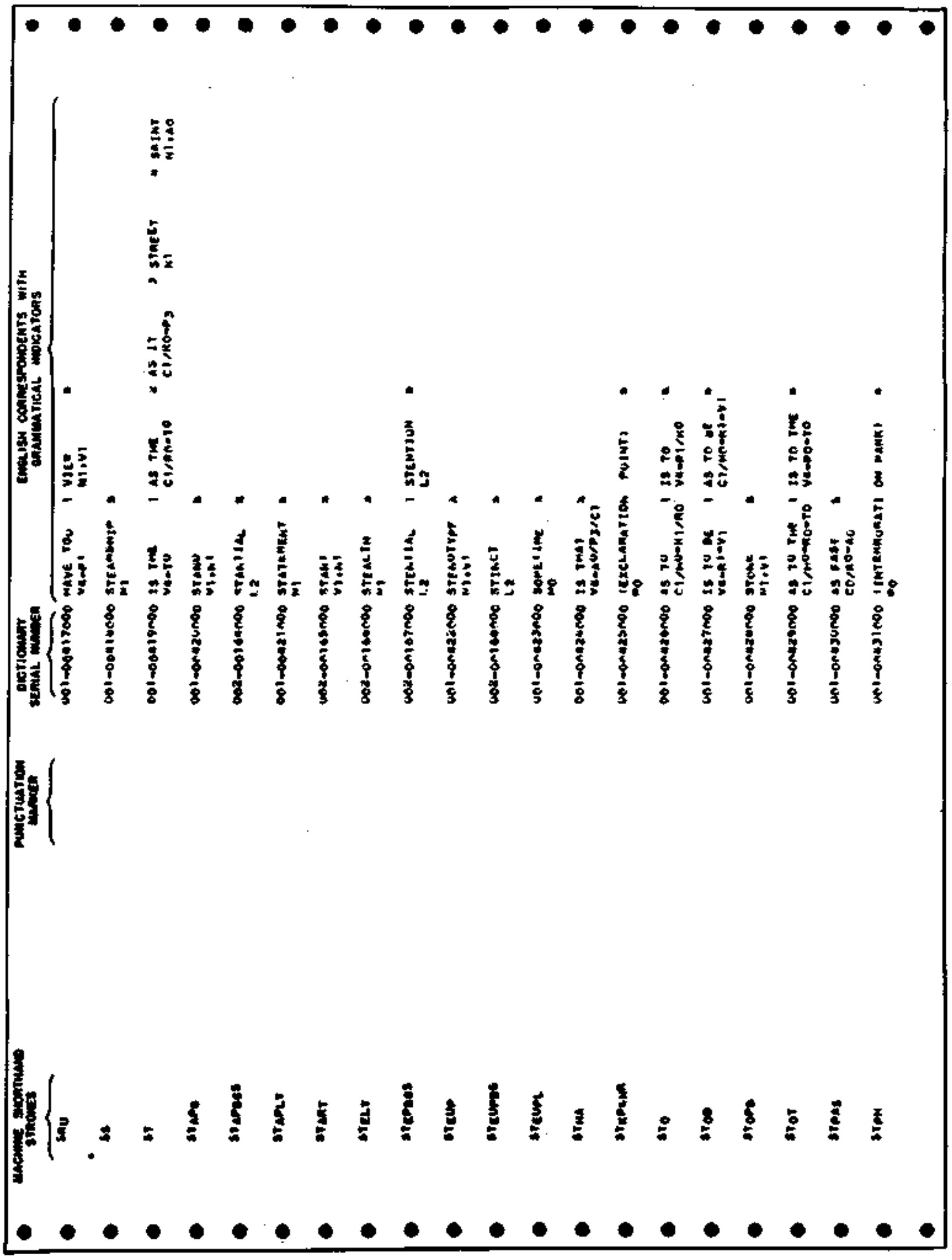
THE PHONEMIC ENGLISH INPUT

To demonstrate the techniques involved, machine shorthand as produced by a stenotype device is used as input. Since the vowel letters and many consonant letters are written according to sound rather than according to the normal English spelling, the machine shorthand units, or *strokes,* represent linguistic units of the spoken language. Each stroke may correspond in written English to one syllable, one complete word, or one phrase consisting of several words. Moreover, because of the limited number of keys available on the stenotype machine, a special transcription system is used which substitutes a set of letters available on the keyboard for each English phoneme to be represented. A number of special shorthand abbreviations, similar to the contractions used in Braille, are also provided to represent certain high frequency words, affixes or phrases.[9]

In order to decode the abbreviations and phrases, a small dictionary is used, including approximately one thousand high frequency shorthand strokes. Each stroke is listed together with a set of English correspondents. Each correspondent in the dictionary is furnished with a set of grammatical indicators. As an example, the shorthand stroke A has correspondents "a" and "an". The grammatical indicators for "a" show that, in English, this correspondent can be an article, a verb prefix, an adjective prefix, a noun prefix, or an adverb prefix. An excerpt from the dictionary of abbreviations and phrases used in the computer experiments is shown in *figure 1.**

To keep the dictionary down to reasonable size, only a relatively small proportion of the many thousands of possible English words and syllables are included in the dictionary. Many output correspondents will therefore not be found in the dictionary during the look-up operation and no grammatical information is available for these correspondents

* A Univac I computer was used to perform the experimental work.



MACHINE SHORTHAND STROKES	PUNCTUATION MARKER	DICTIONARY SERIAL NUMBER	ENGLISH CORRESPONDENTS WITH GRAMMATICAL INDICATORS				
SRU		001-00817000	HAVE YOU V4-P1	1 VIEW N1,V1	*		
SS		001-00818000	STEAMSHIP N1	*			
ST		001-00819000	IS THE V4-TO	1 AS THE C1/R0-T0	2 AS IT C1/RO-P3	3 STREET N1	4 SAINT N1,AO
STAPG		001-00820000	STAND V1,N1	*			
STAPBGS		002-00164000	STANTIAL L2	*			
STAPLT		001-00821000	STATEMENT N1	*			
START		002-00165000	START V1,N1	*			
STELT		002-00166000	STEALTH N1	*			
STEPBGS		002-00167000	STENTIAL L2	1 STENTION L2	*		
STEUP		001-00822000	STEAOTYPE N1,V1	*			
STEUPBG		002-00168000	STINCT L2	*			
STEUPL		001-00823000	SOMETIME MO	*			
STHA		001-00824000	IS THAT V4-AO/P3/C1	*			
STKPWR		001-00825000	(EXCLAMATION POINT) PO	*			
STO		001-00826000	AS TO C1/MO-N1/RO	1 IS TO V4-P1/MO	*		
STOB		001-00827000	IS TO BE V4-R1-V1	1 AS TO BE C1/MO-R1-V1	*		
STOPG		001-00828000	STONE N1,V1	*			
STOT		001-00829000	AS TO THE C1/MO-RO-TO	1 IS TO THE V4-RO-TO	*		
STPAS		001-00830000	AS FAST CO/RO-AO	*			
STPH		001-00831000	(INTERROGATION MARK) PO	*			

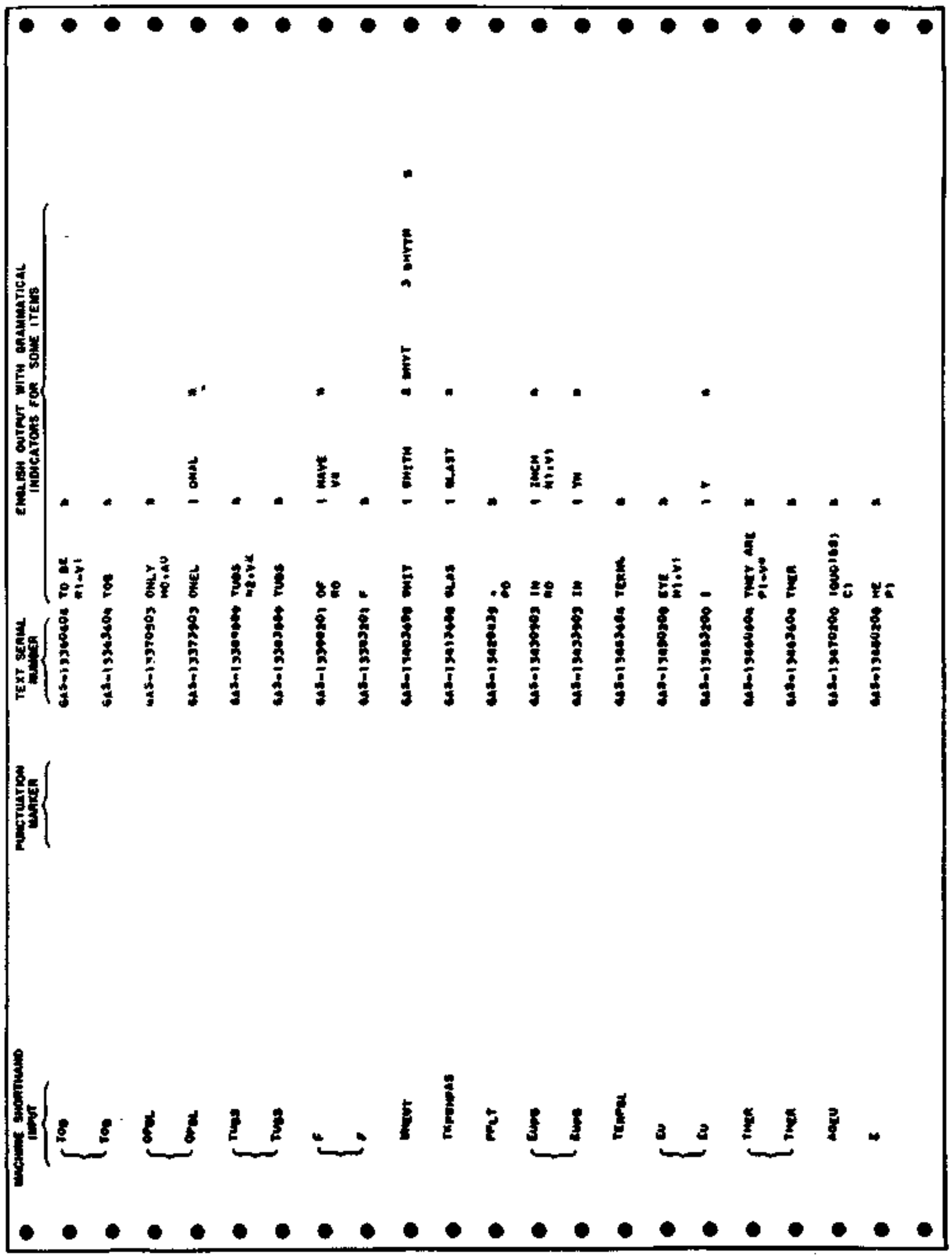
Dictionary of Abbreviations and Phrases
Including Grammatical Indicators
Fig. 1

An output sample showing the transliterated text after substitution of English letter clusters for the corresponding shorthand clusters, and after the dictionary look-up operation is shown in *figure 2.* It will be noted that the English output includes incorrect word forms and multiple correspondents. In some cases both correct and incorrect word forms are generated. In other cases, several morphologically correct forms are produced, and it is necessary to use criteria extracted from the context to choose the appropriate correspondent. A representative sample of the types of multiple correspondents which may arise is shown in *Table 1.*

TABLE 1

MULTIPLE CORRESPONDENCES GENERATED BY DICTIONARY LOOK-UP

Type of Ambiguity	Shorthand	Possible Translations
Morphological	TPHAGS OFRGS PWHREF EUPB TKUS TREUL	naings, nation offertion, offerings blef in dus (industrial) tril
Syntactic	OPGS TKEUG WE URBGT	openings, option dig, dying when, we you shall go it, your account
Semantic	SHRED SPORT TKEUG	shred, sled sport, support, export dying, dyeing



Assembled Shorthand Output Including
Grammatical Indicators
Fig. 2

THE PARTICLE ANALYSIS

The particle analysis is designed to assign syntactic indicators to correspondents not already furnished with grammatical information, and to reduce the ambiguity in the grammatical indicators found during dictionary look-up. This is done by considering in turn every grammatical indicator of each English correspondent attached to a given shorthand stroke. A list of the principal grammatical indicators is given in *Table 2*. The first character of the indicator is used to represent the word type, and the second character distinguishes various cases within each word type. As an example, AS is the indicator used for adjective suffixes, and NP represents noun prefixes.

Shorthand strokes are first partitioned to determine whether one of a set of recognizable suffixes is present. If so, the grammatical indicators derived from the suffix, as shown in *Table 3,* are tested for compatibility with the indicators attached to the stems; correspondents with incompatible stem and suffix indicators are eliminated. Correspondents consisting of several English words are similarly tested, and incompatible phrases such as articles not followed by either adjectives, adverbs, or nouns are eliminated.

The particle analysis then concentrates on those correspondents not provided with grammatical indicators, or provided with particle indications. In particular, correspondents without grammatical indication or with a prefix indication are given a special identification whenever one of the correspondents of the succeeding item is provided with in infix or suffix indicator. Such a prefix can later be attached to the corresponding suffix supplied by the next shorthand item. Similarly, correspondents without grammatical indication, or with an infix or suffix indicator, are modified if the preceding item includes a correspondent with a prefix indication.

A simplified flowchart for the particle analysis is shown in *figure 3,* The sample analysis of *figure* 4 exhibits the grammatical indicators before and after particle analysis (columns 3 and 4). The correspondent of stroke 5, for example, which could not be found in the dictionary is given a special infix marker (L2) because it is followed by an adverb suffix (HS indication), and preceded by a correspondent with prefix indicators (VP,AP). Similarly the correspondent of word 16 is recognized as an adjective prefix, and word 24 as a participle infix, even though no grammatical information was originally available for these two correspondents. In each case the grammatical indicator used as a criterion in the modification process is circled in column 3 of *figure 4* and grammatical indicators of adjacent shorthand items are suitably modified as shown in column 4 of *figure 4*. The L2 marker is used for all infix or suffix particles, and the corresponding shorthand strokes are treated as parts of a single item during syntactic analysis.

It is seen that, as a result of the particle analysis, all but three

TABLE 2

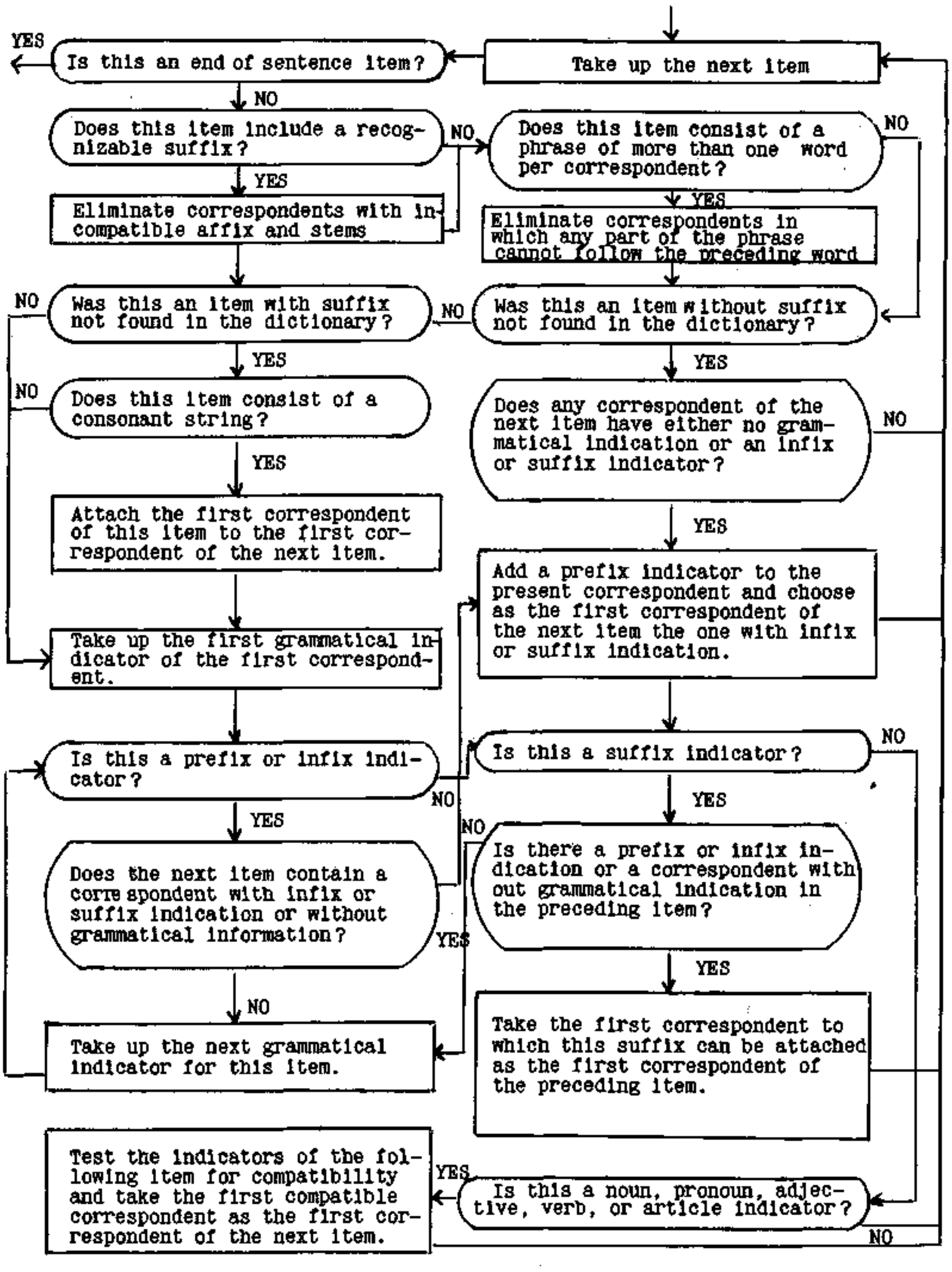
GRAMMATICAL INDICATORS

Grammatical Indicators		Definition
Character	Position	
1	2	
A		Adjectives
B		Participles
C		Conjunctions and Style Indicators
L		Particle Indicators
N		Nouns
P		Pronouns
R		Prepositions
T		Articles
V		Verbs
	1 2 3 4	Indicators marking distinctions within each group, such as N1 for singular nouns, N2 for plural nouns, and N3 for possessive nouns.
	P	Prefix Indicators
	M	Infix indicators
	S	Suffix Indicators
	0	Complete Word Indicators

of the correspondents shown in *figure 4* are furnished with grammatical indicators. Moreover, since particle indicators have been eliminated, the average number of indicators attached to a given correspondent is now much smaller. The probability of ambiguity arising in the predictive analysis is therefore considerably reduced.

TABLE 3
GRAMMATICAL INDICATORS DERIVED FROM SUFFIXES

Shorthand Suffix	English Correspondents	Derived Grammatical indicator
PLTS	ments	N2
	mities	N2
	mates	N2, V2
PLT	ment	N1, V1
	mity	N1
	mate	V1, N1, A0
BGS	ction	N1
	ctious	A0
GS	tion	N1, V1
	ings	N2, V2
	tious	A0
	tial	A0, N1
TS	ities	N2
	ates	N2, V2
D (alone)	ed	B0, V1
G	ing	N1, A0, B0, V1
S (alone)	s	N2, V2
T	ity	N1
	ate	V1, N1, A0



SIMPLIFIED FLOWCHART FOR PARTICLE ANALYSIS

Fig. 3

	Short-hand Original	Selected English Correspon-dents	Grammatical Indicators			Function Selected by Predictive Analysis
			After Dictionary Look-up	After Particle Analysis	After Predictive Analysis	
1	TPH	in	RO	RO	RO	Infinity
2	1938	1938	-	-	N1	Prep. Comp.
3	A	a	TO	TO	TO	Subject
4	EPB	en	VP, AP, BS, VS	HOZ	HO	Adverb Es.
5	TEUR	tir	-	L2	HO	Adverb Es.
6	HREU	ly	HS, AS, NP, AP	L2	HO	Adverb Es.
7	TPHU	new	AP, HO	AO, HO	AO	Adj. Master
8	APB	and	CO, C1	CO, C1	CO	Infinity
9	TKEUF	different	AO	AO	AO	Subject
10	TEUP	tip	-	-	N1	Adj. Master
11	F	of	RO	RO	RO	Prep. Es.
12	HREUT	lit	AP, NP, VP, HP	AOZ	AO	Prep. Comp.
13	SORS	sors	N2, V2	N2, V2	N2	Adj. Master
14	RBGS	,	CO	CO	CO	Infinity
15	T	the	TO	TO	TO	Prep. Comp.
16	TPHRU	flu	-	AOZ	AO	ADj. Master
17	RES	res	NP, AP, VP, AM	L2	AO	Adj. Master
18	EPBT	ent	AS, NS	L2	AO	Adj. Master
19	HRAPL	lam	-	-	N1	Adj. Master
20	P	p	-	L2	N1	Adj. Master
21	RBGS	,	CO	CO	CO	Infinity
22	WAS	was	V4	V4	V4	Pred. Head
23	EUPB	in	RO, HO	BOZ	BO	Verb Comp.
24	TRO	tro	-	L2	BO	Verb Comp.
25	TKUS	duce	VS, NM	L2	BO	Verb Comp.
26	D	had	V4	ZZZ		
		d	BS, NS, VS, AS	L2	BO	Verb. Comp.
27	TOT	to the	RO-TO	RO-TO	RO-TO	Prep. Comp.
28	A	a	TO	ZZZ		
		a	VP, AP, NP, HP	AOZ	AO	Adj. Master
29	PHER	mer	AM, NM	L2	AO	Adj. Master
30	KAPB	can	NP, AP, VP, AS, NS	L2	AO	Adj. Master
31	PUB	public	AO, N1	AN, N1	AO	Adj. Master
32	FPLT		PO	PO	PO	End Sent.

The Recognition of Word Boundaries
Sample Analysis
Fig. 4

THE PREDICTIVE SYNTACTIC ANALYSIS

The method of predictive syntactic analysis originating at the National Bureau of Standards,[10] and modified at Harvard University,[11,12] is based on the notion that, in terms of the human communications system, sentence analysis cannot be a complicated multi-layered combinatorial problem in which the function of a given word is made to depend upon the characteristics of all other words in the sentence. Rather, it is noted that speech emission and reception appear to be quasilinear, one-dimensional processes when viewed as a function of time, and that as a result it must be possible to analyze a sentence in a reasonably linear fashion.

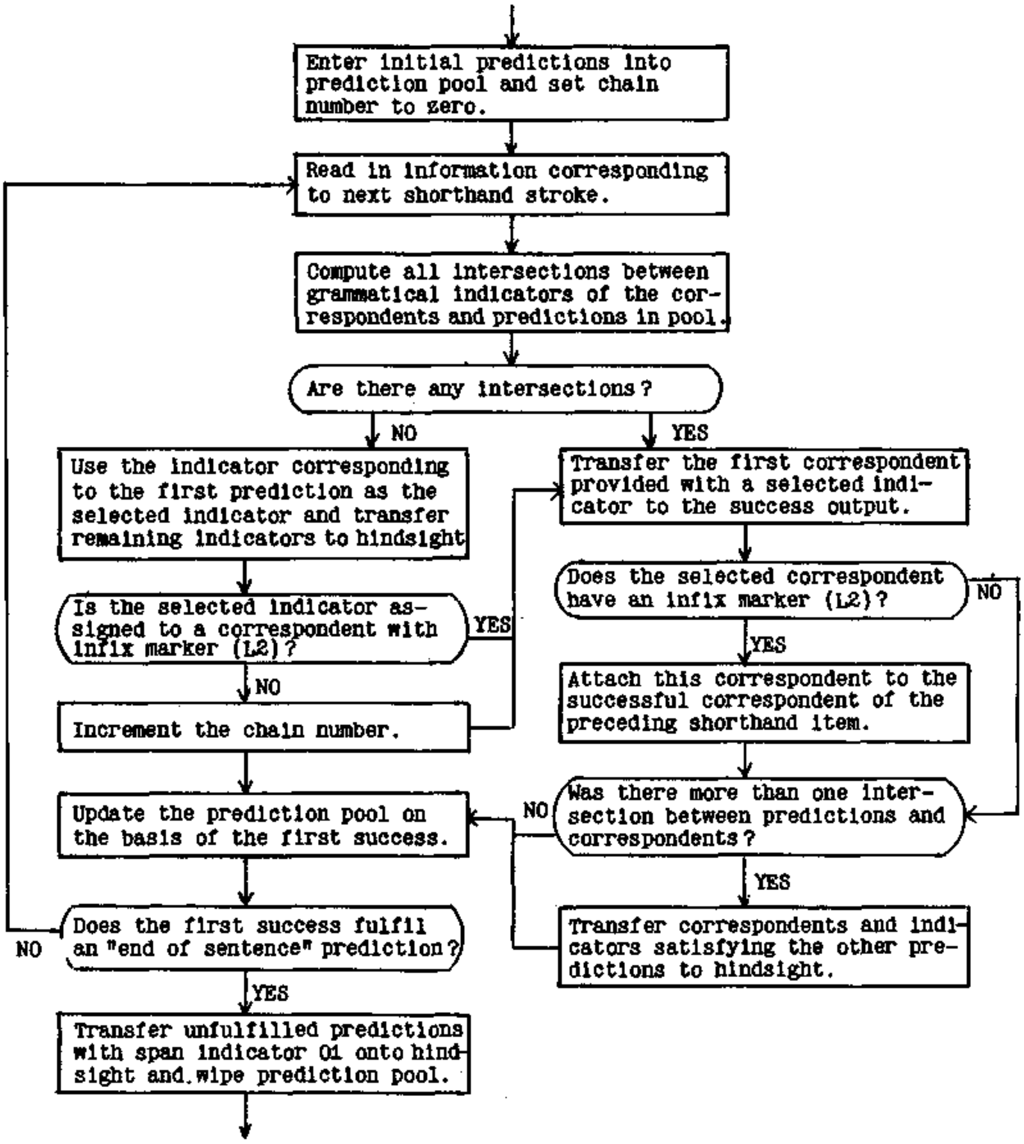
This premise is supported by certain statistical studies which indicate a large degree of regularity in sentence formation,[13] by psychological experiments undertaken to test the memory span of human beings,[14] and by research toward a new model for language structure and sentence production[15].

The method of predictive analysis consists in scanning through a sentence from left to right, one word at a time, while making *predictions* at every point about syntactic structures to be found further to the right. Each word in the sentence is tested to find what previously made prediction it fulfills; this information is used, in turn, to set up further predictions for later structures in the same sentence.

The programme actually used attempts to match the syntactic indicators of each correspondent with the predictions stored in a *prediction pool.* At the start of the sentence, a set of initial predictions including "subject", "predicate", and "end of sentence" is entered into the prediction pool. The first match between one of the indicators and a prediction, in the pool is called a *success*. The corresponding word and its indicator, known as the selected correspondent and selected indicator, respectively, are transferred onto a *success output tape* and the prediction pool is updated by eliminating the selected indicator and adding further predictions. Other correspondents which also match one of the predictions are removed to a storage pool known as *hindsight.*

If no match occurs for a given set of correspondents between grammatical indicators and predictions, a success is forced by accepting the indicator corresponding to the first prediction and the first correspondent as the selected items. When a success if forced, an error index known as the *chain number* is incremented to provide at least a partial indication of the accuracy of the analysis. Different types of predictions are identified by a *prediction span indicator* which is used in part to control the updating of the prediction pool.

A simplified flowchart for the predictive analysis is shown in *fig.5.* In actual practice, the indicators of each correspondent of a given shorthand stroke are tested separately against each of the predictions



Simplified Flowchart for Predictive Analysis

Fig. 5

in the pool. If no match occurs, a success is forced; the chain number is then increased if the selected word is not provided with an infix marker. Words which were previously recognized as particles can thus accept any indicator without penalty.

When the last word in a sentence has been treated, the prediction pool is wiped and essential predictions, identified by an 01 span indicator, are written on the hindsight tape. If no essential predictions are left in the pool at the end of a sentence, and if the chain number is small, indicating that the predictions were generally fulfilled, the analysis is assumed correct. Otherwise, it is possible to refine the analysis by reprocessing a sentence several times, using information originally stored in the hindsight. Span indicators and chain numbers are thus useful not only for error detection but also for error correction in some cases.

The various types of prediction span indicators are listed in *Table 4*. The principal grammatical functions performed by the words in a sentence are shown in *Table 5* together with the grammatical indicators which can *fulfil* each particular function. The functions *predicted* by the grammatical indicators are similarly shown in *Table 6.* For example, a correspondent provided with a noun (N) indicator is acceptable as a noun-adjective complement *(Table 5)* and will in turn predict the occurrence of another noun-adjective complement or of a preposition function *(Table 6).*

TABLE 4.

PREDICTION SPAN INDICATORS

Prediction Span Indicator	Code	Special Action during Analysis
Prediction must be fulfilled by next word or not at all	00	Erase prediction after one word
Prediction must be fulfilled	01	Put prediction in hindsight if not fulfilled
Prediction may be fulfilled several times	02	Do not erase: prediction if fulfilled
Prediction should be fulfilled some time before the end of sentence	03	Erase only when fulfilled or during end wipe

TABLE 5

SYNTACTIC FUNCTIONS FULFILLEDBY GRAMMATICAL INDICATORS

Grammatical Functions	Function Codes	Fulfilling Indicators
Subject of Vert	SUBJCT	A0, T0, Nl, N2, P1, P2, R1, B0
Master	ADJ MAS	Nl, N2, N3, A0, B0
Object of Verb	OBJT V	A0, T0, N1, N2, N3, P2, P3, B0
Predicate Head	PRED H	V1, V2, V3, V4
Noun-Adjective Complement	NADJ CM	N1, N2, N3, B0, A0, R1
Verb Complement	VERB CM	R1, B0, V1, V4
Adverb Master	ADV MAS	A0, B0, V1, V2, V3, V4, H0
Preposition Complement	PREP C	A0, T0, N1, N2, N3, P2, P3, B0
Adverb Function	ADV E	H0
Preposition Function	PREP E	R0
Subclause Subject	SCL SB	T0, A0, N1, N2, P1, P3, B0
Infinity	INFINT	R0, H0, C0, C1
Infinitive Base	INF BS	V1, V4
End of Sentence	END SEN	P0

TABLE 6

SYNTACTIC FUNCTIONS PREDICTED BY GRAMMATICAL INDICATORS

Grammatical Indicators	Grammatical Functions Predicted
A (adjectives)	Master Noun Adjective Complement Adverb Function Preposition Function
B (participles)	Noun Adjective Complement Verb Complement Adverb Function Preposition Function Subclause Subject
C (conjunctions)	Subject of Verb Predicate Head Subclause Subject Infinity
H (adverbs)	Adverb Master
N (nouns)	Noun Adjective Complement Preposition Function
P (pronouns)	Noun Adjective Complement Preposition Function Subclause Subject
R (prepositions)	Preposition Complement Adverb Function Infinitive Base
T (articles)	Master
V (verbs)	Object of Verb Verb Complement Adverb Function

A sample sentence is analyzed in *fig. 4*. Selected indicators and selected grammatical functions are shown in columns 5 and 6, respectively, of *fig. 4*. It will be noted that a single correspondent is chosen for each shorthand stroke; each correspondent is assigned a unique grammatical function, and word parts which must be assembled into complete words are assigned the same grammatical indicator. The correspondent "a", for example is properly recognized as an article, subject of the sentence, in stroke 5, whereas in stroke 28 it is recognized as an adjective which forms a part of the word "a-mer-can".

The success and hindsight outputs actually produced on the Univac computer for the sentence previously considered in *fig.4* are shown in *figs. 6 and 7* respectively. Word boundary indications are included in column 3 of *fig. 6,* showing for example that en-tir-ly, flu-res-ent, lam-p, in-tro-duce-d, etc., are to be attached to form complete words. The final chain number of 03 shows that a success was forced three times. The correct correspondent was chosen in each case, all word boundaries were properly recognized and, except for some minor imperfections in the assignment of grammatical functions, the sample sentence was analyzed correctly.

The updating of the prediction pool is illustrated in *fig. 8.* The pool is operated as a push-down store in such a way that the last prediction entered into the pool is the first one tested for fulfillment. This procedure simplifies the computer programme, economizes storage space, and follows the intuitive notion that the predictions made by the last word analyzed are the ones most likely to be fulfilled by the word which follows. At each step, the selected prediction and all predictions with a 00 span indicator are erased, and new predictions are added to the top of the pool.

A sentence similar to the one analyzed in *figs. 4, 6 and 7* is shown in *fig. 9* before syntactic analysis, including all the multiple correspondents, and excluding any information about the word boundaries. One correspondent must be chosen from each set of multiple correspondents shown in the figure. A comparison of *figs. 6 and 9* shows the improvement obtained.

CONCLUSION

The contextual and syntactic analyses described were used successfully on several samples of shorthand text. Word boundaries were properly recognized and correspondents correctly assigned for over ninety percent of the shorthand strokes. Almost identical programmes would seem to promise equal success with a variety of input languages. A dictionary of certain phoneme clusters including both acoustic and linguistic features must be used for speech input, and a dictionary of the principal contractions for Braille. In each case the recognition of the many input clusters which are not included in the dictionary is left to a particle analysis, and the formation of complete well-formed output sentences is handled by a syntactic process similar to the predictive analysis.

MACHINE SHORTHAND INPUT	ASSIGNED CORRESPONDENT		PARTICLE ATTACH MARKER	TEXT SERIAL NUMBER	ASSEMBLY INDICATOR	ASSIGNED FUNCTION	ASSIGNED INDICATOR	CHAIN NUMBER	PREDICTION SPAN INDICATOR	SERIAL OF PREDICTOR STROKE
*NOTYPE HEAD ER	DUMMY	DUMMY		1291	*					
TPH	IN			1292	*	INFINT	RO	00	02	000
1938	1938			1293	IVF	PREP C	N1	01	01	292
A	A			1295	*	SUBJCT	TO	01	01	000
EPB	EN		-	1296	*	ADVB E	HO	01	00	295
TPUR	TIR		-	1297	IVF	ADVB E	HO	01	00	295
HREU	LY		)	1298	*	ADVB E	HO	01	00	295
TPHU	NEW			1299	*	ADJ MAS	AO	01	01	295
APB	AND			1300	*	INFINT	CO	01	02	000
TKEUP	DIFFERENT			1301	*	SUBJCT	AO	01	00	300
TFUP	TIF			1302	IVF	ADJ MAS	N1	02	03	301
F	OF			1303	*	PREP F	RO	02	00	302
HREUT	LIT			1304	*	PREP C	AO	02	01	303
SORS	SORS			1305	IVEA	ADJ MAS	N2	02	03	304
HRGS	,			1306	IVF	INFINT	CO	02	02	000
T	THE			1307	*	PREP C	TO	02	00	306
TPHRU	FLU		-	1308	IVF	ADJ MAS	AO	02	01	307
RES	RES		-	1309	*	ADJ MAS	AO	02	01	307
EPBT	ENT		)	1310	*	ADJ MAS	AO	02	01	307
HRAPL	LAN		-	1311	IVF	ADJ MAS	N1	03	03	308
P	P		)	1312	IVF	ADJ MAS	N1	03	03	309
KBGS	.			1313	IVF	INFINT	CO	03	02	000
WAS	WAS			1314	*	PRED H	V4	03	01	000
EUPB	IN		-	1315	*	VERB CH	BO	03	03	314
TRO	TRO		-	1316	IVF	VERB CH	BO	03	03	314
TKUS	DUCE		-	1317	*	VERB CH	BO	03	03	314
D	D		)	1318	*	VERB CH	BO	03	03	314
TOT	TO THE			1319	*	PREP C	RO-TO	03	01	318
A	A		-	1320	*	ADJ MAS	AO	03	01	319
PHEP	MER		-	1321	*	ADJ MAS	AO	03	01	319
KAPB	CAN		)	1322	*	ADJ MAS	AO	03	01	319
PUB	PUBLIC			1323	*	ADJ MAS	AO	03	03	320
FPLT				1324	IVF	END SEN	PO	03	01	000

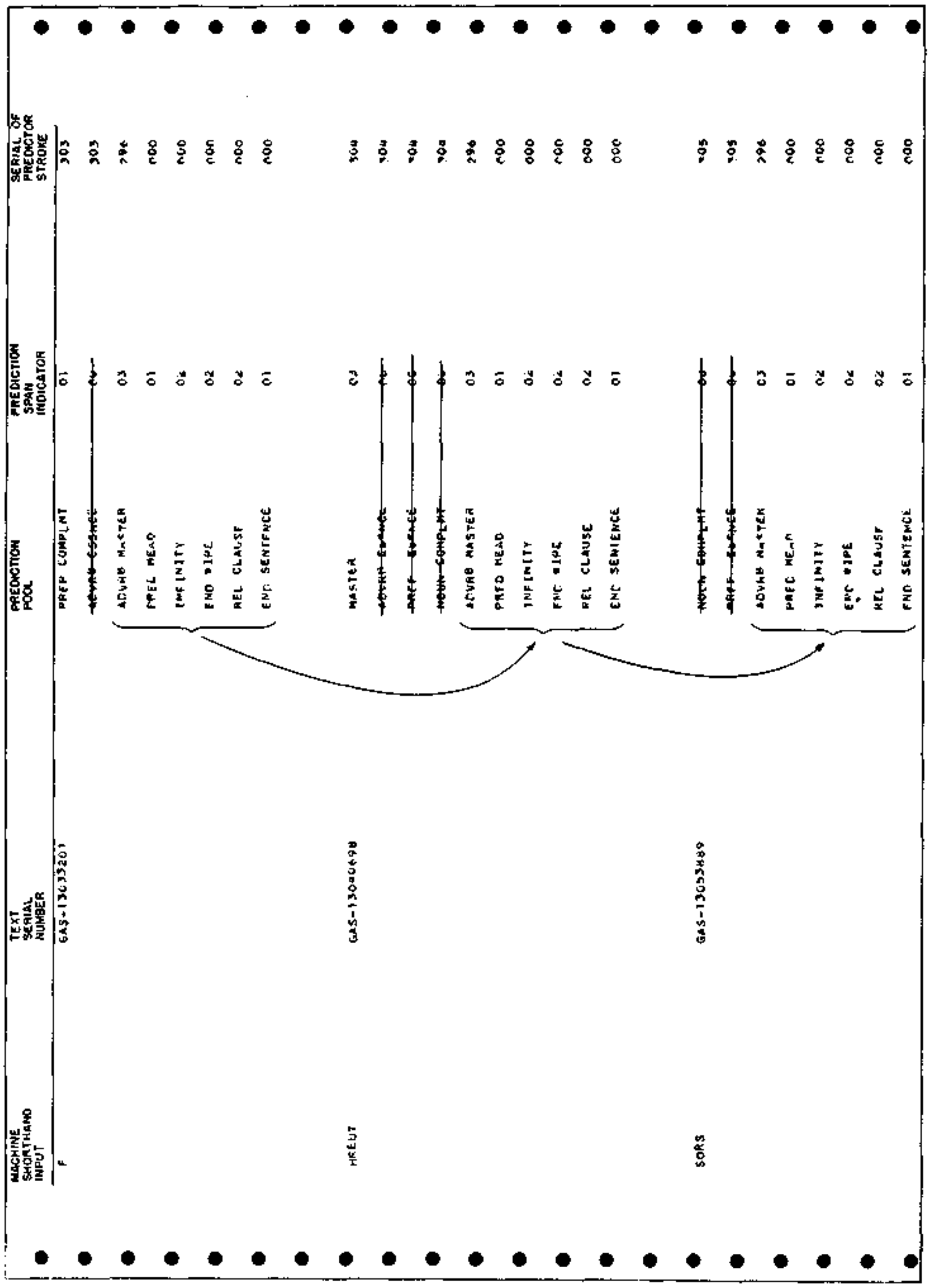
Success Output

Fig. 6

MACHINE SHORTHAND INPUT	ASSIGNED CORRESPONDENT	TEXT SERIAL NUMBER	ASSEMBLY INDICATOR	ASSIGNED FUNCTION	ASSIGNED INDICATOR	CHAIN NUMBER	PREDICTION SPAN INDICATOR	SERIAL OF PREDICTOR STROKE
TPH	CORRES X-LIT N	1292	IVF	SUBJCT	N1	00	02	000
1938	CORRES X-LIT 1939	1293	IVF	PREP C	N1	00	02	000
A	AN	1295	*	SUBJCT	TO	01	01	000
A	CORRES X-LIT A	1295	IVF	NADJ FM	N1	01	02	000
EPB	EN	1296	*	INFINT	MO	01	02	000
EPB	CORRES X-LIT EN	1296	*	ADJ MAS	N1	01	02	000
TPHU	NEW	1299	*	ADV MAS	AO	01	03	296
TPHU	NEW	1299	*	ADV MAS	MO	01	03	296
TPHU	NEW	1299	*	INFINT	MO	01	02	000
TPHU	CORRES X-LIT NU	1299	IVF	ADV MAS	AO	01	02	000
APB	AND	1300	*	INFINT	C1	01	02	000
APB	CORRES X-LIT AN	1300	IVF	ADJ MAS	N1	01	02	000
TYEUF	DIFFERENCE	1301	*	SUBJCT	N1	01	03	300
TYEUF	DIFFERENT	1301	*	ADV MAS	AO	01	03	296
TYEUF	CORRES X-LIT DIF	1301	IVF	SUBJCT	N1	01	02	000
TFUP	CORRES X-LIT TIP	1302	IVF	ADJ MAS	N1	01	02	000
F	HAVE	1303	*	ADV MAS	V8	02	03	296
F	HAVE	1303	*	PRED H	V8	02	01	000
F	OF	1303	*	INFINT	RO	02	02	000
F	CORRES X-LIT F	1303	IVF	NADJ FM	N1	02	02	000
HPEUT	LIT	1304	*	ADV MAS	AO	02	03	296
HPEUT	CORRES X-LIT LIT	1304	*	PREP C	N1	02	02	000
SPRS	ZORE	1305	IVEA	ADJ MAS	N2	02	03	300
SPRS	SORE	1305	IVEA	NADJ FM	N2	02	00	300
SPRS	ZORE	1305	IVEA	NADJ FM	N2	02	00	304
SPRS	SORE	1305	IVEA	ADV MAS	V2	02	03	296
SPRS	ZORE	1305	IVEA	ADV MAS	V2	02	03	296
SPRS	SORE	1305	IVEA	PRED H	V2	02	01	000
SPRS	ZORE	1305	IVEA	PRED H	V2	02	01	000
SPRS	CORRES X-LIT SORE	1305	IVF	ADJ MAS	N1	02	02	000

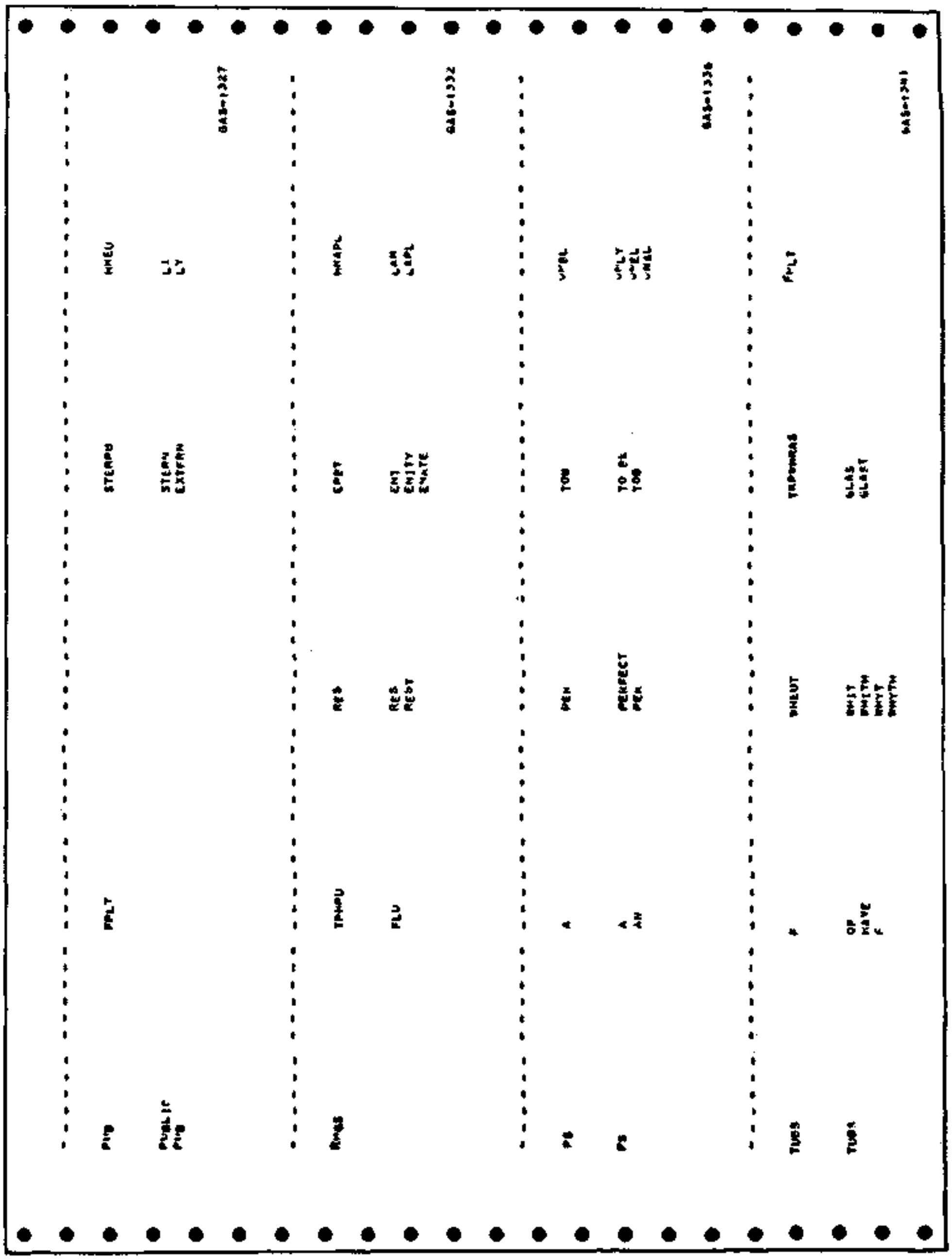
Hindsight Output

Fig. 7



Updating *of* Prediction Pool

Fig. 8



Edited Interlinear Translation
Before Syntactic Analysis

Fig. 9

REFERENCES

1. STEVENS, H. N., "Toward a Model for Speech Recognition", *JASA,* 1960 **32,** No. 1.
2. FRY, D.B. and DENES, P., "An analogue of the Speech Recognition Process", *Symposium on the Mechanization of Thought Processes, National Physical Laboratory,* Her Majesty's Stationery Office, London 1959.
3. PETERSON, G.E., WANG, W. S-Y., and SIVERTSEN, E., "Segmentation Techniques in Speech Synthesis", *JASA,* 1958, **30,** No. 8.
4. LOCHBAUM, C.C., "Segmentation of Speech in Voiced Sounds, Unvoiced Sounds, and Silence", *JASA,* 1960, **32,** No. 7.
5. PARKER-RHODES, A. F., "An Electronic Computer Program for Translating Chinese into English", *MT,* 1956, 3, No. 1.
6. TAKAHASHI, S., WADA, H., TADENUMA, R., and WATANABE, S., "English Japanese Machine Translation", Information Processing, Butterworths Scientific Publications, London, 1960.
7. LAMB S.M., "Segmentation" Proceedings of the National Symposium on Machine Translation, Los Angeles 1960.
8. CLEAVE, J.P., "Braille Transcription", *MT,* 1955. 2, No. 3,
9. SALTON, Gerard, "The Automatic Transcription of Machine Shorthand", *Proceedings EJCC,* December 1959.
10. RHODES, Ida, "A New Approach to the Mechanical Syntactic Analysis of Russian", *National Bureau of Standards, Report No. 6595,* November 1959.
11. SHERRY, M.E. "Predictive Syntactic Analysis", *Proceedings of the National Symposium on Machine Translation,* Los Angeles 1960.
12. OETTINGER, A.G., "Current Research on Automatic Translation at Harvard University", *Proceedings of the National Symposium on Machine Translation,* Los Angeles 1960.
13. BAXENDALE, P., "A Statistical Analysis of the Pattern Structure of the English Sentence", *IBM Research Memorandum RR-MR-27,* San Jose Research Laboratories, September 1958.
14. MILLER, G.A., "Human Memory and the Storage of Information", *IRE Transactions on Information Theory,* 1956 TT-2, No. 3.
15. YNGVE, V.H., "A Model and an Hypothesis for language Structure", *Proc. Amer. Phil. Soc.* 1960, **104,** No.5.