

A METHOD OF TRANSLATING SIMPLE
CHINESE TEXTS INTO ENGLISH BY MACHINE

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
David Djen-Hsien Cheng
1963



This is to certify that the
thesis entitled

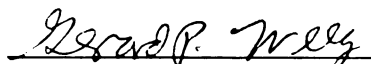
A METHOD OF TRANSLATING SIMPLE CHINESE
TEXTS INTO ENGLISH BY MACHINE

presented by

David Djen-Hsien Cheng

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Electrical
Engineering


Major professor

Date June 28, 1963





RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

FEB 22 1967
6984227

ABSTRACT

A METHOD OF TRANSLATING SIMPLE CHINESE TEXTS INTO ENGLISH BY MACHINE

by David Djen-Hsien Cheng

In translating Chinese to English by machine there are three areas to be considered; namely (1) dictionary storage and retrieval of Chinese characters, (2) selection of the correct meaning of the word and (3) syntax. In view of the magnitude of the task involved, this thesis is confined to cover the first and third areas, for which an automatic translating system is developed. Moreover, the system is primarily designed to translate simple Chinese texts where the writings do not involve complicated syntactic structures.

Using the radical system, a method of storage and retrieval of Chinese characters by machine is described. An evaluation of the radical-oriented look-up system is made. In addition, algorithms are developed to translate two important word order phenomena in Chinese, the idiomatic expression and frame construction. The dictionary storage and look-up program written for the CDC 160-A computer and machine results of a sample translation are also presented.

The syntax system involves the analysis of word order structures in a Chinese sentence and the synthesis into correct counterpart in

David Djen-Hsien Cheng

English. Certain Chinese words are used in a sentence for unique purposes and the translation of such cases can be processed by machine easily. The main syntax translator is derived based on the concept of syntactic unit (synit), which is made of word or string of words bounded by certain relationship. Chinese syntactic structures can be analyzed by studying the relative sequences of the synits. Detection of particular sequences of synits therefore provides the key to successful translation into English. Algorithms for the complete syntax system are developed, and simulated results presented.

A METHOD OF TRANSLATING SIMPLE CHINESE TEXTS
INTO ENGLISH BY MACHINE

by

David Djen-Hsien Cheng

A THESIS

Submitted to the
School of Advanced Graduate Studies of
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering

1963

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Gerard P. Weeg, Professor of Electrical Engineering and Mathematics, for his guidance and encouragement in the development of this thesis.

To his parents, he is ever grateful for their constant encouragement and understanding.

Sincere thanks are extended to Dr. Lawrence W. Von Tersch, Dr. Richard J. Reid, Dr. Howard E. Campbell and Mr. James P. Wang for serving on the Guidance Committee.

TABLE OF CONTENTS

CHAPTER	Page
INTRODUCTION	1
1. Characteristics of a Language	1
2. Goals of Mechanical Translation	2
3. Developments Made in Machine Translation of Chinese to English	3
4. Objectives of the Thesis	3
I. DICTIONARY RETRIEVAL AND STORAGE	7
1. Nature of the Chinese Character	7
2. The Radical System	7
3. Word Look-Up Based on the Radical System	9
4. Evaluation of the Radical Look-Up System	10
5. Dictionary Information Storage	11
6. Some Solution to the Meaning Problem	14
a. Idioms	14
b. Frame Constructions	15
7. An Algorithm of Idiom and Word Look-Up	19
II. SYNTAX	25
1. Introduction	25
2. Recognition Grammar Analysis	26
3. Discussion of the Syntax Flow Chart and Associated Algorithms	29
a. Processing "Past-Tensors"	29
b. Deletion of Words	31
c. Formation of Syntactic Unit ("Synit")	32
d. Synit Sequences Involving 的	42
e. Reformation of Synit of Third Order	49
f. Other Synit Sequences	53
g. English Equivalents	57
4. Conclusion	60

LIST OF TABLES

TABLE	Page
1. Various Locations of Radicals	11
2. Frame Constructions	16
3. Look-Up Results by Hand	24
4. An Illustrative List of Classifiers	31
5. Code Assignment for Synits I	40
6. Code Assignment for Synits II	41
7. A List of Frequently Used Prepositions in Prepositional Synits	55

LIST OF FIGURES

FIGURE	Page
1. Information Box	13
2. Algorithmic Chart for Processing Frame Constructions . .	18
3. Dictionary Information Storage in CDC 160-A	20
4. Algorithmic Chart for Inputting Dictionary Information .	21
5. Algorithmic Chart for Idiom and Word Look-Up	23
6. Syntax Flow Chart	28
7. Algorithmic Chart for Detecting "Past-Tensors"	30
8. Algorithmic Chart for Deletion of 的 Preceded by Adjective	33
9. Algorithmic Chart of Deletion of Classifiers Preceded by Adjectival Pronoun	34
10. Formation of Synits of First Order*	36
11. Formation of Synits of Second Order	37
12. Formation of Synits of Third Order	38
13. Algorithmic Chart for Forming Noun Synit of First Order .	43
14. Algorithmic Chart for Forming Noun Synit of Second Order.	44
15. Algorithmic Chart for Forming Noun Synit of Third Order .	45
16. Algorithmic Chart for Translating a Chinese Phrase whose Sequence of Synits is [V] [N] ₁ [的] [N] ₂	50
17. Algorithmic Chart for Processing Synit Sequences Formed by 的 Preceded by a Noun or Verb Synit	51
18. Algorithmic Chart for Re-formation of Noun Synit of Third Order	52
19. Algorithmic Chart for Translating a Chinese Phrase whose Synit Sequence is [P] [V] [N]	56
20. Algorithmic Chart for Processing Comparative and Superla- tive Words in Chinese	58
21. Algorithmic Chart for Processing Statement of Comparison.	59

LIST OF APPENDICES

APPENDIX	Page
1. The Radical List	61
2. A List of 1500 Chinese Characters	63
3. A Short List of Chinese Words Compiled According to the Radical System	68
4. Classification of Part-of-Speech of the Chinese Word . .	71
5. Look-Up Program and Sample Results Using CDC 160-A Computer	74
6. Simulated Translation of Chinese Passages into English .	79

INTRODUCTION

Language as a means of communication is possibly the most influential factor in the progress of mankind. But expansion of world wide communication has been growing so that it now reaches a point where the chief asset of languages, that of enabling free exchange of information, now becomes an obstacle. As the volume of information in different languages multiplies many fold each day, it is inefficient to trust the translation process to human beings alone. The advent of digital computers opens possibilities toward logical processing of translation in terms of analysis of the source language and synthesis of the target language. This thesis shall deal with two phases of mechanical translation of Chinese texts to equivalents in English.

1. Characteristics of a Language

A natural language is not developed according to a systematic plan. Certainly it does not attribute its growth to the highly stylized structures of logic and reasoning. Rather, a language is a convention that has evolved through centuries of use. It must constantly absorb new ideas and create new words and meanings. Indeed, as a tool of communication to reflect multitudes of facets, languages are developed based on many complex patterns.

2. Goals of Mechanical Translation

The ultimate goal of translating languages by machine such as from Chinese to English is to input to a computer a passage of Chinese and receive as output a lucid English counterpart of the input Chinese. To relate closely with the actual problems, we should aim to accomplish the following:

- (1) Construction of a complete recognition system for Chinese, both in meaning and syntax.
- (2) Development of effective methods to translate Chinese sentences into corresponding English automatically.
- (3) Development of the algorithm and principles of programming for the translation process.

As the problems listed are very much interconnected with each other, the fundamental task becomes that of method of attack. An ideal starting point would be to conduct an exhaustive study of the Chinese language structure, but this is obviously impossible. Furthermore, the established results on Chinese by linguists, compared to the magnitude of the problems, have been meager. In fact, there is yet no established grammatical system that can meet the fundamental requirements of machine-oriented analysis - that of a formal ordering of the structure of Chinese. This lack of linguistic elucidation, mainly due to the fact that linguists have been more concerned with methods of teaching rather than

analyzing the language, certainly does not relieve the existing critical situation.

3. Developments Made in Machine Translation of Chinese to English

A recent report from the National Science Foundation[1] shows only a handful of organizations, all in the U.S.A., that are actively involved in machine translation work from Chinese to English. The major effort in this field comes from the University of California, International Business Machines Corporation, and Ohio State University. At the University of California work has been concentrated on compilation of a set of cross-reference indices for 10,000 Chinese characters, and publication of the results is said to be forthcoming. I.B.M. is also engaged in solutions of encoding problems as well as developing storage devices for the dictionary. The work at Ohio State University is involved with some preliminary studies of the syntactic structure of Chinese. However, other than the mentioned publication, very little published results are in evidence so far.

4. Objectives of the Thesis

In view of the aims and what has been accomplished it is recognized that machine translation work is a formidable task. Indeed, it is so difficult that one can attack only a few isolated problems at a time. But in developing an automatic translator for Chinese, several basic questions must be considered. These are:

How is the Chinese text to be input?

Must the input be performed by one skilled in Chinese linguistics?

Can the computer not only discover the correct meaning of each word but also arrange a meaningful English translation?

To accomplish a meaningful translation the translator must not only use a dictionary to find the meaning of a word, but it must also examine the context in which the word appears. Can a computer do all this?

As a partial solution to some of these problems, this thesis will handle two situations. First, the computer will be made to perform a dictionary search for all the English equivalents of Chinese words, producing not only meanings but also grammatical information. Second, the computer will be presented with the correct English equivalents of the words in a Chinese text and their grammatical uses, and will output a correct translation.

It should be clearly understood that these are two distinct processes. The first process, the dictionary look-up, requires an input of Chinese words. These words can be input by any well-trained clerk, who needs understand no Chinese. The output could conceivably be provided to one familiar with English and only mildly familiar with Chinese, who could then sort out the correct meanings of the passage. In the second

process, the input information to the machine is highly refined, and includes the correct meaning, part-of-speech and other potential syntactic functions of the words. The machine then will proceed to output the English translation.

Clearly there is a gap between the two processes and one which will eventually be closed. It is visualized that a program using both the dictionary look-up and the passage translator should be provided which will on a statistical basis select a meaning for each word in a passage; then attempt to translate the passage on the basis of the word meanings selected together with their parts-of-speech. Then an English speaking person will look at the result and declare to the computer that the result was good or bad. If bad, the computer re-selects the word meaning and tries again. The principle to be employed is similar to the learning machine of Arnold [2].

The translating system in this thesis is designed to translate fairly straightforward Chinese text. The main reason is that these writings have more orderly syntactic structures. The two books constantly used as references are Chinese Reader For Beginners by Shan Wing Chan [3] and Character Text For Mandarin Primer by Yuen Ren Chao [4]. These two books are primarily used by students in beginning Chinese. The attributes of the texts chosen are that they are grammatically correct; require a relatively small vocabulary; and they do not involve complicated structures. Thus the translator will not be of

such a complexity as to overpower a computer; yet on the other hand, the translator will be handling reputable passages which occur in the standard active use of the language.

A word should be given concerning the programs written. The major contribution of the thesis is not so much the programs as it is the algorithms from which the programs sprang. Given the latter, the former can be implemented by any clever programmer. To date, the dictionary program for idiom and word look-up is written and runs. Samples of results of that program are included with the thesis. The second program, the syntactic translator is in the code checking phase, and no results are available. However, in a thesis of this sort, results constitute the "proof of the pudding". Thus, in lieu of the machine results there is included a translation by hand of several passages in Chinese, obtained by strictly following the algorithms presented in Chapter 2.

I. DICTIONARY RETRIEVAL AND STORAGE

1. Nature of the Chinese Characters

Characters or words of the Chinese language derive their present forms through several methods. Two common ones are the pictographic and the ideographic forms. The mixture of sources from which the characters are derived probably leads to the fact that the language does not have a basic system, such as the alphabets in English, by which characters are formed. Rather, Chinese words are unique in construction; through combinations of strokes the possibilities of forms are literally inexhaustible. Since words must be fed into a machine in some code best suited for recognition of the written language, the input of Chinese text to the machine consequently presents a critical problem.

2. The Radical System

During the past thirty years Chinese linguists have come up with almost one hundred different look-up systems. Unfortunately, none has really proved to be superior enough to win any great amount of favor from its users.

Investigations on a system called the radical system has shown that it has features which may eventually prove it to be an effective and unique

look-up system for the Chinese characters. A radical is a character form on the graphemic level somewhat the parallel of alphabetic symbols in English. A Chinese word or character which is not itself a radical consists of two parts, the radical and the auxiliary. The radical is one of the list of 214 radicals (Appendix 1). The auxiliary is the remaining portion of the word, which sometimes can be another radical. In order to look up a word using the radical system, one would first identify the radical associated with the word, and then by going to the list under that particular radical the word in question can be located. The latter step should present no problem unless the list is quite long. In that case, one can divide the list further by the number of strokes each word takes. Resorting to this scheme is not a complicated procedure since the average number of strokes of a Chinese word is approximately ten; which means the counting process, when used, would not be a time consuming job.

Once the radical-oriented vocabulary is compiled, the digital equivalents of the words used as input to the machine can be formulated. The numbering of the digital equivalents shall be decided with special attention paid to the size of storage available in the computer; such a process therefore has to be done arbitrarily. It is conceivable to correspond the equivalent number of the word with the memory location at which the dictionary information for that word is found. Appendix 2 contains a Chinese vocabulary of 1,400 commonly used words, and the listing is based on the radical system. A smaller vocabulary, to be used for

demonstration purposes later, is shown in Appendix 3, and an arbitrary digital equivalent is indicated alongside each word.

3. Word Look-Up Based on the Radical System

Word look-up is thus a series of matching processes. Prior to any machine input one would first locate the word to be translated from the radical list. Then the digital code of that word is used as input to the computer, which in turn proceeds to locate the section of the memory containing all the information on grammar and meanings of that word. This information is then placed where needed for the next stage of the translation process.

As an illustration, assume that a portion of a certain Chinese passage to be translated contains the words

我 回 家 了

From Appendix 3, these 4 words can be located under the radical lists of 戈, 口, 宀 and 丿 respectively. The digital codes for them are 2470, 1730, 2220 and 1100, which are to be input to the machine as unique equivalents to the 4 words. The machine recognizes these 4 digital codes and proceeds to bring out from memory information on the words. It can be seen that if the digital equivalents are exactly the memory locations of the words the retrieving step is made easier. A detailed description of programming procedures using the CDC 160-A Computer is shown in Section 7

of this chapter.

4. Evaluation of the Radical Look-Up System

The look-up system so described does not require the input operator to be familiar with the Chinese language. A person with some intuition on symbols can be trained to recognize the set of radicals. This step of human look-up prior to machine input seems to be unavoidable, at least until automatic pattern recognition schemes are further developed.

While it is ideal that Chinese characters can be looked up by any "symbol-minded" person, efficiency and uniqueness are other essential qualities to be considered in a look-up system. The radical approach gives a unique system, but not a very efficient one. First, 214 radicals are excessive. Next, inconvenience arises from the fact that there is no standard for pinpointing exactly the placement of the radical in the character. The list shown on the next page illustrates the various positions in which a radical can be found in a word.

A recent publication by Tu Hsui-Chih [5] indicates his efforts to revise the radical forms so that the final list contains only 85 radicals. This is a reduction of 129 from the original one and definitely takes a giant step toward a simplified system. Along with the reduced list, Tu is working on a look-up system of characters by their structural compositions. A word is to be categorized into one of five classes: upper-lower, left-right, outer-inner, middle-side, and major-minor. This partitioning

Table I Various Locations of Radicals

Radical in upper part of word:	小	in	尚
lower part of word	一	in	丘
left part of word	二	in	况
right part of word	土	in	壯
middle part of word	目	in	真
Radical split in the middle part of word:	衣	in	衷
split in the middle part of word	行	in	衍
split in the upper part of word	臼	in	舉
split in the lower part of word	人	in	來
Radical in upper-center part of word:	玄	in	率
lower-center part of word	手	in	承
left-center part of word	口	in	喪
right-center part of word	佳	in	獲
upper-left part of word	土	in	報
lower-left part of word	子	in	執
upper-right part of word	勹	in	匏
lower-right part of word	氏	in	坻

of characters provides a means of look-up free from any confusion due to the difficulty of locating the radical in a character. However, all these works are still far from any fruitful result.

5. Dictionary Information Storage

The role of a dictionary is to store all the information pertinent to the meaning of the word and its potential functions in a sentence. This variety of information is to be divided into several fields. The

first field shall contain all possible meanings of the word, either by itself or combined with other specific words. The codes of the meanings are actually represented by the locations of their equivalents in English. The parts-of-speech of the different meanings should also be stored in their respective order. These parts-of-speech constitute the most essential grammatical information of the words, and consequently are extremely useful at the stages of meaning determination and syntax. The classification of the part-of-speech must therefore be as refined as possible so as to allow precise analysis of syntactic structures. Appendix 4 lists all the parts-of-speech and their sub-fields. This classification is not far different from that usually found in a dictionary. One noticeable change is that of shifting a number of words into a category called special terms, which are commonly used words or groups of words whose grammatical roles are too complex to be classified into single part-of-speech categories.

Lastly, one other field must be incorporated to include codes which reflect the potential function of a word in a sentence. This function supplements the part-of-speech to serve in the capacity of detecting word formation in the context of the word. More will be said of this in the chapter on syntax.

The method of storage of the dictionary information can only be decided by the particular memory available. The conversion into bit forms is simple, but the word length of the computer governs the amount of information that can be stored in each word. Generally, one may

conceive of an information box to associate with each meaning of a word. The number of information boxes then equals the number of meanings a word can have. The box content is divided into 3 fields as follows:

Meaning (i.e., Location of English Equivalent)	Part-of-Speech	Potential Syntactic Function
---	----------------	------------------------------

Figure 1 Information Box

The number of meanings of words is obviously varied. Then the storage space allocated for each word should also be varied accordingly, and a means of indicating end-of-list of information boxes of the word is necessary.

From here on strings of Chinese words will be more often considered. A string always ends with a punctuation mark. For uniformity the representation for a string of words is designated by:

$$W_1 W_2 \dots W_k \dots W_p, \text{ where } \begin{array}{l} W_1 = \text{first word of string} \\ W_k = k^{\text{th}} \text{ word of string} \\ W_p = \text{a punctuation mark.} \end{array}$$

The three fields with respect to W_k are abbreviated by

$$\begin{array}{ll} \text{LOC}(W_k) &= \text{location of English equivalent of } W_k \\ \text{POS}(W_k) &= \text{part-of-speech of } W_k \\ \text{FUN}(W_k) &= \text{syntactic function of } W_k \end{array}$$

while the computer word is referred to as

$$\text{INFO}(W_k) = \text{total content of information box of } W_k.$$

This nomenclature will become useful in later sections.

6. Some Solution to the Meaning Problem

Determining the correct meaning of a word from its many possible meanings calls for a thorough contextual analysis of the word. Usually a skilled linguist is required to make the appropriate choice. However, there are two types of word construction in Chinese which are strictly dependent on the rigid order of specific words, and they are (1) idioms or combination words and (2) frame constructions. The meaning of such entities is unique, i.e. if a particular phrase can be recognized as an idiom or a frame construction, its meaning is fixed. Thus, one can incorporate these constructions into the machine program to allow automatic processing and thereby reduce the problem of multiple meaning to some extent.

6a. Idioms

In Bar-Hillel's [6] words, "A given sentence in language L1 is idiomatic with respect to a language L2 to a given bilingual word dictionary from L1 to L2 and to a given list of grammatical rules if, and only if, none of the sequences of the L2 correspondents of the sequences of words of the given L1 sentences is found to be grammatically and semantically a satisfactory translation after perusal of the applicable grammatical rules." Idioms can be removed by extending the dictionary. Their specific patterns of word orders can be stored as part of the word information, and these patterns can provide the necessary diagnostic

clues to insure detection by machine.

For example, the English translation of the word 要 is one of (1) important, (2) to want, and (3) must. We find that in every case when 要 is immediately followed by the word 是, which means (1) to be or (2) right, then the correct meaning of 要是 together is "if".

The digital equivalents of 要 and 是 in Appendix 3 are 4130 and 2500 respectively. The code 2500 must be part of the dictionary information content of the word 要 (4130). There may be other word or words which when preceded by 要 form a single meaning, and the digital codes for those words should also be pre-stored along with that of 是. Therefore, as soon as the word 要 is looked up, the digital code of the word following 要 is also brought in to match with the pre-stored numbers. In the case of 要是, the match is satisfied and the corresponding information box for the meaning of 要是 in combination replaces that of the word 要, and the word 是 need not be looked up.

6b. Frame Constructions

Frame constructions are so named because its members are located at the beginning and the end of a clause separated in the middle by a word or string of words, which is thus framed by the elements of the construction. These constructions usually indicate different types of spatial or temporal relationships. The initial element of a frame

construction is a verb-preposition. The closing element is a preposition or a noun with spatial or temporal connotation. Following is a list of the most commonly used frame constructions:

Table 2 Frame Constructions

SPATIAL FRAME CONSTRUCTIONS:

在 (to be at) ...	中	=	amid	...
	上	=	above	...
	下	=	below	...
	前	=	in front of	...
	後	=	behind	...
	間	=	among	...
	內	=	inside of	...
	外	=	outside of	...
	裏	=	in, inside	...

TEMPORAL FRAME CONSTRUCTIONS:

當 (to be at) ...	時候 (time)	=	when	...
到 (to, at) ...	時候 (time)	=	at the time of	...
在 ...	時候 (time)	=	when	...

Indicators of the frame construction are assigned to both the initial and final elements. For instance, take 在 ... 時候 (when ...) in a clause; the machine first detects the initial element indicator in the word 在, and proceeds to find the final element indicator within the clause. In this case it is found in 時候, and then

the list of all possible frame constructions, which has been pre-stored in the machine, is consulted. Eventually the particular combination of words (i.e., 在 and 時候) is found in the list and the meaning of "when" and its associated information take the place of the meaning of 在, and 時候 is deleted.

Sometimes similar word formations prove to be a false frame construction. The clue lies in members interior to the frame construction. For example, 在 (at) ... 時候 (time) in the list shown would mean "when". Consider now 在甚麼時候, where 甚麼 means "what", this phrase then should be translated "at what time", or in the natural order in which the words appear. Words like 甚麼 constitute the clue, and will be called a frame stopper.

During machine processing, the indicating signs for frame construction can not be found by examining the parts-of-speech of the words. But these signs can be incorporated into the category of potential functions of the words. Assuming the following arbitrary codes:

- 11 for frame construction initial element
- 12 for frame construction stopper
- 13 for frame construction final element,

the flow chart describing the automatic processing is shown in Figure 2.

Using the nomenclature of Section 5 of Chapter 1, we assume the string of words to be examined is represented by $w_1 w_2 \dots w_k \dots w_p$.

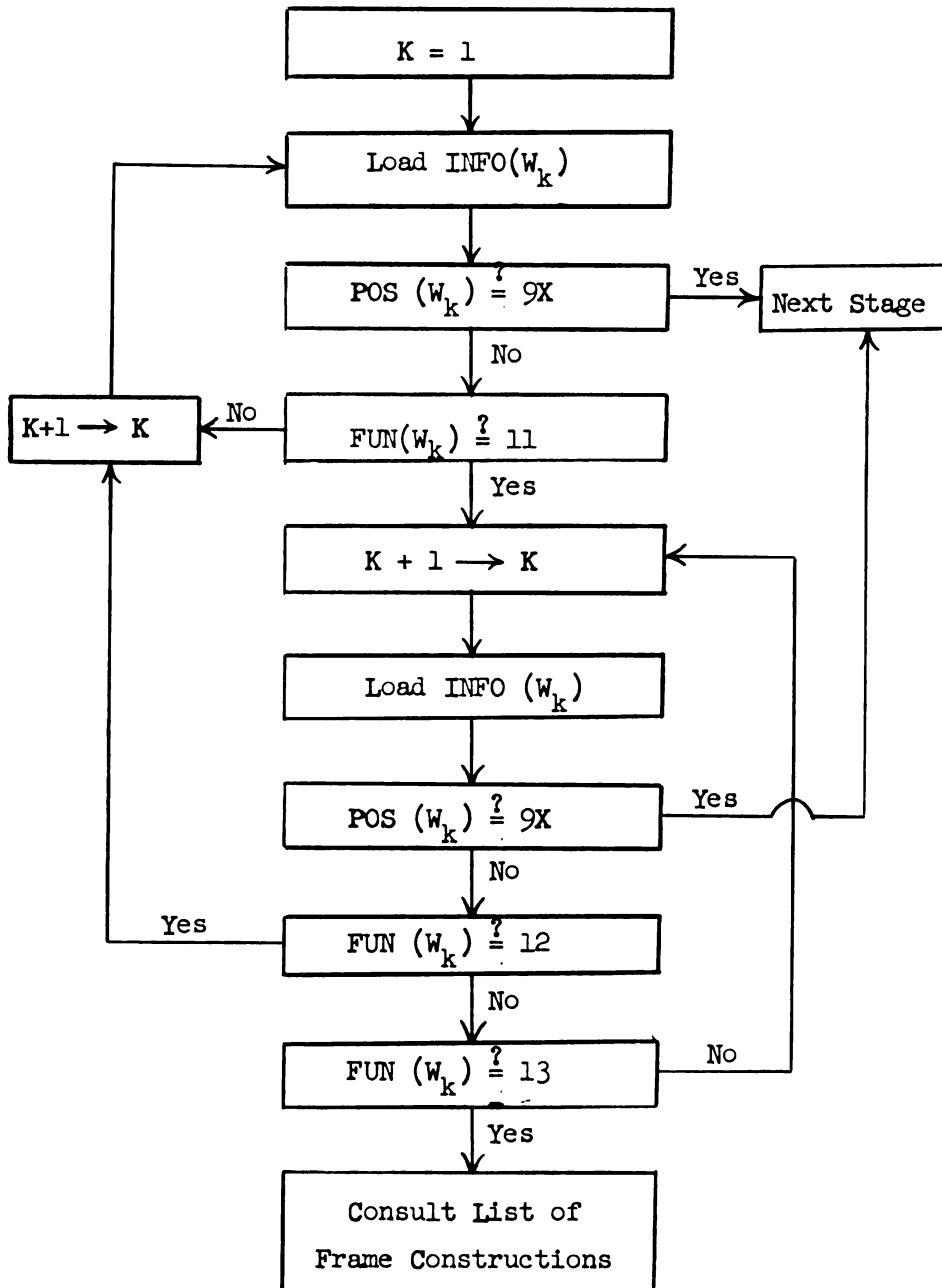


Figure 2 Algorithmic Chart for Processing Frame Constructions

7. An Algorithm of Idiom and Word Look-Up

The system of word storage and look-up designed for the CDC 160-A computer is now described in two parts. The first part deals with storage of dictionary information into memory. The second part describes the look-up procedures.

Each word needs ten octal positions to store its information, which consists of two idioms involving the word and one meaning*. For illustration take the word 學_子 ; from the radical list under 子 in Appendix 3, 學_子 is identified to be numerically equivalent to 2150. The two-word idioms involving 學_子 as the first word are 學校 and 學期 . The numerical equivalents of 校 and 期 are 3050 and 3030 respectively, and the locations of the English equivalents of the idioms are 5110 and 5020 respectively. Both idioms are nouns (numerical code 12), and have no significant syntactic functions. The meaning of 學_子 by itself is "to learn" which is a verb (numerical code 41) and whose English equivalent can be found at location 3020. Hence, the memory locations 2150-2157 inclusive shall store the information associated with 學_子 in the following fashion:

* For convenience, we shall do away with the multiple meaning problem by incorporating only the correct meaning of the word in its context. This does not hamper the basic principles of storage and look-up since any additional meaning of the word merely increases the memory space but the storage and look-up formats are similar. Moreover, the restriction to two idioms clearly does not alter the basic process; but for a small computer, such as the CDC 160-A, does make the memory go farther.

2150	3	0	5	0	Digital Equivalent of 校
2151	0	0*	1	2	Syntactic Function and Part-of-Speech of 學校
2152	5	1	1	0	Location ^{of} English Equivalent of Idiom 學校
2153	3	0	3	0	Digital Equivalent of 期
2154	0	0	1	2	Syntactic Function & Part-of-Speech of 學期
2155	5	0	2	0	Location of English Equivalent of Idiom 學期
2156	0	0	4	1	Syntactic Function and Part-of-Speech of 學
2157	3	0	2	0	Location of English Equivalent of Word 學

Figure 3 Dictionary Information Storage in CDC 160-A

The memory section of 1000-4777 inclusive in CDC 160-A is allocated for storing dictionary information of all the words in Appendix 3. This information must be stored sequentially, that is to say, storage begins at 1000 with the dictionary information of the word 一, continues at 1010 with the word 三, and so on to 4770 with the punctuation mark !. An end of list sign is necessary, and an 0077 is chosen for that purpose. The algorithmic chart for inputting dictionary is shown on the following page.

* 00 designates no syntactic function.

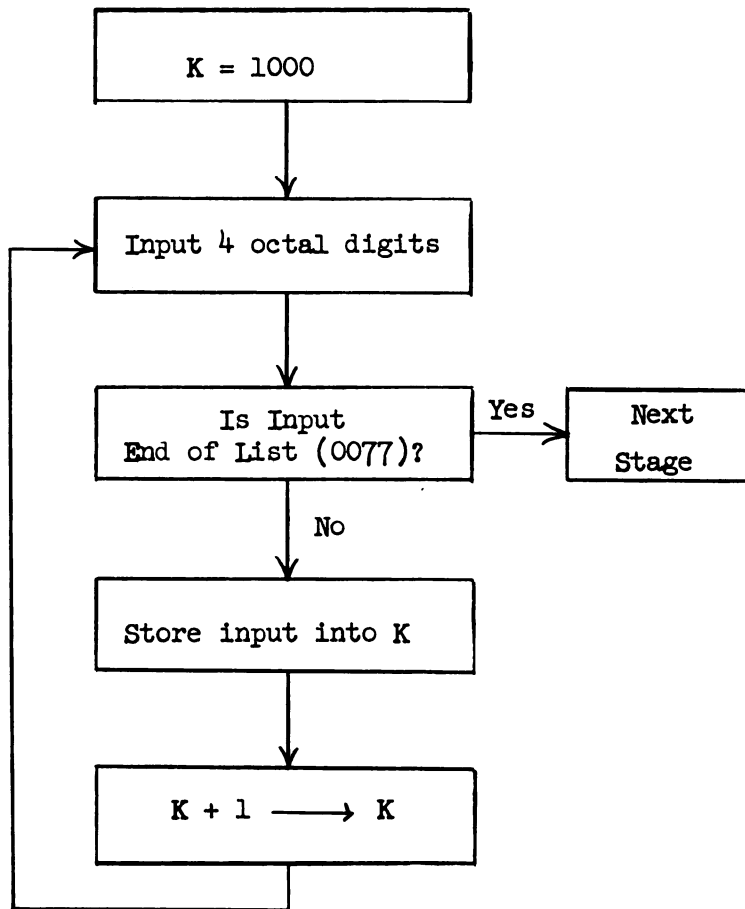


Figure 4 Algorithmic Chart for Inputting Dictionary Information

The look-up process begins with finding the equivalent digital codes of the words to be translated. The method is outlined in Section 3 of this chapter. Suppose the Chinese passage below is to be translated, then from Appendix 3 the equivalent codes are found and indicated directly below the corresponding words.

我	不	知	道	這	個	學	期	在	甚	麼	時
<u>2470</u>	<u>1040</u>	<u>3550</u>	<u>4340</u>	<u>4320</u>	<u>1230</u>	<u>2150</u>	<u>3030</u>	<u>2000</u>	<u>3400</u>	<u>4670</u>	<u>2750</u>
			候	終	止	.					
			<u>1240</u>	<u>3660</u>	<u>3110</u>	<u>4710</u>					

These numerical codes become the input to the machine. As end of list sign is again needed, and an 0077 is again used.

The first step of machine look-up is the search for idioms. If a combination of adjacent words forms a single meaning, it must be found at this stage; for otherwise the single meanings of the words will be output. It should be noticed that the problem of multiple meaning actually lessens with the increasing number of idioms in a passage since these idiomatic meanings are unique. The procedures of idiom look-up is described in the example in Section 6a. The algorithmic chart for the combined idiom and word look-up is shown in Figure 5.

The routine to search frame constructions can be conveniently inserted after the idiom and word look-up. The combined program of dictionary storage and retrieval written for CDC 160-A computer is shown in

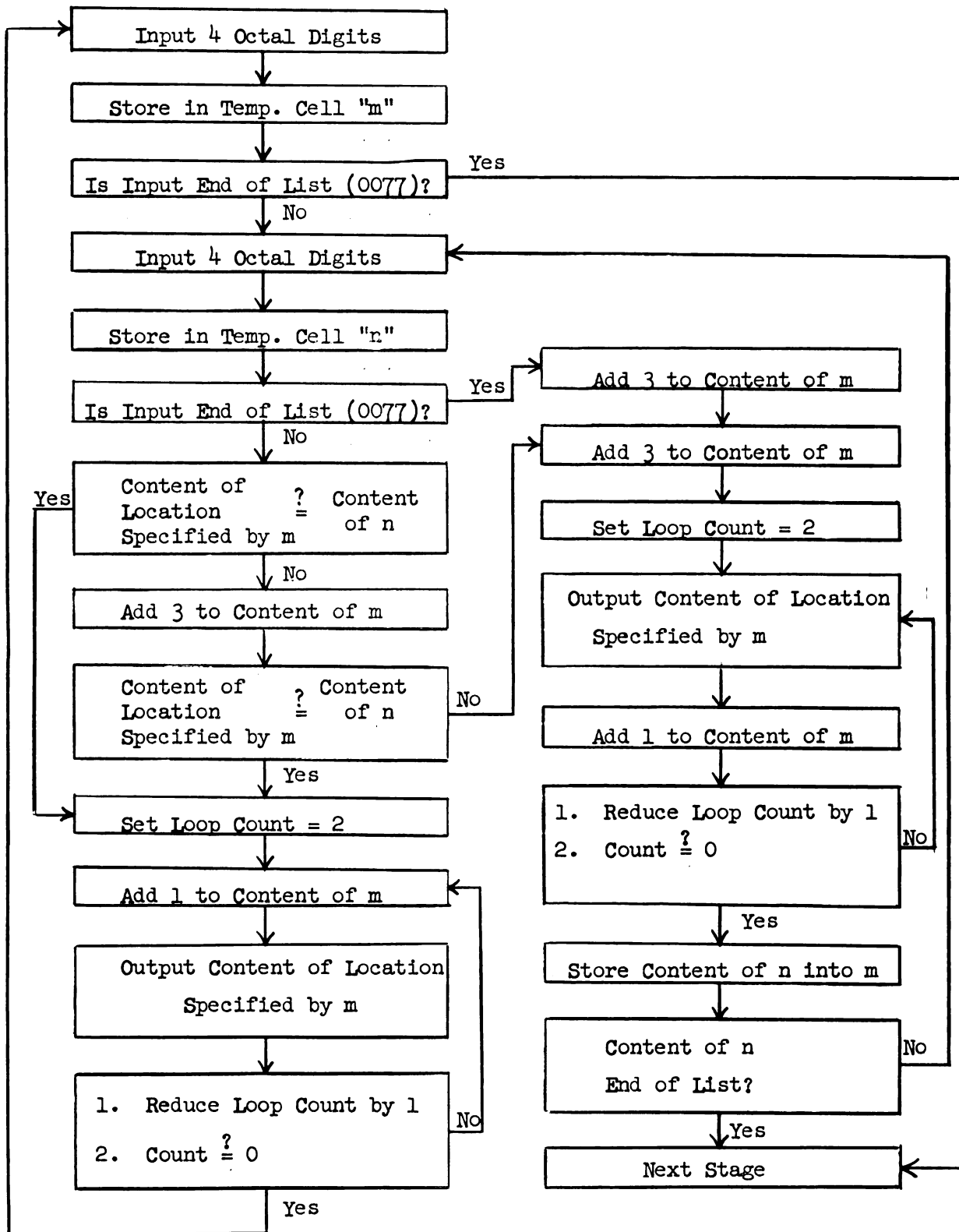


Figure 5 Algorithmic Chart for Idiom and Word Look-Up

Appendix 5, and machine results of translating the given Chinese passage are also included. The results worked out by hand are shown below.

Table 3 Look-Up Results by Hand

Words in Passage	Syntactic function	Part-of-Speech	Location of English Equivalent
我	00	12	2420
不	00	31	3410
知道	00	41	2700
這	00	15	6110
個	00	63	6400
學期	00	12	5020
在	11	44	0150
甚麼	12	82	7030
時候	13	11	6010
終止	00	42	6120
.	00	91	6600

It can be seen that the machine results and hand results are identical.

II. SYNTAX

1. Introduction

Syntax is the process by which words combine to form phrases and sentences according to certain grammatical rules. It is, therefore, that we begin by assuming every sentence has a structure, that is, that each string of words which purports to be a sentence is as a matter of fact a sentence. With the string of Chinese words and its correct meanings and grammatical preperities supplied to the computer, the machine must be made to process these words into proper order for synthesis into English. This means the grammatical contents of the words are to be manipulated while the conceptual contents of the message remains constant throughout the automatic processing. In general, grammatical analysis of a sentence may aim at various results. The method to be used in the analysis may differ considerably, depending on the particular aim at hand. One can set up a generative grammar to produce strings of words which are sentences of a given language. Another type of grammar may be aimed at distinguishing sentences from non-sentences. However, these grammars do not analyze existing features of a sentence. For this, a recognition grammar can be set up which takes a given sentence of a language and finds the syntactic structures of the sentence. The analysis to be taken here

is that of recognition grammar, and assumes that in one way or another there is a decision involved concerning the syntactic relationship of each word string.

2. Recognition Grammar Analysis

The two common methods of describing a sentence are phrase structure analysis and dependency analysis. The former, also known as the immediate-constituent analysis, starts at the syntactic level and considers a step-by-step breakdown into components of increasingly lower order of complexity. The dependency analysis, on the other hand, breaks the sentence up into a tree formation with the predicate as the stem. P. Garvin [7] conceived a method similar to that of the phrase structure analysis with the significant difference being that the minimum units, the words, are taken first and these units are gradually fused into units of higher order of complexity. This thesis follows that approach; in this section a recognition grammar is developed. This leads to an algorithm for translating Chinese to English.

Similar to the process of translation by human beings, syntactic analysis should follow that of look-up and meaning determination of words. As Oettinger [8] has stated, word-by-word translation is a linear approximation to the more sophisticated transformations capable of mapping elements of the source language into elements of the target language. It is generally recognized that these more sophisticated transformations will

be functions, not of isolated words, but of words and their context. However, jumping from word-by-word analysis to an analysis of contextual influence in the broadest sense is not likely to be ideal. Initially restricting analysis to linguistic context, and to a small neighborhood of a word at that, is more likely to lead to results both linguistically significant and concrete.

Therefore, the initial system analysis is to provide for signs that lead first to the readily determinable relations and gradually shift to more complex relations. The initial indicating signs are derived from the part-of-speech codes. From there, groupings of words into possible structural, sequential, meaningful combinations shall take place in terms of the grammatical information contents of the units already determined. The process of recognition is not carried out in an arbitrary manner but in a definite sequence in which completion of the work in one stage of recognition insures the successful translation at the next stage. This combining process is built up until the sentence level is reached.

The flow chart shown on the next page describes the procedures involved in the syntax system.

INPUT: A String of Words $W_1 W_2 \dots W_k \dots W_p$ + their information boxes.

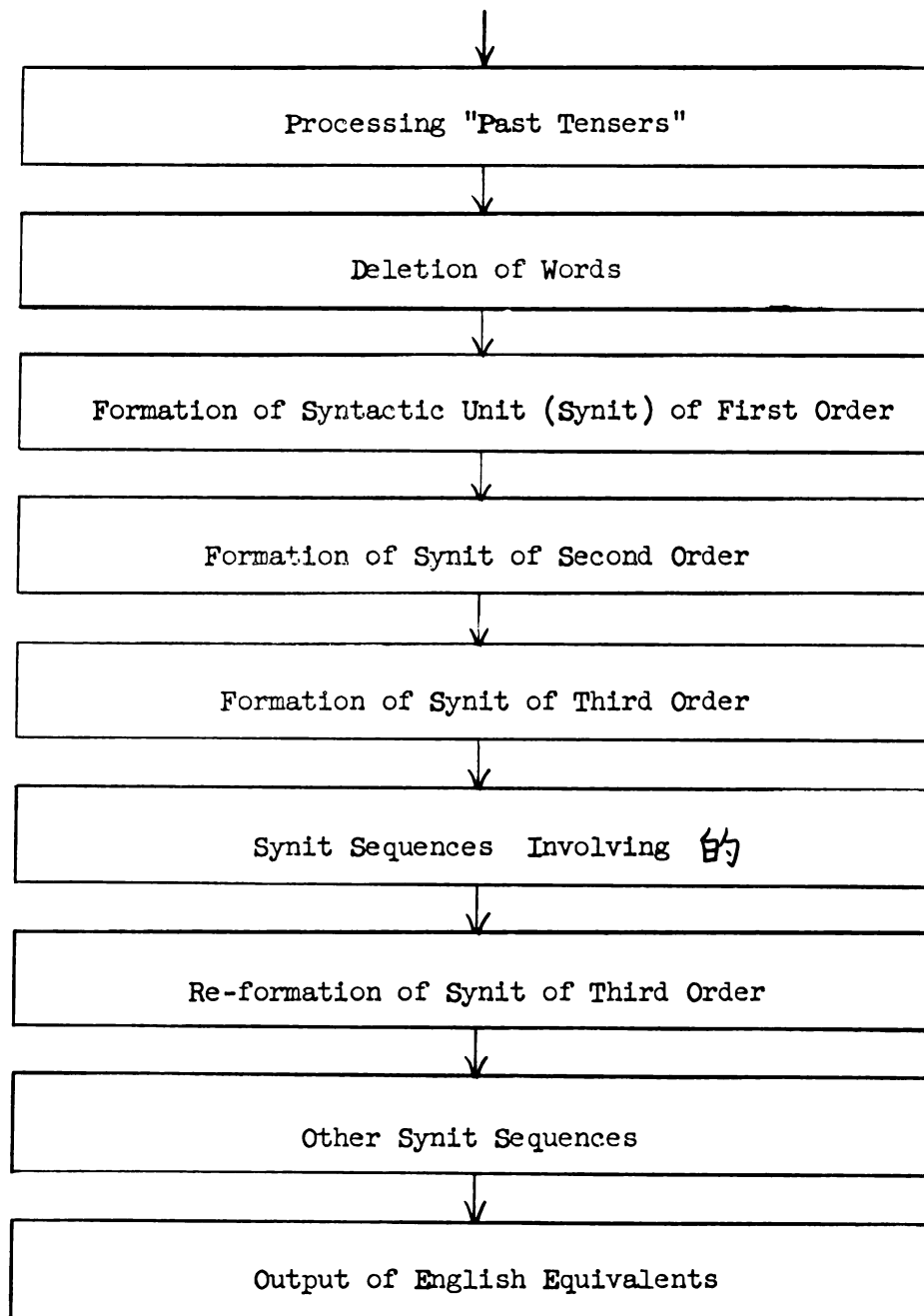


Figure 6 Syntax Flow Chart

3. Discussion of the Syntax Flow Chart and Associated Algorithms

3a. Processing "Past-Tensors"

The fact that Chinese is not an inflected language leads to the necessity of contextual analysis of verbs as to their tenses or aspects.* In one situation, single characters following certain verbs carry the grammatical connotation of past action. There are two words, 了 and 過, commonly used in this sense, which shall be called "past-tensors."

The function of the word 了 in a sentence is many fold, but in the case when it is used to convert a verb into or past tense or completed aspect, it follows immediately after the particular verb. Thus 吃 "to eat" becomes 吃了 "eaten", and 回來 "to return" becomes 回來了 "returned".

When 過 is used as a sign for past tense it is also placed immediately after the verb the tense of which it affects. Thus, "to have seen" in Chinese is 看見過 ("to see" 看見), and "to have eaten" 吃過.

The procedure for making verbs into past tenses is shown on next page.

* Chinese linguists have been debating on whether verb tenses really exist in the Chinese language. Some asserts that the word "aspect" is perhaps more fitting to indicate the temporal order. However, the analysis here is not involved to such depth, and the words "tense" and "aspect" will be used interchangeably.

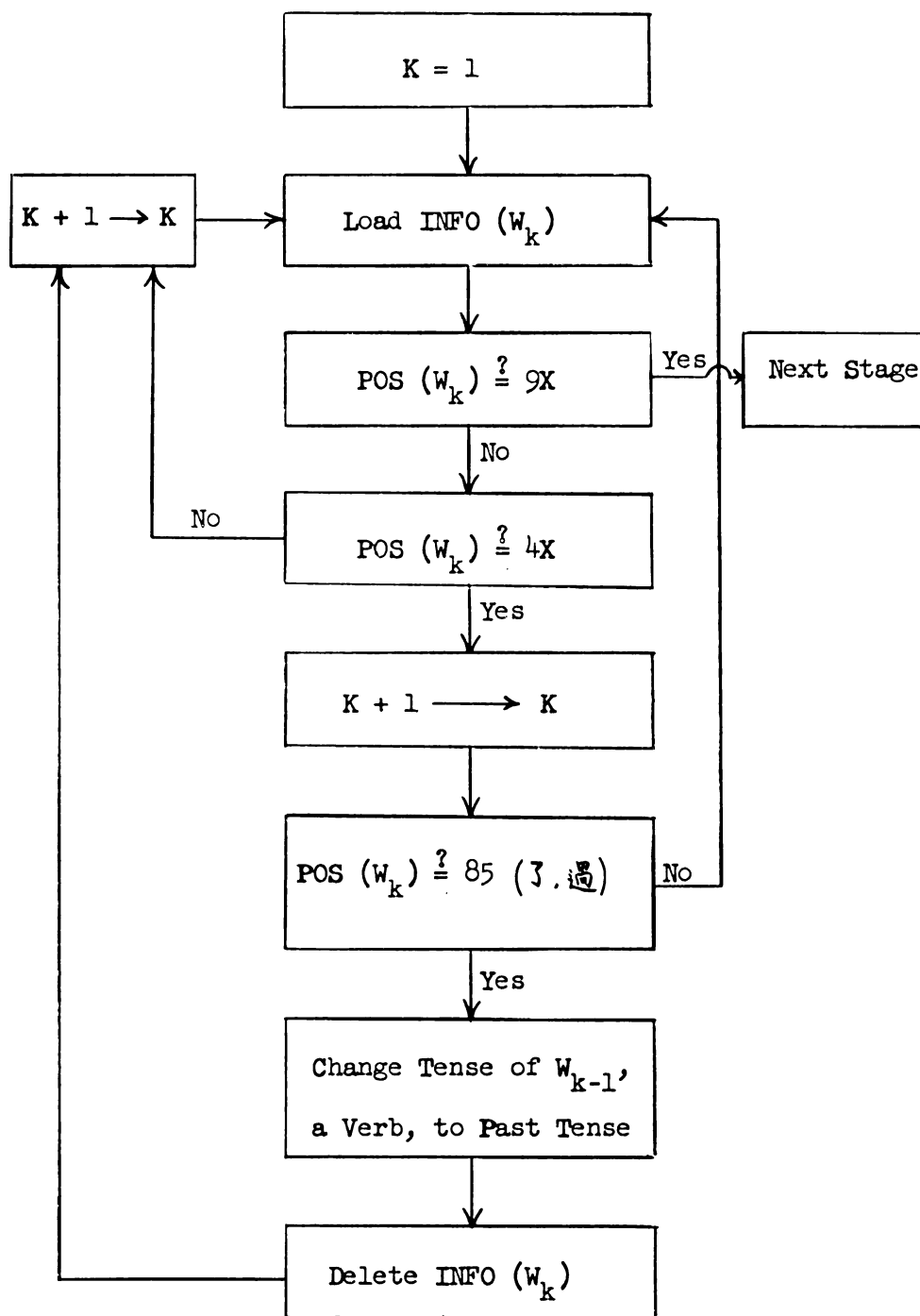


Figure 7 Algorithmic Chart for Detecting "Past-Tensors"

3b. Delection of Words

There are words in Chinese which when translated into English are considered excess baggage. One such case is the usage of 的 after adjectives. The function of 的 here is to carry the modifying tone which is missing when the adjective is used alone in Chinese. Generally it is also grammatically incorrect if the Chinese adjective is not followed by 的. However this is not true in English. Therefore, 的 can be deleted in this case without any misinterpretation.

Another case of excess baggage involves certain usages of units or classifiers. It is a rather interesting phenomenon in Chinese that when a number is used with a substantive, a unit or a classifier must also be included. There are various types of classifiers to associate with the substantives. An illustrative list of these is shown below:

Table 4 An Illustrative List of Classifiers

The classifier for	馬 (horse)	is	匹
"	車 (car)	is	輛
"	信 (letter)	is	封
"	樹 (tree)	is	棵
"	狗 (dog)	is	隻
"	房 (house)	is	所

Thus, "two horses" in Chinese is 二匹馬, and "three cars" become 三輛車, which literally means "two classifier (horse) horses" and "three classifier (car) cars" respectively.

An extension of this phenomenon is the usage of more general terms like the adjectival pronoun "this" and "that" to describe a substantive instead of a specific number. Here, too, the classifier must be used. "This (這) horse" in Chinese is 這匹馬, and "that (那) car" is 那輛車. In the case of specific numbers, the classifier should be retained for the purpose of clearer interpretation. But when used with adjectival pronouns, it can be deleted because the modified terms, like horse and car, do not require the classifiers in English usage.

Algorithmic charts for the deletion processes are shown in Figures 8 and 9.

3c. Formation of Syntactic Units

Analyzing the relations among the components of any particular passage presupposes a prior cataloguing of both the possible significant relations among passage components and the properties of representations that reflect these relations. The relations in question are formal ones, as between adverb and verb, or between noun and adjectival modifier. These relations are typically reflected in the order of the related representations or in structural similarities of the related representations. For example, in Chinese the adjectival modifier always precedes its noun, and adverbs precede verbs. These formally fixed relationships constitute the very information which is needed to construct correspondence with the target language.

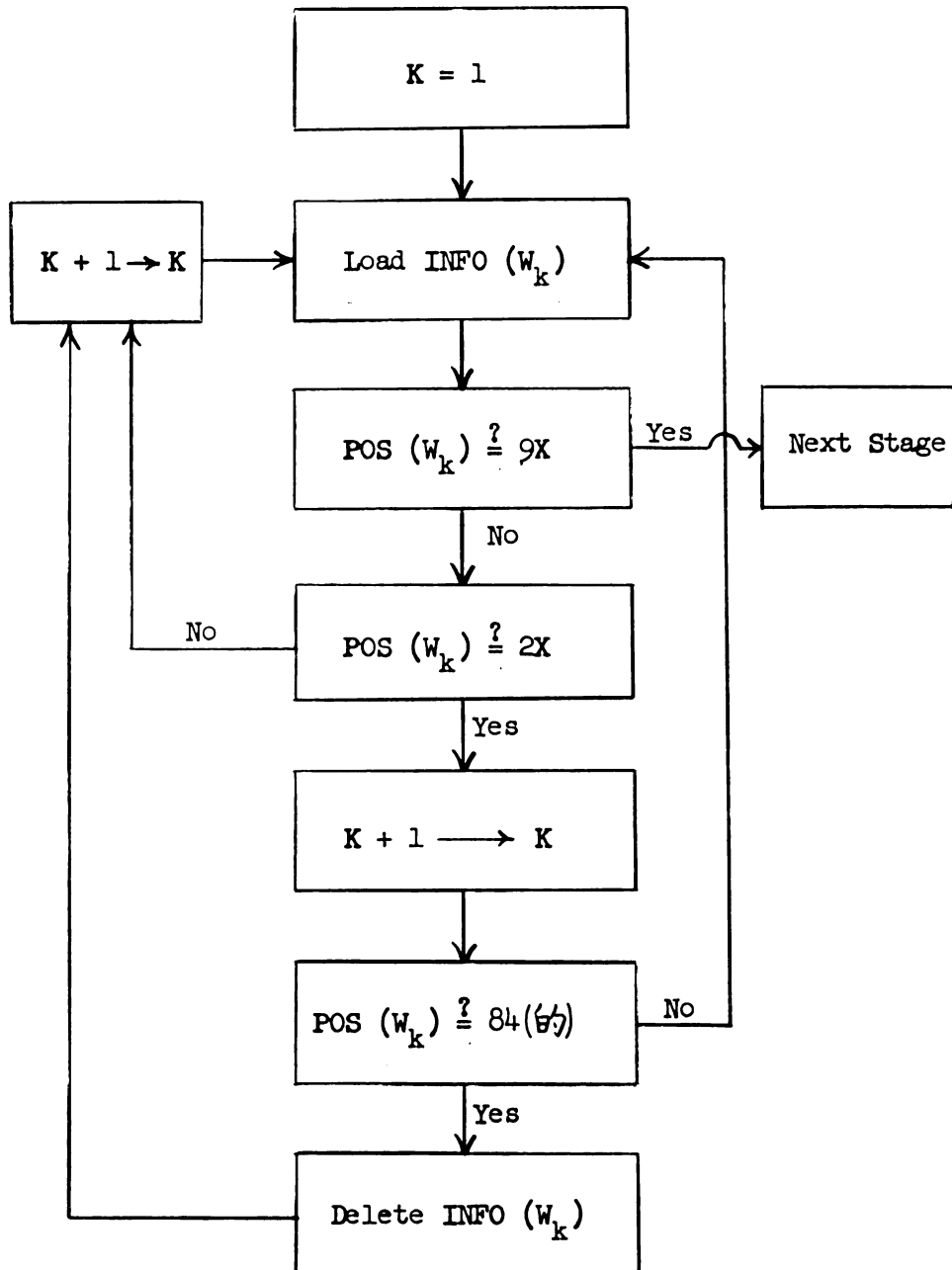


Figure 8 Algorithmic Chart for Deletion of 的 Preceded by Adjective

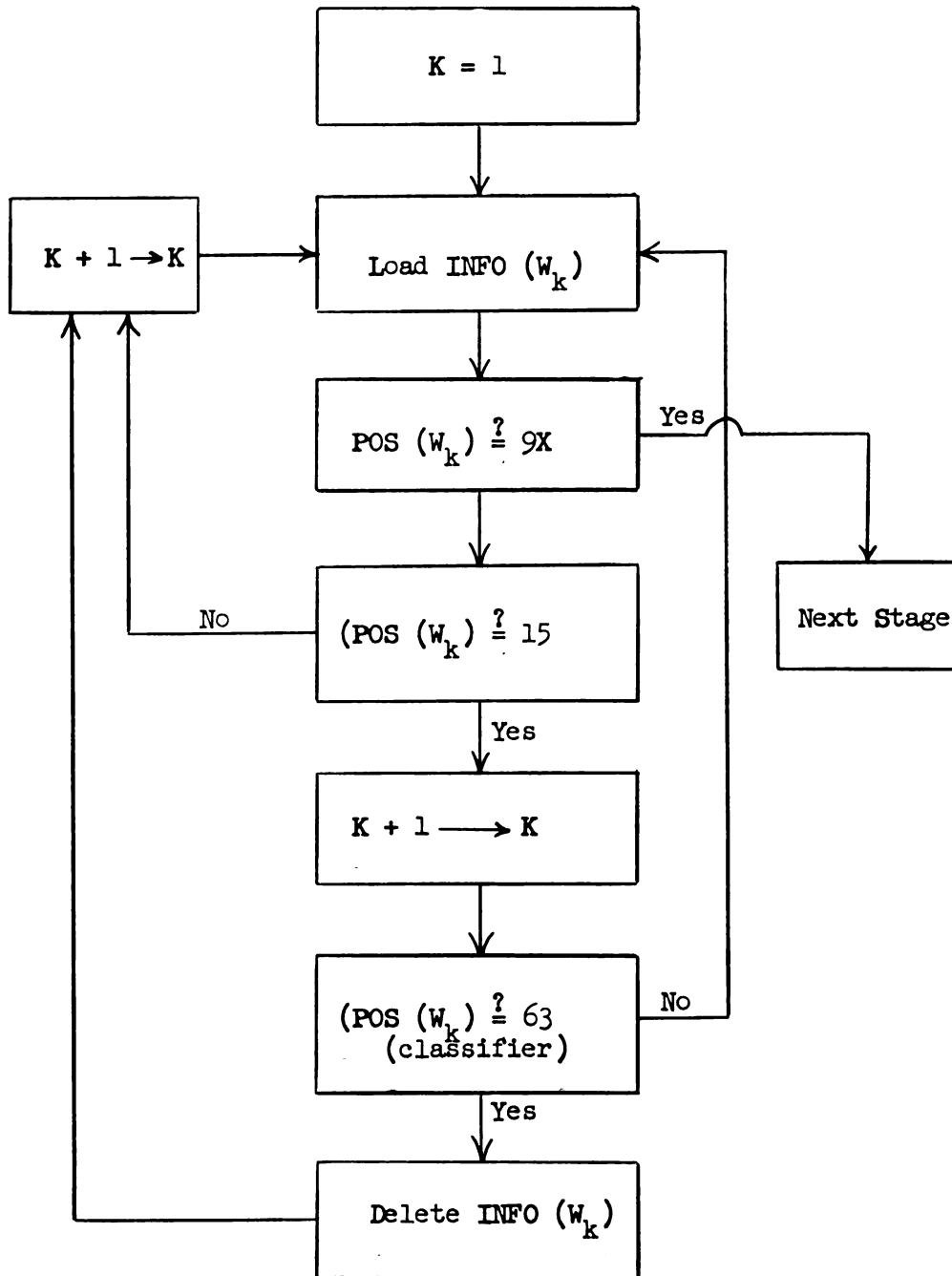


Figure 9 Algorithmic Chart of Deletion of Classifiers Preceded by
by Adjectival Pronoun

Therefore, to establish the relationship among words, we set out to formulate constituent patterns of words in terms of their formal grammatical properties. The immediately accessible ones are their part-of-speech codes. We begin by listing the very simple relationships. When adjacent words satisfy any of the established relationships, they can be bounded to form a unit which shall be called syntactic unit, or "synit" for short. A synit is not necessarily restricted to contain one or two words only; rather it can include a sequence of words as long as the adjacency relationship is satisfied. Gradually the synits of lower orders starting with the first order shall be developed into the composition of synits of higher orders.

In formulating the set of rules to form synits we rely not on linguistic intuition or pre-existing descriptive analysis but on the study of a large number of passages derived mainly from the two books previously mentioned [3] [4]. This has the immediate advantage that one can be sure, even before testing, that one has a set of rules of fairly wide applicability. Furthermore, in dealing directly with the writings, one gains familiarity with the problem areas that may arise and thus is in a better position to resolve the ambiguities. Inadequacies will inevitably occur, but in the light of accumulated results, this empirical approach does provide a working method toward the paramount aim.

The synits of first order are formed by directly searching the part-of-speech codes associated with the words. Assume the following

designations:

Noun-	synit of the first order	- (N)
Adjective-	"	- (A)
Verb-	"	- (V)
Preposition-	"	- (P)
Number-	"	- (#)
Conjunction-	"	- (C)
Special term-	"	- (S)
Punctuation mark-	"	- (\$)

The combination of words, each of which is represented by its part-of-speech code identical with that in Appendix 4 to form synits of the first order, can be made according to the rigid order as follows:

(N)	-	1X 1X 1X
(A)	-	2X 2X 2X 32 2X 32
(V)	-	4X 4X 4X 31 4X 33 4X 33
(P)	-	5X
(#)	-	62 61 63 61 14 53 52 61 14
(C)	-	7X
(S)	-	8X
(\$)	-	9X

Figure 10 Formation of Synits of First Order*

* Each pair of symbols, such as 1X or 4X is to be read as a two digit decimal number whose least significant digit is unspecified.

Thus, for example, number signits of the first order are of the form 61 63, i.e. numerical number followed by a number-associate noun; etc.

Synits of first order are developed into the make-up of synits of higher orders. The following is a list of the synits of second and third orders, derived from synits of first and second orders respectively. Assume double brackets for second-order synit and triple brackets for third-order, i.e. noun synit of second order is represented by ((N)), and third order by (((N))).

((N))	-	(N)
	-	(A) (N)
((V))	-	(V)
((P))	-	(P)
		(P) (N)
((#))	-	(#) (N)
		(#)
((C))	-	(C)
((S))	-	(S)
((\\$))	-	(\\$)

Figure 11 Formation of Synits of Second Order

(((N)))	-	((#)) ((N))
	-	((#))
	-	((N))
(((V)))	-	((V))
(((P)))	-	((P)) ((N))
	-	((P))
(((C)))	-	((C))
(((S)))	-	((S))
(((\$)))	-	((\$))

Figure 12 Formation of Synits of Third Order

It can be seen that there are nine part-of-speech classifications (Appendix 4), and the types of synits of the first order, derived directly from these parts-of-speech codes, is numbered at eight. The number of different types of synits of second and third orders are further reduced to seven and six respectively. The formation of synits is to set the stage for the next step of the syntax process, that of examining sequential ordering of the synits. Therefore, the contents of synits of the highest order must be so constituted that they are compatible with the ensuing needs. For certain, each synit shall represent a distinct entity whose properties are only known through the synit with which it is identified. An examination of the synits of the third order shows that the conjunction, special term, and punctuation mark synits are identical in

content as with the respective synits of first and second orders. In the case of conjunction and punctuation mark synits, their functions are that of subordinating, coordinating, and ending a clause and there is no meaning or concept relationship between these words and other types of words. Special terms, as indicated, are words or groups of words whose grammatical roles are quite complex and therefore should be isolated. The remaining synits of third order are noun, verb, and preposition synits. One may contemplate a further merging of these synits into synits of higher order. However, a closer investigation shows that this will defeat the purpose of synit formation. If the noun synit is combined with the verb synit, then their combination will generalize the conceptual or thought content of the new synit to such an extent that the unique distinction of synit content is lost. For the same reason combining the preposition synit with the verb synit leads to a similar result. The remaining possibility is between noun and preposition synits, but they are both formed from noun synits of second order, therefore the significant relationship between these two synits are already processed during the formation of third order synits. It is then concluded that third order is the upper limit of forming synits. Further evidence will be found in the sections on synit sequences.

The algorithms for formations of synits are now described. The string of Chinese words ending with a punctuation mark is again represented by

$W_1 W_2 \dots W_k \dots W_p$, where W_1 = 1st word of string
 W_k = kth word of string
 W_p = a punctuation mark.

and abbreviations for the different fields of the information box also conform with those previously designated.

Each word shall be preceded by a symbol indicating which type of synit the word belongs to. The symbols for the different orders of synits are arbitrarily assigned numerically as follows:

Table 5 Code Assignment for Synits I

(N) - 11	((N)) - 12	((N)) - 13
(A) - 21		
(V) - 41	((V)) - 42	((V)) - 43
(P) - 51	((P)) - 52	((P)) - 53
(#) - 61	((#)) - 62	
(C) - 73	((C)) - 73	((C)) - 73
(S) - 83	((S)) - 83	((S)) - 83
(\$) - 93	((\$)) - 93	((\$)) - 93

Notice the composition of words in all three orders of conjunction, special term and punctuation mark synits is identical, consequently the code for each is the same.

Each word which is an initial element of a synit is preceded by one of the above symbols, which serves the same purpose as the left-hand bracket. Other words belonging to the same synit also need preceding symbols which are not necessarily identical to that of the initial element. These symbols merely indicate that more words are included in the synit, and are assigned as:

Table 6 Code Assignments for Synits II

Word interior to the noun	- synit - 15
" adjective	- synit - 25
" verb	- synit - 45
" preposition	- synit - 55
" number	- synit - 65
" conjunction	- synit - 75
" special term	- synit - 85

Suppose the words which form a noun synit of the first order are $W_3W_4W_5$ (i.e., $p > 5$). Then they are recognized by the machine as

11 INFO (W_3) 15 INFO (W_4) 15 INFO (W_5).

The symbols for the synits associated with each word, such as 11 and 15, shall be abbreviated in general by SYB W_k .

As each word in the string is processed by the computer, the machine must be provided with three pieces of information in order that the appropriate choice of synit sign preceding the word and the location

where the word is to be stored can be made. These are (1) the symbol of the current synit, (2) first location of the list of $\text{INFO}(W_k)$'s and their preceding SYB_{W_k} 's starting with W_1 , and (3) the location where the current $\text{INFO}(W_k)$ and its associated SYB_{W_k} must be stored. For this purpose, then, three memory locations, to be named s , f , and l , are allocated to store the respective information.

The procedures involved in forming the various synits follow identical patterns. Since the formations of the noun synits embrace more possibilities than others, they are described in detail in the following charts; and formation of other synits can use the same charts with the only difference being that the part-of-speech code and the synit symbol preceding the word are changed wherever applicable. The algorithmic charts for forming noun synit of first, second, and third order are given in Figure 13, 14 and 15 respectively.

3d. Synit Sequences Involving 的

If the same or very similar structural devices are used to represent relations in both source and target languages, the transformation of relation representations would present little difficulty. However, Chinese and English do use different word orders to denote the same grammatical relationships. Therefore, synthesis of Chinese sentences into English requires not only a clear understanding of synit content, but their sequential formations in a sentences as well.

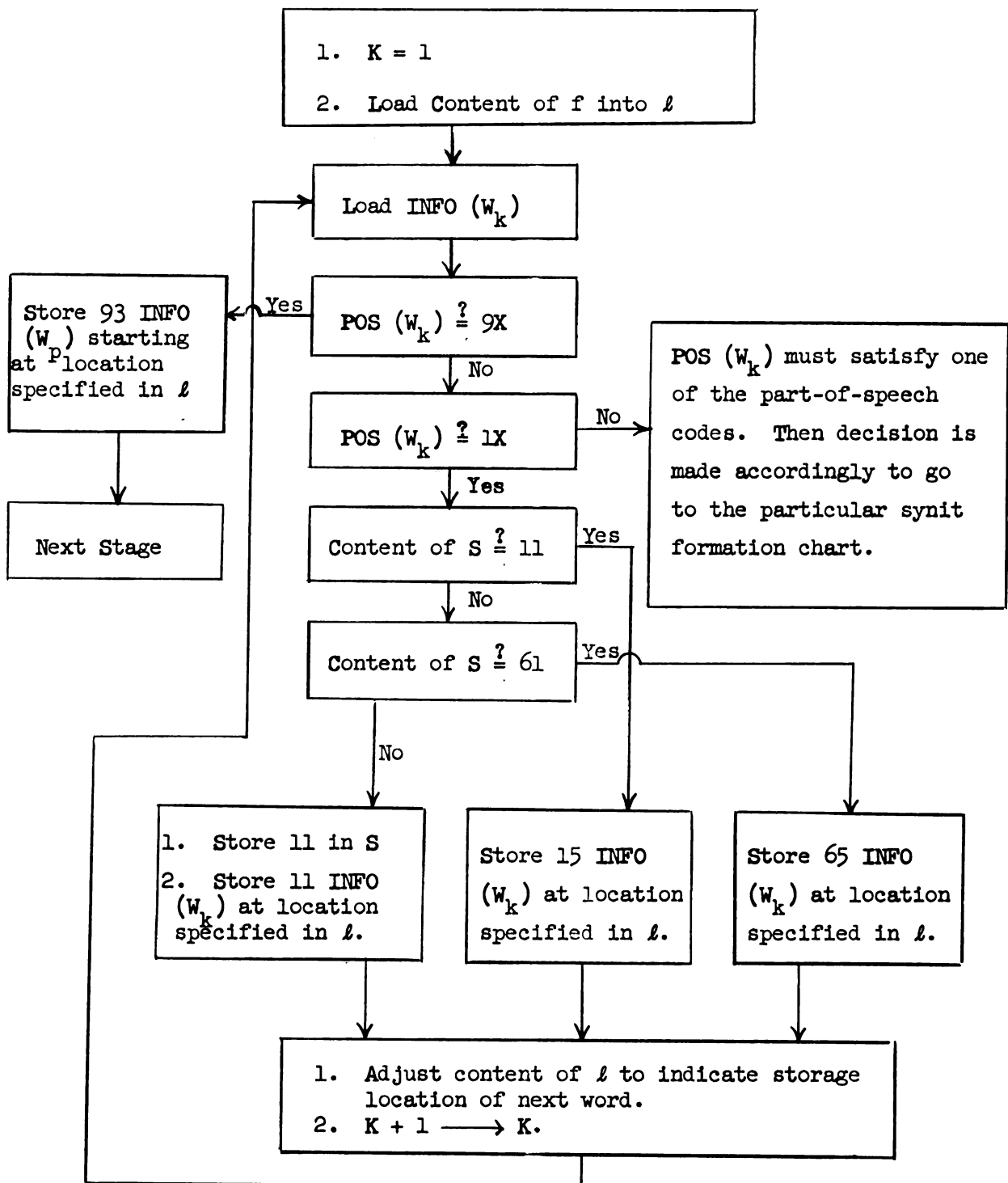


Figure 13 Algorithmic Chart for Forming Noun Synit of First Order

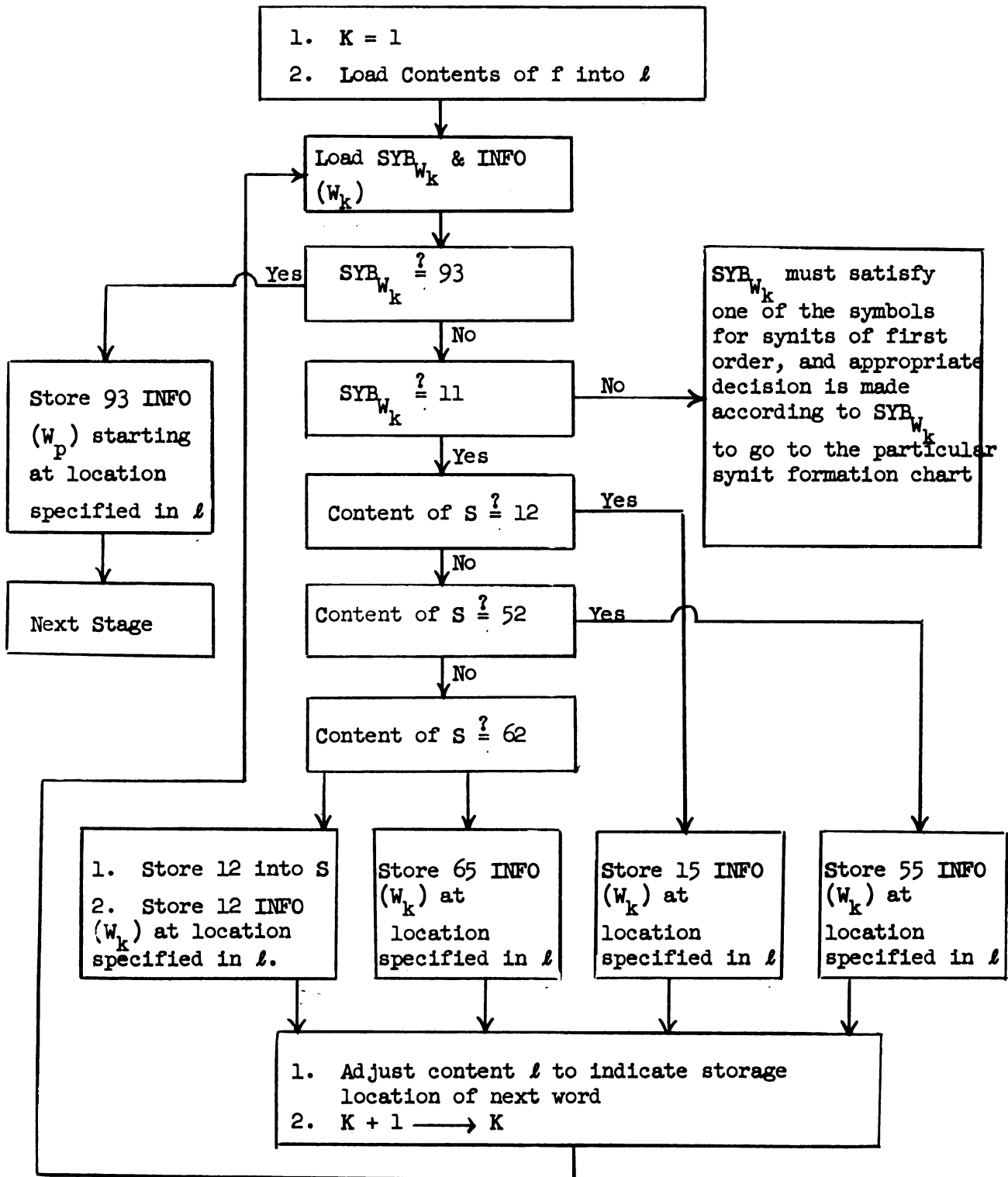


Figure 14 Algorithmic Chart for Forming Noun Synit of Second Order

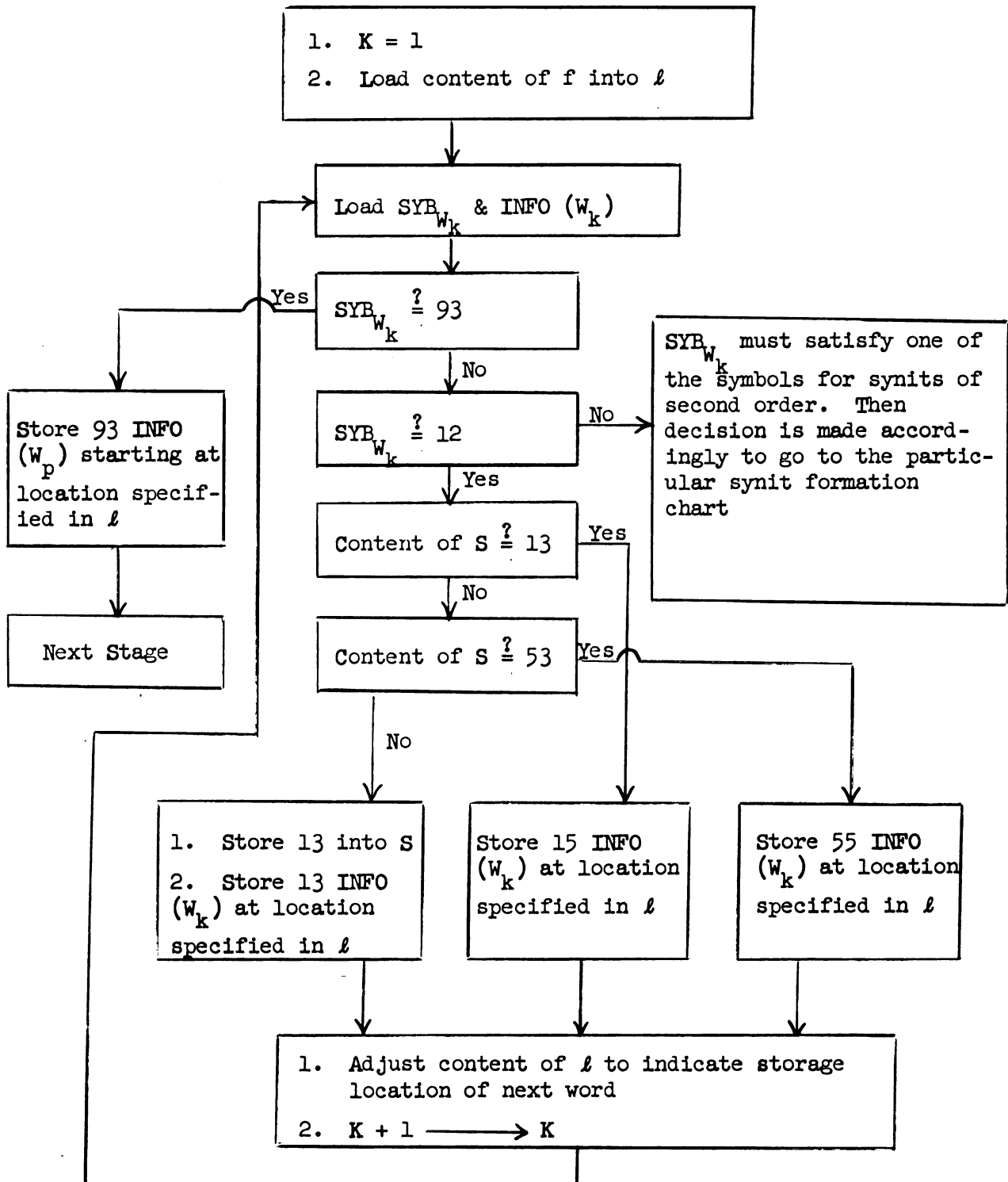


Figure 15 Algorithmic Chart for Forming Noun Synit of Third Order

Synits partition a sentence into localized thought groups, but to weld these different groups into a complete expression requires further investigation of relative sequences of these synits. Additionally, the special terms and the potential syntactic functions of individual words also provide key information toward syntax analysis.

As indicated in Section 5 of Chapter I, a classification of part-of-speech called special terms is necessary due to the complexity of the grammatical uses of these terms. Possibly the most frequently used word among the special terms is the word 的 , which is one of the chief means for the detection of certain relationships among synits in a sentence.

All synits now considered are of third order. ([] is used as equivalent to ((())) .) The particular sequence of synits that is significant is formed as follows:

$$[\text{verb}] [\text{noun}]_1^* [\text{的}] [\text{noun}]_2$$

However, detection of such a sequence of synits is not sufficient to guarantee a synthesis procedure into English. The reason is that not all verb **synits** will satisfy the given formation. Consequently, the information box of the verb in the verb synit must be examined, and in

* Presence is optional in certain cases.

particular the potential syntactic function part of the word. The indicating sign is thus found from the combination of its part-of-speech and its potential syntactic function. In case of more than one verb in the verb synit, it is assumed that the initial verb should be examined, and the assumption is made on the basis that the first element generates the action or thought involved in the verb synit.

The verbs must also be separated into two categories. The first category includes intransitive verbs and verb-prepositions, and the second includes transitive verbs only. The intransitive verb-的 or the verb-preposition-的 sequences are formed by the following order.

1. [verb_{intransitive}] [noun]₁* [的] [noun]₂
2. [verb-preposition] [noun]₁* [的] [noun]₂

An example for each case is shown below:

1. [住在] [美國] [的] [人]
To live in U.S. People

Translation: People who ~~do~~ live in U.S.

2. [在] [南極洲] [的] [一個地方]
To be in Anarctic continent one classifier place

Translation: One classifier place which is in Anarctic Continent

* Presence is optional.

The synthesis procedures take the second noun synit, i.e. [noun]₂ and places it in front of the predicate clause of [verb] [noun]₁ and the word 的 is deleted.

To make the English translation correct, the conjunctive pronoun "which", "who", or "where" must be inserted following the [noun]₂. The choice should be made to agree with the substantive in [noun]₂, but this is difficult if not impossible. We shall merely indicate that a [WH] term is to be inserted to stand for either of "which", "who", or "where".

The 的 in the above case has only one possible function because of the intransitivity of the action involved. If a transitive verb takes the place of the intransitive verb in the same formation of synits then the noun synit [noun]₁ whose presence was optional previously, is now a necessary element to follow the verb. The usage of 的 here can have two interpretations. 的 's function may be that of making the case of the noun synit preceding it possessive. Then synthesis merely involves a process of combining this noun synit as a modifier to the noun synit following 的 ; 的 is deleted but no order change is necessary.

On the other hand, the word 的 may be used to effect a change similar to the previous intransitive verb case. Examples:

1. [訪問] [美國] [的] [學者]
To visit U.S. Scholar(s)

Translation: Scholar(s) [WH] ~~to~~ visit U.S.

2. [教] [英文] [的] [人]
To teach English Person(s)

Translation: Person(s) [WH] ~~to~~ teach English.

It is important that the two different functions of 的 must be distinctly recognized. The proper choice again goes back to the stage of meaning determination where human intervention is necessary.

If by now 的 still remains in the passage without satisfying any of the formations thus far considered then its presence should be located either following a noun synit or a verb synit. In the [noun]- 的 case, the result is to make the noun synit a possessive case, i.e. [noun]'s, and in the verb case the verb is made into past tense. 的 is deleted after the change is completed.

The algorithmic charts for processing the synit sequences involving 的 are shown in Figure 16 and 17.

3e. Re-formation of Synit of Third Order

It is evident that the synit order may be altered after processing all synit sequences involving 的. The possibility arises from the fact that some of the synits which were separated now form adjacent synits which are so related that they can be combined into a single synit according to the rules shown in Section 3c (Figure 15). However, these

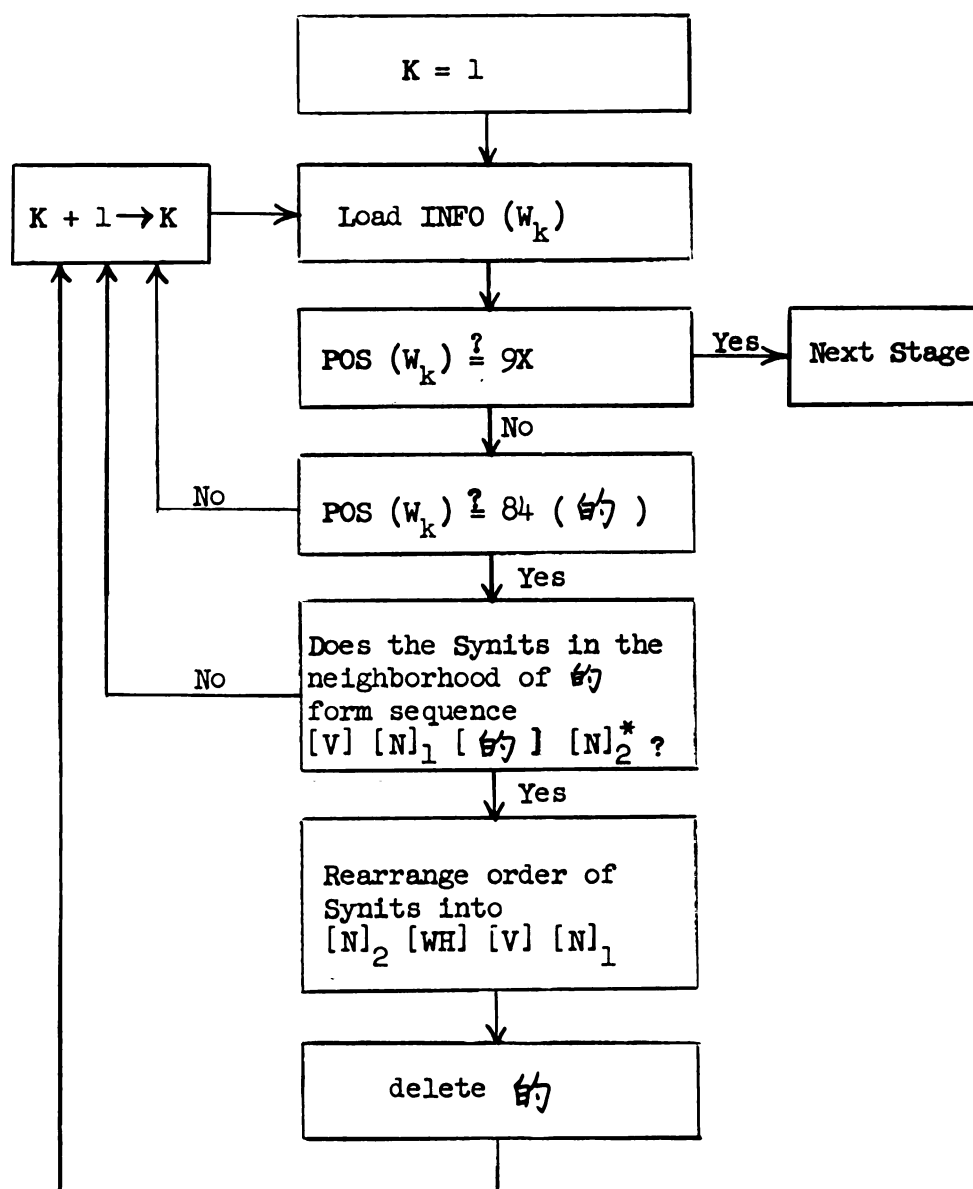


Figure 16 Algorithmic Chart for Translating A Chinese Phrase whose Sequence of Synits is $[V] [N]_1 [的] [N]_2$

* The use of preceding signs to indicate types of synits for words becomes significant here. The detection of the particular sequence of synits, such as $[V] [N]_1 [的] [N]_2$ above, is to be made based on those signs. Subsequent syntax algorithms dealing with synit sequences also employ this scheme.

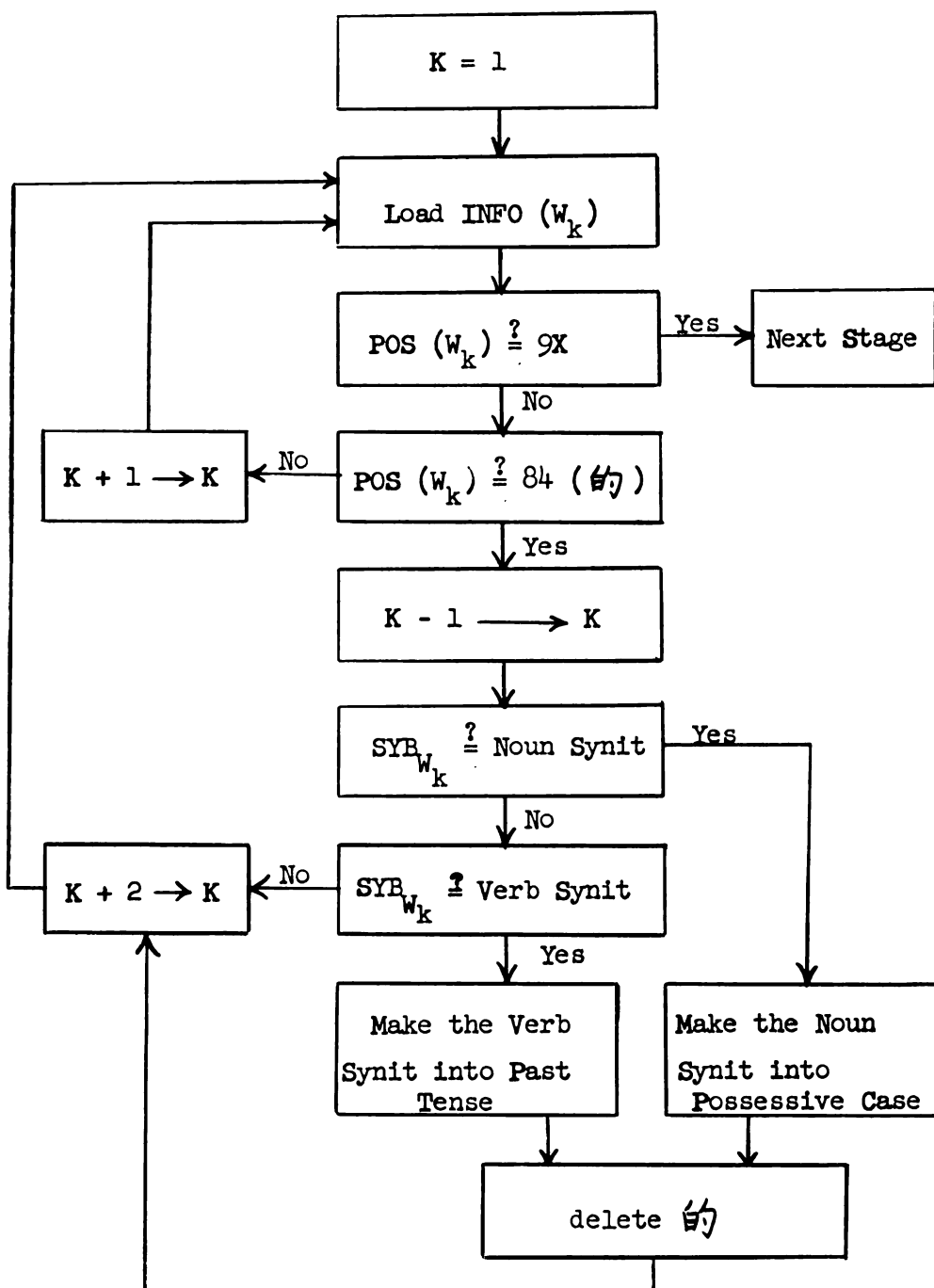


Figure 17 Algorithmic Chart for Processing Synit Sequences Formed by 的 Preceded by a Noun or Verb Synit

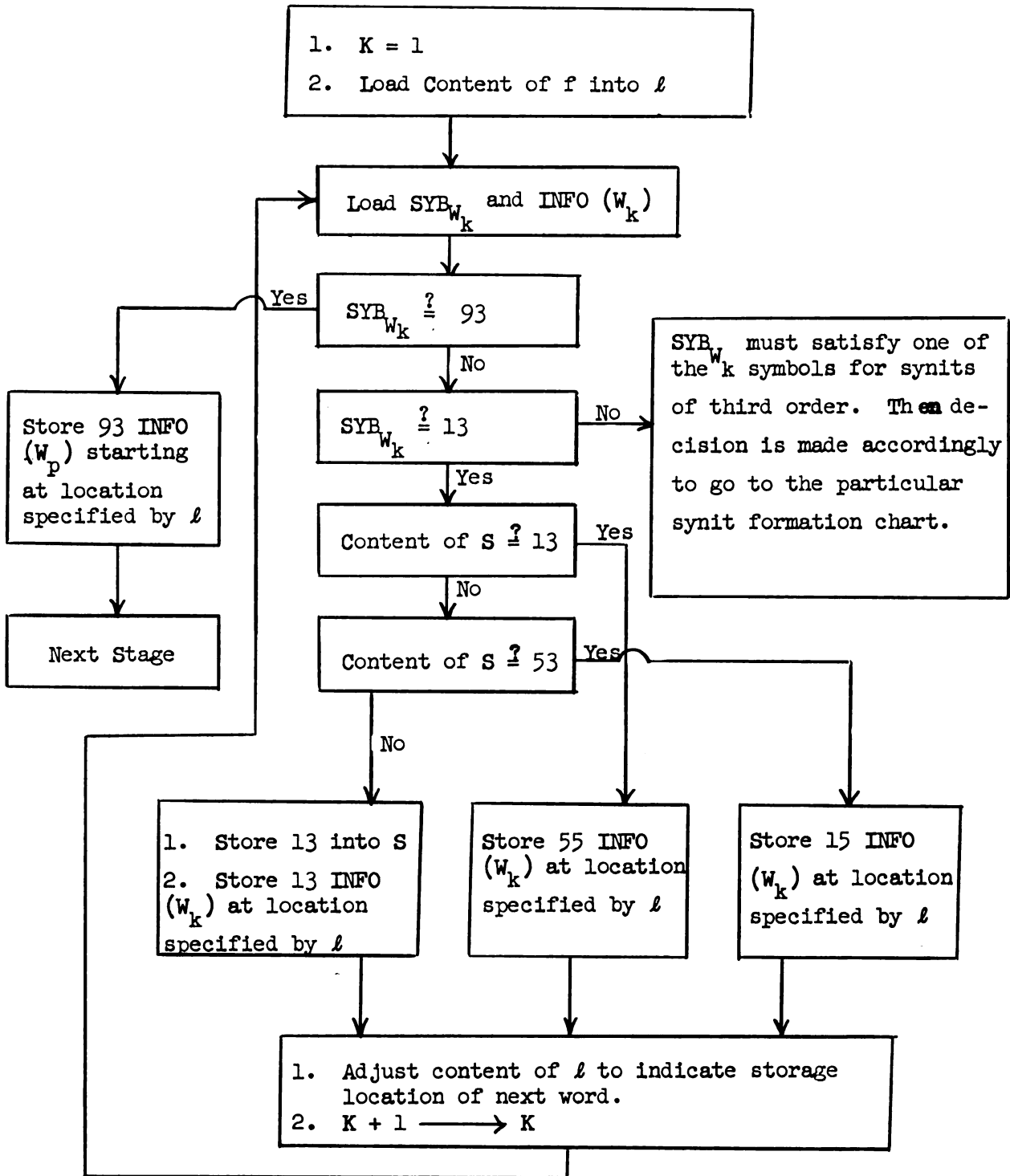


Figure 18 Algorithmic Chart for Re-formation of Noun Synit of Third Order

established rules were based on information of synits of second order. Otherwise, the procedures follow the same format. In Figure 18 the algorithmic chart is given with modifications made to deal with synits of third order only. The formation of the noun synit is again considered to exemplify the general procedure.

3f. Other Synit Sequences

It is obviously not possible to enumerate all the syntactic structures in the Chinese language which are different from English. Even if a certain sequence of synits is detected that effects an order change for synthesis into English, one may well find a similar sequence, in which the words used are different, which does not need such a change. It can be seen that in the synit sequences discussed in the preceding section the order change is made based on two factors; namely the presence of the word 的 and agreement on the syntactic function of the verb. Therefore, it is extremely difficult to establish general rules indicating only the type of synits in a sequence without any regard to specific words.

In view of the irregularities in word structures in Chinese only two significant synit sequences are found to be meaningful when their orders are changed. The first deals with a sequence initiated by a preposition synit. The second involves a statement of comparison. In both cases the syntactic function of certain specific words needs to be examined.

The formation of the prepositional sequence is

[preposition] [verb] [noun]*

Some illustrations are first given.

1. [向 你] [告] [別]
To you To bid Farewell

Translation: To bid farewell to you.

2. [給 我] [買] [的]
For me To buy Past-tenser

Translation: To buy past tenser for me. (or bought)

3. [從 家裏] [來]
From home To come

Translation: To come from home.

The synthesis involves rearranging these synits into [verb] [noun]*
 [preposition]. The flow chart for automatic processing is given in
 Figure 19. The preposition used for this particular function should
 be provided with sign in its potential syntactic function field. A list
 of some of the prepositions more frequently used in the prepositional
 sequences is shown on the next page.

* Presence is optional.

Table 7 A List of Frequently Used Prepositions
in Prepositional Synits

[向	...	] [verb] [noun]
[給	...	] [verb] [noun]
[從	...	] [verb] [noun]
[跟	...	] [verb] [noun]
[對	...	] [verb] [noun]
[被	...	] [verb] [noun]
[用	...	] [verb] [noun]

The statement of comparison usually uses certain comparative signs. The common ones are 更 and 最 . Thus "tall" (高), "taller" and "tallest" in Chinese are 高 , 更高 , and 最高 . However, when a comparison is made in which one member is expressive of inequality, the word 比 is used. 比 is placed before the member expressive of inequality and the measure involved after. Examples:

1. 我 比 他 高
I He Tall

Translation: I taller than he.

or I am taller than he.

2. 比 那 棵 樹 高 三 倍
That Classifier Tree Tall Three Times

Translation: Three times taller than that classifier tree.

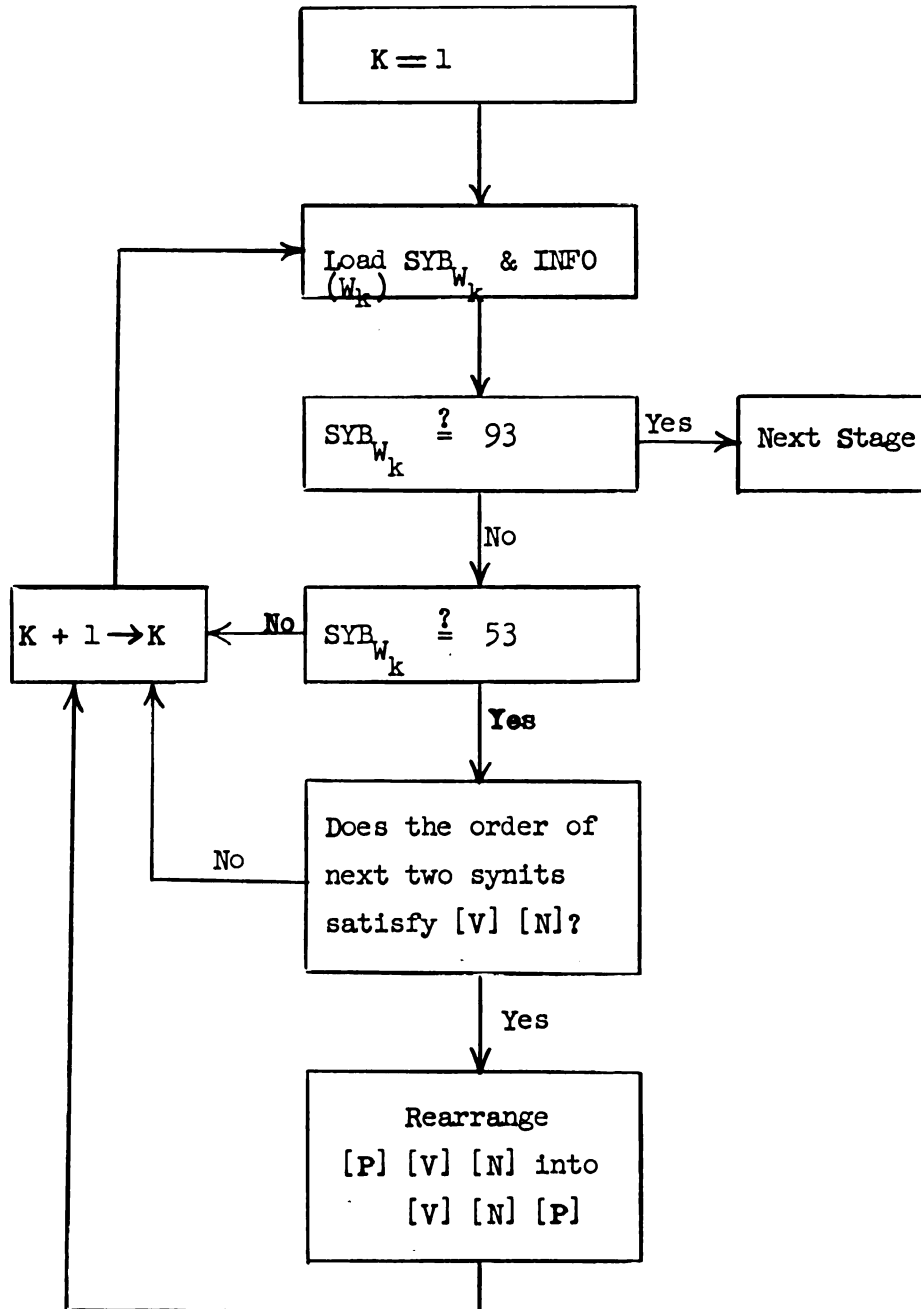


Figure 19 Algorithmic Chart for Translating a Chinese Phrase whose Synit Sequence is $[P] [V] [N]$.

The change of word order that is necessary to give correct translation into English involves the placement of the adjective of comparison in its comparative case and the quantity, if specified, before the member which is compared. In the case as illustrated in Example 1, the word 高 (tall), or in general the adjective of comparison, is made comparative (i.e. taller than) and placed before the element being compared, 他 (he). In a more complicated pattern as shown in Example 2, the additional part of the phrase, i.e. 三倍 (three times), is placed before the comparative adjective "taller than". The word 比 is deleted after word order is changed.

The sequence of synits is not directly responsible for recognition of statement of comparison. The key to detection comes from the syntactic functions of the words 比 and the adjective of comparison. But 比 is a special term and therefore, forms a synit by itself. The remaining words following 比 in the statement of comparison may form one noun synit. However, no ambiguity should arise since the syntactic function in adjective of comparison is first detected and the proper word order change is then made within the synit. The algorithmic charts for processing the statements of comparison are shown in Figures 20 and 21.

3g. Output of English Equivalent

The final stage of translation, that of the output of the English

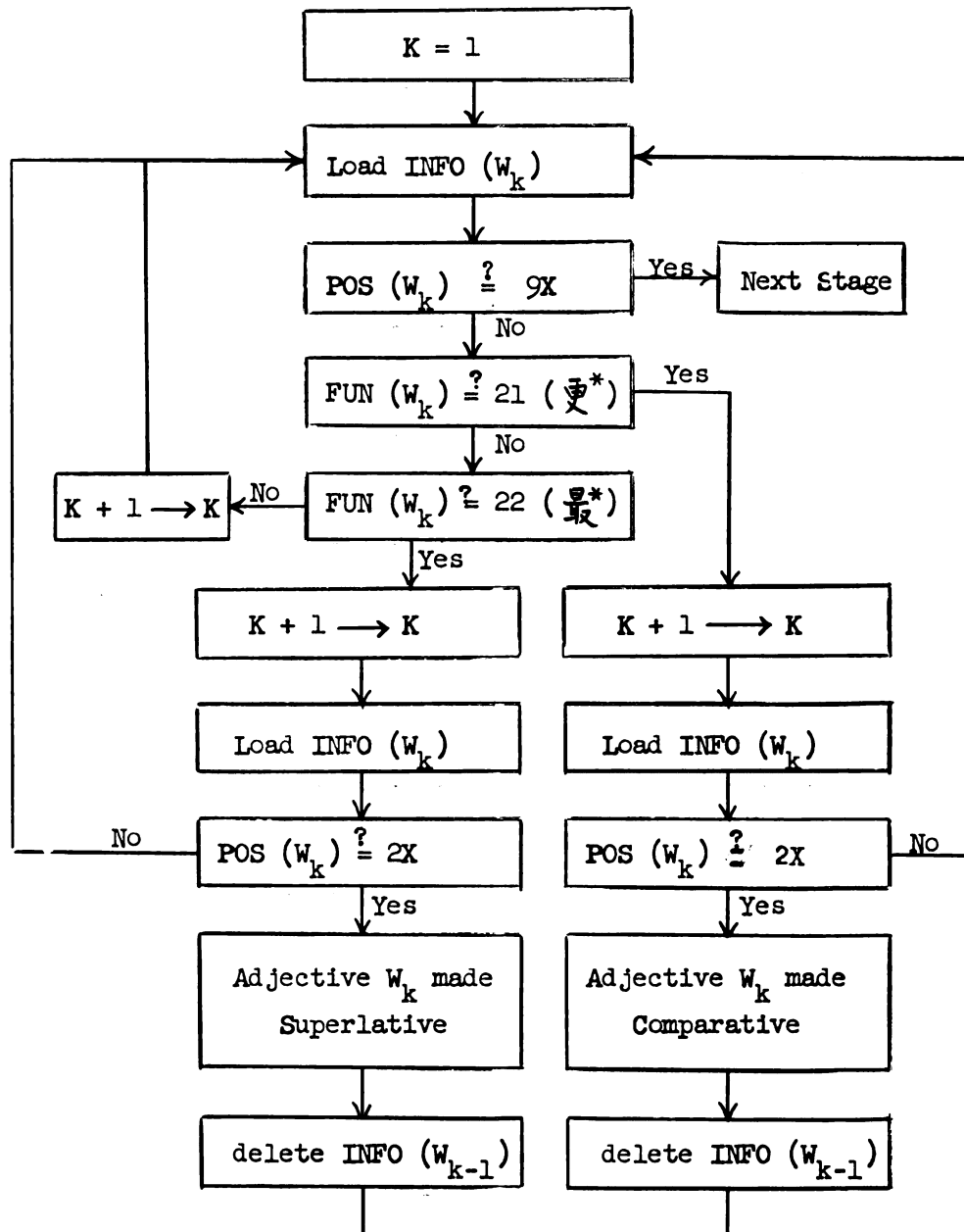


Figure 20 Algorithmic Chart for Processing Comparative and Superlative Words in Chinese

* For convenience, the syntactic functions of 褒 and 贬 in their roles of comparative and superlative cases are arbitrarily assigned as 21 and 22 respectively.

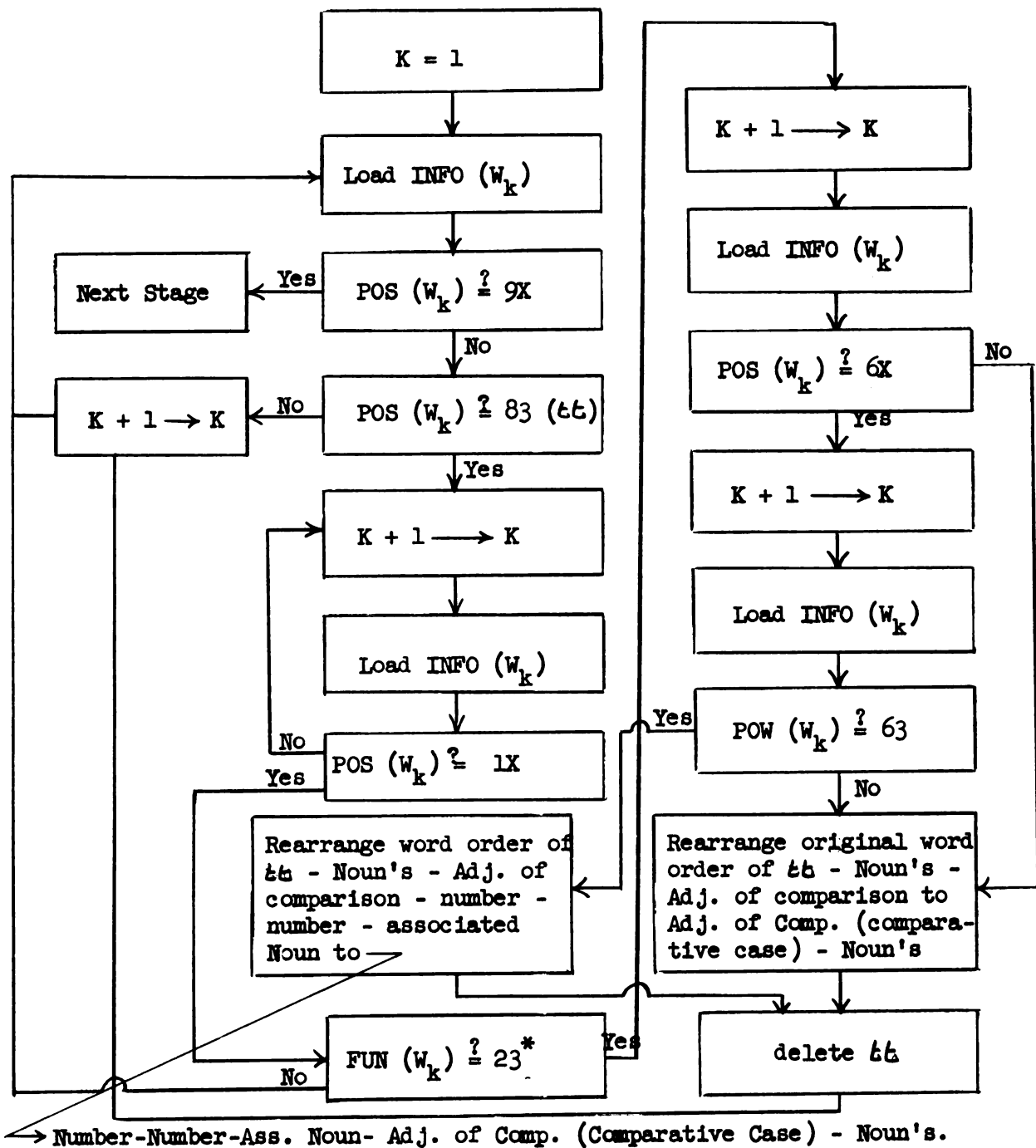


Figure 21 Algorithmic Chart for Processing Statement of Comparison

* The syntactic function of the adjective of comparison is arbitrarily assigned as 23.

counterpart of the Chinese passage, is a matching process since the locations of the English equivalents of the words can be found from their information boxes. Presumably the English words are stored in the memory, and it is only necessary to go to the particular location and bring out the word for an eventual print out. The other fields in the information box, the part-of-speech and potential syntactic function, no longer serve any purpose and can be deleted.

4. Conclusion

Although machine results are not available at this time, the syntax routines have been tested repeatedly by hand by following the algorithmic charts shown in this chapter. Some simulated results worked out by hand are shown in Appendix 6. The passages translated are selected so as to represent the various cases of grammatical patterns in the Chinese language.

Emphasis has been previously made on the fact that the translating system thus developed is designed to translate fairly elementary Chinese texts, as exemplified by the reference books used. There is no doubt that this system will encounter many difficulties when tried upon any text of more complex nature, both in grammar and word usage. However, it is believed that the schemes developed, particularly the formation of synits, are of considerable value toward future work of mechanical translation from Chinese to English.

APPENDIX 1
THE RADICAL LIST

	1	2	3	4	5	6	7	8	9
0	一	丨	丶	ノ	乙	丿	二	ニ	人 ₁
10	儿	入	八	冂	乚	几	凵	刀 _リ	力
20	夕	匕	匚	乚	卜	尸 _尸	厂	厶	又
30	口	口	土	士	又	夕	大	女	子
40	冂	寸	小	尢	屮	山	《 _川	工	己
50	巾	干	么	广	又	艹	弋	弓	乚 _彡
60	彳	心 _忄	戈	户	手 _扌	支	支 _攴	文	斗
70	方	无 _无	日	曰	月	木	欠	止	夕
80	母	比	毛	氏	气	水 _氵	火 _灬	爪	父
90	片	片	牙	牛 _牜	犬 _犭	玄	玉 _王	瓜	瓦
100	生	用	田	疋	疒	死	白	皮	目 _目
110	矛	矢	石	示 _礻	肉	禾	穴	立	竹 _竹
120	糸 _纟	缶	网 _罒	羊	羽	老	而	耒	耳
130	肉 _月	臣	自	至	臼	舌	舛	舟	艮
140	艸 _艹	虎	虫	血	行	衣 _衤	西 _西	見	角
150	谷	豆	豕	豸	貝	赤	走	足 _足	身
160	辛	辰	走 _走	邑 _邑	西	采	里	金	長 _長
170	阜 _阜	隶	佳	雨	青	非	面	革	韋
180	音	頁	風	飛	食	首	香	馬	骨
190	髟	鬥	鬯	鬲	鬼	魚	鳥	鹵	鹿
200	麻	黄	黍	黑	黹	冠	鼎	鼓	鼠
210	齊	齒	龍	龜	龠				鼻

APPENDIX 2

**A LIST OF 1400 COMMONLY USED
CHINESE CHARACTERS**

1	6	以	保	價	典	別	動	南	受	吩	喊	國	壞	奪	李
一	了	令	倍	儘	兼	利	務	博	叛	含	喚	團	33	38	孤
七	事	件	修	儲	13	到	勝	25	叢	吵	喜	32	士	奴	孩
三	7	任	個	10	冊	制	勞	卦	30	吸	喝	土	壯	如	孫
上	二	份	倍	元	再	則	募	26	口	吹	單	地	壽	妙	學
下	五	休	們	兄	冒	刻	勵	印	古	告	嗎	在	34	妹	40
不	互	估	們	充	14	前	勸	危	另	吾	喔	址	35	妻	守
且	况	你	倒	兆	冠	剛	20	却	只	呀	嘆	均	夏	始	安
世	些	伴	候	先	15	刺	勿	卷	叫	呢	嘉	坐	36	姐	完
丟	亞	似	借	光	冬	副	包	卻	召	周	嘖	坡	外	好	宗
並	8	但	倡	克	冰	割	21	卻	可	吧	嘗	城	多	委	官
2	交	佈	假	兒	冲	劍	化	27	台	味	嗽	域	夜	威	宜
中	亦	位	偉	11	決	19	北	厚	史	呵	嘴	執	夢	娘	客
3	京	低	做	16	凡	力	22	原	右	呼	嘿	基	37	娶	宣
主	亮	住	停	凡	17	功	23	厭	司	命	器	堂	大	婆	定
4	9	佛	健	17	出	加	匹	厲	各	和	頓	堅	太	婚	室
乃	人	作	側	兩	函	劣	區	28	合	唉	嚇	堆	夫	婦	害
久	什	佩	偶	12	18	助	24	去	吉	售	唸	堪	央	媽	宴
之	仁	使	傑	八	刀	努	十	29	吊	唱	嚴	報	失	嫁	家
乎	今	來	備	公	分	効	千	又	同	商	屬	場	夾	嫂	容
乘	介	例	傷	六	切	効	卅	及	名	問	31	塊	奇	嫁	宿
5	仍	侍	傾	共	列	勁	升	友	吐	啓	回	塞	奉	嫩	寄
九	他	供	僅	兵	初	勁	午	反	向	喂	四	墳	奏	39	密
也	付	便	像	其	判	勇	半	取	君	喃	因	境	奔	孔	寧
乾	代	俗	僕	具		勉	協		否	菩	圈	壁	套	字	富

寒	就	工	幸	57	微	恢	慾	托	掛	收	斤	晚	服	植	歡
察	<u>44</u> 尺	左	<u>52</u> 幻	引	德	恥	憂	扮	接	改	斯	晝	有	業	<u>77</u> 止
實	尼	巧	幼	弟	<u>61</u> 心	恨	憐	承	推	攻	新	晤	朗	極	正
審	尾	巨	幾	弱	必	恩	憤	抄	提	放	斷	晨	望	榮	此
寫	局	差	<u>53</u> 床	張	忍	恭	憶	把	換	政	<u>70</u> 方	普	朝	概	步
寬	屈	<u>49</u> 己	底	強	志	息	應	投	揮	故	於	景	<u>75</u> 木	樂	武
寶	居	巴	店	彈	忘	悲	懶	抗	援	效	旁	智	未	標	歲
<u>41</u> 寸	屈	巷	府	<u>58</u> 形	忘	悶	懷	折	搞	教	族	暖	本	模	歷
寺	屋	<u>50</u> 市	度	<u>59</u> 彩	忙	情	戀	拋	搖	敏	旗	暗	朵	樣	歸
封	展	布	康	影	忠	惠	<u>62</u> 我	拍	搗	救	<u>71</u> 既	暈	李	樹	<u>78</u> 死
射	屢	希	廚	<u>60</u> 徬	快	惡	或	拒	搶	敗	<u>72</u> 日	暫	材	機	<u>79</u> 段
將	層	師	廟	役	念	惱	戰	拖	摸	敢	旦	暴	村	檢	殺
專	<u>45</u> 山	席	廢	往	忽	意	戲	拚	撕	散	早	曉	東	權	毀
專	<u>46</u> 岩	帶	廣	很	怎	想	載	招	撞	敬	昌	更	林	<u>76</u> 欠	<u>80</u> 母
尋	岸	常	<u>54</u> 建	待	怒	愛	<u>63</u> 戶	拳	擊	敬	明	<u>73</u> 書	果	次	每
對	峯	帽	<u>55</u> 弄	律	怕	慈	房	拾	操	整	昏	曼	枝	欣	毒
導	島	幕	<u>56</u> 建	後	怖	態	所	拿	擔	敵	易	曾	架	歎	<u>81</u> 比
<u>42</u> 小	<u>47</u> 川	幣	<u>57</u> 弄	徐	恒	慌	<u>64</u> 手	指	據	數	星	替	某	歎	<u>82</u> 毛
少	州	幫	<u>58</u> 式	徑	思	慘	才	振	擠	<u>67</u> 文	春	最	查	歎	毫
尖	<u>48</u> 巡	<u>51</u> 平		得	性	慢	打	捉	擦	<u>68</u> 斗	昨	會	校	歎	
尚		年		從	怪	慣	扔	捕	擺	料	時	<u>74</u> 月	根	歌	
<u>43</u> 尤				復	忽	慨		捨	<u>65</u> 支	斜	晉	朋	案	歌	
					恐	慶		掃	<u>66</u>				條		

83	民	84	氣	85	水	永	求	汗	江	沒	沙	河	油	治	沿	泉	冷	泡	泥	注	泰	洋	洗	洲					
124	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
125	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
126	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
127	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
128	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
129	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
130	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
131	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
132	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
133	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
134	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
135	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
136	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
137	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
138	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
139	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝
140	翁	130	肉	133	自	140	花	若	苦	英	茫	茲	茶	荒	莊	莫	菜	華	萬	落	葉	華	蒙	蓋	蔬	薄	藉	藍	藝

藥	裁	訪	諭	152	賞	跳	近	邊	鈞	限	雜	178	183	鬍	麼
蘇	裂	許	誦	象	賄	身	返	163	銀	降	難	179	飛	191	201
141	補	設	諸	豬	賢	躲	追	那	銅	院	離	180	184	鬧	黃
虎	裝	訴	謀	153	賣	159	退	郎	鋒	陣	難	音	食	192	202
處	製	診	誑	貌	質	車	送	邵	錄	除	173	響	飲	193	黏
號	146	証	謎	貓	賺	執	逃	部	錢	陰	雨	181	飯	194	203
虧	西	註	講	154	購	軍	途	郵	錯	陳	雪	頁	餅	鬼	黑
142	要	評	謝	貞	賽	軟	造	都	鏡	陵	雲	頂	養	魔	點
蛋	147	試	謠	負	贈	載	通	鄉	鐘	陸	零	項	餐	195	黨
蜜	見	詩	謹	財	贊	輕	造	鄰	鐵	陽	雷	順	餘	魚	204
蟲	規	話	證	貢	155	輝	連	164	鑰	隊	電	須	館	鮮	205
143	親	該	識	貧	赤	輩	週	酒	168	隔	需	預	185	196	206
血	覺	誇	譯	貨	赫	輪	進	酸	長	險	霧	頓	首	鳥	207
衆	觀	認	議	貪	156	輸	遇	醒	169	隱	靈	頗	186	鷄	鼓
144	148	誓	護	買	走	轉	遊	醫	門	隨	174	領	香	197	208
行	角	語	讀	青	起	160	運	醫	開	171	青	頭	187	鹹	鼠
術	解	誠	變	貴	超	辦	過	165	開	隸	靜	頸	馬	鹽	209
街	149	誤	讓	買	越	辦	道	采	間	172	175	題	駕	198	鼻
衛	言	說	讚	費	趕	辦	達	166	關	佳	非	願	188	麗	210
衛	訂	誰	150	賀	趣	辦	遠	里	170	隻	靠	類	骨	麟	齊
衡	計	課	151	資	157	161	通	重	防	雅	176	顧	體	199	211
145	訊	調	豆	賊	足	農	遞	野	阻	集	面	顯	189	麥	212
衣	討	談	豎	賓	跑	162	選	量	阿	雖	177	182	高	麵	龍
表	訓	請	豐	賜	跟	迎	邀	167	金	雙	革	風	190	200	213
	記	論			路		還				鞋	飄	髮	麻	214

APPENDIX 3

A SHORT LIST OF CHINESE CHARACTERS COMPILED
ACCORDING TO THE RADICAL SYSTEM

The word is listed on the right side of the column. Its associated radical is indicated at the upper left corner of the first word of its list. The numerical equivalent for each word is arbitrarily assigned and appears at the left of the word.*

一 1000	一	1210	來	1420	到	1630	向	2040	天	2250	少	2460	成
1010	三	1220	倍	1430	前	1640	同	2050	太	2260	就	2470	我
1020	上	1230	個	力 1440	加	1650	含	2060	失	工 2270	工	2510	或
1030	下	1240	候	1450	功	1660	告	女 2070	女	已 2300	已	戶 2520	所
1040	不	1250	們	區 1460	區	1670	和	如 2100	如	巾 2310	常	手 2530	手
1050	中	1260	做	十 1470	十	1700	員	好 2110	好	干 2320	平	2540	把
1060	之	1270	像	1500	千	1710	問	始 2120	始	2330	年	2550	拿
1070	也	1300	人	1510	半	1720	喜	媽 2130	媽	度 2340	度	2560	探
1100	了	1310	先	1520	南	口 1730	回	子 2140	字	後 2350	後	2570	接
1110	事	1320	內	1530	原	1740	因	學 2150	學	很 2360	很	2600	改
1120	些	1330	全	1540	去	1750	國	它 2160	它	得 2370	得	2610	教
1130	以	1340	六	又 1550	又	土 2000	在	完 2170	完	快 2400	快	2620	敗
1140	他	1350	共	1560	及	1760	地	定 2200	定	怕 2410	怕	2630	數
1150	住	1360	再	1570	可	及 1770	夏	容 2210	容	愛 2420	愛	2640	文
1160	作	1370	冰	1600	右	夕 2010	外	家 2220	家	想 2430	想	2650	新
1170	你	1400	冷	1610	只	2020	多	寫 2230	寫	慶 2440	慶	2660	方
1200	但	1410	別	1620	吃	大 2030	大	小 2240	小	慢 2450	慢	2670	於

* Using the CDC 160-A Computer, it is conceived that dictionary information for each word takes ten octal memory positions. Hence, starting at location 1000, the digital equivalents are accordingly assigned.

日	母	每	生	給	覺	長	長	黑	黑
2700	3140	比	3410	3670	4150	4430	門	4700	黑
2710	66	氏	3420	3700	喜	門	4440	4710	。
2720	3150	氏	用	3710	4160	4450	開	3770	，
2730	3160	氣	3430	3720	4170	4460	險	4420	；
2740	3170	水	田	羽	4200	4470	雖	3510	：
2750	3200	沒	3440	3730	4210	4500	難	4720	（
2760	3210	洲	3450	老	4220	雨	零	4730	）
2770	3220	海	3460	3740	4230	4510	電	4740	[
3010	3230	深	3500	而	4240	4520	需	4750	]
3020	3240	測	3520	月	4250	4530	非	4760	！
3030	3250	溫	3530	臣	4260	4540	面	4770	！
3040	3260	無	3540	自	4270	4550	音		
3050	3270	然	3550	4010	4280	4560	須		
3060	3280	為	3560	4020	4290	4570	預		
3070	3290	爸	3570	4030	4300	4600	飯		
3080	3300	片	3600	4040	4310	4610	養		
3090	3310	物	3610	4050	4320	4620	驗		
3100	3320	獸	3620	4060	4330	4630	高		
3110	3330	玩	3630	4070	4340	4640	鹽		
3120	3340	現	3640	4080	4350	4650	麗		
3130	3350	甚	3650	4090	4360	4660	麼		
	3360		3660	4100	4370	4670			
	3370			4110	4380				
	3380			4120	4390				
	3390			4130	4400				
	3400			4140	4410				

APPENDIX 4

CLASSIFICATION OF PART-OF-SPEECH OF THE CHINESE WORD

The numerical code for each classification is assigned arbitrarily.*

<u>Nouns (1X)</u>	<u>Sample Words of the Classification</u>
Nouns (11)	教員, 家, 天氣
Pronouns (12)	你, 我們
Proper Nouns (13)	日本
Time-Associated Nouns (14)	年, 月, 星期
Adjectival Pronouns (15)	這, 那
Conjunctive Pronoun (16)	[WH], i.e. which, who, where
<u>Adjectives (2X)</u>	
Adjectives (21)	美麗, 普通
Adjectives of Comparison (22)	大, 高
<u>Adverbs (3X)</u>	
Adverbs (31)	不, 也, 只
Adverbs of Comparison (32)	太, 最
Adverbs of Condition (33)	容易, 勉強
<u>Verbs (4X)</u>	
Transitive Verbs (41)	吃, 拿
Intransitive Verbs (42)	開始, 想
Auxiliary Verbs (43)	會, 要
Verb-Prepositions (44)	在, 當
<u>Prepositions (5X)</u>	
Prepositions (51)	向
Time-Related Prepositions (52)	上, 下, 前, 後

* The two-digit number with X as second number denotes the general category, such as 4X for verb; while a numerical second number denotes the particular sub-field of part-of-speech, such as 42 for intransitive verb.

<u>Numbers (6X)</u>	<u>Sample Words of the Classification</u>
Numerical Numbers (61)	一, 三, 十
Pronominal Numbers (62)	一些, 許多
Number-Associated Nouns (Classifiers) (63)	個, 片, 度
<u>Conjunctions (7X)</u>	
Introductory (Subor.) Conjunctions (71)	如果, 但是, 所以
Coordinate Conjunctions (72)	和, 或
<u>Special Terms (8X)</u>	
Special Elements I (81)	有, 是
Special Elements II (82)	甚麼
比 (83)	
的 (84)	
Past Tenses (85)	了, 過
<u>Punctuation Marks (9X)</u>	
Period - .	
Comma - ,	
Colon - ;	
Semi-Colon - :	
Brackets - (), []	
Quotation Marks - "	
Exclamation Mark - !	

APPENDIX 5

LOOK-UP PROGRAM AND SAMPLE RESULTS
USING CDC 160-A COMPUTER

Master Routine

<u>Function Code</u>	<u>Execution Address</u>	<u>Operand</u>	<u>Comments</u>
JPR		Subroutine I	Input dictionary in- formation
LDD	75		
STM		1000	First Storage location = 1000
SBN	77		End of list check
ZJF	03		
AOB	03		
NZB	10		
JPR		Subroutine I	Input word to be looked up
LDD	75		
STM		0105	Temporary Storage 0105
SBN	77		End of list check
NZF	02		
Go to Next Stage			
JPR		Subroutine I	Input following word for idiom search
LDD	75		
STM		0106	Temporary Storage 0106
SBN	77		End of list check
NZF	05		
LDN	03		
RAM		0105	
NZF	33		
LDM		0105	

<u>Function Code</u>	<u>Execution Address</u>	<u>Operand</u>	<u>Comments</u>
STF	02		
LDM		****	
SBM		0106	Check for first idiom in dictionary
ZJF	12		
LDN	03		
RAM		0105	
STF	02		
LDM		****	
SBM		0106	Check for second idiom in dictionary
NZF	12		
LCN	02		Output idiom information
STD	77		
AOM		0105	
JPR		Subroutine II	
AOD	77		
NZB	05		
ZJB	54		
LDN	03		Output word information
RAM		0105	
JPR		Subroutine II	
AOM		0105	
JPR		Subroutine II	
LDM		0106	
STM		0105	
SEN	77		
NZB	63		
Go to Next Stage			

Subroutine I

<u>Function Code</u>	<u>Execution Address</u>	<u>Operand</u>	<u>Comments</u>
JPI ****	01		This routine inputs 4 octal digits
EXC		4102	
INP	03		and stores in
PJF	02		temporary location 0075
0101			
LDM		0101	
LS3			
SCM		0102	
LS3			
SCM		0103	
LS3			
SCM		0104	
STD	75		
PJB	24		
NJB	25		

Subroutine II

<u>Function Code</u>	<u>Execution Address</u>	<u>Operand</u>	<u>Comments</u>
JFI ****	01		This routine prints out 4 octal digits
LDM		0105	Stored in location
STF	02		specified by the content
LDM		****	of 0105
STD	76		
EXC		4104	
LDD	76		
LPC		7700	

<u>Function Code</u>	<u>Execution Address</u>	<u>Operand</u>	<u>Comments</u>
LS6			
OTA			
LDD	76		
LPN	77		
OTA			
PJB	22		
NJB	23		

For the given Chinese passage:

我不知道這個學期在甚麼時候終止。

Input to Machine:

2470 1040 3550 4340 4320 1230 2150 3030 2000 3400 4670 2750
1240 3660 3110 4710 0077

Output:

00	12	2420
00	31	3410
00	41	2700
00	15	6110
00	63	6400
00	12	5020
11	44	0150
12	82	7030
13	11	6010
00	42	6120
00	91	6600

APPENDIX 6

SIMULATED TRANSLATION OF CHINESE PASSAGES
INTO ENGLISH

The following three passages in Chinese are to be translated into English. The correct English translation is first presented beneath each passage. The translation using the Syntax System developed in Chapter 2 are then carried out in ensuing sections. Each string of words ending with a punctuation mark is first written, followed by the word-for-word translation. Details are stressed on processing the significant syntactic relationships existing in the passages. The trivial parts of the translating process are not indicated.

- I. 我不知道這個學期在甚麼時候終止，也不知道我們的大考在甚麼時候開始。我只知道教員已經說過，"在大考開始以前，我們要把舊的功課溫習好。"

Translation:

I do not know when does this semester end, also do not know when does our final exam begin. I only know that the teacher already said, "Before final examination begins, we must review well the old lessons."

- II. 三年前我到你家向你告別，現在我回來了。

Translation:

Three years ago I went to your home to bid farewell to you, now I returned.

- III. 當天氣太冷的時候，海水都會結冰。但是在南極洲的一個地方，有一片死海；當天氣冷到零下六十度的時候，這片海也不會結冰。去探險的科學家測驗結果，發現它含的鹽是比普通海水多十一倍。所以它不容易結冰。

Translation:

When the weather is very cold, even sea water can freeze. But at one place in the Anarctic Continent, there is a dead sea; when the temperature reaches sixty degrees below zero, this sea does not freeze. The scientists who went there to explore for experimental results, found it contained salt eleven times more than ordinary sea water. Therefore it does not freeze easily.

- I. 1. 我 不 知道 這 個 學期 在 甚麼 時候 終止
I do not, not to know this classifier semester at what time to end

Deletion of word:

Delete 個 (classifier) after 這 (this)

Formation of Synits of first order:

(I) (do not, not to know) (this semester) (at) (what) (time)
(to end)

Formation of Synits of third order:

[I] [do not, not to know] [this semester] [at] [what] [time]
[to end]

Translation:

I do not, not to know this semester at what time to end
do not

2. 也 不 知道 我們 的 大考 在 甚麼 時候 開始
Also do not, not to know we 的 final exam at what time to
begin

Formation of Synits:

[Also do not, not to know] [we] [的] [final exam] [at]
[what] [time] [to begin]

Sequence of synits:

[we] [的] → [we] in possessive case → [our]

Reformation of Synits:

[Also do not, not to know] [our final exam] [at] [what]
[time] [to begin]

Translation:

Also do not, not to know our final exam at what time to begin.
do not

3. 我 只 知道 教員 已經 說 過
I only to know teacher already to say past-tenser

Processing past-tense:

to say past-tenser → said

Formation of Synits:

[I] [only to know] [teacher] [already said]

Translation:

I only to know teacher already said

4. 在 大考 開始 以前
to be at final examination to begin before

Frame Construction:

to be at ... before → before

Formation of Synits

[before final examination] [to begin]

The sequence of synit above is [P] [V], but the syntactic function of the word "before" as found by frame construction will not be so assigned to effect a change in the manner discussed in Section 3f. of Chapter 2. Otherwise translation is incorrect.

Translation:

before final examination ~~to~~ begins

5. 我們要把舊的功課溫習好
We must with old 的 lesson(s) to review well

Deletion of word:

的 after 舊, an adjective

Formation of Synit:

[we] [must] [with old lesson(s)] [to review well]

Sequence of Synit:

[P] [V] [N] (optional) to [V] [N] (optional) [P]
[to review well] [with old lesson(s)]

Translation:

We must ~~to~~ review well ~~with~~ old lesson(s).

II. 1. 三 年 前 我 到 你 家 向 你 告 別
Three year(s) ago I go to, to your home to you to bid farewell

Formation of Synits:

[three year(s) ago I] [go to,to] [your home] [to you]
[to bid] [farewell]

Sequence of Synits:

[P] [V] [N] to [V] [N] [P]
[to you] [to bid] [farewell] [to bid] [farewell] [to you]

Translation:

Three year(s) ago I go to, to your home to bid farewell to you.
went

2. 現在 我 回來 了
Now I to return 了

Processing Past-Tensor

to return } $\longrightarrow$ returned

Synit Formation:

[Now] [I] [returned]

Translation:

Now I returned

- III. 1. 當 天氣 太 冷 的 時候
to be at weather very cold 的 time

Frame Construction:

to be at ... time → when

Deletion of Word:

Delete word 的 after 冷, as adjective

Formation of Synits:

[When] [weather] [very cold]

Translation:

When weather -is- very cold.

2. 海 水 都 會 結冰
Sea water even can to freeze

Formation of Synits:

[Sea water] [even can to freeze]

Translation:

Sea water even can to freeze.

3. 但是 在 南極 洲 的 一 個 地方
But to be in Anarctic continent 的 one classifier place

Formation of Synits:

[But] [to be in] [Anarctic continent] [的]

[one classifier place]

Sequence of Synits:

[verb-preposition] [noun]₁ [的] [noun]₂ → [noun]₂

[WH] [verb preposition] [noun]₁

[one classifier place] [WH] [to be in] [Anarctic continent]

Translation:

But one classifier place ~~[WH]~~ to be in Anarctic continent.
which is in

4. 有 一 片 死 海
There to be one classifier dead sea

Formation of Synits:

[~~there to be~~] [one classifier dead sea]

Translation:

There to be one classifier dead sea.
is

5. 當 天氣 冷 到 零 下 六十 度 的 時候
to be at weather cold to zero below sixty degree(s) 的 time

Frame Construction:

to be at ... time → when

Formation of Synits:

[when] [weather] [cold] [to zero] [below sixty degree(s)]
[的]

Translation:

When weather -is- cold to zero below sixty degrees ~~的~~

6. 這 片 海 也 不 會 結 冰
This classifier sea also do not, not can to freeze.

Deletion of Word:

Delete 片 (classifier) after 這 (this)

Formation of Synits:

[this sea] [also do not,not can to freeze]

Translation:

This sea also ~~do not,not~~ can to freeze
not

7. 去 探險 的 科學 家 測驗
to go to,to to explore 的 science expert(s) to experiment
結果
result

Formation of Synits:

[to go to,to to explore] [的] [science expert(s)]
[to experiment] [result]

Sequence of Synits:

[verb-preposition] [的] [noun] → [noun] [WH] [verb-prep.]
[science expert(s)] [WH] [to go to,to to explore]

Translation:

Science expert(s) [WH] ~~to go to,to~~ to explore to experiment
who went
result

8. 發現 它 含 的 鹽 是 比 普通 海
To find it to contain 的 salt to be 比 ordinary sea
水 多 十一 倍
water more eleven times

Formation of Synits:

[to find] [it] [to contain] [的] [salt] [to be] [比]
[ordinary sea water] [more] [eleven times]

Sequences of Synits:

1. [verb] [的] → verb-past tenser

[to contain] [的] → contained

2. Statement of comparison:

[eleven times] [more than] [ordinary sea water]

Translation:

~~To find~~ it contained salt to be eleven times more than
Found
ordinary sea water.

9. 所以 它 不 容易 結冰
Therefore it do not, not easily to freeze

Formation of Synit:

[Therefore] [it] [do not, not] easily to freeze]

Translation:

Therefore it ~~do not, not~~ easily ~~to freeze~~.
does not

REFERENCES

1. Current Research and Development in Scientific Documentation No. 11.
Chapter 3, "Mechanical Translation", pp. 190-228, National Science Foundation, November, 1962.
2. "A General Compiler Capable of Learning", Richard Fairbanks Arnold,
A Thesis, Master of Science, Michigan State University, 1958.
3. Shau Wing Chan, "Chinese Reader for Beginners", Stanford University
Press, Stanford University, California, October, 1943.
4. Yuen Ren Chao, "Character Text for Mandarin Primer", Harvard University
Press, Cambridge, Massachusetts, 1954.
5. Tu Hsui-Chih (杜學知), China Weekly, pp. 10-12, 漢字排檢與漢學索引,
July, 1962.
6. Yehoshua Bar-Hillel, "Machine Translation of Languages", Edited by
William N. Locke and A. Donald Booth, pp. 183-193, The Technology
Press of the Massachusetts Institute of Technology and John Wiley
and Sons, Inc., New York, October, 1957.
7. Paul L. Garvin, Proceedings of the National Symposium on Machine
Translation", H. P. Edmundson, Editor, pp. 286-292, Prentice-Hall,
Inc., Englewood Cliffs, New Jersey, 1961.
8. Anthony G. Oettinger, "Automatic Language Translation", Harvard
University Press, Cambridge, Massachusetts, 1960.

ROOM USE ONLY

ROOM USE ONLY

MICHIGAN STATE UNIV. LIBRARIES



31293104746676