

FEB 1 1965

Reference book not to be
taken from the library.



Technical Note

254

ON THE PICTORIAL STRUCTURE OF CHINESE CHARACTERS

B. KIRK RANKIN, III,
WALTER A. SILLARS, AND ROBERT W. HSU



U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; cooperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business, and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information. The scope of the Bureau's activities is suggested in the following listing of its four Institutes and their organizational units.

Institute for Basic Standards. Electricity. Metrology. Heat. Radiation Physics. Mechanics. Applied Mathematics. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radio Standards Laboratory: Radio Standards Physics; Radio Standards Engineering.** Office of Standard Reference Data.

Institute for Materials Research. Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.** Office of Standard Reference Materials.

Central Radio Propagation Laboratory.** Ionosphere Research and Propagation. Troposphere and Space Telecommunications. Radio Systems. Upper Atmosphere and Space Physics.

Institute for Applied Technology. Textiles and Apparel Technology Center. Building Research. Industrial Equipment. Information Technology. Performance Test Development. Instrumentation. Transport Systems. Office of Technical Services. Office of Weights and Measures. Office of Engineering Standards. Office of Industrial Services.

* NBS Group, Joint Institute for Laboratory Astrophysics at the University of Colorado.

** Located at Boulder, Colorado.

NATIONAL BUREAU OF STANDARDS

Technical Note 254

ISSUED JANUARY 4, 1965

ON THE PICTORIAL STRUCTURE OF CHINESE CHARACTERS

B. Kirk Rankin, III,
Walter A. Sillars, and Robert W. Hsu

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

On the Pictorial Structure of Chinese Characters

by

B. Kirk Rankin III

Walter A. Sillars

Robert W. Hsu

A grammar of radical combination in Chinese characters has been written. From a sample study, it appears that this grammar is powerful enough to generate 80% of the characters in one of the standard dictionaries. Research leading to the construction of this grammar is embedded in a larger framework for describing the pictorial structure of the characters in detail. The descriptive framework defines five areas of study within the overall study of the characters: radical combination, radical variation, stroke combination, stroke variation, and distinctive features of strokes.

A. Extended Notions of Grammar and Language

Modern linguistics is concerned with the systematic study of (natural) languages. It typically involves the discovery and analysis of certain regularities of various kinds, and the expression of these (and of their interrelationships) in the form of grammars for those languages. Some types of regularities studied by linguistics are the relations among sentences; the constituent structure of sentences in terms of words; word-formation; the constituent structure of words in terms of sound units; and the specification of sound units in terms of sound features. ^{1/}

1/

Those interested in detailed accounts of modern linguistics might consult [1] and [2].

It is our contention that sets of data other than natural languages show many of the same features and regularities as those that are found in natural languages. Some very preliminary work on the analysis of chemical structure diagrams and electronic circuit diagrams has indicated that at least the linguistic notion of constituent structure is quite valuable in describing rather complex diagrams of both types. One other set of data which shares with these two types of diagrams both their "two dimensional" nature and their tractability to linguistic description is the set of Chinese characters. The linguistic analysis of this set is our concern in this report. ^{1/}

To summarize: We view the set of characters as a language. This language shows many regular features similar to features found in natural languages. Accustomed to constructing grammars to exploit and explain these regularities in natural languages, we construct a grammar to account for corresponding phenomena in the language of Chinese characters.

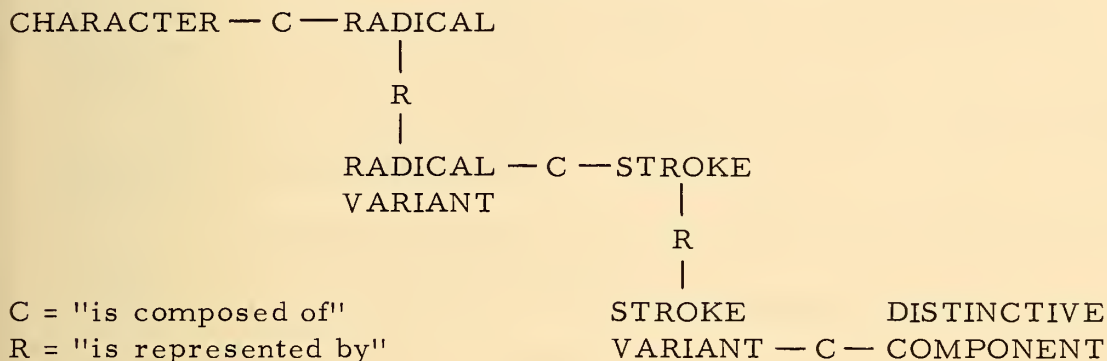
B. The Descriptive Framework

What follows is a general outline of a framework for studying Chinese characters from a linguistic point of view. The framework or model is inferred from two sources: Hockett [3] and Lamb [4]. We

1/

Let it be understood that the set of characters we describe is more or less abstract. What we describe is the manual output of a hypothetical native writer of Chinese as he prints characters. We do not describe either cursive handwriting or actual printed specimens from some font.

are not committed to this particular framework of description. ^{1/} It simply offers us what we want: a convenient framework for presenting our research results; in the long run, it may turn out to be inadequate or in some sense undesirable. The model imposes the following stair-way structure on our language of Chinese characters:



C. Characters are Composed of Radicals

Ours is not, by any means, the only way of describing the structure of Chinese characters in terms of radical-combination. The system in most common use, by both Chinese scholars and Western sinologists, is that of classification according to radical and phonetic. ^{2/}

1/

And we specifically do not pretend to speak officially for the "Stratificational" school of linguistics, which appears to take as basic to their theory the two sources just cited.

2/

The traditional approach described here is essentially a means of getting into a dictionary. It does not pretend to be a detailed study of character structure such as we are offering.

Traditional Chinese scholarship recognizes 214 distinguished characters called radicals, some of which have variant forms. (See the list displayed in Appendix A.) Each character is either itself a radical or may be decomposed into a radical plus a residue which is generally called the phonetic.

The ideal to which the traditional system aspires is a quite simple system. When a character is decomposed into radical and phonetic, the radical gives a clue to the meaning of the character and the phonetic a clue to its pronunciation. The traditional algorithm which identifies the radical in the majority of cases is as follows:

If the character naturally splits into a left side and right side, the left side is the radical; otherwise, if the character naturally splits into a top and a bottom, then the bottom is the radical.

However, there are many characters for which this procedure does not give the correct radical. These characters may be divided into two classes:

1. Those to which the algorithm does not apply, since the character does not have either a natural left-right or top-bottom split, for example, 開, whose radical is 門, and
2. Those for which the algorithm determines a radical other than the correct one, for example, 相, whose radical is 目.

A more elaborate algorithm for identifying the radical is described in Chao [5], Mathews [6], and many other sources, including nearly every Chinese-English dictionary. It runs somewhat as follows:

1. If the character is itself a radical, then it is its own radical.

2. If the character has a natural left-right split and
 - a. If the right side is one of the small number of radicals which can occur on the right, then the right side is the radical; otherwise,
 - b. If the left side is a radical, then the left side is the radical; if not, use the "look-up" procedure in step 5.
3. If neither 1 nor 2 applies, and if the character has a natural top-bottom split and
 - a. If the top is one of the small number of radicals which can occur on the top, then the top is the radical; otherwise,
 - b. If the bottom is a radical, then the bottom is the radical; if not, use the "look-up" procedure in step 5.
4. If none of the above apply, use the procedure for determining the radical of a character whose natural split is into a border and an inside, or which is formed by the superimposition of two or more parts. If this procedure determines the radical, the job is finished; if not,
5. Consult a table of characters whose radicals are difficult to determine by inspection.

There seem to be at least two obvious weak points in this algorithm, namely:

1. The procedure to be followed in step 4 is not explicit; for example, the radical of 開 is 門, while that of 問 is 口

When we consider in addition to these the fact that the list of difficult

characters $\frac{1}{2}$ in Mathews includes over 10% of the total number of characters in this dictionary, it is easy to see why the traditional analysis of characters is cumbersome to use when searching a dictionary for an unfamiliar character.

We feel that the procedure of assigning to each character an associated character chosen from a list of distinguished characters is a useful tool for classifying the characters; however, we also feel that the defects in the traditional classification scheme make that particular system unfeasible for mechanical recognition purposes and not particularly desirable for other purposes. Therefore, we have in Appendix B a list which, when finished, is intended to replace in our work the traditional list of 214 radicals. This list has been arrived at by observing combinations of strokes which recur often enough to prompt us to recognize them as "linguistic" units.

We have written a grammar similar to Chomsky's phrase-structure grammars $\frac{2}{2}$ which seems to be capable of generating approximately 80% of the characters in Mathews [6]. This grammar is exhibited in Appendix B. The manner in which the grammar operates, along with some of the difficulties we have encountered, is illustrated by a small-scale grammar DZ4 given as follows:

1/

That is those characters whose radicals are not obvious.

2/

See Chapter 4 in [7].

The grammar DZ4 consists of a finite number of rewrite rules. These rules are grouped into two classes, syntactic and lexical. The syntactic rules have the form $A = X$ or $A = \alpha (X, Y)$ where A , X and Y are nodes of the grammar and α is one of the three symbols h , v , or s . The lexical rules have the form $A = R$ where A is a node and R is a radical. For our purposes a node is any symbol which occurs as a constituent on the left side of some rule (either syntactic or lexical), and a radical is any symbol which occurs on the right side of a lexical rule. There are no symbols which are both nodes and radicals. The syntactic rules correspond to those rules in a phrase-structure grammar which generate from the top node down to the immediate sub-terminal level; the lexical rules correspond to those which generate the terminal symbols.

We use the customary left-to-right expansion convention, except that all syntactic rules are applied before any lexical rules. The rules of the grammar DZ4 are as follows:

Syntactic Rules

CHAR = COMP; IT 1/

COMP = $v (N, S)$; $h (W, E)$; $s (O, CHAR)$

N = COMP; NT

S = COMP; ST

W = COMP; WT

E = COMP; ET

O = OT

1/ The semicolon separates optional subrules; this line is really an abbreviation for the two rules: CHAR = COMP and CHAR = IT.

Lexical Rules

	WT	ET	NT	ST	OT	IT
十			x	x		x
亅	x					
二						x
口 ^{1/}	x	x	x	x		x
土	x	x	x	x		x
厶			x			
亼			x			
口 ^{1/}					x	
心				x		x
言	x	x				x
門		x		x		x

1/

Of course these two radicals may, in the long run, prove to be variants of a single radical. However, that fact need not prevent us from using them in a sample grammar as two different radicals.

CHAR (= Character) is the initial or top node of the grammar. CHAR can be rewritten as COMP, which eventually gives a complex (i. e., multi-radical) character. It can also be rewritten as IT, a symbol which heads a column in the lexical rules, the column consisting of terminal symbols which can occur in isolation. COMP can be rewritten as a vertical array having top N (North) and bottom S (South), a horizontal array having left W (West) and right E (East), or a surrounding array having border O (Outside) and inside CHAR (any character). N can be rewritten either as COMP (note that this option introduces a potentially infinite loop into the grammar) or as NT, the heading of a column of terminal symbols in the lexical rules which can occur on the top of a vertical array. The situation is similar for the remaining rules except that the border of a surrounding display cannot consist of a complex character.

Some examples of characters which DZ4 can generate are 門 , 個 , 言 , 們 , and 唯 . Although not all of these are in fact in use, they are all well-formed and acceptable to informants.

The steps in the generation of 一個 are:

Syntactic Steps

CHAR

COMP

h (W, E)

h (WT, E)

h (WT, COMP)

h (WT, s (O, CHAR))

h (WT, s (OT, CHAR))

h (WT, s (OT, COMP))

h (WT, s (OT, v (N, S)))

h (WT, s (OT, v (NT, S)))

h (WT, s (OT, v (NT, ST)))

Lexical Steps

h (一 , s (OT, v (NT, ST)))

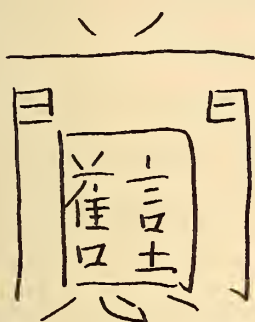
h (一 , s (口 , v (NT, ST)))

h (一 , s (口 , v (+ , ST)))

h (一 , s (口 , v (+ , 口)))

We will now discuss some currently unresolved problems concerning our grammar, specifically as they relate to the grammar displayed in Appendix B although most of our examples are chosen from DZ4.

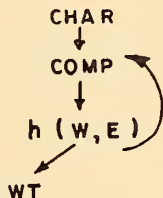
First is the question of recursion. Recall that in DZ4 there is a potentially infinite loop through the node COMP. This has the effect of asserting that there is no upper bound on the size of the characters, so that



is just as well-formed as 門, 言, etc.

We are not satisfied with this feature of the grammar, but neither are we pleased with the thought of putting an arbitrary upper bound on the number of recursions permitted.

The second problem is actually an extreme case of the first. There is nothing in the grammar to prevent an arbitrarily long recursion process which generates exclusively horizontal or exclusively vertical strings. That is, the grammar contains sub-grammars like



which give rise to terminal arrays like 口佳口土土仕言門 which does not satisfy our intuition as to what ought to constitute a well-formed character. Our tentative conclusion is that the grammar should have a device associated with it that would force an alternation between horizontal and vertical rules. For those relatively few characters which allow more than one successive application of a horizontal (or vertical rule) a list would have to be constructed.

Third, the grammars (both the one in Appendix B and DZ4) generate some characters in more than one way. In the appended

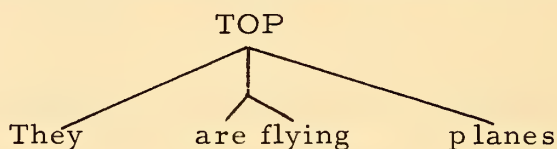
grammar, for example, 謝 is generated both as a combination of 言 and 寸 and as a combination of 言 and 射 .

There is strong motivation for multiple generations in a grammar for generating sentences of a natural language. In such cases we say that the given sentence is syntactically ambiguous. There will, in general, be one semantic interpretation for each syntactic parse of a sentence. To illustrate, consider the following two possible derivations of the sentence

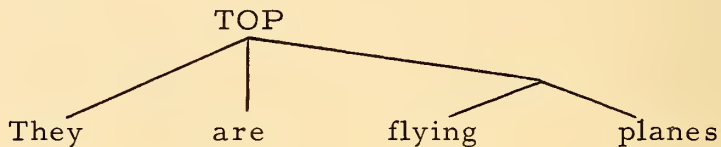
They are flying planes

in some phrase-structure grammar for a subset of English:

1.



2.



However, we are far from certain as to what, if anything, is the analog to diverse semantic interpretations in syntactically ambiguous characters.

At this point we also call the reader's attention to two features of our grammar which we consider to be defects and to indicate how they will be overcome in a future version. These features are

- (a) The apparently arbitrary assignment of certain radicals to certain geometric positions, for example, the assignment of 广 to the north, and
- (b) The listing as single terminals of characters whose immediate constituent structure cannot be expressed in terms of the grammar, for example 東 .

Consider the radical 宀, for example. Within the limits of the types of geometric structures handled by our grammar, there are three possible assignments of 宀 to a terminal class, namely, north, west, and outside. We (somewhat, but not entirely arbitrarily) assign 宀 to the north, 廴 to the south, and, in general, when a radical includes a vertical and a horizontal side we assign it to the horizontal. Besides having the advantage of making a definite decision, this is probably a more natural choice than either of the others, since it seems that the class of radicals which can occur with, say, 宀, has more in common with those that can occur with radicals which beyond doubt belong to the north terminal class, such as 宀, than it has with, say, those which can occur inside 口. For similar reasons, 門 is treated as an outside rather than as two separate radicals, one west terminal and one east terminal.

A still better solution, which is in the process of being incorporated into the grammar, consists of an increase in the number of basic operations and a refinement of the co-occurrence classes. In the next version of the grammar we will consider 問, 這, 因 and 麼 to be generated by four different binary operations, each having its own distinct set of co-occurrence restrictions.

The fact that we do not partition 東 into those constituents, namely, 木 and 日, of which it is traditionally held to be composed, is explained by the fact that the grammar does not have a superimposition operation. However, this is not so much a justification of our choice as an admission that the grammar is incomplete. An examination of the 300 characters in Wang [8] indicates that up to 10% of these would be more naturally treated by means of a superimposition operation than by means of the non-overlapping operations exclusively. To incorporate this operation will require detailed study of

"superimposed" characters with a view to defining in detail new co-occurrence classes.

In addition to these, we wish to mention two further questions which arise in the consideration of grammars for Chinese characters. These questions are of a more general nature than those just considered in that they are not tied to any particular grammar or even to any particular model of language.

The first of these is the question of well-formedness in general. That is, what is the class of characters which we want our grammar to generate? We clearly would not be satisfied with a grammar which generated all the characters in Mathews [6] and no more, for it would have no predictive power and thus would not be able to account for those characters which Mathews omits, those coined since its publication and those which will be coined in the future.

This is a perplexing problem but one which has been encountered previously by linguists working with (spoken) natural languages. We seem to be driven to the classic field technique of informant response. We are, in other words, forced to rely on the intuition of a native writer of Chinese as to what does and what does not constitute a well-formed character. Thus, on the analogy of 他, 她, and 付, we construct 姤 -- which does not occur in Mathews -- and offer it to our informant for his evaluation. The informant accepts it (in this case) as being well-formed, i. e., a reasonable combination of radicals, and therefore the power to generate 姤 is added to our minimal requirements for an acceptable grammar.

The second problem is that of partitioning the set of radicals into classes in such a way that all members of a given class play a similar role in the grammar. In other words, it is the problem of finding an operational definition of radical. We decide whether a group of strokes

is one radical or should be cut into two or more subgroups by the following criterion. We accept a proposed division of a group of strokes into two or more subgroups if and only if for each of the smaller groups resulting from the proposed cut there are many combinations of strokes which can replace it without requiring any change in the remainder of the original group.

For example, we consider 第 to be composed of two radicals (个 and 弟) rather than three since our intuition rebels at the thought of changing one part of 个个 without changing the other. Also we would not divide 言 into 讠 and 口 since, although there are many combinations which can occur over 口, there are few if any other than 口 which can occur under 讠.

However, it must be admitted that not all of the terminals in the grammar in Appendix B were arrived at by this operational definition of radical. The definition problem remains perplexing.

D. Radicals are Represented by Radical Variants

It has been a part of the traditional view of radicals to list several variant forms for some radicals. Thus, for example, 人 and 亻 are said to be variant forms of radical "9"^{1/} and 彳 and 辵 are said to be variant forms of radical "162".

We accept the view that some radicals have variant forms. However, we depart from tradition in that we list many more instances of multivariant radicals than occur in the traditional list. Looked at carefully, the data yield quite a number of instances of radical variation -- many more than are found in the traditional list.

^{1/}

Numbers in quotes refer to radicals of the traditional list. See Appendix A.

A relatively trivial phenomenon is size variation. When a radical such as "86" stands in isolation as a character as in 𠂇, it has, let us say, normal size; when it is embedded in a simple character as in 𠂇, it is smaller. Finally, when it is embedded in a more complex character as in 𠂇, it is much smaller. One of the problems in this research is to determine how many size-variants there are per radical.

A related phenomenon to that of size is that of shape. Sometimes when a radical is embedded in a character, not only is it reduced in size, but also its shape is distorted. For example, note the shape variation of radical "75" in 相 and 木.

Of course, 才 and 木 are still intuitively recognizable as being instances of the same radical. There are other cases which are not so clear-cut. For example, 亻 and 人 are traditionally considered variants of radical "9", but one feels that this is at least an extreme case of radical variation. This is directly analogous to the phenomenon of suppletion in natural language. An example of suppletive forms in English is "good" and the "bett-" and "be-" elements in "better" and "best." These are similar in meaning and noncontrastive in distribution. However, they are phonemically completely dissimilar.

One begins to search for constraints on radical variation. A perhaps sufficient constraint might be that two candidates for being radical variants of the same radical must be pictorially similar. Pictorial similarity ^{1/} remains to be defined precisely, but an idea of our intuition about it can be seen from our tentative solution of the 人 -

1/

The linguist will of course, see in this the analog of phonetic similarity as a constraint on the grouping of complementarily distributed phones into one phoneme.

1 problem. Both 人 and 亻 are grouped into the same radical, because they are both sufficiently described by some abstract structural statement like: Radical "9" consists of a northeast-to-southwest stroke which is joined L-wise to a non-west-to-east stroke. (See the next section for a discussion of L-adjoining.)

A related phenomenon is that of stroke-sharing. Often, when two radicals combine into a character, they partially overlap to the extent that one stroke of each radical occupies the same space. A clear case of stroke-sharing is in the character subpart 𠂇, where 立 and 戈 share a stroke. At this point we mention that the rules to cover cases such as this, i. e., the rules corresponding to the morphophonemic rules in a grammar for a spoken language, are tacitly assumed to be applied in our grammar whenever relevant. ^{1/}

E. Radical Variants are Composed of Strokes

It is clear from even a superficial study of the internal structure of radicals and their variants that there are regularities to be observed in the phenomenon of stroke-adjoinment. For example —, ↓, and ↙ may combine to produce 𠂇, but they may not combine to produce 丅 and 𠂆, for these latter two are not well-formed.

^{1/}

Stroke-sharing is of course closely related to the phenomenon of superimposition mentioned above.

There seem to be four common adjoining operations. First, there is T adjoining as in \top , where $—$ and \lrcorner combine T-wise. Then there is L adjoining: \lrcorner and \rceil combine L-wise to produce ρ . There is also X-adjoining, where two strokes intersect to form a radical or radical subpart. An example is the combination of $—$ and \lrcorner to form \top . Finally, there is an operation which does not actually adjoin two strokes, but which rather places them in a "near" relation; \lrcorner and $—$ are adjoined by the near-operation in $\overset{\cdot}{—}$.

The task here is to devise rules powerful enough to state the internal structure of all the radicals in terms of their constituent strokes and the stroke operations. We have in fact constructed a very preliminary finite-state grammar to represent these stroke-adjoining operations. This grammar, though perhaps promising, does fail on two counts. First, it cannot account for many radicals which seem to have nearly unique structures and, second, it seems to be able to account for many multiradical structures.^{1/} Thus, it seems to be at present both too weak and too strong.

F. Strokes Are Represented By Stroke-Variants

The model does define this area of research. However, we are not sure that this is a particularly fruitful area. There seem, in fact, to be very few instances of strokes with variant shapes. One possibility

^{1/}

Such as 高 , a traditional radical, but decomposable into 一
 口 , 冂 , and 口 .

is a stroke which has the two variants \downarrow and \downarrow . It might be characterized as a "vertical hooked stroke", the direction of the hook being predictable from the environment in which it occurs. Of course, every stroke has size variants, but this is almost a trivial observation.

G. Stroke Variants are Composed of Distinctive Components

It is presumably the case that each stroke-variant can be uniquely specified by a particular combination of distinctive components. Such components and their values would perhaps include:

<u>Ending:</u>	Hooked, Nonhooked (This would distinguish \downarrow from \checkmark)
<u>Direction:</u>	North-South, East-West, etc.: \downarrow versus — .
<u>Size:</u>	Large, Small: — versus — in +

H. Concluding Remark

It is expected that future work will provide much more in the way of results in the last four areas. As future work is pursued, much time will be devoted to work with native informants. The final step will be to express the results of this analysis in the form of a completed formal grammar.

ACKNOWLEDGEMENTS

For reading earlier versions of this paper and for making valuable suggestions, we wish to thank Margaret R. Fox, Ethel C. Marden, and J. Wegstein of the Information Technology Division; and R. A. Kirsch and W. Watt of the Applied Mathematics Division. We are further indebted to Mr. Kirsch for introducing to us the notion of "picture syntax" which underlies our approach to this problem. Finally, we wish to thank C. Chriss and D. Lottridge of the Information Technology

Division for the statistical tests relating to our statement concerning the adequacy of the grammar in Appendix B.

REFERENCES

1. Hockett, Charles F. A Course in Modern Linguistics, Macmillan, New York, 1958.
2. Gleason, H. A. , Jr. An Introduction to Descriptive Linguistics; Holt, Rinehart, and Winston, New York, (Revised Edition) 1961.
3. Hockett, Charles F. "Linguistic Elements and Their Relations," Language, Vol. 37, No. 1; pp. 29-53; 1961.
4. Lamb. Sydney. Outline of Stratificational Grammar, University of California, 1962.
5. Chao, Y. R. Mandarin Primer, Cambridge, Mass., 1961.
6. Mathews, R. H. A Chinese-English Dictionary, Harvard University Press, Cambridge, Mass., (Revised American Edition) 1960.
7. Chomsky, Noam. Syntactic Structures, Mouton and Company, s'Gravenhage, The Netherlands, 1957.
8. Wang, F-Y. Read Chinese, New Haven, 1961.

APPENDIX A

LIST OF TRADITIONAL 214 RADICALS, WITH VARIANT FORMS SHOWN

1	一	15	丶	28	厶	43	尢	54	攴
2	丨	16	几	29	又		兀	55	廾
3	丶	17	凵	30	口		宀	56	弋
4	丿	18	刀	31	凵	44	尸	57	弓
5	乙		リ	32	土	45	𠂇	58	彡
6	丿	19	力	33	士	46	山		彡
7	二	20	勹	34	夕	47	巛	59	彡
8	一	21	匕	35	夕		川	60	彡
9	人	22	匚	36	夕		巛	61	心
	亻	23	匚	37	大	48	工		忄
10	儿	24	十	38	女	49	己		忄
11	入	25	卜	39	子	50	巾	62	戈
12	八	26	冂	40	山	51	干	63	戶
13	冂		巳	41	寸	52	么	64	手
14	冂	27	厂	42	小	53	广		扌

65	支	79	爰	93	牛	106	白	120	系
66	攴	80	毋		牛	107	皮		系
	攴	81	比	94	犬	108	皿	121	缶
67	文	82	毛		彡	109	目	122	网
68	斗	83	氏	95	玄		回		回
69	斤	84	气	96	王	110	矛		冂
70	方	85	水		王	111	矢	123	羊
71	无		灬		王	112	石	124	羽
	无	86	火	97	瓜	113	示	125	老
72	日		灬	98	瓦		示	126	而
73	日	87	爪	99	甘	114	肉	127	耒
74	月		彡	100	生	115	禾	128	耳
75	木	88	父	101	用	116	穴	129	聿
76	欠	89	爻	102	田	117	立	130	肉
77	止	90	井	103	足	118	竹		月
78	彡	91	片	104	疒		竹	131	臣
	彡	92	牙	105	心	119	米	132	自

133	至	148	角	177	卩	194	革	鬼
134	白	149	言	164	酉	195	韋	魚
135	舌	150	谷	165	采	196	韭	鳥
136	舛	151	豆	166	里	197	音	鹵
137	舟	152	豕	167	金	198	頁	鹿
138	艮	153	豸	168	長	199	風	麥
139	色	154	貝	183	長	200	飛	麻
140	艸	155	赤	169	門	201	食	黃
141	艸	156	走	170	阜	202	首	黍
141	虎	157	足	186	卩	203	香	黑
142	虫	158	身	171	隶	204	馬	黼
143	血	159	車	172	隹	205	骨	黽
144	行	160	辛	173	雨	206	高	鼎
145	衣	161	辰	190	雨	207	髟	鼓
	衣	162	辵	174	青	208	鬥	鼠
146	西	163	辵	175	非	209	鬯	鼻
147	見	163	邑	176	面	210	鬲	齊

211

齒

212

龍

213

龜

214

龕

APPENDIX B

Grammars of this type are fully explained on pages 6, 7, and 8.

The numbers which appear in the left-most column of the lexical rules refer to the number of strokes per radical in the list of radicals below the number.

Grammar of radical-combination

Syntactic Rules

CHAR	=	COMP; IT
COMP	=	v (N, S); h(W, E); S(O, CHAR)
N	=	COMP; NT
S	=	COMP; ST
W	=	COMP; WT
E	=	COMP; ET
O	=	OT

LEXICAL RULES

I.	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
ㄥ		x					ㄍ					x	
一			x	x		x	ㄎ				x		x
一			x	x			ㄏ			x			
、		x	x	x			ㄏ			x			
/	x		x	x			ㄏ					x	
、			x				ㄏ						x
一			x				ㄏ				x		
一		x	x				ㄏ			x			x
丿		x		x			ㄏ			x			
乙			x	x			ㄏ			x			x
、				x			ㄏ					x	
丨		x					ㄏ			x			
乙				x		x	ㄏ	x	x	x	x		x
2.							又	x	x	x	x		x
几					x		几		x	x	x		x
《		x					七	x	x		x		x

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
川				X			九				X		X
七	X	X	X	X		X	刀	X	X	X			X
力	X	X	X	X		X	㇇			X			
㇇		X					ハ			X	X		
丁		X	X	X		X	儿				X		
又			X	X		X	㇇			X			
ト		X		X		X	㇇	X			X		X
ニ		X		X		X	ユ			X			
八			X			X	イ	X					
十	X	X	X	X		X	ト			X			
㇇			X				㇇			X			
了				X		X	㇇					X	
マ			X				㇇			X			
㇇			X				㇇	X					
㇇		X	X	X			㇇			X			
㇇	X					X	㇇					X	

3.	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
𠂇		x	x	x		x	又						x
止				x			𠂇						x
儿				x			弓	x					
乃				x		x	才	x					
云			x				口	x	x	x	x		x
彡		x	x	x			士		x	x	x		x
少				x			寸	x	x		x		x
𠂇			x				土	x	x	x	x		x
𠂇						x	丸	x	x	x	x		x
毛		x		x		x	𠂇	x					
久						x	尸			x	x		x
川	x					x	彡	x					
𠂇			x	x			工	x	x	x	x		x
兀				x			女	x	x	x	x		x
𠂇		x					𠂇		x				x
千		x	x			x	上	x	x	x	x		x

广			x			夕	x		x	x		x
口					x	毛		x		x		x
山	x	x	x	x		也		x		x		x
大		x	x	x		彡	x					
上	x					卅				x		
干	x	x		x		又		x				x
么				x		才						x
廿				x		卅			x	x		
子	x	x	x	x		仁			x			
么			x	x		巾	x			x		x
夕	x	x	x	x		丈		x				x
下						万				x		x
小		x	x	x		五			x	x		x
卩	x	x				丰			x	x		x
彡	x					亼			x	x		
山					x	刃		x	x			

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
亻	x						平				x		
彳	x						巾		x				x
巳				x		x	尹		x	x			
亡		x	x			x	丑						x
亅	x						舟						x
己		x		x		x	乏						x
己		x				x	井						x
几		x					予		x				x
之						x	五		x	x			x
4.							夕	x					
午		x				x	云		x				
歹	x					x	天	x	x		x		x
欠	x						元				x		x
辶				x			壬		x		x		x
尸			x				户			x			x
由			x				巨		x				

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
不		X	X			X	𠂇				X		
𠂇				X			𠂇		X		X		
圭			X			X	𠂇			X			
月	X	X		X		X	𠂇	X	X		X		X
无						X	𠂇			X			
𠂇			X				𠂇	X		X	X		X
斗		X		X		X	止	X		X	X		X
𠂇			X				丹				X		
水			X	X		X	𠂇			X			
王	X			X		X	毛				X		X
用	X					X	𠂇			X			
片	X					X	牙	X					X
𠂇				X			𠂇		X		X		X
犬	X	X	X	X		X	欠		X				
手						X	𠂇	X					
𠂇	X						𠂇			X			X

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
市		X				X	止				X		
允	X		X			X	开				X		X
艮		X					半				X		
甘		X					日	X	X	X	X		X
丰			X				斤	X	X	X	X		X
令		X					曰			X	X		X
亏				X			戈		X	X	X		X
斤		X					世		X	X	X		X
气			X				夬		X				
少				X			巴	X	X	X	X		X
丐		X					尤		X		X		X
生			X				夂		X				
亘			X				心				X		X
彡	X						勿	X	X		X		X
丰		X					屯		X				X
夭		X				X	木	X	X	X	X		X

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
石	x			x		x	石		x				
生				x		x	史	x	x		x		x
令	x					x	臣		x		x		x
夫			x				四	x		x	x		x
丞			x	x			立	x	x	x			x
韭			x				瓦		x				x
彡			x				冉				x		
互		x				x	母		x		x		x
业			x				本		x				x
卅			x				必						x
半		x				x	半				x		
禾	x	x	x	x		x	矢	x			x		x
电				x			目	x	x	x	x		x
弗		x				x	央		x	x	x		x
戊					x		四				x		
示	x			x		x	示	x					

WT ET NT ST OT IT

WT ET NT ST OT IT

皮

X

X

勾

X

由

X

X

X

虫

X

低

X

X

X

出

X

乍

X

X

X

X

内

X

六

X

6.

甲

X

身

X

凶

X

X

业

X

X

申

X

X

X

产

X

田

X

X

X

X

X

高

X

心

X

X

瓜

X

X

臣

X

西

X

丙

X

禾

X

疒

X

先

X

X

永

X

X

X

兆

X

X

羊

X

X

吏

X

X

白

X

X

X

X

X

朱

X

X

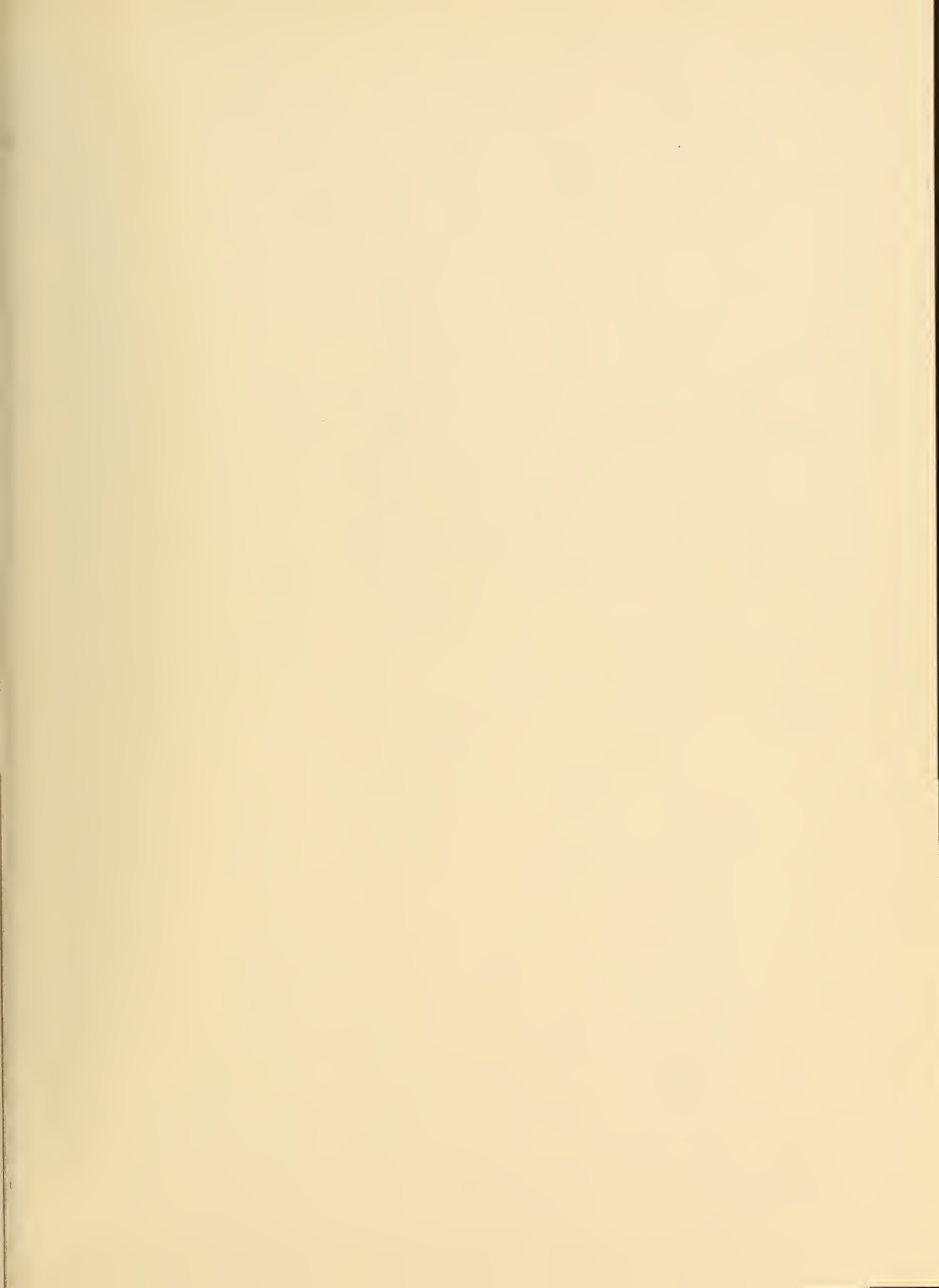
	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
羊		X				X	其			X			
戌		X				X	臣	X	X				X
戔			X				虐			X			X
衣	X	X		X		X	合	X	X	X	X		X
禾				X			圭		X				X
后		X					米		X		X		X
夫			X				亥	X	X		X		X
州		X					虫	X		X	X		X
矢			X				百		X		X		X
夷		X				X	白				X		
羊			X				再		X		X		X
且			X	X		X	自	X		X	X		X
鹵			X				羽		X		X		X
夆	X						夆			X			
丰				X			丰	X	X				X
自			X			X	而				X		X

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
糸	x	x		x		x	局		x				x
卑				x			由			x			
白			x	x		x	甫			x			
耳	x	x	x	x		x	夾		x				x
西			x				丟	x					
关			x			x	回				x		
四			x	x			烏			x			
羊			x	x			兔		x				x
曲			x			x	兩						x
舟	x	x				x	串						x
聿			x	x			西				x		
艮		x	x	x		x	更		x				x
聿			x				辛	x					
勻				x			辛		x				
7.							美			x			
呂	x	x		x		x	兑		x				

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
角	X					X	見		X		X		X
辰			X				足	X					
良	X	X	X	X		X	里		X		X		X
貝	X	X	X	X		X	8.						
采	X	X	X	X		X	非		X		X		X
言	X	X	X	X		X	果		X		X		X
臬				X			馬			X			
豕				X		X	魚				X		
身	X	X		X		X	肅				X		
門	X	X		X	X	X	乖						X
我		X		X		X	戔	X	X				X
豆	X	X		X		X	里			X			
坐				X		X	食	X					
酉	X	X		X		X	雨			X			X
豸	X						革	X					X
車	X		X	X		X	其	X	X	X			X

	WT	ET	NT	ST	OT	IT		WT	ET	NT	ST	OT	IT
甫		X	X			X	面		X		X		X
長		X				X	傘						X
帛				X			幸			X			X
林			X			X	垂		X				X
佳		X	X	X		X	堇		X				X
宙			X				無						X
金	X					X	東						X
東						X	炆			X			
來		X		X		X	齒	X					X
事						X	乘						X
兩		X				X	絲		X	X	X		
9-plus							頁	X	X		X		X
乖						X	馬		X				X
亞						X	魚	X	X		X		X
網			X				鳥		X		X		X
禹		X		X		X	骨	X					X

	WT	ET	NT	ST	OT	IT	WT	ET	NT	ST	OT	IT
黑	X					X						
鹿	X		X			X						
絲						X						
首			X			X						
咕			X									
留			X									
重	X			X		X						
甚						X						
綯			X									
禺		X		X		X						





U.S. DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20230

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE

OFFICIAL BUSINESS

