THE STRUCTURALISTIC APPROACH TO CHINESE GRAMMAR AND VOCABULARY

JANUA LINGUARUM

STUDIA MEMORIAE
NICOLAI VAN WIJK DEDICATA

edenda curat

CORNELIS H. VAN SCHOONEVELD
STANFORD UNIVERSITY

SERIES PRACTICA
VI

1964
MOUTON & CO · THE HAGUE

THE STRUCTURALISTIC APPROACH TO CHINESE GRAMMAR AND VOCABULARY

Two Essays

by

GUSTAV HERDAN

UNIVERSITY OF BRISTOL

1964

MOUTON & CO · THE HAGUE

Printed in Germany by J.J. Augustin, Glückstadt

PREFACE

The use of the term 'structural' in the title of a book makes it advisable to state clearly how that term is to be understood. Modern linguistics suffers rather from the polysemy of its termini technici, such as 'information', 'function', 'structure', to name only these.

I use the term 'structure' in the Saussurian sense, as referring to language as a system of oppositions between linguistic forms. Since this applies to the forms of any specified level of language, it leads to the conception of language as a system of solidarity among the phonemes of a given language, among its vocabulary items, and among the grammatical forms needed for the expression of relations between the vocabulary items. Insofar as structure is amenable to mathematical formulation, we speak of mathematical linguistics.

The comparative rarity of investigations into the mathematical structure of Chinese has, no doubt, as one of its reasons, that a knowledge of both Chinese and mathematical linguistics will only rarely be found together in one head. This was no obstacle in the present case. I happen to have some knowledge of Classical Chinese, and I was greatly assisted by my wife Innes Herdan, whom I may describe as a Chinese scholar. She is the author, under her maiden name Jackson, of *China only Yesterday* (London, Faber & Faber, 1938) and has collaborated with Chiang Yee in his first and best known book, on Chinese painting, *The Chinese Eye* (London, Methuen 1936; reprinted 1960), to name only these two instances of her literary work.

The two essays in this book originally formed part of the manuscript of my book *The Calculus of Linguistic Observations* (The Hague, Mouton & Co., 1962), but the amount of Chinese type required made it necessary to publish them in form of a separate booklet.

Bristol, August 1963. GUSTAV HERDAN

CONTENTS

THE COMBINATORIAL ELEMENT IN CHINESE GRAMMAR AS SUBJECT TO LINGUISTIC DUALITY

I

PRINCIPLES OF CHINESE GRAMMAR

Chinese is fundamentally a language in which each word is isolated, uninflected, and equivalent to a root.

> "Les caractères chinois sont tous monosyllabiques, indeclinables et inconjugables. Il ne sont donc point susceptibles de recevoir ces flexions qui, dans les langues greque et latine, font voir, au premier coup d'oeil, les genres, les cas, et les nombres des noms, les voix, les temps, les modes et les personnes des verbes. Mais, malgré cette absence de flexions, la langue chinoise est, pour un sinologue instruit, aussi claire, aussi intelligible que les langues savantes qui sont richement pourvues des flexions qui lui manquent" (S. Julien, *Syntaxe nouvelle de la langue chinoise*, Paris, 1869).

How is that possible? What is there instead of flexions (bound grammar forms) which is capable of making a language "grammatical"? The answer can be given in one sentence: "The whole of Chinese grammar depends on position." This is the principle adopted by Stanislas Julien, S. J. who for sinology in France is what James Legge is for sinology in England.

Illustrations

(a) Take, for example, the word 治 'to govern'. If this verb is placed before 國, these two characters will signify: 'to govern the kingdom'.

If the word order is reversed, and 國 'kingdom' is put before 治 'to govern', the two words will signify: 'the kingdom is governed'.

If the word 治 is placed after 史 'magistrate', it becomes a substantive, and the two words will signify: 'the administration of magistrates'.

(b) The word 團 'round' can, according to its position, be adjective, noun, verb and adverb.

One says 團扇 'a round fan',
粉團 'a bowl of flour',
團飯 'to put cooked rice into a bowl',
團坐 'to sit in a circle'.

(c) The word 官 'magistrate' can carry four different meanings according to its position, or according to the characters with which it is combined:

建官 'to nominate a magistrate',
中官 'an official of the inner palace' (i.e. a eunuch),
不辭小官 'not to refuse a small magistrature',
能官人 'he can employ men'.

It should be clearly understood that the precision of expression which in the classical languages is achieved by the flexions of nouns and verbs can in Chinese be achieved, to some extent at least, through the mobility of the Chinese characters which are capable of acquiring all sorts of grammatical values according to the position they occupy in the sentence, or phrase, and according to the particular words (characters) which are used in conjunction with them. It is the relative position of the words which determines their role, and gives all desirable clarity to expression.

The fluidity of the Chinese grammar categories and the consequent possibility of an ideogram appearing in almost any position in the sentence is responsible for a high degree of utilisation of combinatorial possibilities in the order of ideograms within the sentence. The mobility of Chinese words is such as to approach the combinability of the letters of the alphabet in a Western language or, as we prefer to say today, of the phonemes of a language. It represents an important factor in reducing the effort needed for expression which, as I have shown (*Language as Choice and Chance*,[1] Sect. 7.2) is the less the smaller the number of irreducible units.

In a Western alphabetical language, say English, we have only 26 letters of the alphabet for embodying all possible words in the language, and 42 phonemes. This is sufficient for the formation of the whole English vocabulary. Similarly, the possibility of grammatical transposition in Chinese means a comparatively small number of basic words needed for expression, and reduces the effort needed for speech and writing. Obviously, a language in which a noun was always a noun, a verb always a verb, an adjective always an adjective, would require three times as many words as Chinese for expressing the same complex of content.

In this respect, combination of unit word forms serves here the same purpose as does that of distinctive unit sounds (phonemes) in the Western languages. This is in full agreement with my thesis (*CLO*, Sect. 15.1) that isomorphy between distant languages, e.g. between English and Chinese, may exist for different levels.

[1] The author's previous books on mathematical linguistics will be referred to as:
LCC for *Language as Choice and Chance* (Groningen, P. Noordhoff, 1956).
TTM for *Type-Token Mathematics* (The Hague, Mouton & Co., 1960).
CLO for *The Calculus of Linguistic Observations* (The Hague, Mouton & Co., 1962).

Adding this isomorphy to those established in *CLO*, Chapter 15, we list our three results of correspondences as follows:

Isomorphies

Western languages	Chinese
binary units	strokes
phonemes (letters)	ideograms
morphemes	syntax

But all this must be taken cum grano salis. Combination as we understand it is a mathematical term for chance variations of order and position, and as such, knows no limits or exceptions. In its application to language, however, there are two kinds of restriction.

(1) We know that not all phoneme combinations are admitted in a language, e.g. of 18 possible German consonants, 17 are admitted in initial and 14 in final position, in 2-phoneme monosyllables.[2] This is also true, to an even higher degree, for word combinations in Chinese. When speaking about the sequence of words in a Chinese sentence, of combination as a chance ordering of the words, we do not mean that every such sequence makes sense, and that, therefore, every such sequence represents a sentence.

(2) Even among the admissible combinations of sounds in a language, i.e. the morphemes as listed in the dictionary, there will be preferred and less preferred ones, the number of words containing the former being more numerous. This quite apart from the differential utilisation of the morphs in speech and writing.

All this must be regarded as interference with the mathematically possible number of combinations, and as reducing numerically the actually occurring ones.

In addition to the restrictions of combinability of words, which Chinese on the syntactic level has in common with Western languages on the phonemic level, there are certain *differences* between phoneme (alphabetic) and ideogram combinations.

(1) In order to be able to say whether a language has been economical in the utilisation of the possible phoneme combinations, or whether it has squandered such possibilities, we need a standard of comparison, and such a standard is provided by the mathematically determined number of possible variations (combinations) for a given number of units (*TTM*, Sect. 7.4).

What are we to take as the number of units corresponding to the 26 letters and 42 phonemes of English, for word combinations in Chinese? The whole

[2] *TTM*, Section 7.4.

dictionary, which may have between 5,000 and 50,000 words or more, according to the completeness of the dictionary? Since the formulae for the number of combinations work with the factorials of numbers, that is, with the continued products $1 \times 2 \times \ldots\ldots\ldots n$, the figures would become so formidable as to be almost useless for our purpose. It is clear that the difference in magnitude of the basic unit numbers makes it necessary to work with comparatively small groups of words for providing the necessary standard of comparison.

(2) There is another, and even more intrinsic difference. The different combinations of phonemes belonging to a given group have no relation to each other over and above that of being random variations (combinations) of the same basic units.

Different sentences, however, consisting of the same ideograms or words, wholly or partly, must stand in some sort of semantic relation to each other. This is what makes their combination different from mere random combination. Sentences consisting of the same words are grammatical transpositions of one another. This is what makes us see them as a whole, the parts of which are systematically related to each other.

These two differences provide us with some hint of where to look for the analogue of the arithmetical combination formulae which are suitable for providing a standard for letter and phoneme combinability in a language. They will help us to find the structural background–for this is the proper function of these formulae–for syntactic combinations in Chinese. What we are looking for is a standard against which to judge the resourcefulness of the language in syntactic combinations. As I shall now show, such a standard is provided by some of the basic configurations of projective geometry, the so-called theorems of duality.

Anyone who has followed my argument carefully will know that I do not mean that these theorems are to be regarded as laws of grammar in the usual sense. This would be as wrong as to declare chance combination to represent the law of morphology. In both cases, mathematics only explains *how it is possible to achieve so much* (the whole vocabulary of a language and all the sentences in a given universe of discourse) *by so little* (a comparatively small number of letters or phonemes and a comparatively small number of ideograms). This is more fully discussed in the following Section.

II

PHONEMIC COMBINATION IN WESTERN LANGUAGES AND SENTENCE STRUCTURE IN CHINESE

As was said above, the difference between the universe of phonemes and that of morphemes (words) may serve as pointers to where to look for the standard form of morpheme combinations.

Considering the isomorphy between Western phonemes (letters) and Chinese radicals as parts of ideograms, the combination of the former into morphs of specified length would provide the pattern for the combination of the latter into sentences. Just as it is an easy matter to calculate the possible number of combinations of phonemes into words containing, say, 3 phonemes, so it is equally feasible to calculate the possible number of sentences to be formed of, say, 3 words, and in general with a finite and comparatively small number of words, say, not more than ten. This would provide the standard against which to judge the semantically feasible or sensible number of sentences of that length which may be formed out of the given vocabulary material.

The second feature to be considered is the element of structure. As explained above, the combinations of words into sentences are not entirely random, but have an element of structure. Whereas the random element belongs to the plane of expression, the structure element as that which ensures the sentence making sense, is the contribution from the plane of content.

If we now consider that the non-random or structural element of the combinations could be expressed in terms of a geometrical figure of some kind, it is clear that what we are looking for must be a pattern in which both chance combination and geometrical combination participate. Suitable figures of this kind are, I suggest, provided by certain configurations of projective geometry, namely the various theorems of geometric duality.

Before demonstrating these theorems, we must be clear about what is meant by "fitting" these configurations to linguistic data.

They are not meant to represent "laws" of language, and especially of sentence structure, in the sense of being the governing principle of language or sentence formation from the start. They do not compel the speaking individual to express himself in any particular way so as to "obey" these laws. In the words of a German phrase: "es war der Sprache nicht an der Wiege gesungen", that its configurations would closely resemble certain geometrical

patterns. That resemblance, such as there is, has grown in time as a result of the economical tendency on the part of the members of a speech community to be sparing, not wasteful, with the number of different vocabulary items. This explains both the close approach to combinatorial methods in sentence formation, and the high degree of interlocking device between sentences. The combination of the combinatorial and structural or interlocking devices on the syntactic level of language is reminiscent of the cross-word puzzle as a combination of combinatorial and interlocking devices on the phonemic level, except that the interlocking device is here only something arbitrary, non-linguistic, added for the sake of the game, whereas on the syntactic level it is a most essential part of the linguistic structure.

What started without any intention at structure, namely the formation of sentences ad hoc, as the need for communication required it, became with time, and prompted by the need for economy in words, a system of solidarity, *as if* governed by structural laws of projective geometry.

What happened, and especially how a primarily non-structural use of words as means of communication grew with time into an interlocking system of words according to their meaning, has been described in an inimitable way by Goethe in one of those comet-like utterances of his which do not seem to have any connection with anything else he has written. It is entitled "Mephistopheles spricht" and has these lines:

> "So wird erst nach und nach die Sprache festgerammelt,
> Und was ein Volk zusammen sich gestammelt,
> Muß ewiges Gesetz für Herz und Seele sein."

How perfectly Goethe, though rather averse to any form of linguistics, has here expressed the fact that what started as a haphazard connection between words became with time the interlocked system of meaning which governs the use of words! That he puts this statement into Mephistopheles' mouth, shows his awareness of the fact that it was not in accordance with orthodox etymology.

III

GEOMETRICAL OR CONFIGURATIVE COMBINATIONS

The following examples from Chinese are chosen so as to illustrate the argument of the previous Section, step by step.

With the ten characters given below (A), thirty-five sentences were formed (B), this being by no means the maximum possible number.

Examples (*A*):

天	好	開	明	性	有	無	中
'Heaven', 'Nature', 'weather'	'good'	'reveal'	'clear', 'clearly'	'disposition'	'to have'	'no', 'not'	'centre', 'among'

眞	假
'true', 'sincere'	'false'

Examples (*B*):

1. 明天好	11. 眞開明	21. 無眞性	31. 眞好天
2. 天性好	12. 假開明	22. 假天性	32. 假好天
3. 好開明	13. 天明中	23. 眞天性	33. 眞有無
4. 有天性	14. 無眞假	24. 明天有	34. 假有無
5. 無天性	15. 有眞假	25. 明天無	35. 明性好
6. 有無中	16. 眞中假	26. 明眞假	
7. 無中有	17. 假中眞	27. 天明無	
8. 有中無	18. 天眞好	28. 天開無	
9. 有好天	19. 假天眞	29. 天性眞	
10. 無好天	20. 有眞性	30. 天性假	

So far the combinations do not seem different from the phoneme combinations which make up the morphemes of a Western language. However, we now proceed to more ambitious patterns. We may try to sort out combinations which by interlocking with each other enable us to view a group of sentences

as one organic whole. From the above ten characters, sets of three can be selected such that each of the ten characters appears in three sentences, and no more. This means that we arrange the ten characters in ten sets of three characters, each character appearing in three sentences, which arrangement satisfies Desargues' theorem (see Sect. IV). One such set as given below under (C) is a close approximation to it. The sentences are now more closely packed, and we may enquire after the "closest possible packing" (see Section IV).

Examples (*C*):

Chinese	*English*
無中有	'There is something in nothing'.[3]
有好天	'It is good weather'.
好開明	'How open-minded he is'.
假開明	'Pretending to be open-minded'.
眞中假	'The true in the midst of the false'.
假中眞	'The false in the midst of the true'.
無眞性	'It is not an upright disposition'.
明性好	'A fair disposition is good'.[4]
有天性	'To have a natural disposition'.
天開無	'Has the sky cleared?'

[3] Usually written 無中生有 'something grows from nothingness'.
[4] A slightly forced expression: one or two more characters are needed for clarity.

IV

THEOREMS OF DUALITY AND CHINESE SENTENCE STRUCTURE

The peculiar type of interlocking of sentences as illustrated is in good conformity with the structure of a famous theorem of projective geometry – Desargues' duality theorem. This does not mean that the relations in a group of sentences are exactly those between the points and lines of Desargues' theorem, but rather that the linguistic pattern approaches the geometrical one closely, and that, therefore, the latter may be regarded as a pattern of "optimum packing" which languages might find useful to aim at, without ever completely reaching it. It represents the ideal for a closely-knit group of sentences.

Desargues' Theorem

One of the fundamental theorems of projective geometry is that due to DESARGUES who, by the way, can be regarded as the discoverer of the principle of geometrical duality.

His theorem is shown in Figure 1. Starting at O, we draw the lines OA, OB, OC. A, B, C can be anywhere on these lines. We then mark three points A', B', C' as shown in the diagram. We join A, B and A', B', which two lines meet in F. In the same way, A, C meets A', C' in E; B, C meets B', C' in D. The points D, E, F lie on a straight line. This result uses only the ideas of straight line and point. It is fully projective, and independent of measurement. The theorem which, as observed above, is known as Desargues' theorem, possesses a remarkable symmetry. Every point in it is as good as any other point. Instead of beginning with O and then bringing in A, B, C and A', B', C' so as to finally arrive at D, E, F, we could have marked any of the 10 points as O and we would have got the same result. The diagram would be the same only that the points are differently named. There are, in fact, 120 different ways of putting in the letters in this diagram without any alteration in the results. 120 is the number of times in which three items could be chosen out of 10 according to the formula $10!/3!\,7! = 720/6 = 120$.

There is perfect democracy among the points and lines of the diagram, which is the characteristic feature of distribution in general. There are 10 lines each having 3 points on it. There are 10 points each having 3 lines through it.

Figure 1 is spoken of as a 10_3 10_3 configuration. As we shall see, there are other diagrams with similar properties, and the question is, what are their general characteristics. In a configuration, each line must have the same number of points on it and each point the same number of lines through it.

Incidentally, Desargues' theorem gives the answer to a well-known nursery puzzle: how to plant 10 trees in 10 rows of 3. This is how we can describe the interlocking pattern of 10 sentences under (C), p. 18.

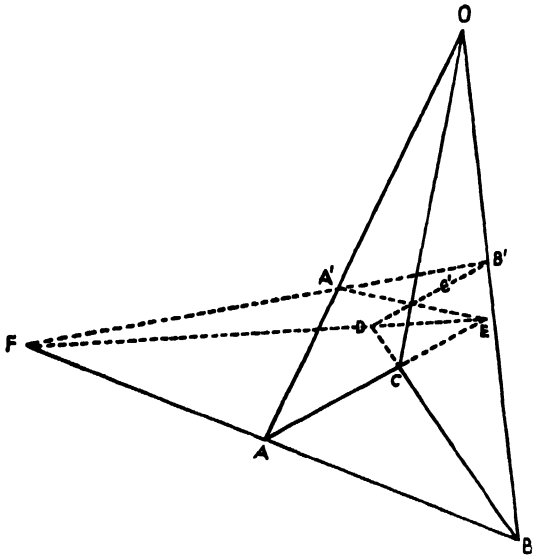

Fig. 1

Incidentally, the close fitting of the sentence group by Desargues' theorem of duality supports my basic thesis in which Type is equated with point and Token with line (*TTM*, Chapter XVII), which demands that the fundamental relations between Type and Token, being those between point and line, should obey the fundamental geometrical laws governing the relations between point and line.

Pappus' theorem

Another theorem independent of angle and length is Pappus' theorem (320 A.D.), which involves 9 points and 9 lines as shown in Fig. 2. We draw 2 lines which may, or may not be parallel and mark anywhere 3 points on each, A, B, C and D, E, F. We then draw the lines A, E and D, B crossing at N, A, F and D, C crossing at M and B, F, C, E crossing at L. We find then that M, N, L are on a line. This diagram is a configuration denoted by 9_3, 9_3. It has 9 points and 9 lines, each line containing 3 points and each point being on 3 lines. Each point is as good as any other point, each line as good as any

other line. There is again complete solidarity between the points, between the lines, and between lines and points, so that the theorem remains true if the terms point and line are interchanged. There are 108 different ways of putting in the letters of this diagram without any change in the statement of the theorem becoming necessary. Today this theorem is considered to be a special case of the theorem of Pascal (see below).

The following language example (D) is a 9_3, 9_3 configuration (theorem of Pappus) of linguistic elements, namely one of Chinese words, but this time of 9 such words forming 9 sentences with 3 words each.

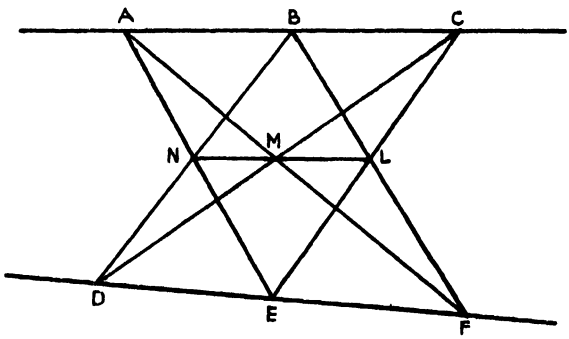

Fig. 2

Examples (D):

Chinese	*English*
ABC 人爲老	'The man is old'.
ANE 人本善	'Man is fundamentally good'.
AMF 人之仁	'As regards man's benevolence'.
BND 爲本性	'It is his fundamental disposition'.
BLF 爲義仁	'He is virtuous and charitable'.
CMD 老之性	'An old (person's) disposition'.
CLE 老義善	'Always virtuous and good'.
DEF 性善仁	'His disposition is good and charitable'.
NML 本之義	'As regards fundamental virtue'.

The letter combinations in the first column of examples (D) and (E) refer to the corresponding letter sequences in figures 2 and 3 respectively.

Pascal's Theorem

The reason why it is comparatively easy to make sentence groups fall into duality patterns is, as shown above, the pronounced separation of grammar

category and grammatical function in Chinese. A character which primarily belongs to a specified grammar category, say noun, can, as a rule, be also used in the function of a verb or adjective. And the same, mutatis mutandis, applies to characters belonging primarily to the other two grammatical categories. The next illustration (Fig. 3) represents this principle of Chinese grammar in form of the theorem of Pascal. Let A, B, C represent 3 Chinese words, A belong to the category of nouns, B to the category of verbs, C to the category of adjectives. Since most Chinese characters, no matter to what grammatical category they originally belong, could be used with different grammatical function, we denote by A′, B′, C′ and A″, B″, C″ two grammatical functions different from the grammatical category to which the word originally belongs. In addition, the grammatical nature of a Chinese word is determined by the order in the sequence of characters and, as will be seen, the configuration according to Pascal's theorem provides also for this through modification of the original grammatical category.

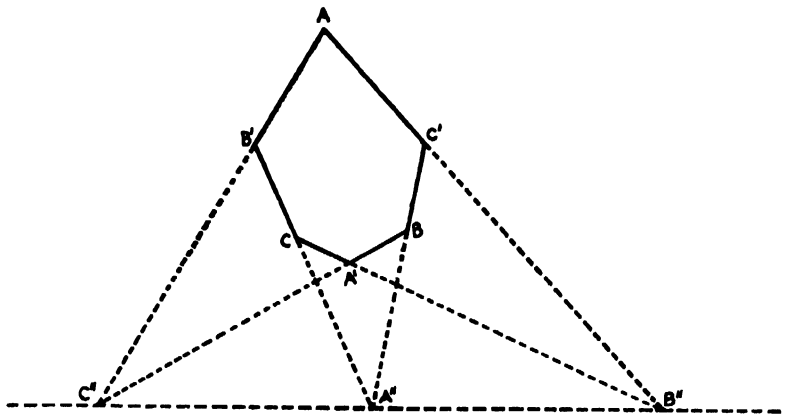

Fig. 3

The sentences of 3 words according to Pascal's configuration are

A B′ C″	C A′ B″
A C′ B″	B A′ C″
B′ C A″	C″ A″ B″
C′ B A″	

The characters A, A′, B, B′, C, C′ occur each in two sentences and the characters A″, B″, C″ each in three sentences. There is, thus, complete solidarity between grammatical categories, grammatical functions and sentences which we have found to be characteristice for any configuration of points and lines. This leaves room for the modification of the position of the character as

required in Chinese grammar, B, C, B′, C′, A″, C″ occurring in 2 different positions each (not so B″, A and A′) according to our scheme.

Examples (*E*):

A 生	*noun*, 'life'.
A′	*adjective*, 'living, growing'.
A″	*verb*, 'to live', 'produce', 'grow'.
B 喜	*verb*, 'rejoice in', 'be happy about'.
B′	*adjective*, 'happy'.
B″	*noun*, 'happiness'.
C 長	*adjective*, 'long, elderly'.
C′	*verb*, 'prolong, grow'.
C″	*noun*, 'length, seniority'.

In the following examples, the translation in brackets is an alternative and more literal version:

A B′ C″	生喜長	'A happy life has length'. (Life delights in being long.)[5]
A C′ B″	生長喜	'Life is prolonged by happiness'. (Life will be happy if long).
B′ C A″	喜長生	'Happy to live long'. (Taking delight in long life.)
C′ B A″	長喜生	'Always love to live'.
C A′ B″	長生喜	'A long-lasting, growing happiness'. (Long life is delightful.)
B A′ C″	喜生長	'Happy with increasing length of life'. (Satisfaction in growing old)
C″ A″ B″	長生喜	'Length of life produces happiness'.[6]

[5] If 喜 is taken to be a verb.
[6] If 長 is in the third tone, the expression 長生 means 'to grow'.

V

LINGUISTIC DUALITY

The importance of the concept of duality for the understanding of language structure has so far escaped the attention of the linguist, although mathematicians have emphasised it. The English mathematician H. W. Turnbull says:

"It is a fact that duality is inherent in language and at all levels; and a fruitful exploration of language is possible with this fact in view."

"It appears in the contrast between the two great language systems of the Chinese and of the West." (H. W. Turnbull, "Mathematics in the Larger Context", *Research*, Vol. 3, No. 5, 1950).

And I have elaborated this in three books of mine, *LCC* (1956), *TTM* (1960), and *CLO* (1962), and particularly I have presented the matter in the Epilogue of the second book as well as I know how. That it should fall on the "blind spot" of the linguist's eye is very regrettable, but it is far too important a matter for an understanding of the Calculus of linguistic observations for me to give up trying.

1

Concepts having originated through complete abstraction as general entities, in contradistinction to individual objects, have a sort of objective existence outside time and space. They do not form part of either the temporal sequence or the spatial configuration of events in the manifold of experience. If, however, they are to enter the world of experience, they must be "degraded" so to speak, to the nature of individual material objects. These physical objects are the words (or the forms, in general) of language. They are the physical signs of the concepts and, as such, the necessary means for fixing the latter (Schopenhauer, quoted in *CLO*, p. 234, 235).

The "degradation" consists in entering time and/or space as the words of language. In the case of the alphabetic languages, such as the Indo-European, where the spoken language is primary and the written secondary, it is the time element which mainly degrades the concepts, whereas in the case of the ideographic languages, such as Chinese, which are primarily written languages, it is the space element which acts as the main degrading factor.

It is through this that language cannot but share certain properties with other activities of the human organism. Most human actions are performed by symmetrical organs (*TTM*, Chpt. 17), e.g. arms, legs, eyes, ears, lungs. Symmetry of action implies space and time. Concepts, being as such outside space and time, have no symmetry until they are "degraded" in the sense described above, i.e. until they have entered space and time as the forms of language. It is then that a sort of symmetry appears. But that symmetry is a peculiar one: it is a symmetry sui generis fully appropriate to the basic units, the concepts. It consists of the reciprocity–a perfectly symmetrical relation–which is known under the name of the Principle of Duality. This kind of symmetry is inherent in the linguistic space-time, which conception of it is in full accord with the modern view of the origin of mathematical duality (Turnbull, *l.c.*). The fact of language having a sort of built-in symmetry, or reciprocity, explains why the speech function in the brain could be one-sided, in spite of speech being one of the most complicated brain functions. Language does not require two symmetrical organs with nerve-centres located in different halves of the brain for producing the reciprocity or duality peculiar to linguistic events, because it is in the very nature of these events that such duality resides. (*TTM*, Sect. 17.3, 17.4).

2

The law of duality in language assumes different forms according to the type of universe, which represents the background of linguistic events against which they must be seen if they are to be understood in their true perspective.

(1) The duality appropriate to the plane of content is Boole's law of duality as the expression of the Principle of Contradiction, in symbols

$$x(1-x) = 0$$
$$x = x^2$$

As Boole has so convincingly pointed out, this is the basic law of logic and, with it, of the content of language. The figure 1 in the basic equation stands here for the manifold of experience or the world, and x stands for a conceptual quality and can take only two values: 1 for presence and 0 for absence. The mathematics on this level is Boolean algebra.

(2) The universe appropriate to the plane of expression is that of discourse, or rather one of the many possible universes of discourse.

Apart from the alternative for a linguistic form, say a word, to occur or not occur in a given universe, the actually occurring linguistic forms may do so with different probabilities between 0 and 1, the extremes standing for non-occurrence and total universe content, respectively. This is in accordance with the frequency definition of a probability as the ratio of the number of

occurrences of a particular event to the number of possible occasions for its occurrence (the universe of discourse).

(3) The third form of the law of duality is the law of geometrical duality. This law governs the Type-Token relation on each particular level of language; phonemic, vocabulary, grammar. This is the law of duality of projective geometry. It was this law in particular which Turnbull had in mind when stressing the duality structure of language. When mathematicians speak of duality it is mostly this variety which they are thinking of.

The universe is here the language level, "Type" the linguistic units appropriate to that level, and "Token" the combinations of these units into the next higher forms, e.g. phonemes into morphemes, words into sentences. Curiously enough, since the point of geometry has the dimension 0 and the line the dimension 1, the alternatives Type and Token on each language level are again represented numerically by the symbols 0 and 1.

My contribution here was to equate linguistic Type to the Point of the duality relation, and linguistic Token to its Line. Moreover, I have shown geometrical duality to be only one variety of the law of duality in language, the other two forms being Boole's law and the probability laws. This has led to the recognition of three distinct universes of mathematical linguistics: the manifold of experience, or "the world" (for Boole's law), the universes of discourse or the statistical universes (for the probability laws of language), and the different levels of language (for the law of duality of projective geometry). What I call the Calculus of linguistic observations takes all these types of universe into account when considering language structure. In so complex and many-sided a subject as language, one cannot expect each of these methods to have full application everywhere. Nor is this the case. We find that it is mostly to phonemes and vocabulary that the second part of the Calculus applies, whereas the third has its proper sphere on the grammatical or syntactic level.

The extension of the duality concept as outlined under (2) may be regarded as the quantification of de Saussure's ideas, especially his langue-parole dichotomy, and that under (3) as the manifestation of Boole's law of duality in the plane of expression.

This is the essence of my teaching, *LCC* (esp. Parts I, II) being essentially an exposition of the former, and *TTM* (esp. Part IV) of the latter.

VI

GRAMMATICAL TRANSPOSITION AS A STYLE CHARACTERISTIC FOR ITS OWN SAKE

Grammatical transposition which so far we have encountered in prose as a powerful syntactic device for increasing the number of possible constructions has been used in Chinese poetry as a device of art, that is, for its own sake. We shall develop our argument by referring to the so-called Language Game of Classical Chinese.

Incidentally, this will be helpful in meeting an anticipated objection to our conclusion of duality theorems as governing the distribution of words in sentences, at least in the Chinese language, namely that such duality patterns are not intrinsic to language, but are only read into it by the mathematician; that they are not operating when words are arranged into sentences by the speaker or writer, but are only the result of selection from accumulated data.

The essence of the Language Game may be expressed like this: starting from an original poem, the order of the sentences, and that of the words within a sentence, are altered so as to give a new poem. It should be clearly understood that what I call the 'original poem' is not just any given poem, but must have already been composed with the subsequent operations in view. Since both the displacement of the original sentences (the columns of the poem in the Chinese script) and the reversal of the characters within a column are mathematical operations, we may say that this shows a mathematical device as operative in *transforming an original poem into its dual.* For this is the desired result: the situation described in the original is, in the transformation, seen from another angle, the two poems being dual to one another in expression and content.

The most amazing poems of this type are those on the Huei Wen T'u (回文圖) or Revolving Chart, by the Lady Su Huei (蘇蕙)[7] of Eastern Ch'in dynasty (東晉). This Chart consisted of 800 characters woven in five colours on a piece of silk, representing love poems written by Lady Su Huei and sent to her husband, General Tou T'ao (竇滔) who was then guarding the Northern Border against the Tartar invasion. If read in different directions, these characters read as different poems. The poem I in our Table I describes this

[7] 蘇蕙璿璣圖康萬民著璇璣圖詩句讀法解人頤.

kind of activity on the part of a Chinese girl, and at the same time gives an example of it, and may thus be regarded as an object lesson on the Chinese Language game.

The reversal of the line implies several transpositions of the parts of speech in the original, almost for every ideogram. E.g. the subject in the first line of the original 羅 'gauze', becomes in the reversed version (last line) the object. What makes this possible is the peculiar nature of Chinese grammar according to which the grammatical category of an ideogram is easily changed according to the syntactic function of the word. Thus an ideogram is a noun, a verb, etc. only predominantly or for most of the time, not altogether and for all of the time, transposition being an essential feature of Chinese sentence structure.

There are really two kinds of transposition implied in the Language Game; the grammatical one of the parts of discourse, and the geometrical one. The result is the reversal in the order of characters within the line.

It is obvious that the sentences are formed with a view to making such transpositions possible, and this is wherein the "game" consists. Only a sovereign command of the language enables one to compose the original poem with a definite view to the transposition necessary for the reversal of characters within a line, and the transposition (in the geometrical sense) of the lines. This makes it clear that transposition is here valued as an art, or at least as an artful device which enhances the enjoyment to be derived from the poem.

I

TRANSFORMATION	ORIGINAL	
合迴別才	永長麗香	(Reading downwards
繡文人女	懷恨錦羅	and from right to left)
綺織幽天	天幽織綺	
羅錦恨懷	女人文繡	
香麗長永	才別迴合	
[V Ƨ Ɔ Ꭰ]	[D C B A]	

II

3rd Transf.	*2nd Transf.*	*1st Transf.*[8]	ORIGINAL
天傳世千	千傳世天	天傳句千	千篇世天
下來人萬	萬來人上	下來分萬	萬分人上
飛錦留愁	詩錦留飛	飛錦章愁	詩字留飛
仙得得成	成得得仙	仙得讀成	成讀得仙
飛留錦詩	愁留錦飛	飛留字詩	愁章錦飛
上人來萬	萬人來下	上人分萬	萬分來下
天世傳千	千世傳天	天世篇千	千句傳天
[V Ƨ B Ꭰ]	[D Ƨ B A]	[V Ƨ Ɔ Ꭰ]	[D C B A]

5th Transf.	*4th Transf.*
傳天篇千	篇千傳天
來下分萬	分萬來上
錦飛字愁	字詩錦飛
得仙讀成	讀成得仙
留飛章詩	章愁留飛
人上分萬	分萬人下
世天句千	句千世天
[Ƨ V C Ꭰ]	[C D Ƨ A]

Table 1

[8] A translation of the poems is given in the Appendix, p. 55.

Translation of poem I in Table I

Vocabulary	*Translation*
香 'fragrant' 羅 'silk' 綺 'gauze' 繡 'to embroider' 合 'to suit', 'harmonize with'	'Fragrant gauze suited for embroidery',
麗 'beautiful' 錦 'brocade' 織 'to weave' 文 'lines', 'characters' 迴 'winding'	'Beautiful brocade woven through with verses in palindrome'.
長 'long' 恨 'sorrow' 幽 'lonely', 'secret' } 'recluse or 人 'man' } refined person' 別 'to part', 'parting'.	'Her unending sorrow is for the refined person who has departed';
永 'forever' 懷 'cherish' } 'passion' 天 'heavenly (gifted)' } 女 'girl' 才 'skill'	'She will cherish for ever her skill as a gifted girl'.

In the transformation the same characters are read in reverse: the translation of the poem is then as follows:

> The gifted girl keeps her passion forever,
> Her secret sorrow for the parted one is endless;
> The palindrome of verses is woven through the brocade's beauty,
> To match the embroidered gauze in its fragrance.

VII

THE LANGUAGE GAME AND THE THEORY OF GROUPS

In order to define the mathematical operations performed in this sort of language game, we must use mathematical symbols. The original poem (I) in Table I has four columns, each representing a sentence containing five characters (words). I shall represent the columns by the letters A, B, C, D, and the inversion of the order of words in a column by inverting the letters for the columns, thus ∀, Ǝ, Ɔ, Ɑ. We begin with eight operations, and indicate the effect of each upon ABCD, the original poem:

S_α: Interchange the first and second, also the third and fourth, BADC.
S_β: Interchange the first and third, also the second and fourth, CDAB.
S_γ: Interchange the first and fourth, also the second and third, DCBA.
S_δ: Stay as you were, ABCD.
D_α: Turn the third and fourth upside down, ABƆƆ.
D_β: Turn the second and fourth upside down, AƎCƆ.
D_γ: Turn the second and third upside down, AƎƆD.
D_δ: Stay as you were, ABCD.

We also use an operator denoted by the sign – which means "turn them all upside down".

We can apply two or more of these operations in succession. For example, $S_\alpha\ S_\beta$ means that, having applied the operation S_β which gives CDAB, we perform on the result the further operation S_α which interchanges the first and second and also the third and fourth. The result is DCBA. This is the same as the result of the single operation S_γ; consequently

$$S_\alpha\ S_\beta = S_\gamma. \tag{3}$$

Applying now the operation —, we can describe the transformation of the poem as being brought about by the operation $-S_\gamma$ on the original poem.

Sometimes, but not always, it makes a difference which of the two operations is performed first. For example:

Taking the result of the operation D_γ, viz. AƎƆD, and performing on it the operation S_α, we obtain ƎADƆ. But taking the result of the operation S_α, viz. BADC, and performing on it the operation D_γ, we obtain BVƆC.

Thus the double operation $S_\alpha D_\gamma$ is not the same as $D_\gamma S_\alpha$. There is, however, a simple relation. ꓭADↃ is obtained by inverting each letter in BVꓒC, that is to say, by applying the operation which we denote by the sign –. Thus

$$S_\alpha D_\gamma = - D_\gamma S_\alpha. \tag{4}$$

Operators related in this way are said to *anticommute*. On examination we find that S_α, S_β commute, and so do D_α, D_β; so also do S_α and D_α. It is only a combination of an S and a D with different suffixes α, β, γ (but not δ) which exhibits anticommutation.

We can make up sixteen different operators of the form $S_a D_b$, where a and b stand for any of the four suffixes α, β, γ, δ. It is these combined operators which chiefly interest us. Eddington (1935) calls them E-operators and denotes them by E_1, E_2, E_3,...E_{16}. They form a *Group*, which means that the result of applying two operations of the Group in succession can equally be obtained by applying a single operation of the Group. In other words: a set of operators such that the product of any two of them always gives an operator belonging to the set is called a *Group*.

The *Theory of Groups* has acquired great importance in theoretical physics. And as Eddington expresses it: "As the result of a game with four letters we have been able to describe a scheme of structure which can be detached from the game and given other applications. When thus detached we find this same structure occurring in the world of physics." To give only one instance: what by the mathematician in the game with four letters is called *anticommutation* of the operators turns out to be the property which the geometer conceives as *perpendicularity of displacements*. This means that we are describing the property of perpendicular directions without using the traditional picture of space.

Returning now to our Table I (I), the original is in our notation ABCD. The transformation has the order of the columns reversed, and is thus DCBA, which is the result of operation S_γ upon the original ABCD. It could also be the result of the operations S_α and S_β carried out in succession, since $S_\alpha S_\beta = S_\gamma$. This shows the different possible operations, which are applied when transforming the original poem, to form a group, according to how a group is defined. Furthermore, since in the transformation every column is turned upside down, the operator – has also been applied, and we can say that the transformation represents the result of the operation $-S_\gamma$ upon the original poem.

The same applies to the first transformation in Table I (II). It represents the result of the action of operator $-S_\gamma$ upon the original ABCD. The second and third transformations, ABꓭD and DBꓭA do not fall into the group pattern, since column C is missing; neither do the fourth and fifth transformations AꓭDC and ꓒCVꓭ. However, this is only a poor selection of the 30 to 40

transformations as variations upon the poem as the original theme, which are possible without adding new words. It would be an easy matter to select those in conformity with the theory of groups.

What can be inferred from all this is that the Language Game consists in letting operators according to the theory of Groups act upon the original poem. Even if this cannot always be performed with the completeness attainable only in the game itself, the identity of the operations is unmistakable. We have thus found the Language Game to be just another application of the mathematical game with four letters, this time not in the world of physics but in that of language.

It is to be emphasised that what we find ourselves using here is a mathematical technique, the theory of groups, which is peculiar to Quantum physics. This is in full agreement with the conclusions reached before (and esp. in *CLO*, Chapters: 10, 11, 18) that if in the quantitative study of the use of vocabulary the aspect of meaning is taken into consideration, the appropriate methods are those of Quantum statistics.

VIII

LANGUAGE GAME AND TOPOLOGY

So far, in the Language Game, the columns as such remained separate and intact. Once a column always a column, and always containing the same characters. The displacement of columns and the reversal in the order of characters left the columns intact.

A more complicated form of the Language Game is one which obliterates the "between" and "within" distinction, i.e. that "between columns" and "within columns". The columns of the derivative poems of the 7×8 matrix of characters–Table II–may be identical with the rows of the original matrix consisting of words from different columns, but in general consist of parts of the original columns read downwards or upside down, and of parts of the original rows read either from right to left or from left to right. The basic matrix of characters is here not a poem, and the derivative poems are now dual pairs: the two outer and the two inner being dual to one another.

Description of the Game

From a 7×8 matrix of characters, four poems of 28 characters each are derived by first *uncoiling* the characters anti-clockwise down to the 28th character, which gives the four columns of poem 1, then continuing in this way from the 29th. character onwards to the end which gives the four columns of poem 2; and then *winding up* the characters in the opposite direction, starting from the middlemost character, i.e. the 56th., which gives the poems 3 and 4 having the same columns as 2 and 1, respectively, but displaced and with the characters in reverse order. In terms of the theory of groups: 3 is the result of applying operation $-S_y$ upon 2, and 4 of applying operation $-S_y$ upon 1.[9]

An appropriate mathematical model for the process of deriving the four poems from the basic matrix seems to be the so-called *Moebius strip*. This is what is known as a *one-sided surface*.

[9] A translation of the poems is given in the Appendix, pp. 55–56.

箋 波 衍

(4)				(3)				(2)				(1)			
嚴	添	樹	簷	階	槐	袖	釵	心	深	悉	陰	迴	梅	篆	開
寒	炭	倚	霜	庭	井	擁	冰	驚	院	悉	陰	文	影	裊	簾
曉	獸	時	綴	點	颭	間	隧	幾	小	寒	雨	鏤	瘦	香	怯
對	爐	看	瓦	雪	風	窗	處	處	窗	風	雪	瓦	看	爐	對
怯	香	瘦	鏤	雨	寒	小	幾	隧	間	颭	點	綴	時	獸	曉
簾	裊	影	文	陰	悉	院	驚	冰	擁	井	庭	霜	倚	炭	寒
開	篆	梅	迴	陰	悉	深	心	釵	袖	槐	階	簷	樹	添	嚴
[A	B	C	D]	[E	F	G	H]	[E	F	G	H]	[A	B	C	D]

Table II

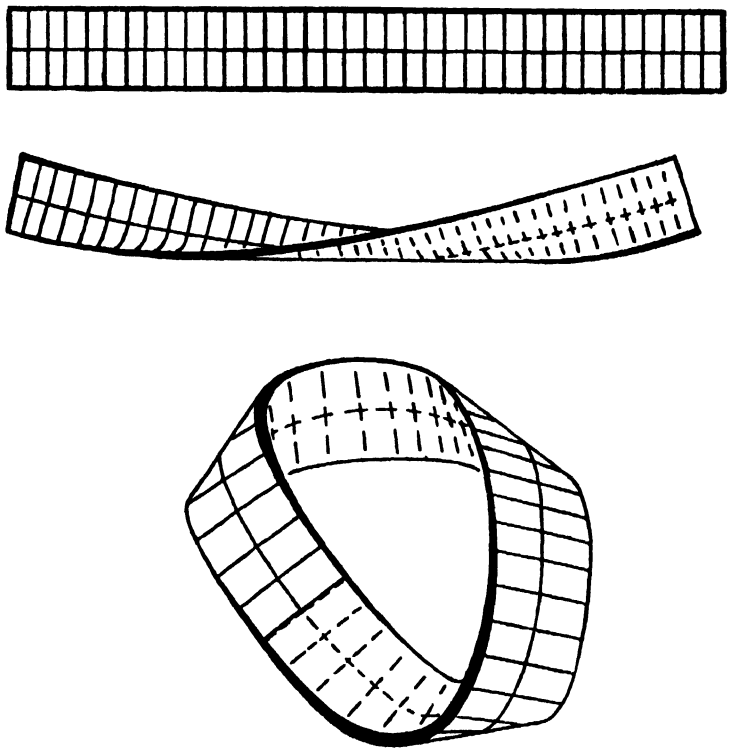

Fig. 4. Forming a Moebius strip.

An ordinary surface has two sides. The two sides of such a surface could be painted with different colours to distinguish them. If the surface is closed, the two colours never meet. If the surface has boundary curves, the two colours meet only along these curves. A bug crawling along such a surface and prevented from crossing boundary curves, if any exist, would always remain on the same side.

Moebius made the surprising discovery that there are surfaces with only *one* side. The simplest such surface is the so-called Moebius strip, formed by taking a long rectangular strip of paper and pasting its two ends together after giving one a half-twist, as in Fig. 4. A bug crawling along this surface, keeping always to the middle of the strip, will return to its original position upside down. The Moebius strip has only one edge, for its boundary consists of a single closed curve. The ordinary two-sided surface formed by pasting together the two ends of a rectangle without twisting has two distinct boundary curves. If the latter strip is cut along the center line, it falls apart into two different strips of the same kind. But if the Moebius strip is cut along this line (shown in Fig. 4) we find that it remains in one piece. It is rare for anyone not familiar with the Moebius strip to predict this behaviour, so contrary to one's intuition of what "should" occur. If the surface that results from cutting the Moebius strip along the middle is again cut along its middle, two separate but intertwined strips are formed.

Now let the distinction between top and bottom of the strip correspond to that between the outside and inside of a column of the basic matrix (Table II), that is, between the columns and the single characters within them. The Moebius strip then means the obliteration of the "column inside–column outside" distinction. When passing from a given column to a character belonging to, and thus being inside of, the neighbouring column, we move like the insect on the Moebius strip. And this may happen at any height of a given column.

But this is exactly what we have been doing when playing this variety of the language game. The columns of the derivative poems 1 to 4 consist each of 7 characters which are partly successive words *within a column* of the basic matrix, and partly words occurring at the same height of *different columns*. The poems are thus composed disregarding the differences of "within" and "between" columns.

This show the Moebius Strip to be the appropriate topological model for this variety of the Language Game, and thus, rather unexpectedly, this branch of Topology as part of our Calculus of language.

IX

LANGUAGE GAME AND REALITY

I am well aware of the Language Game being deprecated by literary scholars, and not for one moment would I think of defending it against such criticism. Nevertheless my contention is that the mere possibility of the Language Game is deeply anchored in the structure of Chinese, and that only a sovereign command of the language would enable one to succeed in the Game.

What decides the quality of a poem is that of the writer as a poet; a bad poet will only produce rubbish, no matter how free and easy his poetic measure, whereas a good poet will create poetry, no matter how many constraints are imposed upon versification. Whatever the criticism of the Language Game as a means of creating poetry, my purpose is not to teach anybody how to write poems in Chinese. It is the exposition of the Calculus of Linguistic Observations, and here I have shown the Theory of Groups (Sect. VII) and a branch of Topology (Sect. VIII) to be part of the Calculus.

Like any other purely mechanical procedure applied to language, the group-theoretical transformation of a poem–even if the original has been designed with such an aim in view–cannot be credited with turning out polished products of linguistic expression, let alone works of art. But the fact of such transformations making sense–and this is undeniable–, and satisfying the native Chinese mind as something worth doing, points to a deeper meaning. There is strong evidence in support of the view that the presentation of a subject from different sides, or angles of view, is the only way for making our mind get a good grasp of it, and so store it in the memory. This activity has as its crude and schematic prototype, or rather as its idealised form, the duality transformations according to the theory of groups. Languages differ from one another with regard to the ease of recognising and using that relation, and so will the different stages of development of a given language. From the data available to us, it would seem that Classical (written) Chinese reveals the relation better than any other language.

The strongest evidence for the view that dualistic transformations are more than a mere game, and, positively, that only by viewing a subject from different angles we are able to make it our own, so to speak, comes from physics. In the course of his analysis of *physical reality*, Eddington arrives at the con-

clusion that it is the synthesis of all possible aspects of nature. This is to be taken literally. By physical reality Eddington means matter, and–quite in agreement with the idealistic philosophers, Berkeley, Kant, Schopenhauer–he has arrived at the conclusion that matter is a creation of the mind. As he expresses it: "matter is the monument erected by the mind to mark the scene of unusual contortions." In simpler terms:

"The immediate perception of the world with one eye is a two-dimensional appearance. But we have two eyes, and these combine the appearances of the world as seen from two positions; in some mysterious way the brain makes the synthesis by suggesting solid relief, and we obtain the familiar appearance of a three-dimensional world. ... The lesson is that a vast number of appearances may be combined into one consistent whole. *Reality is only obtained when all conceivable views have been combined.*"[10]

What corresponds in the sphere of thought–or inner experience–to physical reality or matter, is the emergence of *psychological reality* or thought, in the sense of something that has become the property of our mind, or of ideas becoming our own. As in the physical world, this is accomplished by the synthesis of all conceivable views of the subject. But, as I have shown (*TTM*, Chpt. 17), the feature of symmetry in the perception of the manifold of experience, is in language replaced by the principle of duality. The machinery for this consists in the operations which present the subject from all angles. What is called "The Language Game" of the Chinese consists of such operations, and may be regarded as an automatic application of the very same principle which makes us perceive a physical object as "real".

[10] A. S. Eddington, *Space, Time, and Gravitation* (Cambridge, 1920).

X

THE INTERLOCKING DEVICE AS PART OF THE CRAFT OF FICTION

The combinatory plus interlocking device of syntax has its perfect analogue on the level of literary creation in the closely-knit pattern of relations which forms the nucleus of works of fiction. As an illustration of duality as a device of the craft of fiction, I quote what I said about two major works by Goethe, "Iphigenie" and "Wahlverwandschaften".

> Goethe's "Iphigenie" is permeated with duality. There is the opposition of Iphigenie in Aulis and Iphigenie in Tauris, the first representing the occasion when Iphigenie almost becomes the victim of Diana, the second when her brother Orestes is in the same position.
>
> The earthly relationship of brother and sister between Orestes and Iphigenie is duplicated by its divine counterpart between Apollo and Diana. And there is the cross-relation: the wish to placate Diana was the cause of the threatened sacrifice of Iphigenie, and the wish to please Apollo is the cause of Orestes' implication and threatened sacrifice.
>
> The plot itself is a perfect duality: Apollo's dictum promising Orestes recovery from disease if he brought home the sister from Tauris, which makes Orestes embark on the journey to Tauris, is understood by Orestes as referring to the image of the heavenly sister, Diana, but is found to imply his own living sister, Iphigenie. (*TTM*, p. 277).

> Goethe's "Wahlverwandschaften" starts with the married life of Eduard (E) and Charlotte (C) as an ideal of this type of human relationship. However, soon we learn that each had been married before, and thus been a partner in conjugal life with another companion. Marriage as such is the relation between themselves, and also between each of them with another person.
>
> A more fateful and complete separation of the marital relation from the two individuals, E and C, takes place, or rather is initiated when the "Hauptmann" (H) and Ottilie (O) enter the houshold. E's infatuation with O and C's with H lead to the "marriage through elective affinity" between E and O, on the one hand, and between C and H, on the other,

with the child conceived by C from E as the incarnation of each parent's infatuation with another partner. In appearance, it is the son of O and H. This means the complete duality separation of the original union between E and C: the parents by elective affinity, H and O, have been substituted for the legitimate parents, E and C, without alteration of the essential relation of marriage. Moreover, to make the inverse characteristic of duality complete, the elective father's, H's, contribution comes through C, and the elective mother's, O's, through E, as the parties in the act of conception. (Never mind about what the scientific geneticist with his x and y chromosomes would say to this).

The ever active and rightly so called Mittler may be regarded as the symbol of the conventional view, incapable of separating word and concept. (*TTM*, p. 236).

It is of great importance to realise that the masterpieces of literary art epitomise also the optimum use of language, and in particular of the linguistic device of duality (though this is by no means the only device to be used by the artist). This is in keeping with my contention that the systematic misuse of language is one which has discarded that device (*LCC*, Sect. 15.3; *TTM*, Chpt. XIX A).

THE DEVELOPMENT OF CHINESE VOCABULARY AS A GRADUAL APPROACH TO EFFICIENT CODING

The isomorphy between the binary coding of radical groups in a Chinese dictionary and their stroke number has been established in *CLO*, Chpt. XV on formal, non-linguistic grounds, though it was hinted then that there may be good reasons of a linguistic nature for this phenomenon. I shall now show that the conception of the stroke number of radicals as corresponding today, by and large, to the symbols of binary coding of radical groups in general, and the conception of Chinese writing as a stroke code, receives support from the history of Chinese writing, which again is intimately bound up with the basic structural features of Chinese. If in this part I mention and explain things Chinese well known to the sinologue, it is because I wish to make the contents of this essay accessible also to non-sinologues, the established isomorphy between certain features of alphabetic languages and Chinese being a matter which concerns all structural linguists.

1. The characteristics which give Chinese its peculiar constitution are *monosyllabism* and *isolation*, i.e. it treats the mainly monosyllabic words as isolated units without modifying them according to their function in the sentence. Thus, F.N. Finck (*Die Haupttypen des Sprachbaues*, Leipzig, 1910):

> "There are two peculiarities the co-existence of which gives the structure of this language a very curious character. On the one hand the Chinese words consist, not absolutely always but in the majority of cases, of a single syllable; on the other hand the relation of these monosyllabic words to the whole sentence is not expressed by any marks in the words themselves, but in the first place by a fixed word order, and secondly–and in a less important degree–by the addition of words, the original concrete sense of which is so far faded that they can be used for formal purposes somewhat like our so-called auxiliaries."

With regard to the other characteristic, monosyllabism, the statement that Chinese words are, if not without exception, so yet for the greatest part, monosyllables, must not be misunderstood. In Chinese, as in other languages, we distinguish between simple and compound words (composita), that is,

words consisting of two or more parts, each of which can be used by itself. Chinese is very rich in such compound words. But every simple word consists, as a rule, of a single syllable. In comparison with English, this implies two things: Chinese has no dissyllabic or polysyllabic stemwords, and it has no simple words with derivation affixes. The isolating characteristic, on the other hand, implies that its words cannot be varied by inflexional affixes.

As a result, Chinese words are assembled in a sentence like a set of building blocks of the same size and pattern. That such uniformity has not existed in all phases of Chinese (e.g. in the latter part of the Chou dynasty, 1122–249 B.C.) does not concern us here.

2. Monosyllabism has produced some very remarkable results. The number of pronouncable syllables is not unlimited. The more words the language created, the more difficult it became to prevent them from being similar, if not identical in sound. And this homophony was considerably increased by the Chinese tendency to sound simplification. As far as the development of Chinese can be traced back, we find that tendency. As early as A.D. 500, the language tolerated no more than *one* consonant at the beginning of a word (with the single exception of certain "affricates" (ts, dz, ch, dj) which as phonemes can be regarded as single sounds. In the final sounds of the syllables there was still greater monotony. Just as ancient Greek did not allow other final sounds than a vowel or n, r, s (with x), so every Chinese word at that period had to end in a vowel or in p. t, k, m, n, or ng. As a consequence of these limitations, there were already a great many words in the language which had become identical in pronunciation without having any connection whatever in origin or meaning (B. Karlgren, *Sound and Symbol in Chinese*, London, 1933).

It is true, we find examples of this in other languages. But in European languages instances of this kind are not so numerous as to lead to ambiguity on a large scale. In China, on the other hand, in the period of which we are speaking, the number of homophones must have already become detrimental to the intelligibility of the spoken word. And it was to become still more so. Since that time, the Chinese language has advanced still further on the road of sound simplification, the matter becoming more complicated by the fact that the simplification had not taken place in the same way in all parts of China.

Comparing now modern Mandarin with the language of 500 A.D., we find that the continued process of sound simplification has left very striking traces. Among other things, Mandarin has lost its original p, t, and k at the end of words, which produced many more homophones. As a result, the Chinese language which some five centuries after Christ was already inconveniently rich in homophones has gone from bad to worse as time went on. The modern dialects all have a very meagre stock of phonetically different syllables, with

great numbers of homophones. The mandarin dialect of Peking is one of the poorest with not more than 420 different syllables, many of which are puzzlingly like one another. A foreigner listening to the talk of a Pekingese gets the impression that he has a vocabulary of a few dozen words which he keeps repeating continually.

As a consequence, all the simple words of this highly developed language have to be distributed among these 420 syllables. But the inconvenience of the homophones is considerably alleviated by certain phonetic elements: the musical accents or "tones". Musical accent is a phenomenon that existed in the prehistoric Indo-European language, but it has not survived in the majority of the daughter languages; it is preserved only in Swedish, Norwegian, Serbo-Croatian, and Lithuanian. In Chinese this phenomenon is of great importance. In every Chinese word there is inherent a certain melody, and words otherwise phonetically identical can be distinguished by their different melodies. In Canton, for example, there are no less than nine tones, in the Mandarin of Peking only four: the even tone, the quickly rising tone, the slowly rising tone, and the falling tone.

What is the explanation of this Chinese tendency to sound simplification regardless of the lack of intelligibility of the spoken word, which follows unavoidably in its wake? The answer is not difficult to find. In Chinese it is not the sound which directly leads to the sense of the word, i.e. the concept behind it, but the written symbol, the ideogram. To the Chinese mind, the sound of words is of secondary importance compared with their ideographic differentiation. To develop the system of the ideograms so as to carry every possible meaning, was the main preoccupation of the creative genius of the Chinese language. Intelligibility by the spoken word came in as a bad second. This explanation seems to be preferable to one which would regard the sound simplification as primary and the development of the script as being secondary, an ad hoc affair. But whichever way we prefer to look at it–and for the purpose of this essay we shall not confine ourselves to one hypothesis rather than the other–the fact remains that there is a close inverse relation between the stunted development of differentiation by sound and the highly developed script of Chinese.

3. The development of Chinese vocabulary is dominated by the contrast between the tendency to sound simplification resulting in high degrees of phonetic word uniformity, and the tendency to ideogram variation. Nothing illustrates the latter tendency better than the Chinese classic on painting, 芥子園, 'the Mustard Seed Garden' of the 12th century, in which the pictures of objects in nature, e.g. mountains, trees, are differentiated according to season–a tree in spring, a tree in summer, in autumn, in winter. This is strongly reminiscent of certain Western languages differentiating between verbal forms

according to whether the movement, say, is forward or backward, to the right or the left. What the Western language achieves through variation in sound, Chinese achieves by variation of the picture in the ideogram.

The Chinese script is a very complicated product. First, Chinese characters are, by no means, all formed according to the same principle. There are three main kinds of characters according to the *method* of formation: picture characters, logical compounds and phonetic compounds. Secondly, the writing has, of course, changed technically through the different writing materials, and this has radically changed the *shape* of the characters.

But let us follow the development of the script step by step. The earliest form of Chinese writing consisted in the drawing of objects. A great number of picture characters refer to objects in nature. *Jih*, 日, 'sun' and *yüeh*, 月, 'moon' belong to this class. *Jen*, 'man', is represented by 人 reminding us of Shakespeare's "forked radish". Picture characters were eminently suitable for the representation of concrete objects, and several hundred of them became current. But they were in general unsuitable for the representation of abstract ideas. It is true, a limited number of such ideas could be symbolised, e.g. *i*, 一, 'one', *êrh*, 二, 'two', *san*, 三, 'three', *shang*, 上, 'above', *chung*, 中, 'middle', and the Chinese script has about a hundred of such characters. On the whole it can be said that only a few abstract notions could be satisfactorily rendered by such "indicators". Another method was to let the picture of a concrete object symbolise an abstract idea intimately associated with it. E.g. 'to cross', 'entangle', 'connection', is *chiao*, written originally 爻 and later 交, meaning a man with crossed legs. Yet another method was to borrow the character of a homophone concrete word.

But all these methods, no matter with how much ingenuity they were employed, did not, and could not suffice, when characters had to be devised for the great mass of abstract notions. Further invention became necessary and it took the form of what we call "logical compounds". An obvious form of combination is that of two objects both conspicuous for the quality to be expressed, or possessing it in common. Thus we have *ming*, 明, 'bright', denoted by 'sun and moon'; *chien*, 件, 'individual', 'piece', 'article' expressed by 人, 'man', and 牛, 'ox', exemplifying two individual things.

Another combination is that of two objects whose coexistence would express the idea, e.g. *hao*, 好, 'good, happy', formed by 女 'woman' and 子 'son', indicating a happy relationship.

All these logical compounds of various types, though admirably adapted for expressing words that could not be represented by simple pictures suffered from the disadvantage that they involved the exercise of too much ingenuity of invention–if not arbitrariness. When it became necessary to create characters for thousands of new words, the method broke down because of its difficulty. The Chinese had, therefore, recourse to a new method by means of which new

characters could be invented in practically unlimited numbers. This method led to the variety called "phonetic compounds".

4. It was a kind of phonetic writing, though not as an alphabetic script. The principle of alphabetic writing is to represent the sound of the word without reference to its meaning, and this was entirely alien to Chinese habits of thought. On the one hand, the Chinese were used to the character indicating the meaning of the word, and they were not disposed to abandon this principle entirely. On the other hand, they possessed already, as mentioned above, a few instances of a very primitive phonetic writing, e.g. 𗴂 (萬) 'scorpion', Mandarin *wan*, was used for a homophoneous word '10,000', Mandarin *wan*. But now, perfecting this rather clumsy method they were able to raise it to an ingenious and highly efficient level.

The new characters, the "phonetic compounds", consist of a phonetic portion indicating, or at least suggesting–absolute homophony is not necessary–the pronunciation of the word, and a *significant* part (by Europeans usually called the "radical") indicating, or at least suggesting, the sense of the word or giving the category to which the object belongs. E.g. if we compare the words, 坊, 'district', 紡, 'spin', 訪, 'ask', 鈁, 'kettle', 枋, 'board', we find that they all have the same Mandarin pronunciation, *fang*. In *sense*, the words are far apart, their meanings being indicated by the significant part of the character. In *fang*, 'district', the sense is marked by 土 'earth', in *fang*, 'to spin', by 糸 'silk', in *fang*, 'ask', by 言 'talk', in *fang*, 'kettle', by 金 'metal', and in *fang*, 'board', by 木 'wood'. But in ancient Chinese, the sound of all these words must have been identical with or similar to that of the word 方, *fang*, 'a square', and consequently the latter was inserted into the five characters only to convey that they were pronounced more or less like the word for 'square'.

The phonetic is not necessarily a simple picture character. There are many cases where ideograms in themselves composed of two or more constituents, serve as phonetics in new compounds. Moreover, there is nothing to prevent a character composed of a radical plus phonetic being regarded as a unit and serving as phonetic in a new word.

In this method of composing new characters, the Chinese had discovered a simple and convenient method for creating new ideograms ad lib. By far the greatest part of the Chinese vocabulary–about nine-tenths–are formed in this way.

Observe the fundamental difference between Western phonetic writing and the phonetic compounds of Chinese. In languages with alphabetic script the writing follows more or less closely in the wake of the changes in pronunciation. Now English is notoriously conservative in this respect and sticks to its orthography long after the pronunciation has changed, while German, for instance, has fairly faithfully changed its spelling with changes in pronun-

ciation. In Chinese there has been no modification whatsoever. Once the composition of a character has been fixed, perhaps thousands of years ago, the script was not modified in spite of the extensive changes in pronunciation most of the words have undergone. A consequence of this is that the "phonetic script" of Chinese was very primitive even in ancient times–the phonetic and the derivative seldom being exact homophones.

The two tendencies in the formation of Chinese words, sound simplification and ideogram variation, lead to a very great number of words in the written language, and the need for codification made itself felt at a very early stage of development. Neither of the two aspects of Chinese words, the phonetic or the ideogramic, could provide the ordering element which is necessary for arranging words in a dictionary. Another aspect of the Chinese words was needed for the codification of the vocabulary; or as we say for lexicographic work. And here we come to the third principle in the formation of Chinese words: the stroke number of radicals and phonetics. This is the tendency to stereotype the characters by fixing both the number and order of the strokes needed in writing them. This is used for arranging the vocabulary items in a Chinese dictionary. The basic classification is according to radical and in the order of the stroke number. Under the headings of these radicals all other characters are placed, and they are further arranged with reference to the number of strokes they contain besides the radical.

5. It is very important to realise what the principle of fixed number and order of strokes in a Chinese character meant in the development of Chinese word formation. We have seen how what originally started as pictorial writing has changed in the logical compounds to symbolic writing, in the narrower sense of the term, and how the adoption of phonetic compounds had meant, to a great extent, abandoning pictorial presentation: only the radical of the ideogram remained as a feeble indicator of the sense.

The next step on the road away from pictorial representation was the stereotyping of the characters by fixing the number and order of strokes used in their formation. Such stereotyping could only result in severing the last connection between the pictorial remnant of the ideogram and the sense of the word, because it was no longer the picture of an object which dominated the use of the ideogram but a sort of code of strokes of different number and position. This was no accidental result of using certain writing materials and of the need for dictionary codification, but rather the necessary consequence of the Chinese preference for ideogramic–as against phonetic–representation. As the vocabulary increased, the need for economy in writing had to be satisfied, and the immense variety of ideogram outlines had to be both stereotyped and simplified. The only way for achieving this was the adoption of the stroke code, exactly comparable to the Morse code of telegraphy with which as a,

by and large, binary code the stroke code is in good conformity, as shown in *CLO*, Chpt. XV. Strokes were, of course, always used in writing Chinese characters, but it makes all the difference whether they were used as each scribe thought fit, or whether they were used in a stereotyped manner as a proper stroke code.

In schematic form we may now give the principles governing the development of Chinese word formation as follows:

Homophony of syllable ——→ Variation in Ideograms ——→Simplification and stereotyping of Ideogram shape by adopting a stroke code.

The corrective action of the tendency to ideogram variation against sound simplification, and to ideogram stereotyping against ideogram variation must, of course, be understood to have been exercised partly alongside the development of the tendencies which needed curbing. It should be mentioned that there were also developments in the spoken language itself calculated to curb or contrast the increase in homophonies, and similarly the exuberance of different ideograms was effectively trimmed by language reforms such as that of Ch'in Shih Huang Ti in connection with the "Burning of the Books". But neither of these corrections would have been sufficient, as the subsequent development has shown. Something more radical and lasting was required and provided by the ingenuity in devising new ideograms to counteract the inevitable homophony, and by the development of the script as a system of stereotyped coding symbols, with definite number and order of strokes.

6. *Efficiency of Coding extended to Stroke Number of Phonetic*

It was observed in *CLO*, p. 173, that for completely establishing the stroke number of Chinese ideograms to be in accordance with the principle of efficient coding, it would be necessary to take account also of the stroke number in the phonetic, not only in the radical as we have done so far. This will be done now.

Table III gives the distribution of ideograms in the Chinese dictionary according to the number of strokes of the phonetic. Column 2 was derived from the basic Table 6 of *CLO* (pp. 158–165). It represents the sum of the last columns of that Table.

TABLE III (*LCC*, p. 263)

Distributions of Chinese characters (numbers and relative frequencies) with regard to the number of strokes in the phonetic, according to

No. of Strokes	A) Mathews'[11] Dictionary		B) Fenn's[12] Dictionary		C) Karlgren's[13] Dictionary	
		%		%		%
1	71	0.82	80	1.37	6	0.42
2	241	2.76	219	3.77	38	2.65
3	378	4.34	294	5.07	57	3.98
4	722	8.28	538	9.27	106	7.41
5	**949**	**10.88**	**640**	**11.03**	**122**	**8.53**
6	779	8.93	533	9.14	118	8.25
7	763	8.75	508	8.75	117	9.18
8	**985**	**11.29**	**666**	**11.48**	**144**	**10.07**
9	833	9.55	516	8.90	140	9.79
10	688	7.89	412	7.11	125	8.75
11	590	6.81	367	6.33	114	7.94
12	534	6.12	327	5.64	105	7.34
13	396	4.54	237	4.08	66	4.61
14	246	2.82	148	2.55	50	3.50
15	149	1.71	87	1.50	33	2.31
16	130	1.49	76	1.31	30	2.09
17	89	1.02	47	0.81	19	1.33
18	60	0.68	35	0.60	13	0.90
19	57	0.65	37	0.63	11	0.77
20	17	0.20	11	0.18	7	0.48
21	15	0.17	7	0.12	3	0.21
22	14	0.16	11	0.18	3	0.21
23	2	—	1	—	1	—
24	5	—	—	—	1	—
25	2	—	1	—	—	—
26	—	—	—	—	—	—
27	1	—	—	—	—	—
Total	8711		5798		1429	

[11] R. H. Mathews, *Chinese-English dictionary* (Shanghai, 1931).
[12] C. H. Fenn, *The Five-Thousand Character dictionary* (Peking, 1932).
[13] B. Karlgren, *Analytical Dictionary of Chinese and Sino-Japanese* (London, 1924).

Since a total of 8711 words can be regarded only as a fairly big sample of Chinese words (though by no means as a random sample) the question arises to what extent the relative frequencies could be taken to represent corresponding probabilities. To answer that question, the counting experiment was repeated on another dictionary, the *Five-Thousand* [actually, 5,798] *Dictionary* by C. H. Fenn, containing thus almost 3000 words less then the Mathews' dictionary, or $33\frac{1}{2}\%$ less. But it goes without saying that even the 5798 words whose *number* is common to both dictionaries, will not be the *identical* words in both, since the compilation of a small sized dictionary is very much a matter of selection on the part of its author.

Comparing the two percentage distributions, cols. 3 and 5,we find them very similar.

The similarity of the two distributions is remarkable. A possible explanation of their conformity would be that they depended upon a third distribution which determined, on the whole, their shape. As a distribution underlying that of the ideograms, the distribution of the phonetics as such suggests itself. To test this assumption, a count was made on a dictionary of a different kind, viz., Karlgren's *Analytic Dictionary of Chinese and Sino-Japanese*, which is arranged not according to radicals but according to phonetics, and thus exhibits the frequency of phonetics of different stroke number.

Column 6 shows the frequency distribution of phonetics in the Chinese language according to stroke number. The series of relative frequencies, col. 7, is remarkably similar to those of the ideograms in cols. 3 and 5.

It appears, therefore, that our hypothesis is justified and that the *distribution of ideograms* according to the number of strokes in the phonetic is *determined by that of phonetics* according to the same variable, viz., stroke number. The greater the number of a certain class of phonetics, the greater the number of the ideograms using phonetics of that particular class.

The relation which we encounter here is formally analogous to that between the phoneme distribution of grammatical linguistic forms and the total phoneme distribution (*LCC*, Sect. 6.4). In each case we find a partial distribution imposing its characteristic shape upon the total distribution, be it of phonemes (Western languages) or of ideograms (Chinese).

A remarkable characteristic of these distributions is that they have peaks at 5 and 8 strokes. The reader interested in this particular question of form is referred to G. Herdan, *The mathematical analysis of linguistic behaviour*,[14] where a theory is developed for the explanation of the preference for 5 and 8 stroke phonetics.

The preference for phonetics with 3, 5 and 8 strokes is in agreement with a well-known psychological fact. If we have to count a greater number of similar

[14] G. Herdan, *The mathematical analysis of linguistic behaviour*. Thesis, Library of the Imperial College of Science (London, 1940).

articles, we prefer to count them in groups of 3 or 5, and if the counting has to be done with some speed, often in groups of 8 specimens.

The numbers 3, 5, 8 are also successive terms of the so-called Fibonacci series, if it starts with unity. Each term of the series is the sum of the two preceding ones, and each term divides the distance between its immediate neighbours according to the "Golden Rule" (*De Divina Proportione* of Lucca Paccioli), the ratio of the greater part to the smaller being the same as that of the whole to the greater part. This ratio is, of course, identical with the ratio between any two successive terms of the series, and approaches quickly the limiting value of 1.614.[15]

This shows also the stroke number of the *phonetic* to be governed by the principle of convenience in writing, as required in an effective code.

[15] P. Cattaneo, "Sui numeri di Fibonacci", *Boll. Union e Mathem, Ital.*, 1943.

APPENDIX

To page 29

Translation of Poem II (Table I), original and 1st transformation. The reader could construct for himself transformations 2 to 5, by observing both order and position of the letters A–D.

Original:

The winged fairy up in heaven flies down to earth;
Mortal man is bequeathed a poem woven in brocade:
Each word has its distinct place in the line, and each line in the verse:*
Thousands of poems are thus produced to express endless sorrows.

* (meaning that both, words within a line and lines within a poem, can be combined in many ways.)

1st transformation:

Endless sorrows are expressed in thousands of poems;
Each line reads separately and each word stands distinctly in each verse.
Thus the woven brocade is bequeathed to the human world,
And the winged fairy flies back to heaven.

To page 35

Translation of poems (1) to (4)

(1) Opening the screen, she shrinks from the bitter cold of dawn;
The seal writing-like smoke from the brazier spirals upward when more charcoal is added.
She watches the frail outline of the winter plum, leaning awhile against a tree,
And how the tiles engraved with winding characters ornament the frosty eaves.

(2) Softly, softly the snowflakes light on the stone steps of the front court
Harshly, harshly the freezing wind tosses the locust tree beside the well.
By the little window of the inner apartment she hides (her face) in her sleeve;
Distraught, she lets here and there fall her icy hairpins.

(3) Her hairpins are icy, she drops them here and there, distraught as she is!
She hides behind her sleeve by the window of the little inner apartment.
With the locust tree by the well tossed by the wind, it looks fiercely cold;
The stone steps of the front court are sprinkled with snowflakes falling softly, softly.

(4) The frost on the eaves ornaments the tiles with chiselled winding characters;
Leaning on the tree for a time she gazes at the frail outline of the winter plum.
Adding more charcoal to the brazier the smoke rises up spirally like seal writing;
Facing the bitter cold morning she shrinks from opening the screen.