

## TABLE OF CONTENTS

Preface . . . . .	5
-------------------	---

*BOOK ONE:*  
*LANGUAGE IN THE MASS—THE UNIVERSES OF LANGUAGE STATISTICS*

PART I  
STATISTICS OF PHONEMIC (ALPHABETIC) SYSTEMS

I. FUNDAMENTAL CONCEPTS OF THE CALCULUS OF LINGUISTIC OBSERVATIONS	17
1.1 The definition of the calculus . . . . .	17
1.2 The geometrization of linguistics . . . . .	21
1.3 The universes of statistical linguistics . . . . .	23
1.3.1 Population probabilities of statistical linguistics . . . . .	23
1.3.2 The 'sample space' of statistical linguistics . . . . .	24
1.3.3 The ergodicity of word masses . . . . .	25
II. THE PHONEMIC FREQUENCY DISTRIBUTION . . . . .	29
2.1 The multinomial theorem as the general distribution law of phonemes	29
2.2 The derivation of the entropy from the multinomial theorem . . .	30
III A. THE POSITION OF CHI-SQUARE IN THE SYSTEM OF STATISTICAL LINGUISTICS	32
3.1 The Chi-square distribution as the limiting form of the multinomial theorem . . . . .	32
3.2 The relation between Chi-square and the entropy . . . . .	33
3.3 Advantages of the relation explained in sect. 3.2 for computational purposes . . . . .	34
III B. THE REPEAT RATE OF LINGUISTIC FORMS IN ITS RELATION TO OTHER STATISTICAL PARAMETERS . . . . .	36
3.4 The characteristic $K$ (and $v_m$ ) as the sample statistic of the repeat rate . . . . .	36
3.5 Restrictive condition of the geometrical interpretation of the repeat rate when applied to linguistic distributions . . . . .	38

3.6 An inequality relation between the entropy and the repeat rate (and its sample statistic $K$ ). . . . .	39
IV. TWO MODIFICATIONS OF THE MULTINOMIAL THEOREM . . . . .	41
4.1 The binomial distribution . . . . .	41
4.1.1 The Poisson distribution . . . . .	42
4.2 The lognormal distribution of linguistic units. . . . .	43
4.3 The vocabulary distribution . . . . .	45
V. THE COMBINATORIAL STRUCTURE OF MORPHEMES . . . . .	47
5.1 The patterning of semitic verbal roots . . . . .	47
5.2 Consonant frequencies in Arabic verbal roots . . . . .	47
5.3 Combinatorial analysis applied to the patterning of Arabic verbal root morphemes . . . . .	51
5.4 The relation between the theoretical and observed numbers of root morphemes . . . . .	54
PART II	
VOCABULARY AND STYLO-STATISTICS	
VI. THE STUNTED DEVELOPMENT OF VOCABULARY AND STYLO-STATISTICS THROUGH UNSUITABLE MATHEMATICAL MODELS . . . . .	59
6.1 Theoretical shortcomings of the Zipf and Mandelbrot laws. . . . .	59
6.1.1 The Zipf law . . . . .	59
6.1.2 The Mandelbrot canonical law . . . . .	61
6.1.3 Substitution of 'cost' etc. for 'rank' . . . . .	61
6.2 Shortcomings from the practical angle. . . . .	62
VII. A CRITICAL EXAMINATION OF SIMON'S MODEL OF CERTAIN DISTRIBUTION FUNCTIONS IN LINGUISTICS. . . . .	64
7.1 The assumptions underlying Simon's model as not in accordance with facts . . . . .	64
7.2 Simon's model of certain skew distribution functions tested on linguistic data . . . . .	66
7.3 The conceptual assumptions underlying Simon's model . . . . .	69
7.4 A new model based upon vocabulary connectivity . . . . .	70
7.5 Vocabulary partition as the structure of meaning underlying the word frequency distribution . . . . .	72
VIII. POETRY AND STRUCTURAL LINGUISTICS . . . . .	76
8.1 A critic refuted. . . . .	76
8.2 Variation in Shakespeare's metrical style subjected to stylo-statistical analysis . . . . .	77

8.3 The chronology of Shakespeare's plays and discriminant analysis 79  
 8.4 The recurrence of initial phonemes . . . . . 79  
 8.5 A method for establishing the pattern of repetition of initial line phonemes . . . . . 80  
 8.6 Application to Vergil's *Georgics* . . . . . 83

IX. GENEALOGICAL RELATIONSHIP BETWEEN LANGUAGES—VOCABULARY CORRELATION . . . . . 86  
 9.1 Previous attempts at determination of nearness of linguistic relationship . . . . . 86  
 9.2 Factor analysis as the appropriate method for the problem—the application to Indo-European languages . . . . . 87  
 9.3 Interpretation of results . . . . . 90

PART III  
 SPECTROGRAPHY OF SPEECH ON THE MORPHEMIC LEVEL

X. THE PARALLELISM BETWEEN PHONOLOGY AND VOCABULARY STATISTICS . 95  
 10.1 Phonology and vocabulary statistics: a parallel . . . . . 95  
 10.2 Previous attempts at morpheme analysis . . . . . 96  
 10.3 Independence of the form of word occurrence distribution from content. . . . . 98  
 10.4 Yule's mathematical model for the use of words in a text . . . . 99  
     10.4.1 The quantification of de Saussure's 'valeur' . . . . . 102  
 10.5 Error-reducing device on the morphemic level . . . . . 105  
     10.5.1 Frequency and vocabulary partition . . . . . 107  
 10.6 Common characteristics of the system of phonemic and morphemic oppositions . . . . . 108

XI. SPECTROGRAPHY OF SPEECH (WRITING) ON THE MORPHEMIC LEVEL . . 111  
 11.1 The concept of morpheme spectrography . . . . . 111  
 11.2 The technique of morpheme spectrography . . . . . 112  
 11.3 The method of selective operators . . . . . 116

BOOK TWO:  
 THE RELATION BETWEEN LANGUAGE IN THE MASS AND LANGUAGE IN THE LINE

PART IV  
 THE STATISTICAL THEORY OF LINGUISTIC CODING—INFORMATION THEORY

XII. FUNDAMENTAL CONCEPTS—INFORMATION AS A STATISTICAL CONCEPT . 123  
 12.1 The strategy of inquiry . . . . . 123  
 12.2 Derivation of the entropy as a measure of information . . . . . 124

12.3	Binary coding as optimum strategy of inquiry . . . . .	126
12.4	Binary coding of the English alphabetic system . . . . .	127
12.5	Binary coding of the Cyrillic alphabetic system . . . . .	130
XIII A.	EFFICIENCY OF CODING—THE LAW OF OPTIMAL REDUNDANCY . . . . .	133
13.1	The condition for optimal coding . . . . .	133
13.2	The law of facility for the discrimination of probabilities of linguistic units. . . . .	134
13.3	The logarithmic series in its relation to the 'law of facility' . . . . .	135
XIII B.	BI-VARIATE ENTROPY . . . . .	139
13.4	Bi-variate entropy and the theory of language translation . . . . .	140
13.5	Application . . . . .	142
XIV.	THE ENTROPY AS A DISCRIMINATING PARAMETER—THE RELATION BETWEEN ENTROPY AND LOG-LIKELIHOOD . . . . .	144
14.1	The relation between entropy and log-likelihood . . . . .	144
14.2	Application to assessing the Germanic and Romance components of vocabulary in English writers. . . . .	148
XV.	INFORMATION-THEORETICAL INTERPRETATION OF CHINESE WORD STRUCTURE . . . . .	155
15.1	The relation between Western phonemic (alphabetic) systems and the system of Chinese radicals . . . . .	155
15.2	Binary coding of Chinese vocabulary according to radicals . . . . .	157
15.3	Isomorphy between the binary coding of Chinese characters and stroke number . . . . .	171
PART V		
QUANTUM THEORY OF LANGUAGE		
XVI.	CHANCE, THE EVER-PRESENT ALTERNATIVE . . . . .	179
16.1	Introduction . . . . .	179
16.2	The deterministic view of the use of words and some facts against it . . . . .	180
16.3	Chance, the ever-present alternative . . . . .	181
16.4	Fitting the random partitioning function to the results of empirical vocabulary partitioning. . . . .	184
16.5	Conclusions . . . . .	185
XVII.	VOCABULARY PARTITION AND QUANTUM STATISTICS . . . . .	187
17.1	Common characteristics of Bose-Einstein statistics and the random partitioning function . . . . .	187

17.2 The random partitioning function and transition probabilities  
 between words . . . . . 189  
 17.2.1 Vocabulary association. . . . . 189  
 17.2.2 Vocabulary association from experimental data . . . . . 190  
 17.2.3 Vocabulary association due to random sampling. . . . . 191  
 17.2.4 Application of random partitioning to our data . . . . . 194  
 17.2.5 Vocabulary association and Random Partitioning . . . . . 196  
 17.3 'New Statistics' vs. classical statistics . . . . . 197

XVIII. THE ENTROPY ACCORDING TO BOSE-EINSTEIN STATISTICS AND THE IN-  
 FORMATION VALUE OF WORDS . . . . . 199  
 18.1 The bi-variate distribution of words from a running text . . . . . 199  
 18.2 The dynamics of vocabulary distribution . . . . . 200  
 18.3 The entropy according to Bose-Einstein statistics. . . . . 201  
 18.4 Word counts in their relation to vocabulary, word association and  
 grammar . . . . . 205

XIX. THE QUANTUM THEORY OF LANGUAGE. . . . . 208  
 19.1 Characteristics of the word frequency distribution and quantum  
 numbers . . . . . 208  
 19.2 Word frequency and second quantum number. . . . . 210  
 19.3 The units ('bits') of indeterminacy per word . . . . . 211

*BOOK THREE:*  
*LANGUAGE IN THE LINE—THE UNIVERSES OF LANGUAGE DISCOURSE*

PART VI  
 THE UNIVERSE OF DISCOURSE AS A LINE—LINGUISTIC DUALITY

XX A. DISTRIBUTION IN THE LINGUISTIC SENSE . . . . . 219  
 20.1 The general structure of linguistic distribution . . . . . 219  
 20.2 Relation between the planes of content and expression . . . . . 219  
 20.3 The Law of duality as operating in both planes . . . . . 220  
 20.4 Distribution and likelihood . . . . . 222  
 20.5 Distribution vs. classical phonology . . . . . 223

XX B. THE UNIVERSE OF DISCOURSE AS A LINE . . . . . 226  
 20.6 A one-dimensional universe. . . . . 226  
 20.7 Duality as the principle of grammar . . . . . 227

XXI. DUALITY AND SEPARATION. . . . . 231  
 21.1 Phonological opposition as separation . . . . . 231  
 21.2 Grammar as separation . . . . . 232

21.3	The restless universe of language . . . . .	234
21.4	The importance of pattern in Goethe's 'Wahlverwandtschaften'. . . . .	236
21.5	Duality and Poetics . . . . .	237
<b>XXII.</b>	<b>THE CALCULUS IN RETROSPECT—A SUMMARY . . . . .</b>	<b>240</b>
22.1	Geometrisation in the plane of expression. . . . .	240
22.2	The 'distance' between the planes of content and expression as a criterion for the classification of languages . . . . .	246
 <b>PART VII</b> <b>LANGUAGE AND MATHEMATICS</b>  		
<b>XXIII.</b>	<b>ON THE RELATION BETWEEN LANGUAGE AND MATHEMATICS. . . . .</b>	<b>249</b>
23.1	The Eigenstatistics of language . . . . .	249
23.2	Apriority of language . . . . .	251
23.3	The linear order of linguistic forms as the prototype of the logarithm . . . . .	252
23.4	Eigenstatistics as the error-reducing device of language . . . . .	254
23.4.1	Transition probabilities between letters and phonemes . . . . .	255
23.4.2	Association between words as error-reducing device. . . . .	256
23.4.3	Language deterioration through systematic neglect of error- reducing devices. . . . .	256
<b>XXIV.</b>	<b>MATHEMATICAL LINGUISTICS VERSUS LINGUISTIC PHILOSOPHY . . . . .</b>	<b>259</b>
24.1	The decay of English philosophy . . . . .	259
24.2	The tactics of language . . . . .	260
24.3	The principle of opposition in language on the phonemic, mor- phemic and syntactic level . . . . .	260
24.4	Dictionary meaning of words as precipitate of their use in all works of literature. . . . .	261
24.5	The corrective action of the principle of opposition . . . . .	262
24.6	The lack of style in linguistic philosophy . . . . .	263
	<b>Bibliography . . . . .</b>	<b>265</b>
	<b>Index . . . . .</b>	<b>268</b>