Pleased Keturn to M. Hally Sonderdruck

KRATYLOS

KRITISCHES BERICHTS-UND REZENSIONSORGAN FÜR INDOGERMANISCHE UND ALLGEMEINE SPRACHWISSENSCHAFT

JAHRGANG III. HEFT :

1958

OTTO HARRASSOWITZ · WIESBADEN

KRATYLOS

Kritisches Berichts- und Rezensionsorgan für indogermanische und allgemeine Sprachwissenschaft Im Auftrage der Indogermanischen Gesellschaft herausgegeben von GEORGES REDARD

	V. Nachrichten Kongreß Persönliche Nachrichten	IV. Nekrolog H. H. Schaeder † (W. Eilers)	I. Forschungsbericht Tovar, Antonio: Das Keltiberische, ein neuer Zweig des Festlandkeltischen II. Rezensionen Terrand, G.: Leaguage as Choice and Chance (M. Halle) Literard, G.: Die indoeuropäischen Personalpronomina und die Laryngaltheorie (A. Debrunner) Birwe, R.: Griechisch-Arische Sprachbeziehungen im Verbalsystem (M. Manthoffen) Etudes Myceniennes, Actes du Colloque international sur les textes myceniens (A. Heurbeck) Pangalog, G. E.: Hegitovinianouxoviliouaurog tij Kojirg, I (S. C. Caranzas) Pangalog, G. E.: Hegitovinianouxoviliouaurog tij Kojirg, I (S. C. Caranzas) Nieddenanische Stemmeskunde (E. Schwennungen) Kahler, H.: Gremmetik der Bahasa Indonésia (J. Gonda). Ketturnen, L.: Die Herkunft des Terminativs, Genitivs, Instruktivs und Komitativs (H. Fronk) Ketturn, L.: Terminologie grammaticale du sanskrit (A. Debrunner) Renou, L.: Terminologie grammaticale du sanskrit (A. Debrunner) Petropoulos, D.: La comparaison dans la chanson populaire greoque (S. C. Caratzas) Manier, A.: L'évolution phonétique et les sons du latin ancien (R. Peister) Petropoulos, D.: La comparaison dans la chanson populaire greoque (S. C. Caratzas) Keahe, H.: Germanische Sprachwissenschaft (S. Sondereger) Wolf, A.: Wörterbuch des Rotwelschen (E. Schwenner) Petropoulos, A.: Wörterbuch des Rotwelschen (E. Schwenner) Petropoulos, B.: Survey of the Uralic Languages (T. Sköld) Kealer, B.: Survey of the Uralic Languages (T. Sköld)	ricit 1
1	94	822	88 778 666 67 578 678 678 678 678 678 678 678 678 678 6	

Zuschriften, die die Redaktion betreffen, werden erbeten an Prof. Dr. Georges Redard,
Junkerngasse 23, Bern/Schweiz.
Besprechungsexemplare bitte nur an den Verlag: Otto Harrassowitz, Wiesbaden, Friedrichstr. 14, senden.
Die Zeitschrift erscheint zweimal jährlich mit je einem Heft mit ca. 96 Seiten Umfang, Jahresabonnement 20,— DM; Einzelheft 11,— DM. Die Mitglieder der Indogermanischen Gesellschaft erhalten die Zeitschrift gegen Zahlung ihres Jahresbeitrages an die Gesellschaft.

Gedrucks mis Unterstilleung der Deutschen Forschungsgemeinschaft

hoff, 1956, 8°, XVI, 350 S. Geb. 8\$. Herdan, G.: Language as Choice and Chance. Groningen, P. Noord-

main stream of linguistics as a science. This may seem surprising, on the past; the results obtained have, however, remained outside the dealing with data that have been obtained by a repetitive operation, first sight, since "statistical methods are essentially methods for niques to phenomena of language. Similar attempts have been made in 1.0 The book under review is an attempt to apply statistical tech-

1947, 1. ¹ P. G. Hoel, Introduction to Mathematical Statistics, New York-London

> ces of all kinds—are "obtained by a repetitive operation." and it appears almost self-evident that linguistic phenomena—utteran-

poetical works have been investigated from a quantitative point of 2. A much smaller body of data has been collected on phoneme- or been primarily of three types: 1. A considerable amount of work has letter-frequencies in spoken and written discourse. 3. A fair number of been done on the frequency of occurrence of words in different texts. that have been studied from a quantitative point of view. These have is necessary to inquire into the nature of the linguistic phenomena In order to understand the reason for this paradoxical situation it

set of abstract statements. model like pitching pennies or rolling dice, than those of the equivalent in statistics, for it is often easier to grasp the consequences of a physical presented in terms of physical models, which is a common procedure with respect to word frequencies1. These hypotheses have been Attempts to develop statistical hypotheses have been made primarily with the observed facts (e.g., the word frequency distributions). mine an assumed distribution" (ibid. 201-202). In other words, the more, or a relationship between two or more, of parameters that deteradvanced which is subject to statistical tests. "A statistical hypothesis tools of significant power come into play only when a hypothesis is and utilize statistical techniques to a very modest extent. Statistical hypothesis determines some sort of distribution which can be compared is usually regarded as a statement which specifies the value of one or The overwhelming majority of the above studies are simple counts

cribed in statistical terms, for not every true statement concerning sufficient to demonstrate that some aspect of language can be desthat matter, the relevance of any new research method—it is not prove the relevance of statistical investigations for linguistics—or, for language is necessarily relevant to the science of linguistics Before examining these models, we must note that in order to

statistical interest for his study. Acad. Sees St. Pétersbourg 7, 1913. It is to be noted that Markov claimed only du roman 'Eugène Onègin', illustrant la liaison des épreuves en chaîne, Bull letter frequencies is A. A. Markov's Essai d'une recherche statistique sur le texte logy, Bloomington Ind., 1955, 3—14. — Perhaps the most important study of de communication, Paris, 1953. — H. A. Simon, On a Class of Skew Distribution Functions, Biometrika 42, 1955, 425—440. — C. F. Hockett, A Manual of Phono-Mass., 1949. — B. Mandelbrot, Contribution à la théorie mathématique des jeux ¹ G. U. Yule, The Statistical Study of Literary Vocabulary, Cambridge, 1944. G. K. Zipf, Human Behavior and the Principle of Least Effort, Cambridge

Linguistics, like any other science, is a set of systematically connected statements. (It is the insistence on such systematic connectedness that differentiates scientific collections of facts from agglomerations of odd bits of curious information). If statistical or other mathematical tools are to be employed in linguistics it is essential to show that these tools can handle problems of significance for linguistics, and not just isolated facts of language. This has never been demonstrated. As a matter of fact, I shall endeavor to show that it could not have been done because the statistical models proposed are inadequate for purposes of linguistic description.

The models which underlie all statistical hypotheses of word distribution are essentially urns from which words are selected by the user of the language. A good description of a model of this type is given by Yule in his interesting and valuable book: Let us assume all words in a writer's treasure-chest to be written on tickets: some—the words most rarely used—on only one ticket, some on two, some on three, and so on; some few, words very frequently employed, on several hundreds or thousands of tickets each. We regard the process of writing (it does sound rather ridiculous put in this way) as bearing an analogy to the drawing of tickets out of this chest... (o. c. 48—49).

generating, among others, sentence-types like "The man who said for linguistics, because without them it is impossible to give a rule for generating the utterance. These abstract symbols are, however, crucial which are replaced by concrete word sequences in the process of such abstract intermediate symbols as "sentence", "phrase", ,,verb" only concrete sequences of words and has no provision for introducing of ruling out this preposterous consequence. Even this, however, would scrambled up in any one of a multitude of conceivable ways¹. It is of words occurr in inverse order, or one in which the same words are that ... was still there"; or "If ..., then ..." where "..." stands for lack of abstractness². The treasure-chest model is capable of generating has pointed out, all models of this kind are vitiated for linguistics by a not save a treasure-chest model for linguistic purposes, for, as Chomsky transitional probabilities between words, which would have the effect possible to meet this objection, as Hockett has done, by introducing distinguish a normal English book from one in which the identical set Such a model is fundamentally a negation of syntax, for it does not

any grammatical sentence or sequence of sentences of English regardless of length." 1

For syntax the treasure-chest model is, thus, inadequate as it cannot account for perfectly normal types of utterance. The model is equally inadequate for morphology and phonology, since the description of the utterance in terms of morphemes and phonemes also requires abstract, intermediate symbols². In sum, as far as the three traditional subdivisions of linguistics are concerned, the model is of little value.

The above is not to be construed as a blanket condemnation of statistical studies of language. I do not rule out, by any means, the possibility that statistical hypotheses of deep significance for linguistics may yet be discovered. Nor is it my intention to deny merit to the cited studies, all of which make contributions to our knowledge either by providing new data or by extending statistical theory in important new directions. My criticism is strictly limited to the point that word frequency distributions and the various models proposed to account for them have no demonstrated relevance to linguistics. In the trenchant words of Professor Bull: Neither the number of nouns in the Oxford Dictionary, nor the frequency with which any English nouns is used, are basic linguistic facts. The total number of English nouns is a manifestation of the technological and cultural advancement of speakers of English . . . and the frequency of a noun like aspirin, for example, is simply a reflection of the headache rate of these speakers.

1.1 My major objection to H.'s book is that the question of what relevance his work has for linguistics is never seriously posed. From the perfunctory way in which this problem is treated it would seem that H. was not even aware of its existence. This impression is further borne out by H.'s basic model, which he describes as follows:

¹ I owe this example to R. B. Lees.

² For details, see N. Chomsky, Syntactic Structures, 'S-Gravenhage, 1957, ch. 3; and also his review of C. F. Hockett o.c., IJAL 23, 1957, 23—234.

different words, contains tickets with different letters and a special space symbol — a word being defined as a sequence of letters ending in a space—the Zipf-Mandelbrot straight-line relationship can be derived as a consequence of the fact that words are sequences of letters ending in a space. This is yet another illustration of the utter irrelevance of word counts for linguistics. Cf. G. A. Miller, Some Effects of Intermittent Silence, Amer. Journ. of Psych. 70, 1957, 311—314, which contains the first clear exposition of this result due to Mandelbrot.

² With regard to phonology this point was made by Einar Haugen, The Syllable in Linguistic Description, For Roman Jakobson, 'S-Gravenhage, 1956, 213—221, where it is convincingly demonstrated that the syllable is a unit on a higher level of abstraction than the phoneme or the phone, and that consequently the syllable need not necessarily be marked phonetically.

³ W. E. Bull, Problems of Vocabulary Frequency and Distribution, mimeo-

³ W. E. Bull, Problems of Vocabulary Frequency and Distribution, mimeographed paper presented at the First M.I.T. Conference on Mechanical Translation, June 17—20, 1952.

If we regard language to be the total word-engrams ('la langue' in de Saussure's terminology) plus their probability of occurrence in individual speech, and this as the various ways in which the event 'engram' could happen, together with the relative frequencies of the different engrams in actual use, the conception fulfils all the requirements of what is known as the statistical population of such events, or their statistical universe. Any individual speech utterance ('la parole' in de Saussure's terminology) plays then the role of a sample from that population (3).

A careful reading reveals that this differs from Yule's "treasure-chest of tickets" only in its abstruse language. For H., though not for de Saussure, 'la langue' consists of words (or concepts, which for H. are synonymous with words) with associated probabilities of occurrence, and the mechanism of sentence generation is then identical with that of "drawing of tickets out of the chest". (This is the only meaning that can be attached to H.'s equating an utterance with "a sample from that population".) In view of this all criticisms directed against the as totally unproven H.'s major thesis that "literary statistics (are) an of Statistical Language.

2.0 Statistical techniques of essentially the same type as those just discussed have been applied in the study of second-order effects of language, specifically the structure of metrical verse and the nature of literary style with special reference to questions of disputed authorship. Since H. devotes considerable attention to these problems—cf. part I of his book which is titled "Stylostatistics"—I shall examine them in some detail.

2.1 Metrical verse is produced by imposing an organization on the distribution of one or more linguistic elements—e.g., on the placement of word accents or of word boundaries—which in prose are distributed according to no readily discernible principle, but are rather subject to the influence of a multiplicity of undefined factors. Consequently the distribution in verse of the element(s) in question differs sharply from that in prose. Comparison with the prose (random) distribution—i.e., with the distribution produced by a treasure-chest model—may allow us at times to see these specific deviations more clearly.

Counts and statistical investigations are particularly useful in cases where there is doubt as to the nature of the metrical pattern. Thus, for instance, the study of H.N. Porter, which H. discusses at some length, is primarily an attempt to show that the Greek hexameter is structured not only in terms of the distribution of long and short morae, but also in terms of the distribution of its word boundaries. The latter has been questioned by scholars on the grounds that word boundaries are often not marked phonetically. Porter proves his point by demonstrating "that word-ends occur in certain positions in the line much more

frequently than would be expected *ceteris paribus*." Porter's demonstration might have been made impeccable from a statistical point of view if he had also calculated the frequency of occurrence of word boundaries at different places in the line on the assumption of a random distribution of word lengths. In his case, however, this was not of very great practical importance since the raw data themselves clearly bear out Porter's thesis.

H., therefore, misses the essential point of Porter's study when he states that "the regularity of the occurrence of caesurae in early Greek hexameter may be explained by the influence of grammar as a form element". (125). Since grammar governs prose as well as verse, this statement makes sense only if it is accompanied by a demonstration of how grammar functions differently in prose than in verse. There is no such demonstration in Porter's book, since that was not his hypothesis, and there is also none in H.'s. Instead, H. states that "the preference for caesurae after 3, $5 (5^{1})_{2}$ and 8 syllables is in agreement with a well-known psychological fact. If we have to count a greater number of similar 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". (129).

It is hardly necessary to say that the appearence of the numbers 3, 5, 8 in

to is narroly necessary to say that the appearence of the numbers 3, 5, 8 in two disparate sets of observations proves exactly nothing. The same numbers appear also in the arrangement of leaves on plants³, and as far as I know biologists do not believe that this has any connection with the "well-known psychological fact" just mentioned or vice versa.

2.2 In his discussion of the Romance vocabulary of Chaucer, H. follows Yule in noting that the percentage of words of Romance origin in the vocabulary of a text depends on the length of the text. This can be predicted from the treasure-chest model, since the percentage in the vocabulary of the rarer words—in Chaucer, words of Romance origin are rarer than words of Anglo-Saxon origin—increases with the number of chances the author has at drawing tickets bearing different words (Yule, o.c. 140ff.).

H., however, goes farther than Yule. He suggests that the formula

 $p_v = 10 \log_{10} N$

describes the relationship between the percentage of words of Romance origin in the vocabulary (p_v) of a text and the number of running words (N) in the text, and uses the formula as the basis for attributing to Chaucer a certain astronomical treatise $(17-22^4)$.

In order to be useful in settling questions of disputed authorship the formula must yield values in substantial agreement with those observed when applied to works known to be by Chaucer. Serious deviations in more than a few instances cannot be tolerated. The formula fails to satisfy this requirement, as can be seen

¹ H. N. Porter, The Early Greek Hexameter, Yale CiStud 12, 1951, 10

² Actual calculations of this sort in connection with studies of metrical verse were made by B. Tomaševskij in a number of very instructive essays collected in his book O stixe, Moscow, 1929; see particularly 100—107.

³ D'Arcy W. Thompson, On Growth and Form, Cambridge, 1952, 912—933.
⁴ This passage is a summary of H.'s paper Chaucer's Authorship of the Equatoric of the Planetis', Lg 32, 1956, 254—259.

M. Halle: Herdan, Language

from the following table, which contains the observed and predicted values for about a sixth of the works by Chaucer listed in H.'s Table 3. Indeed, inspection of H.'s Fig. 18 reveals that the observed data are scattered widely and cannot be adequately represented by the straight line he proposes. Calculation of any one of a number of statistical indices of goodness of fit yields the same negative result.

10. Wom. Onc. 173 31.94 21.96 10. Wom. Nob. 225 41.59 23.52 11. Bukton 264 15.45 24.21 16. Fort. 576 39.04 27.60 18. Am. Comp. 744 20.33 28.72 20. Comp. L. 1018 17.76 30.01 30. A. Rv. 3308 21.71 35.19 39. A. Mil. 5140 26.67 37.11 25.59 1 Pars 50.679 20.79 39.34 38	observed predicted	$egin{array}{ccccc} Work & Length & egin{array}{ccccc} ^{0/_0} & Romance & Nu \ Vocabulary & (p_v) & Wo \end{array}$
1.96 27 19 8.52 47 27 4.21 19 30 7.60 98 69 8.72 50 71 1.01 54 91 5.19 165 268 7.11 271 377 3.34 332 625 1.72 1128 625	dicted observed ² predicted	(v_p) Number of Romance Words in Vocabulary

indices of style, the quantity $v_m \cdot v_m$ is said to be "a measure of the 'langue-parole' duality" and is supposed to decrease with time for any given language, "as economy in vocabulary decreases with time." (o. c. 41). To substantiate this offers two tables containing a total of eight data points. (Table 12d, 40 and of different periods:

The second table is titled "Values of the relative fluctuations of the mean vm, for English writings ranging from the 1st century A.D. [sic.] to the 19th century (calculated from Yule's K values)". As an example of 1st century of 14th—15th century English, the Latin authors Thomas à Kempis and Gerson.

There is no need, either for a detailed discussion of correlations in the vocabulary of different authors as an index of style. The conclusions which H. is able to draw from an elaborate statistical study of the vocabulary of Churchill, Halifax, Stresemann, Beneš, Hitler and Stalin are so tame 3 that any labor expended over them is a clear waste of time.

The value of type-token ratios as characteristics of the richness of an author's vocabulary has been seriously questioned by Yule, who noted that this ratio in general would not correspond to the intuitive concept the ratio is intended to explicate (o.c. 97). H. does not advance any arguments to meet this objection.

3.0 About a quarter of the book is devoted to a discussion of what H. terms 'linguistic duality'. Since the idea is apparently original with H. and since a large part of the argument of the book rests on this notion, I quote a few relevant passages in extenso.

"Linguistic duality as a fundamental principle of the use of words or concepts has the character of a binary code not only because it uses two concepts only for the explanation of a great multitude of concepts,—just as in the binary system two elements only are used in combination to replace the series of phonemes, say,—but also because the usefulness or appropriateness of the two concepts used in the explanation of all other concepts concerned has its counterpart in the efficiency of the binary code" (219).

"In judging the code quality of a language ... the multiplicity of duality relations which characterise the language structure must be taken into account. In the alphabetic languages, we have the duality of phoneme and letter, of phoneme and condensed writing (shorthand), and in the non-alphabetic languages, that of sound and symbol ... The richer the structure of a language in dualities the more capable will it be of describing the manifold adequately without onomatopoetic representation in the widest sense of the term. On the contrary the more complex the duality structure, the more will the language be felt as something different from what it describes; it will be valued as a creation of its own, obeying laws of its own. Conversely it is a sign of stunted development in a language when certain fundamental dualities are missing, and of retrograde development if the language has lost some such previously used dualities" (287—288).

These theoretical considerations are cited by H. in explanation of among others, the following "facts":

- a) that, "as the pattern of relations of the outer and inner life grew more complex, the Semitic languages ceased to be adequate tools for linguistic expression of these relations" (288).
- b) that Chinese "although a perfect instrument for lyrical expression has severe shortcomings as an analytical tool for expressing scientific relations" (289).
- c) that none of "the European languages which can boast of a one-to-one correspondence between alphabetic and phonetic symbols such as Spanish, [sic!] Finnish, Czech, have developed a great literature" (144).
- d) that "if we were to interpret what has happened to the German language under Hitler in the light of linguistics, we would say that his work was the obliteration of the fundamental langue-parole duality" (293).

The straight line drawn in Fig. 18 does not represent H.'s formula. The correct line has a smaller slope and passes through the points x = 1, y = 10 and x = 4, y = 40. The same mistake can be found in the figure in H.'s article in Lg.

² Number of vocabulary items of Romance origin from J. Mersand, Chaucer's Romance Vocabulary, Brooklyn, 1937, 75—77, which is also the source of H.'s data.

³ E.g., "Churchill was typical in the use of words constituting the 'political vocabulary' as gathered from the *samples* taken from the six persons". ibid. 57. The rest of the conclusions can be found on 57—59.

4.0 It is my conviction that linguistics, like every science, has much to gain from the utilization of proper mathematical techniques. I believe that future progress in linguistics is dependent on the introduction and wide application of mathematical tools. It is a much deplored fact that the sophistication of most linguists in matters mathematical leaves much to be desired. Thus some of us regard anything in mathematical garb as if it were a revelation from on high, into accepting conclusions he is not in a position to verify. Crude attempts at applying unsuitable mathematical techniques to irrelevant problems, therefore, not only waste the time of those who carry out the research, but also increase the resistance among linguists to developments that represent genuine progress.

Cambridge 39, Mass., USA Massachusetts Institute of Technology

Morris Halle

NEUERSCHEINUNGEN

Peter Brang Puškin und Krjukov

Zur Entstehungsgeschichte der "Kapitanskaja dočka« (Slavistische Veröffentlichungen, Band 14)
1957. 95 Seiten, broschiert 14,— DM

WOLFGANG KASACK

Die Technik der Personendarstellung bei Nikolaj Vasilevič Gogol (Bibliotheca Slavica)

1957. VIII, 170 Seiten, broschiert 16,— DM

WILHELM LETTENBAUER
Russische Literaturgeschichte
2., vermehrte und verbesserte Auflage
1958. Ca. 350 Seiten, Leinen ca. 22,— DM

ROSEMARIE RICHHARDT

Polnische Lehnwörter im Ukrainischen
(Slavistische Veröffentlichungen, Band 15)
1957. XII, 122 Seiten, broschiert 20,— DM

HEIDELBERGER SLAVISCHE TEXTE

Herausgegeben von Dimitrij Tschižewskij und Johann Schröpfer

Aus den verschiedenen slavischen Sprachen werden in jedem Heft zu einem bestimmten Thema Texte geboten, denen knappe Vorbemerkungen und Fußnoten sowie teilweise auch die Übersetzungen beigegeben sind

Heft 1

Russische literarische Parodien

1957. 71 Seiten, kl.-8°, broschiert 5,80 DM

Heft 2

Berufung und Bestimmung des Dichters in der slavischen Dichtung

I. Von den Anfängen bis zur Romantik 1957. 45 Seiten, kl.-8°, broschiert 4,— DM

In Vorbereitung befinden sich: Formalistische slavische Dichtung; Selbstkritik der Slaven; Petersburg und Moskau in der russischen Dichtung; Gray's Elegien in den slavischen Übersetzungen.

OTTO HARRASSOWITZ · WIESBADEN

NEDERSCHEINDNGEN

HANS I. COTTSCHALK AI-Malik al-Kāmil von Egypten und seine Zeit

Eine Studie zur Geschichte Vorderasiens und Egyptens in der ersten Hälfte des 7./13. Jahrhunderts 1958. X, 256 Seiten, broschiert ca. 26,—DM

Fritz Heuser · İlhami Şevket Türkisch-Deutsches Wörterduch

4., unveränderte Auflage verfaßt und herausgegeben von Fritz Heuser 1958. XVI, 618 Setten, Leinen 44,— DM

MIGAX ZHMHMM

Gedenkschrift Paul Kretschmer (2. Mai 1866—9. Marz 1956) Herausgegeben von Heinz Kronasser Band 2. 1957. XL, 256 Seiten und 1 Tafel, Leinen 40.— DM Noch lieferbar: Band I. 1956. VIII, 236 Seiten mit 8 Abb. und 1 Taf., Leinen 36.— DM

KURT MEISSMER TINGANGSSPIACHO TINGANGSSPIACHO LINGANGSPIACHO S., unveränderte Auflage

25., unversance Aunage 25., unversance Aunage 1958. XVI, 132 Seiten, broschiert 14.— DM

Grand Sorter Grammatik und Lesebuch der neugriechischen Volks- und Schriftsprache

4., revidierte Auflage 1958. 157 Seiten, broschiert 12,— DM

Uленсы Sohmoll Die Vorgriechischen Sprachen Siziliens 1958. Ca. 1≜0 Seiten mit zahlreichen Textabbildungen, broschiert ca. 22,— DM

SIBVEIS

Festschrift Hans Kraht zum 60. Geburtstag am 7. Februar 1958 dargebracht von Freunden, Schülern und Kollegen 1958. VIII, 315 Seiten und 1 Tajel, Leinen 36,— DM

OTTO HARRASSOWITZ · WIESBADEN