BOOK REVIEWS

The Psycho-Biology of Language. Pp. ix + 336 and 7 plates. By GEORGE K. ZIPF. Boston: Houghton Mifflin Co., 1935.

This is the third publication by Zipf on the theory and application of his principle of relative frequency in the structure and development of language. His first treatise on the subject[1] applied the theory to accent and phonology, and laid the foundation which he still uses with inconsiderable modifications. It was reviewed, in a generally sympathetic spirit but with regretful refusal to agree that the problems attacked had been solved, by Elise Richter[2] and by Eduard Hermann.[3] The second book was shorter.[4] In it Zipf presented his study of the vocabulary of four Plautine plays, and a phonetic, syllabic, and vocabulary study of 20,000 syllables of connected text in the colloquial Chinese of Peiping; there were hints of a contemplated extension into the field of morphology. It was reviewed by Eduard Prokosch[5] with a severity that bespoke a painful disappointment of his hopes for the possibilities of the new line of inquiry. In the present volume Zipf embraces the whole range of linguistic study and phenomena, from phonemes to 'the stream of speech and its relation to the totality of behavior'. Apparently nothing remains untouched within that range, and the treatment almost uniformly evidences a belief that the author has attained valid formulations. Further, the book is subtitled 'An Introduction to

[1] Relative Frequency as a Determinant of Phonetic Change, Harvard Stud. Class. Phil. 40.1–95 (1929).

[2] ASNS 157.291–6 (1930).

[3] PhW 51.598–603 (1931). Other reviews: Kent, LANG. 6.86–8, 'not convinced . . . despite the statistics'; Meillet, BSL 31. 3.17 (1931) (not available, but Meriggi, IdgJb 16, reports 'ablehnend'); Meriggi, IF 50.246–7 (1932), 'calls attention to neglected factor' but 'exaggerated, almost mechanical utilization' and 'paper phonetics'; Sütterlin, LGRPh 52.241–3 (1931), appreciative; Twaddell, Monatshefte f. deut. Unterricht 21. 230–7 (1929), appreciative and constructively critical.

[4] Selected Studies of the Principle of Relative Frequency in Language (Harvard Univ. Press 1932) viii + 51 and plates 62.

[5] LANG. 9.89–92. Other reviews: Cohen, BSL 33.3.10 f (1932) (not available); Malone, MLN 48. 394 f. (1933) 'deserves credit for taking the first steps' but 'does not seem to realize that his task has just begun.'

Dynamic Philology', and to judge from the text this means a comprehensive survey of an established science written by an adept. We may therefore take the book for a complete though perhaps not the definitive presentation of Zipf's doctrine, and consequently believe that this is a proper time and occasion to attempt a critique of that doctrine, of its substantiation, and of its application.

The thesis, very briefly stated, is that the key to the explanation of all synchronic and diachronic language-phenomena has been found in a statistically established tendency to maintain equilibrium between size and frequency.[5a] Previous critics found the conclusions rash and largely improbable; they placed the blame partly on the introduction of a new technique into linguistic study. If they conceived an unjustly harsh opinion of statistical method in linguistics, the mistake was a natural one, for there was no one to warn them where statistics left off and explanation began except Zipf himself. As the matter now stands, neither the usefulness of statistical method in linguistics nor the value of Zipf's daring and ingenious explanations can be properly appraised, for they have not yet been separated. The separation and the separate appraisals will be the subject of this paper.

Before proceeding with the critique, it is proper to issue a general warning. The statistician avoids the popular concept 'cause and effect' and prefers to work with the concept 'functional interrelation' as it is used in mathematics and natural science, where the word 'function' has a technical meaning. When the statistician is confronted with two variable quantities in a complex of phenomena, he sets himself to observing whether certain values of variable A are associated (in his observation) with the probabilities or possibilities (both a posteriori) that variable B will have certain of its possible values. The two are said to stand in 'functional interrelation' when every possible choice of a value of A is found associated with a restriction of the possible or probable values of B. Two particular cases will be of interest: (1) Each variable is said to be a single-valued 'function' of the other when

[5a] In greater detail: (1) That relatively frequent use of a linguistic unit causes it to be reduced in one or more of its various kinds of magnitude—accent, complexity of articulation, extent in time, number of components, etc.—while relative infrequency of use occasions corresponding enlargements; (2) that this Law of Abbreviation has been established by statistical study; (3) that this Law can serve as the basis of a new science of language; (4) that current techniques of linguistic science thereby become partly obsolete, partly ancillary. These formulations are my own; the corresponding statements and implications in Zipf's writings are scattered and diffused throughout his publications.

the restriction is always to a single value of B, so that a choice among the values of A implies a choice of a certain value of B. (2) The two variables are said to stand in 'statistical correlation' when it is a restriction of the probable values of B, so that the probabilities that B has certain of its possible values are different from what they were before the choice was made. There is a full range of possibilities in strictness of correlation. Evidently a single-valued function is that limit case of correlation which could be called 'perfect'; it is indeed that case which is marked with the statement 'correlation equals one', and the 'correlation' (η) used here is so defined that it can be calculated from numerical data. There has been observed a correlation between the heights of husbands and wives; it is measured (in the United States) by $\eta^2 = 0.20$ approximately, which means that when you meet a stranger in a well-lighted place you have a 12 percent better chance $\left(\sqrt{\frac{1}{1-0.20}} = 1.12\right)$ of guessing his wife's height than if it were too dark to judge his height. But it does not mean—and does not say— that a man's height partly determines his wife's height!

It should be particularly noted that here the variables are called 'A' and 'B' instead of being conventionally marked 'independent' and 'dependent'. Mathematicians know that the employment of the latter terms is arbitrary: either variable may be called 'dependent', for each is by definition equally a function of the other. By avoiding that arbitrariness we are enabled to see the beauty of the technique: we see that talking about functional interrelation does not imply a judgment as to which is the cause, or even a judgment as to whether or not there is any such thing as cause and effect.

On Plates I–III (44) Zipf graphically presents his data on the frequency-distribution of words in Chinese, English, and Latin. Having done some work of this sort myself, and having seen a much larger amount done by another man, I am in a position to judge the validity of this: the data are adequate and are correctly represented by the points plotted. The variables I shall call f (the FREQUENCY or times that the same word occurs in the text studied) and n (the NUMBER of different words which have the same frequency). A glance at the charts suffices to show that there is unquestionably a statistical correlation between n and f, and that apparently it is close to that sort of functional relation which would be represented by a straight line on the chart (namely $nf^a = k$ where k is a constant); further, rough measurement shows that in all three languages the value of a is about 2. For

Zipf this nearness of a to *2* is a discovery of cosmic import: 'But the overwhelming disclosure is this, that the formula for abbreviation is $ab^2 = k$, a formula exactly identical to that of gravitation.'[6] And so he is ready to disregard the possibly significant differences among the three languages, and to find that in all three a equals exactly *2*; I quote, replacing his symbols with mine, from page 41 of the present book: 'Now, the line drawn approximately through the center of the points in each chart represents in each case the formula $nf^2 = k$.' I do not know how Zipf drew his lines; if by eye-measure, then he has excellent eyes, but not good enough to settle a point of such great theoretical importance. Taking the first twenty points for Chinese and the first thirty for Latin (first from above in the tables (26–7), first from below on the charts), which are all the points we dare use for reasons which Zipf properly mentions (43), and applying the laborious but exact method of least-squares, we find that for Latin the best straight line is $nf^{1.988} = k$ and that for Chinese it is $nf^{1.93} = k$. The lines are good fits for the points chosen.[7] It should be noted that a mathematically complete set of points would include not only the extreme points from the data (e.g., $n = 1, f = 514$ in Latin) which lie outside the charts on the upward extension of the left margin, but also points for all the gaps in the series of possible values of f (e.g., $n = 0, f = 60$ in Latin), which points all lie at an infinite distance to the left; since the latter lie below the line while the former lie above it, there is no justification for the statement (42) that 'If one extended the diagonal line on each chart to include these words of great frequency, the line would bend up sharply.' For Latin indeed, the only proper interpretation of Plate IV indicates that the line on Plate III, if extended to the left, would bend DOWN, AND NOT AT ALL SHARPLY. The English line on Plate II would remain fairly straight; the Chinese line on Plate I would bend like the Latin one (cf. footnote 9). Incidentally, the line which Zipf drew on the Chinese chart is about $nf^{1.97} = k$, so that it lies half-way between its true place and the place where he says it is; or, to put it mildly, his line forms a connecting link between data and theory.

In order to determine the significance of the index and of its nearness to the number *2*, we must study the related chart on Plate IV. There we find words ranked in order of frequency. I shall use the symbol r for the RANK of a word, assigning $r = 1$ to the most frequent word. We find, in agreement with Zipf, a close approximation to the functional interrelation $fr = k$; that is, the second-most-frequent word

[6] Zipf, Selected Studies 24.

[7] The fit on Plate I is measured by $\eta^2 = 0.974$. That is, it is 97.4% perfect.

is about half as frequent as the first, the third in order is one-third as frequent as the first, and so on. Now the sum of all the individual word-frequencies must be the LENGTH (L) of the text, and the sum of the relative frequencies $\frac{f}{L}$ equals *one*. In English the relation between rank and frequency is $\frac{f}{L} = \frac{1}{10r}$ approximately, and we have $\frac{1}{10}\sum\frac{1}{r} = 1$, where the summation extends, word by word, through the whole vocabulary of the text. But $1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$ is a divergent series, and only about 12,000 terms are needed to give the required sum *10*. Yet the relative frequency of the most frequent word *the* ought to be the same for any amount of text up to the whole contents of all the English books ever printed and beyond. And our dictionaries recognize over a quarter of a million words instead of the indicated 12,000. It follows that the relation $fr = k$ can hold only for vocabularies of the order of magnitude found in Zipf's samples, and that Dewey's count[8] of 100,000 words of connected texts, with its 10,161 different words, already approaches the limit. Let us therefore replace the divergent series by a convergent series on the assumption (which I shall continue to treat as an assumption) that the lines on Plates I–III must be straight lines and not curves. On that basis the only possible series is $\frac{1}{10}\sum\frac{1}{r^{1+b}}$, where b is a small positive fraction, and then we can have a vocabulary of any size we please, each size corresponding to a particular value of b. An infinite vocabulary corresponds to b a trifle more than *0.10* (since $b\int_1^\infty \frac{dr}{r^{1+b}} = 1$ $[b > 0]$ where the small error made in replacing the series with this continuous function can be closely estimated). If the relation $nf^a = k$ holds, there is a definite correspondence between vocabulary-size and the value of b, and it is: Infinite vocabulary: $b = 0.106$; very large vocabulary (English dictionaries): $b = 0.08$ approximately; about 12,000 words: $b = 0$; still smaller vocabulary: b negative (divergent series). And b cannot be greater than about *0.106*. Now since, by definition of n, r, and f, we can say that $n = -\frac{\Delta r}{\Delta f}$ when $\Delta f = 1$, we can also, with trifling error, say that $n = -\frac{dr}{df}$. Therefore the same b

[8] Godfrey Dewey, Relative Frequency of English Speech Sounds (Harvard Univ. Press 1923).

reappears in the formulas for the lines on Plates I–III, and their indices are not *2* but $1 + \frac{1}{1+b}$. If we replace this with $2 - c$, then c will be nearly equal to b when b is small, and will always be less than b. If the line is straight its index cannot be less than *1.90*.

The reason for this brief excursion into the infinitesimal calculus is of course to get the argument on the record for any competent person to check; if any further excuse is needed, let it be the fact that Karl Verner could have checked it himself. Coming back to Plates I–III, we are now able to say that, if a straight line is a good representation of the relation between n and f, then a will have to be very close to *2* in its formula $nf^a = k$. It will be more than *2* for very small vocabularies, less than *2* for large ones,[9] and cannot be less than *1.90*. If the 'best' straight line has an index less than *1.90*, then it is not as good a fit as some curve would be. A sinuous curve with curvature AT EACH END would counterfeit what we have just found true of straight lines, but that is not what we find on Plates I–III (cf. footnote 7). This closes the use of the explicit assumption made above, and leaves the account balanced.

We can now leave Zipf's 'overwhelming disclosure' and turn to an error which crept in because of it—because he took *2* for a sort of ideal value for the index and correspondingly took the straightness of the line for granted. If the index of the best straight line is less than *1.90*, then a better fit would be a curve that is concave below, as can easily be demonstrated. That is the situation on Plates VI–VII (256), where Zipf nevertheless drew straight lines. The indices of the best straight lines for those charts would have only the one virtue of measuring a sort of average slope of the proper curves. To draw straight lines on those charts, and consider them as any sort of representation of the data, is an absurdity of which no statistician would be guilty—he would recognize it as an absurdity inherent in the fact that the definitions of n and f exclude an equation $nf^a = k$ with a less than *1.90*.

Since there will not be room later for further discussion of Plates VI and VII, I must here discharge the obligation of pointing out several curious errors in them and in the accompanying text. At least 13 points are missing on Plate VI; since some of those which are not missing

[9] The Latin index points to more than the actual 8,437 words; the high-frequency divergence shown on Plate IV is of the right sort to explain this. Strangely enough, the same is true of Chinese, as we learn from the full data (Selected Studies, App. C).

are slightly misplaced, it is hard to identify the missing points, but they seem to be those for *f* ('Peipingese: Number of Occurrences', 257) equals *16*, *26–33*, *36*, *37*, and *50*. Eleven of these fall below the left end of Zipf's line; two lie just above it. Taking the twelve missing points into account, the least-squares line for the first forty points is $nf^{1.55} = k$. This 'best' line has a slope some four degrees different from his line, or swings away from it about 8 mm. (more than a quarter-inch) in the length of his line—since the two lines cross, we add the deviations at both ends. The matter is further complicated by the fact that Zipf gives an erroneous index for the line he drew. His line is $nf^{1.33} = k$, and not $nf^{1.78} = k$ as he states at the top of 258. It is, then, steeper than the line he drew on Plate VII (for which he gives the correct index), and this would have to mean, according to his criteria, that Chinese is more highly inflected than French!

But the data are not such that a comparison of degree of inflection could be based on them. Zipf says (256): 'It is inconsequential for our present purposes that Henmon did not include the relative frequencies of formal prefixes, suffixes, and endings since the high relative frequencies of these would place them above that portion of the curve which is of special interest to us'. There is no warrant for the conclusion he expresses in the words 'since . . . ', and analogy ought to have led him to believe that the opposite was true—that, just as is true of words, certain formative elements were extremely rare. Who would dare say that *-îmes* is 'frequent'? Besides this, Chinese polysyllabism is a sort of synthesis, or aggregation, or 'addition' of morphemes and their meanings, and so is not comparable to the specialization or 'multiplication' (using the word as it is used in symbolic logic) of French inflection.

We are now in a position to appraise Plates I–III. The relation between the frequency of each different word in the Latin, English, and Chinese data on the one hand, and the number of words having the same frequency on the other hand, is correctly (within a measurable margin of error) represented by a straight line on a double-logarithmic chart, by the formula $nf^a = k$ in analytic terms. But the nearness of a to the number *2* is derivative and so-to-say accidental: It is connected with nothing but the straightness of the line and with the fact that the vocabulary is some thousands of words in size, as it would have to be in a sample of any respectable and yet manageable length. In graphical terms, nothing can be significant but the straightness of the line. It might and ought to be asked, even though it did not occur to Zipf to ask the question: Is not perhaps the straightness of the line implied

in the nearness of the index to the number *2*, so that if a curve of that average slope resulted from any count of a sample of different things, then that curve ought to be a straight line? There is a theoretical answer to this (the mathematical reader will already have recognized it in what has gone before), but fortunately we can use instead a practical check. Nearly two years ago the writer (with some assistance) counted the words in two samples of more or less abnormal language, a sample of Basic English and one of Esperanto. The data did not yield straight lines; both lines were curved, and they were curved in opposite directions.[10] As far as we know, then (for this sort of thing can be disproved but never proved), it is possible that straight lines ($nf^a = k$) are uniquely associated with 'natural' language, which Basic English and Esperanto unquestionably are not. But it still remains to be seen whether that is the most appropriate paraphrase of that straightness, and for the present we are left with nothing but the straightness itself.

The lines are straight; the relation between n and f is a power function. Now that may not seem to mean very much, but at least it means just what it says. What the 'philosophical implications' (in more accurate language, the 'possible paraphrases') may be, we are not yet in a position to guess; for the present we are still at the point where we have enough to do in simply arguing about what could be said concerning the fact that the relation of n to f is 'monotone'—that the lines, in going downward, go always to the right—for there seems to be no a priori reason for supposing that there ought to be more hapax legomena than there are of words occurring twice. But if the centuries-long experience of the natural scientist is to count for anything, then the straightness of Zipf's lines will some day prove rich in philosophical implications, and the paraphrases of $nf^a = k$ will be various and frequent in future linguistic works. For straight lines are notoriously among the most valuable discoveries of the scientific observer. Any natural scientist will confirm this. But he will also tell you (if carefully questioned—it is a thing so well known that it is seldom said) that it has not been found profitable to begin the use of each new discovery by using it as the only basis of a new science.

Though there are, as we have seen, plenty of opportunities for inconsistency and neglect of important principles in the mechanics of statistics, the most subtly dangerous errors come in when one begins to paraphrase and argue. As one possible paraphrase of the straightness

[10] This is not surprising: Esperanto shuns metaphor, while Basic English overworks its small vocabulary.

of the lines on Plates I–III, we have $\left(\text{since } n = -\frac{dr}{df}\right)$ the 'harmonic structure' of vocabulary (Plate IV and 45 ff.) according to which the mean interval at which a word recurs in connected text increases regularly with its rank, in order of frequency, among all the different words in that text. Zipf properly inquires whether this is connected with the length or phonetic size of the words, and furnishes a set of tables (26 ff.) showing how the sizes of words are, in his experience, correlated with their frequencies. We might have guessed that long words are generally rare (or that rare words are generally long—it can make no difference which way it is said), but it is one of the virtues of statistics—and not the least of its virtues by far!—that it can be used TO MEASURE THE OBVIOUS. And indeed, whenever the obvious has been measured there is occasion for a goodly amount of philosophising. Zipf is clearly within his rights in setting out to talk about this.

But he forfeits his rights with the first sentence he writes (28), for he starts out by hunting for a causal relation. Of course he finds it, and finds it quickly, for he has already decided[11] what it ought to be. He thoroughly confuses the synchronic and diachronic aspects of language-description, using paraphrases which might well apply in one or the other aspect but would have to be replaced with more complex ones if both were to be covered together. And he refers every use of a short word (where a longer one would be possible) to intention—to a conscious or unconscious striving for economy. Then, of course, the nature and (nota bene) the direction of the causal relation is easily settled: 'This tendency of a decreasing magnitude to result from an increase in relative frequency, may be tentatively named the Law of Abbreviation' (38). There is nothing hard about this. Given an observed functional interrelation it is always possible to 'prove' that one variable is the cause and the other the effect—that is the way homo sapiens, accustomed from earliest childhood to having things 'explained', invariably behaves—and with a modicum of the will-to-believe we quickly arrive at a causal Law, with nothing tentative but the adverb.

Of course it helps if we have first decided that one variable must be the cause of the other.[12] And it helps still more if we summarily decide

[11] Many years ago, that is, as we learn from his Relative Frequency 1, where this illuminating phrase occurs: 'With my *a priori* theory in mind, . . . ' Cf. footnote 21.

[12] As Zipf immediately does (28 f.), forgetting that both together might as well be considered as effects of a third or several other causes, once you start talking about causality. Trained statisticians seldom make that mistake; they are

WHICH is to be the cause of the other, as Zipf does (29) with these words and no more: '. . . because a speaker selects his words not according to their lengths, but solely according to the meanings of the words and the ideas he wishes to convey. Occasionally, of course, out of respect for the youth, inexperience, or low mentality of a particular auditor, a given speaker may seek to avoid long and unusual words. On the other hand, speakers are sometimes found who seem to prefer the longer and more unusual words, even when shorter more usual words are available. Yet in neither case are the preferences for brevity or length followed without respect for the meanings of the words which are selected'. This is intended to show that shortness cannot be the cause of frequency, which conclusion is taken as a demonstration that frequency is rather to be considered the cause of shortness of words. Now the amusing thing here is that, taking the ideas just quoted and remembering that the same meaning can be expressed in words of different lengths (as Zipf repeatedly says throughout the book), it is equally easy to 'prove' what he dismisses, and to use all his illustrations as illustrations of the (apparent) contrary of what he uses them for. The demonstration can safely be left to any reader who likes to talk about talking.[13] The combination of these apparent contraries is not, however, an antinomy—it is really a tautology.

For as far as we know, the two variables in functional interrelation are ambivalent, in that either may be taken as the cause whenever one feels the urge to find a causal law.[14] Taking each choice in turn and combining the two resultant laws, we have a circular 'explanation' instead of the original linear description. That linear description says just what we know: 'Short words are generally frequent words, long words generally rare', with a footnote to say that this formulation is not

---

always on the alert for this their special bête noir, and have given it the pejorative name of 'spurious correlation', under which title the curious will find, in any good book on mathematical statistics, directions for discovering whether there really is any relation between the consumption of imported apples per head and the female cancer death-rate. Zipf states the alternative carefully, but forgets that the Law of the Excluded Middle applies only when the alternatives BY DEFINITION together include all possibilities.

[13] A pursuit which Fritz Mauthner (somewhere in his three-volume Beiträge zu einer Kritik der Sprache) aptly compares to keeping a fire in a wooden stove.

[14] Incidentally, it is precisely that sort of ambivalence which the ideal natural scientist has in mind when he speaks of 'dynamic equilibrium'. He has not demonstrated mutual causality in nature; he has simply perceived a quality of his knowledge.

intended to be in any way different from 'frequent words are generally short words, rare words generally long'. The 'explanation' says more than we know, for it introduces, without justification, the notion that there is a difference between these two formulations—and causality creeps in with the notion that the choice between the two is not arbitrary. The expansion of the linear description into a circular explanation is a tautology, and the Law of Abbreviation, the central principle of Zipf's doctrine, is an arbitrary half of a tautology.[15]

All tautologies aside, Zipf has presented us with one functional interrelation, measured another for us, and offered a number of stimulating paraphrases. As a scientist he may well be proud of having done that much; the rest belongs to omniscience.

The next section of the book deals with 'the form and behavior of phonemes'. Since Zipf's phoneme dates from before the appearance of Twaddell's treatise,[16] it is very hard to follow him sympathetically through the arguments. Fortunately that is unnecessary, for as far as they are valid they can be replaced in terms of Twaddell's procedure. For example, the useful[17] section dealing with 'skewness' (101–6) can be replaced by the use of micro-phonemic sets;[18] the results will then be in terms of what is known—the data, the explicit assumptions, and nothing else. Zipf's arguments depend partly on probably unverifiable guesses about articulation, partly on an ethical fiction;[19] that is, his phoneme is an articulatory norm standing in mutual-causality relation with the speaker's intention.[20]

[15] A great deal more could be said about these things, but fortunately it has been made unnecessary for me to try to say it myself, and I hasten to give all due credit, thanks, and appreciation to Leonard Bloomfield, who in all his works and especially in two recent reviews (LANG. 8.220 and 10.32) has rendered to linguistics the inestimable service of reminding us of the nature of scientific method.

[16] W. F. Twaddell, On Defining the Phoneme, LANGUAGE MONOGRAPH No. 16 (1935).

[17] Eminently useful in that it states a problem and gives us valuable hints for the application of a strict procedure.

[18] Twaddell 61 f.

[19] Similar to the fictions called Free Will and Responsibility. (Hans Vaihinger, Philosophie des Als Ob, Leipzig 1922[8], 59 f.) Zipf holds the speaker responsible for having articulated in a certain fashion under the fiction that he INTENDED to articulate in a certain (perhaps different) fashion. It does not appear that Zipf recognizes this as a fiction, nor does he recognize what is still more important, that an ethical fiction cannot be the basis of a scientific method.

[20] Zipf does not DEFINE his phoneme; characteristically, he 'explains' it. The kernel of his explanation is this: 'The speech-sounds, distributed about the norm

Just as Zipf has 'found' a causal relation between lengths of words and their relative frequencies, so he is prepared to find a causal relation in the same sense[21] between the relative frequencies of phonemes[22] in the stream of speech and their sizes—their 'magnitude of complexity'—so that a rare phoneme would (because of its rareness) be given a complex articulation to make it conspicuous (so that it wouldn't be confused with a well-known common one?), while the commoner ones would be given a less complex articulation. In preparation for the finding of this causal law, he gives arguments to show that aspirated, fortis, and voiced stops are respectively more complex than unaspirated, lenis, and voiceless stops—he is presently going to show us that the latter are more frequent. It should be noted that Zipf's 'complexity' is now entirely articulatory (in contrast to his earlier method, where the word was 'conspicuousness'[23]), and is to be determined solely by observation of the act of speaking. Now in that frame of discourse 'complexity' can only mean 'difficulty', and ought to be measured by the amount of control[24] which the speaker exercises. On that basis, an unaspirated stop is more complex than an aspirated one, since to avoid aspiration the speaker must begin producing voice or else stop the lung-pressure EXACTLY as the stop is released, though he may do the same SOONER OR LATER after release and still produce a true aspirated stop. Again, a voiced stop is easier to manage than a voiceless one, since it does not

---

of the phoneme, give significance to the phoneme, just as the norm of the phoneme gives significance to the speech-sounds which approximate it.' (53.)

[21] Zipf found his Law first for syllables and sounds, and presented it as the Principle of Relative Frequency on page 4 of his Relative Frequency. In the present book he derives it first for words, which is easier to do. Apparently he got the idea of the universality (or versatility) of the Law in the interesting fashion stated in the quotation referred to in my footnote 6.

[22] Space is lacking here for explicit treatment of many theoretical questions—e.g., whether a phoneme 'occurs'—but I shall try to cover them by implication. So much can be said, however: Zipf's difficulties show the necessity of a rigid procedure for determining phoneme-membership.

[23] Relative Frequency 36 f. It was a phonetic—largely acoustic—conspicuousness, and was to be measured principally by the amount of attention it attracted. Here the frame of discourse is SOCIAL. Now in social behavior, practically by definition, conspicuousness is the same thing as rarity—for it is only unusual behavior that attracts special attention. The proper definition of 'conspicuousness' could hardly be anything but 'that which varies inversely with frequency'.

[24] The amount of EFFORT is irrelevant, unless we are prepared to find that more complex phonemes are more frequent in accented syllables than in unaccented ones.

require cessation of voice after a preceding vowel, etc.—a conclusion which would find general favor among Romance scholars.

The other preparation for the introduction of the data is a demonstration of the propriety of associating English [t] with French [t], English [d] with French [d], for statistical purposes, in spite of differences in articulation. The association is done according to resemblances (in that $[t]^e$ is closer to $[t]^f$ than to $[g]^f$ etc.); since the appearance of Twaddell's monograph we must dismiss this, and for quite different reasons than those which Zipf cites as possible objections. I see no reason, however, why Twaddell's method should not be extended to this field, if any good purpose can be served thereby. Then $[t]^e$ could be statistically associated with $[t]^f$ according to the proportion $[t]^e : [d]^e :: [t]^f : [d]^f$ as soon as it has been shown that in each language there is a two-member series of tongue-tip stops. The same cannot be said of an attempt to associate Burmese [t] with Cantonese [t], for the Burmese phoneme is in a three-member series while the Cantonese series has only two members. If this needs support in the minds of those who do not accept Twaddell's procedure, let it be the consideration that very likely Burmese [t] would have special articulatory peculiarities, to keep it apart from BOTH $[t^h]$ and $[t_1]$, which would be unnecessary in Cantonese.[25]

The data which Zipf presents (68–79) are intended to show statistical correlation between two variables OF WHICH ONE IS A CLASSIFICATION (for Zipf admits the impossibility of measuring complexity and does no more than classify into 'more' and 'less' complex). Now that sort of correlation is recognized in statistics, and there are well-grounded methods for attacking the problem. But they are founded on a condition which must never be forgotten: Before using the classes, one must really classify. There must be a classifying procedure founded on data, assumptions, and water-tight logic. What is wanted here is a uniform classifying procedure which, applied to the three articulatory oppositions mentioned, would infallibly determine which member of each opposition is to be called 'more complex' (or called by any other factitious name). When such a procedure is lacking—and I think I have shown that it is lacking here—THE CLASSIFICATION ITSELF must be based on statistical study. It can be done, and has been done repeatedly.[26]

[25] The connection between the two points of view is implied in Twaddell 57.

[26] A classical example is Karl Pearson's article On the probability that two independent distributions of frequency are really samples of the same population, Biometrica 10.85–143 (1914–15).

The result, as the statistician knows, is a circularity which must then be carefully collapsed into a linear description—the reverse of Zipf's favorite procedure. Zipf here chooses the harder way (perhaps the impossible way) and then makes it easier for himself by the standard magician's device of putting the egg in the hat before the audience knows he is going to take anything out of the hat.

In view of all this, the particular uncertainties in the data are beside the point here, and they have already been partly treated by the reviewers of Zipf's previous books. At least one definite improvement ought to be noted: the Chinese stops now have a more acceptable treatment than before. But this is set off by the fact that the Czech data are now cited as being 'from ACCURATE PHONEMIC transcriptions' (74), though the same data are given in his first publication (to which he here refers us for the sources) as representing 'printed letters'.[27] The footnote in the present book contradicts and partly corrects the misstatement with the words ' . . . the conventional Czechish alphabet is practically as accurate a phonemic alphabet as can be devised for Czechish'. The correction is not complete, for there is good warrant for considering Czech (as also Russian and Bulgarian) *b d* (*g*), when final or before voiceless sounds, as belonging to the [p] [t] ([k], [x]) phonemes. When we note that Zipf went to a great deal of trouble[28] to take account of the similar peculiarity in German, we can only conclude that he was not sufficiently on his guard against possibilities of ignorance or inconsistency; we may even be pardoned for suspecting that he simply hadn't been told about the Spanish orthographic equality $v = b$.

There are three more principal divisions of the present book, dealing with 'accent', 'the sentence', and 'the stream of speech and its relation to the totality of behavior'. Since they are based on what has already been discussed, and since the Law of Abbreviation determines the argument throughout, further detailed criticism is hardly necessary. As elsewhere, there is a great deal of stimulating discussion and suggestive formulation. But its claim to permanent scientific value is vitiated by the practice of working out an unambiguous causal explanation from and for each ambivalent correlation. Such 'explanations' may be as

[27] Zipf, Relative Frequency 42 (top) and again (45) at the head of the table. The Czech source was Těsnopisné Rozhledy ('Stenographic Survey' or 'Review'), so that we should expect the data to represent letters and not phonemes, even though Sedláček was the author.

[28] Relative Frequency 52–6.

coherent and logical as you please; they are not scientific demonstrations. As Bloomfield so neatly puts it, they 'short-circuit inquiry'.

In his Introduction (5) Zipf has written: '. . . it is difficult to believe that linguistics has been entirely mistaken in the direction which it gave to language study. Certainly no student of speech-dynamics can for a moment regret the stringency of the historical and comparative disciplines which have provided him with immediately available material'. This bears an implication, unjust both to statistical method and to linguistic science, which must be explicitly denied. Statistics enjoys no qualitative superiority over the best practice of modern linguistics. It is simply a technique for so describing data that their regularities are given consistent labels. The statistician chooses to work mathematically because mathematics, which is definable as 'pure consistency', furnishes a conveniently available array of consistency-patterns. His technique has nothing to do with causality—not even with that mystic mutual causality for which the label 'dynamic equilibrium' has been borrowed from science. Those are limitations of all scientific method. Anything beyond this is not science, however worthy it may be as artistic description—e.g., as 'explanation',—and when those limits are overstepped, science has a right to disclaim the result and refuse any possible blame.

Although we cannot ascribe to the statistical method any sovereign efficacy in linguistics, it by no means follows that a sound use of statistical method is out of place in our discipline. One of the objects of this critique has been the habilitation—under the circumstances one might even say the re-habilitation—of a branch of scientific method, the statistical, as a tool in linguistic study.

MARTIN JOOS

De Hettitische *ḫ*. Pp. 43. By WALTER COUVREUR. (Teksten en Verhandelingen, Nummer 12; Beheer van Philologische Studien.) Leuven, 1935.

Coming from a study of de Saussure's[1] mathematical speculations about the origin of the Indo-European long vowels, Couvreur, like several other scholars, found unexpected confirmation of the theory in the Hittite sound (or sounds) written with the cuneiform *ḫ*-signs. He tells us that his theory was completed before he became acquainted with the views of Kurylowicz, Cuny, and Pedersen, and he differs from their conclusions in several respects.

[1] Ferdinand de Saussure, Mémoire sur le système primitif des voyelles dans les langues indo-européennes 134–84 (1879).