COMPUTERIZED EXPERIMENTS WITH THE FREQUENCY LISTS OF THE FIVE SCROLLS*

To say that the computer is a useful tool in literary criticism and may open new areas of research and new avenues to researchers is merely to state a truism. That it has become a common place must not lead us to neglect it. The amount of literature awaiting exploration is so vast, the preparation of the texts for processing by the computer still so laborious and the cost so high that comparatively little has so far been done in practice. This holds true, especially, for the Hebrew Bible. The purpose of this study then is first to subject a certain part of this book to some, and for the time being even to a rather superficial examination, and later, by concentrating on one specific book, to investigate a few of the problems it poses and to offer a solution for one at least.

The subject matter of the present enquiry is the Five Scrolls, a term which refers to five books, parts of the Hagiographa, i.e. the third section of the Hebrew Bible, which, in their canonical order, are Canticles (1249 words), Ruth (1287 words), Lamentations (1539 words), Ecclesiastes (2988 words) and Esther (2408 words).

With the entire corpus of words, including detailed lexical, gramma~~tical~~ syntactical, ~~morphological and, to some extent, semantic~~

information on magnetic tape, the computer was asked to draw up (a) frequency lists and (b) frequency profiles for each scroll, and then to do the same for two halves of each scroll when approximately divided in the middle. Samples of what the printouts look like are given for the Book of Ruth in Tables 1 and 2. The symbols *N* stand for the total length of the text and *V* for the number of different words therein.

Table 1

Frequency list of Ruth

N°	Word	Absolute Frequency	Cumulative Relative Frequency
	(1)	(2)	(3)
1	ʾāmar	54	4.20
2	lō̄	51	8.16
3	ʾašer	42	11.42
4	ʾet	37	14.30
.		.	.
.		.	.
.		.	.

Table 2

Frequency Profile of Ruth

Absolute Frequency	n° of Words such	Cumulative n°. of Words such		Cumulative Relative Frequency of Words such	
		in *V*	in *N*	in *V*	in *N*
	(1)	(2)	(3)	(4)	(5)
1	155	155	155	45.99	12.04
2	68	223	291	66.17	22.61
3	30	253	381	75.07	29.60
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
54	1	337	1287	100.00	100.00

Table 1 is self-explanatory. Table 2 should be read as follows : There occur in Ruth 30 different words the frequency (*f*) of which is 3. Their number taken together with the number of all other different words of a lower frequency ($f \leqq 3$) amounts to 253 or 75.07 % of the total vocabulary (*V*), while the number of occurrences of these 253 words is 381 or 29.60 % of the text size (*N*).

As interesting as the lists and profiles are, they obviously cannot be presented here in full. Leaving the profiles aside, for the moment, ~~we~~ shall ~~look instead only~~ at the ~~lists and find out what may be~~

learned from the first ten most frequent words in each. They are listed in Table 3 where a distinction is made between grammar words and content words by printing the latter in italics. For the benefit of those who are familiar with the Bible, but rather less so with Hebrew, the approximate equivalents in English are added. Personal, including Divine, names, toponyms etc. are omitted from the counts.

Even before looking at Table 3, we feel confident of what to expect there : grammar words will surely rank highest. In English and French, where grammar words make up about half of any sample of at least 30 words, probably no content word at all would climb as high as rank 10 in the frequency list.[1]
In Hebrew, where the indefinite article is absent entirely and where the definite article and the three most frequent prepositions are bound morphemes which lack autonomous existence, one might expect two or three content words to rank between nos. 5 and 10, at the most. One may further predict the five lists to be almost identical. But a glance at Table 3 proves that these expectations are wrong.

Table 3

*The Ten Most Frequent Words and their Percentages**

Rank	Canticles		Ruth		Lamentations		Ecclesiastes		Esther	
	Word	%	Word	%	Word	%	Word	%	Word	
1	*dōd* (beloved)	2.72	*'āmar* (say)	4.20	lō ' (no)	2.53	'ašer (rel.pron.)	2.97	*melek* (king)	6
2	lō (him, her)	2.64	lō (him,her)	3.96	ʿal (on)	2.47	kī (subord.conj.)	2.92	'et (nota accus.)	3
3	ʿal (on)	2.08	'ašer (rel.pron.)	3.27	kol (all)	2.34	'et (nota accus.)	2.44	'ašer (rel.pron.)	3
4	'et (nota accus.)	1.60	'et (nota accus.)	2.87	lō (him,her)	2.21	kol (all)	2.18	ʿal (on)	2.
5	*'anī*(I)	0.96	kī (subord.conj.)	2.18	kī (subor.conj.)	2.01	lo' (no)	2.18	kol (all)	2.
6	*bat* (girl)	0.96	'el (to)	1.63	*hāyāh* (become)	1.82	gam (too)	1.94	'et (to)	1.
7	*yāfe* (pretty)	0.96	*gā'al* (redeem)	1.63	*bat* (girl)	1.50	*'ādām* (human being)	1.84	*yōm* (day)	1.
8	ma (what)	0.96	*hāyāh* (be,become)	1.63	*yōm* (day)	1.42	*hāyāh* (be,become)	1.74	*'āmar* (say)	1.
9	*ahāvāh* (love)	0.88	*'īš* (man)	1.63	*rā'āh* (see)	1.04	*rā'āh* (see)	1.54	*yehudī* (jew)	1.
10	lō ' (no)	0.88	*ba'* (come)	1.56	*'ōyēv* (foe)	0.97	*'eyn* (neg.part.)	1.47	*ʿāśāh* (do)	1.
Total of *N*		14.64		24.56		18.31		21.22		27.

*Content words in *italics.*

To be sure, the same first glance reveals that there is nothing much exciting about Table 3. What is astonishing, though, is how much the lists differ and how much the ten words of top frequency alone reveal of the contents and genres of the five books. For instance, the argumentative character of Ecclesiastes is shown by the preponderance of grammar words, its speculative tone by the high ranks of *all, human being* and *see*, and its pessimistic attitude to life by two negative particles occupying rank 5 and 10, respectively. Canticles and Esther are different. The number of contents words and their very high ranks attest to the lyrical quality of its love and nature poems in Canticles, while the many content words in Esther point to its dramatic plot. These two observations of course do not by far exhaust what may even now be gathered from the short lists, the less so from their full lengths, but further analysis would lead too far afield. One more detail, however, is worth noting : the total percentages (bottom line) divide the five scrolls into poetry and prose, roughly, but nevertheless reliably.

The same neat division obtains when two other features are considered : the relative frequencies of the *nota accusativi* (*'et*), the use of which is facultative in classical Hebrew, on the one hand, and of the subordinative conjunctions, a measure of hypotaxis reigning in a text, on the other. Both are tabulated in Table 4.

Table 4

Percentages of the nota accusativi and Subordinative Conjunctions

		Poetry			Prose	
	1.	Lamentations	0,39	1.	Ecclesiastes	2.44
nota accusativi	2.	Canticles	1.60	2.	Ruth	2.87
				3.	Esther	3.63
Mean			1.00			4.47
	1.	Canticles	1.28	1.	Esther	4.90
Subordinative	2.	Lamentations	2.73	2.	Ruth	6.07
Conjunctions				3.	Ecclesiastes	6.59
Mean			2.00			8.78

Not only is there no overlapping between the two sets of values, but they are markedly dissimilar : percentages for prose are as much as four times higher than for poetry. These rather simple findings may not be altogether useless to the scholar when he is called upon to distinguish between prose and poetry, a distinction often quite difficult to make in Biblical literature.

Another kind of problem - one of much wider and at times theological implication - which besets Biblical criticism is whether a certain book was written by one or by several authors. In our five cases, Ruth and Esther are universally viewed as homogeneous while opinions widely differ regarding the rest. Let us see what the frequency lists tell us in this matter.

After dividing each scroll in half and drawing up frequency lists for each part, we may calculate vocabulary concentration (*C*). This is usually done by adding up the relative frequencies of the first 50 most frequent words excluding names as above ($C = \sum_{1}^{50} f$) (2). Results are shown in Table 5.

Table 5

Vocabulary Concentration (C)

Scroll	Section	C in %	Probability of Homogeneity in %
		(1)	(2)
Canticles	A	41.92	0.70
	B	40.85	
	A + B	38.03	
Ruth	A	58.12	0.56
	B	59.73	
	A + B	55.32	
Lamentations	A	45.83	0.0001
	B	36.10	
	A + B	36.70	
Ecclesiastes	A	57.52	0.03
	B	53.56	
	A + B	54.35	
Esther	A	55.72	0.05
	B	52.11	
	A + B	53.62	

That this experiment again differentiates between prose and will result in high vocabulary richness in the prose as compared with the poetry can be foreseen. Yet Table 5 contains additional information, that is, the values of *C* found for each section pair. These data enable us to inquire whether *C* remains even within homogeneous books and what the probabilities are that the two halves of each book originate in the same population. Results are given in col. 2. That this probability is high in Ruth may have been taken for granted, for the inner unity of this short story is uncontested. An even higher value obtained for Canticles which one would perhaps not have guessed since this is a collection of unconnected short poems. Ecclesiastes, though generally ascribed to a single author, is a borderline case, probably because of the variegated literary types occurring within it. The same is true suprisingly for Esther, a matter less easy to explain. The lowest probability was found for Lamentations.

We shall tentatively proceed on the assumption that vocabulary concentration remains stable within a homogeneous work. In contrast to all other scrolls, Lamentations displays so startlingly different a vocabulary behaviour that it deserves closer inspection. We shall return to it later.

So far, we have been concerned with the frequency lists and contenting ourselves with examining, first, which words occupy the ten highest ranks and, second, what proportion of the text is taken up by the fifty most frequent words. We did not pay any attention to what happens in these lists from rank 51 onward for one simple reason : these lists are too long to be shown here. Nevertheless, they may be seen in their entirety if, from now on, we decide to disregard which words and what sort of words they contain and to be satisfied with numerically expressed frequencies. By this means, the long frequency lists may be curtailed to frequency profiles of the kind shown in Table 2, which, in the case of Ruth, consists of only 54 lines unlike the corresponding Table 1 of 337 lines.

The purpose of our first experiment with the profiles is to test whether, and if so, to what extent the distributions of word frequencies in the five books resemble each other. The question of similarity vs. dissimilarity must of course not be left to subjective decision on the investigator's part, but should be examined by means of conventional statistics without any *a priori* assumptions. The proper test is the one called chi-square (χ^2). It tells us whether or not two series of frequencies compared with each other were drawn from the same statistical population, and χ^2indicates the probability a of two series being homogeneous and may be found in standard tables in the column of the appropriate number of degrees of freedom. From them, standardized measurements (*d*) for dissimilarity are calculated which will decrease with the increasing probability a of homogeneity. Further details of this procedure may be read in any handbook of statistics(3)and have no place in this enquiry. Table 6 shows what we found.

Table 6

Comparisons between Pairs of Scrolls[*]

n°	Pair	*df*	χ^2	*a*	*d*
		(1)	(2)	(3)	(4)
1.	Esther/Ruth	14	14.06	42%	0.21
2.	Ruth/Ecclesiastes	13	17.58	18%	0.90
3.	Esther/Ecclesiastes	18	29.44	3%	1.91
4.	Ruth/Canticles	10	48.40	<0.5%	6.35
5.	Esther/*Canticles*	14	66.30	<0.5%	9.88
6.	Ruth/*Lamentations*	10	42.67	<0.5%	7.31
7.	Esther/*Lamentations*	14	84.82	<0.5%	13.38
8.	Ecclesiastes/*Canticles*	14	42.12	<0.5%	5.31
9.	Ecclesiastes/*Lamentations*	14	51.13	<0.5%	7.02
10.	*Canticles/Lamentations*	9	18.23	5%	2.18

[*] Poetry in italics.

That vocabulary richness as expressed in frequency profiles does indeed somehow characterize a work of literature or, at least, indicate its literary type is borne out by Table 6. Nearest to each other lie those pairs where books of the same type were compared : Esther/Ruth, Ruth/Ecclesiastes, Esther/Ecclesiastes (prose vs. prose), and Canticles/Lamentations (poetry vs. poetry). On the other hand, the dissimilarity is the widest between Esther, a semi-humorous prose narrative, and Lamentations, a collection of elegiac poetry.

An interesting case is n^{o} 8 : of all cross-type comparisons (n^{os} 4-9) it shows the lowest value ($d = 5.31$), although the two comparands -a collection of sweet love poems and another of musings by a resigned old sage- are surely most dissimilar. Without attaching too much significance to this result, one is, of course, still immediately reminded of what Jewish tradition has to say of Ecclesiastes and Canticles : that they were both composed by King Solomon.

The next step is to examine the proposition that vocabulary richness as indicated by frequency profiles remains stable within the two halves of the same book. This experiment resulted in Table 7.

Table 7

Comparisons between Halves of Scrolls (in decreasing order of a)

Scroll	*df*	χ^2	*a*	*d*
	(1)	(2)	(3)	(4)
Ruth	7	8.08	40%	0.29
Esther	12	15.30	23%	0.67
Ecclesiastes	12	18.88	10%	1.40
Lamentations	12	21.75	< 0.5%	5.30
Canticles	6	27.50	< 0.5%	6.21

The expectation that a verdict of inner homogeneity would emerge from Table 7 for Ruth and Esther was satisfied if we let our argument rest on the one single property under scrutiny, i.e. vocabulary richness : values of *d* in both cases are minimal. For Ecclesiastes, a somewhat higher value obtained, probably caused by the variety of genres contained within this book. In constrast, the linguistic behaviour in regard to vocabulary richness of Canticles and Lamentations is so different in their respective two halves that the odds, that each pair of halves originated in the same statistical population, are less than 0.5% ! Since this is the second time in the course of this enquiry that Lamentations is found to be exceptional, we shall now subject this book to deeper investigation and leave Canticles for another occasion.

Concentrating now on the Book of Lamentations, we shall try to extract from its frequency profiles as much as these may contribute to solving some of the questions asked about the book. But, first of all, they must be posed here.

The book consists of five distinct poems lamenting the fall of Jerusalem and the destruction of the Solomonic Temple by the Chaldeans in 586 B.C.E. Chs. 1,2,4, and 5 bemoan the national ruin while ch. 3 is a personal elegy. According to Jewish, and then Christian, tradition, Jeremiah was their anonymous author (bab.Baba Batra 15a), and since he died shortly after 580, the book seems clearly dated. The Talmud, though, noted for its disinterest in matters of authorship, is not, and probably does

not expect, to be trusted in this case. Its tradition is now unanimously discredited - which is also the point where the consensus of contemporary scholars ceases. The five poems are today variously attributed to one or more, and, at the most, four authors - not to five because chs. 2 and 4 are, in the opinion of all critics, believed to be the work of one person. For mainly impressionistic reasons, these two are said to be the best from the literary viewpoint. And as it is almost axiomatic in Biblical literary criticism that the quality of writing declined after 586, they must necessarily be the earliest. Since ch. 3 is allegedly 'artificial', it follows from the same argument that it must be the latest, and moreover, since it is the only one written in the first person singular, that it is 'foreign'. Another school takes ch. 5 to be climax of the five, although, as again some other critics would have it, it may be dated as late as the Hasmonean era (2nd century B.C.E.).
It has also been suggested that the composition of ch. 1 preceded the final catastrophe by a couple of years when the Jewish kingdom was already on the brink of extinction. The grounds on which these conflicting views are based are to be found in any introduction to the book(4) or in a Biblical encyclopedia(5).
Be that as it may, not one of the proponents of these hypotheses has any explanation to offer with regard to the sequence of the poems as they are found in the collection.

Disregarding, for the sake of objectivity, the alleged affinity between chs. 2 and 4, we compared the five dirges with each other and thus arrived at Table 8.

Table 8

Comparisons between Chapter Pairs of Lamentations

n^{o}	Chapter	Pair	*df*	χ^2	*a*	*d*
			(1)	(2)	(3)	(4)
1.	ch.1	ch.2	4	6.14	0.17	0.76
2.		ch.3	4	12.39	0.01	2.97
3.		ch.4	3	14.96	< 0.005	4.88
4.		ch.5	4	14.91	< 0.005	3.86
5.	ch.2	ch.3	4	1.73	0.78	-0.80
6.		ch.4	3	7.70	0.06	1.92
7.		ch.5	3	13.23	< 0.005	4.17
8.	ch.3	ch.4	2	2.41	0.27	0.21
9.		ch.5	3	6.70	0.09	1.51
10.	ch.4	ch.5	2	3.02	0.17	0.51

Table 8 is most surprising. What emerges from it (see col.4) is that the dissimilarity indices between adjacent chapters are extremely small and that they neatly and gradually increase - almost without a flaw, i.e. excepting n^{o} 3 - until they reach maximum when the first is compared with the last chapter. Thus, if vocabulary richness only is taken into account, there seems no reason any more to regard chs. 2 and 4 as especially related and close to each other or to label ch. 3 as extraneous. In general, Table 8 would point to some sort of homogeneity where chapters placed beside each other resemble each other whereas those that are separated slowly grow more dissimilar. This conclusion, however, does not exploit all the information available in the frequency profiles.

For instance, these profiles may be represented by points, the distances between which are given by indices, a procedure called 'scaling'. Such scaling is preferable when feasible in the smallest possible dimension, i.e. on a straight line, and should such one-dimensional scaling be impossible, in an area, and so on. It is evident that the greater the number of points, the lower the probability of one-dimensional scaling, provided that the dissimilarity indices are random. The exact odds for such scaling are given in Table 9.

Table 9

Probabilities of Low-Dimension Scaling

N° of Points	Lowest Possible Dimension in Pessimal Case (1)	Probability of Scaling on Straight Line in Given Order (2)
3	1 (straight line)	1 : 3
4	2 (area)	1 :15
5	3 (space)	1 :85

Let us now take another look at Table 8 again. The scaling on a straight line is characterized by the relations between the ten indices (*d*). There are altogether 10 x 9/2, i.e. 45, such relations. Among them, two only (the pairs n^{os} 6/7 and 8/9) do not conform to perfect scaling on a straight line while all the remaining 43 do so in full accordance with the present order of chapters. The probability of this event happening by chance is 1 : 34 or three time in one hundred analogous cases, which enforces the suspicion, hinted at above, that the arrangement of the chapters is not arbitrary.

The rather extraordinary findings of Table 8 may conveniently be represented in a more lucid way by a diagram instead of a series of numbers. Let us plot the number of different words of a given frequency on a grid of coordinates, and let the abscisse be the logarithm of this number and the ordinate the logarithm of the frequency. By this means, the profile will show as a series of points which will generally follow a straight line in accordance with Zipf's Law(6) as developed by Mandelbrot(7). The resulting slope of the line chatacterizes vocabulary richness : the steeper the slope, the less rich the vocabulary. An example (for ch. 1) of such graphic presentation is Fig. 1.

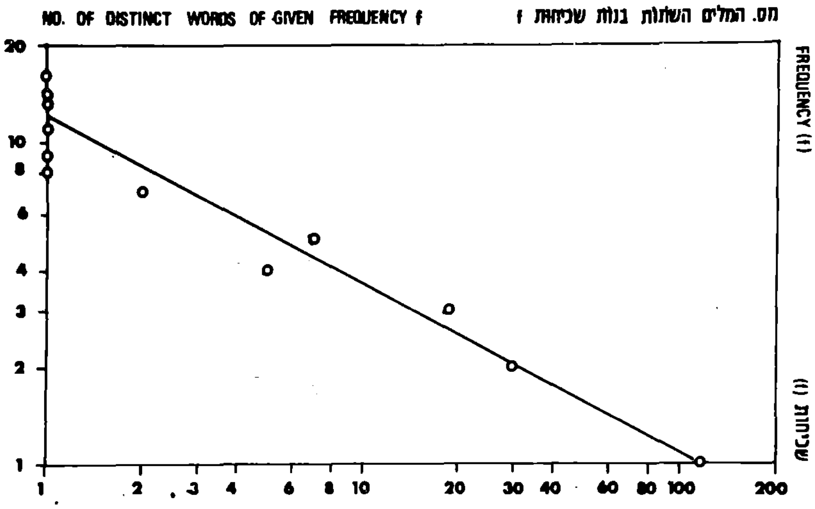

Fig. 1 FREQUENCY SLOPE (Ch. 1) גרף מס. 1 שיפוע-שכיחות (פרק א)

If we add in Fig. 1 the slopes of the remaining chapters, it would only confuse the reader. Instead, one may calculate the slopes by means of special regression analysis. This takes into account that the dispersion of points gradually decreases with the frequency. The five slopes appear in Table 10 with their standard deviations.

Table 10

Frequency Profile Slopes in Lamentations

Chapter	Slope	Standard Deviation
	(1)	(2)
1	0.52	0.03
2	0.48	0.05
3	0.45	0.03
4	0.39	0.03
5	0.38	0.02

The striking feature of Table 10 is that the steepest slope, the expression, as mentioned before, of the lowest vocabulary richness, emerged for ch. 1 and that the slope of each following chapter is less steep than the one of the preceding it - precisely according with the sequence of chapters in the book. The probability of this being a chance event is no more than 1 : 120.

Here, a comment is in order. The findings of Table 10 are in no way a natural corollary of those of Table 8. In fact, the contrary is true : the two are independent. In Table 8, there was no presupposition that the frequency plot would display a certain pattern - linear or otherwise - but in Table 10, it was assumed in the first place that the result would be a straight line, which in fact it was. It follows that these two independent sets of findings corrobate each other, lend force and validity to each other, and, in addition, prove that the slopes correctly represent the respective frequency profiles.

This result cannot be overrated : statistics after all aspires, among its other aims, to arriving at one single distinct numerical value - a function of observations called statistic - permitting us to condense and render the information included in a sample in its totality and all at once.

For better evaluation, Fig. 2 presents the data of Table 10 visually.

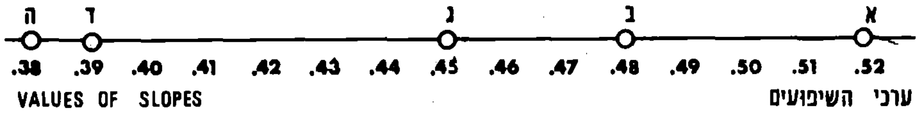

Fig. 2 CHAPTERS ON THE FREQUENCY SLOPE SCALE

גרף מס. 2 הפרקים בסולם שיפועי השכיחות

It must be admitted, though; that Fig. 2 does not tell the entire truth since it neglects the standard deviations which appear in Table 9. This deficiency is corrected in Fig. 3 where the confidence intervals of 95% are shown. Thus, for instance, the 'true' slope (for a homogeneous infinite text) of ch. 1 lies within the interval of ± 2 x 0.03, i.e. between o.46 and 0.58.

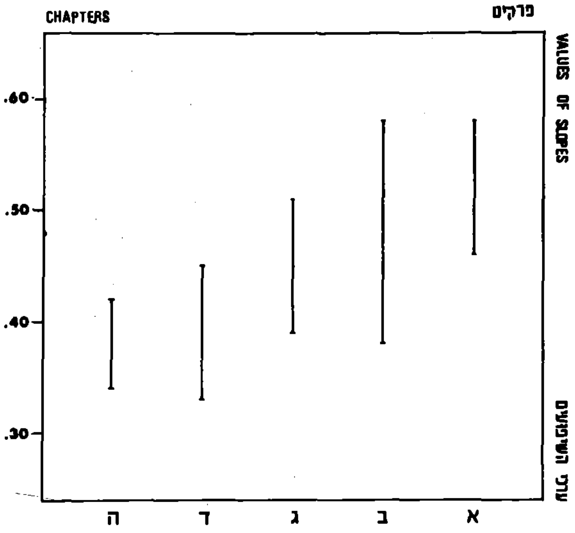

Fig. 3 CONFIDENCE INTERVALS OF FREQUENCY SLOPES

גרף 3. מרווחי־סמך של שיפועי השכיחות

Here we may, at long last, ask what one may deduce from the results of computations detailed so far with regard to three specific problems : (a) Who was the author of Lamentations ? (b) If Lamentations is not the work of one person, how many different hands may be discerned ? (c) Is the sequence of its poems random, and if not, what principle does it follow ?

The various arguments from style for or against Jeremiah's authorship have little value. Inference of two authors from the differences of style between Lamentations and the Book of Jeremiah is a plain *non sequitur* as the same author may make use of many styles, exactly as parallels between the two books are no sign of single authorship for such similarity may be due to influence by or to intentional imitation of another writer. We would be on firmer ground if the two books could be compared taking into account such characteristics of language behaviour as are truly author-specifying by virtue of their unconscious employment by the writer, and one of these may indeed be vocabulary richness(8). But as long as the Book of Jeremiah is not available on magnetic tape with full linguistic analysis of each word, any comparison between the two books is futile and judgment has to remain suspended.

Wether Lamentations was written by one or more poets is another matter. When it is divided into two, both halves differ from each other exactly as much as Canticles differs from Ecclesiastes (see Tables 6 and 7), which would be cause enough to pronounce a verdict of heterogeneity.

On the other hand, there are the small distances between each pair of adjacent chapters (see Table 8). Since it is without doubt not at all impossible for vocabulary richness to change - for certain reasons, one of which will be discussed later not only *inter auctores,* but also *intra auctorem,* one single property is not sufficient for deciding the issue. Incidentally, two more such properties have indeed already been investigated and found to remain constant throughout Lamentations[9]. Hence, single authorship is not precluded.

Let us now consider the structure of the book : did any principle guide the editor, whoever he may have been - perhaps the poet himself ? - when he collocated the poems in the present order, and if so, what principle ?

Two things are certain : the five dirges are arranged in an orderly fashion of increasing vocabulary richness and the odds of this being accidental are extremely low. Since for all practical purposes chance is excluded we must presume intention. But can anybody really fancy that an editor was conscious of a literary property which required complex and computer-aided calculations to bring to light ? Since this possibility is also to be ruled out, we must look for an alternative solution.

A writer is usually influenced by two impulses which are opposed to each other.

One is the urge to give expression to an experience which had a deep impact upon him. This will compel him to concentrate upon it and thus to repeat the same words over and over again. The more he identifies with it and the more acute and immediate the experience is, the more repetitive his language will become. A classic example is King David's lament on the death of his son Absalom (2 Sam 19 : 1) which consists of thirteen words, six of which are taken up by the same two lexical items. Thus, the effect of an author's intense involvement in an emotional experience and the immediacy of it to him is to limit his vocabulary.

The opposite impulse is the writer's wish to render his experience in all its nuances and to achieve maximal *mimesis* of what happened to him and within him. He will then refine his choice of words, carefully weigh each with as much sensitivity as he has and labor long and hard until he finds *le mot juste.* Writing of this kind presupposes proper perspective - the event seen from afar, so to speak. Such subtlety is made possible only by relative aloofness on the part of the writer which again is possible only if enough time has passed between the moment of his experience and his description of it. It may well be that is what Wordsworth had in mind when he wrote, in his prose preface to Lyrical Ballads (1803), "Poetry is the spontaneous overflow of powerful feeling : it takes its origin from emotion recollected in tranquillity". The effect of such distance is a delicately discriminating, or in other words, a rich vocabulary.

How to balance these two diametrically opposed impulses is a problem every writer must grapple with and solve for himself.

If we now recall that the five poems in Lamentations are arranged in ascending order in terms of vocabulary richness, we feel justified in inferring that there is a chronological principle at work here : the first being composed when the catastrophe had just occurred and the writer was still almost 'speechless', the last after time - years or decades - has passed, and the remaining three, one after the other, over the years intervening.

One cannot reproach Biblical scholars for the detachment which prevented them from correctly assessing which of the five poems is the most deeply felt : to do so, one has either to be a poet himself or to share personally in tbe author's grief. Neither of these qualifications is, in general, met in professional critics. No wonder, therefore, that the moving intensity of ch. 1 was argued not by them but by the German-Jewish poet-theologian Franz Rosenzweig (d.1924) who called it 'zu tiefst erschütternd' and by Rabbi Samson Raphael Hirsch (d. 1888) who regarded it as the most 'ergreifend'(10).

Objections to the hypothesis offered may be raised on three points.

Firstly, the claim that the latest dirge - the one of greatest vocabulary richness, usually seen as a positive quality in literature - is the best contradicts the generally accepted opinion that literary excellence declined in the Second Judean Commonwealth.

Secondly, it may be asked how one can rely, when evaluating and even dating a collection of poems, on one single criterion alone ? Thirdly, considerations such as certain - or, better, rather much uncertain - psychological urges at work in the creative mind are so imponderable and incommensurable as not to have a place in a scholarly discussion, even less so when it claims to be based on scientific, that is, in the present instance, statistical methods.

The first objection is easy to refute, since it grows out of a prejudice which itself should have no place in scholarship. Moreover, the time interval between the dates of the first and the last chapter may be so narrow - perhaps no more than a few years - that the terms 'early' and 'late' can not apply at all.

The second reservation is more valid, but does not obtain in this case. The present enquiry did not rely on one criterion, only : on the contrary, comparisons were made of distributions of word frequencies, and these consist of an entire series of values, each representing a different characteristic.

And despite this, we shall soon employ an additional criterion to make a double check.

Finally, paying attention to psychological processes is anything but 'out of bounds' in literary criticism. Furthermore, the new criterion just mentioned, while also 'psychological' to some extent, is indeed measurable.

It is called 'special vocabulary', a term coined and a criterion first used by Herdan. It has already been put to good use in previous authorship studies and means that the *realia* behind a literary work, the emotional situation of the writer, and the factual background of his creation will be reflected in the frequencies of words which are semantically connected with these three aspects of his work. These frequencies may then be employed when one work is compared with another, provided that the two deal with the same subject and are of the same literary type. These conditions are surely met in Lamentations. 'Special vocabulary' will be concerned with nouns only, for verbs cannot be semantically categorized with the same degree of certainty. In our case, we shall define as 'special vocabulary' all those nouns that belong to the semantic groups 'war' and 'grief'.

How is the proportion of 'special' nouns to be calculated ?

Herdan mentions three alternatives[11], namely

(a) p_V = the percentage of *different* 'special' nouns within the number of all *different* nouns;

(b) p_W = the percentage of the occurrences of 'special' nouns within the number of occurrences of all nouns; and

(c) p_C = the percentage of *different* 'special' nouns within the number of occurrences of all nouns;

and arrives at the conclusion that only $p_W$remains constant within the same sample and thus characterizes it.

Following Herdan, we drew up Table 11.

Table 11

"Special" Vocabulary in Lamentations[*]

Chapter	Total n° of Nouns	n°. of War-Nouns	n°. of Grief-Nouns		Standard Deviation of N_{W+G} in %
	N	N_W	N_G	N_{W+G}	s
	(1)	(2)	(3)	(4)	(5)
1	134	29	20	49	4.16
		21.64	*14.93*	*36.75*	
2	180	31	26	57	3.47
		17.22	*14.44*	*31.66*	
3	180	11	30	41	3.04
		6.11	*16.66*	*22.77*	
4	120	12	8	20	3.40
		10.00	*6.66*	*16.66*	
5	60	6	4	10	3.40
		10.00	*6.66*	*16.66*	

[*] Percentages in *italics.*

Col. 4 of Table 11 proves that the writer's preoccupation with war and grief gradually and almost at the same rates lessened from ch. 1 to ch. 5 - *quod erat demonstrandum*.

Table 11 supplies us with the data of an additional criterion which is unquestionably independent of vocabulary richness and yet correlated with it. How great this correlation is - and thus how much the two corroborate each other - is shown in Fig. 4. There, vocabulary richness is plotted on the horizontal axis *X* and values of p_W on the vertical one *Y*. The five points, each standing for one chapter, fall on a perfectly straight line and, what is more, exactly in their proper order in the book !

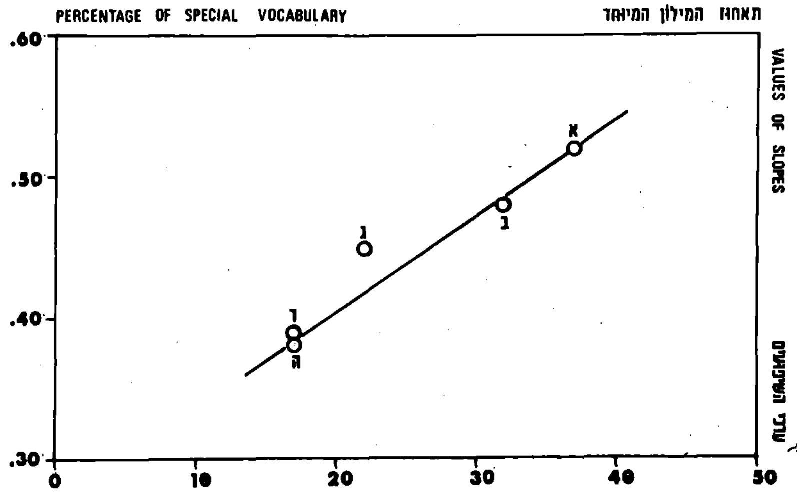

Fig. 4 FREQUENCY SLOPES VS. SPECIAL VOCABULARY

גרף מס 4 שיפועי השכיחות מול המילון המיוחד

The independence of the two variables warrants further examination, to be sure, but for the time being there seems no reason to question it. Since, however, both are in some way associated, we may assume that they are both dependent on a third variable. We cannot think of any other but the factor of time.

Is it perhaps possible, it will be asked, to assign this association to mere chance ? Anticipating the reluctance of Biblical criticism to renounce its preconceived ideas such as that chs. 2 and 4 are 'cognate' and ch. 3 a 'foreign body' - two conceptions which were here shown to be untenable - and further anticipating man's inclination to call 'a chance event' the one that does not conform to his *a priori* conceptions, we calculated the probability of randomness for the two independent variables examined behaving as they do in Lamentations and as shown in Fig. 4 : the odds are 1:14400.

Let us return for a moment to Table 10. There we read that of the five coefficients of slopes (col. 1) one only, namely that of ch. 1, exceeds 0.50. If we assume that Zipf's Law is valid for Lamentations - and all our findings confirmed it - then any slope, the coefficient of which is 0.50 or more, points to a finite vocabulary, even when the length of the text is infinite. In such a case, then, no new words will occur after the text has arrived at a certain length, no matter how long it continues. This has been probed by Mandelbrot who mentions that in all his investigations he has met two cases only where the slope coefficient surpassed the critical value(12). One is the Latin writings of the

German monk Notker, Abbott of St. Gallen (9th cent.) who occasionally wrote in his second language and not in his mother tongue, and apparently had some difficulty in expressing himself(13). The second case is that of publications in Modern Hebrew in then Palestine shortly after the revival of this language, when speakers were still confined to the vocabulary of classical Hebrew only. To these two cases, another can now be added : Lamentations ch. 1. Since one cannot be doubt that its author, contrary to Notker and to early Jewish settlers in Palestine, wrote in his mother tongue, the only explanation for his limited vocabulary is that he wrote this one poem while still suffering the emotions attendant on the fall of the Temple, the city, and the Judean kingdom. Scripture itself testifies to this when it says, in Psalm 137, "How shall we sing ? " and"Let my tongue cleave to the roof of my mouth".

Yehuda T. RADDAY
Moshe A. POLLATSCHEK

NOTES

* The authors reported on this enquiry for the first time in a very short paper read by them at the Seventh World Congress of Jewish Studies, Jerusalem 1977. A slightly longer version is forthcoming in Hebrew in Balšanut. This has been vastly expanded for the present article which includes material never published before.

(1) See P. Guiraud, *Les caractères statistiques du vocabulaire* (Paris : Presses universitaires), p. 32, n.1, and passim throughout, e.g. p.62.

(2) *ibid.*, p. 51.

(3) e.g. in H.M. Blalock, *Social Statistics*^{2} (New York : McGraw-Hill 1972). pp.276-87.

(4) For instance, R. Pfeiffer, *Introduction to the Old Testament* (New York : Harper 1941), pp.720-23.

(5) For.instance, *Encyclopedia Judaica,* vol.10 (Jerusalem : MacMillan 1971), s.v. Lamentations.

(6) This law has been widely discussed. It was first published by G.K. Zipf in *Selected Studies of the Principles of Relative Frequencies in Language* (Boston : Harward University Press 1932).

(7) See B. Mandelbrot, "On the Theory of frequencies and on Related Markovian Models of Discourse," *Structure of Language and its Mathematical Properties,* ed. R. Jakobson (Providence : American Mathematical Society 1961).

(8) For an example of dealing with a problem of dubious single authorship by relying a whole battery of criteria, see Y. T. Radday and D. Wickmann, "The Unity of Zechariah Examined in the light of Statistical Linguistics", *Zeitschrift für die Alttestamentliche Wissenschaft,* 87 (1975), Heft 1, 30-55.

(9) These two criteria are the use of the definite article and the frequency of the particle *and.* Concerning the first, see Y.T. Radday and H. Shore, "The Definite Article - An Author - and/ or Type-specifying Discriminant in Biblical Literature", *Bulletin of the Association for Literary and Linguistic Computing,* 3 (1976), 23-29; and concerning the latter, *id., "And* in Isaiah", *Revue de l'Organisation internationale pour l'Etude des Langues anciennes par Ordinateur,* 2 (1974), 52-67.

(10) Rosenzweig's and Hirsch's appreciations of ch. 1 are here quoted from memory. We were unable to trace the exact references.

(11) G. Herdan, *The Advanced Theory of Language as Choice and Chance* (Berlin-Heidelberg-New York : Springer Verlag 1966), pp.78-83.

[12] Quoted from an oral remark by Mandelbrot as published in *A Critical Review of Thermo-dynamics,* ed. E.B. Stuart et al. (Baltimore : Mono Book Corporation 1970), p. 230.

[13] In medieval literature he is frequently called Notker Balbulus, i.e. the stammerer. Whether this appellation refers to a speech defect or his poor Latin vocabulary is hard to ascertain. In the latter case, it would be a welcome confirmation of Mandelbrot's findings.