

# Stems and Roots

## Some Statistics Concerning the Verbal Stems in the Hebrew Bible\*

Arian Verheij (Amsterdam)

### § 0 Introduction

The *Verbal Stem* is one of the many intriguing features of Biblical Hebrew. Whether there is a system of stems, and if so, what kind of system, are as yet unsolved problems<sup>1</sup>. Statistics have played virtually no role in the discussion so far. The numerical differences between the stems have in fact been taken for granted. The present paper, by its statistical approach, aims at introducing a new element into the debate, rather than adducing evidence in support of any established position. I shall demonstrate in it that the productivity of the stems as attested in the Hebrew Bible is regular.

### § 1 Concepts, the Problem

*Verbal Stem* is seen in the present analysis as a morphological category with several realizations (the *stems* of qal, hifil, etc.<sup>2</sup>), whose distribution<sup>3</sup> is connected with the

---

\* Paper read at the IIIrd International Conference of the *Association Internationale Bible et Informatique* (Tübingen, August 26-30, 1991).

<sup>1</sup> Especially B.K. Waltke – M. O'Connor, *An Introduction to Biblical Hebrew Syntax*, Winona Lake 1990, pp. 354ff (henceforth to be referred to as WoC), – based to a large extent on E. Jenni, *Das hebräische Pi<sup>c</sup>el: Syntaktisch-semasiologische Untersuchung einer Verbalform im Alten Testament*, Zürich 1968 – , contains a very valuable attempt to give an outline of the stems as a system, transferring the discussion from the lexicon to syntax. Cf. also M.H. Goshen – Gottstein, *The System of Verbal Stems in the Classical Semitic Languages*, in: *Proceedings of the International Conference on Semitic Studies Held in Jerusalem, 19-23 July 1965*, Jerusalem 1969, 70-91, who deals with the problem mainly from a diachronical-morphological point of view.

<sup>2</sup> The stems are for the moment assumed to be well-defined morphologically. The discussions concerning passive qal and other variants are left aside.

<sup>3</sup> The distribution of the stems over the Hebrew Bible is not systematically addressed by the classical grammars. GKa § 51f signals the use of nifal instead of passive qal but does not locate this usage in particular books (cf. F.E. König, *Historisch-kritisches Lehrgebäude der Hebräischen Sprache II*, Leipzig 1895, § 121.2.d.g, III, Leipzig 1897, § 99-100; Joüon § 51c). König III § 101 suggests an increase of the passive hitpa'el from earlier to later Biblical Hebrew (cf. WoC § 26.1.3; E.Y. Kutscher, *A History of the Hebrew Language*, edited by R. Kutscher, Jerusalem/Leiden 1982, pp. 36f, 127). A. Kropat, *Die Syntax des Autors der Chronik verglichen mit der seiner Quellen. Ein Beitrag zur historischen Syntax des Hebräischen* (BZAW 16), Gießen 1909, p. 14f, signals a preference in Chronicles for transitive and active stems/verbs, esp. hifil (cf. the diagrams 2.2.1, A2, and A3 in this article, and note 4). Concerning Qumran Hebrew, cf. E.Y. Kutscher, *The Language and Linguistic Background of the Isaiah Scroll (IQIsa<sup>a</sup>)*, Leiden 1974, pp. 358ff, E. Qimron, *The Hebrew of the Dead Sea*

vocabulary of a text. The numbers of verbal forms in each of the stems, found in a specific text (*occurrences*, *O*), and the number of verbal *roots* that are involved (*R*) will serve as parameters for a description of the productivity of the stems as attested in the text.

The question then is whether these parameters are related, statistically. If they are, we will have found a pattern, a regularity in the productivity of the stems.

Our perspective, it should be stressed, will be that of the stems, not that of the roots. The fact that many roots are found in several stems<sup>4</sup> will not play a role in the present analysis. Related questions, such as: how many roots occur in how many stems, which roots occur in which stems, etc. will be equally left unconsidered.

## § 2 Data

### 2.1 *The Entire Hebrew Bible*

E. Jenni and C. Westermann<sup>5</sup> provide figures concerning the use of the verbal stems in the whole of the Hebrew Bible. Their diagram contains, among other data, the basic parameters of *O*(ccurrences) and *R*(oots) just described. The figures from the diagram that are relevant to the present discussion are reproduced below, in an adapted format<sup>6</sup>.

Diagram 2.1 Hebrew Bible

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	49180	(68.8)	1115	44.1	0.04
hi	9370	(13.1)	505	18.6	0.04
pi	6450	(9.0)	415	15.5	0.04
ni	4140	(5.8)	435	9.5	0.02
hitp	830	(1.2)	175	4.7	0.03
other	680	(0.9)	130	5.2	0.04
pu	460	(0.6)	190	2.4	0.01
ho	400	(0.6)	100	4.0	0.04
total	71510	(100 %)			

Scrolls, Harvard Semitic Studies 29, Atlanta 1986, p. 49). On Mishnaic Hebrew, M.H. Segal, *A Grammar of Mishnaic Hebrew*, Oxford 1927 / repr. 1986, p. 63.

<sup>4</sup> This overlap can vary between books, which is one reason for occasional differences in distribution. Comparing the distributions of the stems as presented in the Appendix, one finds that the proportion of hifil in the book of Chronicles (18%) is greater than in any of the other texts (10 to 13%) (cf. note 3). The fact is, however, that the number of roots used in the qal is rather low. This causes a low proportion of verbs in the qal, and therefore a higher proportion of the other stems, especially the hifil.

<sup>5</sup> THAT II, p. 542. Reproduced in WoC § 21.2.3e.

<sup>6</sup> The stems are listed here in descending order of frequency of occurrence; data concerning *hapax legomena* have been omitted; the columns O/R and O/R<sup>2</sup> are not in THAT.

As to a possible correlation between Occurrences and Roots, we see that these figures differ according to almost exactly the same pattern. But the Roots figures vary not nearly as strongly as the Occurrences do. There are about 5 times as many qal occurrences as hifil ones, but the number of roots is only just over twice as large. This applies to the other stems as well. The numbers of occurrences are not directly proportional to the numbers of roots. Or, to put it differently, the average numbers of occurrences per root are not the same for all stems<sup>7</sup>. The O/R column of the diagram makes this clear.

The remarkable thing now is that, statistically, the O/R figures are correlated with the R figures in a linear correlation<sup>8</sup>. The question of § 1 can therefore be answered affirmatively. The more roots are used in a particular stem in the Hebrew Bible, the more often – on average – each of these roots is used in that stem. The correlation can be formalized as  $O/R = qR$ . This implies that the number of occurrences of a particular stem is correlated with the squared number of roots used in this stem:  $O/R^2 = q$ . The fifth column of the diagram shows the values of  $O/R^2$  for each of the stems.

Two things should be considered in order to understand this result properly. In the first place, these figures are obviously not all exactly the same. But I feel confident that they are sufficiently similar to suggest, through statistical interpretation, that the productivity of the stems is described with some accuracy in the formula of  $O/R^2 = q$  or  $O = qR^2$ . And secondly, a statistical correlation as the one just proposed does not prove anything and contains no indication as to any real state of affairs. It is no less and no more than an evaluation and interpretation of figures.

But we now understand, that, if the formula is correct, the number of qal roots may be not per chance „just over twice“ as large as the number of hifil roots, as we said earlier, but, quite regularly, ca.  $\sqrt{5}$  times as large.

## 2.2 Parts of the Hebrew Bible

The discussion so far has been about the Hebrew Bible as a whole, and about all stems. The formula just established seems to contain no reference to any specific stem or any specific text. It therefore looks as if it is valid for any stem in any text. However, the value of  $q$  has been deliberately represented as variable. It was specified so far only for the Hebrew Bible as a whole, and found to be between 0.03 and 0.04. But it may vary under specific conditions. To find out, we have to manipulate the numbers of Roots, which can be done by examining parts of the Hebrew Bible separately. We will then be able to see if the numbers of Occurrences vary accordingly. In doing so we must bear in mind the fact that small numbers are less reliable than larger ones. Deviations will be greater, relatively speaking, in

<sup>7</sup> If they were, with  $y$  occurrences per root for each stem, there would be  $1115y + 505y + \dots = 3065y = 71510$  occurrences. The value for  $y$  would be 23.33, so that we would expect to have some 26014 qal, 11782 hifil, 9682 piel, 10149 nifal, 4083 hitp, 3033 other, 4433 pual, and 2333 hofal forms.

<sup>8</sup> Pearson's correlation coefficient for the R and O/R columns in the diagram is 0.979. The probability of chance is < 1%.

smaller texts than in larger ones. The figures of  $O/R^2$  for the numerically minor stems in separate books are not likely to be constant anyway<sup>9</sup>.

As an example of a separate text we will take a look at the figures for the book of Chronicles<sup>10</sup>.

Diagram 2.2 Chronicles

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	2829	(65.2)	236	12.0	0.05
hi	781	(18.0)	126	6.2	0.05
pi	325	(7.5)	81	4.0	0.05
ni	223	(5.1)	66	3.4	0.05
hitp	121	(2.8)	30	4.0	0.13
other	19	(0.4)	6	3.2	0.53
pu	20	(0.5)	14	1.4	0.10
ho	19	(0.4)	11	1.7	0.16
total	4337	(100 %)			

The diagram shows two things. In the first place, the pattern  $O/R^2 = q$  applies quite neatly to the four most common stems in book of Chronicles. The figures for the less frequent stems are too small to conform to the rule, as we expected. Secondly, we see that the value of  $O/R^2$  is higher in Chronicles than in the whole of the Hebrew Bible. The roots that occur in a certain stem are used more often, relative to their number, in Chronicles than in the entire Hebrew Bible.

The reader is referred to the samples in the Appendix<sup>11</sup> to find that in this respect the book of Chronicles would seem to be representative of prose in general. The poetical works that are listed, on the other hand, show a value of  $O/R^2$  that is lower than the figure for the whole Bible. A root in a certain stem is used roughly twice as often – relative to the squared number of roots – in Genesis, Samuel, Kings, and/or Chronicles, as it is in Psalms and/or Isaiah (which is a mathematical way of suggesting that the vocabulary in poetic texts is relatively more varied than in prose ...). Obviously, the figure for all the material is somewhere between those of the constituent parts.

### § 3 Conclusion

The figures lead us, in my opinion, to an important conclusion: the distribution of the possible realizations of the morphological category of *Verbal Stem* in the Hebrew Bible is regular. The regularity is in the linear correlation between the

<sup>9</sup> The value of  $O/R^2$  is greater than 0.04 if there are less than 25 roots to a particular stem in a particular text, since  $O \geq R$ .

<sup>10</sup> The figures of Diagram 2.2 and the Appendix were retrieved from the computer facilities of the *Werkgroep Informatica*, Faculteit der Godgeleerdheid, Vrije Universiteit, Amsterdam.

<sup>11</sup> The Appendix covers some 45% of the verbs of the Hebrew Bible.

numbers of occurrences of these realizations and the squared numbers of verbal roots used in any not too small corpus of biblical text. The multiplying factor in this correlation, however, is greater in prose than in poetry.

There are quite a number of things these figures do *not* tell us. They do not tell us what *stem* is, apart from a well-known morphological category, a parameter of any Hebrew verbal form. The question of whether and how the grammatical (or semantic) functions of the stems are related to each other is left untouched by our findings. We do not know from them why the numbers of roots differ the way they do, nor is it clear what is behind the regularity we encountered.

These are questions to which a statistical analysis may in the end not be able to provide definitive answers. Such a bird's-eye-view approach may be inadequate to penetrate the inner workings of the system (there would seem to be one, now!) of verbal stems. Yet statistics can provide a sound empirical basis for the discussion and raise it above the level of contingent examples. And even regarding the questions just mentioned, I think that a more refined analysis (taking into account e.g. the overlap between the stems, cf. § 1) will considerably clarify the matter<sup>12</sup>.

*Abstract:*

The number of verbal forms of any particular stem in a specific text is statistically correlated with the squared number of verbal roots used in that stem in that text.

*Address of the author:*

*Dr A.J.C. Verheij, 1e Nassaustraat 11-C, NL 1052 BD Amsterdam, Nederlande*

---

<sup>12</sup> After the completion of this article, J. Hoftijzer delivered his valedictory lecture: *Een Kwesie van Vraagstelling*, Leiden 1991, in which lines for such research are set out.

## Appendix

Diagram A1 Genesis

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	3861	(76.4)	310	12.5	0.04
hi	505	(10.0)	104	4.9	0.05
pi	359	(7.1)	78	4.6	0.06
ni	211	(4.2)	84	2.5	0.03
hitp	66	(1.3)	31	2.1	0.07
other	4	(0.1)	2	2.0	1.00
pu	32	(0.6)	12	2.7	0.22
ho	17	(0.3)	10	1.7	0.17
total	5055	(100 %)			

Diagram A2 Samuel

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	4606	(73.4)	315	14.6	0.05
hi	768	(12.2)	143	5.4	0.04
pi	405	(6.5)	99	4.1	0.04
ni	311	(5.0)	106	2.9	0.03
hitp	130	(2.1)	39	3.3	0.09
other	15	(0.2)	6	2.5	0.42
pu	11	(0.2)	9	1.2	0.14
ho	27	(0.4)	12	2.3	0.19
total	6273	(100 %)			

Diagram A3 Kings

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	4582	(75.4)	282	16.2	0.06
hi	714	(11.8)	134	5.3	0.04
pi	419	(6.9)	86	4.9	0.06
ni	207	(3.4)	66	3.1	0.05
hitp	92	(1.5)	33	2.8	0.09
other	8	(0.1)	1	8.0	8.00
pu	20	(0.3)	13	1.5	0.12
ho	32	(0.5)	9	3.6	0.40
total	6074	(100 %)			

Diagram A4 Genesis + Samuel + Kings + Chronicles

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	15878	(73.0)	530	30.0	0.06
hi	2768	(12.7)	232	11.9	0.05
pi	1508	(6.9)	200	7.5	0.04
ni	952	(4.4)	168	5.6	0.03
hitp	409	(1.9)	86	4.8	0.06
other	46	(0.2)	9	5.0	0.56
pu	83	(0.4)	38	2.2	0.06
ho	95	(0.4)	31	3.1	0.10
total	21739	(100 %)			

Diagram A5 Psalms

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	3486	(60.1)	513	6.8	0.01
hi	883	(15.2)	198	4.5	0.02
pi	892	(15.4)	190	4.7	0.02
ni	310	(5.3)	117	2.6	0.02
hitp	125	(2.2)	57	2.2	0.04
other	49	(0.8)	13	3.8	0.29
pu	47	(0.8)	29	1.6	0.06
ho	9	(0.2)	8	1.3	0.14
total	5801	(100 %)			

Diagram A6 Isaiah

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	3243	(65.5)	497	6.5	0.01
hi	668	(13.5)	189	3.5	0.02
pi	421	(8.5)	156	2.7	0.02
ni	392	(7.9)	159	2.5	0.02
hitp	87	(1.8)	46	1.9	0.04
other	31	(0.6)	15	2.1	0.14
pu	77	(1.6)	55	1.4	0.03
ho	32	(0.6)	23	1.4	0.06
total	4951	(100 %)			

Diagram A7 Psalms + Isaiah

	O	(%)	R	O/R	O/R <sup>2</sup>
qal	6729	(62.6)	697	9.7	0.01
hi	1551	(14.4)	280	5.5	0.02
pi	1313	(12.2)	248	5.3	0.02
ni	702	(6.5)	212	3.3	0.02
hitp	212	(2.0)	88	2.4	0.03
other	80	(0.7)	22	3.6	0.17
pu	124	(1.2)	76	1.6	0.02
ho	41	(0.4)	28	1.5	0.05
total	10752	(100 %)			