The definite article: a Type - and/or author - specifying discriminant in the Hebrew Bible

2 authors, including:

Haim Shore
Ben-Gurion University of the Negev
121 PUBLICATIONS   **829** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Statistical Modeling of Work-Process Duration View project

Project   Simple Approximations for Statistical Distributions View project

Association for Literary and Linguistic Computing

BULLETIN

CONTENTS

THE DEFINITE ARTICLE:

A TYPE- AND/OR AUTHOR-SPECIFYING DISCRIMINANT IN THE HEBREW BIBLE[1]

Y. T. Radday and H. Shore *(Israel Institute of Technology,*

Haifa; Tel Aviv University, Israel)

*   *   *   *

Abstract

The following claims are investigated and statistically substantiated: firstly, the frequency of the definite article in Biblical Hebrew is a rough yardstick for measuring the poeticity of a book as the two are negatively correlated;   secondly, when the *arc-sin* test was applied for uniformity of the use of the article in samples drawn from the same books, a significant result was found in all cases of a clear change of literary type and, in addition, in almost every case where a break in authorship had previously been suspected by Higher Bible criticism.

*   *

All those who are engaged in investigations of disputed or multiple authorship are in constant search of reliable discriminants.   In order to qualify, these should be independent of each other, not necessitated by subject matter, measurable, and tested in control groups.   Moreover, one should insist on such criteria of which the author himself could not have been aware because only then can they not − *a fortiori* − be imitated.   Preferably, they should also be so minute that they may be counted precisely by a computer only.   Finally, they and their uses should be neither dictated by literary type nor by rules of grammar and syntax.   The ideal criterion will therefore be one the use of which is optional, i.e. its realization will in no way impair or change the meaning of the text.   We believe we have discovered another such discriminant in the Hebrew definite article.

While there is no indefinite article in Hebrew, the use of the definite is left almost entirely to the discretion of the writer.   The Hebrew definite article is the morpheme *ha-* proclitically and ligatively bound to the word: *shalom* = peace; *hashalom* = the peace.   A good example of its indiscriminate use is Psalms 103. 11, rendered in the Revised Standard Version 'as *the* heavens are high above *the* earth' where sense and parallelism would have demanded either the use of *ha-* before both nouns or before neither.   In fact, *ha-* occurs before 'earth' while it is absent before 'heavens'.

The article cannot occur before every word, but, more or less as in English, only before nouns, adjectives, and numerals, and never before verbal forms other than participles.   It may, though, also occur before certain toponyms, particularly names of mountains and rivers.   This short introduction to Hebrew should suffice for comprehending how occurrences were counted.

Of the traditionally twenty-four but actually thirty-eight books of the Hebrew Bible, we had eight available on magnetic tape, with every single word analyzed and the exhaustive information numerically encoded and registered.   A considerable number of the remaining books is so short (a few hundred words each) that we were able to scan them manually.   Thus, much of Biblical literature was combed through for the use of the article in its entirety, and the rest in samples of two thousand words each on the average.   Five books were left out: Numbers and Jeremiah,

because they contain a frequently alternating mixture of literary types, Chronicles because of its many and lengthy genealogies, and Ezra and Daniel because they are written partly in Hebrew and partly in Aramaic.

Most books were arbitrarily divided into two equal halves in order to find out whether the frequency of the article remained constant in both. In a few cases where critics ascribed certain sections to different authors, their opinions were taken into account and the books partitioned into three sections. In six other books a clear caesura of literary type occurs at a certain point, hence they were divided accordingly. Table 1 shows the divisions of books in detail.

In Table 1, N is the sample size in terms of words and n the number of potential carriers of the article within N. Asterisks mark those books where the samples make up the entire length of the book. The detailed reasons for dividing a certain book in the way it has been done have no place in the present article.

The next step was counting the occurrences of *ha-* and calculating their relative frequencies. Obviously, they must refer to n and not to N. With these data in hand we proceeded to test statistically whether the use of the article remained uniform in books of unquestioned single authorship; whether it changed from one literary type to another; and whether a significant break was noticeable where critics suspected a break in authorship.

The statistical test applied was as follows. Two samples are given, S_1 and S_2, their sizes being n_1 and n_2, and the times the article occurs in them x_1 and x_2, respectively. Its relative frequencies are

$$p_1 = \frac{x_1}{n_1}, \quad p_2 = \frac{x_2}{n_2}.$$

These two variables are binomially distributed. What we want to know is whether the two samples may have been drawn from the same population, in which case $p_1 - p_2$ is a random deviation from zero. That $p_1 - p_2 = 0$ is the null hypothesis. According to the Central Limit Theorem, the two random variables p_1 and p_2 will be normally distributed, provided that the samples are of sufficient size. The parameters of this normal distribution,

$$U = \text{arc-sin}\sqrt{\frac{x_1}{n_1}}, \quad V = \text{arc-sin}\sqrt{\frac{x_2}{n_2}},$$

are mutually dependent. In order to obtain the variance σ^2 independent of p we apply the so-called *arc-sin* transformation

$$\mu = p \quad \text{and} \quad \sigma^2 = \frac{p \cdot (1-p)}{n},$$

respectively. The variables U and V have the variance

$$\frac{1}{4n_1} \quad \text{and} \quad \frac{1}{4n_2},$$

respectively. This enables us to standardize the two variables U and V in the following way:

$$\underline{z} = \frac{T-\mu}{\sigma} = \frac{(\underline{U}-\mu)-(\underline{V}-\mu)}{\sigma}$$

$$= \frac{\arccos\sqrt{\frac{x_1}{n_1}} - \arccos\sqrt{\frac{x_2}{n_2}}}{\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}}.$$

Allowing for an α error probability of 5%, we get the following instructions:

if $|\underline{z}| < 1.96$, do not reject the null hypothesis of a common population, otherwise reject it;

if $\alpha = 1\%$, substitute 2.58 for 1.96.

We may now return to our actual samples as found in our texts.   (See Table 2.)

Table 2 presents the relative frequency of the article per sample with percentages calculated for $\underline{n}$.   Column 3 shows that frequencies fluctuate very widely: from 73% (Joshua A) down to 2% (Proverbs A), a first and most welcome sign that the article may be a characteristic indeed.   A second and not less welcome detail is that, in many cases, the percentages in samples of the same book lie closely together, e.g. in Leviticus, Micah, Nahum, Psalms, Job, Canticles, Lamentations, Ecclesiastes, Nehemiah, and particularly Haggai and I Samuel.   Closely lying together is of course a loose term and not to be used in statistics.   We therefore calculated the arc-sin variable $\underline{z}$ for each sample pair.   The results are given in columns 4 and 5.   Values of $\underline{z}$ fluctuate between 19.33 (Joshua) and 0.00 (I Samuel).

The next two tables display the values of $\underline{z}$ again, but in a more lucid arrangement.   Table 3 lists those books for which values of $\underline{z}$ below the critical limit were found, in decreasing order of $\underline{z}$.   All books appearing in Table 3 are homogeneous with regard to their literary type and not one among them is ascribed by any scholar to more than one author each, except perhaps I Kings and Psalms.   The average value of $\underline{z}$ is comparatively low, namely 1.048, about half the critical one.

Table 4 lists books the values of $\underline{z}$ of which were found to surpass 1.96, and arranges them in increasing order of $\underline{z}$.   We notice, first of all, that in most, i.e. eleven among fourteen cases, the change in employing the article is so remarkable that $\underline{z}$s exceed even the critical value of 2.58 (see the horizontal line), in other words, even when the more severe yardstick of $\alpha = 1\%$ is applied.   Table 4 is of course the crucial point of the *ha*-experiment: whether the present criterion may serve as a discriminant depends on whether the books mentioned in Table 4 are indeed those where literary genres sharply alternate or where multiple authorships had already been claimed before by scholars on philological, historical, impressionistic, or other grounds.   Our hope that calculations would single out these books for inclusion in Table 4 did materialize – see columns 3-5.

In Habakuk, Exodus, Jonah, Ezekiel, and Joshua, the sudden switch in using the article or not results from an abrupt change in subject matter and/or literary type. Proverbs is extremely composite in character and perhaps origin, and in II Samuel and II Kings material originating in various sources may have been incorporated

which incidentally is the consensus of Biblicists.   In Joel, a second author who
wrote chapters 3-4, and in Zechariah, a second and third author who wrote chapters
9-11 and 12-14, respectively, have long ago been assumed by scholarship.   The
Samson Cycle (Judges, Sample B), the only instance in the entire Biblical canon of
glorified physical strength, is said to be a foreign body altogether - even the
Rabbis were already at a loss to explain its inclusion.   Whether the difference in
'article-behaviour' between the two samples of Genesis is due to a difference in
source or in genre (Sample B, the Joseph Cycle, is the only long short-story in the
Torah) is a matter for philologists to decide.

Finally, there is Zephaniah.   This little booklet of 511 words is, it must be
admitted, generally regarded as one unit, except for a few interspersed passages,
alleged to be later addenda.   Its $z=4.12$, a very high figure, is therefore a sur-
prise.   We found a similar result for Zephaniah, however, when we experimented with
'and' as a discriminant.[2]   There too, the book stood out as the only one among
those generally considered homogeneous where the use of 'and' varied significantly
between the two halves.   Perhaps interpreters and commentators should have a second
look at Zephaniah.

Be it as may, we believe that, despite this single puzzle among almost three
dozen books examined, it has been proved, at least to the satisfaction of statistici-
ans if not of Bible scholars, that the article is a valid and quite powerful dis-
criminant.   Yet it is even more.   It is also a measurement for poeticity, as shown
in Table 5.   Here, the relative frequencies are arranged in decreasing order.
Names underlined refer to entire books, the remaining to sections of a specific genre
in one and the same book.   The general appearance of Table 5 is that not one single
poetic work found its place in the uppermost left corner and no totally prosaic one
at the bottom right.   It is of course understood that poeticity must be measured -
if it can be measured at all - in other, additional and more relevant terms than by
counting the article, but the latter is, in spite of this reservation, a very rough
measure of the poetic character of a work: the lower its relative frequency, the
more poetic the book.

We mentioned at the beginning that the validity and power of a discriminant
should first be tested in control groups before using it in an actual case of
doubted unity.   In fact, this and little else has been done and reported upon so
far in this article in preparation for ultimately employing the *ha*-measure in the
Book of Isaiah, the *locus classicus* of disputed literary unity in the Hebrew
Biblical canon.[3]   Because this book deeply influenced Christianity and its founder
- its impact upon Judaism was, we believe, less strong - the question of its single
or multiple authorship generates much more heat than whether there was one
Zephaniah or two.   For this reason the topic '*The* in Isaiah' deserves detailed
treatment in a paper devoted exclusively to the problem and hopefully to be publish-
ed soon.

Notes

1.   This is a revised version of the first part of a paper read at the ALLC Sum-
     mer School, Cambridge, July 1975.   It is intended here primarily for non-
     Hebraists engaged in literary statistics.   The second half of the paper read,
     i.e. the one concerning the Book of Isaiah exclusively, will be published
     separately elsewhere and deal in a more thorough way with those aspects of
     Hebrew philology necessarily omitted here.

2.   See Y. T. Radday and H. Shore, '*Waw* coniunctivum - A Possible Author- and/or
     Type-Specifying Criterion in Biblical Literature', *Proceedings of the Sixth*

World Congress of Jewish Studies, August 1973 (forthcoming) (in Hebrew, with an abstract in English).

3.   See Y. T. Radday, *The Unity of Isaiah in the Light of Statistical Linguistics*, with a contribution by D. Wickmann, Collection Massorah, second series No.1, edited by G. E. Weil (Hildesheim: Verlag Dr H. Gerstenberg, 1973), 312 pp. (in English).   In this publication the Book of Isaiah was examined for realizations of more than twenty linguistic criteria which, however, did not include the definite article.

Table 1

IDENTIFICATION OF SAMPLES

Book	Sample	No. of Words N	n	Book	Sample	No. of Words N	n
Genesis	A(1-36)	3,286	1,361	*Micah	A(1-3)	588	300
	B(37-50)	1,440	608		B(4-7)	807	429
Exodus	A(1-20,32-34)	1,080	364	*Nahum	A(1-2:7)	230	118
	B(21-24)	540	209		B(2:8-3)	323	177
	C(25-31,35-40)	1,323	739	*Habakuk	A(1-2)	425	192
Leviticus	A(1-13)	791	269		B(3)	204	82
	B(14-27)	1,079	357	*Zephaniah	A(1-2:8)	247	167
Deuteronomy	A(1-15)	810	303		B(2:9-3)	264	168
	B(16-34)	810	315	*Haggai	A(1)	243	99
Joshua	A(1-12)	990	333		B(2)	273	179
	B(13-24)	990	350	*Zechariah	A(1-8)	900	322
*Judges	A(1-12)	4,512	753		B(9-11)	642	265
	B(13-16)	1,603	302		C(12-14)	739	328
	C(17-21)	2,359	596	*Malachi	A(1-2:13)	434	154
I Samuel	A(1-15)	1,000	542		B(2:14-3)	442	147
	B(16-31)	1,000	485	Psalms	A(1-77)	864	403
II Samuel	A(1-12)	1,000	478		B(78-150)	846	372
	B(13-24)	1,000	436	Proverbs	A(1-16)	720	413
I Kings	A(1-12)	1,000	530		B(17-31)	720	467
	B(13-24)	1,000	464	Job	A(3-21)	1,071	466
II Kings	A(1-13)	1,000	439		B(22-42)	1,071	452
	B(14-25)	1,000	551	*Canticles	A(1-4)	655	360
Ezekiel	A(1-39)	1,780	738		B(5-8)	629	358
	B(40-48)	1,448	766	*Ruth	A(1-2)	672	213
Hosea	A(1-3)	544	211		B(3-4)	561	175
	B(4-14)	1,815	670	*Lamentations	A(1,2,4,5)	1,135	549
*Joel	A(1-2)	578	311		B(3)	385	161
	B(3-4)	340	147	*Ecclesiastes	A(1-6:9)	1,483	663
*Amos	A(1-5)	1,160	481		B(6:10-12)	1,483	722
	B(6-9)	842	343	*Esther	A(1-5)	1,645	756
*Obadiah	A(1)	291	115		B(6-10)	1,600	664
*Jonah	A(1,3-4)	544	261	Nehemiah	A(1-7:5)	2,110	992
	B(2)	110	38		B(7:6-13)	2,245	1,016

Table 2

RELATIVE FREQUENCIES AND VALUES OF z

Book	Sample	Freq.	Comparison	z
Genesis	A	29.83	A:B	2.78
	B	23.52		
Exodus	A	32.14	A:B	2.61
	B	21.53	A:C	1.71
	C	27.47	B:C	1.49
Leviticus	A	50.87	A:B	1.49
	B	45.00		
Deuteronomy	A	32.67	A:B	1.55
	B	38.73		
Joshua	A	73.57	A:B	19.33
	B	32.00		
Judges	A	55.11	A:B	2.35
	B	47.02	A:C	0.37
	C	56.38	B:C	2.56
I Samuel	A	21.96	A:B	0.00
	B	22.27		
II Samuel	A	28.62	A:B	2.87
	B	27.52		
I Kings	A	26.60	A:B	1.48
	B	23.28		
II Kings	A	34.17	A:B	3.45
	B	24.50		
Ezekiel	A	16.80	A:B	8.07
	B	34.99		
Hosea	A	15.17	A:B	1.11
	B	12.09		
Joel	A	17.04	A:B	2.43
	B	27.21		
Amos	A	20.58	A:B	1.35
	B	24.78		
Jonah	A	28.74	A:B	4.55
	B	2.63		
Micah	A	12.67	A:B	0.82
	B	10.72		
Nahum	A	13.55	A:B	0.76
	B	11.30		
Habakuk	A	11.46	A:B	2.16
	B	9.76		
Zephaniah	A	33.53	A:B	4.21
	B	15.48		
Haggai	A	37.37	A:B	0.16
	B	37.43		
Zechariah	A	37.26	A:B	4.00
	B	21.51	A:C	0.26
	C	37.80	B:C	4.26
Malachi	A	17.53	A:B	1.48
	B	25.17		
Psalms	A	6.95	A:B	1.03
	B	8.60		
Proverbs	A	1.69	A:B	3.73
	B	6.64		
Job	A	6.22	A:B	0.61
	B	7.08		
Canticles	A	18.33	A:B	0.68
	B	19.83		
Ruth	A	26.76	A:B	1.70
	B	35.43		
Lamentations	A	10.38	A:B	0.71
	B	12.42		
Ecclesiastes	A	26.70	A:B	0.42
	B	27.83		
Esther	A	36.38	A:B	1.93
	B	40.51		
Nehemiah	A	35.68	A:B	0.91
	B	37.50		

Table 3

BOOKS IN DECREASING ORDER OF $z < 1.96$

Book	z	Book	z	Book	z
Esther	1.93	I Kings	1.48	Lamentations	0.71
Exodus A:C	1.71	Amos	1.35	Canticles	0.68
Ruth	1.70	Hosea	1.11	Job	0.61
Deuteronomy	1.55	Psalms	1.03	Ecclesiastes	0.42
Leviticus	1.49	Nehemiah	0.91	Haggai	0.16
Malachi	1.48	Micah	0.82	I Samuel	0.00
		Nahum	0.78		

Table 4

BOOKS IN INCREASING ORDER OF $z > 1.96$

z	Book	Argument *ex auctore*	Argument *ex genere*	Other
2.16	Habakuk		prophecy v. psalm	
2.43	Joel	a Deutero-Joel assumed by critics		
2.35	Judges A:B ⎫	'Samson' (=B) assumed		
2.56	Judges B:C ⎭	by critics to be of a different source		
2.61	Exodus A:B		story v. laws	
2.78	Genesis A:B			?
2.87	II Samuel			?
3.45	II Kings			?
3.73	Proverbs	composite		
4.00	Zechariah A:B	a Deutero-Zechariah assumed by critics		
4.12	Zephaniah			?
4.26	Zechariah A:C	a Trito-Zechariah assumed by critics		
4.55	Jonah		story v. psalm	
8.07	Ezekiel		vision v. description	
19.33	Joshua		story v. geographical lists	

Table 5

RELATIVE FREQUENCIES IN DECREASING ORDER

Narration and Prose	Narration and Poetry	Poetry and Narration	Poetry	Freq. in %
Joshua (narr.)				73.57
Judges				54.09
Joshua (total)				52.27
Leviticus				47.94
Esther				38.45
		Zechariah (Trito-)		37.80
Haggai				37.40
	Zechariah (Proto-)			37.26
Deuteronomy				35.76
Ezekiel (cult)				34.99
	Zechariah (total)			32.90
Exodus (narr.)				32.64
Joshua (geogr.)				32.00
Ruth				30.95
II Kings				28.78
Jonah (narr.)				28.74
Genesis				27.89
Exodus (total)				27.82
Exodus (tabern.)				27.47
			Joel (Deutero-)	27.21
Ecclesiastes				27.17
	Ezekiel (total)			26.06
	Jonah (total)			25.42
I Kings				25.05
			Zephaniah	24.51
			Amos	22.68
II Samuel				22.67
I Samuel				22.10
Exodus (covenant)				21.53
		Zechariah (Deutero-)		21.51
			Malachi	21.33
			Joel (total)	20.31
			Canticles	19.08
			Joel (Proto-)	17.04
			Ezekiel (prophet.)	16.80
		Hosea (biograph.)		15.17
			Obadiah	13.04
		Hosea (total)		12.83
			Nahum	12.43
			Hosea (prophet.)	12.09
			Micah	11.70
			Lamentations	11.40
			Habakuk	10.61
			Psalms	7.78
			Job	6.65
			Proverbs	4.16
			Jonah (psalm)	2.63