THE BALANCE OF PROBABILITY: STATISTICS AND THE DIACHRONIC STUDY OF ANCIENT HEBREW

Jarod Jacobs

E-mail:jarod.jacobs@gmail.com

(Received 02/02/2016; accepted 05/10/2016)

ABSTRACT

In this article, I discuss three statistical tools that have proven pivotal in linguistic research, particularly those studies that seek to evaluate large datasets. These tools are the Gaussian Curve, significance tests, and hierarchical clustering. I present a brief description of these tools and their general uses. Then, I apply them to an analysis of the variations between the "biblical" DSS and our other witnesses, focusing upon variations involving particles. Finally, I engage the recent debate surrounding the diachronic study of Biblical Hebrew. This article serves a dual function. First, it presents statistical tools that are useful for many linguistic studies. Second, it develops an analysis of the *he-locale*, as it is used in the "biblical" Dead Sea Scrolls, Masoretic Text, and Samaritan Pentateuch. Through that analysis, this article highlights the value of inferential statistical tools as we attempt to better understand the Hebrew of our ancient witnesses.

INTRODUCTION

Over the past several years, scholarship in Hebrew linguistics has focused on projects that rely upon large amounts of data. Two specific examples are Hornkohl (2014) and Rezetko and Young (2014). Both of these texts include a wealth of data and the authors interact with these data in various ways. However, their work can be strengthened by the application of robust statistical tools beyond simple graphs and tables. In the following pages, I will present three statistical tools that can help shed light on the current debate in diachronic Hebrew studies, the two sides of which can be represented by Hornkohl (2014) and Rezetko and Young (2014).

First, I briefly place this article in the context of corpus linguistics, which depends upon inferential statistical tools. Second, I present the concept of standard deviations and how it can be applied to the study of ancient languages. Third, I discuss hypothesis testing using significance tests. And fourth, I examine the use of hierarchical clustering. Throughout, I apply the theories I discuss to the example of variations involving particles found in the “biblical” Dead Sea Scrolls (DSS) and to the recent debate surrounding the diachronic development of the Hebrew language.

CORPUS LINGUISTICS

While statistical analysis is utilised in many fields of language study, statistical tools are most natural and necessary in corpus linguistics. Biber, Conrad, & Reppen (1998:4) summarises the essential characteristics of corpus linguistics with the following four points:

> [1] it is empirical, analyzing the actual patterns of use in natural texts; [2] it utilizes a large and principled collection of natural texts, known as a “corpus,” as the basis for analysis; [3] it makes extensive use of computers for analysis, using both automatic and interactive techniques; [4] it depends on both quantitative and qualitative analytical techniques.

The first three of these principles (pattern analysis, based on a corpus, and use of computers) necessitate the fourth, specifically quantitative techniques.[1] When a researcher uses a computer to analyse a corpus, which is often comprised of hundreds of thousands or even millions of words, it is very difficult to see patterns. When one does see a pattern, it is often impossible to know if that pattern truly exists or if the bias of the researcher has made it appear. This is where inferential statistics comes in. Once we have compiled our corpus and gleaned our data from it, we can use statistical tools to both identify possible patterns as well as determine which patterns are relevant and which are not. The first place to start is in the identification of patterns. For that, we can use the concept of standard deviations.

---

[1] While qualitative analytical techniques are also vital, they are beyond the scope of this article.

Standard deviations

The standard deviation of a set of data is the square root of the variance, which is the average of the squared differences between variation rates from the mean. In practical terms, we can use standard deviations to identify aspects of our data that are “normal” and aspects that are “abnormal”. By overlaying a Gaussian Curve (also known as a Bell-Shaped Curve) onto a plot of data and determining the standard deviation for the dataset, we can identify statistically relevant outliers. Ninety-five percent of data that follow a Gaussian Curve fall within two standard deviations of the mean. Any data point that falls outside of the ninety-five percent are considered statistically abnormal. All the data that fall within the curve should be considered statistically normal. Figure 1 presents a visual of this concept:

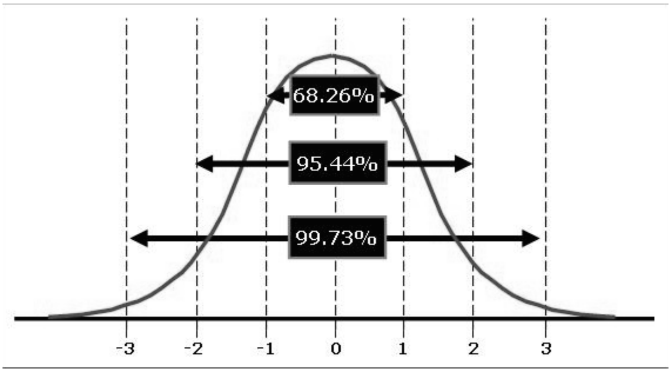

Figure 1
(“Normal distribution”; Six-Sigma-Material.com)

This graph shows that, for normally distributed data, 68.26% of data points fall within one standard deviation, 95.44% within two, and 99.73% within three. Any data point that falls to the left of negative two or to the right of positive two is considered abnormal. The place where a specific data point falls is normally reported as a z-score. Thus a data point that is 2.25 standard deviations to the left of the mean has a z-score of -2.25. A basic example of how this can be used is the average heights of people. While the mean adult height of a male born in Austria in 1950 is about 164 cm, most in that population are not that tall. But, most Austrian males born in 1950 (in fact 95% of all Austrian males born in that year) are between 150 cm and 178 cm (Garcia &

Quintana-Domeque 2007). In other words, ninety-five percent of Austrian males fall within two standard deviations of the mean height. Those outside this range can be considered abnormal (although we would likely use a different word to refer to this section of the population). This type of analysis is particularly useful for the study of the “biblical” DSS when comparing different types of variation rates within the entire corpus. I will turn to an analysis of variations involving particles in the “biblical” DSS to demonstrate the use of standard deviations.

For this analysis, I will focus on the most frequently occurring particles in the Hebrew Bible: the direct object marker (D.O.M.), the definite article (D.A.), the conjunction *vav*, the *he-locale* (H.L.), the interchange between אל and על, and the prepositions ל, ב, and כ. The following table presents the raw data for variations involving these particles:

Table 1

Global Variation Rate	D.O.M.	D.A.	*Vav*	H.L.	Preps
Pluses[2]	66	125	483	40	77
Minuses	22	84	314	27	32
Total Occurrences	1914	4043	9312	207	7120
Percent Increase	3.45%	3.09%	5.19%	19.33%	1.08%
Percent Decrease	1.15%	2.08%	3.37%	13.04%	0.45%
Total Variation	4.60%	5.17%	8.56%	32.37%	1.53%

The pluses row shows the number of places where the “biblical” scrolls contain each particle where the MT does not, while the minuses row shows the opposite. These numbers are normalised against the total number of occurrences of these particles within the “biblical” DSS (found in the fourth row) which results in the variation rates shown in the rows labelled percent increase and percent decrease. Finally, the total variation rate for each particle is listed in the last row.

If we were to simply look at the raw data, we might conclude that the conjunction *vav* is the most statistically relevant word for analysis due to the high number of variations. But the simple step of normalising these data and calculating the variation

2 I use “pluses” and “minuses” here and below to refer to pluses and minuses in the “biblical” Scrolls vis-à-vis the MT.

rates helps us to focus our attention. Of these particles, the H.L. has the highest variation rate by far. When these data are plotted on a graph, the difference between the rates of variation for the H.L. and the other particles becomes clear.

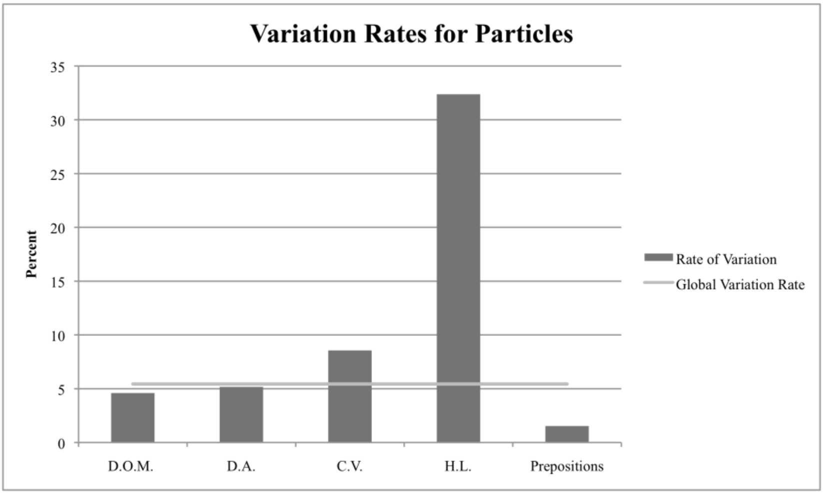

Figure 2

Figure 2 compares the variation rates of particles to the global variation rate (GVR)[3] within the "biblical" scrolls. While the rates of variation for the D.O.M., the D.A., *vav*, and prepositions are relatively close to the GVR, the variation rate for the H.L. is far higher. In order to determine if this difference is statistically valid or just slightly higher than expected, I will calculate the standard deviation for this dataset. Any data point that falls more than two standard deviations away from the mean can be considered statistically relevant. The standard deviation for this set of data is 11.37% and the mean is 9.61%. Thus two standard deviations from the mean is 32.35%. The variation rate for the H.L. is 32.37%, which is more than two standard deviations from the mean.[4] Therefore, we conclude that the high variation rate for the H.L. is statistically relevant and in need of further analysis.

3 Global Variation Rate (5.44%) = All variations including pluses, minuses, and substitutions of words (4,993) divided by the total words in the "biblical" DSS corpus (91,716).

4 While the variation rate for the H.L. is just 0.02% higher than the two standard deviations from the mean, this is enough to conclude that this result is statistically significant.

By using the standard deviation model, we are able to pin-point which variation types are outside the norm, thus helping us to identify issues for further research. Next, I discuss hypothesis testing and significance tests and how they can help analyse linguistic phenomena. While the following discussion is somewhat long and involved, it is essential for understanding and establishing the scientifically rigorous method that I will apply to the language variations in the Hebrew Bible.

Hypothesis testing

In the corpus linguistics approach to language study, a typical research question "take[s] the form: is variable x used differently in corpus A compared with corpus B" (Cantos 2012:103). In order to be answered, this question needs to be stated in the form of a hypothesis that can be tested. Typically this hypothesis states that any difference between the use of variable x in corpus A and corpus B is due to random chance. This is called the null hypothesis and is represented by "H_0". Once the null hypothesis is formulated the alternative hypothesis needs to be stated. Normally this takes a form similar to: variable x is used differently in corpus A compared with corpus B and this difference is not due to random chance. The alternative hypothesis is represented by "H_1". Once these hypotheses are developed, a statistical test is selected to analyse the truth of H_0. If H_0 is found to be false, then H_1 is assumed to be true.

Significance level

When a null hypothesis is analysed, the results are normally presented in the form of a significance level. Everitt (1998:345) defines "significance level" as "the level of probability at which it is agreed that the null hypothesis will be rejected". Significance tests report a percentage, the P-value,[5] which tells "the probability of the observed data (or data showing a more extreme departure from the null hypothesis) when the null hypothesis is true" (Everitt 1998:304). This score ranges from 100 percent to zero percent, with higher P-values telling us that the observed data had a high chance of

5 Also known as the Spearman's *rho* (Everitt 1999:315).

occurring if the H_0 is true and lower values telling us that the observed data had a low chance of occurring if the H_0 is true. The significance level is set by consensus at a specific point, so that if the P-value falls below this level we reject the null hypothesis as false. The significance level is arbitrarily set at five percent in many fields of study, including corpus linguistics and thus that is the benchmark that I utilise in this analysis.

Significance tests

At their most basic level, significance tests are used to test the validity of a null hypothesis. But more precisely a significance test is applied to a set of observations and produces a P-value related to a null hypothesis. The P-value states the probability of the data if the null hypothesis is true. The flipping of a coin helps to demonstrate the application of a significance test. First we develop our hypotheses:

H_0 The coin is fair

H_1 The coin is not fair (it has been tampered with)

Then we gather data. For this example let us assume we ran 100 tests of ten flips. We can then use a significance test, in this case a one-sample t-test, to produce a P-value. Let us assume our P-value came out as 98.7%. This tells us that if the H_0 is true then the observed data has a 98.7% chance of occurring. Since this P-value is far above the five percent significance level, we can accept the H_0 and reject the H_1. Now, let us imagine that we ran the same scenario with a different coin producing a different set of data. We then applied the one sample t-test to the data and got a P-value of 3.5%. This tells us that if the H_0 is true then this new set of data only has a 3.5% chance of occurring. Because this P-value falls below the five percent threshold we should reject the H_0 as untrue. We would therefore accept the H_1 and conclude that the coin is not fair and should be checked for tampering. Thus, the significance test helps us to sift through a large amount of data in order to test hypotheses.

Choosing the appropriate significance test is vital for obtaining reliable results. It is therefore important to understand the assumptions made by the tests being used, the purpose for which the tests were developed, and the possible applications for which they are useful. I will thoroughly examine the main test that I will utilise in this study and that I believe is most helpful for Hebrew linguistics, the Student's t-test developed by William Sealy Gosset in 1908.

The Student's t-test, or simply t-test, is used to assess a hypothesis based on population means. The t-test assumes that the data follow a normal distribution, equal variance, and independent sampling. When a set of data follows a normal distribution, its dispersion follows a Gaussian Curve. The t-test has been shown to produce reliable results even when the assumption of normalcy is not met as long as the dataset contains a large number of samples (less than 100 can prove to be enough at times) (Lumley, Diehr, Emerson, and Chen 2002). Normalcy tests have been used to see if each sample meets the first assumption of the t-test. To meet the assumption of equal variance, both datasets being compared must have equal internal variation within a margin of error. For this study, the f-test[6] has been used to test for equal variance.[7] The assumption of independent sampling states that all samples found in one dataset cannot be included in a second (Romaine 1982:107). This is true of all corpora in this study; no single manuscript is included in two corpora that are being compared. As an example, I would not compare all *tefillin* to all *plene* manuscripts as some of the *tefillin* are written with *plene* orthography. This sort of comparison would cause errors in the statistical calculations due to having some of the same witnesses in each corpus. The t-test is a good tool for this study because all of the data upon which my conclusions are based meet the three assumptions of the test. The corpus is normally distributed (roughly) and contains enough samples to overcome the slight skewness observed in some datasets. The t-test assumes equal variance, but there is an alternative version that allows for unequal variance: the heteroscedastic t-test. This version of the t-test is used throughout because all of the datasets used in this study

---

6 "A test for the equality of the variances of two populations having normal distributions, based on the ratio of the variances of a sample of observations taken from each. Most often encountered in the analysis of variance, where testing whether particular variances are the same also tests for the equality of a set of means" (Everitt 1998:153).

7 For datasets that do not have equal variances, I use the heteroscedastic t-test (which assumes unequal variances), as opposed to homoscedastic t-test (which requires equal variances).

have unequal variance. And, as I have just discussed above, all of the data have been independently sampled.

While I utilise the t-test throughout, there are three forms of this test that can be applied in different contexts. They are the one-sample t-test, the one-tailed t-test, and the two-tailed t-test. Each of these has different applications. I have already utilised the one-sample t-test above, but a further example and some explanation are needed.

The one-sample t-test is used to compare the mean of a single group of observations with a specific value (Altman 1991:183). The coin flip example that has already been used above provides a basic way of looking at this t-test. If we flipped a coin 100 times, recorded the results, and then wanted to know if those results are statistically different from what is expected, we could use the one sample t-test. This test would compare the results of the 100 flips to the expected value of 50/50 (or stated in the form of a mean, 50% heads). If the test produced a P-value of five percent or lower, we would conclude that our test results are abnormal and would question the fairness of the coin that we flipped. Thus the one sample t-test examines the possibility of obtaining the observed results given the expected mean.

The one sample t-test could be utilised in a number of contexts within Hebrew language studies. As an example, if one particular corpus, say all Torah manuscripts, displays a particularly high variation rate of a given type, this test could be applied in order to calculate the probability of that variation rate occurring given the overall variation rate for all manuscripts.

The other two forms of the t-test are related. Both of these tests analyse the means of two independent samples (e.g., the control group and the test group in a medical trial). These tests compute the similarity of each group given the internal variation of each and determines whether they are statistically the same. The Gaussian Curve is used by both tests. The mean of each group is calculated in relationship to the normal distribution of the data. If the mean of one group falls to the extremes of one of the “tails” of the curve, then that group is considered extreme. The use of the Gaussian Curve and the tails of this curve as a reference point gives rise to the names of each form of this test. The one-tailed t-test only calculates the probability of obtaining a statistically relevant result on one half of the curve (one tail). This test is regularly used in medical testing since researchers are often only trying to find out if a new drug produces better results in the test group compared with the control group. As an example, let us say a research group wants to test a new blood pressure medication. They know that the medication is likely to lower blood pressure, but they want to know if it lowers it more than the standard drug on the market. In this case, the one-tailed t-test can be used because the researchers know on which tail they should focus – the reduction of blood pressure. Any significant result for this research project will fall to one tail of the Gaussian Curve.

In contrast, the two-tailed t-test needs to be used when we do not know in which direction to look for significant results. As an example of this, let us return to the medical field. Imagine a group of researchers want to study the impact of coffee upon the well-being of people. They set up two groups. One group will drink coffee every day, the other will not. These researchers are interested in knowing if coffee improves the test group's overall feeling of well-being or if it negatively influences it. After running their experiment and gathering the data, the researchers would use a two-tailed t-test because significant results could be found at either tail of the Gaussian Curve. A non-significant result would show at the peak of the curve, while the researchers would be particularly interested in any result that fell to either tail.

He-locale

Let us return to the *he-locale* (H.L.). As noted above, the variation rate for the H.L. is 32.37% compared with the overall mean for particles of 10.44%. This high rate of variation, statistically speaking, marks these variations as significant and in need of further analysis and explanation.

Numbers 13:22 and Deuteronomy 10:22 are good examples of this relatively common variation:

Num 13:22
4Q27
ויבוא עד חברון ושמה אחימן

MT

ויבא עד חברון ושם אחימן

Deut 10:22

4Q128

תיכה מצרים[ירדו אבו

MT

ירדו אבתיך מצרימהה

Table 2 contains a summary of the data for the distribution of these variations across the “biblical” DSS corpus:

Table 2

Orth. Style	Number of Manuscripts	Number of H.L.s	Number of Pluses	Mean Percent Change (Pluses)	Number of Minuses	Mean Percent Change (Minuses)
Def.	52	82	4	2.05%	11	3.83%
Plene	27	102	36	14.30%	12	7.46%
N.E.D.[8]	179	28	5		0	
Total			45		23	

In the “biblical” DSS, there are 45 instances that have the H.L. where the MT does not.[9] On the other hand, there are 23 cases where the “biblical” scrolls do not have the H.L. where the MT does. These variations are spread over 29 manuscripts, five of which have both pluses and minuses.

Statistical analysis of these variations reveals some significant insights. While there are several different ways to enter into statistical analysis of these data, I will utilise the orthographic style of the manuscripts as a control. This approach has produced excellent results in work by Eibert Tigchelaar (Tigchelaar 2010). Those manuscripts that contain only the defective spelling of לא contain only four of the

8 N.E.D. stands for “Not Enough Data” and includes all manuscripts under 200 words (see Jacobs 2015:83–104 for justification of this choice) as well as those that do not contain any intact occurrence of the negation לא. This row is included in order to be comprehensive, but these data are not utilised for statistical testing. Because of this, and because of the lack of data for these manuscripts, mean percentages of change are not given as these data would simply be misleading.

9 In 38 of these places, the scroll contains the H.L. while the MT and SP do not.

additional H.L.s, an overall increase of only 4.88%. Those manuscripts that contain at least one *plene* spelling of לוא have 36 additional H.L.s, an increase of 35.29%. These numbers suggest a clear distinction between defective and *plene* manuscripts. Statistical testing confirms that this distinction is relevant and not just due to random chance.

In order to test the relevance of the difference between the use of the H.L. within the *plene* and defective manuscripts, I will utilise a two-tailed t-test to compare the corpora. First I will propose a null hypothesis and the corresponding alternative hypothesis.

> H_0 The difference in the use of the H.L. between the two corpora is due to chance.
> H_1 The difference in the use of the H.L. between the two corpora is statistically significant.

By applying a two-tailed t-test to the data a P-value will be obtained that will either confirm or reject the null hypothesis. Table 3 presents the results of the t-test.

Table 3

Compare Means			
Descriptive Statistics			
	Sample Size	Mean	Variance
Def.	52	0.02051	0.0055
Plene	27	0.14304	0.06678
Two-tailed Distribution			
P-value	0.02032	Critical Value (5%)	3.39826

A two-tailed t-test applied to these two corpora results in a P-value of 2.03 percent, well below the 5 percent threshold of statistical relevance. Therefore, we reject our null hypothesis and accept the alternative: The difference between the use of the H.L. in these corpora is statistically significant. The two-tailed t-test thus strongly confirms that the *plene* scrolls contain significantly more pluses of the H.L. than the defective scrolls.

A comparison of the “biblical” scrolls to the MT and Samaritan Pentateuch lends credence to this correlation between *plene* orthography and the pluses of the H.L. Thirty-eight of the 45 pluses found in the “biblical” DSS are not found in any other major Hebrew witnesses. This suggests that these variations where not inherited by the “biblical” scrolls from their *Vorlagen*.

An analysis of the minuses of the H.L. also supports the conclusion that the *plene* scrolls contain significantly more pluses of this particle. When a two-tailed t-test is applied to the data for the minuses with the same hypotheses as above, the following results are found:

Table 4

Compare Means			
Descriptive Statistics			
	Sample Size	Mean	Variance
Def.	52	0.0383	0.01131
Plene	27	0.07463	0.0255
Two-tailed Distribution			
P-value	0.2859	Critical Value (5%)	3.3677

With a P-value of 28.59% and means that are relatively close (especially when compared with those found for the pluses), we can conclude that there is no statistically significant difference between the *plene* and defective scrolls when the minuses of the H.L. are analysed. This stands in clear contrast to the results found for the pluses of this particle. Further analysis of the variations involving the H.L. reveals a more refined picture.

Twenty-one of the 45 pluses of the H.L. and seven of the twenty-three minuses are found on שם. All but one of the pluses are found in manuscripts that contain *plene* readings of לוא. The only exception is the one found in 11Q7, which does not contain לא in any form. In contrast, all but one of the minuses of the H.L. are found in manuscripts that do not contain any *plene* readings of לא. This data is summarised in Table 5:

Table 5

	Pluses to שם	Minuses from שמה
Defective	0	6
Plene	21	1
N.E.D.	1	0

The data in Table 5 show that the manuscripts that utilise the *plene* לוא also contain the long form of שמה, which shows a clear correlation between *plene* readings and the long form of this particle. The distribution of שמה is not random amongst the "biblical" DSS. On the contrary, there is a correlation between *plene* orthography and the long form of שמה. The correlation between this type of variation and the *plene* character of the manuscripts may suggest that the scribes who produced these readings favoured them more for their stylistic feel as opposed to any diachronic change in the Hebrew language that may have impacted their copying. Above I have shown that the H.L. is a plus on שם in the *plene* scrolls far more often than any other word. This suggests that the scribes favoured the long form of שמה. If the pluses of the H.L were more spread across a diverse range of nouns, then a diachronic explanation would be more plausible, but this is not what was found. The preference simply for the long form of שמה is consistent with the use of other long forms, particularly long suffixes such as המה. Thus, variations involving שמה could probably be considered simple orthographic variations as opposed to representing specific syntactic shifts in Hebrew.

Aside from this significant pattern of preference for שמה, there are two other smaller trends of note. The first focuses on the context in which nine of the 23 minuses of the H.L. are found. These particular variations result in the minus of the H.L. from a place-name (there is only one occurrence of a plus of a H.L. to a place-name). This seems likely to have been caused by the loss of the locative force of this suffix, and thus it may have appeared to the scribes as out of place (Muraoka 2000:207).

The second may be explained by what Muraoka calls "fossilized lexemes" (2000:207). He argues that Mishnaic Hebrew lost the H.L. all together, except in a few phrases such as למעלה. In the "biblical" DSS we find three places where the MT has מ/לעמל and the scroll has a plus of the H.L. These few pluses may have developed under the influence of the spoken language or later written register of the scribes, since they may have been more familiar with the long form. This is supported by the fact that the Mishnah only contains the long form. Yet the limited amount of data make any conclusions tentative at best.

The correlation between the *plene* spelling of לׄא and the increased use of the H.L. in the “biblical” DSS may be linked to the diachronic development of Hebrew. Several recent studies have analysed the development of the H.L. over time. In order to explore this possible interpretation of the above discussion, I will turn to those recent investigations to see what light they can shed on this issue.

STATISTICS AND RECENT STUDIES ON THE H.L.

The H.L. has played a significant role in the debate over the diachronic development of the Hebrew language as evidenced by the considerable attention given to this particle in Hornkohl (2014) and Rezetko and Young (2014). An examination of these two discussions will help to further the dual aims of the current article, namely to explore the use of statistics in Hebrew language studies and to analyse the use of the H.L. in biblical texts. However, due to the primary goal of this article being the former of these two, I will focus mainly on how inferential statistical tools can further develop the conclusions of Hornkohl and Rezetko and Young. With that focus in mind, I will limit the depth of my analysis of their work, while highlighting places for further research. For this part of my analysis I will utilise two statistical tools: hierarchical clustering and the t-test. While I have had occasion above to discuss the use of the t-test, I have not developed the application of hierarchical clustering to the Hebrew language. I will briefly introduce this statistical tool here and then will apply it to the discussions found in Hornkohl (2014) and Rezetko and Young (2014).

Hierarchical clustering

Hierarchical clustering is a cluster analysis method that begins with each individual in its own cluster and then combines individuals into ever bigger clusters, until finally all individual members are part of one group (Everitt 1998:7). This process is oftentimes visually presented with a dendrogram. A dendrogram is a tree-like figure that illustrates the successive grouping steps taken in hierarchical clustering. A comparison of dog breed heights provides a straight forward and intuitive example of how this works.

First, we start with our data:

Table 6

Dog Breed	Average Height
Chihuahua	9
Pug	11
Basset Hound	15
Golden Retriever	24
Saint Bernard	35

Using a statistical program such as R, we can calculate the distance[10] between each data point and group them together according to closeness. These calculations produce the following dendrogram:

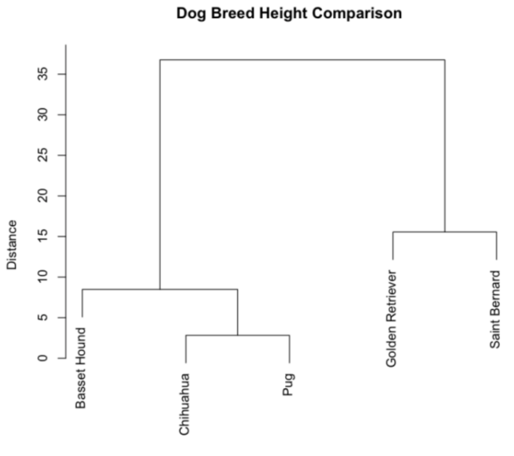

Figure 3

10 Euclidean distances (Everitt 1998:134) and complete linkage clustering (Everitt 1998:82) are utilised in this article.

Since our dataset contains five points, there are four stages to the clustering. The first stage groups together the two dog breeds that have the most similar heights. In this case, Chihuahua and Pug are grouped together. At this point the Chihuahua/Pug cluster is considered one individual, so the next stage groups together the two breeds with the most similar height out of the remaining four options: Basset Hound, Chihuahua/Pug, Golden Retriever, and Saint Bernard. Of these four, the Basset Hound and the cluster Chihuahua/Pug are most similar. In the third stage of clustering we are left with the group Basset Hound/Chihuahua/Pug, Golden Retriever, and Saint Bernard. In this case, the Golden Retriever and Saint Bernard are the most similar and thus they are grouped. We are left with two major groups at this point, the small dogs and the big dogs. The final stage of clustering only has two remaining groups and thus they are joined together. Through this analysis we end up with a dendrogram, which is a helpful visual that can aid us in an examination of dog breed heights.[11]

In the following analysis of recent diachronic research on the H.L., I utilise dendrograms in order to help visualise the data. This visualisation, and the calculations behind it, help to address several issues with which Hebrew linguistics is concerned. The first of these is the lack of a holistic picture of Hebrew language change that does not disregard the fine details. The second is related. As Rezetko and Young have argued, simply grouping together "early" books and "late" books is problematic (2014:391). Hierarchical clustering helps to address this problem by either supporting or rejecting the traditional book groupings. I demonstrate this point in more detail below.

[11] While hierarchical clustering and dendrograms are useful when considering one variable, such as the height of dogs, these statistical tools are most powerful when multiple variables are being examined. Sticking with the example of dog breeds, one could consider multiple variables, such as height, weight, temperament, intelligence, etc. When calculating the similarity of each individual, all these variables are included in the analysis. This is significant because humans can oftentimes see patterns in one or two variables, but when more variables are in play, we need robust statistical tools to undertake a holistic analysis.

Hornkohl (2014) and the H.L.

Hornkohl (2014) applies the traditional methodology developed by Hurvitz to analyse the diachronic development of Hebrew and Jeremiah's place within that history. Hornkohl discusses numerous linguistic features from all parts of speech, including the H.L., which is one of the foci of this article. Hornkohl argues that the H.L. was utilised differently in Hebrew over time. His analysis is divided into two parts. First, he provides a holistic review of the data and argues that when all the occurrences of the H.L. are considered, "Early Biblical Hebrew" (EBH) uses this particle more often than "Late Biblical Hebrew" (LBH) (Hornkohl 2014:207).[12] In support of this claim, Hornkohl references Joosten's (2005: 337–338) and Rezetko's (2013:48–56) work on the subject. The second part of Hornkohl's analysis is on the non-standard use of the H.L. He argues that there is an increase of the non-standard[13] use of the H.L. in LBH books. Both of these arguments can be developed further by utilising hierarchical clustering as well as by applying the t-test. I will start with Hornkohl's first argument, that the H.L. is used less often in the LBH books when compared to the EBH books. Then, I will turn to the non-standard use of the H.L.

Hornkohl groups together Genesis-Kings for EBH and Qohelet-Chronicles for LBH. The grouping of books in this manner has been called into question on a number of occasions. By utilising hierarchical clustering, we can determine if these books actually do group together when considering all the occurrences of the H.L.[14] The

---

[12] Hornkohl includes Genesis through Kings in the EBH corpus and Qohelet through Chronicles in the LBH corpus (2014:5).

[13] "According to the general approach adopted here, in its standard use the suffix in question indicates destination, direction, or orientation while deviations from this rule are to be explained as exceptions [or non-standard usages]" (Hornkohl 2014:205).

[14] For this analysis, I utilised Hornkohl's data found on pages 205 and 206 of AHP. While Hornkohl's data are difficult to reproduce, since he references several other works instead of simply providing a chart containing the occurrences of the H.L., it seems that it is exhaustive and correct. Also problematic is Hornkohl's separation of books like Samuel into two texts. I follow Hornkohl's lead here, since I am attempting to show how inferential statistical tools can be used to build upon current research. However, future research should consider treating Samuel, Kings, and Chronicles as one text each.

following is a chart of all the occurrences of the H.L. in the Hebrew Bible, totalling 1094.[15]

Table 7

Book	Total H.L.s	Total Words	Ratio	Book	Total H.L.s	Total Words	Ratio	Book	Total H.L.s	Total Words	Ratio
Gen	138	28627	0.0048	Ezek	118	26000	0.0045	Ps	10	24993	0.0004
Exod	70	23563	0.0030	Hos	2	3122	0.0006	Job	7	10788	0.0006
Lev	32	16863	0.0019	Joel	2	1304	0.0015	Prov	2	8783	0.0002
Num	92	23026	0.0040	Amos	2	2774	0.0007	Ruth	2	1799	0.0011
Deut	66	19994	0.0033	Obad	0	387	0.0000	Song	2	1656	0.0012
Josh	120	14525	0.0083	Jonah	3	980	0.0020	Eccl	4	4155	0.0010
Judg	52	14051	0.0037	Mic	1	1894	0.0005	Lam	0	1977	0.0000
1 Sam	61	18895	0.0032	Nah	0	728	0.0000	Esth	0	4574	0.0000
2 Sam	42	15663	0.0027	Hab	2	890	0.0022	Dan	9	8716	0.0010
1 Kgs	42	18564	0.0023	Zeph	0	1021	0.0000	Ezra	2	5572	0.0004
2 Kgs	51	17244	0.0030	Hag	2	867	0.0023	Neh	2	7852	0.0003
Isa	23	22765	0.0010	Zech	4	4432	0.0009	1 Chr	36	15662	0.0022
Jer	49	29665	0.0017	Mal	0	1175	0.0000	2 Chr	44	19644	0.0022

When the Euclidian distances are calculated for these data and hierarchical clustering is preformed, the following dendrogram is produced:

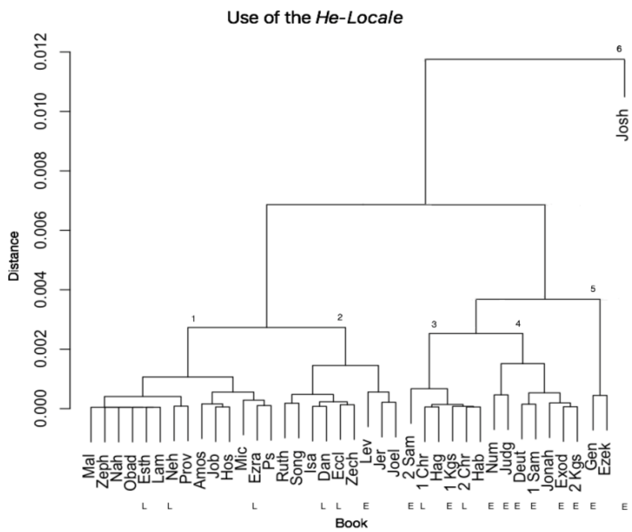

Figure 4

[15] Hornkohl notes that there are "approximately 1090 cases of the suffix in the Bible" (2014:205).

There are six main clusters in this dendrogram. The first on the left, labelled "1", contains all the books with either zero occurrences or very few occurrences of the H.L. The further to the right a cluster is found, the higher the ratio of H.L.s to the number of total words.[16] Thus Joshua, with its ratio of 0.0083 H.L.s per word, has the highest density of this particle in the Hebrew Bible. Across the bottom of this dendrogram, I have labelled all EBH books with an "E" and all LBH books with an "L". We can observe that in cluster 1 (the cluster containing the books with the lowest ratio of H.L. to total words) there are only LBH books. In clusters 2 and 3, there is a mix of EBH and LBH books. In clusters 4, 5, and 6 (those clusters containing the books with the highest ratio of H.L. to total words) there are only EBH books. This analysis gives some (but not unequivocal) credence to grouping these sets of books together.

Further analysis of these two groups and their use of the H.L. can be accomplished by applying the t-test discussed above. Utilising the data in Table 7, the following results of the t-test are obtained:

Table 8

Compare Means				
Descriptive Statistics				
	Sample Size	Mean	Standard Deviation	Variance
EBH	11	0.00364	0.00173	0.0000029933
LBH	7	0.00101	0.00093	0.0000008648
Two-tailed Distribution				
P-value	0.00072	Critical Value (5%)	2.11991	

Since the t-test produced a P-value of 0.07%, well below the 5% threshold of significance, we can conclude that there is a statistically valid difference between the use of the H.L. in the EBH books when compared to the LBH books. Thus, the application of hierarchical clustering as well as the t-test has lent some support to Hornkohl's conclusion that there was a diachronic development from EBH to LBH on

[16] A more refined picture might be able to be attained by normalising the number of H.L.s to place names or place names with verbs of motion. However, the general picture offered by normalizing to the number of words is sufficient to demonstrate the value of these statistical tools.

the whole in the use of this particle. However, aside from the core EBH and LBH books, it is remarkable that many conventionally-dated "early" and "late" books do not align chronologically, especially within the Latter Prophets (e.g., Zephaniah, Nahum, Amos, Hosea, Micah, Jonah, and Ezekiel), as well as texts often considered exilic such as Jeremiah and Ezekiel.

After analysing all the occurrences of the H.L. in the Hebrew Bible, Hornkohl turned to what he calls the non-standard use of this particle. He argues that while the general tendency is for the H.L. to be used less over time, the non-standard use actually increases. In support of this conclusion, Hornkohl gives the following statistics (structured as the percent of non-standard uses out of all of the occurrences):

> Torah: approximately 17 percent (63/395)[17] or under 10 percent excluding Numbers.
> Former Prophets: 14.1 percent (51/361).
> Latter Prophets: 41.7 percent (88/211)[18] or 30.9 percent excluding Ezekiel.
> Poetic Books (Psalms, Job, Proverbs, and Song of Songs): 81.9 percent (18/22).[19]
> Core LBH material: 34.4 percent (33 out of 96).

From these data, Hornkohl concludes that there is a clear shift towards the non-standard use of the H.L. in later phases of the Hebrew language. However, the presentation of Hornkohl's data may account for how this picture comes together. By simply grouping together books of the Bible, Hornkohl smooths over variation in the use of the H.L. in individual books (although he does highlight two cases he considers unusual, Numbers and Ezekiel). Again, I will turn to hierarchical clustering to see if the groups he proposes are valid.

---

17 Hornkohl (2014:209) notes that this number is a bit misleading due to the large number of non-standard H.L.s in Numbers (a total of 24).

18 Again, Hornkohl (2014:209) notes that these numbers are misleading due to a high number of non-standard uses of the H.L. in Ezekiel (a total of 33).

19 Hornkohl (2014:210) notes that this high ratio is due to the poetic style of these books.

The following chart reproduces Hornkohl's data for the non-standard use of the H.L.[20]

Table 9

Book	Total N-stan	Total Words	H.L. N-s Ratio	Book	Total N-stan	Total Words	H.L. N-s Ratio	Book	Total N-stan	Total Words	H.L. N-s Ratio
Gen	9	28627	0.00031	Ezek	63	26000	0.00242	Ps	11	24993	0.00044
Exod	18	23563	0.00076	Hos	1	3122	0.00032	Job	4	10788	0.00037
Lev	25	16863	0.00148	Joel	0	1304	0.00000	Prov	1	8783	0.00011
Num	6	23026	0.00026	Amos	0	2774	0.00000	Ruth	1	1799	0.00055
Deut	3	19994	0.00015	Obad	0	387	0.00000	Song	2	1656	0.00121
Josh	24	14525	0.00165	Jonah	0	980	0.00000	Eccl	3	4155	0.00072
Judg	9	14051	0.00064	Mic	0	1894	0.00000	Lam	0	1977	0.00000
1 Sam	1	18895	0.00005	Nah	0	728	0.00000	Esth	0	4574	0.00000
2 Sam	2	15663	0.00013	Hab	2	890	0.00225	Dan	0	8716	0.00000
1 Kgs	7	18564	0.00038	Zeph	0	1021	0.00000	Ezra	1	5572	0.00018
2 Kgs	7	17244	0.00041	Hag	0	867	0.00000	Neh	0	7852	0.00000
Isa	10	22765	0.00044	Zech	0	4432	0.00000	1 Chr	14	15662	0.00089
Jer	13	29665	0.00044	Mal	0	1175	0.00000	2 Chr	15	19644	0.00076

When the Euclidian distances are calculated for these data and hierarchical clustering is preformed, the following dendrogram is produced:

[20] For this analysis, I utilised Hornkohl's data (2014:204–206). Hornkohl's data are challenging to reproduce. At one point he notes there are 66 occurrences of a specific type of the non-standard use of the H.L., but goes on to list 67 occurrences (2014:205, n. 69); at another point he refers his readers to "the concordances", and at another he writes "see below" but actually seems to be referring to a footnote on the previous page. After sorting through these references and piecing together Hornkohl's data, I have accounted for 252 total non-standard H.L.s, while he notes there are 255.

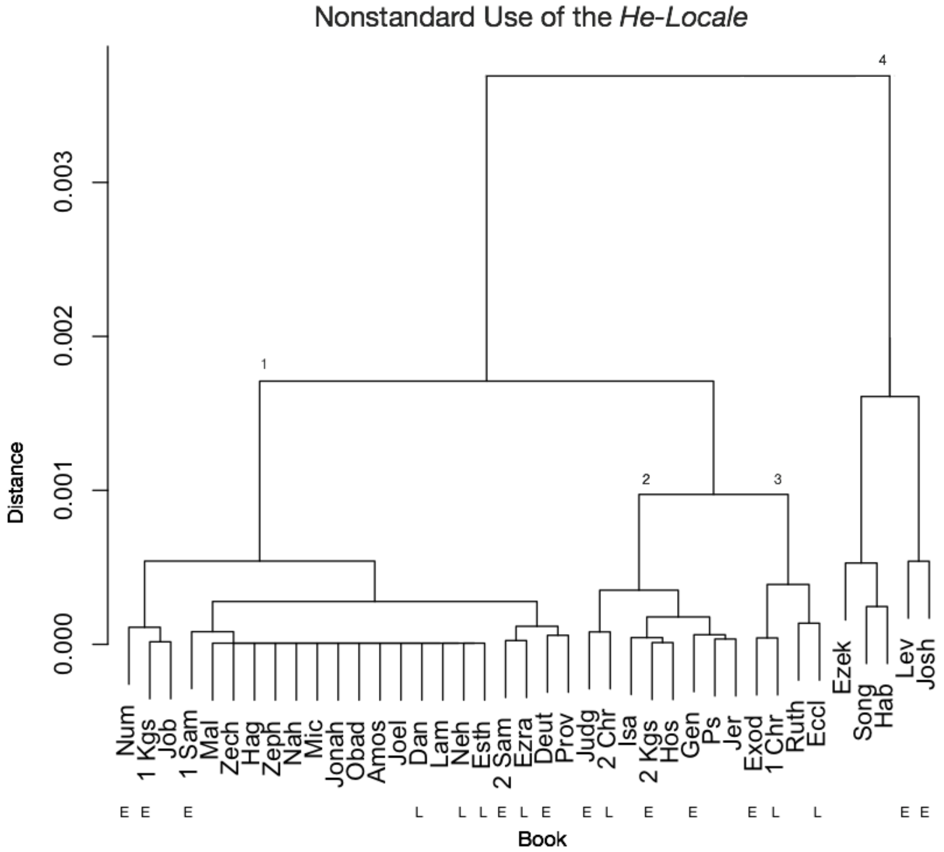

Figure 5

This dendrogram has four main clusters. The first on the left, labelled “1”, contains those books which have the lowest ratio of non-standard H.L.s to total number of words. The ratio increases the further right the clusters appear on the dendrogram, with cluster 4 having the highest. At the bottom of this dendrogram, I have again labelled the EBH and LBH books. This dendrogram shows that the EBH and LBH books are mixed. Three of the four (1, 2, and 3) main clusters contain both EBH and LBH books. Further, the more generic groupings of Torah, Former Prophets, Latter Prophets, and poetic books break down in this dendrogram. We find the Psalms clustered with Genesis, 2 Kings with Isaiah, Chronicles with Judges, and Song of Songs with Leviticus.

The above analysis suggests that the books of the Hebrew Bible should not be grouped when analysing the nonstandard use of the H.L. But, if one were to group them, a t-test[21] could be applied to see if there are any statistically significant differences between them. The following chart presents the results of a t-test when applied to the occurrences of the non-standard H.L. in the EBH and LBH books.

Table 10

Compare Means				
Descriptive Statistics				
	Sample Size	Mean	Standard Deviation	Variance
EBH	11	0.00066	0.00057	0.0000003257
LBH	7	0.00044	0.00052	0.0000002673
Two-tailed Distribution				
P-value	0.41973	Critical Value (5%)	2.14479	

The P-value for this t-test is 41.97%, well above the 5% threshold of significance. This result forces us to conclude that there is no statistically significant difference in the use of the non-standard use of the H.L. in the EBH and LBH books. Hornkohl (2014:212–217) supports one of his central assertions, the transitional nature of Jeremiah's language, by appealing to the book of Jeremiah's use of the non-standard H.L. However, the above analysis has shown that the non-standard use of the H.L. is not a distinguishing feature between EBH and LBH books and thus it should not be used to help characterise the linguistic nature of Jeremiah.

Rezetko and Young (2014) also analyse the occurrences of the H.L. in our ancient Hebrew sources. Below, I explore their results by applying hierarchical clustering and the t-test as I have done with Hornkohl's work above. Then, I critically compare and contrast their varying conclusions. Finally, I conclude by reflecting on the place of statistics in further research on the diachronic development of Hebrew.

---

[21] However, it should be noted that while we must group together books to run a t-test, the individual data points for each book are still analysed. The t-test does not simply compare the mean of one group with the mean of the other. It also accounts for the individual values of each book and their impact upon the whole.

Rezetko and Young (2014) and the H.L.

Rezetko and Young (2014) also address the use of the H.L. within ancient Hebrew, but they take a different approach to Hornkohl. Instead of focusing on all of the occurrences of this particle, they work with only those that are attached to nouns. They further refine their study of this particle by focusing upon only the verb בוא followed by a place of destination. From these data, they compare the use of the H.L. attached to the place of destination with alternate expressions such as the use the prepositions אל or ב. The following table reproduces their data (Rezetko & Young 2014:380–383).

Table 11

Book	Noun w/ H.L.	Noun w/out H.L.	Book	Noun w/ H.L.	Noun w/out H.L.	Book	Noun w/ H.L.	Noun w/out H.L.
Gen	28	32	Ez	1	28	Prov 1–9, 30–31	0	1
Ex	7	24	Hos	0	4	Prov 10–29	0	2
Lev	0	18	Am	0	4	Ruth	0	8
Num	2	24	Ob	0	2	Song	0	2
Deut	0	32	Jon	1	3	Qoh	0	1
Josh	2	18	Mic	0	5	Lam	0	2
Judg	5	31	Nah	0	1	Esth	0	8
Sam	24	87	Hab	0	2	DanH	0	7
1 Kgs 1–2 Kgs 23	13	65	Zech	0	6	EzH	0	8
2 Kgs 24–25	0	5	Mal	0	1	Neh	0	7
Isa 1–39	1	17	Pa	0	19	ChronSyn	1	12
Isa 40–55	0	2	Pb	0	2	ChronNonSyn	4	31
Isa 56–66	0	1	Job 3:1–42:6	0	6	DSS	2	55
Jer	8	60						

After a very thorough and statistically grounded analysis of these data, Rezetko and Young (2014) conclude:

1. Verbs of motion generally follow similar patterns with regard to their use of the verbal and prepositional complements.
2. These use patterns are generally applicable over the whole Bible, with two major exceptions: ל with place names is virtually restricted to Ezra–Nehemiah and Chronicles, and the density of the ה- directive is considerably greater in the prose

passages of the Pentateuch than in the prose passages of the rest of the books.[22] Thus, Rezetko and Young (2014) see some patterns emerging out of their analysis, but they do not conclude that this feature of Hebrew separates "early" and "late" books. This conclusion can be developed further by analysing their data through hierarchical clustering as well as with a t-test.

The following dendrogram is produced from Rezetko and Young's data after they have been normalised to the total number of words within each book. Also, the production of this dendrogram is different from those above, because there are two variables involved: occurrences of בוא with a H.L. attached to a place of destination and בוא with a place of destination with an alternative preposition.

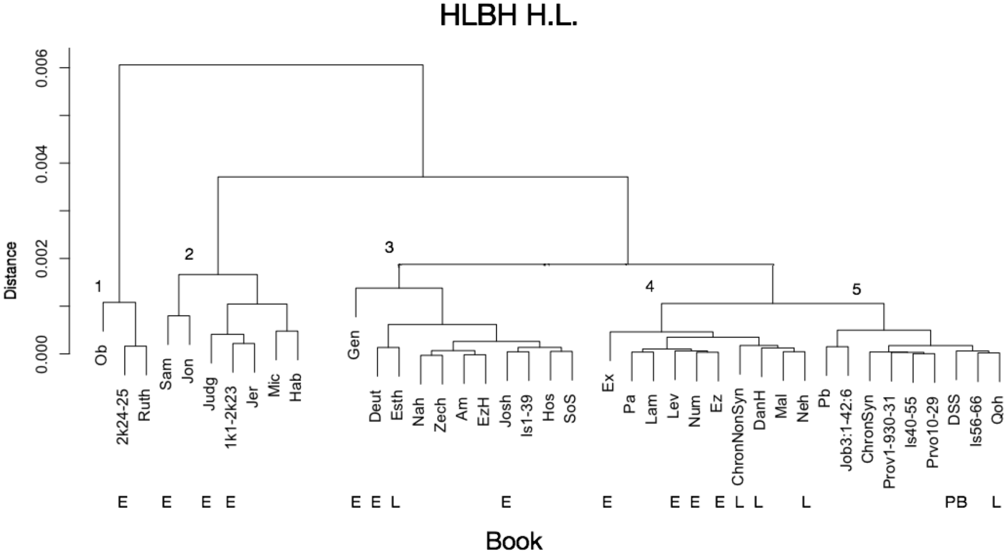

Figure 6[23]

[22] Rezetko and Young (2014) set their analysis of the H.L. within the context of a previous study done by Austel (1969). After correcting for some of his methodological mistakes, Rezetko and Young's conclusions align closely with his (2014:391).

[23] EzH = the Hebrew sections of Ezra. Pa = Psalms 1–18; 20–27; 29–32; 34–39; 41:1–13; 42–44; 46–62; 64–71; 72:1–17; 73–74; 76–102; 105; 108; 110; 114–115; 118; 120–122; 127; 130–132; 134; 138–142; 144:1–11; 149–150. Pb = 19; 28; 33; 40; 45; 63; 75; 104; 106:1–46; 107; 109; 111–113; 116; 126; 128; 135; 137; 143; 146–148 (Rezetko and Young

This dendrogram shows that Rezetko and Young's data group into five main clusters. Generally speaking the clusters on the left, beginning with number 1, contain more occurrences of בוא plus place of destination without a H.L. per word then those clusters on the right. The occurrences of בוא plus place of destination with a H.L. influence the clustering of these books to a lesser extent due to the number of texts that do not contain this feature (26 total).

From this dendrogram we can see that the EBH texts are concentrated on the left, while the LBH texts and the DSS (labelled "PB" for post-biblical) are concentrated on the right. Esther (often considered a LBH book) is one major exception to this, being found in cluster 3 with EBH books. However, as noted above with Hornkohl's data, several books, such as some of the Latter Prophets, do not align along conventional "early" and "late" divisions.

By applying the t-test to Rezetko and Young's data, we find that there is a statistically valid difference between the EBH and LBH books. The t-test needs to be run twice, once for each of the variables being considered. For בוא with H.L. attached to place of destination, the t-test produces the following results:

Table 12

Compare Means				
Descriptive Statistics				
	Sample Size	Mean	Standard Deviation	Variance
EBH	11	0.00029	0.00033	0.0000001089
LBH	5	0.00003	0.00004	0.0000000017
Two-tailed Distribution				
P-value	0.035	Critical Value (5%)	3.6622	

The t-test produces a P-value of 3.5% for these data, below the 5% threshold of significance. Similar results are found for בוא plus place of destination without a H.L.

2014:249–250).

Table 13

Compare Means				
Descriptive Statistics				
	Sample Size	Mean	Standard Deviation	Variance
EBH	11	0.00179	0.00102	0.0000010359
LBH	5	0.00085	0.0005	0.0000002465
Two-tailed Distribution				
P-value	0.02392	Critical Value (5%)	3.52262	

The above table shows a P-value of 2.4%, again well below the 5% threshold of statistical significance. While there is a clear difference between these two corpora, this difference is not consistent with a diachronic interpretation as a close look at the means for each corpora will show:

Table 14

H.L.	EBH	0.00029
	LBH	0.00003
No H.L.	EBH	0.00179
	LBH	0.00085

This table supports the conclusion that there is a significant difference between these two corpora on the whole. However, a closer look shows that the EBH books contain more occurrences per word of both features. If the use of בוא with the H.L. was a marker of early texts, we would have expected the alternative feature to increase over time while the dominant early feature would have decreased. Since these data do not align well with a diachronic explanation, further analysis (which is beyond the scope of this study)[24] is needed to explore other possibilities, such as style, topic, and genre.

Critical analysis of Hornkohl (2014) and Rezetko and Young (2014)

Both Hornkohl (2014) and Rezetko and Young (2014) have their strengths and their weaknesses. These cannot be fully discussed here, but I will offer some reflections dealing with methodological issues raised above, as well as possible avenues of

[24] For further discussion see Jacobs (2015) and the references quoted there.

inquiry for future development.

The above analysis of Hornkohl (2014) offered the following insights and conclusions. First, the application of hierarchical clustering showed that the grouping of EBH and LBH books for the analysis of all of the occurrences of the H.L. was a relevant step. Second, hierarchical clustering showed that the same grouping was not appropriate for the nonstandard use of the H.L. Taken together, these two points highlight the importance of not categorically clumping books together. Corpus studies need to be based on principled corpora made up of texts with similar characteristics (Biber et al. 1998:4). The grouping of texts should be supported by evidence for each feature of Hebrew language studies.

The application of t-tests as well as hierarchical clustering concluded with three main points concerning Hornkohl (2014) and Rezetko and Young (2014): first, the use of the H.L. in the EBH books is statistically different than in the LBH. Second, the nonstandard use of the H.L. does not show a significant difference between these two corpora. Finally, בוא plus place of destination with a H.L. is used statistically more often in EBH than in LBH. Similarly, the alternative linguistic structure, בוא plus place of destination with a preposition, is also used statistically more often in those corpora, thus pointing to a non-diachronic explanation of these differences. If diachronic development had caused these differences, we would have expected the constructions with the H.L. to be more prominent in EBH books while the alternative feature would have been dominant in the LBH books. But, the data presented above do not support that conclusion.

The analysis presented in this article supports one of the two main conclusions presented in Hornkohl (2014), namely that overall, H.L. is used more often in the EBH books then in the LBH books. Hornkohl (2014:207) concludes that this difference is due to diachronic development in the use of this particle. This is supported in part by Hornkohl's discussion of an alternate feature, motion verb plus ל plus destination (2014:219–226). Hornkohl concludes that this feature is used more often in LBH books when compared with EBH books. A t-test applied to his data[25] confirms that

[25] See Hornkohl (2014:220).

this difference is statistically relevant by producing a P-value of 4.01%, below the 5% threshold. This result suggests a shift away from the H.L. to this alternate feature. However, there are some methodological issues that force us to regard this conclusion as tentative. First, the preposition ל is not the only alternative feature to the H.L. As Rezetko and Young develop, other prepositions are used in Hebrew for the variable "to come to x" such as אל, עד, and ב (Rezetko & Young 2014:378). When all the possible alternatives to the H.L. are taken into account, this difference no longer exists (as the test case in Rezetko & Young (2014) on the motion verb בוא shows). When only the preposition ל is considered, there is a difference between EBH and LBH. However, as noted in Rezetko & Young (2014:390–391), only a few LBH books contain a significant amount of this variant, namely Ezra, Nehemiah, and 2 Chronicles. But, since there are only seven books total in Hornkohl's corpus of LBH the three with high rates make the group's mean significantly higher than EBH books. This leads to the second methodological problem with Hornkohl's analysis.

The small number of texts analysed in Hornkohl (2014) forces us to regard Hornkohl's conclusions as tentative. Biber shows that ten texts of each type need to be compared in order to attain statistical relevancy (1993:243–57). Hornkohl includes nine texts in his EBH corpus and only six in his LBH corpus. Both of these fall short of providing a statistically relevant sample size for analysis.

The final reason the analysis of the H.L. in Hornkohl (2014) should be considered tentative is the linguistic principle of "noise".[26] While, Hornkohl attempts to control for time by grouping EBH books together and LBH books together, he fails to consider issues of genre (such as the differences between Leviticus, Numbers, Qohelet, and Chronicles). Also, differences in style (say between Samuel/Kings and Chronicles) could have resulted in their different usages of the H.L. Creating corpora that account for these possibilities might help to identify time as the primary cause of the difference.

[26] "It is important to note that in order to carry out comparisons we should use similar corpora whenever possible, that is, corpora that are alike or comparable in size, content, and design, in order to avoid too much 'noise' and prevent or reduce the interaction of too many variables" (Cantos 2012:104).

Turning now to Rezetko and Young (2014), we find a number of places for further development. As Rezetko and Young note, their analysis of the use of the H.L. is only preliminary (2014:391). To add support to their conclusion, more verbs of motion would need to be analysed. As mentioned above, due to restricting their analysis to only the motion verb בוא, there are 26 out of 35 texts that contain zero tokens of this verb plus place of destination with a H.L. In contrast, only 10 of 35 texts contain zero occurrences of the alternate construction. That leaves the alternate feature to overly influence the results of their study. The addition of more motion verbs would alleviate this issue.

While the use of descriptive statistics in Rezetko and Young (2014) is thorough and creatively applied to the analysis of the H.L., inferential statistics are necessary to confirm their conclusions. As an example, Rezetko and Young (2014:387) write, "Genesis, Exodus, Judges, Samuel, and Kings stand out as the books that make regular use the [*sic*] [H.L.] ... [While] all the other biblical and DSS writings evidence much higher absolute and relative frequencies of the non-*he* variants." Their data support this conclusion: Genesis, Exodus, Judges, Samuel, and Kings contain 77 out of the 99 H.L. plus בוא occurrences, while the other texts contain 399 of the 643 alternative variant. While these raw numbers appear to be convincing, a thorough analysis using a significance test, such as the t-test presented above, would help to validate their statistical relevance.

Hornkohl (2014) and Rezetko and Young (2014) have advanced our understanding of the H.L. in biblical and nonbiblical texts. Further analysis using inferential statistical tools has shown that the H.L. is utilised in several distinct ways in the Biblical Hebrew corpus. Also, a close examination of the H.L. in the "biblical" DSS has shown distinctive patterns of usage in different types of manuscripts, particularly the *plene* scrolls. While these are promising results, further research is needed to provide confirmation and to clarify areas that have proven to be inconclusive.

FUTURE RESEARCH

In this article, I have highlighted three statistical tools that can be utilised by Hebrew linguists in order to test their data for significance: standard deviations, the t-test, and hierarchical clustering. While there are many different statistical tools that we could and should draw upon, the ones discussed here are some of the most useful for several reasons. First, they are basic. By this I mean that both standard deviations and t-tests are necessary to begin many types of statistical analysis. They provide a basis from which to build. Second, although the theory and math behind these three tools are somewhat complicated, computer applications such as Microsoft Excel and the statistical software program R provide easy access to them. Third, they are appropriate for many types of Hebrew linguistics research, such as has been seen in Hornkohl (2014) and Rezetko and Young (2014). And finally, they are complimentary. Standard deviations and hierarchical clustering can help us to identify patterns within large amounts of data and the t-test can examine those patterns for significance.

While research on the diachronic development of Hebrew has taken great strides in the past decade, especially with the integration of historical linguistic methodologies (Rezetko and Young 2014) and refinement of the traditional approach (Hornkohl 2014), there is of course still room for improvement. In the above, I have shown that the application of robust statistical tools can help us to build upon, correct, and refine past research while opening new avenues of exploration.

One particular area that could benefit from statistical tools is the analysis of multiple variables. In order to demonstrate the application of statistical tools as clearly as possible, the above analysis has been restricted to one (at times two) variables. Further research on the development of Hebrew over time should take into account many variables, not separately, but together. Hierarchical clustering is especially useful in that realm. More advanced statistical tools also need to be employed. Tools such as scatter diagrams[27] (and their multivariant analogue, biplots), analysis of

---

[27] "A two-dimensional plot of a sample of bivariate observations. The diagram is an important aid in assessing what type of relationship links the two variables" (Everitt 1998:334).

variance (ANOVA),[28] seriation,[29] and others need to be regularly applied when analysing large sets of data with multiple variables.

However, there is often one big challenge: navigating the world of statistics. Many Hebrew linguists who have been trained in biblical studies, Hebrew language, and even linguistics, have not been exposed to advanced statistical tools. Fortunately, academia is full of statisticians. Future work on the Hebrew language could benefit strongly from the interaction of Hebrew linguists with statisticians. Such interdisciplinary work is rewarding and vital to further our understanding of the diachronic development of the Hebrew language.

BIBLIOGRAPHY

Altman, D G 1991. *Practical statistics for medical research*. New York: Chapman & Hall.
Austel, H 1969. Prepositional and non-prepositional complements with verbs of motion in Biblical Hebrew. PhD dissertation. Los Angeles: University of California.
Biber, D 1993. Representativeness in corpus design, *LLC* 8:243–257.
Biber, D, Conrad, S & Reppen, R 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
Cantos, P G 2012. The use of linguistic corpora for the study of linguistic variation and change: types and computational applications, in Hernández & Silvestre 2000:99–122.
Everitt, B 1998. *The Cambridge dictionary of statistics*. Cambridge: Cambridge University Press.
Garcia, J & Quintana-Domeque, C 2007. The evolution of adult height in Europe: a brief note, *EHB* 5:340–349.
Hernández, J M & Silvestre J C C 2012. *The handbook of historical sociolinguistics*. Cambridge, MA: Wiley-Blackwell.
Hornkohl, A D 2014. *Ancient Hebrew periodization and the language of the book of Jeremiah: the case for a sixth-century date of composition*. SSLL 74. Leiden: Brill.
Jacobs, J T 2015. The balance of probability: a statistically based analysis of the linguistic character of the "biblical" Dead Sea Scrolls. PhD dissertation. Manchester: The University of Manchester.
Joosten, J 2005. The distinction between Classical and Late Biblical Hebrew as reflected in syntax, *HS* 46:327–339.
Lumley, T, Diehr, P, Emerson, S & Chen, L 2002. The importance of the normality assumption

[28] "The separation of variance attributable to one cause from the variance attributable to others" (Everitt 1998:11).

[29] Seriation studies were suggested by Dean Forbes (see Forbes, this section). Seriation can be defined as: "The problem of ordering a set of objects chronologically on the basis of dissimilarities or similarities between them" (Everitt 1998:341).

in large public health data sets, *ARPH* 23:151–169.
Metso, S, Najman, H and Schuller, E (eds) 2010. *The Dead Sea Scrolls: transmission of traditions and production of texts*. STDJ 92. Leiden: Brill.
Muraoka, T 2000. An Approach to the Morphosyntax and Syntax of Qumran Hebrew, in Muraoka and Elwolde 2000:193–214.
Muraoka, T & Elwolde, J F (eds) 2000. *Diggers at the well: Proceedings of a Third International Symposium on the Hebrew of the Dead Sea Scrolls and Ben Sira*. STDJ 36. Leiden: Brill.
Rezetko, R 2013. The Qumran Scrolls of the book of Judges: literary formation, textual criticism, and historical linguistics, *JHS* 13:1–69.
Rezetko, R & Young, I 2014. *Historical linguistics and Biblical Hebrew: steps toward an integrated approach*. ANEM 9. Atlanta: Society of Biblical Literature Press.
Romaine, S 1982. *Socio-historical linguistics: its status and methodology*. Cambridge Studies in Linguistics 34. Cambridge: Cambridge University Press.
Six-Sigma-Material.com, n.d. Normal distribution. Available: http://www.six-sigma-material.com/ Normal-Distribution.html. [2015/10/15.]
Tigchelaar, E J C 2010. Assessing Emanuel Tov's 'Qumran scribal practice', in Metso, Najman and Sculler 2010:173–208.