A Proposed Search For Long ELSs in the Hebrew Bible

by Randall Ingermanson and Ed Sherman

Introduction

This is a proposal for a joint research project between the Bible Code Digest team (www.biblecodedigest.com) and Dr. Randall Ingermanson (www.rsingermanson.com).

Ed Sherman, leader of the Bible Code Digest team and author of the book *Breakthrough: Encountering the Reality of the Bible Codes,* is an advocate for the validity of some codes.

Randall Ingermanson is author of *Who Wrote the Bible Code?,* which presents evidence that the total amount of information encoded in the Hebrew Bible is quite small and could be zero. Ingermanson has not seen any evidence that convinces him of the validity of the codes, though he remains open to the possibility.

Recently, Dr. Nathan Jacobi, of the Bible Code Digest team, has discovered a 22-letter ELS which Sherman believes to be an authentic code. In the October issue of his newsletter, he quoted odds of about 1 in 14 trillion that this ELS could be due to random chance.

Dr. Ingermanson does not accept this calculation, arguing that the finding lacks suitable scientific controls.

Both Sherman and Ingermanson agree that there is much to be gained in designing and conducting scientific experiments that test the potential existence (or lack thereof) of real ELSs regarding a broad range of subjects. Such tests would contrast sharply with experiments that focus only on a very narrow topic, such as the Great Rabbis Experiment.

Ingermanson has proposed the following controlled experiment to search for other long ELSs in the Hebrew Bible. Sherman has suggested some revisions. The details of the experiment will be agreed on by Ingermanson and the Bible Code Digest team in advance of the experiment. The calculations will be performed jointly and published on both their web sites.

The Hypothesis

The hypothesis we are testing is that "Long meaningful ELSs occur more frequently in the Hebrew Bible than one would expect by random chance."

There are several problems with defining the meaning of this hypothesis in an objective way. What is the precise definition of a "long meaningful ELS"?

First, we will define a "long ELS" to be an ELS composed of one or more Hebrew words with a total length of at least 10 letters. The reason for requiring at least 10 letters is that we expect such ELSs to be quite rare. In this experiment, we are not interested in analyzing the huge numbers of short ELSs in the Bible.

But how does one define "meaningful?" Roughly speaking, there are two levels of "meaningfulness." The first level is that the ELS has to consist of "real" Hebrew words. The second level is that the "real" words need to make sense grammatically. Let's look at these in turn.

There is no perfect way to define a complete list of "real" Hebrew words, but there is an obvious objective choice. We will construct a list containing all Hebrew (and Aramaic) words that occur in the Koren text of the Hebrew Bible. Words joined by a *maqqeph* will be considered to be two words. The advantage of using this definition is that it is purely objective. There is no wiggle room here. We have no control over the Koren text. The disadvantage is that we risk missing some authentic codes that would contain words not found in the Bible. Since the vocabulary of the Bible is reasonably large, we feel that an Encoder would have little trouble expressing at least some of his thoughts using this word list. We're willing to risk missing a few codes if we can squeeze out all wiggle room. The word list will be made publicly available for inspection on our web sites.

Likewise, there is no perfect way to define a complete list of all "grammatically correct" Hebrew phrases. Again, if we restrict attention to only those phrases that actually occur in the Koren text of the Bible, we gain objectivity, at the risk of missing a great many possible codes. Again, we're willing to take that risk in order to eliminate all possible wiggle room.

Technical Issues

There are some technical problems that need to be addressed also. What are the shortest ELSs to be searched for? Are backward skips included? What about the "short-range order" found in all Hebrew texts? Should "large" skips be counted with "small" skips? Should "long sub-ELSs" be counted? What if a long ELS is ambiguous—that is, it can be broken into words in more than one way? What is our control group of texts? What about subtle "medium-range order" that occurs in texts?

We'll look at each of these in turn.

How Short Can a "Long ELS" Be?

Professor Robert Haralick has recently produced a program that searches for long ELSs. Preliminary results show that there are large numbers to be found, especially at lengths less than 20 or so. We will count all ELSs of length 10 or more.

Backward Skips

Since backward ELSs are generally considered to be as good as forward ELSs, we will use both forward and backward ELSs in this study. When we speak here of ELSs at a skip of 10, for example, we really mean at either a skip of +10 or –10. This definition will be implicit in the rest of this document.

The Short-Range Order Issue

Short-range order occurs in all texts using a phonetic alphabet. Ingermanson showed in his book *Who Wrote the Bible Code?* that the behavior of ELSs at skips less than about 10 is extremely nonrandom for representative English and Hebrew texts. There is a measurable nonrandomness at skips between 10 and 20, and some residual nonrandomness for skips up to about 50, but this is probably negligible for our purposes. For this experiment, we agree to use only ELSs that occur at skips of 20 or more.

The Large-Skip/Small-Skip Issue

The Bible Code Digest team notes that ELSs at "small" skips (say, less than 100) are more highly prized by codes searchers than ELSs at "large" skips (anything more than 100). The reason is that small-skip ELSs are rarer, for obvious reasons. We therefore agree that in the initial search for long ELSs, we will keep information on the skip at which each long ELS was found. When we analyze the data, we will consider ELSs in three groups:

1) ELSs with skips between 20 and 100 inclusive.
2) ELSs with skips between 101 and 1000 inclusive.
3) All ELSs with skips greater than 1000.

The experiments to be described below will be performed on each group and results will be reported for all three variations of each experiment.

The Long Sub-ELS Issue

We will define a "long sub-ELS" to be a long ELS which occurs within a longer ELS. It could happen that an Encoder would encode a long ELS which might be extended by chance. For example, an Encoder working on an English text might encode "The Cat In The Hat" at some skip. By chance, this might be followed by the word "Dog," leading to the even longer (but meaningless) ELS "The Cat In The Hat Dog." There are two possibilities:

1) Count each sub-ELS that meets our criteria for a long ELS. In the example given, we would count each of the following sub-ELSs as hits, since each have 10 or more letters: "The Cat In The," "Cat In The Hat," "In The Hat Dog," "The Cat In The Hat," and "Cat In The Hat Dog."
2) Count only ELSs that are "maximal"—they are not contained in any larger ELS.

Our experiments will have two variations, one for each of these possibilities.

The Ambiguity Problem

It will often happen that a long ELS can be broken into words in more than one way. As an example, a six-letter ELS might turn out to be three real 2-letter words and also two real 3-letter words. We will count it once for each way it can be broken into real words. This implies, however, that the variance of long ELSs will not be equal to the expectation value. In general the variance will be larger. We will therefore measure both the expectation value and the variance empirically, using random control texts.

The Control-Text Issue

It is important to define a suitable set of random control texts. Obviously, if we are studying Genesis, we should compare results to random permutations of the letters in Genesis. However, as Ingermanson discussed in the Technical Appendices of his book (freely available in PDF format on his web site), one should randomize texts so as to retain the large-scale local variations in letter frequencies that occur in texts. The correct randomization procedure is as follows:

Break the text into bins of a certain size *B*. Scramble the letters within each bin. Dr. Ingermanson's software CodeCracker (available for free on his web site) does this randomization and we will use this.

The pseudo-random number generator requires an integer seed. Since our results should be independent of this seed, we will use the values, 1, 2, 3, . . ., *R*, where *R* is the number of random texts we need for the experiment.

What size *B* should we choose for the bins? We'll argue in the next section that bins should be smaller than the shortest skip at which we're searching for ELSs.

The Subtle Medium-Range Order Issue

In an interesting series of articles on Mark Perakh's Web site, Brendan McKay and Mark Perakh have used "Letter Serial Correlations" to investigate the short-range, medium-range, and long-range order of texts in a variety of languages. Certain patterns have emerged that cover scales from dozens to hundreds to thousands of letters. Our randomization procedure will control for these patterns at skip sizes larger than the bin size *B*. However, for skips smaller than *B*, our randomization procedure will effectively wash out any medium-range order. So for skips smaller than the bin size, our random control texts will be "more random" than the corresponding Biblical texts.

We will therefore choose bins half the size of the smallest skip at which we are searching for ELSs. To ensure good randomization, we agree to use bins of at least size 10. (This is one reason that we are limiting our search to skips larger than 20.) To verify that this correctly controls for medium-range order, we will run our experiments first on some non-Biblical documents in a variety of languages.

Experiment 1

Our first experiment will use the first definition of "long meaningful ELSs."

We will examine each of the Hebrew texts (Koren edition) of Genesis, Exodus, Leviticus, Numbers, Deuteronomy, the full Torah, and Isaiah. Because of the unusual statistical properties of Numbers 7, we will also analyze the text "MixedNumbers" as defined in Dr. Ingermanson's book, as well as a "Mixed Torah" having "MixedNumbers" in place of Numbers. For each text, we will construct the complete list of all ELSs and the complete list of "maximal ELSs" with the following properties:
1) The ELS has 10 or more letters.
2) The ELS consists entirely of "real" Hebrew words found at least once in the Koren text of the Hebrew Bible.
3) The ELS occurs at a skip within a given range. We will perform three variations on the experiment: skips of 20 to 100, 101 to 1000, and 1001 to infinity.

We will do the same search on 100 randomized texts, where the randomization is as defined above, with a bin size of half the minimum skip size and random number seeds going from 1 to 100.

We will then investigate the following pair of hypotheses:
1) Alternate Hypothesis: The Bible texts contain significantly more long ELSs than the random texts.
2) Null Hypothesis: The number of long ELSs found in each Hebrew texts is not statistically different from the number of long ELSs found in the corresponding random texts.

We will test these hypotheses individually on each text and for each acceptability criterion as follows.

For each text, we will construct a histogram of the acceptable long ELSs by length for the aggregate results of all the randomizations of that text. That is, we'll count the number of 10-letter ELSs, 11-letter ELSs, and so on, that occurred in the random texts. The final bin will be the first number *n* for which we have not found at least 10 *n*-letter ELSs. We will then count all ELSs having *n* or more letters in that bin. If an ELS occurs more than once, we will count it multiple times in the histogram. The random texts will also give us an estimate of the variance in each bin.

For each bin *k*, we will then compute the number of *k*-letter ELSs expected in the Biblical text according to the Null Hypothesis (that is, we'll use the empirically measured mean for the random texts), and we will use the empirically measured variance. We will use these values to compute the Z-value for each of the set of bins 10 through *n* for the ELSs found in the Biblical text. Thus, for each bin, we can compute a p-value using the normal distribution.

It is tempting to use the chi-squared statistic to combine the results for all bins into a single p-value. We believe this will be valid for the case of maximal ELSs, since we expect the bins to be independent (an issue we will verify empirically on non-Biblical texts first). However, this would be incorrect for the case where we accept sub-ELSs—the bins are not completely independent. For this case, we'll use the Bonferroni inequality—we'll multiply the best p-value by the total number of bins to obtain the p-value for the distribution of bins.

Acceptance criterion: In keeping with the usual conservative requirements, we will only reject the Null Hypothesis if the computed p-value is less than or equal to .001. We will fail to reject the Null Hypothesis if the computed p-value is greater than .001.

For each variation of the experiment, the histogram and results of the calculation for each bin, including its individual p-value, along with the combined p-value for all bins, will be posted on our web sites.

Experiment 2

In our second experiment, we will repeat the calculations, but with one change in the definition of the "successful" ELSs:
1) The ELS has 10 or more letters.
2) The ELS actually occurs as a sequence of words somewhere in the Koren text of the Hebrew Bible.
3) The ELS occurs at a skip within a given range. We will perform three variations on the experiment: skips of 20 to 100, 101 to 1000, and 1001 to infinity.

Requirement (2) above is intended to enforce some level of "grammatical meaningfulness." Again, this has the advantage that it is purely objective, since we have no control over the Koren text. It has the disadvantage that we will miss a large number of possible meaningful phrases not found in the Hebrew Bible. Again, we're willing to take that risk in order to eliminate all possibility of wiggle room.

The statistical analysis will proceed as in Experiment 1. Results will be posted on both our web sites.

Implementation Details

The Bible Code Digest team and Dr. Ingermanson will agree on all details before any computations are begun. The software to compute the lists of long meaningful ELSs will be written in Java by Dr. Ingermanson. The source code and the working Java program will be publicly available on his web site so that anyone can verify the correctness of his algorithms and can check that the published results are correct. Results will be posted on both web sites and in the Bible Code Digest Newsletter and Dr. Ingermanson's Bible Code Newsletter.

Interpretation of Results

We've made every effort to design an objective experiment. However, it's possible that we will still reach slightly different conclusions at the end. We agree in advance to publish up to three sets of conclusions in both our newsletters on both our web sites:

1) Conclusions we agree to jointly.
2) Conclusions of the Bible Code Digest team.
3) Conclusions of Dr. Ingermanson.

Acknowledgments

We thank Roy Reinhold, Keith York, Mark Perakh, Harold Gans, and Brendan McKay for their comments on this document.