Component Analysis of Long Torah Code Phrases
Description of the Data

Art Levitt
15 August 2006

Figure 1. The analyzed phrase, with bin Laden as *anchor*.

1. Introduction

This document details the data used in analyzing the Torah Code of Figure 1. This phrase contains three Hebrew components, surrounding our *anchor*, bin Laden. The components are translated as follows:

1. I will name you *Destruction*
2. Cursed [is bin Laden]
3. Revenge belongs to the Messiah

These components are referred to below as the "destruction", "cursed" and "revenge" components, respectively.

Our method measures the significance of each component, by comparing its relevance to that of thousands of candidate competitors.

Even a cursory examination of these candidates (for example `ra2a.htm`), which were constructed as discussed in the following sections, shows how rare it is to find truly competitive phrases. The following entries are typical of the great majority, in their questionable degrees of intelligibility, and/or obvious non-relevance to the *bin Laden* theme. These are candidate competitors for the component "*I will name you destruction*":

- You disabled their pollen
- He fed from nutcrackers
- We will bribe a spider

This rarity is quantified by our method, yielding the highly significant p-levels detailed in the conclusion below.

2. Summary of the method

The method consists of two parts:

1. Data collection,
2. Analysis.

The data collection method is summarized in sections 4 and 5. The analysis method is described fully in [1].

The essential steps of the analysis method are summarized as follows. For each component:

1. Use the data collection method to create a large set of candidates which will attempt to compete with the component.

2. A single initial reviewer rejects the irrelevant and/or unintelligible candidates, subject to the correction applied in (6). In this first implementation, the author performed this review, listing all rejected cases in the web pages linked from this document.

3. List all surviving candidates, which are considered to be the viable competitors. Mix the original component in an unmarked location among these competitors.

4. Use a set of reviewers, under double blind protocol, to rate all members of the list for intelligibility and relevance to the anchor.

5. Assign an initial outcome for the component.

6. Correct the outcome with a "gatekeeper" factor, which adjusts for the fact that the rejections contain a portion that could have been viable competitors, but were denied access in step (2) (at the starting "gate"). This factor is derived by the same simulation technique described in [2] (a study which analyzed the same bin Laden phrase without splitting it into components – obtaining a similar high significance).

The final step in [1] combines the results of the three components with the Fisher statistic.

The whole process is done in two modes:

1. *Component Analysis* (CA) follows the above steps using the original Hebrew. CA details are given in section 4.

2. *Relevance Analysis* (RA) follows the above steps using English language concepts rather than the original Hebrew. RA details are given in section 5.

3. Safeguards for Obtaining Accurate Estimates

Four factors help to ensure accuracy, as follows:

1. The reviewers rate each candidate competitor with a score from 0 to 5, with 5 meaning completely relevant, 0 meaning irrelevant (or unintelligible), and intermediate values indicating intermediate judgements along this scale. We tabulate only the number of reviewers assigning a score of 5 for each competitor. In this way, we consider only the most definite "votes" (opinions of relevance).

2. The simulation technique mentioned above helps ensure that the viable competitors are not understated. Those competitors which are borderline relevent/intelligible have a good chance of being rejected in the initial "gatekeeper" review, but many of them may have been viable. The simulation therefore inflates the number of observed competitors receiving n votes, as appropriate, for all values of n. We designate n as the *popularity level*. The inflating done by the simulation uses a simple probabilistic approach which naturally reflects the fact that most of the would-be candidates that were prevented from competing at the starting gate are in fact borderline cases. Likewise, few of the really competitive phrases tend to be rejected at the starting gate; accordingly, the simulation inflates the results for these to a lesser degree.

   The simulation adjusts for the gatekeeper effect by determining the initial, or *inherent*, popularities. For example, if 80 % of the candidates have inherent popularities of only 2 (out of say 20 reviewers), this means that if we would let the full group decide on the candidate's viability as a competitor, about 10 % of them would classify it as viable (more than that if the reviewers are lenient, but we can think of it as 10 % to be conservative). But we have only a single reviewer guarding the gate, not the full set. Therefore the candidate has only one chance, and it has roughly a 90 % chance of being rejected. This implies that the number of observed competitors (post-gatekeeper) at popularity 2 is probably only 10 % of what we would see if our initial gatekeeping review was more robust.

   The simulation accounts for the above as follows. It sets a random starting value for each inherent popularity level. It then simulates the outcome of the gatekeeper's decision for each candidate at that level. In our example, it will choose roughly 10 % of the candidates at inherent popularity 2 (based on a random number between 0 and 1 falling in the range less than 0.1). This 10 % will then be subjected to a simulation of the full set of r reviewers to determine the simulated final popularity level of each candidate. The final levels are compared to the observed levels. As a result of the comparison, the number of candidates at each inherent popularity level is adjusted upward or downward accordingly. The adjustments and recalculations are repeated over hundreds of iterations until the outcome of the simulation approaches and stays near a closest fit to the observed popularity levels.

   A look at the observed and simulated figures in the tables of this document, reflects the expected probabilistic trend, in which most of the inflating occurs at the lower popularity levels.

3. A review of the web pages containing the rejected cases, referenced from this document, verifies that the simulation results are reasonable. In particular, the lists of rejected cases are divided into a set that is clearly irrelevant and a much smaller set that is border-

line. As it turns out, the simulations typically inflate the competitors at the lower popularity levels sufficiently, so that all of the borderline cases could be permitted through the gate rather than being rejected. Of course, there is no doubt much variation among readers as to which particular cases are considered borderline, but the total number of such cases does not vary greatly among most reviewers. Further, the simulation is conservative in that it would actually be sufficient to accommodate a portion of the borderline cases rather than all of them.

4. We use a large set of reviewers: 28 for the CA method and 10 for the RA method.

The combination of these factors work together to enable quantification of the subjective notion of relevance.

4. The CA (Hebrew) Results

All data is available for viewing in the web links provided. The results are tabulated here for each of the three components, with the boldfaced final numbers used as inputs to [1].

4.1 The "destruction" and "revenge" components

As described in [1], the candidate competitors for these two Hebrew components (being multi-word components) are generated by extracting ELSs from the Hebrew Bible, at random anchor locations. All such candidates must consist of words that exist on our lexicon of over 100,000 Hebrew words. A total of 12,500 such ELSs were extracted. While the words comprising this set are all valid, their combination is quite often unintelligible, as reflected in the following results:

1. The phrases that were discarded by the "gatekeeper" decision are listed here: `ca02.htm`. As shown, 613 of these, while intelligible, had only borderline relevance. They were therefore prevented from passing the gate, as were the additional 11,787 cases that are listed in the unintelligible category.

2. The remaining 100 phrases comprised the set of viable competitors. These phrases and the reviewers' ranks for them are listed here: `ca01.htm` Four of these 100 were actually borderline but were permitted to compete as controls which we expected not to be popular (highlighted in yellow). Added to the list of 100 (but mixed in) were nine additional phrases: 6 control phrases that we expected to score well (highlighted in light green) the two real phrases being studied (in bright green), and an extra real phrase related to another bin Laden code (bright blue - not part of our current focus, but mentioned here for completeness).

   The actual number of phrases scoring as well or better than the *destruction* and *revenge* phrases are given in table 1. Also tabulated there are the adjusted values for each given popularity level (from the simulation). Popularity n indicates that n reviewers assigned the highest allowable relevance score (5) to the phrase. So, for example, 13 phrases were observed to receive the highest allowable relevance score from 5 reviewers. But the simulated number of phrases receiving such a score is the inflated number, 74.

Table 1. Results of the gatekeeper simulation, for the Hebrew *destruction* and *revenge* components.

Popularity	Observed	Simulated
0	1	N/A
1	7	302
2	8	125
3	11	106
4	22	157
5	13	74
6	9*	37
7	13	53
8	4	14
9	3**	6
10	7	19
11	1	2
12	2	4
13	1	2
14	0	0

$*$ includes the Hebrew phrase *revenge to the Messiah*
$**$ includes the Hebrew phrase *I will name you destruction*

The *revenge* phrase received 6 highest relevance scores (its popularity is therefore 6). The sum of the simulation numbers for all popularity levels greater than 6 (= 100), plus 19 (half of all level 6's, i.e. the original plus the 37 simulated), yields **119 out of 12,500** as our final rank for this component.

The *destruction* phrase had a popularity of 9. The sum of the simulation numbers for all popularity levels greater than level 9 (= 27), plus 3.5 (half of all level 9's, i.e. the original plus the 6 simulated), yields **30.5 out of 12,500** as our final rank for this component.

4.2 The "cursed" component

The "cursed" component competed against all 1640 roots in the Hebrew Maskilon dictionary. Almost all of

these roots were screened out prior to the "gate" by being categorized as not directly describing bin Laden. This set is displayed here, sorted by word within category: `cursa.htm`, and again here (`cursb.htm`) sorted by order of appearance in the dictionary. From the 18 viable roots, we derived appropriate nouns or adjectives to create a similar context to the original phrase: a single word describing bin Laden. The 18 viable roots yielded 19 derivations in Hebrew and 16 in English, as indicated on the first web page above. Reviewers indicated their favorite three choices from the set of 19 (Hebrew) or 16 (English). The former set is tabulated here: `ca03.htm`.

Table 2. Results of the gatekeeper simulation, for the Hebrew *cursed* component

Popularity	Observed	Simulated
0	3	N/A
1	5	213
2	2	33
3	1	10
4	2	15
5	2	12
6	1	4
7	1	4
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	1	2
14	0	0
15	0	0
16	0	0
17	0	0
18	1*	0
19	0	0

$*$ Includes the Hebrew component for *cursed/damned*

The original Hebrew synonym for "cursed/damned" had a popularity of 18, as tabulated in table 2. No competitor scored better than the original (not even when considering the inflated simulation results). Therefore our rank for this component is **0.5 out of 1640**.

5. The RA (English) Results

As we did for CA above, all data for RA is referenced and tabulated for each component, with boldfaced results used as inputs to [1].

5.1 The "destruction" component

5.1.1 Forming the candidate competitors

To generate candidate competitors for this component, we collected the first two verbs, as available, from every page of the 990-page 2002 Oxford English Dictionary, forming set 1. We collected the first two nouns, as available, from the same dictionary, forming set 2.

For the nouns in set 2, we embellished them with definite and indefinite articles, possessive pronouns, and leading connectors at rates similar to those found in a random sampling of Hebrew ELSs. For example, a leading letter "vav" in Hebrew (meaning "and"), occurs in the first position of roughly 10% of all ELSs. Therefore we added the word "and" in front of 10% of the nouns.

Likewise, for each verb in set 1, we added leading connectors and pronouns, and modified the tense, at rates similar to those found in a random sampling of Hebrew ELSs.

Next, we expanded set 1 by repeating it almost six times, in order to obtain exactly 8,000 entries; and similarly, for set 2. Finally, we combined the two expanded sets, forming 8,000 unique rows which served as our 8,000 candidate competitors.

5.1.2 Rejecting the non-competitors

The list of all non-competing candidates is here: `ra2a.htm`

This list includes:

1. 128 phrases that were worth a second look, but were ultimately rejected.

2. An additional 7,796 that were rejected out of hand.

The remaining 76 viable competitors and the reviewers' ranks for them are listed here: `ra1a.htm`. Four of these 76 were actually borderline but were permitted to compete as controls which we expected not to be popular (highlighted in yellow). Added to the list of 76 (but mixed in) were three additional phrases: 2 control phrases that we expected to score well (highlighted in light blue) and the real phrase (in green). The actual number of phrases scoring as well or better than the real phrase are given in table 3. Also tabulated there are the adjusted values for each given popularity level (from the simulation).

The real phrase had a popularity of 3. The sum of the simulation numbers for all popularity levels greater than 3 (= 13), plus 8 (half of all level 3's, i.e. the original and the 15 simulated), yields **21 out of 8000** as our final rank for this component.

Table 3. Results of the gatekeeper simulation, for the English *destruction* component

Popularity	Observed	Simulated
0	21	N/A
1	20	316
2	21	123
3	7*	15
4	6	11
5	1	1
6	1	1
7	0	0

$*$ Includes the English *destruction* component

5.2 The "revenge" component

Generating candidate competitors for this component was similar to the procedure used for the *destruction* component, except we needed only nouns. We built two initially identical sets from the dictionary nouns. We then duplicated and embellished each set, as described above for the *destruction* component. Our final combining of the two sets yielded 8,000 candidate competitors.

Table 4. Results of the gatekeeper simulation, for the English *revenge* component

Popularity	Observed	Simulated
0	25	N/A
1	36	349
2	19	274
3	7	44
4	2	11
5	1*	2
6	0	0

$*$ Includes the English *revenge* component

The list of all non-competing candidates is here: `ra2b.htm`. This list includes:

1. 83 phrases that were worth a second look, but were ultimately rejected.

2. An additional 7,830 that were rejected out of hand.

The remaining 87 viable competitors and their ranks from the reviewers are listed here: `ra1b.htm`. Two of these 87 were actually borderline but were permitted to compete as controls which we expected not to be popular (highlighted in yellow). Added to the list of 87 (but mixed in) were 7 additional cases: two cases ("boat traveler" and "job certificate") which were added as extra, uninteresting controls (also in yellow; their short format matches the "sin/crime" entry described below); three control phrases which we expected to score well (highlighted in light blue and green); the real phrase (bright green) and an extra real phrase ("sin/crime"), related to another bin Laden code (bright blue - not part of our current focus, but mentioned here for completeness). The actual number of phrases scoring as well or better than the real phrase are given in table 4. Also tabulated there are the adjusted values for each given popularity level (from the simulation).

The real phrase had a popularity of 5. The sum of the simulation numbers for all popularity levels greater than 5 (= 0), plus 1.5 (half of all level 5's, i.e. the original and the 2 simulated), yields **1.5 out of 8000** as our final rank for this component.

5.3 The "cursed" component

The 16 derived English entries described above in the CA discussion were rated by our reviewers as follows: `ra03.htm`. The starting number of 1640 Hebrew roots was used here, as it was in CA.

Table 5. Results of the gatekeeper simulation, for the English *cursed* component

Popularity	Observed	Simulated
0	5	N/A
1	4*	78
2	3	17
3	1	10
4	0	2
5	1	0
6	2	1
7	0	2
8	0	0

$*$ Includes the English *cursed* component

The English word "cursed" had a popularity of 1, as tabulated in table 5. The sum of the simulation numbers for all popularity levels greater than 1 (= 32), plus 39.5 (half of all level 1's, i.e. the original and the 78 simulated), yields **71.5 out of 1640** as our final rank for this component.

6. Two problems leading to unexpected results

Two unanticipated problems caused the final p-values to be less significant than may have been the case without these problems:

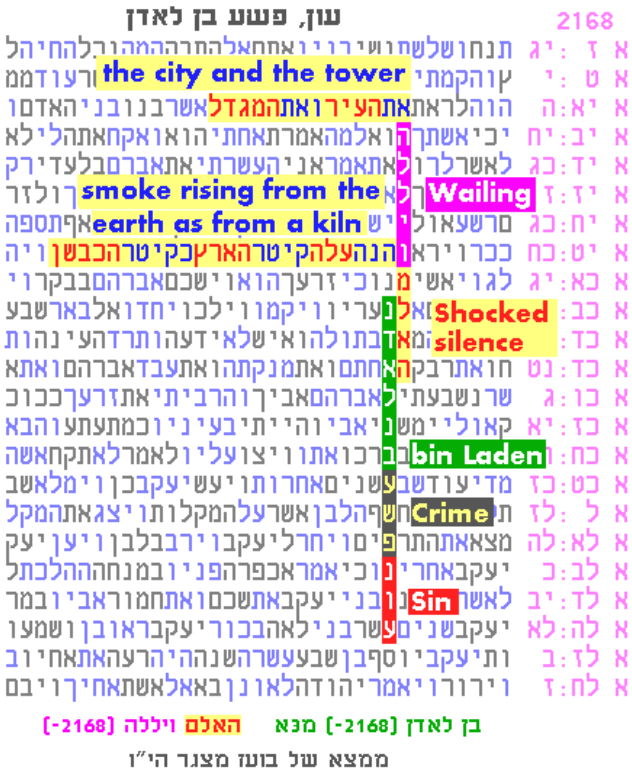

Figure 2. A second phrase with bin Laden as anchor.

Figure 3. A third code that ties the first two together.

1. For RA, it appears that our choice of *cursed* as the English translation, rather than *damned*, caused a less significant result. The latter is a more accurate translation, and in fact both words were part of the Hebrew (CA) study. For the Hebrew, the synonym with the connotation of *damned* was by far the most popular choice among all competitors, while the other synonym was much less popular.

2. For CA, the Hebrew reviewers were given two sets to rate, one for this bin Laden study and a second set for a study of Messiah codes. In each set, they were rating relevance to the theme for that set. It appears likely that many reviewers decided that the *revenge* component was not relevant to bin Laden, because they assumed that it was actually more relevant to the Messiah theme. It was not made clear to the reviewers that the two sets should be treated totally independently. In fact, the English reviewers, who were given only the bin Laden set, did not have this confusion and the Messiah component was extremely popular in that group.

7. Conclusion

These results, as indicated in [1], combine to yield a final p-level of $p = 4.0 \times 10^{-7}$ for the CA method (about 1 in 2.5 million) and $p = 1.1 \times 10^{-6}$ for the RA method (about 1 in 0.9 million). In addition, the current results should not be viewed in isolation. Figures 2 and 3 show two other significant examples related to the code under study.

References

[1] A. Levitt. Component analysis of Torah Code phrases. In *Proceedings of the 18th International Conference on Pattern Recognition*, August 2006.

[2] A. Levitt, N. Bombach, H. Gans, R. Haralick, L. Schwartzman, and C. Stal. Long phrases in Torah Codes. In *Proceedings of the 2nd International Interdisciplinary Conference on Fundamental and Applied Aspects of Speech and Language*, Belgrade, Serbia and Montenegro, November 2004. The Institute for Experimental Phonetics and Speech Pathology.