DRAFT. This edition: April 7, 1998. First edition: March 3, 1998.

On the Witztum-Rips-Rosenberg Sample of Nations

We study the Witztum-Rips-Rosenberg (WRR) sample of nations (see [WRR2, WRR3]) and find clear evidence that their results were obtained by selective data manipulation and are therefore invalid. Our tool is the study of variations - we vary the sample of nations in many ways, and find that the variations are almost always "worse" than the original. We argue that the only way this can be possible is if the original was "tuned" in one way or another. Finally, we show that "tuning" is a sufficiently strong process that can by itself produce results similar to WRR's.

Dror Bar-Natan
Institute of Mathematics
The Hebrew University
Giv'at-Ram, Jerusalem 91904
Israel.
drorbn@math.huji.ac.il

Brendan McKay
Department of Computer Science
Australian National University
Canberra, ACT 0200
Australia.
bdm@cs.anu.edu.au

Shlomo Sternberg
Department of Mathematics
Harvard University
Cambridge, MA 02138
USA.
shlomo@abel.math.harvard.edu

1. Introduction
   1.1. Some technical notes
   1.2. Acknowledgements
2. The choice of prefixes
   2.1. How good is their story?
   2.2. Why Vilna Gaon on Job?
   2.3. "The people of X"
3. "Tuning" on War and Peace
   3.1. The first story - Leaders
   3.2. The second story - Celestial Guardians
   3.3. The third story - Genesis ain't a War Epic
   3.4. Stories, Stories, Stories
4. More evidence
   4.1. The plural form
   4.2. The identification of the nations
      4.2.1. On orthography and on back translation
      4.2.2. A bit on Targum Yonatan
      4.2.3. Other identifications of the nations
      4.2.4. More on Genesis, chapter 10
   4.3. Other nice experiments
      4.3.1. The relation to the text
      4.3.2. The identifications of the nations and countries
      4.3.3. An alternate form for "the language of X"
   4.4. Some global choices
5. How much more have we done?
   5.1. Disclaimer
6. Conclusions
   6.1. A Personal note by the first author
7. Appendix: Notes on the Metric
   7.1. The permutation test on control texts
   7.2. Discussion
8. Bibliography

1. Introduction

Readers who tend to shy away from mathematics are assured that they can skip everything that appears too technical and still understand practically everything that really matters.

Let us start with a very brief review of the results of Witztum, Rips, and Rosenberg. WRR argue that there is a "code" hidden in the Hebrew Bible by means of "equidistant letter sequences". That is, they argue that if one reads the Hebrew text of the Bible (or just the book of Genesis alone) in equal skips, reading only every 7th or 19th or 666th letter and ignoring all spaces, the resulting stream of letters is far from random. More specifically, they argue that often one finds within this stream a meaningful word or a meaningful phrase, and that related words and phrases found in this way tend to be found unusually close to each other. They provide anecdotal examples, but acknowledge that serious study cannot be founded on anecdotes, and thus they proceed to describe (in [WRR1]) a reasonable (though far from perfect, see section 7) statistical procedure for measuring the significance of the phenomenon.

The details of WRR's procedure appear in their paper, and are too lengthy to repeat in full in this introduction. So we give only a digest. First, they define a certain notion of "distance", c(w,w'), between a pair of Hebrew words w and w', by finding some occurrences of w and w' in the text as equidistant letter sequences (ELSs), and by measuring in some specific way the distance between such occurrences of the word w and of the word w'. Programs for computing c(w,w') are available and are not difficult to use. For example, it takes our primary computer about 30 seconds to find that c(המהרש"ל,כסלו כ"ב(י"))=4/125. The distances c(w,w') are by construction rational numbers between 0 and 1, with denominator at most 125. So we find that the distance between ELSs of המהרש"ל, a designation of Rabbi Shlomo Luria, and the ELSs for his date of death, 12th of the Hebrew month of Kislev, is nearly as small as it can be. WRR argue that this phenomenon is recurring. They demonstrate it by constructing (in a certain systematic manner), a list of 32 famous rabbis, each with a list of designations, and by computing their distances to their dates of birth and death, written in several reasonable forms. The result is a list of distances (163 distances, to be precise). Some of these distances are very small (i.e., near 0) and some are very big (i.e., near 1), but overall, these distances tend to be small.

How small? Again the details are in [WRR1] and we only present a summary. WRR give this question two answers. The answer that is most relevant to us is formed by multiplying together those 163 distances and then scaling the product in a particular way to form a value between 0 and 1 called P_2(names, dates) (or simply, "the P_2 score") where "names" stands for the list of designations of rabbis and "dates" stands for the various dates associated with each rabbi. The result was originally intended by WRR to denote a probability, and would be quite astonishing if that were true:

$$P_2(\text{names, dates})=1.15\ 10^{-9}.$$

The reason why this is not a probability is that the method used to compute it from the 163 distances is based on some false assumptions. In order to avoid this problem, WRR introduced a second method of giving a numerical significance value to the apparent smallness of the 163 distances. They generate one million random permutations of the list of dates, and compute P_2(names,permuted dates) one million times. If P_2(names, dates) turns out to be smaller than P_2(names,permuted dates) in most cases, then we know that names of rabbis tend to be closer to their own birth and death dates than to the birth and death dates of other rabbis. So we don't need to interpret the smallness of P_2(names, dates) directly; instead, we simply see what proportion of the random permutations are smaller than P_2(names,dates). This proportion turns out to be exceedingly small - 3 of 1,000,000, to be precise. (We say that the "permutation rank" of the list of names and the list of dates is 4/1,000,000). This is a rather impressive quantitative measure of the proximity of ELSs for words with related meaning, and

Witztum-Rips-Rosenberg conclude that it cannot be due to chance.

The biggest problem with this argument is that it still depends on some human input, which may have been tuned to generate the impressive final result. For example, one has to choose which designations to use for each rabbi. Maybe many "bad" designations (those that appear far from their corresponding dates) were omitted? Maybe several "good" designations were inserted artificially even though they are hardly ever used? If any of these were done, no wonder the 163 distances measured are unnaturally small as a group! In a future publication we plan to detail these problems (and a few others) in greater depth. (An intermediate partial report in [BM] demonstrates that there is enough flexibility in the definition of the data to allow an equally strong result to be "cooked up" for War and Peace.)

When one of us (Bar-Natan) confronted Prof. E. Rips with these problems, he gave several answers (to be discussed elsewhere), but also added that anyway, since the publication of [WRR1], he and his colleagues had generated several new samples that produced highly significant results and that involved no human input at all (beyond the choice of general topic). He suggested to Bar-Natan that he study these samples too. The purpose of this paper is to report our findings in relation to the "Nations Sample", the one that Rips told Bar-Natan to study first, presumably because he thought it was in some sense the best.

In brief, the Nations Sample is the following: (Full details are in [WRR2, WRR3]).

As a first list of words, take the 68 names of descendants of Noah, as they appear in [Genesis, chapter 10] (there are 70 descendants listed in that chapter, but two of the names repeat - שבא and חוילה). As a second list, of words with related meaning, take exactly the same 68 names, only with the prefixes עם (nation of), ארץ (country of), שפת (language of), and כתב (script of) appended in front of each one (plus some additions, see below). To justify their choice, they refer to a tradition that each of the original 70 nations of the world (according to the Bible, the descendants of those 70 persons) has four attributes: its name, its country, its dialect, and its script. In addition to these four attributes, WRR sometimes associate with each nation X a few more words: the construct יX, meaning "the people of X", and newer names for the nation of X and the country of X. These associations are made only for some nations, subject to some rules that we will review (and criticize) in section 4. The lists they get are reproduced in table 1. Within this paper, we divide their list of 'related words' into two parts: the "regular part", consisting of columns 2-5 of table 1, and the "irregular part" consisting of columns 6-8 of that table.

Once again, when distances are computed between the ELS occurrences of the descendants of Noah and the 'related words' in table 1, the results are astonishing. The resulting P_2-score is $2.59817\ 10^{-6}$, and the permutation rank is even more impressive - according to our computations it is 15 out of one billion (10^9), according to [WRR2] it is 4 out of one billion, and according to [WRR3] it is 17 out of one billion! Again, WRR argue that this cannot be due to chance.

We note that there is some difference of methodology between [WRR1] and [WRR2, WRR3]. In [WRR1], the distance measure c(w,w') is defined by searching for occurrences of both w and w' as ELSs, and by computing (in an appropriate way) the distance between these occurrences. In [WRR2, WRR3], the distances are computed (in much of the same way) between ELS occurrences of w' (with skip greater in absolute value than 1) and straight or reverse occurrences of w (i.e., occurrences with skips +1 or -1). All the words w that are considered in [WRR2, WRR3] are names of descendants of Noah, and by the nature of their definition, they all appear in [Genesis, chapter 10] with skip +1. This difference of methodology explains the "Relation to the Text" in the title of [WRR2] - what [WRR2] seem to prove is that there is a relationship between the "coded" occurrences of the words w' in columns 2-8 of table 1 and the "straight" occurrences of the words w in the first column of that table. For a technical reason, WRR restrict the definition of c(w,w') to words w' that have between 5 and 8 letters. No such restriction is imposed on the word w.

In this paper we criticize those impressive results of Witztum, Rips, and Rosenberg, and show them to be invalid. We show that there is much more choice in making up table 1 than meets the eye; enough choice to generate a similar result in several control tests. More importantly, we show that whenever WRR had to choose, they somehow knew to make the choice 'correctly' - namely, to make the choice that makes their result appear more significant. We find this difficult to accept if the data was correctly prepared without peeking at it first. In section 2 below we present a strong piece of evidence that such 'tuning' was done: we measure the P_2 score of the same list of 68 names with 136 other lists, obtained by adding 136 different reasonable prefixes to these names. We get 136 scores (that seem to be pretty random, as a whole), and it turns out that 3 of the 4 prefixes that WRR use, עם, כתב and שפת, are the top 3 scorers in our list of 136! This is a rather improbable result; we believe it suggests that Witztum, Rips, and Rosenberg cheated and tried (either directly or indirectly, see section 2) many possible prefixes before deciding which four to use. They've added the fourth prefix, ארץ, only because they needed to build a reasonable story around the other three.

Could it be that WRR first had the story and only then chose the prefixes? Even if so, the fact that there are many other very good possible prefixes whose P_2 success is rather low, means that in this context there is no unusual proximity of ELSs for nations and related notions. But anyway, we explain that this possibility is rather unlikely. The question boils down to "how many as-good stories are there?". If there are many, how come they've picked just the right one? And so we spend the rest of section 2 criticizing their story. The following section, section 3, shows that there are very many good stories indeed. In fact, so many that we could choose one that got a permutation rank of $5/10^8$ on a control text! (The regular part of their sample, involving only the four prefixes, gets a permutation rank of only $566/10^8$).

In section 4 we present some more evidence that WRR tuned their data to get the results they wanted. In section 4.1 and section 4.2 we consider a few variations of the irregular part of their sample. In all cases we find that the variations we considered are 'worse' than the original, suggesting that the original was somehow optimized. In section 4.3 we report on a few experiments that by the WRR logic should have been successful, but aren't. In section 4.4 we report on some global choices WRR had in arranging their list. There we find some variations of the WRR experiment that improve its results slightly, but it seems that testing these variations was almost impossible given the technical means that were available to WRR.

Following a brief disclaimer in section 5, we discuss our results and present our conclusions in section 6.

1.1. Some technical notes

Unless otherwise noted, the computations reported in this article were carried out using the program WRR gave us, `els2.c`. We modified it to work under Unix and re-wrote the permutation test part, but made no modifications to the main part of the code. The text of Genesis and the text of War and Peace we used were also given to us by WRR, though we've modified the text of War and Peace to get the text WGP, as described in section 3.

1.2. Acknowledgements

We are sad to report the untimely death on March 21, 1998, of Dr. Michael Weitzman of the University of London. Michael was an authority on the Aramaic scriptures. His criticisms, advice and encouragement were of great value to us. We will miss them just as we miss him.

We also wish to thank Maya Bar-Hillel, Alec Gindis, Gil Kalai, Michael Sokoloff, and the others who helped, for their suggestions and support.

2. The choice of prefixes

Let X be a name of a nation, as in [Genesis, chapter 10]. The WRR results are that ELSs for X tend to appear in close proximity to ELSs for Xעם, Xארץ, Xשפת, and Xכתב. If this is so, it is unnatural to believe that this phenomenon is restricted to these four prefixes. We would expect that X would also appear in close proximity to Xבני, to Xמלך, to Xאבי, and to X itself, without any prefix. With that in mind, we made a list of some 129 possible prefixes, and tested them all. Only prefixes in the length range 2-4 were considered. Longer prefixes produce very few distances, as attaching them to a nation name nearly always gives a word longer than 8 letters or a word with no ELS. Later on we tried a few additional prefixes, and we decided that it is better to disregard a few prefixes whose original inclusion was based on an error (misspelling, irrelevance). We were left with some 136 reasonable prefixes, shown in table 2. (The prefixes we decided to disregard are shown in table 3, to allay any suspicion that we were selective in our choice of prefixes.) We then computed the P_2 score corresponding to each one of the table 2 prefixes. The results of this computation are shown in table 4. The prefixes in that table are ordered by increasing P_2 score. Some have a very good P_2 score, some have a bad P_2 score. Overall, it seems that the P_2 scores observed are distributed rather evenly in the relevant range (between 0 and 1), and the corresponding permutation ranks (also shown in table 4) exhibit behavior not unlike what one would expect (see section 7 for a detailed analysis).

This by itself is a significant piece of evidence against the work of WRR. Essentially, we have just performed 132 attempted reproductions of the WRR experiment (4 of our 136 prefixes are the 4 used by WRR). If there were any truth in the phenomenon of codes, we'd expect to see many "good" P_2 scores and permutation ranks (values near 0), and only a few bad ones (values near 1). This is not what we get, and hence we view our 132 attempted reproductions of the WRR experiment as 132 failures.

But this is not all. Take a closer look at table 4; the biggest question mark regarding WRR's work is right there. If they did not have some prior knowledge, how does it happen that three of their four prefixes are the top three successes in table 4?

It is pointless to compute the odds of that happening by chance alone; some prefixes are clearly "better" than others. If codes do exist in the book of Genesis, they are probably more likely to speak of עם גמר (the nation of גמר) than of שיר גמר (the song of גמר). But why don't the codes speak about obvious and often-used constructs such as בני גמר, (people of גמר) מלך גמר (king of גמר), אל גמר (God of גמר) or צבא גמר (army of גמר)? Can one seriously say that כתב גמר (script of גמר) is more likely to be coded than דגל גמר (flag of גמר), זרע גמר (seed of גמר, offspring of גמר) or חטאי גמר (sins of גמר)? And who would have guessed that the Vilna Gaon would himself suggest the prefixes נמוס (laws of, manners of) and לבוש (dress of, clothing of), that a-priori seem somewhat artificial? (See section 2.2). While we cannot put a number on it, it is clear that the chance of picking the top three prefixes from within so many possibilities without any prior knowledge is very small.

(Just to get a feel for the scale: if there are just 50 prefixes and they are all equally likely, the probability of choosing four at random and finding that three of them are the top three is 1 in 4,900. If there are 100 such prefixes, the probability drops to 1 in 40,425.)

So maybe some prior knowledge was available to the authors of [WRR2, WRR3]? We can't tell exactly how they gathered such knowledge. Maybe they stared at rectangular-shaped printouts of the book of Genesis for a very long time, until they found a few anecdotal appearances of עם גמר or שפת מגוג, and only then they decided to turn it into a full scale sample? Maybe they found ELSs for, say, ארפכשד, continued them backward and found that they tend to be preceded by כתב, and only then they've decided to use the prefix כתב in their sample? Maybe they simply made up a table like our table 4, decided to use עם, כתב, and שפת, and then cooked up a good story around these prefixes and added the prefix ארץ to make the

story look better?

This brings us to the question, how good is the story? True, we have 132 failed replications and thus at least as many question marks, but maybe the wonder isn't with the non-existent success of the many prefixes in table 4, but rather in the nice story woven around the top prefixes in that table? After all, maybe the mere fact that such a nice story can be sewn around the top successes is by itself a miracle? The experiment is not a quantitative proof of that, and putting probabilities next to it (as WRR attempt to do) is silly, but still, isn't it impressive enough? So our next task is to try and see how good the story really is, and how hard it is to make up stories just as good.

2.1. How good is their story?

Here's the story, as WRR tell it in [WRR2]:

> Jewish tradition further tells us ([VG1]), that a nation has the following four characteristics:
> 1. its name.
> 2. its country.
> 3. its language.
> 4. its script.

Let us take a closer look at the relevant section in [VG1], a commentary on the book of Job by the Vilna Gaon:

> ידוע כי העמים בכתבי קדש היו נבדלים בארבעה ענינים. א, בשמותיהם. כי כל עם נקרא על שם אביהם הראשון אשר יוצאים ממנו כמו כוש מצרים פוט כנען וכדומה מהעמים אשר עד היום הזה נקראים כל זרעם מצרים אדומי הכל על שם הראשון. ב, בשם מדינותיהם. כי שמות המדינות נקראים ג"כ בשם המיסדם כמו ארץ מצרים. הנה נקראת המדינה ע"ש המיסד אותה וכן ארץ כוש ארץ כנען וכדומה. ג, בלשון. כי לכל א' הסכימו לשון מיוחד ונקרא לשון מצרי לשון כוש. הד, בכתב. אשר לכל א' יש כתב מיוחד. והן ארבעה ענינים הלה שאמרנו המצאם מפורש במאמר בראשית יו"ד אלה בני שם למשפחותם ללשונותם בארצותם לגויהם. משפחותם הוא ההבדל הא' מה שהאומה בכללה נקראת על שם ראש המשפחה אשר ממנו נחצבה. ללשונותם הוא ההבדל הג' שהם מתחלפים כלם בלשונות מתחלפות שונות זה מזה. בארצותם הוא ההבדל הב' ויסוב על שמות מדינותם. וההבדל הד' בכתב הוא מרומז ג"כ במאמר למשפחותם לגויהם שכל אומה נקראת גוי בפ"ע באין מחסור דבר.

In plain English (our translation),

> It is known that nations in the holy scriptures are distinguished by four characteristics: 1) by their name. For each nation is called after their first father from whom they descend, viz. Cush Mitzraim (Egypt), Phut, Canaan, etc., so that until the present day they are called Egyptians, Edomites, all on account of the originator. 2) By the name of their country. For the names of the countries are also called after the founders, like the land of Egypt - it is named after its founder. Similarly the land of Cush, the land of Canaan and so on. 3) By dialect. For each nation agreed upon a special dialect which is called the Egyptian dialect, the dialect of Cush. 4) In script. As each nation has its own script. These four characteristics you can find explicitly in Genesis chapter 10: "These are the sons of Shem according to their families, their dialect, in their lands, according to their nationalities." The phrase "their families" is the first distinction - that the nation as a whole is named after the head of the family from which it was hewed. "Their dialect" is the third distinction - that they all differ with distinct dialect for each. "In their lands" refers to the second distinction - and turns on the names of their countries. The fourth distinction - in the script - is also hinted at in the passage "their families , their nations" - that each and every people is called a "nation" unto itself without lacking anything.

Let us see what choice of prefixes would seem to follow if we accept the Vilna Gaon's defining characteristics of nationhood. The third characteristic is dialect - לשון is the word that the Vilna Gaon uses to describe dialect, and it is the word that the Torah uses. But instead, WRR use the word שפת. Notice that that the prefix שפת is very successful for their purposes (see table 4). Indeed, the words שפה and לשון are not completely synonymous. According to the commentary of Rabbi S. R. Hirsch [Hi] to this passage, שפה should be translated as "language" while לשון means "dialect".

Now let us turn to the first criterion - description by "name". The most reasonable constructions suggested by the Vilna Gaon's explanation would be to apply the prefix שם (name of), or to use the plain name of each nation without any prefix. Both possibilities give poor results. Instead, WRR chose to use the prefix עם, which only makes sense if it refers to the actual forefather - the nation of Gomer, for example. But WRR use it even as a prefix to the plural forms - the nation of the Ludites, in which case it has nothing to do with the Vilna Gaon's first criterion. Furthermore, the Torah uses the word גוי for nation, not עם. The Vilna Gaon in a commentary related to this chapter [VG2, on Isaiah 1:4] makes a point of explaining the distinction in meaning between these two terms. The prefix גוי performs much worse than עם, so again it seems that the choice made by WRR to use עם was to optimize their outcome, not as a result of any objective criterion.

There is a bit of additional "story" in [WRR2]. They say they choose the prefixes עם, ארץ, and שפת because they appear as prefixes in "Biblical expressions", and that using כתב as a prefix would make a "Biblical formation" where a "Biblical expression" does not exist. This very much seems like an ad-hoc excuse, and different and as good ad-hoc excuses could have been invented to support the prefix לשון (used by the Vilna Gaon in the above quote derived from [Genesis, chapter 10]), the empty prefix, or, for that matter, almost any other prefix. Furthermore, it is hard to take seriously the "Biblical expression" excuse, that applies to only 3 of the 4 prefixes and that applies just as well to many other prefixes that fail miserably. We also note that the issue whether a word is Biblical or not didn't seem to matter when WRR assembled columns 7 and 8 of table 1, where even misspelled non-Hebrew words are used.

In summary, the prefixes chosen by WRR are but poorly supported by the story they bring in support of them. Prefixes more obviously suggested by their story do much worse.

2.2. Why Vilna Gaon on Job?

Why look at [VG1], the Vilna Gaon's commentary on the book of Job, and not at [VG2], in his commentary on [Isaiah 1:4]? There the Vilna Gaon says:

> אחר המבול נתחלקו לשבעים אומות אם בסיבת פירוד משפחה ומשפחה לעצמה,
> או מחמת לשונות מופלגות, או ארצות מופלגות, או נימוסים מופלגים,
> וז"ש למשפחתם בארצותם ללשונותם לגויהם. ...

In short, the Vilna Gaon says that after the Deluge the people split into 70 nations, and lists four reasons: different families (the prefix משפחת, "family of"), different dialects (the prefix לשון), different countries (the prefix ארץ), and different laws (the prefix נמוס). Running the permutation rank program on Genesis with these four prefixes, the rank we get is 6/100, about 10,000 times worse than the rank the four WRR prefixes get, even though we share one prefix with them. Using the alternative spelling נימוס instead of נמוס the score is 9/100 - even worse. Using מנהג instead of נמוס we do a bit better and get a score of 3/100 (still 5,000 times worse than WRR), but we cannot justify this replacement either by reading the Vilna Gaon quote shown above or by reading the Bible, where the meaning of מנהג is "driving" or "charioteering", and not "laws" or "manners". Were WRR just lucky to choose the Job commentary over the Isaiah commentary?

In yet another place, his commentary on the book of Esther [VG3], the Vilna Gaon writes

...כי האומות נחלקים בד' דברים, בכתב ולשון, וארץ - ..., ובמלבושים.

.. because the nations differ by 4 attributes, by script כתב and dialect לשון, and country ארץ - ..., and by dress מלבוש.

Here we are led to consider the four prefixes כתב, לשון, ארץ, and מלבוש (or לבוש). With two prefixes in common with WRR, the permutation rank here is a mere 8/1000. Replacing מלבוש by the shorter and equivalent לבוש, we get 89/1000. Replacing it by the common בגד, we get 31/1000. Were WRR just lucky to choose the Job commentary over the Esther commentary?

Another possibility is to read one of the verses on which the Vilna Gaon bases his commentaries, [Genesis 10-5]:

מאלה נפרדו איי הגויים בארצתם, איש ללשנו, למשפחתם, בגויהם.

The King James translation of this verse is

By these were the isles of the Gentiles divided in their lands; every one after his tongue, after their families, in their nations.

The prefixes suggested here are אי (isle of), ארץ (country of), לשן (tongue of, meaning dialect of), משפחת (family of), and גוי (nation of). The permutation rank is 130/1000, another failure.

2.3. "The people of X"

Why did WRR consider the column labeled יםX in table 1? If X is a name of a country, the plural form of X is usually יםX, and in Hebrew it is a reference to the people of that country. If the country is ישראל, the people of that country are the ישראלים. The Vilna Gaon almost certainly makes no reference to that form (he only writes מצרים which is usually vocalized as mitsrayim (in the Bible), not mitsrim, and is not a plural, and אדומי, which is a singular (though in one source we've seen [VG1] quoted with אדומים; we think אדומי is more authentic, but we could be wrong). So it's not clear why the column יםX is at all present in table 1. It's neither "the name of X", nor "the nation of X", nor is it required by reading [VG1]. Anyway, there are at least two equally valid ways to refer to "the people of X": Xבני and Xאנשי. Looking at table 4, we see that neither one is a great success.

3. "Tuning" on War and Peace

Let us now see how easy (or hard) it is to invent stories like WRR's so as to get similar results on another text. There is a slight problem; for the distance function c(w,w') to be defined, the word w has to appear in the text with no skips. But very few texts other than Genesis contain in them words such as פתרסים. So we had to create our own text. We took an initial segment of a Hebrew translation of Tolstoy's War and Peace of the same length as Genesis, and replaced a part of it by chapter 10 of Genesis, so as to ensure that all the relevant names will indeed be found. We placed the Genesis graft in War and Peace in its original location in Genesis, from letter number 11,570 to letter number 12,692, and called the resulting text WGP (War Genesis Peace). The text WGP is available at http://cs.anu.edu.au/~bdm/dilugim/Nations/WGP.txt. (It is worthwhile to note that the Genesis piece inserted in WGP comprises only 1.4% of its total length, and fully contains only about 0.02% of the ELSs in WGP). We then computed the P_2-score of the prefixes in table 2. The results are in table 5. Finally, we went fishing for stories that could justify building a sample around some of the top prefixes in table 5. Here are three such stories (an additional brief story is in section 4.3.1):

3.1. The first story - Leaders

The most obvious story is to say that each nation has leaders, and hence we should expect a phrase like מלך מגוג (king of מגוג) to appear not very far from מגוג. Hence we take our prefixes to be מלך, מלכי, שר, שרי, ראש, and ראשי (king of, ruler of, leader of, in both singular and plural). Running the permutation test program with these prefixes as input and on the text WGP we get a score of 149/1,000,000. Not as good as the WRR choice of four prefixes, whose Genesis score is about 6/1,000,000, but nice enough for a first attempt.

There is a small problem here. Could it be that the surprising correlation we just found on WGP is due to the "G" (Genesis) part that we grafted in? If so, then the same list of 6 prefixes should do at least as well on the original Genesis. But its Genesis score is only 393,814/1,000,000, relieving our worries.

3.2. The second story - Celestial Guardians

We now explain how the choice of the prefixes מלכי, שר, עיר, and אלהי can be derived from the commentary of the great medieval scholar, Rabbi Moses ben Nachman, known as the Ramban (and also as Nachmanides to the English speaking public). In his commentary to [Leviticus 18:25] he discusses the celestial beings who represent and supervise the Nations of the world. We quote him at length. In the following quote the above four prefixes are clearly used, and the Ramban explicitly makes reference to our chapter in Genesis - the Table of Nations - in his commentary (we follow Chavel's translation [Na]):

> But the secret of the matter is in the verse which states "When the Most High gave to the nations their inheritance, when He separated the children of men, He set the borders of the People" etc. "For the portion of the Eternal is His people" etc. [Deuteronomy 32:8-9]. The meaning thereof is as follows: "The Glorious Name" [ibid. 28:58] created everything and He placed the power of the lower creatures in the higher beings, giving over each and every nation, "in their lands, after their nations" [Genesis, 10:31] some known star or constellation, as is known by means of astrological speculation. It is with reference to this that it is said. "which the Eternal thy G-d hath allotted unto all the people", [Deuteronomy 4:19] for He allotted to all nations constellations in the heavens, and higher above them are the angels of the Supreme One whom he placed as Lords (שרים) over them, as it is written, "But the prince (שר) of the kingdom of Persia withstood me" [Daniel 10:13] and it is written there, "lo the prince (שר) of Greece shall come" [ibid, 20]. They are called "kings" (מלכים) as it is written [there] "and I was left over there beside the kings (מלכי) of Persia" [Daniel 10:13]. Now "the Glorious Name" [Deuteronomy 28:58] is "G-d of gods (אלהים), and Lord of lords" [Deuteronomy 10:17] over the whole world. But the Land of Israel, which is in the middle of the inhabited earth, is the inheritance of the Eternal, designated to His name. ...
>
> ... nations over whom He appointed princes (שרים) and other celestial powers (אלהים), ...
>
> ... since He is the G-d of gods Who rules over all, and He will in the end "punish the host of the high heaven on high", [Isaiah 24:21] removing the celestial powers and demolishing the array of the "servants", and afterwards He will punish "the kings of the earth upon the earth" [Ibid.]. This is the meaning of the verse stating "The matter is by the decree of 'irin' (עירין) (the wakeful ones) and 'sh'elta' (the sentence) by word of the holy ones" [Daniel 4:14], meaning, the matter that was decreed on Nebuchadnezzar [that he be driven from men and he eat grass as oxen etc.] is the pronouncement of the guarding angels and the sentence of the word of the holy ones, who have ordained on the powers emanating from them that it be so. They [the angels] are called irin (עירין) [literally: "the wakeful ones"] because from their emanations proceed all the powers that stir all activities, similar to that which it says, "and behold 'ir' (עיר) (a wakeful one) and a holy one came down from heaven. He cried aloud and said thus: Hew down this tree" etc. [Daniel 4:10-11]. ...

> ... This is the meaning of the expression, "and he will go astray after the foreign gods (אלהי) of the land", ...

Notice that he uses the phrase מלאך (angel) as the supervisor or שר of the constellation of each nation. In this connection he uses the phrase שר יון, for example, to illustrate his principle. He maintains that these supervisory angels can also be referred to by the prefix מלכי and also illustrates it by referring to מלכי פרס. He also refers to these guardian angels as אלהים and as עירין. For this reason, we may feel fully justified in choosing as our prefixes מלכי, שר, עיר, and אלהי. Notice that these four words all appear explicitly (three of them as prefixes) in the above quotation. When a word appears in the quotation in both singular and plural form, we always prefer the (shorter) singular form. When it appears only in plural form, we can't know if it is reasonable to use it in singular form and hence we keep it in plural form.

Running our program again, using these four prefixes on the text WGP, we get a permutation rank of $5/10^8$, about 100 times better than the WRR prefixes on Genesis! Once again our prefixes get a boring result on Genesis ($310{,}817/10^6$), so our result is not due to the G piece in WGP.

3.3. The third story - Genesis ain't a War Epic

Our third story is admittedly the weakest, but it is instructive and fun. And after all, what's life without a smile here and there?

Consider the prefixes חיל (army of, corps of), גדוד (battalion of), חרב (sword of), and אלפי (thousands of, family of, part of the tribe of). The first three are clearly war-related. The fourth also feels war-like to the Hebrew-trained ear. That feel is confirmed by the verse

"ויקרבו אל משה הפקדים אשר לאלפי הצבא שרי האלפים ושרי המאות",

from [Numbers 31:48].

> "And the officers which were over thousands אלפי of the host, the captains of thousands, and captains of hundreds, came near unto Moses".

It is quite un-surprising that such a war-like lineup of prefixes does well on the war-epic-based WGP, getting a permutation score of 6/1000. It is just as expected that on Genesis, a book of creation and holiness, this lineup fails miserably, getting a permutation score of 999,460/1,000,000.

3.4. Stories, Stories, Stories

In this section we have shown some stories that produce good scores in WGP. None of them are perfect, but at least one is as good as WRR's (who didn't even use the actual prefixes their story gave).

Of course, the real question is not how many successful stories can be found for Genesis or WGP. The real question is: faced with a set of prefix performances which are essentially random, what is the probability that a successful story can be woven around some of the best prefixes? It is a very difficult question, but our success with WGP and the experience it gave us suggests that the probability is quite good. The number of potential stories is so vast that one of them is almost certain to be a winner.

Obviously we could also use our data to concoct additional stories that give good results for Genesis. Readers should remember that our case is that finding good stories is easy for any text.

4. More evidence

4.1. The plural form

[WRR2] explains the column labeled יX in table 1 as follows:

> Concerning the expressions "יX" [meaning: the people of X], we are aware of the fact that only valid linguistic formations should be used; therefore we use these expressions if, at least, their singular form יX appears in the Hebrew Bible. For instance, for גמר (item 1) we do not take גמרים, since the expression גמרי does not appear in the Hebrew Bible; so we cannot know whether it has any meaning at all. The check was done using Even-Shoshan's New Concordance of the Bible (Even-Shoshan, 1981).

The decisions WRR made in this paragraph seem completely arbitrary. Why not take all plural forms? In most cases no Hebrew speaker would have any difficulty deciding what the correct plural gentilic form of a given nation's name is, even if no hints are given in the Bible. Alternatively, why not require that the plural appears explicitly in the Bible? And why remove the definite article from nations 36-44 (from היבוסי through החמתי)? This they did without even telling us about it even though they kept it except in this place. In addition, it seems that WRR did not even fully comply with the rule they stated above, for they have taken אשורים as the plural form of אשור. The word אשורים does not appear in the Bible, and the word אשורי appears there once, in reference to the tribe of אשר. When the people of אשור are mentioned in the Bible, they are called אשורם. Finally, there is a problem with the treatment of לודים. See below.

It turns out that if WRR's mistake is corrected or if any of their decisions is reversed, it always weakens their result. Is it reasonable to assume that WRR knew which column to choose if they didn't try some of the other possibilities? Here are the relevant scores in a tabular form: (the raw data is in table 6)

Table 7: Permutation test scores (out of 10^6) for the various choices regarding the plural form.

	WRR	WRR without אשורים	WRR with אשורם	WRR without לודים	Common Sense Extension	Explicit Plural	In Chapter	Keep ה	All With ה
Number of words considered	33	32	33	32	62	23	19	33	33
Score	4,836	28,534	44,661	14,683	119,448	70,361	133,070	18,745	609,227

Let us explain the columns of this table.

- The first column contains WRR's choice for plurals. That choice gets a permutation test score of $4{,}836/10^6$ when isolated from the rest of WRR's sample.
- The second column is the same, only with the incorrect "אשורים" removed. In the third column "אשורים" is replaced by the correct plural, "אשורם".
- There seems to be an error in the treatment of the form לודים. Nation number 26 is called לודים, while nation number 48 is called לוד. As the text clearly identifies them as different nations, there's no justification (following the WRR rules) for using the plural form לודים for nation number 48 - the word לודי does not appear in the Bible at all, and the word לודים appears just twice, in both cases in reference to nation number 26. So how can we be sure (in WRR's words) that the plural form of לוד is indeed לודים? The column titled "WRR without לודים" is the same as the column labeled "WRR", only

with the form לודים removed for nation number 48.

- The column labeled "Common Sense Extension" is what we consider as the common sense extension of the WRR choice of plurals to the part of the list for which WRR provided no plurals. There are only six nations for which the decision on the plural is not clear for a Hebrew speaker or determined by precedents in the WRR column; these are the six nations whose name ends with a ה. Common Hebrew practice does not determine the correct plural in these cases, so we left them out. In Hebrew, there are three standard forms for the name of the people of a country whose name ends with a ה. Either the ה gets replaced by a ים (i.e., the people of קנדה are the קנדים), or by an אים, or by a נים (אמריקה and אמריקאים or אמריקנים). There is also a less common (but Biblical!) possibility of replacing the ה with תים (עזה and עזתים). Using these four possibilities uniformly, we get scores of 160,251, 151,285, 135,122, and 158,741 out of 10^6, respectively.
- The column labeled "Explicit Plural" is about those plurals that appear explicitly in the Bible as plurals. Note that the fact that the singular form is Xי does not guarantee that the plural form is Xים. For example, the singular form אדמי appears in the Hebrew Bible several times (for instance, in [Deuteronomy, 23:8]). The word אדמים appears in the Bible several times, but not as the plural of אדמי. On the other hand, the plural אדמיים appears once [1 Kings, 11:17] and the plural אדומים appears twice [2 Chronicles, 25:14 and 28:17].
- To produce the plural forms that appear in the column labeled "In Chapter", we applied the same rules that WRR used, only that we restricted the search for the forms Xי and Xים to [Genesis, chapter 10], just to be absolutely certain that when such a form appears it pertains to nations and nothing else.
- In the column labeled "WRR", WRR stripped the definite article from the names numbered 36 to 44, without giving any justification for doing so, even though these definite articles were not removed in forming columns 1-5 of table 1. Just as well they could have kept the definite article for these names, or prepended the definite article to all names that didn't have them already. The columns labeled "Keep ה" and "All With ה" contain these two possibilities.

4.2. The identification of the nations

[WRR2] explains the columns labeled "new nation name(s)" and "new country name" in table 1 as follows:

> In Talmudic literature we find various new names for nations and countries, and we cannot decide in favor of one against the other. We preferred to use a single source. The Aramaic translation of Genesis that gives the largest number of new names is Targum Yonatan. We used Targum Yonatan as printed in Torah Shelemah (Kasher, 1929) (with Torah Shelemah's corrections according to the Ginsburger manuscript and others). ...
>
> ... Targum Yonatan is a translation into Aramaic. We need the identifications of the nations, but we do not intend to check whether the Aramaic appears non-randomly as ELS's. Thus, where the Hebrew formation exists, we used it alone.
>
> ... Here, as in [WRR1], we used the grammatical orthography - ktiv dikduki.

In this section we wish to study how reasonable was WRR's choice of a source, how faithfully they followed it, and whether the idea of looking for the "new" identifications of the nations and the countries was sound to start with.

4.2.1. On orthography and on back translation

"Grammatical Orthography", supposedly used in [WRR2, WRR3], was defined in [WRR1] by reference to the entry for "כתיב" in the Hebrew-Hebrew dictionary [ES]. The definition given there for grammatical orthography (ktiv dikduki) is (our translation) "some words are spelled using defective spelling, כתיב חסר, and some using plene spelling, כתיב מלא, according to the rules of the masoretic grammar [the grammar used in the masoretic Bible] ... In our time, grammatical orthography is common only in vocalized texts [texts with nikud]. The orthography in this dictionary is grammatical orthography ...". In other words, the notion of grammatical orthography is simply not defined for words in foreign languages such as Aramaic. Also, when WRR write "where the Hebrew formation exists...", what exactly do they mean? Clearly, there is some choice in deciding what is the precise Hebrew formation parallel to some Aramaic word. So overall, WRR left themselves much freedom in transcribing Ginsburger's Targum Yonatan. Let us see how they used this freedom. In so far as we could tell, WRR obtained their list of identifications from Ginsburger's Targum Yonatan (as printed in [Ka]) using the following procedure:

1. For any given nation, look at the identification(s) in the text of Targum Yonatan and the identification(s) in the Ginsburger footnotes. Ignore all identifications that differ from the original nation name only by a minor spelling change (עירק versus ערק, for example).
2. If the identification(s) in the text and in the footnotes differ, the footnotes take precedence.
3. Several identifications are given in the Aramaic plural form, with the letters "אי" appended at the end of the word. If the identification at hand is a new country name, WRR keep the suffix "אי" in place. If it is a new nation name, they translate the suffix into the Hebrew form "ים".
4. Finally, the following 7 corrections are made (some, presumably, because of the back translation into Hebrew, and some, presumably, to comply with grammatical orthography): המדיי gets replaced by המדי, מקדיניא by מקדוניא, אכייא by אכיא, ערביא by ערב, לובאי by לוב, and פנטפוליטיה by פנטפוליטי, and מווריטינוס gets altogether dropped. (The last two corrections are irrelevant to the end result because the words involved are longer than 8 letters).

We've tried several variants of the WRR procedure. First, some of the 7 corrections WRR made seemed unnatural. Why is המדיי replaced by המדי? The original Targum Yonatan is not vocalized with the exception of a very few vowel signs, but in [Ka], the word המדיי is vocalized, and there's a Hiriq under the first Yud. In this vocalization, it is senseless to remove the second Yud. Why is מקדיניא replaced by מקדוניא but אפריקי is not replaced by אפריקא or אפריקה and טרסס is not returned to the form תרשיש? Why is אכייא replaced by אכיא? The word אכייא is not vocalized in [Ka] and neither אכייא nor אכיא appear in the Bible, so the change is at least optional. Why is לובאי replaced by לוב but סמראי is not replaced by סמר? In each case, one may come up with some excuse, but one could also come up with an equally valid excuse for making a different decision.

As it turns out, undoing any of the first 5 of the WRR corrections weakens their result (and the last 2 are irrelevant, as noted above):

Table 8: Permutation test scores out of 10^6.

WRR (columns 7 and 8 of table 1)	turning המדי back to המדיי	turning מקדוניא back to מקדיניא	turning אכיא back to אכייא	turning ערב back to ערביא	turning לוב back to לובאי
4,277	11,891	11,872	5,737	5,599	9,105

It is interesting to note that the job of back-translating Targum Yonatan into Hebrew was carried out earlier, for completely different purposes, by Rieder [Ri2]. We have extracted the Hebrew formations out of [Ri2], and kept the Aramaic where [Ri2] gave no parallel Hebrew formations, getting a new list of identifications. This list (which has a better claim to be "correct" than the WRR list, as it was made in advance by an independent professional) is shown in the column titled "WRR with Rieder's Hebrew (literal)" in table 9. The column titled "WRR with Rieder's Hebrew (modified)" is the same, only that we've attempted to convert Rieder's original spellings into what appears to be the WRR notion of correct spelling - namely, we've reduced double Yuds into single Yuds and used the word both with and without the vowel "א" when it appears as a "mater lectionis" (see [WRR1]). The column titled "WRR - Modified Ginsburger" contains the original WRR choices. The results of the permutation test for these lists of identifications are shown in table 10. The two "Rieder's Hebrew" lists get results which are about 100 times weaker than the WRR result!

Note. WRR often prefer to spell geographical names with a final א, rather than with a final ה. When the final ה's in the two "Rieder's Hebrew" columns are replaced by א's uniformly (except for ברבריאה, which ends with a ה even in the "WRR" column), the permutation test scores are slightly better but they are still statistically insignificant and around 50 times worse then the WRR score. For the "literal" column the score is 157,726, and for the "modified" column it is 230,659.

Table 10: Permutation test scores (out of 10^6) for other choices of spellings and back translations.

WRR - Modified Ginsburger	WRR with Rieder's Hebrew (literal)	WRR with Rieder's Hebrew (modified)	WRR, reducing וו	WRR with ים	WRR with אי	WRR with ה
4,277	401,722	462,379	10,036	4,752	9,036	9,325

The other columns in table 9 and table 10 are:

- In the columns labeled "WRR, reducing וו" we took the WRR list of identifications and replaced the two instances of a double ו by a single ו. From the perspective of grammatical orthography (whatever that means for non-Hebrew words), this is just as justified as the replacement יי to י practiced twice by WRR. Notice that according to the vocalization in Kasher [Ka], the וו's are pronounced with an "ah" sound, so the second ו in each case is unnecessary.
- Recall that WRR translated the אי suffixes to the Hebrew ים suffixes only for the identifications the countries. Though Rieder [Ri2] does the same, we still feel that the distinction between countries and nations in this context is artificial; זינגאי is certainly not a Hebrew name of a country or a nation, and if זינגאי is acceptable, then נסיוטאי is just as reasonable. Thus in the columns labeled "WRR with ים" we've uniformly replaced all אי suffixes by ים, and in the columns labeled "WRR with אי" we've uniformly made the opposite choice.
- The Aramaic suffix "אי" often contains in it the definite article, in which case the Hebrew formation parallel to the Aramaic נסיוטאי is הנסיוטים. Thus in the columns labeled "WRR with ה" we added the definite article ה in all cases where WRR made the אי to ים conversion.

A quick glance at table 10 shows that these additional reasonable spelling variations give results weaker than WRR's. How did WRR know which precise translation scheme was best for them?

After this section was completed, Doron Witztum issued a paper [Wi] in which he gave the following clarification of his 'ktiv dikduki' rule:

> The original quote reads, "For words in Hebrew, we always choose what is called the grammatical orthography ...". Note that we specifically say "words in Hebrew," not "Hebrew words" - that is, any word which has been rendered into Hebrew, even if derived from a foreign language, is to be written in grammatical orthography. The only expressions which do not fall under this rubric are words derived from languages which themselves use Hebrew characters, such as Yiddish and Ladino, because these languages do not need to be rendered into Hebrew. This rule was followed consistently in the construction of both published lists regarding all foreign names.

We sincerely wonder if Mr Witztum really believes what he wrote, as both the English phrase "words in Hebrew", and the Hebrew phrase "מלים בעברית" (which is how the rule appears in [WRR3]) mean "words belonging to the Hebrew language", which does not include foreign words merely transliterated into the Hebrew script. Nevertheless, our present interest in the quotation concerns the reason given for excluding Yiddish and Ladino. It is obvious that Aramaic fits the same conditions perfectly. Therefore, according to Mr Witztum's own interpretation, whether or not it makes sense, we should not apply 'ktiv dikduki' to Aramaic words, nor in fact modify their spelling in any way. Obviously this contradicts what he did in [WRR2].

In any case, the primary issue is not whether it is defensible to apply 'ktiv dikduki' to Aramaic words. The primary issue is whether it was required of WRR to do that. The answer is clearly that they had a choice, as they had not considered Aramaic words before (to our knowledge) and their rule can be argued to apply or not very easily. As we have demonstrated, they made the decision in the way most favorable to themselves and then applied it in the most favourable way, even to the extent of making advantageous errors.

4.2.2. A bit on Targum Yonatan

What is "Targum Yonatan"? The Talmud relates that Yonatan ben Uziel, a student of Hillel, fashioned an Aramaic translation of the Prophets. It makes no mention of any translation by him of the Torah. So all scholars agree that this Targum is not due to Yonatan ben Uziel. Indeed, de Rossi [de Ro] (16th century) reports that he saw two very similar complete Targumim to the Torah, one called Targum Yonatan Ben Uziel and the other called Targum Yerushalmi. A standard explanation, cf. for example [Go] is that the original title of this work was Targum Yerushalmi, which was abbreviated to תי", and these initials were then incorrectly expanded to Targum Yonatan which was then further incorrectly expanded to Targum Yonatan ben Uziel. For this reason, scholarly publications generally refer to the work as Targum Pseudo-Yonatan.

The first of these manuscripts cited by de Rossi is thought to have been the basis of the first printing in Venice (1591) where the false title Targum Yonatan ben Uziel is used. The second manuscript - the only known one to still exist - is in the British Museum and was published by Ginsburger in 1903. This formed the basis for Rabbi Kasher's editorial notes in his Torah Shelemah. Unfortunately, Ginsburger's work was unreliable, so the whole basis of the choices of WRR is without foundation. More on this later.

The so-called Targum Yonatan ben Uziel is more than a mere translation. It includes much Aggadic material collected from various sources as late as the Midrash Rabbah as well as earlier material from the Talmud. So it is a combination of a commentary and a translation. In the portions where it is pure translation, it often agrees with the Targum Onqelos.

As to the date of its composition, this is a matter of dispute. The majority opinion, on the basis of much internal evidence, is that it cannot date from before the Arab conquest of the Middle East despite incorporating some older material. For example, Ishmael's wife is called by the legendary Arabic name Fatimah. Such evidence is summarised in [Ma]. As an upper bound, it is referred to (perhaps for the first time) in 15th century commentaries. Gottleib [Go] puts the time of composition toward the end of the

eighth century. On the other hand, since the Geonim are unfamiliar with it, and Rashi doesn't mention it, Rieder [Ri1] puts the composition some time after Rashi, perhaps during the period of the crusades. The one surviving manuscript was probably written in the 16th century and is an unknown number of generations removed from the original.

Now to Ginsburger's "corrections" which were incorporated by Rabbi Kasher into his Torah Shelemah. Rieder (Leshoneinu vol 32. p.298) demonstrates that Ginsburger's edition is full of errors and incorrect copying and that it confused the printed version with manuscript and vice versa. Rieder writes as follows in the introduction to his edition [Ri1]:

> The many errors in the edition of Ginsburger have tripped up many investigators in their work. Outstanding among these is Rabbi M. M. Kasher, in his Torah Shelemah where he copied the errors of Ginsburger or his "corrections" as if they were actually in the manuscript.

He goes on to give many startling examples of such false "variants".

The third author would like to record here his fond memory of many pleasant and instructive conversations he had with Rabbi Kasher over 40 years ago. Rabbi Kasher was a man of great warmth, intelligence, and wit. In a work of such encyclopedic proportions as the Torah Shelemah, it was inevitable that he had to rely on the scholarship of others, rather than to examine all manuscripts himself. It is unfortunate that he was tripped up by Ginsburger.

4.2.3. Other identifications of the nations

To remedy the poor choice of the Ginsburger "corrections" made by WRR, we have also run the permutation program using the Rieder edition of the Targum Yonatan, and using the Targum Yonatan manuscript printed in [Cl]. The identifications of countries and nations in the Rieder edition of Targum Yonatan and in [Cl] are presented in the columns labeled "Rieder" and "Clarke", respectively, in table 11 (note that we've treated the אי suffixes in the same way as WRR). The columns labeled "Modified Rieder" and "Modified Clarke" in table 11 are the same, only that we've tried to back translate Rieder and [Cl] to Hebrew and use "grammatical orthography" following the precedents in [WRR2, WRR3]. The column labeled "Ginsburger" in table 11 contains the source of the Ginsburger "corrections" as it appears in [Ka], and the column labeled "Modified Ginsburger" is the same, only modified as in [WRR2, WRR3] - in other words, it is simply the WRR choice of identifications.

We have also run the permutation program on several other plausible lists of identifications (well, at least no less plausible then WRR's choice). These lists are also printed in table 11, under the columns labeled "Targum Yerushalmi" and "Modified Targum Yerushalmi" (Targum Yerushalmi as printed in [Ka]), "Talmud Yerushalmi" and "Modified Talmud Yerushalmi" (Talmud Yerushalmi as in the Responsa database [Re]), "Talmud Bavli" and "Modified Talmud Bavli" (Talmud Bavli as in [Re]), and "Midrash Rabbah" and "Modified Midrash Rabbah" (Midrash Rabbah as in [Re], Vilnius version).

The results of these runs of the permutation program are printed in table 12. The rows labeled "#" in that table contain the number of identifications appearing in each source.

Table 12: Permutation test scores (out of 10^6) for the various country and nation identifications.

Source:	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
#	39	39	34	22	16	10	22
Score:	66,445	721,658	686,283	54,081	870,221	453,783	482,268
Source:	WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
#	38	37	32	22	16	10	22
Score:	4,277	353,231	285,627	180,365	104,791	85,065	563,935

This table clearly indicates that the decision WRR made (to use Ginsburger's Targum Yonatan) worked to their benefit.

One other thing that we tried was to collect all identifications appearing in Kasher [Ka]. After all, WRR explicitly write "in Talmudic literature we find various new names for nations and countries, and we cannot decide in favor of one against the other", and that they took Ginsburger's Targum Yonatan because it "gives the largest number of new names". So clearly, WRR think that the more the better, and clearly, as [Ka] is the source they've used, [Ka] was available for them. So according to their logic, taking everything that appears in [Ka] should give better results than just taking the identifications in Ginsburger's Targum Yonatan.

Oddly enough, the opposite turns out to be true, as shown in table 13, with the corresponding raw data in table 14. In table 14, the column titled "Kasher" contains the identifications appearing in Kasher in their literal form (except that we've treated the אי suffixes like WRR do), the column titled "Modified Kasher" is the same only that we've used the WRR precedents for back translation and grammatical orthography, and in the column titled "Modified Kasher, reducing וו" we've further reduced all occurrences of וו to ו, as explained in section 4.2.1.

Table 13: Permutation test scores (out of 10^6) for the identifications appearing in Kasher's "Torah Shelemah" ([Ka]).

	WRR - Modified Ginsburger	Kasher	Modified Kasher	Modified Kasher, reducing וו
#	38	138	128	127
Score:	4,277	251,140	9,615	27,120

We note that the Kasher data fully contains the Ginsburger data (by definition: the Ginsburger data appears in the Kasher book), and hence any optimization that WRR may have performed on the Ginsburger data (the choice of spelling rules and of back translation, the mere fact it was chosen, and chosen over other possible sources, the treatment of the אי suffixes) improves also the scores for the Kasher data. So it is of interest to note that if the Ginsburger data is removed from the three Kasher columns in table 14, the permutation scores drop to 586,547, 127,056, and 214,454 of 10^6, respectively.

For the amusement of our readers and to help them appreciate the wealth of data in Kasher's book that WRR ignored, we include one page out of Kasher's book as figure 1. (The PostScript version of this figure is available separately from the main PostScript file, at http://cs.anu.edu.au/~bdm/dilugim/Nations/KasherPage.ps).

4.2.4. More on Genesis, chapter 10

Chapter 10 of Genesis, known as the Table of Nations, is a unique document in all of ancient literature. Much effort throughout history has been devoted to identifying the nations named in the table. For a partial survey of the literature up to 1971, see the bibliographical references at the beginning of [We, Chapter 10]. A computer search has yielded many many more references. It is certainly not our intent to give a survey of this vast field, nor are we competent to do so. But a few remarks are in order.

For some of these nations, the identification is certain. Thus Yavan (יון) is clearly Ionia, and Madai (מדי) is clearly Media since these names have survived unchanged down to our time. Some identifications are almost universally agreed upon, such as Gomer (גמר) being identified with the the Cimmerians (Kimmeroi) in the classical authors and Gimirrai in the cuneiform texts. Some suggestions are intriguing - for example Westerman's observation ([We]) that Japeth (יפת) (the name of one of the sons of Noah) agrees phonetically with the Greek Iapetos, one of the Titans in Greek mythology, the son of Ouranos and Gaia. Some of the identifications of the nations are in dispute, and some nations are completely unidentified.

The organizing principle of the list is also a matter of scholarly discussion. For example, [AARS, Kr, WF] suggest that the division is geographical: three spheres of peoples and lands which meet in the region of the Holy Land. Other organizing principles - by way of life, or by profession, for example - have been suggested. Some of the names seem to be place names, some names of tribes or peoples, and some names of individuals.

Suppose a (partial) list of identifications has been worked out in a given period and in a given language. In the course of time, this list of identifications will inevitably be corrupted, especially if the copyist does not speak the language and is completely unfamiliar with the geographical terms. For example, the printed version of the Babylonian Talmud (Yoma 10a) identifies Madai and Yavan as "Macedonia and 'as its name implies'". This is clearly a mistake. Yavan should have been identified as Macedonia, and the name Madai=Media is unchanged. Obviously the copyist or the typesetter, not understanding what he was dealing with, simply transposed the two identifications. (This particular error was caught by Rabbi Joel Sirkes, see Hagahot haBah ad locum.) In other places, where we do not know the correct identification of either the Hebrew original or the Aramaic translation, undetected errors are unavoidable. This is why the WRR experiment using names from the Targum Yonatan was meaningless from the start, even without the gaffe of using Ginsburger's incorrect version.

4.3. Other nice experiments

4.3.1. The relation to the text

The title of [WRR2] contains the phrase "the relation to the text". But instead of measuring the relation to the text, their distance function c(w,w') measures the relation to the text and to the reversed text - in the computation of c(w,w') the word w is searched for with both skip +1 (text) and skip -1 (reversed text). This means, for example, that in the computation of c(מש, ארץ מש), not only occurrences of the word מש in the text of Genesis are considered, but also all 519 occurrences of the word שם, whose meaning is completely different. Thus the phrase "the relation to the text" is false advertising, and should have been replaced by "the relation to the text and the reversed text".

If reverse occurrences of the word w are not considered in the computation of c(w,w'), in most cases this does not change the value of c(w,w') by much, because most names in column 1 of table 1 simply do not appear in the reversed text anywhere. But it is still amusing to note that the honest "relation to the text" result is $25/10^7$ (computed using our own programs), about 150 times weaker than "the relation to the text and the reversed text" result computed by WRR.

At this point, when textual matters are at the height of our concerns, we are lead to consider the textual residues of the ancient nations, their books and their letters, and the semantic information we can gather out of these texts, the language, script, and orthography of these nations. Thus we are led to consider the prefixes ספר (book of), מכתב (letter of), שפת (language of), כתב (script of), and כתיב (orthography of). True to the current context, we test only the relation to the true text and not the relation to the reversed text. Doing so, we get a permutation rank of 19/1,000,000 for WGP. For Genesis, the permutation rank is 338/1,000,000, consistent with the fact that two of the prefixes we use here are taken from the WRR sample.

Note that we have found a reasonable experiment that gets quite a good result in two different texts simultaneously. The reader may choose to infer that achieving a good result in only one text is not much of a challenge.

Could it be that the WGP success is at least partially due to codes that lie entirely within chapter 10 of Genesis and have nothing to do with War and Peace? This would be the case, and it would weaken our point considerably, if it turns out that there are many ELSs for expressions of the form pX entirely within chapter 10 of Genesis, where p is one of the 5 prefixes listed above and X is one of the 68 descendents of Noah. But it turns out that the number of such ELSs in chapter 10 of Genesis is precisely zero.

4.3.2. The identifications of the nations and countries

Column 2 of table 1 contains phrases of the form Xעם, "the nation of X", and column 7 of that table contains the "modern" identifications (at least according to WRR) of these nations. Pretending to believe WRR, we expect that ELS occurrences of the phrases in column 2 should appear close to ELS occurrences of the words in column 7. As most phrases in column 2 do not appear in the text or the reversed text, the appropriate distance function to use here is the one used in [WRR1]. The permutation rank one gets is 39/100 (on 7 measured distances) - completely insignificant. Similarly it makes sense to compute the permutation rank of column 2, Xעם, against column 6, Xים, using the distance function of [WRR1]. The result is 78/100 (on 20 measured distances). Finally, column 3, Xארץ against the identifications of the countries in column 8, gives 16/100 (on 5 measured distances).

4.3.3. An alternate form for "the language of X"

In Hebrew, just as the construct Xים refers to "the people of X", so does the construct Xית refer to "the language of X" (for example, when in [WRR3] WRR refer to the language in which Targum Yonatan is written, they call it "ארמית"). This calls for another experiment - run column 1 of table 1 against a variant of column 6 in which all ים suffixes get replaced by ית. The result is 68/100 - a failure.

4.4. Some global choices

This section is different in nature from the previous ones (and from the following ones) in that it is the only one in which we report on permutation ranks that are meaningful only if evaluated using at least a billion permutations, a number so large that according to WRR (in [WRR3]) it was at the edge of what their computers could do. (Specifically, they write in a comment following table 6 of [WRR3] that they couldn't compute the P_1 permutation rank (see [WRR1]) with a billion permutations because they lacked computer power). Thus, in all likelihood, WRR could not have performed the experiments reported below.

Some of the nations in [Genesis, chapter 10] appear there with some grammatical prefixes and/or suffixes, and not in the raw form. Namely, nations 36-44 (from היבוסי through החמתי) appear with both the definite article prefix ה and the possessive suffix י, and nations 13, 14 and 26-33 (כתים, דדנים, and from כפתרים to לודים) appear with the plural suffix ים. It makes sense to experiment with the raw nation names (without the prefixes and suffixes). Here are some results:

Table 15: Permutation test scores (out of 10^{10}) for some global choices. (For the 3 largest scores we used only 10^9 permutations and then renormalized).

WRR	Deleting ה and י for #36-44	Deleting ים for #13, 14, 26-33	Deleting all grammatical prefixes/suffixes	Ignoring repetitions (see below)	P_1 permutation rank
146	17,710	28 (137)	3,420 (12,030)	46	6

Comments:

1. Deleting the ים suffixes turns לודים into לוד and דדנים into דדן. לוד and דדן are already nation names, and so the total number of nation names reduces to 66. The numbers not in parenthesis in table 15 result from a computation that ignores this fact. The way to treat this fact is suggested by the way WRR treated the fact that the names שבא and חוילה each appear twice in [Genesis, chapter 10] - they just reduced the number of entries in their lists from 70 to 68. Doing the same we arrive at a list of 66 names, and to compute the permutation rank we have to choose random permutations out of 66! possibilities. The "corrected" scores are presented (when applicable) in table 15 in parenthesis.
2. In addition we checked what happens if the original WRR experiment is re-evaluated completely ignoring repetitions of nation names, namely, using a total of 70 names. The results are on the column labeled "Ignoring repetitions" of table 15.
3. Finally, we computed the P_1 permutation rank of the original WRR sample. The result is reported in the column "P_1 permutation rank".

In table 15 we find the only examples of modifications of the WRR experiment that improve its results. In view of the large number of modifications we examined, all that these exceptions indicate is that WRR did not "play" with these particular parameters (remember also that "playing" would have required more computer power than they had, by their own testimony).

5. How much more have we done?

In this context, very little. We've reported all the prefixes that we tried. We've reported all variations of the original WRR experiment that we tried, with the exception of a small number of variations that originated in an error (misreading of some source texts, typos, etc.). None of the variations that we didn't report changes the picture in any way; none gave extreme values, and none did better than the original WRR list.

The text WGP that we used for some of our examples was the second (of the two) that we considered (for that purpose). Before switching to WGP we played briefly with the same verse permuted version of Genesis that WRR used as a control text. But very quickly we realized that the fact that the text locations of the 68 names of nations in the permuted Genesis are evenly spread changes its statistical properties in a significant way and makes it a very poor control (see a further discussion in section 7).

5.1. Disclaimer

We made every effort to ensure that this article is free of substantial errors. But no scientific piece of work of comparable size can be entirely error-free. If you find any errors in this article, please let us know and we will fix them in the next edition (and add your name to the acknowledgements...).

6. Conclusions

WRR used the prefix שפת rather than the prefix לשון used both in [Genesis, chapter 10] and in [VG1], even though the two prefixes have different meaning. This deviation from the originals is both highly unnatural and highly beneficial to WRR's purposes. It seems that a similar game was also played with the prefix עם. This by itself is a clear indication that WRR looked at the data before assembling their final sample, and this by itself is sufficient to invalidate their results.

But as we have seen in section 2, this is not all. WRR managed to "guess" 3 of the most successful prefixes without being distracted by some much more "tempting" guesses such as בני or מלך, which are used in the Bible hundreds of times, without being distracted by other sources who wrote on chapter 10 of Genesis, without being distracted by other writings of the Vilna Gaon on that same chapter, and without letting the precise language used in [VG1] interfere. And as shown in section 4, almost every other aspect of their experiment is perfect or near perfect. Is this believable?

Reuven tosses a coin 20 times in a row, and 20 times in a row he claims that he guessed right which way it will fall, before it actually fell. Reuven clearly proved something to a significance level of 2^{-20}. What did Reuven prove?

Finally, we know from section 3 and section 4.3.1 and from the experience we gained in [BM] that deliberate cheating can produce results that appear to be highly significant without really being so.

6.1. A Personal note by the first author

I wish to express special thanks to Prof. David Kazhdan of Harvard University for his insistence (in public and in person) that the WRR claims on codes in the Book of Genesis deserve serious study, whatever the conclusions may be.

Much as I thank David for introducing me to this subject, and much as I learned from (but not enjoyed!) working on it, it is time for me to move on (and back) to other things, or else the harm already done to my career will grow fatal and my kids will go hungry. So except for some commitments I have already made (the most important of which is moral support to Brendan), I quit from this subject. This paper is sure to get a response from WRR, and I (and my co-authors) may get personally attacked. My future lack of response should not be interpreted as my inability to respond, but rather as my concern for my family.

I first approached the results of WRR with a great deal of scientific (and personal) curiosity. The results of [WRR1], and even more so, of [WRR2, WRR3], seemed very solid, and, if true, they would have been one of the most significant human discoveries ever. As the findings reported here have unfolded, my curiosity turned into anger, and over the 5-6 months that we spent writing this paper, my anger got replaced by fond memories of the good old times before all of that, when I was studying honest mathematics. It is a good time to quit.

7. Appendix: Notes on the Metric

This appendix is more technical in nature than the rest of the paper, and is intended for the more mathematically and/or statistically oriented readers. In this appendix we study some of the technical features of the distance function c(w,w') and some aspects of the statistical methodology used in [WRR2, WRR3]. We show that c(w,w') has some undesired features that lead to a failure of the statistical analysis of [WRR2, WRR3]. Thus, even if the serious problems uncovered in the main part of this paper were not present, one could not take the significance result reported by WRR, of some parts in a billion, as anything meaningful.

The basic problem should be clear to anyone who studies table 1. The Nations Sample is about the c(w,w')-distance between the words in the column labeled X and the words in the other columns. But there is an obvious structural correlation between these two sets of words; most of the words w' in columns 2-8 are just extensions of the corresponding words in the first column. Therefore their lengths are correlated and their letter-contents are correlated. Thus there is a clear structural difference between the identity permutation and any other permutation; anybody (even a computer) can pick out the identity permutation from within a billion random permutation by looking at completely structural properties of the lists, completely independently of the text, without the need of any understanding of the historical or theological context.

It is the job of the distance function c(w,w') and the statistical analysis to filter out this effect. They should have been designed in such a way that the obvious structural peculiarity of the identity permutation would not give it any a-priori advantage (or disadvantage) over a randomly chosen permutation. Unfortunately, c(w,w') and P_2 do not do their job properly, and the identity permutation remains distinguished. This undermines the permutation test and renders its results invalid.

In section 7.1 we show the results of some very brief tests we did on control texts. A distinct non-uniformity in the permutation ranks is noted, especially at the extremes. However, Genesis does not seem at all exceptional from this point of view. In section 7.2 we offer some preliminary discussion.

7.1. The permutation test on control texts

We begin by summarising the behaviour of the permutation text for Genesis and WGP, applied for our 136 prefixes. The raw data for some of this table appears in table 2.

Table 16: Statistics of various permutation ranks for the 136 prefixes in the texts Genesis and WGP. (All numbers computed using our program, not WRR's).

	Genesis			WGP		
bin	original	restricted	forward	original	restricted	forward
0	21	23	17	23	25	28
1	19	17	21	11	10	12
2	12	9	14	17	13	10
3	12	13	17	12	14	10
4	10	11	6	14	12	12
5	9	12	9	11	15	7
6	11	10	14	5	5	14
7	16	14	17	11	10	8
8	14	16	7	15	14	13
9	12	11	14	17	18	22
min	536	493	1465	527	475	3429
mean	463298	463125	457162	473084	477535	481131
max	990821	990921	996492	997088	998727	997716

1. The columns marked "original" concern the metric as WRR define it. The nation names may appear forwards and backwards, and there is no restriction on the placement of the ELS of the second word. The columns marked "restricted" are the same, except that ELSs lying entirely inside the section of the text containing Chapter 10 of Genesis are ignored. The columns marked "forward" are like "original" except that the nation names must appear in the forwards direction.
2. The first ten rows show the numbers of prefixes for which the rank order lies in each of bins 0,1,...,9, where bin i contains permutation rank orders 100000i+1 through 100000(i+1) out of a million.
3. The last three rows show the smallest, average, and maximum rank orders out of a million.

Several things can be seen from the table. There does seem to be an excessive number of ranks in the smallest bin, and the minimums appear to be smaller than expected. (For 136 independent uniform variables, the expected minimum is 7299.)

One might imagine in that evidence of codes, but that would be unfounded. A minimum of 500 or less is not very unlikely (almost 6.6% if the ranks are uniform and independent). Moreover, the minima for Genesis and WGP are about the same, and the tiny differences between the "original" and "restricted" distributions show that to be not primarily due to the Genesis 10 portion of WGP. Standard goodness-of-fit tests fail to show any difference between the distributions for Genesis and WGP, even though they do verify that the overall non-uniformity is probably real.

To investigate whether non-uniformity and exaggerated extremes are the norm, we ran the same tests on 10 control texts. We generated each control text by randomly permuting the order of the words within each verse of Genesis except for the verses in Chapter 10. Chapter 10, where all nation names are found in a specific fixed order, was left alone in order to make the comparison with Genesis and WGP more meaningful.

The results are tabulated below. ELSs within Chapter 10 are ignored, just as for the "restricted" columns in the previous table. (However, as for the previous table, this restriction makes hardly any difference.)

Table 17: Statistics of the permutation ranks for the 136 prefixes in ten control texts. (All numbers computed using our program, not WRR's).

bin	Text 0	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Text 7	Text 8	Text 9
0	15	19	22	23	18	17	14	13	19	19
1	13	13	8	15	12	6	13	18	13	18
2	17	13	16	15	8	18	13	15	11	13
3	13	14	7	9	7	16	18	10	12	11
4	9	7	13	7	10	12	9	13	7	7
5	16	11	13	12	16	5	12	8	10	8
6	10	7	9	22	22	10	8	13	18	10
7	9	13	20	10	9	13	11	15	16	10
8	13	15	15	14	14	23	15	13	11	13
9	21	24	13	9	20	16	22	18	19	27
min	5314	3991	1584	608	874	4979	180	6184	6247	2481
mean	504905	517197	496493	458654	534142	519963	522828	505910	512851	514246
max	999807	999923	999141	997583	999864	999849	999229	997025	991129	998778

The most obvious characteristic of these distributions is their inconsistency. A few are near-uniform, but others are skewed markedly in the positive or negative direction. Both the minima and maxima appear exaggerated.

Several things should be clear from these results. Firstly, WRR's assumption of uniformity in the rank orders is unfounded, as many of the texts give profoundly non-uniform distributions. More importantly, the non-uniformity may be more pronounced at the extremes of the distribution where WRR measure their "significance levels".

7.2. Discussion

Given the absence of any real evidence of a non-chance phenomenon, it is hard to see much justification for the months-long study that would be needed to understand the permutation test properly in this context. However, we can make a few pertinent observations.

Take the artificial example of two nation names, ABCD and EFGH, and two prefixes, WX and YZ. If EFGH has many ELSs (compared to perturbed ELSs) with a letter close to an appearance of ABCD in the text, then both c(ABCD,WXEFGH) and c(ABCD,YZEFGH) have an enhanced probability of being small. Similarly for the opposite extreme. Thus, c(ABCD,WXEFGH) and c(ABCD,YZEFGH) will tend to be positively correllated. This applies to each pair of prefixes, but the degree of correllation will depend in a complicated manner on the lengths of the prefixes and the letters they contain.

This correllation between the prefixes could partly explain the overall non-uniformity evident in the tables above. It also undermines the WRR result severely, as they use four prefixes at once and they will tend to reinforce each other. However, it is not clear whether the correllation adequately explains the actual shape of the distribution or the exaggerated extremes.

Another contribution to the confusion is the natural clumpiness of letter distributions through the text. If w and w' have letters in common, this clumpiness will tend to make c(w,w') smaller than expected. Obviously, this favors the identity permutation especially.

Other factors are operating as well, but we don't pretend to know their effect. Our lack of understanding is illustrated by the absence of any plausible theory to explain the exaggerated maximum scores in our experiments. However, we are not embarrassed by this situation. The embarrassment belongs to WRR, who used their defective methods without even acknowledging that there were questions to be asked about them.

A final point to make is that even if Genesis gave profoundly different results from randomly-generated control texts, we would not know the reason. Anyone with a small amount of Hebrew can see at a glance that Genesis is not random, so why should we expect it behave as if it is? Natural texts have many statistical properties that random texts don't, including clumpiness of letter distributions at all scales, short-range correllations, and perhaps [ASEAS] long-range correllations. How do we know that WRR's ELS experiment is not just measuring some property of natural texts? (Recall that in WRR's experiment the identity permutation is structurally distinct!) The question suggests using other natural texts as controls, but why should we expect those to behave quantitatively like Genesis? The lack of suitable controls for computer analysis of biblical texts is a well known dilemma. See [Bar] for a discussion.

8. Bibliography

[AARS]
Y. Aharoni, M. Avi-Yonah, A. F. Rainey and Z. Safrai The MacMillan Bible Atlas (third edition), MacMillan, New-York 1995.

[ASEAS]
M. Amit, Y. Shmerler, E. Eisenberg, M. Abraham and N. Shnerb, Language and Codification Dependence of Long-range Correlations in Texts, Fractals 2 (1994) 7-13.

[BW]
M. Bar-Hillel and W. A. Wagenaar, The Perception of Randomness, Advances in Applied Mathematics 12 428-454 (1991).

[BM]
D. Bar-Natan and B. McKay, Equidistant Letter Sequences in Tolstoy's "War and Peace", draft, 1997.

[Bar]
D. J. Bartholomew, Probability, Statistics and Theology, Journal of the Royal Statistical Society A 151 (1988) 137-178.

[Cl]
E. G. Clarke, Targum Pseudo-Jonathan of the Pentateuch: text and concordance, Ktav Publishing House, Hoboken New-Jersey 1984.

[ES]
A. Even-Shoshan, A new dictionary of the Hebrew language, Kiryat-Sefer, Jerusalem 1989.
[Go]
Z. Y. Gottleib, Targum Yonatan ben Uziel al haTorah, Melila 1 (1944) 26-34.
[Hi]
S. R. Hirsch, Hamishah humshe Torah, English translation by M. Ha-Levi, Judaica Press, Gateshead 1976.
[Ka]
M. M. Kasher, Torah Shelemah, (1929).
[Kr]
E. G. Kraeling, Rand McNally Bible Atlas, Rand McNally, New York 1956.
[Ma]
M. Maher, Targum Pseudo-Jonathan: Genesis, Volume 1B of The Aramaic Bible, The Liturgical Press, Collegeville 1992.
[Na]
The Nachmanides, הרמב"ן, Commentary on the Torah, פרוש התורה, translated and annotated by C. B. Chavel, Shilo, New-York 1971-76.
[Pf]
C. F. Pfeiffer, Baker's Bible Atlas, Baker Book House, Grand Rapids, 1961.
[Ri1]
D. Rieder, Targum Yonatan ben Uziel al hamishah humshe Torah, hu'tak miktav yad London, British Museum Add. 27031, im hearot vetikunim me'et David Rieder, haAkademyah haAmerikanit lemada'e haYahadut, Jerusalem 1974.
[Ri2]
--, Targum Yonatan ben Uziel al haTorah, meturgam lelvrit im be'urim zionei mekorot umakbilot, Jerusalem 1984.
[Re]
The Bar-Ilan Responsa CD-ROM, version 5, Bar-Ilan University, Israel.
[de Ro]
dei Rossi Azariah ben Moses, Meor Enayim, Makor, Jerusalem 1969/70.
[VG1]
The Vilna Gaon, A Commentary on the Book of Job, באור על ספר איוב, ספר דבר אליהו.
[VG2]
--, Aderet Eliyahu, אדרת אליהו.
[VG3]
--, A Commentary on Megilat Esther, ספר מגילת אסתר.
[We]
C. Westerman, Genesis 1-11, a Commentary, English translation, Augsburg Publishing House, Minneapolis 1984.
[WRR1]
D. Witztum, E. Rips and Y. Rosenberg, Equidistant Letter Sequences in the Book of Genesis, Statistical Science 9-3 (1994) 429-438.
[WRR2]
--, -- and --, Equidistant Letter Sequences in the Book of Genesis: II. The Relation to the Text, preprint, circa 1995. Retyped copy available at http://cs.anu.edu.au/~bdm/dilugim/Nations/WRR2/.
[WRR3]
--, -- and --, צפן חבוי בדילוג שווה בספר בראשית: מובהקות סטטיסטית של התופעה, (Hidden codes in Equidistant Letter Sequences in the Book of Genesis: The Statistical Significance of the Phenomenon), Preprint accompanying a lecture given by E. Rips to the Israeli Academy of Sciences, 1996.

[Wi]
D. Witztum, A Refutation Refuted, Part 2. Available as a PDF file.
[WF]
G. E. Wright and F. V. Filson, The Westminster Historical Atlas to the Bible, The Westminster Press, Philadelphia 1956.

The WRR Sample of Nations: Table 1

The full table of nations: The first column contains the names of the descendents of Noah (X's). Columns 2-8 contain the "related words": the regular constructs Xעם, Xארץ, Xשפת and Xכתב in columns 2-5, the plural construct Xים in column 6, and the externally-provided identifications of the nations and countries in columns 7 and 8 respectively.

The full table of nations.

	1	2	3	4	5	6	7	8
	X	עסX	ארץX	שפתX	כתבX	Xים	new nation name(s)	new country name
1	גמר	עם גמר	ארץ גמר	שפת גמר	כתב גמר			אפריקי
2	מגוג	עם מגוג	ארץ מגוג	שפת מגוג	כתב מגוג			גרמניא
3	מדי	עם מדי	ארץ מדי	שפת מדי	כתב מדי	מדים		המדי
4	יון	עם יון	ארץ יון	שפת יון	כתב יון	יונים		מקדוניא
5	תבל	עם תבל	ארץ תבל	שפת תבל	כתב תבל			יתיניא
6	משך	עם משך	ארץ משך	שפת משך	כתב משך			מוסיא
7	תירס	עם תירס	ארץ תירס	שפת תירס	כתב תירס			תרקי
8	אשכנז	עם אשכנז	ארץ אשכנז	שפת אשכנז	כתב אשכנז			אסיא
9	ריפת	עם ריפת	ארץ ריפת	שפת ריפת	כתב ריפת			פרכוי
10	תגרמה	עם תגרמה	ארץ תגרמה	שפת תגרמה	כתב תגרמה			ברבריאה
11	אלישה	עם אלישה	ארץ אלישה	שפת אלישה	כתב אלישה			אלס
12	תרשיש	עם תרשיש	ארץ תרשיש	שפת תרשיש	כתב תרשיש			טרסס
13	כתים	עם כתים	ארץ כתים	שפת כתים	כתב כתים	כתים		אכיא
14	דדנים	עם דדנים	ארץ דדנים	שפת דדנים	כתב דדנים	דדנים		דרדניא
15	כוש	עם כוש	ארץ כוש	שפת כוש	כתב כוש	כושים		ערב
16	מצרים	עם מצרים	ארץ מצרים	שפת מצרים	כתב מצרים	מצרים		
17	פוט	עם פוט	ארץ פוט	שפת פוט	כתב פוט			אליחרק
18	כנען	עם כנען	ארץ כנען	שפת כנען	כתב כנען	כנענים		
19	סבא	עם סבא	ארץ סבא	שפת סבא	כתב סבא	סבאים		סיניראי
20	חוילה	עם חוילה	ארץ חוילה	שפת חוילה	כתב חוילה			הינדיקי
21	סבתה	עם סבתה	ארץ סבתה	שפת סבתה	כתב סבתה			סמראי
22	רעמה	עם רעמה	ארץ רעמה	שפת רעמה	כתב רעמה			לוב
23	סבתכא	עם סבתכא	ארץ סבתכא	שפת סבתכא	כתב סבתכא			זינגאי
24	שבא	עם שבא	ארץ שבא	שפת שבא	כתב שבא	שבאים	זמרגר זמרגד	
25	דדן	עם דדן	ארץ דדן	שפת דדן	כתב דדן	דדנים	מזג	
26	לודים	עם לודים	ארץ לודים	שפת לודים	כתב לודים	לודים	ניווטים	
27	ענמים	עם ענמים	ארץ ענמים	שפת ענמים	כתב ענמים	ענמים	מריוטים מרסיוטים	
28	להבים	עם להבים	ארץ להבים	שפת להבים	כתב להבים	להבים	ליווקים	
29	נפתחים	עם נפתחים	ארץ נפתחים	שפת נפתחים	כתב נפתחים	נפתחים	פנטסכינים	
30	פתרסים	עם פתרסים	ארץ פתרסים	שפת פתרסים	כתב פתרסים	פתרסים	נסיוטים	
31	כסלחים	עם כסלחים	ארץ כסלחים	שפת כסלחים	כתב כסלחים	כסלחים	פנטפוליטי	
32	פלשתים	עם פלשתים	ארץ פלשתים	שפת פלשתים	כתב פלשתים	פלשתים		
33	כפתרים	עם כפתרים	ארץ כפתרים	שפת כפתרים	כתב כפתרים	כפתרים	קפודקים	
34	צידן	עם צידן	ארץ צידן	שפת צידן	כתב צידן	צידנים		
35	חת	עם חת	ארץ חת	שפת חת	כתב חת	חתים		
36	היבוסי	עם היבוסי	ארץ היבוסי	שפת היבוסי	כתב היבוסי	יבוסים		
37	האמרי	עם האמרי	ארץ האמרי	שפת האמרי	כתב האמרי	אמרים		
38	הגרגשי	עם הגרגשי	ארץ הגרגשי	שפת הגרגשי	כתב הגרגשי	גרגשים		

The full table of nations.

	1	2	3	4	5	6	7	8
	X	עמX	ארץX	שפתX	כתבX	Xים	new nation name(s)	new country name
39	החוי	עם החוי	ארץ החוי	שפת החוי	כתב החוי	חוים		
40	הערקי	עם הערקי	ארץ הערקי	שפת הערקי	כתב הערקי	ערקים		
41	הסיני	עם הסיני	ארץ הסיני	שפת הסיני	כתב הסיני	סינים	אנתוסים ארתוסים כפרוסים	
42	הארודי	עם הארודי	ארץ הארודי	שפת הארודי	כתב הארודי	ארודים	רודסים	
43	הצמרי	עם הצמרי	ארץ הצמרי	שפת הצמרי	כתב הצמרי	צמרים	חומצים	
44	החמתי	עם החמתי	ארץ החמתי	שפת החמתי	כתב החמתי	חמתים	אנטיוכים	
45	עילם	עם עילם	ארץ עילם	שפת עילם	כתב עילם			
46	אשור	עם אשור	ארץ אשור	שפת אשור	כתב אשור	אשורים		
47	ארפכשד	עם ארפכשד	ארץ ארפכשד	שפת ארפכשד	כתב ארפכשד			
48	לוד	עם לוד	ארץ לוד	שפת לוד	כתב לוד	לודים		
49	ארם	עם ארם	ארץ ארם	שפת ארם	כתב ארם	ארמים		
50	עוץ	עם עוץ	ארץ עוץ	שפת עוץ	כתב עוץ			
51	חול	עם חול	ארץ חול	שפת חול	כתב חול			
52	גתר	עם גתר	ארץ גתר	שפת גתר	כתב גתר			
53	מש	עם מש	ארץ מש	שפת מש	כתב מש			
54	שלח	עם שלח	ארץ שלח	שפת שלח	כתב שלח			
55	עבר	עם עבר	ארץ עבר	שפת עבר	כתב עבר	עברים		
56	פלג	עם פלג	ארץ פלג	שפת פלג	כתב פלג			
57	יקטן	עם יקטן	ארץ יקטן	שפת יקטן	כתב יקטן			
58	אלמודד	עם אלמודד	ארץ אלמודד	שפת אלמודד	כתב אלמודד			
59	שלף	עם שלף	ארץ שלף	שפת שלף	כתב שלף			
60	חצרמות	עם חצרמות	ארץ חצרמות	שפת חצרמות	כתב חצרמות			
61	ירח	עם ירח	ארץ ירח	שפת ירח	כתב ירח			
62	הדורם	עם הדורם	ארץ הדורם	שפת הדורם	כתב הדורם			
63	אוזל	עם אוזל	ארץ אוזל	שפת אוזל	כתב אוזל			
64	דקלה	עם דקלה	ארץ דקלה	שפת דקלה	כתב דקלה			
65	עובל	עם עובל	ארץ עובל	שפת עובל	כתב עובל			
66	אבימאל	עם אבימאל	ארץ אבימאל	שפת אבימאל	כתב אבימאל			
67	אופר	עם אופר	ארץ אופר	שפת אופר	כתב אופר			
68	יובב	עם יובב	ארץ יובב	שפת יובב	כתב יובב			

The WRR Sample of Nations: Table 2

Comments:

1. In this table we've restricted our attention to prefixes whose length is 2-4 letters, for the reasons explained in the text.
2. The prefixes labeled with a "+" sign at the "Meaning" column were added as this article was typed up, and were not a part of the initial list of prefixes.
3. We emphasize again that not all prefixes listed below are of equal importance. Some, such as בני or מלך are very natural and are used extensively in biblical sources. Others are used less extensively. We warn our readers against cutting the list of prefixes into "more important prefixes" and "less important prefixes" according to some arbitrary criteria. The human eye is surprisingly good at detecting patterns, even within random data (see e.g. [BW]). Once you have stared at the data shown here for a sufficiently long time, you may fool yourself into thinking that you've discovered a pattern which is "extremely unlikely to be due to chance alone". If you insist on attempting such a subdivision, the only reliable way to do so is by turning to a third party who has not seen the numerical data on this table and on table 4, and by phrasing your request to him/her in completely general terms.
4. We emphasize that we did not suppress any prefixes that we have examined. Barring unlikely technical glitches, every prefix (other than the 5-letter מלבוש, משפחת and נימוס) we ever tried is here or in table 3.

The prefixes ordered lexicographically.

Prefix	Meaning	Genesis Rank	WGP Rank
	the empty prefix	116	100
אב	father of, progenitor of	127	62
אבי	father of, progenitor of	63	17
אדמת	land of	110	92
אוכל	food of	101	33
אומת	nation of	76	63
אופי	character of	96	16
אחזת	+ estate of, property of	83	60
אי	isle of, see Genesis 10-5	86	117
איי	isles of, see Genesis 10-5	89	38
איל	lord of, magnate of	51	132
אילי	plural of איל	21	82
איש	person of, man of	9	104
אכל	food of	27	55
אל	god of	50	74
אלה	god of, celestial power of	32	26
אלהי	plural of אלה	18	5
אלוף	head of, leader of	109	88
אליל	idol of, false god of	42	80
אלף	thousand of, family of, part of the tribe of	62	108
אלפי	plural of אלף	135	50
אמת	+ nation of	106	128
אנשי	people of, men of	33	35
ארץ	country of	23	72
אשם	guilt of	81	46
אשמת	guilt of	66	12

The prefixes ordered lexicographically.

Prefix	Meaning	Genesis Rank	WGP Rank
אשת	+ woman of	24	18
בגד	cloth of, dress of	118	130
בגדי	plural of בגד	34	45
בירת	capital of	122	98
בית	+ house of, family of, dynasty of	128	85
בן	+ son of, native of	133	93
בנות	+ plural of בת	55	110
בני	plural of בן	71	99
בת	+ daughter of, female native of	17	21
בתי	+ plural of בית	129	115
גבול	border of	60	56
גבור	hero of	15	122
גבר	man of, hero of	31	123
גברי	plural of גבר	91	127
גדוד	battalion of	136	20
גוי	nation of	12	78
גזע	race of	37	54
גר	convert of	87	131
גרי	plural of גר	124	22
דבור	speech of	79	42
דגל	flag of	99	73
דגלי	plural of דגל	77	121
דין	justice of, law of	107	96
דיני	plural of דין	54	59
דם	+ blood of	114	57
דמי	+ plural of דם	117	9
דרך	way of	111	103
דרכי	plural of דרך	92	84
דת	religion of, law of	58	64
דתי	laws of	28	102
זהב	gold of	104	67
זכות	right of, privilege of	65	112
זמן	time of, period of, fate of	13	107
זמר	song of, melody of	43	49
זמרת	variation of זמר	69	13
זרע	seed of, offspring of	49	75
חבל	region of, district of	88	124
חטא	sin of	74	95
חטאי	plural of חטא	36	116
חיל	army of, corps of	115	94
חילי	plural of חיל	6	37
חק	law of	38	40
חקי	plural of חק	100	19
חרב	sword of	125	8
טבע	nature of, character of	121	66
כהן	priest of	47	97
כהני	plural of כהן	19	29

The prefixes ordered lexicographically.

Prefix	Meaning	Genesis Rank	WGP Rank
כוח	alternative spelling of כח	57	76
כח	force of, vigour of	7	44
כחות	plural of כח	103	125
כסף	money of	48	86
כספי	plural of כסף	39	58
כתב	script of	2	43
כתיב	orthography of	35	23
לאם	nation of	61	79
לבוש	dress of, clothing of	120	113
לשון	dialect of	64	114
לשן	dialect of	59	71
מאכל	food of, diet of	52	36
מושב	place of, seat of	130	47
מזל	luck, destiny, constellation of	102	136
מטבע	coin of	30	91
מכתב	letter of, writing of	85	30
מלאך	angel of	72	39
מלך	king of	41	105
מלכי	plural of מלך	56	2
מלכת	+ queen of	45	68
מלל	speech of	126	3
ממשל	rule of, government of	95	111
מנהג	manner of, custom of	4	31
מס	tax of, levy of	75	32
מסי	plural of מס	46	87
מעות	money of	90	109
מקום	place of	44	41
מראה	appearance of, view of	8	77
נמוס	+ manners of, customs of, laws of	70	24
נס	flag of, banner of, miracle of	40	135
נסי	plural of נס	108	48
נשות	+ plural of אשת	73	65
נשיא	president of, head of	5	119
סוף	end of, conclusion of	29	133
סמל	symbol of, emblem of	78	52
סמלי	plural of סמל	132	89
ספר	book of	67	25
עוון	alternative spelling of עון	82	90
עון	sin of	123	81
ע"ז של	idolatry of	98	101
עיר	city of, angel of	94	4
עם	nation of	1	14
ערי	towns of	112	51
פשע	crime of	105	83
פשעי	plural of פשע	68	6
צאצא	child of, descendant of	26	1
צבא	army of, troops of	84	70

The prefixes ordered lexicographically.

Prefix	Meaning	Genesis Rank	WGP Rank
ראש	head of, leader of	20	61
ראשי	plural of ראש	93	10
רוח	spirit of	80	118
רכב	+ chariot of	11	15
רקוד	dance of	119	27
שדה	+ field of	10	126
שדות	+ plural of שדה	14	28
שיר	song of, hymn of, music of	53	134
שירי	plural of שער	131	120
שם	+ name of, fame of	134	53
שמות	+ plural of שם	16	34
שפת	language of	3	11
שר	ruler of, head of, minister of, prince of	113	7
שרי	plural of שר	97	106
שרש	root of	25	69
תושב	resident of	22	129

The WRR Sample of Nations: Table 3

Warning: The computation of the numbers in the columns labeled "Permutation rank" is extremely complicated. For reasons of efficiency we had to rewrite the code for computing c(w,w'), and the results reported in these columns may be very slightly different than if we had used the original WRR program.

The prefixes we disregarded.

Prefix	Meaning	P_2 in Genesis	Permutation rank (out of 10^6) in Genesis	P_2 in WGP	Permutation rank (out of 10^6) in WGP
אומה	misspelled אומת	0.353483	946134	0.203476	711904
אחזה	misspelled אחזת	0.721375	756609	0.389444	301846
בול	stamp of	0.342712	533645	0.951279	972178
בולי	plural of בול	0.958761	952078	0.169267	200260
חוק	misspelled חק	0.999081	977964	0.954404	890243
חוקי	misspelled חקי	0.706497	559334	0.330676	160335
לאום	misspelled לאם	0.0787993	239085	0.730964	565149
מכס	custom of, duty of	0.795042	188012	0.456631	160932
מכסי	plural of מכס	0.840942	910733	0.817495	727363
מעה	misspelled מעות	0.24885	160897	0.657419	500210
ע"ז	idolatry (not "idolatry of")	0.0888923	40244	0.615443	211661

The WRR Sample of Nations: Table 4

Warning: The computation of the numbers in the column labeled "Permutation rank" is extremely complicated. For reasons of efficiency we had to rewrite the code for computing c(w,w'), and the results reported in this column may be very slightly different than if we had used the original WRR program. For the same, results reported in this column may be slightly inconsistent with results reported elsewhere in this paper, where the original WRR program was used.

Comment: The entries in this table are sorted by their P_2 score and not by the more reliable permutation rank. Computing all permutation ranks was simply not feasible at the time of writing of [WRR2], and we believe they could not have had this data available. To get these numbers we had to run our more efficient program for several days on about 50 very fast workstations in parallel, using the second author's scheduling tool autoson. (The computation of just the P_2 score without computing permutation ranks was easily feasible at the time of writing of [WRR2]. We made this computation on a single CPU using the program supplied to us by WRR, long before we were able to compute the permutation ranks.) Sorting by permutation ranks, שפת, עם and כתב are number 1, 3, and 9 respectively. Not as good as 1-2-3, but still way too good to be a chance hit.

The prefixes ordered by increasing P_2 score on Genesis.

Genesis Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
1	עם	nation of	0.0119782	6175
2	כתב	script of	0.0154729	42139
3	שפת	language of	0.0167725	536
4	מנהג	manner of, custom of	0.0482328	75435
5	נשיא	president of, head of	0.0521407	287171
6	חילי	plural of חיל	0.0533556	82120
7	כח	force of, vigour of	0.0543148	92503
8	מראה	appearance of, view of	0.0592405	32997
9	איש	person of, man of	0.0805091	2139
10	שדה	+ field of	0.081064	118958
11	רכב	+ chariot of	0.084107	83087
12	גוי	nation of	0.0871378	102459
13	זמן	time of, period of, fate of	0.105434	42084
14	שדות	+ plural of שדה	0.114518	19913
15	גבור	hero of	0.115053	165933
16	שמות	+ plural of שם	0.115651	269493
17	בת	+ daughter of, female native of	0.12688	199425
18	אלהי	plural of אלה	0.133538	54473
19	כהני	plural of כהן	0.137136	515791
20	ראש	head of, leader of	0.164057	344663
21	אילי	plural of איל	0.166803	64860
22	תושב	resident of	0.170673	112436
23	ארץ	country of	0.176805	82530
24	אשת	+ woman of	0.187643	452918
25	שרש	root of	0.189707	103651
26	צאצא	child of, descendant of	0.189729	16704
27	אכל	food of	0.196157	321710
28	דתי	laws of	0.19987	299330
29	סוף	end of, conclusion of	0.202499	305248

The prefixes ordered by increasing P_2 score on Genesis.

Genesis Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
30	מטבע	coin of	0.212442	166068
31	גבר	man of, hero of	0.238762	151527
32	אלה	god of, celestial power of	0.243384	161608
33	אנשי	people of, men of	0.257713	232531
34	בגדי	plural of בגד	0.259907	97049
35	כתיב	orthography of	0.280834	144275
36	חטאי	plural of חטא	0.297355	135886
37	גזע	race of	0.305544	371713
38	חק	law of	0.340336	844631
39	כספי	plural of כסף	0.342699	86522
40	נס	flag of, banner of, miracle of	0.354192	91661
41	מלך	king of	0.35584	281726
42	אליל	idol of, false god of	0.357117	124102
43	זמר	song of, melody of	0.385473	570668
44	מקום	place of	0.395972	197409
45	מלכת	+ queen of	0.400825	143567
46	מסי	plural of מס	0.406045	7902
47	כהן	priest of	0.449381	705860
48	כסף	money of	0.462919	695886
49	זרע	seed of, offspring of	0.466294	447203
50	אל	god of	0.473938	391624
51	איל	lord of, magnate of	0.478264	116177
52	מאכל	food of, diet of	0.497015	203657
53	שיר	song of, hymn of, music of	0.498342	179397
54	דיני	plural of דין	0.505132	685798
55	בנות	+ plural of בת	0.506676	930618
56	מלכי	plural of מלך	0.507817	497716
57	כוח	alternative spelling of כח	0.508809	631113
58	דת	religion of, law of	0.516599	72295
59	לשן	dialect of	0.516989	765353
60	גבול	border of	0.517919	676500
61	לאם	nation of	0.542476	122390
62	אלף	thousand of, family of, part of the tribe of	0.546674	812903
63	אבי	father of, progenitor of	0.552237	859528
64	לשון	dialect of	0.555848	298878
65	זכות	right of, privilege of	0.555934	762899
66	אשמת	guilt of	0.556714	396914
67	ספר	book of	0.559084	313562
68	פשעי	plural of פשע	0.559515	801434
69	זמרת	variation of זמר	0.579658	701240
70	נמוס	+ manners of, customs of, laws of	0.583041	203717
71	בני	plural of בן	0.588154	354627
72	מלאך	angel of	0.590789	837107
73	נשות	+ plural of אשת	0.594379	207989
74	חטא	sin of	0.604537	532389
75	מס	tax of, levy of	0.611794	248979
76	אומת	nation of	0.614179	279027

The prefixes ordered by increasing P_2 score on Genesis.

Genesis Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
77	דגלי	plural of דגל	0.615957	522896
78	סמל	symbol of, emblem of	0.630773	658010
79	דבור	speech of	0.640026	655979
80	רוח	spirit of	0.642974	771622
81	אשם	guilt of	0.645628	144676
82	עוון	alternative spelling of עון	0.650594	636614
83	אחזת	+ estate of, property of	0.658322	571898
84	צבא	army of, troops of	0.671741	422587
85	מכתב	letter of, writing of	0.681591	221995
86	אי	isle of, see Genesis 10-5	0.683857	341205
87	גר	convert of	0.686306	62959
88	חבל	region of, district of	0.688237	778963
89	איי	isles of, see Genesis 10-5	0.692233	136006
90	מעות	money of	0.703272	433695
91	גברי	plural of גבר	0.706495	544000
92	דרכי	plural of דרך	0.707976	868229
93	ראשי	plural of ראש	0.715351	427358
94	עיר	city of, angel of	0.721183	368140
95	ממשל	rule of, government of	0.734041	729211
96	אופי	character of	0.735826	753121
97	שרי	plural of שר	0.736456	349603
98	ע"ז של	idolatry of	0.745627	970273
99	דגל	flag of	0.76363	687090
100	חקי	plural of חק	0.769939	926929
101	אוכל	food of	0.781038	805747
102	מזל	luck, destiny, constellation of	0.782135	486478
103	כחות	plural of כח	0.785669	535477
104	זהב	gold of	0.787238	814631
105	פשע	crime of	0.789287	447249
106	אמת	+ nation of	0.793969	527593
107	דין	justice of, law of	0.796568	761291
108	נסי	plural of נס	0.799767	632074
109	אלוף	head of, leader of	0.800194	764038
110	אדמת	land of	0.803608	480728
111	דרך	way of	0.806929	797123
112	ערי	towns of	0.807486	958515
113	שר	ruler of, head of, minister of, prince of	0.8114	730219
114	דם	+ blood of	0.812034	878426
115	חיל	army of, corps of	0.817688	916551
116		the empty prefix	0.818587	707578
117	דמי	+ plural of דם	0.8274	460308
118	בגד	cloth of, dress of	0.832618	778238
119	רקוד	dance of	0.835076	362500
120	לבוש	dress of, clothing of	0.844589	947535
121	טבע	nature of, character of	0.852564	671436
122	בירת	capital of	0.885983	654089
123	עון	sin of	0.888212	866002

The prefixes ordered by increasing P_2 score on Genesis.

Genesis Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
124	גרי	plural of גר	0.899475	850617
125	חרב	sword of	0.907717	947356
126	מלל	speech of	0.908192	796914
127	אב	father of, progenitor of	0.908487	800602
128	בית	+ house of, family of, dynasty of	0.914892	818314
129	בתי	+ plural of בית	0.918177	733968
130	מושב	place of, seat of	0.930393	834604
131	שירי	plural of שער	0.931032	970804
132	סמלי	plural of סמל	0.945328	582368
133	בן	+ son of, native of	0.960413	987776
134	שם	+ name of, fame of	0.962751	922350
135	אלפי	plural of אלף	0.973716	990821
136	גדוד	battalion of	0.997802	966635

The WRR Sample of Nations: Table 5

Warning: The computation of the numbers in the column labeled "Permutation rank" is extremely complicated. For reasons of efficiency we had to rewrite the code for computing c(w,w'), and the results reported in this column may be very slightly different than if we had used the original WRR program. For the same reason, results reported in this column may be slightly inconsistent with results reported elsewhere in this paper, where the original WRR program was used.

Comment: The Hebrew letter צ is rare in WGP, and hence the prefix צאצא appears there as an ELS infrequently. In fact, it appears only in conjunction with a single nation, חת. Thus its appearance at the top of the table below is of limited interest.

The prefixes ordered by increasing P_2 score on WGP.

WGP Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
1	צאצא	child of, descendant of	0.0147059	29493
2	מלכי	plural of מלך	0.0490328	6392
3	מלל	speech of	0.0515043	76223
4	עיר	city of, angel of	0.0795474	6081
5	אלהי	plural of אלה	0.0804652	16470
6	פשעי	plural of פשע	0.0902369	101395
7	שר	ruler of, head of, minister of, prince of	0.0948236	527
8	חרב	sword of	0.0965665	59124
9	דמי	+ plural of דם	0.128857	105261
10	ראשי	plural of ראש	0.147151	70390
11	שפת	language of	0.17324	34777
12	אשמת	guilt of	0.179418	271934
13	זמרת	variation of זמר	0.183104	116515
14	עם	nation of	0.185365	26776
15	רכב	+ chariot of	0.189169	52158
16	אופי	character of	0.190546	390890
17	אבי	father of, progenitor of	0.201887	250670
18	אשת	+ woman of	0.21395	162759
19	חקי	plural of חק	0.215311	98882
20	גדוד	battalion of	0.230883	68510
21	בת	+ daughter of, female native of	0.232916	282390
22	גרי	plural of גר	0.2369	686918
23	כתיב	orthography of	0.243217	18956
24	נמוס	+ manners of, customs of, laws of	0.246203	216174
25	ספר	book of	0.263237	102686
26	אלה	god of, celestial power of	0.267271	76311
27	רקוד	dance of	0.271113	316725
28	שדות	+ plural of שדה	0.287646	542533
29	כהני	plural of כהן	0.294472	509971
30	מכתב	letter of, writing of	0.294774	220622
31	מנהג	manner of, custom of	0.295612	312939
32	מס	tax of, levy of	0.300691	17306
33	אוכל	food of	0.325132	80178
34	שמות	+ plural of שם	0.327151	780927
35	אנשי	people of, men of	0.330403	80744

The prefixes ordered by increasing P_2 score on WGP.

WGP Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
36	מאכל	food of, diet of	0.338634	77735
37	חילי	plural of חיל	0.340745	85108
38	איי	isles of, see Genesis 10-5	0.35262	61694
39	מלאך	angel of	0.356812	159112
40	חק	law of	0.358252	167084
41	מקום	place of	0.364509	228413
42	דבור	speech of	0.369673	566568
43	כתב	script of	0.388551	254356
44	כח	force of, vigour of	0.391945	131456
45	בגדי	plural of בגד	0.39259	796517
46	אשם	guilt of	0.406398	216173
47	מושב	place of, seat of	0.406959	324927
48	נסי	plural of נס	0.411148	420959
49	זמר	song of, melody of	0.41807	481418
50	אלפי	plural of אלף	0.42367	72503
51	ערי	towns of	0.425085	511566
52	סמל	symbol of, emblem of	0.432327	339751
53	שם	+ name of, fame of	0.432522	675421
54	גזע	race of	0.4327	421233
55	אכל	food of	0.43398	371027
56	גבול	border of	0.43809	187492
57	דם	+ blood of	0.441202	464535
58	כספי	plural of כסף	0.442749	342254
59	דיני	plural of דין	0.449257	351317
60	אחזת	+ estate of, property of	0.461236	183760
61	ראש	head of, leader of	0.46141	57030
62	אב	father of, progenitor of	0.474792	206199
63	אומת	nation of	0.475212	557983
64	דת	religion of, law of	0.481659	286052
65	נשות	+ plural of אשת	0.49032	355460
66	טבע	nature of, character of	0.496889	444142
67	זהב	gold of	0.498803	930535
68	מלכת	+ queen of	0.547765	620533
69	שרש	root of	0.55207	410107
70	צבא	army of, troops of	0.55566	744524
71	לשן	dialect of	0.571554	232297
72	ארץ	country of	0.576227	283986
73	דגל	flag of	0.57898	213805
74	אל	god of	0.579054	316359
75	זרע	seed of, offspring of	0.580861	316652
76	כוח	alternative spelling of כח	0.609068	828166
77	מראה	appearance of, view of	0.609789	962044
78	גוי	nation of	0.610946	458369
79	לאם	nation of	0.614703	331320
80	אליל	idol of, false god of	0.63256	487622
81	עון	sin of	0.643686	522374
82	אילי	plural of איל	0.653476	227578

The prefixes ordered by increasing P_2 score on WGP.

WGP Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
83	פשע	crime of	0.656973	973643
84	דרכי	plural of דרך	0.660326	730895
85	בית	+ house of, family of, dynasty of	0.664562	287905
86	כסף	money of	0.671525	787673
87	מסי	plural of מס	0.679375	831358
88	אלוף	head of, leader of	0.679392	493195
89	סמלי	plural of סמל	0.680779	491239
90	עוון	alternative spelling of עון	0.697615	545941
91	מטבע	coin of	0.703522	575924
92	אדמת	land of	0.710502	686037
93	בן	+ son of, native of	0.715456	884632
94	חיל	army of, corps of	0.719289	133636
95	חטא	sin of	0.719531	547564
96	דין	justice of, law of	0.724938	780354
97	כהן	priest of	0.726514	267711
98	בירת	capital of	0.730573	744892
99	בני	plural of בן	0.730969	510473
100		the empty prefix	0.747399	627426
101	ע"ז של	idolatry of	0.767362	757358
102	דתי	laws of	0.774299	466636
103	דרך	way of	0.77883	442505
104	איש	person of, man of	0.783905	411107
105	מלך	king of	0.805124	871241
106	שרי	plural of שר	0.820516	281799
107	זמן	time of, period of, fate of	0.823103	908041
108	אלף	thousand of, family of, part of the tribe of	0.824563	816501
109	מעות	money of	0.83183	427909
110	בנות	+ plural of בת	0.83872	859438
111	ממשל	rule of, government of	0.852308	535770
112	זכות	right of, privilege of	0.855516	740668
113	לבוש	dress of, clothing of	0.856127	856709
114	לשון	dialect of	0.859666	880893
115	בתי	+ plural of בית	0.886635	826611
116	חטאי	plural of חטא	0.893336	889203
117	אי	isle of, see Genesis 10-5	0.895175	987036
118	רוח	spirit of	0.900598	952706
119	נשיא	president of, head of	0.900725	991585
120	שירי	plural of שער	0.908971	964625
121	דגלי	plural of דגל	0.91158	850378
122	גבור	hero of	0.913064	772745
123	גבר	man of, hero of	0.914002	886605
124	חבל	region of, district of	0.92879	883412
125	כחות	plural of כח	0.93096	987894
126	שדה	+ field of	0.933808	952237
127	גברי	plural of גבר	0.947163	770513
128	אמת	+ nation of	0.949779	882582
129	תושב	resident of	0.962662	949614

The prefixes ordered by increasing P_2 score on WGP.

WGP Rank	Prefix	Meaning	P_2	Permutation rank out of 10^6
130	בגד	cloth of, dress of	0.977759	980459
131	גר	convert of	0.985814	993763
132	איל	lord of, magnate of	0.991214	961856
133	סוף	end of, conclusion of	0.991296	872877
134	שיר	song of, hymn of, music of	0.994582	992970
135	נס	flag of, banner of, miracle of	0.996048	997088
136	מזל	luck, destiny, constellation of	0.997267	959411

The WRR Sample of Nations: Table 6

Alternative choices for the plural form column.

	X	WRR	WRR without אשורים	WRR with אשורם	WRR without לודים	Common Sense Extension	Explicit Plural	In Chapter	Keep ה	All With ה
1	גמר					גמרים				
2	מגוג					מגוגים				
3	מדי	מדים	מדים	מדים	מדים	מדים			מדים	המדים
4	יון	יונים	יונים	יונים	יונים	יונים	יונים		יונים	היונים
5	תבל					תבלים				
6	משך					משכים				
7	תירס					תירסים				
8	אשכנז					אשכנזים				
9	ריפת					ריפתים				
10	תגרמה									
11	אלישה									
12	תרשיש					תרשישים				
13	כתים	כתים	כתים	כתים	כתים	כתים	כתים	כתים	כתים	הכתים
14	דדנים	דדנים	דדנים	דדנים	דדנים	דדנים	דדנים	דדנים	דדנים	הדדנים
15	כוש	כושים	כושים	כושים	כושים	כושים	כושים		כושים	הכושים
16	מצרים	מצרים	מצרים	מצרים	מצרים	מצרים	מצרים		מצרים	המצרים
17	פוט					פוטים				
18	כנען	כנענים	כנענים	כנענים	כנענים	כנענים	כנענים		כנענים	הכנענים
19	סבא	סבאים	סבאים	סבאים	סבאים	סבאים	סבאים		סבאים	הסבאים
20	חוילה									
21	סבתה									
22	רעמה									
23	סבתכא					סבתכאים				
24	שבא	שבאים	שבאים	שבאים	שבאים	שבאים	שבאים		שבאים	השבאים
25	דדן	דדנים	דדנים	דדנים	דדנים	דדנים	דדנים		דדנים	הדדנים
26	לודים	לודים	לודים	לודים	לודים	לודים	לודים	לודים	לודים	הלודים
27	ענמים	ענמים	ענמים	ענמים	ענמים	ענמים	ענמים	ענמים	ענמים	הענמים
28	להבים	להבים	להבים	להבים	להבים	להבים	להבים	להבים	להבים	הלהבים
29	נפתחים	נפתחים	נפתחים	נפתחים	נפתחים	נפתחים	נפתחים	נפתחים	נפתחים	הנפתחים
30	פתרסים	פתרסים	פתרסים	פתרסים	פתרסים	פתרסים	פתרסים	פתרסים	פתרסים	הפתרסים
31	כסלחים	כסלחים	כסלחים	כסלחים	כסלחים	כסלחים	כסלחים	כסלחים	כסלחים	הכסלחים
32	פלשתים	פלשתים	פלשתים	פלשתים	פלשתים	פלשתים	פלשתים	פלשתים	פלשתים	הפלשתים
33	כפתרים	כפתרים	כפתרים	כפתרים	כפתרים	כפתרים	כפתרים	כפתרים	כפתרים	הכפתרים
34	צידן	צידנים	צידנים	צידנים	צידנים	צידנים	צידנים		צידנים	הצידנים
35	חת	חתים	חתים	חתים	חתים	חתים	חתים		חתים	החתים
36	היבוסי	יבוסים	יבוסים	יבוסים	יבוסים	יבוסים		יבוסים	היבוסים	היבוסים
37	האמרי	אמרים	אמרים	אמרים	אמרים	אמרים		אמרים	האמרים	האמרים
38	הגרגשי	גרגשים	גרגשים	גרגשים	גרגשים	גרגשים		גרגשים	הגרגשים	הגרגשים
39	החוי	חוים	חוים	חוים	חוים	חוים		חוים	החוים	החוים
40	הערקי	ערקים	ערקים	ערקים	ערקים	ערקים		ערקים	הערקים	הערקים
41	הסיני	סינים	סינים	סינים	סינים	סינים	סינים	סינים	הסינים	הסינים
42	הארודי	ארודים	ארודים	ארודים	ארודים	ארודים		ארודים	הארודים	הארודים

Alternative choices for the plural form column.

	X	WRR	WRR without אשורים	WRR with אשורם	WRR without לודים	Common Sense Extension	Explicit Plural	In Chapter	Keep ה	All With ה
43	הצמרי	צמרים	צמרים	צמרים	צמרים	צמרים		צמרים	הצמרים	הצמרים
44	החמתי	חמתים	חמתים	חמתים	חמתים	חמתים		חמתים	החמתים	החמתים
45	עילם					עילמים				
46	אשור	אשורים		אשורם	אשורים	אשורים			אשורים	האשורים
47	ארפכשד					ארפכשדים				
48	לוד	לודים	לודים	לודים		לודים	לודים		לודים	הלודים
49	ארם	ארמים	ארמים	ארמים	ארמים	ארמים	ארמים		ארמים	הארמים
50	עוץ					עוצים				
51	חול					חולים				
52	גתר					גתרים				
53	מש					משים				
54	שלח					שלחים				
55	עבר	עברים	עברים	עברים	עברים	עברים	עברים		עברים	העברים
56	פלג					פלגים				
57	יקטן					יקטנים				
58	אלמודד					אלמודדים				
59	שלף					שלפים				
60	חצרמות					חצרמותים				
61	ירח					ירחים				
62	הדורם					הדורמים				
63	אוזל					אוזלים				
64	דקלה									
65	עובל					עובלים				
66	אבימאל					אבימאלים				
67	אופר					אופרים				
68	יובב					יובבים				

The WRR Sample of Nations: Table 9

Comment: To save space, we only display those nations for which the fields in the table are non-empty.

Variations on the spelling of names taken from Ginsburger's Targum Yonatan.

	X	WRR - Modified Ginsburger	WRR with Rieder's Hebrew (literal)	WRR with Rieder's Hebrew (modified)	WRR, reducing וו	WRR with ים	WRR with אי	WRR with ה
1	גמר	אפריקי	אפריקה	אפריקה	אפריקי	אפריקי	אפריקי	אפריקי
2	מגוג	גרמניא	גרמניה	גרמניה	גרמניא	גרמניא	גרמניא	גרמניא
3	מדי	המדי	מדיה	מדיה	המדי	המדי	המדי	המדי
4	יון	מקדוניא	מקדוניה	מקדוניה	מקדוניא	מקדוניא	מקדוניא	מקדוניא
5	תבל	יתיניא	יתיניה	יתיניה	יתיניא	יתיניא	יתיניא	יתיניא
6	משך	מוסיא	אוסיה	אוסיה	מוסיא	מוסיא	מוסיא	מוסיא
7	תירס	תרקי	תרקיה	תרקיה	תרקי	תרקי	תרקי	תרקי
8	אשכנז	אסיא	אסיא	אסיא	אסיא	אסיא	אסיא	אסיא
9	ריפת	פרכוי	פרכוי	פרכוי	פרכוי	פרכוי	פרכוי	פרכוי
10	תגרמה	ברבריאה	ברבריאה	ברבריאה	ברבריאה	ברבריאה	ברבריאה	ברבריאה
11	אלישה	אלס	אלס	אלס	אלס	אלס	אלס	אלס
12	תרשיש	טרסס	טרסס	טרסס	טרסס	טרסס	טרסס	טרסס
13	כתים	אכיא	אכזייה	אכזיה	אכיא	אכיא	אכיא	אכיא
14	דדנים	דרדניא	דורדניה	דורדניה	דרדניא	דרדניא	דרדניא	דרדניא
15	כוש	ערב	ערב	ערב	ערב	ערב	ערב	ערב
17	פוט	אליחרק	אליחרק	אליחרק	אליחרק	אליחרק	אליחרק	אליחרק
19	סבא	סיניראי	סיניראי	סיניראי	סיניראי	סיניירים	סיניראי	סיניראי
20	חוילה	הינדיקי	הנדקי	הנדקי	הינדיקי	הינדיקי	הינדיקי	הינדיקי
21	סבתה	סמראי	סמראי	סמראי	סמראי	סמרים	סמראי	סמראי
22	רעמה	לוב	לובאי	לובאי	לוב	לוב	לוב	לוב
23	סבתכא	זינגאי	זינגאי	זינגאי	זינגאי	זינגים	זינגאי	זינגאי
24	שבא	זמרגד זמרגר	זמרגד	זמרגד	זמרגד זמרגר	זמרגד זמרגר	זמרגד זמרגר	זמרגד זמרגר
25	דדן	מזג	מזג	מזג	מזג	מזג	מזג	מזג
26	לודים	ניוטים	גיוואטים	גיוואטים גיווטים	ניוטים	ניוטים	ניוטאי	הניוטים
27	ענמים	מרסיוטים מריוטים	מרטיוטים	מרטיוטים	מרסיוטים מריוטים	מרסיוטים מריוטים	מרסיוטאי מריוטאי	המרסיוטים המריוטים
28	להבים	ליווקים	ליווקים	ליווקים	ליוקים	ליווקים	ליווקאי	הליווקים
29	נפתחים	פנטסכינים	פנטסכינים	פנטסכינים	פנטסכינים	פנטסכינים	פנטסכינאי	הפנטסכינים
30	פתרסים	נסיוטים	נסיוטים	נסיוטים	נסיוטים	נסיוטים	נסיוטאי	הנסיוטים
31	כסלחים	פנטפוליטי	פנטפוליטים	פנטפוליטים	פנטפוליטי	פנטפוליטי	פנטפוליטי	פנטפוליטי
33	כפתרים	קפודקים	קפודקים	קפודקים	קפודקים	קפודקים	קפודקאי	הקפודקים
41	הסיני	אנתוסים ארתוסים כפרוסים	אנתוסים כפרוסים	אנתוסים כפרוסים	אנתוסים ארתוסים כפרוסים	אנתוסים ארתוסים כפרוסים	אנתוסאי ארתוסאי כפרוסאי	האנתוסים הארתוסים הכפרוסים
42	הארודי	רודסים	רודסים	רודסים	רודסים	רודסים	רודסאי	הרודסים
43	הצמרי	חומצים	חומצים	חומצים	חומצים	חומצים	חומצאי	החומצים
44	החמתי	אנטיוכים	אנטכויים	אנטכוים	אנטיוכים	אנטיוכים	אנטיוכאי	האנטיוכים

The WRR Sample of Nations: Table 11

Comment: To save space, we only display those nations for which the fields in the table are non-empty, and to ease reading, we placed an asterisk (*) next to the WRR choice in each row.

Alternative choices for the identifications of the nations and countries.

#	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
X	*WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
1	אפריקי	אפריקי	אפריקי	אפריקי	גרממייה	גרממיא	אפריקה
גמר	*אפריקי	אפריקי	אפריקי	אפריקי	גרממיה	גרממיא	אפריקה
2	גרמניא	גרמניא	גרמניא	גרמניא	גותייא	קנדיא	גירמניא
מגוג	*גרמניא	גרמניא	גרמניא	גרמניא	גותיא	קנדיא	גירמניא
3	המדיי	המדיי	המדיי			מקדוניא	
מדי	*המדי	המדי	המדי			מקדוניא	
4	מקדיניא	מקדינייא	מקדינייא	מקדוניא	אווסוס		מקדוניא
יון	*מקדוניא	מקדוניא	מקדוניא	מקדוניא	אווסוס		מקדוניא
5	יתיניא	יתיניא	יתיניא	יתניא	וותנייה	אונייקי	איסניא
תבל	*יתיניא	יתיניא	יתיניא	יתניא	ותניה	אוניקי	איסניא
6	מוסיא	אוסיא	אוסיא	אנסייא	מוסייא	מוסיא	תוניה
משך	*מוסיא	אוסיא	אוסיא	אנסיא	מוסיא	מוסיא	תוניה
7	תרקי	תרקי	תרקי	תרקי	פרס תרקא	פרס תרייקי	פרס תורקי
תירס	*תרקי	תרקי	תרקי	תרקי	פרס תרקא	פרס תריקי	פרס תורקי
8	אסיא	אסיא		אסייא	אסייא		אסייא
אשכנז	*אסיא	אסיא		אסיא	אסיא		אסיא
9	פרכוי	פרבוי		פרכון	הדיית		הדייף

Alternative choices for the identifications of the nations and countries.

#	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
X	*WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
ריפת	*פרכוי	פרבוי		פרכון	הדית		הדיף
10	ברבריאה	ברבריאה	תורגמא	ברבריאה	גרמניקיה		גרמנייא גירמניקייא
תגרמה	*ברבריאה	ברבריאה	תורגמא	ברבריאה	גרמניקיה		גרמניא גירמניקיא
11	אלס	אלס	אלס	אלסטרסוס	אלסטרסס		אלסיטרוס
אלישה	*אלס	אלס	אלס	אלסטרסוס	אלסטרסס		אלסיטרוס
12	טרסס	טרסס	טרסס				
תרשיש	*טרסס	טרסס	טרסס				
13	אכייא	אכזיא	אכזיא		אביה		איטלייא
כתים	*אכיא	אכזיא	אכזיא		אביה		איטליא
14	דרדניא	דרדניא	דרדניא	דודניא	דרדנייה		דרדנייא רודנים
דדנים	*דרדניא	דרדניא	דרדניא	דודניא	דרדניה		דרדניא רודנים
15	ערביא	ערביא	ערביא				
כוש	*ערב	ערב	ערב				
17	אליחרק	אליחרק	אליחרק				
פוט	*אליחרק	אליחרק	אליחרק				
19	סיניראי	סיניראי	סיניראי				
סבא	*סיניראי	סיניראי	סיניראי				
20	הינדיקי	הינדיקי	הינדיקי				
חוילה	*הינדיקי	הינדיקי	הינדיקי				

Alternative choices for the identifications of the nations and countries.

#	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
X	*WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
21	סמראי	סמראי	סמראי			סקיסתן	
סבתה	*סמראי	סמראי	סמראי			סקיסתן	
22	לובאי מווריטינוס	לובאי מווריטינוס	לובאי מווריטינוס				
רעמה	*לוב	לוב	לוב				
23	זינגאי	זינגאי	זינגאי			סקיסתן	
סבתכא	*זינגאי	זינגאי	זינגאי			סקיסתן	
24	זמרגד זמרגר	זמרגד	זמרגד				
שבא	*זמרגד *זמרגר	זמרגד	זמרגד				
25	מזג	מזג	מזג				
דדן	*מזג	מזג	מזג				
26	ניווטים	ניווטים	ניווטים				
לודים	*ניווטים	ניווטים	ניווטים				
27	מרסיוטים מריוטים	מריוטים	מריוטים	מריוטים			
ענמים	*מרסיוטים *מריוטים	מריוטים	מריוטים	מריוטים			
28	ליווקים	ליווקים	ליווקים	פנטפולטים			
להבים	*ליווקים	ליווקים	ליווקים	פנטפולטים			
29	פנטסכינים	פנטסכינים	פנטסכינים	לוסטים			
נפתחים	*פנטסכינים	פנטסכינים	פנטסכינים	לוסטים			

Alternative choices for the identifications of the nations and countries.

#	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
X	*WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
30	נסיוטים	נסיוטים	נסיוטים	פילוסים			
פתרסים	*נסיוטים	נסיוטים	נסיוטים	פילוסים			
31	פנטפוליטיה	פנטפוליטי	פנטפולוטי	פנטסכנים			
כסלחים	*פנטפוליטי	פנטפוליטי	פנטפולוטי	פנטסכנים			
33	קפודקים	קפודקים קפודים	קפודים	קפודקים			
כפתרים	*קפודקים	קפודקים קפודים	קפודים	קפודקים			
39				טריפולים			חלדין
החוי				טריפולים			חלדין
40							ארקא ארקא-דליבנן
הערקי							ארקא ארקא-דליבנן
41	אנתוסים ארתוסים כפרוסים	אנתוסים	אנתוסים	כפרוסים			ארתוסייה
הסיני	*אנתוסים *ארתוסים *כפרוסים	אנתוסים	אנתוסים	כפרוסים			ארתוסיה
42	רודסים	לודסים לוטסים	לוטסים	אנטרידנים	רודוס		ארווד
הארודי	*רודסים	לודסים לוטסים	לוטסים	אנטרידנים	רודוס		ארוד
43	חומצים	חומצים	חומצים	חמצים	חמץ		חמץ
הצמרי	*חומצים	חומצים	חומצים	חמצים	חמץ		חמץ

Alternative choices for the identifications of the nations and countries.

#	Ginsburger	Rieder	Clarke	Targum Yerushalmi	Talmud Yerushalmi	Talmud Bavli	Midrash Rabbah
X	*WRR - Modified Ginsburger	Modified Rieder	Modified Clarke	Modified Targum Yerushalmi	Modified Talmud Yerushalmi	Modified Talmud Bavli	Modified Midrash Rabbah
44	אנטיוכים	אנטכוים אנטיוכי	אנטיוכי	אנטוכיא	חמת		פיפני
החמתי	*אנטיוכים	אנטכוים אנטיוכי	אנטיוכי	אנטוכיא	חמת		פיפני
46		אתור	אתור			סילק	
אשור						סילק	

The WRR Sample of Nations: Table 14

Comment: To save space, we only display those nations for which the fields in the table are non-empty.

The identifications appearing in Kasher's "Torah Shelemah".

	X	WRR - Modified Ginsburger	Kasher	Modified Kasher	Modified Kasher, reducing ו
1	גמר	אפריקי	אפריקי גרממיא גרממייה גרמנייא	אפריקי גרממיא גרממיה גרמניא	אפריקי גרממיא גרממיה גרמניא
2	מגוג	גרמניא	גותייא גינתא גיתייה גרמניא גרמנייא קנדיא שיטים	גותיא גינתא גיתיה גרמניא קנדיא שיטים	גותיא גינתא גיתיה גרמניא קנדיא שיטים
3	מדי	המדי	המדאי המדיי המדן מקדוניא	המדאי המדי המדן מקדוניא	המדאי המדי המדן מקדוניא
4	יון	מקדוניא	אווסוס אוכיסוס מוקדון מוקדוניא מקדוניא מקדיניא	אווסוס אוכיסוס מוקדון מוקדוניא מקדוניא	אוכיסוס אוסוס מוקדון מוקדוניא מקדוניא
5	תבל	יתיניא	אונייקי איסוניא ותנייה טובר יתיניא יתינייה יתניא	אוניקי איסוניא ותניה טובר יתיניא יתיניה יתניא	אוניקי איסוניא ותניה טובר יתיניא יתיניה יתניא
6	משך	מוסיא	אוסיא אנסייא מוסיא מוסייא מוסייה מושכי סוסיא	אוסיא אנסיא מוסיא מוסיה מושכי סוסיא	אוסיא אנסיא מוסיא מוסיה מושכי סוסיא
7	תירס	תרקי	פרס תראסיה תרייקי תרקא תרקי	פרס תראסיה תריקי תרקא תרקי	פרס תראסיה תריקי תרקא תרקי
8	אשכנז	אסיא	אסיא אסייא	אסיא	אסיא
9	ריפת	פרכוי	דיפת הדייף הדיית חדיב פרכווי פרכוי פרכון פרסווי	דיפת הדיף הדית חדיב פרכווי פרכוי פרכון פרסווי	דיפת הדיף הדית חדיב פרכוי פרכון פרסוי
10	תגרמה	ברבריאה	ברבריאה ברברייא גירמניקייא גרמונקיא גרמניא גרמנייא גרמניקיה	ברבריא ברבריאה גירמניקיא גרמונקיא גרמניא גרמניקיה	ברבריא ברבריאה גירמניקיא גרמונקיא גרמניא גרמניקיה
11	אלישה	אלס	אלס אלסו אלסטרסוס אלסטרסס	אלס אלסו אלסטרסוס אלסטרסס	אלס אלסו אלסטרסוס אלסטרסס
12	תרשיש	טרסס	טרסוס טרסס תרסס	טרסוס טרסס תרסס	טרסוס טרסס תרסס
13	כתים	אכיא	אבזווייה אביה איטליא איטליון אכזייא אכייא	אבזוויה אביה איטליא איטליון אכזיא אכיא	אבזויה אביה איטליא איטליון אכזיא אכיא
14	דדנים	דרדניא	דדניא דודניא דורדניא דרדנייא דרדנייה רודנים	דדניא דודניא דורדניא דרדניא דרדניה רודנים	דדניא דודניא דורדניא דרדניא דרדניה רודנים
15	כוש	ערב	ערביא	ערב	ערב
17	פוט	אליחרק	אליחרוק אליחרק	אליחרוק אליחרק	אליחרוק אליחרק
19	סבא	סיניראי	סיניראי	סיניראי	סיניראי
20	חוילה	הינדיקי	הינדיקי הנדקי	הינדיקי הנדקי	הינדיקי הנדקי
21	סבתה	סמראי	סמדאי סמראי סקיסתן	סמדאי סמראי סקיסתן	סמדאי סמראי סקיסתן
22	רעמה	לוב	לובאי מווריטינוס מוריטינוס	לוב	לוב
23	סבתכא	זינגאי	זינגאי סקיסתן	זינגאי סקיסתן	זינגאי סקיסתן
24	שבא	זמרגד זמרגר	זמדגד זמרגד זמרגר	זמדגד זמרגד זמרגר	זמדגד זמרגד זמרגר
25	דדן	מזג	מזג	מזג	מזג
26	לודים	ניוטים	גיוטים ניוטים	גיוטים ניוטים	גיוטים ניוטים
27	ענמים	מרסיוטים מריוטים	מרטיוטים מריוטים מרסיוטים	מרטיוטים מריוטים מרסיוטים	מרטיוטים מריוטים מרסיוטים
28	להבים	ליווקים	ליווקים פנטפוליטס	ליווקים פנטפוליטס	ליוקים פנטפוליטס
29	נפתחים	פנטסכינים	לוסטים פנטסכינים	לוסטים פנטסכינים	לוסטים פנטסכינים

The identifications appearing in Kasher's "Torah Shelemah".

	X	WRR - Modified Ginsburger	Kasher	Modified Kasher	Modified Kasher, reducing ו
30	פתרסים	נסיוטים	נסיוטים פילוסים	נסיוטים פילוסים	נסיוטים פילוסים
31	כסלחים	פנטפוליטי	פנטסכנים פנטפולוטי פנטפוליטיה	פנטסכנים פנטפולוטי פנטפוליטי	פנטסכנים פנטפולוטי פנטפוליטי
33	כפתרים	קפודקים	קפודקים קפוטקים	קפודקים קפוטקים	קפודקים קפוטקים
39	החוי		חלדון טריפולים	חלדון טריפולים	חלדון טריפולים
40	הערקי		ארקא ארקא-דליבנן	ארקא ארקא-דליבנן	ארקא ארקא-דליבנן
41	הסיני	אנתוסים ארתוסים כפרוסים	אנטוסים אנתוסים ארתוסייה ארתוסים כפרוסים	אנטוסים אנתוסים ארתוסיה ארתוסים כפרוסים	אנטוסים אנתוסים ארתוסיה ארתוסים כפרוסים
42	הארודי	רודסים	אנטרידנים ארווד לוטסים רודוס רודסים	אנטרידנים ארווד לוטסים רודוס רודסים	אנטרידנים ארוד לוטסים רודוס רודסים
43	הצמרי	חומצים	חומצים חמץ חמצים	חומצים חמץ חמצים	חומצים חמץ חמצים
44	החמתי	אנטיוכים	אנטוכיא אנטיוכים אנטכוים פיפניה	אנטוכיא אנטיוכים אנטכוים פיפניה	אנטוכיא אנטיוכים אנטכוים פיפניה
46	אשור		אשוראה אתור אתוראה אתורייה נח סילק	אשוראה אשוריה נח סילק	אשוראה אשוריה נח סילק

© Copyright (1998) Dror Bar-Natan drorbn@math.huji.ac.il, Brendan McKay bdm@cs.anu.edu.au, and Shlomo Sternberg shlomo@math.harvard.edu.