

The vocabulary of the Hebrew Bible

Avner Ramu



In February 2011 the Israeli Postal service issued this stamp which portrays the sources of the Modern Hebrew language. The Hebrew Bible is acknowledged as the most ancient source of this language.



***Sefer Hashorashim* (Book of roots) written by David Kimhi (1160-1235) printed in Venice in 1546.**

This Book is a part of a larger book named **מכלול** (*michlol*) written by Rabbi David Kimhi. It was one of the earliest attempts to compose a dictionary of biblical Hebrew.

The number of words in the Hebrew Bible is 305,441. Of these, 4,828 are Aramaic words while the rest are of the Hebrew language.

Most biblical words appear more than once and they can be grouped as belonging to certain distinct words. Andersen and Forbes indicated that several kinds of decisions have to be made when classifying and ordering this word stock. These decisions require the identification of all the words that are judged to be sufficiently similar in some lexical sense to be considered “the same”, and the choice of a generalized form (*lemma*) to represent the group (1).

The words are generally divided to lemmas according to the similarity in their consonantal form and sameness of meaning. Yet there are words that have similar consonantal form but appear to have several different meanings. One has to decide in such cases whether their meanings are different enough for separating them to several lemmas (e.g. the word **את** - “*eth*” which 742 times means: “with”, and is an object marker in another 11,123 appearances). Obviously many of these judgments are arbitrary and subjective (1). While Andersen and Forbes count in the Hebrew Bible 9,980 lemmas, Jenni and Westermann determined that in the Mandelkern Concordance there are only 8,250 distinct words (2).

The biblical word-type distribution is similar to that found in most languages, where about half of the vocabulary is made of noun lemmas, about one quarter is adjectives, and one seventh is the verbs. It was indicated by Rabin that the ~8,000 lemmas vocabulary of the Hebrew Bible probably represents less than one third of the wordlist of the Hebrew language spoken in biblical times (3). For example, archeological studies revealed that cats became an integral element of the human household as early as the fifth millennium BC (or even earlier). It is also known that since the 3rd millennium BC most of the wheat and barley produced in Egypt and Mesopotamia was used for the production of beer. Yet there are no known words in the Hebrew Bible that refer to this animal and to this beverage. Rabin suggested that as the Bible deals with a restricted number of themes, its authors had no reason to use most of the spoken Hebrew words. However, it is also possible that the vocabulary of Hebrew Bible contains many other independent lemmas that have not been recognized as such.

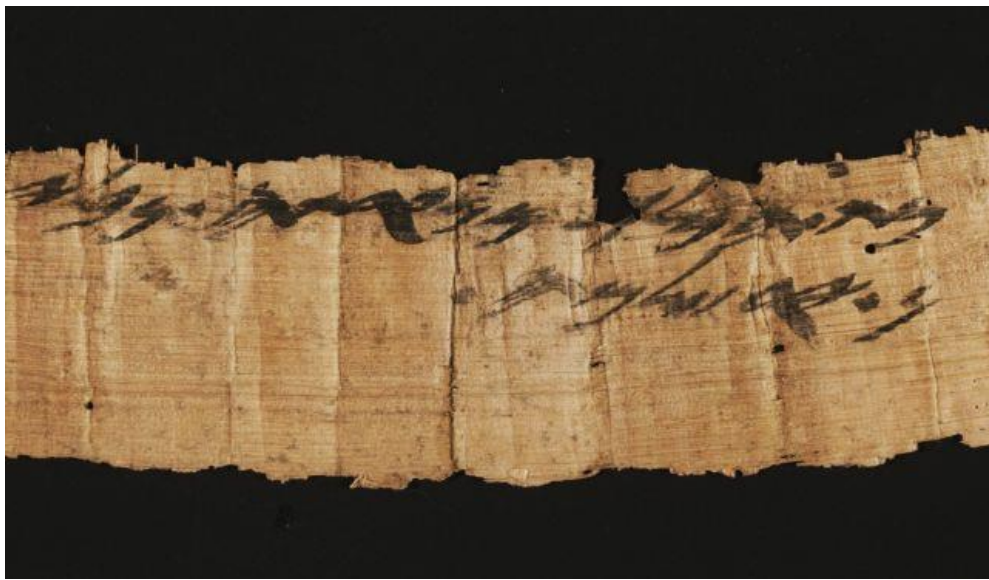
The largest lexical class of the Biblical Hebrew language is nouns which according to Andersen and Forbes account for 153,511 biblical words. They include ~3,678 common noun lemmas (2,159 masculine; 1,138 feminine; and 387 gender-free lemmas) which account for 107,578 noun words. However, most scholars think that the Hebrew language does not have gender-free noun (or pronoun) lemmas (e.g. unlike the English language, there is no “it” in Hebrew). Therefore we may assume that Andersen and Forbes defined as “gender-free” lemmas, those lemmas that are treated sometimes as masculine objects while on other occasion as feminine objects. However, such changes may reflect change in view in different historic periods, unevenness in the degree of familiarity with the Hebrew grammar among the biblical scribes, or simply spelling errors.

According to Andersen and Forbes, 12 common noun lemmas, each appearing more than 1,001 times: **עיר, אב, דיבר, יד, עס, בית, פנים, איש, יום, מלך, ארץ, בן** (son, king, land, day, man, face, house, people, hand, word, father, and city, respectively), account for 24.4% of the common noun biblical words. Another 19 common noun lemmas, each appearing between 501 and 1,000 times: **הר, גוי, בת, מים, ראש, לב, אח, דרך, כהן, נפש, אשה, עבד, שם, עין, שנה** (year, eye, name, servant, woman, soul, priest, way, brother, heart, head, water, daughter, people, mountain, human being, voice, mouth, and life, respectively) account for 11.7% of the common noun words. In addition, 45 other common noun lemmas, each appearing between 251 and 500 times, account for 14.2% of the common noun words. There are also 70 common noun lemmas, each appearing between 126 and 250 times, which account for 11.6% of the common noun words.

Many lemmas such as: גדול (big), קטן (little), חכם (wise), ראשון (first), אחרון (last) were classified by Andersen and Forbes as “common nouns. Yet it seems that it is more accurate to assign them to other lexical classes (in particular: as adjectives and adverbs).

There are also 2,726 noun lemmas that are proper names, and they appear 38,806 times. They include 55 divine names that appear 10,587 times. The names: יהוה, אלוה, אלהים (YHWH, Eloha, Elohim) account for 90.5% of all the divine name words in the Bible.

There are also 1,581 human names which appear 18,919 times. The most numerous among them are: ישראל, דויד, יהודה, משה, שאול, לוי, יעקב, אהרון, שלמה, פרעה, אברהם, יהושע (Israel, David, Judah, Moses, Saul, Levi, Jacob, Aaron, Solomon, Pharaoh, Abraham, and Joshua, respectively).



An 8th century BC papyrus containing the word: ירשלמה – “to Jerusalem.”

In the Hebrew Bible there are 1090 geographical and ethnic names which appear 9,194 times. The most prevalent among them are: מצרים, ירושלים, פלשת, בבל, מואב, ארם, ציון, אשור, אדום, עמון, גלעד, שמרון (Egypt, Jerusalem, Philistia, Babylon, Moab, Aram-Syria, Zion, Asshur-Assyria, Ammon, and Gilead, respectively).

The Bible also contains 46 distinct numerical nouns which appear 7,127 times. Eight numbers which include: “one” to “five”, “ten”, “hundred”, and “thousand” account for 61% of all the numerical words of the Bible. Among the Hebrew numerical words, the number “eleven” is the oddest. It appears as אחד עשר (masculine, 3 times) or as עשתי עשר (masculine, 8 times), and as אחת עשרה (feminine, 10 times) or as עשתי עשרה (feminine, 10 times).

Pronouns are defined as “words that can replace nouns.” According to Andersen and Forbes there are 8 Demonstrative pronouns (2,651 appearances), 4 Relative pronouns (5,973

appearances), 13 Interrogative pronouns (2,315 appearances), and 13 Personal pronouns (which together with the Suffix pronouns appear 50,591 times). Seventeen of these pronouns: **זה, זאת, אלה, אשר, מה, מי, אני, אנכי, אתה, את, הוא, היא, אנחנו, אתם, אתן, הם, הן** (This (m), this (f), these, which, what, who, I (+ me), I, you (m), you (f), he, she, we (+ us), you (m), you (f), they (m + them), they (f + them), accordingly), account for 97.1% of all the appearances of the pronoun lemmas.

Another lexical class is verbs. There are 73,957 verb words which represent ~1,570 lemmas.



A reprint (Berlin, 1896) of “The Book of roots” by Jonah ibn Janah (c. 990 - c. 1050). It is a critique and expansion of the work of Juda ben David Hayyuj, the founder of systematic Hebrew grammar studies, on the Hebrew tri-literal root.

Verb words are identified by roots, but it is not always easy to determine the verb’s genuine root. Since the Middle-Ages it is assumed that the roots of almost all the Hebrew verbs (and of many nouns, adjectives, and adverbs) are tri-literal. Yet it is currently thought that in biblical times there were also many bi-literal roots, that were “converted” in the middle ages to tri-literal roots by adding a third letter (mostly a vowel letter, and less often by doubling one of the bi-literal root letters).

Jenni and Westermann counted in the Mandelkern Concordance 71,510 verb words which they assigned to 1,565 distinct verbs. The most frequent (5,378 times) are those belonging to the root **אמר** (*amr*) - "say."

The number of verbs originating from 12 roots, each one appearing over 1,000 times: **אמר, היה, עשה, בא, נתן, הלך, ראה, שמע, דבר, ישב, יצא, שוב** (say, be, make, come, give, walk, see,

hear, speak, sit, go out, and return, respectively) account for 34.2% of all the verb words in the Bible. Verbs from other 12 roots: (**קח, דע, עלה, שלח, מת, אכל, קרא, נשא, קמ, שמ, עבר,** , respectively) appear between 501 and 1,000 times each, and account for 12.6% of all the verb words in the Bible. Verbs from other 19 roots appear between 251 and 500 times each, and account for 9.6% of all the verb words in the Bible. Verbs from other 56 roots which appear between 126 and 250 times each, account for 13.8% of all verb words in the Bible. The remaining 1,466 distinct verb roots (93.67% of all the verb lemmas), each with less the 126 appearances, account for the remaining 29.8% of all the verb words of the Bible (2).

The Hebrew verb system is a set of morphological distinctions or *binyanim* and the frequency of verb words among them is:

קל	<i>Qal</i>	49,180
נפעל	<i>Nipha'l</i>	4,140
פעל	<i>Pie'l</i>	6,450
פעל	<i>Pua'l</i>	460
הפעיל	<i>Hiphi'l</i>	9,370
הפעל	<i>Hopha'l</i>	400
התפעל	<i>Hithpae'l</i>	830
15 apparent other	<i>binyanim</i>	680

There are also few lemmas that were defined as “Quasi-verbal forms” and appear 2187 times. The words: **יש** - “exists”, **אין** - “not exist”, **הנה** - “behold”, and **עוד** - “still” account for 99.2% of the Quasi-verbal forms appearances.

Adverbs are defined as words that modify verbs. Andersen and Forbes defined 49 lemmas as “Adverbs” and assign to them 5,775 appearances. However, as indicated above, many of their “Common nouns” are actually adverb words. Furthermore, in Hebrew the attachment of the letter **ב** (b) as prefix to certain common nouns turns them to adverbs (and adjectives).

However, other words that contain the prefix letter **ב** (b) are considered as Prepositions, which are defined as words that link nouns or pronouns to other types of words. Other types of Prepositions are obtained by attaching the prefix letters **כ** (k), **ל** (l), or **מ** (m) to other words. Of the preposition words, there are four: **על** - “upon” (5,864 appearances), **אל** - “unto” (5,508 appearances), **עד** - “until” (1,216 appearances), and **עם** - “with” (1,117 appearances), which together account for 71.7% of all the appearances of the Bible’s preposition words. Another six words: **את** - “with”, **אחרי** - “after”, **מן** - “from”, **תחת** - “under”, **בין** - “between”, and **תוך** - “inside” account for another 19.2% of all the appearances of the Bible’s preposition words.

The remaining biblical words belong to lexical classes such as Conjunctions. Two words: **כי** - “because” (4,409 appearances), and **אם** - “if” (1,032 appearances) account for 83.1% of all the appearances of the biblical conjunction words.

The two most common Hebrew negative words are: **לא** (*lo*) - “no” or “not” or “don’t”, which appears 5,258 times, and its mirror image **אל** (*ul*) which is mentioned 735 times.

Finally, the words **את** (*eth*) which is an object marker (and has no equivalent in the English language) with 11,123 appearances, and the word **כל** (*kol*) - “all” which appears 5,516 times, are the 2nd and 8th most numerous words of the Hebrew Bible.

The distribution of lemmas according to the frequency of their appearance is as follows:

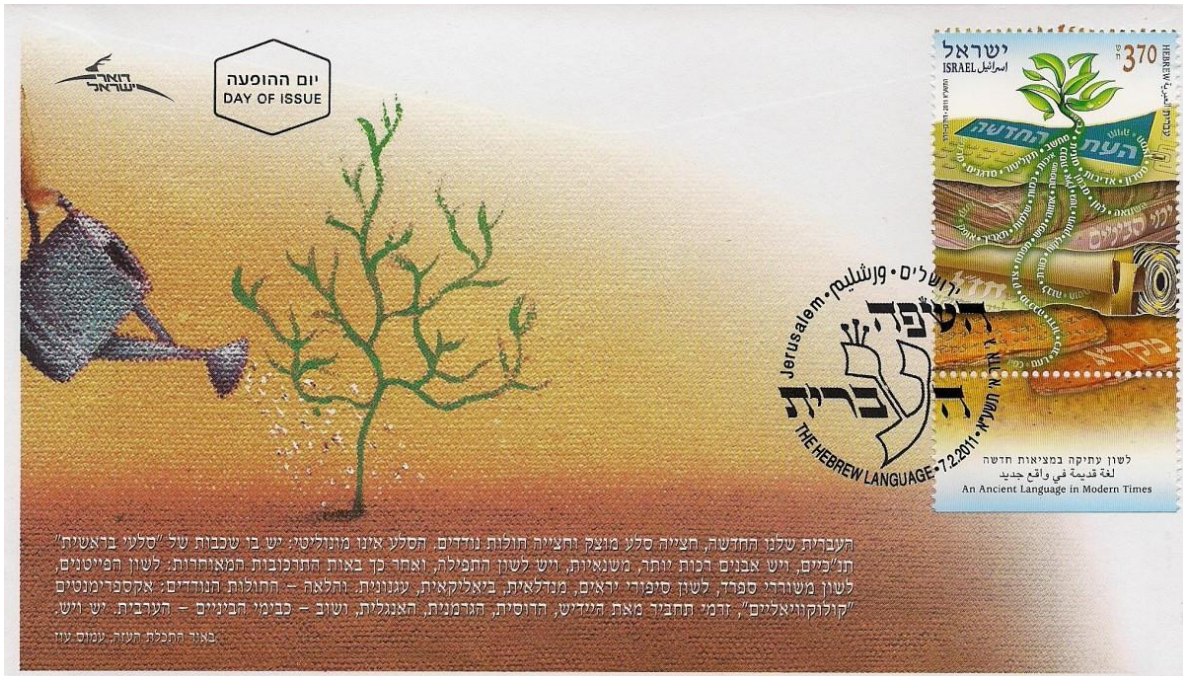
There are	2	lemmas	which	appear	>10,001	times,		combined	appearances	26,945
	10	“	“	“	between	5,001	and	10,000	times,	“
	10	“	“	“	between	2,501	and	5,000	“	“
	14	“	“	“	between	1,251	and	2,500	“	“
	45	“	“	“	between	626	and	1,250	“	“
	72	“	“	“	between	316	and	625	“	“
	119	“	“	“	between	158	and	315	“	“
	1393	“	“	“	between	20	and	157	“	“
	~8350	“	“	“	<20	times,			“	“
										~25,000

We may conclude that 2.73% of the biblical lemmas make 80.71% of all the words of the Hebrew Bible. On the other hand, 83.72% of the biblical lemmas make only 8.19% of the word stock of the Hebrew Bible. Obviously, our understanding of the exact meaning of a lemma is directly related to its appearance frequency in the Bible. Furthermore, the probability that a lemma is the result of scribal error is inversely related to its occurrence in the Bible.

There are about 1500 words that appear in the Hebrew Bible only once (**מילים יחידאיות** - Hapax legomena). Naftali Herz Tur-Sinai indicated that several of them are “ghost words” resulting from spelling errors (4), and Frederick Greenspahn claimed that there are no more than 322 absolute Hapax legomena (5). Several other words, although appear in the Hebrew Bible more than once, occur exclusively in single verses or in separate verses that appear to be copies of a single verse (6). The meaning of more than a few of such once-appearing words is not sure or known.

References:

1. F I Andersen and A D Forbes: The Vocabulary of the Old Testament. Editrice Pontificio Instituto Biblico. Rome 1989.
2. E Jenni and C Westermann. Theological Lexicon of the Old Testament. Hendrickson Publishers. Peabody, MA. Vol 3. Pp 1435-1447. 1997.
3. C. Rabin. A short history of the Hebrew language. Pub. Jewish Agency. Jerusalem. 1973.
4. N. H. Tur-Sinai. Halashon Vehasefer. 1954. In Hebrew.
5. F. E. Greenspahn. Hapax legomena in biblical Hebrew. Scholars Press. Chico, California. 1984. p. 33.
6. H. R. Cohen. Biblical Hapax Legomena in the light of Akkadian and Ugaritic. Scholars Press 1978.



The official Israeli Postal service “First day envelope” of the “Hebrew language” stamp. 2011.