Evidence From Computational Linguistics for the Concept of Biconsonantal Etymons in Hebrew

Avigail Stekel (avigailste@ariel.ac.il)
School of Computer science, Ariel University, Israel

Moshe Stekel (moshe.stekel@ariel.ac.il)
School of Computer science, Ariel University, Israel

Amos Azaria (amos.azaria@ariel.ac.il)
School of Computer science, Ariel University, Israel

Abstract

This paper explores the hypothesis of the historical evolution of Semitic morphology from biconsonantal (2C) etymons, to triconsonantal (3C) roots, which make up the majority of words in Biblical Hebrew as well as in other Semitic languages such as modern Hebrew, Arabic, etc. The rules for reducing the 3C roots to their 2C etymons are provided in detail. We use BHSA, a manually annotated corpus of the Hebrew Bible, and Word2Vec, a method for converting words to a vector representing their semantic meaning, to study the hypothesis of evolution from 2C etymons to 3C roots in biblical Hebrew. Namely, we show that words in Hebrew with different roots, that might have originated from the same 2C etymons form a denser cluster than random sets of words of the same size. These differences are statistically significant and strongly support the hypothesis of evolution from 2C etymons to 3C roots.

Keywords: Semitic morphology, Semitic semantics, etymon, biconsonantal, triconsonantal, word2vec, semantic similarity.

Introduction

Natural language and the ways of its evolution occupy many researchers, as they combine many other branches of science, such as history, anthropology, psychology, and even mathematics and computer science. Natural languages are ancient, and the origins of their development are shrouded in history; yet, they persist within us, and their complexity intrigues our desire for knowledge and understanding, sparking our imagination.

In this paper, we discuss the concept of etymon. This concept was first mentioned by (Gesenius, 1817) and denotes a biconsonant morphological particle. Over the past two centuries, there has been a lively debate among linguists about the viability of this concept. Below are some of the most important milestones of this discussion.

To better understand the concept of etymon from the linguistic perspective, it is worth referring to the works of Georges Bohas ((Bohas, 1977), (Bohas, 2016)) and the Theory of Matrices and Etymons (TME) developed by him. This theory offers a three-level view of morpheme formation:

- The first level is the **matrix** (μ) which is a non-ordered pair of vectors of phonetic features (such as [+nasal], [-nasal], [+continuant], etc.), where each pair is associated with a generic notion representing a mental experience. For example, in Arabic, the matrix {[±nasal], [+continuant]} is linked to every term related to the nose (such as respiration, sniffing, etc.).
- The second level is the **etymon** (ε) which is a non-ordered pair of concrete consonants "implementing" a given matrix. For example, the {m, š} implements the {[±nasal], [+continuant]} matrix, since m $\in$ [±nasal] and š $\in$ [+continuant].
- The third level is the **radical** (R) which is an etymon developed by diffusion of the final consonant, prefixation, or incrementation (initial, internal, and final) and including at least one vowel, recorded in the lexicon or provided by the morphological mechanisms of the language, and developing the notional invariant of the matrix/etymon. At this level the lexicon and morphology meet.

In the field of psycholinguistics, (Boudelaa & Marslen-Wilson, 2001) used two psycholinguistic experimental techniques, cross-modal immediate repetition priming, and masked morphological priming, in order to test whether etymons can act as lexical units. For example, during the cross-modal immediate repetition, the participants were examined on different types of {prime, target} pairs, where the prime and the target are related to each other etymologically and semantically, etymologically but not semantically, only phonetically completely unrelated. The results clearly showed that two words sharing an etymon do facilitate each other both in cross-modal and masked priming even though they do not share a root.

On the other hand, (Bentin & Frost, 2001) criticized this experiment, arguing their point of view with traditional arguments against the theory of etymons, such as the complexity of identifying an etymon from a given root, as well as empirical arguments, conducting an experiment with Arabic-speaking students, which encourages participants to recognize the etymons of words they know well. The results of this experiment were less satisfactory.

This paper examines biconsonantal etymons as a morphological phenomenon in Biblical Hebrew. Biblical Hebrew is a Semitic language, and as such its roots consist only of consonants. The vowels, serve as the glue between the consonants and are not stable when listing the different syntactic forms of the root. An illustrative description of this situation is the metaphor of consonants as the pegs of a tent, and vowels as the tent sheets. A Semitic language first hammers the pegs and only then lays out the different sheets that may change depending on the weather (the context). To illustrate this characteristic of the Semitic languages, we review several words derived from the root KTV denoting the English root “write”: **K**o**T**e**V** = is writing, **K**a**T**a**V** = wrote, **K**a**T**u**V** = written, **KT**i**V**a = writing. Even the highly interconnected forms of the root show a prominent level of flexibility, if not total freedom, in selecting vowels along with high adherence to the three consonants of the root.

Most of the roots of Biblical Hebrew have three consonants. Several indicators in Biblical Hebrew and its derivative, Modern Hebrew, confirm the evolution of biconsonantal (2C) etymons into triconsonantal (3C) roots. One indicator is the presence of a minority of 2C roots alongside the overwhelming majority of 3C roots in Biblical Hebrew. The next indicator is the presence of the “hollow" and the “disabled" subgroups of Hebrew roots, in which the third consonant is simply missing, and the root is made up of two consonants and an auxiliary semi-consonant. Another indicator is the presence of 2C etymons that do not change their central meaning when the third consonant is added to them to form a 3C root. An example of such an etymon is the Semitic etymon PR whose primary meaning is “fruitfulness", a word that most succinctly describes a process of multiplication and expansion. Indeed, the addition of a third consonant does not harm this basic meaning and only adds a semantic color to it. It is reasonable to assume that during the formation of the Proto-Semitic language, these additional consonants were added as suffixes. Figure 1 lists the various roots derived from the PR etymon and their meanings[1], expressing the various shades of expansion: explosion, proliferation, exaggeration, separation, etc.

Other, somewhat less convincing examples, include the etymon *QṢ*, which primarily means “edge". It is assumed that some roots with semantics close to “cut" have evolved from this etymon, including *QṢV* = allot, *QṢR* = reap, *QṢṢ* = chop. Some etymons hold similar meanings. For example, the etymon *GZ* is semantically close to *QṢ* and its derivative roots mean additional shades of “cut".

Another indication of the evolutionary development of Semitic roots is the presence of quadriconsonantal (4C) roots in the Hebrew Bible. The most famous Biblical 4C roots are:

- ŚMʾL = left.
- ṬʾṬʾ = sweep with a broom.
- ŠḤWH = bow.
- DRDR = thistle.
- GLGL = wheel.

Many 4C roots have developed by duplicating a 2C root. These roots usually mean an atomic repetitive action (Vera, 2021). This repetitive pattern is also used in modern Hebrew. Other 4C roots are developed by adding a new letter to a 3C

- PRʾ = wild, savage (one who cannot control his instincts and they explode out).
- PRG = a flower with many seeds, spread out by the wind.
- PRD = separate (from one to many)
- PRH = proliferate
- PRZ = erase boundaries, exaggerate
- PRḤ = flower, flourish, prosper
- PRṬ = break down into pieces
- PRK = break gradually
- PRM = unstitch
- PRS = cut into pieces
- PRʿ = ruffle
- PRṢ = burst
- PRQ = take to pieces, unpack
- PRR = crumble, granulate
- PRŠ = retire, secede
- PRT = detail

Figure 1: Various roots derived from the PR etymon.

root. These patterns highlight the combinatorial nature of the formation of Semitic roots and words by using 1C “bricks" and 2C “etymons" and applying either concatenation, duplication, or permutation to generate new words.

However, it can be argued that many arbitrary sets of words, not necessarily related to the same etymon or root, tend to show some level of semantic correlation because the semantic aspects of words are very numerous, and the words of any arbitrary small set are expected to overlap on a subset of these aspects. For example, the words "**P**a**Ṭ**i**Š**=hammer, ma**SM**e**R**=nail, and maʿ**D**e**R**=hoe" have numerous semantic aspects in common, as they all belong to the semantic category of craft tools, even though they come from different roots and etymons. But even randomly selected words are supposed to be correlated to each other by some subset of semantic aspects.

In this paper, we show the high feasibility of a gradual evolutionary process and establish the evolution of Semitic syllables, or more correctly biconsonantal (2C) roots (etymons) [2], also known as “gates" in the Jewish religious literature [3] to triconsonantal (3C) roots. Specifically, we use the Biblia Hebraica Stuttgartensia Amstelodamensis (BHSA), a manu-

[1]The Hebrew to Roman transliteration used along this paper was introduced by the Hebrew Academy in 2006 replacing their 1953 conventions.

[2]Vowels do not participate in the formation of Semitic roots.

[3]The term “gate" comes from the fact the two consonants are seen as the two columns/jambs of an actual gate

ally annotated corpus of the Hebrew Bible, and Word2Vec, a method for converting words to vectors representing their semantic meanings, to study the hypothesis of evolution from biconsonant etymons to roots in biblical Hebrew. We empirically show that the semantic relation between words obtained from the same biconsonant etymon is higher than that of a randomly selected set of words. These differences are statistically significant and strongly support the hypothesis of evolution from biconsonant etymons to roots.

To summarize, the main contribution of this work is the empirical confirmation of the controversial theory of the gradual formation of semantic structures from simple to more complex.

Related Work

The works of (Bohas, 2016), (Boudelaa & Marslen-Wilson, 2001) and (Bentin & Frost, 2001), which are most related to the concept of etymon, have been discussed in the introduction of this paper.

In addition, (Greenberg, 1949) compares the comparative reconstruction of natural languages and the genetic analysis. Comparative reconstruction is a comparison of forms between different languages to find unidirectional or mutual influence, while genetic analysis is a classification of languages into language families. Our hypothesis belongs to the category of comparative analysis.

An early source that expands on the “consonant-etymon-root" evolution is the Kabbalistic “Book of Creation", dated to the Tannaitic period (Westcott, 2018). The book, which describes the formation of language in the Platonic concepts of celestial abstraction, examines in detail the formation of language in stages, starting from the stage of the formation of letters, moving on to the stage of combining pairs of letters into biconsonant etymons and ending with the stage of combining triplets of letters by selecting a pair of letters from the previous step and a single letter and applying combinatorial operation, such as concatenation and permutation.

In addition, there are many modern works on linguistics that describe and discuss the phenomenon of interest to us. Some of them refer to this phenomenon in the context of comparative linguistic reconstruction, which is a practice of establishing features of the ancestor of two or more related languages, belonging to the same language family, by means of the comparative method.

(Ehret, 1995) assumes the relations between 2C and 3C roots and applies tools for reconstructing the ancient forms and paradigms of Proto-Afroasiatic. Another source is (Orel & Stolbova, 1995), which also deals with the reconstruction of the Proto-Afroasiatic language. These works differ in the comparative lexicon offered by each of them. There are a few interesting nuances between (Ehret, 1995) and (Orel & Stolbova, 1995) on what manipulations should be considered legal when matching 2C roots and their 3C relatives. For example, (Ehret, 1995) states that the third consonant is always a fossilized suffix, so it is added to the end of the 2C root. In contrast, the manipulations allowed by (Orel & Stolbova, 1995) are more flexible, and so each 2C root spans a larger space of 3C roots.

A recent paper closely related to our discussion is the work of (Zeldin, 2021), in which he discusses the evolution of Hebrew and other Semitic roots and presents many interesting aspects of it. (Agmon, 2010) expands the discussion to the coincidence of the evolution of material with the evolution of language on the time axis. (Ratcliffe, 2013) studies the notion of the linguistic reconstruction itself. Ratcliffe states that the evaluation of a proposed reconstruction must go beyond qualitative evaluations of individual proposed cognate sets and incorporate quantitative tools for evaluating the probable degree of chance matches within the reconstruction as a whole. He presents several examples of such quantitative tools. The method proposed in our paper is a quantitative tool based on advanced Machine Learning techniques.

In this paper, we utilize word2vec to support the hypothesis of biconsonantal etymons in Biblical Hebrew. Word2vec (Mikolov, Chen, Corrado, & Dean, 2013), is a machine learning algorithm for training natural language models. Word2vec is trained on a text corpus and maps the vocabulary of this corpus into a collection of vectors, also known as word embeddings. Each entry in the resulting vector qualifies as a semantic feature of this word. Furthermore, the distances between word vectors in their vector space successfully reflect the semantic similarities between the original words. For example, subtracting the vector associated with the word “Man” from the vector associated with the word “King”, and then adding back the vector associated with the word “Woman” results in a vector that is very close to the vector associated with the word “Queen”.

The evaluation method

Detailed description

The main idea of our experiment is to either prove or disprove (or at least not prove) the hypothesis of the "2C to 3C" evolution, which is the second part of the entire hypothesis of the "1C -> 2C -> 3C" evolution, by using statistical means, in particular advanced Machine Learning techniques. The hypothesis of "2C -> 3C" evolution would be confirmed if we achieved a prominent level of statistical significance in finding high within-cluster semantic similarity when clustering the words of the target corpus by their biconsonantal etymons.

Our algorithm

Our algorithm obtains an annotated corpus and returns the mean of within-group means of pairwise similarities of words grouped according to their 2C etymons. The algorithm works as follows:

- Train a word embedding model on the corpus.
- Iterate through the corpus words, and for each word infer the 2C etymon from the root.

- Group the words by the 2C etymon, so that every root is represented by a single word inside the group of its 2C etymon
- Convert the words to their embeddings.
- For each group evaluate the list of the pairwise similarities between the vectors within the group and return the mean.

The actual code used in this paper is publicly available at: https://github.com/mstekel/gate-theory.

The corpus

We used the Biblia Hebraica Stuttgartensia Amstelodamensis (BHSA) corpus, which is an open Hebrew Bible Database using text-fabric. It contains the text of the Hebrew Bible, augmented with linguistic annotations, compiled by the Eep Talstra Centre for Bible and Compute, etc. diversity Amsterdam (Winther-Nielsen, 2019).

Before passing the corpus to our pipeline, we ran the following preprocessing:

- First, we filtered out words other than verbs, nouns, adverbs, or adjectives because the other parts of speech carry less contextual information.
- We then normalized the words by converting them to their lexemes.

The word embedding algorithm

As a word embedding algorithm, we used **word2vec**. This algorithm works in a reasonable amount of time and provides good contextual correlations between the resulting word embeddings, even when trained on small corpora (Stekel, Azaria, & Gordin, 2022). This algorithm uses a neural network to train a language model. It supports two different training modes—continuous bag-of-words (CBOW) or continuous skip-gram. In CBOW training mode, the task is to predict a masked word from its context (not necessarily the entire sentence, but rather a window of words surrounding it, when the length of the windows is a hyper-parameter). In the continuous skip-gram mode the task is reversed—to predict the context (again, a continuous window of surrounding words). The word embeddings are compiled from the output vector of one of the last layers of the neural network.

Finding Biconsonantal Etymons

Inferring the 2C etymon from a 3C root works as follows:

1. Look for W Y H N ᾿ M T (the order matters). If found, delete the first occurrence and return the result.
2. Otherwise, delete the third letter and return the result.

The rationale behind this logic is that the W Y H N ᾿ M T letters are considered unstable in Hebrew. These letters are often used as "reading servants" (vowel markers) or as prefixes or suffixes. The specified order W Y H N ᾿ M T determines the level of instability from higher to lower and therefore, the most unstable found letter should be removed first. If no unstable letter was found, remove the third one, assuming the tendency of generating new words by adding suffixes. The final assumption will be discussed in Section .

Results

Our dataset contains 69484 lexemes from the aforementioned BHSA corpus, annotated by 435 distinct roots. Applying our method of finding the 2C etymon from a given 3C root has resulted in 109 different 2C etymons. We have obtained 42 different 2C etymons with 2 3C roots per etymon, 33 with 3 roots per etymon, 19 with 4 roots per etymon, 9 with 5 roots per etymon, a single etymon with 6 roots, and 2 etymons with 7 roots (see Table 2). The global mean of the within-group means (where a group is the group of words belonging to a single 2C etymon) of the pairwise similarities between the words has been evaluated to 1.019, while the global mean of the within-group means for groups of the corresponding sizes but populated by randomly selected words has been evaluated to 1.068. These differences are statistically significant ($p < 0.02$; using a student t-test). Therefore, we could conclude that the semantics of words with similar 2C etymons are closer to each other than random words.

We have also compared the global mean of per-etymon within-group means above to the mean of means similarly obtained from groups of words that belong to the same root within each group. As one might expect, grouping the words by their roots has resulted in a much lower global mean pairwise similarity of 0.934. See the results summarized in Table 1.

Table 1: This table presents the average distance between words originating from the same root, compared to the average distance between words originating from the same 2C etymon, and finally the distance between randomly selected words. The results obtained confirm our hypothesis about the semantic similarity of triconsonantal roots originating from the same 2C etymons.

The same 3C root	The same 2C etymon	Random words
0.934	1.019	1.068

We conclude this section by providing four examples of pairs of words with the same 2C etymon but with different roots. Two of the pairs have a much greater than the average semantic distance, and the two have a less than the average semantic distance. Namely, tables 3, 4, 5 and 6 list the four pairs of words derived from the same 2C etymon. We note that the large semantic distances between the first two words may be due to the low occurrence of these words in the corpus, such that each instance is associated with a different context. Therefore, the word2vec embedding may not be able to accurately infer their meanings.

Table 2: The number of root members of the different 2C etymons.

Number of roots per etymon	Number of etymons	Mean Similarity
2	42	0.99
3	33	1.02
4	19	1.08
5	9	1.00
6	1	0.90
7	2	1.09

Table 3: Example of two words with the same 2C etymon (ʾŠ) but different roots. As noted, the distance between these words is much higher than the average distance between words with identical 2C etymons. Interestingly, a semantic relation between these words is expressed in other sources—traditional Jewish texts associate the lack of divine harmony between man and woman with fire. In addition, the myth of Prometheus associates fire with the virtue of man. However, these associations do not appear in the Hebrew Bible, and therefore, the word2vec embedding model could not capture them.

Hebrew	English	The Root	The 2C Etymon
אֵשׁ	fire	ʾŠ	ʾŠ
אִשָּׁה	woman	ʾNŠ	ʾŠ
Similarly compared to overall mean:			135%

Table 4: Example of two word with the same 2C etymon (ṬF) but different roots. The distance is between the two vectors that represent these words. The connection between these words is clear, as balm drips, so do babies drip and patch things. The inability of the word2vec model to capture this connection is due to the rarity of the word "balm" in the Bible—it occurs only once.

Hebrew	English	The 3C Root	The 2C Etymon
נָטָף	dripping balm	NṬF	ṬF
טַף	babies	ṬNF	ṬF
The distance compared to the average:			137%

Conclusions

The experiment described in this paper empirically substantiates the hypothesis of the evolution of Semitic morphology from two-consonant roots to three-consonant roots. This is achieved by demonstrating a statistically significant difference between the distance of vectors representing the semantics of words of different roots derived from the same biconsonantal root and those of random sets of words of the same size. The distances between the semantics of words are measured using the word2vec language model trained on BHSA, a manually annotated Hebrew Bible corpus. This experiment sheds light on the evolution of Semitic morphology, which, in turn, can reflect the evolution of Semitic semantics and serve as evidence for understanding the processes of formation of human language and thought.

Table 5: Example of two word with the same 2C etymon (ŚV) but different roots ("old age" vs. "satisfied"). The close distance is justified semantically since ŚV (= an aged man) is also called "satisfied of days"/"fed of days" in Hebrew (ŚVʿ = fed).

Hebrew	English	The 3C Root	The 2C Etymon
שָׂבָע	satisfied	ŚVʿ	ŚV
שֵׂיבָה	old age	ŚYV	ŚV
The distance compared to the average:			58%

Table 6: Example of two word with the same 2C etymon (Šʾ) but different roots. The distance is between the two vectors that represent those words. The distance here is small because even in the Bible the first word may be interpreted as "noisy" in addition to "calm". In other words, these are either etymologically related synonyms or antonyms.

Hebrew	English	The 3C Root	The 2C Etymon
שַׁאֲנָן	calm	ŠʾN	Šʾ
שָׁאוֹן	noise	ŠʾH	Šʾ
The distance compared to the average:			40%

Future work

There are several directions for future work. Some focus on improving the quality of our algorithm and data set to show even better results, while others focus on solving further morphological evolutionary patterns using the same technique.

Optimize the algorithm by changing the underlying assumptions

In the course of the experiment, we had to proceed from some assumptions that seemed most likely (see below). However, these assumptions are not absolute, and we can try to improve the results by questioning them and trying new directions and examining alternative evolutionary patterns. For example, our proposed method for inferring 2C etymons from 3C roots can be revised such that, if no unstable letters are found, the letter to be removed to find the etymon is not necessarily the last one. If we refer back to the PR etymon, we can assume that roots like PʾR, PTR, or PʿR evolved from this etymon, although this evolutionary pattern adds a letter to the middle of the etymon, not to the end. Moreover, sometimes the first letter of a root may be claimed to come from a prefix. At least one example of such a formation is widely known and obvious—the TRM=donate root originates from the RÛM=raise/donate root and the T prefix. A second evolutionary pattern worth examining is anagram/metathesis, i.e., rearranging the letters of a given word to form a new word. For example, there seems to be a relation between the roots *PRZ*=spread and *PZR*=scatter. Similarly, there seems to be a relation between KVŚ and KŚV, both mean "a sheep".

More patterns to be examined are assimilation, voicing voiceless phonemes or unvoicing voiced phonemes. As we already mentioned, there seems to be a relationship between QṢ and GZ (Everyone can convince himself by comparing *QṢṢ*=chop to *GZZ*=shear or *QṢR*=harvest to *GZR*=cutout).

Finding larger corpora

Using a larger corpus may improve results. We may either add more root annotations to BHSA (currently only about a quarter of words have root annotations) or attempt to use the Modern Hebrew corpus, even though it may be less efficient in finding the correct contextual correlations. Using Tannaic Hebrew and Talmudic Hebrew is another way of expanding the corpus.

Testing more etymological hypotheses

Another direction for future work is to investigate etymological phenomena similar to 2C-to-3C evolution. For example, we may investigate the formation of quadriconsonantal roots (roots made up of four consonants). A brief overview of quadriconsonantal roots can highlight the following patterns of formation:

- 2C etymon duplication, like in the *GLGL*, *DRDR*, or *VMVM* roots. For example, the graph in Figure 2 illustrates various evolution branches of the *GL* etymon to triconsonant roots as well as a quadriconsonant root *GLGL*.

- A combination of two different 2C etymons. For example, the rare biblical root ḤŠML, which denotes a shade of light and is used in modern Hebrew to denote electricity, is sometimes explained as a combination of two 2C etymons: ḤŠ (to be silent) + ML (to speak).

- Using paradigms from other Semitic languages to create Hebrew roots. For example, the quadriconsonantal root ŠKLL developed from the triconsonantal root KLL using an Akkadian paradigm of prepending Š to the root (this pattern has become widespread in modern Hebrew, as you can see in the roots ŠḤZR, ŠḤPL, and others).

- Adoption of roots of foreign languages, for example, the root HNDS is of Persian origin.

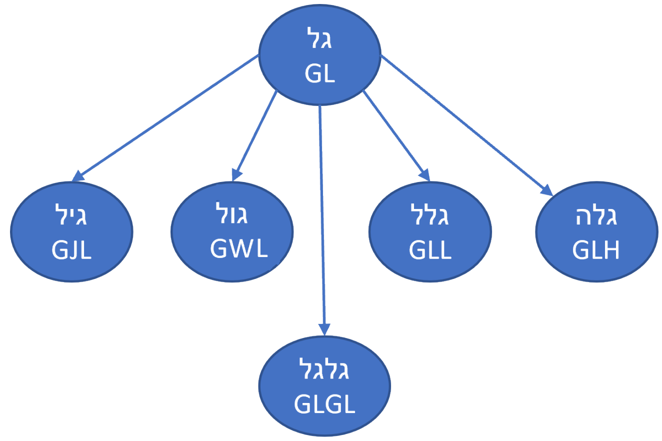

Figure 2: The evolution branches of the *GL* etymon

References

Agmon, N. (2010). Materials and language: Pre-semitic root structure change concomitant with transition to agriculture. *Brill's Journal of Afroasiatic Languages and Linguistics*, *2*(1), 23 - 79. Retrieved from `https://brill.com/view/journals/aall/2/1/article-p23_2.xml` doi: https://doi.org/10.1163/187666310X12688137960669

Bentin, S., & Frost, R. (2001). Linguistic theory and psychological reality: A reply to boudelaa & marslen-wilson. *Cognition*, *81*(1), 113–118.

Bohas, G. (1977). 'iltiqā'u s-sākinayn et problèmes connexes. *Bulletin d'études orientales*, *29*, 73–84.

Bohas, G. (2016, 06). Organisation and consequences of the ?teory of matrices and etymons? (tme).

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the arabic mental lexicon. *Cognition*, *81*(1), 65-92. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0010027701001196` doi: https://doi.org/10.1016/S0010-0277(01)00119-6

Ehret, C. (1995). *Reconstructing proto-afroasiatic (proto-afrasian): Vowels, tone, consonants, and vocabulary*. University of California Press. Retrieved from `https://books.google.co.il/books?id=cN9yUe3AgQIC`

Gesenius, W. (1817). *Gesenius' hebrew grammar* (edited and enlarged by Kautzsch ed.; E. Kautzsch, Ed.). Göttingen: Vandenhoeck & Ruprecht.

Greenberg, J. H. (1949). Studies in african linguistic classification: Iii. the position of bantu. *Southwestern Journal of Anthropology*, *5*(4), 309–317. Retrieved 2023-01-22, from `http://www.jstor.org/stable/3628591`

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Orel, V., & Stolbova, O. (1995). *Hamito-semitic etymological dictionary: Materials for a reconstruction*. E.J. Brill. Retrieved from `https://books.google.co.il/books?id=dFUOAAAAYAAJ`

Ratcliffe, R. R. (2013). *On calculating the reliability of the comparative method at long and medium distances: Afroasiatic comparative lexica as a test case.* Tokyo University of Foreign Studies. Retrieved from https://benjamins.com/catalog/jhl.2.2.04rat doi: https://doi.org/10.1075/jhl.2.2.04rat

Stekel, M., Azaria, A., & Gordin, S. (2022, june). Word Sense Induction with Attentive Context Clustering. *Journal of Data Mining & Digital Humanities*, *NLP4DH*. Retrieved from https://jdmdh.episciences.org/9665 doi: 10.46298/jdmdh.9175

Vera, A. (2021, May). *The Development of Quadriliteral Reduplicated Roots as an Aspectual Marker in Modern Hebrew.* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.4749508 doi: 10.5281/zenodo.4749508

Westcott, W. W. (2018). *Sepher yetzirah*. Lulu. com.

Winther-Nielsen, N. (2019). Interfacing the hebrew bible: past, present and future applications for the bhsa. *HIPHIL Novum*, *5*(2), 143-152.

Zeldin, A. (2021). The semitic root evolution (cultural and historical aspect). *Uchenye Zapiski Kazanskogo Universiteta. Seriya Gumanitarnye Nauki*, *163*(4-5), 47–66.