Introduction (2 minutes)

Good afternoon. As was just mentioned, my paper is titled “The Menzerath- Altmann Law and Ancient Hebrew: Does the Bible Break the Law?” In this paper, I will first describe the Menzerath-Altmann Law (also referred to as the MAL), its history, as well as its application to various languages. I will then discuss the corpora and tools that I have used for this research. Next, I will present a summary of the data that I have collected and the main results of my analysis. Finally, I will close with a look toward future research.

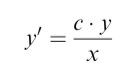
Let me begin by sharing the motivation behind this project. Currently, I am interested in three research areas: general linguistic theory, computational linguistics, and the Hebrew language. In order to bring these three areas together, I decided to look into foundational laws of linguistics, such as Martin’s law, rank-frequency laws, the Poitrowski law of language change, and the MAL. I chose to work with the MAL mainly because the tools we have available to us in Hebrew studies allow for a robust analysis of this law. Further, my interested in computational linguistics drove me to work with the MAL because this law cannot be thoroughly analyzed without the assistance of a computer. The main reason for this is the large amount of data that is necessary to undertake such an analysis. For this study, as an example, I analyzed every sentence, clause, phrase, word, and letter in the Biblia Hebraica Stutgartensia, the War Scroll, and the Community Rule; a total of over 2 million data points. Of course, analyzing such a large dataset requires computational linguistic tools. But before I get into that more, I will turn to the MAL itself.

MAL (4 minutes)

The Menzerath-Altman Law was proposed by Paul Menzerath in 1928 when he applied it to phonology. As an example, his work showed that “The longer a word (measured in terms of the number of syllables it consists of) the shorter (on average) the syllables of the given word.”

In 1980, Gabriel Altmann generalized Menzerath’s work, showing that it applied to all linguistic levels. Altmann stated “the longer a language construct, the shorter its constituents.” Altmann and others suggested that this is true due to the linguistic principle of minimizing memory. It is thought that as a language construct gets longer, its constituent parts necessarily need to be smaller to allow for comprehensible communication.

Not only did Altmann generalize this law to cover all linguistic levels, he also derived its mathematical form. Altmann described the law through a functional model and differential equation, seen here:



The following function presents the solution to this differential equation:

*y* = *axbe*−*cx*

Due to strong empirical evidence, this simplified form of the function is thought to be just as accurate:

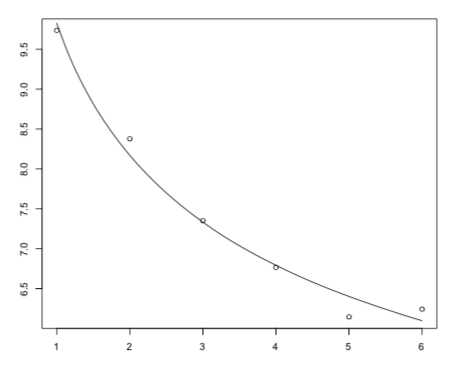
*y* = *axb*

Here, y is the mean size of the constituents, while x is the size of the construct. As an example, if we were analyzing the relationship between word length and syllable length, x would be the length of the word and y would be the mean length of the syllables in that word. The constants a and b are parameters and are different for each text that is analyzed and are therefore considered to be text characteristics. Here, some examples will be helpful to elucidate the application of this equation to linguistic data.

Köhler was the first to empirically test the MAL on the sentence level when he analyzed short stories and philosophical texts in English and German. Here we see a summary of the data he collected:

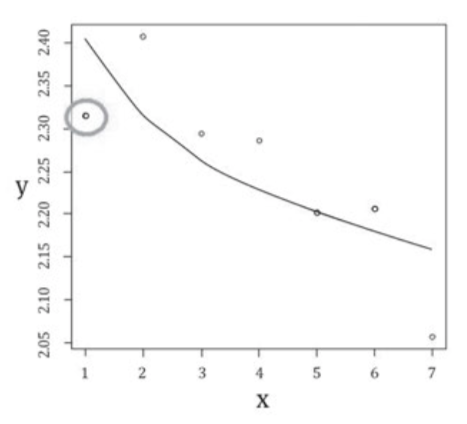
|  |  |
| --- | --- |
| Sentence Length (in clauses) | Mean Clause Length (in words) |
| 1 | 9.7357 |
| 2 | 8.3773 |
| 3 | 7.3511 |
| 4 | 6.7656 |
| 5 | 6.1467 |
| 6 | 6.2424 |

In the left-hand column, we see the length of the sentence measured in the number of clauses that it has. In the right-hand column is the mean length of the clauses measured in number of words. As we can see, the shortest sentences (those with only one clause) tend to have the longest clauses averaging about 9.7 words. Sentences that have two clauses, average clause lengths of about 8.4 words. And as the sentences get longer, the average clause length gets smaller. The inverse relationship of this data can clearly be seen in this visualization:



Here the x axis shows sentence length and the y axis, clause length. The dots represent the actual data collected, while the line presents the data fitted to the MAL differential equation. One can think of this line as the idealized representation of the MAL given the actual data. The closer the actual data is to this line, the better the MAL fits the text that is being analyzed. This fit can be calculated by using a determination coefficient denoted as R2. The R2 for Köhler’s data is 0.9858, telling us that the texts he analyzed very closely fit the MAL, thus supporting the validity of the law. Let me present a different example briefly.

Benesová and her colleagues analyzed a Czech text from the Olomouc Speech Corpus consisting of a 23-minute dialogue between four speakers. The results are summarized in this table:



The goodness-of-fit for this regression curve is a R2 of 0.5710. Thus, the fit is relatively good, but not nearly as close as the first example. Further, particular points do not follow the MAL’s assumption that each observation will be lower than the last. As we can see here, where point two is higher than point one, breaking the MAL. While there is no agreed upon R2 value that is a threshold for confirming the MAL, this result of 0.5710 is relatively low.

These are just two examples of the many studies which have shown different languages follow the MAL. The two examples given here, represent the spectrum of results. Some languages and levels show rather poor fits to the MAL, such as the second example, while most languages fit very closely to this law as was seen with the first example.

Corpus (1.5 minutes)

Now I will turn to the study at hand. My analysis utilizes the BHS as well as 1QM and 1QS. I focus on three linguistic levels: sentence, clause, and phrase. For the sentence level, I measure the length of the sentence by the number of clauses that it contains and the length of those clauses by the average number of phrases in each clause within the sentence. As an example, if a sentence has four clauses and 12 phrases, the length of that sentence is tagged as 4 and the length of the four clauses is tagged as 3 each. I gathered and analyzed data for every text in the BHS, but due to the challenges of identifying sentences within poetry, I will base my conclusions only on the narrative texts.

For the clause level, I measure the length of the clause by the number of phrases that it contains and the length of those phrases by the average number of words in each phrase within the clause. For the phrase level, I measure the length of the phrase by the number of words that it contains and the length of those words by the average number of letters in each word within the clause. For both the clause and phrase levels, I include all the narrative and poetic texts of the BHS.

Tools (1.5 minutes)

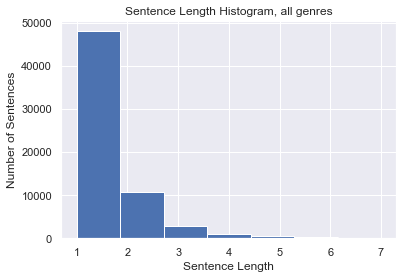
For my analysis, I utilize the Eep Talstra Center for Bible and Computer linguistically tagged database of the BHS and 1QM and 1QS. For over 30 years, the ETCBC has developed and maintained advanced syntactic databases of the Hebrew Bible and other ancient texts. One of the main advantages of utilizing the ETCBC databases is that they are open source, allowing research conducted with those databases to be reproducible – one of the hallmarks of good scientific inquiry. Another advantage is that the ETCBC databases are accessible with the programming languages, Python and R which are commonly used in computational linguistics. Utilizing a programming language allows us to gather and process large amounts of data and to utilize statistical packages to analyze that data. As an example, I used Python’s SciPy and Math packages to fit my data to the MAL differential equation, a task that would be rather arduous if done by hand. Further, programming languages also assist in reproducibility as the code which was used to work with the data can be made publicly available. The code for this project, as well as all of the data, visualizations and this paper can be found on my github page – github.com/jarodjacobs.

Data

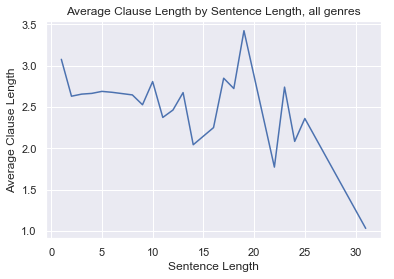
As I mentioned previously, I collected over 2 million data points for this analysis. Before turning to my results, I will briefly summarize this data and highlight a few points. Here I will focus on the BHS, bringing in the Qumran scrolls in the concluding section of this paper.

First, I will explore the sentence level. For this level, I measure the length of a sentence by the number of clauses it contains. The length of the clauses is calculated by counting the number of phrases in the sentence and dividing that number by the number of clauses in the sentence. This provides us with the average clause length.

There are a total of 63,570 sentences in the BHS. As can been seen by this histogram:

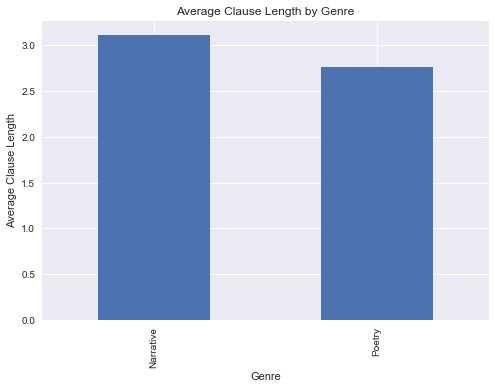


I have excluded sentences that are longer than 7 clauses from this analysis for two reasons. First, many of the longer sentences contain lists, making them artificially lengthy. As an example, the longest sentence, which is found in 2 Samuel 23:24-39, contains 31 clauses and is a list of David’s thirty mighty men. This sentence is clearly an outlier and has an undue impact upon the data. The second reason I have excluded sentences longer than 7 clauses is because of their relative infrequency. As this histogram shows, there are nearly 50,000 sentences with one clause, over 10,000 with two and almost 3,000 with three clauses. However, all together, there are only 165 sentences that are 8 clauses or longer. Including these long sentences would have given this small group a large impact on the results that is unwarranted. This impact can be seen in the following chart:



Here we see sentences by length on the x axis and their average clause lengths on the y axis. Notice that after a sentence length of 7, the data peaks and dives sharply. This is due to the lack of data for each sentence length causing their averages to be rather extreme.

Also, of importance is the separation of narrative and poetic texts. Due to the difficulty in identifying sentence units in poetry, I have excluded the poetic books from this analysis focusing instead on the narrative books where sentence boundaries are clearer. However, here I will briefly highlight a couple things regarding the data from the poetic books. As this chart shows, the poetic books have an average sentence length nearly one clause shorter than the narrative books:



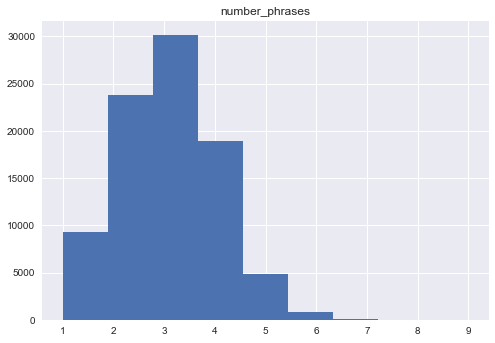
When plotted according to length of sentences, we see that the poetic books consistently have less clauses, while generally the data of the two corpora follow similar trends as seen here:



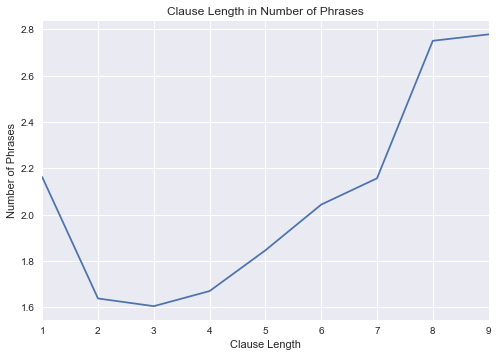
The blue line represents the data for the narrative books and the green the poetic books. We see that both have a significant drop in sentence length from 1 to 2 with the poetic books raising in length from 3 to 4 and both flatten out after that.

Now I will turn to a summary of the second level of my analysis – clauses. Here, I measure the length of a clause by the number of phrases that it contains and the length of the phrases by counting the number of words in the clauses and dividing that number by the number of phrases, which returns an average phrase length.

This analysis is based on the 88,000 total clauses in the BHS. These clauses follow a rough Gaussian curve, centered around a clause length of 3 phrases, as seen in this histogram:

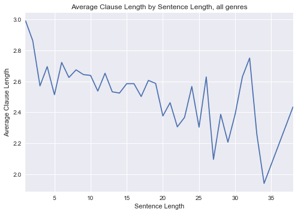


The length of clauses in the BHS ranges from 1 to 9 and for this analysis, I include all of these even though clauses with 8 and 9 phrases occur relatively rarely. I include these because they closely follow the pattern set by clauses with shorter lengths as can be seen in this graph:



For the final level of analysis for this project, I focus on phrases. I calculate the length of a phrase by counting the number of words that it has and the length of the words by counting the number of letters in the phrase and dividing that number by the number of words, return an average word length for the phrase.

There are over 253 thousand phrases in the BHS, made up of over 426 thousand words and nearly 1.2 million letters. For this analysis, I exclude phrases with lengths over 20 words, because that is where the average lengths begin to vary dramatically as seen in this graph:



Results

With this summary of the data in place, I will now turn to an analysis of the MAL and ancient Hebrew. To do this I will work through the three levels first by giving an overview of all the BHS texts together. Then I will highlight some individual books as well as 1QM and 1QS.

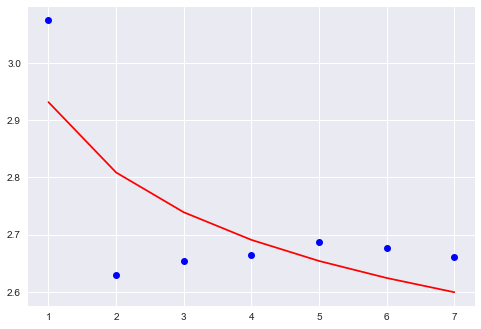
For the sentence level, as this data shows, we see a swift drop in average phrases between sentences that have 1 clause and sentences that have 2. From there we find the average clause length going up very slightly for the next three sentence lengths. Then we see the two longest sentence groups having lower average clause lengths.

|  |  |
| --- | --- |
| **Sentence Length** | **Average Clause Length** |
| 1 | 3.075193 |
| 2 | 2.629325 |
| 3 | 2.654567 |
| 4 | 2.663388 |
| 5 | 2.687727 |
| 6 | 2.676617 |
| 7 | 2.660881 |

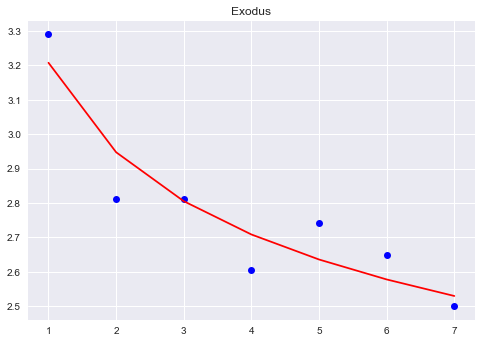
This graph clearly shows these trends:



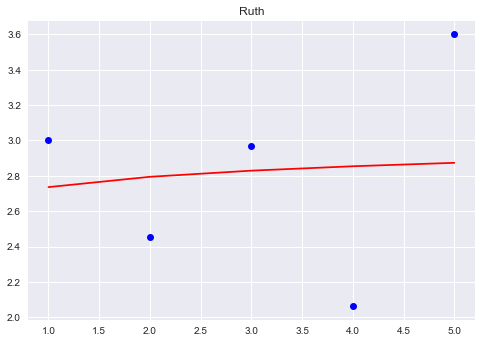
Just from the raw numbers we can see that at the sentence level, biblical Hebrew does not closely follow the MAL. The first steep drop is consistent with this law, but the steady rise in clause length after that for the next three lengths is not expected. This next chart presents what the MAL would have predicted given the data. The red line is data fitted to the MAL differential equation:



The R2 score for this data is 0.501. This reveals a relatively poor fit between the collected data and the MAL. Of the biblical narrative books, Exodus proves to have the closest fit to the MAL with an R2 of 0.843. This chart shows that the data for Exodus is indeed relatively close to the predicted values from the MAL differential equation:

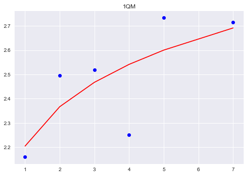


Only sentences with four clauses fall out of line with the expected steady decrease in clause length. On the other end of the spectrum, we find that the book of Ruth has the worst fit to the MAL with an R2 of 0.001. This poor fit is highlighted in this graph where we do not see a steady downward trend in clause length:



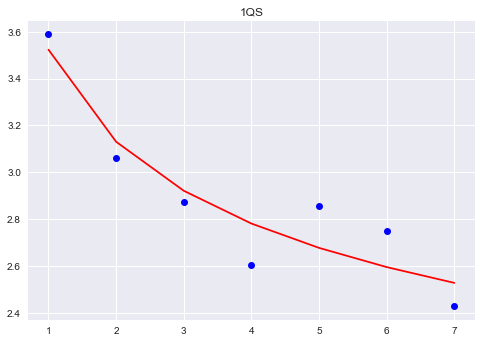
Overall, only four of the biblical narrative books have an R2 over 0.7: Genesis, Exodus, Deuteronomy, and Judges.

Now I will take a quick look at the two Dead Sea Scrolls covered in this analysis. 1QM has an R2 of 0.538, which while being on the low side still suggests an okay fit with the MAL. However, as can be seen with this chart:



The data from this scroll does not follow the main assertion of the law which is that clause length should descend as sentence length increases. In general, we see the opposite here.

1QS presents the opposite results. With at R2 of 0.857, it has a better fit to the MAL than all of the biblical narrative books. This graph highlights the closeness of the fit as well as the general trend of decreasing clause length as sentence length increases:

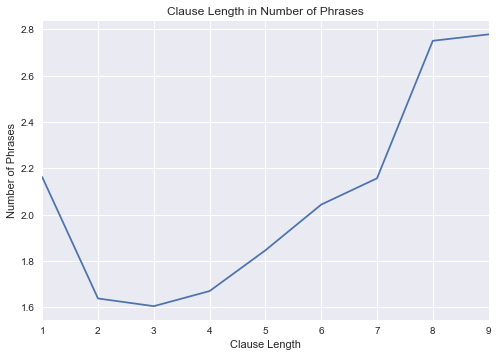


With the exception of a small portion of the biblical narrative books as well as 1QS, these results for the sentence level generally show a rather poor fit to the MAL. With that conclusion, I will now turn to the clause level.

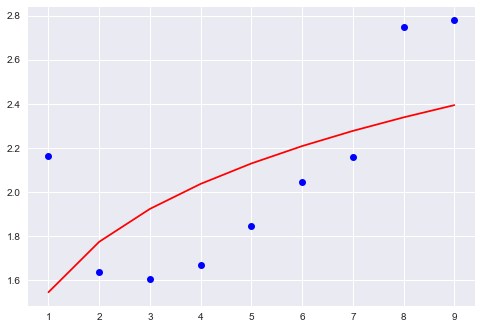
The data for the 88,000 clauses in the BHS follow a rather unexpected pattern as we see here:

|  |  |
| --- | --- |
| **Clause Length** | **Phrase Length** |
| 1 | 2.162078 |
| 2 | 1.63794 |
| 3 | 1.604729 |
| 4 | 1.669904 |
| 5 | 1.845819 |
| 6 | 2.043144 |
| 7 | 2.156522 |
| 8 | 2.75 |
| 9 | 2.777778 |

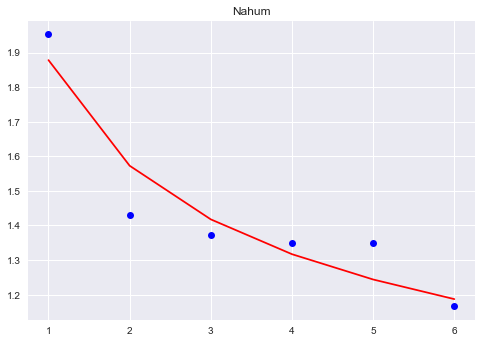
The data starts out with a drop from clauses that have one phrase to clauses that have three, but from there the phrase length steadily increase – the opposite of what the MAL predicts. This chart visually highlights this well:



As we would expect from this data, the phrase level has a poor fit with the MAL, having an R2 of 0.294 and not following the downward trend we expect for longer clauses:

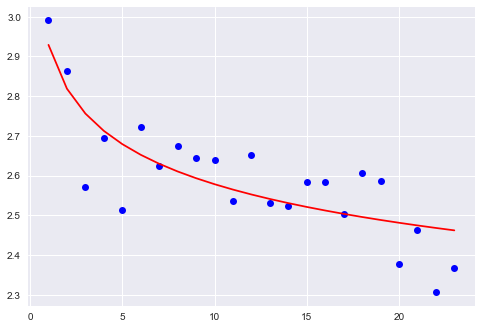


Of the 39 biblical books, only six have an R2 above 0.50: Joel, Nahum, Habakkuk, Psalms, Ecclesiastes, and 1 Chronicles. Of these only Habakkuk, presents a steady downward trend in clause length.

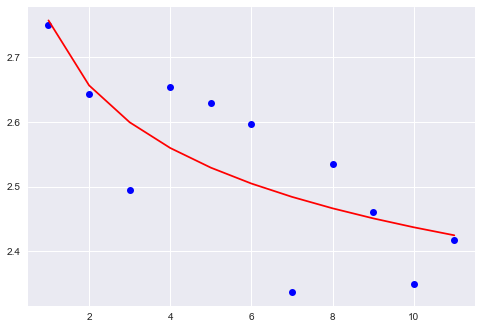


Similarly, both 1QM and 1QS present poor fits to the MAL, with R2s of 0.333 and 0.164 respectively.

The above analysis has shown that generally speaking, the Hebrew at the clause leve in the BHS, 1QM, and 1QS – with some exceptions – does not follow the MAL. This picture shifts as we turn to the phrase level. The R2 for the phrase level is the highest of the three levels – 0.655. While this is high compared to the other levels in Hebrew, it is still relatively low. However, when the data is fitted to the MAL differential equation, we see that the overall trend is a steady downward slope – as Hebrew phrases get longer, the words in those phrases tend to get shorter.



We find a similar result in 1QM and 1QS:



Overall, this shows a steady downward trend and with an R2 of 0.601, the results are a relatively good fit to the MAL.

I’ll take a moment now to summarize these results before moving onto some conclusions. The sentence level showed a loose fit to the MAL with an R2 of 0.501. The clause level revealed a worse fit with an R2 of 0.294 and finally, the phrase level showed the best fit with an R2 of 0.655.

With those results in place, the question at hand is why does Hebrew not fit better to the MAL? A robust answer to this question will require some more research, but for now I will discuss one possible theory.

For decades, scholars have debated over the nature of biblical Hebrew. Was it ever similar to a spoken dialect? Was it an artificial literary language? Are different phases of ancient Hebrew more “natural” than others? All these questions have long histories and prominent scholars argue both sides of their answers. I will not go too deep into this topic here, but the results of this analysis suggest that we may not be working with a natural language. The MAL has been shown to be true for many natural languages, both spoken and written. As one of the examples I discussed earlier showed, not all natural texts align well with the MAL, but a long history of research shows that most do. Since the MAL is a linguistic law that was developed by observing natural languages, it is obvious to ask: should we consider a language that breaks this law to by artifical? My analysis of course does not answer this question directly, but the results do align with the conclusion that some forms of ancient Hebrew way have been artificially developed. The data specifically supports this conclusion when considering the three linguistic levels I have analyzed for this paper. Both the sentence and clause levels did not fit will with the MAL, while the phrase level showed more consistency with this law. One might expect an artificial literary language to be overly stylized at the sentence and clause levels, while staying truer to the natural language at the phrase level. As an example, the Bible’s propensity for lists is not something we find in spoken conversation often. These lists have a large impact on the sentence and clause levels, while not playing as large a role in the phrase level. Further, an artificial literary language would need to draw from the natural language’s lexicon to build phrases, but could form clauses and sentences in ways that we would not often see in the spoken register. This possible interpretation of my results is just conjecture at this point. However, further study may be useful in proving or disproving it.

Of particular interest would be analyses of both Mishnaic Hebrew, which is thought by many to be closer to a natural form of Hebrew than biblical dialects. Also, working with modern spoken Hebrew could provide some clear answers to this question and as far as I know, a comprehensive study on the MAL and any form of Hebrew has never been done. Working with modern Hebrew is possible at this time as there are some robust natural language processing tools built for this language. Yet, using computational linguistics with rabbinic texts is further off. To that end, my research partner, Martijn Naaijer, and I, along with colleagues at Andrews University and the ETCBC have recently started working on a project that aims to using machine learning to computationally tagged ancient Hebrew texts. We intend to use artificial neural networks to teach a computer model how to accurately analyze ancient Hebrew at the morphological and syntactical level. This project will greatly speed up our ability to tag ancient Hebrew texts, providing much needed access to the linguistic data of the Dead Scrolls and rabbinic literature. The resulting database will of course be made open source and free of cost through the ETCBC. Projects, such as the one presented here, will be able to take major strides forward as this new set of data becomes available. We look forward to presenting on our progress in the coming years.

Thank you