APPLYING THE COHERENCE-BASED GENEALOGICAL METHOD (CBGM) TO
THE TEXT OF THE OLD TESTAMENT: AN EVALUATION

ADVISOR KENNETH CHERNEY

BY
DEAN ELLIS
APRIL 14, 2018
## CONTENTS

INTRODUCTION ................................................................................................................................. 1  
   The Benefits of a New Method ........................................................................................................ 2  

METHODS .......................................................................................................................................... 6  
   Manuscript Alignment ...................................................................................................................... 6  
   Pre-Genealogical Analysis ................................................................................................................ 7  
   Coherence Based Genealogical Analysis ......................................................................................... 8  

RESULTS: ........................................................................................................................................... 9  

DISCUSSION: ...................................................................................................................................... 13  
   The Current State of CBGM Analysis of Old Testament texts ..................................................... 13  
   Current Challenges to Implementing CBGM in the Old Testament .............................................. 13  
   Strengths and Weaknesses Inherent in this Method ....................................................................... 14  

BIBLIOGRAPHY ................................................................................................................................ 15
INTRODUCTION

A new method is being used in evaluating the general epistles in the New Testament. This new method, called the Coherence-Based Genealogical Method (CBGM), has led the publishers of the New Testament to elevate variants that were in the footnotes into the main text, and move variants that were in the main text into the footnotes. Is the CBGM method qualified to warrant these changes? Part of this paper is aimed at evaluating this method’s strengths and weaknesses as it assesses textual transmission. This is the first time this method will be applied to the Old Testament.

Currently, the Old Testament manuscripts are categorized as either belonging to the Masoretic and Proto-Masoretic texts, the Pre-Samaritan and Samaritan Pentateuch manuscripts, the Old Greek manuscripts or belonging outside these groups. Will the CBGM method support the formation of these categories or propose another set of categories that works better? During this study, we will strive to evaluate these groupings.

One of the goals of this project is to evaluate the relationship between a manuscript and the text it contains. The assumption is that an ancient text may very well be present in a manuscript that was recently copied. For instance, a late manuscript, like Leningradensis B19A, can embody an early form of the text, and vice versa, if “late” means “centuries removed from the author”. Another goal is to propose the text flow of these manuscripts using CBGM. The text flow is a diagram that shows the transmission of a text.
The Benefits of a New Method

How does CBGM work? Peter Gurry in his book *A Critical Examination of the Coherence-Based Genealogical Method in New Testament Textual Criticism* sets out to explain the principles behind this method. This is extremely important as up until recently this method has been described by many critics as a black box. In this book, Peter Gurry explains the principles and procedures behind the method. The breakdown of the CBGM method follows. Genealogy is established by evaluating the percent agreement between two manuscripts. If the percent agreement is high enough then there is a genealogical relationship between the manuscripts. This approach becomes much more complex when trying to reconstruct the genealogy of multiple witnesses at the same time. This is where the CBGM, as a computer-assisted algorithm, comes into play.

The CBGM method makes four initial assumptions. It assumes the scribe made copies with fidelity\(^1\). Secondly, it assumes that the scribe used an additional source when he strayed from the main source. Thirdly, it assumes that a scribe used few rather than many sources\(^2\). Finally, it assumes that the witnesses are closely related to all the sources that were used. With these four assumptions in hand an overview of the method is provided below.

The first step is for the Bible scholar to draw up as many local stemmata as possible for the corpus of the text being analyzed\(^3\). The local stemmata are the variants within texts. The

---


corpus of the text is the alignment of all texts. This alignment is made by aligning the witnesses and then evaluating variants that are considered by the editor to be significant. Of the significant variants the editor proposes the relationship between the variants. These relationships can be modified as the CBGM analysis continues.

The second step uses the proposed stemmata to compare the ratio of the prior and posterior readings between two witnesses. The prior and posterior readings are determined when the editor evaluates the sum of each variant. This ratio is used to suggest which of the two witnesses is the ancestor and which is the descendent. This helps suggest a flow for the texts.

The third step determines the smallest number of ancestors to explain the text in the proposed descendent. As multiple ancestors will be present for each descendent, the challenge is to identify the key ancestors in the formation of the descendant. In short, this is done by using the smallest number of ancestors that will explain the witness that is identified as the descendent.

The fourth step evaluates the ancestors for each witness and then combines all of them to make a large map of the flow of the text. This is often referred to as the global stemma. The global stemma also suggests the flow of the text through the witnesses and identifies the arch-text.

Throughout the description of these four steps the term stemmata were introduced. A stemma represents a relationship between manuscripts. Three stemmata are identified in the CBGM. The first is the local stemma, which represents the relationship at the variant reading level. The next is the substemma which represents the relationship between a manuscript and its

---

descendants. The final stemma is the global stemma, which is the overall representation of the relationships between all the manuscripts.

This relationship has been identified up until this point as the agreement between the manuscripts. Manuscripts that agree more closely with one another have a stronger relationship. A suitable synonym that can be used is coherence. In the CBGM method there are two types of coherence that are identified.

The first type of coherence is pre-genealogical coherence. This is the most important type of coherence. Using this coherence, the relationship between two witnesses is determined based on their agreement with one another. Pre-genealogical coherence is defined by Gurry as the agreement between two witnesses at all points of comparison. This is often given as a percent agreement based on the total number of places the texts agree. This type of coherence simply evaluates how closely related two texts are. While this type of coherence shows how close two witnesses are, it does not suggest which of the manuscripts is the ancestor and which is the descendant. It does not suggest which text came first. In short, it gives coherence but not genealogical coherence. It does show that witnesses are related without giving a substemma, a text flow.

This is where genealogical coherence comes into play. While pre-genealogical coherence says that two witnesses are related, genealogical coherence gives the editor more information about the relationship between the witnesses. Genealogical coherence proposes the flow of the texts. Text will either propose a text flow or suggest that there is no direct relationship with each

---

6 Gurry, A Critical Examination of the Coherence-Based Genealogical Method in New Testament Textual Criticism, 51
text. To complicate matters, genealogical coherence may also indicate that a text has multiple ancestors.

This ability to not only look at the relationship between witnesses but also to reconstruct their relationships gives the CBGM method an immense potential when it comes to studying the Old Testament. So how did this method come about?
METHODS

To accomplish the task of applying CBGM to the Old Testament there is a need to develop the software to assist in performing this analysis by aligning texts and then performing the iterative steps in the analysis with the help of an editor and preferably editors.

Logos Bible Software was used in this study to extract the facsimiles of the digitized manuscripts. One of the many benefits of using this software to extract digitized facsimiles is found in the ability to export selected chapters and verses.

Python and R programming languages were selected to create the software. Python performs all the upfront handling of the manuscripts that are exported from Logos Bible Software. At this stage the pointing is removed The Hebrew is transliterated into the Latin alphabet for further processing for use in the R programming language. R takes the manuscripts that have been processed by Python and performs the alignment. After the alignment is calculated the Hebrew characters are then restored. The coherence and genological relationship is determined based on the alignments.

Manuscript Alignment

To align manuscripts for CBGM analysis each manuscript is aligned in a pairwise manner with every other manuscript. For 19 manuscripts under consideration this would produce 361 different pairs of alignments to analyze. One of the many benefits of designing the alignment algorithm
from the ground up is the ability to take the steps of the alignment out of the black box to be able to visualize each step of the alignment.

When two manuscripts are aligned the words in each manuscript are compared with every other word in the other manuscript. If there is a match, then the word is given a score based on its length. Longer words have higher scores. If the words do not match, then the word is broken down into letters for further analysis. This allows the program to calculate the similarity of each word.

The final step determines the best alignment. This is achieved by working through the alignment matrix file from the end of each manuscript. The program works backwards and looks for the sequence matches. This is best visualized using an Excel spreadsheet.

**Pre-Genealogical Analysis**

After aligning the manuscripts with each other the pre-genealogical analysis may begin. The editor begins by determining which texts are related to one another. This is done by comparing all the texts with one another to calculate the average percent agreement with each manuscript. The average percent agreement for all texts being analyzed is calculated and then may be used as a cutoff to determine if two texts have a relationship with one another. The points of variation are then evaluated by the editor.

Emanuel Tov does not consider synonymous readings as being significant for further analysis.\(^7\) He does include Linguistic and Content variants as being significant for variant

---

\(^7\) Tov, *Textual Criticism of the Hebrew Bible*, 264.
The analysis of the variants by the editor is of primary importance as these decisions are used in determining the flow of the genealogical coherence of the text. Once the flow of a text is decided all the possible ancestors of a text are then considered.

**Coherence Based Genealogical Analysis**

With the pre-genealogical analysis complete, the genealogical analysis is ready to run. The pre-genealogical analysis may be adjusted by the editor at any time to incorporate new insights gained through the genealogical analysis. This is important as the method itself is constantly improving.

The program takes each pair of sequences that has a percentage agreement higher than the mean and then continues to look at the overall text flow for the sum of the local stemmata. For instance, if 9 variants, stemmata, are proposed by the editor to move from text A to text B while 1 variant is proposed to move from text B to text A, then there is a strong suggestion that text A is the ancestor and text B is the descendent. If, on the other hand, 4 passages are proposed to have come from text B while 6 from text A, then the determination of the flow of the text is not as confident.

---

RESULTS:

The most crucial component of the pre-genealogical analysis is the alignment of the manuscripts under consideration. A solid and reliable alignment will produce a solid and reliable CBGM output. Figure 1 shows two aligned texts from Deuteronomy 5:1-10, with word matches colored from red, indicating a weak word match, to green, indicating a strong word match. A quick glance over the alignment matrix reveals a diagonal line. This line is indicative of the best text alignment match.

![Figure 1. Example of Text Alignment between BHS and LXX.](image)

The alignment matrix is then used to determine the best possible word matches. The result is an alignment between two texts. An example is seen Table 1. The consensus text represents the relationship between the texts. The consensus text has a ‘+’ if the two words are an exact match, a ‘~’ if the words do not match, and a ‘M’ if the words are 75% similar. The
agreement criteria, currently set at 75%, is useful for detecting words with Plene and Defective spellings.

<table>
<thead>
<tr>
<th>BHS</th>
<th>LXX</th>
<th>Consensus Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>בְּרֵאשִׁית</td>
<td>הָיוּ</td>
<td>+ + ~ ~ + + + ~ ~ ~</td>
</tr>
<tr>
<td>לְאָדוֹן</td>
<td>שָׁלוֹשׁ</td>
<td>~ ~ + + + ~ ~ ~</td>
</tr>
</tbody>
</table>

*Table 1. Example of an alignment between BHS and LXX.*

Following the alignment of the manuscripts the pre-genealogical and genealogical coherence can be calculated. Table 2 shows the results from the CBGM run. The first two columns list the two manuscripts being compared. The third column shows the percentage match between the two manuscripts. The average percent agreement for all manuscripts analyzed was 23%. The standard deviation was 0.21, indicating that 95% of the results ranged from 23% to 44% agreement. The average percent agreement was used to set the criteria a genealogical relationship. The fourth column lists the number of word matches. The low number of word matches is heavily dependent on how fragmented the text is. The fifth column lists the shortest manuscript in the analysis. The sixth, seventh and eighth columns show the genealogical coherence between the two manuscripts. The Ancestor (A) column shows how many variants suggested that the first manuscripts variants gave rise to the second manuscripts variants. The Descendant (D) column shows the number of variants where the second manuscript is thought to have given rise to the first manuscript. The Undetermined (U) column lists the number of variants that could not be classified as being either an Ancestor or Descendent.
Table 2. Genealogical Coherence Parameters.

A visual representation of the results in Table 2 are seen in Figure 2. This representation is referred to as a text flow diagram. The line weights indicate the strength of the relationship between manuscripts. The LXX and BHS have the strongest relationship. The arrow indicates that the text in the LXX preceded the text in the BHS. A dotted line indicates that the relationship between the manuscripts is weak. The relationship between the BHS and Qphyl J manuscript is week and no coherence, text flow, could be determined with the current dataset. The text in Qphyl B, A, and L appear to have been preceded by the text of the BHS. The following point must be reemphasized. The assumption is that an ancient text may very well be present in a manuscript that was recently copied. For instance, a late manuscript, like
Leningradensis B19A, can embody an early form of the text, and vice versa, if “late” means “centuries removed from the author”. The LXX and BHS appear to be more closely related to each other than to the Samaritan Pentateuch.

*Figure 2. Text Flow Diagram showing the overall relationship between the texts.*
DISCUSSION:

The main components relating to CBGM that will be discussed includes: the current state of CBGM analysis of Old Testament texts, current challenges to implementing CBGM in the Old Testament, strengths and weaknesses inherent in this method and recommendations for future research.

**The Current State of CBGM Analysis of Old Testament texts**

So far, we can align Old Testament texts using a new alignment algorithm. This allows the editor to determine pre-genealogical coherence as well as genealogical coherence. Based on the genealogical coherence the editor is then able to create a text flow diagram that proposes the overall transmission of a text from manuscript to manuscript. This will provide a novel way to assess the current groupings of the manuscripts.

**Current Challenges to Implementing CBGM in the Old Testament**

What is the best way to handle fragmented texts? The algorithm attributes longer variants as ancestors and gaps in the text as indicators of being descendants. There are many situations where these rules are not valid. If the gaps in a fragmented text are caused by the integrity of the manuscript, then this rule will not apply.

The current algorithm analyzes manuscript pairs. An improvement to the alignment algorithm will include aligning all the manuscripts together to get a master alignment. This will produce a *corpus* of text would then allow the software algorithm to distinguish between variants
that differ between two manuscripts and variants that exist in the corpus of text but are similar between the two manuscripts being analyzed.

The current algorithm will also benefit from applying the principle of parsimony to the texts. This is not a major focus as the number of manuscripts available for analysis is limited to begin with.

**Strengths and Weaknesses Inherent in this Method**

Strengths include the ability to align Old Testament texts and determine their relationship with one another. Another strength of this method includes the use of two parameters, agreement and a sum of the variants supporting an ancestor or descendent role, to evaluate the relationship between the manuscript and ultimately the text. The fact that dating information and geographical information are excluded give further insight into the textual relationship and not the manuscript relationship.

Weaknesses in the method includes the calculation points that introduce subjectivity into the method. One area where subjectivity may bias the results includes the editor’s input in determining whether each variant is an ancestor or descendent. To mitigate the first weakness, a more thorough statistical model may be sought to draw the line between related and unrelated manuscripts.

As a new method in Old Testament textual analysis the software that has been developed for this project also will have many opportunities to be optimized and grow. As the dataset grows further software changes may be needed to adapt to new discoveries.

Strengths and weaknesses considered, the CBGM has tremendous potential in the study of Old Testament manuscripts. It is able to combine the questions generated by the editor with
computer assisted software to take a meta-genomic look at Old Testament texts. This will in turn give the editor a high-level picture of the relationship between texts.

BIBLIOGRAPHY


