TC: A Journal of Biblical Textual Criticism 28 (2023): 111–126

The First Computer-Generated Greek New Testament

Alan Bunning, Center for New Testament Restoration

Abstract: A plausible Greek New Testament text can be automatically generated by a computer program using statistical analysis and algorithms that weigh the earliest manuscript data in a manner simulating a reasoned-eclecticism approach. This method offers several substantial advantages by providing a consistently weighed text that is openly transparent, without any theological bias, and scientifically reproducible, and the results are very similar to our best modern critical tests. This initial accomplishment could have a number of future implications for the field of textual criticism regarding advances in the use of statistics and algorithms for further refinements in the production of critical texts.

Introduction

It has often been said that textual criticism is both an art and a science.[1] The unfortunate reality, however, is that the process has mostly been dominated by the art part. One group of scholars will examine all of the variant readings for a particular passage and then choose the reading that they think best explains how the other readings may have occurred.[2] But the problem is that another group of scholars does the exact same thing, and they come to a completely different conclusion. Consequently, there has long been a desire to increase the scientific aspects of textual criticism. Text-critical canons such as Bengel's twenty-seven principles and Griesbach's fifteen rules could perhaps be considered an early forerunner to this sentiment, providing a set of guidelines based on assumed probabilities to guide the selection of variant readings in a more logical fashion.[3] Likewise, the genealogical method often associated with Karl Lachmann back in the nineteenth century "originated from the need to base reconstruction on scientific and objective criteria, reducing as far as possible the subjectivity of the editors."[4] Scholars of the twentieth century such as Dom Henri Quentin, Sir Walter W. Greg, Archibald A. Hill, and Vinton A. Dearing considered several statistical approaches to textual criticism, but they were

[1] This well-known mantra was presumably derived from A. E. Housman's quote: "Textual criticism is a science, and, since it comprises recension and emendation, it is also an art. It is the science of discovering error in texts and the art of removing it." A. E. Housman, "The Application of Thought to Textual Criticism," *Proceedings of the Classical Association* 18 (1921): 68.

[2] Johann Jakob Griesbach has been credited with the rule followed by many textual critics: "The reading is to be preferred as the original which best explains the existence of all other." Eldon J. Epp and Gordon D. Fee, *Studies in the Theory and Method of New Testament Textual Criticism* (Grand Rapids: Eerdmans, 1993), 181.

[3] Johann Albrecht Bengel, *Gnomon Novi Testamenti* (Tubingen: Johann Heinrich Philipp Schramm, 1742); Johann Jakob Griesbach, *Novum Testamentum Græce, Textum ad fidem Codicum Versionem* (London: Halae Saxonum, 1796).

[4] Paolo Chiesa, "Principles and Practice," in *Handbook of Stemmatology: History, Methodology, Digital Approaches*, ed. Philipp Roelli (Berlin: de Gruyter, 2020), 86.

fairly limited in scope without the aid of a computer.[5] There have also been many other types of statistical analysis providing a more objective basis for understanding scribal habits and comparing variant units in manuscripts. Unfortunately, most of these efforts have had to be done by hand, using only a few select manuscripts over relatively small passages of Scripture as a sample size, from which the rest could then be extrapolated. E. C. Colwell and E. W. Tune foresaw the need for computers to get involved in textual criticism way back in the 1960s: "We are working in a period when the data for textual criticism will inevitably be translated into mathematics. In fact it is doubtful that NT textual critics can really hope to relate all of the data now available to them without the aid of computers."[6]

There have since been several examples of computer-assisted research over the decades in fulfillment of this sentiment, such as the Coherence-Based Genealogical Method (CBGM) developed by Gerd Mink and the cladistics approach used by Stephen Carlson for the book of Galatians.[7] But despite a popular misunderstanding, techniques like the CBGM do not "provide a means of automating the reconstruction of the initial text," as they are merely considered to be tools to help in the subjective decision-making process.[8] Part of the reason for this is due to the significant amount of genealogical corruption in the data. Many of the earliest witnesses are clearly seen to be doing their own textual criticism, copying from multiple witnesses already available to them. But despite some of its shortcomings,[9] the work of the CBGM was particularly valuable in the sense that this work had to be done in order to know that this was the case, demonstrating that most of the earliest witnesses do not have direct genealogical relationships to each other.

Even with these technological advances, the crux of the matter is that textual criticism has still been largely treated as an art, with scholars viewing scientific statistical analysis as merely suggestions to help guide their subjective decisions.[10] That is why some of our best modern critical texts, even those with similar philosophies considering the same evidence, still disagree with each other in thousands of places.

Computer-Generated Text

The ultimate result of applying science to textual criticism was envisioned years ago in the automatic creation of a computer-generated text without any human subjectivity. Yet despite our best efforts we were "nowhere near having computer tools that can algorithmically produce

5 Bruce Metzger, *The Text of the New Testament*, 3rd ed. (Oxford: Oxford University Press, 1992), 163–69.

6 E. C. Colwell and E. W. Tune, "Variant Readings: Classification and Use," *JBL* 83 (1964): 255–56.

7 Gerd Mink, "The Coherence-Based Genealogical Method—What Is It About?" (online paper, Münster: Institut für Neutestamentliche Textforschung, 2002), https://www.uni-muenster.de/INTF/Genealogical_method.html; Stephen C. Carlson, "The Text of Galatians and Its History" (PhD diss., Duke University Graduate Program in Religion, 2012).

8 Klaus Wachtel, "Towards a Redefinition of External Criteria: The Role of Coherence in Assessing the Origin of Variants," in *Textual Variation: Theological and Social Tendencies? Papers from the Fifth Birmingham Colloquium on the Textual Criticism of the New Testament*, ed. David C. Parker (Piscataway, NJ: Gorgias, 2008), 127.

9 Stephen C. Carlson, "A Bias at the Heart of the Coherence-Based Genealogical Method (CBGM)," *JBL* 139 (2020): 319–40; Jarrett W. Knight, "Reading between the Lines: 1 Peter 4:16, MS 424, and Some Methodological Blind Spots in the CBGM," *JBL* 138 (2019): 899–921.

10 For example, the results of the CGBM were not followed by the Nestle-Aland 28th edition editorial committee at 2 Pet 3:10 because the CGBM does not make up conjectures. Instead, the committee made up their own reading, which is not supported by any Greek manuscript.

a stemma and a critical text from a bundle of scanned manuscripts."[11] But that is no longer the case. The Statistical Restoration (SR) represents the first computer-generated Greek New Testament. All the earliest manuscript evidence is fed directly to a computer program as raw data, and the *most probable text* is generated based on statistical analysis and algorithms. The SR was created according to the principles of Scientific Textual Criticism, which represents a fundamental paradigm shift from the traditional methods of textual criticism. Subjective textual decisions are replaced with objective statistical and computational methods, rooted in the fields of data science and computer science. The motivation, rationale, limitations, and implications for this approach are described in *Restoration of the New Testament: The Case for Scientific Textual Criticism*;[12] discussion and answers to common objections will not be repeated here. Instead, this paper will focus on the details of how the SR was created, and it is merely the first example of a critical text meeting the scientific criteria of objectivity, plausibility, transparency, and reproducibility outlined in that book. Computers have been used before for different aspects of text criticism related to the Greek New Testament, but the SR endeavored to reflect the most probable text based on data-driven processes that were designed to simulate *a reasoned-eclecticism approach actually used by scholars*, weighing both external and internal evidence. Accordingly, the SR serves as a proof-of-concept demonstrating that a plausible computer-generated text can be produced that yields a satisfying result when compared to our best modern critical texts.

From its initial conceptualization, the SR took almost two decades to complete. It began innocently enough with the creation of the Scientific Greek New Testament Interlinear (SGNTI) project in 2003. That project's goal was to provide a computer-generated collation of the earliest Greek manuscripts in an interlinear format. Using the original electronic transcriptions created over the life of that project, the Bunning Heuristic Prototype (BHP) Greek New Testament was created by hand in November 2012 as a preliminary template to approximate the results of what could foreseeably be produced in a computer-generated Greek New Testament.[13] This was done for the purpose of anticipating what types of problems might be encountered in writing such a computer program. Soon after, these initiatives were absorbed into the Center for New Testament Restoration (CNTR),[14] established in 2013. The CNTR's charter was to apply advanced computational and statistical methods, rooted in the fields of data science and computer science, to the field of textual criticism.

The basis of an algorithm for a computer-generated text was first discussed in the CNTR Project Description in 2016 and later updated to include a basic formula in 2018. Using that formula as a starting point, the first version of the program was written. This resulted in the first computer-generated Greek New Testament on 1 October 2020. The formula underwent a number of successive iterations using a data-driven approach until it arrived at the current algorithm. A number of technical breakthroughs had to occur along the way in order to accomplish this feat, including the automatic determination of variant unit boundaries and their

---

11 Philipp Roelli, *Handbook of Stemmatology: History, Methodology, Digital Approaches* (Berlin: de Gruyter, 2020), 5.

12 Alan Bunning, *Restoration of the New Testament: The Case for Scientific Textual Criticism* (West Lafayette, IN: Biblical Worldview Publishing, 2022).

13 The BHP is an open-licensed Greek New Testament that is currently used as the basis for the unfoldingWord Greek New Testament (UGNT) to assist in Church-Centric Bible Translation (CCBT). This text was later released in 2017 and was used to create the unfoldingWord Greek New Testament (UGNT), which has since been translated into a number of other languages. See https://github.com/Center-for-New-Testament-Restoration/BHP.

14 http://greekcntr.org.

relationships to each other, the classification of homophones based on the orthographical-priority method, and rating the statistical reliability of manuscripts against the corpus of data.[15] Based on those innovations, the beta version of the text was released on 20 September 2021. After a few more minor tweaks, the final version of the text was released on 17 October 2022 and called the "Statistical Restoration Greek New Testament."

The SR offers several improvements compared to most other modern critical texts:

- The SR replaces the subjective theological bias of human editors with the use of *objective* statistical and computational methods. The meaning of words was not considered when making textual decisions. Instead, external and internal evidence was objectively weighed. As a result, the SR provides a plausible text built on a statistical scientific method.
- The SR is based on all the early extant manuscripts dated before 400 CE, including all the continuous-text manuscripts, as well as quotations from amulets, inscriptions, and other writings.[16] This data was not readily available as a complete dataset until the creation of the CNTR collation.[17] Since the SR only considers extant evidence, it does not contain any conjectural emendations that are found in some other critical texts. Only *actual readings* found in manuscripts were considered.
- The SR weighs the manuscript data in a *consistent* manner that is not possible by human editors. The computer can accurately process complex statistical relationships that cannot be kept track of or discerned by human intuition. The computer can make the exact same decisions when given the same conditions, whereas humans are often swayed by unconscious biases and may not remember what they did on previous occasions.
- The SR was built on processes that are openly inspectable, verifiable, and reproducible, which provides a *transparent* basis for its evaluation. When combined with the CNTR collation, each textual decision can be publicly scrutinized and judged based on its own merits. The probability of each word is displayed along with the data that it was directly derived from, which can be drilled down all the way to the actual manuscripts themselves.
- The SR can be updated immediately whenever new manuscript evidence is found or new assessments are given to the existing manuscripts. It does not take years to assemble a committee, painstakingly go through all the manuscript evidence by hand, and then vote on each variant reading. The SR can be regenerated in less than a minute reflecting all of the latest evidence. It can also be reprogrammed to try out new theories or provide other analyses, giving *immediate* feedback with very little associated cost.
- The SR comes with *both* Koine Greek orthography representative of the early manuscripts and the traditional modern orthography, including accents, capitalization,

15 Alan Bunning, "Scientific Definition of Variant Unit Boundaries" (paper presented at 2022 Annual Midwest Regional Meeting of the Society for Biblical Literature, virtual, 5 February 2022); Bunning, "Orthographic Priority for Interpreting Homophones in New Testament Manuscripts" (paper presented at 2021 Annual Meeting of the Society of Biblical Literature, San Antonio, TX, 22 November 2021); Bunning, "Corpus-Based Statistical Measurements of Textual Reliability for New Testament Manuscripts" (paper presented at 2022 Annual Midwest Regional Meeting of the Society for Biblical Literature, virtual, 5 February 2022).

16 Bunning, *Restoration of the New Testament*, §1.2.1.3.

17 https://greekcntr.org/collation/index.htm.

and punctuation. There are several places where every early manuscript is in agreement with how a word is spelled, which is different from the canonical spellings shown in most modern critical texts and lexicons. The Koine Greek orthography also includes *nomina sacra*[18] that presumably indicate the deity but are not included in other critical texts.

- The SR comes complete with several additional electronic resources, including Enhanced Strong Numbers (ESN), morphological parsing, and English context-sensitive glosses developed by the CNTR. Such resources normally have to be manually added later when a critical text is released, but they are generated *automatically* with the SR text because they are already encoded in the CNTR database for every possible variant that could be chosen.
- The SR has been publicly released under open-source licenses,[19] which will allow others to build on the work and contribute other improvements to serve the needs of the global church. The text is released under the Creative Commons Attribution 4.0 International License (CC BY 4.0), and the source code is released under the GNU General Public License 3.0 (GPLv3). This is particularly significant in that it satisfies the need to provide an open-licensed modern critical text based on the early manuscript evidence, with a process that is fully *accessible* to the public.

The SR text is released in several different data formats including Unified Standard Format Markers (USFM), Tab Separated Values (TSV), and Manuscript Encoding Specification (MES). More detailed information about the specific fields can be found in the CNTR Technical Reference.[20]

Infrastructure

The generation of the SR relies on the infrastructure of the CNTR relational database and a series of computer programs that were specifically designed for textual criticism. The CNTR database was created from scratch from original electronic manuscript transcriptions and currently contains over 1.5 million words with data from 201 early witnesses. This dataset contains all the most important variant readings in the New Testament, including all of the earliest Greek witnesses from extant manuscripts up to 400 CE, both continuous texts (class 1 data) and other Scripture quotations (class 2 data), as well as several major critical texts that were included for reference purposes.[21] This transcription data is relationally tied to metadata, lexical, morphological, syntactical, and other forms of data, which enables advanced data analysis that has never before been possible. For example, the painstaking counting of certain scribal habits that used to be done by hand can now be completed in seconds by a single database query.

In addition to this data, the CNTR database provides several advanced features for textual criticism not available in any other computer platform. First, the CNTR database contains collation alignment data, which provides an easy and consistent way to compare texts regardless of orthographical differences. The CNTR collation alignment is based on distinct lexical/

18 *Nomina sacra* is Latin for "sacred names" and was a scribal practice where frequently occurring divine names were often represented by an abbreviation of two or more overlined letters.

19 https://github.com/Center-for-New-Testament-Restoration/SR.

20 Alan Bunning, "CNTR Technical Reference," Center for New Testament Restoration, 1 June 2022, https://greekcntr.org/resources/technical.pdf.

21 1885 Westcott and Hort (WH), 2012 Nestle-Aland 28th edition (NA), 2010 Society of Biblical Literature (SBL), 2018 Robinson/Pierpont, 2014 King James Textus Receptus (KJTR), and 1550 Stephanus (ST).

morphological/phonological word forms, which compares words phonetically according to a standard set of rules governing phonemes, while ignoring other orthographical differences such as elision, movable *nu* or *sigma*, *nomina sacra*, and other abbreviations. The collation alignment data was generated without reference to any base text by using two different algorithms—a maximum text was created as a template containing all known variants for each verse using a recursive longest common sequence first algorithm, and then each witness was aligned to this template using a nonrecursive longest common sequence algorithm considering multiple sequences. Second, the CNTR database contains fields that mark objective boundaries of the variant units. Two sets of boundaries were established based on whether variant words were partially dependent or fully independent of each other as determined by a complex computer algorithm. These boundaries take into account words supplied in lacuna and identified by *vid*, and homophones that are interpreted by an orthographical-priority approach.[22] Third, the CNTR database contains statistical information such as the statistical reliability of witnesses compared to the entire corpus and the textual affinity between witnesses based on their variant readings, which are discussed in more detail below. More details about the CNTR database can be found in the CNTR Technical Reference.[23]

From start to finish, the entire process to create the computer-generated text utilized several different programs that were implemented in stages so that the results could be checked after each step. These programs were all written in JavaScript using Structured Query Language (SQL) to query the CNTR relational database:

1. Collation alignment algorithm
2. Variant pattern identification
3. Orthographical probability algorithm
4. Variant unit boundaries algorithm
5. Textual reliability and textual affinity statistics
6. Computer-generated text algorithm

With some extra work, there would be nothing preventing all of these programs from being combined into one turnkey solution, thus achieving the holy grail of scientific textual criticism, where all of the electronic transcriptions are fed into one program and it automatically recreates the original autographs without human intervention. As it is, the results of the first two computer-assisted steps were slightly tweaked by hand, which otherwise could have been accomplished by additional processing. But what is being emphasized in this paper, is the final algorithm, which creates the computer-generated text from a static infrastructure, requiring no human intervention in the decision-making process.[24] The infrastructure itself is not predisposed to any particular outcome.

Algorithm

The algorithm behind the SR is modelled on a form of reasoned eclecticism that attempts to approximate the thought process of modern textual critics by use of a computer program. Reasoned eclecticism is the normal method used by scholars for reconstructing the reading of

22 Bunning, "Orthographic Priority." *Vid* is an abbreviation for the Latin word "videtur," which means "as it seems." It indicates that there is sufficient evidence to support a variant reading that was missing in a manuscript.

23 Bunning, "CNTR Technical Reference."

24 Obviously, humans had to be involved in creating the infrastructure itself by setting up the database schema, loading the data, running the programs, etc.

an original text by considering both external and internal evidence based on the compilation of multiple sources. The eclectic approach used by textual critics today is perhaps the most scientific approach when considering the nature of errors, since it recognizes that scribes do not always make the same mistakes in all the same places. An error can occur anywhere by anyone and be passed down by anyone. The eclectic approach is well suited to winnowing out these errors. Indeed, a common textual criticism exercise repeated in classrooms every year demonstrates to students how errant and mutilated copies of a text can be used to accurately reconstruct the original text using the eclectic method.[25]

By utilizing an eclectic approach, someone might think that this algorithmic approach may be more likely to result in an artificial text that "rapidly degenerates into one possessing no support among manuscript, versional, or patristic witnesses."[26] However, the SR text does at least have manuscript support because the text was generated directly from the early manuscript evidence. The fact of the matter is that *all* the major critical texts are eclectic texts to various degrees, regardless of whether they favor Alexandrian or Byzantine readings, since none of the major critical texts are merely copies of an existing manuscript. Indeed, even the early scribes can be seen doing their own textual criticism as they cross through words or change them to other variant readings.[27]

As previously mentioned, the initial idea for the algorithm underwent many improvements through successive iterations using a *data-driven approach* until arriving at the current algorithm. Several kinds of algorithms were tried in numerous configurations, which produced slightly different texts, but most were all in the same ballpark with no more than a few hundred words different. The goal was to find an algorithm that would best approximate the textual critics' use of reasoned eclecticism, but only with regard to objective internal and external evidence. The biggest challenge in designing the algorithm was trying to program the computer to systematically do what a textual critical would naturally do by intuition.

The resulting algorithm weighs each variant reading within a variant unit by considering a combination of internal and external evidence. The *external evidence* is a major component of the algorithm, which is weighed by the following formula:

$$\mathit{external}(r) \propto (\mathit{reliability}(r) \cdot c_1) + (\mathit{earliness}(r) \cdot c_2) + (\mathit{support}(r) \cdot c_3)$$

A variant reading (r) is evaluated and the constants (c_1, c_2, c_3) are used to weigh the relative importance of each component. The final constants used were 1.22, 1, and .7 respectively. This formula is meant to simulate the considerations given by textual critics where readings that are earlier, more reliable, and have more support (not by counting copies but by statistical diversity) are given more weight. But instead of subjective impressions, these variables are weighed with precise statistical accuracy based on objective criteria:

---

[25] One example of this was conducted by Ryan Haines with The Gospel Training Ground, "Textual Criticism Experiment: Final Results!," 24 August 2018, https://www.youtube.com/watch?v=ht8bRQWhfQXw. The results of the experiment were that there was "not one single difference in the wording" but only minor differences in punctuation, capitalization, and paragraph breaks; none of which "changed the meaning or the wording or what was written."

[26] Maurice A. Robinson, "New Testament Textual Criticism: The Case for Byzantine Priority," *TC* 6 (2001), http://jbtc.org/v06/Robinson2001.html.

[27] Bunning, *Restoration of the New Testament*, §1.2.3.2.

- *reliability*: The statistical reliability of each witness is based on its relative statistical relationship to the entire corpus of data.

  $$reliability(r) \propto \sum_{w \in r} rating(w)^{c4}$$

  The reliability rating for a witness (w) is analogous to sports indexes, which provide relative power rankings based on win-loss records, strength of schedule, and margins of victory. No one can prove that one team is better than another team, but there is an objective way to statistically rank the teams based on their body of work. The same is true of manuscripts whose ratings were determined based on four different measurements of their singular readings against the entire corpus, resulting in an overall reliability rating for each manuscript.[28] The *rating* of each witness was computed by a separate program, recorded in the CNTR database, and then retrieved for this calculation. A constant ($c4$) of 9.08 was used to stretch the numbers in scale, giving greater weight to the readings supported by the more reliable manuscripts. The resulting value for each reading was scaled as a percentage in proportion to ($\propto$) the total amount of all readings.

- *earliness*: The date ranges of the witnesses were determined by experts in the field using standard practices from paleography, occasionally aided by other techniques.

  $$earliness(r) \propto \sum_{w \in r} \left(maxDate - \frac{\frac{date1 + date2}{2}}{maxDate - minDate}\right)^{c5}$$

  The average of the date range (shown by the $date1 + date2 / 2$) is used for each witness (w) and adjusted as a percentage between the earliest and latest date of all of the witnesses. While an early manuscript is not necessarily more accurate than a later manuscript, it provides *prima facie* evidence of when a reading existed in time. Logically speaking, later manuscripts that could possibly be copies of earlier manuscripts do not have the same weight as those that definitely *are* early manuscripts. The date is important when combined with the support function, for if a later reading has no support from an early manuscript, then it is at least suspect because it could have been made up centuries later. If it is merely a copy of other early manuscripts, then its vote is less useful. A constant ($c5$) of 1.16 was used to slightly stretch the numbers in scale giving greater weight to the readings supported by the earlier manuscripts. The resulting value for each reading was scaled as a percentage in proportion to ($\propto$) the total amount of all readings.

- *support*: The statistical textual affinity between witnesses ($w1$, $w2$) is used to provide an indication of how well a reading is attested based on its relative diversity.

  $$support(r) \propto max(\forall(w_1, w_2) diversity(r, w_1, w_2)^{c6}) + (1 - min(\forall(w_1, w_2) affinity(r, w_1, w_2)^{c6}))$$

  This part of the algorithm is rather nonintuitive and difficult to explain, but it seeks to determine the level of support for a variant reading while eliminating statistical redundancy between its witnesses. The textual *affinity* ($\forall$) for each set of witnesses was computed by a separate program that compared the variant readings of each

[28] Bunning, "Corpus-Based Statistical Measurements."

> witness to every other witness (similar to pregenealogical coherence of the CBGM) on a book-by-book basis and recorded in the CNTR database, and then retrieved for this calculation. Obviously, there is no point in simply counting the number of manuscripts that support a reading, for the number of copies made from copies does not make the reading more correct! Instead, each reading receives proportionate weight based on the maximum diversity between its witnesses (plus the opposite of the minimum affinity between its witnesses). If witnesses often disagree with each other but agree on a reading, that gives the reading greater significance. But if the witnesses are merely close copies of each other, the number of additional copies that bear witness to a reading does not receive much weight. This avoids the CBGM's problem of genealogical corruption among the early manuscripts because it gives the proper weighted percentage of diversity to a witness *regardless of which direction the copying may have occurred*. In the extreme case that all of the witnesses were exact copies of each other, the support would be zero (representing the same genealogical branch), and if all of the witnesses did not agree on anything other than the one variant, the support would be nearly 100 percent (representing a completely different genealogical branch). Of course, neither extreme is found among the earliest data, so all witnesses fall on a continuum somewhere in between. In other words, the more diverse the witnesses, the more weight they are given when they agree. However, having a large number of witnesses with similar reading can still receive some weight because of the accumulation of small values. A constant (*c*6) of 0.16 was used to slightly stretch the numbers in scale, giving greater weight to the readings with the most diversity of support. The resulting value for each reading was scaled as a percentage in proportion to (∝) the total amount of all readings.

Each of these three external factors (reliability, earliness, and support) was expressed as a fraction between 0 and 1. After they are combined together in the formula, the resulting value for each reading is scaled as a percentage in proportion to (∝) the total amount for all readings. Thus, each reading is given a final value between 0 and 1, representing its overall percentage of likelihood. A rating of 1 would mean that a reading has perfect reliability, the earliest possible date, and the greatest diversity of support, but that, of course, never occurs.

The algorithm does not consider external evidence alone, but also includes a hybrid form of *internal/external evidence* that considers the internal probability that each particular word belongs within a variant unit in relation to its external evidence. For example, depending on how the collation is arranged, the word κωφους could be found at a lot of different positions in this variant unit at Matt 15:30:

…			χωλους				κυλλους	τυφλους	κωφους	…
…			χωλους		κωφους	τυφλους	κυλλους			…
…			χωλους	τυφλους			κυλλους			…
…	κωφους		χωλους	τυφλους			κυλλους			…
…			χωλους	τυφλους			κυλλους		κωφους	…
…	κωφους	τυφλους	χωλους				κυλλους			…
…			χωλους	τυφλους	κωφους		κυλλους			…

Table 1: Variant Unit at Matthew 15:30

While κωφους may not be a compelling choice at any single position, the word obviously should be included, and thus the algorithm ensures that the chosen reading includes that word somewhere because of its frequent occurrence. Rather than simply rely on the word frequency across all the readings in the variant unit, the hybrid aspect is that the word is also weighted in proportion to its external evidence. This essentially simulates a textual criticism process

$$internal(r) \propto \sum_{w \in r} external(t)$$

that asks, How often does a word appear across the variant readings, and how weighty are the witnesses in which it appears? Thus, given the words (t) at each position within a variant unit, each variant reading (r) is rated based on the accumulated external probabilities for the words that it contains, as represented by the following formula:

Here, the same external evidence formula is used as before, but this time, the individual words (t) of the variant readings are fed into it, instead of variant readings as a whole (r). This is then aggregated across each variant reading according to the word probability. In other words, a witness list is created for each word position in a variant unit (regardless of word order), then the words at each position are individually weighed and accumulated by the external evidence based on their witnesses, and finally each variant reading is evaluated *as a whole* according to the accumulated weight of the words that it contains. Although the words are weighed individually within each variant reading, the selected variant reading is chosen as a whole, and thus the text does not contain any "Frankenstein monster" variant readings cobbled together from various words that never previously existed together as a unit. In essence, the algorithm assesses the probability of whether each word belongs there and, if so, which words and in which position (in the case of word order differences).

Another more traditional type of internal evidence involves variant patterns. The CNTR database records the type of variant patterns in variant units, such as conflations, homeoarcton, homeoteleuton, et cetera. Such information can then be leveraged to make more precise selections based on the different situations. For example, in the SR, readings that were singular conflations were simply eliminated from consideration, being considered to be extremely unlikely candidates. Such situations could have been dealt with in a wide variety of ways depending on the circumstance, but the SR included this one situation to demonstrate the potential future development of this category of processing.

The variant reading (r) with the highest overall score based on the internal evidence (which also considers external evidence) was selected for each variant unit (u). If there is a tie (which can occur if there are only word order differences), the algorithm breaks the tie by selecting the variant reading that has the highest external evidence alone. The portions of the New Testament that do not contain any variant readings are automatically included in the resulting text and thus are not processed by the algorithm.[29]

When all of these components are combined, the obvious question is: how much and to what degree should each factor be applied? Textual critics intuitively weigh factors like earliness, reliability, and diversity of witnesses together, but they do not have precise values for them, nor do they do so consistently. As you can see, the resulting algorithm contains several constants impacting both the weighting ($c1$, $c2$, $c3$) and scaling ($c4$, $c5$, $c6$) of these components. The selection of different sets of constants would obviously result in different texts being produced. But if given the additional goal to make a reasonable text that closely matches our best

[29] Approximately 76 percent of the words of the Greek New Testament are not involved in a variant unit among all the manuscripts found in the CNTR database.

modern critical texts, it turns the selection of these constants into a purely scientific endeavor, similar to approaching the curve of an asymptote. Thus, the constants in the algorithm were computationally calibrated by varying every combination of the values in batches of one thousand runs at a time, each time comparing the result to a target text and selecting the values that produced the closest resulting text. It actually mattered very little if the Nestle-Aland, Society of Biblical Literature, or Tyndale House critical text were chosen as the general target, for they are all close enough that the variance is fairly insignificant. If the algorithm approached the asymptote of one of those texts, it would simultaneously approach the asymptote for the other texts (but it would never be possible to exactly match any of them). Thus, the resulting computer-generated text also falls within that same range of variance, for all of them are in the same general ballpark compared to other critical texts. This first release of the SR was calibrated to the BHP critical text mentioned above, which had been created precisely for this purpose.[30]

Orthography

The orthography of the SR is provided in both Koine Greek and Medieval Greek forms. The Koine Greek orthography does not contain any accents, punctuation, or capitalization and contains the majority spellings of words expressed at each location in the early manuscripts, taking into account how words are spelled in manuscripts that are missing from a location. *Nomina sacra* are specified for words in the locations where there is unanimous consent among the manuscripts that include this feature. The Medieval Greek orthography contains accents, punctuation, or capitalization and canonical spellings that emerged later in the Middle Ages, which is the form shown in most modern critical texts and lexicons. The accents, punctuation, capitalization, and spelling were initially seeded by a computer-assisted process that looked for commonality across several different critical texts and applied various metrics for where there were disagreements. All of these were later manually adjusted by hand as needed.

Data Limitations

The CNTR database currently contains only class 1 and class 2 data, which is the best data currently available electronically, but it is certainly not all the data. That data also predominantly reflects only one geographical region (Egypt). While many Byzantine readings are included, the resulting text tends to be more Alexandrian in nature, just like our best modern critical texts, which heavily weigh the significance of the earliest extant manuscripts. The algorithm is actually oblivious to any text-type theories but simply retrieves the data and processes it with blind statistical analysis. It is simply more rational to process the actual evidence that we have than to rely on theories based on evidence that we do not have. The large volume of information from class 3 data (church father quotations) and class 4 data (foreign versions) has not been fully utilized in the field of textual criticism, and its future addition to the CNTR database should greatly improve the statistical accuracy of the computer-generated text.[31] While class 3 and class 4 data is generally of lesser value, it still contains many early readings from multiple

30 It should be noted that the BHP is also within the same range of variance, which in itself is only about five hundred words different from the Nestle-Aland text (depending on how one counts word differences).

31 Bunning, "CNTR Technical Reference," §2.2.

geographical regions that are important for understanding the nature of the original text and its transmission.

There are some places where the statistical difference in selecting one reading versus another is very close (e.g., 51 percent to 49 percent). While that is adequate for indicating a preference, it is inadequate for establishing any level of confidence in such readings. This, however, is no different from the number of close calls that are made among textual critics, especially since textual critics often disagree with each other. It is expected that future inclusion of additional data will help minimize the number of such close calls. But in any case, the CNTR displays the statistical calculations made for each variant reading online, which makes the process both open and inspectable. This is an improvement over critical texts which were created behind closed doors with no stated justifications.

In acknowledgement of these limitations, the algorithm has been equipped with additional *expert-assist* and *expert-override* options that can independently be turned on or off as desired. These options allow a reading selected by the algorithm to be changed or influenced based on readings from the major critical texts as a type of safety net. When either of these options are turned on, there is an option to have the alternate readings placed in single square brackets to indicate areas of possible concern.

When the expert-assist option is turned on, if there are only two early witnesses and they disagree with each other, the support of the major critical texts that match either of the two early readings is also considered. The algorithm would otherwise perform poorly in such cases, because unless the internal evidence is a deciding factor, the external evidence would always weigh one manuscript above the other (i.e., one manuscript would always be earlier, more reliable, etc. than the other). The algorithm essentially needs at least three witnesses for best results, and this feature is thus considered to be essential and is always turned on. When the expert-assist feature is applied, it currently affects about 213 words. As more data is added, this number will automatically be reduced.

When the expert-override option is turned on, the reading selected by the algorithm is replaced with the unanimous consensus of all of the critical texts. The proposition here is that, if the major critical texts ranging in diversity from Westcott and Hort to the King James Textus Receptus unanimously agree on a reading that is different from the one selected by the algorithm, then that is surely worthy of consideration. If a reading is chosen with the expert-override option, that does not necessarily mean that the reading chosen by the algorithm was wrong, for the readings it chooses are always plausible, being directly derivable from the earliest manuscript evidence. For example, at Eph 1:1 all of the major critical texts include the reading εν εφεσω (although some texts include the phrase in brackets, indicating some uncertainty), but the reading would have been left out of the SR. The inclusion of that reading in the other critical texts is arguably an example of harmonization where the words were later added to provide symmetry with the other Pauline epistles, but it was clearly missing from the three earliest and rather weighty manuscripts. When the expert-override feature is applied, it affects 345 variant units.

The computer-generated algorithm in a sense can be programmed to be aware of its own limitations and can highlight these areas of concern resulting from expert-assist and expert-override. But even with all these options turned off, the SR still provides a reliable and consistent text that reflects the earliest manuscript evidence based on objective statistical analysis. It is possible that the SR text could contain some errata, but that does not necessarily mean that the algorithm is deficient. It could simply be that the data was not properly encoded or that there was a programming bug in the implementation of the algorithm.

Results

The SR text is only about 1 percent different (1,458 words) from the Nestle-Aland 28th edition, depending on how word differences are counted. Some of those perhaps should not have been counted, for they were merely orthographical differences; the words have the same morphology and make no translatable difference. And some of those were merely word transpositions where all of the words were still the same but just in a different position. In addition, all of the Nestle-Aland's bracketed words indicating questionable readings were counted as being present (but could have alternatively been counted as being absent). With these caveats, the number of different words were as follows:

Additions:	340
Omissions:	562
Substitutions:	556

It is worth noting for comparison purposes that the SR was about 0.6 percent different (825 words) from the BHP, which was the manual prototype specifically created for the purpose of approximating the results of a computer-generated text. The fact that the resulting text is slightly different from the BHP illustrates exactly why the actual creation of the computer-generated text was necessary. As Colwell and Tune pointed out, it is just too difficult for a human to precisely know all of the variables involved and to be *consistent* in weighing the data.

It is difficult to argue against the SR's text, for almost every reading chosen in the SR is backed up by at least one major critical text. For example, if a supporter of the Nestle-Aland text argues that the SR's reading of *σκοτίᾳ* is incorrect at Matt 4:16, one can point out that the SR agrees with both the Society of Biblical Literature and Tyndale House texts there. If a supporter of the Society of Biblical Literature text argues that the SR's reading of *χάριτι* is incorrect at Heb 2:9, one can point out that the SR text agrees with the Nestle-Aland and Tyndale House texts there. The SR does not exactly match any existing critical text, but as described below it is within the same range of variance as our best modern critical texts, which bolsters confidence in the resulting text. Any perceived difficulties in the SR text should not warrant any different treatment than some of the controversial readings already present in the other critical texts.[32] Since our best critical texts disagree anyway, why not let the computer settle the matter in a more objective manner based on scientific statistical analysis?

Other Considerations

The SR is only one implementation of a method designed to simulate a reasoned-eclecticism approach. Other variables and other algorithms could also have been considered. There were a number of other objective data that originally looked promising for consideration in this algorithm, but when analyzed using the data-driven approach, they proved to be ineffective and were ultimately discarded:

- It was originally thought that adding geographical location from where the manuscripts originated would be useful so that multiple witnesses coming from the same region would not be given an exaggerated weight. However, when statistics were examined regarding this concept, there was no observable correlation between the locations where the manuscripts were found and their likelihood to contain a similar text. This corresponds with the observation that many different Byzantine readings

32 Bunning, *Restoration of the New Testament*, §2.3.2.

are already found in early Egyptian manuscripts. Other analysis similar to this has already caused many textual critics to abandon the traditionally held geographical text-types theories.[33]

- It was originally thought that the scribal writing quality (professional hand, reformed documentary hand, documentary hand, or common hand) and class of data would be useful factors in determining a witness's textual reliability. But when the statistics were analyzed, it was discovered that they were only weakly correlated with the textual reliability rating, and the inclusion of this was not helpful to the process.
- It was originally thought that genealogical data would be needed to help weigh the evidence, so that supporting witnesses that were derivatives of each other would not be given a reading an exaggerated weight. But as previously discussed, the results of the CBGM demonstrate why that is not practical for generating the earliest form of the text, since most of the early manuscripts are not direct genealogical descendants of each other. Instead, the support function algorithm was devised to overcome the problem of genealogical corruption by weighing percentage of diversity between manuscripts so that the direction does not matter regarding which manuscript was copied from another manuscript.
- Some might be inclined to think that adding later manuscripts to the database might be useful as well, but it would have little effect on the resulting text as previously mentioned. If the later manuscript contains a new variant reading that was not found in any previous manuscript, then it would carry little weight because of its later date against the united testimony of all of the earlier manuscripts that contradict it. And if it merely added support behind an earlier variant reading, then it would be redundant and add little weight because of the support function. In oversimplified terms, if a reading does not have any early support, it cannot be trusted, and if it already has early support, then its vote is not needed. That is not to say that later manuscripts are not valuable for other aspects related to exploring the textual tradition.

There were also a number of other types of algorithms that were considered but were likewise discarded because they resulted in greater deviation away from our best critical texts. The calibrating process mentioned above would be particularly well-suited for replacement by artificial intelligence (AI), whereby the program could be fed the corpus of early data and then be asked to do whatever possible to approximate our best modern critical texts. Ultimately, if one thousand monkeys typing on typewriters could produce a random block of code that made the computer-generated text a closer match to these texts, then perhaps that code should also be considered. But so far it appears that using rational algorithms designed to mimic a textual critic's intuition seems to work best. Alternatively, a *dejure* approach could be considered for constructing a text from an a priori agreed upon set of rules, instead of this defacto approach, which assumes that our best critical texts should be emulated.

Obviously, the algorithmic approach is not limited to a single solution, for it is merely a tool that could be used to produce any number of different computer-generated texts. Thus, an algorithmic approach in general is subjective, since any type of algorithm could be used to produce all sorts of texts. Indeed, a similar process could be used to generate a Byzantine majority text or a textus receptus text by using different algorithms, data, and constants. But even with that line of thinking, it would still be less subjective than what has been done in the past,

---

33 In particular, the Coherence-Based Genealogical Method (CBGM) has convinced some "to abandon the concept of text-types traditionally used to group and evaluate manuscripts." Tommy Wasserman and Peter Gurry, *A New Approach to Textual Criticism: An Introduction to the Coherence-Based Genealogical Method*, RBS 80 (Atlanta: SBL Press, 2017), 7.

for the underlying definitions and processes would depend on more precise and objective criteria that are testable and repeatable. One advantage of using a computer-generated text is that this type of subjectivity exists at a higher level and does not apply to the selection of individual variant readings, which can be influenced by theological bias and inconsistent reasoning. The algorithmic approach enforces objective *consistency* across the text, preventing the process from being gamed by trying to pick certain favored readings. Thus, if someone tried to tweak the algorithm so that one particular reading was chosen, it would simultaneously cause several other readings not to be chosen. Indeed, minor changes to the weighing of earliness can change whether the longer ending of Mark is included or not, but it also would correspondingly change many other readings that would not necessarily be wanted.

Future Improvements

The algorithm utilized here is by no means fully optimized, and others may indeed be able to find superior algorithms in the future, for many other ideas that were thought of have not yet been tried. Several areas have already been noted where the process could be improved for later releases:

- Subvariants that represent smaller changes within a variant unit could be processed separately and then weigh in on the result as a whole.
- Precise metrics could be created for each type of variant pattern, allowing the probabilities within each pattern to be treated differently.
- Other dependent variant units that exist beyond consecutive verses could be identified by a computer algorithm.
- Variant readings that spanned multiple verses were decided one verse at a time, when they could be more efficiently processed together.
- Other types of internal evidence could be processed such as word frequencies across the entire text or rating the harder readings by a rubric.
- The different algorithmic steps could be combined into a single program, making it closer to a pure turnkey solution.
- The minimal dataset could be greatly expanded, particularly by adding the church fathers and foreign versions data.

There are also at least three different areas for which AI would be well suited to make improvements to some of the existing processes: a more detailed identification of the different variant patterns related to the demarcation of the variant unit boundaries, a more exhaustive calibration to ensure the best of all possible weightings were considered (as previously mentioned), and a more probabilistic approach to the identification of supplied readings designated as *vid*.

The SR is planned to continue to be developed and improved, and periodical snapshots of it will serve as future releases. The textual choices and associated probabilities will obviously change as new data is added and the algorithms improve. Similar to the development of software, when a new edition of the SR is officially ready for release, it will replace the existing SR text, and then the next developmental version will begin. As with other versions of the Greek New Testament, a revised edition does not necessarily mean the previous text was bad, only that it represents the latest scholarship based on our best current knowledge. In addition to being able to produce a complete text, it should be noted that the SR's algorithms can also be used as a tool for evaluating other critical texts and informing textual decisions in general.

Conclusion

It is expected that the release of this first computer-generated text may have a profound impact on the field of textual criticism that could reverberate for decades. The use of computers in the field of textual criticism has been grossly underleveraged, but with the emergence of a vast number of electronic transcriptions and a number of computer-based projects such as the CNTR, this is beginning to change. The fact that a computer program given its stated limitations was capable of producing a very reasonable text certainly challenges the thinking in a number of areas concerning the importance of later manuscripts, the value of genealogical data, and the reliance on text critical canons, as discussed in the previously mentioned book.[34] Indeed, the fact that the resulting SR text is so similar to the Nestle-Aland text is quite surprising if not confounding to some. One of the reasons for this is that the eclectic methodology used by the Nestle-Aland text is similar to the eclectic methodology used by the SR. That is, the SR does much of what the editors of the Nestle-Aland text were perhaps trying to do, but it does it more consistently and with more accurate data. As with the disagreements between other critical texts, this issue is not really about proving something is right or wrong, but when compared with the early manuscript evidence, the SR surely presents a reasonable text.

With the field of textual criticism splintering across more and more subjectively created critical texts, it was perhaps inevitable that a more objective solution would be sought using a more scientific basis. Regardless of the reception that this particular computer-generated text receives, it is expected that the SR will open the door to all sorts of other statistical analysis and computer processing as this is just the tip of an iceberg. The potential applications of computer science and artificial intelligence to the field of textual criticism may result in further refinements that could propel these concepts far beyond what has initially been accomplished here. It is conceivable that arguments over which critical text is better today may one day be replaced with arguments over which algorithm is better. The future of textual criticism may eventually be rooted in the fields of data science and computer science, and the SR is just one early example of that.

34 Bunning, *Restoration of the New Testament*.