Statistical Restoration Greek New Testament

Alan Bunning
Center for New Testament Restoration

Statistical Restoration Greek New Testament
October 12, 2022

Biblical Worldview Publishing
West Lafayette, Indiana

Introduction

The Statistical Restoration (SR) is the first computer-generated Greek New Testament, which was created by Alan Bunning in 2022 for the Center for New Testament Restoration (CNTR)[1]. The SR is the long-planned successor to the Bunning Heuristic Prototype (BHP) which was manually created in 2012 to anticipate what types of problems might occur in the creation of a computer-generated text. The BHP was later publicly released in 2017 under an open license at the request of unfoldingWord[2] and subsequently translated into several different languages. The SR is intended to be the replacement for the BHP, as the estimation of what the computer might do, has now been replaced by what the computer actually did.

Methodology

The SR was created in conformance to the principles of Scientific Textual Criticism, requiring objectivity, plausibility, transparency, and reproducibility.[3] Scientific Textual Criticism is a relatively new concept being championed by the CNTR, which eliminates subjective textual decisions influenced by theological bias, and replaces them with objective scientific processes based on data science and computer science. The SR endeavored to reflect the most probable text based on statistical analysis and algorithms designed to simulate a reasoned-eclecticism approach actually used by scholars, weighing both external and internal evidence. The external evidence was weighed based on manuscript reliability, earliness, and diversity of support. The internal evidence was weighed based on word probabilities within a variant unit and consideration of variant patterns. A number of breakthroughs had to occur along the way in order to accomplish this feat, including the automatic determination of variant unit boundaries and their relationships to each other,[4] the classification of homophones based on the orthographical-priority method,[5] and rating the statistical reliability of manuscripts against the corpus of data.[6] The details of the statistical analysis and algorithms used to create the SR are described in the paper tentatively entitled, "The First Computer-Generated Greek New Testament".[7]

Advantages

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, but instead, external and internal evidence was objectively weighed. As a result, the SR provides a rational text built on an objective scientific standard.

- The SR is based on all the early extant manuscripts dated before 400 AD, which includes all the continuous-text manuscripts, and quotations from amulets, inscriptions, and other writings.[8] This data was not readily available as a complete dataset until the creation of the CNTR collation.[9] Since the SR only considers extant evidence, it does not contain any conjectural emendations that are found in some other critical texts, but only *actual readings* found in manuscripts.
- 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 which 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 the previous time. Accordingly, the SR does not contain any eccentric readings that are found in some other critical texts.[10]
- The SR was built on processes that are openly inspectable, publicly verifiable, and scientifically reproducible, providing an inherent basis for a *trustworthy* text. When combined with the CNTR collation, each textual decision can be independently 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 analysis, 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, and punctuation. There are several places where every early manuscript is in agreement with how a word is spelled, which is different than the canonical spellings shown in most modern critical texts and lexicons. The Koine Greek orthography also includes *nomina sacra*[11] which often give clear indication to the deity of Christ, but is 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. 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 is released under open-source licenses which will allow others to build on the work and contribute other improvements to serve the interests of the global Church. The text is released under the Creative Commons Attribution 4.0 International License (CC BY 4.0)[12] and the source code is released under the GNU General Public License 3.0 (GPLv3).[13] This is particularly significant in that it satisfies the need to provide an open-licensed modern critical text, based on the best manuscript evidence available, with a process that is fully *transparent* and *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). Please consult the CNTR Technical Reference for more detailed information about specific fields.[14]

Limitations

While the SR offers several improvements over other critical texts, it also has some deficiencies related to its current dataset:

1. The SR was only based on all the early manuscript data up to 400 AD, which is the best data currently available electronically, but predominantly reflects only one geographical region (Egypt).
2. The data set is relatively sparse in a few different places, only containing two early witnesses for some verses.[15]

Because of these deficiencies, this first version of the SR utilized expert-assist and expert-override features which consulted the readings of other major critical texts where the evidence was insufficient. These deficiencies are planned to be addressed by the inclusion of church father data and foreign versions in future editions. Such data was not intentionally excluded, but simply does not currently exist as a complete dataset of actionable electronic transcriptions that can be fully utilized.

The SR will 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.

Significance

Despite its limitations, the SR offers perhaps the closest reflection of the earliest extant manuscripts using a scientific method, and thus, arguably the closest reflection of the original autographs so far. It cannot be proven that a statistically probable text is always right, but basing textual decisions on science surely presents no less of a reasonable text. And the ability to provide an objective scientific text represents great strides in being able to defend the integrity and accuracy of God's word. *The SR is not an end, however, but merely a starting point to a process that will be improved with additional data and better scientific processing*. The release of the SR serves as a proof-of-concept demonstrating a rational computer-generated Greek New Testament could be produced, yielding satisfying results when compared to our best modern critical texts. Other more sophisticated endeavors with superior algorithms are expected to follow in the future. Providing a scientifical computer-generated text that is open to public scrutiny represents a huge step forward toward providing the accuracy and textual authority deserving of the New Testament.

Credits

- Jesus Christ, who is the Lord of my life and if it were not for Him, none of this would have ever transpired.
- My wife Joanne and children Regan and Anna, who patiently allowed me the time to commit thousands of hours on such a massive project.
- Dr. David Dilling ThD, PhD of the Kensington Theological Academy, who not only taught me Greek, but also proofed much of the accompanying morphology.
- Volunteers from unfoldingWord and other individuals all over the world who helped with various aspects of the CNTR project.

---

[1] http://greekcntr.org.

[2] https://www.unfoldingword.org.

[3] Alan Bunning, *Restoration of the New Testament: The Case for Scientific Textual Criticism*, Center for New Testament Restoration, 2022.

[4] Alan Bunning, "Scientific Definition of Variant Unit Boundaries", Textual Criticism and Papyrology section, 2022 Midwest Region Society for Biblical Literature (Virtual meeting), February 5, 2022.

[5] Alan Bunning, "Orthographic Priority for Interpreting Homophones in New Testament Manuscripts", Biblical Lexicography section, 2021 Society of Biblical Literature Conference (San Antonio, TX), November 22, 2021.

[6] Alan Bunning, "Corpus-Based Statistical Measurements of Textual Reliability for New Testament Manuscripts", Textual Criticism and Papyrology section, 2022 Midwest Region Society for Biblical Literature, (Virtual meeting), February 5, 2022.

[7] This paper is not currently available yet, but is being readied for possible publication.

[8] Alan Bunning, *Restoration of the New Testament: The Case for Scientific Textual Criticism*, § 1.2.1.3, Biblical Worldview Publishing (West Lafayette, IN), 2022.

[9] https://greekcntr.org/collation/index.htm.

[10] Alan Bunning, *Restoration of the New Testament: The Case for Scientific Textual Criticism*, § 2.3.2, Center for New Testament Restoration, 2022.

[11] 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.

[12] https://creativecommons.org/licenses/by/4.0

[13] https://www.gnu.org/licenses/gpl-3.0.en.html.

[14] Alan Bunning, "CNTR Technical Reference", June 1, 2022; https://greekcntr.org/resources/technical.pdf

[15] There are 124 verses in 1 Timothy, 2 Timothy, Titus, Philemon, and Revelation that are limited to only two early witnesses.