

***Center for New Testament Restoration
Technical Reference***

By Alan Bunning

Rough Draft
January 15, 2021

Copyright © 2021 by Alan Bunning. All rights reserved. This rough draft is a work in progress and intended for review purposes only. Distribution without the author's prior consent is prohibited.

Table of Contents

1. Introduction

- 1.1 Database*
- 1.2 Website*

2. Witnesses Acquisition

- 2.1 Terminus Ad Quem*
- 2.2 Data Classes*
- 2.3 Identification*
- 2.4 Metadata*

3. Transcriptions

- 3.1 Graphemes*
 - 3.1.1 Early Greek Manuscripts*
 - 3.1.2 Critical Texts*
- 3.2 Manuscript Encoding Specification*
 - 3.2.1 Textual Spacing*
 - 3.2.2 Character Condition*
 - 3.2.3 Supplied Words*
 - 3.2.4 Abbreviations*
 - 3.2.5 Scribal Corrections*
 - 3.2.6 Editorial Marks*
- 3.3 Accuracy*

4. Linguistic Parsing

- 4.1 Words*
 - 4.1.1 Word Boundaries*
 - 4.1.2 Alternate Spellings*
 - 4.1.3 Misspellings*
 - 4.1.4 Unknown words*
- 4.2 Lexicography*
 - 4.2.1 Lemmas*
 - 4.2.2 Lexemes*
 - 4.2.3 Enhanced Strong's Numbers*
- 4.3 Morphology*
- 4.4 English Glosses*

5. Structural Components

- 5.1 Versification*
- 5.2 Collation Alignment*
- 5.3 Slot Assignments*
- 5.4 Variant Units*

Appendix – Database Design

1. Introduction

This document describes the various standards, procedures, and technical details of the Center for New Testament Restoration (CNTR) project which is dedicated to the pursuit of scientific textual criticism. Much of this data is technical in nature and requires a knowledge of computers, New Testament Greek, and/or textual criticism. Those who have questions about any of the details of the project are encouraged to consult this document first as many of the issues commonly raised are addressed below. This information will also be beneficial to researchers and developers who are interest in obtaining information in various formats for data exchange. A general description of the project and its goals are discussed in the CNTR Project Overview document.¹

1.1 Database

All data used in the CNTR project has been incorporated into an SQL accessible relational database which allows the data to be easily stored and retrieved for a variety of purposes. The database was built from scratch to store various types of data for the specific purpose of supporting research and analysis in textual criticism. The CNTR database currently contains over 1.5 million words from 196 witnesses and contains all of the most important variant readings in the New Testament – it includes all of the earliest extant Greek manuscripts up to 400 AD, both continuous texts (class 1 data), amulets, inscriptions, and other quotations (class 2 data); and also six major critical texts² which contains any variant readings beyond those which the experts thought were important.^{§2.1}

The database was developed from the bottom up in stages using a data-driven approach which correspond to the various sections of this document:

1. Witness Acquisition – the sources of data were acquired along with their accompanying metadata.
2. Manuscript Transcriptions – electronic transcriptions were made for each witness.
3. Linguistic Parsing – the transcriptions were lexically, morphologically, and syntactically parsed.
4. Structural Components – advance structures were added to enable contextual processing.

The CNTR database provides several advanced features for textual criticism not available in any other computer platform, enabling several types of advanced data analysis that have never before been possible. The power of a database seems to be poorly understood and grossly underutilized by the current generation of textual critics. For example, the painstaking counting of certain scribal habits that used to be done by hand can now be completed by in seconds by a single SQL query. Variant units can be compared and statistically analyzed across all manuscripts. Morphological word forms or orthographical tendencies can even be examined across manuscripts or isolated to any geographical region.³ The CNTR database is not tied to any particular type of research as many different kinds of data analysis can benefit from accessing the CNTR database.

1.2 Website

The CNTR website (<https://greekcntr.org>) is written in compliance with the HyperText Markup Language (HTML) version 5 and Cascading Style Sheets (CSS) version 3 standards. The website code is currently maintained by hand without the aid of any web design software. All programming elements of the project are written in JavaScript in compliance with the ECMAScript version ECMA-262 edition 11 standard. All programming elements that run outside of the browser are also written in JavaScript using

Node.js. For example, the CNTR collation and transcripts are static webpages that are generated offline directly from the CNTR database using Node.js. Philosophically, the CNTR project tries not to use any of the frameworks built on JavaScript because their popularity constantly changes,⁴ they are each fundamentally different languages,⁵ they carry additional overhead, and the project is currently not complicated enough to realize significant benefits from them.

¹ Alan Bunning, *CNTR Project Overview*, Center for New Testament Restoration: 2021.

² 1885 Westcott and Hort (WH), 2012 Nestle-Aland 28th edition (NA28), 2010 Society of Biblical Literature (SBL), 2018 Robinson/Pierpont, 2020 King James Textus Receptus (KJTR), and 1550 Stephanus (ST).

³ For example, one simply query was done for a researcher at Tyndale House which lists all forms of *nomina sacra* and equivalent words at the same positions in all early manuscripts ordered by verse and then manuscript which he said saved them hundreds of man hours.

⁴ Such frameworks usually don't last long with each one given way to the latest fad – jQuery, Angular, Backbone, React, Svelte, etc. To commit to any of them is to get stuck in the past as programmers coming out of college ten years from now will have never heard of them. Pure JavaScript however will last as it is the most used programming language built-in to all browsers, and will still be needed in the future to run all of those frameworks 😊.

⁵ The fact that JavaScript allows self-modifying code enables these frameworks to be built upon it, but such derivative languages are no longer JavaScript – they are different languages with new syntaxes that each carry their own learning curves. They are not forms of JavaScript any more than C and Python are the same because they both are translated to the same set of CPU instructions.

2. Witnesses Acquisition

The CNTR project attempts to include the text of all witnesses which contain any portion of the New Testament up to the *terminus ad quem* (ending date) of 400 AD. This includes the text from inscriptions, amulets, talismans, and quotations from church fathers written on any material (papyrus, parchment, ostraca, stone, etc.). Each corrector of a manuscript counts as a separate witness to the text of the New Testament, provided that the corrections occurred prior to the *terminus ad quem*. It is important to note that a witness is not any less important just because its text is fragmentary, since those fragments were once part of a complete manuscript. The CNTR collation also includes some of the more significant Greek critical texts for comparison purposes.

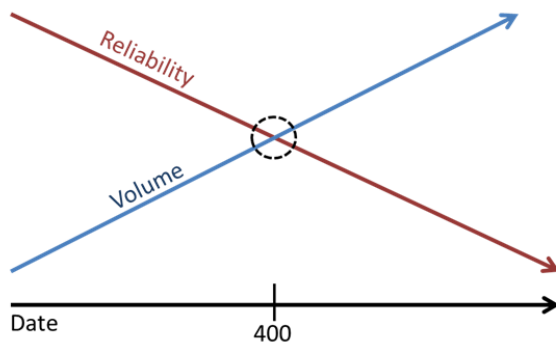
2.1 Terminus Ad Quem

In order to be included as a witness, the CNTR project has adopted a *terminus ad quem* of 400 AD. The “early Christian” era is generally accepted by many disciplines to be the time from Jesus Christ’s ascension around 30 AD up through the Council of Nicea in 325 AD. Around that time frame, there was a noticeable change where many Greek manuscripts had truly become eclectic texts, simply borrowing between different sets of variant readings among the available source materials. For example, there are several instances where the scribes were not merely making corrections to their handwriting, but were crossing out words and supplying other variant readings. Textual critic Kurt Aland states:

“It was assumed in the early period there were several recensions of the text (cf. von Soden), or that at the beginning of the fourth century scholars at Alexandria and elsewhere took as many good manuscripts as were available and applied their philological methods to compile a new uniform text (this was the view of our fathers, and is still that of many textual critics today as well).”¹

Aland considers early manuscripts to be those dated “no later than the third/fourth century, for in the fourth century a new era begins.”² He contends that under Constantine’s influences, “the Byzantine church molded it to the procrustean bed of an ecclesiastically standardized and officially prescribed text.”³ Bruce Metzger concurs that after “Christianity received official sanction from the State” during the fourth century, “it became more usual for commercial book manufacturers, or scriptoria, to produce copies of the books of the New Testament.”⁴ Indeed, by the end of the *terminus ad quem*, this change in textual nature is readily apparent in later manuscripts such as Codex Bezae (GA 05) and Codex Washingtonianus (GA 32). Later manuscripts are also more likely to have crossed geographical boundaries of textual transmission and become mixed as they gained greater dispersion over time, making it impossible to know what textual lineage they may have been copied from. For example, the Coptic Orthodox Church of Alexandria ironically no longer uses the “Alexandrian” text which was once native to their region, but now prefers a Textus Receptus text.⁵

As a more practical matter, a *terminus ad quem* of 400 AD also just also happens to provide a good tradeoff between earliness and coverage. Ideally, the *terminus ad quem* should be set as early as possible to minimize later corruptions of the text. Yet it must also be late enough to provide a sufficient amount of material. If we were to consider only class 1 data,^{§2.2} for example, a *terminus ad quem* of 300 AD would not provide enough material to even cover all the verses of the New Testament. But moving the *terminus ad quem* to 400 AD, however, provides sufficient coverage of all the verses of the New Testament while minimizing the amount of later corruptions.⁶

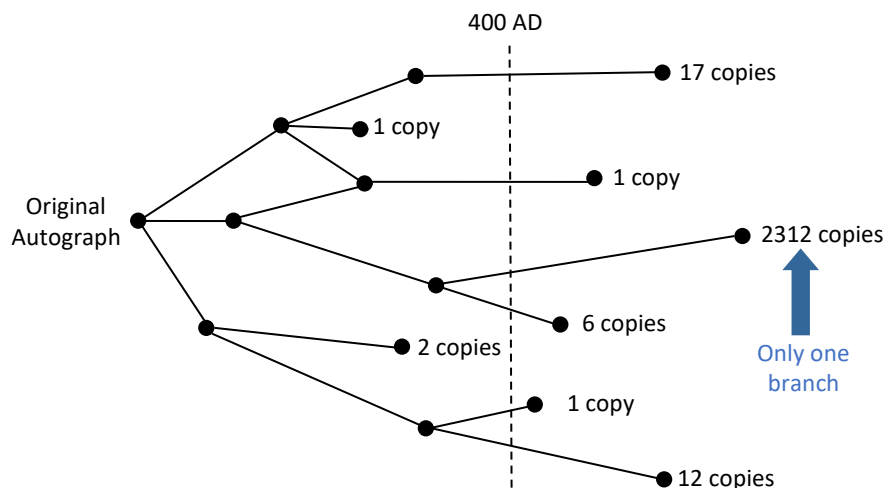


Aland points out that “practically all the substantive variants in the text of the New Testament are from the second century”,⁷ which is also confirmed by both Ernest C. Colwell⁸ and George D. Kilpatrick.⁹ If this is true, then any variant worthy of consideration was already on the table well before 400 AD.¹⁰ Indeed, all the variants chosen for the text of the Nestle-Aland 28th edition exist in extant manuscripts before 400 AD except for about 24 words (and some of those might be better viewed as errors on their part). Clearly, 400 years is long time of textual transmission and provides a large window for textual corruption to occur as evident by the large number of variants produced. And once a new error is introduced, it tends to continue to be propagated forever.

The dates assigned to manuscripts usually involve a certain amount of subjectivity and so the sources of the dates used by the CNTR are listed along with other pertinent information influencing its assessment. Most of the manuscripts are dated based on paleography. That is, the handwriting of a manuscript is compared to the paleographic features of other documents that are more precisely dated by known historical events. The field of paleography for New Testament manuscripts is not an exact science and thus dates spanning no less than 50 years have typically been assigned beginning on 25 year boundaries. A few of the later manuscripts included would normally be assigned to the fifth century, but the general rule was adopted that a manuscript would be included if at least one expert dated it to the fourth century. Thus, the *terminus ad quem* is somewhat of a fuzzy line where anything that could have *possibly* have been fourth century is included. Erring on the side of inclusion was preferable over the possibility of excluding valuable data in some expert’s opinion.

The CNTR currently has no aspirations to go beyond this *terminus ad quem*. That is because any *new* variant reading that occurs after that date must be viewed as highly suspect compared to the sufficient amount of earlier textual material that contradicts it. On what rational basis would you go against the collective testimony of all the earlier sources to adopt a new variant reading from a later witness? It is always amusing to read articles written by textual critics fawning over the new discovery of a manuscript dated to the ninth century. But what value does such a late manuscript possibly have toward determining the original text of the New Testament? If the manuscript contains a new variant reading that was not found in any previous manuscript, then it should automatically be rejected because of united testimony all of the earlier manuscripts that contradict it. And if it merely adds support behind an existing variant reading, then it adds nothing new to the debate because of its late date. In other words, if a later manuscript does not have any early support, it cannot be trusted, and if it already has early support, then its “vote” is not needed. What authority can a later variant reading have that is opposed by *every* early New Testament text that was used by the early Church in multiple geographical areas?

The informed textual critic needs to understand that *the number of times a variant reading appears in later manuscripts is irrelevant, especially if it doesn’t exist in any of the earliest manuscripts!* In terms of stemmatics, it is important to identify the earliest branches of the text, and if possible, their geographic origin, not how many times a branch is copied after that.



Again, Aland contends that only “manuscripts which derive from the third/fourth century or earlier have inherent significance, i.e., those of the period *before* the development of the great text types.”¹¹ Obviously, if a scribe made an error and it was then copied a zillion times, that would not make it correct. Thus, any type of headcount approach of manuscripts such as those advocated by Byzantine Majority text proponents is invalid for scientific textual criticism. This does not mean that some Byzantine readings may not be early or correct, but only that they cannot be chosen on the basis of medieval popularity. Ironically, the majority of texts we possess up till the 9th century are mostly Alexandrian in nature, and now the “majority text” used today is again Alexandrian in nature because of the popularity of the Nestle-Aland text and its basis for the majority of modern Bible translations used by the Church. There is absolutely no reason to exclude copies of manuscripts from the counts just because they came after the printing press!¹² The printing press was really no different than the early scriptoriums that were formed to mass produce copies of the New Testament, except that the printing press could do it better and faster. Clearly, the number of times something is copied without weighing its genealogical relationship to earlier texts establishes nothing.¹³

While it is possible that a later manuscript could contain readings that were directly copied from an earlier manuscript that had been lost, it is just as likely that it could have been copied from a manuscript that was made the previous year! There is simply no way to tell. Aliens from another planet could have obtained the originals directly from the apostles and given them to a monk in the ninth century, but we have no evidence to believe that either! A later manuscript reading must first be able to demonstrate that it has early support before it would be relevant. The informed textual critic needs to understand that *apart from any additional documentation* regarding its origin, a manuscript from the ninth century has absolutely no advantage in determining the original text of the New Testament than a critical text made in the nineteenth century! A ninth century manuscript carries no more weight *because there is no way to distinguish if the scribe was making his own eclectic text, or if he was simply trying to copy an earlier manuscript*. Indeed, a monk who produces a manuscript in the ninth century may arguably have been in a worse position than those who create critical texts today, since he may have had access to fewer manuscripts or have been limited to manuscripts from only one geographical region. All that a manuscript from the ninth century can tell you is what one individual thought the text of the New Testament was in the ninth century!

Of course, this does not mean that every early manuscript necessarily contains a more accurate text than every later manuscript; for after all, an early scribe could have made mistakes in copying his text. For example, Codex Vaticanus is dated later than manuscript P46, yet it is often viewed as being more accurate; and Codex Bezae Cantabrigiensis is a relatively early manuscript and it is considered to be wildly inaccurate compared to almost any other manuscript. But what it does mean is that all unknown variables being considered equal, the *corpus* of early manuscripts by probability has had less opportunity for multiple generations of copying mistakes to be expressed than later manuscripts. Thus, they provide earliest snapshots of the text in time and place, which provide important clues for how the text was transmitted. It

is important to examine all of the earliest sources first, before endeavoring to surmise what later manuscripts may have been copied from, regardless of how many of them copied a particular form. In the realm of science, the *prima facie* evidence of an early manuscript must necessarily outweigh unsubstantiated speculations assigned to later manuscripts. There is a big difference from something that might have been copied from an early manuscript and something that *is* an early manuscript! A later manuscript may have possibly been copied from a 2nd century manuscript, but it cannot take precedence over a manuscript that *is* a 2nd century manuscript. The CNTR dataset of early witnesses will always have inherent value because the readings that *are* early will always take precedence over later readings that *might* be early. Indeed, apart from additional evidence, it is entirely *illogical* to favor unsubstantiated speculations or conspiracies theories over the hard *prima facie* evidence.

In summary, a *terminus ad quem* of 400 AD was not set as an arbitrary cut-off date, it is a date *well after* the point after which there should be sufficient early material from all data classes,^{\$2.2} so that the addition of any later data would simply be irrelevant for any rational basis for textual criticism when compared with the wealth of the earlier data spread across multiple geographical regions. (Of course, the CNTR would have no objection to adding another century's worth of data to the collation, that is, if someone else is willing to do the work. 😊)

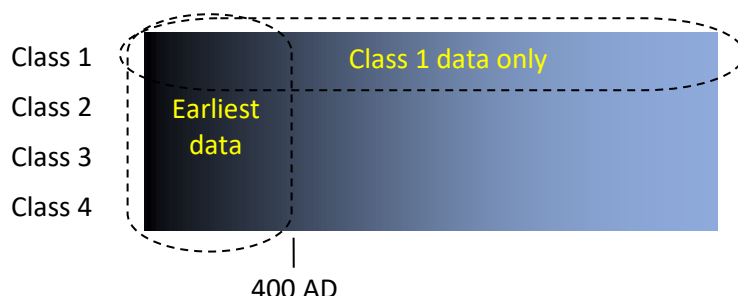
2.2 Data Classes

The earliest witnesses necessary for establishing the original autographs of the Greek New Testament can be categorized into six classes of data according to their relevance:

1. Greek copy of New Testament book(s) written as a continuous text from an extant manuscript dated before the *terminus ad quem*. This data represents the most reliable evidence for it was written by scribes with the intent of passing down copies of the New Testament.
2. Greek quotation of the New Testament in an extant manuscript dated before the *terminus ad quem*. This data is not as reliable since the author could be quoting the Scripture from memory or only making an allusion to a passage. For this project, a quotation is differentiated from an allusion in that it must contain at least four matching words unless there is some other supporting context.
3. Greek quotation of the New Testament attributed to a “time-stamped” source dated before the *terminus ad quem*, but retrieved from an extant manuscript dated after the *terminus ad quem*. This data is not as reliable as class 2 data since the quotations of Scripture could have been altered by a later scribe to match the wordings of a different text (although this is considered less likely). The church fathers were well aware that their writings could be corrupted by scribes in subsequent copies.¹⁴ Indeed, textual criticism is often needed to reconcile different versions of a church father's work. (A variant reading in a church father quotation cannot be dismissed, however, if the meaning of the variant was being discussed in the work!) There is also a danger that some of these works could be pseudepigraphal in nature which would make them useless since they would not really be “time-stamped” then.
4. Foreign language translation of the New Testament contained in an extant manuscript dated before the *terminus ad quem*. These are also “early” manuscripts, but they cannot be used to establish the text of the *Greek* New Testament. A back-translation from these manuscripts cannot provide the precise wording of the Greek text, but they can be used to show support for or against particular variants.
5. Foreign language quotation of the New Testament in an extant manuscript dated before the *terminus ad quem*. This data has the same reliability problems as class 2 data as well as the translations problems as class 4 data.
6. Foreign language quotation of a of the New Testament attributed to a “time-stamped” source dated before the *terminus ad quem*, but retrieved from an extant manuscript dated after the *terminus ad quem*. This data has the same reliability problems as class 3 data as well as the translations problems as class 4 data.

The first two classes of data represent the best *prima facie* evidence for establishing the original Greek text of the New Testament and for the first time in history have been provided in a complete collation by the CNTR. The other classes of data beyond these have a progressively lesser value towards reconstructing the text of the New Testament, but will be included at a later date if possible. Other categories of manuscripts such as minuscules and lectionaries (with the exception of GA ℓ 1604) are not included because these later texts are dated after the *terminus ad quem*.

One limited approach has traditionally been to focus almost entirely on class 1 data, including manuscripts of much later dates, with only a limited awareness of the other classes of data. A much better approach for reconstructing the earliest form of the New Testament, however, would be to start with an exhaustive set of the earliest possible data for all classes of data.



Although all class 1 and class 2 data has already been compiled by the CNTR up to the *terminus ad quem*, the inclusion of the other classes of data would be crucial for any reconstruction of the original autographs of the New Testament. The class 3 data, for example, contains very early data from all geographical regions and could more than triple the amount of data currently in the CNTR collation.¹⁵ A complete set of this data has never been collated before and therefore has not been properly utilized by any critical text. Daniel Wallace points out, “If it could be determined what kind of text they used when they quoted from the New Testament, such information would naturally be highly valuable. But textual critics do not usually give much weight to the church fathers.”¹⁶ Critical texts that have placed an overemphasis on class 1 data present a distorted view of the New Testament because about 95% of that data comes from Egypt which only represents one geographical region. Readings from all the so-called “text-types” are found in Egypt before the *terminus ad quem* which tends to debunk the geography-based text-type theories,¹⁷ but that does not indicate the nature of the textual transmission in other geographical areas. Byzantine priority theories claim that a Byzantine text-type reflects the original autographs, but their early manuscripts were not preserved as well as they were in Egypt because of environmental conditions and that is why the Byzantine manuscripts hold later dates.¹⁸ Obviously, textual criticism should not be decided based solely on the weather! Geographical analysis done with the inclusion of the other classes of data will either prove or disprove that notion. As it stands, adherents to the “Alexandrian” and “Byzantine” text-type positions are largely based on faith. The “Alexandrian” proponents have faith that the earliest Egyptian manuscripts are representative of other geographical areas, and the “Byzantine” proponents have faith that the majority of later manuscripts are representative of other early manuscripts that we don’t have. Without a comprehensive analysis of the geographical distribution of all the early classes of data, the field textual criticism has largely been shooting in the dark.

2.3 Identification

Each CNTR witness containing portions of the Greek New Testament has been assigned a unique identifier according to its class of data. Class 1 identifiers consist of Gregory-Aland numbers prefixed with “GA” that were first introduced by Caspar Rene Gregory,¹⁹ and then updated by Kurt Aland,²⁰ and are now maintained at the Institut für Neutestamentliche Textforschung (INTF) at the University of Münster.²¹ These represent manuscripts which originally contained one or more books of the New Testament written

on papyrus (represented by the letter “P”), parchment (represented by the number “0”), or ostrakon (represented here by the letter “O”). The GA numbers previously assigned to ostraca²² are no longer included in the “Liste Handschriften”.²³ In a few instances, some manuscripts that were assigned more than one GA number have been combined because they were later determined to have originally been part of the same manuscript. In such cases, the manuscript is classified by the first GA number followed by the “+” sign to indicate that it includes other manuscripts:

GA P15+ = P15 + P16
GA P49+ = P49 + P65
GA P64+ = P64 + P67
GA P77+ = P77 + P103
GA 029+ = 029 + 0113 + 0125 + 0139
GA 059+ = 059 + 0215

Class 2 identifiers consist of Trismegistos numbers prefixed with “TM”. Trismegistos numbers represent content from several databases and have become the defacto numbering system for all documents of antiquity.²⁴ The manuscripts in this category are not copies of the New Testament books, but contain citations of the Greek New Testament from any source including inscriptions, amulets, talismans, and “church fathers” quotations. Some manuscripts that were originally (mis)assigned GA numbers have been reclassified to this category since they were clearly not copies of books of the New Testament:

TM 61715 = GA P7 (a Christian writing probably by a church father)
TM 61868 = GA P10 (a writing exercise)
TM 62312 = GA P12 (an isolated quotation in private correspondence)
TM 61709 = GA P50 (an amulet or talisman)
TM 61839 = GA P62 (a selection of Scripture passages)
TM 61695 = GA P78 (an amulet or talisman)
TM 61873 = GA P99 (a glossary of unconnected words and phrases)
TM 61645 = GA P80 (a commentary on Scripture)
TM 61795 = GA 0192 or GA-11604 (quotations from the earliest known lectionary)
TM 61914 = GA 0212 (a harmonization of the Gospels known as the Diatessaron)
TM 61871 = GA O24 (a Scripture quotation written on ostrakon)

If an identifier is followed by another character, it denotes another scribe that corrected the manuscript. The “^” symbol indicates a correction by the original scribe who wrote the manuscript and thus should be considered to be a part of the original intent of the manuscript. After all, the original author should be allowed to correct his own mistakes! Subsequent corrections to the manuscript by other hands, however, are treated as separate witnesses and are assigned a letter (i.e. a, b, c), provided that the corrections themselves still fall within the *terminus ad quem*. All other corrections done after the *terminus ad quem* have been ignored. Corrections made centuries later by other hands are no more useful in determining the original autographs than if someone today were able to get their hands on one of those manuscripts and decided to make their own changes to it. The critical texts that have been included for comparison purposes have been assigned their own unique mnemonic abbreviations.

2.4 Metadata

Detailed metadata has been compiled and placed in a database for each witness used in the CNTR project. Such data is useful for conducting searches to provide new insights that have never been possible before. For example, queries such as “display the geographical distribution of all of the copies by John

progressing by date”, “display the relationship between geographical distribution and certain variant readings”, or “display all the copies of Matthew that have so many columns or lines per page” are now possible for the first time and can be answered in seconds. Original research had to be done to collect much of this data, often discovering mistakes made in other publications (but perhaps introducing some new ones as well ☺). The following information has been collected for each witness:

- Aliases:** References made to the same manuscript in other catalogues or numbering systems. The Trismegistos (TM) number of the artifact is also included. Brackets depict which leafs belong with which alias.
- Description:** The material, dimensions, number of extant leafs, columns, lines and characters, and the languages used. The dimensions depict the largest fragment of the manuscript followed by an estimate for the reconstructed manuscript in parenthesis if they are different. The number of lines reflects the highest number of visible lines per page (followed by an estimate for the reconstructed manuscript in parenthesis if they are different). The number of characters reflects the average number of characters per line within 50% of the most frequent line character counts (when given sufficient data).
- Publication:** The first time the text of manuscript was published with either a transcription or images of the complete manuscript. Some manuscripts have been published numerous times, but only the first of these is listed here. The Online Computer Library Center (OCLC) number of the publication is also listed. Brackets depict which leafs were first published by which source.
- Origination** The date and place of origination. After the date is listed, the source for the estimate has been given followed by any other pertinent information influencing its assessment. The original name of the geographic location is listed followed by the modern name and the Getty Thesaurus of Geographic Names (TGN) number.
- Provenance:** The history of the artifact detailing the transfer of ownership. The bold headings indicate the library or museum where the extant manuscript is currently located. Brackets depict which leafs were transferred to which owners. Occasionally the earlier provenance is somewhat speculative in nature.
- Scribes:** The scribes that wrote or edited verses on the manuscript along with their dates (provided that they fall within the *terminus ad quem*). The original scribe is rated by his hand writing quality: professional hand (well-formed letters adhering to upper and lower boundaries with even spacing), reformed documentary hand (approaching a professional look but slightly less polished), documentary hand (common for legal documents or correspondence written for others with inconsistent letters and spacing), or common hand (untrained hand barely able to write). One might assume that a professional scribe may have gone to greater lengths to ensure an accurate copying process than an untrained person writing with a common hand for their own personal use.
- Literature:** For quotations of Scripture, the literary work or genre and the author are listed. The date of the literature is also provided when the identity of the work is known.
- Content:** The verses are listed in the order contained in the manuscript, noting that the order of New Testament books can vary from manuscript to manuscript. The list of omissions specifies verses that were explicitly missing from the manuscript. Verses listed in brackets were omitted or included by one of the hands.
- Notes:** Any other pertinent notes of interest about the manuscript.

Transcript: The date that the electronic transcription was created and last modified. Beyond the images that were used, any other works that were consulted are listed. Also listed are any electronic transcriptions that it was later compared against for improved accuracy. Those that were the first electronic texts publicly available obviously could not be compared with anything electronically and thus are more likely to contain errors.

¹ Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 50, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987.

² Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 56, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987.

³ Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 69, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987.

⁴ Bruce Manning Metzger. *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 14, At the Clarendon Press: Oxford, United Kingdom, 1968.

⁵ Coptic Orthodox Church of Alexandria, Diocese of Los Angeles, "The Holy Scripture"; <https://lacopts.org/orthodoxy/our-faith/the-holy-bible>, accessed May 28, 2018.

⁶ With only class 1 data, there are some verses in the books of 1 Timothy, 2 Timothy, Titus, Philemon, and Revelation that are limited to two witnesses. The inclusion of the other data classes would more than remedy this situation.

⁷ Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 290, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987

⁸ "...for the overwhelming majority of readings were created before the year 200". Ernest C. Colwell, *Studies in Methodology in Textual Criticism of the New Testament*, p. 55. E.J. Brill: Leiden, Netherlands, 1969.

⁹ "Apart from errors which can occur anywhere as long as books are copied by hand, almost all variants can be presumed to have been created by 200 AD." George D. Kilpatrick, "The Bodmer and Mississipp Collection of Biblical and Christian Texts", *Roman and Byzantine Studies*, Vol. 4, No.1, p. 42, Winter 1963.

¹⁰ Some substantive variants are not found before 400AD if limited to the class 1 and class 2 data currently shown in the CNTR collation, but may be found in the other classes of data.

¹¹ Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 104, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987.

¹² Consider if an archeologist 1000 years from now conducted an excavation and found that majority of the manuscripts from this time period contained the Nestle-Aland text or were translated from the Nestle-Aland text. Would that therefore make its text correct? And if it remains the majority text 1000 years from now, would you conclude that the word "ΟΥΧ" was original to 2nd Peter 3:10, even though it was not invented by the INTF until 2015.^{§Error! Reference source not found.} Would you conclude that it must have copied directly from another early Greek manuscript that we no longer possess since there are so many copies of it in existence?

¹³ Some have made a similar error by using a headcount approach with the earliest class 1 manuscripts in the CNTR collation to arrive at a base text without weighing geographical distribution, textual affinity, and textual reliability, and without consideration of the other classes of data.

¹⁴ Irenaeus wrote, "If, dear reader, you should transcribe this little book, I adjure you... to compare your transcript and correct it carefully by this copy from which you have made your transcript. This adjuration likewise you must transcribe and include in your copy." (Eusebius, *Ecclesiastical History*, 5.20.2; G. A. Williamson tr., *The History of the Church from Christ to Constantine*, p. 227, Barnes & Noble Books: New York, 1995.) Jerome wrote, "If then you find errors or omissions which interfere with the sense, these you must impute not to me but to your own servants; they are due to the ignorance or carelessness of the copyists, who write down not what they find but what they take to be the meaning, and do but expose their own mistakes when they try to correct those of others." (Jerome, *Epistulae*, 71.5; Phillip Schaff ed., *Nicene and Post-Nicene Fathers of the Christian Church*, series 1 and 2, Christian Literature: New York, 1886-1889.)

¹⁵ An estimate from Biblidex.com indicates that church father quotations up to 400 AD would contain approximately 147,279 total verses covering 89.4% of the New Testament. A large but unknown number of these, however, are merely allusions.

¹⁶ Daniel Wallace, “The Majority Text and the Original Text: Are They Identical?”, *Bibliotheca Sacra*, 148, p. 151-169, 1991.

¹⁷ “The simple fact that all these papyri, with their various distinctive characteristics, did exist side by side in the same ecclesiastical province, that is, in Egypt, where they were found, is the best argument against the existence of any text types, including the Alexandrian and the Antiochian [Byzantine]”. Kurt Alan, “The Significance of the Papyri for Progress in New Testament Research” in J. Phillip Hyatt ed., *The Bible in Modern Scholarship*, p. 325-346, Abingdon Press: New York, 1965.

¹⁸ Maurice A. Robinson, “New Testament Textual Criticism: The Case for Byzantine Priority”, *TC: A Journal of Textual Criticism*, vol. 6, 2001; <http://rosetta.reltech.org/TC/v06/Robinson2001.html>, accessed May 13, 2016.

¹⁹ Caspar Rene Gregory, *Die griechischen Handschriften des Neuen Testaments*, J.C. Hinrichs'sche Buchhandlung: Leipzig, Germany, 1908.

²⁰ Kurt Aland, et al. eds., *Kurzgefaßte Liste der griechischen Handschriften des Neuen Testaments*, 2nd revised ed. Walter de Gruyter & Co.: Berlin, Germany, 1994.

²¹ Institut für Neutestamentliche Textforschung (INTF), University of Münster, “Fortführung der Handschriftenliste”; http://www.uni-muenster.de/NTTextforschung/KgLSGII06_12_12.pdf, accessed November 26, 2012.

²² E. v.Dobschütz, “Zur Liste der NTlichen Handschriften”, *Zeitschrift für die neutestamentliche Wissenschaft und die Kunde der älteren Kirche*, no. 32, p.188, 1933.

²³ Institut für Neutestamentliche Textforschung (INTF), University of Münster, “Fortführung der Handschriftenliste”; <http://intf.uni-muenster.de/vmr/NTVMR/ListeHandschriften.php>, accessed November 26, 2012.

²⁴ Trismegistos, <http://www.trismegistos.org>, accessed September 6, 2012.

3. Transcriptions

CNTR electronic transcriptions endeavor to faithfully represent the text of various types of witnesses to the New Testament including any spelling or grammatical errors that the texts contain. These transcriptions have taken many years to create and have been compared with other transcriptions when possible for improved accuracy.^{§3.3} These transcriptions coupled with the CNTR metadata allow for detailed orthographical studies that were never before possible. If you discover any manuscript transcription errors, please report them through the CNTR website.

3.1 Graphemes

The CNTR transcriptions use standard Unicode characters to represent the individual letters. When working with different character sets, it is important to recognize the difference between a character and a font that is applied to a character. In the beginning, personal computers only used ASCII characters and there were no such things as fonts for there was only one way that they could be displayed. Later, as word processing programs evolved, the ability to apply different fonts to characters began to emerge. Thus, the ASCII character “a”, for example, could be made to *look* differently by applying different fonts such as Ariel “a” or Courier New “a”. Later, Greek fonts were developed that could make English letters look like Greek letters. For example, the Microsoft Symbol font¹ makes the character “a” *look* like the Greek alpha “α”, even though the underlying character is just an English “a”. When Unicode was developed, however, each international character (Greek, Hebrew, Chinese, etc.) was assigned its own unique character code. Thus, it was no longer necessary to use a font to make English characters *look* like Greek characters because each Greek character had been assigned its own unique character code. Thus, the characters “a”, “α”, and “ⲁ” in Unicode are all completely different characters and are generated by typing in different keystroke combinations. The sets of fonts were then extended to cover these new international codes so that now the Unicode Greek alpha character (which is not the English letter “a”) can be rendered by different fonts such as Ariel “α” or Courier New “α”.

There are a number of different ways that Unicode characters can be represented which can also be confusing, but UTF-8 and UTF-16 are the most common formats. Characters that are not able to be represented by a particular font often appear as a square box “□”. The CNTR Character Tool feature is a useful utility for converting Greek characters between different ASCII and Unicode formats.² Using the tool, a transcription can be entered using only the ASCII characters available on the standard keyboard (Symbol+) and then converted to Unicode.

3.1.1 Early Greek Manuscripts

The CNTR transcriptions of the early Greek manuscripts (including all class 1, class 2, and class 3 data) are represented in Koine Greek which is quite different than the Medieval Greek that is commonly displayed in the critical texts. The transcriptions for class 2 and class 3 data only include the portions of verses that are part of the Greek New Testament. The Unicode letters representing the Greek alphabet in these transcriptions are shown in the following table using the CNTR KoineGreek font.³

Glyph	Character	UTF-8	UTF-16	Description	Symbol+
α	alpha	CEB1	03B1	Greek small letter alpha	a
β	beta	CEB2	03B2	Greek small letter beta	b
γ	gamma	CEB3	03B3	Greek small letter gamma	g

Δ	delta	CEB4	03B4	Greek small letter delta	d
ε	epsilon	CEB5	03B5	Greek small letter epsilon	e
Ζ	zeta	CEB6	03B6	Greek small letter zeta	z
Η	eta	CEB7	03B7	Greek small letter eta	h
Θ	theta	CEB8	03B8	Greek small letter theta	q
Ι	iota	CEB9	03B9	Greek small letter iota	i
Κ	kappa	CEBA	03BA	Greek small letter kappa	k
Λ	lambda	CEBB	03BB	Greek small letter lambda	l
Μ	mu	CEBC	03BC	Greek small letter mu	m
Ν	nu	CEBD	03BD	Greek small letter nu	n
Ξ	xi	CEBE	03BE	Greek small letter xi	x
Ο	omicron	CEBF	03BF	Greek small letter omicron	o
Π	pi	CF80	03C0	Greek small letter pi	p
Ρ	rho	CF81	03C1	Greek small letter rho	r
Σ	sigma	CF83	03C3	Greek small letter sigma	s
Τ	tau	CF84	03C4	Greek small letter tau	t
Υ	upsilon	CF85	03C5	Greek small letter upsilon	u
Φ	phi	CF86	03C6	Greek small letter phi	f
Χ	chi	CF87	03C7	Greek small letter chi	c
Ψ	psi	CF88	03C8	Greek small letter psi	y
Ω	omega	CF89	03C9	Greek small letter omega	w

The CNTR transcriptions do not differentiate between the medial sigma and final sigma characters because there was no such distinction present in the original Greek manuscripts.⁴ Originally, there was only one type of sigma, the lunate sigma, which does not resemble the modern uppercase, lowercase, or final sigma, but had a crescent shape “C”. (Note that there is a separate Unicode character specifically designated for the lunate sigma (UTF-16 03F2) but that character is not used in the CNTR transcriptions.)⁵ The modern sigma and the concept of a final sigma were not added until centuries later. The use of final sigmas today can be misleading for they appear to signify the definite end of the word when there was no such distinction in the original manuscripts.

In addition to the normal Greek alphabet, the following special characters and ligatures are used in some early Greek transcriptions:

Glyph	Purpose	UTF-8	UTF-16	Description	Symbol+
◆	unknown character	EFBFD	FFFD	replacement character	#
ˉ	terminating nu	C2AF	00AF	Macron	N
Ϛ	numeric value for 6	CF9B	03DB	Greek small letter stigma	j
ϛ	numeric value for 90	CF9F	03DF	Greek small letter koppa	v
Ϝ	numeric value for 1000	CDB5	0375	Greek lower numeral sign	l
Ϟ	ligature for TP (staurogram)	E2B3A8	2CE8	Coptic symbol tau rho	R
Ϡ	ligature for MOY	EE8081	E001	private use character	M
Ϝ	ligature for KAI	CF97	03D7	Greek kai symbol	K

The unknown character “◆” represents a letter that was clearly present, but it could not be made out at all. This often occurs when a letter was erased or rubbed out by a corrector and then overwritten. These characters are distinguished from the supplied characters in lacunae for those characters were not present, while the unknown characters were present but simply not decipherable.

The terminating nu character “ˉ” is one example of the use of the overline in early Greek manuscripts (it is also used for abbreviations).^{§3.2.4} In this case, a line was placed over the last letter at the end of a column

to imply the present of the letter nu that was left off to save space. This scribal mark can even occur if a column breaks at a nu in the middle of a word, and so it should not be confused with the grammatical concept of “movable nu”. This character is represented by the implied nu character (̸̣) in the collation since the formatting of column breaks is not retained in that format.

A ligature is a character which represents multiple letters that have been combined together to form a single glyph. There are many more kinds of ligatures than the ones shown above, but these are the only ones found before the *terminus ad quem*. Most other transcriptions do not retain these characters at all, but spell out the letters they represent in full. For purposes of textual criticism, however, it is important to maintain the correct character spacing in the manuscripts whenever possible. If the letters were spelled out in full, it would alter the correct line lengths of a column.

The CNTR transcriptions of the early Greek manuscripts do not include accents, punctuation, or capitalization because Koine Greek did not include any of these features:

- Transcriptions of the early Greek manuscripts do not contain any diacritical marks (acute accent ᾱ, grave accent ᾷ, circumflex accent ᾶ, rough breathing ᾰ, smooth breathing ᾱ, dieresis ᾶ, iota subscript ᾱ, or elision ᾰ) because they were not present in the original autographs of the New Testament. Diacritical marks were reputedly first introduced by Aristophanes of Byzantium in order to help preserve the pronunciation of ancient Greek⁶, but they were not widely used in Greek texts until hundreds of years later.⁷ (While some marks such as the dieresis, iota subscript, and elision existed at that time, they were rarely used and not reflected in the early New Testament manuscripts in any consistent fashion.) It would be impractical to try to add diacritical marks to a badly fragmented manuscript since the recessive position cannot be determined when the ending of a word is missing, enclitics and proclitics cannot be determined when the adjacent word is missing, and heteronyms cannot be properly marked when the context of the sentence is missing. Applying diacritical marks to a text where they didn’t exist can often impose a bias on the text which forces a specific interpretation between words that are heteronyms.
- Transcriptions of the early Greek manuscripts do not use any punctuation marks because they were not present in the original autographs of the New Testament. Aland points out that the original texts “naturally also lacked punctuation”.⁸ The authors of the New Testament did not use any periods, question marks, commas, semicolons, quotation marks, etc. in their writings because such punctuation did not become prevalent in Greek texts until hundreds of years later.⁹ Later efforts to standardize punctuation such as paragraph marks, pauses, or stops, based on markings in some early manuscripts, could perhaps show an earlier understanding of the text, but this does not necessarily give any indication at all of what the original authors wrote. In keeping with this idea, Metzger states: “The oldest manuscripts (P^{66, 75*} ⲕ* A B) have no punctuation here, and in any case the presence of punctuation in Greek manuscripts, as well as in versional and patristic sources, cannot be regarded as more than the reflection of current exegetical understanding of the meaning of the passage.”¹⁰ Like diacritical marks, punctuation marks can often bias the interpretation of the text by externally influencing the connection of phrases and ideas.¹¹ There is also wide range of disagreement in many verses between editors who have added punctuation to their Greek texts.
- Transcriptions of the early Greek manuscripts do not contain any capitalization because it did not exist in the original autographs of the New Testament. There was no such thing as uppercase and lowercase characters because there was only a single form for each letter which was written in a majuscule (or uncial) script. Around the 9th century, a cursive minuscule script emerged which eventually developed into the Greek lowercase letters used today. This minuscule script was used because the more compact style used less parchment and could be written more quickly.¹² The earliest minuscule Greek text with a known date is the Uspenski Gospels (MS 461) bearing the year 835.¹³ Even then, however, there was still no concept of capitalizing words; it was merely a change in the style of script. The practice of capitalization developed even later in the Middle Ages as the first letter of a word was sometimes capitalized to provide a form of emphasis in some formal documents. The modern uppercase letters were derived from the majuscule script (with a few letters borrowed from an even older archaic script

used for inscriptions¹⁴), and the lowercase letters were derived from the minuscule script. Thus, any capitalization present in the Greek New Testament texts today is due entirely to an editor’s subjective additions. Arguments over which words should be capitalized to refer to a deity are entirely speculative and bias the text since such distinctions did not exist in the original Greek manuscripts.

3.1.2 Critical Texts

The CNTR transcriptions of the critical texts are rendered in Medieval Greek which is the usual way Greek is displayed in most printed materials. Unlike the early Greek manuscripts, the critical texts contain diacritical marks, punctuation and capitalization that were added to Greek texts during the middle ages. This is not the same as modern Greek, which now only has one type of accent.^{§Error! Reference source not found.} Sometimes a single character can be represented in more than one way using Unicode, and in those cases the CNTR transcriptions always uses normalized characters.¹⁵ For example, the Greek alpha character with an acute accent “ᾱ” can be represented in Unicode as a single normalized character (Greek small letter alpha with tonos, UTF-16 03AC) or two decomposed characters (Greek small letter alpha, UTF-16 03B1) and (combining acute accent, UTF-16 0301). The CNTR Character Tool provides the ability to convert between normalized and decomposed characters and these functions are also available in the Javascript and Python programming languages. The following punctuation marks may appear in the critical text transcriptions:

Glyph	Purpose	UTF-8	UTF-16	Character Description
'	elision	CABC	02BC	modified letter apostrophe
¶	paragraph	C2B6	00B6	pilcrow sign
,	comma	2C	002C	comma
;	semicolon	C2B7	00B7	middle dot
:	colon	3A	003A	colon
–	dash	E28093	2013	en dash
.	period	2E	002E	full stop
!	exclamation mark	21	0021	exclamation mark
;	question mark	3B	003B	semicolon
(open parenthesis	28	0028	left parenthesis
)	close parenthesis	29	0029	right parenthesis
`	open single quote	E28098	2018	left single quotation mark
'	close single quote	E28099	2019	right single quotation mark
“	open double quote	E2809C	201C	left double quotation mark
”	close double quote	E2809D	201D	right double quotation mark
◊	alternative versification	E28B84	22C4	diamond operator

Sometimes other Unicode characters are used in critical texts to represent some of these punctuation marks, and in those cases, they were converted to follow this standard.

3.2 Manuscript Encoding Specification

The CNTR has developed the Manuscript Encoding Specification (MES) to represent the textual features commonly found in New Testament manuscripts. Alternative data formats such as Text Encoding Initiative (TEI) Extensible Markup Language (XML) and JavaScript Object Notation (JSON) were evaluated and previously implemented but were not easily utilized by the vast majority of users. Thus, a new system was developed with the goal to create a system that is easy to use for the average user and

requires no specialized software for either data entry or data manipulation. Those who have the expertise to use TEI XML or JSON can easily convert the MES data to those formats, while the vast majority of users who lack that technical ability can work directly with the data in almost any editor or spreadsheet program. The main advantage of MES is that it is much more compact and less complicated to parse than those other formats. It is also advantageous because the same data format can be used for both inputting manuscript transcriptions and outputting data for distribution. The philosophy behind the MES was to use specific ASCII characters found on the keyboard for quickly marking up the text which contains Unicode letters and punctuation:

Category	Glyph	Purpose	ASCII	Character Description
Textual Spacing	\	page break	5C	backslash
		column break	7C	vertical bar
	/	line break	2F	forward slash
	&	line remnant in lacuna	26	ampersand
	*	verse remnant in lacuna	2A	asterisk
Character Condition	%	character damaged	25	percent sign
	^	character missing	5E	circumflex accent
Supplied Words	~	word supplied	7E	tilde
	+	word supplied by vid	2B	plus sign
Abbreviations	=	nomina sacra	3D	equals sign
	\$	numeric abbreviation	24	dollar sign
Scribal Corrections	*	original scribe uncorrected	2A	asterisk
	^	original scribe correction	5E	circumflex accent
	a	second scribe correction	61	Latin small letter A
	b	third scribe correction	62	Latin small letter B
	{	begin edited text	7B	left curly bracket
	}	end edited text	7D	right curly bracket
Editorial Marks	_	altered word division	5F	underscore
	[begin questionable text	5B	left square bracket
]	end questionable text	5D	right square bracket
	+	verse present	2B	plus sign
	-	verse absent	2D	hyphen-minus sign
	"	direct quotation	22	quotation mark
	'	incorporation	27	apostrophe
	`	allusion	60	grave accent

The application of these codes is explained in more detail in the sections below. One benefit of MES is that any of these markup characters can easily be deleted in an ordinary editor, leaving only certain codes of interest, or just the text if so desired. When deleting such codes, note that the symbols “*^” have two uses and when they represent scribal corrections they are followed by “{”. Most of the MES symbols are only used for the early Greek manuscripts and the symbols “_ []” only appear in the critical texts. MES currently uses all of the visible ASCII keyboard symbols except “@<>?”.

3.2.1 Textual Spacing

There are several different symbols that are used to indicate breaks or gaps in the text which primarily provide the instructions to recreate the layout of the text in a manuscript format:

- The “\ | /” symbols indicate that a break in the text has occurred and the text continues on the next page, column, or line respectively. If these symbols are followed by a number, it indicates how many of those designated units are skipped. If these symbols are followed by a “0”, it indicates that an unknown number of those units have been skipped.
- The “&” symbol indicates that there was a gap in some of the characters on a line. If it is followed by a number, it indicates approximately how many characters are missing.
- The “*” symbol is used to indicate a gap of unknown text relative to the verse in the collation. It indicates that the missing words in a gap are unknown, which is distinguish from a gap where the text is known to be omitted from the text.

The transcriptions for class 2 data only include the portions of verses that are part of the Greek New Testament and use the same symbols to skip over other parts of the text.

The manuscript transcriptions are numbered in order reflecting the number of *extant* leafs of the entire manuscript, including any additional works besides the New Testament portions. The leaf has been selected as the basic numbering unit because every manuscript page always has a front (recto) and a back (verso). Scrolls are numbered by their columns on the front and columns on the back. The Inventory, Provenance, and Publication fields make reference to the applicable extant leafs when the entire manuscript does not reside in the same location, was not transferred as a single unit, and/or was published piecemeal by different sources. If additional leafs of a manuscript are found later, the transcription numbering will be updated accordingly. Other units involving folded pages such as folios and quires may prove useful in understanding a manuscript’s construction, but are often unknown and are not included here. It important to note that the leaf numbers do not necessarily correspond to any page numbers that might appear on the manuscript. Such page numbers are inadequate as a reference system for several reasons: some manuscripts are not numbered at all, some leading or trailing pages may not be numbered, some pages have been misnumbered, and some pages were numbered by later scribes and placed into a different order.

3.2.2 Character Condition

The CNTR transcriptions follow the normal trinary transcription standard based on the Leiden convention¹⁶ and its subsequent refinements^{17,18} which differentiates the clear text from the damaged text and missing text (lacuna). CNTR transcriptions present these distinctions using the following conventions:

- A character not followed by anything is *present* and is displayed in normal readable text (X).
- A character followed by a “%” means it is *damaged* but traces of the character remain (analogous to an underdot) and is displayed with a gray background (X̣).
- A character followed by a “^” means it is *missing* but supplied by surrounding context (analogous to square brackets) and is displayed as an inverse character (X̅).

The implementation of this trinary standard can be subjective and is often inconsistently applied by different transcriptionists:¹⁹ A transcriptionist’s criteria for designating an underdot may range anywhere from “damaged” to “small trace” to “uncertain”. In practice, many transcriptionists do not always evaluate an individual character in isolation from its surrounding context. For example, a transcriptionist who is familiar with the text may place an underdot under the trace of a letter that makes up part of an anticipated word, while another transcription would use a question mark since the character could not be identified on its own. In addition to this, there are other factors which influence a transcription’s quality:

- Some transcriptions were done by inspecting the original manuscript, while others were done from photographs of varying quality.

- Some transcriptions made use of more sophisticated forensic techniques, while others were limited to visual inspection.
- Some transcriptions were done when a manuscript was in better condition, while others were performed years after its discovery and the manuscript had begun to degrade.

As to the latter point, consider how much the manuscript GA P72 has degraded over time through three successive images:



A transcriptionist who examined the third image which is the most recent (and highest resolution) would show several letters as damaged and missing that were completely visible to the earlier transcriptionist who examined the first image (of lower resolution). Indeed, among the many transcriptions consulted for this project, there was often inconsistency in the condition of the damaged letters (and the CNTR transcriptions are not particularly consistent either). The idea that a consensus could be reached among scholars for the level of certainty of every letter in every manuscript is highly impractical.

As a result, the CNTR transcriptions tend to use an *optimistic* transcription standard. That is, if another transcription specified greater certainty for a particular letter, then the CNTR transcription would often be altered to accept that designation, trusting that they had access to better images or technology. That is, *provided that there was no disagreement on what the letter in question actually was*. Thus, any debates would only be over the condition of particular letters, but not what the letters actually were. If someone were to contest a questionable letter in order to advocate a new reading, they would have to go back and examine the original manuscript anyway to see if any remnants of the letter would be consistent with their proposed reading. Any disagreements between transcriptions over the identity of a particular character were resolved by examining the images of the extant manuscripts.

3.2.3 Supplied Words

Many of the early Greek manuscripts contain lacunae (or gaps) in the text where a damaged portion of a manuscript was missing or illegible. Missing words were only supplied in the CNTR transcriptions where the extent of a lacuna was sandwiched between words that did not extend beyond consecutive verses. Thus, there is no single verse in the CNTR transcriptions that consists entirely of supplied words. Obviously, the larger the lacuna, the more speculative the supplied reading becomes. For this reason, there was no attempt to supply long speculative reconstructions to the imagined edges of a manuscript which could have contained any reading.

The CNTR transcriptions try to supply the most probable for lacunae by examining the equivalent readings from other texts while attempting to retain the manuscript’s conventions of spelling and abbreviations in a manner that is consistent with its column spacing. Although the supplied words for lacunae may be speculative in nature, sometimes they are highly probable when a limited amount of space or grammatical factors indicates a certain variant reading while excluding others. In some apparatuses these readings are marked as ^{vid} which stands for *videtur* in Latin, meaning “apparently”. All supplied words are displayed with inverse characters (⌘) with different background colors to display various types of situations. The following symbols are used to indicate the various distinctions of supplied words:

- The “~” symbol indicates that the word was supplied. If the supplied word is not part of a variant unit it is expected text shown in inverse black, and if it is in a variant unit it is shown in inverse slate.
- The “+” symbol indicates that the word was supplied by vid. It can only be used if the word is in a variant unit and is displayed in inverse maroon.
- The “+~” symbols used together indicate that the word was supplied by transpositional vid, which means that the word was highly probably but does not necessarily belong in the order shown. It can only be used if the word is in a variant unit and is displayed in inverse brown.

If any of these symbols are used preceding a word that contains some characters and some missing characters (denoted by the “^” symbol), then then the missing characters take on those global attributes and are displayed in the associated inverse colors.

3.2.4 Abbreviations

Most early Greek manuscripts contain abbreviations of words which consist of a line drawn over some letters. Such abbreviations are not normally depicted in the critical texts.²⁰ Of particular importance are abbreviations of *nomina sacra* which is a Latin term meaning “sacred names”. An ordinary word such as “son” would be written out fully in other contexts, for example, but then abbreviated with an overline if it was used in reference to the “Son of God”. Such abbreviations are preceded by the “=” symbol in the transcriptions. The most common examples of early *nomina sacra* include:

Lemma	Abbreviations
ΑΝΘΡΩΠΟΣ	ΑΝΟC , ΑΝΟΙ , ΑΝΟΥ , ΑΝΩΝ , ΑΝΩ , ΑΝΟΙC , ΑΝΟΝ , ΑΝΟΥC , ΑΝΕ
ΘΕΟΣ	ΘC , ΘΥ , ΘΩ , ΘΝ , ΘΕ
ΚΥΡΙΟΣ	ΚC , ΚΥ , ΚΩΝ , ΚΩ , ΚΝ , ΚΕ
ΙΗΣΟΥC	ΙC , ΙΗ , ΙΗC , ΙΥ , ΙΗΥ , ΙΥ , ΙΝ , ΙΗΝ
ΙΕΡΟΥCΑΛΗΜ	ΙΗΛΜ , ΙΛΗΜ
ΙCΡΑΗΛ	ΙΗΛ , ΙCΛ
ΠΑΤΗΡ	ΠΡ , ΠΗΡ , ΠΡΕC , ΠΡC , ΠΡΩΝ , ΠΡΙ , ΠΡΑCΙΝ , ΠΡΑ , ΠΡΑC , ΠΕΡ
ΠΝΕΥΜΑ	ΠΝΑ , ΠΝΑΤΑ , ΠΝC , ΠΝΑΤΩΝ , ΠΝΙ , ΠΝΑCΙ , ΠΝΑ , ΠΝΑΤΑ
CΤΑΥΡΟC	CΤΡΟC , CΤΡΟΥ , CΤΡΩ , CΤΡΟΝ
ΥΙΟC	ΥC , ΥΙC , ΥΥ , ΥΙΥ , ΥΩ , ΥΙΩ , ΥΝ , ΥΙΝ , ΥΕ

XPICTOC	\overline{XC} , \overline{XPC} , \overline{XY} , \overline{XPY} , $\overline{X\omega}$, $\overline{XP\omega}$, \overline{XN} , \overline{XPN} , $\overline{X\epsilon}$, \overline{XPe}
---------	---

Although evidence of *nomina sacra* is present in most of the earliest manuscripts, it is unknown if they were present in any of the original autographs.²¹ After the 2nd century, this practice was expanded to include more words causing some to doubt whether the practice truly denoted sacred names or whether they were simply common abbreviations:

Lemma	Abbreviations
$\overline{\Delta\alpha\gamma\iota\Delta}$	$\overline{\Delta\alpha\Delta}$
\overline{MHTHP}	\overline{MHP} , \overline{MPC} , \overline{MPI} , \overline{MPA}
$\overline{OY\text{P}\alpha\text{NOC}}$	\overline{OYNOC} , \overline{OYNOI} , \overline{OYNOY} , $\overline{OYN\omega N}$, $\overline{OYN\omega}$, \overline{OYNOIC} , \overline{OYNON} , \overline{OYNOYC} , \overline{OYNE}
\overline{CWTHP}	\overline{CHP} , $\overline{C\omega P}$, \overline{CPC} , \overline{CPI} , \overline{CPA}

Numbers were also commonly abbreviated with a line written over a Greek letter, where each letter of the Greek alphabet also represents a numerical value. Such abbreviations are preceded by the “\$” symbol in the transcriptions.

Letter	Value
$\overline{\alpha}$	1
$\overline{\beta}$	2
$\overline{\gamma}$	3
$\overline{\delta}$	4
$\overline{\epsilon}$	5
$\overline{\zeta}, \overline{\varsigma}$	6
$\overline{\zeta}$	7
$\overline{\eta}$	8
$\overline{\theta}$	9

Letter	Value
$\overline{\iota}$	10
$\overline{\kappa}$	20
$\overline{\lambda}$	30
$\overline{\mu}$	40
$\overline{\nu}$	50
$\overline{\xi}$	60
$\overline{\omicron}$	70
$\overline{\pi}$	80
$\overline{\rho}$	90
$\overline{\rho}$	100

Letter	Value
$\overline{\rho}$	100
$\overline{\varsigma}$	200
$\overline{\tau}$	300
$\overline{\upsilon}$	400
$\overline{\phi}$	500
$\overline{\chi}$	600
$\overline{\psi}$	700
$\overline{\omega}$	800
$\overline{\lambda}$	900
$\overline{\iota\alpha}$	1000

The stigma and koppa characters were no longer used as letters in the Greek alphabet at that time, but they still retained their numerical values and were used for that purpose.

3.2.5 Scribal Corrections

Many of the early Greek manuscripts contain corrections either by the original scribe and/or later scribes who made subsequent changes to the manuscripts. These are encoded in the CNTR transcriptions by the “*” symbol preceding the original uncorrected reading in curly brackets “{}”, followed by the corrector symbol “^ab” preceding each subsequent correction in curly brackets. An omission is indicated by curly brackets with no contents in between. For example, in John 3:33 there is the correction:

*{} ^{τοϋτο} α{ουτοσ}

which indicates that the word was omitted by the original scribe, and then corrected by the original scribe to be “τοϋτο” and then corrected by a later scribe to read “ουτοσ”. Corrections are only included in the CNTR transcriptions that fall within the *terminus ad quem*.^{\$2.1}

3.2.6 Editorial Marks

The CNTR transcriptions contain several editorial marks which obviously were not part of the text of the Greek manuscripts, but provide additional information:

- The “_” symbol is used in the critical texts to indicate that a word in the original publication was divided in order to conform to CNTR word division conventions.^{§4.1.1} If a word begins with an underscore, then that word was originally appended to the end of the previous word. If an underscore appears in the middle of a word, then the word was originally divided into two words.
- The “+–” symbols indicate that an entire verse were present or absent respectively if they are not accompanied by any other letters. If they are followed by a ~ without any other accompanying letters it means that this was determined by vid.
- The “[]” symbols are used in some of the critical texts to indicate that words or letters in a reading are suspect. “A secondary reading consisting of the omission of words retained in the primary reading is marked by simple brackets [] in the text, enclosing the omitted word or words.”²²
- The “””” symbols are used to indicate the beginning and end of direct quotations, incorporations, and illusions respectively, but this feature has not been implemented yet. These symbols would only be used in class 2, class 3, class 5, and class 6 data.

The double bracketed text used in the 1885 Westcott and Hort, 2010 Society of Biblical Literature, and Nestle-Alan 28 texts was not retained since it was not considered to be part of the original text. “Double brackets [[]] have therefore been adopted...where the omitted words appeared to be derived from an external written or unwritten source, and had likewise exceptional claims to retention in the body of the text...”²³ Likewise, the colophons used at the ends of the 2005 Byzantine Textform texts were not retained because they were not considered to be part of the original text.

3.3 Accuracy

There is nothing special about a computerized transcription that makes it inherently immune from errors, and in fact all of the same types of common scribal errors were made at one time or another in the creation of the CNTR transcriptions. Consider that professional transcription standards which allow 1 error per 20,000 characters²⁴ would amount to about 37 errors in the New Testament. And not only that, but there were new types of *electronic* scribal errors that don’t have names, such as accidentally deleting a character by hitting the Del key instead of the PgDn key which is right below it, hitting stray characters when copying and pasting, and introducing artifacts from computer processing. Thus, whenever possible, the CNTR transcriptions were compared against electronic texts from other sources. The CNTR transcriptions often differ from other transcriptions in a number of ways: internal formatting codes used, transcription starting and stopping points; Unicode characters used; determinations of clear, damaged, or missing characters; scribal hand identification and inclusion; words reconstructed in lacuna; and occasionally there is not agreement on the words themselves! Thus, each transcription was normalized to a standard format before comparing them with other transcriptions. Any differences between the CNTR transcriptions and these other sources were resolved by examining images or publications of the original extant manuscripts. This comparative methodology does not guarantee that all CNTR transcriptions are without errors, but it greatly improves their accuracy compared to the other texts that were examined. In the cases where three or more original electronic transcriptions were compared, it is highly likely that those texts do not contain any transcription errors. Many CNTR transcriptions, however, were the first electronic transcriptions made publicly available so it is possible that they may still have some errors since there were no other electronic

transcriptions to compare them with at the time. One advantage of computerized transcriptions is that when errors are corrected, then tend to stay corrected.

While it is understandable that there would be some disagreements between the electronic transcriptions of manuscripts, what was unexpected is that the electronic transcriptions of the *modern* Greek critical texts also contained errors, and at a much higher rate! For example, most of the well-known modern Greek critical texts that were obtained from the Internet and various Bible programs contained errors! For example, here are the number of errors that were discovered when transcribing the 1885 Westcott and Hort text:

Errors	Source	Checked	Media
1015	Maurice Robinson	2003	Bible program
1000	Vincent Broman	2003	Internet
555	John Carras	2007	E-mail
278	Perseus Digital Library	2007	Internet

Similar errors were found in almost every other critical text found on the Internet and various Bible programs! Several errors were even found in the Nestle-Aland 28th online text hosted on the organization's own website!²⁵ Unfortunately, these same errors continue to find their way into different websites and Bible programs as copies are merely being made of these same errant copies. While some of this analysis was done several years ago, it would not be surprising if most of the copies available on various websites and Bible programs still contain the same errors.

Such errors have proliferated a new form of *electronic textual criticism*, whereas the lineage of an electronic manuscript can often be determined by examining a few verses that contain these known errors. It should be noted that *having multiple copies of an electronic text is not the same thing as having multiple unique transcriptions*. For example, many different downloads of the 1550 Stephanus text that are available today are all derivatives from the same transcription source²⁶ containing an orthography that is greatly different from the original text. Instead of complaining about all of the textual variants caused by scribes who made careless copies by hand, perhaps the critics ought to be concerned about all of these *new* electronic textual variants that are being propagated all over the world on the Internet! Ironically, errors were passed down by hand because there were no computers, and now errors are passed down by the computers!

¹ Microsoft Typography. <http://www.microsoft.com/typography/fonts/font.aspx?FID=15&FNAME=Symbol&FVER=2.00>, accessed May 9, 2007.

² <https://greekcntr.org/charutil/index.htm>

³ This font was created for the CNTR by Alan Bunning and is available for free download under the CC BY-SA 4.0 license at <https://greekcntr.org/downloads/KoineGreek.ttf>.

⁴ Nick Nicholas, "Greek Unicode Issues", sect. 6.1.1, July 29, 2006; <http://www.tlg.uci.edu/~opoudjis/unicode/unicode.html>; accessed November 26, 2012.

⁵ Using the regular sigma code for the lunate sigma allows text to easily be cut and paste into other programs without have to do a conversion. Sorting with the lunate sigma code would also have been difficult to manage. Philosophically, there was only one sigma that existed between rho and tau and that was the lunate sigma, so it can be argued that a separate code for the modern sigma should have been the one that was added out of sequence.

⁶ "Aristophanes of Byzantium", *The Columbia Encyclopedia*, 6th ed., Columbia University Press: New York, NY, 2006; <http://www.bartleby.com/65/ar/AristphByz.html>, accessed June 8, 2007.

⁷ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 227, Hodder & Stoughton: London, England, 1919.

⁸ Kurt and Barbara Aland, Erroll F. Rhodes tr., *The Text of the New Testament*, 2nd ed., p. 287, William B. Eerdmans Publishing Co.: Grand Rapids, MI, 1987.

⁹ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 242, Hodder & Stoughton: London, England, 1919.

¹⁰ Bruce M. Metzger, *A Textual Commentary On The Greek New Testament*, 4th rev. ed., p. 167, United Bible Societies: 1994.

¹¹ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 245, Hodder & Stoughton: London, England, 1919.

¹² Bruce Manning Metzger. *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 11, At the Clarendon Press: Oxford, United Kingdom, 1968.

¹³ Bruce Manning Metzger. *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 66, At the Clarendon Press: Oxford, United Kingdom, 1968.

¹⁴ This archaic script used for inscriptions is sometimes referred to as “capitals” because some of the modern capital letters (alpha, epsilon, xi, sigma and omega) letters were borrowed from it after the fact. But this is somewhat of a misnomer because they were not actually capital letters at that time; there was only a single form for each letter and there was no such thing as uppercase and lowercase letters.

¹⁵ Unicode® Standard Annex #15: Unicode Normalization Forms; <https://unicode.org/reports/tr15>.

¹⁶ International Congress of Orientalists. “Section Autonome des Papyrologues”, *Actes du XVIIIe Congrès international des Orientalistes, Leiden 7-12 septembre, 1931*, E.J. Brill: Leiden, Netherlands, 1932.

¹⁷ Sterling Dow, *Conventions in Editing: A Suggested Reformulation of the Leiden System (Greek, Roman, and Byzantine Study Aids 2)*, Duke University: Durham, NC: 1969.

¹⁸ Hans Krummrey and Silvio Panciera. “Criteri di edizione e segni diacritici.” *Tituli*, vol. 2, p.205-215, 1980.

¹⁹ Herbert C. Youtie, “Text and Context in Transcribing Papyri”, *Greek, Roman, and Byzantine Studies*, 7:251-258, 1966.

²⁰ The Bunning Heuristic Prototype (BHP) text was the first modern critical text to include *nomina sacra* in 2017; https://greekcentr.org/downloads/BHP_Introduction.pdf

²¹ Bruce Manning Metzger. *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 14, At the Clarendon Press: Oxford, United Kingdom, 1968.

²² Brooke Foss Westcott and Fenton John Anthony Hort, *The New Testament in the Original Greek*, vol. 2, p. 291, The Macmillan Company, Cambridge, 1881.

²³ Brooke Foss Westcott and Fenton John Anthony Hort, *The New Testament in the Original Greek*, vol. 2, p. 296, The Macmillan Company, Cambridge, 1881.

²⁴ Digital Library Production Service (DLPS), University of Michigan. “Text Encoding and Transcription Standards”; <http://www.umdl.umich.edu/docs/encodingstandards.html>, accessed June 8, 2007.

²⁵ For example, as of May 2016, the online Nestle-Aland 28th Edition had lost all of its double-brackets around words, left off an opening bracket at John 21:23, and introduced an undocumented textual change at James 5:10; and this was from their own text contained on their own website (<http://www.nestle-aland.com/en/read-na28-online>)!

²⁶ Maurice A. Robinson, *Stephens 1550 TR*. “The Stephens 1550 Textus Receptus edition, as printed in the George Ricker Berry Interlinear Greek NT volume.” <http://www.byztxt.com/download/STV-TR.ZIP>, accessed February 6, 2004.

4. Linguistic Parsing

Several additional features have been added to the CNTR transcriptions as aids to the reader that were not present in the original manuscripts. The CNTR was the first to lexically and morphologically parse all the earliest manuscripts and the major critical texts to a single uniform parsing scheme. This included about 35,000 words in early variants that had never been parsed before. Displaying the CNTR collation of texts in an interlinear format with lexical entries, morphological parsing, Enhanced Strong's Numbers, and contextual English word glosses allows those who have no knowledge of Greek to not only observe the textual differences, but to understand what the variants mean in a way that has never been possible before. Here is an example of some of the additional features described in the subsections below:

Word	Medieval Word	Medieval Lemma	ESN	Syn	Morph	Gloss	Punct	Caps
ΟΥΤΩC	οὕτως	οὕτως	37790	D.	thus	¶	P
ΓΑΡ	γάρ	γάρ	10630	CC	for		
ΗΓΑΠΗΣΕΝ	ἠγάπησεν	ἀγαπάω	250	VT	IAA3...S	loved		
Ο	ὁ	ὁ	35880	EANMS	-		
ΘΕ	Θεός	θεός	23160	N.NMS	God		G
ΤΟΝ	τόν	ὁ	35880	EAAMS	the		
ΚΟΣΜΟΝ	κόσμον	κόσμος	28890	N.AMS	world	,	
ΩCΤΕ	ὥστε	ὥστε	56200	CS	so_that		
ΤΟΝ	τόν	ὁ	35880	EAAMS	the		
ΥΙΟΝ	Υἱόν	υἱός	52070	N.AMS	Son		G
ΤΟΝ	τόν	ὁ	35880	EAAMS	-		
ΜΟΝΟΓΕΝΗ	μονογενῆ	μονογενής	34390	ARAMS	only_begotten		
ΕΔΩΚΕΝ	ἔδωκεν	δίδωμι	13250	VT	IAA3...S	he gave	,	
ΙΝΑ	ἵνα	ἵνα	24430	CS	in_order_that		
ΠΑC	πᾶς	πᾶς	39560	RINMS	everyone		
Ο	ὁ	ὁ	35880	RDNMS	-		
ΠΙCΤΕΥΩΝ	πιστεύων	πιστεύω	41000	VI	PPA.NMS	believing		
ΕΙC	εἰς	εἰς	15190	P.A..	in		
ΑΥΤΟΝ	αὐτόν	αὐτός	8460	RP	...3AMS	him		
ΜΗ	μή	μή	33610	D.	not		
ΑΠΟΛΗΤΑΙ	ἀπόληται	ἀπόλλυμι	6220	VI	SAM3...S	may perish	,	
ΑΛΛ	ἀλλ'	ἀλλά	2350	CC	but		
ΕΧΗ	ἔχη	ἔχω	21920	VT	SPA3...S	may_be having		
ΖΩΗΝ	ζωήν	ζωή	22220	N.AFS	life		
ΑΙΩΝΙΟΝ	αἰώνιον	αἰώνιος	1660	AAAFS	eternal	.	

Each word in the CNTR transcriptions is accompanied by its lexeme and its morphology in order to uniquely identify each occurrence of a word in a text:

- If the lexeme were missing, it would not be possible to distinguish between identical word forms that come from different lexemes. (Is ΗΛΙΟΥ from ΗΛΙΑC or ΗΛΙΟC?)
- If the morphology were missing, it would not be possible to distinguish between different conjugations or declensions of a word that result in the same word form. (Is ΤΕΚΝΟΝ nominative or neuter?)
- If the word were missing, it could not be reconstructed from the lemma and the morphology alone, because of orthographical differences. (Was the original word ἄλλ or ἄλλᾱ?)

Only the lexeme and the parsing, however, are necessary for assigning a unique “translational unit” to each entity, since the orthography of the underlying Greek word would be irrelevant.

Such additional features are sometimes subjective as there are often alternative interpretations, particularly with morphological parsing and word divisions, which correspondingly affect the meaning. Such a massive undertaking has never been done before, so please report any errors that you find that they might be duly corrected.

4.1 Words

A word is the smallest unit of syntax in a language that cannot be broken into smaller units capable of independent use. The CNTR transcriptions employ the use of word divisions as a convenience to the reader, even though they were not present in the original autographs of the New Testament. There were no spaces between words in the original Greek manuscripts, for they were written *scriptio continua* meaning that all words were run together without any consideration of word boundaries. This was not especially difficult to read, however, as the syllables in the words could be sounded out phonetically and the end of the words could usually be identified by either a vowel or the consonants Ν, Ρ, or C.¹ Consider the following example in English:

ITISNOTANYMOREDIFFICU
LTFORYOUTOREADTHISTH
ANITWASFORTHEGREEKS

Someone may recognize unintended words in this text such as “red”, “tore”, and “twas”, but then the sentence would not correctly follow rules of grammar and the context would not make any sense. On rare occasions, however, there are situations where the divisions of the words are ambiguous. For example, an English phrase such as, “GODISNOWHERE” could be read as either “God is nowhere” or “God is now here” resulting in radically different meanings. Also “BREAKFAST” could be one word which refers to a morning meal, or two words that indicate easting for the first time after a period of fasting. The correct meaning, of course, would be determined by the surrounding context. Examples in the New Testament Greek include “ἄλλοιC” vs. “ἄλλ οιC” (Mark 10:40), “οἰδαμεν” vs. “οἰδα μεν” (Rom. 7:14), “μαρὰν ἑα” vs. “μαρὰν ἑα” (1Cor. 16:22), and “ὁμολογοῦμεν ὡC” vs. “ὁμολογοῦμεν ὡC” (1Tim. 3:16).² Translation of words should never be decided on whether the editor chose to divide a word or not, but by the surrounding context on a case by case basis.

4.1.1 Word Boundaries

The CNTR transcriptions follow the standard linguistics rule for word divisions that *words should be divided into the smallest units possible that can stand alone as individual words without sacrificing any loss of meaning*. One common mistake that some have made is to parse the Greek words down to their smallest possible units in a “concordant method”³ which sacrifices the meaning of compound words. For example, consider the English word “understand” where the words “under” and “stand” can stand alone

separately, but when joined together have a completely different meaning (“to comprehend something” not “standing beneath something”). In this case, dividing the compound word into its apparent roots using a concordant method would clearly be invalid for it would result in a loss of meaning. Likewise, there are many examples where dividing a Greek compound word into its apparent roots would cause a loss of meaning, or worse, provide the wrong meaning. For example, if the word “ΑΠΟΚΡΙΝΟΜΑΙ” were divided into the words “ΑΠΟ” (meaning “from”) and “ΚΡΙΝΟΜΑΙ” (meaning “to judge”) it would literally mean “to judge from” which is quite different than its actual meaning “to answer”.

On the other hand, there are many compound words which can be split apart without any significant loss of meaning. For example, the English word “homeschool” can be divided into the words “home” and “school” which still means the same thing in either representation. There are many examples where the editors of the critical texts have divided words in an inconsistent manner which do not necessarily have any effect on the meaning. Some of these words include: “ΑΝΑ ΜΕΘΩΝ”, “ΑΠ ΑΡΤΙ”, “ΑΡΑ ΓΕ”, “ΔΙΑ ΠΑΝΤΟΣ”, “ΔΙΑ ΤΙ”, “ΕΙ ΓΕ”, “ΕΚ ΠΕΡΙΣΤΟΛΗΣ”, “ΕΞ ΑΥΤΗΣ”, “ΙΝΑ ΤΙ”, “ΚΑΙ ΤΟΙ ΓΕ”, “ΜΗ ΓΕ”, “ΜΗ ΠΩΣ”, “ΜΗ ΠΟΤΕ”, “ΜΗ ΤΙ(Σ)”, “ΟΥΚ ΕΤΙ”, and “ΤΟΥΤ ΕΣΤΙ(Ν)”. A.T. Robertson points out that “no very clear principles in this matter can be set forth”.⁴ In many cases, the differences in word divisions are not meaningful, since there was no space in the Greek and thus there was never any difference to them! For now, the CNTR transcriptions typically break these words according to the conventions of most of the critical texts, but this is entirely arbitrary and needs to be studied more thoroughly at a later date. Word divisions in non-conforming critical texts have been altered according to match this convention and the underscore character is used to show where the original word divisions were.^{§3.2.6}

4.1.2 Alternate Spellings

In Koine Greek, words were spelled phonetically and a word was not considered to be a spelling error as long as the specified letters led to the same pronunciation (which makes the words homophones). Thus, examples in English like “he fetched a pale of water” or “his faced turned pail” would have both been acceptable. Unlike critical texts which usually only show one sanitized form of a word, the CNTR database contains all forms of the words as they exist “in the wild”. Indeed, as A.T. Robertson states, “Among the strictly illiterate papyri writers one can find almost anything.”⁵ There were clearly preferred spellings that were derived by popular use, but there were not fixed spellings of canonical words as they are now found in modern lexicons. Friedrich Blass states that there was “no one fixed orthography in existence, but writers fluctuated between the old historical spelling and a new phonetic manner of writing.”⁶

An alternative spelling is defined here then as word that is a homophone of another word in adherence to standard set of common phonetical substitutions.⁷ The set of phonetic rules include the following sound-alike letter combinations:

Α = ΑΑ, Β = ΒΒ, Δ = ΔΔ, Γ = ΓΓ, Θ = ΘΘ, Κ = ΚΚ, Λ = ΛΛ, Μ = ΜΜ, Ν = ΝΝ, Π = ΠΠ, Ρ = ΡΡ, Σ = ΣΣ, Τ = ΤΤ,
Γ = Κ, Γ = Χ, Κ = Χ, Δ = Θ, Δ = Τ, Θ = Τ, Ζ = Σ, Μ = Ν, Π = Φ,
ΓΞ = Ξ, ΤΘ = ΤΤ, ΜΠ = Π, ΜΨ = Ψ, ΝΒ = ΜΒ, ΝΓ = ΓΓ, ΝΚ = ΓΚ, ΝΛ = ΛΛ, ΝΜ = ΜΜ, ΝΞ = ΓΞ, ΝΠ = ΜΠ,
ΝΧ = ΓΧ, ΝΦ = ΜΦ, ΝΨ = ΜΨ, ΝΣ = Σ, ΝΤ = ΤΤ, ΝΖ = Ζ,
Α = ΑΥ, Ε = ΑΙ, Ι = ΕΙ, Ι = Η, Η = ΕΙ, Ο = Ω, Υ = ΟΙ, Υ = ΟΥ, ΟΙ = ΟΥ, Ω = ΩΙ, Η = ΗΙ

In addition to these rules, other orthographical considerations must be taken into account when parsing words such as movable nu or sigma, elision, *nomina sacra*, numeric abbreviations, etc.

In some cases, the alternative spelling of one word can be the same as a completely different word. For example, the two different words “ΚΑΙΝΟΣ” (meaning “new”) and “ΚΕΝΟΣ” (meaning “empty”) are both alternative spellings of each other according to the common phonetical substitution “ΑΙ” = “Ε”, and both spellings are used for both words in the early manuscripts (1Cor. 15:58, 2Cor. 6:1, Gal. 2:2, Eph. 4:5, Phil. 2:16, Col. 2:8, Jas. 2:20, 2Pet. 3:13, Rev. 2:17, Rev. 3:12, Rev. 21:1, 21:5), and in some cases the meaning is ambiguous (Eph. 4:5, Col. 2:8, 2Pet. 3:13, Rev. 21:1).

When parsing homophones in these cases, the CNTR used an *orthographical priority* approach in which the meaning of the word would first be determined by its canonical spelling if possible (i.e. “ΚΑΙΝΟC” means “new” and “ΚΕΝΟC” means “empty”) as long as that word could make sense in context, otherwise the meaning of a homophone was considered. For example, Codex Washingtonianus (032) always spells the homophones “ΤΠΙC” and “ΤΡΕΙC” as “ΤΠΙC” regardless of meaning, and thus the word meaning must be determined by its usage in context. Likewise, words that have conjugation endings that sounded the same such as -ΤΑΙ and -ΤΕ are treated using the same approach. To designate the meaning of a word solely by its canonical spelling would not only disregard the phonetical aspects of Koine Greek, but it would force many nonsensical readings which could have easily been explained by orthographical differences. Other examples of similar homophones include: ΓΕΝΕCΙC vs. ΓΕΝΝΗCΙC (Matt. 1:18, Luke 1:14), ΙΔΗΤΕ vs. ΕΙΔΗΤΕ/ΕΙΔΗΤΑΙ (Matt. 9:6, 24:33, Mark 2:10, 13:14, Luke 5:24, 12:54, 21:20, 21:31), CΥNHΛACCEN vs. CΥNHΛΛACCEN (Acts 7:26), and ΤΠΙC vs. ΤΡΕΙC (Matt. 26:34,75, Mark 14:30,72, Luke 22:34,61, John 13:38, Acts 10:16, 11:10, 2Cor. 11:25, 12:8).

4.1.3 Misspellings

A misspelled word contains inserted, omitted, substituted, or transposed characters so that it is not an exact phonetical match with any canonical word form, yet can still be recognized by its contextual position. Such words are encoded in the CNTR database in a separate field with the precise character transformations that occurred in order to consider the word to be misspelled. These different types of transformations can then be easily queried which is useful in studying specific scribal habits in one manuscript or across multiple manuscripts. The internal codes used in the CNTR database are as follows:

- + Inserted characters
- Omitted characters
- > Dittography (inserted characters of a repeated pattern)
- < Haplology (omitted characters of a similar sounding syllable)
- : Substituted characters
- ~ Transposed characters

When confronted with a series of ambiguous letters, the CNTR transcriptions follow a principle of *scribal deference* in which the scribe was first given every benefit of the doubt to have produced valid words that could make sense in context (considering alternative spellings and word boundaries), before a word was considered to be a spelling error. For example, in Matthew 4:13, most of the texts contain “ΠΑΡΑΘΑΛΑCCIΑΝ” as a substantive adjective, but both Codex Washingtonianus (032) and the uncorrected hand of Codex Sinaiticus (01) specify “ΠΑΡΑΘΑΛΑCCΑΝ” which is not an attested word. Rather than consider that two different scribes had both made the same phonetical spelling error by omitting an iota, the letters instead were interpreted as two separate words giving the prepositional phrase “ΠΑΡΑ ΘΑΛΑCCΑΝ” with “ΘΑΛΑCCΑΝ” used as a noun. Surprisingly, these types of examples were encountered several times among the CNTR transcriptions. Alternative word spellings due to different dialects are not considered misspellings but are attested words that are represented with different lemmas within the same lexeme.

4.1.4 Unknown words

An unknown word contains one or more unknown characters^{[§Error! Reference source not found.](#)} and cannot be identified as any word or even a misspelling of a word in any known variant and cannot be speculated as to what it might mean. These words therefore have an unknown lexeme (encoded as 99999) but are still given their own column position in the CNTR collation. There are only a handful of such words throughout all of

the CNTR transcriptions and it is possible that some of them could match words found in variant readings that we do not yet possess.

4.2 Lexicography

Lexicography is a part of linguistics that is concerned with compiling and editing dictionaries for a language. It should be noted that there are many words in the CNTR transcriptions that were not covered by any previous lexicon, including the LSJ⁸ and the BDAG⁹ lexicons.

4.2.1 Lemmas

A lemma represents the abstract form of a word and is typically used as the headword to a lexical entry. Every word in the CNTR transcriptions has been identified with its associated lemma so that the definition of every word can be looked up in a Greek lexicon. The rules governing what constitutes a separate lexical entry vary between different Greek lexicons and are often applied inconsistently within the same lexicon.¹⁰ The CNTR lexicon attempts to adhere to a set of consistent rules which include some of the following conventions:

- Homophones which are alternative spellings of the same word are not given separate lexical entries since they are phonetically pronounced the same and represent the same word with the same meaning,^{§4.1.2} (but the most popular alternative spellings are still retained in a separate field). For example, “ΠΕΙΛΑΤΟΣ” and “ΠΙΛΑΤΟΣ” are simply two different spellings of the same word.
- Homophones that were derived from different etymologies are given separate lexical entries because they are different words with different meanings that just happened to be pronounced the same. For example, “ΒΑΤΟΣ” has two different lexical entries because one is a masculine noun of Hebrew origin referring to a liquid measure, and the other is a feminine noun of Hellenistic origin referring to a thorn-bush.
- Words that follow different paradigms are given separate lexical entries. For example, “ΜΑΡΙΑ” and “ΜΑΡΙΑΜ” represent two different lemmas because “ΜΑΡΙΑ” is declinable but “ΜΑΡΙΑΜ” is not, but both would be assigned to the same lexeme.
- Words that represent different dialects of the same word are given separate lexical entries since they are not phonetically pronounced the same. For example, “ΑΙΤΙΩΜΑ” and “ΑΙΤΙΑΜΑ” represent two different lemmas of the same lexeme because they are not phonetical equivalents (much like “ain’t” is often given a separate entry in English dictionaries for “isn’t”).
- Adverbs which were formed from neuter adjectives are given their own separate lexical entries. For example, “ΔΕΥΤΕΡΟΣ” (which is a determiner meaning “second”) and “ΔΕΥΤΕΡΟΝ” (which is an adverb meaning “secondly”) are given separate entries. A.T. Robinson points out that grammarians have generally neglected the adverb and discusses the common formation of adverbs from neuter adjectives.¹¹ Morphologically speaking, adverbs are not declinable and thus do not belong under their related adjective entries which are declined. (This is *not* the same as substantive adjectives which are equally declinable.)
- Comparatives and superlatives are given their own separate lexical entries. Morphologically speaking, suffixes such as “ΕΤΕΡΟΣ” or “ΤΑΤΟΣ” are not declensions, nor are they especially different than other types of adjective suffixes such as “ΗΤΟΣ”, “ΟΙΟΣ”, “ΙΝΟΣ”, or “ΙΚΟΣ” which form other adjectives that are given separate lexical entries. Neither are comparatives and superlative forms predictable since other suffixes can be used to form them and some of the forms are quite irregular.

- Diminutives are given their own separate lexical entries. There are several suffixes that are used to form diminutives, but not every noun that ends in “ION” is necessarily a diminutive. Swanson’s list of diminutives was consulted for establishing distinct lexical entries.¹²

Subcategories for nomina sacra, numeric abbreviations, and irregular principal parts are also recorded in the CNTR database. All words (and misspelled words) in the CNTR transcriptions were assigned to the associated lemmas taking into account the orthographical priority^{§4.1.2} and scribal deference^{§4.1.3} principles described above.

4.2.2 Lexemes

A lexeme represents all word forms which cover the same semantic domain(s) of meaning. Each lemma is associated with the proper lexeme, and multiple lemmas can be associated with the same lexeme. For example, the following lemmas are all associated together with one lexeme:

Lemma	Word	Count	Form
ΙΕΡΟCΟΛΥΜΑ	ΙΕΡΟCΟΛΥΜΑ	198	Canonical
	ΙΕΡΟCΟΛΥΜΟΙC	85	Canonical
	ΙΕΡΟCΟΛΥΜΩΝ	61	Canonical
	ΕΙΕΡΟCΟΛΥΜΑ	2	Alternative spelling
	ΕΡΟCΟΛΥΜΑ	2	Alternative spelling
	ΙΕΡΟCΟΛΜΑ	1	Misspelling
	ΙΕΡΟCΟΛΟΜΟΙC	1	Misspelling
	ΙΕΡΟCΟΛΥCΟΛΥΜΑ	1	Misspelling
	ΙΕΡΟΥCΑΛΥΜΗΜΑ	1	Misspelling
	ΡΟCΟΛΥΜΑ	1	Misspelling
ΙΕΡΟΥCΑΛΗΜ	ΙΕΡΟΥCΑΛΗΜ	236	Canonical
	ΕΙΕΡΟΥCΑΛΗΜ	2	Alternative spelling
	ΙΕΡΟΥΕΡΟΥCΑΛΗΜ	1	Misspelling
ΙΑΗΜ	ΙΑΗΜ	142	Abbreviation
	ΙΑΗ	1	Abbreviation misspelling
ΙΗΛΑΜ	ΙΗΛΑΜ	33	Abbreviation

In this example, four distinct lemmas have been grouped together as one lexeme that all cover the same range of meanings. The various lemmas grouped together in a lexeme could come from different etymologies as long as the semantic domain(s) of meaning is the same.

Again, it should be stressed that there was not necessarily one correct form of a word, and thus it is not necessary to designate one particular lemma to represent the entire lexeme. A.T. Roberston stresses, “There has never been a fixed orthography for the Greek tongue at any stage of its history.”¹³ The frequency count showing the number of occurrences among the early manuscripts was used to help determine the difference between the preferred forms and alternative spellings. This does not mean that any particular word form was necessarily “wrong”, but it does indicate the most popular forms of usage. A special lexeme was also created to represent unknown words.

4.2.3 Enhanced Strong’s Numbers

While any numbering system could have been used to represent the lexemes, the CNTR has created a system using Enhanced Strong’s Numbers (ESN) to represent the meanings of the words in the CNTR

transcriptions. Although the definitions from the Strong's Exhaustive Concordance¹⁴ are inadequate for scholarly use for a number of reasons, the Strong's numbering system itself has become the de facto standard and is extensively used in many other reference works and Bible programs. Thus, in order to maintain some "backward compatibility" with Strong's numbers, the ESNs were created by modifying the Strong's numbering system in the following manner:

1. Numbers that were assigned to redundant or derivative forms of the same lexeme were eliminated and replaced by the number of the preferred lexical form.

Number	Lemma	Eliminated Numbers
40	ΗΓΙΟC	39
71	ΑΓΩ	33
150	ΔΙCΧΡΟC	149
159	ΔΙΤΙΟC	158
217	ΑΛΑC	251
483	ΑΝΤΙΛΕΓΩ	471
568	ΑΠΕΧΩ	566, 567
681	ΑΡΤΩ	680
757	ΑΡΧΩ	756
846	ΑΥΤΟC	848
872	ΑΦΟΡΑΩ	542
934	ΒΑCΙΛΕΙΟC	933
1182	ΔΕΚΑΤΟC	1181
1210	ΔΕΩ	1163
1427	ΔΕΚΑΔΥΟ	1177
1473	ΕΓΩ	1691, 1698, 1700, 3165, 3427, 3450, 2248, 2249, 2254, 2257
1487	ΕΙΠΕΡ	1512
1492	ΕΙΔΩ	2396, 2400, 2467
1510	ΕΙΜΙ	1488, 1498, 1511, 1526, 2070, 2071, 2075, 2076, 2077, 2252, 2258, 2277, 2468, 5600, 5607
1520	ΕΙC	3391
1667	ΕΛΙCΩ	1507
1826	ΕΧΕΙΜΙ	1832
1893	ΕΡΕΙΠΕΡ	1897
2068	ΕCΘΙΩ	5315
2207	ΖΗΛΩΤΗC	2208
2280	ΘΑΔΔΑΙΟC	3002
2414	ΙΕΡΟCΟΛΥΜΑ	2419
2455	ΙΟΥΔΑC	2448
2500	ΙΩCΗC	2499
2909	ΚΡΕΙΤΤΩΝ	2908
3004	ΛΕΓΩ	2036, 2046, 4483
3017	ΛΕΥΙ	3018
3062	ΛΟΙΠΟC	3064
3187	ΜΕΙΖΩΝ	3185, 3186
3189	ΜΕΛΑC	3188
3398	ΜΙΚΡΟC	3397
3403	ΜΙΜΝΗCΚΩ	3415
3441	ΜΟΝΟC	3440
3588	Ο	5120
3739	ΟCΤΕΡ	3746

3748	ΟCΤΙC	3755
3778	ΟΥΤΟC	5023, 5025, 5026, 5124, 5125, 5126, 5127, 5128, 5129, 5130
4239	ΠΡΑΥC	4235
4240	ΠΡΑΥΤΗC	4236
4302	ΠΡΟΛΕΓΩ	4277, 4280
4308	ΠΡΟΟΡΑΩ	4275
4387	ΠΡΟΤΕΡΟC	4386
4413	ΠΡΩΤΟC	4412
4496	ΡΙΠΩ	4495
4556	CΑΡΔΙΟΝ	4555
4566	CΑΤΑΝ	4567
4569	CΑΥΛΟC	4549
4771	CΥ	4571, 4671, 4675, 5209, 5210, 5213, 5216
5036	ΤΑΧΥC	5035
5225	ΥΠΑΡΧΩ	5224
5275	ΥΠΟΛΕΙΠΩ	5277
5306	ΥCΤΕΡΟC	5305

2. Numbers that were assigned to lemmas in improper lexical form were updated if there was not an associated number for the proper lexical form.

Number	Lemma
5504	ΕΧΘΕC
3062	ΛΟΙΠΟC
4287	ΠΡΟΘΕCΜΙΑ
4992	CΩΤΗΡΙΟC

3. Numbers that were assigned to compound words and phrases were eliminated if they were separated according to the CNTR word division rules.^{\$0} Some of these entries, however, will be retained in the CNTR lexicon as “see” entries pointing to the separated words.

Number	Lemma	Substituted Numbers
534	ΑΠΤΑΡΤΙ	575, 737
697	ΑΡΕΙΟC ΠΑΓΟC	6965, 3803.5
1275	ΔΙΑ ΠΑΝΤΟC	1223, 3956
1302	ΔΙΑΤΙ	1223, 5101
1489	ΕΙΓΕ	1487, 1065
1499	ΕΙ ΚΑΙ	1487, 2532
1508	ΕΙ ΜΗ	1487, 3361
1509	ΕΙ ΜΗΤΙ	1487, 3385
1513	ΕΙ ΠΩC	1487, 4458
1527	ΕΙC ΚΑΘ ΕΙC	1520, 2596, 1520
1536	ΕΙ ΤΙC	1487, 5100
1768	ΕΝΝΕΝΚΟΝΤΑΕΝΝΕΑ	1752.5, 1767
1888	ΕΠ ΑΥΤΟΦΩΡΟC	1909, 847.5
2444	ΙΝΑΤΙ	2443, 5101
2534	ΚΑΙ ΓΕ	2532, 1065
2568	ΚΑΛΟΙ ΛΙΜΕΝΕC	2570, 3040
2651	ΚΑΤΑΜΟΝΑC	2596, 3441
3362	ΕΑΝ ΜΗ	1437, 3361
3363	ΙΝΑ ΜΗ	2443, 3361
3364	ΟΥ ΜΗ	3756, 3361

3378	ΜΗ ΟΥΚ	3361, 3756
3387	ΜΗΤΙC	3361, 5100
3569	ΤΑ ΝΥΝ	3588, 3568
3603	Ο ΕCΤΙ	3739, 1510
3381	ΜΗ ΠΩC	3361, 4459
3386	ΜΗΤΙ ΓΕ	3385, 1065
3801	Ο ΩΝ ΚΑΙ Ο ΗΝ ΚΑΙ Ο ΕΡΧΟΜΕΝΟC	3588, 1510, 2532, 3588, 1510, 2532, 3588, 2064
5123	ΤΟΥΤ ΕCΤΙΝ	3778, 1510
5516	ΧΖC	1812, 1835, 1803

4. Numbers that were assigned to words that were not used in the CNTR transcriptions were eliminated. The numbers 2717 and 3203-3302 were eliminated because they were not originally used in the Strong's Greek Dictionary.

Number	Lemma
1418	ΔΥC
1970	ΕΠΙΠΝΙΓΩ
2312'	ΘΕΟΛΟΓΟC
2526'	ΚΑΘΟΛΙΚΟC
2858	ΚΟΛΑCCAΕΥC
2997	ΛΑCΧΩ
3390	ΜΗΤΡΟΠΟΛΙC
3505	ΝΕΡΩΝ
3620	ΟΙΚΟΔΟΜΙΑ
3818	ΠΑΚΑΤΙΑΝΗ
4452	ΠΩ
5104	ΤΟΙ

5. After this, almost 400 new numbers were assigned to words used in the CNTR transcriptions that were not found in the Strong's Greek Dictionary. Others have dealt with this problem by placing all of the words at the end, giving them higher numbers in a specific range, but this places them out of sorted order and any new words added in the future would continue to be placed out of order. To remedy this problem, the CNTR database merely adds another digit to the right of *all* Strong's numbers to allow new lexemes to be added. For example, Strong's number 2424 for ΙΗCΟΥC becomes the number 24240. By making use of this additional digit, new lexical entries are then inserted into the correct places while maintaining Strong's sorted order. For example, three new words have been inserted between the existing Strong's words 647 ΑΠΟCΤΑCΙΟΝ and 648 ΑΠΟCΤΕΓΑΖΩ as follows:

Number	Lemma
6470	ΑΠΟCΤΑCΙΟΝ
6473	ΑΠΟCΤΑCΙC
6475	ΑΠΟCΤΑΤΕΩ
6477	ΑΠΟCΤΑΤΗC
6480	ΑΠΟCΤΕΓΑΖΩ

The words are displayed in the CNTR website using a decimal point: 647, 647.3, 647.5, 647.7, 648.

It should be pointed out that not all lemmas will be in sorted order using this scheme since different orthographical forms of lemmas of the same lexeme may result in different alphabetical positions. There were also several entries in Strong's that were not numbered in alphabetical order to begin with (2858, 2994, 3451, 4211, 4696, 3569, 5506). The number itself is not especially important as it merely represents a lexical unit of meaning, so the fact that most of the lemmas associated with the number will be in sorted order is an incidental advantage.

4.3 Morphology

Morphology is a part of linguistics that studies patterns of word-formation and attempts to formulate rules that model the implementation of a language. There are several different morphological schemes with different grammatical categories that have been applied to the Greek New Testament over the years. And within any given scheme, there are many word forms that are ambiguous which can be parsed in multiple ways, resulting in more than one possible meaning. The general approach of the CNTR parsing scheme is to keep the syntactical, morphological, and lexical attributes of a word separate. For example, the lexical attributes of a word (such as whether a word is a number, abbreviation, indeclinable, diminutive, etc.) do not change depending on a word's particular occurrence in a text and thus do not need to be encoded with the morphology for each word. Such lexical elements could be *displayed* with the morphology in computer programs, but in that case, they merely need to be retrieved from the lexicon. Likewise, the syntactical attributes of a word related to sentence structure are also distinct from the morphological form of a word and are kept separately. Ultimately, other parsing scheme's such as Robinson's¹⁵ and Tauber's¹⁶ can be reconstructed from the morphological parsing system used here without data loss, and indeed, both of them have been compared with the CNTR's texts resulting in greater accuracy.

The CNTR transcriptions are encoded with a parsing scheme where the morphological form of each Greek word in the New Testament is represented by a nine-character code:

Syntactical		Morphological						
Role	Type	Mood	Tense	Voice	Person	Case	Gender	Number
Noun	Substantive adj.					Nominative	Masculine	Singular
	Predicate adj.					Genitive Dative Accusative Vocative	Feminine Neuter	Plural
Adjective	Ascriptive					Nominative	Masculine	Singular
	Restrictive					Genitive Dative Accusative Vocative	Feminine Neuter	Plural
determiner	Article Demonstrative differential Possessive Quantifier Number Ordinal Relative interrogative				1st 2nd 3rd	Nominative Genitive Dative Accusative Vocative	Masculine Feminine Neuter	Singular Plural

pRounoun	Demonstrative Personal rEflexive reCiprocal Indefinite Relative inTerrogative				1st 2nd 3rd	Nominative Genitive Dative Accusative	Masculine Feminine Neuter	Singular Plural
Verb	Transitive Intransitive Linking Modal Periphrastic	Indicative iMperative Subjunctive Optative iNfinitive Participle	Present Imperfect Future Aorist pErfect pLuperfect	Active Middle Passive	1st 2nd 3rd	Nominative Genitive Dative Accusative Vocative	Masculine Feminine Neuter	Singular Plural
Interjection	Exclamation Directive Response	iMperative	Aorist	Active	2nd			Singular Plural
Preposition	Improper					Genitive Dative Accusative		
aDverb	cORrelative							
Conjunction	Coordinating Subordinating cORrelative							
parTicle	Foreign Error							

The syntactic function of a word, of course, determines which of the other morphological attributes may apply. Closed function morphemes such as determiners, pronouns, and conjunctions have detailed grammatical subtypes that are different from open content morphemes such as nouns, adjectives, adverbs, and verbs which could be categorized by semantic domains of meaning obtainable from the lexicon.¹⁷ The CNTR parsing scheme offers greater linguistic distinctions that are not found in other parsing schemes:

- The substantive noun subtype is used for marking adjectives that are used substantively. Sometimes this can lend itself to a difference in translation. For example, “*ἀμαρτωλός*” is usually translated as “sinful” when used as an adjective and “sinner” when used as a substantive adjective (instead of “sinful [one]”). Likewise, the predicate noun subtype is used for marking predicate adjectives for similar reasons.
- The ascriptive adjective subtype is used for marking descriptive adjectives that may occur either before or after the noun they modify. The restrictive adjective subtype is used for distinguishing the article/noun/article/adjective pattern which could be translated either as a substantive (i.e. “Jesus, the Nazarene”) or as a regular adjective (i.e. “the Nazarean Jesus”).
- Determiners are a relatively new concept in the field of linguistics,¹⁸ but are significantly different than adjectives in both syntax and meaning.^{19,20} Concerning syntax, determiners can occupy syntactical positions that do not apply to descriptive adjectives. For example, you could say, “some happy people”, but not “happy some people”. Concerning meaning, determiners are typically not gradable and cannot form comparatives or superlatives. For example, you could say “very happy”, “happier”, or “happiest”, but not “very some”, “somer” or “somes”.
- Determiners are also distinct from pronouns in that a word normally used as a determiner only becomes a pronoun when it is used substantively. For example, the word “*οὗτος*” is always marked as a

demonstrative pronoun in many parsing schemes, but actually it is only a demonstrative pronoun when it is used substantively (i.e. “This is where I live.”) and is a demonstrative determiner when it modifies a noun (i.e. “I live in this house.”).

- The correlative adverb subtype is used for marking the head of paired correlative conjunctions such as “either...or”, “both...and”, etc. This concept is based on some general syntactical analysis in linguistics²¹ along with the observation that the words that serve as the head of correlative conjunctions are normally considered adverbs.
- The transitive verb subtype is used to mark verbs that have an explicit direct object (normally indicated by the accusative case²²) or could be a clause. The direct object must exist syntactically, supplied ideas do not count. Passive verbs are normally marked as intransitive unless an associated direct object is present.
- The modal verb subtype is used for marking verbs that are coupled with another infinitive verb. This is not necessarily how modality is traditionally understood linguistically, although there is a great amount of overlap between the concepts.

For those who prefer a simpler parsing system, it is quite easy to convert these codes to broader parsing categories. For example, if someone preferred to consider all of the determiners to be adjectives, they can easily change all of the different “E” codes to the single “A” code, but the converse is not true – it is a one-way street. It is easy to lose information and reduce the CNTR codes to Robinson’s or Tauber’s parsing system, but their codes cannot be converted to this parsing system because they lack the necessary granularity of information.

Here are some examples of this morphological parsing system:

Word	Lemma	Syn	Morph	Description
ΑΓΑΠΑΙC	ΑΓΑΠΗ	N. DFP	dative feminine plural noun
ΑΓΙΟΥ	ΑΓΙΟC	A. GNS	genitive neuter singular adjective
Ο	Ο	EA NMS	nominative masculine singular definite article determiner
ΠΑΝΤΑC	ΠΑC	EQ AMP	accusative masculine plural quantifier determiner
ΕΜΕ	ΕΓΩ	RP	. . . 1A.S	first person accusative singular personal pronoun
ΕΜΑΥΤΟΥ	ΕΜΑΥΤΟC	RR	. . . 1GMS	genitive masculine singular first person reflexive pronoun
ΗΛΘΟΝ	ΕΡΧΟΜΑΙ	V.	IAA3 . . P	indicative aorist active third person plural verb
ΛΕΓΩΝ	ΛΕΓΩ	V.	PPA . NMS	present active nominative masculine singular participle
ΕΚ	ΕΚ	P. G . .	preposition (used with genitive case)

Words formed by krasis were parsed primarily to reflect the force of the second word. For example, “ΚΑΓΩ” is considered a type of personal pronoun rather than a type of conjunction. Neuter adjectives which share an adverbial form are parsed as adjectives if they are preceded by an article or preposition.

There are also orthographical considerations similar to those previously discussed^{§4.2.2} regarding words that are homophones. The issue this time is not in identifying the lemma, but identifying the correct morphological form of the word. For example, there are hundreds of verbs with the ending of “ΤΑΙ” or “ΤΕ” containing the common phonetical substitution “ΑΙ” = “Ε” that are used interchangeably. While the context may indicate a preference between a third person singular word or a second person plural word, there are occasions where the choice is ambiguous. The CNTR’s approach was always to choose the morphological form according to the preferred spelling (i.e. “ΤΑΙ” is third person singular and “ΤΕ” is second person plural). Again, this occasionally makes nonsensical readings which could have easily been explained by orthographical differences, but it preserves all the possible variant readings which was deemed to be more important.

4.4 English Glosses

Each Greek word in the CNTR transcriptions has been associated with the closest equivalent English word(s) according to its usage and context, essentially providing an English/Greek interlinear for all CNTR transcriptions. In some Bible interlinear programs, only one general gloss is assigned to each lexical entry, but the CNTR assigns a separate context-sensitive gloss for each different meaning of a lexical entry. It is a classic mistake by novice students to assume that every Greek word can be represented by the meaning of a single English word. Such a “concordant method”²³ often fails to acknowledge the diverse wealth of meanings that can be contained in a single homonym. For example, what single meaning would you ascribe to the word “draft”? Does it involve a leaky door, military conscription, or a preliminary writing? Similarly, in Greek the word “ΚΟΣΜΟΣ” could mean “adornment”, “world”, or “universe” which cover three different senses of meaning.

On the other hand, some Bible interlinear programs and translations use multiple synonyms for a single word in an inconsistent manner, even when it contains the same usage and meaning. They just use whatever word sounds good in one particular context without considering what words they previously used in the exact same context. For example, the King James Bible translates “ΘΑΠΤΙΣΙΣ” as “afflicted”, “affliction”, “anguish”, “burdened”, “persecution”, “tribulation”, and “trouble”, when one or two glosses could have sufficed for all of them. The BDAG seems to be guilty of this problem as well as it often invents numerous unnecessary subcategories for a word that really only has a few senses of meaning. For example, the entry for “ΠΡΟΦΗΤΗΣ” has six major categories (which probably should have been subcategories or examples of usage) covering a full page of material when in all cases the word could simply be translated as “prophet”. Many other large BDAG entries could be reduced to just two or three different senses of meaning.²⁴

To address both these issues, the CNTR has selected English glosses according to the following principles:

1. Identify the fewest number of unique senses possible for each Greek word. These typically correspond to the top-level meanings of a normal lexical entry. For example, the word “ΚΟΣΜΟΣ” would have three different senses, and “ΠΡΟΦΗΤΗΣ” would only have one sense as cited in the examples above.
2. The word form of each English gloss should reflect the same part of speech as the associated Greek word. In other words, Greek verbs should be rendered in English verbs, adjectives should be rendered as adjectives, adverbs should be rendered as adverbs, etc.
3. Reuse the same English components (roots, prefixes, and suffixes) to correspond to the equivalent Greek components whenever possible. For example, “ΚΗΡΥCΣΩ”, “ΚΗΡΥΞ”, and “ΚΗΡΥΓΜΑ” are associated with “proclaim”, “proclaimer”, and “proclamation” respectively.
4. Assign different English glosses to different Greek words whenever possible, as long as there could be a discernable difference in meaning. For example, “ΚΕΙΜΑΙ” is associated with “lie”, while “ΚΑΤΑΚΕΙΜΑΙ” is associated with “lie down” even though “down” is seemingly redundant to us in English.
5. If two different Greek words are synonyms in meaning, associate them with two equivalent English synonyms whenever possible. For example, “ΒΡΕΦΟΣ” is associated with “baby” and “ΠΑΙΔΙΟΝ” is associated with “infant”, even though they are essentially interchangeable in both Greek and English. This then makes little difference in meaning, but can still indicate that a different Greek word was used by looking at the glosses.

English glosses then are *consistently* assigned to each Greek word so that the same Greek word is *always* used to represent the same sense of meaning according to its usage and context. These glosses are also directly tied to the CNTR lexicon, typically representing the top-level meanings for each lexical entry. Subcategories showing the different senses and usages of a word within each meaning may be further

delineated. Reliance on such glosses, of course, can lead to a very stilted translation, but are still adequate for conveying the correct semantic domains of meanings for most words.

¹ Bruce Manning Metzger, *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 13, At the Clarendon Press: Oxford, United Kingdom, 1968.

² Bruce Manning Metzger, *The Text of the New Testament: Its Transmission, Corruption, and Restoration*, 2nd ed, p. 13, At the Clarendon Press: Oxford, United Kingdom, 1968.

³ A. E. Knoch, “The Concordant Method”, Concordant Greek Text, Concordant Publishing Concern: 1975; http://www.scripture4all.org/help/isa2/Articles/The_Concordant_Method/The_Concordant_Method.htm, accessed April 14, 2017. The Concordant Method is not without value, however, as its glosses for the root meanings in some cases may better express how Greeks actually thought about the meaning of some words that are not easily translated into English.

⁴ A. T. Robertson, *A Grammar of the Greek New Testament in the Light of Historical Research*, 3rd ed., p. 244, Hodder & Stoughton: London, England, 1919.

⁵ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 181, Hodder & Stoughton: London, England, 1919.

⁶ Friedrich Blass, *Grammar of New Testament Greek*, p. 6, MacMillan and Co.: New York, NY, 1898.

⁷ F. Blass and A. Debrunner, trans. Robert W. Funk, *A Greek Grammar of the New Testament and Other Early Christian Literature*, p. 13-20, University of Chicago Press, Chicago and London, 1961. Carl Darling Buck, *Introduction to the Study of the Greek Dialects*, p. 17-43, Ginn and Company: Boston, 1910. Randall Buth, “Notes on the Pronunciation System of Koiné Greek”, 2012; <https://www.biblicallanguagecenter.com/koine-greek-pronunciation>, accessed August 11, 2018.

⁸ Henry George Liddell, Robert Scott, and Henry Stuart Jones, *A Greek-English Lexicon*, 9th ed., Clarendon Press: Oxford, 1925.

⁹ W. Bauer, F. W. Danker, W. F. Arndt, and F. W. Gingrich, *A Greek-English Lexicon of the New Testament and Other Early Christian Literature (BDAG)*, 3rd ed., University of Chicago Press: Chicago, 2000.

¹⁰ For example, for the conventions listed, the BDAG sometimes has a separate lexical entry, sometimes tucks the category under an existing lexical entry, and sometimes has a “see” entry – all in an inconsistent manner. Other lexicons usually fare worse.

¹¹ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 292-295, Hodder & Stoughton: London, England, 1919.

¹² Donald C. Swanson, “Diminutives in the New Testament”, *Journal of Biblical Literature*, vol. 77, no. 2, p. 134-151, June 1958.

¹³ A. T. Robertson, *A Grammar of the Greek New Testament In the Light of Historical Research*, 3rd ed., p. 177, Hodder & Stoughton: London, England, 1919.

¹⁴ James Strong, *Strong’s Exhaustive Concordance of the Bible*, Abingdon-Cokesbury Press: New York, NY, 1890.

¹⁵ Robinson, Maurice A, PhD. “The Online Greek New Testament Declension Codes For Nouns, Adjectives, Prepositions, Conjunctions And Particles” and “The Online Greek New Testament Parsing Codes For Verb-Related Forms”, 27 July 2004; <http://kotisivu.dnainternet.net/jusala/RP2005/PARSINGS.TXT>.

¹⁶ J. K. Tauber ed., *MorphGNT: SBLGNT Edition*, version 6.12 [Data set], 2017; <https://github.com/morphgnt/sblgnt>, accessed April 26, 2017. Tauber parsing scheme was originally derived from Center for Computer Analysis of Texts, University of Pennsylvania. “United Bible Societies 1992 3rd Ed.” ftp://unboundftp.biola.edu/pub/CCAT_Parsed_NA26.zip, accessed October 25, 2003.

¹⁷ Department of Linguistics, *Language files: Materials for an introduction to language and linguistics*, 11th ed., Ohio State University Press, 2011.

¹⁸ Leonard Bloomfield, *Language*, p. 203-206, George Allen & Unwin LTD: London, 1933.

¹⁹ Noah D. Frederick, “The Syntax of Determiners in Attic Greek: A Theoretical Approach”, The Department of Classical Studies, University of Michigan, Winter 2009; https://deepblue.lib.umich.edu/bitstream/handle/2027.42/63950/frederick_noah_2009.pdf, accessed April 16, 2017.

²⁰ Richard Faure, “Determiners”, *Encyclopedia of Ancient Greek Language and Linguistics*, vol. 1, p. 442-446. E.J. Brill: Leiden, Netherlands, 2014.

²¹ Janne Bondi Johannessen, “The syntax of correlative adverbs”, *Lingua*, vol. 115, p. 419-443, 2005.

²² Exceptions were made for relatives pronouns affected by attraction.

²³ A. E. Knoch, “The Concordant Method”, *Concordant Greek Text*, Concordant Publishing Concern: 1975; http://www.scripture4all.org/help/isa2/Articles/The_Concordant_Method/The_Concordant_Method.htm, accessed April 14, 2017.

²⁴ See entries such as “ΔΥΝΑΜΙΣ”, “ΟΙΚΟΣ”, “CΥΝΑΓΩΓΗ”, “ΤΕΚΝΟΝ”.

5. Structural Components

There are a number of structural components in the CNTR database built upon the other basic data elements which provide advanced capabilities for textual criticism. These components enable the efficient retrieval and processing of verses, collated data, and variant units for different types of research and analysis.

5.1 Versification

The CNTR processes the collation and variant units on a verse basis. The CNTR transcriptions follow the definitive standard in versification for the New Testament which was first introduced by Robert Estienne (also known as Roberti Stephani and Robertus Stephanus) in his 1551 publication of the Greek New Testament.¹ His son Henry wrote of this endeavor: “As the books of the New Testament has been already divided into the sections (*themata*) which we call chapters, he himself subdivided them into those smaller sections, called by an appellation more approved of by others than by himself, *versicles*....He accomplished this division of each chapter on his journey from Paris to Lyons, and the greater part of it *inter equitandum*.”² Some jest that the bumpiness of riding on horseback surely caused his pen to jerk resulting in some very awkward verse boundaries, but a more natural inference of “*inter equitandum*” is that the work was done while resting at inns along the road.³ The chapter divisions previously mentioned had been established in the 13th century by Stephen Langton, Archbishop of Canterbury.⁴

Bibles today attempt to follow Estienne’s versification scheme for the most part, yet all of them deviate from it in various places. The first complete Bible in English to contain verses was the Geneva Bible, for example, and it too strays from Estienne’s standard in several places. There is not any consistency to these deviations as every Bible translation seems to make up a few new verse boundaries of its own. Different electronic editions of the same text often contain different versification schemes depending on which software it was derived from. While differences in versification do not diminish the text, it has caused much confusion over the years and makes comparisons of verses among different texts somewhat problematic for Bible programs.

The methodology used by the CNTR for recovering the verse boundaries originally specified by Estienne involved constructing a master list of discrepancies between the verse boundaries used in various Bibles. These discrepancies were then resolved by consulting an original copy of Estienne’s 1551 text on reserve in the Special Collections at the University of Chicago. Out of all of the Bibles examined, the Nestle-Aland text was the closest in adhering to Estienne’s versification standard: “The verse divisions...with a few exceptions, are identical with the verse divisions first introduced by Stephanus in his 1551 edition of the Greek New Testament, and are widely adopted in modern editions.”⁵ Yet, even the Nestle-Aland text still deviated from Estienne’s standard in the following verses: Matt. 12:49, Luke 24:45, Acts 2:47, 1 Cor. 7:33, Eph. 5:13, and Rev. 6:1. There was also an instance where the text of the verses was switched at Phil 1:16-17.

The implementation of Estienne’s versification scheme, however, was not always straightforward. When a variant reading *not present* in Estienne’s text fell on a verse boundary, for example, it could be placed either at the end of the previous verse or at the beginning of the next verse. In such cases, the versification of the Nestle-Aland text was followed since it carried the honor of being the closest to adhere to Estienne’s standard. Estienne’s 1551 text also contained several printing errata which have been corrected as follows:

- Matthew 12:50 was added where an indentation in the Latin text indicated a new verse, but the number was missing as the Greek did not have a corresponding indention.

- Mark 11:33 was realigned to the text indentation as it had been placed too low.
- Luke 16:31 was added where an indentation indicated a new verse but the number was missing.
- John 4:53 had been repeated twice and thus the second reference was renumbered as 4:54.
- John 13:38 had been skipped and thus 13:39 was renumbered as 13:38.
- Acts 23:25-26 had shared one line and were split into two separate verses.
- Acts 23:35 was added where an indentation indicated a new verse, but the number was missing.
- Acts 24:19-20 had shared one line and so 24:21-28 were renumbered 24:20-27.
- 1 Thes. 2:12 was realigned to the text indentation as it had been placed one line too low.
- 2 John 1:12 had been juxtaposed one line below 1:11 and so it was realigned to the indentation and 1:13 and 1:14 were renumbered as 1:12 and 1:13 leaving a total of 13 verses.

Differences in verse divisions found in the critical texts have been altered to conform to Estienne's standard for use in the CNTR project.

There, of course, were no verse numbers in the original Greek manuscripts. Verse numbers are merely supplied here for reference and should never be construed to have any relationship with the beginning of sentences or phrases. "The Stephanic verses have met with bitter criticism because of the fact that they break the text into fragments, the division often coming in the middle of the sentence, instead of forming it into convenient and logical paragraphs, an arrangement which has seldom found favor. But their utility for reference outweighs their disadvantage."⁶ Because of this, the careful reader should always consider the verses before and after any reference to obtain the proper context.

A verse is only included in a CNTR text where there is at least one character present in the extant manuscript (even if it is partially damaged). An explicit verse omission occurs when the previous and following verses run together consecutively and the words for the verse in-between are missing. The following verses do not yet appear in any Greek extant manuscript found before the *terminus ad quem*: Mark 15:28, Acts 8:37, 24:7, 28:29, Rom. 16:24.

5.2 Collation Alignment

The CNTR collation of early manuscripts was generated to alleviate the myriad of problems that stem from using apparatuses commonly found in the critical texts. The CNTR collation currently contains a complete set of variants for the class 1 and class 2 data (along with several critical texts for comparison) showing orthographical differences, scribal corrections, and the condition of characters, while not being tied to any particular base text. Each column in the CNTR collation is assigned a CollationID and represents a distinct lexical/morphological/phonological word form. All of the words in a column represent homophones of each other according to a standard set of phonetic rules,^{§4.1.2} ignoring any orthographical differences such as elision, movable nu or sigma, *nomina sacra*, numeric abbreviations. Misspelled words that deviate from these phonetic rules, but could be no other plausible word, are placed in the intended column and the nature of the errant characters is encoded as previously discussed.^{§4.1.3} Words with the same lexeme and morphology but different pronunciation (i.e. ΕΔΙΔΟΥΝ vs. ΕΔΙΔΟCΑΝ) are considered to be separate lexical/morphological/phonological forms and therefore placed in a separate column.

The collation was generated by using three different computer algorithms. First, the maximum text was created as a templet containing all known variants for each verse by using a recursive longest common sequence first algorithm without reference to any base text. Second, each individual witness was then aligned to this templet using a non-recursive longest common sequence algorithm considering multiple sequences. By utilizing the CNTR parsing information, a separate column was thus generated for each lexical/morphological/phonological word form and the words in all texts were linked to the appropriate column in the CNTR database. There are often several different ways that variants can be aligned in a given verse, so the database was manually adjusted to minimize the number of columns, which is also not

necessarily unique. For example, the variant “ΙΗCOY XPICTOY” versus “XPICTOY ΙΗCOY” in 1st Corinthians 1:1 could have had a column alignment of either “ΙΗCOY XPICTOY ΙΗCOY”

ΠΑΥΛΟΣ	ΚΛΗΤΟΣ	ΑΠΟΣΤΟΛΟΣ	ΙΗCOY	XPICTOY	...
ΠΑΥΛΟΣ	ΚΛΗΤΟΣ	ΑΠΟΣΤΟΛΟΣ		XPICTOY	ΙΗCOY ...

or “XPICTOY ΙΗCOY XPICTOY”

ΠΑΥΛΟΣ	ΚΛΗΤΟΣ	ΑΠΟΣΤΟΛΟΣ	XPICTOY	ΙΗCOY	...
ΠΑΥΛΟΣ	ΚΛΗΤΟΣ	ΑΠΟΣΤΟΛΟΣ		ΙΗCOY	XPICTOY ...

and one is not necessarily any better than the other. Third, the CNTR website collation is displayed in collapsed mode which was achieved by using a gap-reducing algorithm that eliminates space and attempts to align words with similar form or meaning even though they may have different morphologies.

Any statistical processing of columns, however, is done with the data in uncollapsed mode. One such application is the autoparse program which automatically parses any Greek New Testament manuscript by aligning it with the existing collation and obtaining the associated lexical, syntactical, and morphological information from the appropriate column according to its context. If the text to be parsed contains a new variant reading, then those words are added to the appropriate place in the collation and flagged in the output so that the associated parsing data can be added manually. Thus, the CollationIDs are not static since new CollationIDs are created and reordered to incorporate the words of any new variant reading.

5.3 Slot Assignments

The concept of textual slots was invented by the CNTR to facilitate the sharing of different types of data between any two Greek New Testament texts regardless of their textual variants. Thus, the data annotations, morphological analysis, syntax trees, etc. created by a researcher using one particular base text can quickly be transferred to another researcher using a different base text. Unlike the CollationIDs which change with the addition of new variants, each word in the CNTR database is also assigned a SlotID which will never change. That way researchers who are using the textual slots will never have to change their data or mappings. Another benefit of SlotIDs is the ability to compare any two Greek texts for word differences or word order differences while ignoring any orthographical differences.

A SlotID is defined as a unique lexical/morphological word form where *word order does not matter* within variant transposition units. A transposition unit (TU) is defined as a region of textual variation which extends to cover all mutually exclusive uses of a word form. A complex TU could involve several words that switch positions. As new variant words are found in other texts, a TU could potentially be expanded in size, but the SlotIDs previously assigned would not change. (As a side note, a variant unit used in the context of textual criticism will often be larger than a TU, as there are other contextual considerations and several TUs may be linked by dependencies. For example, if a text has a particular reading in TU A, it may necessarily have an associated reading in TU B.)

To illustrate this concept, consider again the example of 1st Corinthians 1:1 where there is a TU of “ΙΗCOY XPICTOY” vs. “XPICTOY ΙΗCOY”:

GA P46	175-224	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΧΡΥ	ΙΗΥ	...
GA 01*	325-360	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΑΥ	ΧΥ	...
GA 01	325-360	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΥ	ΧΥ	...
GA 03	325-349	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΧΥ	ΙΥ	...
GA 02	375-499	ΠΑΥΛΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΥ	ΧΥ	...
WH	1885	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΗCOY	ΧΡΙCΤOY	...
NA28	2012	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΧΡΙCΤOY	ΙΗCOY	...
SBL	2010	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΧΡΙCΤOY	ΙΗCOY	...
RP	2005	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΗCOY	ΧΡΙCΤOY	...
KJTR	2016	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΗCOY	ΧΡΙCΤOY	...
ST	1550	ΠΑΥΛΟΣ ΚΛΗΤΟΣ ΑΠΟΣΤΟΛΟΣ	ΙΗCOY	ΧΡΙCΤOY	...

Each unique lexical/morphological word form would be assigned a unique SlotID regardless of order as follows:

SlotID	Words
de2	ΠΑΥΛΟΣ
4Hf	ΚΛΗΤΟΣ
p34	ΑΠΟΣΤΟΛΟΣ
Ynr	ΧΡΙCΤOY, ΧΡΥ, ΧΥ
kcG	ΙΗCOY, ΙΗΥ, ΙΥ, ΑΥ

Notice that SlotID Ynr represents the words ΧΡΙCΤOY, ΧΥ, and ΧΡΥ ignoring any orthographical differences. The words in each text are then tagged with the correct SlotID as follows:

Texts	SlotIDs
P46, 03, NA28, SBL	de2 4Hf p34 Ynr kcG
01*, 01, WH, BYZ, KJTR, ST	de2 4Hf p34 kcG Ynr
02	de2 p34 kcG Ynr

Notice that the first set of manuscripts has the order Ynr kcG while the second set of manuscripts has the order kcG Ynr. The 02 text does not have SlotID 4Hf because it does not contain that word. Once the SlotIDs have been assigned, one researcher can then easily transfer various data attributes to another researcher regardless of orthographical differences or differences in word order. Obviously if a word is missing in one text, the attributes cannot be transferred.

Unlike the CollationIDs, the same SlotIDs may be assigned to different word forms provided that they still represent the same lexical/morphological meaning in a TU. Thus, words with different lemmas in the same lexeme such as “ΜΑΡΙΑ” and “ΜΑΡΙΑΜ” would be assigned the same SlotID and different word forms with the same morphological parsing such as “ΕΔΙΔΟΥΝ” and “ΕΔΙΔΟCΑΝ” would be assigned the same SlotID. Words that have different morphological parsings are always assigned to different SlotIDs. But a single word form that could be *interpreted* by more than one morphological paradigm will only receive one SlotID. Thus, while the “-ετε” ending could be *interpreted* to be either indicative or imperative, there is only one word form and thus there would only be one SlotID assigned.

The SlotIDs were initially populated by a computer algorithm that identified TUs by finding words with the same lexical/morphological parsing that were mutually exclusive between two or more columns, and then these were later hand checked for false positives. In the future, additional layers can be added to the SlotIDs to distinguish different morphological orthographical forms.

5.4 Variant Units

The CNTR database contains various types of information regarding variant units (also called variation units). A variant unit is defined here as a contiguous region of textual change which can be further subdivided by word order differences. The boundaries of the variant units were automatically marked by a separate computer program, and then manually adjusted for context in a small number of cases. The database also records whether or not a word is completely represented by all witnesses in a column (i.e. whether it is involved in a variant reading or not). It also contains alignment data that determines which words can be considered to be substitutes for each other in the collation. Significant statistical research is beginning to be done with this variant unit data, and the processes are continuing to be refined as a work in progress.

¹ Robert Estienne, Desiderius Erasmus, and Andreas Osiander. *Apanta Ta Tes Kaines Diathekes. Nouum Iesu Christi D.N. Testamentum: cum duplici interpretatione*, 2 vols, Ex officina Roberti Stephani, Geneva, Switzerland, 1551.

² Henry Estienne, Robert Estienne, and Theodore de Beze. *Concordantiæ Testamenti Novi, Græcolatinæ. Nunc primum plenæ editæ: & diu multumque desideratæ, vt optimæ duces ad veram vocum illius interpretationem futuræ. In his quid præstitum sit, præfixa ad lectorem epistola docet*, Ex typographieio Henri Stephani: Geneva, Switzerland, 1594.

³ John McClintock and James Strong, “Verse”, *Cyclopaedia of Biblical, Theological, and Ecclesiastical Literature*, vol. 10, p. 762, Harper & Brothers: New York, NY, 1894.

⁴ George William Gilmore, *New Schaff-Herzog Encyclopedia of Religious Knowledge*, vol. 2, p. 113, Baker Book House: Grand Rapids, MI, 1954.

⁵ Kurt Aland, et al. eds., *The Greek New Testament*. 3rd corrected ed., p. xi, United Bible Societies: Stuttgart, Germany, 1983.

⁶ George William Gilmore, *New Schaff-Herzog Encyclopedia of Religious Knowledge*, vol. 2, p. 114, Baker Book House: Grand Rapids, MI, 1954.

Appendix – Database Design

