

# Greek Linguistic Databases

## Overview, Recent Work, and Future Prospects

Rick Brannan

Biblical Scholarship: Data Types, Text, Language and Interpretation

Lorentz Center, Leiden, Netherlands

February 8, 2012

How have databases of Greek material transformed? These materials used to focus on the word as unit of analysis, providing access to morphological data. This began to change in 2006 with Logos Bible Software's release of the *OpenText.org Syntactically Analyzed Greek New Testament*. Logos has made other syntactic analyses available, including the *Lexham Syntactic Greek New Testament* and the *Cascadia Syntax Graphs of the New Testament* (based on work done by the Asia Bible Society). Logos has also released the *Lexham Discourse Greek New Testament*, an application of discourse grammar to the entire Greek New Testament as well as an associated grammar, *Discourse Grammar of the Greek New Testament*, which provides an extensive explanation of the framework and terminology used in the discourse grammar analysis of the Greek New Testament.

A wealth of material and perspective is available through these sorts of data sets, but they are still relatively new. It is entirely possible that students are more capable with these syntactic and discourse-level databases than their professors. This paper will provide a basic overview of each of these resources as well as various methods to access, view and use the analysis they offer.

## Introduction and Historical Background

In the world of Greek texts, the electronic availability of texts is not new. Projects focused on encoding Greek text go back to the 1970s<sup>1</sup> with projects like the *Thesaurus Lingua Graeca* (TLG)<sup>2</sup> and the Perseus Digital Library.<sup>3</sup> In 1976, the GRAMCORD Institute undertook the creation of what they called a "grammatical concordance"<sup>4</sup> of the Greek New Testament, annotating grammatical qualities of each word in the Greek New Testament and providing a tool to use in the analysis and searching of this type of data. An analysis of the Septuagint was completed by the Center for Computer Analysis of Texts (CCAT) at the University of Pennsylvania.<sup>5</sup>

In the following years, Bible Software companies (Logos,<sup>6</sup> BibleWorks,<sup>7</sup> Accordance<sup>8</sup> and others) licensed data

---

<sup>1</sup>The TLG was founded in 1972. <http://www.tlg.uci.edu/about/history.php>

<sup>2</sup>The TLG website: <http://www.tlg.uci.edu/>

<sup>3</sup>The Perseus website: <http://www.perseus.tufts.edu/hopper/>

<sup>4</sup>The GRAMCORD Institute website: <http://gramcord.org/>

<sup>5</sup>Version 1.0 was released on November 7, 1986 according to: <http://ccat.sas.upenn.edu/gopher/text/religion/biblical/0-readme.txt>

<sup>6</sup>Logos Bible Software's web site: <http://www.logos.com>

<sup>7</sup>BibleWorks' web site: <http://www.bibleworks.com/>

<sup>8</sup>Accordance's web site: <http://www.accordancebible.com/>

from GRAMCORD, from Timothy and Barbara Friberg,<sup>9</sup> James Tauber,<sup>10</sup> and from Maurice A. Robinson and William G. Pierpont.<sup>11</sup> Soon software providers began to undertake their own analyses of the grammar of the Greek New Testament.

Analysis, while still at the word level, began to expand in scope. The GRAMCORD analysis began to associate a noun case notation with prepositions. This was a method to associate the prepositional object with its preposition, even if they were not adjacent. Additionally, analyses began to move from providing a strict analysis (e.g. “conjunction”) to providing further information on types and classes of things (e.g. “conjunction, coordinating, copulative”). Analysis began to take on a subjective, instance-based nature. Classification thus began to be morpho-syntactic.

Other innovations were made as well, though all still essentially at the level of the word. The Friberg analysis devised a system of '+' and '-' notation to connect some terms that function together as a pair:

A plus sign (+) immediately before or after a tag indicates a close relationship between the word associated with the tag and another word, as in cases of verbal periphrastics. The sign appears on the side of the tag on which the pairing occurs. A minus sign (-) precedes a relative pronoun tag when there is no overt antecedent in the text.<sup>12</sup>

The parties creating these annotations were bumping against the boundary of the word, trying to get past it but still constrained by its limits.

## Recent Work in Linguistic Annotation of the Greek New Testament

In the early years of the 21<sup>st</sup> century, Stanley Porter, Jeffrey T. Reed, Matthew Brook O'Donnell, Randall Tan and Catherine Smith commenced the OpenText.org project.<sup>13</sup> It was formally introduced on the OpenText.org web site on September 22, 2005.<sup>14</sup> Details of the analysis had been published earlier in various sources including conference papers and published dissertations.<sup>15</sup>

Logos Bible Software released an implementation of the *OpenText.org Syntactically Analyzed Greek New Testament* in 2006. The previous year we had commenced work on the *Lexham Syntactic Greek New Testament*,<sup>16</sup> an analysis taking a traditional approach much like Reed-Kellogg stick diagramming, the initial portion of this analysis (Hebrews through Jude) was also released in 2006.<sup>17</sup> This analysis of the New Testament was completed in 2010.

In the mid-2000s, the Perseus project began to create treebanks of Latin and Greek material, now known as *The*

---

<sup>9</sup>Timothy and Barbara Friberg's analysis of the Greek New Testament was published in print as the *Analytical Greek New Testament*, published by Baker Book House in 1980.

<sup>10</sup>James Tauber's edit and revision of CCAT's Greek New Testament work: <http://morphgnt.org>

<sup>11</sup>Robinson's byztxt.com site: <http://byztxt.com/download/index.html>

<sup>12</sup> Barbara Friberg et al., vol. 1, *Analytical Greek New Testament : Greek Text Analysis* (, Baker's Greek New Testament library Cedar Hill, Texas: Silver Mountain Software, 2001). This system is more fully explained in appendix 3.8 of the book.

<sup>13</sup>The project web site (<http://www.opentext.org>) has a copyright date of 1998-2005; however the earliest article available on the site is dated January 11, 2001. <http://opentext.org/resources/articles.html>

<sup>14</sup>Matthew Brook O'Donnell, “Introducing the OpenText.org Syntactically Analyzed Greek New Testament”. <http://opentext.org/resources/articles/a8.html> Accessed: January 18, 2012.

<sup>15</sup>Dissertations include those of Jeffrey T. Reed, *A Discourse Analysis of Philipians* (1997) and Matthew Brook O'Donnell, *Corpus Linguistics and the Greek of the New Testament* (2005).

<sup>16</sup>More info here: [http://blog.logos.com/2005/12/greek\\_syntax\\_in/](http://blog.logos.com/2005/12/greek_syntax_in/) Product page at Logos.com: <http://www.logos.com/logos3/new/lexhamSGNT>

<sup>17</sup>Logos also released a syntactic analysis of the Hebrew Bible in 2006: *The Hebrew Bible: Andersen-Forbes Phrase Marker Analysis*.

*Ancient Greek and Latin Dependency Treebanks*.<sup>18</sup> While both the *Iliad* and the *Odyssey* have been analyzed, the Greek New Testament has not been analyzed.

In March of 2008, Logos released the *Lexham Discourse Greek New Testament* and its counterpart, the *High Definition New Testament: ESV Edition*.<sup>19</sup> This applies discourse grammar according to the framework of Stephen H. Levinsohn and Steven Runge to the entire text of the Greek New Testament. It identifies devices that work above the word level as well as implementing a propositional outline of the entire Greek New Testament. In 2010 Logos released a grammar explaining the framework called *Discourse Grammar of the Greek New Testament*.<sup>20</sup> It was also published in print by Hendrickson Publishers. A video series explaining the concepts has been released as well.<sup>21</sup>

In November 2009, Logos released the *Cascadia Syntax Graphs of the New Testament*, derived from a dynamic treebank project developed by the Asia Bible Society.<sup>22</sup> This analysis is unique in that it is built upon a computer-readable Greek grammar that is used by a parser to generate syntactic trees. These trees are reviewed, corrected and refined through an editorial process; the corrections and edits become a knowledgebase that informs the parser on future editions. In November 2010, the *Cascadia Syntax Graphs of the New Testament: SBL Edition* was released, offering analysis of a different edition of the Greek New Testament. In January 2012, after receiving updated analysis data from the editors, Logos released significant updates to both of these syntactic analyses.

In March of 2010, the PROIEL<sup>23</sup> project announced that it had annotated the gospels of the Greek New Testament.<sup>24</sup> As of the writing of this paper (January 2012), the PROIEL annotation looks to also include data for Acts and the Pauline epistles, with the balance of the Greek New Testament under review.<sup>25</sup>

In November of 2010, Oak Tree Software released the first portion of a grammatical and syntactical analysis of the Hebrew Bible and the Greek New Testament for the Accordance program.<sup>26</sup>

In January of 2012, Jana E. Beck released a demonstration edition of a syntactic annotation of the Gospel of Matthew. It is part of a larger project, the Penn Parsed Corpora of Historical Greek.<sup>27</sup>

These databases<sup>28</sup> were and are revolutionary because they put the focus of analysis beyond simple annotation of

---

<sup>18</sup> "The Ancient Greek and Latin Dependency Treebanks," <http://nlp.perseus.tufts.edu/syntax/treebank/> Accessed January 24, 2012.

<sup>19</sup> Announcement: [http://blog.logos.com/2008/03/study\\_the\\_nt\\_like\\_never\\_before/](http://blog.logos.com/2008/03/study_the_nt_like_never_before/) Product page at Logos.com: <http://www.logos.com/product/3888/lexham-discourse-greek-new-testament-bundle>

<sup>20</sup> Product page at Logos.com: <http://www.logos.com/product/4599/discourse-grammar-of-the-greek-new-testament-a-practical-introduction-for-teaching-and-exegesis>

<sup>21</sup> *Introducing New Testament Discourse Grammar*. <http://www.logos.com/product/7076/introducing-new-testament-discourse-grammar-video-series>

<sup>22</sup> Wu, Andi, and Randall Tan. *Cascadia Syntax Graphs of the New Testament*. Logos Bible Software, 2009.

<sup>23</sup> "PROIEL: Pragmatic Resources in Old Indo-European Languages," <http://www.hf.uio.no/ifik/english/research/projects/proiel/> Accessed January 24, 2012.

<sup>24</sup> "[Corpora-List] New Testament corpus," <http://mailman.uib.no/public/corpora/2010-March/010402.html> Accessed January 24, 2012.

<sup>25</sup> This analysis uses Tischendorf's Eighth edition of the Greek New Testament, prepared by Dr. Maurice A. Robinson with morphology integrated and edited by Dr. Ulrik Sandborg-Petersen. Online: <http://files.morphgnt.org/tischendorf/>

<sup>26</sup> Product page at AccordanceBible.com: <http://www.accordancebible.com/store/details/?pid=GNT-T.syntax>. As of February 2012, data for the New Testament is complete.

<sup>27</sup> More information available at: <http://www.ling.upenn.edu/~janabeck/greek-corpora.html>

<sup>28</sup> The week this paper was presented I was made aware of another syntactic analysis of the Greek New Testament, Dr. Ulrik Sandborg-Petersen. I will update this portion once I have the available information.

words and enabled database users to consider higher level structures (phrases, clauses, and sentences) built up from the word level data.

However, these sorts of databases are also inherently complex: Complex to understand the frameworks that produced them. Complex to display the information to users of the databases. Complex to enable searching of the myriad relationships and structures within the databases. Complex to translate exegetical and grammatical concepts routinely found in commentaries and exegesis of the New Testament into terminology and concepts used in the databases.

## Syntactic Database Requirements: Seeing and Searching

While there are problems with simply annotating syntactic data, once one has a syntactic analysis one has the further problems of searching through the data, and of seeing (or visualizing) the data. In order for people to use a syntactic database, they need to have some idea of the features it annotates, the text the annotations apply to, and the relationships between each feature.<sup>29</sup>

In our work on syntactic analyses of the Hebrew Bible and the Greek New Testament, we have found that in order to have any hope of actually using a syntactic analysis, users need to be able to see something that visually represents the relationships and structures encoded in an analysis. When they have a picture, they can begin to understand and work with the analysis. With a text that only annotates morphological information, this is relatively simple using either popups on hover, an interlinear display, or both. With a text that could have many layers of analysis, however, simple text has limitations.

## Seeing the Data

In 2005, when Logos was beginning our work on syntactic databases, we had the good fortune of becoming familiar with the work of Francis I. Andersen and Dean A. Forbes, particularly with their work in using directed acyclical graphs<sup>30</sup> to visualize the relationships in the *Andersen-Forbes Phrase Marker Analysis* (AFPMA). As the visualizations of their analysis relied on these sorts of graphs, it was necessary to implement this solution to properly handle the visualization of their data. Once the capability was available we decided to apply the same visualization to our available Greek data (at that time the OpenText.org analysis) to see if the benefits of the directed graph approach would carry over.

In a blog post from November 2005, my colleague Eli Evans described a directed graph, and gave some examples:

A graph is a convenient way to show connections between things. Flow charts, for example, are a specialized kind of graph. So are corporate organizational charts. Sometimes people draw graphs when they are brainstorming, by putting words in little bubbles and then drawing lines between them to connect them.

...

When you put arrows on the lines, you're making a **directed graph**, which is especially useful for making things like flowcharts and corporate org charts. ... This sort of arrangement is also useful for labeling the parts of a clause:

---

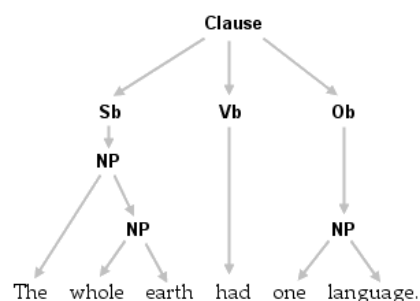
<sup>29</sup> I work for Logos and have extensive knowledge of the Greek syntactic analyses produced for Logos Bible Software. My examples will only refer to analyses available from Logos Bible Software. There are other options (e.g. Oak Tree Software's analysis in Accordance), but as I am unfamiliar with these options I will not be able to explain or discuss them in detail.

<sup>30</sup> An explanation of the use of graphs for syntax data in Logos Bible Software is available in a series of blog posts:

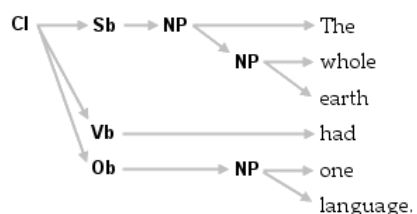
"What's a Syntax Graph Anyway?": [http://blog.logos.com/2005/11/whats\\_a\\_syntax/](http://blog.logos.com/2005/11/whats_a_syntax/)

"Syntax: Why Graphs?": [http://blog.logos.com/2005/11/syntax\\_why\\_grap/](http://blog.logos.com/2005/11/syntax_why_grap/)

"Syntax: Why Graphs? Part II": [http://blog.logos.com/2005/11/syntax\\_why\\_grap\\_1/](http://blog.logos.com/2005/11/syntax_why_grap_1/)



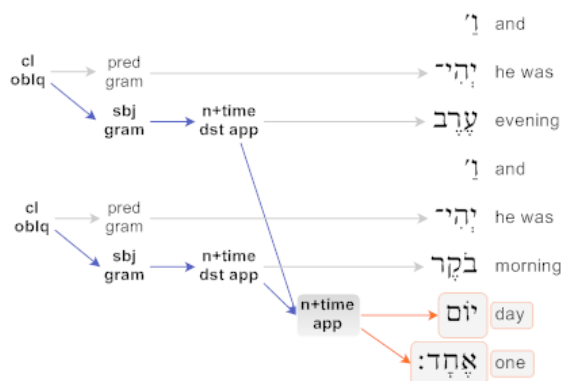
This is a fine way to diagram a sentence; but while it works well for short sentences, this presentation does not lend itself well to displaying entire texts, such as the Bible. If, however, we turn the graph on its side so that the text reads from top to bottom along the right-hand margin, we get a graph like this:



Now we can have an infinitely long running text that reads from top to bottom, and that's just what we have when we look at a Logos Bible Software syntax graph.<sup>31</sup>

The above is both a directed graph and a tree. While trees are nice in the laboratory, when language is out in the wild, it is not always so easy. There are places where an analysis may “tangle” its lines. Trees (like the above) tend to avoid tangling in their presentation. When tangling occurs, there are two choices: rearrange the words, or re-analyze the text.

In Andersen and Forbes' application of directed acyclical graphs, the text is allowed to run in its text order (from top to bottom) and the graphs are allowed to tangle as necessary to represent the structures of the language. One complex example is found in Genesis 1.5, which shows a multi-dominant structure:<sup>32</sup>



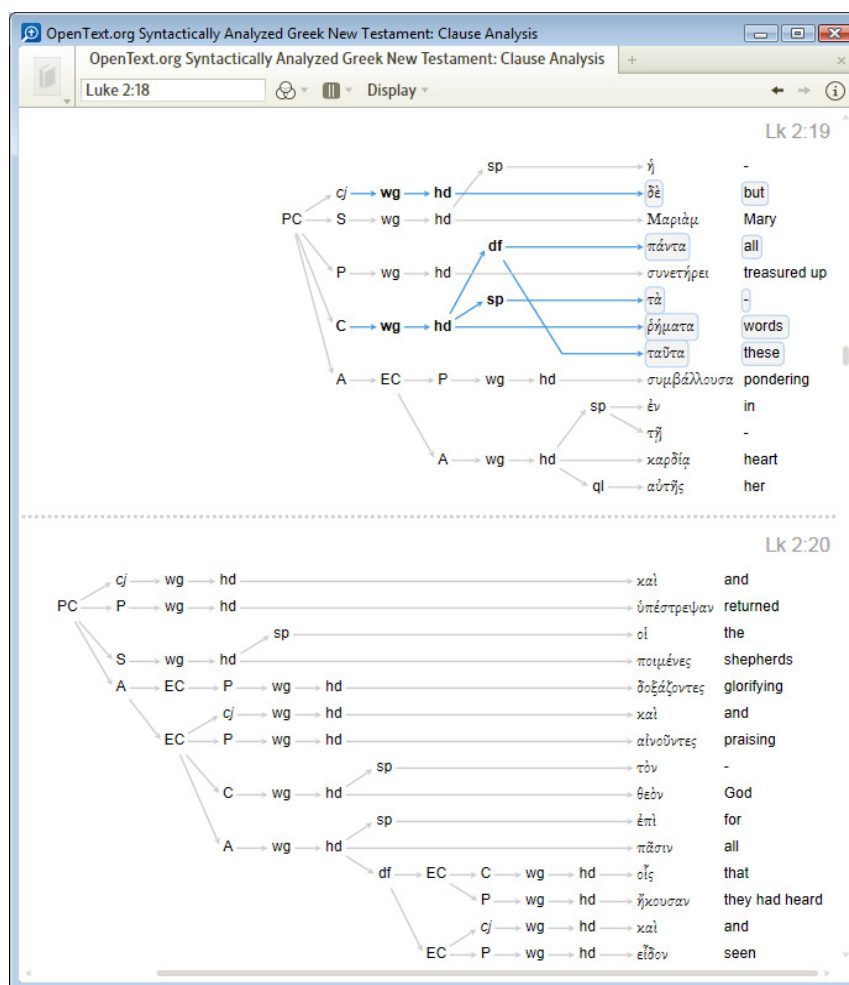
So supporting the concept of directed acyclical graphs allows the Logos Bible Software representation to reflect, in a visual manner, the complexity of the language as analyzed.

While the *Andersen-Forbes Phrase Marker Analysis* takes full advantage of these capabilities, other syntax analyses do not. The analyses of the Greek New Testament do at times “tangle,” but they do not express multidominance as

<sup>31</sup> Eli Evans, “What’s a Syntax Graph Anyway?” [http://blog.logos.com/2005/11/whats\\_a\\_syntax/](http://blog.logos.com/2005/11/whats_a_syntax/) Accessed January 23, 2012.

<sup>32</sup> Eli Evans, “Syntax: Why Graphs? Part II” [http://blog.logos.com/2005/11/syntax\\_why\\_grap\\_1/](http://blog.logos.com/2005/11/syntax_why_grap_1/) Accessed January 23, 2012.

AFPMA does. Below is an example from the *OpenText.org Syntactically Analyzed Greek New Testament* of Luke 2.19, with a postpositive δε and some further tangling due to the analysis of the clause's Predicator and Complement.



Luke 2.19 in OpenText.org.

The primary purpose of these graphs in Logos Bible Software is to express the complex relationships within each clause visually in a way that communicates the relationships to the user. A user can see the structures within the graph, and see which graph components and which words are contained in the structure.

## Searching the Data

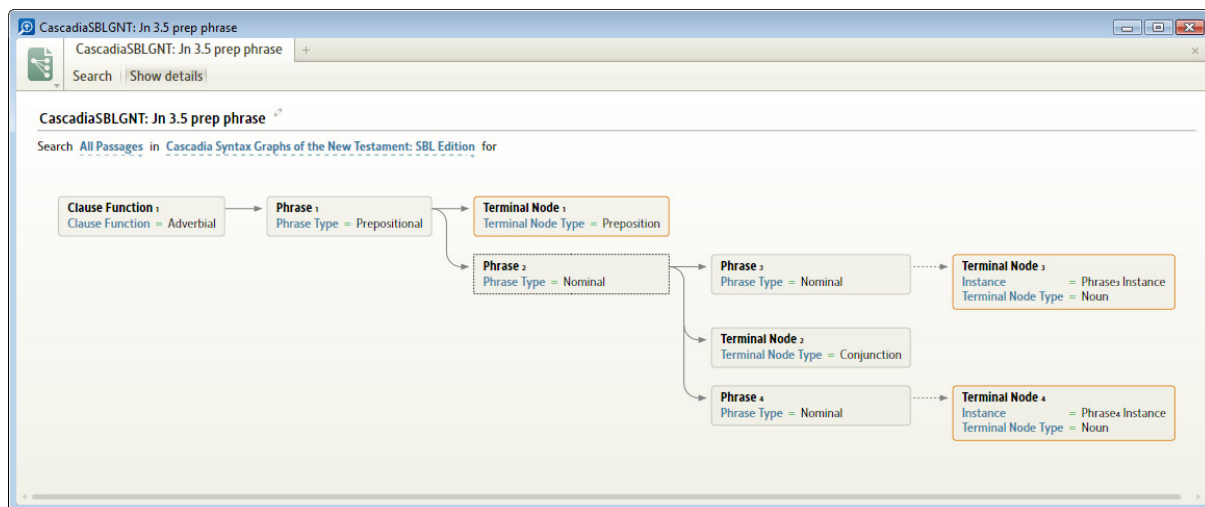
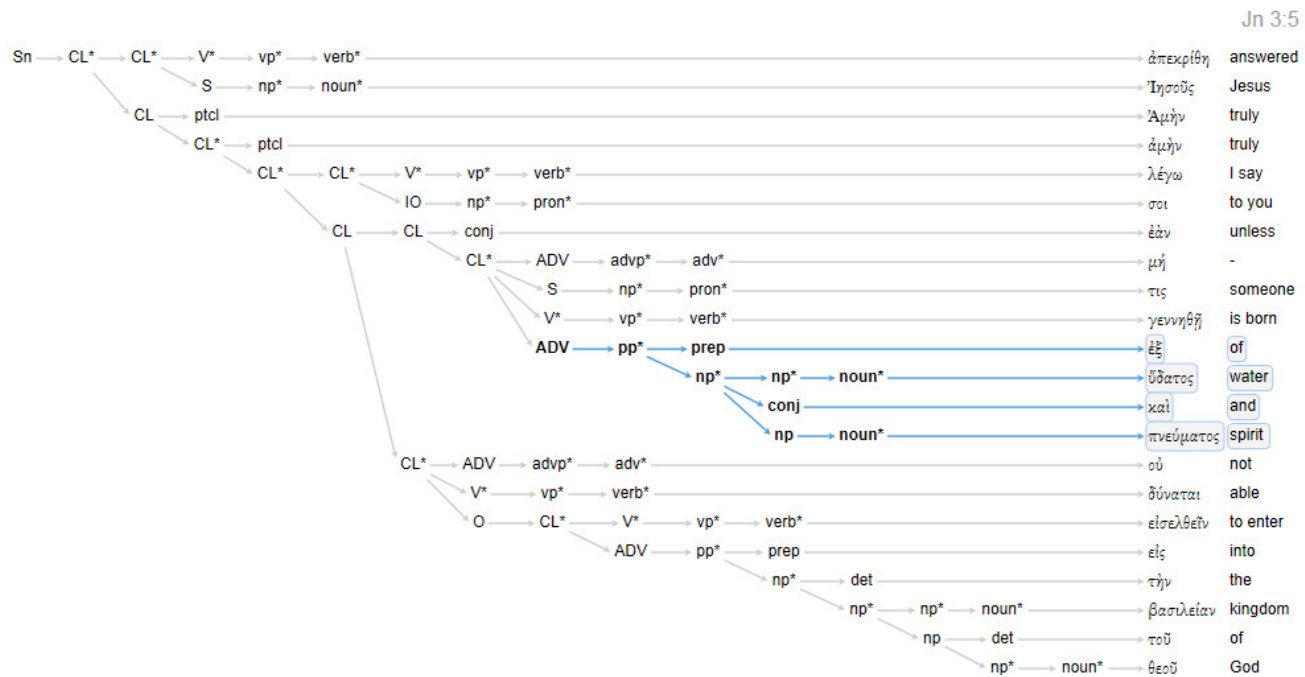
Seeing the data is one portion of the equation; searching the data is another. Modeling complex relationships and then searching them is a difficult problem. Fortunately for us at Logos, the ground was prepared by Ulrik Sandborg-Petersen's Emdros,<sup>33</sup> which accurately bills itself as a "text database engine for storage and retrieval of *analyzed or annotated text*."<sup>34</sup> With Logos Bible Software, we use Emdros to generate the database that is actually delivered to users and searched. We also include the relevant portions of Emdros within our PC and Mac versions of Logos Bible Software to search the databases.

We have created a query editor that allows one to visually create a graph like those discussed above. We create an

<sup>33</sup> "emdroS: A Corpus Query System" <http://emdroS.org/> Accessed January 23, 2012.

<sup>34</sup> "emdroS: A Corpus Query System: What is Emdros?" <http://emdroS.org/whatis.html> Accessed January 24, 2012.

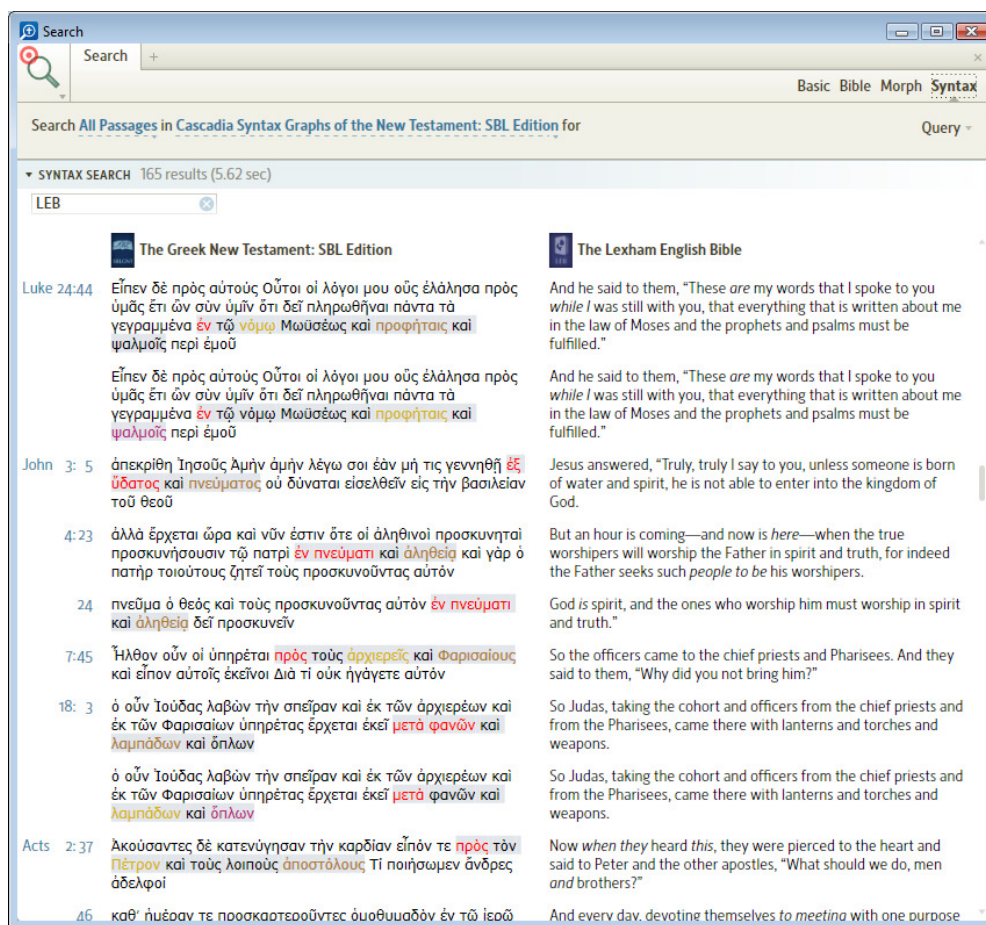
an example, awhile back I was asked about the prepositional phrase in Jn 3.5. Below is the graph<sup>35</sup> for Jn 3.5 (with the prepositional phrase portion in question highlighted). Following that is a query (of the *Cascadia Syntax Graphs of the New Testament: SBL Edition*) that will locate the basic structure.



The portion of a clause (or clause component) that functions adverbially, which is a prepositional phrase. The prepositional object is compound, two noun phrases joined by a conjunction.

<sup>35</sup> Screen captures are from the Windows version of Logos Bible Software 4.5.





So we use Emdros behind the scenes to accomplish this. After determining a data model appropriate for the way Emdros functions, we convert source data into the necessary MQL. We use tools provided with Emdros to compile the database from the MQL. We then create a resource for Logos Bible Software that includes the MQL database and other metadata that Logos needs; this is what is delivered to users.

## Using Syntactic Analyses in Logos Bible Software

There are several levels of usage of this data in Logos Bible Software. Usage can range from casual browsing of the syntax graphs to creation of complex queries to locate particular structures. In reality, though, most usage is toward the lower end of the spectrum, even below simply browsing a syntax graph.

## Grammatical Relationships and Preposition Use

While many users have a desire to learn about how particular passages are structured and the differing syntactic contexts which particular words and ideas are used, few users are able to utilize syntax databases to this extent. So since we initially released syntax databases in 2006, we have tried to build knowledge of syntax into the typical user workflow.

Users like to do “word studies.” They like to start with a Greek, Hebrew or Aramaic word and see how else that word is used in the Bible. In Logos Bible Software<sup>36</sup> we have a feature called “Bible Word Study.” A user can begin in an English or Greek or Hebrew text and use this feature to learn about the word in context.

<sup>36</sup> Windows, Macintosh, iOS and Android versions



As an example, consider the word βασιλεία (typically translated “kingdom”) in Mark 1.15. In Logos Bible Software, the user can right-click the word (or, in Bibles with a reverse interlinear alignment, the English translation of the word) and select “Bible Word Study,” which will automatically generate a report with several different sections of information. Two sections are of interest for our context, one is called “Grammatical Relationships” and the other is “Preposition Use.”

## Grammatical Relationships

Grammatical Relationships provides information on the use of the subject word (here βασιλεία)<sup>37</sup> in different syntactic contexts:

Bible Word Study | βασιλεία

βασιλεία

Add

GRAMMATICAL RELATIONSHIPS

Subject of ...

▶ εἰμί	◀	eimi	be, exist	20
▶ ἐγγίζω	◀	engizō	draw near; approach	6
▶ ἐρχομαι	◀	erchomai	come; go	5
▶ ὁμοιῶ	◀	homoioō	make like; compare	4
▶ γίνομαι	◀	ginomai	be born, become	3

more »

Object of ...

▶ κληρονομῶ	◀	klēronomeō	inherit; acquire	5
▶ λαμβάνω	◀	lambanō	take	3
▶ κηρύσσω	◀	kēryssō	announce; proclaim	3
▶ εἶδον	◀	eidon	see	3
▶ ποιέω	◀	poieō	make; do	2

more »

Adjunct of ...

▶ εἰσερχομαι	◀	eiserchomai	enter; come into	12
▶ πίνω	◀	pinō	to drink	3
▶ εἰμί	◀	eimi	be, exist	3
▶ καλέω	◀	kaleō	call	3
▶ ἐρχομαι	◀	erchomai	come; go	2

more »

Used Adjectivally

▶ θεός	◀	theos	God; god	69
▶ οὐρανός	◀	ouranos	heaven	32
▶ αὐτός	◀	autos	self; he	10
▶ σύ	◀	sy	you, your	5
▶ ἐμός	◀	emos	my, mine	3

more »

Modified by Adjective ...

▶ μυστήριον	◀	mystērion	secret	3
▶ εὐαγγέλιον	◀	euangelion	good news	3
▶ υἱός	◀	huios	son	2
▶ κληρονόμος	◀	klēronomos	heir; beneficiary	1
▶ ῥάβδος	◀	rhabdos	rod, stick	1

more »

So βασιλεία is the subject of a clause with the verb of ἐγγίζω 6 times in the New Testament, according to the *OpenText.org Syntactically Analyzed Greek New Testament*. If that section of the report is clicked on, it will be expanded with the text of each of those 6 hits, like below:

<sup>37</sup> A video demonstrating Grammatical Relationships using βασιλεία is available on YouTube: <http://youtu.be/MWBDukofiRk>



εκ, απο and περι (with genitive) have been found to occur with βασιλεια as object.

The screenshot shows the 'Bible Word Study | βασιλεια' window. The 'PREPOSITION USE' section features a circular diagram with 'έν' (en) at the center. Arrows point to various prepositions: 'ἀνά' (ana), 'ἐπί' (epi), 'παρά' (para), 'ἐκ' (ek), 'κατά' (kata), 'ὑπό' (hupo), 'διὰ' (dia), 'πρός' (pros), 'σύν' (syn), 'μετά' (meta), and 'ἀντί' (anti). A legend on the right lists: 'περί G', 'ὑπέρ G', 'ὑπό G', and 'κατά A'. Below the diagram, the results for 'εις' (eis) in the accusative case are listed:

Reference	Text (Greek)	Text (English)
Mt 5:20	Λέγω γὰρ ὑμῖν ὅτι ἐάν μὴ περισσεύσῃ ὑμῶν ἡ δικαιοσύνη πλεῖον τῶν γραμματέων καὶ Φαρισαίων οὐ μὴ εἰσέλθῃτε εἰς τὴν βασιλείαν τῶν οὐρανῶν	For I tell you, unless your righteousness exceeds that of the scribes and Pharisees, you will never enter the kingdom of heaven.
Mt 7:21	Οὐ πᾶς ὁ λέγων μοι κύριε κύριε εἰσελεύσεται εἰς τὴν βασιλείαν τῶν οὐρανῶν ἀλλ' ὁ ποιῶν τὸ θέλημα τοῦ πατρὸς μου τοῦ ἐν τοῖς οὐρανοῖς	"Not everyone who says to me, 'Lord, Lord,' will enter the kingdom of heaven, but the one who does the will of my Father who is in heaven.
Mt 18:3	καὶ εἶπεν ἁμῖν λέγων ὑμῖν ἐάν μὴ στραφῇτε καὶ γένησθε ὡς τὰ παιδία οὐ μὴ εἰσέλθῃτε εἰς τὴν βασιλείαν τῶν οὐρανῶν	and said, "Truly, I say to you, unless you turn and become like children, you will never enter the kingdom of heaven.
Mt 19:23	Ὁ δὲ Ἰησοῦς εἶπεν τοῖς μαθηταῖς αὐτοῦ ἁμῖν λέγω ὅτι πλούσιος δυσκόλως εἰσελεύσεται εἰς τὴν βασιλείαν τῶν οὐρανῶν	And Jesus said to his disciples, "Truly, I say to you, only with difficulty will a rich person enter the kingdom of heaven.
Mt 19:24	πάλιν δὲ λέγω ὑμῖν εὐκοπώτερον ἐστὶν κάμηλον διὰ τρυπήματος ραφίδος διελθεῖν ἢ πλούσιον εἰσελθεῖν εἰς τὴν βασιλείαν τοῦ θεοῦ	Again I tell you, it is easier for a camel to go through the eye of a needle than for a rich person to enter the kingdom of God."

In the above graphic, the instances of εις plus an accusative βασιλεια are listed; the listing is activated by clicking on the graphic.

The data from both Grammatical Relationships and Preposition Use comes from a syntactically annotated text,<sup>39</sup> but the user had no direct interaction with the analysis—no queries were written—in order to retrieve the information. Instead, we have pre-determined the sorts of information we can glean from a syntactic analysis, and provide that information to the user in the context of the Bible Word Study report.<sup>40</sup>

## Using Query Forms

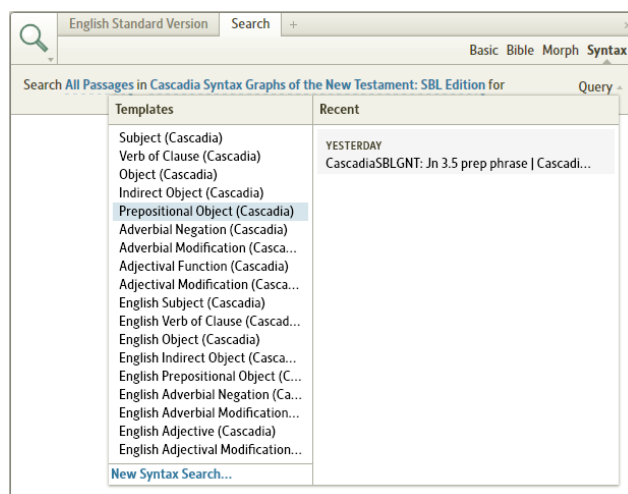
Another method to ease in to searching syntactic data is to provide templates for the most common sorts of queries that users might desire to run. Logos Bible Software provides templates that are accessible by two different methods. The first method is Query Forms.<sup>41</sup>

In the Syntax Search dialog, there is a Query dropdown list. This list contains a set of templates (on the left) as well as access to other queries you have written (on the right).

<sup>39</sup> The OpenText.org analysis for the Greek, the Andersen-Forbes analysis for the Hebrew and Aramaic.

<sup>40</sup> Alternately, users can create their own custom reports using these sections, so a user specifically interested in this type of data could create a specialized report containing only the Grammatical Relationships and Prepositional Use sections.

<sup>41</sup> A video demonstrating query forms is available on YouTube: <http://youtu.be/dmar7jHT4hQ>



When a template (such as “Prepositional Object”) is selected, a textbox appears. In this case, if we type βασιλεια in the box, we will be searching for where βασιλεια is the object of a preposition. Hit enter, or click the “Go” button, and the search will run.

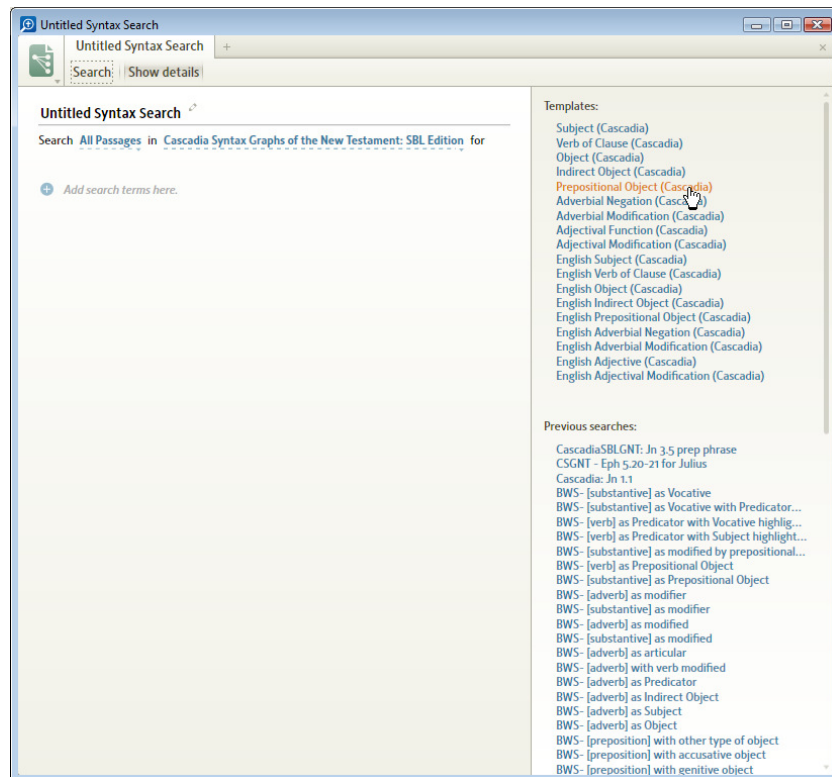


A complex search involving prepositional phrases is run without needing to understand the intricacies of the theories behind the syntactic analysis. The notion of a prepositional object is fairly common and fairly easy to understand for those who have had some training or done some reading in grammar. These syntax query forms are delivered along with the syntax database, giving users an easier way to locate the more common structures — typically structures reported by Grammatical Relationships.

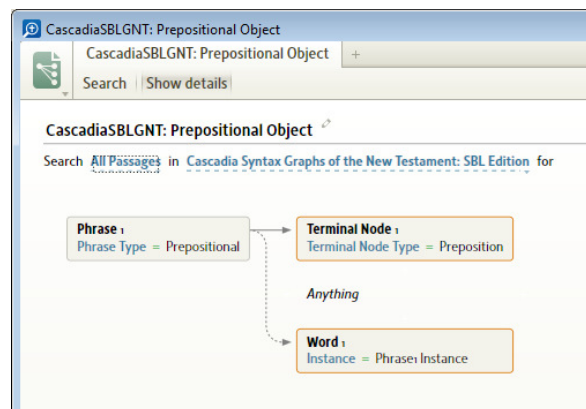
## Syntax Search Templates

Another method to search a syntactic analysis without writing a new query is to use a template.<sup>42</sup> The same information used to drive the Query Forms is also available as a query, ready to be loaded in the Syntax Search Query Editor. Using the File menu, create a new, empty Syntax Search Document. On the right side of the new query is a list of available templates. If the “Prepositional Object (Cascadia)” template is selected, it will use the template to form the initial query.

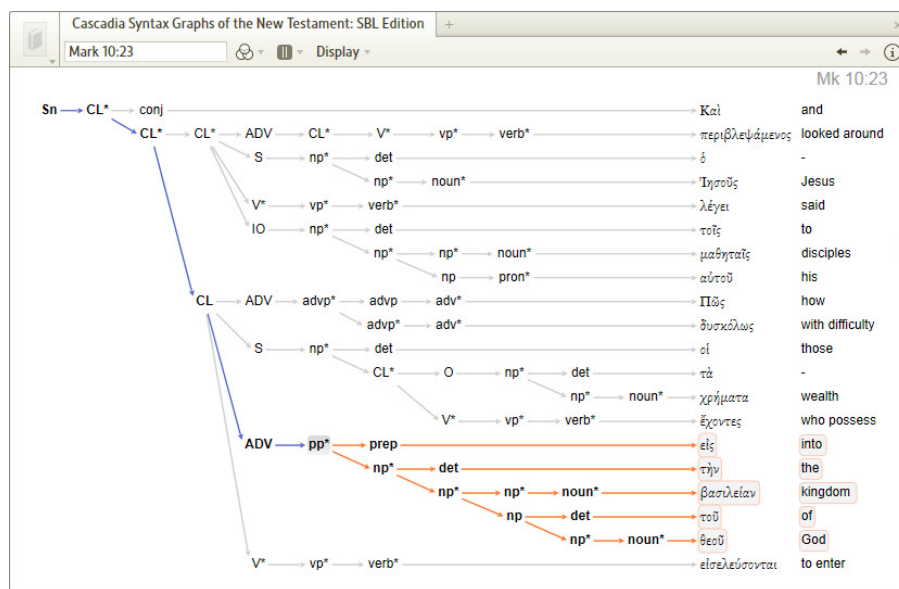
<sup>42</sup> A video demonstration of Syntax Search Templates is available on YouTube: <http://youtu.be/VJ2mjyxb-Ko>



This will insert the necessary query information into the query canvas:

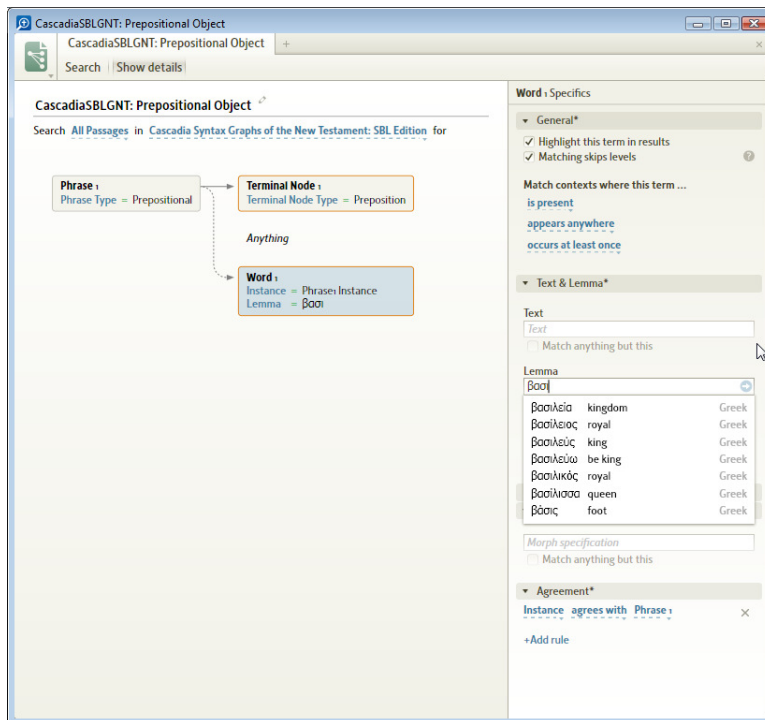


Important to note here is how the query is formed, and how that roughly matches the representation of the query in a syntax graph. As an example, here is the prepositional phrase εις την βασιλειαν in Mk 10.23:



Note how the “pp” is equivalent in structure to the syntax query’s phrase object with specified type of “prepositional.” In the graph, one arrow points directly to the “prep,” represented in the syntax query by the terminal node object with type preposition. Another arrow in the graph points to a noun phrase structure which has the object (with article). The query handles this by noting with a dotted line that structure levels are skipped until the specified word object is found.<sup>43</sup>

To modify the query, just select the word object and add the lemma or whatever other criteria you desire to locate.



By exploring and experimenting with the available templates, one can begin to understand how queries are best

<sup>43</sup> If there is time, I will explain in more detail the “Instance” agreement, which is key to compact queries that span levels in the section below about the Cascadia databases.

formed and begin to build knowledge of how to utilize a syntactic analysis.

## Descriptions of Syntactic Analyses of the Greek New Testament Available in Logos Bible Software

Important to using a syntactic analysis is an understanding the approach and terminology used. One need not fully agree with the theory or framework behind the analysis, one must only understand the analysis enough to reliably predict or comprehend how structures are marked.

In this light, short descriptions of each analysis are presented below.

### Cascadia Syntax Graphs of the New Testament (CSGNT)

The *Cascadia Syntax Graphs of the New Testament* (CSGNT) is based on the work of the Asia Bible Society's Greek Syntactic Treebank Project.<sup>44</sup> As previously mentioned, the CSGNT is unique in that it is built upon a computer-readable Greek grammar that is used by a parser to generate syntactic trees. These trees are reviewed, corrected and refined through an editorial process; the corrections and edits become a knowledgebase that informs the parser on future editions.

The analysis is essentially from the clause level down through the word level, though clauses are grouped and combined into higher-level structures called "sentences."<sup>45</sup> The CSGNT is not innovative in terminology, it is instead rigorous in the application of the grammar to the entire corpus and its refinement.<sup>46</sup>

There are four primary levels of analysis:

- Sentence and Clause
- Clause Function
- Phrase
- Terminal Node

Words are represented as nodes, but the word-level data comes from the Logos Bible Software morphological analysis. This includes lexical and morphological information, Louw-Nida semantic domain annotation, and English glosses.

Further note that the framework behind CSGNT is head-driven phrase structure grammar (HPSG). The notion of "head-driven" is important, because, as Aubrey notes,

... not only are phrases headed by their lexical counterpart (i.e. nouns head noun phrases), but also the grammatical and morpho-syntactic information encoded in the lexical entry for a given inflectional form is carried from the terminal node level all the way up the headed phrase structure. This includes, subject agreement, noun phrase agreement, valency, and any number of other grammatical features.<sup>47</sup>

In the Logos implementation of the CSGNT, this means that the lexical and morphological properties of the head of a given node structure are encoded not only at the word level, but propagated up with each node as

---

<sup>44</sup> There are *Cascadia* analyses of both NA27 and SBLGNT.

<sup>45</sup> Though note "The label 'Sentence' itself has been used for the lack of a clearly preferable alternative term." Wu, A., & Tan, R. K. (2009; 2009). *Cascadia Syntax Graphs of the New Testament: Glossary*. Logos Research Systems, Inc.

<sup>46</sup> Much of the following discussion is based directly on Wu, A., & Tan, R. K. (2009; 2009). *Cascadia Syntax Graphs of the New Testament: Glossary*. Logos Research Systems, Inc.

<sup>47</sup> Mike Aubrey, "The Theory behind the Cascadia Syntax Graphs," online: <http://evepheso.wordpress.com/2010/03/21/the-theory-behind-the-cascadia-syntax-graphs/>



appropriate. If one is searching for a particular word used as a subject, one only need specify the word at the Clause Function node in the query, and need not worry about sifting through the structures between the Clause Function and the Terminal Node or even the Word node.

## Sentences and Clauses

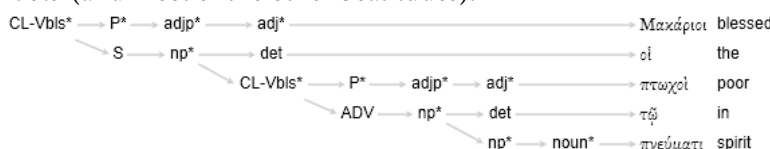
In the CSGNT clauses may contain other clauses, clause functions, and terminal nodes that represent conjunctions. In the Logos Bible Software representation, the Sentence is the highest clause node, and contains a single clause that contains the rest of the grammatical structure. On sentences, the CSGNT Glossary notes:

Clear groupings of clauses have been connected together into sentences, especially when subordination or ellipsis are involved. However, the identification of sentences has not undergone the same vigorous process of testing as the identification of structures at the clause level and below.<sup>48</sup>

Apart from Sentences, there are four types of clauses annotated:

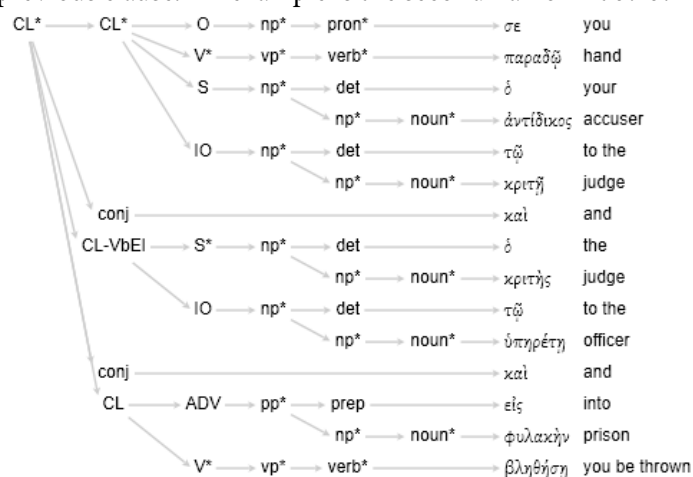
**Clause (Verbal).** These are the most commonly occurring clause structures and contain a verbal element as the head of the clause.

**Verbless Clause.** These clauses have no explicit verbal element and instead typically consist of Subject and Predicate clause functions. Some sort of verbal element is usually implied, typically some form of εἰμι (“to be”). An example would be Mt 5.3 (and most of the other beatitudes):



Verbless Clause in Mt 5.3

**Verb Elided Clause.** These clauses have no explicit verbal function but imply the “carrying over” of a verbal function, typically from a previous clause. An example is the second half of Mt 5.25:

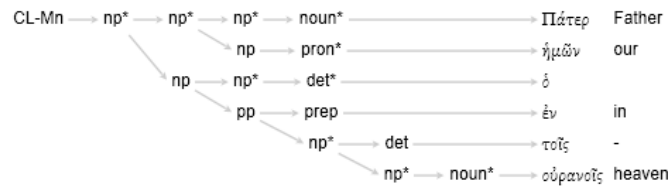


Mt 5.25b

**Minor Clause.** These clauses have no predication; they have no assertions or propositions. Tan notes “They function interpersonally (vocative direct address or interjection to gain attention/alert) or textually (left-dislocated focus nominal phrases, e.g. as for the game, I did not get to watch it).”<sup>49</sup> One example is the first clause of the Lord’s Prayer as recorded in Mt 6.9:

<sup>48</sup> CSGNT Glossary

<sup>49</sup> CSGNT Glossary



Mt 6.9b

## Clause Functions

What CSGNT labels “Clause Function,” or the elements which function within a clause, other analyses label clause components or constituents. These are the primary components of a clause, such as subject, verb and object. Clause Functions contain Phrases and may contain other Clauses.

Most clause functions have labels that clearly describe their function. These include:

- Adverbial Function
- Indirect Object Function
- Object Function
- Subject Function
- Verbal Function

Some labels require further explanation. These include:

- **Second Object Function.** When a verb takes two objects, the second object is noted with the Second Object Function. These are typically double accusatives and object complements.
- **Predicate Function.** A constituent that makes an attribution or identification about the Subject of a non-verbal (verbless or verb-elided) clause.<sup>50</sup>
- **Verbal Copula Function.** A constituent that facilitates an attribution or identification about the Subject. This typically corresponds to a “to be” verb in English, which indicates a relation between the Subject and the Predicate rather than any action/event.<sup>51</sup>

Because the Cascadia analysis is head-driven, each clause function, in addition to delimiting the words and phrases included in the structure, grammatical information on the head term is included with the clause function node. This means that one could search for a Clause Function of type “Subject” where the head term is a particular word (e.g. Ἰησοῦς).

## Phrases

Phrases are the structures that typically stand between clause functions and terminal nodes, though it is possible for phrases to contain clause structures. Below are the types:

- Adjectival Phrase
- Adverbial Phrase
- Nominal Phrase

<sup>50</sup> CSGNT Glossary

<sup>51</sup> CSGNT Glossary

- Numeral Phrase
- Prepositional Phrase
- Verbal Phrase

Nominal Phrases include phrases headed by adjectives or pronouns when the adjective or pronoun function substantively, like nouns.<sup>52</sup> Prepositional Phrases are formally headed by the object of the preposition, not the preposition itself.

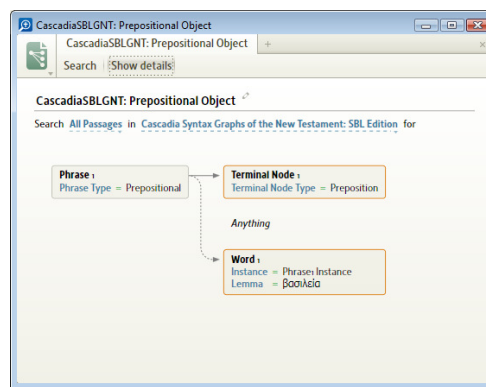
## Terminal Nodes

Terminal nodes are the bottom-most nodes of the analysis and have a one-to-one relationship with the words of the edition. They have a direct relationship with the word they represent. Type is typically determined by the part of speech of the word the terminal node contains. The following types of terminal nodes are available:

- Adjective
- Adverb
- Conjunction
- Determiner (Article)
- Interjection
- Noun
- Numeral
- Preposition
- Particle
- Pronoun
- Verb

## Using Head Term Data in Searches


Each node's head term (from Sentence nodes to Terminal nodes, and all nodes in between) also encodes an instance relationship, directly associating the head node with the word instance it represents. A good example of this is the search for a prepositional phrase with the object of βασιλεια which was illustrated earlier.



<sup>52</sup> CSGNT Glossary

In the above, note that the Word object has “Instance = Phrase<sub>1</sub> Instance” as a constraint. This constrains the word in this lookup to be the instance that is also the head term of the Phrase object. Each word object has a unique “Instance” identifier. When a node such a Phrase or Clause Function encodes head term information, it also encodes the Instance identifier of the head term. When a node specifies instance agreement with a word node, it means there is a direct relationship created between the two nodes.

The most direct benefit to specifying instance agreement is in search hit highlighting. When reporting syntax search hits, there are two levels of highlighting. First, the entire structure matched (so, in this case, the prepositional phrase) is given a light grey highlight. Second, any further nodes with highlighting specified (noted by the yellow border around the node, specified by right mouse click or through menu selection) will highlight the necessary words in a different color in the search results. This means that results can specify what is highlighted, like below. The entire structure has a shaded background, the preposition is highlighted with red text and the prepositional object is highlighted with blue text.


The Greek New Testament: SBL Edition

**Matt 5:19** ὃς ἐὰν οὖν λύσῃ μίαν τῶν ἐντολῶν τούτων τῶν ἐλαχίστων καὶ διδάξῃ οὕτως τοὺς ἀνθρώπους ἐλάχιστος κληθήσεται **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν ὃς δ' ἂν ποιήσῃ καὶ διδάξῃ οὗτος μέγας κληθήσεται **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν

ὃς ἐὰν οὖν λύσῃ μίαν τῶν ἐντολῶν τούτων τῶν ἐλαχίστων καὶ διδάξῃ οὕτως τοὺς ἀνθρώπους ἐλάχιστος κληθήσεται **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν ὃς δ' ἂν ποιήσῃ καὶ διδάξῃ οὗτος μέγας κληθήσεται **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν

**20** λέγω γὰρ ὑμῖν ὅτι ἐὰν μὴ περισσεύσῃ ὑμῶν ἡ δικαιοσύνη πλεον τῶν γραμματέων καὶ Φαρισαίων οὐ μὴ εἰσέλθητε **εἰς τὴν βασιλείαν** τῶν οὐρανῶν

**7: 21** Οὐ πᾶς ὁ λέγων μοι Κύριε κύριε εἰσελεύσεται **εἰς τὴν βασιλείαν** τῶν οὐρανῶν ἀλλ' ὁ ποιῶν τὸ θέλημα τοῦ πατρὸς μου τοῦ ἐν τοῖς οὐρανοῖς

**8: 11** λέγω δὲ ὑμῖν ὅτι πολλοὶ ἀπὸ ἀνατολῶν καὶ δυσμῶν ἤξουσιν καὶ ἀνακλιθήσονται μετὰ Ἀβραάμ καὶ Ἰσαάκ καὶ Ἰακώβ **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν

**11: 11** ἀμὴν λέγω ὑμῖν οὐκ ἐγγίγνεται ἐν γεννητοῖς γυναικῶν μείζων Ἰωάννου τοῦ βαπτιστοῦ ὁ δὲ μικρότερος **ἐν τῇ βασιλείᾳ** τῶν οὐρανῶν μείζων αὐτοῦ ἐστίν

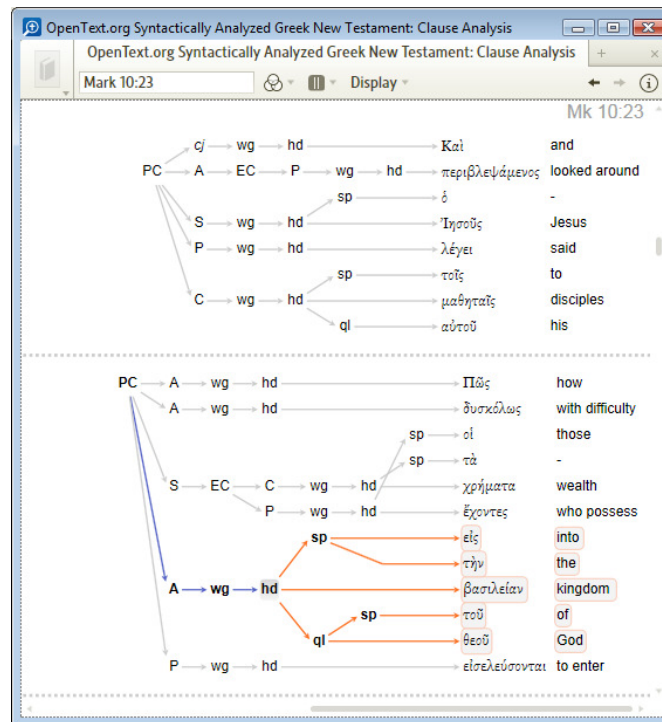
**13: 41** ἀποστελεῖ ὁ υἱὸς τοῦ ἀνθρώπου τοὺς ἀγγέλους αὐτοῦ καὶ συλλέξουσιν **ἐκ** τῆς βασιλείας αὐτοῦ πάντα τὰ σκάνδαλα καὶ τοὺς ποιοῦντας τὴν ανομίαν

**43** Τότε οἱ δίκαιοι ἐκλάμπουσιν ὡς ὁ ἥλιος **ἐν τῇ βασιλείᾳ** τοῦ πατρὸς αὐτῶν ὁ ἔχων ὥτα ἀκουέτω

The instance data allows for a more visually descriptive display of search hits.

## OpenText.org Syntactically Analyzed Greek New Testament (OpenText.org)

The Logos Bible Software implementation of the OpenText.org Syntactically Analyzed Greek New Testament was released in 2006. Ostensibly based on Hallidayan Systemic Functional Linguistics (SFL), the model is simple but innovative in terminology. That is, there under 20 terms to learn, but they are terms that are not commonly used in traditional approaches to the grammar and syntax of the Greek New Testament. Students with a traditional understanding of the Greek of the New Testament have had some difficulty with being able to use this analysis, particularly the word group level. Below is an example of Mark 10.23, to compare with the same verse in the Cascadia analysis presented above.



With OpenText.org, there are three levels of annotation: Clause level, Word Group level, and Word level. The word level annotation amounts to a morphological analysis, so it will not be discussed here. For the basics of the definitions, I am relying on the *The OpenText.org Syntactically Analyzed Greek New Testament: Glossary*.<sup>53</sup>

## Clause Level

There are three types of clauses in the OpenText.org annotation.

- **Primary Clause:** Usually a clause with a finite verb that stands on its own. It is an independent clause. Primary clauses may contain secondary clauses and embedded clauses.
- **Secondary Clause:** A dependent or subordinate clause. Examples include relative clauses, adverbial clauses, and instances of genitive absolute.
- **Embedded Clause:** A clause with a non-finite verb (thus participial or infinitive clauses), typically embedded within a clause component.

Clauses consist of clause components. There are six types of clause components:

- **Subject:** The grammatical subject of the clause.
- **Predicator:** The verbal element of the clause.
- **Complement:** The element that completes the predicator of the clause.
- **Adjunct:** An element that provides circumstance to the clause. These are typically adverbs, negators, and prepositional phrases.
- **Conjunction:** An element that contains a word that provides a linking relationship to previous or following clauses.

<sup>53</sup> Porter, S., O'Donnell, M. B., Reed, J. T., Tan, R., & OpenText.org. (2006; 2006). *The OpenText.org Syntactically Analyzed Greek New Testament Glossary*. Logos Research Systems, Inc.

- **Addressee:** Typically consists of a word group with a head term in the vocative case.

## Word Group Level

The Word Group level annotation is an attempt to describe the relationships within the clause component. Each word group consists of a head term and optional modifiers. There are five types of modifiers:

- **Specifier:** Specifiers are typically articles or prepositions when the prepositional phrase functions adverbially.
- **Definer:** Definers are typically adjectives or appositional structures.
- **Qualifier:** Qualifiers are typically words in the dative or genitive cases.
- **Relator:** A relator is the preposition of a prepositional phrase that functions adjectivally within the current word group.
- **Connector:** A connector is a conjunction that functions within the current word group or which connects two word groups.

## Lexham Syntactic Greek New Testament (Lexham SGNT)

The Lexham SGNT is a fairly traditional analysis of the Greek New Testament.<sup>54</sup> It is primarily concerned with structures that are typically found in stick (or “Reed-Kellogg”) diagrams. It also provides information on the “syntactic force” (e.g. objective genitive or subjective genitive) of each non-verbal term.<sup>55</sup> For terminological basis of these syntactic force annotations, it relies primarily on BDF and Wallace’s *Greek Grammar Beyond the Basics*. It has the largest terminological base, so terminology will not be discussed in any depth here. It is as well the most subjective of the syntactic analyses available in Logos Bible Software.

In Logos, this analysis is represented by four different resources:

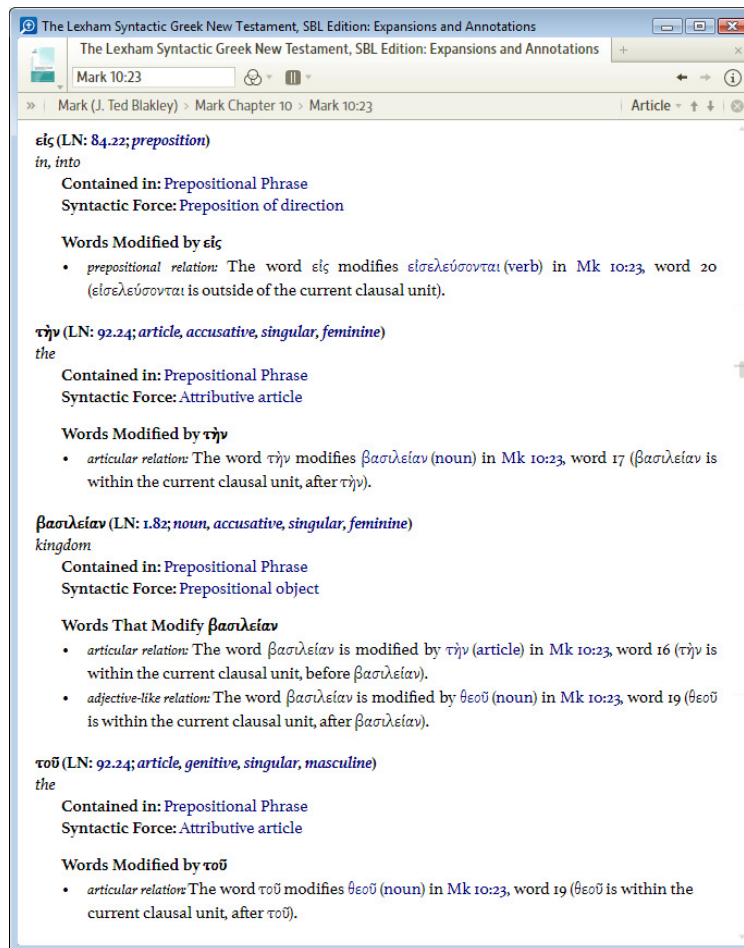
- **The Lexham Syntactic Greek New Testament:** An edition of the Greek New Testament formatted in a block-outline style, representing the sentences, clauses, and phrases.
- **The Lexham Syntactic Greek New Testament: Sentence Analysis:** A graph style visualization of the Greek New Testament
- **The Lexham Syntactic Greek New Testament: Expansions and Annotations:** A resource that is essentially a grammatical and syntactic commentary on every individual word of the Greek New Testament. Information on syntactic force, containing structures, morphology, and other data are listed and revealed in this resource.
- **The Lexham Syntactic Greek New Testament: Glossary:** A glossary of terminology with references to major reference grammars where applicable.

The *Expansions and Annotations* resource is perhaps the most innovative and useful in that it provides a sort of exhaustive version of Zerwick and Grosvenor’s *A Grammatical Analysis of the Greek New Testament*. Below is an example from the sample passage of Mark 10.23:

---

<sup>54</sup> Both the NA27 and SBLGNT have been analyzed.

<sup>55</sup> A possible future enhancement would be to pursue an analysis of verbs and their function/force (e.g. ingressive aorist) and transitivity.



Each word includes information on morphological properties, a Louw-Nida annotation, information on containing structures, information on syntactic force notations, and information on words that are in some sort of modification relationship (adjectival, adverbial, article, prepositional, and others) with the current word.

## Future Directions

The future directions of these sorts of databases can be summed up with two words: “refinement” and “discourse.”

### Refinement

Data sets that encode syntactic relations are, in the scope of things, still relatively new. We are still learning how to use them, how to enhance them, and how to present them to users.

The *Cascadia Syntax Graphs* analysis, as implemented by Logos Bible Software, is a good example of this. I had implemented both the Lexham SGNT and OpenText.org material in 2006. Insights from that experience and from interaction on the Logos forums on the use of these data sets were valuable in 2009 when implementing the CSGNT. These previous experiences led directly to the use of the “instance” agreement technique described above, among other things.

The Asia Bible Society (producers of the CSGNT) are as well a good example. They have recently (January 2012) provided revised and updated data of their analyses (NA27 and SBLGNT) which have in turn been compiled and delivered to our users.



## **Discourse Analysis**

In addition to refinement, many producers of syntactic data are also interested in using syntactic analyses as a basis for further work in discourse analysis; much like morphological analysis has been used as a foundation for syntactic analysis.