Locating New Testament Cross-References: Some Stragegies

Rick Brannan, Information Architect, Logos Bible Software
(rick@logos.com)

Marginal cross-references have long been a feature of several Bibles in print. Each of the myriad versions has some edition with "marginal cross-references" or "center-column cross-references". Yet electronic editions, apart from those reproducing data available in printed editions, have not done a good job of complementing the text with relevant cross-references. Most electronic editions of Bibles are centered on the words of the text and not its presentation or on supplying ancillary data to help in the study of the text.

NO TECH

Someone else has already done the work; if it is in the public domain or licensable, then you can use their work.

LOW TECH

The "low-tech" approaches typically involve reshuffling and processing of available data sources. These can be at the "section" (multiple verse) level, the verse level, the term level, or even at a thematic level.

Section-to-Section

Gospel Parallels: Eusebian Canon Tables

Where the Eusebian canons appear in print, the numbers representing canon boundaries typically occur in the margin of the running text of the New Testament. So one could follow the margins and create reference ranges for each canon reference. Such data allows one to write a program to read the canon table data, define pericope boundaries and associations, and convert the data into something more usable in an electronic environment.

OT Quotes in the NT

When Old Testament text is quoted, cited or alluded to in the New Testament, knowledge of the source citation is handy (to say the least). Listing these citations as cross references makes sense as the source context and phrasing may be exegetically valuable.

Thematic

Some resources, such as *Nave's Topical Bible* (Nave's) and *Torrey's New Topical Textbook* (Torrey's) are available for this purpose. The original editions of both texts are in the public domain; either could serve as a source of data for topical cross-references. Nave's has a huge amount of topics with relatively long lists of references while Torrey's is more brief (630+ top-level topics, 23,000+ sub-topics, more than 21,000 unique references by my calculation); to my mind this makes Torrey's perhaps the better choice to serve as a basis for topical cross-reference data for a given passage.

Term-to-Term

Louw-Nida Semantic Domain

If each word in the Greek New Testament is tagged with its LN identifier, words could be concorded by their semantic sense. This could provide interesting fodder for cross-references, particularly for infrequently-used senses or other items of significance.

Person and Place Names

These are typically people and places (ranging from buildings and streets to cities and larger geographic regions like countries and even mountain ranges or bodies of water). Internally at Logos we have databases of people and places that are being used in all sorts of ways; these datasets are forming the basis of something we're calling the Bible Knowledgebase. Sean Boisen's *New Testament Names* database is available online at SemanticBible.com; OpenBible.info has a database of Bible places available as well. Alternately; Louw-Nida domain 93 is all about People (subdomain A) and Places (subdomain B); so they have already identified and provided at least some references for all people and places in the New Testament.

MO' TECH

The "Mo' tech" approach involves computing term significance or similarity between textual groups (verses, pericopes, etc.). In all cases, the text used as basis of comparison is the UBS[4] edition of the Greek New Testament.

Word-to-Word Cross References

Statistically Improbable Words

In any given text, some words occur more frequently; other words occur less frequently. There are statistical measures that can help in evaluating these words. This approach compares the use of a word in a given book of the NT with its usage across the whole NT. It highlights words that have an improbable frequency in a given book and words that have an improbable infrequency in a given book.

Phrase-to-Phrase Cross References

N-grams and Repeated Word Groups

Many times, cross references involve lexical "hooks"—similar phrases or keywords. One method to examine, therefore, is repetition of similar phrases. This method examines each verse (as defined by the NA/UBS text) and builds N-grams of each verse. The N-grams are then compared to each other and where they match a cross-reference is generated. Instead of acting on English text, the underlying Greek text is used as basis. In addition, instead of the inflected form of the Greek word, the dictionary form is used, making the matching a bit more flexible.

Common Substrings

This is a variation to the N-gram appraoch, comparing each verse to every other verse in the New Testament, keeping track of the number of characters held in common between the two verses. Verses that have many characters in common (usually phrases) are considered related. This is similar to but not exactly the same as the N-gram approach mentioned above.

Using Three-Word Phrases

This approach is another variation on the N-gram and similar substring approaches. Instead of searching out the longest common substring between two verses; it instead compiles all three-adjacent-word combinations (tri-grams), building a concordance of three-word phrases. This method could be used to generate potential cross-references using three-word phrases as the point of commonality.

Using Modifiers in OpenText.org SAGNT

Programmatically delimiting the extent of a phrase or clause is somewhat unreliable, but when a corpus has been annotated with such boundaries, relying on clause and phrase boundaries is possible. In the *OpenText.org Syntactically Analyzed Greek New Testament*, the phrase boundary work has been done in the word group annotation level. Even better, modification relationships within each word group have also been annotated. Grouping not only where the same word(s) occur, but where they occur in similar modification relationship to each other holds more promise than the typical adjacent-word N-gram approach for similar size groups of words.

Section-to-Section Cross References

In the area of section-to-section cross-references, one method that may have promise is the use of a vector similarity algorithm to establish similarity between 'documents' (here sections/pericopes). Unfortunately, I did not have time to explore this option in depth so I cannot report on it here.

The approach does hold promise. The idea is to use document comparison measures to determine how similar documents are; results would then be grouped by similarity. This should, in theory, cause the most similar pericopes to group together; and the relationships between these different sections/pericopes could be interesting.

CONCLUSION

Hopefully this paper has shown that there are several sources to consult and exploit when it comes to considering cross-references for the New Testament. The ideas mentioned in this paper are only the tip of the iceberg; there are surely a myriad of possibilities, both in the creation and analysis of potential datasets.

In the creation of a cross-reference database, however, one item has become clear to me as I've considered the different approaches in this paper. Cross-references are helpful, but recording the reason for the cross-reference can also be just as helpful. How many times have you followed a cross-reference only to come across a passage that doesn't seem to have much of anything to do with the referring passage? Recording the reason for the cross-reference—be it because of similar words, similar phrasing, common names or places, common themes—is something that automated methods can begin to accomplish. As new and updated cross-reference databases are created and enhanced for the next century of Bible study, schemes for communicating the *what* (reference) along with the *why* (reason) should be part of the equation.