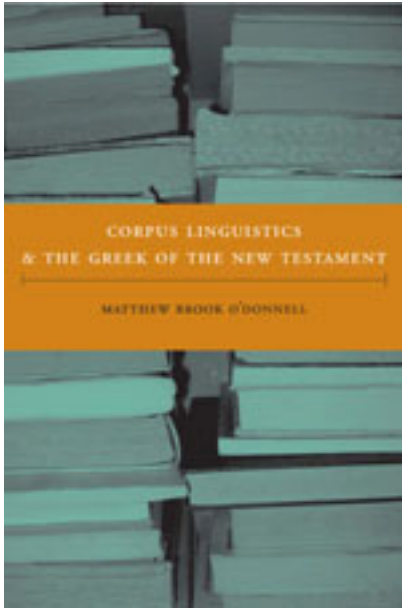


RBL 04/2007



**O'Donnell, Matthew Brook**

***Corpus Linguistics and the Greek of the New Testament***

New Testament Monographs 6

Sheffield: Sheffield Phoenix Press, 2005. Pp. xxii + 552.

Hardcover. \$95.00. ISBN 1905048114.

Paul Elbert

Church of God Theological Seminary  
Cleveland, Tennessee

Corpus linguistics is the computer-assisted method of studying a naturally occurring language when collected into a representative electronic sample. It has the potential to aid in the discovery, corroboration, and understanding of original meaning and authorial intent. The data sample or corpus must be appropriately annotated (or encoded) with computer-recognizable information reflecting linguistic features that are lexical (indicating the Greek lexeme of a word), grammatical (parts of speech and other elements), syntactical (clause elements and types), and semantic (semantic domains). This approach potentially allows for a closer inspection and realization of connective thought via the identification of linguistic elements and patterns.

The Greek language under consideration here is that written by native and nonnative speakers throughout the Hellenistic and Roman worlds from the fourth century B.C.E. to the fourth century C.E. (for considerations pertaining to such a temporal grouping within the continuum of the Greek language, see Chrys C. Caragounis, *The Development of Greek and the New Testament: Morphology, Syntax, Phonology, and Textual Transmission* [WUNT 67; Tübingen: Mohr Siebeck, 2004]). Naturally, for the ultimate success of the method, the texts “must be classified and grouped in a manner that allows for representative statements to be made” (O'Donnell, 137). The goal in *Corpus Linguistics* is to study the

selected Greek on its own merits, which is not, for example, how students are often taught to compose Greek sentences from English ones, where going from a weakly to a strongly inflected language and from a relatively fixed word order to a relatively flexible one might contribute to a new, modern “Hellenistic” dialect/register or even a new language (8). In this vein of affording objective consideration to the language, perhaps it goes without saying that the method of computer-aided textual analysis being developed here is, by its very nature, not indebted to hidden presuppositions requiring prooftexting to “make-it-fit” or to the a priori imposition of philosophical speculation upon a New Testament text to “make-it-fit” as an unarticulated interpretive tool.

O’Donnell reviews linguistic theory, noting that although his method is primarily quantitative, it can allow for the qualitative, namely, the ambiguity inherent in human language (24). In this way, the computer-assisted analysis of Hellenistic Greek, in as much depth as possible, can provide access to data that may be otherwise unobservable but that might in certain instances, such as the interdependent relationship of style and grammar, lead to a reliable understanding of what an author is actually doing to express thought (cf. M. Stubbs, *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture* [Oxford: Blackwell, 1996]; S. Bayer et al., “Theoretical and Computational Linguistics: Toward a Mutual Understanding,” in *Using Computers in Linguistics: A Practical Guide* [ed. J. M. Lawler and H. A. Drys; London: Routledge, 1998], 231–55).

O’Donnell points out that significant new linguistic discoveries are possible via careful examination of unannotated texts, as those now in the *Thesaurus Linguae Graecae*, an electronic thesaurus at the University of California at Irvine that includes virtually all ancient Greek texts surviving between Homer (eighth century B.C.E.) and 600 C.E., as well as a large number of texts deriving from the period between 600 C.E. and the fall of Byzantium in 1453. Such linguistic-oriented discoveries that O’Donnell alludes to may be illustrated, for example, by earlier and recent efforts based on the unannotated Greek text of Luke-Acts (S. H. Levinsohn, *Textual Connections in Acts* [SBLMS 31; Atlanta: Scholars Press, 1987]; D. Mealand, “Luke-Acts and the Verbs of Dionysius of Halicarnassus,” *JSNT* 63 [1996]: 63–86; P. Elbert, “An Observation on Luke’s Composition and Narrative Style of Questions,” *CBQ* 66 [2004]: 98–109). However, O’Donnell stresses that the quantitative examination of annotated texts may quite realistically be expected to “facilitate the discovery of new linguistic facts, unobservable to the armchair linguist. This is particularly the case for an epigraphic language such as Hellenistic Greek” (29).

In “Previous Studies and the History of Corpus Linguistics and Authorship Attribution,” (38–84), O’Donnell reviews grammatical work relative to both Hellenistic Greek and the papyri (the work of Moulton-Turner, Robertson, Porter, and Mandilarus comes to the fore), but the work of Gignac on the grammar of the papyri seems to be overlooked. After

a brief overview of some corpus-based linguistic studies, we find an excursus on “Stylistics, Stylometry and Authorship Studies” (85–101). Here work on style and statistics relative to various linguistic elements and their frequency is helpfully reviewed as it applies to authorship questions. Regarding the Pastorals, O’Donnell finds that apparent statistical precision can often be meaningless, producing only a psychological effect, not a scientific basis for any assured conclusions. Particularly interesting, however, is his assessment of A. Kenny, *A Stylometric Study of the New Testament* (Oxford: Clarendon, 1986), who found a high correlation (using criteria such as noun frequency, conjunction occurrence, and tense-form distribution) between Colossians and Ephesians. The Pastorals also correlate highly with each other; 1 and 2 Timothy correlate significantly with other Paulines; while Titus, due to its short format, is undetermined.

In general, as to previous statistical stylistics applied to authorship, “many of these studies have rested upon inadequate linguistic and statistical foundations. They have frequently failed to make the transition from observing differences of numerical significance (in, for example, *hapax legomena*, sentence length, dependent genitives, and common conjunctions) to understanding and applying their linguistic and contextual significance.” In a later excursus entitled “Lexical Measures in Authorship Studies” (387–94), O’Donnell supports this assertion with two heuristic tables, “Vocabulary Frequency Data for Pauline Letters” and “Comparison of Unique Types in Texts in Pauline Corpus with Romans Excluded.” This evidence suggests that it is curious why the counting of unique words or *hapax legomena* has “gained a place of such prominence in discussion of the authorship of New Testament documents” (383). Even an initial analysis of adjectives unique to the Pastorals, when compared to the ten-letter Pauline corpus, suggests that looking at collocations of words and not just extracted words is the way forward (392–93). Indeed, inspection of this preliminary evidence that O’Donnell presents suggests that some New Testament scholarship has gotten well ahead of itself. Therefore, I think we would be wise to concur that “Future studies must utilize more complex multivariate statistical processes and base their criteria and interpretations on a sound linguistic framework” (101).

The multivariate technique of which O’Donnell speaks would be able to examine many variables within a data sample. Perhaps this approach might be applied to the Thessalonian correspondence as compared to the Pauline corpus, finding a quantitative as well as a qualitative way to include linguistic factors bearing on “theological content,” together with sensitivity to identical vocabulary and features such as particles, adverbs, and/or verbal affixes. As to the unannotated text of 1 and 2 Thessalonians, I would call attention to three previous studies and then suggest a complementary procedure for such a multivariate computer-aided examination of these letters with respect to the Pauline corpus. I. H. Marshall (“Pauline Theology in the Thessalonian Correspondence,” in *Paul*

and Paulinism: Essays in Honour of C. K. Barrett [ed. M. D. Hooker and S. G. Wilson; London: SPCK, 1982], 173–83) finds five theological concepts to be absent from the Thessalonian epistles and eight constant factors therein. He also finds five points of similar style that they share and concludes that, “If 2 Thessalonians is regarded as a kind of explanatory appendix to 1 Thessalonians, the comparative lack of Pauline theology in it receives a satisfactory explanation” (181). Independently, D. J. Clark (“Structural Similarities in 1 and 2 Thessalonians: Comparative Discourse Anatomy,” in *The Intertextuality of the Epistles: Explorations of Theory and Practice* [ed. T. L. Brodie, D. R. MacDonald, and S. E. Porter; NTM 16; Sheffield: Sheffield Phoenix Press, 2006], 196–207), based on discursive similarities, suggests that “It seems improbable that a pseudonymous writer would go to the lengths of structural analysis and imitation that the letters actually display, whereas the same writer could easily retain in his subconscious mind the main outline of a recently written earlier letter” (202). O. Leppä (“2 Thessalonians among the Pauline Letters: Tracing the Literary Links between 2 Thessalonians and Other Pauline Letters,” in Brodie, MacDonald, and Porter, *Intertextuality of the Epistles*, 176–95) concludes that “2 Thessalonians should be studied with the possibility in mind that the author might have been acquainted with the other undisputed Pauline letters as well as with 1 Thessalonians and 1 Corinthians” (194). Clearly there is a need for input from a quantified corpus-linguistic procedure that can comment on how reliable or appropriate the grounds for authorial dispute really are, based on texts and not on theories into which the texts are forced to fit. This must be complemented by another external procedure that begins with a corpus of texts attributed to one writer, such as Euripides, Aristides, or others, and identifies a set of reliable results showing both the flexibility and the unity of that corpus. This would serve as an external control on results obtained similarly by using Paul, enabling one to work outward to comment credibly on letters that are claimed by some to be attributed to his name but not written by him.

In “Building a Corpus for the Linguistic Analysis of Hellenistic Greek” (166–272) and “Tools and Concepts for the Corpus-Based Analysis of Hellenistic Greek” (273–313), chapters 4 and 5, O’Donnell illustrates the various types of annotation suitable to Greek texts for the purposes of quantitative linguistic study. The annotation must of course be devised to allow for the maximum electronic recognition of relevant linguistic features. Here a clear and detailed example of discourse annotation is shown with respect to a personal second-century letter, P. Mich. VIII 491, including word-group, clause, and paragraph level encoding (132, 169–201, 212–14, 253–72). This enables the readers to see exactly what is involved. The corpus of 512,301 words O’Donnell proposes as initially representative of a working corpus of Hellenistic Greek (164–65), within which P. Mich VIII 491 falls, is worth noting. It consists of the Greek New Testament, LXX Judges, 1 Maccabees, Didache, Shepherd of Hermes, Letters of Ignatius, Strabo’s geography,

Epictetus' dissertations, Polybius' history, Plutarch's *Cato*, Josephus's *Life*, Philo's *Moses*, Arrian's *Anabasis*, Diodorus Siculus's history, 2 Esdras, Cassius Dio's history, Apollodorus's library, various documentary papyri (30,000 words worth but not individually identified), and inscriptions.

O'Donnell offers four exemplary applications of his method and procedure. As to textual criticism, he shows how to annotate the textual variation in words attributed to Jesus in Mark's Gospel (274–93). This process allows manuscripts to be compared for similarities and differences and could shed light on getting to the best judgment on the weight of the variant reading (see also S. E. Porter and M. B. O'Donnell, "The Implications of Textual Variants for Authenticating the Activities of Jesus," in *Authenticating the Words of Jesus* [ed. B. D. Chilton and C. A. Evans; NTTTS 28.1; Leiden: Brill, 1998], 121–51).

With regard to source criticism, O'Donnell draws attention to the famous "dotplot" technique, a two-dimensional visual scheme of graphing discourse and displaying the coincidence of words (298–301). He dotplots Matt 5:3 versus Luke 6:20b and Matt 15:27b versus Mark 7:28b, suggesting that Luke could have improved on Matthew and that Matthew could have improved on Mark, but much more work of this sort would be necessary to confirm this. Dotplots of the passion predictions in the Synoptics (306) suggest the absence of a single source; that is, there was not one underlying passion prediction drawn upon by all three Evangelists. A dotplot of Jude versus 2 Peter (308) suggests the nonexistence of a documentary relationship between these texts. It would be useful here to have a similar plot, for the important purpose of contrast and external control, of Apollonius's *Argonautica* versus Homer, given that at certain points in the *Argonautica* characters demonstrate and use awareness of events recorded in Homer, implying the existence of a documentary relationship (see V. Knight, *The Renewal of Epic: Responses to Homer in the Argonautica of Apollonius* [MnS 152; Leiden: Brill, 1995]). Lastly, dotplotting the sixteen chapters of the Didache (310–12) suggests a source for chapter 3.

In addition to applications to textual and source criticism, the third area where computer-aided analysis may fruitfully contribute is in lexicography. O'Donnell illustrates this potential via the New Testament words for resurrection, ἐγείρω and ἀνίστημι (340–50), where their slight differences in function and meaning can be highlighted. Further discussion focuses on "Causality and the Greek Voice System" with respect to these two verbs and leads to "An Example: The Third Passion Prediction in the Synoptic Gospels" (376–85). In this example, comparison of the Evangelists within Mark 10:32–34; Matt 20:17–19; and Luke 18:31–34 utilizes identification of clausal structure with respect to the interesting usage of passive and middle forms of ἐγείρω and ἀνίστημι. O'Donnell attempts to expand upon the "atomistic tendencies found in most NT commentaries—

i.e., the practice of dealing with one word or group of words in isolation of its larger discourse context” (377). He seeks an answer, based on clausal analysis, for the reason Matthew changes Mark's intransitive future middle (ἀναστήσεται, “he will rise,” Mark 10:34e) to a future passive (ἐγερθήσεται, “he will be raised,” Matt 20:19b), while not specifying the external agent. O’Donnell suggests that Matthew reduces the explicit role of the participants (chief priests, scribes, and Gentiles [τὰ ἔθνη]) and that therefore Matthew’s chain of infinitives, to ridicule, to whip, and to crucify (Matt 20:19a), lacks clarity, adding that “It is not clear who is the agent and cause of these processes because of the infinitive forms” (381). Hence, Matthew changes ἀναστήσεται to ἐγερθήσεται and, thus, by contrast is “clearly indicating the external cause of the expected process of resurrection” (381).

However, Matthew’s infinitive chain is not all that unclear, making it difficult to follow this argument. Members of τὰ ἔθνη would seem easily to satisfy the subject of the verbal ideas in Matthew’s infinitives. Should this be the case, perhaps Matthew chose his future passive ἐγερθήσεται for another reason. If Matthew anticipated that his readers, from preaching and/or from past or eventual reading of 1 Corinthians or Romans, would fill in for themselves that God or the Holy Spirit raised Jesus from the dead (as at 1 Cor 15:12–17 with five uses of the perfect passive of ἐγείρω and/or Rom 1:4: ἐξ ἀναστάσεως νεκρῶν), Matthew could compose with such an active readership in mind, knowing that readers would supply the external agent of his ἐγερθήσεται for themselves. After all, as far as the assumption of active readers is concerned, Matthew drops the storyline concerning a major prophecy about his main character (3:11d), perhaps aware that his readers could fill this in for themselves. Further, contemporary progymnasmatic instruction in narrative composition advised students, in the interest of conciseness, and certainly not neglecting clarity, to leave out what seems to be already understood. Lastly, since Luke’s editing of Mark retains ἀναστήσεται (Luke 18:33b), O’Donnell nicely suggests (385) that in this case the string of six future passive verbs Luke has put together in the immediately preceding clauses could supply a passive contextual connotation informing Luke’s preservation of Mark’s original middle form.

The fourth application that O’Donnell presents, “Grammar and Discourse Analysis” (398–487), is his seventh and final chapter. Here we reach a *Zielpunkt* of the method being treated via the presentation of examples of corpus-based discourse studies in Jude (406–25, 486–87) and via a model for corpus-based discourse analysis of Philemon (444–85). Before reviewing these two efforts, I would draw attention to a valuable excursus sandwiched between them: “Cohesion, with Reference to Romans 1:18–28” (426–43). Whether or not one is interested in exploring textual annotation and computer-assisted methods of interpretation, this excursive gem could be beneficial both to exegetes, as found in a complete commentary format on this passage (cf. R. Jewett, *Romans*

[Hermeneia; Minneapolis; Fortress, 2007], 148–83), and to New Testament scholars in general as they seek to be considerate of reasonable hermeneutical techniques focusing on contiguous personification, plot, repetition, theme, coherence, and authorial integrity, as found in suggestions pertaining to discursive and narrative expression (see A. Berlin, “A Search for a New Biblical Hermeneutics: Preliminary Observations,” in *The Study of the Ancient Near East in the Twenty-First Century: The William Foxwell Albright Centennial Conference* [ed. J. S. Cooper and G. M. Schwartz; Winona Lake, Ind.: Eisenbrauns, 1996], 195–207). Here O’Donnell demonstrates how to lay out the text of Rom 1:18–28 with attention to contiguous linguistic elements and patterns, thereby affording interpreters a panoramic view of Paul’s thinking in the Greek language. This is highly recommended.

Now that D. Biber, S. Conrad, and R. Reppen have employed a grammatically annotated corpus to study features of scientific discourse (see their *Corpus Linguistics: Investigating Language Structure and Use* [Cambridge: Cambridge University Press, 1998]), O’Donnell is keen to apply a similar method to Jude and Philemon. Of course, discourse analysis by its very nature seeks to study textual units longer than a sentence, assuming coherent thought on the part of an author who is afforded a proper measure of competence to produce cohesive ties between linguistic elements. O’Donnell’s applications to Jude show that the writer reaches an informational peak at verse 17, highlights the emotions and attitudes attributed to recipients of the letter and localizes the plot, neatly identifying participant-reference chains inherent in the composition. As to Philemon, a critique of O’Donnell’s detailed discussion is beyond the present scope, but two points should be made. The difficult intent of verse 6 may be clarified by looking at word-group relations linked to that verse (468–70). Inspection of these clauses lowers the active role of Philemon and offers insight into other participants in the wider discourse. In addition, O’Donnell is able to reach a conclusion with assurance that has not been pressed home by some students of the letter, namely, that Paul’s “request to Philemon is clearly for the transfer of Onesimus to him” (484).

In conclusion, *Corpus Linguistics* might offer a challenging read to some, but, as an armchair introduction, it is well worth the effort because what O’Donnell is describing is able to effectively serve New Testament scholarship with an exciting and practical method that complements traditional exegesis and hermeneutics of unannotated New Testament texts. O’Donnell provides an excellent comprehensive bibliography as well as indices of authors and references.