

OpenText.org: the problems and prospects of working with ancient discourse

Matthew Brook O'Donnell, Stanley E. Porter and Jeffrey T. Reed
University of Surrey Roehampton and OpenText.org

1. Introduction

The vast majority of studies in corpus linguistics have focused upon contemporary usage of modern languages. However, although there have been a number of studies of the earlier periods of some of these languages, such as Old English and Old French, they have tended to adopt the methods developed for modern languages. In our theoretical paper (Porter and O'Donnell 2001), we have explored the methodological challenges and questions posed by the study of an epigraphic language, such as Hellenistic Greek.¹ In particular, we have found the need for closer attention to the criteria used in the compilation of a corpus, the integration of the levels of annotation applied to a corpus, and maintaining a focus upon both traditional referential access and narrative or sequential text-analysis.² As a result, our approach is more than simply computer-aided text analysis, but in some ways fulfils the goals of the originators of corpus-based linguistics. It does this by performing a full analysis of the language, utilizing a structured corpus, rather than analyzing just a small portion of a much larger corpus. The textual orientation of classical and New Testament scholarship is compatible with and, in fact, requires a micro-pattern analysis, and the reading and analysis of full texts. Rather than placing it outside the scope of corpus linguistics, we argue that this perspective offers new avenues for the discipline, ones that we are exploring in OpenText.org.

OpenText.org is a web-based initiative dedicated to creating resources for the linguistic analysis of Hellenistic Greek, and especially the Greek of the New Testament, in collaboration with interested scholars around the world. OpenText.org aims to make use of the insights and methods of corpus linguistics, specifically in terms of building a representative corpus of Hellenistic Greek, richly annotated according to a functional discourse model. This bottom-up discourse model relies upon the notion of levels of formal analysis. We parse all forms, beginning with morphology, but categorize from the word group up. We have found that the word group constitutes the smallest meaningful unit for discourse analysis. The increasingly higher levels are those of the clause, paragraph and discourse. One of the major principles of our annotation scheme is to mark features at the level of discourse at which they function. This has required the development of level-reflective and level-sensitive notational categories, as certain elements such as a conjunction may operate either at the word group, clause or paragraph level. These notational categories serve in a horizontal dimension both to specify the function of the individual element and to indicate its relationship to other elements within the structural unit. Each structural unit then constitutes a minimal unit at the next highest level of discourse. Each level of analysis builds upon the previous level, and thus analysis at a particular level can reach down to include features from lower levels. This vertical dimension of analysis is also segregated according to the components of register, field, tenor and mode. As a result, various features at a given level will serve different register meta-functions. This schematization provides a means for moving from the elements of text to the context of situation. The intended result is a complex calculus of features and functions that enables the analysis of the discourse. At the heart of this method, therefore, is annotation. More than that, we have found that the process of annotation itself constitutes a major part of the analytical process, raising questions as to the function of components of discourse within their respective units and the nature of text itself as these elements constitute the discourse (DeRose *et al.* 1990; Renear *et al.* 1996; Leech 1994).

This paper presents two examples utilizing this methodology, drawn from the Greek New Testament. The first demonstrates how the OpenText.org annotation model can facilitate a full discourse analysis of the letter to Philemon. We have consciously selected this text for a number of reasons. The letter to Philemon is a small, complete text that encapsulates an interesting moment in discourse. It allows for us to present a synchronic view of this text, while still facilitating analysis of individual textual components. The result is, we think, a presentation of most of the major features of discourse as they are contained and displayed in our discourse model. We unfortunately will not be able to present all of these dimensions here. Though the shortest Pauline letter, Philemon has

¹We are defining Hellenistic Greek as that Greek written by native and non-native Greek speakers throughout the Hellenistic and Roman worlds from approximately the fourth century B.C. to the fourth century A.D.

²The third element aligns corpus linguistics with discourse analysis, sharing a common concern for the analysis of real language usage, the observation of patterns of linguistic usage, and the (quantitative) filtering of large amounts of data. On referential and narrative methods of access to corpus data, see McCarty 1996.

received considerable discussion in scholarly circles; however, the major focus of this discussion has been the attempt to reconstruct the probable context of situation that gave rise to the discourse (see Pearson 1999 for an overview of major positions). Few of these attempts are predicated upon a thorough analysis of the text as discourse, at least as we are defining and using the concept here. We think that insight can be gained into the context of situation of this provocative and intriguing New Testament text through close analysis of the elements of discourse, particularly the questions of the relationships among the major participants (Paul and Philemon) and of the function that Onesimus plays in these relationships. We think that our discourse model pushes interpretation forward by pointing out textual relationships that previous interpreters have often overlooked. In that sense, questions of the tenor of discourse, among others, are specifically addressed in this example. The second example illustrates the use of semantic-domain annotation of the book of Revelation to explore structure, specifically through the identification of cohesive units. In this example, rather than examining a single discourse from a variety of perspectives, we are examining one particular component of register over a larger and potentially more complex text. The structure of the book of Revelation suggests a set of complex semantic relations often linked with shifts in text-type. Our analysis suggests that through a variety of textual means the author has structured discourse themes to progress the thematic content of the book. This thematic content is introduced through lexis. Analysis of the semantic-domain patterns over the length of the discourse shows how the thematic material is not only introduced and treated in individual units, but how it is used to cohere the individual paragraph units and the entire discourse. Whereas the instance from Philemon examines a number of levels in order to discuss the tenor of discourse, this examination treats broad patterns of semantic-domain structure to say something about field and mode of discourse.

2. Philemon and participant structure

As we mentioned above, our method of analysis begins with the word group, and here we present three word groups from the book of Philemon. We present these because each illustrates important elements, before moving to analysis of a single set of clauses that constitute part of a paragraph. Under the field component of register the semantic relationships between words in a word group are annotated. We have defined four forms of modification, and one type of conjunction. The four modifiers are: specifier (*sp*), including articles and prepositions; definer (*df*), adjectives and appositional words; qualifier (*ql*), genitive and dative modifiers; and preposition (*pr*), a prepositional phrase that modifies a substantive. The conjunction or connection (*cn*) relationship is used to join two words within the group. It is helpful to visualize these semantic relationships through a series of nested boxes. Each box represents a word with slots for each of the relationships below the word. The boxes for modifiers (and their associated modifiers) are drawn within the relative slot of the word they modify. Thus word groups can be represented through the application of a recursive process.

The first is word group 1 (see fig. 1) in the book of Philemon, part of what is traditionally called the epistolary salutation, ‘Paul, prisoner of Christ Jesus’. This word group consists of four words, with the head term being *Παῦλος* in the nominative, or what might be called the subject case, but this implies analysis beyond the word group. Syntagmatically, the three words follow in sequential order the word *Παῦλος* (syntagmatic order is indicated by the word identifier, e.g., w10), but there are two types of semantic relations indicated. The first is that of definer, in which *δέσμιος* (prisoner) defines who Paul is. This relationship often describes what is traditionally called an epexegetical relationship that restates the head term. The relationship of *Ἰησοῦ* (Jesus) and *Χριστοῦ* (Christ) is also one of definition. However, the relation between *Χριστοῦ* (Christ) and *δέσμιος* (prisoner) is one of qualification. This is not a relationship that defines but qualifies who the servant is—he belongs to Christ (modifiers in the genitive case in Greek, as this one is, often have qualifier relations). This is a fairly straightforward example of a word group. This first word group is annotated in XML in the following manner and visualized in figure 1.

```
<wg:group id="wg1" head="w1">
  <w id="w1">Παῦλος</w>
  <w id="w2" modify="w1" rel="define">δέσμιος</w>
  <w id="w3" modify="w2" rel="qualify">Χριστοῦ</w>
  <w id="w4" modify="w3" rel="define">Ἰησοῦ</w>
</wg:group>
```

w1 Παῦλος						
sp	df				ql	pr
w2 δέσμιος						
	sp	df	ql		pr	
w3 Χριστοῦ						
		sp	df	ql	pr	
w4 Ἰησοῦ						

Fig. 1. Word group 1 ‘Paul, prisoner of Christ Jesus’

The second word group is number 3 (see fig. 2; appendix fig. 8 for annotation), ‘Philemon, the beloved and fellow worker of us’. It is also part of the salutation of the book of Philemon, except that here the head term, *Φιλήμονι* (Philemon), is in the dative case, or what might be called a complement case, that is, a case in which a complement can occur.³ In word group 3 we have a slightly different kind of semantic relationship among the elements, one that is not linear as word group 1 is. Here, the head term is defined by two separate modifiers, both adjectives, *ἀγαπητῶ* (beloved) and *συνεργῶ* (fellow worker). The first of these is preceded by a specifier, the article (*τῶ*), and the second is followed by a qualifier, the pronoun *ἡμῶν* (our).

w9 Φιλήμονι							
sp	df				ql	pr	cn
w11 ἀγαπητῶ							
	sp	df	ql	pr			
w10 τῶ							
w12 καί							
w13 συνεργῶ							
	sp	df	ql	pr			
w14 ἡμῶν							

Fig. 2. Word group 3 ‘Philemon, the beloved and fellow worker of us’

These two modifiers are connected in their modifying function by a conjunction (*καί*), operating at the word-group level.⁴ Our word-group analysis is economical in so far as distribution of elements is concerned, so we only analyze the article as occurring as a specifier of the first and the pronoun as a qualifier of the second adjective. The annotation principle behind this is to indicate each modifying word as connected to the one other word in the group to which it is most closely attached in terms of grammatical function.

So far, from the structure of the word groups, one might well think that these groups are somewhat similarly structured, and probably perform similar functions. Word group 54, the third example to consider (see fig. 3; appendix fig. 9 for annotation), consists of a significantly larger number of words than in either word group 1 or 3, ‘Onesimus, the then useless to you, but now useful to you and to me’. In some ways, the structure here is similar to word group 3. There is a head term, *Ὀνήσιμον* (Onesimus), in the accusative case, another of what might be called the complement cases (again extending analysis beyond the word group). There are also two definers, both adjectives. The first is *ἄχρηστον* (useless) and the second *εὐχρηστον* (useful). Each of these, however, has a number of further modifiers. *ἄχρηστον* has a specifier, the article *τόν*, and two qualifiers, *ποτέ* (then) and *σοι* (you), the first an adverb and the second a pronoun. The second definer, *εὐχρηστον*, has three qualifiers, the adverb *νυνί* (now), and the pronouns *σοί* (you) and *ἐμοί* (me). There are a

³ However, at the level of the word group this is not significant, since here we are analyzing relations of the words within the word group, not its relations outside of it.

⁴ The word *συνεργός* can be used as a substantive, in which case the phrase could be analyzed as consisting of two separate word groups, ‘Philemon the beloved’ and ‘fellow worker of us’, joined at the clause level by the conjunction *καί*. The head term of the first word group, *Φιλήμονι*, has a single definer, *ἀγαπητῶ*, and the second, *συνεργῶ*, a qualifier *ἡμῶν*. The annotation scheme can allow for alternative analyses where necessary.

number of instances of parallelism here worth noting. First is the structural parallelism with the two definers; the second is the temporal alternation between past and present; the third is the use of the common root of the adjectives; and the fourth is the repetition and expansion of scope indicated by the pronouns. Some of this parallelism emerges more clearly because of the use of the word group conjunctions. The first, *δέ* (but), joins the two defining phrases, while the second, *καί* (and), joins two qualifiers. The parallelism, as well as the sheer bulk of the construction, seems to give significant weight, at least at the level of the word group, to this particular construction.

w145 Ὀνήσιμον						
<i>sp</i>	<i>df</i>			<i>ql</i>	<i>pr</i>	<i>cn</i>
	w149 ἄχρηστον					
	<i>sp</i>	<i>df</i>	<i>ql</i>	<i>pr</i>		
	w146 τόν		w147 ποτέ			
			w148 σοι			
	w156 εὐχρηστον					w151 δέ
	<i>sp</i>	<i>df</i>	<i>ql</i>	<i>pr</i>	<i>cn</i>	
			w150 νυνί			
			w153 σοι			
				w154 καί		
			w155 ἐμοί			

Fig. 3. Word group 54 ‘Onesimus, the then useless to you, but now useful to you and to me’

These three word groups are important for a number of reasons. The first is that these are the word groups in which the major participants of the discourse of the book of Philemon are introduced, Paul, Philemon and (according to most scholars) Onesimus, a point we will come back to below. There are a number of other participants also mentioned in Philemon (e.g. Christ Jesus, who is, theology aside, in the position of a qualifier of a definer in relation to the head term, Paul; Timothy; Apphia; Archippus; and ‘the church in your house’; etc.; see fig. 6), but, as we would see if we analyzed more of the discourse, these are not given prominence as the others are. Each of the major participants is introduced in a word group in which there is significant modification, which is not found in the word groups for the other participants (see word groups 2, 4 and 5, where the modification is noticeably less). For each major participant, there is a grammaticalized reference by name, and then appropriate modification to indicate their role and relation. Subsequent participant reference in the discourse is made by a combination of grammaticalized, reduced and implied reference. For word groups 3 and 54, there is also internal modification that references Paul (and in word group 54 Philemon as well), thus further indicating participant relations at the word-group level. On the basis of this evidence, it appears that Onesimus is at least as important, at least at the level of the word group, as any of the other major participants, and certainly more important than the minor ones. That is consistent with traditional examination of Onesimus (word group 54, see fig. 3) within this Pauline letter by the vast majority of New Testament scholars. Though our treatment of annotation at the word-group level has focused upon the interpretative and discourse significance of individual word groups, it should be clear how this detail of annotation applied across a larger corpus of texts could be utilized using both the traditional referential retrieval paradigm (i.e. searching for particular words, participants or grammatical features in a certain modification position) and more discourse oriented narrative approaches.⁵

The next level of analysis is that of the clause. Within the book of Philemon, there are 47 clausal units, arranged within 5 paragraphs. Paragraph one consists of two clauses. The first has six word groups, two of which are presented above. They form the two major groups that constitute the structure of clause one. In paragraph three, however, there are 16 clauses. We wish here to analyze

⁵ The annotation of semantic relationships (specifier, definer, qualifier and preposition) within a word group is just one element of the field of discourse. Other features marked at the word-group level include semantic domains (field), part of speech and lexical information (mode) and participant reference (tenor). For detailed specification of the OpenText.org annotation model for each of the levels of discourse and the associated XML DTDs, see <http://www.opentext.org/specifications>.

clauses 15-24, since it is within this complex of clauses that the word group that introduces Onesimus into the discourse is found. Clausal structure, in our analysis, consists of subject, predicate, complement and adjunct. The predicate is the major unit of the Greek clause, but whether the subject is grammaticalized, reduced or implied in relation to that subject, and its placement in first or later position in the clause are very important for the information structure. We analyze this information structure in terms of prime and subsequent elements in the clause. Once the clauses have been analyzed, we examine the relations of the clauses at the paragraph level. A primary clause usually has a finite verb form or, in the absence of a finite verb, a word group functioning similarly. These clauses are used to convey the main thread or backbone of the discourse. Secondary clauses are usually relative, participle and infinitive clauses. These clauses are used to add further specification at the present point in the discourse, and do not progress the discourse in the horizontal plane. The paragraph level, under the mode of discourse, is the appropriate point in the model to analyze the connection and relationship between clauses. These connections determine the level of a clause. For example, a clause with a finite verb form that might be classified as primary, but is clearly connected to a secondary clause (e.g. a relative clause), is classified as functioning at the secondary level. The appendix contains the clause and paragraph level analysis of Philemon 10-14. Figure 4 presents a visual representation of this annotation (see appendix, fig. 10 for XML annotation of clauses). Clauses are represented by boxes. The single primary clause (c15), *παρακαλῶ σε περὶ τοῦ ἐμοῦ τέκνου*, is placed on the left-hand side of the diagram. All secondary level clauses connected to this clause are joined to the right-hand of this main clause. Other clauses not connected directly to c15 but to another of the secondary clauses are placed to the right of the clause to which they connect, unless they are embedded within the clause (e.g. clause c23 *ποιῆσαι* is a complement of clause c22) where they are positioned within the box of the clause in which they are embedded.

c15 παρακαλῶ σε περὶ τοῦ ἐμοῦ τέκνου	c16 ὃν ἐγέννησα ἐν τοῖς δεσμοῖς Ὀνήσιμον τὸν ποτέ σοι ἄχρηστον νυνὶ δὲ σοὶ καὶ ἐμοὶ εὐχρηστον	
	c17 ὃν ἀνέπεμψά σοι αὐτόν	c18 τοῦτο ἔστιν τὰ ἐμὰ σπλάγχνα
	c19 ὃν ἐγὼ ἐβουλόμην c20 πρὸς ἐμαυτὸν κατέχειν	c21 ἵνα ὑπὲρ σοῦ μοι διακονῇ ἐν τοῖς δεσμοῖς τοῦ εὐαγγελίου
	c22 χωρὶς δὲ τῆς σῆς γνώμης οὐδὲν ἠθέλησα c23 ποιῆσαι	c24 ἵνα μὴ ὡς κατὰ ἀνάγκην τὸ ἀγαθόν σου ἦ ἀλλὰ κατὰ ἐκούσιον

Fig. 5. Display of clause level and connections for Philemon 10-14

What is important for this paper is the analysis of word group 54 within such a paragraph analysis. Clause 15 is the primary clause of the unit that consists of clauses 15-24, with a series of four secondary clauses related to it, three of them linked by relative pronouns (clauses 16, 17 and 19). Clause 15, 'I urge you concerning my child', refers to the three participants that we mentioned before: Paul, this time implied through the use of the first-person singular verb; Philemon, referred to by use of a reduced pronoun; and a figure cited in a prepositional phrase (adjunct) as 'my child', not yet named. The author then defines this child in the three relative secondary clauses. The first one has complement-predicate-adjunct-complement structure, with a split complement because of the relative pronoun (*ὃν*). Again, Paul is the implied subject of the clause, but it is here that Onesimus is introduced as the complement of this secondary clause. A number of further statements are made about this 'child', Onesimus, though without using his name, in the further secondary relative clauses all with Paul as the implied subject (clauses 17 and 19) and the further secondary clauses (clauses 18, 20, 21).

In other words, even though Onesimus has often been analyzed as a major participant in the book of Philemon, with much secondary discussion concerned with who he is and how he relates to Paul and Philemon, the discourse structure does not confirm this analysis. Paul especially, but also Philemon, are the major participants, and their relationship is in fact the major element of the tenor component. This is confirmed by how many times Paul and Philemon, whether in grammaticalized,

reduced or implied form, are the subjects or complements of the primary clauses, but also secondary clauses.

Participant	Primary clause	Secondary clause	Total
1 Paul	26	14	40
2 Jesus Christ	9	3	12
3 Timothy	2	0	2
4 Philemon	21	13	34
5 Us	4	0	4
6 Apphia	2	0	2
7 Archippus	2	0	2
8 You (plural)	4	0	4
9 God	3	0	3
10 The Saints	1	1	2
11 Onesimus	4	9	13

Fig.6 Summary of participant reference according to clause level in Philemon

Figure 6 shows a summary of the number of references to each of the participants annotated in the discourse. Onesimus is confined for the most part to peripheral status as a complement of secondary clauses, often in reduced or implied form (9 of the 13 references to Onesimus occur in secondary clauses and 8 of these references occur in the complement slot of the clause). He is not even introduced by name the first time he is referred to, as are the other major participants, but is referred to in a reduced form by means of a noun, 'child', that fulfils a discourse function of characterizing Onesimus in relation to the major participants. The implications of this analysis for interpretation of the discourse are significant. The focus of future analysis will need to be upon the major participants as supported by the discourse itself, Paul and Philemon, recognizing that, as important as Onesimus may be in terms of a catalytic function in their interpersonal relations, his function is secondary to the primary relation. Further, instead of the status of Onesimus as a slave or a runaway, and whether or not he took property that was not his, the major issue seems instead to be how it is that Paul as apostle relates to Philemon as significant figure in the church to which Paul writes. References to Paul in the letter according to clause component are: subject 8x, predicate 16x, complement 12x and adjunct 6x. For Philemon, these figures are: subject 1x, predicate 9x, complement 17x and adjunct 7x. These basic figures require filtering according to the causality (voice of verbal forms, indicating actor or patient status) and position within a word group (see above). Some have noted that Paul uses a number of discourse techniques in his communication with the church there, although wanting to hold back from characterizing these techniques as at all manipulative or forcefully persuasive (Wilson 1992; Fitzmyer 2000; cf. Porter 1999). However, this notion seems to be much closer to what the discourse supports rather than examining this letter as an exposition of slave and master relations in the ancient world, in which Paul has an incidental interest.

3. Revelation, semantic-domains and cohesion

The book of Revelation is much larger than Philemon. Philemon has a total of about 335 words compared to the approximately 9850 in Revelation. This extra length provides enough discourse scope for development of a number of different patterns. The way several of these patterns relate to each other is what we would like to examine in this second example.

As discussed in the introduction, a common problem for both discourse analysis and traditional referential corpus linguistics is the analysis and summarization of large amounts of data in order to carry out a full analysis of even a single feature of discourse, such as lexical cohesion, over even a medium sized discourse, such as the book of Revelation. Considerable amounts of information must be tracked. The standard reference for the study of cohesion in discourse is the work of Halliday and Hasan (1976). This has served as the basis for a number of recent discourse annotation schemes (Wilson and Thomas 1997) and theoretical studies (Hoey 1991). In addition, a number of more computationally orientated studies (Morris and Hirst 1991; Kozima and Furugori 1994) have developed algorithmic approaches to the study of lexical cohesion. One of the ways the study of Hellenistic Greek, particularly the Greek of the New Testament, is more advanced than comparable study of English is in the availability of a semantic-domain lexicon (Louw and Nida 1988; Nida and

Louw 1993). This lexicon distinguishes 93 broad semantic domains, covering all of the semantic fields of the Greek found in the New Testament. Within these 93 domains are a large number of sub-domains. Words are categorized within these domains and sub-domains, with some words being found within as many as four or five domains, except for such function words as prepositions, which are categorized in as many as six domains. It is not appropriate to discuss the shortcomings of this important work, although one of the results of our study is ideally to construct a more principled set of domains that are more closely linked to patterns of New Testament usage. However, in the meantime, this tool has proved invaluable in the study of lexis and cohesion in the New Testament.

For our study of lexical cohesion in the book of Revelation, we focus upon content words only (verbs, nouns, adjectives, and adverbs). Each of these word types in the text was annotated with its major domain number from the lexicon, with no attempt to disambiguate multiple classifications. We then made use of a simple algorithmic technique to identify the number of semantic chains found within a fifty word window measured at intervals of ten words throughout the text. The first window begins at word 1 and extends to word 50, the second window at word 11 extending to word 60, etc. At each interval, the number of domains exhibiting chains of five or more words in the subsequent window are noted. For example, the first window (words 1-50) has two chains, one from domain 33, communication, and one from domain 93, names of persons and places. The following window (words 11-60) has only a single chain from domain 33. In a number of ways this fits the pattern that one might expect at the beginning of a book. We have found that domain 33 is probably the most common domain in New Testament books, especially the epistles, and indicates what one might expect in a theological text, that is, the desire to communicate is activated, and often this chain is maintained through much of the discourse. It is also worth noting, though not necessarily of great significance, that the beginning introduces at least some of the people or places involved in the action. This is a relatively low-level use of vocabulary. Figure 7 contains a line-plot with the position in the discourse in terms of word number on the x-axis and the number of significant semantic chains (domains with five or more words) at each point indicated on the y-axis. This graph, through the alternation of peaks and troughs of semantic chain usage, provides a helpful macro-view of lexically cohesive units within the book of Revelation (and perhaps insights into thematization), and serves as a tool to highlight sections for closer micro-level analysis. The genre of the book of Revelation is a complex one, since it has a mix of text-types. The peaks and troughs of the semantic chains are coordinated with the major structural divisions of the text-types. For example, within a new text-type, such as the letter section, a recurring pattern is the use of a semantic chain that peaks and then recedes. This seems to be a consistent pattern, in which the new text-type activates a set of related terms. The text-type continues, even though the semantic chains may not continue, before a new text-type activates a new set of domains.

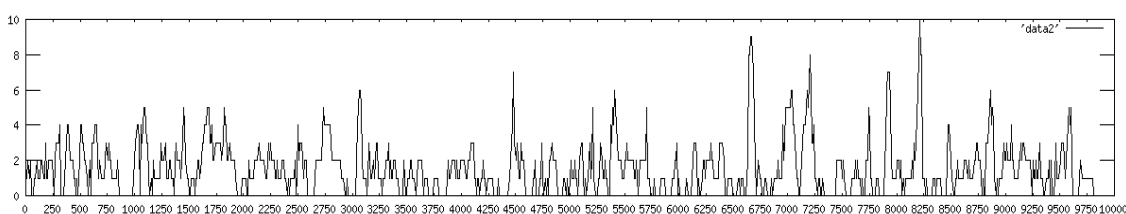


Fig. 7 Lexical cohesion at 10 word intervals across the book of Revelation

Two micro-level analyses serve to illustrate the kinds of observations that emerge from this close study. The first is from Revelation 4 and the other Revelation 16. The window beginning at word 1611 (corresponding to Rev. 3.22, the last verse of chapter 3) seems to mark a significant shift in the discourse. The preceding sample windows have had only one or two semantic chains. One of these, domain 33 (communication), is relatively low-level in significance, but the other chains anticipate the next structural unit, by introducing words from semantic domain 12 (supernatural beings and powers) and domain 37 (control, rule). This confirms a similar finding concerning the way in which grammatical features such as tense-form and voice anticipate and mark transitions in discourse (Biber, Conrad, and Reppen 1998; Porter 1994). Beginning with word 1611, we find a long chain of domain 12 that extends over 200 words. A number of additional but shorter chains are also activated at this point at the opening of the unit. These include a continuation of semantic domain 37, and the introduction of domains 85 (existence in space), 6 (artefacts), 60 (number), and 14 (physical events and states). Shorter chains dispersed throughout this paragraph (extending from words 1611 to 1931) include semantic domains 2 (natural substances), 4 (animals), 67 (time), 57 (possess, transfer, exchange), 41 (behavior and related states). These shorter chains are usually between two and four

windows long, a maximum of 80 words. This is a semantic-chain based description of the unit. The scene itself in the book of Revelation marks a transition from the epistolary sections (the so-called letters to the seven churches) to the apocalyptic vision that constitutes much of the rest of the book. Most commentators would see these as the two major sections of the text. The first scene of this apocalyptic section is the dramatic appearance of the heavenly throne room. In this room, there is a majestic throne surrounded by other thrones with heavenly beings seated upon them. These beings have crowns and jewels and golden lamps. These two different descriptions are complimentary. The latter is an aesthetically based account of the various elements depicted. The semantic-chain description is an attempt to quantify and even explain the aesthetic account by classifying the data that make up the account.

The second section is in many ways similar to the first. Here we are concerned with the unit from word 6621 to word 6751, which is preceded by a 200 word section with virtually no significant semantic chains. In contrast to the section discussed above, there is no preparatory transitional material that anticipates this unit. The semantic chains in this section come from semantic domains 15 (linear movement), 13 (be, become, exist, happen), 14 (physical events), 41 (behavior and related states), 57 (possess, transfer, exchange), 78 (degree), 79 (features of objects), 85 (existence in space), and 91 (discourse markers). The semantic domains in the Louw-Nida lexicon are arranged along a continuum, with contiguous domains having overlapping semantic features. Thus a given word may be classified within several contiguous domains. That is the case in this particular episode, as demonstrated by the domains activated. The result is that the significance of these chains may be less than these data first appear, since a number of the chains activated may contain the same word. In other words, there are probably 4 or 5 major semantic areas activated here, rather than the 7 or 9 noted above. Also in contrast with the example above, this intense confluence of semantic chains here is relatively short. Attention to the text reveals that the section marks the climax of a unit beginning around word 6300 (Rev. 16.1), which begins: 'I heard a loud voice from the temple saying to the seven angels, "Go pour out the seven bowls of wrath upon the earth"'. Then follows the account of each of the seven angels distributing the contents of their bowls over the earth. This section describes the action and result of the seventh angel, completed with a voice from the throne saying that 'It is done' (Rev. 16.17). This is followed by dramatic environmental disruption, including lightening, thunder, and earthquakes. Commentators have noted the climactic nature of this event in the discourse, which can be confirmed by the pronounced accumulation of semantic chains.

We introduced this part of our paper by talking about how such semantic chain patterning applied to textual cohesion. Cohesion is a concept that seems to be useful on at least two levels. One is in terms of the macro-patterns of usage that unite an entire discourse. We have seen that there are a number of semantic chains that are activated throughout the book of Revelation, such as domain 33 (communication). There are a number of domains that function at this low level to activate a number of basic concepts that unite the discourse. Cohesion also seems to function in terms of the relationship between various units or paragraphs in the discourse. Part of the cohesive function of semantic chains is to activate accumulations of chains at appropriate times. These serve the function of delineating the units of the discourse by closing and opening units, and marking transitions. The boundary here between cohesion and prominence therefore is not a firm one, since those semantic chains that are marked as prominent because of their relatively infrequent but high-level activation only become prominent when they are seen in relation to the semantic chains that cohere the discourse.

4. Conclusion

In this paper, we have tried to show some of the practical differences that working with an epigraphic language such as Hellenistic Greek makes for corpus linguistics. The kind of micro-analysis that we have briefly introduced in the above two examples seems to be required if one is to utilize the limited corpus size from the ancient world. Micro-analysis allows for close scrutiny of a finite set of elements, but each is seen to function within a variety of levels of discourse. The definition and classification of these levels enables each element to provide maximal data for interpretation. Such micro-analysis can only take place, however, if the corpus of texts is richly annotated to provide the largest amount of information as possible. In the first example, that of Onesimus and his relation to Paul and Philemon, we saw that at one level—that of the word group—Onesimus seems to be grammaticalized in similar fashion to the other two major participants. However, when the word group in which he is introduced is placed in the larger frame of being a component of paragraph structure, his prominence fades, and he is seen to be relegated to a peripheral participant role. In the examples of Revelation, we have examined a more narrowly circumscribed set of features in terms of the entire discourse. Here we have noted that by using an intensive semantic-domain study, and correlating this with paragraph boundaries, one can observe

how semantic chain shifts are coordinated from paragraph to paragraph. The result is a clearer demarcation of how the major subject matter of each paragraph unit is shifted and developed, but also how cohesion is both created within the individual paragraph units and extended over the entire discourse. One of the facts of corpus-based studies as usually conceived is that they can generate huge quantities of data for analysis, and this is thought to be a desirable feature, allowing more precise generalizations to be reached on the basis of a larger sample surveyed. One of the complaints about discourse analysis is that it generates too much data to study within the confines of a single analysis. These two are not necessarily reconcilable, especially when a non-finite corpus is involved. Even for an epigraphic language such as Hellenistic Greek, an abundance of data can be generated for analysis. In the light of the finite corpus size, however, these data are to be desired, and to be maximized for their use in corpus-based discourse studies.

References

- DeRose S J, Durand D G, Mylonas E, Renear A H 1990 What is text, really? *Journal of Computing in Higher Education* 1(2): 3-26.
- Fitzmyer J A 2000 *The letter to Philemon*. New York, Doubleday.
- Halliday M A K, Hasan R 1976 *Cohesion in English*. London, Longman.
- Hoey M 1991 *Patterns of lexis in text*. Oxford, Oxford University Press.
- Kozima H, Furugori T 1994 Segmenting narrative text into coherent scenes. *Literary and Linguistic Computing* 9(1): 13-19.
- Leech G 1994 Corpus annotation schemes. *Literary and Linguistic Computing* 8(4): 275-281.
- Louw J P, Nida E A 1988 *Greek-English lexicon of the New Testament based on semantic domains*. 2 vols.; New York, United Bible Societies.
- McCarty W L 1996 Peering through the skylight: towards an electronic edition of Ovid's *Metamorphoses*. In Hockey S, Ide N (eds), *Research in humanities computing 4: selected papers from the ALLC/ACH conference, Christ Church, Oxford, April 1992*. Oxford, Clarendon Press, pp 240-262.
- Morris J, Hirst G 1991 Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics* 17: 21-48.
- Nida E A, Louw J P 1993 *Lexical semantics of the Greek New Testament*. Atlanta, Scholars Press.
- O'Donnell M B 1999 The use of annotated corpora for New Testament discourse analysis: a survey of current practice and future prospects. In Porter S E, Reed J T (eds), *Discourse analysis and the New Testament: results and applications*. Sheffield, Sheffield Academic Press, pp 71-116.
- Pearson B W R 1999 Assumptions in the criticism and translation of Philemon. In Porter S E, Hess R S (eds), *Translating the Bible: problems and prospects*. Sheffield, Sheffield Academic Press, pp 253-280.
- Porter S E 1994 *Idioms of the Greek New Testament*. Sheffield, Sheffield Academic Press.
- Porter S E 1999 Is Critical Discourse Analysis Critical? An Evaluation Using Philemon as a Test Case. In Porter S E, Reed J T (eds), *Discourse analysis and the New Testament: results and applications*. Sheffield, Sheffield Academic Press, pp 47-70.
- Porter S E, O'Donnell M B 2001 Theoretical issues for corpus linguistics raised by the study of ancient languages. In *Proceedings of Corpus Linguistics 2001*.
- Renear, A H, Mylonas E, Durand D 1996 Refining our notion of what text really is: the problem of overlapping hierarchies. In Hockey S, Ide N (eds), *Research in humanities computing 4: selected papers from the ALLC/ACH conference, Christ Church, Oxford, April 1992*. Oxford, Clarendon Press, pp 263-280.
- Wilson A 1992 The pragmatics of politeness and Pauline epistolography: a case study of the letter of Philemon. *JSNT* 48: 107-19.
- Wilson A, Thomas J 1997 Semantic annotation. In Garside R, Leech G, McEnery A (eds), *Corpus annotation: linguistic information from computer text corpora*. London, Longman, pp 53-65.

Appendix: annotation examples

```
<wg:group id="wg3" head="w9">
  <w id="w9">Φιλῆμον</w>
  <w id="w10" modify="w11" rel="specify">τῶ</w>
  <w id="w11" modify="w9" rel="define">ἀγαπητῶ</w>
  <w id="w12" join="w11" to="w13" rel="connect">καὶ</w>
  <w id="w13" modify="w9" rel="define">συνεργῶ</w>
```

```

    <w id="w14" modify="w13" rel="qualify">ἡμῶν</w>
</wg:group>

```

Fig. 8. Annotation of word group 3 (Philemon 1b)

```

<wg:group id="wg54" head="w145">
  <w id="w145">Ὁνήσιμον</w>
  <w id="w146" modify="w149" rel="specify">τόν</w>
  <w id="w147" modify="w149" rel="qualify">ποτέ</w>
  <w id="w148" modify="w149" rel="qualify">σοι</w>
  <w id="w149" modify="w145" rel="define">ἄχρηστον</w>
  <w id="w150" modify="w156" rel="qualify">νυνί</w>
  <w id="w151" join="w149" to="w156" rel="connect">δέ</w>
  <w id="w153" modify="w156" rel="qualify">σοί</w>
  <w id="w154" join="w153" to="w155" rel="connect">καί</w>
  <w id="w155" modify="w156" rel="qualify">έμοί</w>
  <w id="w156" modify="w145" rel="define">εὐχρηστον</w>
</wg:group>

```

Fig. 9. Annotation of word group 54 (Philemon 10b-11)

```

<cl:clause id="c15" level="primary">
  παρακαλῶ σε περὶ τοῦ έμοῦ τέκνου
</cl:clause>

<cl:clause id="c16" level="secondary" connect="c15">
  ὃν έγέννησα έν τοῖς δεσμοῖς Ὁνήσιμον τόν ποτέ σοι ἄχρηστον νυνί δέ σοί και
  έμοί εὐχρηστον
</cl:clause>

<cl:clause id="c17" level="secondary" connect="c15">
  ὃν άνέπεμψά σοι αυτόν
</cl:clause>

<cl:clause id="c18" level="secondary" connect="c17">
  τουτο ἔστιν τὰ έμά σπλάγχνα
</cl:clause>

<cl:clause id="c19" level="secondary" connect="c15">
  ὃν έγώ έβουλόμην
  <cl:clause id="c20" level="secondary" connect="c19">
    πρὸς έμαντόν κατέχειν
  </cl:clause>
</cl:clause>

<cl:clause id="c21" level="secondary" connect="c19">
  ἵνα ὑπέρ σου μοι διακονῆ έν τοῖς δεσμοῖς του εὐαγγελίου
</cl:clause>

<cl:clause id="c22" level="secondary" connect="c15">
  χαρίς δέ της σῆς γνώμης οὐδέν ήθέλησα
  <cl:clause id="c23" level="secondary" connect="c22">
    ποιῆσαι
  </cl:clause>
</cl:clause>

<cl:clause id="c24" level="secondary" connect="c22">
  ἵνα μή ώς κατά ανάγκην τὸ άγαθόν σου ἦ αλλά κατά έκούσιον
</cl:clause>

```

Fig. 10. Annotation of clause level and connection (Philemon 10-14)