Introduction

The study of the Babatha archive,¹ a cache of documents found in the Cave of Letters near the Dead Sea, raises a number of questions.² These include: the linguistic nature of the Greek found in these documents of a Nabatean woman fleeing Rome in the early second century, the extent of linguistic interference in these Greek documents, and the relation of the papyri from Palestine to those of Egypt. This study addresses several important questions raised by examination of the Babatha archive. However, before such large-scale comparisons can be made as the questions above suggest, one must have thoroughly examined the texts involved. By subjecting this small, discrete corpus of twenty-three Greek papyri to analysis, we hope to gain further linguistic insight into this particular set of documents in the Greek language, especially as it was used in the eastern Mediterranean of the first several centuries CE.

The analysis of these Greek texts is part of a larger project, known as the OpenText.org project (http://www.opentext.org), which is attempting to create heavily annotated Greek texts representing the range of literary text-types of the Greco-Roman world.³ For the first paper on adopting this approach, forty-five papyrus letters, with a total of 3,341 words (average 74 words per papyrus document), were annotated (referred to in this paper as the structured corpus), and constituted the basis for the publication that this one follows.⁴ The twenty-three Greek papyri of the Babatha archive add 4,344 words to this expanding corpus (average 189 words per papyrus document), and provide the basis of the data of this study.⁵

¹ N. Lewis (ed.), The Documents from the Bar Kokhba Period in the Cave of Letters: Greek Papyri (Jerusalem 1989).
² For a study that places this archive in context, see S.E. Porter, "The Greek Papyri of the Judaean Desert and the World of the Roman East," in S.E. Porter and C.A. Evans (eds.), The Scrolls and the Scriptures: Qumran Fifty Years After (Sheffield 1997) 293–316, esp. 311–314.
⁵ There are more letters in the Babatha archive, but several are too fragmentary for annotation and analysis.
Papyrological Studies, Corpus Linguistics, and Register Theory

This paper utilizes principles from corpus linguistics to study the Greek of the Babatha archive. Papyrologists are long familiar with the term "corpus," associating it with groups of documents that come from either a particular place (e.g. Oxyrhynchus) or a particular archive (e.g. the Zenon papyri). Corpus linguistics is related to such corpora in that it is concerned with developing representative databases of naturally occurring language. One of the earliest such corpora was developed by the University of Michigan linguist Charles Fries, who compiled a corpus of around 250,000 words from transcribed telephone conversations, in order to study how native English speakers actually used language in conversations related to day to day activities. Fries made a number of important discoveries that went contrary to expectations on the basis of his analysis of language in use.

Recent work in Papyrology has increasingly recognized the need to move beyond analysis of individual texts to collections of texts gathered for a variety of purposes. However, there are limitations to the study of ancient languages, such as the Greek of the Hellenistic and Roman periods, from the standpoint of the usual linguistic methods, which are dependent upon the intuitions and responses of native informants. This is the place where the use of a robust linguistic method, in conjunction with a corpus-based study, can be of benefit to the linguistics of ancient languages. The lack of native informants, "rather than causing despair should make more pressing the need to reevaluate constantly the interpretative models employed and to rely more heavily upon formal linguistic features of the extant corpus." Furthermore, corpus based principles of data collection can overcome limitations in the available data, so as to create corpora that facilitate maximal comparison and contrast of the available data.

Corpus linguistics, as a field of study, is itself a constantly developing method of analysis, rather than a specific linguistic theory. As a result, corpus linguists distinguish between an archive – an ad hoc assemblage of data, such as the documents discovered at Oxyrhynchus – and a corpus, in which there are governing principles of structure and representativeness. This distinction is crucial for the study of ancient Greek, where we can never hope to have a corpus that includes massive samples that attempt to capture the entire language (a monitor corpus). Instead, the OpenText.org project is attempting to gather a sample corpus based upon a structured selection of texts established through external and internal criteria. The use of a structured and representative corpus of texts allows for better comparative quantification of results that positively move beyond the kinds of concordance-based results typical of archival searching.

---

6 This interdisciplinary approach builds upon linguistic research first presented at the "Buried Linguistic Treasure: The Potential of Papyri for Greek and Latin Language Research Colloquium," held at Christ Church, Oxford in 2006, by applying its corpus based approach to the Babatha archive. The published form of this paper has appeared as: S.E. Porter and M.B. O'Donnell, "Building and Examining Linguistic Phenomena in a Corpus of Representative Papyri," in T.V. Evans and D.D. Obbink (eds.), The Language of the Papyri (Oxford 2010) 287–311. I draw directly upon this paper in the material that follows, including using wording, categories of discussion, and results.

7 See O'Donnell, op.cit. (above, n. 3) 1. This is the best introduction to the subject of corpus studies for the Greek of the ancient world.


The inclusion of a papyrus archive to the corpus adds significantly to the sample range of the corpus, but also adds problems of its own, in that the texts are not structured, are not truly a representative sample, and are limited in a variety of other ways (not least in terms of the fragmentary nature of many of the texts, which either limits the data or results in questionable reconstructions).

The OpenText.org project has pioneered the development of heavily annotated computer searchable Greek texts that are annotated according to various levels of linguistic structure. Level-based annotation increases the level of precision of annotation, and hence has the benefit of encouraging a variety of more precisely defined searches and, ideally, more widely usable data and potentially helpful conclusions. The annotation in the OpenText.org project is (to date) made at the levels of morphology, word, word group, clause, and clause type.

For the purpose of this study, I am using a register-based analysis that Matthew Brook O’Donnell and I have been pioneering in the field of ancient-text studies. We have noted before that in early papyrological studies there was optimism that the study of these ancient documents would provide great insights into the broader world in which they were created. George Milligan, who was deeply involved in lexical studies using the (then) recently discovered papyri, compiled a volume of Greek papyri and noted their significance for those interested in the ancient world, especially in the areas of language, the letter form and the wider social and religious environment. This optimism has been squelched, however, due to the overwhelming amounts of material that have been uncovered since, and the sense of urgency to get these documents published before turning to larger and broader questions. Some have raised doubts whether the kinds of insights being sought are even achievable from the kinds of materials available, that is, private documents that reflect set forms of expression. Such a perspective is far too pessimistic, especially in the light of the computer-based resources that are increasingly available for the study of the ancient world. Initial ventures into the study of the papyri indicate that there are substantive and measurable linguistic differences across a corpus of texts that justify their study both as individual texts and, more importantly, as a collection of texts. Further, register analysis has been developed as a rigorous and robust method of linguistic analysis that is able to analyze texts of any size, as it is able to use finite data for maximal results. Finally, register analysis is a multivalent model that is able to differentiate and then re-integrate various linguistic components, so as to be able to analyze a variety of grammatical, literary and social factors by means of a single heuristic framework.

Register analysis, a form of functional linguistics, addresses transient linguistic varieties, that is, varieties of language according to use in a given context of situation. Any communicative act, including producing documentary papyri such as letters or legal documents, performs along two major functional axes: linguistic behavior and sociolinguistic context. Register analysis works from the standpoint that the use of language varies on the basis of the authorial situation and that such language use indicates the compositional situation, with each one informing the other. Register as a functional notion establishes the constraints on language usage, rather than prescribing specific lexico-grammatical instantiations. The three

---

11 G. Milligan, Selections from the Greek Papyri (Cambridge 1910) xxvii–xxxii.
registers of discourse are field, tenor and mode. These three registers have correlate semantic metafunctions in terms of the ideational, interpersonal and textual dimensions of language function. This indicates the means by which actual instances of language usage relate to the semantic component.

Register Analysis of the Babatha Archive

In this section, I will treat each of the three registers separately. As part of the annotation of a structured corpus of papyri, an effort was made to categorize the letters on the basis of a socially based status classification. John White identifies essentially three levels of social interaction between the author and recipient of a given letter. Letters may be addressed between those of equal social status, to a person of higher social status, or to a person of lower social status. On the basis of this grid, the Babatha archive, with its twenty-three letters, contains eleven letters that are addressed to a person of a higher social status, six letters addressed to a person of a lower social status, and six letters addressed between those of equal status. Although it might prove interesting to retain these three social status categories in the analysis of the register of the letters in the Babatha archive, it was thought that there are too few letters in each category to provide a basis for the results, and so such analysis has not been performed.

Before presenting and analyzing the results of the study of these documents, I note three significant problems that affect analysis of the Babatha archive, and potentially any similar collection of documents. (1) There are many texts that are fragmentary at least in part. Many of these documents, while they have survived largely intact, are not without damage. As a result, there are many places where lettering is no longer visible. Rather than working to reconstruct such texts, I have attempted to decipher readings where there is some text visible, but have not attempted where there is none, so the results do not include any calculations based upon reconstructions. (2) There are a number of double documents in the Babatha archive. In fact, fourteen of the twenty-three documents are double documents. These double documents provide an interesting set of analytic circumstances. On the one hand, there is duplication of data, in the sense that the doubled documents often provide two instances of the same linguistic phenomenon, but the two uses do not provide further new information. However, analysis of the double documents also shows a number of instances where the second document differed from the first, often in terms of specific grammatical phenomena. This justifies the inclusion of the double documents as they are. (3) This corpus is one that is rightly called an archive, as we did not choose what to include, but history and the ravages of time spent in a cave made the decision. However, these documents are intended to form part of a growing corpus of papyri in the OpenText.org structured corpus, and so their particularities will be seen within this larger representative corpus of documents.

Mode of Discourse and Textual Semantic Component

The mode of discourse activates the textual metafunction. The structuring factors of a discourse include a variety of elements. In terms of the Babatha archive, I wish to consider conjunctions, and focus in terms of thematized elements and ordering of clausal elements.

---

Conjunctions, though often treated as if they all function on the same linguistic level, actually function at a variety of linguistic levels. Linguistic annotation of levels of discourse allows for tagging of the papyri so as to distinguish use of conjunctions at the word group, clausal or paragraph levels. A richly annotated database such as this allows for study of conjunctions in terms of frequency of occurrence and level of function. In this instance, as we did for the structured corpus of papyri, I wish to examine the use of clausal level conjunctions. In the Babatha archive, there are 209 instances of clausal level conjunctions. The most frequently used conjunctions appear in the following order of frequency: καὶ (55x), δὲ (40x), εἰ (21x), ἕαυ (14x), καθός (8x), ὅτι (6x), and ὡς (5x). Asyndeton is used at the clausal level 32x. There are a number of other conjunctions that appear 1–4 times each: ἵνα, καθάπερ, διό, δέν, ὡστε, οὕτως, ὁπόταν, ὁποῦ, ἐπεὶδὴ, and ἐπεί. Two widely used conjunctions are not used in the Babatha archive: οὖν and γάρ. The structured corpus of papyri has the following most common clausal level conjunctions: καὶ (99x), δὲ (39x), ἵνα (23x), οὖν (20x), γάρ (17x), ὡστε (13x), ἕαυ (12x), ὅτι (11x), εἰ (10x), ἐπεί (7x), and ὡς (6x).

There are some interesting observations to make on the basis of comparison of the Babatha archive with usage of conjunctions in the Greek of the New Testament and the structured corpus.

(1) Relative frequency of conjunctions. In comparison with the Greek of the New Testament, the distribution of clausal conjunctions in the Babatha archive is roughly proportionate. In the Babatha archive, the conjunctions καὶ and δὲ are the most frequent (26% and 19% respectively), with καὶ more frequent. In the structured corpus, the conjunction καὶ appears in 38% of the instances, with δὲ in 15% of the instances (of the most frequent conjunctions). There are no instances of οὖν and γάρ in the Babatha archive, while οὖν and γάρ are the fourth and fifth most frequent in the structured corpus. In the Babatha archive, the conjunctions εἰ and ἕαυ are the third and fourth most frequent, probably because of the legal content of several of the documents, while in the structured papyri corpus they are the seventh and ninth most frequent.

(2) Implications of the use of the conjunctions. There are several implications of the results of these two studies regarding conjunctions. The first is that, in the structured papyri corpus, which reflects the use of Greek in personal letters from Egypt, there is a much higher use of the conjunction καὶ than there is in the documents written for the Nabataean woman living east of Palestine. As was stated in our earlier paper on this topic, despite the contentions of some scholars, paratactic conjunctions were more frequent in Greek of the Hellenistic and Roman periods, especially in documentary texts, than some scholars want to recognize.

(3) With regard to any supposed Semitic nature of the Babatha archive, this Semitic influence cannot be established on the basis of the use of conjunctions, as the type and distribution of conjunctions appears to be less susceptible to Semitic analysis than in the structured corpus.

---


Greek discourse focuses select material on the basis of the ordering of elements in their respective linguistic units, with the element in the first slot receiving such focus. Each discourse level has its own structure for marking elements. At the clause level, this focal placement is called thematization, and relates to the ordering of thematic and rhematic material. The clausal elements consist of Subject (S), Predicator (P), Complement (C), and (for optional modifying elements) Adjunct (A). When one of these elements is placed first (discounting conjunctions), this element is thematized.

I wish to analyze two patterns of ordering of clausal elements in the Babatha archive, in relation to the Greek of the New Testament and especially in relation to the structured corpus.

(1) Frequency of thematized elements. The most frequent pattern of thematized order in the Greek New Testament (from greater to lesser frequency) is: P > A > S > C (Predicator is more frequent than Adjunct than Subject than Complement). In the structured papyrus corpus, on the basis of 398 instances, the order is: P > C > A > S. In the Babatha archive, on the basis of 565 instances, the frequency of thematized elements is: P > A > C > S. The instances and percentages are as follows:

<table>
<thead>
<tr>
<th>Order of thematized clausal elements</th>
<th>Structured corpus</th>
<th>Babatha archive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P (174x, 43%)</td>
<td>P (565x, 34%)</td>
</tr>
<tr>
<td></td>
<td>C (90x, 25%)</td>
<td>A (555x, 33%)</td>
</tr>
<tr>
<td></td>
<td>A (83x, 21%)</td>
<td>C (387x, 23%)</td>
</tr>
<tr>
<td></td>
<td>S (54x, 13%)</td>
<td>S (164x, 9%)</td>
</tr>
</tbody>
</table>

Several observations are worth making regarding this. The P and A elements are virtually identical in frequency, so it is difficult to know whether PACS or APCS is most frequent in the Babatha archive. If the A is removed (as it is an optional element) the P > C > S ordering is seen to be the fundamental frequency ordering in both the Babatha archive and the structured corpus. In the structured corpus, the A appears more frequently than the C, although the fundamental frequency (without the A) is the same. In the Babatha archive, it appears that the adjunctive elements (A) are more explicitly discussed than things (C) or agents (S). The subject, once grammaticalized or otherwise established, becomes less thematically important than the other elements of the clause. Adjuncts are frequently used in the introduction to the text in the Babatha archive as part of the legal introduction to the document. This certainly helps to account for the high frequency of A elements.

(2) Ordering of clausal elements. When the individual clausal elements are isolated, and their frequency of specific ordering patterns considered, the structured corpus of papyri and the Babatha archive are very similar in results. The only noticeable difference is that the Babatha archive tends to have more extreme results than the structured corpus.

<table>
<thead>
<tr>
<th>Patterns of clausal element ordering</th>
<th>P &gt; C</th>
<th>C &gt; P</th>
<th>P &gt; S</th>
<th>S &gt; P</th>
<th>S + P</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured corpus</td>
<td>176 (57%)</td>
<td>100 (43%)</td>
<td>32 (47%)</td>
<td>36 (53%)</td>
<td>68 (16%)</td>
<td>370 (84%)</td>
</tr>
<tr>
<td>Babatha archive</td>
<td>192 (77%)</td>
<td>83 (23%)</td>
<td>66 (44%)</td>
<td>80 (56%)</td>
<td>148 (25%)</td>
<td>421 (75%)</td>
</tr>
</tbody>
</table>

There are several observations to be made on the basis of these data. In terms of the relation of P and C elements, for both the structured corpus and the Babatha archive P > C ordering occurs more
frequently than C > P ordering. The Babatha archive, however, has a greater frequency of the P > C construction than the structured corpus. In terms of the P and S elements, both corpora have a greater frequency of S > P ordering than P > S ordering, with the frequency very similar. The Babatha archive more frequently has both the S and P elements in a clause than does the structured corpus. Thus, as expected, the structured corpus has a larger number and frequency of clauses with the P without an explicit S, although for both corpora the P appears alone in roughly 75–85% of all clauses.

**Tenor of Discourse and Interpersonal Semantic Component**

The tenor of discourse activates the interpersonal metafunction. The tenor of discourse is concerned with participant structure, that is, who is involved in the discourse and their roles and relations, defined both intra-textually and extra-textually. In this section I identify and discuss participant interaction and reality. This differentiation of person roles and function is determined on the basis of participation in terms of person and number grammaticalization, and attitude in terms of mood forms.

1. Participation (Person and Number), indicated by the use of the person and number system.

<table>
<thead>
<tr>
<th></th>
<th>1st Singular</th>
<th>1st Plural</th>
<th>2nd Singular</th>
<th>2nd Plural</th>
<th>3rd Singular</th>
<th>3rd Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured corpus (578x)</td>
<td>153 (26%)</td>
<td>61 (12%)</td>
<td>227 (39%)</td>
<td>34 (6%)</td>
<td>89 (17%)</td>
<td>14 (2%)</td>
</tr>
<tr>
<td>Babatha archive (272x)</td>
<td>62 (23%)</td>
<td>15 (6%)</td>
<td>29 (11%)</td>
<td>2 (1%)</td>
<td>145 (53%)</td>
<td>19 (7%)</td>
</tr>
</tbody>
</table>

In the structured corpus of papyri, the second person singular is the most frequent, followed by the first person singular, and then the third person singular and first person plural, with the second person plural and then third person plural used only a relatively few times. By contrast, the Babatha archive uses the third person singular most frequently, followed distantly by the first person singular, second person singular, third person plural and first person plural, with the second person plural only used twice (1%). In both corpora, the singular person is used more frequently than the plural. The first person singular is used with equal frequency in each corpus (26%), but the structured corpus uses the second person (especially singular) more frequently, while the Babatha archive uses the third person (especially singular) more frequently.

2. Attitude (Mood Forms)

<table>
<thead>
<tr>
<th></th>
<th>Indicative</th>
<th>Imperative</th>
<th>Subjunctive</th>
<th>Optative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured corpus (368x)</td>
<td>200 (56%)</td>
<td>99 (26%)</td>
<td>65 (17%)</td>
<td>4 (1%)</td>
</tr>
<tr>
<td>Babatha archive (272x)</td>
<td>208 (76%)</td>
<td>13 (5%)</td>
<td>51 (19%)</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

For both the structured corpus and the Babatha archive, the indicative appears most frequently, though the Babatha archive has a greater percentage of instances of the indicative. The optative is used only 4x (1%) in the structured corpus, and not at all in the Babatha archive. The ordering of frequency of the mood forms is:
Structured corpus: Indicative > Imperative > Subjunctive (> Optative)
Babatha archive: Indicative > Subjunctive > Imperative

The structured corpus has a more directive attitudinal semantics than does the Babatha archive (26% Imperatives vs. 5%). Both have a roughly equal frequency of projective attitude (Subjunctive: 17% and 19% respectively).

Some observations can be made regarding participation and attitude.
(1) Singular number. The use of the singular number is significant. In the structured corpus, it appears that the personal letters are geared around a "you (singular) and I" axis of interchange. In the Babatha archive, in the light of the legal and litigious nature of these documents, it is not surprising to find a "he/she (third person singular) and I" axis of interchange, as Babatha defends herself against her disputants.

(2) Participation and attitude. Regarding the correlation of participation and attitude, the second person singular is used most frequently in the structured corpus, along with the imperative of the non-indicative mood forms. By contrast, in the Babatha archive, the indicative mood form is used most frequently by a significant amount (even higher percentage than in the structured corpus), which is at least consistent with the "he/she and I" axis of interchange.

(3) Social status. In our work on the structured corpus, we made several observations regarding correlation of social status to participation and attitude. The Babatha archive introduces a new, significant element in its frequent use of the third person. This may be typical of legal documents, in which the "facts" of the situation are narrated. This would also probably correlate with the very high frequency of indicative mood forms used.

Field of Discourse and Ideational Semantic Component

The field of discourse activates the ideational metafunction. The field of discourse is concerned with the subject matter and purpose of the communication, and may involve either intra-linguistic or extra-linguistic items. Both the transitivity network and the lexicon through semantic domain organization are important for establishing the field of discourse. In this section, two factors of the field of discourse are included for analysis: clausal relations and aspect and causality.

Above it was noted that clauses consist of four major components. Here we examine the clauses and their relationship to each other. Clauses in Greek may be categorized in complexes, in which there are primary, secondary and embedded clauses. Primary clauses carry the main line of argument, while secondary and embedded clauses are developmental off-line clauses. Embedded clauses are formed around infinitives and participles, and are, as the title implies, embedded within either primary or secondary clauses. Secondary clauses are clauses that are conjunctively connected to other clauses.

---

17 I recognize that there are other means by which "directives" or "commands" may be formulated in Greek, but this analysis is formally based, as a means of establishing quantifiable data as the basis of semantics, before moving to pragmatics.
There are a number of differences to note between the structured corpus of personal letters and the Babatha archive.

(1) Clauses per text. One is that there is a larger number of clauses per text in the Babatha archive. We already noted above that the average size of a document in the Babatha archive is significantly larger than the letters in the structured corpus.

(2) Distribution of clauses. More significant than the number of clauses per text is the distribution of clauses. The structured corpus is predominantly structured around primary clauses, with smaller numbers of embedded and secondary clauses: Primary (66%) > Embedded (18%) > Secondary (15%). The Babatha archive is structured significantly differently, with embedded clauses being the most numerous, followed by primary and secondary clauses in nearly equal numbers: Embedded (45%) > Primary (28%) > Secondary (27%). In the structured corpus, the social status did not affect the distribution of clause types. This indicates that the structured corpus and the Babatha archive are structured differently on the basis of a need to structure the ideas differently. Many of the legal background statements of the Babatha archive, often placed at the beginning of the text, are placed in embedded clauses (including genitive absolute constructions).

Verbal aspect describes the author’s perspective on depiction of the verbal process, and causality indicates the means by which actions are performed. These are important to the transitivity network and indicate the contours of the action. The following patterns are observable.

(1) Aspect (tense-form) is a morphologically based semantic category that structures action as perfective (aorist tense-form), imperfective (present and imperfect tense-forms) or stative (perfect and pluperfect tense-forms).

<table>
<thead>
<tr>
<th></th>
<th>Structured corpus (543x)</th>
<th>Babatha archive (524x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfective</td>
<td>220 (40%)</td>
<td>166 (32%)</td>
</tr>
<tr>
<td>Imperfective</td>
<td>226 (42%)</td>
<td>275 (55%)</td>
</tr>
<tr>
<td>Stative</td>
<td>97 (18%)</td>
<td>83 (13%)</td>
</tr>
</tbody>
</table>

Both of the corpora of papyri have the imperfective aspect as predominant, although in the structured corpus this is only marginally greater. Both have the following pattern of usage: Imperfective > Perfective > Stative. The pattern in the Babatha archive is closer to what one would expect in the expositional epistolary text type, with the imperfective aspect as predominant. Neither corpus uses the stative aspect frequently, which aspect is reserved for semantic frontgrounding.

(2) Causality (voice form) is a morphologically based semantic category that structures causation in terms of active, middle and passive causal forms. In the previous study of the structured corpus, we did not differentiate middle/passive forms for the present/imperfect and perfect/pluperfect voice forms, as I have done here.
<table>
<thead>
<tr>
<th></th>
<th>Active</th>
<th>Middle</th>
<th>Mid./Pass.</th>
<th>Passive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured corpus (575x)</td>
<td>401 (68%)</td>
<td>21 (4%)</td>
<td>134 (24%)</td>
<td>19 (4%)</td>
</tr>
<tr>
<td>Babatha archive (540x)</td>
<td>302 (60%)</td>
<td>76 (13%)</td>
<td></td>
<td>162 (27%)</td>
</tr>
</tbody>
</table>

There is no way to prejudge the accuracy of such a hypothesis, but it appears that, if the Babatha archive is representative in its distribution of usage, approximately two-thirds of the middle and passive forms are passives. In this study, the active voice is predominant in both corpora, in which the subject of the verb is the causal agent. The Babatha archive uses the voice forms in frequency as one might expect: Active > Passive > Middle.

**Conclusion**

There is much more that can and should be done with the data gathered from this study. The results are very preliminary and not large, but they provide further papyrological data to be added to that already gathered from the structured corpus of letters. The linguistically-based findings from examination of the Babatha archive, especially when compared with the findings from the structured corpus of papyri, are that the Greek of the Babatha archive, rather than being out of harmony with other Greek of the time, especially as found in other papyri, maintains similar parameters of usage. Register analysis provides a linguistically sensitive tool for the differentiation of various linguistic indicators found within the Greek language, so as to sharpen analysis. The use of register analysis enables the identification of various linguistic features, and quantification of their usage within a corpus. In this instance, the Babatha archive, in relation to the structured corpus of letters already studied, as well as the Greek of the New Testament, forms the basis of this examination. Register analysis provides a tool for the data of usage to be correlated with semantic metafunctions, so as to begin to go beyond the accumulation of raw data and to begin to understand the meaning and significance of these data. Comparison of the Babatha archive with the Greek papyri from Egypt as represented in the structured corpus shows that the major register indicators are relatively constant, although there are some differences that are perhaps explainable on the basis of the content and purpose of the respective corpora. These similarities draw further lines of connection between Egypt and Palestine closer, especially in terms of linguistic evidence.