



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

A Quantitative Analysis of the Aramaic Qumran Texts

John M Starr

PhD

University of Edinburgh

2013

Declaration

I declare

- (a) that the thesis has been composed by me, and
- (b) that the work is entirely my own, and
- (c) that the work has not been submitted for any other degree or professional qualification.

Signed

Abstract

Ḥirbet-Qumran lies about 15km south of Jericho. Between 1947 and 1956 eleven caves were discovered that contained thousands of fragments, mainly of prepared animal skin, representing approximately 900 texts many of which are considered copies. Over 100 of these texts are in Aramaic, though many are short fragments. The provenance of these texts is uncertain, but the fact that copies of some of them are known from beyond Qumran indicates that not all the Aramaic texts found at Qumran are likely to have originated there, though Qumran may have been a site where they were copied. Hitherto, this Aramaic corpus has been referred to as a single entity linguistically, but increasingly it is recognized that heterogeneity of textual features is present. Such heterogeneity may provide clues as to the origins of different texts. The current classification of Qumran texts is in terms of the cave they were found in and the order in which they were found: it does not provide any information about textual relationships between texts. In this thesis I first review the literature to identify suitable quantitative criteria for classification of Aramaic Qumran texts. Second, I determine appropriate statistical methods to classify edited Aramaic Qumran texts according to quantitative textual criteria. Third, I establish ‘proof of principle’ by classifying Hebrew bible books. Fourth, I classify the Aramaic Qumran texts as a corpus without reference to external reference texts. Fifth, I investigate the alignment of this internally-derived classification with external reference texts. Sixth, I perform confirmatory analyses to determine the number of different text-type groups within the corpus. Seventh, I use this classification to allocate previously unclassified texts to one of these text-types. Finally, I examine the

textual characteristics of these different text-types and discuss what this can tell scholars about individual texts and the linguistic development of Aramaic during Second Temple Judaism as reflected by Qumran.

Table of Contents

Acknowledgements.....vi
LIST OF TABLES.....x
LIST OF ILLUSTRATIONS.....xx

Chapter

1. Classification of Qumran texts.....1
Present state of classification of Qumran texts.....1
Qumran texts and Scripture: a basis for classification?.....6
Can the Qumran texts be considered as a corpus?.....14
Present state of classification of the Qumran Aramaic corpus...19
2. Statistical approaches relevant to Qumran Aramaic texts.....22
A background to statistical methods.....24
Emanuel Tov and the conventional text-critical approach.....29
Andrew Morton, quantitative approaches to New Testament...33
texts and implications for Qumran

	<i>Frank Moore Cross and the quantification of variant readings</i>	36
	<i>Qumran texts considered as a corpus</i>	39
	<i>Yehuda Radday and statistical analysis of selected Hebrew</i>	42
	<i>bible texts</i>	
	<i>Summary</i>	49
3.	Textual classification criteria.....	50
	<i>Stylistic criteria</i>	51
	<i>Corpus linguistic approaches</i>	79
	<i>Criteria for comparison with non-Qumran texts</i>	90
4.	Aramaic textual criteria.....	102
	<i>Lexical criteria</i>	107
	<i>Morphological criteria</i>	110
	<i>Syntactic criteria</i>	128
	<i>Summary</i>	130
5.	Materials examined.....	132
	<i>Accordance computerised texts</i>	132
	<i>Bibleworks computerised texts</i>	146
	<i>Comprehensive Aramaic Lexicon computerised texts</i>	146

6.	Statistical methods.....	159
	<i>Power calculations – general background.....</i>	160
	<i>Effect sizes in the Hebrew bible.....</i>	166
	<i>An exploratory Principal Components Analysis of the Hebrew bible.....</i>	177
	<i>Vocabulary richness and translated texts.....</i>	186
	<i>Summary.....</i>	190
7.	Description of Aramaic Scrolls.....	191
	<i>Hierarchical cluster classification using raw rates of textual criteria.....</i>	201
	<i>Hierarchical cluster classification using standardised rates of textual criteria.....</i>	207
	<i>Comparison of raw and standardised classifications.....</i>	213
	<i>Characterisation of typical text groups.....</i>	218
	<i>Realigning the Qumran Aramaic texts.....</i>	222
	<i>Internally situating other major Qumran Aramaic texts.....</i>	225
8.	External textual correlates.....	230
	<i>Targum Onkelos.....</i>	231
	<i>Targum Jonathan.....</i>	240
	<i>Targum Neofiti.....</i>	249
	<i>Targum Pseudo-Jonathan.....</i>	256
	<i>Qumran Biblical and Targumic Aramaic.....</i>	264
	<i>Aramaic documents from Ancient Egypt.....</i>	267

	<i>Aramaic documents from Ancient Egypt and their relationship to Qumran Aramaic scrolls</i>	277
	<i>Non-Qumran Judean Desert Aramaic texts</i>	291
	<i>Non-Qumran Judean Desert Aramaic texts and their relationship to Qumran Aramaic scrolls</i>	300
	<i>Babatha Archive Aramaic texts</i>	309
	<i>Summary of exploratory analyses of internal and external Aramaic text types</i>	318
9.	Description of syntactic features of Aramaic scrolls.....	337
	<i>Correlations between syntactic features</i>	358
	<i>Hierarchical clustering by syntactic features of the Qumran Aramaic scrolls</i>	361
	<i>Analysis of variance by syntactic features between Qumran Aramaic scrolls clusters</i>	365
	<i>Relationships between syntactical variable pairs in syntactical cluster groups of the Qumran Aramaic Scrolls</i>	374
	<i>Comparison of syntactic features in the Qumran Aramaic scrolls with the Targums and Ancient Egyptian documents</i>	382
	<i>Comparison of syntactic features in the Qumran Aramaic scrolls with the Judean Desert corpus and Babatha archive</i>	386
	<i>Summary of findings from syntactic data</i>	392
10.	Confirmatory models of Qumran Text types.....	395
	<i>Latent Class Analysis with known classes</i>	396
	<i>Confirmatory LCA for Enochic – Aramaic Levi text types</i>	398
	<i>Confirmatory LCA for Tobit text types</i>	399

	<i>Summary</i>	400
11.	Extending the classification of Qumran texts.....	402
	<i>Latent class analysis with training</i>	403
	<i>Characterisation of the extended text types</i>	409
	<i>A textual classification</i>	427
	<i>Non-aligned scrolls</i>	428
	<i>External validation of the morphological factors</i>	434
	<i>Lexical characteristics of different text types</i>	437
	<i>Summary</i>	443
12.	Discussion.....	444
	<i>Limitations to interpretation</i>	444
	<i>Implications of the classification</i>	449
	<i>Implications for our understanding of Second Temple Judaism</i>	460
	<i>Qumran and its Aramaic texts</i>	466
	<i>Future work</i>	468
	<i>Conclusions</i>	469
	 Bibliography.....	 471
	 Appendix	
	1. Glossary of statistical terms.....	A-1
	2. Supplementary analyses of the Hebrew Bible.....	A-8

Acknowledgements

Being a part-time student, this thesis has had a lengthy gestation. As such, it has developed its own form along the way. I had originally thought to classify all the Qumran scrolls, but it soon became clear that the vast number of Hebrew scrolls would have resulted in a thesis far too long. That massive undertaking had to be put aside, although the work presented in this thesis hopefully signals the way ahead. Indeed, I originally came to study the Qumran material quite by accident; I had originally intended to classify early Christian literature. It was the encouragement of my principal supervisor, Professor Timothy Lim, that set me on this task and, strangely, I have gradually come to see his wisdom, since much early Christian literature, even though it is written in Greek, occurs against the backdrop of oral traditions in Aramaic and the kaleidoscope of theological understanding that comprises Second Temple Judaism. Professor Lim has gently encouraged me throughout my studies and I would like to express my strongest gratitude to him for both his patience and inspiration. I would also like to thank Professor April McMahon, my second supervisor, for all her help and for sticking with me as a student despite her move to Aberystwyth to take up the arduous task of being Vice-Chancellor.

Of course, there are many people whose influence I would like to acknowledge because, to a greater or lesser extent, they have helped shape my thinking and thus what has emerged from my studies of the Aramaic scrolls at Qumran. There are, I think, two strands that stand out. First, those who have inculcated into me attention to scientific

method, this thesis owes not a little to this way of thinking, and here I would especially like to thank Professor Lawrence Whalley who took me on as an untried research fellow and displayed the erudition that comes from a mind open to a wide field of scholarship. Second, I am grateful to my lecturers and supervisors on the History of Art course I took in my third year as an undergraduate. They helped to open my eyes to a different world and bring me in contact with writers whose insights have stayed with me since then. Pertinent to this thesis is the work of Michel Foucault. At the start of his 'Order of Things', he explains how the book arose from a passage he came across in a Chinese encyclopaedia in which it is written, 'animals are divided into: (a) belonging to the Emperor, (b) embalmed, (c) tame, (d) sucking pigs, (e) sirens, (f) fabulous, (g) stray dogs, (h) included in the present classification, (i) frenzied, (j) innumerable, (k) drawn with a very fine camelhair brush, (l) *et cetera*, (m) having just broken the water pitcher, (n) that from a long way off look like flies'. Foucault was astounded by the way of thinking that could devise such a taxonomy and it starts him off on a quest to understand how the way we classify things has changed over the last few centuries. This thesis betrays my way of thinking and I am aware that it reflects a scientific positivism that privileges data. I hope that this does not reflect an uncritical attitude, and indeed there is much discussion about what types of data should be considered that requires an *a priori* theoretical framework: even Sir Isaac Newton, one of the greatest of scientists, recognized the important theoretical work of predecessors by quoting Bernard of Chartres' dictum of 'standing on the shoulders of giants'. Hence, the weaknesses in this thesis are totally attributable to my choices as to which giants I have chosen to stand on,

not the giants themselves, whilst I would not have been able to produce this classification without a massive amount of work by highly erudite scholars in the field.

About halfway through ‘The Order of Things’ in Chapter 5, Foucault discusses the ‘System’ and the ‘Method’ as understood in the Classical age. He explains that the ‘Method’ consists of starting from a fixed object, a text for example, and then relating everything else to this fixed object in terms of all the differences. This is common in biblical textual criticism where there is a ‘typical’ text and other texts are classified in terms of their variants from that text. Opposed to this was the ‘System’ where only certain characteristics of a set of objects were considered and a taxonomy developed in terms of these characteristics. With the ‘System’, the choice of different characteristics would produce a different taxonomy. Although less familiar to biblical scholars perhaps, the ‘System’ is frequently used in scientific endeavours. Although each approach reflects a different mindset, as it were, Foucault asserts that they ‘are simply two ways of defining identities by means of the general grid of differences’. I hope that this thesis can, at least in a small way, reflect this greater unity of purpose in the task of classification.

Finally, and most importantly, I would like to thank my wife, Claire, and my children, John, Robert, Toby and Gabriel, for all their patience and understanding with me as I have spent many, many hours working on what must have seemed to them an extremely obscure and hardly relevant topic to today’s world. Unfortunately for my children, Aramaic, or even Qumran, do not feature in today’s school curricula, so they have not

even had the advantage of expert help with homework or school projects. Perhaps that has been just as well.

List of Tables

For each Table the number before the decimal point indicates the Chapter it appears in and the number after the decimal point indicates its sequential position in the Chapter. Note that some Tables are large and run over more than one page due to formatting regulations. These Tables are marked with an asterisk (*) and can be conveniently viewed at www.qumran-aramaic.co.uk.

- | | |
|------------|--|
| Table 3.1 | Mean number of variants compared with the BHL with 95% confidence intervals for each of the pesharim taken from Lim (1997). |
| Table 6.1 | Number of words, Hebrew and Aramaic, per book as calculated by Bibleworks 6. |
| Table 6.2 | Power (%) to detect true frequency differences in texts of different word counts compared to a 1,000 text with a frequency of 16 words at $\alpha=.05$. |
| Table 7.1* | Parts of speech count and total words for Qumran Aramaic scrolls considered in <i>a priori</i> hypotheses testing. |
| Table 7.2 | Number of Qumran Aramaic scrolls considered in <i>a priori</i> hypotheses testing containing zero, 1-5 or more than 5 occurrences of each criterion. |
| Table 7.3 | Mean, median and range of rates per 1,000 words for the 22 criteria occurring in at least 10% of the 74 Qumran Aramaic scrolls identified as of adequate length. |

Table 7.4	Medians of rates per 1,000 words for the 22 criteria occurring in the different typical text type groups.
Table 8.1	Parts of speech count and total words for the Targum Onkelos by book.
Table 8.2	Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Onkelos.
Table 8.3	Parts of speech count and total words for the Targum Jonathan by book.
Table 8.4	Mean, median and range of rates per 1,000 words for the 22 criteria occurring in Targum Jonathan.
Table 8.5	Parts of speech count and total words for the Targum Neofiti by book.
Table 8.6	Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Neofiti .
Table 8.7	Parts of speech count and total words for the Targum Pseudo-Jonathan by book.
Table 8.8	Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Pseudo-Jonathan.
Table 8.9	Ratios (presence/absence) of Biblical group specific features in the Targums.
Table 8.10*	Parts of speech count and total words for the Egyptian Aramaic papyri.
Table 8.11	Mean, median and range of rates per 1,000 words for the 22 criteria occurring in the Aramaic Documents from Ancient Egypt.
Table 8.12	Rates per 1,000 words for the 22 criteria occurring in the Qumran ` Aramaic scrolls most closely associated with the typical Ancient

Egyptian documents together with mean rates for the typical Ancient Egyptian documents.

- Table 8.13* Absolute counts of parts of speech and total words for the non-Qumran Judean Desert Texts.
- Table 8.14 Rates per 1,000 words for parts of speech for non-Qumran Judean Desert texts of adequate length for statistical analysis.
- Table 8.15 Means and 95% confidence intervals of rates per 1,000 words for parts of speech by site in the non-Qumran Judean Desert texts.
- Table 8.16 Rates per 1,000 words for the 22 morphological criteria occurring in the non-Qumran Judean Desert texts by site.
- Table 8.17 Mean (95% confidence intervals of the mean) rates per 1,000 words for the 15 of the 22 *a priori* morphological criteria occurring in the non-Qumran Judean Desert texts comparing Wadi ed-Daliyeh to the three other sites grouped together as south Judean Desert texts. P-values <0.05 indicate significant differences between text types. 95% confidence intervals are given for non-zero means with zero fixed as the lowest value for the lower confidence interval.
- Table 8.18 Absolute counts of parts of speech and total words for the CAL Babatha Archive Texts with more than 50 words. (N.B. Total words also include number of determined nouns).
- Table 8.19 Rates for parts of speech per 1,000 words for the CAL Babatha Archive Texts with more than 50 words. (N.B. Total words also include number of determined nouns).

- Table 8.20 Rates per 1,000 words for the 22 morphological criteria occurring in the Babatha archive Aramaic texts.
- Table 8.21 Scheffe's test identifying homogeneous sets for Aphel rates in the various text types. Sample sizes and mean rates per 1,000 words for each text type are displayed.
- Table 8.22 Scheffe's test identifying homogeneous sets for Haphel rates in the various text types.
- Table 8.23 Scheffe's test identifying homogeneous sets for relative particle rates in the various text types.
- Table 8.24 Scheffe's test identifying homogeneous sets for *dy* rates in the various text types.
- Table 8.25 Scheffe's test identifying homogeneous sets for *yt* rates in the various text types.
- Table 8.26 Scheffe's test identifying homogeneous sets for determined state nouns rates in the various text types.
- Table 8.27 Scheffe's test identifying homogeneous sets for *dnh* or *dn'* rates in the various text types.
- Table 8.28 Scheffe's test identifying homogeneous sets for *'lh* rates in the various text types.
- Table 8.29 Scheffe's test identifying homogeneous sets for hitpeel rates in the various text types.

- Table 8.30 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of aphels within total aphels and haphels plus one.
- Table 8.31 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of haphels within total aphels and haphels plus one.
- Table 8.32 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion *dy* to relative particles plus one.
- Table 8.33 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of determined state nouns within total nouns.
- Table 9.1 Counts of infinitives in absolute and construct states in the Qumran Aramaic scrolls for the common stem forms.
- Table 9.2 Frequencies of the 15 of the 85 lemmas that occur at least five times as a peal infinitive construct in the non-biblical Qumran Aramaic scrolls.
- Table 9.3 Word position in sentence of infinitive construct verbs according to stem in the Aramaic Qumran scrolls.
- Table 9.4* Word position in sentence of *hwh* in the Aramaic Qumran scrolls.
- Table 9.5* Rates per 1,000 words of verbs as first word in the clause or preceded only by a particle together with verbs as the final word in the clause in the Qumran Aramaic scrolls.

Table 9.6*	Rates per 1,000 words of nouns in the construct state together with nouns in determined or absolute state followed by a relative particle in the Qumran Aramaic scrolls.
Table 9.7	Rates per 1,000 words of nouns in determined or absolute state followed by a relative particle in Targums Onkelos, Neofiti and Pseudo-Jonathan.
Table 9.8	Rates per 1,000 words of nouns in the construct state in Targums Onkelos, Neofiti and Pseudo-Jonathan.
Table 9.9*	Rates per 1,000 words of nouns in the construct state together with nouns in determined or absolute state followed by a relative particle in the Ancient Egyptian documents.
Table 9.10*	Rates per 1,000 words of all passive participles together with <i>t</i> -stem active participles in the Qumran Aramaic scrolls.
Table 9.11	Rates per 1,000 words of all passive participles together with <i>t</i> -stem active participles in Targums Onkelos, Neofiti and Pseudo-Jonathan.
Table 9.12*	Rates per 1,000 words of all passive participles together with <i>t</i> -stem active participles in the Ancient Egyptian documents.
Table 9.13	Means, medians and ranges of syntactic variables rates per 1,000 words in the Qumran Aramaic corpus.
Table 9.14	Spearman correlation coefficients between syntactic variables in the Qumran Aramaic corpus. * indicates $p < .05$ (statistically significant association).
Table 9.15*	Low level and high level cluster membership according to syntactical features

- Table 9.16 Sample size (N), means and 95% confidence intervals of the means for each of the syntactical variables by low level cluster group in the Qumran Aramaic scrolls. Note negative values are ‘theoretical’.
- Table 9.17 Sample size (N), means and 95% confidence intervals of the means for each of the syntactical variables by high level cluster group in the Qumran Aramaic scrolls. Note negative values are ‘theoretical’.
- Table 9.18 Verbal position at beginning or end of clauses presence or absence counts for all the Qumran Aramaic scrolls.
- Table 9.19 Passive participle and *t*-stem active participle presence or absence counts for the Qumran Aramaic scrolls.
- Table 9.20 Corresponding rates between Qumran Aramaic cluster groups and the Ancient Egyptian documents and the Targums. A plus (+) sign is inserted for each of the four syntactical variables (nouns in construct state, nouns followed immediately by relative particle, passive participles and *t*-stem active participles) where rates coincide.
- Table 9.21 Rates per 1,000 words for syntactical variables in the Aramaic texts from the Judean Desert Corpus.
- Table 9.22 Rates per 1,000 words of the six syntactic criteria in the Babatha Archive.

- Table 11.1* Class probabilities for each Qumran Aramaic scroll (n=68) per text type: Enochic – Aramaic Levi (EA); Enochic, Four Kingdoms (F); New Jerusalem (N); Tobit/Visions of Amram/Giants (Tam); Biblical (B); Targumic (Targ). Text type in **BOLD** indicates newly classified. Text type in *ITALICS* indicates previously classified and class membership fixed with 100% certainty. Text type in (PARENTHESES) indicates classification less certain, <90% probability of any class membership.
- Table 11.2* Means (95% confidence intervals) for rates of the 22 *a priori* morphological criteria in the different text type groups of scrolls.
- Table 11.3* Text types which have criteria means that are not part of a homogeneous set, so that text types have mutually exclusive high and low rates on Scheffe’s ranges. Empty cells denote text types not excluded from homogeneous sets with either the text types with high and low rates. Medium denotes homogeneous sets which differ from high or low text type mean rates.
- Table 11.4 Frequencies when pairs of text type groups are within the same homogeneous set for the 11 morphological criteria with more than one homogeneous set.
- Table 11.5 Mean proportion of haphels to aphels plus haphels plus one with 95% confidence intervals of the mean.
- Table 11.6 Mean proportion of *l*-prefixed imperfects to all third person masculine singular imperfects plus one with 95% confidence intervals of the mean.

- Table 11.7 Means by text type groups for proportion of third person masculine plural *-w* endings within *-w* or *-w'* (plus unity) as homogeneous subsets (1-4) by Scheffe's ranges.
- Table 11.8 Mean proportion of *m*-prefixed pe'al infinitive constructs within all pe'al infinitive constructs plus one with 95% confidence intervals of the mean.
- Table 11.9 Mean rates for text type groups for proportion of *dy* within total relative particles (plus one) as homogeneous subsets by Scheffe's ranges.
- Table 11.10 Varimax rotated component matrix for the five standardized variables *aphel/haphel* (Haph_prop), *ithpeel/hithpeel* (Hith_prop), *m*-prefixed/not *m*-prefixed pe'al infinitive construct verbs (M_peal_inf_prop), third person masculine plural *-w/-w'* endings (m3p_w_prop), and *dy*/not *dy* relative particles (dy_prop).
- Table 11.11 Mean factor scores and standardized *l*-prefixed imperfect proportion values for the six extended Qumran text types.
- Table 11.12 Spearman correlation coefficients (p-values) between summary alternative forms criteria and other morphological criteria that differentiate the extended text types.
- Table 11.13 Class probabilities for each Qumran Aramaic scroll (n=73) per text type: Enochic – Aramaic Levi (EA); Enochic, Four Kingdoms (F); New Jerusalem (N); Tobit/Visions of Amram/Giants (Tam); Biblical (B); Targumic (Targ). Text type in **BOLD** indicates newly classified. Text type in *ITALICS* indicates previously classified and class membership

fixed with 100% certainty. Text type in (PARENTHESSES) indicates classification less certain, <90% probability of any class membership.

- Table 11.14 The twenty most frequent common nouns listed in order (tied 20th places included) for each text type group corpus. Nouns common to all five groups are not shown in **bold**.
- Table 11.15 The number of different common nouns shared between text type groups, excluding those nouns shared by all groups.
- Table 11.16 The twenty most frequent common verbs listed in order (tied 20th places included) for each text type group corpus. Verbs common to all five groups are not shown in **bold**.
- Table 11.17 The number of different common verbs shared between text type groups, excluding those verbs shared by all groups.
- Table 11.18 Sum of the number of different common nouns and verbs shared between text type groups, excluding those nouns and verbs shared by all groups.

List of Figures

For each Figure the number before the decimal point indicates the Chapter it appears in and the number after the decimal point indicates its sequential position in the Chapter.

Note that some Figures are large and run over more than one page due to formatting regulations. These Figures are marked with an asterisk (*) and can be conveniently viewed at www.qumran-aramaic.co.uk.

- Figure 6.1 Frequencies of Hebrew bible books grouped by word count.
- Figure 6.2 Frequencies of Hebrew bible books with word counts less than 5,000 by word count.
- Figure 6.3 Total number of words versus total number of different lemmas in Septuagintal (LXX) biblical books.
- Figure 6.4 Vocabulary richness of Hebrew Masoretic Text (MT) versus Septuagint (LXX) vocabulary richness for individual biblical books.
- Figure 7.1 Frequencies of total word counts for Qumran Aramaic scrolls considered in *a priori* hypotheses testing.
- Figure 7.2 Hierarchical cluster dendrogram of 22 criteria raw rates per 1,000 words in the 74 Qumran Aramaic scrolls identified as of adequate length.
- Figure 7.3 Two dimensional scaling using the Euclidean distance between the 22 *a priori* criteria for the 74 Qumran Aramaic scrolls.

- Figure 7.4* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls using raw rates of the 22 *a priori* criteria.
- Figure 7.5 Hierarchical cluster dendrogram of 22 criteria standardised rates per 1,000 words in the 74 Qumran Aramaic scrolls identified as of adequate length.
- Figure 7.6 Two dimensional scaling using the Euclidean distance for standardised rates between the 22 *a priori* criteria for the 74 Qumran Aramaic scrolls.
- Figure 7.7* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls using standardised rates of 20 *a priori* criteria.
- Figure 7.8 Box plot of total nouns and determined state common nouns in the five typical text type groups.
- Figure 7.9 Box plot of total verbs and y-prefixed imperfect verbs in the five typical text type groups.
- Figure 7.10 Hierarchical clustering dendrogram of 68 Qumran Aramaic scrolls using standardised rates of 20 *a priori* criteria.
- Figure 8.1* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the five books of Targum Onkelos (Gen_O, Exo_O, Lev_O, Num_O, Deut_O) using raw rates of the 22 *a priori* criteria.
- Figure 8.2* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with Targum Onkelos books using standardised rates of 20 *a priori* criteria.

- Figure 8.3* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the 21 books of Targum Jonathans (suffixed _J) using raw rates of the 22 *a priori* criteria.
- Figure 8.4* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the 21 books of Targum Jonathans (suffixed _J) using standardised rates of 20 of the 22 *a priori* criteria.
- Figure 8.5* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the five books of Targum Neofiti (Gen_N, Exo_N, Lev_N, Num_N, Deut_N) using raw rates of the 22 *a priori* criteria.
- Figure 8.6* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with Targum Neofiti books using standardised rates of 20 *a priori* criteria.
- Figure 8.7* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the five books of Targum Pseudo-Jonathan (Gen_PsJ, Exo_PsJ, Lev_PsJ, Num_PsJ, Deut_PsJ) using raw rates of the 22 *a priori* criteria.
- Figure 8.8* Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with Targum Pseudo-Jonathan books using standardised rates of 20 *a priori* criteria.
- Figure 8.9 Two dimensional scaling using the Euclidean distance between the 22 *a priori* criteria for the 64 Aramaic Documents of Ancient Egypt.
- Figure 8.10* Hierarchical clustering dendrogram of the 64 Aramaic Documents of Ancient Egypt using raw rates of the 22 *a priori* criteria.

- Figure 8.11* Hierarchical clustering dendrogram of the 64 Aramaic Documents of Ancient Egypt using standardised rates of the 22 *a priori* criteria.
- Figure 8.12* Hierarchical clustering dendrogram of the typical Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls of discrete text types using standardised rates of the 22 *a priori* criteria.
- Figure 8.13* Hierarchical clustering dendrogram of the non-core Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls of discrete text types using standardised rates of the 22 *a priori* criteria.
- Figure 8.14* Hierarchical clustering dendrogram of the typical Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the 22 *a priori* criteria.
- Figure 8.15* Hierarchical clustering dendrogram of the non-core Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the 22 *a priori* criteria.
- Figure 8.16* Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with the Qumran Aramaic scrolls using standardised rates of the 22 *a priori* criteria.
- Figure 8.17 Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls aligned to a text type using standardised rates of the 22 *a priori* criteria.

- Figure 8.18* Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls not aligned to any text type using standardised rates of the 22 *a priori* criteria.
- Figure 8.19* Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls not aligned to any text type using raw rates of the 22 *a priori* criteria.
- Figure 8.20* Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls not aligned to any text type using standardised rates of the 22 *a priori* criteria.
- Figure 9.1 Scattergram of frequencies per 1,000 words of nouns in the construct state against nouns followed by a relative particle labelled by Qumran Aramaic scroll.
- Figure 9.2 Two dimensional scaling of rates of syntactic features in the Qumran Aramaic scrolls.
- Figure 9.3 Hierarchical cluster dendrogram of the Qumran Aramaic scrolls using standardised values of the syntactic variables as clustering criteria.
- Figure 9.4 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group A of the Qumran Aramaic scrolls.
- Figure 9.5 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group A, omitting low level cluster group c, of the Qumran Aramaic scrolls.
- Figure 9.6 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group B of the Qumran Aramaic scrolls.

- Figure 9.7 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group C of the Qumran Aramaic scrolls.
- Figure 9.8 Scattergram of rates of nouns in the construct state versus rates of nouns followed immediately by a relative particle in high level cluster group A of the Qumran Aramaic scrolls.
- Figure 9.9 Scattergram of rates of nouns in the construct state versus rates of nouns followed immediately by a relative particle in high level cluster group A, excluding zero values, of the Qumran Aramaic scrolls.
- Figure 9.10 Scattergram for passive participles versus *t*-stem active participles in high level cluster group A of the Qumran Aramaic scrolls.
- Figure 9.11* Hierarchical cluster dendrogram of the Qumran Aramaic scrolls together with Babatha Archive texts using standardized values of the syntactic variables as clustering criteria.
- Figure 11.1 Scree plot of principal components extraction from five standardised variables *afel/hafel*, *itpeel/hitpeel*, *m*-prefixed/not *m*-prefixed *peal* infinitive construct verbs, third person masculine plural *-w/-w'* endings, and *dy/not dy* relative particles.
- Figure 11.2 Loading plot of Varimax rotated component matrix for the five standardised variables *afel/hafel* (Haph_prop), *itpeel/hitpeel* (Hith_prop), *m*-prefixed/not *m*-prefixed *peal* infinitive construct verbs (M_peal_inf_prop), third person masculine plural *-w/-w'* endings (m3p_w_prop), and *dy/not dy* relative particles (*dy_prop*).

Figure 11.3 Two dimensional scaling using Euclidean distances of the relationship between the six extended text types derived from the two factor scores and standardised *l*-prefixed imperfect proportion values.

1. Classification of Qumran texts

Present state of classification of Qumran texts

Hirbet-Qumran lies about 15km south of Jericho. Between 1947 and 1956 eleven caves were discovered that contained thousands of fragments, mainly of prepared animal skin, representing approximately 900 texts many of which are considered copies.¹ Although many of the fragments comprise only short sections of texts, several complete scrolls were discovered. The subsequent publication of the edited texts was not without controversy,² but the rate of publication of volumes of *Discoveries in the Judaean Desert* containing the texts increased significantly under Emanuel Tov as editor-in-chief. The primary classification criterion for classifying these texts for publication was the cave number where the fragment was discovered. There are eleven caves at Qumran, thus the texts are primarily denoted as 1Q, 2Q, ..., 11Q. The numbering reflects the temporal sequence in which the caves were discovered. Geographically caves 4-10 are clustered close together about 1km west of the northern reaches of the Dead Sea with caves 1 and 2 near to each other about 1km further north. Cave 11 is about half a kilometre north of caves 1 and 2 with cave 3 the northernmost by another 300m or so. There is wide variation in the number of fragments discovered in each cave. Caves 9 and 10 have yielded only a single text each whilst over 500 texts have been found in cave 4.

¹ E. Tov (2001a), p102 states that many of the texts represent multiple copies of the same composition. However, for the Aramaic texts at Qumran, the subject of this thesis, such duplication is rare.

² See G. Vermes (2004), Introduction for a personal account.

The current primary classification thus represents the scrolls as archaeological artefacts; the texts, themselves, are not taken into consideration.

Of course, with hundreds of scrolls being found in some caves, a further, secondary classification was necessary to allocate each scroll a unique identity. This secondary classification assigns each scroll a number placed after the “Q” for Qumran. The numbers were allocated sequentially, not according to any more detailed archaeological criteria, but rather in the order that the scrolls were found. The ordering reflects editorial bias in that scholars chose to edit, and thus publish, some scrolls before others.

Moreover, the ordering reflects that of the Hebrew bible. Hence scrolls which contain texts that are found in the Hebrew bible are usually numbered according to the order found in the Hebrew bible and are often given primacy. So, 1Q1 is a Genesis text, 1Q2 is an Exodus text, 1Q3 is a Leviticus text, and so on. The same principal is followed for Caves 2, 4, 5, 6, 8 and 11.³ It is noteworthy that primacy is given to the place of discovery over textual content; it would have been possible to group all the Genesis scrolls together, for example, with Cave number used as a secondary designation.

The relationship between the current classification and the Hebrew bible is not without its problems. First, only the minority of scrolls found at Qumran relate directly to the Hebrew bible texts: 174 of the 900 or so Qumran scrolls were identified with books from

³ See Encyclopedia of the Dead Sea Scrolls (2000) Provisional List of Documents from the Judean Desert, p1013ff for this classification in detail.

the Hebrew/Aramaic bible and were designated as “biblical” texts.⁴ Beyond these texts the classification is arbitrary. Secondly, some scrolls contain material that relates to more than one book of the bible, but the ordering refers only to the majority text. Some scrolls contain texts that come from what in terms of the Hebrew/Aramaic canonical order would be disparate parts of the bible. This suggests that if there were connections between different biblical books, these might not always be identical to those determined later when codices were put together. Thirdly, Qumran texts precede any known Hebrew biblical canon. One can speak of canonical Hebrew scriptures at the time of Qumran, even if this is limited to the Torah, but not of a Hebrew bible *per se*. In publishing the list of biblical texts, Tov admits that the definition does not necessarily relate to what he considers the ‘biblical corpus’ or ‘canonical conceptions’ of the persons who deposited the scrolls, but rather to a later Hebrew canon. Indeed, scrolls that relate to so-called apocryphal texts are excluded from the ‘biblical’ list, even though some of these texts appear in the Septuagint (e.g. Tobit). However, the same ‘biblical’ classification includes scrolls written in Greek that are designated with the sub-classification ‘LXX’ indicating that they relate to the Septuagint.⁵ Tov, in a footnote, considers that this sub-classification is, on balance, appropriate. A further complication in defining Qumran fragments as ‘biblical’ is that some, such as *tefillin* and *mezuzot* contain segments of Hebrew scripture rather than being parts of an entire book.

⁴ E. Tov (2001b). Twenty-six or possibly twenty-eight additional “biblical” scrolls were found elsewhere in the Judean desert.

⁵ ‘Septuagintal’ classification can also be applied to non-Greek scrolls such as 4QJerB that are written in Hebrew but text-critically ‘Septuagintal’.

Vermes accepts the classification of some of the scrolls as ‘biblical’ according to Tov’s criteria, adding the further categories of ‘Apocrypha’, ‘Pseudepigrapha’ and ‘Sectarian’.⁶ This last category comprises texts Vermes describes as being ‘composed or revised’ by the Qumran Community and, apart from the Damascus Document, unknown from any other source. There is, by this definition, some conceptual overlap between the ‘Sectarian’ group with what Vermes considers Pseudepigrapha that are known only from Qumran. Vermes’ definition is also problematic in terms of criteria for composition or revision. T. Lim, whilst pointing out the pluriformity of biblical texts at Qumran, argues that ‘the Qumranians were not mere commentators, but also in some sense authors of the biblical texts’.⁷ This raises the question as to how authoritative the different texts were to the Qumran community, a question I consider in the next section.

For the present, each Qumran text has a unique classification in terms of cave and publication order, and beyond this scholars have grouped them into different broad categories. Furthermore, each text is usually given a label at publication. For “biblical” texts, this is usually that of the book which the text most closely relates to. For other scrolls, there may be other texts known from elsewhere which a particular scroll resembles (e.g. Enoch) and they are labelled accordingly. For many scrolls, the label reflects the editor’s own judgment as to what would be informative about the scroll’s content as a general descriptor. Further characterisation is added to some of the Qumran fragments. Some texts are designated ‘*paleo*’ because they are written in a paleo-Hebrew

⁶ G. Vermes (2004), p11.

⁷ T. Lim (1997), p. 109.

script rather than the Assyrian ('square') script. Tov argues that these texts reflect ancient traditions 'since they were probably copied from texts that were also written in that script'.⁸ Other texts are sub-classified according to the material they were written on. Implications for textual relationships of the paleo-Hebrew script are limited because only texts of the Torah and Job are represented. Implications for relationships between texts relating to the material used are even less clear.

In summary, the present state of classification of Qumran texts is somewhat arbitrary, depending on order of discovery or publication, or anachronistic, depending on notions of 'bible' and 'canon' that do not pertain to the period during which the texts were written and deposited.⁹ Nevertheless, there is agreement that a proportion of Qumran fragments are related to texts found separately from Qumran whilst others are possibly unique to Qumran. It may be thought that, as McMahon and McMahon state, "It must be confessed that classification in and of itself does not immediately sound particularly gripping",¹⁰ but such classification is an essential prerequisite to the understanding of the relationships between the texts found at Qumran.

⁸ E.Tov (2001a), p.106.

⁹ This is a consistent theme that arose out of the Hampton Court conference on the Judaeian texts held in 2000 CE whose proceedings were published as *The Bible as Book* (2002).

¹⁰ A. McMahon & R. McMahon (2005), p1.

Qumran texts and Scripture: a basis for classification?

Although the terms biblical and canonical are misleading anachronisms, the notion that some of the Qumran texts can be classed as ‘Scripture’ carrying some authority is perhaps more tenable. If we could determine which texts at Qumran were considered as ‘Scripture’, this would provide the basis for a classification that could replace the current “biblical” and “non-biblical” categories. Lim argues that the concept of ‘Holy Scripture’ is equally applicable to the Qumran commentaries and Pauline letters.¹¹ Bruce Metzger asserts

From the first day of its existence the Christian Church possessed a canon of sacred writings – the Jewish Scriptures, written originally in Hebrew and widely used in a Greek translation called the Septuagint.

and explains the label ‘Scripture’ because citations from it were introduced by the formula ‘it stands written’ (*γεγραπται*).¹² Greek texts, which have been designated as ‘Septuagintal’ have been found at Qumran, but it is unclear whether the Qumran community used these more commonly than their Hebrew/Aramaic counterparts, especially since the latter are far more numerous. The preference of the Qumran community for Hebrew/Aramaic over Greek texts can be inferred not only from the observation that Greek texts comprise only around 3% of the ‘biblical’ fragments, but also from John Wevers’ observation that ‘the scribes probably knew their Hebrew text

¹¹ T. Lim (1997) titles his monograph accordingly.

¹² B. Metzger (1987), p2.

much better than the LXX, and inadvertently added text to agree with a longer Hebrew text.’¹³

Wevers is careful not to state explicitly that the scribes responsible for these Greek texts were from the Qumran community since it is quite possible that these scrolls originated outside of Qumran. The corollary of this is that just because a text is discovered at Qumran this does not necessarily imply that it had any authority for the community. What is also clear from the Greek texts is that ‘Scripture’ as we know it remained textually fluid during the Qumran epoch with scribes feeling free to amend received texts where they felt appropriate. Lim argues that similar flexible approach to Scripture was shown by Paul.¹⁴ Similarly, it cannot be assumed that just because a particular text is not found at Qumran (e.g. Esther) that it was not considered as Scripture by the community; such texts may have failed to survive the nineteen centuries necessary for discovery.

Despite these caveats there is considerable evidence that many texts were considered authoritative. First, multiple copies of several ‘biblical’ texts were recovered. Most numerous are Psalms (36), Deuteronomy (30), Isaiah (21), Genesis (19-20), Exodus (15) and Leviticus (12).¹⁵ These six books also provide 82% of all New Testament citations

¹³ J. W. Wevers (2005), p23 concludes this after collation of the Qumran Pentateuch Greek texts.

¹⁴ T. Lim (1997) also cautions against the assumption made by Metzger, for example, that Christian Old Testament Scripture was necessarily Septuagintal.

¹⁵ E. Tov (2001a), p104-5 as of 2000 CE.

of the Old Testament.¹⁶ This lends contemporary support to the likely widespread authority of the Torah together with Psalms and Isaiah. Secondly, commentaries (Pesharim) have survived from Qumran on some other 'biblical' texts. The majority of these relate to the minor prophets. Thirdly, other authorities, such as the grandson of ben Sirach, indicate that a group of texts was recognised as authoritative comprising the Law, the Prophets and other writings.¹⁷ However, a problem lies in determining what comprised these 'other writings' not just for ben Sirach, but also for communities such as the early Christians and Qumran.

The situation for the early Christians, roughly contemporaneous with the Qumran community, is a little clearer and thus, perhaps, can illuminate issues pertinent to Qumran. To preface this discussion, it is helpful to deal first with some terminological issues. Bruce Malina in his critique of the Pontifical Biblical Commission document, *The Jewish People and their Sacred Scriptures*, cautions against received views where labels such as 'Jewish' and 'Christian' are retrojected to the period of the Qumran community.¹⁸ He perceives that

The use of the terms "Jews" and "Christians", implying and asserting a constant identity between first-century Eastern Mediterranean Israelite groups referred to in biblical documents and contemporary groups called "Jews" and "Christians"...to be a fundamental error.

He argues that ancient labels from an Israelite perspective depended on whether the Israelite was speaking in Judea, Galilee or Perea or outside these areas. Such issues of

¹⁶ Novum Testamentum Graece. 27th ed., 2001. Calculated from appendix IV, loci citati vel allegati.

¹⁷ Sirach Prologue 1:1.

¹⁸ B. Malina (2005), p287ff.

expressing identity are pertinent to Christians such as Paul, but are less problematic for the Qumran community with its Judean sectarian perspective. If the Qumran community could consider itself as the 'true Israel' (e.g. 4Q239), so could the community associated with Matthew's gospel. The label 'Christian', by contrast, originated outside Israel (Acts 11:26) and, apart from Acts, is used only once (1Peter 4:16) elsewhere in the New Testament. To label the 1st century CE group of Jesus' followers as 'Christians' is not to deny their own perceptions of Judean identity in contrast with other peoples considered as *ta ethnē* (τὰ ἔθνη). It is arguable that in this regard the identity of early Christians living in Judea was closer to the Qumran community than to the vast majority of Christians in the 21st century.

A major criterion that would have distinguished Judean Christians from Qumran sectarians is the acceptance of the authority of Jesus. Metzger contends that even at a very early stage the words of Jesus, handed down in oral tradition, would have had an authority equal to, and at times greater, than the Torah.¹⁹ Moyise points out the parallels between Jesus and the Qumran community's 'Teacher of Righteousness'.²⁰ In time, however, it was not only Jesus' words that were considered Scripture by early

¹⁹ B. Metzger (1987), pp2-3.

²⁰ S. Moyise (2001) pp 10-11. In particular Moyise highlights that both figures are seen by their communities within an eschatological context. It is unlikely that either the Qumran or Christian communities contemplated texts they considered as scripture would be the subjects of academic study 1900 years later. Although Moyise is right in pointing out the parallel between the 'Teacher of Righteousness' and Jesus, he is in danger of overplaying the role of the Teacher of Righteousness for the Qumran community since he is mentioned in only a limited number of texts found at Qumran.

Christians, Paul's letters also began to be considered authoritative.²¹ Lüdemann discusses the difference between texts, such as Paul's letters, having authority for certain early Christian groups and achieving canonical status.²² Even when New Testament canons were established, differences remained between Eastern and Western churches with regard to which books were included let alone establishing a single authoritative text. Like Lüdemann, Étienne Trocme in his book *The Childhood of Christianity* argues that the various texts that eventually were brought together to form the New Testament were originally associated with specific Christian communities that each had a distinct understanding and practice of Christianity.²³

The pluriformity of early Christianity parallels that of contemporary Judaism if, indeed, it can be considered separately in the 1st century CE. Fragments discovered at Qumran may themselves reflect the diverse nature of second Temple Judaism since there is no *a priori* reason to assume that only texts considered authoritative for the community were deposited in the caves. One question that arises here is whether the texts deposited in individual caves reflect particular traditions. For early Christian communities one indication of which writings were considered authoritative is their citation by other authors. The overlap between New Testament citations of books from the Old Testament

²¹ B. Metzger (1987), p4.

²² G. Lüdemann (1996) views the formalisation of a Christian canon as a reaction against 'heresy' typified by Marcion. The notion that Scripture, as a written record, is a way of defining boundaries may have been more important for communities that were not physically bounded or separate such as Qumran. This is one sense in which scripture may be labelled 'Holy'.

²³ É. Trocme (1997).

has already been mentioned. This overlap may also be seen to extend to the lack of any citations from Esther in the New Testament and the failure to discover any texts related to Esther at Qumran. The Nestle-Aland 27th edition of the *Novum Testamentum Graece* (NA27) considers there to be four allusions to Esther: Revelation 4:5, Luke 10:13 and two separate verses in Mark 6:23.²⁴ Of these, the allusion in Revelation 4:5 to Esther 1:1 is considered definitely Septuagintal. Josephus, by contrast, draws on the more extensive Septuagint version in his retelling of the story of Esther in Book 11, chapter 6 of his *Jewish Antiquities*.²⁵ Hence it cannot be assumed that 1st century Christians drew largely on Greek texts of Old Testament books whilst Judeans drew on the Hebrew/Aramaic texts. Indeed, NA27 records no Septuagintal citations or allusions to Ecclesiastes (Qoholeth). This is unsurprising since the Greek translation is unlikely to have been undertaken before the 2nd century CE.²⁶ Thus Metzger's assertion that Christians possessed a canon of sacred writings in Greek from the very first (see footnote 12 above) requires some qualification. Moreover, Menken argues that the precise form of citation of Old Testament passages in early New Testament texts influenced later versions of the Septuagint.²⁷ Thus the model of Old Testament texts (e.g. the Septuagint) influencing what outsiders might consider sectarian writings (here the New Testament) cannot be assumed. 'Old' and 'New' scripture evolved together.

²⁴ *Novum Testamentum Graece*. 27th ed., 2001, pp 784-5.

²⁵ Josephus in Loeb Classical Library: *Josephus* vols. 7-9. Ed. G.P. Goold. Cambridge MA: Harvard University Press, 1963-1965.

²⁶ J. K. Aitken (2005).

²⁷ M. J. J. Menken (2002) gives John 2:17 citing Ps 69:10 influencing Vaticanus and Sinaiticus as an example.

One result of this process of old and new scripture evolving together is that some old texts considered authoritative enough to cite fail to be included in the eventual canon. A well known example is the citation from Enoch in Jude 14. Other references alluded to, though not directly cited, and noted by NA27 include Jubilees, Martyrdom of Isaiah, Psalms of Solomon, Assumption of Moses, Apocalypses of Baruch and Elijah, and various Testaments of the Twelve Patriarchs. Such citations cannot be categorised in terms of the anachronistic category of ‘non-biblical’ texts, but have to be considered as Scripture from the perspective of the communities associated with the books making the citations or allusions.

For the Qumran community, then, relevant Scripture might include not only the Torah, Psalms etc., but also Enoch and Jubilees, texts that are not conventionally classified as ‘biblical’. There are more Qumran texts related to Enoch, for example, than many of the books of the Hebrew/Aramaic bible. Vermes classifies these texts as ‘biblically based apocryphal works’,²⁸ but there is considerable overlap with the ‘non-biblical’ citations in the New Testament. Even for biblical books such as Psalms the boundaries of Scripture are unclear. 11QPsa contains fragments that relate to Ps 151 in the Greek Psalter and others that are known in Syriac. The Peshitta Old Testament was adopted by eastern Christians in much the same way that the Septuagint came to have authority in the west with similar alterations of Old Testament texts to conform more closely to New

²⁸ G. Vermes (2004), p537ff.

Testament theology,²⁹ although the Eastern New Testament canon was more restricted than the Western canon. Projecting trajectories from Qumran in the inter-testamental period forward to fourth century writers such as Aphrahat and Ephrem is speculative, but different strands of Judaism detectable in Qumran texts do resonate with other writers.

The Enochic traditions found in Jude, the true Israel tradition that bears similarities to Matthew (Matt 19:28) and later Syriac fathers,³⁰ have already been mentioned. Another tradition that is found in Ephrem is that around Melchizedek. Murray considers that Ephrem is following a midrashic tradition, known from the Palestinian Targum, that Melchizedek is Noah's son Shem.³¹ This focus on Melchizedek, who is mentioned only twice in the Hebrew/Aramaic bible, is prominent in Hebrews with regard to Jesus' priesthood. 11Q13 describes Melchizedek in messianic terms referring to Sons of Heaven, Sons of Light and Sons of God and contrasting him to his demonic counterpart Belial. Such traditions strengthen the impression of the Qumran community's eschatological orientation. The corollary of this is that any division of Qumran texts in terms on apocalyptic or non-apocalyptic genres is likely to be an external projection since the community would be likely to read all scripture with an eschatological

²⁹ R. Murray (2006), p17 argues for a trajectory from Qumran through rabbinic Judaism in the east through to Syriac theologians such as Aphrahat though he notes the opposed views of Brock, Neusner and Sandmel.

³⁰ R. Murray (2006), p174.

³¹ R. Murray (2006), p180.

hermeneutic. For example, 11Q13 not only cites Daniel, a typical apocalyptic book, but also Leviticus (Lev 25:9) on the instructions about sounding the trumpet to proclaim the Jubilee.

If external judgements of genre may be inappropriate to determine the Qumran community's own understanding of how different texts related to each other, similarities to literature found beyond Qumran remains useful in placing Qumran texts in the wider context of contemporary eastern Mediterranean literature. Unfortunately, apart from biblical texts there is a paucity of such literature to facilitate this task. For the present, then, a classification based on a definition of 'scripture' is unworkable. Nevertheless, a review of what might have been considered as 'scripture' at Qumran and beyond indicates that a complementary approach to classification in terms of assumed 'biblical' or 'scriptural' texts is thus necessary.

Can the Qumran texts be considered as a corpus?

Classification of Qumran texts into those that relate to others found beyond Qumran and those that are unique to Qumran raises the question of how appropriate it is to view the Qumran scrolls as a single corpus: in archaeological terms are we dealing with one body or two, or perhaps even more? We might hypothesise textual relationships that mirror those of the Christian bible where Old Testament and New Testament texts were brought together. The difficulty of such a hypothesis is that it is projected back from a canonical

perspective. If we were to be looking at the relationship between early Christian literature and the Hebrew/Aramaic scriptures during say the 2nd century CE such textual homogeneity would be less clear as the Marcionite movement illustrates. A further complication, as we seek to organise the ‘bones’ into one or more ‘skeletons’ is that many of the pieces are missing. As Vermes puts it,³²

After all this time, we are still not certain that we have collated the whole evidence correctly or interpreted it properly. Questions continue to arise in the mind and there is still no way to be sure of the answers.

It is only recently that the whole body of evidence has been laid out in public view. Thus far it has appeared rather like rabbits being produced out of a magician’s hat and each text placed according to where it was found and to what other texts it might relate. We are now at a point where different methods may be used to improve the collation of the evidence and thus its proper interpretation. The situation has much in common with questions about the Pauline corpus in the New Testament. Scholars have long questioned whether this is a single body and were also aware of missing pieces (Col 4:16). It is only over the last five decades or so that this question has been able to be addressed using quantitative methods.

³² G. Vermes (2004), p 26 further notes the absence of documents that provide a systematic framework for the Qumran sect’s constitution. Despite this he takes the ‘sectarian’ documents as a corpus identified with the Qumran community as a single entity, though later, p 34ff he discusses differences between those who resided permanently at Qumran and others who formed a ‘community’ beyond its physical boundaries.

Unlike the Pauline corpus which is accepted as being as complete as possible, until recently scholars were aware of ‘missing pieces’ of the Qumran corpus in terms of texts that remained to be published. However, under Emanuel Tov as editor-in-chief of *Discoveries in the Judaean Desert*, the corpus is now essentially complete, although re-editing of texts is likely to continue for some time. Classification of the corpus is now ready for the next phase to take it beyond that based on cave and publication order. But how might this be done? Tov, himself, suggests a way forward:

The classification of the Qumran texts remains a difficult assignment. The texts should not be grouped by cave, since the contents of each individual cave were not homogeneous. Nor should they be classified by origin (copied by the Qumranites/brought from outside), since this distinction is neither firmly secured nor sufficient. Nor should the texts be classified by date, by palaeographical or codicological criteria, since none of these criteria are firm. Probably the best criterion for classification is according to textual character, even though this criterion is also problematic.³³

Accordingly, Emanuel Tov has pioneered the classification of Qumran texts according to textual criteria. He has particularly focused on different text types according to how closely Qumran texts align with the Masoretic text, Samaritan Pentateuch or retroverted Septuagint.³⁴ As he states,

³³ E. Tov (2002), p152.

³⁴ Tov classifies these as proto-MT, pre-Samaritan and texts close to the presumed Vorlage of the LXX.

The principle behind this classification is the recognition that all texts can be grouped according to the degree of closeness to MT, LXX, or SP without accepting the claim that these three texts are the central pillars (recensions, texts, text types, etc.) of the biblical texts.³⁵

In specifying that the collection and arrangement of texts depends on the degree of closeness of one datum/set of data to another, Tov's principle is ideally suited to a statistical approach. Indeed, in the appendix to his article Tov presents statistics for the 'de luxe' scrolls that includes measures of margins, number of lines, height and correction rates per line.³⁶ Statistical approaches have also formed the basis of many investigations into the unity of the so-called Pauline corpus referred to above. I will discuss how statistics have been used in various corpora in the next chapter. But first, there are two major difficulties in adopting a classification based on textual criteria for the entire Qumran corpus.

The first difficulty is that grouping of Qumran scrolls according to external texts (e.g. MT, LXX etc) can only be applied to the minority of the Qumran corpus. For most Qumran scrolls there are no suitable external texts available for comparison. In this sense, the 'biblical' scrolls enjoy a privileged position. It is not unreasonable, therefore, that scholars have focused so much attention on these. Nevertheless, to set aside classifying the 'non-biblical' texts where textual criteria are available, albeit different in nature from those used to make a comparison with external texts, may mean that highly valuable insights into how the different texts relate to each other will never be realized.

³⁵ E. Tov (2002), p152.

³⁶ E. Tov (2002), pp158-159.

This is a key point. Textual criteria can refer to features typical of a specific ‘text type’ (e.g. proto-Masoretic), but also to the use of language (e.g. vocabulary, morphology, syntax etc). As noted above, for a large proportion of the Qumran corpus we are without any ‘standard’ texts, so we need to rely on the linguistic features to classify texts textually. Such a classification will necessarily be about relationships between texts *within* the corpus and draw on shared textual criteria. Texts that display substantial shared textual criteria can thus be termed ‘text types’ on the basis of these typical features as against closeness to some external standard text. The second difficulty arises from the need for shared textual criteria: grouping of ‘non-biblical’ Qumran scrolls can only be language specific. This is not to preclude non-textual classifications in terms of genre or content, but textual criteria cannot be applied across languages. In other words, shared textual criteria will differ between texts in Hebrew, Aramaic or Greek. If Tov’s suggestion of using textual criteria as the optimal approach to classifying scrolls found at Qumran is to be followed, and such criteria would be of limited use if they could only be applied to ‘biblical’ scrolls, then there are at least three separate corpora that need to be considered separately, those in Hebrew, Aramaic and Greek. There are also a few scrolls where the language cannot be determined. All three corpora are possible objects of scholarship; which one to start with is somewhat arbitrary. However, there are relatively few Greek texts at Qumran so that statistical approaches, such as those I will use in this thesis, are unlikely to be successful. The Hebrew corpus, on the other hand, is by far the largest. It is often useful to pilot new approaches on smaller amounts of data before undertaking colossal analyses. The Aramaic corpus offers a compromise: it is neither too big nor too small. It contains some ‘biblical’ texts, but also a wide range of other genres,

the only drawback is that it is not thought to contain any texts that might be designated as 'sectarian'. So, to strike a 'happy medium', it is with the Qumran Aramaic corpus that I will begin and make the object of this thesis.

Present state of classification of the Qumran Aramaic corpus

There are around 120 texts written in Aramaic found at Qumran. This number is approximate because there are a few short fragments where it cannot be determined whether they are written in Aramaic or Hebrew. The Aramaic texts were found in caves 1, 2, 4, 5, 6 and 11 only. In terms of subject matter, they can be thought of in a few broad categories.³⁷

- 1) Texts that contain Aramaic sections also found in the Books of Ezra and Daniel in the Hebrew bible (4Q112, 4Q113, 4Q115, 4Q117);
- 2) Texts that are Aramaic versions of parts of Tobit found in the Septuagint (4Q196, 4Q197, 4Q198);
- 3) Texts that are Aramaic versions of parts of the Books of Enoch and Jubilees, cited as authoritative in the New Testament (1Q23, 1Q24, 4Q201-4Q212, 4Q530, 4Q531, 4Q532, 6Q8);

³⁷ This summary classification is somewhat arbitrary, based largely on content and genre, and follows that given in G. Vermes (2004).

- 4) Texts that are closely related to books of the Hebrew bible (1Q20, the Genesis Apocryphon; 4Q156, Targum of Leviticus; 11Q10, Targum of Job);
- 5) Texts that are related to patriarchal figures of the Hebrew bible (1Q21, Testament of Levi; 4Q213-4Q214b, Aramaic Levi; 4Q537, Testament of Jacob; 4Q538, Testament of Judah; 4Q539, Apocryphon of Joseph B; 4Q540 and 4Q541, Apocryphon of Levi; 4Q542, Testament of Qahat; 4Q543-4Q548 and possibly 4Q549, Visions of Amram);
- 6) Texts that relate to Daniel or court tales during the exilic period (4Q242, Prayer of Nabonidus; 4Q243 and 4Q244, Pseudo-Daniel; 4Q550, 4Q550a and 4Q550c, Proto-Esther);
- 7) Texts that relate to a vision of the New Jerusalem (2Q24, 4Q554, 5Q15, 11Q18);
- 8) Miscellaneous, relatively short texts, often with apocalyptic content.

It is clear that without the labels provided by the various editors, the primary and secondary classification by cave and order of publication are not very informative about how the texts might relate to each other. Indeed, there is not even anything to indicate that these texts are in a different language from the bulk of Qumran texts since the Aramaic texts appear as short sequences in the midst of those in Hebrew. The categorization above highlights two features of the Qumran Aramaic texts: first, though there are unsurprisingly few 'biblical' texts, many relate to biblical material, and second there are no texts that relate directly to the rules, liturgy etc of the community at Qumran. From the classification and descriptions, it is hard to judge the provenance, Qumran or beyond, for most of the scrolls written in Aramaic found at Qumran. The

scrolls form a single corpus in that they were all discovered at Qumran, but beyond that the present classification is unable to tell us anything. In particular, it is of no use if we wish to understand the textual relationships between the different scrolls.

2. Statistical approaches relevant to Qumran Aramaic texts

How might we go about investigating the textual relationship between Aramaic texts found at Qumran? Apart from portions of Daniel, all the texts differ, so the well-established methods of comparing variants are not open to us. Instead we have two broad, and not necessarily exclusive, options. The first would be to make some qualitative judgement about similarity, but it is unclear which criteria we might choose, especially if we are to eschew any quantitative influences. The second is to take a quantitative approach, deciding to count one or more specific textual characteristics, then comparing how often they occurred in different texts. Quantitative approaches are well established in biblical studies, though their utility depends on an understanding of statistics.

The Oxford English Dictionary defines statistics as,¹

In early use, that branch of political science dealing with the collection, classification, and discussion of facts bearing on the condition of a state or community. In recent use, the department of study that has for its object the collection and arrangement of numerical facts or data, whether relating to human affairs or to natural phenomena.

The *sine qua non* of any statistical approach is the availability of numerical data.

Numerical data have, in some ways at least, been integral to biblical texts from an early

¹ Shorter OED (1973).

date. Chapter divisions, and hence indexing, are found in the fourth century CE Codex Vaticanus and further numerical classification followed with the Eusebian canon tables.² Relevant to the Hebrew/Aramaic bible, though from a later date, were the various numerical data listed in the Masorah.³ However, the invention of the electronic computer has allowed a major advance in the ability to apply statistical approaches to texts. The earliest example of a computer-assisted approach to textual statistics is that of A. Q. Morton whose counts of *kai* in various works, including the Pauline corpus and the Shepherd of Hermes, were published half a century ago in 1964.⁴

This early work provided *kai* counts for samples of texts, but did not perform any formal statistical tests to estimate the probability that these counts were associated with different authors. Some later authors involved with computer-assisted studies of biblical texts have sought to distinguish between what they term ‘mere enumeration’ and ‘statistics’ that involves hypothesis testing.⁵ To make this distinction using these terms is to misunderstand the nature of statistics since, as the OED definition confirms, statistics includes the ‘collection and arrangement of numerical facts’. It is thus useful here, as a preliminary matter, to describe what is understood by ‘statistics’ in this thesis.⁶

² B. M. Metzger & B. D. Ehrman (2005), p34ff.

³ P. H. Kelley, D. S. Mynatt & T. G. Crawford (1998).

⁴ A. Q. Morton & J. McLeman (1964).

⁵ K. J. Neumann (1990), p13ff. takes this line citing Morton for support.

⁶ I apologise to those readers who are well versed in statistical methods, but despite their long and eminent history, statistical approaches remain unfamiliar to some scholars in the field of Biblical studies. In addition to the following section, I have also provided a glossary of statistical terms to provide an easy point of reference (Appendix 1).

A background to statistical methods

When we are faced by data such as the Qumran Aramaic texts, it is often helpful to be able to describe it to others without replicating the raw data. It is the use of statistics that allows us to do this for quantitative data. For example, we might want to describe some text by explaining that the body of the text (the letters) appears in the middle of the page surrounded by margins. Statistically we could describe this in terms of measures of letter density according to page position: the margins would have a low letter density, the centre would have a high density. Similarly we could give the mean number of letters per line, word, paragraph etc. This is not 'mere enumeration' because we have had to choose specific criteria that we think are useful in describing the text. To emphasise, all statistical approaches require *a priori* value judgments; the utility of the statistics produced depends on those initial value judgments. Hence, a statistical approach to the Qumran Aramaic texts requires some initial value judgments about which textual criteria are important to count. If the wrong criteria are chosen, no amount of statistical sophistication will be of avail, the results will still lack utility.

Once we have decided which quantitative criteria are useful, that is which variables we are going to measure, we can use these variables to perform statistical hypothesis testing. Simply put, statistical tests fall into two broad categories: those that test for differences between variables and those that test for associations. In fact, such tests are often the two sides of the one coin, as it were, since measuring how far apart two things are (tests of

difference) can also be thought of as measuring how close together they are (tests of association). Thus a statistical approach is useful when a research question can be expressed in terms of quantitative differences or similarities between variables (e.g. textual features) of interest.

It follows from these considerations that statistics necessarily deal with issues of probability and error. First, there is always some likelihood of error occurring when something is measured. Even something as simple as counting the number of letters on a page demonstrates this. Secondly, sometimes not all the data that would be useful are available. For example, one might be interested in calculating the mean number of letters per word for a book,⁷ but unfortunately some accident has left a few of the words unreadable. Thus only a sample, albeit a large one, of the total number of possible words in the book is available. The mean number of letters cannot then be known for certain, but it can be estimated from the available sample. Depending on sample size and the variation of letters per word in the sample, not only can a mean be estimated for the whole book, but also, importantly, some idea of the likely error of this estimate can be calculated. In this example, the greater variation in letters per word and/or the smaller the sample size, the greater the probable error in estimating the mean number of letters per word. At this point it is also worth noting that the way samples are selected may affect the likelihood of detecting differences. Returning to the example of letter density,

⁷ The mean number of letters per word is calculated by summing all the letters for all of the words and dividing this grand total by the number of words in the book. Means are thus one type of 'average'; they are probably the most common type of average used in statistical analyses.

if samples are selected by dividing each page in quarters symmetrically, all four samples are likely to contain similar portions of margin, whereas if samples are selected at random from the page but using the same quarter page sample frame, significant differences in letter density are more likely to be detected.

In one sense, once statistical methods are applied to real data, they relate to the field of epistemology. As former US Defense Secretary Donald Rumsfeld famously said,⁸

[T]here are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know.

Statistics takes what is known (data) to estimate the extent of what is unknown (margin of error). Moreover, it describes the margin of error in terms of probability: a 5% chance that the margin of error will be to this extent, a 1% chance that it will be to this extent etc. The relationship between probability and the margin of error depends on the distribution of the data.

The distribution that reflects the variation of a variable, as well as its mean, is thus an important statistical measure. Observation over long periods has shown that many variables exhibit similar patterns of distribution. One common distribution is called a 'normal' (sometimes termed 'Gaussian') distribution. This is a symmetrical distribution about the mean value with frequencies of observed variable values diminishing and

⁸ In a statement to the press, February 2002.

tailing off the further from the mean they are. An example would be women's heights. Most women will have heights close to the mean height in their population with only a few women being either very tall or very short. One measure of the amount of variation of a variable is called the standard deviation.⁹ The larger the standard deviation, the wider is the distribution in absolute terms. For a variable with a normal distribution, 95% of the data will lie within approximately two standard deviations either side of the mean. This mathematical property of the normal distribution allows statistical testing of differences between two samples. For such testing to be performed, the assumption that the two samples' data come from normal population distributions is required. The first formal stage of testing is first to state a null hypothesis that the two samples come from the same population. Next the means are compared to see if they lie within two standard errors of each other. The standard error is sample size dependent as it equates to the standard deviation divided by the square root of the sample size. Thus if the sample consist of only one observation, the limits equate to two standard deviations. But as more observations are available, the more certain it becomes that the mean lies close to their values. If the means are more than two standard errors apart, the null hypothesis is rejected at a statistical significance level of 5% (i.e. the difference could occur by chance

⁹ Formally the standard deviation is given by the equation

$$\sigma = \sqrt{E[(X - \mu)^2]}.$$

where σ is the standard deviation, X is the variable and μ is the mean value of X . In practice this means subtracting the mean from each value of X , squaring it and summing the total for each value of X , then finally taking the square root of the total sum. The total sum, before the square root is taken to calculate the standard deviation, is termed the variance. Both the standard deviation and the variance are measures of how much variation there is in the values of any observed variable and are often used to express the size of statistical effects (the amount of influence of one variable on another).

only once in every twenty comparisons) and the samples are considered statistically different.

Of course, not all variables are normally distributed, the example of letters per word is a good instance of this, but there are various statistical methods that allow description and testing of data with different distributions. It can be seen, therefore, that data description is not 'mere enumeration', but an essential preliminary statistical task prior to proceeding to statistical testing. It is also critical to be aware of the important role error plays in statistics. Choosing a significance level of 5% means that we are dealing with probabilities, not certainties. On some occasions a more stringent level of significance may be appropriate, perhaps 1%. If one performs lots of comparisons drawing samples from the same population, then by chance a few will be deemed statistically significant. The more comparisons that are made, the greater the chance of a statistically significant result. The potential error in interpreting such comparisons as probably being different is termed Type 1 statistical error. The counterpart is having so few data in a sample that the estimate of the mean and distribution has a wide margin of error (i.e. the standard error will be large). Hence comparisons to a sample from a population with a different mean may not be found to be statistically significant. This situation is termed Type 2 statistical error. Such errors can be adjusted for to some degree by altering the level of significance used as the criterion for difference. However, it is usually better practice to decide *a priori* what level of significance is thought important to a specific comparison and design data collection accordingly. Sometimes, however, this luxury is not afforded us due to the paucity of available data. The Qumran Aramaic texts, unfortunately, provide many instances of this.

The issues surrounding Type 1 and Type 2 statistical error may appear esoteric, but they are essential in any statistical analysis. They address two situations that Donald Rumsfeld failed to cover in his famous epistemological edict. First, sometimes statistics can lead us to think we know things that in fact are unknown (Type 1 statistical error). Secondly, we can sometimes know things that, because of statistics, we think we don't know (Type 2 statistical error). The following review of how statistics have been used for Qumran texts and in other areas of biblical studies, will find little mention of Type 1 and Type 2 statistical error being considered but, nevertheless, their importance should not be dismissed.

Emanuel Tov and the conventional text-critical approach

In the previous chapter, I noted the shift in classification criteria applied to the Qumran scrolls from the initial archaeological, through publication, to Emanuel Tov's pioneering of textual characteristics. This shift had heralded the introduction of statistical approaches to the Qumran texts. Tov focused attention on the so-called biblical scrolls to begin with, comparing Qumran texts to external "standard" texts. He has persisted with this approach. Tov, in his revised edition of his *Textual Criticism of the Hebrew Bible*,¹⁰ continues to argue for the Masoretic text as the principal standard with which manuscripts should be compared. He contends that for many of the texts found at

¹⁰ E. Tov (2001a).

Qumran there are few differences compared with the Leningrad Codex (BHL) of some thousand years later.¹¹ From this observation he defines a group of consonantal proto-Masoretic texts, not all derived from Qumran. He notes that the BHL is only one of a group of Masoretic texts and that a single archetype of the Masoretic text may never have existed.¹² Nevertheless, the careful preservation of this text type over many centuries does offer a relatively fixed external standard for statistical comparison. Such a comparison will inform about how similar to or different from a specific text from a period one thousand years later is the particular Qumran text in question. Using a fixed external standard for classification, however anachronistic, is at least more textually informative than classifications based on which cave texts were found in. As cited above, Tov has used three external standards for classification: the proto-Masoretic (relating to the BHL), the proto-Samaritan (relating to the Samaritan Pentateuch (SP)) and the Septuagintal (LXX). However, his classification of Qumran texts does not fall into three corresponding groups, but five. He adds one group of texts ‘written in the Qumran Practice’¹³ and another that he describes as ‘Non-Aligned’ (which is, in fact, a miscellaneous category not of just one text type)¹⁴ His classification scheme, therefore, is not purely externally derived, but also depends on internal comparisons between Qumran texts. Tov thus provides a precedent for basing classification of Qumran scrolls

¹¹ Two examples he reproduces photographically come from 1QIsa^b, but these still contain 13 of what Tov describes as ‘minor differences’ (excluding orthography) in a total of 26 verses, Tov (2001a), p30.

¹² E. Tov (2001a), p25. I refer to the Masoretic Text following convention, but this is not to reject Tov’s observation of plurality of Masoretic texts.

¹³ E.Tov (2001a), p114.

¹⁴ E.Tov (2001a), p116.

on textual criteria, applying statistical methods to Qumran texts and making textual comparisons between Qumran texts without reference to external “standard” texts.

However, he does not bring these three methods together. Beyond the Masoretic text, the limitations of using an external standard text become increasingly obvious.

The comparison with the Samaritan Pentateuch parallels that with the BHL. The number of texts that can be compared is more limited because only five of the ‘biblical’ texts are available for comparison, though comparison by analogy is possible with other Samaritan texts that cover Joshua, Judges, Samuel, Kings and 2Chronicles.¹⁵

Comparison with the LXX is far more problematic, however. First, as noted above in the discussion of the Greek Qumran texts, the LXX, more so than the Masoretic text, represents a heterogeneous group. Some scholars seek to make a distinction between the Old Greek (OG) and the LXX.¹⁶ Secondly, and more obviously, there is the issue of how to compare Hebrew and Greek texts. Commonly, comparison is made with ‘the presumed Hebrew source’.¹⁷ The reliability of such presumed Hebrew sources is debatable. B. Wright demonstrates that such a procedure would be open to considerable error in the case of ben Sirach.¹⁸ Perhaps the degree of error associated with comparisons with translations explains why Tov considers only 5% of Qumran biblical

¹⁵ E. Tov (2001a), p81.

¹⁶ R. T. McLay (2003) for example, though he cautions against the over-categorization of many texts as kaige-recensions.

¹⁷ Thus E. Tov (2001a), p115. Tov states explicitly that the Vorlage of the LXX was a single biblical text, p116, allowing such direct comparisons by translation.

¹⁸ B. J. Wright (1989).

texts to be close to the LXX, and these to show considerable heterogeneity.¹⁹

Fortunately, Wright shows that the translation technique for a fair proportion of books is more tightly rule-based than that of ben Sirach, so there remain possible methods to address this issue of comparison for many texts.

Similar issues of translation arise for ‘non-biblical’ texts found at Qumran where equivalent texts, albeit in other languages, have been found beyond Qumran. The issue of which external standard is appropriate is even more problematic here. A leading group that would have claims to be considered as scripture is the Jubilees material.²⁰ The Ethiopic manuscripts that provide full texts are over 1,500 years younger. Moreover, given the diversity of Qumran Jubilees texts, the appropriate standard may be some prototype that underpins not only this material but also other related texts such as that in Jacob of Edessa’s ‘Jewish Histories’.²¹ Hence, for the Aramaic texts found at Qumran, the conventional text critical approach of comparing versions is not viable, and retroverting equivalent texts from other languages fails to provide a suitable external standard. An alternative approach is thus necessary, one that has already been suggested by Tov’s use of comparisons between Qumran texts without resort to any external texts. Such approaches are now well established in the field of biblical studies research, if not for Qumran texts.

¹⁹ E. Tov (2001a), p116.

²⁰ C. Hempel (2000).

²¹ S. P. Brock (1978).

Andrew Morton, quantitative approaches to New Testament texts and implications for Qumran

One way forward, if no appropriate external standard is available that corresponds with the text being considered is to compare more general stylistic features rather than to quantify textual variants. This approach was pioneered by Morton as discussed with regard to *kai* usage above. The initial research question Morton sought to answer using stylistic criteria was whether all the so-called Pauline corpus was written by Paul. Morton assumed *a priori* that Galatians, or later Galatians and 1Corinthians, were genuine Pauline works and so could be used as standards to compare other putative Pauline letters to by using a variety of stylistic variables. Kenny has carefully demolished the validity of this approach for the Pauline corpus, by showing that if other sections of Galatians are compared using Morton's criteria, these are designated non-Pauline.²² The implication is that either the criteria were inappropriate to determine what is Pauline or the chosen standard is inappropriate. Neumann found similar problems with the choice of external standard texts when he examined the Pauline corpus.²³ His original plan to use a variety of contemporary non-biblical texts had to be abandoned because these differed from the Pauline letters far more than the letters differed between themselves. A similar problem of choice of criteria and/or standard (normative) texts, though not formally tested statistically, is one possible reason why Tov finds a grouping of non-aligned texts at Qumran.

²² A. Kenny (1986), p101ff.

²³ K. J. Neumann (1990).

Neumann provides an extensive, critical review of statistical approaches to scripture up to 1990. He summarises three authorship models that utilise standard texts: the consistency model, the population model, and the resemblance model.²⁴

The consistency model tests whether a particular text is consistent with an author by comparing it to known works of that author. In the context of the Qumran texts an example would be comparing the text to the BHL. The population model compares a particular text with a range of contemporary authors to test to which one the text most closely fits. This is the method adopted by Neumann, though eventually non-biblical standard texts are restricted to 1 Clement and Ignatius, neither of which is generally thought of as candidate authors of disputed Pauline letters.

The resemblance model tests which of two alternative standard texts a particular text most closely resembles. An example in the context of Qumran would be to test whether a Pentateuchal text was closer to the BHL or SP. Authorship does not have to equate with a specific known person, but is really an expression of text type in terms of style. Occasionally this text type might be an artificially constructed standard. An example is Güting and Mealand's investigation of asyndeton in Paul.²⁵ First they construct a critical

²⁴ K. J. Neumann (1990), p19 following R. Wachal.

²⁵ E. W. Güting & D. L. Mealand (1998). The way the statistics are performed relies on defining these pluses and minuses, though in reality the choices involved in constructing an artificial standard text from the individual manuscripts is equivalent to a purely within manuscript group comparison.

text based on all available manuscripts of Romans, 1Corinthians and 2Corinthians, then for each manuscript they count how many added (pluses) or omitted (minuses) asyndeta there are compared with the constructed text.

The population model adopted by Neumann allows him to use discriminant analysis. Briefly, this relies on a first step of comparing known standard texts to establish which quantifiable stylistic variables significantly discriminate between them. This produces what is termed a weighted equation, in which some variables are more certain discriminators (i.e. more highly weighted) than others. The next step is to apply this weighted equation to the text where authorship is uncertain to determine which of the standard population texts it is closest to. Unfortunately, for Qumran texts there are very few suitable external texts that can act as ‘standards’. Eugene Ulrich argues the term ‘standard text’ is not applicable to any particular text type of the Second Temple period.²⁶ Julio Trebolle-Barrera considers that the term ‘standard text’ can be applied to the Torah, Isaiah, Twelve Prophets and Psalms, but not for other ‘biblical’ Qumran texts.²⁷ His evidence for this contention comprises the number of preserved copies of specific books, the relative homogeneity of texts and referencing by the Qumran pesharim. Although the number of preserved texts may reflect their scriptural authority, this does not account for chance survival of texts or for the possibility that authoritative texts may not have standard forms. With regards to textual homogeneity, Trebolle-

²⁶ E. C. Ulrich (2000), p85 contends that the only candidate for a ‘standard text’ is the so-called ‘proto-Masoretic’ text which he sees as being retrojected from ‘the mind of the modern person’.

²⁷ J. Trebolle-Barrera (2000), p98ff.

Barrera himself points out the presence of proto-Samaritan texts at Qumran. The issue of pesharim, like that of the number of surviving texts, may indicate the authority of particular biblical books at Qumran, but not of any standard text of these. George Brooke asserts that it is the interpretation rather than the forms of text present at Qumran which is distinctive.²⁸

Although Morton's pioneering work has subsequently, been shown to be flawed, it was seminal in developing statistical approaches to biblical texts based on textual characteristics. In particular, two key points have emerged: first, a statistical approach can ensure that valid criteria are chosen to test similarities and differences between texts and secondly, that establishing a 'normative' set of texts within a larger group facilitates comparisons.

Frank Moore Cross and the quantification of variant readings

Despite these caveats, comparisons with external text standards are represented within the literature. An example germane to this methodological discussion is Cross and Saley's *A Statistical Analysis of the Textual Character of 4QSamuel^A (4Q51)*.²⁹ This publication demonstrates that Tov is not the only longstanding Qumran scholar who has adopted a statistical approach to Qumran texts. Cross and Saley's approach is similar to that of Güting and Mealand in that a critical text is constructed and then variant readings

²⁸ G. J. Brooke (2000), p119.

²⁹ F. M. Cross & R. J. Saley (2006).

of any text element (not just a specific feature such as asyndeton) classified as ‘superior’, ‘inferior’ and ‘unclassified’. This grouping of all variants assumes that equal weight should be given to all of them, an assumption which Cross and Saley do not provide any justification for. The data they present, excluding reconstructed readings, shows 83 superior, 26 inferior and 29 unclassified variants for 4QSamA 1Sam and 2Sam1:1-9:13 compared with the Old Greek translation and 40 superior, 20 inferior and 15 unclassified readings compared with the Masoretic Text. Cross and Saley argue that greatest weight should be given to inferior variants: 19% of all variants versus the Old Greek and 26.7% versus the Masoretic Text. Although entitled a ‘statistical’ paper, Cross and Saley do not formally test whether this is significantly different, but a simple calculation provides a χ^2 of 2.44 which is non-significant at the 5% significance level. In other words, the null hypothesis that inferior variants do not differ between the Old Greek and the Masoretic Text cannot be rejected. Moreover, there is considerable uncertainty in the classification itself. Suppose so-called ‘unclassified’ readings could be classified as superior or inferior, this could affect the proportions greatly. Cross and Saley analyse 2Sam10:1-24:25 separately to allow comparison with the so-called Kaige recension as well as to the Lucianic recension and the Masoretic Text. Statistical analysis is, however, far more problematic here because there are only two variant readings, excluding reconstructions, against the Kaige recension, and 32, only four of which are ‘inferior’, against the Masoretic Text. Cross and Saley rely on ‘superior’ readings, which they have already argued are less discriminatory, and reconstructions that cannot be considered as objective data. Yet Cross and Saley conclude,

In summary, the evidence allows for only one compelling conclusion, that 4QSam^a stands firmly rooted in the Hebrew textual tradition reflected in the Old Greek, with only a minimum of cross-fertilization detectable with the textual tradition which was to develop into the Proto-Rabbinic and Rabbinic Recensions.

These criticisms show the methodological pitfalls of this kind of quantitative approach, and this does not even bring into consideration the difficulties of translation technique in 1 and 2Samuel.³⁰

In summary, any one external text relates to only a proportion of what might be considered scripture at Qumran, only two are available in Hebrew and just one in Aramaic (portions of Ezra and Daniel), and these BHL (or other Masoretic texts) and the SP are not contemporary with the Qumran material. The use of external standard texts as a primary method of classification is thus far from secure and not practicable for the Aramaic texts found at Qumran. This is not to say that comparison with external texts is impossible or unhelpful, but that initial classification of texts using only Qumran materials is the most appropriate first step prior to comparing any identifiable text groups with known text traditions.

³⁰ B. G. Wright (1989), p49.

Qumran texts considered as a corpus

Making comparisons between a group of similar texts to elucidate relationships is the basis of much text-critical scholarship.³¹ Statistical approaches to this task using large amounts of data have been facilitated by electronic computers. Kenny provided a thoughtful computer-assisted study of New Testament texts that illustrates some basic statistical tests that can inform about classification, in this case in terms of authorship.³² He calculates Pearson correlation coefficients³³ between stylometric variables to determine similarity between New Testament books and χ^2 values³⁴ to test for differences. He concludes that,

Some of the problems studied, such as the Pauline canon, would be suitable for treatment by more advanced methods such as cluster analysis... Though such analysis would not, I believe, be likely to alter any of the conclusions presented here, it would undoubtedly permit a more sophisticated and graphic method of presenting the evidence.³⁵

³¹ B. M. Metzger & B. D. Ehrman (2005).

³² A. Kenny (1986).

³³ This correlation coefficient is used when variables have a normal distribution and is named after the British statistician Karl Pearson. It supposes that two variables have a linear relationship such that $y=bx+a$ where y and x are the two variables, b is the gradient of the line and a is the intercept. If the observed data are plotted and the best line estimated mathematically (i.e. a and b estimated from the observed values of x and y), the correlation coefficient expresses how close the observed data group around that line. It is thus a measure of the strength of the relationship.

³⁴ The Chi square value is a measure of difference used for data that are counted in different categories (e.g. ten oranges, thirteen pears etc).

³⁵ A. Kenny (1986), pp 121-122.

Subsequent scholars have taken up this challenge. Therefore some consideration of the limitations of relying on correlation coefficients illustrates why more sophisticated, multivariate methods are necessary.

Kenny totals up scores of 96 features for each letter in the Pauline corpus for 50 word sections of each letter. Sample sizes for each letter vary according to length, Philemon providing six samples, Romans 142. He calculates Pearson correlation coefficients between each of the thirteen books (Hebrews is not included). The 78 resulting Pearson coefficients range from 0.854 to 0.982, most being greater than 0.95 and all greater than 0.90 except for those with Titus. The issue is not whether these correlation coefficients are significant; they are, but this may be the case for any comparison of two Greek letters from that period. The issue is whether there are significant differences between the correlation coefficients. With such large values and such relatively small sample sizes, none of the correlation coefficients differ from each other significantly. Even if they had, making comparisons between 78 coefficients produces a total of 3081 comparisons (hypotheses tested). By chance 154 of such comparisons should be significant at the 5% level – an example of Type 1 statistical error. Hence there is a need for statistical techniques that can describe similarities between variables more parsimoniously.

Güting and Mealand present relationships between manuscripts graphically using correspondence analysis.³⁶ This plots each manuscript in 2-dimensional space according to specific criteria, here according to the tendency to add or lose particles at the beginning of sentences. Correspondence analysis is statistically related to χ^2 testing in that it deals with variables that are counts of different categories rather than data that are normally distributed. Standard correspondence analysis uses χ^2 to estimate the distance between cases (e.g. manuscripts). A similar, but more powerful, approach is available for variables that are normally distributed, factor analysis.³⁷ The commonest factor analysis method is to extract principal components which represent those traits that variables have in common. For example, a principal components analysis of New Testament letters might detect a Pauline trait (the use of specific vocabulary, word positioning, rhetorical devices etc), a Johannine trait (paratactic style) etc. The analysis can determine how many traits parsimoniously explain the data, produce loading plots³⁸ displaying relationships between cases (e.g. New Testament letters) or variables and manipulate these plots to show how cases or variables relate to specific traits or variables.

³⁶ E. W. Güting & D. L. Mealand (1998), pp 106-109.

³⁷ Factor analysis is a statistical method that provides a simplified summary of related variables. It identifies common factors that these related variables share. For example, different meals might share common factors of spiciness, sweetness, acidity etc. even though the specific ingredients will differ between them. Factor analysis can be orthogonal, where the factors are unrelated to each other, or oblique, where the extracted factors are related to each other to some degree. For an introductory explanation see D. Child (2006).

³⁸ A loading plot displays how each variable in a factor analysis 'loads' onto each extracted factor. A factor loading indicates how close a variable is to a factor (0 indicates no relationship, 1 indicates the variable and factor are identical).

Yehuda Radday and statistical analysis of selected Hebrew

bible texts

Although New Testament Greek texts have received the most sophisticated statistical approaches using textual criteria, the Hebrew bible has not been entirely left behind.

Yehuda Radday was an early, and still one of the few, scholars to apply statistical analysis of stylistic features to the Hebrew bible. He first used correlation coefficients and χ^2 testing to test the linguistic unity of Isaiah.³⁹ Some elaboration of these techniques was applied to a similar analysis of Zechariah,⁴⁰ but this study was criticised on methodological grounds some time later by Portnoy and Peteren⁴¹ who suggested that multivariate analysis of variance⁴² would be more appropriate. Radday considered using analysis of variance in his work on Judges,⁴³ but rejected this approach because some of the stylistic variables he chose had very low values. In fact, the absolute size of the mean of any variable is not an issue in analysis of variance, but homogeneity of variance between groups tested is a criterion that is required. That is the spread of values of a

³⁹ Y.T. Radday (1970).

⁴⁰ Y. Radday & D. Wickmann (1975).

⁴¹ S. Portnoy & D. Peteren (1984).

⁴² Variance, as defined above in footnote 9, is a measure of the variability in an observed variable. Analysis of variance is a statistical technique which estimates what proportion of that variability is due to another variable. For example, the variance in height may be accounted for, in part, by male or female sex and by parental height. Multivariate analysis of variance estimates the amount of variance in several variables taken together explained by a variable. For example, the amount of variance explained in height, shoe size and demi-span explained by male or female sex.

⁴³ Y. T Radday *et al.* (1977).

variable such as word length has to be similar in the different sections of a text such as Judges to fulfill the mathematical assumptions behind the statistical procedure. In reality, for large sample sizes, this criterion is fairly robust. Finally, in 1985 Radday published a similar statistical study investigating the authorship of Genesis.⁴⁴ By now he was using more advanced statistical techniques such as factor analysis and cluster analysis as mentioned by Kenny. Unfortunately, the methodology was again severely criticised.⁴⁵

Cluster analysis is a promising statistical method of describing similarities between texts. The technique consists of estimating distances between texts according to a basket of variables. There are several ways of calculating such distances depending on the types of variables involved (interval, ordinal, binary etc). Texts can then be grouped into clusters according to distance. This can be done hierarchically (i.e. in order of distance) or by pre-specifying the number of clusters. Hierarchical models might be particularly applicable to a situation where there was a family tree of texts (e.g. New Testament Greek manuscripts). Pre-specified clusters may be more appropriate when texts are hypothesised to belong to a number of different text types (e.g. different dialect versions of the same Vorlage). The utility of cluster analysis, and indeed factor analysis, is that it not only describes internal relationships between texts, but that cluster membership can be used to look for associations with external 'standard' texts.

⁴⁴ Y. T. Radday et al (1985).

⁴⁵ P. Davies (1986).

Multivariate analysis of variance suggested as a suitable method by Portnoy and Peteren, unlike factor analysis and cluster analysis, tests for significant differences rather than significant associations. This form of analysis of variance is termed multivariate because there is more than one dependent variable. Effects of factors or covariates are tested for significant effects on the shared variance of the dependent variables. For example, the effect of text (factor) on word length, syllable number, percentage of nouns in the construct state etc. (dependent variables) adjusted for text length (covariate). There is no assumption that dependent variables are independent, an important consideration when investigating style. Multivariate analysis of variance can test not only for significant effects on the shared variance of dependent variables, but also if there are significant differential effects between dependent variables. For example, there may be significant differences in the way texts are associated with word length compared with nouns in the construct state.

Multivariate analysis of variance, like univariate analysis of variance, has a few drawbacks. First, as previously mentioned, there is an assumption of homogeneity of variances of dependent variables. This may not always be the case for textual analyses. There are several ways around this. One would be to use equivalent ordinal analyses,⁴⁶ but this sacrifices considerable power since it ignores information actually available (i.e.

⁴⁶ Ordinal variables contain data that are ranked (1st, 2nd, 3rd etc) rather than having absolute values. Data that have absolute values can, of course, be allocated a rank (highest through lowest) to transform them into ordinal data. A different set of statistical analyses are appropriate to ordinal data.

it ranks data rather than uses the observed values). Another way is to try to manipulate the data by transforming it mathematically (e.g. by logarithmic or exponential functions) so that distribution criteria for analysis of variance are fulfilled. This is a traditional approach, but not all data are transformable. More recently random effects models have been applied to address this constraint.

Random effects models assume that the sample, for example Qumran texts labeled by cave, is a random sample from a larger population. The means and distributions may therefore differ between groups (e.g. caves, where one cave may contain texts with a higher proportion of a particular textual characteristic). Adjusting for these random effects, the contribution of factors to the variance of a dependent variable can be estimated. This model is consistent with the assumption that extant Qumran texts are only a sample of a larger group that were deposited. Random effects can be incorporated into general linear models used for analysis of variance. However, it is more common to incorporate them into what are termed mixed linear models. These models use data more efficiently because, unlike analysis of variance that can only use data from cases (e.g. samples of text) where all the data are available, missing data values can be adjusted for in the model. Again, this may be advantageous where there are missing sections of text or values are omitted because of uncertainty. Mixed models also make hierarchical or so-called multi-level modeling easier. For example, we could test whether cave number was important by checking for its effect at a level above that of the individual texts found in it.

Random effects are also important to consider if a multi-level, sometimes known as hierarchical, analysis is to be performed. If there were strong *a priori* reasons to suppose that the cave a text was found in was associated with textual characteristics, such an approach might be warranted. This is because textual characteristics might be open to influence at different ‘levels’, for example by individual scribes and also by cave. Although not directly concerned with textual classification, this is a suitable place to review a statistical study by H. Michael Fried who looked at multiple copies as a criterion to estimate the number of text repositories that comprise the Qumran corpus.⁴⁷ Fried concludes that 40 independent repositories were combined and, if his contention is correct, it may be that each repository has distinct textual characteristics that would represent a level within any analysis. However, there are a few methodological problems with the analysis. First, Fried uses the ‘titles’ of texts to define multiple copies. But for non-canonical works these are often arbitrary as the texts themselves do not usually bear a title. Secondly, he calculates proportions of multiple copies for caves 3, 7 and 8 which he includes in the analysis. However, cave 3 has only ten texts, cave 7 has three texts and cave 8 five texts as he lists them: numbers far too small to use conventional statistics with. Thirdly, there are several implicit, unjustified assumptions made in the modeling process. Fried’s initial model gives equal weighting to each cave in calculating the overall proportions of multiple copies and multiple titles (i.e. he first calculates the proportion for each cave then takes the mean of these proportions) which is a very unusual method. It would be conventional to calculate the overall proportion by

⁴⁷ H. M. Fried (2010).

summing multiple copies across all caves and then dividing by the total number of texts. His next model is only a slight improvement, still treating each cave independently. His third model treats cave 4 separately and combines data from the other caves and hence enumerates multiple copies across caves rather than within each. In other words, the tacit assumption that independent repositories cannot be placed in more than one cave is no longer held. Having shown that the data for these other caves now resemble those of cave 4, instead of concluding that repositories need not be cave-specific, Fried concludes that the findings must imply that cave 4 is made up of a number of separate repositories. Indeed, Fried goes further to assert that the data support each cave, except cave 4, constituting a single independent repository. He estimates the mean size of this repository to be 19 texts from the mean catalogue size of the eight caves apart from cave 4: he does not look at how wide a confidence interval would be. From this he can calculate that cave 4 consists of 30 separate repositories. But this would be far fewer if the mean size were, say, 25 texts, and far more if mean size was 13 texts. Even with 19 texts, this would mean cave 1 comprised two repositories. Moreover, Fried fails to consider differential attrition: that is, a greater proportion of scrolls may have survived in some caves than in others. This would be likely to underestimate the size of original repositories. These analyses illustrate how a series of assumptions used without caution can lead to unjustified conclusions albeit presented with great precision. It is a warning against over-simplistic models, Fried's analyses were carried out by hand from an Excel spreadsheet, and thus provide a stimulus to consider more sophisticated statistical approaches.

These more sophisticated statistical techniques have almost exclusively been applied to Greek biblical texts so far. For example, O'Donnell applied hierarchical cluster analysis to the New Testament corpus to show that when the Didache is added it clusters with James followed by Revelation.⁴⁸ However, Semitic texts, because of their trilateral root nature, do provide the option of one sophisticated statistical approach that would be far more difficult to apply to Greek texts: Markov chain analysis. Thus far, the statistical methods I have reviewed have required the assumption that the observed data, though possibly related, are independent of each other. This property is not necessarily the case for a piece of text which, for example, has to be read in a specific order. Each word or letter of the text can be thought of as a separate 'event' linked to both the preceding letter or word and the subsequent letter or word, thus forming a chain. Markov chain analyses depend on this property of ordered dependence of data.

An informative Markov chain analysis of the Hebrew bible was performed by Abraham Boyarsky and Pawel Góra.⁴⁹ They chose the first three letters of each word, excluding vowels inside words, and looked at which three letters followed. If the word had fewer than three letters, the missing letters were represented by spaces. From this they estimated the mean number of words that could follow any other word in the Hebrew bible to be 106.8 with a standard deviation of 138.8 words. They also estimated the entropy, maximum entropy and thus the relative entropy for the Markov chain (i.e. the text of the Hebrew bible). Entropy in this textual sense is a measure of the uncertainty of

⁴⁸ M. B. O'Donnell (2000).

⁴⁹ A. Boyarsky & P. Góra (2000).

the sequence of letter triplets representing words. It is thus a measure of the average information content of random text samples. Low values indicate greater certainty so that if a text sequence was entirely predictable, entropy would be zero. Boyarsky and Góra's analysis is purely illustrative; it is questionable whether the Hebrew bible with its many discontinuities can be considered as a corpus in this way. Nevertheless, it provides a note of caution when considering whether to include suggested readings for lacunae in Hebrew texts given that there would, on average, be a choice of over one hundred words.

Summary

In this chapter I have surveyed a range of statistical approaches that might be relevant to investigating the relationships between the Aramaic texts found at Qumran. The conventional text critical methods have been set aside as impractical, limitations of simple bivariate statistics have been noted, and more complex analyses previously used with New Testament texts which might be suitable for Qumran texts have been reviewed. In brief, several mathematical methodologies are promising, but whether complex or simple, all rely on selecting the correct variables, in this case textual characteristics, if they are to be informative.

3. Textual Classification Criteria

The choice of textual criteria is the key element of any textual classification. Although different statistical approaches may influence which texts are considered to be closer to or further away from each other during the classification process, it is the textual criteria chosen to classify the Aramaic Qumran texts statistically that are likely to have the major effect on outcome. At the outset it is essential to comprehend that the most common method of textual classification employed in biblical studies, classification based on variants of the same underlying text, cannot be applied to the Qumran Aramaic scrolls. This conventional approach requires the measurement of textual variants between related texts, especially where standard texts are represented by multiple copies. But apart from some Aramaic portions of the biblical books of Daniel, each Aramaic scroll comprises a unique text.

An alternative strategy to determine appropriate textual criteria, well tested in other languages and which has been applied to the New Testament,¹ is the corpus linguistic approach. This attempts to classify texts first on stylistic grounds. However, a major disadvantage of this approach is the sample sizes needed for each text under consideration to achieve adequate power to detect significant differences. That is, if a particular stylistic feature is thought to be important for classification, but only occurs infrequently, it will only be of use for relatively long texts in which it can occur often enough to allow comparison between texts. From the point of view of classifying the

¹ M. B. O'Donnell (2005).

Qumran Aramaic texts, we are forced to make do with what we have: we cannot go and find some more lengthy texts just because a specific textual criterion does not occur frequently enough in the texts we have. The nature of the texts determines which textual criteria are appropriate. Despite this limitation, this stylistic method has one major advantage over the textual variants classification: it allows comparison with texts outside the corpus on stylistic grounds which can help place the corpus texts in context.

Stylistic criteria

The quantitative classification of texts on stylistic grounds, sometimes known as stylometrics, is well established in the field of biblical studies of Greek and, to a lesser extent, Hebrew texts, but not for Aramaic. Thus there is a general background literature that may illuminate the question of which criteria are suitable for statistical approaches in Aramaic, but a scarcity of specific guidance. In view of this scarcity of guidance for Aramaic texts, it is useful to consider stylistic criteria in more general terms as applied to other languages. For example, Kenneth Neumann's extensive overview of past statistical studies considers stylistic criteria according to seven categories: lexical, morphological length, syntactic length, morphological category, syntactic category, and non-grammatical variables.² These categories are helpful, but because Neumann is concerned with the authorship of the Pauline epistles, many of the studies he reviews are of Greek grammatical features that are not of direct relevance to Semitic languages such as Hebrew and Aramaic.

² K. Neumann (1990), pp 23-114.

For similar reasons, the more recent case frame and semantic case-relations studies of Danove³ and Wong⁴ are unable to inform about Aramaic stylistic criteria to any great extent. Bearing these caveats in mind, I will consider each of Neumann's stylistic categories in turn and review how these may be applicable to Aramaic or Hebrew which, as a cognate language, is likely to be informative.

Lexical criteria

Traditionally, lexical studies focused on the occurrence of rare words, especially *hapax legomena*, as a stylistic criterion. Such data were easier to handle in a non-computerised world, but their rarity renders them difficult for statistical purposes with small text samples. With computerisation of texts, examination of frequently used words has become possible and thus statistical comparison of vocabulary between different texts. This can be seen as a paradigm shift in the application of lexical criteria to textual classification. Morton's work on *kai* in the Pauline corpus discussed in the previous chapter, for all its flaws, signaled this shift, though it took quite some time before it was applied to the Hebrew bible.

Radday illustrates this transition from traditional to more complex lexical criteria in his authorship studies of various books of the Hebrew bible published about ten years after

³ P. L. Danove (2001).

⁴ S. S. M. Wong (1997).

Morton's pioneering work. In his study of the Book of Judges,⁵ the only lexical criteria amongst the thirty-eight textual criteria he considers are *hapax legomena* and *dislegomena*. In other studies⁶ he introduces measures of vocabulary richness (the number of 'types' or lemmas divided by the square root of the number of 'tokens' or words) and concentration (the percentage of the whole text covered by the fifty most frequent words). Using these two criteria Radday is able to compare text authorship using the Sichel distribution. This distribution was originally developed by Sichel⁷ and used by him to model *hapax dislegomena*.⁸ The Sichel function estimates the number of times a particular word (type) will occur in a text of given length. Pollatschek and Radday have shown that only two parameters, α and θ , are necessary to describe the Sichel function distribution in full with the slope of the head of the distribution being described by α alone, whilst θ alone describes the slope of the tail.⁹ Sichel estimated α by the ratio between the number of *hapax legomena* types and the total number of types (different words) in the text and θ by the ratio between the total number of words (tokens) and the number of types. However, iterative model fitting of the optimal distribution approximating to the actual data distribution (i.e. type counts) can estimate these parameters without the need to count *hapax legomena*. It is easiest to perform such modeling using a single part-of-speech category, usually nouns. The key point is that using such functions results in a shift from classifying texts according to whether they

⁵ Y. T. Radday (1977).

⁶ For example Y. Radday & D. Wickmann (1975).

⁷ H. S. Sichel (1975).

⁸ H. S. Sichel (1986) for tokens that occur exactly twice in the corpus.

⁹ M. Pollatschek & Y. T. Radday (1981) and as applied in Y. T. Radday *et al* (1985).

contain specific lemmas to classifying texts in terms of the diversity of lemmas each contains. Although computers render this feasible, they do not address the question as to whether such an approach is valid.

The question of validity is very important since other lexical measures have been developed along similar lines. For example, a simpler, and more commonly used, vocabulary criterion is Yule's K that can be considered a measure of vocabulary richness based on the Poisson distribution. This may be less reliable than the Sichel distribution, but Sichel¹⁰ considered it relatively stable independent of text size rendering it useful for smaller samples. The Sichel distribution may provide estimates of the rates of a particular word (type) which in short texts will be less than one for most words (types), hence it is not very useful unless quite long texts are being considered. Neumann reviews related indices in his overview, including a measure of entropy.¹¹ Entropy is an interesting entity with regard to lexical stylistic criteria because it assumes that each textual element, in this case a word (lemma) can predict the next. For example, if an English text started "Once upon a" it would be likely to continue next with "time". Radday's lexical approach considers textual elements as independent, unlike Boyarsky and Góra's Markov chain methods discussed in Chapter 2, which were applied to three letter sequences. Once text is considered as a sequence or chain, the importance of considering syntactic criteria is clear. To illustrate this, measures of vocabulary richness, concentration, etc. would remain unaltered given the same words however they were

¹⁰ H. S. Sichel (1986).

¹¹ K. Neumann (1990), pp 30-31.

distributed in the text. Thus two texts using the words in different orders would be judged textually identical on these criteria that assume that the units of measurement (words) are independent of each other. This is akin to the situation when Eric Morecambe is playing the Grieg piano concerto with Andre Previn conducting. Previn complains to Morecambe that he is not playing the right notes, to which Morecambe responds that he is playing the right notes, but not necessarily in the right order.

One constraint on which words may follow any specific word is the semantic domain of the section of text. Radday recognises this in his study of Isaiah, empirically dividing the vocabulary into special groups based on their subject matter.¹² However, his method lacks the linguistically rigorous approach of Porter and O'Donnell in their analysis of Romans.¹³ They are fortunate in dealing with a New Testament text to be able to draw on the Louw-Nida lexicon¹⁴ for carefully defined semantic domains rather than having to construct these themselves. Porter and O'Donnell examine word frequencies in Romans which they describe as,

On the one hand, a very primitive and blunt device for exploring the meaning of Romans, but, on the other hand, still quite revealing.¹⁵

¹² Y. T. Radday (1973), pp 204-214.

¹³ S. E. Porter & M. B. O'Donnell (2000).

¹⁴ J. P. Louw (1988).

¹⁵ S. E. Porter & M. B. O'Donnell (2000), p161.

They note that with 93 domains listed in the Louw-Nida lexicon, such semantic information is not practicable at the verse level, though some patterns appear at the chapter level that allows cluster analysis. What Porter and O'Donnell build towards is a lexical classification that extends beyond words taken in isolation. However, they criticise other attempts to move beyond single words that rely on a Chomskyan framework of formal grammar.¹⁶ Specifically, they charge David Black's work on the linguistics of New Testament Greek¹⁷ with being unsatisfactory with this regard. What Black is careful to describe in some detail, however, is how textual meaning is linked not only with lexical forms but also with syntactical features, for example those associated with rhetorical language.

Graeco-Roman rhetorical language is not a framework that can be readily applied to the Qumran texts. However, as Black notes, the point of rhetorical devices is often to emphasise certain aspects of the text.¹⁸ Emphatic devices, though of a different nature, are not uncommon in Hebrew texts. Gesenius cites interrogative and desiderative

¹⁶ S. E. Porter & M. B. O'Donnell (2000), p158 referring to grammatical structure as conceived by Noam Chomsky in *Aspects of a Theory of Syntax* (Cambridge: MIT Press, 1965) in which he makes a distinction between deep and surface structures and in which there is a semantic component. In fact, Chomsky lies in a long tradition of endowing sentences with meaning by adding a component to a formal system; for example, Alfred Tarski in his seminal work on truth theory (e.g. *The Concept of Truth in Formalized Languages*. In: *Logic, Semantics, and Metamathematics*, 2nd ed., John Corcoran (ed.), Indianapolis, Hackett, 1982). Semantics tends to rely on some notion of truth. I do not intend to discuss the philosophical underpinnings of semantics in this thesis since it is somewhat distant from the formal statistical approach to classification that I am going to pursue, although the figure of Tarski demonstrates how close language and mathematics can be.

¹⁷ D. A. Black (1995).

¹⁸ D. A. Black (1995), p132.

sentences as examples of rhetorical devices in Hebrew.¹⁹ Amongst other types of sentences that Gesenius considers alongside these are those that contain an oath. These are interesting lexically because they are formulaic with the individual words being interdependent, and thus clearly breaking assumptions of statistical independence. Seow provides a list of such oath formulae that are either introduced by the Niphal of *šb'* or *hy* plus the subject being sworn by. Maledictory oaths (curses) follow the formula *kh-y* 'Sh-ly 'thym *wkh ywsp* or similar followed by a clause (the apodosis) introduced by *ky* **י** **י** for positive conditions or 'm for negative conditions.²⁰ Arnold and Choi provide a comparable, somewhat briefer summary in their Guide to Biblical Hebrew Syntax.²¹ Dobson treats these issues more fully in his typically nuanced fashion.²² For example, he points out that not all occurrences of *hy-yhwh* are part of an oath formula (e.g. 2Sam 22:47) which has important implications for computerised processing of electronic texts. Since there are relatively few oaths compared to total text in the Hebrew bible, a few incorrect classifications might have a major impact on the data. Thus computerised procedures can be used to screen for potential oath occurrences, but these need to be checked manually.

¹⁹ Gesenius (1910), §150 & 151.

²⁰ C. L. Seow (1995), p304ff.

²¹ B. T. Arnold & J. H. Choi (2003), pp 188-189.

²² J. H. Dobson (1999), pp 186-190.

Tony Cartledge provides a detailed survey of vows in the Hebrew bible.²³ He makes a clear distinction between promises, oaths and vows.²⁴ He describes oaths as promises strengthened by a curse; that is, if the promise is not fulfilled, the curse will come into force. By contrast vows start with a request made by a human, usually to God, with an attached promise conditional on the request being granted. Cartledge notes that the full vow formula includes a verb or noun based on the root *ndr*. He identifies 31 occurrences of the verbal form and 60 occurrences of the nominal form in the MT.²⁵ Cartledge proposes that conditional vows occur frequently in poetic books such as Psalms, but that both the form and content of poetic vows differ from those that occur within narrative.²⁶ In particular, Cartledge asserts that the conjunction 'm ׀ ׀ does not occur as part of vows in Psalms, though it can be part of oaths.²⁷ Instead, the 'm plus imperfect is replaced by a third person imperative or jussive and a connective *w* to the apodosis never precedes the simple imperfect verb. There are forty-two occurrences of 'm in Psalms, the majority of which do not fulfil vow formulae. However, Ps 132:12 does have a vow using 'm, though spoken by God rather than to God; conditional promises by God having the force of vows. Ps 81:9-14 also forms a vow made by God interspersed by an historical observation. Nevertheless, the implications of Cartledge's study are that computerised classification may not only include false positives when simple criteria are used, but may also miss many actual occurrences of a phenomenon. If this is the case for

²³ T. W. Cartledge (1992).

²⁴ *ibid.*, pp14-18.

²⁵ *ibid.*, pp138-142.

²⁶ *ibid.*, p152ff.

²⁷ *ibid.*, p154.

fairly set formulae such as oaths, the difficulties for lexical studies extending beyond single words for less fixed situations are likely to be considerable. This brings us back to Porter and O'Donnell's criticism of an approach based on formal grammar that ignores the semantic context. Lexical criteria are limited in this respect. For example, vocabulary richness based on lemmas fails to account that the same lemma may have several different functions depending on the semantic context, yet these will not be taken into account if only the form is considered when quantifying this.

Indeed, beyond the relatively clear link between lexeme and meaning in the context of stereotypical formulae such as oaths, lexical-semantic relationships in classical Hebrew are less well defined. Groom reviews a number of biblical Hebrew lexical semantic studies and finds that they either have a very narrow focus or use eclectic methods often based on the investigator's own intuition of semantic fields that may not reflect those of the writers of biblical texts.²⁸ Statistical studies of lexeme association frequencies may provide a starting place for a more systematic approach for classical Hebrew. However, given the limited size of the Qumran Aramaic corpus, such an approach is unlikely to be worthwhile.

Even though lexical criteria do not appear very reliable candidates as a statistical basis of textual classification, perhaps because they are familiar from non-quantitative biblical studies scholarship, they are the ones chosen by the few scholars who have used a

²⁸ S. Groom (2003), pp116-130.

statistical approach to classify Qumran texts. The first example is in a PhD thesis of 1999 by B. J. Conway. Conway examined the Thanksgiving Hymns and used some rudimentary statistical tests to ascertain whether the texts could be considered homogeneous in terms of lexical criteria.²⁹ Conway is aware of Radday's stylometric approach and also of Portnoy's criticism of Radday's methods which was discussed in Chapter 2. He predefines his hypotheses to be tested in terms of comparing 'Teacher' versus 'non-Teacher' sections, comparing central core columns to those preceding and following, and comparing the columns preceding the central core with those following it. Conway applies Student-t and Mann-Whitney U tests for all these comparisons, calculating the statistics manually.

The first criterion Conway considers is that of common roots. He counts the frequencies of roots in each section to be compared and limits his analysis to those that occur at least seven times. He does not explain or justify his choice of this frequency threshold. He thus arrives at eighteen different roots that occur a total of 200 times. In his thesis he first proceeds to a qualitative description of the semantic domains associated with these roots in the text, though this is not underpinned by any specific methodology such as discourse analysis. He then calculates the number of occurrences of these roots in each hymn and asserts that because the maximum value is within three standard deviations of the mean that 'we can conclude that all the values...come from the same population'.³⁰ This is, unfortunately, a misunderstanding by Conway since, though values drawn from

²⁹ B. J. Conway (1999).

³⁰ B. J. Conway (1999), Section 4.4.

a normal distribution are likely to fall within three standard deviations of the population mean, the inference that values lying within this range are from a normal distribution cannot be asserted. For example, there may be two distinct populations with means fairly close, but each with limited variance, where a similar situation may occur when sampling from both populations. Moreover, Conway then performs formal statistical testing on root frequencies without adjusting for the different lengths of the Thanksgiving Hymns and sections being compared. However, he does adjust for the number of words later.³¹ Following this, Conway considers the frequencies of ‘action’ words in the different sections, but here he does not perform any formal statistical testing stating that, ‘A simple verbal analysis is more pertinent’.³² Finally, in terms of lexical criteria, Conway considers words that are associated with metaphor and imagery. He identifies seven that are used,

Sufficiently often to suggest that they are the basic building blocks on which the Hymnist built his poetry³³

Conway then asserts that these metaphors and images occur in clusters within the text although he does not support this with any evidence such as formal cluster analysis. Conway does draw on one computerised database during his thesis, one that records citations of and allusions to the Hebrew Bible found in the Thanksgiving Hymns. However, this database has been constructed in a loose way and supplemented by

³¹ B. J. Conway (1999), pp141-145.

³² B. J. Conway (1999), p167.

³³ B. J. Conway (1999), p180

various scholars. Conway classifies the data as: A) exact quotations; B) clear link to biblical text, more than two or three words, but not an exact quotation of the MT; C) two word phrases; D) possible allusions that could be ascribed to common Hebrew usage. To provide enough data to use Student-t and Mann-Whitney U tests, Conway combines categories A, B and C to find that the central core columns differ in biblical citation/allusion from the preceding and following columns, that is 'Teacher' versus 'non-Teacher' material.³⁴

Conway's thesis is a preliminary attempt at a statistical approach to classifying a limited number of Qumran texts. However, its lack of methodological rigour and the rudimentary statistics used are unsuitable as a foundation for more extensive and careful classifications. Moreover, it is concerned with Hebrew texts and is thus of only limited assistance in evaluating suitable lexical criteria for statistical classification of Aramaic texts.

Ian Young presents some quantitative data, though he does not formally test these statistically, in his article critiquing linguistic criteria for dating Qumran texts using Peshier Habakkuk as an example and considering the exile of the sixth century BCE as the watershed between Early Biblical Hebrew (EBH) and Late Biblical Hebrew (LBH).³⁵ First he looks at various lexical features and concludes,

³⁴ B. J. Conway (1999), pp243-246. Similarly Bobby Chum (2000) uses some simple statistical methods to examine style of the "Teacher Hymns" of 1QH.

³⁵ I. Young (2008).

Since, therefore, most LBH linguistic features are also found in EBH texts and/or are not characteristic even of all LBH texts, we can argue that the large majority, if not all, LBH features are not reliable indicators of “lateness” in a chronological sense.³⁶

Despite this statement, he then accepts that classification into EBH or LBH can be undertaken according to the proportion of linguistic features and their clustering, though he admits that conclusions will be ‘modest’. He also widens the linguistic criteria appropriate for such classification from purely lexical to morphological (e.g. proportion of hiphil compared with qal verbal forms and the use of verbal suffixes). Young proposes a criterion to assess whether what he terms an ‘accumulation’ of linguistic features has occurred to a degree that would indicate that a text was LBH. His two major linguistic features are the hiphil/qal ratio and the frequency of the preposition ‘ל rather than ’ל and other LBH features which are scored if they occur more than five times in an EBH form or at a ratio of 10:1 or greater in a 500 word text sample.³⁷ Using this criterion, he concludes that PHab aligns linguistically with EBH rather than LBH in contrast with its chronological dating. He argues that the author of PHab was steeped in EBH style and thus used this for this particular composition reflecting an underlying stylistic conservatism. This is one possible interpretation.

³⁶ *ibid*, p7.

³⁷ Young (2008) p18 argues for a 500 word sample size as opposed to the 1,000 word sample size recommended by Biber for English on a pragmatic basis. However, he does not provide any power calculations to justify this or to estimate likely Type 2 statistical error. For an extended discussion of this see Chapter 5.

Another interpretation is that the linguistic criteria chosen by Young to classify EBH and LBH are sub-optimal being limited in number and prone to the influence of potential archaizing. A more sophisticated stylistic analysis of a wider corpus of Qumran texts might suggest better alternatives. Nevertheless, Young's application of diachronic linguistic criteria for the statistical classification of Hebrew Qumran texts suggest that a similar approach in Aramaic, which has a far more extensively documented linguistic history, may be useful. Young's study also illustrates that classifications based solely on lexical criteria necessarily overlook a vast range of non-lexical information useful for textual classification, in particular morphological criteria.³⁸

Peter Coxon uses lexical criteria to date Aramaic texts in a semi-quantitative way, although not truly statistical, with an aim to decide on the dating of Daniel.³⁹ He chooses five verbal and two noun pairs with similar meanings (i.e. synonyms), but where one root is thought to be used more frequently in early Aramaic texts and the other in later texts. Of these seven pairs, he examines the occurrence at Qumran for three of them: *Sym/šwh* (to set, make), *b'h/bqr* (to ask, seek), and 'ys/'ns (man, mankind). Coxon does not provide exhaustive counts, but he notes three occurrences of *Sym* at Qumran against 14 of *šwh*. Similarly, he notes two occurrences of *b'h* versus at least four of *bqr* and more than three of 'ys versus at least fourteen of 'ns. These small numbers of

³⁸ Young & Rezetko (2009) extend the analysis using mainly lexical criteria to an extensive range of Hebrew biblical texts, but do not extend the methodology significantly.

³⁹ P. Coxon (1978).

occurrences of relatively common lemmas illustrate why a lexical approach is unlikely to be useful for statistical classification of Qumran texts.

Word length criteria

Word length represents a halfway-house between lexical and morphological criteria. In English we are familiar with the relationship between the word length and style: so-called 'plain English' encourages the use of words with fewer than three syllables, for instance. Semitic languages, unlike English, are based on three letter roots which leads Radday to note that word length in terms of letters, syllables or morphemes is of limited utility in statistical studies of Hebrew texts.⁴⁰ Nevertheless, he found this criterion useful in Isaiah and produced a ground-breaking study in 1970.⁴¹ An advantage of this criterion is that word length is more easily analysed statistically since there is no reason to expect interdependence of words in terms of syllables except for metrical poetry. A disadvantage, therefore, is that it may just detect differences in poetic style. Hence the differences in the mean number of syllables per word between Isaiah 1-39 and 40-66 that Radday attributes to different authors could equally reflect a deliberately chosen poetic style. Kenny provides several examples of how authors' styles change over time.⁴² Radday's study of Isaiah was pioneering and its statistical imperfections need to be considered within that context. However, similar techniques continue to be used.

Houk, writing in 2002, categorises words in different sections of Genesis according to

⁴⁰ Y. Radday & D. Wickmann (1975), p33.

⁴¹ Y. T. Radday (1970), p321.

⁴² A. Kenny (1986).

the number of syllables they contain.⁴³ As O’Keefe demonstrates Houk’s method and analysis is deeply flawed.⁴⁴ Similar criticisms apply to studies of repeated syllable patterns, phonemes per syllable etc. Germane to Genesis, for example, is the literary device of using poetry as heightened speech.⁴⁵ It would be perverse to argue that such short poetic sections had somehow been inserted by a different author. Houk provides a brief response to O’Keefe in his paper on Lamentations that employs similar methods,⁴⁶ but this is far from a refutation.

Syntactic length

It is not only words that can have their length measured, sentences can also be long or short indicating different styles; again, plain English requires sentences to be as brief as possible. Syntactic length is therefore another potential textual criterion which would be easy to quantify and compare using electronic texts. Neumann sums up the usefulness of syntactic length as a stylistic criterion succinctly,⁴⁷

The measurement of syntactic structure lengths, especially sentences, is frequently attempted, sometimes apparently with success, at other times with little or no success.

⁴³ C. B. Houk (2002).

⁴⁴ R. A. O’Keefe (2005).

⁴⁵ R. Alter (1981), p4.

⁴⁶ C. B. Houk (2005).

⁴⁷ K. J. Neumann (1990), p57.

A major reason why there is such variation in success is the difficulty of defining syntactic units. Sometimes these are easily defined. An example would be the acrostic psalms. Tov's *Textual Criticism of the Hebrew Bible* includes various plates of Qumran texts. Syntactic units are easily identified for the Psalms Scroll from Cave 11.⁴⁸ However, the units are less obvious for the large Isaiah scroll from Cave 1.⁴⁹ The transcription of the text includes verse numbers inserted at the appropriate points taken from the BHL. Some of these are aligned to how the Qumran text is set out (for example verse 3 follows a *pršh stwmh*), but many are not. It may be, therefore, that Qumran scribes and readers had a different syntactic understanding of various texts than the Masoretes some centuries later. This issue would be worth investigation, but it would require a massive effort to evaluate the syntactic structure for the entire Qumran corpus critically. Fortunately, electronic texts for the Aramaic Qumran corpus have been tagged to delineate syntactic units but, as yet, the search algorithms do not allow a direct count of syntactic unit length. What the search engines are well designed to do is to detect and quantify specific parts of speech, that is morphology.

⁴⁸ E. Tov (2001a), plate 8.

⁴⁹ E. Tov (2001a), plates 4-5. Odil Hannes Steck's (1998) monograph on the text divisions of 1QIsa takes the petuhot as delimiters of larger reading units and only notes setumot in supplementary material and concludes that there are two text division systems, an earlier one based on spaces and empty lines and a later one based on marginal signs.

Morphological category

As might be expected for a highly inflected language like Greek, morphological category criteria have been used widely in New Testament stylistic studies. On the other hand Radday, investigating the authorship of Hebrew texts, is less convinced of their utility.⁵⁰ Neumann includes sentence position of specific parts of speech in this criterion.⁵¹ Unfortunately, this involves decisions about where syntactic units start and finish which, as discussed above, are difficult for Qumran texts. It also illustrates the overlap between morphological and syntactic criteria, discussed below. Some parts of speech, however, are useful in delimiting syntactic units. Despite his misgivings about morphological stylistic criteria, Radday includes some of these. For example, he considers emphatic particles and subordinate conjunctions, in his study of Judges,⁵² along with less syntactically-linked features: co-ordinative conjunctions, object marker, definite article, proportion of bound pronouns to the sum of free and bound pronouns, percentages of nouns in absolute and construct states, percentages of finite verbs and non-finite/active/passive verbs, and the sum of autonomous prepositions, demonstrative pronouns and personal pronouns. Hebrew morphology is not so very different from Aramaic so, as with Young's suggestion of drawing on diachronic change to generate suitable textual classification criteria, different morphological forms deserve consideration.

⁵⁰ Y. T. Radday (1973), p136.

⁵¹ K. J. Neumann (1990), pp 61-62.

⁵² Y. T. Radday (1977).

Syntactic category

The problems defining syntactic units in Qumran texts have already been noted.

Punctuation is the commonest method used in computer-assisted textual studies to determine syntactic category. For New Testament studies these are usually determined from standard critical texts since the earliest original texts have little in the way of punctuation. This situation is similar for the Qumran texts, and notably in contrast to the detailed information available for Masoretic texts.

At the simplest level, Neumann considers word transition frequency to be a syntactic criterion.⁵³ Radday uses the frequency of this transition of one word category to another in the following word in his studies. Portnoy and Peteren criticise Radday's statistical procedures for analysing these word transitions because Radday assumes that these are independent, which they are not.⁵⁴ Portnoy and Peteren recommend the use of multivariate statistical approaches to address this issue. Word transition data also lend themselves to Markov chain-type modeling, though this would be complicated because it would not be a chain of binary states (i.e. there are more than two types of parts of speech) which is the conventional situation for Markov chain models. As a natural extension to word transition, Neumann also includes patterns of word order in this category.⁵⁵ However, there is a degree of overlap with criteria concerning sentence position of specific parts of speech that Neumann places within the morphological

⁵³ K. J. Neumann (1990), p90.

⁵⁴ S. Portnoy & D. Peteren (1984).

⁵⁵ K. J. Neumann (1990), p92.

category (see above). Patterns of word order may be more discriminatory in Greek than Hebrew texts because word order is less constrained in Greek than Hebrew so that individual authors have a greater choice of style to serve, for example, rhetorical purposes. Aramaic is less restrictive in terms of possible word order than Hebrew, so looking at the relative positions of subject, verb and object may be relatively more informative.

Jean Carmignac provides an early example of a word order analysis that includes an Aramaic Qumran document, the Genesis Apocryphon (1Q20).⁵⁶ He notes that the infinitive always precedes its object in the Sefire inscriptions, occurs before its object 15 times and after its object 8 times in the Elephantine texts, 23 times before and 28 times after in biblical texts, and 7 times before, once after in 1Q20. These data are not consistent with a purely diachronic explanation of word order. The Sefire inscriptions are from what is now modern Syria, the Elephantine papyri from Egypt, and the geographical provenance of the Aramaic portions of Daniel and Ezra uncertain, though a more Eastern influence would be expected, so some geographical factors may also be at work.

Non-grammatical variables

This is Neumann's final category of stylistic criteria. For Greek texts these generally relate to rhetorical devices both of structure (e.g. chiasmus) and sound (e.g. alliteration).

⁵⁶ J. Carmignac (1966).

Other criteria that fall within this category include the use of figures of speech, clausal structures (that overlaps with syntactical category measures) and indices of interruption. Neumann is pessimistic about the utility of these variables for authorship attribution,⁵⁷

While some of these indices ought to be tested for their value, most of them will probably not be helpful in this and other authorship questions since they result from conscious efforts to produce a certain effect.

There is a paucity of data to inform how helpful these variables are for Hebrew texts let alone ones in Aramaic. Including such untried criteria in a statistical classification based on textual characteristics would be both speculative and, even worse, may devalue the importance of more established and validated criteria.

Combined approaches

The preceding review based on Neumann's textual stylistic categories introduces the strengths and limitations of different types of stylistic features. Although some studies, such as Güting's and Mealand's⁵⁸ focus on a single stylistic criterion, most adopt a combined approach drawing on variables across a range of stylistic categories. The bulk of biblical studies that use a combined approach relate to New Testament texts.

O'Donnell provides a short review of these,⁵⁹ including Neumann's study of the Pauline letters, and concludes,

⁵⁷ K. J. Neumann (1990), p114.

⁵⁸ E. W. Güting & D. L. Mealand (1998).

⁵⁹ M. B. O'Donnell (2005), pp 85-101.

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 to understanding and applying their linguistic and contextual significance, i.e. to say that two different authors were responsible for two or more text samples. Future studies must utilize more complex, multivariate statistical processes and base their criteria and interpretation on a sound linguistic framework.

O'Donnell thus advocates a systematic theoretical basis for the *a priori* choice of stylistic variables in any combined approach. He criticises Neumann for the 'unsystematic structure' and 'highly eclectic nature' of his work and Mealand, though he approves of his statistical methods, for a lack of 'detailed justification' of his chosen stylistic features.⁶⁰ A similar accusation of eclecticism could be leveled at Radday's studies of the Hebrew bible.

The general approach adopted by O'Donnell has already been discussed in the section on semantic criteria. In his monograph *Corpus Linguistics and the Greek of the New Testament*,⁶¹ O'Donnell is able to set out his methods in detail. He emphasises the need for a representative corpus, the use of linguistic annotation, the discovery of linguistic

⁶⁰ M. B. O'Donnell (2005), p100.

⁶¹ M. B. O'Donnell (2005). This monograph is over 500 pages long and explores various applications of corpus linguistics. The theoretical approach is informative about how corpus linguistics can be applied to biblical texts though, because the texts are in Greek, the specific criteria selected are not necessarily applicable to Hebrew texts.

variables and the use of appropriate statistical methods to aid interpretation. He adopts a systemic linguistics approach following Michael Halliday.⁶²

Since Halliday rejects theoretically differentiating between spoken and formal written language (*parole* and *langue*), he does not need to impute a formal written system of Hellenistic Greek grammar. Debates and disputes surrounding formal grammar and its relationship to everyday language have also arisen with respect to Hebrew texts discovered at Qumran. Joshua Blau, for example, takes what he terms as a ‘conservative view of the language of the Dead Sea scrolls’.⁶³ Blau asserts that Qumran Hebrew

Reflects basically the latest stage of biblical (literary) language, exposed to the influence of the spoken vernaculars, viz. Aramaic and some sort of Middle Hebrew, which later crystallized as Mishnaic Hebrew.

Against this view of Qumran Hebrew being closely aligned to normative Judaeans linguistic influences, William Schniedewind suggests that it represents a reaction against normative culture, describing it as an ‘antilanguage’.⁶⁴ Schniedewind argues that Qumran Hebrew’s social function would reflect the relationship between the Qumran community and ‘normative’ Judaism that led to avoidance of Aramaic and colloquial language, classicizing tendencies, distinctive orthography and palaeography, and the use of specific code and symbolic terminology. Schniedewind hypothesises that archaic

⁶² M. A. K. Halliday (1993).

⁶³ J. Blau (2000).

⁶⁴ W. M. Schniedewind (1999).

features were introduced in response to the sect's belief that 'the pattern of language was ordained from the very creation of the world',⁶⁵ likening this to the use of 'thee' and 'thou' by the Quakers. Sue Groom contends that biblical Hebrew, itself, varies according to region and time of origin, and that Qumran Hebrew can thus be subsumed together with biblical Hebrew under the broader category of Classical Hebrew.⁶⁶ Other views include Shelomo Morag's⁶⁷ who reviews various lexical-semantic relationships to conclude,

They show that QH is an entity in itself, not an interim stage between Biblical and Mishnaic Hebrew. QH plausibly goes back to the dialectical spread of the language in the period of the First Temple, or to an early stage of the emergence of the Hebrew dialects in the post-biblical period.

Elisha Qimron also supports the contention that Qumran Hebrew represents a *spoken* dialect.⁶⁸ He asserts this because the Qumran corpus contains unique morphological forms and vocabulary,⁶⁹ though the latter is unsurprising given the extensive nature of the corpus and some of its specialized interests. Any approach, like O'Donnell's, that obviates the need to decide whether Qumran Hebrew has a written or spoken basis avoids making many contentious assumptions.

⁶⁵ W. M. Schniedewind (2000), p255.

⁶⁶ S. Groom (2003), pp10-13.

⁶⁷ S. Morag (2000).

⁶⁸ E. Qimron (2000) states 'Admittedly, most of the DSS are literary works, but their grammar should be considered as reflecting the spoken language of their scribes', p244.

⁶⁹ E. Qimron (1986), pp117-118.

Apropos Aramaic texts found at Qumran, superficially the position may seem different to that of Hebrew because Aramaic was the lingua franca of the region over several centuries. However, it is possible that written Qumran Aramaic texts were formalised and do not represent any specific spoken dialect. Moreover, if Qumran Aramaic texts were copied at Qumran, this may have led to some harmonization possibly including some influence of Hebrew forms as Hebrew texts predominate amongst those texts that have been found.

The next distinction that O'Donnell discards by using Halliday's approach is that between lexical forms and grammar. Instead the lexicogrammar of language is considered as a continuum with grammar at one end and lexeme at the other. Overlaps between Neumann's various categories of stylistic criteria (see above) indicate the usefulness of this paradigm. This is probably even more relevant to Hebrew and Aramaic than Greek, especially with regard to the verbal system where certain roots occur in some verbal patterns but not others.⁷⁰ Moreover, it is difficult to invoke any distinction between lexical form and grammatical form for some parts of speech, such as particles.

⁷⁰ M. B. O'Donnell (2005), pp 63-67 discusses the relevance of the concept of lexicogrammar to verbal patterning in Hellenistic Greek.

Also germane to a quantitative approach to textual comparison is the notion of grammar being probabilistic that O'Donnell expounds,⁷¹ though this is largely a formal underpinning of assumptions common to most stylistic studies. O'Donnell is keen, however, to claim that the observed probabilities of textual elements can inform scholars about underlying semantic aspects of the text. Whether this assertion also applies to Aramaic has not been tested statistically.

The final theme O'Donnell takes from Halliday's system is the idea of register. Register comprises field, the semantic subject of the text, tenor, the relationship between those writing and reading texts, and mode, the form or genre of the text. O'Donnell asserts that, 'Although register is a semantic concept it finds *realization* in morphological and grammatical forms, and is thus quantifiable'.⁷² As noted above in O'Donnell's work with Porter on the semantic clustering of New Testament books and, separately, chapters in Romans, he uses Biber's multidimensional register system to infer semantic information. These dimensions are: 1) informational vs involved production; 2) narrative vs non-narrative concerns; 3) explicit (situation-independent) vs situation-dependent reference; 4) overt expression of persuasion and 5) abstract vs non-abstract style.⁷³ Biber developed these dimensions following factor analysis of English texts and they have been applied to an Early Modern English corpus as well as Hellenistic Greek. However, it is unclear whether they are applicable to Qumran texts. Nevertheless, some of the

⁷¹ M. B. O'Donnell (2005), p31.

⁷² M. B. O'Donnell (2005), p33.

⁷³ M. B. O'Donnell (2005), pp 82-83.

linguistic features (e.g. noun-verb frequency, past-tense verbs) appear appropriate for Hebrew and Aramaic studies. Moreover, some features of Hebrew and, to a lesser extent, Aramaic lend themselves to measures of the narrative versus non-narrative dimension. Such measures may be helpful in testing whether any statistical textual classification is influenced by textual register and thus to genre, semantic content etc.

Biber's approach is also invoked by Donald Parry's quantitative descriptive study of the so-called 'non-biblical' Qumran texts.⁷⁴ Parry states that the object of his paper is:

To provide a corpus-based description of the nonbiblical Qumran texts by presenting a few of their linguistic structures and variants. The description is multifeature/multidimensional, examining the corpus's contents, contentive and function words, and the top-ranked lexical units (i.e., fifty top-ranked nouns, verbs, adjectives, prepositions). The paper does not examine all the linguistic forms, nor does it exhaust an examination of any particular structure.⁷⁵

Parry achieves these aims by counting tagged texts from the Dead Sea Scrolls Electronic Library which was in its final stages of preparation when the analyses were performed and thus may not represent exact final figures. Parry's concluding notes summarise the data and suggests future directions for research. He does not seek to conceptualise the data in terms of register nor does he provide any justification for the specific linguistic

⁷⁴ D. W. Parry (2006), p217.

⁷⁵ *ibid.*

variables quantified. There are two further methodological flaws that are pertinent. First, Parry accepts the non-biblical texts as a corpus which, as has been discussed, is not justified. Secondly, and more importantly, although he excludes the small number of Greek texts, he combines Hebrew and Aramaic texts in his analysis hence, implicitly, accepts the linguistic equivalence of their features. He lists Hebrew and Aramaic lexical forms separately in the high-frequency words table,⁷⁶ but he does not make any distinction for grammatical forms.⁷⁷ To justify this he would first need to show that there were no significant differences between the two languages' use of grammar in the corpus he is examining. Parry fails to do this. Doubtless, Aramaic and Hebrew are closely related Semitic languages, but they are far from identical: the Hebrew preterite is a good example of a major difference.⁷⁸ This assumption that Aramaic and Hebrew share an identical grammar also causes a problem when he comes to compare the linguistic features of the Qumran texts he has studied with the Hebrew bible taken in entirety. It is unclear how much of any difference is attributable to the considerably greater proportion of Aramaic texts in the Qumran corpus. Nevertheless, despite these problems, Parry's paper indicates that a corpus linguistic approach to the Qumran texts is feasible, though whether there is statistical power to adopt this approach to the Aramaic texts on their own is not demonstrated.

⁷⁶ *ibid*, pp230-231.

⁷⁷ *ibid*, pp228-229.

⁷⁸ See T. Muraoka & M. Rogland (1998) for a discussion of the preterite in Aramaic.

Corpus linguistic approaches

As illustrated in the previous section, using a combination of different textual or linguistic criteria to classify texts often occurs in the context of studying an entire corpus. Linguistic criteria encompass a range of analyses beyond the morpho-lexical ones I have focused on thus far. An understanding of the theoretical framework of O'Donnell's corpus linguistics studies is essential to understand how he is able to apply it to the New Testament corpus if we are to gain some idea of its potential transferability to the Qumran corpus. The areas of application O'Donnell chooses to illustrate in his monograph are:

- a) Textual criticism;
- b) Source criticism;
- c) Lexicography – words for resurrection;
- d) Semantic clustering;
- e) Causality and the Greek voice system;
- f) Discourse analysis of Jude – grammar, semantics, participants;
- g) Cohesion – Romans 1:18-28.
- h) Corpus-based discourse analysis of Philemon.

Applications f), g) and h), therefore, are not whole corpus analyses, but are small-scale examples of what might be achieved if resources were available to undertake corpus-wide studies. What is striking in this list of applications is the considerable overlap with

traditional exegetical models.⁷⁹ The principal difference is the use of computer readable texts and the extensive annotation of these that is required for computer-assisted analyses. The section on application d) is essentially a reworking of O'Donnell's Romans and New Testament semantic clustering essays that have already been discussed.

O'Donnell's approach to textual criticism follows well-established methods.⁸⁰ These will be discussed in the section on criteria for comparison with external texts (below). The major methodological issues that O'Donnell brings into sharp focus relate to electronic manuscript annotation. In many ways there is an overlap between textual criticism and source criticism because textual variants may indicate separate sources. For a single Qumran 'biblical' text, for example, some sections may be thought closer to a Septuagintal than a proto-Masoretic *Vorlage* where these are thought of as community-related sources. Radday's work on Genesis, Zechariah, Judges, Isaiah, etc. is essentially computer-assisted source criticism. Some Old Testament books lend themselves to this method. For example, the distinct LXX and MT versions of Jeremiah can be inter-woven with traditional historico-critical biblical scholarship.⁸¹ O'Donnell employs dotplot as a graphical way of displaying similarities between texts that may point to common

⁷⁹ For example G. D. Fee (2002) who stresses the social context, considers textual variants, lexicogrammar issues, has detailed flow within and beyond sentences and relates these to non-biblical texts when examining letter structure or rhetorical devices etc.

⁸⁰ For example B. M. Metzger & B. D. Ehrman (2005) who discuss the advantages of computerisation, pp 240-246 listing: collection, recording and storage of data; data presentation; statistical analyses; and hypertext possibilities. They also summarise significant ongoing projects in the area.

⁸¹ For example C. J. Sharp (2003).

sources.⁸² The examples he provides are relatively short: the healing of the leper pericope and Passion predictions from the Synoptic Gospels; comparison of 2 Peter and Jude; and a brief overview of the Didache, focusing on chapters 2 and 3. The dotplot provides a convenient way of initial identification of sections of text which might relate to common sources and require further investigation. In the Didache O'Donnell argues that certain dotplot patterns may indicate incorporation of early liturgical forms into the text.

O'Donnell's discussion of the lexicography of different words for *resurrection* is, perhaps, the first application that demonstrates the added value of using a corpus linguistics approach. He notes that translations in isolation of *εγεῖρω* are open to biases that a corpus-wide approach would militate against, stating that 'lexicographical choices are being driven by theology'.⁸³ Again, this is an example of the dynamic between lexical semantics and grammar; another less theologically-loaded lexeme may have had its passive form always translated as a passive rather than deciding that the form represents a middle verbal voice and translating with an active form. Computerised corpora are well-suited to producing concordances listing individual lexemes and the frequencies of their morphological forms. However, like Fee in his monograph on exegesis,⁸⁴ O'Donnell asserts that lexical data extend beyond the word under consideration. The words surrounding the word being considered, the collocation, are

⁸² M. B. O'Donnell (2005), pp 297-312.

⁸³ M. B. O'Donnell (2005), p324.

⁸⁴ G. D. Fee (2002).

relevant data that are also easily investigated using electronic means. A major methodological issue when using collocation is deciding on limits and this may involve determining the limits of a clause in unpunctuated texts. For New Testament texts O'Donnell finds that a four word window on either side of the keyword, ignoring clause and verse boundaries, to be the most useful for semantic-based collocational analysis.⁸⁵ It is unclear whether this window is applicable to Qumran texts. O'Donnell also applies the four word window to grammatical collocation. Greek is morphologically richer than Hebrew so four words may not capture as much grammatical information if applied to Qumran texts. Parunak argues that measuring the distance in words between a fixed number of occurrences is statistically preferable to a fixed test window.⁸⁶ He applies cluster analysis techniques to these after suitable data transformation. He also claims to apply time series analyses across the text to investigate cluster patterns further, but provides no details of the method used.

O'Donnell's next area of application, which follows on from his collocational analysis of words for resurrection, is that on causality and the Greek voice system. This is quite distant from corpus linguistic issues that might apply to Qumran Aramaic texts.

As noted above, applications f), g) and h) are small scale rather than corpus-wide with two of them illustrating discourse analyses requiring extensively annotated electronic

⁸⁵ M. B. O'Donnell (2005), p346. Nor is it clear that this window is applicable to non-NT Greek texts such as LXX texts. However, systematically derived semantic domains are unavailable for the LXX.

⁸⁶ H. V. D Parunak (1984).

texts that are not available at present for the Qumran corpus. However, the application of corpus linguistic methods to measuring cohesion is something that may be of relevance to Qumran texts. O'Donnell examines cohesion in Romans 1:18-28. The use, or deliberate lack of use, of conjunctions as cohesive devices has already been mentioned in the work of Güting and Mealand's investigating asyndeton in Paul.⁸⁷ The deliberate use of asyndeton as a rhetorical device is an illustration of the distinction O'Donnell is keen to point out between cohesion and coherence.⁸⁸ Moreover, conjunctions can acquire different meanings in different contexts: O'Donnell gives the example of the word 'and' as used in the nursery rhyme 'Jack fell down and broke his crown'. This breadth of semantic range that pertains to certain conjunctions is of relevance to Aramaic, especially the conjunction *waw*. Nevertheless, whatever the semantic nuance of a particular conjunction, they act as cohesive devices marking flow within a text. O'Donnell provides a diagram of cohesive devices placed along the spectrum of lexicogrammar with conjunctions being placed in the centre with lexical cohesion (clusters of lexis) at one end and reference (substitution and ellipsis) at the other.⁸⁹ O'Donnell suggests that one way of examining cohesion in substantial pieces of text is to look at participant-referent chains. He considers that this method may be applied successfully to non-narrative texts. Practically, collocation analysis centred upon a participant's name as the keyword may provide a useful approach. Extensively annotated texts that are able to identify specific participants from pronouns can provide a

⁸⁷ E. W. Güting & D. L. Mealand (1998).

⁸⁸ M. B. O'Donnell (2005), p426ff.

⁸⁹ M. B. O'Donnell (2005), p431.

more complete picture. At present, such texts are not available electronically for Qumran.

Some time has been spent examining O'Donnell's work because it provides a systematic corpus linguistic approach to biblical texts, albeit from the New Testament, that suggest methodologies that might be usefully applied to Qumran texts and, equally importantly, methodologies that are either flawed or not applicable and therefore should not be considered.

O'Donnell's is not the only voice advocating a corpus linguistic combined approach to textual comparison of biblical texts, but he provides the most worked-out examples.

Vincent de Caën, in his polemic paper *Hebrew Linguistics and Biblical Criticism: a Minimalist Approach*, also argues strongly for corpus-based, non-theological analyses.⁹⁰

He examines the entire Hebrew bible, but without using a combination of different linguistic variables or formal statistical testing, in his paper on pausal forms of the second person singular independent pronouns.⁹¹ de Caën does, however, recognise that his study would have been improved if he had been able to take into account statistical associations preceding the complementisers *ky* and *'m*.

⁹⁰ V. de Caën (2001).

⁹¹ V. de Caën (2005).

Kirk Lowery also adopts a linguistic approach to a biblical Hebrew text, Judges.⁹² He, too, provides largely a descriptive computer-aided statistical account without any formal statistical testing. Although his index of cohesion, that relies on pronoun counts, together with other structural indices such as major/minor participants, foreground/background sentences, prove disappointing, Lowery makes valid cautionary points about the imposition of Indo-European syntactic categories on Semitic languages. Surprisingly, Lowery makes no reference to Radday's work on Judges published eight years previously, which may reflect the fragmented nature of early computer-assisted linguistic research on the Hebrew bible. Robert Holmstedt also demonstrates advantages of a linguistic approach to the Hebrew bible corpus, though not computer-assisted, in his thesis on the relative clause.⁹³ He follows a Chomskyan minimalist programme approach, advocated by de Caën, to show limitations of conventional grammatical approaches. In many ways Holmstedt simplifies the conceptualisation of the relative clause in biblical Hebrew, at the same time arguing that a synchronic rather than diachronic view can be taken for the relative particle 'šr. Identification of an easily identifiable and relatively common textual element with a limited semantic range and which is unaffected by the period when the text was composed appears to provide a useful criterion for any statistical analysis of Hebrew texts, even though Holmstedt does not pursue this in his thesis. The position for the Aramaic equivalent, *dy*, is a little more complex semantically but, nevertheless, could be a promising criterion for textual classification.

⁹² K. E. Lowery (1985).

⁹³ R. D. Holmstedt (2002).

Karen Masterson combines a number of linguistic and statistical approaches in her analysis of Semitic texts.⁹⁴ First she examines the combinatorial structure of Hebrew roots found in Genesis and Syriac roots from the New Testament Peshitto. She compares observed frequencies of single letters found in combination at I,II and I,III and II,III positions within each 3-letter root with frequencies expected if such combinations were random. Unsurprisingly many Hebrew letter combinations occur more frequently or less frequently than expected by chance, though Masterson does not test this formally using, for example, χ^2 tests.

Masterson's next approach builds on Lowery's study of Judges.⁹⁵ She uses information theory to examine grammatical forms within a clause rather than letters within a root. Here the grammatical parts of speech were considered over a span of six words so that the degree of constraint a particular part of speech exercises over the next word (1,2), the word after that (1,3) etc. to the sixth word in the string was calculated to give a coefficient of cohesion. Masterson shows that chapter 5 of Judges, a poetic section, has lower coefficients of cohesion at (1,2) and (1,3) and higher coefficients of cohesion at (1,4), (1,5) and (1,6) than the other prose chapters. She infers that this is due to word order being more flexible in Hebrew poetry. What is clear, however, is that the coefficients of cohesion are fairly similar for these relatively large blocks of text and that a particular part of speech has little constraining effect beyond two words (1,3) in

⁹⁴ K. J. Masterson (1994).

⁹⁵ K. E. Lowery (1985).

Judges. Masterson then extends this approach to the remaining books of the Hebrew bible. She uses a different classification of morphological forms that was available for the Westminster Hebrew Morphological Database. She calculated the first order entropy together with the conditional entropy and the coefficient of constraint for adjacent parts of speech (1,2) to those six words apart (1,7) averaged for each book. As Masterson explains,⁹⁶

The first order entropy calculates the entropy or amount of information taking into account the relative frequencies of the grammatical types but not the effects of neighboring grammatical types.

This measure, therefore, is similar to those used in many statistical analyses of morphological forms that assume independence of data. The conditional entropy, by contrast, takes into account the effect of the index word's part of speech on the dependent word's part of speech. For example, if the index word is an adjective, conditional entropy is a measure of the likelihood of the next word in a (1,2) bigram being a particular part of speech, such as a noun, given the overall relative frequencies for different parts of speech in the book. The coefficient of constraint is the complement of conditional entropy and represents a measure of redundancy. Redundancy correlates with predictability and thus is protective against copying errors.

A key feature of these entropy and constraint measures is that entropy tends to decrease, and constraint increase, with longer texts. One way to correct for this that Masterson

⁹⁶ K. J. Masterson (1994), pp103-104.

notes, though does not apply, is to select the same size sample of text from different biblical books to adjust for bias due to book length. Nevertheless, what is germane to deciding on criteria for comparison for the Qumran corpus is that coefficients of restraint are very similar from (1,3) to (1,7) and that differences between books are quite small. The highest (1,2) coefficient of constraint is 0.18 for Psalms and the lowest 0.09 for 1Kings. Since the coefficient is measured on a logarithmic scale, this represents an eight-fold difference, though the length of Psalms compared with 1Kings may account for a proportion of this difference. It is interesting that a poetic book, like Psalms, is more constrained grammatically than a largely prose narrative book like 1Kings. This suggests that Masterson's findings for Judges may not hold throughout the Hebrew bible corpus. Moreover, twenty-two of the thirty-nine books have coefficients 0.11-0.13, representing only a 60% difference in redundancy. There is thus a trade-off between selecting samples of text that are large enough to reduce the variance of entropy/redundancy measures for the mean of the sample against selecting samples so large that they fail to catch important heterogeneity of constraint on word order within a book. Nevertheless, this approach is superior to one based solely on an assumption that all parts of speech are independent as Masterson shows in her analysis of this measure by book where differences between books are far less than those detected by the coefficient of constraint.⁹⁷

⁹⁷ K. J. Masterson (1994), pp106-108.

Masterson's next analysis is also germane to Qumran texts: she compares the redundancy of vocalised and non-vocalised text of Genesis.⁹⁸ In these calculations bias due to the number of words is not a factor, but because the vocalised text has many more characters, bias due to the number of letters is important. Hence the first order entropy by word for the non-vocalised text was 8.37, substantially lower than that for the vocalised text which was 9.63. Again this measure is on a logarithmic scale so that the difference is over an order of magnitude with the vocalised text carrying more information. However, when the same calculations are made per letter, the situation is reversed with the vocalised text having 20% more redundancy than the non-vocalised. This is because first the vocalised text is longer and secondly there are 38 possible 'letters' in the vocalised text compared with 22 in the non-vocalised text. The extra redundancy represents the error-correcting effect introduced by vocalising the text of Genesis. A similar consideration is necessary in any similar analyses of Qumran texts that may have fuller orthography than non-vocalised counterparts.

Masterson's final, and by far the most lengthy, analysis concerns the collocation of specific lexemes in the Pentateuch.⁹⁹ In essence this addresses the number of idioms or stereotypic phrases in different texts where words occur together more frequently than expected by chance. Unsurprisingly, her analyses demonstrate widespread occurrence of such collocations in the Pentateuch. However, the cultural context within which

⁹⁸ K. J. Masterson (1994), pp109-116.

⁹⁹ K. J. Masterson (1994), pp117-373.

language is used may influence semantic domains,¹⁰⁰ something which is difficult to take account of in the case of Aramaic texts found at Qumran. Further complications are introduced when rhetorical devices, such as ‘artificial Hebrew’ are present.¹⁰¹

Masterson’s work, then, is of undoubted relevance if the task was to classify the longer, Hebrew texts found at Qumran, but most of the Qumran Aramaic texts are too short for the analyses she uses so, with some reluctance, they must be set aside and left for the far more extensive task of classifying the Qumran Hebrew corpus. Nevertheless, her work and that of other scholars engaged in linguistic analyses, does illustrate that a corpus-based approach is feasible as opposed to a detailed, but piecemeal text-by-text analysis. That is, instead of examining each Aramaic scroll on its own and comparing it to other texts, a corpus of Qumran Aramaic scrolls could be considered together if appropriate common textual (and/or linguistic) features can be identified. The disadvantage of such an approach is that if a specific scroll has some highly idiosyncratic textual characteristics, these would not feature in the overall classification which would focus more on what textual features scrolls had in common.

Criteria for comparison with non-Qumran texts

As already noted, opportunities for direct comparison of the Aramaic texts found at Qumran with similar texts found elsewhere are extremely limited. This is not the case

¹⁰⁰ C. H. J. van de Merwe (2006).

¹⁰¹ R. Holmstedt (2006).

for Hebrew texts and historically, comparison with non-Qumran texts such as Masoretic Text or Samaritan Pentateuch text types has drawn on established textual criticism criteria. Although these comparisons are not directly relevant to textual classification methods suitable for the Aramaic texts, they do provide valuable insights into the limitations of statistical approaches that have already been applied to Qumran texts.

For textual comparisons, Tov provides the standard resource that is relevant to Qumran texts.¹⁰² He classifies textual characteristics in terms of orthography, morphology, contextual adaptations and scribal practices before proceeding to consider variants. These textual characteristics can be considered as related to the stylistic criteria considered in the previous section. Tov notes that Qumran texts should be compared with evidence of all other relevant texts though, in practice, comparison is often only with the MT, SP and LXX.¹⁰³ Tov notes that including the textual characteristics as criteria to determine variants results in many variants between Qumran texts and the MT (specifically the BHL), though different Qumran fragments will tend to vary from the MT in terms of one type of textual characteristic (e.g. morphology) whilst others will differ in terms of another textual characteristic (e.g. shorter theophoric names).

In a paper written before all the Qumran texts were published and prior to his more nuanced classification, Tov provided quantitative data for a number of features that he

¹⁰² E. Tov (2001a).

¹⁰³ E. Tov (2001a), p112.

considers differentiate ‘sectarian’ Qumran scrolls and ‘biblical scrolls.’¹⁰⁴ Cross and Saley’s previously discussed *A Statistical Analysis of the Textual Character of 4QSamuel^A (4Q51)*¹⁰⁵ is an example of this text-critical approach. The lack of explicit criteria to determine ‘superior’ readings has already been noted. However, further examination of their methodology is germane to this section.

Cross and Saley argue that corrupt or inferior readings can be considered the text’s ‘bad genes’ and that these ‘bad genes’ are the best criteria for determining which family a specific text belongs to. They consider that ‘superior’ readings, here ones that agree with the Old Greek text rather than the MT, also play a part in determining which family a text may relate to. Cross and Saley consider some ‘superior’ readings to arise because of MT scribal *parablepsis* and these can comprise whole paragraphs. Other such extra material they determine to be ‘inferior’ because they are ‘idiosyncratic’. The concepts of ‘genes’ and ‘families’ seem like an attractive metaphor, but does this approach hold up methodologically?

First, the Hebrew alphabet is far more complex than the base pairs that make up DNA and RNA. The Hebrew alphabet has twenty-two letters as opposed to the four nucleic acids that comprise DNA. Accordingly, the number of words, or indeed three letter roots, far exceeds the number of amino acids that can be coded. Genes are thus composed of a large number of base pairs to provide a sequence of amino acids that

¹⁰⁴ E. Tov (1986).

¹⁰⁵ F. M. Cross & R. J. Saley (2006).

code for the protein gene product. To talk of 'bad genes', then, is a confused terminology. There are several ways genetic variation occurs, but most commonly this occurs secondary to single nucleotide polymorphisms (SNPs). This is where a specific point in the sequence can have one of two alternative amino acids present. Such SNP polymorphisms may or may not affect the functionality of the gene. This is because only part of the gene, termed the *exon*, codes for the protein gene product. Other parts of the gene, termed the *intron*, may act as gene regulators or as redundant material in case other sections of the gene are lost. Typically redundant DNA occurs at the end of chromosomes, termed telomeres, since these parts are at particular risk of being lost during DNA replication. Loss of DNA may thus not be of any importance (cf text *parablepsis*), may result in a shift of reading frame for amino acid coding or complete loss of gene translation. In terms of analogy with textual variants, the question as to whether SNP polymorphisms are counted as equivalent to a word, a morpheme or a letter arise. Moreover, there may be debate as to whether a variant affects an exon-equivalent (i.e. affects what the text is expressing) or an intron-equivalent (has no effect on the text's meaning). Thus, attractive as the gene analogy appears superficially, it can actually make the situation more confused rather than clearer, so should be used carefully. Perhaps, avoiding such analogies and describing these variants as 'Leitfehler', 'corrupt readings', 'indicative errors' or 'errores significativi' would be preferable.

To return to statistical studies comparing Qumran texts, Vegas Montaner examined the relationship between another Qumran text, 1QpHab, and old Greek biblical versions in

an early computer-assisted study.¹⁰⁶ He provides computer-generated tables of variants from 1QpHab in Habakkuk citations in the commentary itself, the Hebrew text of the Murabba'at 88 manuscript, MT, LXX manuscript families, the Greek Dodekapropheton scroll from Nahal Hever and Hexaplaric readings. Vegas Montaner proposes relationships between texts by inspection of counts rather than with any formal statistical tests with statements such as 'it seems evident that' and 'is obviously out of the question'.¹⁰⁷ Still he concludes that

The traditional way of classifying the Qumran documents by trying to fit them into diverse categories of textual types (in our case, Masoretic or Septuagintal) becomes here unadequated (sic)¹⁰⁸

Lim also presents enumerations of Qumran textual variants comparing pesharim materials with the MT.¹⁰⁹ His categorisations are more concise into orthographic differences and/or substantive variants, though he admits that, 'Subjective judgement no doubt plays a role'. Such subjectivity is not an aspect of genetic studies as noted above and Lim makes no use of this analogy. However, there are similarities between genetic and textual variants that suggest that statistical and informatics methods used in genetic studies may be of use in textual comparison. One similarity is that just as the genome is distributed across a number of chromosomes, so texts may comprise distinct sections (e.g. biblical books). Some gross genetic abnormalities occur at the chromosomal level (e.g. trisomy 21 or chromosome 14:21 translocations in Down's syndrome) with the rest

¹⁰⁶ L. Vegas Montaner (1989).

¹⁰⁷ *ibid*, p112.

¹⁰⁸ *ibid*, p318.

¹⁰⁹ T. Lim (1997), p72ff.

of the chromosomes remaining intact. Lim's pesharim variants provide an example of the importance of taking into account the location of variants, at least at this level.

Lim provides lists of variants for six Qumran Isaiah fragments, three Psalms fragments, two fragments from Hosea and Zephaniah, and one fragment from Micah, Nahum and Habakkuk.¹¹⁰ Lim is interested in whether these variants arise from *Vorlagen* that differ from the MT or represent deliberate exegetical alteration of texts. If we assume that Psalms, Isaiah and Minor Prophet *Vorlagen* for the pesharim are on different scrolls, but that the Minor Prophets *Vorlage* is a single scroll, as seems likely from the surviving Qumran texts then, by analogy, these *Vorlagen* can be thought of as different chromosomes. The individual Minor Prophet books can be thought of as genes on a single chromosome. We can first compare 'chromosomes' (i.e. Isaiah, Psalms and the Minor Prophets) using univariate general linear modeling to find that, after adjusting for the number of words in each section, there is no significant difference in variant rate between them ($F=1.22$, $p=.33$). However, when we compare 'genes' (i.e. each book), there is a significant difference ($F=6.71$, $p=.009$, partial $\eta^2=.83$).¹¹¹ The estimated marginal means (i.e. variants adjusted for number of words) with 95% confidence

¹¹⁰ T. Lim (1997), p90 summarised in Table 8 in terms of variants per number of words.

¹¹¹ Some introduction to statistical tests was presented in the previous chapter. The P-value represents probability; here $p=0.009$ means that this difference could have occurred by chance 9 in 1,000 times. It is thus taken as significant at the conventional significance level of $p=0.05$. The F-value is the raw statistic calculated by an analysis of variance statistical test which here represents the amount of variability in variant rates that occurs between books on the same scroll. This value is not very informative on its own, but can be used to produce partial η^2 which is a measure ranging between zero and one representing the proportion of variability explained.

intervals are shown in Table 3.1. Formally, all the confidence intervals overlap except those for Habbakuk. This suggests that the higher variant rate in Habbakuk cannot be attributed to a different *Vorlage*, but to subsequent alteration, be that deliberate (exegetical) or accidental (scribal error). Note that it is the text section (i.e. book), not the individual variants, that can be thought of as a ‘gene’.

book	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Isaiah	24.6	1.35	21.4	27.7
Hosea	24.7	2.39	19.2	30.2
Micah	22.8	3.38	15.0	30.6
Nahum	36.2	3.46	28.2	44.1
Habbakuk	56.1	5.00	44.6	67.5
Zephaniah	24.4	2.53	18.6	30.3
Psalms	24.1	1.92	19.7	28.5

Table 3.1 Mean number of variants compared with the BHL with 95% confidence intervals for each pesharim taken from Lim (1997).

Such studies of variants can be extended beyond the pesharim. For the Minor Prophets Vegas Montaner provides an extensive list of variants from a wide range of Qumran texts.¹¹²

At the furthest point on this excursus, the analogy between textual comparison and genetic studies does extend to the lower level of individual variants which can be considered as the equivalent of single nucleotide polymorphisms (SNPs). A good example of this is manuscript profiling used extensively in the International Greek New

¹¹² L. Vegas Montaner (1980).

Testament Project (IGNTP).¹¹³ Manuscript profiling is based on specifying specific points in texts where significant discriminatory variants occur in known texts. This avoids detailed comparison of the entire text. Similar methods are used in genetic analyses where high density SNP analyses have been performed for the whole genome by the HapMap project.¹¹⁴ The authors state:

These data document the generality of recombination hotspots, a block-like structure of linkage disequilibrium and low haplotype diversity, leading to substantial correlations of SNPs with many of their neighbours. We show how the HapMap resource can guide the design and analysis of genetic association studies, shed light on structural variation and recombination, and identify loci that may have been subject to natural selection during human evolution.

An equivalent analysis for Hebrew textual variants would be very useful in determining sampling strategies and thus exact criteria for textual comparison for Qumran fragments. This raises the question as to how to define such ‘hotspots’. Should this be, for example, within the family of manuscripts that comprises the Masoretic Text? Or should this be against more distinct families of texts such as the Samaritan Pentateuch and the presumed *Vorlagen* of Old Greek texts? The answer to this question depends on which external texts we wish to compare the Qumran texts with. Pragmatically, the variants in different MT manuscripts may well have arisen some time after the Qumran texts were written rendering any comparisons inappropriate. Thus for the purposes of a computer-aided analysis, the MT can be considered as a single text with the BHS most readily

¹¹³ E. J. Epp & G. D. Fee (1993) summarise this and other theoretical approaches to the New Testament.

¹¹⁴ The International HapMap Consortium (2005).

available in a suitably tagged/marked up format. Pragmatic considerations also dictate that the Samaritan Pentateuchal and Septuagintal texts chosen represent whole families of manuscripts with many variants. However, for the Greek texts two clear text types are extant for portions of the Old Testament. A choice arises as to whether hotspots for one Old Greek text type should be defined against the MT or against the alternative Old Greek text type if these do not coincide. In view of this, two separate profiles based on distinct hotspots may need to be generated.

This brief excursus has demonstrated that statistical analyses can be applied to compare Qumran texts with texts found elsewhere. Although comparison of textual variants is not possible for the Aramaic texts found at Qumran, it illustrates a proof of principle that Qumran texts can be situated within a wider linguistic context using statistical methods.

External comparisons with non-Hebrew texts

As a preliminary, a brief discussion of comparing Qumran Hebrew texts with non-Qumran texts in other languages will provide some continuity with the previous section. Here the issue of translation technique needs to be addressed. Benjamin Wright has shown that the translation technique for a fair proportion of books is fairly tightly rule-based, allowing comparisons between Qumran texts and presumed Hebrew *Vorlagen*.¹¹⁵ Furthermore, some translation rules are more variable than others. Choosing hotspots

¹¹⁵ B. J. Wright (1989) takes an interlinear approach to LXX translation rather than taking it as Greek literature in its own right such as in the La Bible d'Alexandrie collection.

that are represented by more invariable Hebrew-Greek equivalents reduces the probability that any variant is due to translation technique. Concordance statistics can facilitate this choice. Any lemma need to be of sufficient frequency to provide an adequate sampling density for text profiling otherwise significant textual differences may be missed because of type 2 statistical error. For example, if we examine word frequencies in the LXX version of the book of Habbakuk, since the Qumran pesharim text appears to vary significantly from the MT, the commonest lemma, excluding the definite article, is *καί* with 96 occurrences. The next most frequent are *αὐτοῦ* (*j* (n=60) and *ἐν* (n=33) out of a total of 585 words and 1105 occurrences. Hence, comparison of an even relatively frequent lemma may only sample around 3% of any text. Prepositions such as *ἐν* may not be the optimal rule-based choice. Wright provides a list of frequencies of *ἐν* compared with other prepositions as a putative Semitic translation feature that supports the hypothesis that the choice of prepositions is influenced by translation technique.¹¹⁶ Text profiling as a method to compare Qumran texts with external sources may thus have limited applicability to shorter fragments where there are few external texts with which to make a comparison.

Moving on to texts of relevance to the Aramaic scrolls found at Qumran, there is the Enochic literature which may have had an authoritative status on a par with the so-called ‘biblical’ texts at Qumran; Psalm 151 and the Testament of Levi are other examples of ‘authoritative’ texts. Indeed, the survival of translations argues, *de facto*, for the

¹¹⁶ B.J. Wright (1985).

importance attributed to these ‘non-biblical’ texts prior to the canon being fixed.

Gabriele Boccaccini argues that a diachronic understanding of Enochic influences at Qumran is essential to the classification of Qumran texts.¹¹⁷ Mark-up texts suitable for computer-assisted analysis are becoming increasingly available thanks to scholars such as Ian Scott. For example critical editions of Greek texts of 1 Enoch together with its equivalent Qumran fragments are accessible.¹¹⁸ Critical editions of Greek texts of Jubilees are also available online.¹¹⁹ Unfortunately, Qumran fragments of these books are of limited extent, making statistical comparison with external texts problematic, though the same could be said of some books of the Hebrew bible such as Obadiah. Moreover, translation technique issues with such limited original language resources available overshadow any decisions about potential variants.¹²⁰ Larson has made a careful study comparing the surviving Aramaic Enochic texts found at Qumran and the Greek Enochic corpus.¹²¹ He notes that though the translation technique is similar to the LXX, there are passages that have a more literary nature with the presence of words having connections with Greek mythology.

¹¹⁷ G. Boccaccini (2006), p37 states that ‘The essential problem consists in finding the correct criteria to classify the [Qumran] material’. However, his classification would imply a category of texts not found at Qumran, specifically the later Enochic literature. He cites the Testaments of the Twelve Patriarchs in support of this hypothesis, despite 4Q538 Testament of Judah, 4Q539 Apocryphon of Joseph B and 4Q215 Testament of Naphtali all being identified within the Qumran corpus.

¹¹⁸ K. M. Penner & I. W. Scott (2006). The online critical pseudepigrapha project also allows display of critical variants with an interactive critical apparatus.

¹¹⁹ D. M. Miller & K. M. Penner (2006).

¹²⁰ J. R. Davila (2005) provides an extensive discussion of the relationship between Greek pseudepigraphic texts and their putative Hebrew or Aramaic *Vorlagen*.

¹²¹ E. Larson (2000).

In summary, the lack of equivalent texts in Aramaic and the difficulties posed by inferring the original Aramaic from translations render external comparisons on the basis of textual variants non-feasible at present. Therefore, if external comparisons are to be made, they will need to be done on the basis of textual features that reflect what might be termed style. Ideally, such stylistic criteria should be common enough at Qumran and in other Aramaic texts to facilitate classification. That is, any textual variables should occur with adequate frequency both within the Qumran corpus and in other Aramaic texts to allow adequately-powered statistical analyses. There may be some textual criteria peculiar to the Aramaic texts found at Qumran and not featured in non-Qumran texts. These would be helpful for an internal classification of Qumran Aramaic texts but, in themselves, would not allow any wider conclusions to be made about the affinities of Qumran texts to others beyond Qumran. These considerations are essential when reviewing potential Aramaic stylistic criteria on which to base any textual classification.

4. Aramaic textual criteria

Although the vast majority of texts found at Qumran are in Hebrew, a number of those that are in Aramaic, such as the Enochic fragments discussed above, allow external comparison with a wider range of epigraphic and textual sources.¹ Such comparisons have been undertaken for many years for the Aramaic sections of the biblical books of Ezra and Daniel.² Historico-critical approaches have driven studies to date the text of Daniel in particular and have resulted in a classification in terms of diachronic ‘dialects’.³ Gerhard Hasel reviews the long history of the debate around dating the Aramaic sections of Daniel.⁴ He draws on Qumran Aramaic documents in his own dating. First, he cites evidence for a first century BCE date of 1Q20, the Genesis Apocryphon, and then draws upon other scholarship which suggests the Aramaic in Daniel is from an earlier date than 1Q20. Next he refers to several articles that place 11Q10, the Targum of Job, earlier than 1Q20, perhaps the second half of the second century BC, but later than Daniel. Although Hasel is focusing on Daniel, he implicitly recognizes that there are different diachronic Aramaic ‘dialects’ within the Qumran corpus itself. However, he does not provide any criteria within his article that allow these dialects to be distinguished from each other.

¹ I have, in fact, already touched on one study by Peter Coxon (1978) above when discussing lexical criteria.

² H. H. Rowley (1929).

³ J. A. Fitzmyer (1979) provides the most widely used diachronic classification: Old Aramaic (900 - 700 BCE), Official Aramaic (700 - 300 BCE), Middle Aramaic (300 BCE – 200 CE), Late Aramaic (200 – 700 CE), Modern Aramaic (700 CE – the present).

⁴ G. F. Hasel (1981).

Hasel's datings are not universally accepted. For example, Ursula Schattner-Rieser dates both 1Q20 and 11Q10 at the turn of the era, whilst dating others such as the Four Kingdoms texts (e.g. 4Q553), to the Hellenistic period and "Proto-Esther", 4Q550a-d, in the Persian period.⁵ Disputes over dating require clear criteria for their resolution. Collating a consensus about which textual criteria are subject to diachronic change is therefore an essential first step.

Hasel also asserts that the Aramaic of Daniel is of Eastern origin and this opens up another dimension for classification. Indeed, some scholars use a theoretical diachronic development of the language to explain differences in regional 'dialects'.⁶ Such classifications for Qumran texts have been the subject of criticism by Michael Wise⁷ whose arguments are relevant to a proposed statistical classification. Wise opposes linguistic analysis as a dating method for Aramaic texts found at Qumran. He argues that the evidence points to the Aramaic texts being written by many hands. He considers that some of the fragments represent personal copies and that the scribes responsible would not necessarily have the Aramaic they were copying as their first language. He contends that this constitutes a form of *diglossia*. Hence, inconsistent textual variants may result

⁵ U. Schattner-Rieser (2004), p25. Similarly, she points out that palaeography can be influenced by Hebrew, citing an instance when the word for God is spelt with paleo-Hebrew characters in 4Q243, pseudo-Daniel. She concludes that palaeographic dating is of limited value, pp30-31 and is useful only in conjunction with other dating criteria.

⁶ E. M. Cook (1992).

⁷ M. O. Wise (1992).

from scribal errors rather than represent transitional phases between dialects. Statistical methods are ideal to test Wise's hypothesis. Although scribal errors will be non-random (i.e. some forms are more likely than others to be incorrectly copied), if this is the result of 'personal' copies, classification of texts will find significantly more of these variants in such fragments rather than those that appear more 'official'. It is thus still worthwhile attempting to define suitable textual variables for linguistic statistical analyses on Aramaic texts.

Wise's theory is far from being unopposed. Michael Sokoloff launches a strong rebuttal in a footnote to an article titled, 'Qumran Aramaic in Relation to the Aramaic Dialects'.⁸ Nevertheless, it is arguable that texts in Aramaic found at Qumran were not necessarily produced at Qumran, and thus to speak of Qumran Aramaic as some homogeneous entity is questionable. Statistical tests are able to determine to what extent the Qumran Aramaic corpus can be considered as representing a single 'dialect'. To apply such texts, the first step is to identify suitable textual criteria. From the previous discussion in this chapter, the focus will necessarily be on lexical, morphological and syntactic criteria since these are those that have both a relevant body of scholarship and can be reliably extracted from current electronic texts.

⁸ M. Sokoloff (2000) footnote 13 first rejects Cook's methodology then rebuts three of Wise's pieces of linguistic evidence. However, he fails to deal with the principal point of scribal errors being mistaken for dialectical shifts.

In the general classification of Aramaic, Qumran Aramaic is considered as a single entity often combined with nearby texts and labeled as “Dead Sea Scrolls”.⁹

Chronologically it is designated as Middle Aramaic along with other ‘dialects’, the major ones being Nabatean, Palmyrene, Hatran and Edessan. Since some of the Aramaic scrolls found at Qumran may have been brought to Qumran from elsewhere or be copies of older texts, it is not possible to assign a *terminus a quo* to these manuscripts, but for Qumran the *terminus ad quem* is c.70 CE. Nabatean ranges from the first half of the second century BCE to the fourth century CE after which it becomes gradually subsumed by Arabic, the first dated Palmyrene inscription is from 44 BCE and the latest 279/80 CE, dated inscriptions for Hatra range from 44 BCE to 238 CE, and Edessan (Old Syriac) can be dated from 6 BCE to 243 CE. In summary, the other Middle Aramaic forms of Aramaic are generally later than the Aramaic found at Qumran and, according to Fitzmyer’s chronological classification (see above), stray into the Late Aramaic epoch.

With regard to geography, the Nabatean kingdom was centered around Petra in modern day Jordan about 100km south of Qumran, Palmyra in modern day Syria, north east of Damascus, Hatra is in modern north west Iraq, south west of Mosul, and Edessa is in the south east of modern day Turkey. Hence Edessan and Hatran are the most eastern geographically, Nabatean and Qumran most western. In view of this considerable chronological and geographical extent it is, perhaps, unsurprising that Qumran Aramaic is treated as a monolithic entity.

⁹ For example S. E. Fassberg (2008).

Chronology and geography are two dimensions of classification that Aaron Koller proposes for the Qumran Aramaic texts.¹⁰ To these he adds two further dimensions of genre and ideology. He only provides one linguistic criterion for genre, the choice of direct object marker and argues that a prefixed *l* indicates a literary composition. Unfortunately, this is also a criterion Koller considers to be typical of Eastern texts. He also notes that the usage may just reflect register within a text rather than the overall genre.¹¹ Koller's fourth linguistic dimension is titled 'ideological'. However, he fails to specify the criteria that relate to this dimension, although he alludes to the possibility of an archaizing tendency in using the verbal system. Again, such criteria would overlap with those of the chronological dimension. Thus, there is an emerging recognition amongst scholars that the Aramaic found at Qumran is far from homogeneous, and that chronological and geographical dimensions may be inadequate to explain this homogeneity, but there is, as yet, no agreement as to how many extra dimensions there are and what they might be. Again, this is a task ideally suited to statistical analyses.

As noted in Chapter 2, statistical analyses are not only helpful in establishing which differences between texts are unlikely to be due to chance, and thus facilitate testing for textual heterogeneity, but they also are able to establish which associations are unlikely to have happened by chance. This provides for some measure of homogeneity of the Aramaic corpus at Qumran.

¹⁰ A. Koller (2011).

¹¹ Ibid., p211.

Lexical criteria

In his study of the Aramaic of Daniel, Zdravko Stefanovic uses some elementary statistics to correlate the Aramaic section of the MT of Daniel with several Old Aramaic inscriptions. He does not perform formal statistical tests, but does report percentages of common word-roots.¹² He reports percentages of common word-roots that range from 56% with the Hadad and Panammu inscriptions to as high as 90% with the Bir-Hadad inscription. He concludes that the Aramaic of Daniel contains a significant amount of material that is similar to Old Aramaic texts, yet also concludes that Old Aramaic, itself, is far from uniform and may comprise three or four dialects.¹³ But it is the data from the Tell Fakhriyah inscription that is most telling with regard to the validity of this criterion. Of the 95 word-roots that are common to Daniel and the inscription, only 20 are found in both parts, 38 are found in part 1 of the inscription only and 37 only in part 2. Hence the Aramaic of Daniel is equally close, and arguably closer, to both parts of the inscription than they are to each other in terms of shared vocabulary.¹⁴ Shared vocabulary, therefore, does not seem likely to be sensitive enough for textual classification at Qumran where, to complicate things, there is considerable variation in text length. It is,

¹² Z. Stefanovic (1992), pp47-61.

¹³ *ibid*, p108.

¹⁴ The first part of the inscription is around 20% shorter than the second part, but even adjusting for this the rates are not significantly different. The word length of the Aramaic section of Daniel also has to be taken into account when making this comparison, but vocabulary richness tends to be predicted better by the square root of the length of text rather than the text length itself. In the case of Daniel, we do not know whether to adjust by text length, square root of text length or some other value which makes statistical analyses less than robust.

perhaps, more suitable to *post hoc* characterisation of any different text types identified by other criteria.

An alternative lexical approach is to identify words that are unique in Aramaic texts found at Qumran and seek to correlate them with external cognates. This might allow a classification of texts in terms of various ‘influences’. Greenfield and Sokoloff provided such a list at a time, 1992, when not all the Aramaic material was yet published.¹⁵ Correlations include Hebraisms, Middle Persian, Egyptian and Greek. The difficulty is determining to what extent these influences represent the original compositions and how much later scribal alterations. Franz Rosenthal suggests that loan words may help trace the historical development of Aramaic, but also indicates that certain languages are associated with particular spheres; for example, he suggests that Greek loan words often belong to the cultural sphere.¹⁶ Nevertheless, if some of the Aramaic texts were composed at Qumran, it is perhaps these that might show the greatest influence from Hebrew which was the usual language for sacred writings and may have been the spoken vernacular as well.¹⁷ However, it may be easier to identify such Hebraisms by morphological than by lexical means. This is, in part, because the influence is far from unidirectional. For example, the lemma *raz*, is a term for mystery in both Hebrew and Aramaic and is thought to have first been borrowed by Aramaic from Persian and later

¹⁵ J. C. Greenfield & M. Sokoloff (1992) are informed by their view that the texts were composed at Qumran. The relationships they identify might fit easier with the Qumran Aramaic texts comprising a collection of texts, not all originating at Qumran.

¹⁶ F. Rosenthal (1983), p57-59.

¹⁷ S.E. Fassberg (1992).

adopted by Hebrew.¹⁸ In addition, as Schattner-Rieser points out in her overview of the grammar of Aramaic at Qumran, some of the texts, such as 11Q10 the Targum of Job, are based on Hebrew originals and read more like translations.¹⁹

At a more detailed level, Jan Joosten examined the changing semantic use of the preposition *qdm* from Biblical texts through Qumran texts to the Targums.²⁰ Such word studies are difficult to translate into statistical hypotheses not only because they require semantic judgements, but also because the frequency of any single lexeme is usually too low to provide adequate statistical power to test hypotheses. However, a statistical classification of the Aramaic texts at Qumran would provide a helpful framework to inform such word studies. The statistical limitations posed by the study of any single lexeme extend, naturally, to orthography as a potential stylistic criterion. In any case, Muraoka notes in his *Grammar of Qumran Aramaic*, that spelling conventions are of no use as criteria for dating of Qumran Aramaic texts.²¹

¹⁸ S. I. Thomas (2009), Appendix I. A comprehensive picture can be gained in the case of Akkadian influence where there is considerable overlap between Biblical Aramaic and Biblical Hebrew lemmas thought to be influenced as detailed by H. ben Yosef Tawil (2009).

¹⁹ U. Schattner-Rieser (2004), p29. In this section she also notes that it is more appropriate to speak about the Aramaic at Qumran rather than the Aramaic of Qumran since the Aramaic at Qumran is not textually homogeneous. Émile Puech (2011) provides a highly critical ten page book review of this *Grammar* published seven years after its appearance. He points out that the grammar was written before all the Aramaic texts had been officially published and that the thesis, on which the monograph is based, was submitted six years earlier in 1998.

²⁰ J. Joosten (2010).

²¹ T. Muraoka (2011), p5.

Morphological criteria

Unlike the situation for Greek and Hebrew texts, there are no sophisticated statistical stylometric studies on morphological aspects of Qumran or Biblical Aramaic.

Morphological stylistic criteria that pertain to Hebrew cannot be blindly assumed to apply to Aramaic despite the similarity of the languages.²² However, there is a considerable body of literature that proposes morphological criteria that may be useful for classification of Aramaic texts. As discussed above, morphological criteria may encompass some lexical, syntactic and semantic aspects, but those scholarly publications that primarily focus on morphological variables will be considered in this section.

Cook provides a table of what he describes as diagnostic features for differentiating between the different dialects he proposes: Hatran, Edessene, Palmyrene, Qumran and Nabatean.²³ These are the common Middle Aramaic dialects and, as noted above, most overlap but Hatran, Edessene, Palmyrene and Nabatean are generally a little later than

²² Here Michel Foucault's critique in *The Order of Things* (2002), is illuminating. First, in Chapter 4 he outlines the rise of the 'new epistemological domain that the Classical age called "general grammar", which promoted similar classifications across languages in terms of articulations (morphosyntax) and order with Hebrew taken as the most archaic in the new taxonomy of languages. Then, in Chapter 8, he notes the early nineteenth century developments where the primacy of roots for classifying languages was displaced by the 'grammatical totality' (p306) which allows comparison between languages without needing to refer to any representational element; the inflectional system (morphology) is a major component of this comparison so that 'words are characterized in the first place by their morphology and by the totality of the mutations each of their sounds is capable of undergoing' (p321). Foucault considers that we are still in an era of formalisation, hence the continuing focus on morphology as a criterion for linguistic classification.

²³ E. M. Cook (1992), p8.

Qumran. Hence there is a degree of chronological as well as geographical diversity. He lists eleven essentially morphological variants:

1. Relative pronoun: *dy/d-*
2. Male plural pronominal suffix: *-hwn/-hwm*
3. Demonstrative plural pronoun: *hlyn/'ln/'lh*
4. Male singular demonstrative pronoun: *hdyn/hn'/dnh/dn*
5. Male plural independent pronoun: *hnw/hnwn/'nwn/hmwm/'nw*
6. Presence of direct object marker *yat*
7. Infinitive of derived stems; *m-w/'-h/*
8. Male plural nominal ending in emphatic state: *-e/-ayya*
9. Imperfect prefix: *l-/y-/n-*
10. Masculine singular possessive suffix on masculine plural nouns: *-yhy/-why*
11. Attested jussive.

Cook produced this list at a time when not all the Aramaic texts found at Qumran had been published. What he considered the norm for Qumran Aramaic may need revision. Nevertheless, an initial examination of the distribution of these variants in Qumran Aramaic texts will help to inform which of them are worth pursuing in formal analyses (e.g. if a feature is constant for all Qumran texts, it can be discounted). Another pronominal suffix variant *-kh* may also be worth exploring despite its relative low

frequency.²⁴ The same may be the case for the third person singular feminine pronominal suffix *-h'/hh/-h* variant suggested by Ursula Schattner-Rieser.²⁵

L. Diez Merino's study of the adverb in Qumran Aramaic suggests that this part of speech may be useful for classification because its ending may take several forms.²⁶ Again, initial statistical exploration should help decide whether this feature is worth pursuing. Michael Sokoloff seeks to develop and refine Cook's work on relating the Aramaic texts found at Qumran to a range of dialects.²⁷ Sokoloff proposes yet another classification of Aramaic texts, perhaps an indication of the lack of consensus as to how these might be classified. He proposes five categories: biblical Aramaic; texts in Jewish Literary Aramaic (in which he includes Qumran texts); legal documents; Judean Aramaic; and Nabatean Aramaic. The classification is primarily based on where texts were discovered: Sokoloff does not take into account the possibility that texts found at any particular location may have originated from a variety of regions and therefore not represent a homogeneous dialect. He then further describes Jewish Literary Aramaic diachronically as a transitional stage between Official Aramaic and what he terms Middle Western Aramaic. He outlines eight morphological features that he considers

²⁴ E. Qimron (1992), unlike Cook, had access to unpublished Aramaic texts at Qumran, but still found only 38 occurrences of this form.

²⁵ U. Schattner-Rieser (2000).

²⁶ L. Diez Merino (1992) provides a clearly written summary of this part of speech that is not easily defined.

²⁷ M. Sokoloff (2000).

allow comparison with other Aramaic dialects and thus may be helpful for text classification. These are:

- 1 Independent third person feminine plural forms of verbs and pronouns;
- 2 The prefix *l-* in third male singular and third person plural forms of *hwy* - this relates to Cook's ninth criterion;
- 3 The second male singular suffix *-kh*: this was proposed by Qimron (see above);
- 4 The third feminine singular pronominal suffix *-h'* as proposed also by Schattner-Rieser (see above);
- 5 The second masculine singular perfect ending *-th, -t'*;
- 6 The second feminine singular pronominal suffix *-ky*;
- 7 *(h)afel* forms;
- 8 *(h)itpeel* forms.

It is unclear how useful these criteria will be for statistical analysis because Sokoloff considers his third criterion, *-kh* endings, to be common, though Qimron could find only 38 occurrences. 'Common' for textual scholars may not equate to 'common' for the purposes of statistical tests. Cook, himself, provides an updated list of morphological criteria.²⁸ These comprise:

²⁸ E. M. Cook (1998).

- 1 Independent pronoun first person singular 'n'/'nh;
- 2 Independent pronoun first person plural 'nhn'/'nhnh;
- 3 Independent pronoun second person masculine plural 'ntn'/'ntwn;
- 4 Independent pronoun third person masculine plural *hmwn'/'nwn*;
- 5 Suffixed third person plural masculine personal pronoun *hwm/hwn*;
- 6 Demonstrative pronoun 'near' masculine singular *dnh/dn*;
- 7 Demonstrative pronoun 'near' feminine singular *hd'/'dh/d*';
- 8 Demonstrative pronoun 'near' plural 'lyn'/'ln;
- 9 Relative pronoun *zy/d-/dy*;
- 10 Perfect verbs third person plural masculine ending -w'/'-w;
- 11 Perfect verbs second person singular masculine ending -th/'-t'/'-t;
- 12 Perfect verbs second person plural masculine ending -twn/tn;
- 13 Perfect verbs first person plural ending -nh/'-n';
- 14 *Pael* infinitive presence/absence of preformative *m*;
- 15 (*h*)*afel* forms;
- 16 *itpeel* and *itpaal* preformative *ht'/'t*.

Others, such as Jonathan Choi, have used some of these morphological features in an attempt to relate the Aramaic MT text of Daniel to non-biblical Aramaic texts.²⁹

Stefanovic,³⁰ also in relating the Aramaic of Daniel to non-biblical inscriptions, highlights the following morphological variants, several of which are suggested by other scholars (see above):

²⁹ J. Choi (1994).

³⁰ Z. Stefanovic (1992), p80ff.

- 1 The precative with *l*;
- 2 The demonstrative pronoun *z't*;
- 3 Retention of aleph with addition of pronominal suffixes;
- 4 Preformative *l* with jussive precative;
- 5 Prefix *m-* for *pe'al* infinitives;
- 6 Absent *-n* on second and third masculine plural jussive precative;
- 7 Syncope of *h* in the causative imperfect;
- 8 Spelling of *'yt/yt*;³¹
- 9 Prefixed *h-* on various verbal forms;
- 10 Forms of the relative pronoun;
- 11 Forms of the masculine demonstrative pronoun;
- 12 Presence of emphatic state nouns.

Steven Fassberg has made a series of suggestions of morphological criteria that are distinctive and relevant to Qumran. In 2004 he suggested the second male singular suffix *-kh* as the most distinct feature, considering other features such as final nasalization and derived stems with *m-* prefixes as rare.³² However, by 2010 he identified features of the verbal system that displayed diachronic change and also differed whether the text was literary or vernacular.³³ These are:

³¹ M. L. Folmer provides a detailed review of the *nota obiecti* in Aramaic later than Qumran, but does not cover the Qumran texts themselves.

³² S.E. Fassberg (2004).

³³ S.E. Fassberg (2010).

- 1 Imperfect *hwh* prefixed by *l-*;
- 2 Perfect verbs second person singular masculine ending *-t'/-t*;
- 3 A few *plene* spellings of the Peal imperfect and infinitive with [o] rather than [a];
- 4 Infinitives of derived stems with prefix *m-*;
- 5 Perfect verbs third person plural masculine ending [-u] or [-un];
- 6 Perfect verbs third person plural feminine ending in an aleph .

Again, these, with the exception of the rare *yiqtol/yiqtal* and *miqtol/miqtal* variants, have been covered by previous scholars.

Holger Gzella in a paper focusing on just a few case studies, primarily of 1Q20 (Genesis Apocryphon) and 11Q10 (Targum of Job), proposes some textual criteria for dating the Qumran Aramaic texts themselves.³⁴ First he notes the difficulty of this given the paucity of references within the corpus to dateable external events³⁵ and that palaeography can only provide a *terminus ad quem*.³⁶ He concludes that 'For more conclusive results, one therefore still depends on an interpretation of the development of linguistic features'.³⁷ He notes that in general linguistic studies assume that Aramaic

³⁴ H. Gzella (2009).

³⁵ *ibid*, p62.

³⁶ *ibid*, p63.

³⁷ *ibid*, p63.

develops in a linear way³⁸ and that the Aramaic corpus from Qumran does not form a ‘linguistic unity’.³⁹ As he considers dating the texts he notes that all existing studies interpret the ‘linguistic facts’ historically, geographically or socially.⁴⁰ With regard to geography, he considers this as either Eastern or Western. With regard to social classification, he proposes the idea of what he calls different ‘registers’, ‘standard versus ‘sub-standard’.⁴¹ He toys with the hypothesis of Aramaic *diglossia* which he conceives of in terms of a spoken vernacular and written literary form.⁴²

Although Gzella considers what he calls ‘inconsistent spellings’, most of these can be considered morphological (e.g. *dy* versus *d*): as noted in Chapter 3, distinctions between different linguistic categories are fuzzy. His morphological criteria overlap with those proposed above. He does not produce a table of counts nor perform any statistical analyses, but uses his experience of the texts to conclude that linguistically 11Q10 precedes 1Q20. Though limited, Gzella’s study represents an important development in this area of scholarship because it is the first to focus on textual features to date Qumran texts without reference to other Aramaic corpora.

Ursula Schattner-Rieser takes Gzella’s approach one step further. She has proposed a more comprehensive set of ‘archaic’ features that she believes will facilitate dating of

³⁸ *ibid*, p63.

³⁹ *ibid*, p64.

⁴⁰ *ibid*, pp65-66.

⁴¹ *ibid*, p75.

⁴² *ibid*, p77.

the Aramaic texts at Qumran on linguistic grounds.⁴³ She notes phonological shifts that predate Qumran reflected in orthography are relatively rare except for *zayin – daleth* shift ‘généralement dans les pronoms’,⁴⁴ and the *he - aleph* shift affecting various verbal stems. She also makes the point that some uncommon orthographical peculiarities are more ‘Hebraisms’ than ‘archaisms’,⁴⁵ and thus are unsuitable for dating texts. The predominant criteria she applies to date texts are based on archaic forms of pronouns:

- 1 *hmwn, hy, hu;*
- 2 *’lh, za, dnh/dn’;*
- 3 *dkn, dk;*
- 4 *zy.*

In addition, like Fassberg and others, she notes the prefix *l-* for some forms of *hwh*, and the use of *’l* retained as an archaic particle of negation with a jussive construction rather than the usual *l’* found in later Aramaic. Schattner-Rieser concludes that most of the Aramaic texts are written in an Aramaic style later than Official (Imperial) Aramaic tending towards a Targumic style. She notes, however, that the Book of Enoch and Aramaic Levi texts contain distinct archaic textual features and that the Tobit texts also have several relatively archaic forms.⁴⁶

⁴³ U. Schattner-Rieser (2010).

⁴⁴ *ibid*, p103. Schattner-Rieser designates *zy/dy* as a relative pronoun in her grammar 2004, noting that this is the only pronoun where an archaic form is found at Qumran (p35).

⁴⁵ *ibid*, p106.

⁴⁶ *ibid*, p118.

Schattner-Rieser's paper was discussed by some leading Qumran Aramaic scholars: Steven Fassberg, Moshe Bernstein, Daniel Stökl, John Collins, Hugo Antonissen, Florentino Garcia Martinez and Émile Puech. The discussion raised some important methodological points.⁴⁷

Fassberg considered that her use of features from texts prior to Qumran to facilitate dating of Qumran texts is convincing and 'holds promise as a reliable yardstick for better determining the date of the compositions'.⁴⁸ However, he raised the question of how one can determine the difference between a genuinely archaic feature and one that is intentionally 'archaizing', especially since archaizing occurs in some Hebrew texts of this period, including some found at Qumran. He also wondered about the dating of the Aramaic sections of Daniel and Ezra found at Qumran, something not examined by Schattner-Rieser in her paper. Bernstein raised another methodological point: whether some texts, such as the Genesis Apocryphon, are composites and so require more than one date? Stökl wondered whether the findings could be due to two contemporaneous dialects co-existing, one with more archaic tendencies. Puech was concerned that Schattner-Rieser's methods did not allow her to date the texts, but rather the date of the copying of the texts present at Qumran. He was also concerned that archaic and non-archaic forms of pronouns occur side-by-side in several texts and noted that in

⁴⁷ I will not discuss minor points relating to individual forms that some of the scholars took exception to, either their inclusion or exclusion.

⁴⁸ *ibid*, p119.

archaeology it was usual to date a stratum by the most recent artefacts found in it, not by the earliest.

Taking the final point first, Schattner-Rieser responded by asserting that her aim was not to propose a precise dating. She did not rebut Puech because his argument relies on a false analogy, linguistic features cannot be compared with dateable artefacts such as coins, because she, herself, describes the texts in terms of strata.⁴⁹ However, Puech's point carries some force: simply relying on archaic features for dating is inadequate: a rigorous approach would require consideration of the balance between archaic and later features in a text. Statistical analyses provide such rigour.

Relying solely on archaic features also runs the danger of the method succumbing to Fassberg's archaizing criticism. Schattner-Rieser responds to this by giving the example from the Enoch palimpsest where *zy* has been corrected to *dy* deliberately during copying to avoid an archaic form presumably present in the *Vorlage*. However, it is unsafe to generalise from a single example, particularly as many scribes were involved with copying texts and others may have had the opposite inclination. On the other hand, it would be a considerable feat to archaize all the features of a text, so using a combination of textual criteria may provide some protection against drawing incorrect inferences. This is where studies that rely on a single feature are likely to be less reliable.

⁴⁹ *ibid*, p118.

In the published discussion, Schattner-Rieser does not comment on the dating of the Aramaic sections of Daniel and Ezra at Qumran. As seen above, there has been considerable interest in dating the Masoretic Text versions of these biblical sections. In fact, the versions at Qumran represent by far the earliest dateable versions of these texts because the Masoretic text is based on manuscripts many centuries later. 4Q117 is relatively short and has two fragments in Aramaic. There are only four differences with the Masoretic text, two of which relate to verbal singular/plural forms where the Masoretic text is grammatically correct. The two other variants relate to *heh – aleph* alternate endings: *nhr'* in 4Q117 for MT Ezra 4:10 *nhrh* and *mdynt'* in 4Q117 for MT Ezra 6:2 *mdynth*. Interestingly, in Ezra 4:11 both MT and 4Q117 have *nhrh*. Now, according to Schattner-Rieser's criteria, this would make 4Q117 later than the MT as *aleph* replaced *heh* as a determined state marker. Moreover, this cannot be a question of Hebrew influence since Hebrew marks the determined state of nouns very differently. The two principal explanations are that the Masoretic scribes have deliberately archaized the *Vorlage* changing *aleph* to *heh*, matching Ezra 4:11 or that the 4Q117 copyist introduced a later textual feature in 4:10 but not in 4:11. There is no way of knowing which of these explanations actually happened, though the second would be consistent with much that is known about the copying of texts, both *nhr'* and *mdynt'* being fairly common nouns. There are eight fragments of Daniel that were found at Qumran, three of which are in Aramaic and more than a few words: 4Q112 (4QDan^a), 4Q113 (4QDan^b) and 4Q115 (4QDan^d). Eugene Ulrich lists and reviews variants that differ from the

MT.⁵⁰ There are forty-three variants listed for the Aramaic section of 4Q112, sixteen for 4Q113 and eleven for 4Q115. Of these, there are nine *heh* – *aleph* differences for 4Q112, two for 4Q113 and one for 4Q115. *Aleph* is present in the Qumran Daniel texts where the MT has *heh* in nine of these twelve variants. In the other three, the MT has *aleph* and the Qumran text *heh*. For this single feature, then, on balance by Schattner-Rieser's criteria, the Qumran texts would be dated later than the MT, but it is noteworthy that there are instances where the trend is in the opposite direction. Schattner-Rieser considers that this orthographical shift was possible because both *heh* and *aleph* became quiescent in the final position in a word, in which case we are looking at scribal preference uninformed by the pronunciation of the word in a period where the orthography was not fixed. We only have a proportion of the Aramaic text of Daniel available from Qumran so a nine – three *aleph* – *heh* balance may just reflect chance survival of specific verses, but if it is representative, it would imply that the texts are probably from a similar period with the Qumran text being a little later. This would be consistent with Ulrich's positioning of both texts within the same early 'edition'.⁵¹ Essentially then, this method does not provide so much a chronological date for Qumran texts, but rather a textual stylistic dating: the MT in these terms has an Aramaic style much older than the chronological date of the extant copies.

Some caution is required, however, when interpreting such morphological data.

Kutscher provides an instructive insight from Samaritan Aramaic with regard to the

⁵⁰ E. Ulrich (2001) does not provide an exhaustive list, but judges which are textually important: there are an adequate number for the purposes of discussion.

⁵¹ Ibid, p 582.

phenomenon of guttural weakening.⁵² His analysis of the Samaritan Targum to Genesis 43 shows that none of the six words where *het* is replaced by *ayin* appear in the Pentateuch whereas of the fifteen words where this weakening does not occur, eleven are found in the Pentateuch and a further three are very common in Aramaic. He concludes that orthography of Pentateuchal words was preserved even if pronunciation had changed. Hence, a mixture of morphological features may represent specific archaic orthographic influences. This is particularly germane to texts such as 11Q10, the Targum of Job, which have a Hebrew antecedent.

Returning to the discussion of her paper, Schattner-Rieser responds to Bernstein's point about the homogeneity of the Genesis Apocryphon by saying that it is 'clearly written in a unified language and surely by one and the same person'.⁵³ In fact, the Genesis Apocryphon is one of the few Aramaic texts at Qumran where it would be possible to treat different sections separately and still retain adequate statistical power. However, this is not an enterprise for an initial statistical classification, but rather a future secondary analysis. The question of homogeneity overlaps with both Puech's archaeological argument, in that a text composed of material from different *Vorlagen* needs to be dated as a text by the most recent features, perhaps those that provide some unification of the disparate parts, and also Fassberg's concern about relying solely on archaic features to date texts.

⁵² E. Y. Kutscher (1976), pp75-76. Kutscher's monograph perhaps typifies the state of studies in this field some time ago when specific Aramaic dialects were mooted based on evidence from one or two texts with scholars disagreeing among themselves as to which text was 'best'.

⁵³ Schattner-Rieser (2010), p122.

Stökl based his two dialect argument on an example from German dialects. Schattner-Rieser responded by pointing out that in an archaic dialect archaic features are found throughout whereas in a text like the Genesis Apocryphon the presence of archaic features ‘seems to have been forgotten by the modern copyist’.⁵⁴ This assertion weakens her argument for using archaic features to date texts because the method depends on the assiduousness of scribes. In other words, given two scribes with different tendencies to retain archaic features, the same *Vorlage* could give rise to two copies dated quite differently according to Schattner-Rieser’s methods. One way to guard against this is to avoid giving too much weight to just a few archaic features, but to try to have a wide range of textual criteria for dating purposes.

This problem of how much Schattner-Rieser’s method dates the *Vorlagen* compared with the actual Qumran text is at the heart of Puech’s first point. In response Schattner-Rieser pointed out that her dating is not precise, but she clearly is more interested in determining a *terminus a quo* than a *terminus ad quem*: she refers to textual features at Qumran that are close to those found at the Jewish colony at Elephantine in Egypt in the fifth century BCE. Many of the Aramaic texts at Qumran relate to ‘biblical’ figures and episodes, especially from Genesis, so archaic sources may have influenced texts implicitly if not explicitly. Hence, this is probably a question that requires consideration after a classification has been produced to determine how robust any diachronic dimension of that classification might be.

⁵⁴ *ibid* p122.

Subsequent to both Schattner-Rieser's Aramaic Grammar and her article, Takamitsu Muraoka published his extensive Grammar of Qumran Aramaic.⁵⁵ He defines Qumran Aramaic in an "inclusive" fashion to include texts from a wide region of the Judean desert.⁵⁶ Nevertheless, he recognises the considerable geographical and diachronic diversity of the texts included in this definition and notes that some texts are more 'colloquial' than others. His argument for this inclusive approach is that, "one could hardly say that such and such a feature is unique for this genre, time or location, and could not have occurred in any other genre, any other time or anywhere else".⁵⁷ The idea of tendencies towards particular uses which he implies supports the use of statistical methods in textual classification.

Much of Muraoka's Grammar consists of lists of instances illustrating particular linguistic features; he does not set out to characterise the lexical and morphological diversity within the corpus he is describing. Nevertheless, he does comment on some features along the way and when he does, his comments are usually in agreement with the observations of other scholars as set out above. Examples include the diachronic aspects of the relative pronoun,⁵⁸ forms of the disjunctive personal pronouns,⁵⁹

⁵⁵ T. Muraoka (2011).

⁵⁶ *ibid*, pXXV.

⁵⁷ *ibid*, pXXVIII. Here he contrasts his position with that of Schattner-Rieser whom he states holds more fixed views of when and where particular linguistic elements were used.

⁵⁸ *ibid*, p50, though not invariantly p193.

⁵⁹ *ibid*, p38.

conjunctive pronouns,⁶⁰ demonstrative pronouns,⁶¹ determined state noun forms,⁶² *m*-prefixed non-G infinitives,⁶³ haphel forms,⁶⁴ and hitpeel forms.⁶⁵ He also notes the difficulties of determining which forms are jussive⁶⁶ and argues for an Eastern origin of 11Q10.⁶⁷

Finally, some scholars have adapted morphological stylistic criteria found to be useful in Hebrew biblical texts for use with biblical Aramaic texts.⁶⁸ However, not only may such criteria not be applicable (see the introduction to this section), but the relative brevity of the Aramaic sections of Ezra and Daniel is somewhat limiting for formal statistical testing of criteria that either require lengthy text segments or rely on infrequent features.⁶⁹

⁶⁰ *ibid*, p43.

⁶¹ *ibid*, p47.

⁶² *ibid*, p53.

⁶³ *ibid*, p104.

⁶⁴ *ibid*, pp109-110 noting that these may be 'bibliacizing'.

⁶⁵ *ibid*, pp111-112.

⁶⁶ *ibid*, p101.

⁶⁷ *ibid*, p160 and p242.

⁶⁸ M. B. Shepherd (2006) relates the different forms of Aramaic verbs to different dimensions of text register. J. C. Kesterson (1984) studied verbal usage in the Serakim and the Damascus document finding it mostly paralleled biblical Hebrew usage.

⁶⁹ Together these sections comprise 1947 common nouns, 295 proper nouns, 76 pronouns, 317 adjectives, 400 verbs and 863 particles so that for criteria based on pronouns only, for example, these sections may lack statistical power.

The above discussion indicates that the interpretation of the morphological criteria is far from simplistic; nevertheless, scholars provide a good number of possible morphological criteria that they consider suitable for classification of Aramaic found at Qumran, and there is a fair degree of agreement between them. Some criteria, taken on their own, have to be considered with caution because there is risk of a degree of circularity in characterising features of Aramaic based on *a priori* conclusions about the dating of certain texts. Moreover, these selections are eclectic rather than systematic so that their utility needs to be established formally. A major question that arises at this point, and which will also be of relevance to analyses, is whether to treat these criteria all equally. One reasonable approach would be to eliminate those criteria which fail to discriminate between Qumran Aramaic texts, but for those criteria which are discriminatory the position is less clear. Conventionally, statistical analyses would treat all criteria equally unless there was some *a priori* belief that some were more important than others. It would be possible to decide that one scholar's recommendations had greater authority than another's and so give preference to his or her choices. Another possibility would be to prioritise those criteria that are selected by several scholars. However, this is a narrow, albeit evolving, field of scholarship, so that the evidence to give greater weight to some criteria over others *a priori* is insubstantial. Decisions about weighting can be left until preliminary statistical analyses have been performed and provide objective information for guidance.

Syntactic criteria

Knowledge of the syntax is important in any language, and in Aramaic where the syntax in particular reflects the history of the language most faithfully, it is of crucial significance. I have taken cognizance of this fact by *not* writing a special discussion of Biblical Aramaic syntax, since I know that few beginning students of a language ever take the trouble of reading the portion of the grammar dealing with syntax.⁷⁰

Syntactic criteria, which necessarily overlap with morphological and semantic features, may prove helpful in classifying Aramaic texts. Scholars have proposed several syntactic variants that they consider help to differentiate between dialects. Prominent among syntactic features that might be considered is the position of the verb with respect to subject and object. Takamitsu Muraoka provides an extensive list of 184 verbs with their preferred position with respect to the object and other parts of speech in his article on verbal rection in Qumran Aramaic.⁷¹ His study examines far more than verbal position. It expounds the various methods by which a verb may be linked to its complement in the Aramaic texts found at Qumran that were available, dealing with verbal suffixes, prepositions, the direct object marker *yt*, and the ‘datival’ and ‘accusativ’ uses of *lamed*. These connecting features may be useful for statistical study with verbs that occur frequently enough to provide adequate power for statistical testing. Verbal rection is also promising for comparison with external texts. For example, Muraoka provides an overview of what he terms the morphosyntax of the infinitive in

⁷⁰ F. Rosenthal (1983), p1.

⁷¹ T. Muraoka (1992).

Targumic Aramaic that examines associated prepositions.⁷² Similarly, Kaddari examines the position of the existential verb *hwh* in the Elephantine papyri and Biblical Aramaic.⁷³

For purposes of comparison with external texts, and also as possible criteria for internal classification, Stefanovic details several syntactical features that he considers significant in defining the relationship between the Aramaic of Daniel with a range of Old Aramaic inscriptions.⁷⁴ The key features can be summarised as:

- 1 Stereotypical formulae often paralleling those in other languages;
- 2 Construct chain versus *zy* constructions;
- 3 Verbal position, especially of the infinitive;
- 4 Shifts between first person and third person report.

A further pair of syntactic features where a choice can be made, just as with construct chains versus the use of noun followed by a relative particle, is use of the passive participle versus the use of a *t*-stem verb. Both these options allow for expression of a range of non-active voice. There is a wide variety of opinion about what contributes to this choice in individual texts,⁷⁵ but one possibility is that syntactic textual style makes a contribution and so is a candidate for statistical analysis. However, as Na'ama Pat-El

⁷² T. Muraoka (1983).

⁷³ M. Z. Kaddari (1983). A brief article, but with 70 occurrences of *hwh* in Biblical Aramaic, promising for formal statistical testing.

⁷⁴ Z. Stefanovic (1992), pp95-107.

⁷⁵ T. Li (2008) reviews these opinions in a study of the non-active participle in the Aramaic sections of Daniel. T. Muraoka (2011), p172 considers that for participles the passive represents a perfective aspect whereas the *t*-stem represents the imperfective aspect.

notes in her discussion on the use of *zy/dy* as a syntactic marker of the historical development of Aramaic, 'syntax is very sensitive to influence by other languages':⁷⁶ again, this is relevant to those Aramaic Qumran texts that have close links with Hebrew texts.

Muraoka makes a case for the *l* versus *yt* object marker forming another alternative pair with *l* being associated with Eastern texts and *yt* with Western and Official Aramaic.⁷⁷ Unfortunately, as he himself notes, usage depends on the specific verb used⁷⁸ so this limits opportunities to compare texts to relatively few instances. This illustrates how much more syntactic criteria depend on semantic underpinnings than morphological criteria so that conclusions will necessarily be more tentative.

Summary

Morphological and syntactic features appear more promising as criteria for the internal classification of Aramaic texts found at Qumran and comparison with external standard texts than lexical criteria. From previous studies there are over 20 morphological features that are worth exploring, but fewer than ten syntactic features. Combining these features is likely to improve discriminatory power: a single feature may not be that useful, but the concurrence of different features may well be. The dividing lines between lexical, morphological and syntactic features is somewhat blurred, and this may well

⁷⁶ N. Pat-El (2008), p56.

⁷⁷ T. Muraoka (2011), p217.

⁷⁸ *ibid*, p214.

result from the classical approach of trying to impose some universal grammar on Aramaic. For the purposes of electronic texts, a pragmatic categorization can be based on those features which can be evaluated at the word level versus those that require the context of the phrase or sentence. For practical purposes, the former can be considered lexical or morphological, the latter syntactic. Since lexical criteria appear to be the least promising for textual classification of Aramaic texts at Qumran, it is sensible to retain these for *post hoc* characterisation of any text types that might emerge from a statistical classification based on morphological and syntactic criteria.

10. Confirmatory Models of Qumran Text Types

Hierarchical cluster analysis is an exploratory statistical method used, as with the Qumran Aramaic scrolls, where there are no *a priori* assumptions about relationships between cases. Although hierarchical cluster analysis provides information about statistical differences (formally distances) between individual scrolls which can be displayed as a dendrogram, it does not, of itself, indicate specific clusters. These have to be inferred by inspection. In view of this, hierarchical cluster analysis cannot provide any information as to how well these inferred clusters describe the relationship between cases (the scrolls). Secondary analyses using analysis of variance is required to determine what statistical differences exist between the different clusters. To limit the arbitrary nature of inferring clusters by inspection, the exploratory analyses in the previous three chapters used three techniques:

- 1) Hierarchical cluster analysis was performed using raw and standardized values of the variables (morphological or syntactic criteria), and only those clusters which emerged in both were retained.
- 2) Clusters that emerged on morphological criteria were compared with those produced by the more limited syntactic criteria.
- 3) Analysis of variance was performed to test for differences between clusters to check that these differed significantly.

The morphological analyses proceeded in two stages. First, clustering was confined to the Qumran Aramaic corpus alone. Secondly, analyses were extended to include known external text types. The first stage resulted in six clusters, but analysis of variance suggested that two of these were not significantly different from each other and could therefore be combined as a single text type. These clusters were: 4Q208, 4Q209, 4Q213a – Enochic (E) and 4Q213b, 4Q539, 5Q15 – Aramaic Levi, Apocryphon of Joseph B, New Jerusalem (A). The second stage suggested that 4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547 – Tobit, Visions of Amram (T) might also include 4Q538 (Testament of Judah), 4Q550c (proto-Esther), 4Q570 (Aramaic R) and 6Q8 (Giants) because of all these scrolls show morphological similarity to Babatha Archive Aramaic texts. Hence there might be a maximum of eight different text types or a minimum of six different text types clearly identified within the Qumran Aramaic corpus. Deciding exactly how many different clearly identifiable text types are present is essential before a probabilistic classification of the remaining Qumran Aramaic scrolls can be undertaken. Confirmatory models using Latent Class Analysis (LCA) are able to provide the necessary information to make this decision.

Latent Class Analysis with known classes

As detailed in Chapter 6, Statistical Methods, LCA is a specific type of finite mixture modeling. The fundamental hypothesis underpinning LCA is that latent classes explain the relationships among the observed dependent variables. Usually LCA is applied when the number of classes is unknown and thus parallels the more familiar statistical

technique of exploratory factor analysis where the number of factors is unknown; exploratory factor analyses have already been encountered in Chapter 7. By contrast, confirmatory LCA can specify known classes, paralleling confirmatory factor analyses where the number of latent factors and their relationship to observed variables is pre-specified. The purpose of confirmatory LCA is to test which pre-specified model fits the observed data best. LCA estimates what is termed “goodness-of-fit” measures for any specific model and these goodness-of-fit measures can be compared between models to decide which model is a better explanation of the observed data. In Mplus, the statistical package used here, information criteria provide goodness-of-fit measures. The measures reported, Akaike's information criterion (AIC) and Bayesian information criterion (BIC), estimate the amount of information lost when using a specific number of latent classes to describe how the observed data relate to each class. Hence, higher AIC and BIC values represent poorer model fit. Since the text types are not in any order, but represent unordered categories, relationships between the observed variables, in this case the 22 *a priori* morphological criteria, and the latent classes are based on multinomial logistic regression estimates (see Chapter 6 for a detailed explanation). Since the exploratory analyses treated all textual variables as independent, a simple mixture model retaining this structure was applied. In essence, the model tests the extent the different pre-specified classes explain the covariance between variables. For a so-called saturated model where there is one latent class for each case (i.e. scroll), all covariance is explained and both AIC and BIC are at a minimum. Thus it is expected that as the number of pre-specified latent classes increases, information criteria values will fall. If, therefore, AIC is higher for a model with more latent classes than another, the model

with fewer classes provides a better statistical explanation of the observed data and should be preferred. The difference between AIC and BIC is that BIC adjusts for the number of parameters in the model to reduce the bias towards over-fitting, that is opting for more complex models than necessary.

Confirmatory LCA for Enochic - Aramaic Levi text types

Initial exploratory analyses using morphological criteria confined to the Qumran corpus only identified twenty-nine from the 74 Aramaic scrolls with 50 words or more as belonging to specific text types:

4Q208, 4Q209, 4Q213a – Enochic (E)

4Q213b, 4Q539, 5Q15 – Aramaic Levi, Apocryphon of Joseph B, New Jerusalem (A)

4Q204, 4Q205, 4Q531, 4Q552, 4Q553, 4Q558 – Enochic, Four Kingdoms etc (F)

2Q24, 11Q18, 4Q210, 1Q21, 4Q554 – New Jerusalem (N)

4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547 – Tobit, Visions of Amram (T)

4Q112, 4Q113, 4Q115, 4Q117. – Biblical (B)

Post hoc testing indicated that there were no statistically significant differences between clusters E and A. Hence, there are two alternative models: one with six known classes and one, where E and A are combined with five known classes. Since at this stage none of the other 35 Aramaic scrolls had been placed into a text type group, these were excluded from the confirmatory analysis to determine whether five or six classes best explains the classification of the 29 scrolls of known text type.

For the uncombined model with six known classes $AIC = 3179$ and $BIC = 3397$. For the model with five known classes, treating E and A as a single text type, $AIC = 3168$ and $BIC = 3354$. Both AIC and BIC are better (i.e. less information lost) for the model where text types A and E are combined. This confirms that these text types should be treated as a single entity.

Confirmatory LCA for Tobit text types

Exploratory analyses that included external texts identified a further Targumic text type and also suggested that some scrolls not aligned to a specific text type hitherto, because of their similarity to Babatha Archive Aramaic texts, might also be included in the Tobit text type group. There were thus 38 of the 74 Qumran Aramaic scrolls included in this analysis with 4Q538, 4Q550c, 4Q570 and 6Q8 either treated as a separate Tobit/Babatha-like class or combined with the Tobit – Visions of Amram group: 4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547. For the uncombined model with $AIC = 4427$ and $BIC = 4762$. For the model with : 4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547 and 4Q538, 4Q550c, 4Q570, 6Q8 treated as a single text

type, AIC = 4460 and BIC = 4758. Thus AIC is lower in the uncombined model, but BIC is lower, albeit marginally, in the combined model. In view of the danger of overfitting, the BIC values should be preferred to the AIC values and this supports including 4Q538, 4Q550c, 4Q570 and 6Q8 as part of a Tobit – Visions of Amram text type.

Summary

Confirmatory LCA supports six known text types covering 38 of the 74 Qumran

Aramaic scrolls:

4Q208, 4Q209, 4Q213a, 4Q213b, 4Q539, 5Q15 – Enochic, Aramaic Levi

4Q204, 4Q205, 4Q531, 4Q552, 4Q553, 4Q558 – Enochic, Four Kingdoms etc

1Q21, 2Q24, 4Q210, 4Q554, 11Q18 – New Jerusalem

4Q196, 4Q197, 4Q213, 4Q538, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547, 4Q550c, 4Q570,
6Q8 – Tobit, Visions of Amram, Giants

4Q112, 4Q113, 4Q115, 4Q117 – Biblical

1Q24, 4Q214, 4Q214b, 4Q318, 4Q548 – Targumic

The text type labels are arbitrary, seeking to capture major components of each group, but inspection shows that both Enochic and Aramaic Levi texts occur in several clusters. The Biblical and New Jerusalem groups are more consistent with this regard. The Targumic label is applied to indicate affinity with an external text type and the epithet of “Babatha-like” could be applied to the Tobit – Visions of Amram – Giants group in a similar manner, only that the Babatha Aramaic texts are not so close morphologically as the Targums are to their group.

However, this is not the final stage of classification and determining the probability of the remaining non-aligned 36 Qumran Aramaic scrolls belonging to known text types may illuminate relationships between scrolls further and suggest a different set of labels.

11. Extending the Classification of Qumran Texts

So far the possibility of different text types within the Qumran Aramaic corpus has been explored based on morphological and syntactic features. It is worth recalling the major choices that may have influenced classification so far:

- 1) Only texts with fifty words or more have been included on the basis of power calculations indicating that morphological criteria would be inadequately powered to detect differences between text types for shorter texts. This means that even if shorter texts contain important information, they will not have informed the classification and, moreover, cannot, themselves, be classified reliably on its basis.
- 2) Only morphological criteria that were present in at least 10% of the remaining 74 scrolls with fifty or more words were used to classify text types. This means that if one of the 27 criteria excluded indicated a discrete text type, this would not be revealed in the current analyses. Such a text type would, by definition, comprise seven scrolls or fewer and if distinct, these would necessarily be among the thirty-six scrolls so far unaligned.
- 3) The thirty-eight scrolls that cluster on exploratory analyses into statistically distinct groups and confirmed as six text types with latent class analysis have been assigned definitely to a text type. That is none of them has been considered to be 90% or 95% probably a specific text type for example, but all assigned to belong to their text type with 100% probability. In a sense this

arises from the whole concept of text types; they require a critical mass of typical texts to have any value otherwise each text becomes its own unique text type, which for the Qumran corpus would mean 74 text types. This defeats the purpose of textual classification. Nevertheless, it is worth noting that this differs from some external index such as in which cave a particular scroll was discovered.

The next stage where classification of the Qumran Aramaic texts is extended should help decide how justified the assumptions made in 2) and 3) are.

Latent class analysis with training

In the previous chapter Latent Class Analysis (LCA) was used to confirm the number of classes (text types) identified during the exploratory cluster analyses. These LCAs were confined to the Qumran Aramaic scrolls which had clustered into a discrete group. In summary, 38 of the 68 scrolls clustered into six discrete classes. This left 30 scrolls so far non-aligned. Thus exploratory techniques had only managed to classify just over half of the corpus. This is not surprising, however. First, in view of the exploratory nature of the analyses, considerable caution had been exercised in deciding the extent of a cluster and checking its robustness by using both raw and standardized data. This means that there may well be some scrolls that might also have been included in a text type group, but did not quite meet strict inclusion criteria. Secondly, mixed or intermediate text

types are possible. Some scrolls, for example, may fall between two different text types and thus not fall clearly into either. Thirdly, some scrolls may not belong to any of the text types identified, nor have close affinities with any of them. Again, these scrolls would be likely to remain non-aligned. LCA can now be used to investigate these possibilities.

This analysis to extend the classification of the Qumran Aramaic scrolls is predicated in knowing first how many classes are present and secondly knowing that some scrolls definitely are members of each class. The measurement model is the same as that used in the confirmatory LCA: multinomial logistic regression into the known classes using the 22 *a priori* morphological criteria. For this model, six classes are specified and the 38 scrolls with known membership designated as such. Using the Training command in Mplus, the model is estimated and a probability of class membership derived for each of the 30 scrolls where this is unknown. A probability of class membership is estimated for each scroll for each class with the sum of probabilities across each class equaling one. Values close to zero indicate a low probability of class membership; values close to one indicate a high probability of class membership. Table 11.1 displays these probabilities. Setting an arbitrary threshold of 90% to determine text type membership, 27 of the 30 previously non-aligned scrolls have been classified into a definite text type.

Scroll	EA	F	N	Tam	B	Targ	Text type
1Q20	.998	.000	.000	.001	.000	.000	EA
1Q21	.000	.000	1.000	.000	.000	.000	<i>N</i>
1Q23	.000	1.000	.000	.000	.000	.000	F
1Q24	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
2Q24	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q112	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q113	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q115	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q117	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q156	.000	.000	.000	.000	.000	1.000	Targ
4Q196	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q197	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q198	.000	.577	.305	.118	.000	.000	(F)
4Q201	.000	.941	.000	.059	.000	.000	F
4Q202	.000	1.000	.000	.000	.000	.000	F
4Q203	.000	.003	.989	.007	.000	.000	N
4Q204	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q205	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q206	.000	1.000	.000	.000	.000	.000	F
4Q207	.000	1.000	.000	.000	.000	.000	F
4Q208	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q209	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q210	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q211	.999	.000	.000	.000	.000	.000	EA
4Q212	.000	.015	.339	.646	.000	.000	(Tam)
4Q213	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q213a	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q213b	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q214	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q214b	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q242	.014	.000	.001	.000	.000	.985	Targ
4Q243	.000	.017	.000	.982	.000	.000	Tam
4Q246	.000	.000	.023	.000	.000	.977	Targ
4Q318	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q529	.000	.000	1.000	.000	.000	.000	N
4Q530	.000	1.000	.000	.000	.000	.000	F
4Q531	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q534	.000	.000	.000	1.000	.000	.000	Tam
4Q535	.000	.000	1.000	.000	.000	.000	N
4Q536	.000	.000	.998	.002	.000	.000	N
4Q537	.000	.000	.000	1.000	.000	.000	Tam
4Q538	.000	.000	.000	1.000	.000	.000	<i>Tam</i>

4Q539	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q540	.000	.000	.999	.001	.000	.000	N
4Q541	.001	.000	.001	.998	.000	.000	Tam
4Q542	.000	.192	.026	.782	.000	.000	(Tam)
4Q543	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q544	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q545	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q546	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q547	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q548	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q550	.000	.001	.999	.000	.000	.000	N
4Q550a	.000	.000	.999	.001	.000	.000	N
4Q550c	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q552	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q553	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q554	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q556	.000	.000	.996	.004	.000	.000	N
4Q558	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q560	.923	.000	.000	.000	.000	.077	EA
4Q561	.000	.000	.000	1.000	.000	.000	Tam
4Q562	.000	.000	1.000	.000	.000	.000	N
4Q570	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
5Q15	1.000	.000	.000	.000	.000	.000	<i>EA</i>
6Q8	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
11Q10	.000	.000	.000	1.000	.000	.000	Tam
11Q18	.000	.000	1.000	.000	.000	.000	<i>N</i>

Table 11.1 Class probabilities for each Qumran Aramaic scroll (n=68) per text type:

Enochic – Aramaic Levi (EA); Enochic, Four Kingdoms (F); New Jerusalem (N); Tobit/Visions of Amram/Giants (Tam); Biblical (B); Targumic (Targ). Text type in **BOLD** indicates newly classified. Text type in *ITALICS* indicates previously classified and class membership fixed with 100% certainty. Text type in (PARENTHESES) indicates classification less certain, <90% probability of any class membership.

Several inferences can be drawn from these results. First, the successful allocation of the remaining scrolls into known text types militates against the likelihood of other, unknown text types accounting for anything but a negligible number of scrolls. The rejection of 27 of the 49 morphological criteria is thus justified by this analysis. Secondly, nearly all the non-aligned texts have very high probabilities of class membership indicating that the exploratory criteria were very strict. This is reassuring in terms of allocating scrolls with certainty to known text types during the exploratory phase; that is the criteria were strict enough to justify this.

Thirdly, there is very little evidence of intermediate or transitional text types; only 4Q212 and 4Q542 would be consistent with this. Fourthly, only 4Q198 poses a difficulty for classification with features that might suggest affinity with three different text types. Fifthly, all text types acquired new members except the Biblical text type. This is also reassuring because the Biblical scrolls would, *a priori*, be considered to be distinct from the other Aramaic scrolls. If the classification was not robust and at random, some non-aligned scrolls would have been expected to have been allocated to this class.

This extended classification has resulted in the following groups of scrolls (leaving their arbitrary descriptive label unchanged for the present):

Enochic – Aramaic Levi (EA)

Nine scrolls, 9,590 total words:

1Q20, 4Q208, 4Q209, 4Q211, 4Q213a, 4Q213b, 4Q539, 4Q560, 5Q15.

Enochic, Four Kingdoms (F)

Twelve scrolls, 8,005 total words:

1Q23, 4Q201, 4Q202, 4Q204, 4Q205, 4Q206, 4Q207, 4Q530, 4Q531, 4Q552, 4Q553,
4Q558.

New Jerusalem (N)

Fourteen scrolls, 3,794 total words:

1Q21, 2Q24, 4Q203, 4Q210, 4Q529, 4Q535, 4Q536, 4Q540, 4Q550, 4Q550a, 4Q554,
4Q556, 4Q562, 11Q18.

Tobit/Visions of Amram/Giants (Tam)

Eighteen scrolls, 9,260 total words:

4Q196, 4Q197, 4Q213, 4Q243, 4Q534, 4Q537, 4Q538, 4Q541, 4Q543, 4Q544, 4Q545,
4Q546, 4Q547, 4Q550c, 4Q561, 4Q570, 6Q8, 11Q10.

Biblical (B)

Four scrolls, 2,806 total words:

4Q112, 4Q113, 4Q115, 4Q117.

Targumic (Targ)

Eight scrolls, 1,447 total words:

1Q24, 4Q156, 4Q214, 4Q214b, 4Q242, 4Q246, 4Q318, 4Q548.

Characterisation of the extended text types

An important first step is to test whether the different text types differ in terms of part-of-speech composition as this would indicate that morphological criteria differences may relate to the proportion of a specific part of speech and possibly necessitate adjustment in further analyses. A multivariate general linear model with the proportions of all principle parts of speech indicated an overall significant difference between text types ($F_{295, 25}=2.28$, $p=0.001$, partial $\eta^2=0.16$). Univariate statistics indicated this was largely due to significant effects of proportion of verbs ($p<0.001$), pronominal suffixes ($p=0.001$) and particles ($p=0.001$); differences between text types for other parts of speech were not significant. *Post hoc* analyses using Scheffe's ranges¹ indicated that Targ text type and Tam text types differed from each other in terms of proportion of

¹ Scheffe's ranges provide 95% confidence intervals for the estimated mean of a variable adjusted for both other variables in the statistical model and the number of comparisons being made (i.e. correcting for Type 1 statistical error with regards to the number of hypotheses being tested).

Criterion	EA	F	N	Tam	B	Targ
Determined nouns	71 (60, 82)	80 (71, 90)	90 (81,98)	38 (30, 45)	143 (126,159)	129 (118, 141)
Afel	20 (14, 26)	12 (7.6, 17)	4.8 (0.3, 9.2)	11 (7.1, 15)	5.0 (0, 14)	14 (8.1, 20)
Hafel	0.04 (0, 3.0)	0 (0, 2.6)	2.6 (0.2, 4.9)	2.7 (0.7, 4.8)	17 (13, 22)	2.5 (0, 5.6)
Itpeel	2.9 (0, 6.5)	7.3 (4.2, 10)	7.5 (4.6, 10)	4.7 (2.1, 7.2)	0.7 (0, 6.1)	4.3 (0.4, 8.1)
Hitpeel	0 (0, 0.6)	0 (0, 0.5)	0 (0, 0.5)	0.2 (0, 0.6)	7.4 (6.5, 8.3)	0 (0, 0.6)
<i>m</i> -prefixed Peal infinitives	1.8 (0, 4.8)	9.1 (6.4, 12)	2.1 (0, 4.5)	9.4 (7.2, 12)	2.5 (0, 7.1)	2.4 (0, 5.6)
Peal infinitives	1.8 (0, 4.7)	9.1 (6.6, 12)	2.1 (0, 4.5)	9.8 (7.7, 12)	2.5 (0, 7.0)	2.4 (0, 5.5)
<i>l</i> prefixed imperfects	0.06 (0, 4.2)	1.5 (0, 5.0)	6.6 (3.3, 9.8)	7.2 (4.3, 10)	1.1 (0, 7.3)	3.8 (0, 8.1)
<i>y</i> prefixed imperfects	1.7 (0,17)	8.8 (0, 22)	25 (13, 38)	21 (11, 32)	15 (0, 38)	29 (13, 45)
3fp verbs	0.02 (0, 0.9)	1.0 (0.3, 1.7)	0 (0, 0.7)	0.9 (0.3, 1.5)	0.2 (0, 1.5)	0 (0, 0.9)
Perfect 3mp ending <i>-w</i>	5.1 (0.2, 10)	22 (18, 27)	3.8 (0, 7.8)	9.1 (5.6, 13)	6.3 (0, 14)	0.6 (0, 5.8)
Perfect 3mp ending <i>-w'</i>	0.1 (0, 0.7)	1.1 (0.6, 1.6)	0 (0, 0.5)	0.04 (0, 0.5)	0 (0, 0.9)	0 (0, 0.6)
Retained aleph before suffix	1.1 (0.02, 2.2)	0.2 (0, 1.1)	0 (0, 0.9)	0.5 (0, 1.3)	0 (0, 1.6)	1.4 (0.3, 2.6)
2ms pronominal suffix <i>-kh</i>	0.6 (0, 4.3)	0.8 (0, 4.0)	3.2 (0.3, 6.2)	3.4 (0.7, 6.0)	1.0 (0, 6.6)	0 (0, 3.9)
3fs pronominal suffix <i>-h'</i>	2.3 (1.3, 3.2)	0.09 (0, 0.9)	0 (0, 0.7)	0.3 (0, 1.0)	0.2 (0, 1.6)	0 (0, 1.0)
Relative particles	16 (7.1, 25)	32 (25, 40)	45 (38, 52)	27 (21, 33)	50 (37, 63)	15 (5.9, 24)
Relative particle <i>dy</i>	13 (6.0, 21)	29 (23, 36)	42 (36, 48)	27 (22, 32)	50 (39, 61)	15 (7.2, 23)
Direct object marker <i>yt</i>	1.2 (0, 2.6)	0.07 (0, 1.3)	1.2 (0.03, 2.3)	0.4 (0, 1.4)	0 (0, 2.1)	0 (0, 1.5)

<i>dnh</i> or <i>dn'</i> forms	0.1 (0, 2.2)	0.1 (0, 2.0)	1.3 (0, 3.0)	2.0 (0.5, 3.5)	6.0 (2.8, 9.2)	2.3 (0.1, 4.6)
Demonstrative pronoun <i>'lh</i>	0.02 (0, 0.6)	0.2 (0, 0.7)	0.4 (0, 0.9)	0.2 (0, 0.6)	1.7 (0.8, 2.6)	0 (0, 0.6)
Demonstrative pronoun <i>'lyn</i>	0.2 (0, 2.5)	5.7 (3.7, 7.7)	1.1 (0, 3.0)	0.7 (0, 2.2)	0.7 (0, 4.2)	2.8 (0.4, 5.3)
Demonstrative pronoun <i>'ln</i>	0.2 (0, 2.5)	4.0 (2.0, 6.0)	0.7 (0, 2.6)	0.5 (0, 2.1)	0.2 (0, 3.7)	2.8 (0.4, 5.3)

Table 11.2 Means (95% confidence intervals) for rates of the 22 *a priori* morphological criteria in the different text type groups of scrolls.

verbs (0.098 versus 0.16), but neither from any of the other text types. Similarly Tam texts had a higher mean proportion of pronominal suffixes (0.12) compared to Targ (0.07) or B (0.068), but otherwise there were no significant differences. The overall significant difference between text types persisted ($p=0.023$) even after adjusting for the proportion of verbs. For particles Targ had a higher proportion (0.48) than either EA (0.40) or Tam (0.39), but otherwise there were no significant differences in text types. Again the overall significant difference between text types persisted ($p=0.009$) after adjusting for the proportion of verbs.

Whereas text type explained 16% of the variance in part-of-speech composition, it explained 72.3% of the differences in the 22 *a priori* morphological criteria ($p<0.001$). Univariate statistics indicated significant differences for all except rates of *yt* ($p=0.62$), *'lh* ($p=0.052$), feminine third person plural verbs ($p=0.11$), masculine second person singular suffix *-kh* ($p=0.54$), *y*-prefixed imperfects ($p=0.082$), *itpeels* ($p=0.13$), *'ln* ($p=0.058$), and retained alephs ($p=0.32$). Thus 14 of the 22 criteria showed significant

differences on their own, with two further criteria being near-significant. Table 11.2 shows mean rates and 95% confidence intervals of the mean for all 22 morphological criteria by text type. Table 11.3 shows which text types have criteria means that are not part of a homogeneous set, so that text types have mutually exclusive high and low rates on Scheffe's ranges. On this very conservative comparison, 11 of the 22 criteria have more than one homogeneous set. Distinct rates occur for 11 of the 22 criteria for the EA group, 5 out of 22 for the F group, 6 out of 11 for the N group, 7 out of 11 for the Tam group, 9 out of 11 for the B group, and 8 out of the 11 for the Targ group. Hence, the EA

Criterion	EA	F	N	Tam	B	Targ
Determined nouns	Medium	Medium	Medium	Low	High	High
Afel	High				Low	Low
Hafel	Low	Low	Low	Low	High	Low
Itpeel						
Hitpeel	Low	Low	Low	Low	High	Low
<i>m</i> -prefixed Peal infinitives	Low			High		
Peal infinitives	Low			High		
<i>l</i> prefixed imperfects						
<i>y</i> prefixed imperfects						
3fp verbs						
Perfect 3mp ending - <i>w</i>	Low	High	Low	Low	Low	Low
Perfect 3mp ending - <i>w'</i>						
Retained aleph before suffix						
2ms pronominal						

suffix <i>-kh</i>						
3fs pronominal suffix <i>-h'</i>						
Relative particles	Low		High		High	Low
Relative particle <i>dy</i>	Low		High		High	Low
Direct object marker <i>yt</i>						
<i>dnh</i> or <i>dn'</i> forms	Low	Low			High	
Demonstrative pronoun <i>'lh</i>	Low			Low	High	Low
Demonstrative pronoun <i>'lyn</i>						
Demonstrative pronoun <i>'ln</i>						

Table 11.3 Text types which have criteria means that are not part of a homogeneous set, so that text types have mutually exclusive high and low rates on Scheffe's ranges. Empty cells denote text types not excluded from homogeneous sets with either the text types with high and low rates. Medium denotes homogeneous sets which differ from high or low text type mean rates.

group is most distinct and the F group least distinct morphologically of the six text types. Table 11.4 shows how many times pairs of text type groups are within the same homogeneous set for the 11 criteria with more than one homogeneous set. The totals indicate considerable overlap on many of the morphological criteria between text types

	F	N	Tam	B	Targ
EA	8	8	8	4	10
F		10	9	6	9
N			9	8	8
Tam				5	9
B					6

Table 11.4 Frequencies when pairs of text type groups are within the same homogeneous set for the 11 morphological criteria with more than one homogeneous set.

Hence, such text types would not have been identifiable using a single criterion approach: it is the combination of morphological criteria that differentiates between text types just as it is the combination of ingredients, rather than any single ingredient, that differentiates between recipes. A summary of the distinctive features of these extended text types is:

- EA High Afel rates, low Peal infinitive rates, low relative particles rates.
- F High Perfect 3mp ending *-w* rates, low *dnh* or *dn'* rates.
- N High relative particles rates
- Tam High Peal infinitive rates, low determined nouns rates.
- B High determined nouns rates, high Hafel rates, high Hitpeel rates, high relative pronoun rates, high *'lh* rates, high *dnh* or *dn'* rates.
- Targ High determined nouns rates, low Afel rates, low relative particles rates.

Since several distinct features relate to verbal forms and the proportion of verbs differs significantly between text types, these raw descriptions, although they summarise the nature of the text types accurately, need to be considered in the context of possible forms. That is, low or high rates of specific verbal forms may just indicate high or low rates of verbal forms in general within a text type. This is also true for relative particles and particles in general, though less so because the majority of particles are *waw*.

Verbal forms and text types

	Mean	Lower 95% CI	Upper 95% CI
F	0	0	0.10
EA	0.004	0	0.12
Targ	0.08	0	0.21
Tam	0.13	0.05	0.22
N	0.14	0.05	0.24
B	0.64	0.46	0.83

Table 11.5 Mean proportion of hafels to afels plus hafels plus one with 95% confidence intervals of the mean.

First hafel and afel forms were considered as alternative morphological choices. The proportion of hafels was calculated as hafels/(hafels + afels + 1) to avoid potential division by zero (i.e. in scrolls with neither hafels nor afels). There was a significant difference between text types ($p < 0.001$) with the B group having distinctly high rates on Scheffe's ranges, although less conservatively both Tam and N groups had mean rates above the upper 95% confidence interval of the means of both EA and F groups (Table 11.5). Similarly, the B group had a distinctly high proportion of itpeels ($p < 0.001$) with the 95% confidence intervals of the means of all the other text type groups overlapping each other's means.

	Mean	Lower 95% CI	Upper 95% CI
EA	0.008	0	0.12
B	0.05	0	0.23
F	0.09	0	0.19
Targ	0.10	0	0.23
Tam	0.20	0.12	0.28
N	0.24	0.15	0.33

Table 11.6 Mean proportion of *l*-prefixed imperfects to all third person masculine singular imperfects plus one with 95% confidence intervals of the mean.

For *l*-prefixed imperfects both *y*-prefixed and *n*-prefixed forms were included in the denominator in addition to unity. There was a significant difference between groups ($p=0.023$), though this did not show up on Scheffe's ranges. On 95% confidence intervals, Tam and N groups had higher rates (Table 11.6). Next, the proportion of third person masculine plural *-w* endings within *-w* or *-w'* (plus unity) calculated. Again this was significant ($p<0.001$), but four subsets emerged on Scheffe's ranges (Table 11.7). Finally, the proportion of *m*-prefixed peal infinitives within all peal infinitives (plus unity) was calculated and showed significant differences between text type groups ($p<0.001$). There was only a single homogeneous subset of means on Scheffe's ranges, but less conservatively on 95% confidence intervals of the mean F and Tam groups have high rates, Targ, N and EA low rates and B an intermediate mean rate (Table 11.8).

Text type (N)		Subset			
		1	2	3	4
Targ	8	.06			
N	14	.28	.28		
EA	9	.35	.35	.35	
Tam	18		.64	.64	.64
B	4			.74	.74
F	12				.82

Table 11.7 Means by text type groups for proportion of third person masculine plural *-w* endings within *-w* or *-w'* (plus unity) as homogeneous subsets (1-4) by Scheffe's ranges.

	Mean	Lower 95% CI	Upper 95% CI
Targ	0.19	0	0.41
N	0.20	0.03	0.37
EA	0.27	0.05	0.48
B	0.49	0.18	0.81
F	0.67	0.49	0.86
Tam	0.68	0.53	0.83

Table 11.8 Mean proportion of *m*-prefixed peal infinitives within all peal infinitives plus one with 95% confidence intervals of the mean.

Relative particles and text types

An identical approach was used for relative particles form as with verbal forms. The proportion of *dy* within all relative particles (plus one) was calculated. Again there were significant differences between text type groups ($p=0.001$), with two homogeneous subsets emerging on Scheffe's ranges (Table 11.9). The order is not dissimilar to that of the proportion of *hafels*.

Text type (N)	Subset	
	1	2
EA	.59	
Targ	.60	
F	.76	.76
N	.83	.83
Tam	.88	.88
B		.96

Table 11.9 Mean rates for text type groups for proportion of *dy* within total relative particles (plus one) as homogeneous subsets by Scheffe's ranges.

Determined state nouns and text types

Use of a noun in a determined state is not necessarily an alternative to using one in either the absolute or construct states, and the proportion of total nouns did not differ significantly between text types. Hence it is unsurprising replicating the general linear

model for the proportion of determined state nouns to total nouns (plus unity) made no difference to the distribution of text types across Scheffe's ranges subsets to that found for determined state nouns alone.

A synthesis of alternative forms

Setting aside pronouns and determined state nouns, the other criteria which differ in rates between the text types provide six pairs of alternative forms: *afel/hafel*, *itpeel/hitpeel*, *l*-prefixed/*y*-prefixed imperfect verbs, *m*-prefixed/not *m*-prefixed peal infinitive verbs, third person masculine plural *-w/-w'* endings, and *dy*/not *dy* relative particles. Making sense of the different, but sometimes similar, ordering of texts by proportions of each of these pairs as shown in Tables 11.5-11.9 is far from easy by inspection alone. Thus it may be helpful to search for some summary underlying factors to clarify the relationships. Furthermore, the different proportions are difficult to compare directly because adding one to the denominator has a larger effect when the other denominator term is low (i.e. for relatively infrequent features) compared with when it is high (i.e. for relatively frequent features). Standardizing these variables makes comparison more comprehensible.

Entering the six standardized variables into a principal components analysis, a conventional initial method of summarizing data, extracted three principal components with an eigenvalue > 1.0. However, the third unrotated principal component accounting

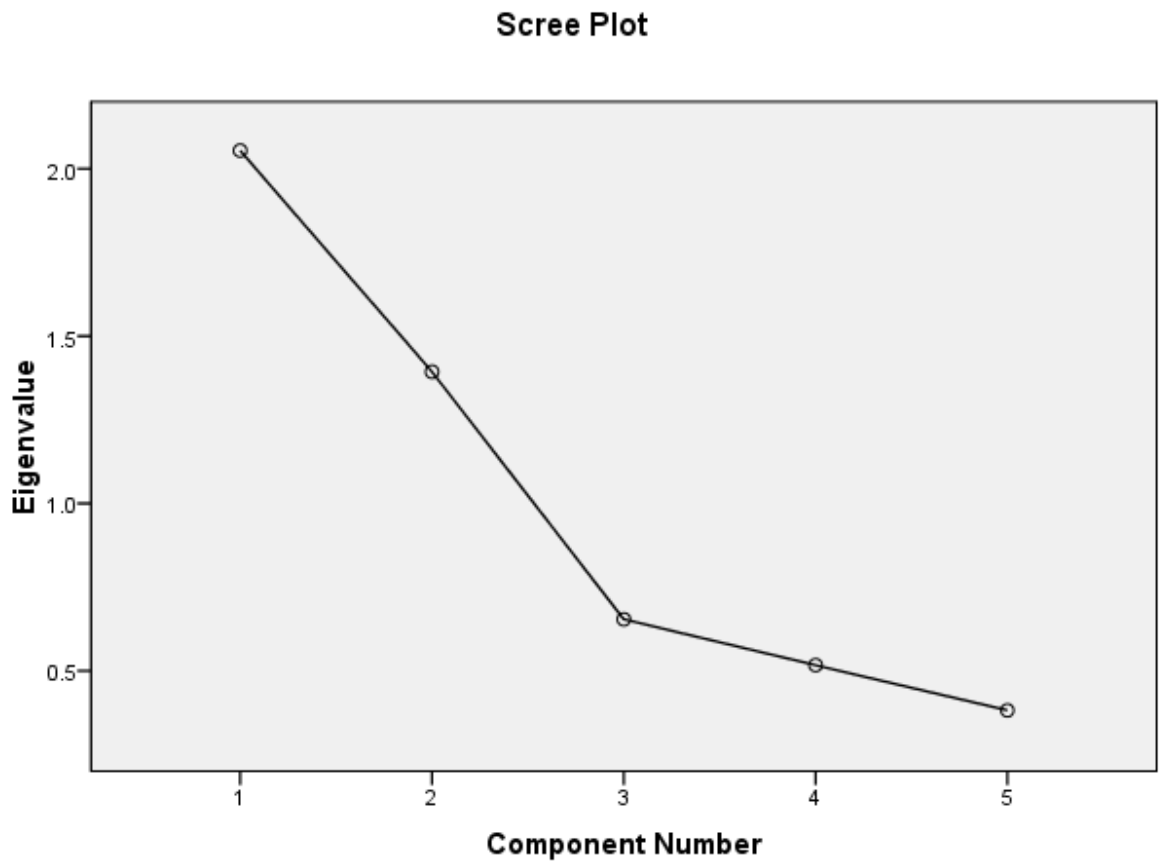


Figure 11.1 Scree plot of principal components extraction from five standardized variables *afel/hafel*, *itpeel/hitpeel*, *m*-prefixed/not *m*-prefixed *peal* infinitive verbs, third person masculine plural *-w/-w'* endings, and *dy*/not *dy* relative particles.

for 17.2% of variance had a loading component of 0.97 for the proportion of *l*-prefixed imperfects with none of the other loading components $>|0.22|$. Hence, the principal components do not aid the intelligibility of the proportion of *l*-prefixed imperfects since it appears unrelated to the other variables, so this was removed to be considered as

separate from the other morphological alternatives. For the remaining five standardized variables, two principal components with eigenvalues >1.0 explained 68.9% of shared variance. The scree plot (Figure 11.1) indicates that a two component solution is optimal. Varimax rotation produced a comprehensible component matrix structure (Table 11.10) displayed as a loading plot in Figure 11.2. This shows that H-stem verbal

Rotated Component Matrix^a

	Component	
	1	2
Zscore(Haph_prop)	.042	.898
Zscore(Hith_prop)	.112	.870
Zscore(m3p_w_prop)	.816	.027
Zscore(dy_prop)	.675	.274
Zscore(M_peal_inf_prop)	.819	-.028

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Table 11.10 Varimax rotated component matrix for the five standardized variables *afel/hafel* (Haph_prop), *itpeel/hitpeel* (Hith_prop), *m*-prefixed/not *m*-prefixed *peal* infinitive verbs (M_peal_inf_prop), third person masculine plural *-w/-w'* endings (m3p_w_prop), and *dy*/not *dy* relative particles (dy_prop).

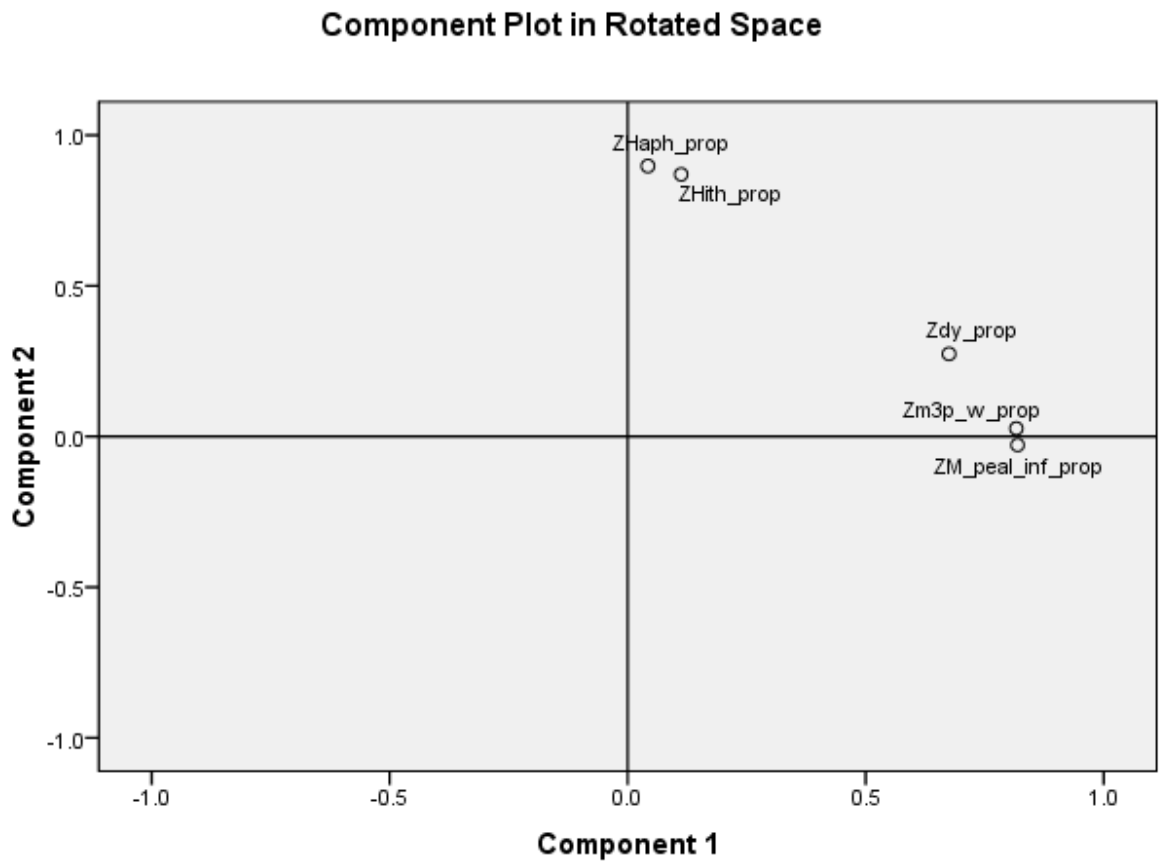


Figure 11.2 Loading plot of Varimax rotated component matrix for the five standardized variables afel/hafel (Haph_prop), itpeel/hitpeel (Hith_prop), *m*-prefixed/not *m*-prefixed peal infinitive verbs (M_peal_inf_prop), third person masculine plural *-w/-w'* endings (m3p_w_prop), and *dy/not dy* relative particles (*dy_prop*).

forms provide one common factor with the other three criteria the other common factor representing typical Qumran forms (see Table 11.10). Hence, Schattner-Rieser gives *-w*

as the standard Qumran form for the third masculine plural perfect verb ending,² *m-* prefix as the standard Qumran form for the Peal infinitives,³ and *dy* as the standard Qumran form for the relative pronoun, though she allows *d* as a late form.⁴

Hence, the first principal component can be said to reflect ‘typical’ Qumran forms, the second component represents proportion of H-verbal forms, and standardized *l*-prefixed imperfect proportion provides a further, separate dimension. Considering these three descriptors allows a more informative comparison of the different text types (Table 11.11). Figure 11.3 displays the two dimensional scaling using Euclidean distances of the relationship between the six extended text types derived from the two factor scores and standardized *l*-prefixed imperfect proportion values. This shows that the B text type is quite distinct from the others, in keeping with previous comparisons

	Typical Qumran form factor	H-stem verb factor	<i>l</i> -prefixed imperfect proportion
EA	-0.68	-0.43	-0.71
F	0.72	-0.58	-0.28
N	-0.53	0.06	0.54
Tam	0.62	-0.11	0.33
B	0.35	3.28	-0.47
Targ	-1.17	-0.15	-0.19

Table 11.11 Mean factor scores and standardized *l*-prefixed imperfect proportion values for the six extended Qumran text types.

² U. Schattner-Rieser (2004), p67 Table II.

³ *ibid*, (2004), p71.

⁴ *ibid*, (2004), p65.

(Tables 11.2-11.4), that F and Tam text types are relatively close, and that EA, N and Targ text types form another, slightly looser grouping.

These summary measures can now be compared with the three other morphological criteria that showed differences on Scheffe's ranges between text types. Table 11.12 shows Spearman correlation coefficients with p-values between summary alternative forms criteria and other morphological criteria that differentiate the extended text types. Of note is that the *l*-prefixed imperfect proportion has no correlation with any of the other morphological criteria whereas the H-stem factor correlates significantly and positively with all three. The typical Qumran form factor correlates significantly, and negatively, with determined state nouns only.

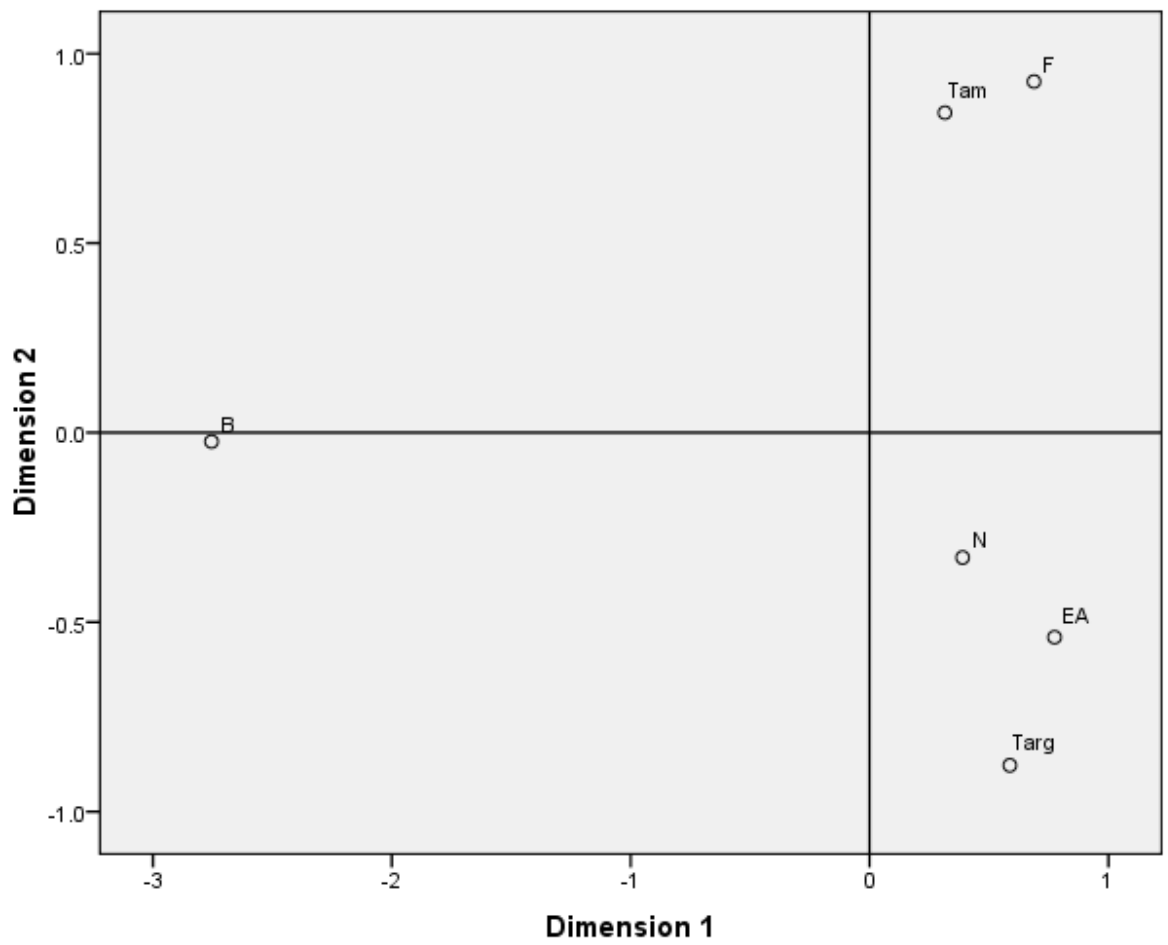


Figure 11.3 Two dimensional scaling using Euclidean distances of the relationship between the six extended text types derived from the two factor scores and standardized *l*-prefixed imperfect proportion values.

	Determined state nouns	' <i>lh</i>	<i>dn'</i> or <i>dnh</i>
Typical Qumran form factor	-0.33 (0.007)	0.15 (0.24)	-0.03 (0.79)
H-stem verb factor	0.41 (0.001)	0.35 (0.004)	0.45 (<0.001)
<i>l</i> -prefixed imperfect proportion	-0.09 (0.46)	0.01 (0.94)	0.03 (0.04)

Table 11.12 Spearman correlation coefficients (p-values) between summary alternative forms criteria and other morphological criteria that differentiate the extended text types.

Hence, *l*-prefixed imperfects constitute a distinct discriminator between text types (putatively indicating a geographical influence with higher rates associated with “Eastern Aramaic”), H-stem forms relate positively to '*lh* and *dn'* or *dnh* forms (putatively indicating a diachronic influence with higher rates indicating earlier style texts), and the rate of determined state nouns relates to the two summary factors in opposite ways.

A textual classification

The preceding analyses allow a textual provisional classification in terms of space and time plus what has been termed “typicality”. The latter factor catches the sense that Qumran Aramaic texts reflect to a greater or lesser extent a mainstream development of what has been termed “Official Aramaic”.

- EA Late, Western, but not typically mainstream Aramaic style texts.
- F Late, Western, typically mainstream Aramaic style texts.
- N Early, Eastern, not typically mainstream Aramaic style texts.
- Tam Middle in age, Eastern, typically mainstream Aramaic style texts.
- B Early, Western, typically mainstream Aramaic style texts.
- Targ Middle in age, neither predominantly Western nor Eastern, quite untypical of mainstream Aramaic style texts.

Non-aligned scrolls

LCA with training classified 65 of the 68 scrolls used in clustering, only 4Q198, 4Q212 and 4Q542 failed to be assigned membership of a text type group with at least 90% probability. There are also six further scrolls that were excluded from exploratory cluster analysis at an early stage: 4Q214a, 4Q244, 4Q532, 4Q549, 4Q557 and 4Q559. These texts were distant from others in the cluster analyses and also relatively short, all fewer than 120 words. Indeed 4Q557 had only 24 words, so would be excluded on the 50 words minimum length criterion for morphological variable statistical analysis and will not be considered further. As a first step towards seeking to characterize these eight remaining scrolls, a consideration of their morphological features in the light of the above textual classification.

4Q198

LCA estimated that this scroll most probably belonged to text type F (57.7% probability), perhaps to text type N (30.5%), possibly to text type Tam (11.8% and which might have been expected on content considerations alone), but highly unlikely to belong to any other text type. On the three classifying descriptors, Q198 scores 0.03 on the H-stem factor suggesting that it is neither an early nor late text similar to both Tam and Targ text types, 0.34 on the typicality factor indicating that it is fairly mainstream, and -0.75 on the proportion of *l*-prefixed imperfects suggestive of a Western text-style in contrast with the Tam group. With this in mind, it might be considered a “proto-F” text-style. It may have arisen due to Western modification during copying of an Eastern Tam text-style text or may represent an original Western Tobit tradition. Indeed, the Qumran Tobit texts within the Tam group, 4Q196 and 4Q197, may represent Eastern modification during copying of now-lost Western texts. In either case, 4Q198’s text-style demonstrates that it is spurious to attribute a certain geographical provenance to the Tobit texts found at Qumran.

4Q212

4Q212 is labelled as one of the Enochic scrolls. At 699 words it is fairly long and thus failure to classify is unlikely to reflect poor statistical power. LCA indicated it was 64.6% likely to be of a Tam text type, 33.9% likely to belong to the N text type and highly unlikely to belong to any other text type. It has an H-stem factor score of -0.53 consistent with a late text style. It scores low on *l*-prefixed imperfects at -0.63 indicative

of a Western text. It scores high at 0.65 on typicality indicating a mainstream Aramaic style. These scores would favour an F text type, at odds with LCA which considered this only 1.5% likely. This suggests that a combination of other, less common features are inconsistent with an F text type; a relatively low rate of third person masculine plural *-w* endings and high rate of feminine third person plural verbs may be part of the explanation.

4Q542

4Q542 is commonly known as the Testament of Qahat. At 467 words it, too, should be easy to classify statistically. LCA suggested that it was most likely of a Tam text type (78.2%), but could be F (19.2%). Chances of it belonging to the text types were small-to-negligible. It scores 0.47 on the H-stem factor consistent with an early text. It scores 1.07 on *l*-prefixed imperfects proportion standardized score strongly suggestive of an Eastern text. It scores 0.69 on the typicality factor indicating that it has a mainstream style. These data would fit with labeling 4Q542 as a “proto-Tam” text.

4Q214a, 4Q244, 4Q532, 4Q549, and 4Q559

Scroll	EA	F	N	Tam	B	Targ	Text type
1Q20	1.000	.000	.000	.000	.000	.000	<i>EA</i>
1Q21	.000	.000	1.000	.000	.000	.000	<i>N</i>
1Q23	.000	1.000	.000	.000	.000	.000	<i>F</i>
1Q24	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
2Q24	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q112	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q113	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q115	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q117	.000	.000	.000	.000	1.000	.000	<i>B</i>
4Q156	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q196	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q197	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q198	.000	.567	.278	.154	.000	.000	(F)
4Q201	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q202	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q203	.000	.000	1.000	.0007	.000	.000	<i>N</i>
4Q204	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q205	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q206	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q207	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q208	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q209	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q210	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q211	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q212	.000	.071	.177	.753	.000	.000	(Tam)
4Q213	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q213a	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q213b	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q214	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q214a	1.000	.000	.000	.000	.000	.000	EA
4Q214b	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q242	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q243	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q244	.000	.001	.999	.000	.000	.000	N
4Q246	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q318	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q529	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q530	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q531	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q532	.000	1.000	.000	.000	.000	.000	F
4Q534	.000	.000	.000	1.000	.000	.000	<i>Tam</i>

4Q535	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q536	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q537	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q538	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q539	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q540	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q541	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q542	.000	.090	.094	.815	.000	.000	(<i>Tam</i>)
4Q543	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q544	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q545	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q546	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q547	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q548	.000	.000	.000	.000	.000	1.000	<i>Targ</i>
4Q549	1.000	.000	.000	.000	.000	.000	EA
4Q550	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q550a	.000	.000	.1.000	.000	.000	.000	<i>N</i>
4Q550c	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q552	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q553	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q554	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q556	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q558	.000	1.000	.000	.000	.000	.000	<i>F</i>
4Q559	1.000	.000	.000	.000	.000	.000	EA
4Q560	1.000	.000	.000	.000	.000	.000	<i>EA</i>
4Q561	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
4Q562	.000	.000	1.000	.000	.000	.000	<i>N</i>
4Q570	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
5Q15	1.000	.000	.000	.000	.000	.000	<i>EA</i>
6Q8	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
11Q10	.000	.000	.000	1.000	.000	.000	<i>Tam</i>
11Q18	.000	.000	1.000	.000	.000	.000	<i>N</i>

Table 11.13 Class probabilities for each Qumran Aramaic scroll (n=73) per text type:

Enochic – Aramaic Levi (EA); Enochic, Four Kingdoms (F); New Jerusalem (N);

Tobit/Visions of Amram/Giants (Tam); Biblical (B); Targumic (Targ). Text type in

BOLD indicates newly classified. Text type in *ITALICS* indicates previously classified

and class membership fixed with 100% certainty. Text type in (PARENTHESES)

indicates classification less certain, <90% probability of any class membership.

These scrolls did not cluster consistently in the exploratory analyses. However, cluster membership is now more extensive. Repeating the training LCA entering class membership as determined by the previous training LCA, but keeping 4Q198, 4Q212 and 4Q242 as undetermined, 4Q214a, 4Q244, 4Q532, 4Q549, 4Q557 and 4Q559 were now assigned text types with very high probabilities (Table 11.13). There were minor shifts in class membership probabilities for 4Q198, 4Q212 and 4Q542, but none exceeded 90%. The final classification is therefore:

EA Late, Western, but not typically mainstream Aramaic style texts

1Q20, 4Q208, 4Q209, 4Q211, 4Q213a, 4Q213b, 4Q214a, 4Q539, 4Q549, 4Q559,
4Q560, 5Q15

F Late, Western, typically mainstream Aramaic style texts

1Q23, 4Q201, 4Q202, 4Q204, 4Q205, 4Q206, 4Q207, 4Q530, 4Q531, 4Q532, 4Q552,
4Q553, 4Q558

+ 4Q198 (proto-F)

& probably 4Q212

N Early, Eastern, not typically mainstream Aramaic style texts

1Q21, 2Q24, 4Q203, 4Q210, 4Q244, 4Q529, 4Q535, 4Q536, 4Q540, 4Q550, 4Q550a,
4Q554, 4Q556, 4Q562, 11Q18

Tam Middle in age, Eastern, typically mainstream Aramaic style texts

4Q196, 4Q197, 4Q213, 4Q243, 4Q534, 4Q537, 4Q538, 4Q541, 4Q543, 4Q544, 4Q545,
4Q546, 4Q547, 4Q550c, 4Q561, 4Q570, 6Q8, 11Q10

+4Q542 (proto-Tam)

B Early, Western, typically mainstream Aramaic style texts

4Q112, 4Q113, 4Q115, 4Q117

**Targ Middle in age, neither predominantly Western nor Eastern, quite atypical of
mainstream Aramaic style texts**

1Q24, 4Q156, 4Q214, 4Q214b, 4Q242, 4Q246, 4Q318, 4Q548

External validation of the morphological factors

Although both clustering and LCA use all 22 *a priori* morphological criteria that occur in at least 10% of the scrolls for classification, three key factors emerged that explained 86.1% of shared variance of the alternative morpho-syntactic forms. These were the proportion of H-form verbal stems hypothesized to represent a diachronic factor, the presence of *l*-prefixed third person masculine singular imperfects, hypothesized to represent a geographical factor, and a “typical Qumran form” factor. Although these

factor labels have face validity, some very limited, independent data are available to provide a measure of external validity checking.

First, radiocarbon dating has been performed for a small proportion of the Qumran scrolls. There are difficulties with this dating because of interference from chemicals used to preserve some of the scrolls which results in a wide range from earliest to latest probable dates. The following Aramaic scrolls have radiocarbon dates assigned:⁵

4Q542	290 BCE, 395-181 BCE (2-sigma range to account for ¹⁴ C fluctuations).
4Q213	175 BCE, 344-324 or 203-53 BCE
4Q208	145 BCE, 172-48 BCE
1Q20	63 BCE, 89 BCE-118 CE

4Q542 was classified as an early text, 4Q213 as being in the middle of the period of Qumran Aramaic scrolls, and both 4Q208 and 1Q20 are all classified as late texts in the same text type group. There is some overlap of radiocarbon dates, but 4Q542 is definitely earlier than 4Q208 and 1Q20, and likely to be earlier than 4Q213. 4Q213 is likely to be earlier than 4Q208 and 1Q20. Thus radiocarbon dating data are consistent with the diachronic factor extracted from morpho-syntactic analyses.

⁵ G. Doudna, Greg (1998).

Palaeographic dates are less reliable as an independent variable because they rely on scholarly opinion.⁶ 4Q542 is dated palaeographically to 125-100 BCE, 4Q213 to 50-25 BCE, 1Q20 to 30 BCE – 30 CE, but 4Q208 is dated to 225-175 BCE.⁷ Hence the palaeographic dating is at odds with the radiocarbon dating for two of the four scrolls: 4Q542 and 4Q208. Nevertheless, apart from 4Q208, palaeographic dates are consistent with the diachronic factor extracted from morpho-syntactic analyses.

Secondly, *l*-prefixed imperfects have been denoted a marker of Eastern Aramaic texts by scholars.⁸ The *y-/l-/n-* alternative prefix was identified by most of the scholars whose views contributed to the *a priori* morphological criteria drawn up in Chapter 4. Hence, external validity underpinned this criterion from the outset in the analysis.

Unlike the diachronic and geographical factors, the “typical Qumran form” factor was not informed by pre-existing assumptions about Aramaic texts at Qumran. There is a related background to this factor in terms of “Imperial” or “Official” Aramaic from the preceding epoch and the delineation of ongoing dialects contemporaneously. What is being brought out by this factor is a cluster of related morphological characteristics that typify Qumran texts which lie in the tradition of “Imperial” or “Official” Aramaic (e.g.

⁶ F. M. Cross (1961) is still taken as the authoritative account for palaeographic dating of Qumran texts, though this focuses mainly on Hebrew texts and dates from a time when many of the Aramaic texts were yet to be published.

⁷ J. VanderKam & P. Flint (2002), Chapter 2, Table 2.2.

⁸ A. Rubin (2007) provides a thorough review of the area. Also H. L. Ginsberg (1970).

Daniel and Ezra). Hence only twenty scrolls are in the atypical text type groups EA and Targ. One potential independent source to validate this factor is scribal practice as described by Tov.⁹ Unfortunately, Tov does not consider most of the Aramaic scrolls, and of those he does consider (1Q20, 4Q196, 4Q213, 4Q213a, 4Q214, 4Q530, 4Q542, 4Q546, 4Q547) he only states an opinion for 1Q20. This he considers not representative of typical Qumran scribal practice, so on this very limited count, the classification of 1Q20 in the EA group (late Western, but not typically mainstream Aramaic style texts) is supported.

Lexical characteristics of the different text types

The text types are best described according to date, geography and Qumran idiosyncrasy, since these provide the major dimensions of the textual classification, rather than the arbitrary labels that reflect scroll membership at an early stage of cluster analysis that is now extended. However, the textual classification relied on morphosyntactic features, largely relating to particles, pronouns and suffixes, rather than lexical features because of a lack of statistical power with regard to the frequency of individual lexemes given the length of many of the scrolls. Now, treating each text type group as a single entity, some information about lexical textual characteristics is available as it relates to the more content-type parts of speech, nouns and verbs. Setting aside the Biblical text type group which has been extensively studied as a corpus

⁹ E. Tov (2004), Appendix 1.

EA	F	N	Tam	Targ
seventh part	all	all	all	all
all	earth	door	son	son
half	son	cubit	earth	blood
daytime	eternity	width	father	hand
son	tree	reed, rod	God	altar
remnant	man	day	brother	God
door	lamb	spirit, wind	eternity	earth
earth	water	south	year	darkness
cubit	day	ris (700 feet)	woman, wife	province
king	dream	earth	day	light
light	mountain	east	king	bull
width	fire	west	face, nose	king
eternity	deed	block	hand	eternity
year	giant	name	man	people
heavens, sky	watcher	son	house	truth
reed, rod	judgment	priest	wisdom	peace
woman, wife	east	measurement	people	virgin
lord	flesh	bread, food	word, thing	kid, Capricorn
man	divination	eternity	name	balances
house	generation	heavens, sky	spirit, wind	scorpion
	hand		peace	archer
				foot
				name
				year

Table 11.14 The twenty most frequent common nouns listed in order (tied 20th places included) for each text type group corpus. Nouns common to all five groups are not shown in **bold**.

already, the twenty most frequent common nouns, in English translation, are given in order for each text type group corpus in Table 11.14. Four nouns (all, earth, son, eternity) are frequent in all groups. Excluding these four nouns that occur frequently in all text type groups, Table 11.15 shows the number of different common nouns shared between groups, which ranges between one and seven from a possible sixteen.

	EA	F	N	Tam	Targ
EA		1	5	5	2
F	1		2	2	1
N	5	2		4	1
Tam	5	2	3		7
Targ	2	1	1	7	

Table 11.15 The number of different common nouns shared between text type groups, excluding those nouns shared by all groups.

EA	F	N	Tam	Targ
go in	see	be	be	be
be	be	measure	say	come up
see	say	see	see	say
enter	make, do	call, name	know	see
shine	untie, allow	go out	make, do	untie, allow
increase	bear, beget	say	call, name	make, do
say	learn, teach	go in	reply, return	come, bring
cover	arise, stand	complete	hear	declare
adhere	go in	give	come, bring	hit, strike
come, bring	perish, destroy	make, do	speak	salt
be dark	know	divide	take	call, name
be like, put	go	arise, stand	answer	wash
rule	descend	perish, destroy	arise, stand	be like, put
give	go out	come, bring	go in	trample
declare	come, bring	go around	go out	go
diminish	understand	be like, put	ask	sprinkle
make, do	bless	eat, accuse	build	know
come up	declare	build	eat, accuse	write
bear, beget	complain	seek	bless	take
take	call, name	lead	seek	go in
reply, return	change	rise	die	incline, pray
		know	come up	thunder
		bear, beget	command	
		come up		
		find, be able		

Table 11.16 The twenty most frequent common verbs listed in order (tied 20th places included) for each text type group corpus. Verbs common to all five groups are not shown in **bold**.

Table 11.16 shows the twenty most frequent common verbs, in English translation, which are given in order for each text type group corpus. Six verbs (be, see, say, go in, come or bring, make or do) are frequent in all text type corpora; Table 11.17 shows the number of different common nouns shared between groups which ranges between two

and seven from a possible fourteen. Table 11.18 shows the sum totals of Tables 11.15 and 11.17, and indicates that quantitatively all groups are fairly similar lexically.

	EA	F	N	Tam	Targ
EA		2	4	3	4
F	2		6	5	5
N	4	6		7	3
Tam	3	5	7		2
Targ	4	5	3	2	

Table 11.17 The number of different common verbs shared between text type groups, excluding those verbs shared by all groups.

	EA	F	N	Tam	Targ
EA		3	9	8	6
F	3		8	7	6
N	9	8		10	4
Tam	8	7	10		9
Targ	6	6	4	9	

Table 11.18 Sum of the number of different common nouns and verbs shared between text type groups, excluding those nouns and verbs shared by all groups.

Each text type corpus includes nouns and verbs that occur relatively frequently within that corpus (i.e. ranked in the top 20 by frequency) but do not occur in the top twenty of any other corpus. These are:

EA seventh part, half, daytime, remnant, lord, enter, shine, increase, cover, adhere,
be dark, rule, diminish

F	tree, lamb, water, dream, mountain, fire, deed, giant, watcher, judgment, flesh, divination, generation, descend, understand, complain, change
N	south, ris, west, block, priest, measurement, bread or food, measure, complete, divide, go around, lead, rise, find or be able
Tam	father, brother, face or nose, wisdom, people, word or thing, hear, speak, answer, ask, die, command
Targ	blood, altar, darkness, province, bull, truth or righteousness, virgin, kid or Capricorn, balances, scorpion, archer or bowman, hit or strike, salt, wash, trample, sprinkle, write, incline or pray, thunder

As a passing comment, if these distinctly frequent words were used to indicate genre, Tam has six out of twelve words relating to speech, suggestive of narrative. N has nine, possibly ten, words out of fourteen relating to space, suggestive of a geographically descriptive genre. In Targ words with cultic associations predominate. The nature of F is less clear, but there is a tendency towards words common in apocalyptic literature. Finally, the EA corpus is the least easily classifiable, but two themes of light-darkness and ruling are present, which might be conveniently labelled as a dualistic-themed genre such as found in heroic literature. Given the proposed dating of the different corpora, this would be consistent with narrative and cultic texts giving way to more apocalyptic and dualistic texts as Second Temple Judaism develops in the West.

Summary

These final analyses confirm that in terms of textual characteristics Qumran Aramaic is heterogeneous with six distinct text types into which nearly all the substantial scrolls classify with high probability. The classification is robust because it was derived by one exploratory statistical method, hierarchical cluster analysis, and confirmed by another, latent class analysis. The text types have distinct characteristics that can be summarized by three factors: diachronic, geographical and Qumran idiosyncratic. Not only do these factors have face validity and internal psychometric (statistical) validity, but they also are supported to a limited extent by validation against independent, external correlates. The robustness of the classification supports the *a priori* textual criteria that were chosen and has identified six key alternative morpho-syntactic forms that would facilitate a less statistically-intensive classification of Aramaic texts of the period.

12. Discussion

The foregoing analyses demonstrate that the Aramaic texts found at Qumran can be considered as a corpus textually because nearly all the texts relate closely to other texts found there. However, this corpus is textually heterogeneous with six distinct text types emerging that can be characterised in terms of time, place and what we might think of as mainstream Qumranic style. Such an outcome is statistically satisfying, that is it manages to describe the data parsimoniously and comprehensibly, but before I briefly explore its implications, it is essential to provide a brief resumé of caveats.

Limitations to interpretation

Factors that limit the interpretation of the classification fall into three major areas. First, limitations relating to the data, that is the texts themselves. Second, limitations relating to the criteria chosen to classify the texts. Third, the methods used to classify the texts. I will address each of these areas in order.

Limitations of the data

First, we only have a limited amount of textual data available. We know that there was far more original material from the state of those texts that have come down to us where there are significant lacunae. It is possible that with more material we might be able to elucidate further text types or some of the discontinuities between text types may have

been filled in causing them to merge. Here, as throughout this thesis from Chapter 1, I am using the term “text type” in its sense of representing a latent textual (linguistic) trait that is discrete and characterisable rather than in the sense of conforming to a standard text (such as the Masoretic Text, the Septuagint etc). Using this definition of text type, the classification can only relate to what we know and it is difficult to conjecture as to what we do not know. The classification would be more certain if we can assume that what has been lost is, in statistical terminology, missing at random. For example, suppose there were Aramaic texts in a different style, perhaps reflecting a distinct provenance, that some factions strongly disagreed with and therefore destroyed, such texts would not be ‘missing at random’. On the other hand, if lacunae occurred within the fragments we have purely by chance, such data lost would be ‘missing at random’. I have not analysed where lacunae occurred and whether this might have introduced some sort of bias; this is really a task to be looked at in more detail on a text-by-text basis which is beyond the scope of this thesis.

Second, I have relied on the electronic form of the texts available in Accordance. These are based on the critical editions that are now available for the corpus; something that was not the situation not so many years ago. Nevertheless, these critical editions may still be open to revision. Muraoka’s grammar, for example, is sprinkled with instances where current readings are called into question.¹ Nevertheless, it is unlikely that such revisions will make major changes to proportions of textual variables used in the

¹ T. Muraoka (2011). The extensive footnotes bear testimony to the extent of uncertainty that remains despite the availability of critical editions of the texts.

classification undertaken in this thesis. Another source of error occurs in transcription from the published texts to the electronic versions. Given that these are used by many scholars, it is likely that most mistakes have been corrected as they are detected during use: the advantage of electronic texts over paper ones is that such corrections can be made as they are required rather than having to produce a new edition when enough corrections merit it.

Thirdly, nearly all the electronic texts I have used have already been grammatically tagged, and this is a further area where errors might occur. In fact, the major issues are more to do with interpretation of particles and occasionally verbal forms.² The principal textual criterion which this applies to is the particle *dy* which can have a variety of functions. Thus the counts for *dy* as a lexeme and the counts for the relative pronoun do not always equal each other. Usually it is obvious where *dy* is acting as a relative pronoun, but morphological categories derived from Indo-European languages do not always capture the semantic range.

Limitations of criteria used for classification

In many ways the limitations of the classification criteria derive from the characteristics of the texts (e.g. their length) and the quantitative approach to classification. In some

² This particularly applies to jussive forms which are morphologically indistinguishable from indicative forms in unvocalised texts in many instances – see T. Muraoka (2011), p101. Hence using jussive forms as a textual criterion for classification would have been likely to have introduced considerable error into the analyses.

senses the final choice represents a compromise. By choosing morphological and syntactic variables, I necessarily excluded very short texts where these would have inadequate power for a statistical distinction between texts to be likely to be detected. There remain, therefore, around fifty very short fragments that were excluded, some of which may or may not be in Aramaic. On the other hand, if I had restricted the classification to just a few far longer texts (e.g. 1Q20, 11Q18 etc) it may have been possible to use lexical criteria in addition. However, there are too few of these longer texts to make such a classification informative. A corollary is that I had to choose some cut off for the frequency of occurrence of a criterion to decide whether to retain it in the analyses. I fixed this cut off at 10% of texts to contain a specific form. There are two points that I considered. First, the group size of any text type, minded that having a text type group with just a few texts making it up may not be very helpful in terms of classification. Second, that the absence of forms would cluster texts together and this clustering would be likely to reflect the length of the texts. As it turned out, by excluding less frequent forms, the shorter texts have not clustered all together. Nevertheless, some of these less common criteria may be informative about textual affiliations, it is just that the degree of uncertainty about such conclusions would necessarily be high. It is also possible that scholars will identify novel criteria to differentiate Aramaic texts textually and that these might impact on the classification that has been presented in this thesis. However, such criteria would have to have quite large effects to weigh against the 22 criteria used and should there be such textual criteria which would outweigh these other criteria, it is surprising that they have not emerged hitherto.

Limitations of statistical methods

There is one major consideration that is a caveat to interpreting the statistical classification: the choice of a hierarchical clustering algorithm for the exploratory phase. This was chosen *a priori* on the assumption that any different Aramaic text types would derive from common text type predecessors. This is a standard assumption in linguistic studies. Explicitly, I have assumed that the text types that have emerged in this corpus of Aramaic texts found at Qumran can be traced back to some common ‘ancestor’ and that different forms of Aramaic did not spring up independently. It is reassuring, however, that latent class profiling, which is a non-hierarchical classification method, strongly supports the exploratory findings. Moreover, for nearly all the texts, the probability of class membership is close to 100% so that it is unlikely that we are dealing with a spectrum of texts with imperceptible differences between them. If we had been artificially imposing distinct textual classes on a continuum, we would have expected far more texts to have exhibited 60%-40%, 70%-30% probabilities of membership of two different classes. Other limiting issues, in particular the issue of statistical power, I have addressed prior to undertaking the analyses and need no repetition here.

Implications of the classification

Although the present classification answers one important research question - are the Aramaic texts found at Qumran textually homogenous? – and in demonstrating distinct text types no longer allows us to speak of ‘Qumran Aramaic’ as some single entity without qualification, the classification, hopefully, will prove useful beyond this point of departure to explore the new textual landscape. It is likely that a thorough exploration would take some time, so in this general discussion I will provide what I see as a general outline. This outline will survey implications for future classifications, text type group membership, and what this might mean in the broader context of Second Temple Judaism.

Implications for future classifications

This thesis demonstrates that a multivariate statistical approach is feasible for classification of Aramaic texts found at Qumran and should therefore be considered for other corpora. It is essential to be aware that classifications based on just one or two criteria would have failed to detect the heterogeneity present. On the other hand, the five alternative choice criteria explained just under 70% of the variance available from the 22 criteria classification, indicating that a relatively limited set of criteria are likely to be adequate for most purposes. Should further texts emerge from this epoch, these criteria would be likely to align them fairly well.

The classification can be considered multi-dimensional; again not something that could be achieved using just one or two criteria. The dimensions of space (geography) and time (date) are not unexpected, but the third dimension of ‘typicality’ perhaps deserves a little more discussion. Setting aside textual variation due to date or geographical provenance, typicality captures texts that align with a ‘mainstream’ Aramaic textual style. This style can hardly be typified as ‘literary Aramaic’ since extended literary compositions such as the Genesis Apocryphon (1Q20) are not part of these groups. The biblical text group, largely made up of court tales, is ‘typical’, however. The style may therefore reflect a scribal tradition related to that responsible for the biblical texts. We can recognise it intuitively because Aramaic grammars often take these biblical texts as their norm. Perhaps it might be thought of as the ‘official’ style as I suggested in the previous chapter crossing boundaries of time and space. If we describe such texts in this way, we may wish to consider the corollary that atypical texts are in some sense ‘unofficial’. They are not situated within the mainstream textually and this might provide insights into them beyond situating them in time and space.

Before turning to a discussion of the text-type groups and their characteristics, there is one further implication for future classifications and this is a methodological one. Using general mixture modelling, and specifically latent class or profile analysis advances linguistic classification in two important ways. First it provides a measure of model fit, that is how well the classification describes the observed data, which allows one possible classification to be compared with another. In the present case, determining the number of different text type groups that afforded the optimal classification. Secondly, I was

able to use the ‘training’ function to allocate texts that had previously failed to be captured in the classification to specific text type groups. This is potentially a very powerful classification tool. Moreover, it was able to identify a small number of texts that did not fit neatly into any one of the text type groups. The paradigm set out here of an initial exploratory phase to produce a limited classification where there was a high degree of certainty of group membership, confirming this statistically and then using this framework to include further texts is one that may well be useful in other contexts. In fact in other corpora, such as the New Testament, the certainty of authorship of some of the books is very high and this could then be used to determine the probability of other books being by the same author.

Implications of text type group membership

The classification groups texts according to date, geographical association and what I have termed typicality. Although the dating of texts corresponds well with the fairly scant evidence from carbon dating, the classification does not ascribe absolute dates to any group, rather the text groups are dated relative to each other. Importantly, this dating provides a stylistic terminus *ad quem* rather than a terminus *ab quo*. That is, the text type of the copy of the text found at Qumran is dated, but not the text itself which might have originated centuries before for all that can be inferred from the extant copy. Hence, for example, Enochic texts that belong to a late text type group, may relate to original texts that are older than some texts with an early text type dating. What is of interest is that such texts were still being copied at a later period and that the text was fluid enough for

its style to be modified. By the same token, an early text might have been copied at a later period, but for various reasons the copyist left the style unaltered; this might be the case for the Biblical text type group, for instance. After all, if we had found Masoretic texts of Daniel at Qumran that come from many centuries later, they would have been classified as early texts stylistically. Although it is possible that such texts have been deliberately written in an archaic style, a more simple conclusion is that these texts originate from an early period and their form has become fixed. Thus the text type groups provide a picture of scribal activity that probably stretches over two or three centuries.

The same type of considerations applies to geographical provenance as those that apply to dating. The text type of a particular text reflects scribal activity and may relate to the scribe's geographical affiliation as much as to the original provenance of the text.

Combining geographical associations with temporal sequence may provide insights into a group of texts origins, however. The same may hold for text type 'typicality' if we use this as a marker of 'official' or 'unofficial' scribal style. It may be helpful to illustrate these points with a few examples.

Biblical texts

First, let us consider the biblical text type as the simplest example. This has turned out to be quite distinct from any of the others. We might infer that this indicates that in general scribes during our period of interest were not striving to copy a biblical style to enhance the authority or authenticity of their texts. This is reassuring with regards to potential

archaising tendencies of scribes and probably means that we are fairly safe to assume the dating as it stands. The Biblical text type is early, Western and typical in style. The events related in the texts may be of Eastern origin, but the texts are not. They are written in an official style perhaps reflecting an official scribal school. All these features would be consistent with Daniel, and probably also Ezra, being recognised as official scripture by the time of Qumran, at latest 70 CE.

Tobit

Secondly, let us look at Tobit texts (4Q196, 4Q197 and 4Q198) since Tobit is represented in the Septuagint, though not the Masoretic Text. 4Q196 and 4Q197 belong to the text group which is middle period in date, Eastern and typical in style. This would place Tobit as being stylistically later than the Biblical text type group, suggesting that it was not fixed in form quite so early, yet retaining a geographical style consistent with its provenance. The style is 'official' which again we might take as reflecting that the texts were being copied by people who had been part of an official scribal school. 4Q198 is stylistically just a little later, is still in the official style, but is westernised. We might infer that the texts are being taken up more widely geographically within official scribal circles. Of course, it might be theoretically possible that all three Tobit texts were originally western in style and that significant scribal alterations occurred to adapt them to Eastern audiences (4Q196 and 4Q197) and in parallel 4Q198 was updated to a more modern Aramaic style in the West. This seems a more complicated explanation and there is no evidence to favour it over the simpler East to West transmission.

Postdiluvian testaments and visions

Thirdly, let us look at what are thought to be fairly early texts: 4Q534 and 4Q537 through to 4Q548, a collection of testaments and visions relating to patriarchal, postdiluvian figures mostly with priestly associations. 4Q535 and 4Q536, 4Q Aramaic N and 4Q Aramaic C are short fragments that are difficult to categorise in terms of content, so will be set aside for present. Nine of the thirteen texts (4Q534, 4Q537, 4Q538, 4Q541, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547) fall within the Mid-date, Eastern, typical style text group similar to that of the presumed early Tobit texts. 4Q542, the Testament of Qahat, is very similar, but somewhat earlier in style. 4Q540, the Apocryphon of Levi, is also stylistically earlier, Eastern, but not written in a typical mainstream Aramaic style. 4Q548, which may or may not be part of a Vision of Amram, is of a similar date to the main group, but is neither Eastern nor Western in style and, like 4Q540, is not in mainstream Aramaic. Lastly 4Q539 is late, Western and not typically mainstream Aramaic in style. In general, then, these texts represent mainstream scribal activity in an Eastern style at a date similar to that of the Tobit texts with 4Q542 a slightly earlier example. If any subsequent scribal copying has occurred, this has not altered the essential textual character of these scrolls. 4Q540 is earlier still than 4Q542 in the form we have it, Eastern, but not in a mainstream style. It has been labelled as an Apocryphon rather than a Testament or Vision and perhaps this also suggests that it does not quite belong with the other testamentary texts. Interestingly, it is in the same text type group as 4Q535 and 4Q536. Similarly, 4Q548 is textually distinct from the other Vision of Amram texts which may indicate that it is not really part of this group. It is from a similar date, but in an atypical style and with some Western stylistic influences

evident. This suggests quite independent scribal activity: it is in a text type group that stylistically resembles the targums, and which I will return to. Finally, 4Q539 is quite different textually, being late and Western and atypical. Like 4Q540 it has been called an Apocryphon indicating that there is evidence of scribal activity for this genre over a long period.

As we have seen, Tobit, the patriarchal testaments and the Visions of Amram are all textually similar. Table 11.14 shows the nouns that are common in these texts: they relate to people and family, God, time, spirit, peace and wisdom. Notably, this is the only text type group where wisdom occurs frequently. Table 11.15 lists the distinctive verbs and around half of these relate to conversation, probably influenced by the narrative nature of Tobit.

Enochic texts

Next we will consider texts related to that important antediluvian figure, Enoch. These texts are often grouped as the Astronomical book (4Q208 – 211), the Book of the Watchers (4Q201 and 4Q202; also sections of 4Q205 and 4Q206), the Book of Dreams (4Q207; also sections of 4Q205 and 4Q206), the Epistle (4Q211) and the Book of Giants (1Q23, 1Q24, 4Q203, 4Q530-532). I have not listed smaller fragments that pertain to the Enochic corpus at Qumran that have not been included in the classification. Of note is that none of these texts are included in the Tobit-postdiluvian patriarchal text type group which suggests that the Enochic corpus is quite distinct from this group. Thirteen of the sixteen texts are in the Western style, one, 1Q24 is neither Western nor Eastern, and two

(4Q203 and 4Q210) are distinctly Eastern. Interestingly, it is these two Eastern style Enochic texts that display the earliest textual style of the entire corpus. Moreover, the thirteen Western texts are also the latest in date. Interestingly, although the vast majority of Enochic texts are late and Western in style, with regards to typicality slightly fewer, ten out of sixteen, can be considered to be in a mainstream official style. The three earlier texts are all in an atypical, unofficial style. 4Q203, part of the Book of Giants, and 4Q210, part of the Astronomical Book are the earliest textually with 1Q24, a small fragment from the Book of Giants, slightly later. Textually, then, the evidence supports the Astronomical Book and the Book of Giants as existing at an earlier date than the other sections of Enoch, but the fact that other texts from the Astronomical Book are in a later style indicates that scribal activity and alterations were occurring over an extended period. We cannot be certain about the earliest dates of the other sections, only that they either arose later than either the Astronomical Book or the Book of Giants or that there was continuing scribal activity on these sections and we have not been left with anything in an earlier style.

Bringing this together, the most parsimonious explanation is that some sections of the Enochic corpus, the Astronomical Book and the Book of Giants, originated fairly early in the East with unofficial scribal activity, but that the material was taken up and transformed in the West, and possibly added to there, again mainly by unofficial scribal schools, but that at a later date manuscripts begin to appear in a more mainstream, official Aramaic style. The Eastern origin of the Book of Giants is consistent with it

containing Babylonian mythic material, that of the Astronomical Book is consistent with the Babylonian astronomical influence it manifests.

Enochic material makes up most of the Late Western typically mainstream Aramaic text type group, the other members being 4Q552 and 4Q553, both Four Kingdom texts and 4Q558 labelled as a Vision. This suggests that the scribal tradition interested in Enochic writings may also have been interested in these three shorter texts (there is some overlap between 4Q552 and 4Q553) which might all be characterised as apocalyptic.

Aramaic Levi texts

A small proportion of the Enochic texts are part of the late, Western, but not typically mainstream text type group. Three of the Aramaic Levi texts also are classified in this group: 4Q213a, 4Q213b and 4Q214a. Two of the other three Aramaic texts, 4Q214 and 4Q214b, are in the middle-dated, neither Western nor Eastern, atypical Aramaic text type group. The remaining text, 4Q213, is also of an intermediate date, but Eastern and with a typical mainstream style. This diversity indicates that scribal activity on these texts was not confined to a single time or place. Such a pattern would be consistent with the views of scholars such as James Kugel who hypothesise that Aramaic Levi is “stitched together” from at least two source documents and is not that early in date.³ If we were to speculate on the development of these texts, and this would involve

³ J. Kugel (2007). This Discussion is not the place to enter into an extensive analysis of various speculative constructions, but it is worth noting that the textual data are consistent with any final Aramaic Levi Document comprising several layers of editing. It is worth noting, however, that Kugel considers 4Q213 to be the oldest part of the document.

considerable speculation given that we have so few texts available on which to base any speculation, then we might conjecture that the initial section of the corpus arose in Eastern official scribal schools, but was added to over time by Western scribes writing in more unofficial styles. It is perhaps relevant that the initial section, 4Q213, describes an important role for Enoch, so perhaps the textual affinities between this corpus and the Enochic corpus at Qumran are not so unlikely.

New Jerusalem texts

The Aramaic New Jerusalem texts in the classification comprise 2Q24, 4Q554, 5Q15 and 11Q18. Three of these are early Eastern, not typically mainstream in style. 5Q15 is late Western and again not typical mainstream Aramaic. This is a small corpus of closely related books that are likely to have a fairly early date of original composition, but where there is some evidence of later Western scribal activity. This spatio-temporal pattern resembles that of Tobit, the Enochic corpus and the Aramaic Levi Document, but the New Jerusalem texts all display an ‘unofficial’, non-mainstream Aramaic style.

The ‘Targumic’ group

The text type group that most clearly represents the atypical, ‘unofficial’ style is the ‘Targumic’ group which is otherwise difficult to characterise in terms of date or geography. The group comprises:

1Q24, 4Q156, 4Q214, 4Q214b, 4Q242, 4Q246, 4Q318, 4Q548

I have already considered 1Q24 with the Enochic texts, 4Q214 and 4Q214b as part of the Aramaic Levi Document and 4Q548 along with the Visions of Amram. This leaves 4Q156 (Targum of Leviticus), 4Q242 (Prayer of Nabonidus), 4Q246 (Aramaic Apocalypse) and 4Q318 (Brontologion). This text type group has a distinctive lexical character (see Table 11.14) with common nouns unique to this group being: blood, altar, virgin, darkness, truth, bull, kid, balances, scorpion, archer, and province. The list is suggestive of cultic and possibly magical concerns. These texts have not undergone westernisation at a later date as far as the extant scrolls bear witness, though this cannot be excluded. The Aramaic style is quite unlike that found in mainstream, official text types.

4Q156 concerns altar rituals which are remarkably similar to those described in 4Q214 and not too dissimilar to those of 4Q214b. 4Q246's apocalyptic themes are reminiscent of those of 4Q548 and possibly also the very fragmentary 1Q24. 4Q242 is thought to be part of a court tale related to the book of Daniel. It describes a possible ritual exorcism and may thus share a focus on what might be termed the magical with 4Q318.

The Genesis Apocryphon and the Targum of Job

There remain two large texts that have not been considered so far, 1Q20 (the Genesis Apocryphon) and 11Q10 (the Targum of Job) since they are not conventionally classified with the other corpora so far reviewed. 1Q20 is textually late, western and not typically mainstream in style. It is in the same text type group as some of the Enochic

texts, some of the Aramaic Levi Document texts and 4Q539 labeled as an Apocryphon of Joseph. The relationship with Enochic texts is not unexpected given the Enochic content of 1Q20's initial columns. The relationship to 4Q213a, 4Q213b and 4Q214a would indicate similar late western scribal activity on narrative subjects related to the Book of Genesis. Together with the Targumic texts, this strengthens the idea that the Aramaic Levi Document is a mosaic of different texts: some focusing on priestly altar rituals and some on parabiblical narratives related to Levi. 4Q539 can also be viewed as parabiblical material related to a major figure from Genesis.

11Q10 is classified along with the Tobit and the postdiluvian testamentary texts as of intermediate date, Eastern and typically mainstream in style. Although it is labeled as a Targum, it shares little in common textually with later targums. There is no evidence at Qumran for any later western scribal activity on this text, but since we only have one substantial copy of this text and this only represents around 15% of the likely original text, one cannot be certain about this.

Implications for our understanding of Second Temple Judaism

In seeking to make inferences about Second Temple Judaism from the Aramaic texts found at Qumran, the limitations already noted need to be borne in mind. In particular, textual style does not necessarily inform about either the time or the place a text was written or copied. For example, some scribes may have persisted in using 'early' styles

until a late date and similarly scribes located in the West may still have used an Eastern style perhaps handed down to them by former generations. This is illustrated by the style of the Babatha archive which we can both date and place, but which uses an archaic style similar to our own contemporary experience of legal documents. It would thus be possible to hypothesize that all the texts classified into their different groups were actually written over a relatively short period in a single place such as Qumran or Jerusalem. However, at least with regards to date, radiocarbon evidence, albeit limited, would be against such a hypothesis. The textual classification presented here is consistent with the available radiocarbon dating, so can probably be given some weight in helping to indicate dates of texts relative to each other. If the texts that we have from Qumran reflect Second Temple Judaism to any degree, then there is a tendency for a shift over time from an Eastern to a Western style: the biblical books are an exception to this, perhaps either originally having a Western style or adopting this at an early point. It is, of course, quite possible that there was a lag between scribal movement from east to west and the resulting change in textual style. Nevertheless, this pattern suggests that over time the scribal centre had shifted with the main focus established in the west well before the end of the epoch covered by Qumran.

Given this east – west shift, it is useful to examine the early, Eastern, unofficial text style group as a possible point of origin for this trajectory. This group comprises 1Q21, 2Q24, 4Q203, 4Q210, 4Q244, 4Q529, 4Q535, 4Q536, 4Q540, 4Q550, 4Q550a, 4Q554, 4Q556, 4Q562, and 11Q18. There are three New Jerusalem texts (2Q24, 4Q554, 11Q18), two Proto-Esther texts, Astronomical Enoch (4Q210) and two Book of Giants (4Q203,

4Q556), Aramaic fragments N, C and D (4Q535, 4Q536, 4Q562), the Testament of Levi (1Q21) and an Apocryphon of Levi (4Q540) and the Words of Michael (5Q529).

Unsurprisingly, given the number of New Jerusalem texts, nouns related to measurement are common in this text group, but priest, bread and heavens are also group-specific.

There are several implications. First, the New Jerusalem and courtly tales material suggests an exilic perspective consistent with an early, eastern, non-mainstream style.

Secondly, stylistically the Astronomical Book of Enoch and the Book of Giants appear as relatively early texts arising in an eastern milieu: that this group of texts was expanded later in the west, stylistically, indicates the importance of this strand of Second Temple Judaism. The Words of Michael (4Q529) may form a link between the New Jerusalem and Enochic texts in that its major themes relate to angels and the building of a new city. Similarly, stylistic findings strengthen the hypothesis that 4Q535 and 4Q536 belong to the Enochic corpus. Thirdly, the two short texts 1Q21 and 4Q540 are stylistically close and raises the possibility that, even though they were found in different caves, they relate to a common source and worth considering together. This leaves 4Q562 stylistically aligned as early and eastern, but difficult to place with other texts in the group.

Continuing to project a tentative, and no doubt incomplete, picture of Second Temple Judaism from the Aramaic texts found at Qumran, having started from an exilic viewpoint we can try to map out the endpoint of this trajectory as represented by the late, Western, mainstream texts. This group comprises 1Q23, 4Q201, 4Q202, 4Q204, 4Q205, 4Q206, 4Q207, 4Q530, 4Q531, 4Q532, 4Q552, 4Q553, and 4Q558. The group is mostly

made up of Enochic texts (1Q23, 4Q210, 4Q202, 4Q204, 4Q205, 4Q206, 4Q207, 4Q530, 4Q531, 4Q532), along with two Four Kingdom texts (4Q552, 4Q553) and 4Q558 (so-called 'Vision'). This group has been discussed above under the 'Enochic' group. The implication is that there was strong interest, at least as far as the evidence from Qumran can tell us, in this apocalyptic viewpoint in late western scribal circles. Whether such circles were centered around Qumran or elsewhere it is not possible to determine.

Along with the late, Western typically mainstream text type group there is another group that must be placed near the end of the trajectory: the late, Western, atypical Aramaic style group. This group is more diverse and comprises 1Q20, 4Q208, 4Q209, 4Q211, 4Q213a, 4Q213b, 4Q214a, 4Q539, 4Q549, 4Q559, 4Q560, and 5Q15. The diversity of style within a late Western context would be consistent with texts deriving from several scribal groups. Again there is an Enochic group (4Q208, 4Q209, 4Q211), but also an Aramaic Levi group (4Q213a, 4Q213b, 4Q214a), a New Jerusalem text (5Q15), the Genesis Apocryphon (1Q20) and three short texts (4Q549, 4Q559, 4Q560) that do not appear to belong to any wider group. The Enochic group may represent a stage of textual transmission one step short of full mainstream assimilation of these texts. I have also discussed how the Aramaic Levi documents appear at this stage to result from a stitching together of two different sources: scrolls discovered at Qumran provide us with a window onto seeing Judaistic texts in the making. The evidence from Qumran demonstrates that Judaism was far from static, but was developing new positions by bringing together different strands from the past. 4Q549 and 4Q559 are both genealogies

that mention Aaron so perhaps might be speculatively allocated to the same scribal school. In some ways 1Q20 can appear as if it, too, is formed from two separate parts: an antediluvian, Enochic section and a postdiluvian patriarchal section. 4Q560 is quite enigmatic, labelled as an 'Exorcism', it appears to stand alone. It suggests an unaligned strand of Second Temple Judaism for which there is limited evidence at Qumran.

Having traced the stylistic points of departure and arrival, it remains to fill in what may be viewed as intermediate or transitional texts. The first text type group, which includes some fairly large texts, comprises 4Q196, 4Q197, 4Q213, 4Q243, 4Q534, 4Q537, 4Q538, 4Q541, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547, 4Q550c, 4Q561, 4Q570, 6Q8, 11Q10. It is Eastern and typically mainstream in style. The group comprises two Tobit texts (4Q196, 4Q197), five Visions of Amram texts (4Q543 to 4Q547) and four other postdiluvian patriarchal texts (4Q213, 4Q534, 4Q537, 4Q538, 4Q541), the Targum of Job (11Q10), one Book of Giants text (6Q8), two Persian Court Tales texts (4Q243, 4Q550c), a Physiognomy/Horoscope (4Q561) and a small fragment (4Q570). It therefore contains very limited Enochic material. Nouns and verbs that occur relatively frequently within that corpus (i.e. ranked in the top 20 by frequency) but do not occur in the top twenty of any other corpus for this group concern family relations, communication and wisdom. In fact, wisdom is a common theme in a large proportion of these texts, so that the Second Temple Judaism scribal school or schools which produced or copied them may have had wisdom as a major perspective. Another theme common at least in the Visions of Amram and Tobit is endogamy which may indicate

that these texts have some continuity with the exilic viewpoint of the earlier eastern style texts.

Finally, there is the text-type group with the most difficult to place style, though interestingly this style is close to that of the targums providing an intriguing possible trajectory evolving post-Second Temple Judaism. Similarly intriguing, is the stylistic similarity to the Wadi ed-Daliyeh texts that have a fourth century BCE Samaritan provenance. This affinity fits with the texts being neither predominantly Eastern nor Western in style and, speculatively, they might be considered the ‘missing link’ between early Samaritan texts and the later Palestinian targums. This relatively small group comprises 1Q24, 4Q156, 4Q214, 4Q214b, 4Q242, 4Q246, 4Q318, and 4Q548. One of these texts, 4Q156, is the Targum of Leviticus and two, 4Q214 and 4Q214b, are sections of Aramaic Levi. I have discussed above how these fragments relate to each other in terms of priestly ritual. The rest, though stylistically similar, are quite diverse in subject matter. 1Q24 is part of the Book of Giants, 4Q242 is a court tale about an exorcism, 4Q246 is an apocalyptic text, 4Q318 the Brontologion, and 4Q548, which is not stylistically grouped with the Visions of Amram, is also apocalyptic in nature. Hence, in contrast to the stylistically contemporary Eastern mainstream texts that represent a wisdom strand in Second Temple Judaism, these texts suggest continuity and development of an apocalyptic strand most likely by a scribal group interested in priestly ritual.

Qumran and its Aramaic texts

In surveying implications for Second Temple Judaism I have used Qumran like a prism to project the different colours of which it was comprised. Now I turn to look at the prism itself. If we include the Biblical text group, it is noticeable how the mainstream style text type groups, of whatever age, were largely deposited in Cave 4: Just three (1Q23, 6Q8 and 11Q10) out of a total of 38 texts are not from Cave 4. By contrast, six out of 35 atypical texts are not from Cave 4. Formally this is a statistically significant difference ($p < 0.05$), but the small numbers indicate caution in interpretation. Looking at text type groups individually would not facilitate interpretation at this level due to even smaller numbers of non-Cave 4 texts. Another reason to be careful about interpretation of this distribution is that a fair proportion of atypical style texts were also deposited in Cave 4. Nevertheless, it would be fair to infer that in general Caves 1, 2, 5, 6 and 11 contain proportionately more atypical texts than expected. There is no clear difference between Caves in terms of date or place of their texts' styles. This suggests that deposition of texts in Caves was not sequential; that is, texts were placed in one cave, this was then closed and texts were then placed in another cave, and so on. Either all caves were being used contemporaneously over a long period or, more likely, the scrolls were deposited in various caves fairly late. At least as far as the Aramaic scrolls are concerned, the most parsimonious explanation is that a large, relatively mainstream or official collection was placed in Cave 4 and that less mainstream collections were distributed amongst the other caves. Such a scenario would fit well with Cave 4

representing some official library; perhaps that from the Jerusalem Temple, as Norman Golb has argued.⁴ However, Golb contends that the Copper Scroll is an important Temple document and this was found in Cave 3. Thus, for the “Jerusalem hypothesis” to hold, we would have to contend that the main Temple library was deposited in Cave 4 and more contentious and less ‘official’ documents were hidden in other caves. On the other hand, Cave 4 might represent the putative Qumran library with groups from beyond Qumran depositing scrolls in other caves or possibly less official scrolls at Qumran being hidden there.

We cannot be certain if any of the scrolls as we have them originate from beyond Qumran: they may all have been copied there retaining their original text styles. Similarly, none of the scrolls may have been written or copied at Qumran as far as the text style evidence takes us. However, if we are to look for any scrolls that were written or modified at Qumran, those from the late, Western typical text type group would appear the most likely candidates, especially since all but one were found in Cave 4. If so, late Qumran scribes were strongly aligned with an Enochic strand of Second Temple Judaism which had a strong apocalyptic focus.

⁴ N. Golb (1995).

Future work

The implications for future classification using similar methodology to that I have applied to the Aramaic Qumran texts have already been mentioned. In fact, I have already begun some preliminary work of this sort with the Hebrew bible (see Appendix) that acted as a ‘proof of concept’. This initial exploration might be expanded, perhaps by breaking larger biblical books into smaller sections and performing latent class analysis as a second stage. Most obvious, though a project requiring far greater resources, would be to undertake a similar classification of the Hebrew Qumran texts. This classification could then be compared to the one I have produced for the Aramaic texts. In some ways such a classification would be easier because there is a larger body of scholarship on suitable textual criteria, but in others it will be more difficult, in particular the paucity of suitable non-Qumran texts as external correlates.

Robust classification, although essential, is only a departure point for future scholarship on the implications of the text groups and, along with this, Second Temple Judaism which I have touched on to point out potential ways forward in the previous sections. Far more in depth study is required to tease out and test these relationships. However, this classification provides a framework for such studies and, at the very least, textual affinities and differences will need to be kept in mind as an important context when examining individual scrolls. There is also the question of the very short, fewer than fifty words, Aramaic texts found at Qumran. The statistical approaches I have presented here are unsuitable for their classification. Possibly stochastic methods focusing on letter

sequences or very common features may have something to offer. How helpful it will prove to be able to align these brief fragments within an overall classification is unclear since, by their nature, such texts yield very little informative content: it may just be a case of intellectual curiosity to attempt to include them.

Conclusions

At the end of the introductory chapter I wrote,

The scrolls form a single corpus in that they were all discovered at Qumran, but beyond that the present classification is unable to tell us anything. In particular, it is of no use if we wish to understand the textual relationships between the different scrolls.

The new textual classification questions whether the scrolls truly form a single corpus. First, there are six distinct text types present with diverse styles both in terms of date and geographical affinity so that it is no longer possible to speak of ‘Qumran Aramaic’ without considerable qualification. Second, texts can be classified according to whether they display an ‘official’, mainstream style with proportionally fewer ‘unofficial’ style texts coming from Cave 4. This raises the real possibility that Cave 4 texts form a distinct entity or library. Third, the classification now provides a key to textual relationships not only between scrolls, but also with Aramaic texts from beyond Qumran allowing each scroll to be placed within a wider context of Second Temple Judaism. Fourth, the combination of exploratory and confirmatory phases coupled with a training

phase to expand the classification has proved to be feasible and statistically robust: it provides a paradigm for future classifications of texts that should prove valuable given the increasing number of electronic texts that are becoming available.

5. Materials examined

This chapter describes the electronic texts examined to obtain raw data for the statistical analyses. All of the Qumran Aramaic texts and the majority of the non-Qumran Aramaic texts were available in the Accordance Bible Software package. A small number of non-Qumran Aramaic texts were not available in Accordance, but were kindly made available by Professor Stephen Kaufmann from the Comprehensive Aramaic Lexicon project. Finally, for the purposes of power calculations, Hebrew Bible text data were extracted from the Bibleworks software package.

Accordance computerised texts

The primary source of tagged computerised texts was the Accordance 8 Bible Software: Oak Tree Software, 2008 (<http://www.accordancebible.com/index.php>). Two primary text modules were searched to provide statistical data: 1) the Qumran Sectarian Manuscripts (Hebrew, tagged), compiled and tagged with lemmas and grammatical information by Martin Abegg and 2) the Dead Sea Scrolls Biblical Manuscripts edited and tagged by Martin Abegg. The Qumran module contains all of the non-biblical Qumran texts. The Biblical Manuscripts module contains all the biblical Qumran texts together with biblical texts discovered elsewhere in the Judaean desert.

The texts are morphologically tagged which allows searching by morphological forms in addition to lexical (inflected or non-inflected forms) searching. The Accordance

program provides text-specific counts of terms searched for as well as highlighting terms found within the text itself.

The available morphological search terms are searchable by gender and number:

- 1) Verbs are searchable by person, stem, aspect and state. In addition verbs with apocopated endings can also be identified.
- 2) Nouns are searchable by state and type (common, proper, Gentilic)
- 3) Adjectives are also searchable by state.
- 4) Pronouns are searchable by class (independent or interrogative).
- 5) Pronominal Suffixes.
- 6) Particles are searchable by type (definite article, conjunction, adverb, interrogative particle, interjection, negative particle, direct object marker, preposition, and compound preposition article).

Texts can also be searched lexically by either lexical (lemma) or inflected form. Hence, searches can comprise:

- 1) Direct counts: these are generated by Accordance by book or other user-set limits of text range.

- 2) Manual counting of variants of morphological forms highlighted in text (e.g. verbal endings).
- 3) Direct counts of specific lexical or inflected forms where these are easy to identify as unique terms.
- 4) Manual counting of inflected forms where there may be more than one word with a very similar form in an unpointed text.

Aramaic texts were considered separately from Hebrew texts. Those from the Qumran Sectarian Manuscripts module with at least 50 words were:

1Q20	Genesis Apocryphon
1Q21	Testament of Levi
1Q23 and 1Q24	Enoch Giants
2Q24	New Jerusalem
4Q156	Targum of Leviticus
4Q196 to 4Q198	Tobit
4Q201 to 4Q212	Enoch
4Q213 to 214b	Aramaic Levi
4Q242	Prayer of Nabonidus
4Q243 and 4Q244	Pseudo-Daniel
4Q246	Aramaic Apocalypse
4Q318	Brontologion
4Q529	Words of Michael
4Q530 to 4Q532	Book of Giants
4Q534	Noah

4Q535	Aramaic N
4Q536	Aramaic C
4Q537	Testament of Jacob
4Q538	Testament of Judah
4Q539	Apocryphon of Joseph B
4Q540 and 4Q541	Apocryphon of Levi
4Q542	Testament of Qahat
4Q543 to 4Q548	Visions of Amram
4Q549	Work mentioning Hur and Miriam
4Q550, 4Q550a and 4Q550c	Proto-Esther
4Q552 and 4Q553	Four Kingdoms
4Q554	New Jerusalem
4Q556	Book of Giants
4Q557 and 4Q558	Vision
4Q559	Biblical Chronology
4Q560	Exorcism
4Q561	Physiognomy/Horoscope
4Q562	Aramaic D
4Q570	Aramaic R
5Q15	New Jerusalem
6Q8	Giants
11Q10	Targum of Job
11Q18	New Jerusalem

The labels are conventional rather than intrinsic to the texts: there are debates about the appropriateness of several of these and I will refer to them by classification number including the text label where this might be helpful.

Those from the Biblical Manuscripts were Aramaic sections of Daniel (2:4b to 7:28) and Ezra (4:9-11 and 5:17-6:5).

4Q112	Daniel
4Q113	Daniel
4Q115	Daniel
4Q117	Ezra

In addition to the Qumran texts, those from Accordance Targum modules were examined as external correlates, namely

Targum Onkelos – Pentateuch

Targum Jonathan – Prophets

Targums of the Writings

Targum Neofiti – Pentateuch

Targum Esther Sheni

Targum Pseudo-Jonathan – Pentateuch

The dates and places of composition of these Targums is unclear due to scribal accretions. Those of Onkelos and Jonathan, though they may have originated in Palestine, are considered to have an Eastern (Babylonian) character whilst the Targums Neofiti and Pseudo-Jonathan are considered more Western (Palestinian) in character, though possibly later in date.¹ With regard to the analyses in this thesis assumptions about 'Western' or 'Eastern' dialectical tendencies are not relevant. The Targums of the Writings are heterogeneous in style with that of Proverbs thought to be distinctly late and Eastern. Specific module details are:

Accordance TARG-T: Contains both Targum Onkelos and Targum Jonathan, fully pointed, based on manuscripts provided to the CAL by Bar Ilan University, checked against the standard edition of Alexander Sperber. TARG-T also contains the following Targums to the Writings: Targum Psalms (based on P. Lagarde's *Hagiographa Chaldaice*, Targum Job (based on the edition published by D. Stec), Targum Proverbs (based on Lagarde), Targum Ruth (based on the edition of Derek Beattie), and Targum Qohelet (Ecclesiastes), based on Ms. Paris, Bibliothèque Nationale 110, checked against the edition of P. Knobel, Targum Song of Songs from the text published by A. Sperber, the First Targum to Esther from the Paris 110 MS., Targum Lamentations from Sperber's text, and the Targum to Chronicles according to Codex Vaticanus Urbinati Ebr. 1, originally published by R. Le Déaut.

Accordance TARG2-T: Contains the full text, completely tagged, of Targum Neofiti, based on a fresh collation of the MS, carried out at the CAL project.

¹ See R. Le Déaut (1978-1981), M. McNamara (1992-1997) and P. Alexander (1992).

Accordance TARG3-T: Contains the full text, completely tagged, of Targum Pseudo-Jonathan, based on the only surviving MS, British Museum Add. 27031.

To provide external correlates predating Qumran Aramaic texts, the morphologically tagged Accordance module (TAD-T) "The Textbook of Aramaic Documents from Ancient Egypt by Bezalel Porten and Ada Yardeni, 4 volumes, Jerusalem, 1986-1999" was examined. This module includes a large collection of Aramaic documents, mainly letters and legal papers, from the 5th century BCE, mainly from Elephantine Island in the Nile near Aswan, composed by Jews in exile during the time of Ezra and Nehemiah. There are 64 separate texts in this collection divided into four volumes covering 1) Letters, 2) Contracts, 3) Literature, Accounts and Lists, and 4) Ostraca and Assorted Inscriptions.

The electronic text was prepared so that the line reference system attempts to follow the printed text as closely as possible, in which each book is a manuscript, and each chapter is a page or column, and each verse is a line number. For consistency, simplicity, and to avoid duplicate references, the following line reference modifications were made:

1. A3.1 (second Letter) lines 1-8 are now lines A3 1.8-15.
2. B8.2X.1-6 are now lines B8 2.35-40
3. C1.1 (first part) is now a separate book C0, with Column 2 (Plate L) as chapter 1 and Column 3 (Plate H) as chapter 2.
4. C1.1 fragments a1-h3 at the end are now respectively lines C1 1.223-244.

5. C2.1c1.1-3 is now C2 1.1-3; C2.1c3.1-4 is now C2 2.1-4.
6. C2.1c4 through C2.1c11 is now C2 3.1-79.
7. C3.2c1 through C3.2c8 is now C3 2.1-8; C3.3c2.1 through C3.3c2.17 is now C3 3.1-17.

8. C3.7A2-7A3 is now C37A 1-3.
9. C3.7B1-7B3 is now C37B 1-3.
10. C3.7C1-7C2 is now C37C 1-2.
11. C3.7D1-7D4 is now C37D 1-4.
12. C3.7E1-7E2 is now C37E 1-2.
13. C3.7K1-7K4 is now C37K 1-4.
14. C3.7G1-7G4 is now C37G 1-4.
15. C3.7F1-7F3 is now C37F 1-3.
16. C3.7J1-7J3 is now C37J 1-3.
17. C3.7J1-7J2 (VERSO) is now C37JV 1-2.
18. C3.7F1-7F3 (VERSO) is now C37FV 1-3.
19. C3.7G2-7G5 (VERSO) is now C37GV 1-5.
20. C3.7K2-7K3 (VERSO) is now C37KV 1-3.
21. C3.7E1-7E2 (VERSO) is now C37EV 1-2.
22. C3.7D1-7D3 (VERSO) is now C37DV 1-3.
23. C3.7B1 (VERSO) is now C37BV 1.
24. C3.7A1 (VERSO) is now C37AV 1.
25. C3.7H1-7H2 is now C37H 1-2.

26. C3.7L2 is now C37L 1-2.
27. C3.8.1-8 is now C38 1-8.
28. C3.8.3 fragments are now C38 3.14-19.
29. C3.8.4.18a is merged into C38 4.18.
30. C3.9-29 is now C3A 9-29 with C3A 1-8 added as filler references.
31. C3.9.20-24 is now C3A 9.21-25 respectively, because of a numbering error.
32. C3.28.51a is merged into C3A 28.51.
33. C3.28.54a-h is merged into C3A 28.54.
34. C3.28 fragments at end are now respectively C3A 28.129-132.
35. C4.2 fragments a1-e2 at end are now respectively C4 2.14-30.
36. C4.9 column 2.1-6 at end are now respectively C4 9.7-12.
37. D1.6 fragments a1-c2 are now respectively D1 6.1-4.
38. D1.17 fragments d1-d2 are now respectively D1 17.13-14.
39. D1.26 fragments b1-b6 are now respectively D1 26.5-10.
40. D1.31 fragments b1-b5 are now respectively D1 31.2-6.
41. D1.33 fragments b1-d4 are now respectively D1 33.3-14.
42. D1.34 fragments a1.1-e.4 are now respectively D1 34.1-20.
43. D2.7 fragments b1-b2 are now respectively D2 7.3-4.
44. D2.10 fragments a1-d1 are now respectively D2 10.13-19.
45. D2.14 fragments c1-d3 are now respectively D2 14.3-7.
46. D2.29 fragments b1-b4 are now respectively D2 29.5-8.
47. D2.31 fragments b1-b4 are now respectively D2 31.3-6.

48. D2.32 fragments b1-b2 are now respectively D2 32.3-4.
49. D3.2 fragments b1-c1 are now respectively D3 2.7-10.
50. D3.3 fragments b6-b8 are now respectively D3 3.6-8.
51. D3.9 fragments b1-b3 are now respectively D3 9.4-6.
52. D3.24 fragments b1-c2 are now respectively D3 24.2-7.
53. D3.38 fragments b1-e2 are now respectively D3 38.3-12.
54. D3.39 fragments b1-b6 are now respectively D3 39.4-9.
55. D3.41 column 2 numbers are now respectively D3 41.4-6
56. D3.46 fragments b1-e6 are now respectively D3 46.4-19.
57. D3.47 fragments b1-b11 are now respectively D3 47.6-16.
58. D4.15 fragments b1-b3 are now respectively D4 15.3-5.
59. D4.23 fragments b1-b2 are now respectively D4 23.3-4.
60. D5.33 fragments c1-c2 are now respectively D5 33.4-5.
61. D5.41 fragments b1-ww1 are now respectively D5 41.2-81.
62. D5.52 fragments a.1-k.1 are now respectively D5 52.1-16.
63. D5.54 fragments b1-c4 are now respectively D5 54.6-12.
64. D6.3 fragments b1-11 are now respectively D6 3.11-23.
65. D6.4 fragments b1-i1 are now respectively D6 4.3-13.
66. D6.5 fragments b1-f1 are now respectively D6 5.4-10.
67. D6.6 fragments b1-n1 are now respectively D6 6.4-20.
68. D6.7 fragments Ia1-Od3 are now respectively D6 7.1-21.
69. D6.8 fragments b1-n1 are now respectively D6 8.2-30.
70. D6.9 fragments b1-11 are now respectively D6 9.4-29.

71. D6.10 fragments Ia1-Og1 are now respectively D6 10.1-18.
72. D6.11 fragments b1-d1 are now respectively D6 11.2-5.
73. D6.12 fragments b1-p1 are now respectively D6 12.2-19.
74. D6.13 fragments b1-n1 are now respectively D6 13.3-15.
75. D6.14 fragments b1-p3 are now respectively D6 14.4-25.
76. D13.2 Side B.a-c are merged into D13 2.14.
77. D19.2,b1-b2 are now respectively D19 2.3-4.
78. D22.9,b1-b2 are now respectively D22 9.3-4.
79. D23.1.5A-5B are now respectively D23 3-4.
80. D23.1.16A.1-16C.6 are now respectively D23 16.1-10.

Of note is that the Comprehensive Aramaic Lexicon project classifies TAD A1 under Imperial Mesopotamian dialect, and TAD A6 and TAD C2.1 as Imperial Persian Administration dialect.

Finally a limited number of mostly short texts were available from the JUDEAN-T module from “Discoveries in the Judean Desert (Oxford: Clarendon Press, 1955-) and Talmon, Shemaryahu. *Masada VI, Yagael Yadin Excavations 1963-1965, Final Reports*. Jerusalem: Israel Exploration Society, 1999”. This module was also edited by Martin G. Abegg, Jr. Very brief texts that comprise only names, one or two words or just fragments of words were excluded. This left the following texts (with the DJD source of the transcriptions) for consideration:

- WDSP1 Gropp, DJD XXVIII Deed of slave sale
- WDSP2 Gropp, DJD XXVIII Deed of slave sale
- WDSP3 Gropp, DJD XXVIII Deed of slave sale
- WDSP4 Gropp, DJD XXVIII Deed of slave sale
- WDSP5 Gropp, DJD XXVIII Deed of slave sale
- WDSP6 Gropp, DJD XXVIII Deed of slave sale
- WDSP7 Gropp, DJD XXVIII Deed of slave sale
- WDSP8 Gropp, DJD XXVIII Deed of slave sale
- WDSP9 Gropp, DJD XXVIII Deed of slave sale
- WDSP10 Gropp, DJD XXVIII Deed of pledge of slave
- WDSP15 Gropp, DJD XXVIII Deed of house sale
- WDSP18 Gropp, DJD XXVIII Deed of slave sale
- Mur8 Milik, DJD II Cereals and vegetables
- Mur10 Milik, DJD II Palimpsest accounts and abecedary
- Mur18 Milik, DJD II Dated in the second year of Nero Caesar
- Mur19 Milik, DJD II Writ of divorce dated 111 CE
- Mur20 Milik, DJD II Marriage contract dated 117 CE
- Mur21 Milik, DJD II Marriage contract undated
- Mur23 Milik, DJD II Deed of sale dated 132 CE
- Mur25 Milik, DJD II Deed of sale of land dated 133 CE
- Mur26 Milik, DJD II Deed of sale
- Mur27 Milik, DJD II Deed of sale

- Mur28 Milik, DJD II Property deed
- Mur31 Milik, DJD II Deed fragments
- Mur32 Milik, DJD II Deed
- Mur33 Milik, DJD II Deed
- Mur72 Milik, DJD II Ostraca text ?genre
- Sdeir2 Yardeni, DJD XXXVIII Debt contract
- XHev/Se7 Yardeni, DJD XXVII Deed of sale
- XHev/Se8 Yardeni, DJD XXVII Deed of sale of a house in Kefar Baru in the days of
Bar Kokhba
- XHev/Se8a Yardeni, DJD XXVII Sale of a house by Hadad to Ele'azar
- XHev/Se9 Yardeni, DJD XXVII Sale of property
- XHev/Se9a Yardeni, DJD XXVII Unclassified
- XHev/Se10 Yardeni, DJD XXVII Receipt for payment of fine
- XHev/Se11 Yardeni, DJD XXVII Marriage contract
- XHev/Se12 Yardeni, DJD XXVII Receipt
- XHev/Se13 Yardeni, DJD XXVII Waiver
- XHev/Se21 Yardeni, DJD XXVII Deed of sale
- XHev/Se22 Yardeni, DJD XXVII Deed of sale
- XHev/Se23 Yardeni, DJD XXVII Deed of sale
- XHev/Se24 Yardeni, DJD XXVII Deed
- XHev/Se24a Yardeni, DJD XXVII Deed
- XHev/Se25 Yardeni, DJD XXVII Deed
- XHev/Se26 Yardeni, DJD XXVII Deposits and barley

- XHev/Se31 Yardeni, DJD XXVII Deed
- XHev/Se32 Yardeni, DJD XXVII Deed
- XHev/Se33 Yardeni, DJD XXVII Unclassified fragment
- XHev/Se34 Yardeni, DJD XXVII Deed
- XHev/Se40 Yardeni, DJD XXVII Unclassified fragment
- XHev/Se50 Yardeni, DJD XXVII Deed of sale
- 34Se3 Morgenstern, DJD XXXVIII Wadi Seiyal 12 tiny fragments, only fragment 2
has partially legible lines

Mur = Murabba'at.

Sdeir = Wadi Sdeir.

WDSP = Wadi ed-Daliyeh.

XHev/Se = Nahal Hever/Nahal Se'elim (Wadi Seiyal) uncertain cave.

34Se = also Wadi Seiyal 12 tiny fragments, only fragment 2 has partially legible lines.

As noted above, a few of these texts can be dated later than Qumran (e.g. Mur19, Mur20, Nahal Hever texts) and some probably earlier (Wadi ed-Daliyeh representing Samaritan papyri from the time of Alexander the Great). These texts are mostly not contemporaneous with Qumran (Mur18 may be an exception), though there may be some overlap for the very oldest Qumran Aramaic texts (e.g. biblical Daniel and Esther). They represent a very narrow range of genres, mainly legal documents. They are considered together purely on geographical criteria.

Bibleworks computerised texts

Bibleworks 6.0 was used for exploratory statistical analyses of the BHS using the *Groves-Wheeler Westminster Hebrew OT Morphology* database. The Qumran biblical manuscripts are not available from Bibleworks. Morphological searching criteria are essentially identical to Accordance except the compound particle is treated as two separate words: preposition and definite article.

Comprehensive Aramaic Lexicon computerised texts

These are available at <http://cal1.cn.huc.edu/> but can only be viewed as one verse at a time. Searching is essentially limited to lexical forms on the website though morphological data are available by requesting parsing of each individual verse. Since the planned statistical analyses relied on morphosyntactical features the data available from the website was inadequate. In view of this Professor Stephen Kaufman, Hebrew Union College, Cincinnati, USA kindly supplied the morphological codes as plain text files of the Comprehensive Aramaic Lexicon (CAL) Palestinian Aramaic texts. These comprised:

50300 Contracts and Financial Obligations Mur 8, 27, 32, 31, 33, 18, 23, 28,

25

50301 Marriage and Divorce Mur 19, 20, 21

50302 Miscellaneous Mur 34, 35 ,72

50303 Documentary Texts from Nahal Hever XHev/Se 7, 8, 8a, 9, 9a, 10, 11,
12, 13, 14, 21, 22, 23, 24, 24a, 25, 26, 27, 28, 31, 32, 33, 34, 35, 37, 50

50304 Babatha Archive: Subscriptions and Signatures to Greek texts pYadin
12, 14, 15, 16, 17, 18, 19, 20, 21, 22, (23), 26, 27

50305 Babatha Archive: Legal Texts 5/6Hev 7 (pYadin 7), 8 (8), 10 (10), 42
(42), 43 (43), 47 (47)

50306 Ostraca and Jar Inscriptions from Masada Masada 391, 398, 399, 401,
402, 408, 413, 414, 415, 420, 421, 422, 423, 425, 426, 432, 438, 440, 455, 458,
460, 461, 462, 463, 464, 468, 469, 470, 471, 480, 481, 482, 483, 502, 503, 504,
506, 507, 512, 513, 544, 546, 552, 553, 558, 559, 560, 561, 562, 563, 564, 565,
566, 567, 569, 571, 572, 573, 574, 575, 576, 577, 578, 582, 583, 613, 657, 667,
668, 670, 671, 674, 698

50307 Various Ostraca from Israel David 1, 2, Jerusalem 1, Aroer 1

50350 Reports on Delivery of Goods

50550 Bar Koseba Letters 5/6 HevEp 1 (PYadin 54), 2 (63), 4 (53), 8 (50), 10 (58), 11 (56), 14 (55), 15 (57)

50551 Letters from Masada Masada 554, 555, 556

The following files were listed on the CAL website and sent by Professor Kaufman but contained no data:

50351 34Se papDeed ar No data on CAL

50352 Documents from Ketef Jericho No data on CAL

50353 Wadi Sdeir 2 (WS 2) No data on CAL

50552 Mur.72 No data on CAL

Fortunately, texts for these files were available from the Accordance JUDEAN-T module as were those for CAL 50300-50303.

The Ostraca and Jar Inscriptions from Masada (50306) mostly consist of single names. None approach the minimum length of 50 words to be adequate for statistical analyses. This also applies to 50307, 50350 and the Letters from Masada (50551). Of the Babatha

Archive: Subscriptions and Signatures to Greek texts (50304), the following texts (by pYadin number) have at least 50 words: numbers 17, 18, and 20. Similarly, for Babatha Archive, Legal Texts 5/6Hev (50305), the following texts have at least 50 words: numbers 7, 8, 10, 42, 47A, and 47B. Finally, for the Bar Koseba Letters 5/6 HevEp (50550), only letter number 1 has at least 50 words and, unfortunately the CAL plain text morphological code file did not match the text as available from the CAL website. In view of this discrepancy, this text was excluded from further consideration.

The final CAL materials were thus:

50304 Babatha Archive: Subscriptions and Signatures to Greek texts

50304017 pYadin17

Text as per *The Documents from the Bar Kokhba Period in the Cave of Letters: Aramaic and Nabatean Signatures and Subscriptions*, ed. Y. Yadin and J. C. Greenfield, Jerusalem, Israel Exploration Society, 1989, p. 73; plates 15 and 16.

Cf. *Ergänzungsband of Die aramäische Texte vom Toten Meer (ATTME)*, p. 177.

autograph attestations of a Greek deposition, date: 21 February 128

50304018 pYadin18

Text as per *The Documents from the Bar Kokhba Period in the Cave of Letters*:

Aramaic and Nabatean Signatures and Subscriptions_, ed. Y. Yadin and J. C. Greenfield, Jerusalem, Israel Exploration Society, 1989, p. 79-80; plates 18 and 19.

Cf. ATTME, p. 178.

autograph attestations of a Greek marriage contract

date: 5 April 128

50304020 pYadin20

Text as per The Documents from the Bar Kokhba Period in the Cave of Letters:

Aramaic and Nabatean Signatures and Subscriptions_, ed. Y. Yadin and J. C. Greenfield, Jerusalem, Israel Exploration Society, 1989, p. 91; plates 23 and 24.

Cf. ATTME, p. 180.

duties in a Greek deed of inheritance and subscriptions

date: 19 June 130

50305 Babatha Archive: Legal Texts

Text as per Yigael Yadin, Jonas C. Greenfield, Ada Yardeni, and Baruch A. Levine, The Documents from the Bar Kokhba Period in the Cave of Letters, Jerusalem: Israel Exploration Society, Institute of Archaeology, Hebrew University, and Shrine of the Book, Israel Museum, 2002.

50305007 5/6Hev 7 (pYadin 7)

Text as per Yardeni, DBKP, pp. 80-88; fig. 4-7 (pp. 76-79).

Deed of Gift

50305008 5/6Hev 8 (pYadin 8)

Text as per Yardeni, DBKP, p. 113; fig. 8 (p. 112).

Purchase Contract

50305010 5/6Hev 10 (pYadin 10)

Text as per Yardeni, DBKP, p. 126; fig. 9-12 (pp. 122-25).

Babatha's Ketubba

50305042 5/6Hev 42 (pYadin 42)

Text as per Yardeni, DBKP, p. 144; fig. 13 (p. 145).

Lease Agreement

50305047 5/6Hev 47 (pYadin 47)

Text as per Yardeni, DBKP, p. 162; fig. 15-16 (pp. 160-61).

Sale Contract

This contract is composed of two sections (hence, 47A and 47B), written by two different scribes, in different directions (Yardeni, p. 157).

Although restricted in genre, these texts are valuable in that they are dateable with a clear Palestinian provenance. Some legal texts (e.g. pYadin 7 and 10) contain duplicated

sections, although the duplication is not absolutely exact. They were discovered in the Cave of Letters near the Dead Sea to the south of Qumran.

The CAL files have an idiosyncratic morphological coding unrelated either to that in BibleWorks or Accordance (the BibleWorks *Groves-Wheeler Westminster Hebrew OT Morphology* equivalents are given for clarity in the second column below):

CAL	BW	meaning
N01	nc-se	singular absolute or determined
N02	nc-sd	singular determined
N03	nc-pa	plural absolute
N04	nc-pc	plural construct
N05	nc-pd	plural determined
A	a---	adjective
A01	a-se	singular absolute or determined
A02	a-sd	singular determined
A03	a-pa	plural absolute
A04	a-pc	plural construct
A05	a-pd	plural determined
I	Pi	interjection
I01	Pi	
I03	Pi	

P	p---	Pronoun
P01	p---	independent
P02	p---	
P03	p---	
V	v	verb
V01	vN	peal
V02	vM	pael
V03	vB	(h) afel
V04	vF	itpeel
V05	vS	itpaal
V06	vE	ettaphal
V07	v1	pay/w(el
V08	v2	reduplicated
V09	v3	quadrilateral
V10	v4	ethpaiel
V11	v5	reflexive reduplicate
V12	v6	reflexive quadrilateral
V01C	vN	
a	Pd	adverb
a01	Pd	singular absolute or determined
a02	Pd	singular determined
a03	Pd	plural absolute
a04	Pd	plural construct

a05	Pd	plural determined
c	Pc	conjunction
c01	Pc	
c03	Pc	
d	nd---	divine name
d01	nd---	
d02	nd---	
d05	nd---	
n	an--	numeral
n01	anse	singular absolute or determined
n02	ansd	singular determined
n03	anpa	plural absolute
n04	anpc	plural construct
n05	anpd	plural determined
p	Ppx	p = preposition
l-	Ppl	l- (p.) = to, for
p01	Ppi	independent
p02	Pps	with pronominal suffix
p03	Ppp	proclitic
PN	np---	proper name
GN	np---	geographical name

Inspection shows that there are some limitations to this scheme. First, the state of nouns and adjectives (and adverbs) is unclear in the singular person. For example N01 is stated to designate singular nouns in the absolute or determined state and N02 is stated to designate singular nouns in the determined state. Inspection of the data files reveals that N02 does designate singular nouns in the determined state. N01 thus represents singular nouns in either the absolute or construct state. There is a clear morphological distinction between the absolute and construct state for feminine singular nouns in Aramaic with construct forms ending with *taw*. Masculine singular nouns are indistinguishable morphologically between absolute and construct state; however, the construct state is easily identifiable in context. With this regard, nouns with pronominal suffixes have been considered to be in the construct state. Secondly, as just noted, there is no separate coding for masculine and feminine forms. Thirdly, and a minor issue, there are no separate categories for dual forms. Dual forms are generally stereotypical and usually feminine, limited to parts of the body and a few other lexemes that have survived; it is far commoner in Ugaritic, for example, though Old Babylonian uses the dual in much the same limited situations to Aramaic and does not use the dual form at all for adjectives.

The coding of verbs is similarly limited. Verbal stem/theme is provided, but not 'tense' (finite and non-finite), person or number. Particles are also not classified adequately for the present purposes of morphological analyses. For example, there is no separate indication of relative particles.

The raw CAL plain text files were therefore recoded to provide:

- 1) state, number and gender for all nouns and adjectives;
- 2) tense, number and gender for all verbs;
- 3) classification of particles as conjunction, adverb, interrogative, interjection, negative, object marker, preposition or relative.

The Westminster Hebrew Morphology coding, as adapted for Aramaic, was used:

Position 1	Part of speech P-particle, p-pronoun, n-noun, a-adjective, v-verb
Position 2	Subclassification

Particles

c – conjunction, d – adverb, g – interrogative, i – interjection,
n – negative, o – object marker, p – preposition, r – relative

Pronouns

i – independent, s – suffix

Nouns

c – common, p – proper

Adjectives - This field set to zero

Verbs

A – afel, B – hafel, H – hishtafel, S – hitpaal, F – hitpeel, G – hitpolel,

I – ishtafel, L – itpeel, P – polel, R – shafel, N – peal, M – pael, O – peil,

D – hofal, Q – safal, K – itpaal, V – itpoel

Position 3 Gender

m – male, f – female, c- common

This field set to zero for particles

Position 4 Number

Nouns and adjectives s – singular, p – plural

Verbs and pronouns 1, 2 or 3

This field set to zero for particles

Position 5 State

Nouns and adjectives

a – absolute, c – construct, d – determined

Verbs

p – perfect, i – imperfect, v – imperative, c – infinitive construct,

a – infinitive absolute, P – participle, s – passive participle

This field is set to zero for other parts of speech

CAL does not count a determined state noun as two words, unlike Accordance. To ensure comparability, word counts for CAL texts were adjusted for the number of determined state nouns and adjectives so that these were equivalent to Accordance counts. To ensure equivalence, proper nouns were categorized as being in the absolute or emphatic state despite being definite, and thus “determined”, by nature.

A few further adjustments are also required. CAL morphology has a separate category for verbs with quadrilateral roots. In practice such verbs can still be considered to conform to a standard root formation (peal, pael etc) and have been designated as such. *dnh* is designated as an adjective in BibleWorks Westminster Hebrew Morphology coding; in converting CAL coding it is retained as a pronoun in all contexts. *kl* is considered as a pronoun in CAL files, but as a noun in construct state in BibleWorks Westminster Hebrew Morphology coding, and the latter has been chosen here. *dy* is always considered as a conjunction in CAL. This uniform classification of *dy* as a conjunction follows Ursula Schattner-Rieser.² However, *dy* is not always used as a subordinating conjunction in these texts, so the classification set out by Rosenthal in his Grammar³ has been followed.

² U. Schattner-Rieser (2004), p96. But *dy* is also classified as a relative pronoun and as a form of genitive construction on p115.

³ F. Rosenthal (2006). *dy* is classified as a relative pronoun except after verbs meaning to know, inform etc, used as “so that” to introduce final clauses and to introduce direct speech (section 86) when it is classified as a conjunction, or as a genitive construction (section 48).

6. Statistical methods

Having outlined statistical approaches relevant to Aramaic texts at Qumran in Chapter 2, I now turn to describing the specific statistical methods to be used. I will use data from the Hebrew Bible (Masoretic Text) as ‘proof of principle’ to ascertain which statistical analyses are likely to be adequately powered and yield informative information. The Hebrew bible, here, will be used because it is familiar and thus results obtained can be easily set in the context of substantial scholarship. Moreover, although it is almost entirely written in Hebrew, textual criteria are better worked out for statistical use in this language. Hence we have well-studied texts and fairly well-evaluated textual criteria that can determine whether proposed statistical methods are likely to be applicable. It would be difficult to argue for the use of methods which are not particularly informative for the Hebrew bible texts, for less-explored areas such as Qumran Aramaic texts.

Appropriate statistical methods depend on the hypothesis under consideration (e.g. testing for significant differences or associations) and the nature and quantity of data being tested. Variables can be described as interval, for which additive and multiplicative rules apply (e.g. money), ordinal, for which multiplicative and additive rules do not apply but which can be ranked (e.g. positions in a race), and categorical, which are essentially counts of any particular category (e.g. nouns, verbs and adjectives) but cannot be ranked. Different statistical tests have been designed to handle these various data types. These tests can be grouped into those that test for differences between variables and those that test for associations. The quantity of

data available (not the values of individual data), known as the sample size, is important in determining the statistical approach. If the sample size is too small, it is unlikely that a statistical test will be able to detect a difference even if such a difference truly exists. It is therefore important to consider the power of a statistical test to detect significant differences in a given sample size prior to using them. This initial step involves power calculations.

Power calculations – general background

Power calculations determine the probability (β) of failing to detect a true difference or association of a given amount (effect size, often termed f) in a sample of a given number (N) at a given level of significance (α). The level of significance chosen is arbitrary. Conventionally α is set to .05, a 5% probability that an effect is detected by chance when no true effect of that size exists. This means that if twenty hypotheses are being tested, one will return a significant result by chance. Hence, the more hypotheses there are to be tested, the lower α should be set if chance significant results are not wanted. Inferring that a significant result is meaningful when it is likely that it may have occurred by chance is termed Type 1 statistical error.

The choice of β is also arbitrary. $1-\beta$ is termed power, and when multiplied by 100 (probabilities lie between zero and one), can be considered the percentage likelihood of detecting an effect of a given size if it is truly there in the population being sampled. A β of .2, for example, would mean that a true effect would be detected 80% of the time. Of course, by chance, 20% of the time it will not be detected. When

this occurs it is termed Type 2 statistical error. The lower the power, the more likely Type 2 statistical error will occur. Significance level (α) and power ($1-\beta$) are interdependent: reducing α results in a greater N being required to achieve the same power.

Effect size is considered differently depending on whether the variables being tested are interval, ordinal or categorical. Categorical variable effect sizes are usually considered in terms of proportions. Ordinal variable effect sizes are usually considered in terms of differences or association of ranks. Interval data effect sizes are usually considered in terms of the variance (variability) of the data measured by the standard deviation. For data that are normally distributed, less than 5% of the population fall beyond two standard deviations from the mean. Similar to α and β , effect size may be an arbitrary choice when no prior data are available. However, often data from other related studies are available that inform about likely effect sizes.

Book	Number of words	Book	Number of words
Genesis	28,832	Ecclesiastes	4,240
Exodus	23,913	Song of Songs	1,702
Leviticus	17,208	Isaiah	23,184
Numbers	23,356	Jeremiah	30,104
Deuteronomy	20,284	Lamentations	2,026
Joshua	14,597	Ezekiel	26,434
Judges	14,161	Daniel	8,745
Ruth	1,805	Hosea	3,174
1Samuel	19,079	Joel	1,325
2Samuel	15,737	Amos	2,814
1Kings	18,791	Obadiah	392
2Kings	17,392	Jonah	986
1Chronicles	15,814	Micah	1,915
2Chronicles	19,945	Nahum	749
Ezra	5,606	Habakkuk	906
Nehemiah	7,958	Zephaniah	1,044
Esther	4,646	Haggai	884
Job	10,955	Zechariah	4,509
Psalms	25,388	Malachi	1,193
Proverbs	8,923		

Table 6.1 Number of words, Hebrew and Aramaic, per book as calculated by Bibleworks 6.

Sometimes sample size (N) is the outcome of interest in power calculations when collecting data prospectively. However, in the case of textual analyses, N is already fixed. Table 6.1, for example, shows the number of Hebrew words per biblical book. Figure 6.1 displays these data as a histogram that reveals that just under half of the books have fewer than 5,000 words and above this there is a wide spread of words per book. Figure 6.2 displays word length for those books with fewer than 5,000 words. Many have word lengths around 1,000 words. Hence, although the mean number of words per book in the Hebrew bible is just over 11,000, for the purposes of power calculations an N of 1,000 is more appropriate if we want to compare books. Even then, Obadiah is likely to be problematic.

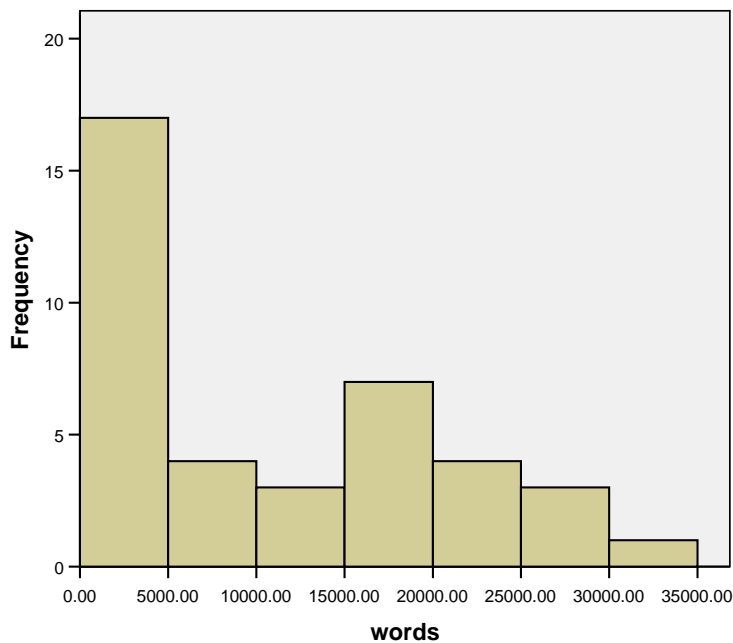


Figure 6.1 Frequencies of Hebrew bible books grouped by word count.

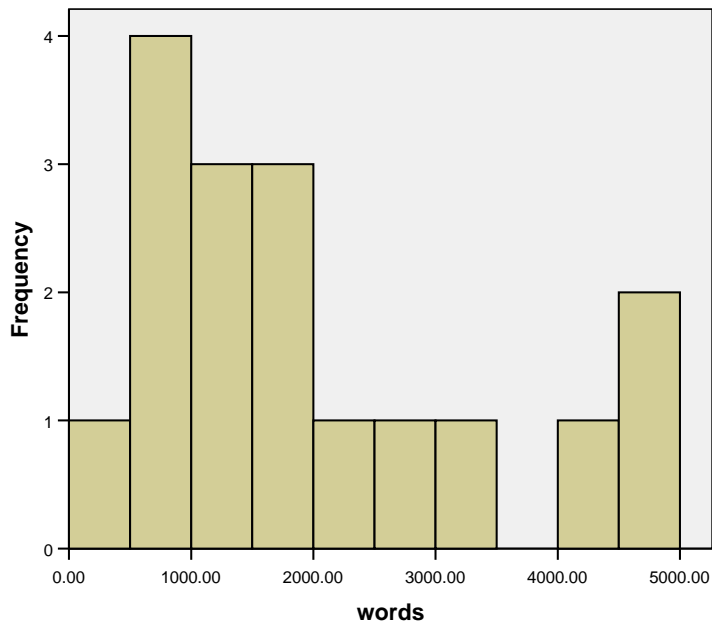


Figure 6.2 Frequencies of Hebrew bible books with word counts less than 5,000 by word count.

Shorter texts like Obadiah necessitate a trade off between Type 1 and Type 2 statistical errors. If we were comparing a single feature in Obadiah with the other thirty-eight books of the Hebrew bible, we would expect 2 significant results to occur by chance with $\alpha=.05$. We might at .01 reduce the likelihood of Type 1 statistical error. If we were comparing each of the thirty-nine books with each of the others in turn, there would be a total of 741 comparisons and $\alpha=.001$ would be more appropriate. But such a low α would reduce the power to detect a given effect size considerably, and this is particularly a problem with smaller sample sizes. Imagine that not only Obadiah, but Obadiah chapter 2 to chapter 10 also exists and was

similar to the known text of Obadiah, we might easily dismiss a true difference between Obadiah that was attested by Obadiah 2 to 10 if we set a too low α . That is, a feature might occur too infrequently in Obadiah that differences between this frequency and those of other biblical books might occur by chance, but if that feature occurred at a similar frequency in our 'newly-discovered' Obadiah 2-10, the total count would be higher and thus less likely to occur by chance. In small sample sizes small effects that influence infrequent events will not show up: different statistical methods are required for what are called 'stochastic' processes. An alternative to risking rejecting a true hypothesis because significance levels have been set too stringently is to set $\alpha=.05$ and accept that some significant results will occur by chance. Since statistical analysis is only a preliminary step, the significant results could be examined and those that were inconsistent with other data could be rejected.

The final step in performing power calculations is often the most difficult: determining what represents an important effect size. When approaching a set of data it is helpful to have some information from other relevant data for which effect sizes are available. For the Qumran corpus the best candidate, for reasons discussed above, is the Hebrew bible.

Effect sizes in the Hebrew bible

As discussed in Chapters 3 and 4 variables whose effect size is of interest can be considered according to seven categories: lexical, morphological length, syntactic length, morphological category, syntactic category, and non-grammatical variables.¹ For the purposes of power calculations there is no need to repeat the discussion about how these categories are inter-dependent. The categories will be considered in turn.

Lexical

There are two main quantitative approaches testing lexical criteria: frequency of individual lexemes and the number of different lexemes within a given text. In the Hebrew bible 200 words account for 55%, 500 words 65%, and 1000 words 75% of all words.² The 200th most frequent word occurs 164 times, out of a total of 430,706 words. The commonest word, excluding commonly prefixed *w*, *b* and *l*, is the direct object marker 't, which occurs 6,718 times equivalent to 1.56% of all words.

Hence, in our typical sample size of 1,000 words, the commonest word would be expected to occur 16 times, and the 55th most frequent word in the Hebrew bible exactly once. Taking the best case with an expected frequency of 16 and testing at $\alpha=.05$ against texts with 1,000, 5,000, 10,000 and 20,000 words with a frequency difference of 4 per 1,000 words, 8/1000 etc. up to 16/1000 (i.e. the comparison text will have either a frequency of zero or 32) shows that even a doubling of expected

¹ K. Neumann (1990), pp 23-114.

² <http://foundationstone.com.au/FoundationStone.html>

frequency, approximately the maximum rate, which is found in Exodus, will only be detected two-thirds of the time against texts of 1,000 and that there is relatively little gain in power for texts beyond 5,000 words (Table 6.2). Lexeme frequency comparisons are therefore unlikely to be useful in around one third of Hebrew bible books due to Type 2 statistical error. That is it is unsafe to assume that the failure to detect a significant difference means that no true difference exists, at least in terms of a doubling of frequency.

Frequency difference per 1,000 words	N=1,000	N=5,000	N=10,000	N=20,000
4	10.3%	14.0%	14.9%	15.5%
8	25.0%	38.1%	40.1%	42.6%
16	66.0%	86.5%	89.2%	90.5%

Table 6.2 Power (%) to detect true frequency differences in texts of different word counts compared to a 1,000 text with a frequency of 16 words at $\alpha=.05$.

The statistical properties of vocabulary richness have already been discussed in Chapters 3 and 4. Detailed consideration for individual books of the Hebrew bible is set out in Appendix 2

These results can now be used for power calculations for tests of vocabulary richness after total word count has been adjusted for (i.e. the tests are not just tests of book length which determines a large proportion of the variance of vocabulary richness).

Setting $\alpha=.05$ and allowing for 10% error variance, a sample size of 1,000 still provides 80.3% power to detect a 0.1% and 97.9% power to detect a 0.2% contribution to variance by a dependent variable of interest. At $\alpha=.01$, there is still 92.2% power to detect a 0.2% contribution to variance.

In summary, effect sizes based on Hebrew bible data suggest that vocabulary richness is likely to be an adequately powered lexical criterion, but that frequency counts of individual words is not, perhaps with the exception of 't and possibly 'šr, but these possible exceptions may be more usefully considered under the syntactic category.

Morphological length

The methodological limitations of counting the number of letters or syllables per word in Hebrew is discussed in Chapter 3. However, it is possible that letter counts may have more to offer in Qumran texts where *plene* forms are more common. Not all Hebrew words are likely to display *plene* forms, the common words 't and 'šr are good examples. For most words the difference in length will be the result of an extra *w* representing a vowel or occasionally a final *n*, though a few words may have two extra letters added in this way. On the whole, then, the expected distribution will be binomial because the variation depends on whether a letter is added or not to each word, rather like tossing a coin. But in addition there will be variation due to vocabulary choice and morphology (e.g. a higher proportion of third person masculine plural verb forms or feminine verb forms compared to third person

masculine singular verb forms). These sources of variation may cancel each other out; for example, a text may have fuller orthography, but choose a vocabulary with shorter words and have a high proportion of third person masculine singular verb forms. Variation due to verb forms can be identified by morphological tagging and *plene* spellings by counting the number of *w* preceded and followed by a consonant. It is likely that other variance in morphological length will correlate with specific morphological variables. For example, frequent use of the independent prepositions *mn* (709 total occurrences in Hebrew bible), *ʿl* (3568 total occurrences in Hebrew bible) and *ʿd* (1095 total occurrences in Hebrew bible) will reduce mean morphological length.

In summary, morphological length analyses can largely be subsumed by morphological analyses. Analysis of *plene* forms may have something to offer, but there are many sources of potential variation and there are no helpful data that help to allow an estimation of likely effect size relevant to Qumran texts. Hence, such analyses are probably best left as secondary to see if the proportion of *plene* forms differs between different groups of texts that have been classified on more certain criteria.

Syntactic length

Qumran text materials are not currently in an advanced enough electronic form to allow this kind of analysis as outlined in Chapter 3.

Morphological

Some idea of likely effect sizes can be gained by considering Radday's morphological criteria: co-ordinative conjunctions, object marker, definite article, proportion of bound pronouns to the sum of free and bound pronouns, percentages of nouns in absolute and construct states, percentages of finite verbs and non-finite/active/passive verbs, and the sum of autonomous prepositions, demonstrative pronouns and personal pronouns.³ Detailed 'proof of concept' analyses are provided in Appendix 2 but, in summary, most of Radday's suggested morphological criteria are likely to be adequately powered for use with texts of around 1,000 words and some can be used with confidence with much shorter fragments.

Syntactic

As noted in Chapter 3, the lack of punctuation in Qumran texts necessitates word category transition approaches that require statistical methods that do not assume independence of data. Standard power calculations are thus not applicable.

Non-grammatical

As discussed in Chapter 3, measures are poorly developed for electronic texts in Hebrew and may be of limited utility.

³ Y. T. Radday (1977).

Implications for combined approaches

The cross-over between Neumann's categories has led to lexicogrammatical approaches which link observed probabilities of textual elements with the semantic aspects of texts. Oath and vow formulae are examples where associations between textual elements relate to semantics of a formulaic clause rather than the individual lexemes. This approach to semantics is adopted by the Semantics of Ancient Hebrew database⁴ where the use of lexemes in recognised formulae and the frequency of other associated lexemes are commented on. From a statistical perspective, however, because the frequencies of most lexemes are too low for conventional analyses, tests of associations with other lexemes will inevitably be underpowered except when comparing large texts.

Beyond associations with individual lexemes, combined approaches may draw on Halliday's concept of register, which is thought to be quantifiable.⁵ The initial step in determining register dimensions is either to perform a factor analysis or principal components analysis⁶ on the textual corpus or to adopt dimensions from other studies. The difference between a factor analysis and principal components analysis is that a factor analysis assumes some underlying structural relationship of the variables where a latent trait, or factor, explains shared variance. Principal

⁴ http://www2.div.ed.ac.uk/research/sahd/lexeme_index.html.

⁵ M. B. O'Donnell (2005), p33.

⁶ These are related statistical methods for looking at what several variables have in common. These are traits or qualities, for example drinks may have alcohol content, sweetness, colour, etc as shared traits.

components analysis does not make such assumptions and looks at both shared variance and that of individual variables. At this point, therefore, an *a priori* structure could be proposed, but for exploratory analyses it is probably safer not to make this assumption. Hence a principal components analysis is the method of choice to explore possible register dimensions. Clearly the textual variables that are entered into the factor analysis are fundamental to the dimensions that are derived. These textual variables must be quantifiable, but they must also have validity if the register dimensions are to be meaningful. Principal components analysis provides a measure for the construct validity of the components extracted and external validity can then be determined by comparison with dimensions derived from other text corpora. Content validity is a more problematic aspect to determine. It is important, however, that the components include aspects of register that scholars are generally agreed are important semantic dimensions.

A further issue is the selection of samples included in the analysis. Biblical books may contain a variety of register dimensions, a mixture of narrative and non-narrative elements, for example. Some books may be thought to have clear sections with different authorship. Moreover, books are of widely differing lengths. If particular textual features (variables) are closely associated with longer texts, should this be weighted by splitting these texts into several smaller units for separate entry into the factor analysis, or should each book be given equal weight by ignoring its length? The advantage of splitting books for the Hebrew bible is that it would increase the number of cases. Ideally there should be at least 100 cases for a principal components analysis or 5 times the number of variables, whichever is larger (here it

amounts to essentially the same sample size). The number of Hebrew bible books falls far short of this. It may thus turn out that the components will not be clearly identified and this method may prove unworkable using each book as a single case. For Qumran Hebrew 'biblical' texts this dilemma is eased by the presence of multiple copies, often of the longer books, that can each be counted as a unit for principal components analysis. For Qumran Aramaic texts it is probably appropriate to use frequencies (or percentages) to explore which textual variables to include in a principal components analysis of Qumran texts as these texts tend to be shorter and including variables which occur infrequently will mean that many individual texts will have zero values for these.

There are a few other assumptions that are necessary for principal components analysis that need to be considered. First, the variables must be interval-level measurements. Counts and percentages fulfil this criterion. Secondly, each case must provide one score for each variable and the cases represent a random sample of all possible cases. This criterion is fulfilled: the cases are totally representative of the population. Thirdly, the relationship between variables should be linear. There is no reason to assume an alternative (e.g. quadratic, exponential) relationship here. Fourthly, variables should be normally distributed. Around half of the variables clearly deviate from a normal distribution, but in practice principal components analysis, which is built on Pearson correlations, is robust to most deviations. Finally, each pair of variables should have a bivariate normal distribution. Again principal components analysis is robust to violations of this assumption. It is thus reasonable to proceed with an exploratory analysis.

As noted in Chapter 3, O'Donnell suggests various textual criteria that can form the basis of a quantitative approach. These include noun to verb ratios, proportion of past tense verbs, and the relationship between proper nouns and pronouns. The marking of common nouns as determinate by the use of the definite article is a criterion that may discriminate between qualitative, abstract and non-abstract register.⁷ A further suggested dimension, overt expression of persuasion, may be associated with specific verbal aspects such as the imperative, jussive and cohortative. The dimension of explicit (situation-independent) versus situation-dependent reference is more difficult to determine in Hebrew because of the wide semantic field of the conjunction *w*. However, the particle *'šr* occurs 4,837 times in the Hebrew bible and may be a suitable textual criterion to quantify situational reference as discussed in the context of Holmstedt's work.⁸

These textual criteria are only a start if a reasonable number of dimensions of register are desired. Several variables are required per component for data reduction in this way to be parsimonious. Moreover, dimensions that may not be common in contemporary literature may be important in Hebrew texts. An example of such a dimension would be that of vow or oath formulae. However, as noted in Chapter 3, full oath and vow formulae are relatively sparse and not always easy to define. Oaths and vows can be subsumed under the larger register of conditional-unconditional

⁷ B.T. Arnold & J.H. Choi (2003), p28ff.

⁸ R. D. Holmstedt (2002).

including ideas of promise, contract and covenant. The preposition *'m* is likely to be a useful measure of conditionality. To increase the number of variables included in the principal components analysis, the preposition *lw*, *lw'* or *lwl'* could be included as used in unreal conditions.⁹ However, these only occur a few times, less than once per book on average, so are unlikely to contribute to the dimension structure.

Similarly *my* can be considered in its idiomatic use, but there are even fewer occurrences than *lw*.¹⁰ de Caen's work indicated that in addition to *'m*, the complementizer *ky* is worthy of consideration.¹¹ Such dimensions may be helpful in moving away from register becoming too closely aligned with genre.

Hence, lexical or morphological criteria of single forms provide around a dozen variables for principal components analysis. Collocational textual criteria are thus likely to be required for enough variables to provide reasonable dimensionality. As has been demonstrated, collocational criteria will need to be morphologically rather than lexically based to have adequate power. One criterion arises from Cartledge's study on vows in the Hebrew bible: an infinitive absolute followed by a finite verb, typical of emphatic statements.¹² Groom considers collocation as an aspect of textual cohesion.¹³ Groom suggests other textual criteria related to cohesion that could

⁹ J. H. Dobson (1999), p192.

¹⁰ *ibid.*, pp185-186.

¹¹ V. de Caën (2005).

¹² T. W. Cartledge (1992).

¹³ S. Groom (2003), p139.

provide variables for an exploratory principal components analysis of the Hebrew bible.

The first of these criteria is reference which Groom illustrates by the relationship between personal pronouns and proper nouns to which they refer.¹⁴ Groom's example connects a masculine plural pronoun with a masculine singular proper noun *yS r'l* which forms part of the formulaic 'sons of Israel'. Hence, morphological association may be indirect in terms of gender and number and it may not always be easy to decide to which proper noun a specific pronoun refers: collocation within verses is a pragmatic, approximate measure. Another criterion of cohesion that Groom proposes is that of junction. This essentially comprises the variety of conjunctions that link phrases or clauses together some of which may be disjunctive, adversative, subordinate etc. Special cases of these are the particles *hnh* and *whnh* that can be considered separately. Other cohesive devices, particularly common in poetry, such as rhythm, alliteration, parallelism, chiasmus etc. are more difficult to extract from electronically tagged texts; omission of such variables from principal components analyses is a limitation, especially if cohesion rather than register is the outcome of interest.

We now have around seventeen variables for an exploratory principal components analysis of the Hebrew bible that will then inform us about which variables to

¹⁴ *ibid.*, pp138-139.

consider when performing principal components analyses on Qumran Hebrew texts. Dobson's review of sentence and clause sequences suggest a few more,¹⁵ though he notes that the context is crucial so that these criteria may be less closely aligned to register. His first, not very common sequence, is perfect verb followed by perfect plus *w* that Dobson considers occurs in passages where events are closely linked in background or reflective passages. His second is perfect followed by preterite, typical of narrative. Dobson's other categories relate less to register than to how past, present and future actions are expressed by the Hebrew tense system. Incorporating these sentence or clause sequence variables into the principal components analysis (for convenience specifying the triggering verb to occur in the first three words of the sentence and the sequential verb anywhere in the same verse) provides around twenty variables, an adequate number to derive a reasonable number of register dimensions.

An exploratory principal components analysis of the Hebrew bible

Having identified a reasonable spread of variables from previous studies and linguistic analyses, an exploratory principal components analysis can be performed that will hopefully suggest register dimensions more appropriate for Qumran texts. The validity of these dimensions can be assessed in the context of the extensive literature on the Hebrew bible. For the purposes of this exploratory analysis, set out in Appendix 2, the book of Daniel was excluded because it has a significant proportion written in Aramaic. In summary, the Hebrew bible register determined by

¹⁵ J. H. Dobson (1999), pp221-230.

principal components analysis of the 22 textual variables counted on a book-by-book basis comprises five main dimensions:

- 1) Historical - determinate past action with subjects and objects;
- 2) Rhetorical – persuasive and evidential;
- 3) Personal exclamation with a conditional tendency;
- 4) Reflective – background text of a non-personal nature;
- 5) Passive – non-narrative text that refers to the effects of external agents

Cluster analysis

Cluster analysis is essentially descriptive; it does not formally test hypotheses. It attempts to group together variables or cases that are similar to each other. Principal components analysis, considered above, is able to group variables together, so cluster analysis is more commonly used to group cases together. There are two main types of cluster analysis: hierarchical and non-hierarchical. Both depend on calculating the ‘distance’ between cases. This calculation is based on the values of variables associated with each case. A simple example of such a calculation would be to take the square of the difference between the values of a variable for two cases, as case A and case B, then the square of the difference for the next variable of interest, and so on until all variables have been considered, take the mean of these squared differences and compare it to the mean squared differences for comparison between case A and case C, case A and case D, and so on for all the cases. Then repeat this for each case. This provides a set, or matrix, of differences between cases for all the variables of interest. It is clear that these computations cannot be done manually for

anything but the simplest of situations. Once this matrix of differences has been calculated, the clustering algorithm can be applied. Non-hierarchical clustering is most useful when the number of clusters is already known and the purpose is to classify cases into these known clusters. In a sense it is just an extension of discriminant analysis. The clustering algorithm seeks to place cases in the preset number of clusters according to how similar to each other the cases are. In fact, for this method, distances between all the cases do not need to be calculated, instead a few cases are chosen to start with, they are placed into initial clusters, then the rest of the cases are tested to see which cluster they are closest to in turn. This requires less computation.

Hierarchical cluster analysis does require all the distances between cases to be calculated, and so is less efficient computationally. It is appropriate when the number of clusters is unknown *a priori*. The algorithm starts with each case in a separate cluster and combines the closest clusters step-by-step, taking the least distance between clusters as the criterion for combination, until all the cases are combined into one cluster. The process can be displayed as a binary tree or dendrogram with cases listed vertically next to their nearest neighbours and the distances between clusters indicated by the length of the horizontal line joining clusters together: the longer across the page, the further apart the clusters. Hierarchical clustering does not produce a definitive list of clusters, rather this has to be inferred by inspection of the dendrogram. Hierarchical cluster analysis is thus an exploratory technique.

From the description of hierarchical clustering just given, some important considerations arise. First, if the scale of different variables differs greatly, this can have a major effect on clustering. For example, if we wished to cluster together towns and cities, we might have latitude measured in degrees, so never more than ninety, but population size might be as high as several millions. Given that we are calculating distances on the mean of the sum of squares, this would bias the clustering to be more dependent on population size rather than latitude. Sometimes such weighting, as it is known, may be appropriate, but when it is not, it is best to equalise the scales of all variables by standardizing them by expressing their values in terms of the number of standard deviations they lie from the mean and setting the means of all variables to zero. Secondly, the type of distance calculation is important; it may be that not all variables are on an interval scale, some might be binary so calculating mean squared differences would be inappropriate. In fact, although there are several methods to calculate differences, it is also possible to consider shared variance between variables as a measure of similarity (or difference) instead for variables measured on an interval scale: this is essentially the same as using the correlation coefficient between the different variable values. Thirdly, the order in which cases are considered may be important if cases are very similar to each other; in this situation next door neighbours may just reflect the order in which cases were considered. In this situation, a top-down (divisive) rather than bottom-up (agglomerative) approach may be better.¹⁶ In a top-down approach all cases start in a

¹⁶ R. H. Baayen (2008) reviews divisive algorithms in his book on analysing linguistic data using the R statistical package. He also reviews a related divisive procedure, Classification and Regression Trees also known as recursive partitioning though notes that this can be difficult to generalise (p150).

single cluster and this cluster is split according to the greatest distance between cases or clusters. Top-down algorithms are similar in many ways to the non-hierarchical approaches outlined above in that they partition cases, only applied in an iterative fashion to produce a hierarchical structure.¹⁷

Although there are no formal measures of how good a description of the data a hierarchical cluster analysis provides, those that are less helpful will often show agglomeration on a case-by-case basis with an even progression in distances, whereas more informative analyses will show distinct clusters linked by small distances then larger distances between each of these clustered groups. Moreover, the robustness can be checked by comparing standardised and non-standardised analyses or by using different distance measures. This may help determine whether some sections of the hierarchical structure are more likely to represent true clusters than others.

Multidimensional scaling

Multidimensional scaling is another way to display differences or similarities between variables or cases. For variables, like cluster analysis, it resembles principal components analysis in that the communalities could be used to produce plots of how variables relate to each other. Alternatively, multidimensional scaling can produce a

¹⁷ A. McMahon & R. McMahon describe a related structure called a Fitch tree in their analysis of sixty German cognates. They also describe a Network structure using the same data, which is similar to a two-dimensional multi-dimensional scaling analysis as described in the next section.

plot of how cases relate to each other in a multidimensional space. Just like cluster analysis, this plot relates directly to the distances between cases as measured from a given set of variables of interest. Practically, multidimensional scaling uses two- or three-dimensional space otherwise interpretation is very difficult; this restriction is therefore somewhat arbitrary. Unlike cluster analysis, there are measures of how well the multidimensional scaling model fits the data, which can help decide whether the model is useful. Again, like cluster analysis, there are no formal statistical tests performed. Moreover, the procedure is not designed to produce clusters or groups that can be classified, although the distribution of cases within the two- or three-dimensional space may suggest some clustering.

A useful measure of how well the multidimensional scaling model fits the data is stress, of which there are various indices. In simple terms, stress functions estimate the fit of the model to the data. For example, suppose we have a data set of values for distances between towns. If we have all these distances accurately, a two-dimensional model will look like the geographical distribution of the towns on a map. The stress of the model will approach zero. Now, suppose we only have the distances to the nearest ten miles. Sometimes 42 miles will be measured as 40 miles and 46 miles as 50 miles etc. This will result in the model having difficulty replicating the geographical distribution, though it should still approximate to it fairly well; stress will be above zero. This illustrates the importance of error and chance sampling. In addition, missing values will also increase stress. It is important to note that changing the reference frame, such as moving from two- to three-

dimensional models, influences stress values making comparison of model fit problematic.

Latent class analysis

Latent class analysis is a special case of finite mixture modelling.¹⁸ Latent class analysis is useful where heterogeneity in observed variables depends, in part, on the presence of latent classes. Just as in principal components analysis where the components do not directly correspond to any single variable but describe some trait common to different variables, latent classes do not necessarily correspond to a variable within the data set, though a variable denoting a latent class is allowed in certain circumstances as discussed below. Perhaps a useful example would be the situation with anthropometric variables. Suppose we have a population of men and women and we recorded height, demi-span, foot length and index finger length. We would expect that, on average, men would score higher on these variables than women. Now suppose we have lost the key that tells us which of the cases are men and which are women. We might analyse the data blind to sex. But we could use latent class analysis to see if any latent variables, in this case, sex accounted for heterogeneity in the observed variables. If our data were fairly representative, we would expect a two class solution with one class having higher scores across all variables than the other. Note that the latent class variable may be categorical. Technically, this kind of model is described as a measurement model, in this case a multivariate¹⁹ regression model describing the relationships between the observed

¹⁸ For a comprehensive description, see L. K. Muthén & B. O. Muthén (2010), Chapter 7.

¹⁹ Multivariate means the dependent term in the regression equation comprises more than one variable.

variables and the categorical latent variables (classes). A measurement model is essentially hypothesis-free and thus is exploratory in nature. Importantly, the analysis estimates the probability for each case of belonging to a particular class. We can imagine that for anthropometric data, short people will have a high probability of belonging to the class that will be associated with being female and tall people a high probability of belonging to the class that will be associated with being male. For those of medium height, probabilities may be somewhere in between, taking account, of course, of the other anthropometric variables scores.

Along with a measurement model, mixture modelling also allows structural models to be built. These are more confirmatory where certain parts of the model are pre-specified such as including a covariate in the model which may have effects on both the latent variables and the directly observed variables. Other relationships, such as those between latent variables can also be pre-specified. Different structural models can be compared to see which one fits the data best.

In addition to pre-specifying relationships between variables, it is also possible to use training data to facilitate classification. For example, in the anthropometric data set example, suppose we knew the sex for half the cases, we could use this information to fix their latent class membership and thus improve the prediction of class membership of the other half of the cases where sex is unknown. Again, goodness of fit statistics are available to inform us about how well the model fits the observed data.

There are various goodness-of-fit criteria to judge how well a model fits the observed data. The ones that will be of main interest in this thesis are termed information criteria. The basic information criterion is called the Akaike information criterion (AIC). It describes how much information is lost compared to the complete data. An equivalent concept is to consider the principal components extracted for the Hebrew bible textual variables. The first five principal components explained 67.9% of the total shared variance of all 22 variables, so 32.1% of information is lost in any model using just these five principal components. If we were modelling these data using these principal components, they would constitute five parameters in the model. The AIC takes account of the number of parameters in any model it is calculated for: more complex models with greater numbers of parameters will tend to have higher AICs. Since AIC is a measure of information lost, the lower it is the better, so models with fewer parameters should be preferred. There is a Bayesian version of the AIC termed the Bayesian information criterion (BIC). Which has a bigger penalty for the number of parameters than the AIC; it tends to be preferred when the number of parameters for a model is large. In addition, there is a sample-size adjusted BIC which penalizes small sample sizes relative to large ones. Therefore, depending on the complexity of the model and the sample size some information criteria may be preferred to others when judging between two models as to which fits the data best.

Latent class analysis can be both an exploratory and confirmatory technique. Its advantage over cluster analysis is that it provides goodness-of-fit measures. It is similar to non-hierarchical cluster analysis in that it partitions cases into a discrete

number of classes that have to be pre-specified rather than a truly exploratory technique such as hierarchical cluster analysis which does not make such assumptions. In practice, it is helpful to use more than one technique for classification: usually starting with an exploratory analysis then using a confirmatory analysis to establish the optimal model. Therefore, as proof-of-concept, a hierarchical cluster analysis of the Hebrew bible books using the five principal components to keep models simple is outlined in Appendix 2.

Vocabulary richness and translated texts

A major task after the internal classification of Qumran texts is to test how closely particular clusters relate to external standard texts. The previous exploratory analyses of the Hebrew bible books based on the MT are likely to be helpful here. However, comparison with translated texts, specifically those in Greek, is more problematic, as discussed in Chapter 3. Textual variables may not be portable between Hebrew and Greek, however closely a translator tried to keep to the Hebrew *vorlage*. We have seen, for example, that Ecclesiastes has a dominant rhetorical dimension to its register. Rhetorical devices in Greek may derive from the Hebrew or relate to common usage in Greek.²⁰ Fortunately, the preceding analyses demonstrate that vocabulary richness is quite distinct from morpho-syntactic features typical of Hebrew. It is also independent of register dimension. A corollary is to ask whether vocabulary richness might be a useful measure to determine closeness of Qumran texts to those in the Septuagintal group.

Figure 6.3 shows the relationship between the total number of words and the number of different lemmas in each Septuagint book equivalent to the MT. Unlike the MT, the relationship is essentially linear. There are around five books with fewer different lemmas than would be expected and one book with more lemmas than expected.

Figure 6.4 shows the relationship between vocabulary richness of the MT against vocabulary richness of the LXX for individual books. The same linear

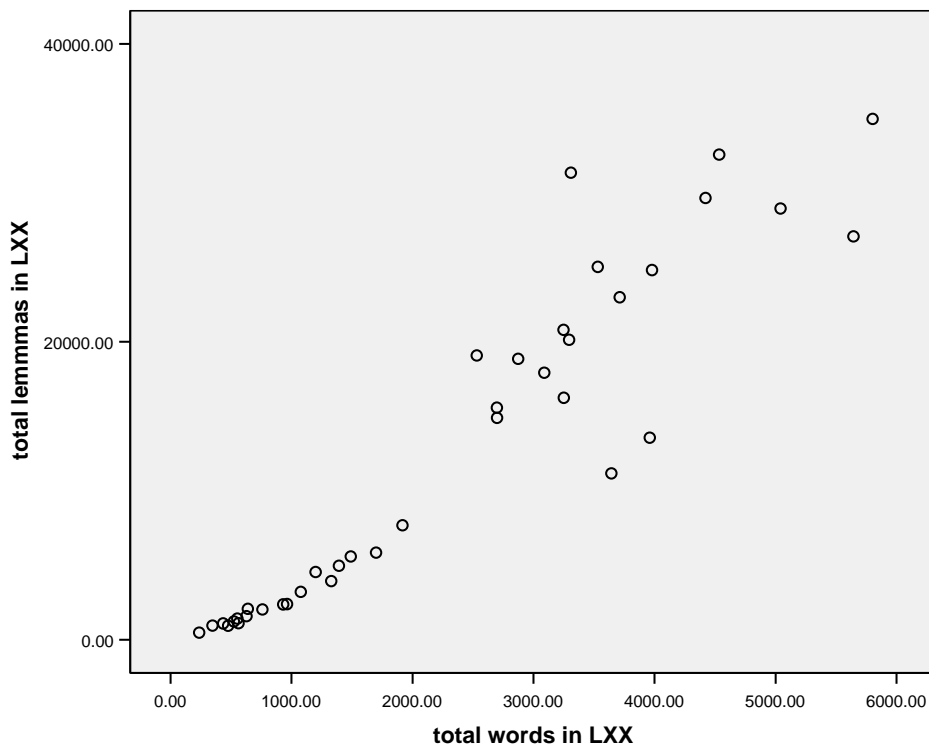


Figure 6.3 Total number of words versus total number of different lemmas in Septuagintal (LXX) biblical books.

²⁰ J. K. Aitken (2005).

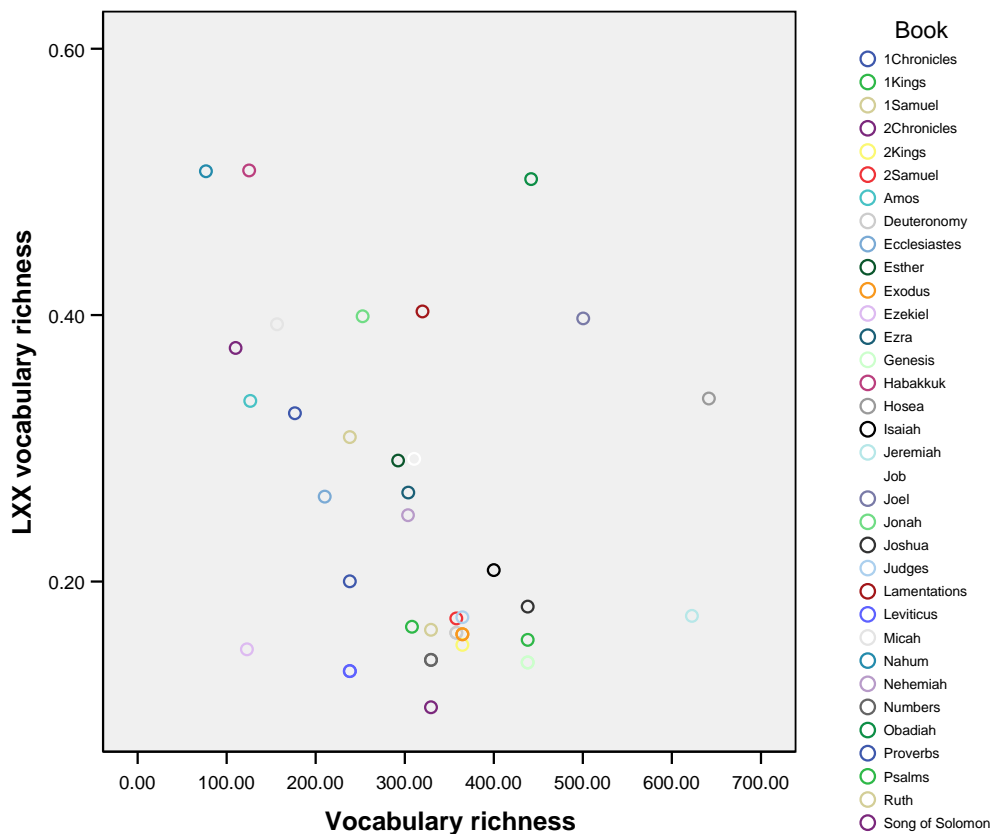


Figure 6.4 Vocabulary richness of Hebrew Masoretic Text (MT) versus Septuagint (LXX) vocabulary richness for individual biblical books.

relationship is seen, but the scatter is greater and it is in the opposite direction. The correlation coefficient is $-.31$ ($p=.051$) indicating that around 16% of vocabulary richness of the LXX is explained by MT vocabulary richness, but the richer the MT vocabulary, the less rich the LXX vocabulary for any given book. Ezekiel stands out as the LXX book with a particularly lower than expected vocabulary richness, whilst Obadiah, Joel and Hosea have a higher LXX vocabulary richness than expected. The reasons for this not only comprise how specific lemmas are translated (i.e. use of more than one Greek lexeme for a Hebrew one or vice versa), but also the use of Greek definite articles, prepositions, etc. that are separate words in Greek but bound as prefixes in Hebrew.

Multiple linear regression could be used to tease out the different influences on LXX vocabulary richness. The derived equation for a LXX book is:

$$\begin{aligned} \text{Number of different Greek words} = & (0.127 \times \text{Total Number of words of LXX book}) \\ & + (0.002 \times \text{Total Number of words of MT book}) \\ & + (0.075 \times \text{Number of different Hebrew words}) \\ & + 492 \end{aligned}$$

The regression terms explain 85.7% of variance in the number of different Greek words. However, the 95% confidence intervals are quite wide. The upper bound is:

$$\begin{aligned} \text{Number of different Greek words} = & (0.147 \times \text{Total Number of words of LXX book}) \\ & + (0.027 \times \text{Total Number of words of MT book}) \\ & + (0.248 \times \text{Number of different Hebrew words}) \\ & + 991 \end{aligned}$$

And the lower bound is:

$$\begin{aligned} \text{Number of different Greek words} = & (0.107 \times \text{Total Number of words of LXX book}) \\ & - (0.023 \times \text{Total Number of words of MT book}) \\ & - (0.099 \times \text{Number of different Hebrew} \\ & \text{words}) - 7 \end{aligned}$$

Practically, however, there are clear discrepancies between the total number of words for LXX versus MT versions of books and far less between the total lemmas.

Vocabulary richness therefore reflects more the Greek style (use of separate definite articles, prepositions etc) than limitation of lexicon to draw on. Hence, despite its promise, vocabulary richness is unlikely to be helpful as a measure for aligning Qumran text clusters with Septuagintal texts.

Summary

Having set out possible statistical approaches to a textual analysis of the Aramaic texts found at Qumran, checking which variables would be suitable in terms of statistical power, and undertaking ‘proof-of-concept’ analyses using data from the Hebrew bible, it is now possible to start the analysis of the Aramaic texts themselves. The initial phase is to describe the textual data to establish which variables will have adequate statistical power, and next to investigate possible clustering of texts based on selected variables: first morphological and then syntactic. This exploratory stage will include relating these analyses to non-Qumran Aramaic texts to provide external correlates. Finally, any possible clusters will be subjected to confirmatory analysis with latent class profiling.

7. Description of Aramaic Scrolls

The total number of words and total counts for each part of speech for the seventy-four Aramaic scrolls considered in the *a priori* hypotheses testing are set out in Table 7.1. The mean number of words for all 74 scrolls was 496, but the distribution was highly skewed to the right with a median count of 233.5 words (Figure 7.1) and 73% of scrolls having total word counts less than 500 words. The proportion of parts of speech is shown in Figures 1 – 6 in Appendix 2. A typical scroll would thus comprise 27.2% nouns, 13.6% verbs, 8.8% pronominal suffixes, 6.7% adjectives, 41.9% particles and 1.7% pronouns, though the distributions vary considerably between scrolls (see Appendix 2 Figures 1 – 6). The proportions have important implications for the *a priori* hypotheses. In particular, those hypotheses relating to pronominal variants are likely to be underpowered for the majority of texts. Even for a scroll which has a total word count around the mean of 500 words, it will only have eight or nine pronouns. By contrast, those criteria that relate to nouns or verbs have far superior power to detect true differences between scrolls.

Scroll	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
1Q20	1431	687	514	233	2306	77	5248
1Q21	50	24	12	10	70	4	170
1Q23	28	13	4	11	41	1	98
1Q24	15	1	4	0	36	0	56
2Q24	84	22	15	40	127	0	288
4Q112	534	289	151	88	808	33	1092
4Q113	344	213	95	56	562	13	980
4Q115	137	74	42	43	193	9	498
4Q117	125	37	21	17	146	4	236
4Q156	48	14	15	5	75	0	157
4Q196	266	169	138	38	389	24	1024
4Q197	196	138	131	31	332	27	855
4Q198	38	18	11	2	51	1	121
4Q201	331	175	106	78	466	18	1174
4Q202	248	121	72	52	375	15	883
4Q203	65	38	31	6	130	6	276
4Q204	574	345	207	137	1014	28	2305
4Q205	117	82	42	48	264	4	557
4Q206	216	153	57	85	430	9	950
4Q207	20	16	9	10	52	0	107
4Q208	149	74	22	73	252	0	570
4Q209	586	306	101	342	949	3	2287
4Q210	128	55	33	54	251	2	523
4Q211	31	10	4	23	52	0	120
4Q212	202	100	55	30	279	33	699
4Q213	121	64	59	16	171	10	441
4Q213a	65	23	26	7	75	6	202
4Q213b	26	18	16	6	45	6	117
4Q214	38	18	15	4	69	1	145
4Q214a	20	17	13	9	54	1	172
4Q214b	44	22	12	7	84	3	152
4Q242	41	24	11	8	61	7	231
4Q243	83	32	16	13	83	4	171
4Q244	25	12	5	1	37	1	387
4Q246	62	26	16	4	62	1	178
4Q318	56	5	0	106	220	0	684
4Q529	51	18	15	9	82	3	638
4Q530	174	116	65	32	287	10	343

4Q531	175	97	52	32	266	16	71
4Q532	17	21	3	4	36	0	188
4Q534	100	53	44	19	124	3	354
4Q535	15	13	5	8	29	1	136
4Q536	48	28	21	11	78	2	144
4Q537	95	62	35	18	142	2	68
4Q538	24	25	21	7	58	1	458
4Q539	45	23	17	7	47	5	467
4Q540	16	12	6	4	30	0	497
4Q541	122	90	45	21	177	3	499
4Q542	126	51	56	13	215	6	497
4Q543	123	74	64	29	187	20	347
4Q544	126	70	54	25	199	25	396
4Q545	142	74	59	27	186	9	207
4Q546	92	55	49	13	128	10	98
4Q547	98	61	40	25	164	8	52
4Q548	73	21	11	11	87	4	248
4Q549	39	14	16	4	43	1	181
4Q550	35	13	11	2	37	0	184
4Q550a	18	6	6	1	20	1	906
4Q550c	56	45	21	14	101	11	140
4Q552	33	33	23	9	74	9	244
4Q553	38	26	21	10	82	7	99
4Q554	306	51	60	133	356	0	220
4Q556	34	20	5	11	64	6	96
4Q557	9	2	2	1	10	0	88
4Q558	58	38	18	18	107	5	803
4Q559	49	5	0	19	12	0	104
4Q560	29	14	6	2	45	3	2540
4Q561	40	28	33	33	80	6	740
4Q562	24	11	7	6	46	2	498
4Q570	21	10	7	4	43	3	236
5Q15	297	34	59	155	258	0	157
6Q8	23	24	10	5	40	2	1024
11Q10	662	470	316	68	978	46	855
11Q18	230	63	58	84	302	3	121

Table 7.1 Parts of speech count and total words for Qumran Aramaic scrolls

considered in *a priori* hypotheses testing.

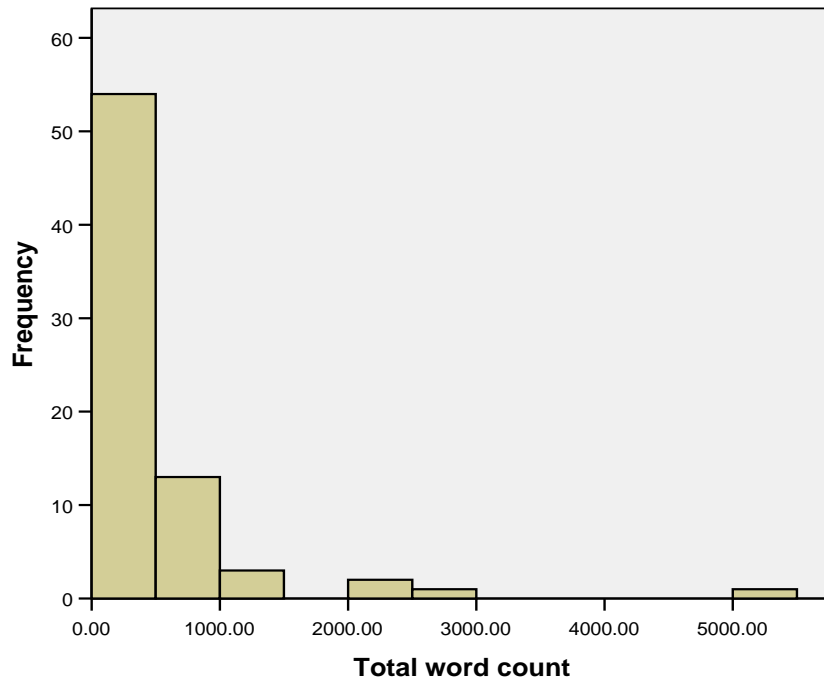


Figure 7.1 Frequencies of total word counts for Qumran Aramaic scrolls considered in *a priori* hypotheses testing.

The frequencies of each of the *a priori* criteria are shown in Table 7.2. Consistent with the parts of speech proportions, some forms were identified in only a small minority of the scrolls. At this stage the empirical data necessitate the exclusion of criteria that lack adequate power to discriminate between text types. This is not to say that these criteria are unimportant linguistically, just that they are inadequate for a quantitative analysis. The cut off point for exclusion is arbitrary, here a frequency of occurrence in at least 10% of the scrolls has been chosen as clusters of text types based on a lower cut off

Criterion	Number of scrolls with zero occurrences	Number of scrolls with 1 - 5 occurrences	Number of scrolls with > 5 occurrences
Determined nouns	1	10	63
Afel	12	43	19
Hafel	52	18	4
Itpeel	27	37	10
Hitpeel	69	4	1
Itpaal	39	27	8
Hitpaal	70	3	1
<i>m</i> -prefixed Peal infinitives	29	30	15
Peal infinitives	28	31	15
<i>l</i> prefixed imperfects	38	31	5
<i>y</i> prefixed imperfects	18	30	26
<i>n</i> prefixed imperfects	73	1	0
3fp verbs	59	13	2
Perfect 2ms ending <i>th</i>	67	4	3
Perfect 2ms ending <i>t'</i>	72	2	0
Perfect 1pl ending <i>n'</i>	66	7	1
Perfect 1pl ending <i>nh</i>	72	2	0
Perfect 2mp ending <i>tn</i>	73	1	0
Perfect 2mp ending <i>tw</i>	67	7	0
Perfect 3mp ending <i>w</i>	24	34	16
Perfect 3mp ending <i>w'</i>	65	7	2
Retained aleph before suffix	65	8	1
2ms pronominal suffix <i>kh</i>	60	11	3
2fs pronominal suffix <i>ky</i>	72	1	1
3fs pronominal suffix <i>ha</i>	64	9	1
mp pronominal suffix <i>hwn</i>	16	31	27
mp pronominal suffix <i>hwm</i>	72	2	0
ms suffix on mp nouns <i>yhy</i>	73	1	0
ms suffix on mp nouns <i>why</i>	35	34	5
Relative particles	4	27	43
Relative particle <i>dy</i>	5	26	43
Direct object marker <i>yt</i>	65	9	0
<i>d'</i>	59	14	1
<i>dh</i>	72	2	0
<i>dnh</i> or <i>dn'</i> forms	56	17	1
3fp independent pronouns	72	2	0
1s independent pronoun <i>'n'</i>	72	2	0
1s independent pronoun <i>'nh</i>	40	25	9

1p independent pronoun 'nhn'	66	8	0
1p independent pronoun 'nhmh	71	3	0
2mp independent pronoun 'ntwn	65	9	0
2mp independent pronoun 'ntn	72	2	0
mp independent pronoun 'hmwn	71	3	0
mp independent pronoun 'hmw	72	2	0
mp independent pronoun 'nwn	46	28	0
mp independent pronoun 'nw	74	0	0
Demonstrative pronoun 'lh	64	10	0
Demonstrative pronoun 'lyn	49	21	4
Demonstrative pronoun 'ln	56	15	3

Table 7.2 Number of Qumran Aramaic scrolls considered in *a priori* hypotheses testing containing zero, 1-5 or more than 5 occurrences of each criterion.

could have seven scrolls as maximum membership. Relatively infrequent forms have a greater probability of occurring in scrolls with more words. To adjust for the potential bias of scroll length, rates per 1,000 words were calculated for the following criteria where at least two alternate forms occurred in at least 10% of the scrolls or the form occurred in at least 10% of the scrolls if no clear alternative (e.g. 3fp verbs):

Determined nouns

Afel

Hafel

Itpeel

Hitpeel

m-prefixed Peal infinitives

Peal infinitives

l prefixed imperfects

y prefixed imperfects

3fp verbs

Perfect 3mp ending *w*

Perfect 3mp ending *w'*

Retained aleph before suffix

2ms pronominal suffix *kh*

3fs pronominal suffix *h'*

Relative particles

Relative particle *dy*

Direct object marker *yt*

dnh or *dn'* forms

Demonstrative pronoun *'lh*

Demonstrative pronoun *'lyn*

Demonstrative pronoun *'ln*

This narrowed the number of *a priori* hypotheses to test between the 74 scrolls to 22.

Table 7.3 shows the mean, median and range of rates for these 22 criteria. The rates vary considerably from criterion to criterion and for most the median differs markedly from the mean. These differences in rates are likely to underpin the relationships between raw

Criterion	Mean	Median	Minimum, Maximum
Determined nouns	77.6	74.8	0, 151.2
Afel	12.6	11.3	0, 41.7
Hafel	2.9	0	0, 26.3
Itpeel	5.1	4.3	0, 28.2
Hitpeel	0.4	0	0, 12.7
<i>m</i> -prefixed Peal infinitives	5.8	4.3	0, 41.7
Peal infinitives	6.0	4.4	0, 41.7
<i>l</i> prefixed imperfects	4.0	0	0, 27.3
<i>y</i> prefixed imperfects	17.4	9.1	0, 102.9
3fp verbs	0.6	0	0, 12.4
Perfect 3mp ending <i>w</i>	9.4	6.2	0, 74.1
Perfect 3mp ending <i>w'</i>	0.5	0	0, 12.4
Retained aleph before suffix	0.4	0	0, 10.0
2ms pronominal suffix <i>kh</i>	1.8	0	0, 29.0
3fs pronominal suffix <i>ha</i>	0.6	0	0, 8.8
Relative particles	29.8	29.3	0, 104.2
Relative particle <i>dy</i>	28.4	28.2	0, 72.9
Direct object marker <i>yt</i>	0.8	0	0, 23.5
<i>dnh</i> or <i>dn'</i> forms	1.6	0	0, 13.8
Demonstrative pronoun <i>'lh</i>	0.3	0	0, 5.3
Demonstrative pronoun <i>'lyn</i>	2.0	0	0, 18.7
Demonstrative pronoun <i>'ln</i>	1.6	0	0, 18.7

Table 7.3 Mean, median and range of rates per 1,000 words for the 22 criteria occurring in at least 10% of the 74 Qumran Aramaic scrolls identified as of adequate length.

analyses indicate that clustering scrolls using these 22 criteria will be disproportionately weighted on the four variables with high mean rates. The resulting dendrogram is shown in Figure 7.4. The further down the dendrogram, the less close to each other are the scroll clusters. With the caveat that there was no *a priori* reason to assume that determined nouns, relative particles, *dy* relative particles and *y* prefixed imperfects were in any way superior criteria for distinguishing text types than any of the other criteria, we can proceed to explore the cluster classification structure.

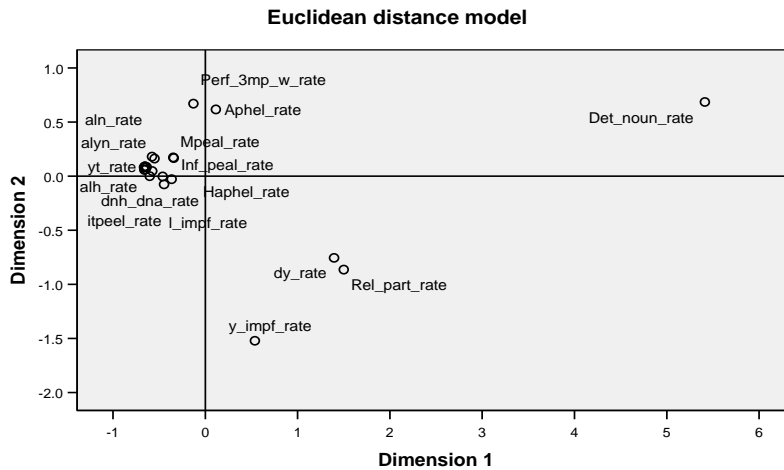


Figure 7.3 Two dimensional scaling using the Euclidean distance between the 22 *a priori* criteria for the 74 Qumran Aramaic scrolls.

Hierarchical cluster classification using raw rates of textual criteria

The dendrogram in Figure 7.4 shows the agglomeration of clusters for the 74 Aramaic scrolls working from left to right based on the absolute values for the 22 textual criteria. The distance between individual scrolls and between clusters of scrolls is represented graphically by the length of horizontal lines connecting them. The order of agglomeration of clusters is indicated by the order the scrolls appear in vertically at each horizontal point. Hence initially small clusters of scrolls are linked by short horizontal lines and then these small clusters are joined together and also with other individual scrolls moving horizontally to the right until all the scrolls are joined. A good classification will have several groups joined together by short horizontal lines and then this small number of groups joined together by long horizontal lines. Such an arrangement would indicate clearly discrete categories.

Low level clusters

Working our way down, following the order of agglomeration, small low level clusters of scrolls can be identified. The first low level cluster comprises 4Q208, 4Q209 and 4Q211, all Enochic texts. Associated with these are 4Q213a and 4Q213b, Aramaic Levi texts, 5Q15 New Jerusalem and 4Q539 the Apocryphon of Joseph B. Related, though not quite so closely, is 4Q214a, another Aramaic Levi text.

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
4Q208	19	↓↘				
4Q209	20	↓↘↓↘				
4Q211	22	↓↘ ↔				
5Q15	69	↓↘↓↘↓↘↓↘↓↘↓↘				
4Q213a	25	↓↘↓↘	↔			
4Q213b	26	↓↘ ↔	↔	↔		
4Q539	44	↓↘↓↘	↔	↔		
4Q214a	28	↓↘↓↘↓↘↓↘↓↘↓↘	↔	↔		
1Q23	3	↓↘↓↘↓↘↓↘↓↘	↔			
4Q203	14	↓↘↓↘↓↘	↔	↔		
4Q535	40	↓↘↓↘↓↘↓↘↓↘	↔			
4Q550a	56	↓↘↓↘↓↘↓↘	↔	↔		
4Q202	13	↓↘	↔	↔	↔	
4Q553	59	↓↘↓↘	↔	↔	↔	↔
4Q205	16	↓↘ ↔	↔	↔	↔	↔
4Q201	12	↓↘↓↘↓↘↓↘	↔	↔	↔	↔
4Q530	36	↓↘ ↔	↔	↔	↔	↔
1Q20	1	↓↘ ↔	↔	↔	↔	↔
4Q531	37	↓↘ ↔	↔	↔	↔	↔
4Q542	47	↓↘↓↘	↔	↔	↔	↔
4Q204	15	↓↘ ↔	↔	↔		
4Q558	63	↓↘ ↔	↔	↔		
4Q552	58	↓↘↓↘	↔	↔		
4Q243	31	↓↘↓↘↓↘↓↘	↔	↔	↔	↔
4Q537	42	↓↘↓↘↓↘↓↘	↔	↔		
4Q244	32	↓↘↓↘↓↘↓↘↓↘	↔	↔		
4Q550	55	↓↘↓↘↓↘↓↘	↔	↔		
2Q24	5	↓↘	↔	↔	↔	
11Q18	72	↓↘↓↘	↔	↔	↔	
4Q210	21	↓↘ ↔	↔	↔	↔	
1Q21	2	↓↘↓↘↓↘↓↘	↔	↔	↔	
4Q554	60	↓↘ ↔	↔	↔	↔	↔
4Q242	30	↓↘↓↘	↔	↔	↔	↔
4Q560	65	↓↘↓↘↓↘	↔	↔	↔	
4Q206	17	↓↘↓↘↓↘↓↘↓↘↓↘↓↘	↔	↔	↔	
4Q207	18	↓↘↓↘↓↘	↔	↔		
4Q549	54	↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘	↔	↔	↔	
4Q559	64	↓↘↓↘↓↘	↔	↔	↔	
4Q541	46	↓↘↓↘↓↘↓↘↓↘↓↘↓↘	↔	↔	↔	
11Q10	71	↓↘↓↘↓↘↓↘↓↘	↔	↔	↔	
4Q538	43	↓↘↓↘	↔	↔	↔	
6Q8	70	↓↘ ↓↘↓↘	↔	↔	↔	
4Q550c	57	↓↘↓↘	↔	↔	↔	↔
4Q196	9	↓↘	↔	↔	↔	↔
4Q197	10	↓↘	↔	↔	↔	↔
4Q213	24	↓↘↓↘	↔	↔	↔	↔
4Q544	49	↓↘ ↔	↔	↔	↔	↔
4Q547	52	↓↘ ↓↘	↔	↔	↔	↔
4Q543	48	↓↘ ↔	↔	↔	↔	↔

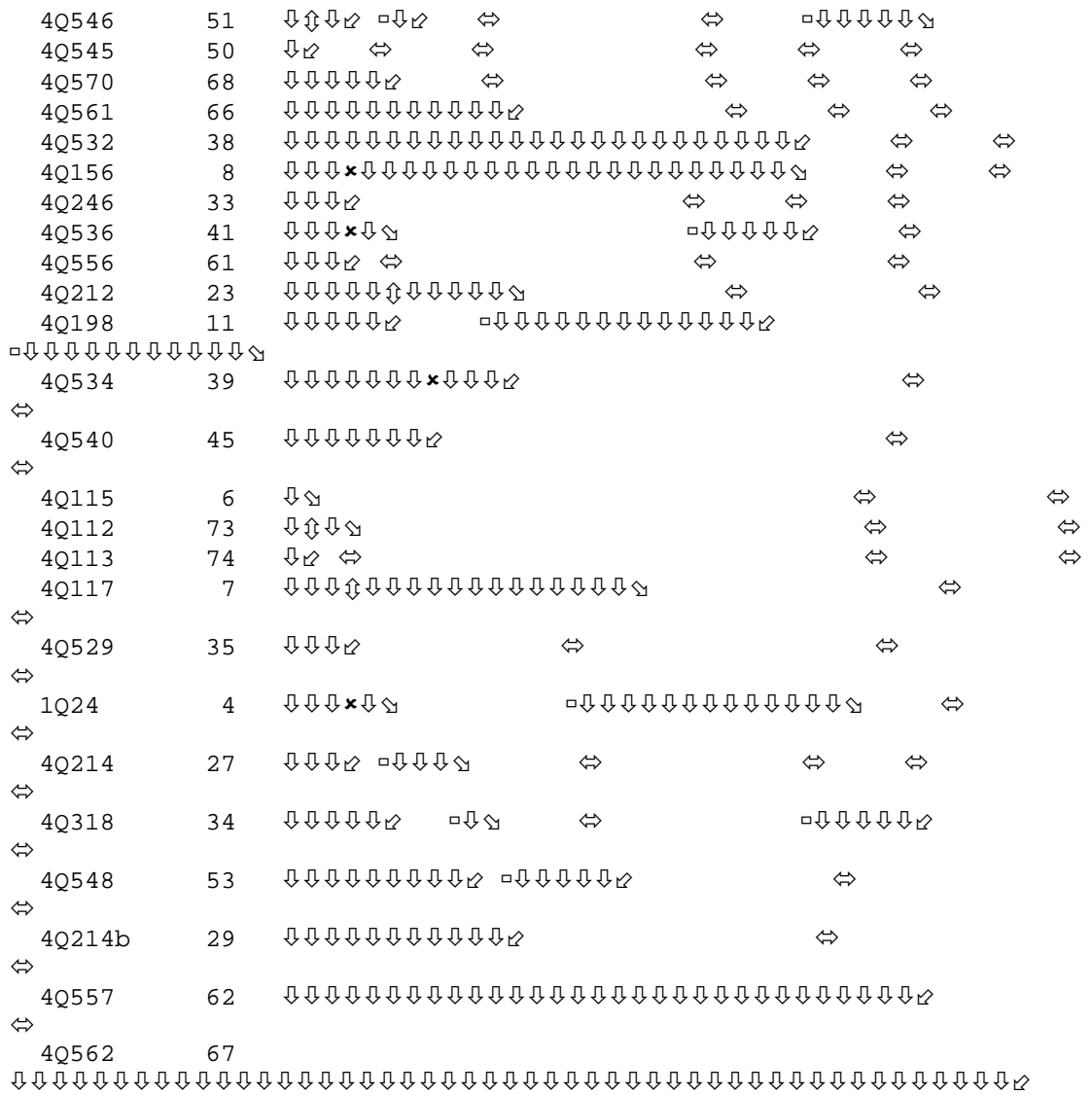


Figure 7.4 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls using raw rates of the 22 *a priori* criteria.

The next low level cluster comprises 4Q202, 4Q205, two Enochic texts, and 4Q553, a Four Kingdoms text. 4Q552, another Four Kingdoms text is associated with these as is the low level cluster of 4Q201 and 4Q530, Enochic texts, and another low level cluster of 1Q20, 4Q204, 4Q531, 4Q542 and 4Q558. This latter group is a mixture of a major text, the Genesis Apocryphon, and two Enochic texts, the Testament of Qahat and 4Q558 designated a Vision.

Three scrolls from different caves make up the next low level cluster: two New Jerusalem texts, 2Q24 and 11Q18, together with an Enochic text 4Q210. This group is associated with a scroll from yet another cave, 1Q21, Testament of Levi and 4Q554, another New Jerusalem text.

Tobit scrolls form part of the next low level cluster, 4Q196 and 4Q197, together with two Visions of Amram texts, 4Q544 and 4Q547, and an Aramaic Levi text, 4Q213. This cluster is closely related to 4Q543, 4Q545 and 4Q546, all Visions of Amram texts.

Further down is another low level cluster, 4Q112, 4Q113 and 4Q115, all biblical Daniel texts. These are closely related to 4Q117, Ezra, and 4Q529, Words of Michael.

The rest of the texts do not cluster together so closely as these five groups that could be characterized as:

- 1) Aramaic Levi – Enochic
- 2) Four Kingdoms – Enochic
- 3) New Jerusalem
- 4) Visions of Amram – Tobit
- 5) Biblical Daniel - Ezra

Intermediate level clusters

The next level of clustering occurs with slightly more distance between texts, but it may have the advantage of pulling together texts that are generally similar, but might differ on a few criteria only. Some will form around the low level clusters identified above.

The first intermediate level cluster comprises 1Q23, 4Q203, 4Q535 and 4Q550a, an Enochic group with Proto-Esther. These scrolls relate to the Four Kingdoms – Enochic low level cluster.

The next intermediate cluster comprises a disparate group, 4Q538, 4Q550c and 6Q8, which relate to the Visions of Amram-Tobit low level cluster.

A further intermediate level cluster is made up of 4Q198 (Tobit), 4Q212 (Enoch), 4Q534 (Noah), 4Q536 (Aramaic C), 4Q540 (Apocryphon of Levi) and 4Q556 (Book of Giants). This intermediate level cluster does not appear to relate to any of the low level clusters.

A final intermediate level cluster includes 1Q24 (Enoch Giants), 4Q214 (Aramaic Levi), 4Q214b (Aramaic Levi), 4Q318 (Brontologion) and 4Q548 (Visions of Amram). This group relates to some degree to Biblical Daniel – Ezra, though its content is closer to the Aramaic Levi – Enochic low level cluster.

High level clusters

By definition, these provide coarse groupings of texts. It is important to note that some individual texts, such as 4Q562, appear to form a cluster on their own. This indicates that either the criteria being applied are inadequate to characterize them or that they are indeed quite distinct from the rest of the texts. Given this, three high level clusters emerge:

- 1) All the low level clusters except Biblical Daniel – Ezra
- 2) The Tobit, Noah, Apocryphon of Levi - Enochic group, joined by 4Q156 (Targum of Leviticus) and 4Q246 (Aramaic Apocalypse)
- 3) The Biblical Daniel - Ezra, low level cluster plus the Aramaic Levi – Enochic intermediate cluster.

We might characterise these groups as 1) ‘mainstream Qumran Aramaic’, 2) ‘Heterogeneous content including Tobit’, 3) ‘Biblical’ text types. The next step is to see

if these clusters are robust. One way is to re-cluster using standardised scores of the criteria variable rates so that all 22 *a priori* criteria are given equal weight.

Hierarchical cluster classification using standardised rates of textual criteria

The purpose of standardisation is to give equal weight to all criteria. Figure 7.5 shows the hierarchical cluster dendrogram for the standardised rates of the 22 *a priori* criteria. This indicates that *m*-prefixed infinitive peal rates are closely related to total infinitive peal rates and that *dy* relative particle rates are closely related to total relative particle rates. The two-dimensional scaling confirms this (Figure 7.6). Including these pairs would result in over-weighting, so the total infinitive peal and relative particle rates were excluded at this stage. Figure 7.6 also shows that other criteria are distributed fairly evenly around a central point consistent with relatively equal weightings. The two variables that are more central 3fp verbs and 2ms suffix *-kh*. Figure 7.7 presents the hierarchical cluster classification using the remaining 20 criteria. The dendrogram indicates that six scrolls are quite distant from the others: 4Q214a, 4Q244, 4Q532, 4Q549, 4Q557 and 4Q559. The lengths of these texts are 114, 81, 81, 117, 24 and 85 words respectively. The likelihood is that they are too short for there to be enough occurrences of the 20 criteria variant forms to classify them adequately. In view of this, these six short scrolls are excluded from the standardised rates hierarchical clustering and will be considered later once exploratory analyses are able to establish distinct clusters.

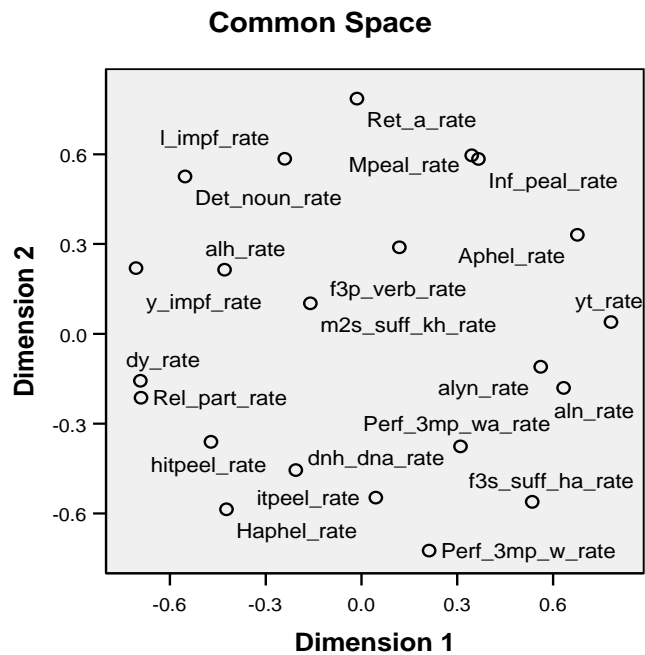


Figure 7.6 Two dimensional scaling using the Euclidean distance for standardised rates between the 22 *a priori* criteria for the 74 Qumran Aramaic scrolls.

Low level clusters

Working down the dendrogram several low level clusters can be discerned:

- 1) 4Q208 (Enoch), 4Q209 (Enoch), 4Q213b (Aramaic Levi), 4Q543 (Vision of Amram)

2) 4Q196 (Tobit), 4Q197 (Tobit), 4Q213 (Aramaic Levi), 4Q544, 4Q546, 4Q553 (Four Kingdoms), 4Q558 (Vision).

3) 1Q24 (Enoch Giants), 2Q24 (New Jerusalem), 4Q210 (Enoch), 4Q552 (Four Kingdoms), 4Q554 (New Jerusalem).

Elsewhere in the dendrogram the distances between scrolls are somewhat larger and arguably relationships cannot be designated low level.

Intermediate level clusters

The three low level clusters group together to form a single intermediate level cluster dominated by Enochic, New Jerusalem and Four Kingdom scrolls together with the Vision of Amram.. Amongst these three level clusters are several other scrolls almost as closely related: 4Q538 (Testament of Judah) and 6Q8 (Giants) form the first pair, then a group comprising 4Q198 (Tobit), 4Q204 (Enoch), 4Q205 (Enoch), 4Q531 (Giants), 4Q545 (Vision of Amram) and 4Q547 (Vision of Amram). To these we can add 4Q211 (Enoch), 4Q529 (Words of Michael) and 4Q550 (Proto-Esther) This whole group can be considered as a single intermediate level cluster.

In view of the very gradual increases in distances for the rest of the agglomeration, it is difficult to be certain of any further intermediate clusters. However, after a little gap a biblical intermediate level cluster appears, 4Q112, 4Q113, 4Q115 and 4Q117.

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
4Q208	19	↓×↓				
4Q209	20	↓	↓			
4Q213b	26	↓×↓		↔		
4Q543	48	↓		↔		
4Q538	43	↓×↓	↓	↓	↓	↓
6Q8	70	↓		↔		
4Q544	49	↓×↓		↔		
4Q546	51	↓	↓			
4Q553	59	↓×↓		↔		
4Q558	63	↓		↔	↔	
4Q196	9	↓		↔	↔	
4Q213	24	↓	↓	↓		↔
4Q197	10	↓		↔		
4Q531	37	↓	↓	↓	↓	↓
4Q545	50	↓	↓	↓		↔
4Q547	52	↓	↓	↓	↓	↔
4Q204	15	↓×↓		↔	↔	
4Q205	16	↓	↓		↔	
4Q198	11	↓	↓	↓		↔
1Q24	4	↓×↓		↔		
4Q554	60	↓		↔	↔	
2Q24	5	↓×↓		↔		
4Q210	21	↓	↓		↔	
4Q552	58	↓	↓	↓		↔
1Q21	2	↓×↓		↔		
11Q18	72	↓		↔	↔	↔
4Q318	34	↓	↓	↓		↔
4Q529	35	↓	↓	↓	↓	↔
4Q550	55	↓	↓	↓	↓	↔
4Q211	22	↓	↓	↓	↓	↔
4Q530	36	↓	↓	↓	↓	↔
4Q550c	57	↓	↓	↓	↓	↔
1Q23	3	↓	↓	↓	↓	↔
4Q560	65	↓	↓	↓	↓	↔
4Q201	12	↓	↓	↓	↓	↔
4Q542	47	↓	↓	↓	↓	↔
4Q202	13	↓	↓	↓	↓	↔
4Q156	8	↓	↓	↓	↓	↔
4Q246	33	↓	↓	↓	↓	↔
4Q212	23	↓	↓	↓	↓	↔
4Q550a	56	↓	↓	↓	↓	↔
4Q540	45	↓	↓	↓	↓	↔
4Q213a	25	↓	↓	↓	↓	↔
4Q539	44	↓	↓	↓	↓	↔
1Q20	1	↓	↓	↓	↓	↔
11Q10	71	↓	↓	↓	↓	↔
4Q214	27	↓	↓	↓	↓	↔
4Q243	31	↓	↓	↓	↓	↔
4Q556	61	↓	↓	↓	↓	↔

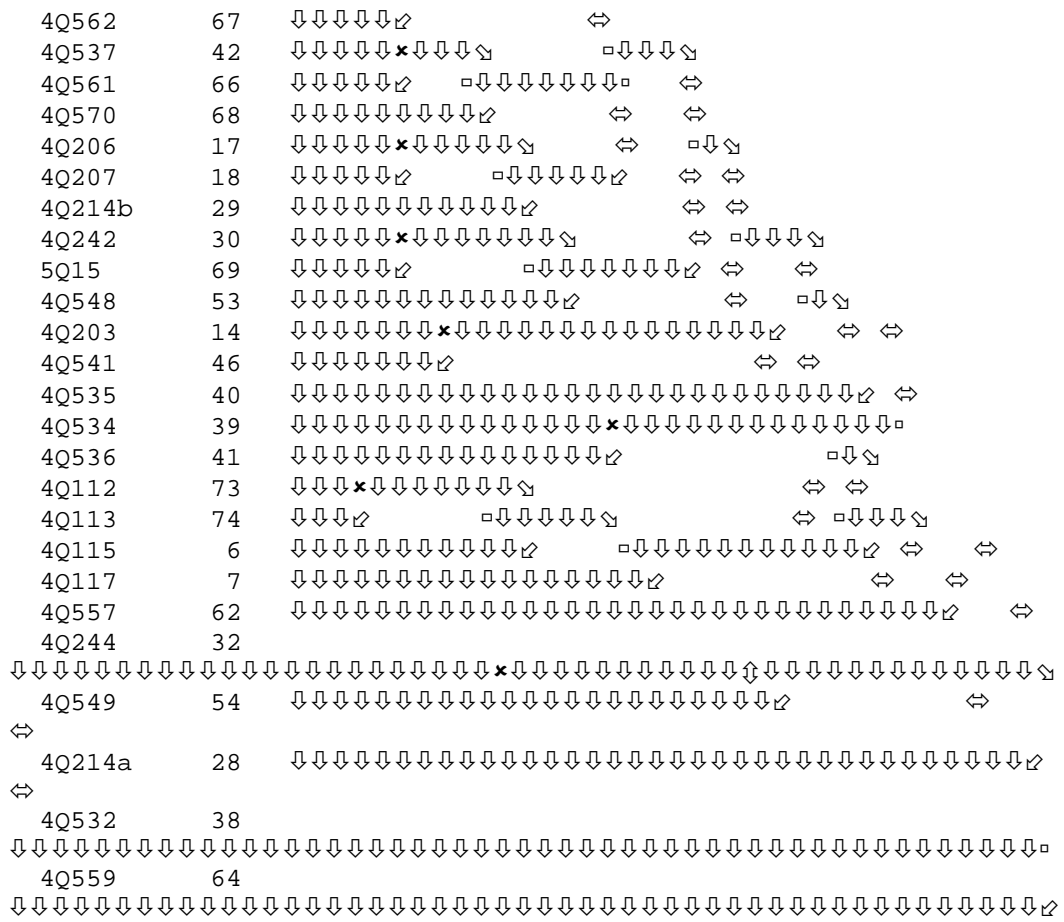


Figure 7.7 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls using standardised rates of 20 *a priori* criteria.

High level clusters

It is difficult to be convinced of the presence of any distinct high level clusters. The agglomeration proceeds relatively evenly in a stepwise manner incorporating those scrolls that do not fall into obvious clusters along the way. Hence the nine intermediate clusters are probably the most distinct groups arising from this analysis. The first three all have a substantial contribution from Enochic texts. After come two pairs of closely related intermediate level clusters. Finally there is a biblical group and a final loosely associated Enochic group.

Comparison of raw and standardised classifications

The classifications that are being compared are one in which all classification criteria are given equal weight (standardised rates) and one in which classification criteria are weighted according to the amount of data available (i.e. criteria that occur rarely are given little weight whilst frequently occurring criteria are weighted more heavily, raw rates). The standardised rates include two fewer criteria and classify six fewer scrolls.

First, the place in the raw rates' classification of the six scrolls, 4Q214a, 4Q244, 4Q532, 4Q549, 4Q557 and 4Q559, which failed to cluster in the standardised rates' classification is worth examining. 4Q214a, an Aramaic Levi scroll, is placed close to the first Enochic – Aramaic Levi low level cluster in the raw rates' classification. 4Q244, though not part of a low level cluster, is incorporated in the high level cluster that

contains the five low level clusters. 4Q532 is relatively unrelated to other scrolls in the raw rates' classification, but falls within the mainstream Qumran Aramaic high level cluster as does 4Q549. 4Q557 agglomerates at the antepenultimate step whilst 4Q559 also falls within the mainstream Qumran Aramaic high level cluster. In summary, both classifications find 4Q557 to be distinct from the vast majority of other scrolls, whilst the raw rates' classification places the five other scrolls excluded by the standardised rates' classification within the mainstream Qumran Aramaic group. This probably reflects the fuller use of data by the raw rates classification.

Secondly, it is worth examining the fate of the low level clusters identified by the raw rates' classification in the standardised rates' classification. The first low level cluster comprises 4Q208, 4Q209, 4Q211, 4Q213a, 4Q213b, 5Q15 and 4Q539. 4Q208, 4Q209 and 4Q213b remain together, but 4Q211 is not included in any of the low level clusters, but is amongst them as part of the related intermediate level cluster. 4Q213a and 4Q539 also remain together, but far further down the dendrogram. The suggestion that the raw rates' classification low level cluster has separate components would be consistent with this situation.

The second low level grouping comprises 1Q20, 4Q201, 4Q202, 4Q204, 4Q205, 4Q530, 4Q531, 4Q542, 4Q552, 4Q553, and 4Q558. 1Q20 is associated with 4Q213a and 4Q539 from the previous grouping, whilst 4Q201 and 4Q202 tag on at the end of the first the intermediate cluster. 4Q204 and 4Q205 remain together close to 4Q531 and to a lesser extent 4Q552, 4Q553 and 4Q558 within the first intermediate cluster. 4Q542 is a little

further down, but still within what might be considered the mainstream Qumran Aramaic texts. Hence a main Four Kingdoms – Enochic grouping persists with other texts moving away.

The next grouping to be considered is that provisionally labelled as New Jerusalem texts: 2Q24, 11Q18, 4Q210, 1Q21, and 4Q554. These texts remain very close together in the standardised rates' classification.

A mixture of mainly Tobit and Visions of Amram scrolls form the next raw rates' low level cluster: 4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546 and 4Q547. Again these scrolls remain fairly closely grouped in the standardised rates' classification.

Finally there are the biblical scrolls 4Q112, 4Q113, 4Q115 and 4Q117 together with 4Q529 the Words of Michael. The biblical scrolls continue to form a close group, but 4Q529 moves to be amongst the mainstream Qumran Aramaic texts.

In summary, the low level clusters remain fairly robust to the use of raw or standardised rates for classification, though not completely intact. The intermediate level clusters are mainly built around the low level clusters except the one made up of 4Q198 (Tobit), 4Q212 (Enoch), 4Q534 (Noah), 4Q536 (Aramaic C), 4Q540 (Apocryphon of Levi) and 4Q556 (Book of Giants). 4Q534 and 4Q536 remain closely related, 4Q212, 4Q540 and 4Q556 a little more distinct in keeping with the intermediate level of clustering, whilst 4Q198 moves into the mainstream Qumran Aramaic group. Again, this shows that the

different clustering criteria have not made a great difference, but that intermediate level clusters are likely to contain greater heterogeneity of text types.

The classifications indicate that the following text groups can be considered typical:

4Q208, 4Q209, 4Q213a – Enochic (E)

4Q213b, 4Q539, 5Q15 – Aramaic Levi, Apocryphon of Joseph B, New Jerusalem (A)

4Q204, 4Q205, 4Q531, 4Q552, 4Q553, 4Q558 – Enochic, Four Kingdoms etc (F)

2Q24, 11Q18, 4Q210, 1Q21, 4Q554 – New Jerusalem (N)

4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547 – Tobit, Visions of Amram (T)

4Q112, 4Q113, 4Q115, 4Q117. – Biblical (B)

Variable	E	A	F	N	T	B	p-value
Determined nouns	68.2	66.0	75.3	104.2	40.0	144.4	<.001
Afel	17.5	17.1	14.8	5.5	10.1	5.8	.042
Hafel	0	0	0	0	0	18.6	.001
Itpeel	0.4	0	8.2	1.4	6.2	0.5	.18
Hitpeel	0	0	0	0	0	6.1	<.001
<i>m</i> -prefixed Peal infinitives	4.4	0	11.0	0	12.8	3.0	.003
Peal infinitives	4.4	0	11.0	0	12.8	3.0	.003
<i>l</i> prefixed imperfects	0	0	1.8	3.5	2.0	0	.20
<i>y</i> prefixed imperfects	0	0	9.1	2.2	8.1	16.7	.026
3fp verbs	0	0	0	0	0.6	0	.14
Perfect 3mp ending <i>w</i>	0	8.5	15.3	3.5	9.4	6.1	.012
Perfect 3mp ending <i>w'</i>	0	0	0	0	0	0	.11
Retained aleph before suffix	0	0	0	0	0	0	.35
2ms pronominal suffix <i>kh</i>	0	0	0	0	0	0	.84
3fs pronominal suffix <i>ha</i>	0	0	0	0	0	0	.50
Relative particles	4.4	19.9	41.1	35.3	21.2	48.1	<.001
Relative particle <i>dy</i>	0	19.9	25.5	35.3	21.2	48.1	<.001
Direct object marker <i>yt</i>	0	0	0	0	0	0	.65
<i>dnh</i> or <i>dn'</i> forms	0	0	0	0	0	5.7	<.001
Demonstrative pronoun ' <i>lh</i>	0	0	0	0	0	1.5	.068
Demonstrative pronoun ' <i>lyn</i>	0	0	2.4	0	0	0.5	.65
Demonstrative pronoun ' <i>ln</i>	0	0	0	0	0	0	.84

Table 7.4 Medians of rates per 1,000 words for the 22 criteria occurring in the different typical text type groups.

Characterisation of typical text groups

Having identified some provisional typical text groups the next stage is to test in what ways they are succinct, that is which of the *a priori* criteria distinguishes between them? The number of scrolls in each of these groups is very small to expect there to be statistical differences unless these are very large. Table 7.4 shows the medians for each of the 22 *a priori* variables used in the raw rates clustering for each typical text type groups together with the significance level for difference between groups. The p-value is computed exactly for a Kruskal-Wallis test¹ that the six groups comes from the same general text type group, though for some variables the asymptotic p-value is given where an exact value could not be calculated by iterative methods. Of the 22 variables, 11 show significant differences between groups indicating that there are substantial differences between these typical text type groups in terms of the *a priori* criteria. The large number of zero medians, although not telling the whole story, indicates that in many instances a particular feature is not characteristic of that specific text type. Closer inspection of Table 7.4 reveals that the differences between groups E and A are small as suggested by the clustering structure. This is formally confirmed by Mann-Whitney U-testing² where

¹ The Kruskal-Wallis test is a one-way analysis of variance that uses ranked (ordinal) data rather than interval data. It is thus suitable for data that deviate from a normal distribution (termed non-parametric) as is the case for the 22 textual variables.

² The Mann-Whitney U-test is another non-parametric test (i.e. used for data that deviate from a normal distribution). It is equivalent to the parametric Student's t-test. Essentially, it ranks the data across the two groups to be compared then determines how probable it is that the rank places of one group are the same

no significant differences are present for any of the 22 variables between these groups. In view of this, it is parsimonious to combine these typical text type groups as in the original low level clustering based on raw rates. Inspection of Table 7.4 also reveals that the 'B', biblical text type group, appears distinct from the other groups on many of the criteria. Re-running the Kruskal-Wallis test on the four remaining groups, 'EA', 'F', 'N' and 'T', results in the differences between groups for *hafels*, *dnh* or *dn*', and *hitpeels* no longer reaching significance. These results indicate that the heterogeneity is not attributable to Biblical group text types only. The salient characteristics for the significant differences for these typical text types may be summarized as follows:

Enochic, Aramaic Levi - high *afel*, absent *hafel*, few imperfects, few relative particles;
Enochic, Four Kingdoms - high *afel*, absent *hafel*, high 3mp Perfect *-w* ending, high relative particles but relatively small proportion of these *dy*, use of *'lyn* as demonstrative pronoun;
New Jerusalem – low *afel*, absent *hafel*, high *l-* prefixed imperfects, low *y-* prefixed imperfects;
Tobit – low determined state nouns, presence of 3fp verbs;
Biblical – high determined state nouns, low *afel*, high *hafel*, low *itpeel*, high *hitpeel*, high *y-* prefixed imperfects, high relative particles represented by *dy*, *dnh* or *dn*' forms present, *'lh* used as demonstrative pronoun.

as those of the other group. Hence equal ranking would give a probability of 1.0 whereas if the ranks of all the members of one group were higher than any of the members of the other group, the probability would tend to zero. As usual, a probability of less than 0.05 is taken as statistically significant.

Some of these characteristic features may relate to the general composition of the texts in terms of parts of speech. Figure 7.8 shows the relationship between determined state

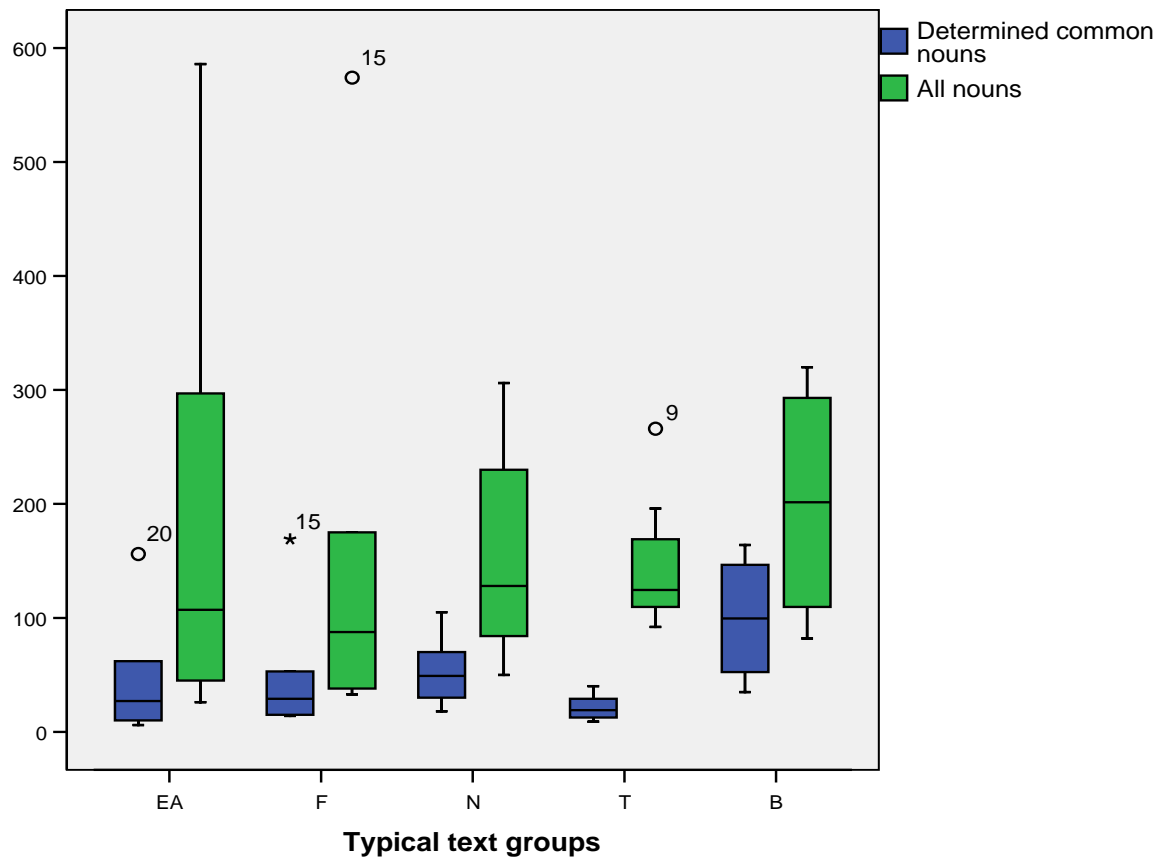


Figure 7.8 Box plot of total nouns and determined state common nouns in the five typical text type groups.

nouns and all nouns for the different groups. The total numbers are similar across all groups ($p=.77$), but the Biblical group has a higher proportion of determined state common nouns, so that this characteristic feature is robust to potential compositional

differences. The same applies to *y*-prefixed Imperfects and total verb count ($p=.23$, Figure 7.9). Characterisation may be aided by examining those criteria that were not selected for inclusion in clustering because there were too few scrolls where they were

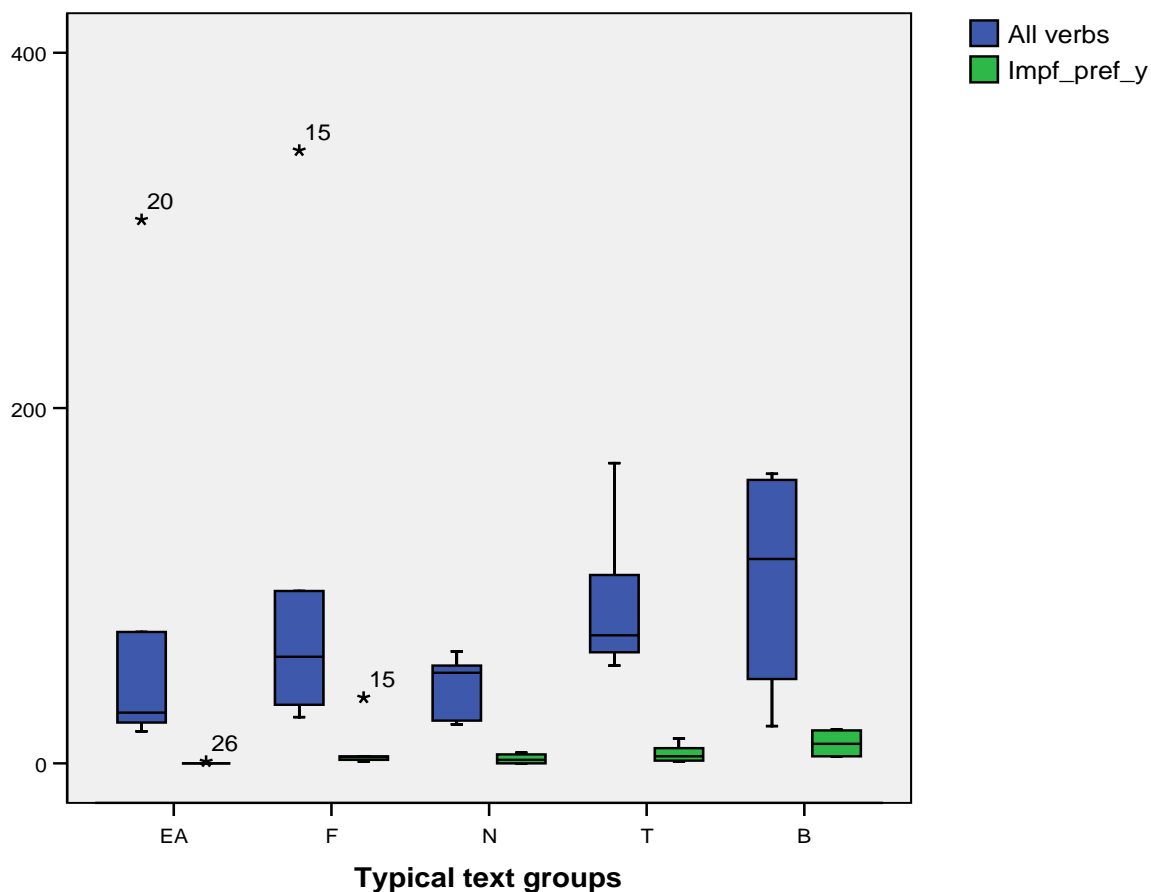


Figure 7.9 Box plot of total verbs and *y*-prefixed imperfect verbs in the five typical text type groups.

present. Of these there are significant differences between the five typical text type groups for hitpaal ($p=.001$) and first plural pronoun '*nHn*' ($p=.016$). Hitpaal forms occur exclusively in the Biblical group whilst '*nHn*' occurs only in the Enochic - Four Kingdoms and Tobit groups.

The h-forms of verbs, presence of *dnh* and *dn'* together with *'lh* forms would be consistent with the Biblical group representing an older text type. By contrast, the Enochic – Four Kingdoms group does not have h-forms of verbs, has a low proportion of *dy* as a relative particle and uses *'lyn* as a demonstrative pronoun, consistent with a younger text type. However, the other groups do not have a profile that fits unequivocally into a diachronic paradigm for Aramaic morphology. This raises the possibility of the influence of dialects plus or minus a diachronic element that means that Qumran Aramaic cannot be considered a single homogeneous entity in terms of morphological criteria.

Realigning the Qumran Aramaic texts

The identification of distinct text types within the Qumran Aramaic corpus allows a reevaluation of the relationship of these texts one to another. Hitherto scrolls have been aligned in terms of cave and presumed content, but the textual criteria do not always support this. First, and probably easiest, there appears to be a 'biblical' group, at least in terms of Aramaic texts, even though the concept of 'bible' had not come into existence at the time of Qumran. It is thus justifiable to use the anachronistic labels 'biblical' and 'non-biblical' with regard to the Aramaic scrolls. The biblical scrolls display archaic features. One possibility is that there has been deliberate 'archaizing' of the texts. However, such archaizing is inconsistent; for example, *afels* occur alongside *hafels*. A similar argument applies to the idea of 'hebraizing' of the texts. Another possibility is

that the texts were originally written at a time of transition between forms and that we do not have adequate data to determine when an *afel* form was used rather than a *hafel*. A further possibility is that the texts are composites of archaic and younger sections or an archaic text has been redacted at a later date.

By contrast to the Biblical group, the EA group displays ‘younger’ features. The scrolls tentatively fall into two sub-groups. First 4Q213a, 4Q213b and 4Q539 comprise two Testament of Levi scrolls and what is termed an Apocryphon of Joseph B. However, the latter is only very loosely related to the Hebrew Apocryphon of Joseph B and its phrase ‘my father Jacob’ is shared by 4Q213a. This raises the possibility that 4Q539 would be better labelled as another Testament of Levi fragment. Secondly, 4Q208, 4Q209 and 5Q15 are labelled as two Astronomical Enoch scrolls and a New Jerusalem scroll. Two other New Jerusalem scrolls comprise part of another distinct text type group. All three are made up largely of text that describes spatial or temporal measures.

The New Jerusalem text type group does not include 5Q15, but does contain three New Jerusalem texts from different caves: 2Q24, 4Q554 and 11Q18. The fact that these all have similar textual features would be consistent with a common origin despite their disparate locations of discovery. Together with these three scrolls is a scroll from yet another cave, 1Q21, labelled as the Testament of Levi. Finally there is 4Q210, another Astronomical Enoch scroll, though textually distinct from 4Q208 and 4Q209. This typical text group thus stands in apposition to the previous one. They both share common content, Astronomical Enoch and New Jerusalem, but appear to be in different

'voices'. This suggests that this group may represent texts in the same Enochic movement that produced the EA group scrolls either earlier or later. The relatively low proportion of *y*-prefixed imperfects would be more consistent with this group coming before the EA group, but a single criterion cannot be considered decisive. Alternatively, these two groups of Enochic texts were written or copied by scribes from different backgrounds possibly around the same time.

The Four Kingdoms text type represents yet another distinct Enochic group. 4Q204 and 4Q205 are more than fragmentary Enochic texts. 4Q531 is part of the Book of Giants. By contrast 4Q552, 4Q553 and 4Q558 are much shorter, the first two labelled as Four Kingdoms texts and the last as a Vision. The characteristic features, notably the low proportion of *dy* forms of the relative particle, would be consistent with a younger group of texts, perhaps of similar age to the EA group. Although it would be going too far to say that 4Q531, 4Q552, 4Q553 and 4Q558 are part of the same Enochic corpus as 4Q204 and 4Q205, the statistical evidence indicates that they are likely to share a similar provenance.

The final typical text type group comprises two Tobit scrolls, what can now be seen as a homogeneous group of Vision of Amram scrolls and 4Q213, part of the so-called Testament of Levi. The textual criteria support neither an early or late date for these texts. A pre-Qumran date has been proposed and its relationship to the Vision of Amram

noted.³ Interestingly 4Q213 is textually distinct from 4Q213a and 4Q213b that have features that would suggest a later date of composition or, perhaps more probably, scribal redaction/copying. A non-Qumran date for Tobit is also highly likely and the morphological data would support a similar pre-Qumran dating and/or region of origin. This group is distinct in being neither biblical nor Enochic, but can be considered as pre-Qumranic Aramaic texts.

Internally situating other major Qumran Aramaic texts

Four of the seven scrolls with over 1,000 words – 4Q112, 4Q196, 4Q204 and 4Q209 – have already been aligned. But this leaves three other long scrolls – 1Q20 (Genesis Apocryphon), 4Q201 (Book of Enoch) and 11Q10 (Targum of Job) – to be situated in relationship to the typical text type groups. 1Q20, the Genesis Apocryphon, is closely related to 4Q204, 4Q531, 4Q542 and 4Q558 in the raw rates dendrogram (Figure 7.2) and to 4Q213a and 4Q539 in the standardised rates dendrogram (Figure 7.10). 4Q204, 4Q531 and 4Q558 are all members of the Enochic - Four Kingdoms text type group whilst 4Q213a and 4Q539 both belong to the Enochic - Aramaic Levi group. We have noted that these text types are fairly similar, but that the Enochic – Four Kingdoms group have fewer *dy* relative particle forms and more *'lyn* demonstrative pronoun forms. 1Q20 has 32 relative particles per 1,000 words, 31 of which are *dy* and an *'lyn* rate of 1.3 similar to its *'ln* rate and higher than its *'lh* rate. It also has a very low *hafel* rate of 0.4 per 1,000 words,

³ Jonas C. Greenfield, Michael E. Stone & Esther Eshel *The Aramaic Levi Document*. Leiden: Brill, 2004.

C A S E	0	5	10	15	20	25
Label	Num	+-----+-----+-----+-----+-----+				
4Q208	19	↓×↓↘				
4Q209	20	↓↘ □↓↘				
4Q213b	26	↓×↓↘ ↔				
4Q543	45	↓↘ □↓↘				
1Q24	4	↓×↓↘ ↔ ↔				
4Q554	56	↓↘ □↓↘ ↔				
2Q24	5	↓×↓□ □↓↘				
4Q210	21	↓↘ ↔ ↔ ↔				
4Q552	54	↓↓↓↘ ↔ ↔				
1Q21	2	↓×↓↘ ↔ ↔				
11Q18	66	↓↘ □↓↓↓↘ ↔				
4Q318	32	↓↓↓↘ ↔				
4Q196	9	↓×↓↘ ↔				
4Q213	24	↓↘ □↓↘ ↔				
4Q553	55	↓×↓□ ↔ ↔				
4Q558	58	↓↘ ↔ ↔ ↔				
4Q197	10	↓↓↓↘ ↔ ↔				
4Q544	46	↓↓↓×↓↘↘↘ ↔				
4Q546	48	↓↓↓↘ ↔ ↔ □↓↘				
4Q531	35	↓↓↓↓↘ □↓□ ↔				
4Q547	49	↓↓↓↓↓↓↓□ ↔ ↔				
4Q545	47	↓↓↓↓↓↓↓↘ ↔ ↔				
4Q538	40	↓×↓↓↓↓↘ ↔ ↔				
6Q8	64	↓↘ ↔ ↔ ↔				
4Q198	11	↓↓↓↓↓↓↓↘↘↘ ↔				
4Q560	59	↓↓↓↓↓↓↓↘ ↔				
4Q529	33	↓↓↓↓↓×↓↓↓↓↓↘↘↘				
4Q550	51	↓↓↓↓↓↘ ↔ ↔				
4Q204	15	↓↓↓×↓↓↓↓↓↘ ↔ ↔				
4Q205	16	↓↓↓↘ □↓□ □↓↘				
4Q530	34	↓↓↓↓↓↓↓↓↘ ↔ ↔ ↔				
4Q542	44	↓↓↓↓↓↓↓↓↓↘ ↔ ↔				
4Q211	22	↓↓↓↓↓↓↓↓↓×↓↘↘↘ □↓↓↘↘				
4Q550c	53	↓↓↓↓↓↓↓↓↘ ↔ ↔				
4Q212	23	↓↓↓↓↓↓↓×↓↓↘↘ ↔ ↔				
4Q550a	52	↓↓↓↓↓↓↓↘ □↓↓↘↘ ↔				
4Q156	8	↓↓↓↓↓×↓↘ ↔ ↔				
4Q246	31	↓↓↓↓↓↘ □↓↓↘↘ ↔				
4Q540	42	↓↓↓↓↓↓↓↘ ↔				
4Q214	27	↓↓↓↓↓×↓↓↓↓↓↓↓↘↘ ↔				
4Q243	30	↓↓↓↓↓↘ □↓↓↓□				
4Q556	57	↓↓↓↓↓↓↓×↓↓↓↓↓↘↘ ↔				
4Q562	61	↓↓↓↓↓↓↓↘ □↓↓↘↘				
11Q10	65	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓□ ↔				
4Q537	39	↓↓↓↓↓↓↓×↓↓↘↘ ↔ ↔				
4Q561	60	↓↓↓↓↓↓↓↘ □↓↓↓↓↓↘↘ ↔				
4Q570	62	↓↓↓↓↓↓↓↓↓↓↓↘ ↔				
4Q213a	25	↓↓↓↓↓×↓↘ ↔				

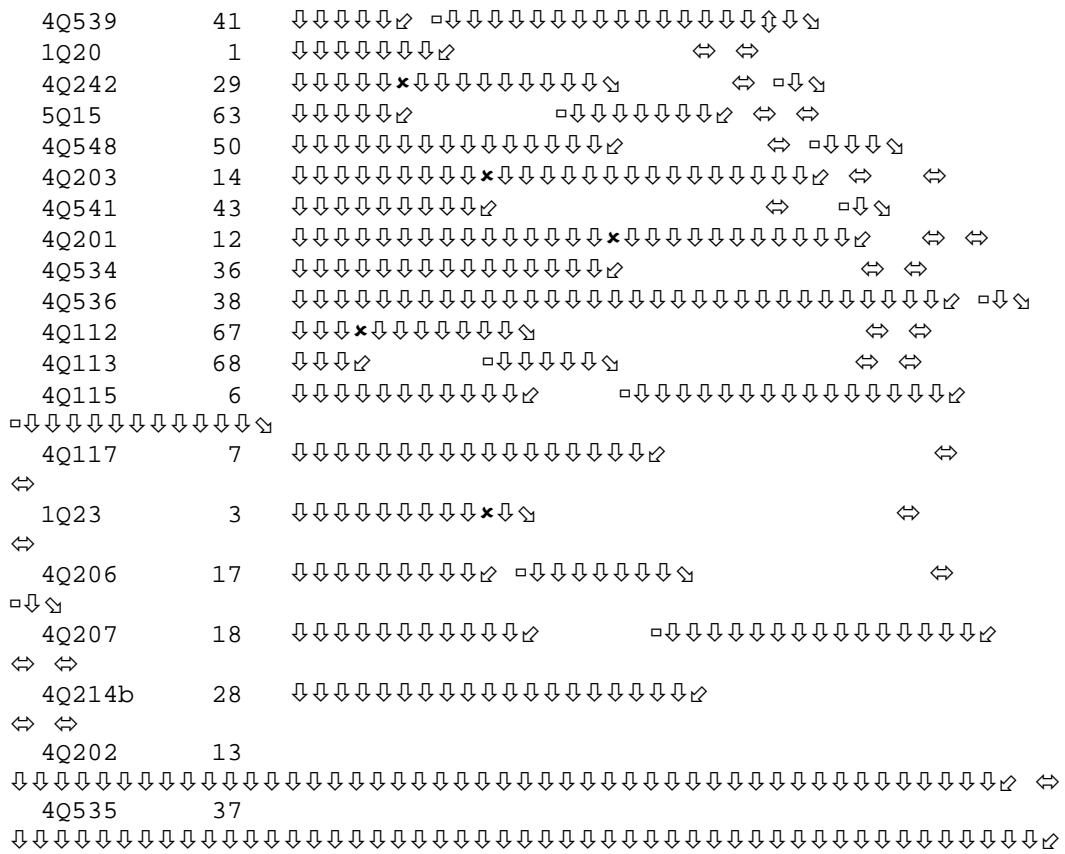


Figure 7.10 Hierarchical clustering dendrogram of 68 Qumran Aramaic scrolls using standardised rates of 20 *a priori* criteria.

though this is much lower than its *afel* rate of 11 per 1,000 words. These data suggest that it has a slightly 'earlier' text type than the Enochic - Four Kingdoms group, but its close relationship to both EA and F text types suggests that it is part of the Enochic mainstream.

4Q201 is part of the Book of Enoch and so, unlike 1Q20, would, *a priori*, be expected to be situated close to at least one of the three Enochic text types. By raw rate criteria it is closest to 4Q530 (Book of Giants), but also to 1Q20, 4Q204, 4Q531, 4Q542 and 4Q558 (Figure 7.2). By standardised rates criteria it clusters quite late and is closest to 4Q534 (Book of Noah), though this is not the case when all 74 scrolls are considered. 4Q201 has a relatively low proportion of *dy* relative particles, 20 out of 27 per 1,000 words, and an *'lyn* rate of 6 per 1,000 words compared with 1 per 1,000 words for *'lh* and zero for *'ln*. It thus appears best situated with the Enochic – Four Kingdoms text type group.

11Q10, on the other hand, does not cluster closely to other scrolls using raw or standardised rates (Figures 7.2 and 7.10). This may reflect the translation aspect of the text. It has, for example a high *hafel* rate of 22 per 1,000 words compared with an *afel* rate of 2 per 1,000 words, but a low *dnh* or *dna* rate close to zero, an *'lyn* rate higher than either *'lh* or *'ln* rates and an absence of *hitpeel* forms whilst *itpeels* are present. Hence there is a mixture of archaic and younger features. As such, it does not align easily with any of the text type groups. This finding flags up the need to try to

situate all the Qumran Aramaic texts in a broader Aramaic context because they may not fit neatly into a purely corpus-derived classification of Qumran Aramaic texts. As Ursula Schattner-Rieser comments on the longer Qumran Aramaic texts,

Le cycle d'Enoch...se compose de nombreux fragments et unités séparés, écrits par des mains différentes. La langue des fragments énochiens ne forme pas un ensemble homogène, et les particularités linguistiques y sont très prononcées. Le Targoum de Job de la grotte 11...s'agit cependant d'un texte mutilé, qui suit son modèle hébreu de très près.⁴

⁴ Ursula Schattner-Rieser. *L'araméen des manuscrits de la mer Morte*. Lausanne : Editions de Zèbre, 2004, p109.

8. External textual correlates

Having identified various text type groups within the Qumran Aramaic corpus the next step is to examine how these groups relate to Aramaic texts beyond Qumran. A range of texts presents itself, though their relevance is open to debate. Ideally external texts should have a clear provenance in terms of date and location of composition. Many dateable texts that are likely to be contemporaneous with Qumran material are short because they are inscriptions and hence not useful for conventional statistical methods. Other longer texts that may originate during the same period are not easy to date because of later scribal accretions. The foremost examples of such non-dateable texts are the Targums.¹ As noted for 11Q10, Targums represent a peculiar text type in themselves because of their translation nature where Hebraisms and older textual features may occur because of the Vorlage. Associations with Qumranic text types may reflect this tendency rather than indicate common textual origins. Importantly, however, the criteria used for correlation are not those which would be chosen to classify the Targums themselves. For example, *n*-prefixed third person imperfect verbal forms are rare in the Qumran Aramaic corpus, but relatively common in some of the Targums and might be chosen to classify the Targums themselves, but is not one of the correlates being tested: Alinda Damsma provides a recent review of textual criteria that are more relevant to these later writings.² These correlates represent features that clearly differentiate Qumran Aramaic texts and

¹ Here I will use the Anglicized plural Targums to avoid confusion between the Hebrew targumim and the Aramaic targumin.

² A. Damsma (2008) reviews early Jewish dialect features in a Targumic Tosefta to Ezekiel.

hence may well have been present in any form of the Targums at the time when the Qumran scrolls were written or copied. Thus one caveat is that Targums may have included such features at one time, but these may have been removed subsequently during later redaction and scribal copying.

Book	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
Genesis	9211	5221	3450	1353	14888	488	34611
Exodus	7908	3768	2321	1136	12860	286	28279
Leviticus	5505	2590	1781	655	9396	261	20188
Numbers	8610	3199	1968	1526	11701	238	27242
Deuteronomy	6221	3577	2984	898	10876	297	24853

Table 8.1 Parts of speech count and total words for the Targum Onkelos by book.

Targum Onkelos

Targum Onkelos is thought to be amongst the oldest Targums, originating in Palestine but reaching its final form in the east under Babylonian influence. However, this conventional view has not been tested statistically and depends on scholarly opinion.

Table 8.1 shows Targum Onkelos' parts of speech composition for each book of the Pentateuch. Table 8.2 shows the rates per 1,000 words for each of the 22 *a priori* criteria used in the raw rates clustering of the Qumran Aramaic scrolls. General inspection of

Criterion	Genesis	Exodus	Leviticus	Numbers	Deuteronomy
Determined nouns	119	138	156	126	135
Afel	15	14	19	10	15
Hafel	0.1	0.5	0	0.1	0.2
Itpeel	6	6	8	5	7
Hitpeel	0	0	0	0	0
<i>m</i> -prefixed Peal infinitives	8	7	6	7	13
Peal infinitives	8	7	6	7	13
<i>l</i> prefixed imperfects	0	0	0	0	0
<i>y</i> prefixed imperfects	10	21	56	23	30
3fp verbs	1.5	1.2	0.6	1.1	1.0
Perfect 3mp ending <i>w</i>	13	10	3	14	4
Perfect 3mp ending <i>w'</i>	0	0	0	0	0
Retained aleph before suffix	0.1	0.3	0	0.2	0.6
2ms pronominal suffix <i>kh</i>	0	0	0	0	0
3fs pronominal suffix <i>h'</i>	1.2	1.4	1.1	1.4	0.6
Relative particles	34	40	42	40	47
Relative particle <i>dy</i>	0	0	0	0	0
Direct object marker <i>yt</i>	28	34	36	25	24
<i>dnh</i> or <i>dn'</i> forms	0	0	0	0	0
Demonstrative pronoun <i>'lh</i>	0	0	0	0	0
Demonstrative pronoun <i>'lyn</i>	2	1	1	2	11
Demonstrative pronoun <i>'ln</i>	0	0	0	0	0

Table 8.2 Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Onkelos.

Table 8.2 reveals that Targum Onkelos has features closer to the Four Kingdoms text type than the Biblical text type despite its 'biblical' content. The hafel rate is very low,

unlike 11Q10, but otherwise the textual profile appears to be not that distant from this Qumran Targum. However, formal clustering using raw rates (Figure 8.1) of the Qumran scrolls with the five books of Targum Onkelos shows that this initial inspection fails to consider how this Targum relates to the Qumran Aramaic corpus on all 22 criteria.

First, and reassuringly, the five Targum Onkelos books cluster closely together. This indicates that, assuming a common source for the five books of Targum Onkelos, the hierarchical cluster analysis is able to identify closely related text types. Hence, we can infer that, although the sources of the Aramaic Qumran scrolls are unknown, the identified clusters are likely to be closely related according to the textual criteria used. Secondly, the core members of the Qumran Aramaic text types remain closely clustered, though the Enochic, Four Kingdoms group is perhaps a little more dispersed. Again this is reassuring because introducing external texts into the clustering procedure has not perturbed the text type groups identified by purely internal comparison within the Qumran corpus. This finding supports the robustness of these text types as distinct groups. Thirdly, of all the identified text type groups, Targum Onkelos is most closely related to the Biblical group. This would not have been predicted by initial inspection, but statistically is the case. Fourthly, the Targum Onkelos group agglomerates late in the agglomeration schedule indicating that it is relatively unrelated to the bulk of the Qumran Aramaic scrolls. This indicates that on the 22 textual criteria Targum Onkelos is distinct in nature from the majority of the Qumran Aramaic scrolls. Fifthly, and a

4Q208	19	↓ ↘
4Q209	20	↓ ↕ ↓ ↘
4Q211	22	↓ ↗ ↔

5Q15	69	↓↓↓↓↑↓↓↓↓↓↓↓
4Q213a	25	↓×↓□ ⇔
4Q213b	26	↓⇨ ⇔ □↓⇨
4Q539	44	↓↓↓⇨ ⇔ ⇨
4Q214a	28	↓↓↓↓↓↓↓↓↓↓⇨ ⇨
1Q23	3	↓↓↓↓↓×↓⇨ ⇨
4Q203	14	↓↓↓↓↓⇨ ⇔ ⇨
4Q535	40	↓↓↓↓↓×↓□ ⇨
4Q550a	56	↓↓↓↓↓⇨ □↓⇨ ⇨
4Q202	13	↓⇨ ⇔ ⇨ □↓⇨
4Q553	59	↓⇨↓⇨ ⇔ ⇨ ⇨ ⇨
4Q205	16	↓⇨ ⇔ ⇨ ⇨ ⇨ ⇨
4Q201	12	↓×↓⇨↓⇨ ⇨ ⇨ ⇨ ⇨
4Q530	36	↓⇨ ⇨ ⇨ ⇨ ⇨ ⇨ ⇨
1Q20	1	↓⇨ ⇨ ⇨ ⇨ ⇨ ⇨ ⇨
4Q531	37	↓□ ⇨ ⇨ ⇨ ⇨ ⇨ ⇨
4Q542	47	↓⇨↓□ □↓⇨ □↓⇨ ⇨
4Q204	15	↓□ ⇨ ⇨ ⇨ ⇨
4Q558	63	↓⇨ ⇨ ⇨ ⇨ ⇨
4Q552	58	↓↓↓⇨ ⇨ ⇨ ⇨
4Q243	31	↓↓↓↓↓□ ⇨ □↓↓↓↓↓⇨
4Q537	42	↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q244	32	↓↓↓↓↓×↓↓↓□ ⇨ ⇨
4Q550	55	↓↓↓↓↓⇨ ⇨ ⇨ ⇨
2Q24	5	↓⇨ ⇔ ⇨ ⇨
11Q18	72	↓⇨↓⇨ ⇨ ⇨ ⇨
4Q210	21	↓⇨ ⇨ ⇨ ⇨ ⇨
1Q21	2	↓×↓⇨↓⇨ ⇨ ⇨ ⇨
4Q554	60	↓⇨ ⇨ □↓↓↓⇨ ⇨ □↓↓↓↓↓⇨
4Q242	30	↓↓↓⇨ ⇨ ⇨ ⇨ ⇨
4Q560	65	↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q206	17	↓↓↓↓↓×↓↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q207	18	↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q549	54	↓↓↓↓↓×↓↓↓↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q559	64	↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q541	46	↓↓↓↓↓×↓↓↓↓↓↓↓⇨ ⇨ ⇨ ⇨
11Q10	71	↓↓↓↓↓↓↓↓⇨ ⇨ □↓↓↓↓⇨
4Q538	43	↓×↓⇨ ⇨ ⇨ ⇨
6Q8	70	↓⇨ □↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q550c	57	↓↓↓⇨ ⇨ ⇨ ⇨ □↓↓↓↓↓⇨
4Q196	9	↓⇨ ⇨ □↓⇨ ⇨ ⇨
4Q197	10	↓□ ⇨ ⇨ ⇨ ⇨
4Q213	24	↓⇨↓⇨ □↓↓↓⇨ ⇨ ⇨ ⇨
4Q544	49	↓□ ⇨ ⇨ ⇨ ⇨ ⇨
4Q547	52	↓⇨ □↓⇨ ⇨ ⇨ ⇨ ⇨
4Q543	48	↓⇨ ⇨ ⇨ ⇨ □↓⇨ ⇨ ⇨
4Q546	51	↓⇨↓⇨ □↓⇨ ⇨ ⇨ □↓↓↓↓↓⇨
4Q545	50	↓⇨ ⇨ ⇨ ⇨ ⇨ ⇨ ⇨
4Q570	68	↓↓↓↓↓⇨ ⇨ ⇨ ⇨ ⇨
4Q561	66	↓↓↓↓↓↓↓↓↓↓↓↓⇨ ⇨ ⇨ ⇨
4Q532	38	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓⇨ ⇨ ⇨

corollary of this distinctiveness, Targum Onkelos is dissimilar to 11Q10, the Qumran Aramaic Targum of Job, in terms of the *a priori* textual criteria. Sixthly, aside from the Biblical group, a further five Qumran scrolls appear to be relatively closely related to Targum Onkelos, more so than the Biblical group, namely 1Q24 (Enoch Giants), 4Q214 (Aramaic Levi), 4Q214b (Aramaic Levi), 4Q318 (Brontologion), and 4Q548 (Vision of Amram). None of these scrolls are core members of the identified Qumran Aramaic text types. Notably, 4Q214 and 4Q214b are not core members of the Enochic – Aramaic Levi group neither is 4Q548 a core member of the Tobit – Visions of Amram group.

Figure 8.2 shows the hierarchical clustering dendrogram for the 74 Qumran scrolls and Targum Onkelos books using standardised rates and excluding total relative particle counts and total Peal infinitive counts as previously. The agglomeration is more gradual suggesting that this is a less robust description of relationships between texts. Again the Targum Onkelos books cluster reassuringly together but some of the text type groups identified by internal comparison are a little more fragmented. Using standardised rates Targum Onkelos is no longer close to the Biblical group. The closest Qumran Aramaic scroll is 11Q10, the Targum of Job.

In considering the differences between the raw and standardised rates clusters, the balance between using all the empirical data to weight the criteria and artificially making all the criteria equal has to be addressed. The individual books of the Targum Onkelos are much longer than any of the Qumran Aramaic scrolls, so the chances of missing or underestimating the incidence of textual features is far less than for the Qumran corpus.

For example, the absence of *l*-prefixed imperfect verbal forms is unlikely to be because the text was too short to expect one to occur. Using raw rates privileges those textual features that occur more frequently, but on the other hand there is no *a priori* reason to assume that all the textual criteria are of equal importance. Since those features that occur more often allow more precise estimation of associations between texts, there is an argument to prefer raw rates if there is a discrepancy between raw and standardised rates results. Certainly in this case the raw rates clustering leaves the Qumran text types more intact than the standardised rates. Mann-Whitney tests of the differences between the Biblical group (4Q112, 4Q113, 4Q115, 4Q117) and the Targum Onkelos books for the 22 variables used in clustering shows no significant difference for only nine of the criteria with 13 being significantly different (rates for *afel*, *hafel*, *dy*, *yt*, *m*-prefixed peal infinitives, all peal infinitives, *dn'* or *dnh*, *'lh*, *'ln*, 3 feminine plural verbs, *l*-prefixed imperfect verbs, *itpeel*, and retained alephs). Notably common forms such as determined nouns, total relative particles and *y*-prefixed imperfections show no significant differences and in the raw rates clustering the effects of these outweigh those of the less common features that are significantly different. Considering 11Q10, eleven of its criteria rates fall outside the range of rates of the five Targum Onkelos books. Thus it clusters closer to the Targum Onkelos group when standardised rates are used. However, it cannot be reckoned that close to this external group on these textual criteria.

4Q208	19	↓*↓↘
4Q209	20	↓↘↔
4Q213b	26	↓↘↔
4Q543	48	↓↘↓□
4Q211	22	↓↘↘↓↘
2Q24	5	↓*↓□↔
4Q210	21	↓↘↔↔
1Q24	4	↓*↓□↔
4Q554	60	↓↘↔↔
4Q552	58	↓↘↓↘↔
4Q553	59	↓↘↔
4Q558	63	↓↘↓↘↘↔
4Q196	9	↓□↔↔
4Q213	24	↓□↘□
4Q197	10	↓↘↔↔↘↓↘
4Q544	49	↓↘↓↘↔↔
4Q546	51	↓↘↓↘↘↔↔
4Q547	52	↓↘↓↘↓↘□↔
4Q538	43	↓*↓↘↓↘↔
6Q8	70	↓↘↔↔
4Q530	36	↓↘↓↘*↓↘↔
4Q550c	57	↓↘↓↘↘↔
4Q204	15	↓*↓↘↔
4Q205	16	↓↘↘↓↘↘↓↘
4Q531	37	↓↘↓↘↘↓↘□↔
4Q198	11	↓↘↓↘↓↘↘↘↔↔
1Q21	2	↓*↓↘↔↔
11Q18	72	↓↘↘↓↘↓↘↓↘↓↘
4Q318	34	↓↘↓↘↘↔↔↔
4Q529	35	↓↘↓↘↓↘↓↘*↓↘↔↔
4Q550	55	↓↘↓↘↓↘↘↔↔
4Q560	65	↓↘↓↘↓↘↓↘↓↘↓↘↔
4Q545	50	↓↘↓↘↓↘↓↘↓↘↓↘↓↘
4Q201	12	↓↘↓↘↓↘↓↘↓↘*↓↘↔
4Q542	47	↓↘↓↘↓↘↓↘↓↘↘↓↘↘
4Q202	13	↓↘↓↘↓↘↓↘↓↘↓↘↘↔↔
4Q212	23	↓↘↓↘↓↘↓↘*↓↘↓↘↘↔↔
4Q550a	56	↓↘↓↘↓↘↘↘↘↘↔
4Q156	8	↓↘↓↘↓↘↘↘↘↘↔↘↘
4Q246	33	↓↘↓↘↓↘↓↘↓↘↓↘↘↘↔↔
4Q540	45	↓↘↓↘↓↘↘↘↘↔↔
4Q213a	25	↓↘↓↘*↓↘↘↘↔↔
4Q539	44	↓↘↓↘↘↘↓↘↓↘↓↘↓↘↘↘↘↘↘↔
1Q20	1	↓↘↓↘↓↘↘↘↘↔
11Q10	71	↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↘↘
Gen_O	75	↓↘↔↔
Num_O	78	↓↘↔↔
Exo_O	76	↓↘↓↘↘↘↘↘↘↔↔
Deut_O	79	↓↘↘↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↘↘↘↘↘↘↘↔
Lev_O	77	↓↘↓↘↘↘↘↘↘↔
4Q214	27	↓↘↓↘↓↘↓↘*↓↘↓↘↓↘↓↘↓↘↘↘↘↘↘↘↘↔

Ultimately, statistical procedures are generally a stimulus to develop explanations for the findings. Here we might consider that the raw rates suggest an affinity of Targum Onkelos for the Biblical text type whilst the standardised rates indicate that there may also be an affinity for the Qumran Aramaic Targum of Job. Targum Onkelos relates to both: to the Biblical group in terms of common textual features and to 11Q10 on less common textual features, but is distinct in at least half of the features from both. What is notable, however, is that these non-Qumran texts are stylistically closer to a small group of the Aramaic texts found at Qumran than other Qumran Aramaic texts are to them. That is, some Qumran Aramaic texts resemble Targumic texts more closely than they resemble other Qumran texts in terms of the 20 textual criteria.

Targum Jonathan

Targum Jonathan is often considered together with Targum Onkelos. Table 8.3 shows the parts of speech composition for each of its books. Table 8.4 shows the mean, median and range of rates for the 22 *a priori* criteria in Targum Jonathan. Formal hierarchical clustering (Figure 8.3) using raw rates shows that all the Targum Jonathan books cluster together. Again this is reassuring: the fact that none of the 21 Targum Jonathan books cluster more closely with any of the Qumran Aramaic scrolls than to each other indicates that the cluster algorithm works well with these *a priori* criteria to identify discrete text types. Unlike Targum Onkelos, the Targum Jonathan books are not closely related to the

Qumran Biblical group but to a loose cluster comprising 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548, the five non-biblical scrolls that Targum Onkelos was also most closely

Scroll	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
Joshua	5286	1984	1201	570	7833	136	17010
Judges	4618	2748	1394	634	7703	223	17320
1Samuel	5991	3697	1872	891	10331	316	23098
2Samuel	5346	2892	1519	632	8292	238	18919
1Kings	6521	3130	1665	828	9865	254	22263
2Kings	6124	3119	1447	641	9319	212	20862
Isaiah	9215	6012	3333	1247	14305	459	34571
Jeremiah	10895	6184	3388	1044	16907	532	38950
Ezekiel	9441	4942	3412	1348	14591	519	34253
Hosea	1532	970	643	83	2314	83	5625
Joel	554	278	142	50	740	17	1781
Amos	1039	622	274	132	1582	41	3690
Obadiah	160	83	52	16	222	6	539
Jonah	290	208	106	43	562	27	1236
Micah	818	465	269	80	1213	31	2876
Nahum	359	211	135	34	459	16	1214
Habakkuk	417	269	167	56	669	31	1609
Zephaniah	429	229	119	44	594	5	1420
Haggai	300	133	51	58	448	17	1007
Zechariah	1620	961	464	243	2552	82	5922
Malachi	396	277	165	56	700	60	1644

Table 8.3 Parts of speech count and total words for the Targum Jonathan by book.

related to on raw rates clustering. Figure 8.4 shows the dendrogram using standardised rates of the *a priori* criteria excluding total relative particles and total Peal infinitive verbs. Again all the Targum Jonathan books cluster together quite distantly from the Biblical group. In the standardised rates clustering the Targum Jonathan books relate to several Qumran scroll clusters but notably not 11Q10, the Qumran Aramaic Targum of

Job. Text types that are included in the Qumran Aramaic scrolls group clustering closely to Targum Jonathan on standardised rates include the Enochic – Four Kingdoms

Criterion	Mean	Median	Minimum, Maximum
Determined nouns	125	125	109, 159
Afel	17.5	17.2	10.4, 29.7
Hafel	0.2	0.06	0, 1.9
Itpeel	10.1	9.5	3.2, 19.9
Hitpeel	0	0	0, 0
<i>m</i> -prefixed Peal infinitives	8.5	8.5	4.9, 11.9
Peal infinitives	8.5	8.5	4.9, 11.9
<i>l</i> prefixed imperfects			
<i>y</i> prefixed imperfects	22.5	23.3	5.5, 39.5
3fp verbs	1.5	1.2	0, 4.2
Perfect 3mp ending <i>w</i>	15.7	15.4	5.7, 33.1
Perfect 3mp ending <i>w'</i>	0	0	0,0
Retained aleph before suffix	0.1	0.06	0, 0.7
2ms pronominal suffix <i>kh</i>	0	0	0, 0
3fs pronominal suffix <i>h'</i>	1.8	1.6	0, 6.6
Relative particles	41.4	41.9	31.4, 57.5
Relative particle <i>dy</i>	0.2	0	0, 2.0
Direct object marker <i>yt</i>	15.1	14.8	3.7, 31.3
<i>dnh</i> or <i>dn'</i> forms	0.008	0	0, 0.06
Demonstrative pronoun <i>'lh</i>	0	0	0, 0
Demonstrative pronoun <i>'lyn</i>	0.6	0.4	0, 3.2
Demonstrative pronoun <i>'ln</i>	0	0	0, 0

Table 8.4 Mean, median and range of rates per 1,000 words for the 22 criteria occurring in Targum Jonathan.

1Sa_J	80	↓☒
2Sa_J	81	↓□
Jdg_J	79	↓↑↓☒
Jos_J	78	↓□ ↔
2Ki_J	83	↓☒ □↓☒
Jer_J	85	↓✕↓□ ↔
Mal_J	98	↓☒ ↔ ↔
1Ki_J	82	↓☒ ↔ ↔
Hag_J	96	↓□ ↔ ↔
Eze_J	86	↓↑↓☒ ↔
Jon_J	91	↓☒ □↓↓↓↓↓↓↓↓☒
Oba_J	90	↓☒ ↔ ↔
Zep_J	95	↓□ ↔ ↔
Hos_J	87	↓□ ↔ ↔
Amo_J	89	↓↑↓↓↓↓□ ↔
Zec_J	97	↓□ ↔ ↔
Isa_J	84	↓□ ↔ ↔
Mic_J	92	↓☒ ↔ □↓↓↓↓↓↓↓☒
Nah_J	93	↓✕↓☒ ↔ ↔ ↔
Hab_J	94	↓☒ □↓☒ ↔ ↔
Joe_J	88	↓↓↓☒ ↔ ↔
1Q24	4	↓↓↓✕↓☒ ↔ ↔
4Q214	25	↓↓↓↓☒ □↓↓↓☒ ↔ ↔
4Q318	32	↓↓↓↓↓↓☒ □↓☒ ↔ □↓↓↓↓↓↓↓↓☒
4Q548	51	↓↓↓↓↓↓↓↓↓↓☒ □↓☒ ↔ ↔
4Q214b	27	↓↓↓↓↓↓↓↓↓↓☒ ↔ ↔
4Q112	71	↓✕↓☒ ↔ ↔ ↔
4Q115	99	↓☒ ↔ ↔ □↓↓↓↓☒
4Q113	72	↓↓↓□ ↔ ↔ ↔
4Q117	100	↓↓↓↑↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓☒ ↔ ↔ ↔
4Q529	33	↓↓↓☒ ↔ ↔
4Q557	60	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓☒ ↔
4Q206	15	↓↓↓↓↓↓✕↓↓↓↓↓↓↓☒ ↔ ↔
4Q207	16	↓↓↓↓↓☒ ↔ ↔
4Q208	17	↓✕↓☒ ↔ ↔
4Q209	18	↓☒ ↔ ↔ ↔
4Q211	20	↓↓↓↓↑↓☒ ↔ ↔
5Q15	67	↓↓↓↓☒ □↓↓↓☒ ↔ ↔
4Q213a	23	↓↓↓☒ ↔ ↔
4Q213b	24	↓↓↓↑↓☒ □↓☒ □↓↓↓↓↓↓☒
4Q539	42	↓↓↓☒ ↔ ↔ ↔
4Q214a	26	↓↓↓↓↓↓↓↓↓↓☒ ↔ ↔ ↔
1Q23	3	↓↓↓↓↓↓✕↓☒ ↔ ↔ ↔
4Q203	12	↓↓↓↓↓☒ ↔ ↔ ↔
4Q535	38	↓↓↓↓↓↓✕↓□ ↔ ↔ ↔
4Q550a	54	↓↓↓↓↓☒ □↓☒ ↔ ↔ ↔
4Q202	11	↓☒ ↔ ↔ □↓☒ ↔ ↔
4Q553	57	↓↑↓☒ ↔ ↔ ↔
4Q205	14	↓☒ ↔ ↔ ↔
4Q201	10	↓✕↓↑↓↓↓↓□ ↔ ↔ ↔
4Q530	34	↓☒ ↔ ↔ ↔ ↔
1Q20	1	↓☒ ↔ ↔ ↔ ↔
4Q531	35	↓↑↓↓□ ↔ ↔ ↔
4Q542	45	↓☒ ↔ ↔ □↓☒ ↔ ↔

4Q204	13	↓x↓□	⇔	⇔	⇔	⇔	⇔
4Q558	61	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q552	56	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q243	29	↓⇔⇔⇔x↓⇔	⇔	⇔	⇔	⇔	⇔
4Q537	40	↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q244	30	↓⇔⇔⇔x↓⇔	⇔	⇔	⇔	⇔	⇔
4Q550	53	↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
2Q24	5	↓⇔	⇔	⇔	⇔	⇔	⇔
11Q18	70	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q210	19	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
1Q21	2	↓x↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q554	58	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q242	28	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q560	63	↓⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q549	52	↓⇔⇔⇔x↓⇔⇔⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q559	62	↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q541	44	↓⇔⇔⇔⇔x↓⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
11Q10	69	↓⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q538	41	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
6Q8	68	↓⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q550c	55	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q196	7	↓⇔	⇔	⇔	⇔	⇔	⇔
4Q197	8	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q213	22	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q544	47	↓x↓□	⇔	⇔	⇔	⇔	⇔
4Q547	50	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q543	46	↓x↓□	⇔	⇔	⇔	⇔	⇔
4Q546	49	↓⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q545	48	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q570	66	↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q561	64	↓⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q532	36	↓⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q156	6	↓⇔⇔x↓⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q246	31	↓⇔⇔⇔	⇔	⇔	⇔	⇔	⇔
4Q536	39	↓⇔⇔⇔⇔⇔	⇔	⇔	⇔	⇔	⇔

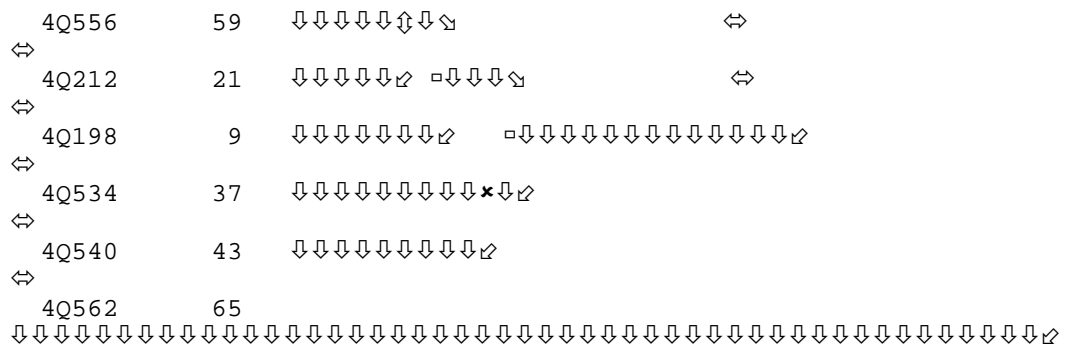


Figure 8.3 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the 21 books of Targum Jonathans (suffixed _J) using raw rates of the 22 *a priori* criteria.

1Sa_J	75	↓↘
2Sa_J	76	↓□
1Ki_J	77	↓□
2Ki_J	78	↓□
Hag_J	91	↓↕↓↘
Jer_J	80	↓□ ↔
Eze_J	81	↓□ ↔
Mal_J	93	↓□ ↔
Amo_J	84	↓□ ↔
Zec_J	92	↓↘ ↔
Jon_J	86	↓↘↓↕↓↘
Isa_J	79	↓✕↓□ ↔
Hab_J	89	↓↘ ↔ ↔
Hos_J	82	↓↘ ↔ □↓↘
Joe_J	83	↓↕↓□ ↔ ↔
Oba_J	85	↓↘ ↔ ↔ ↔
Zep_J	90	↓↘↓↘↘ ↔ □↓↘↘↘
Jos_J	73	↓✕↓↘↘↘ ↔ ↔ ↔
Jdg_J	74	↓↘ ↔ ↔ ↔
Mic_J	87	↓✕↓↘↘↘↘↘↘ ↔ ↔
Nah_J	88	↓↘ ↔ ↔
4Q204	13	↓✕↓↘ ↔ ↔
4Q205	14	↓↘ □↓↘↘↘ ↔ ↔
4Q531	35	↓↘↓↘↘ ↔ □↓↘ ↔ ↔
4Q206	15	↓↘↓↘↘↘↘↘↘↘ ↔ ↔ ↔
1Q21	2	↓✕↓↘ ↔ ↔ ↔
11Q18	70	↓↘ □↓↘ ↔ ↔ ↔
4Q318	32	↓↘↓↘↘ ↔ □↓↘ ↔ ↔
4Q550	53	↓↘↓↘✕↓□ ↔ ↔ ↔
4Q552	56	↓↘↓↘↘ ↔ ↔ ↔ ↔
4Q529	33	↓↘↓↘↘↘↘ ↔ ↔ ↔ ↔
1Q24	4	↓✕↓↘ ↔ ↔ ↔
4Q554	58	↓↘ ↔ ↔ ↔ ↔
4Q208	17	↓✕↓↕↓↘ ↔ ↔ ↔
4Q209	18	↓↘ ↔ ↔ ↔ ↔ ↔
2Q24	5	↓↘ ↔ ↔ □↓□ ↔
4Q210	19	↓↕↓□ ↔ ↔ ↔ ↔
4Q211	20	↓□ ↔ ↔ ↔ ↔ ↔
4Q213b	24	↓↘ ↔ ↔ ↔ ↔ ↔
4Q543	46	↓↘↓↘↘ ↔ ↔ ↔ ↔
4Q553	57	↓↘ ↔ ↔ ↔ ↔
4Q558	61	↓↕↓↘↘ ↔ ↔ □↓↕↓↘↘
4Q196	7	↓□ ↔ ↔ ↔ ↔ ↔ ↔
4Q213	22	↓□ □↓□ ↔ ↔ ↔ ↔
4Q197	8	↓↘ ↔ □↓↘ ↔ ↔ ↔
4Q544	47	↓↘↓↘↘ ↔ ↔ ↔ ↔
4Q546	49	↓↘↓↘↘ ↔ ↔ ↔ ↔
4Q547	50	↓↘↓↘↓↘↘ ↔ ↔ ↔
4Q538	41	↓✕↓↘↓↘↘ ↔ ↔ ↔
6Q8	68	↓↘ ↔ ↔ ↔ ↔
4Q530	34	↓↘↓↘✕↓□ ↔ ↔ ↔
4Q550c	55	↓↘↓↘↘ ↔ ↔ ↔ ↔
4Q198	9	↓↘↓↘↓↘↘ ↔ ↔ ↔
4Q560	63	↓↘↓↘↓↘↓↘↓↘↘ ↔ ↔ ↔
4Q213a	23	↓↘↓↘✕↓↘ ↔ ↔ ↔
4Q539	42	↓↘↓↘↘ ↔ □↓↘↓↘↘↘↘ ↔

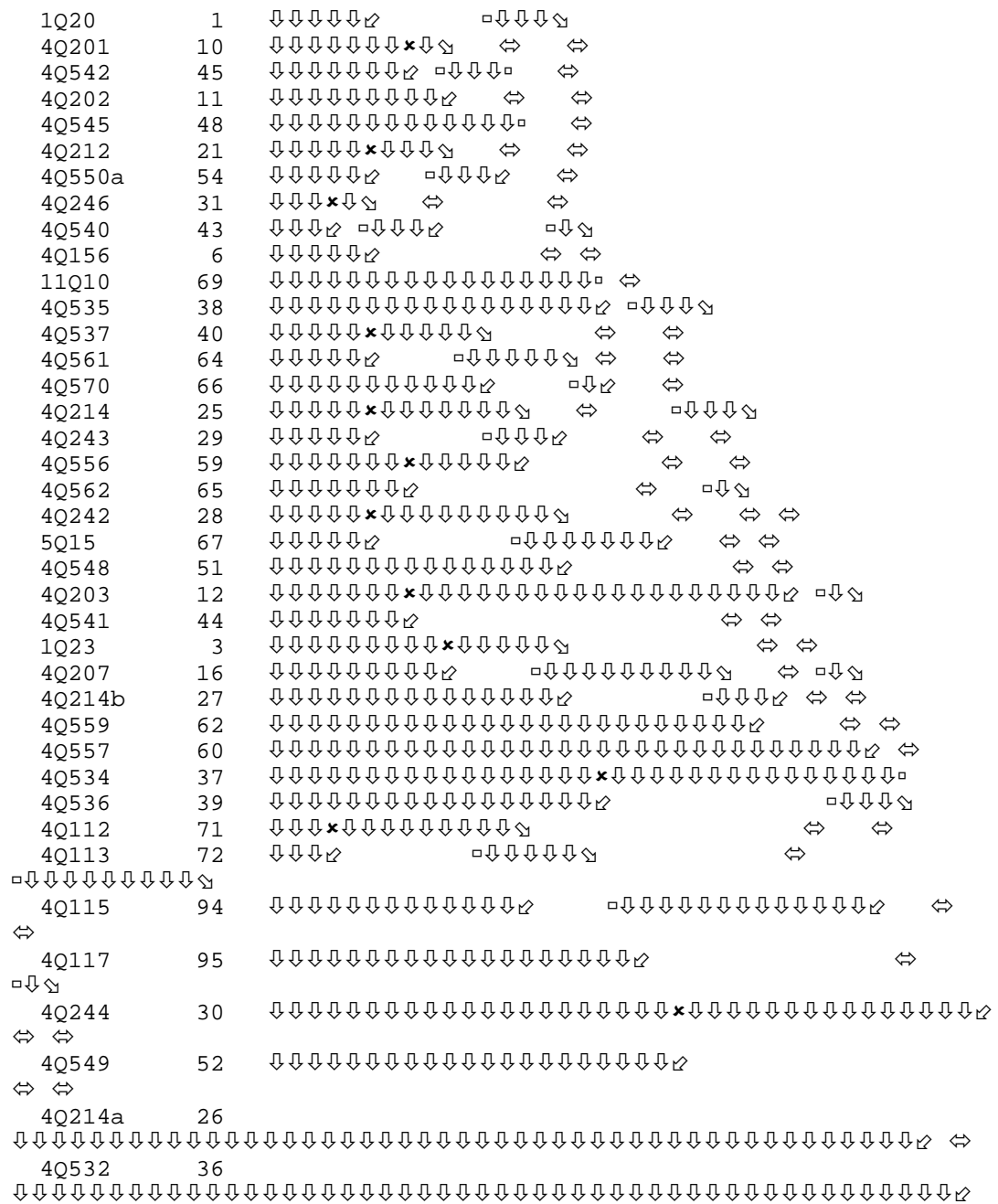


Figure 8.4 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the 21 books of Targum Jonathans (suffixed _J) using standardised rates of 20 of the 22 *a priori* criteria.

group, the New Jerusalem group, the Enochic group and the Tobit – Visions of Amram group. In short, unlike Targum Onkelos, Targum Jonathan is closer to mainstream Qumran Aramaic texts and less closely related to either Qumran Biblical Aramaic scrolls or the Qumran Aramaic Targum of Job. Interestingly, the 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 group is related to both Targum Onkelos and Targum Jonathan, at least in terms of raw rates criteria.

To examine how similar the 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 group is to Targum Jonathan, differences in raw rates between this group and the Targum Jonathan books were tested using Mann-Whitney U-tests. Significant two-tailed differences were found using exact iterative estimates on rates for relative particles ($p < .001$), *dy* form of relative particle ($p = .004$), direct object marker *yt* ($p < .001$), *m*-prefixed Peal infinitives ($p < .001$), total Peal infinitives ($p < .001$), third person feminine plural verbs ($p = .001$), *l*-prefixed imperfects ($p = .031$), *itpeels* ($p = .015$), *'ln* demonstrative pronoun plural form ($p = .031$), and *-w* ending of third person masculine plural perfect forms ($p < .001$). In short, significant differences in rates were found for ten of the criteria, but not for the other twelve. The failure to find significant differences may reflect type 2 statistical error given the small sample size.

To provide a contrast to this number of differences, Targum Onkelos was compared with Targum Jonathan on the same 22 criteria. Of note is that these criteria were mooted for non-Targumic texts and that other criteria are likely to have been selected if a classification of Targums was being pursued. However, these Targums are

conventionally considered to be relatively close textually. Significant differences between Targum Onkelos and Targum Jonathan were found for rates of *yt* ($p=.001$, Targum Onkelos higher), *'lyn* ($p=.022$, Targum Onkelos higher) and *itpeels* ($p=.034$) only. One significant difference at $p<.05$ between the Targums would have been expected by chance. The dissimilarity between the 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 group and Targum Jonathan is greater in terms of the number of criteria that are significantly different than between Targum Onkelos and Targum Jonathan, but is similar to that between Qumran text type groups (Table 7.4).

At this stage we can also examine how close the 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 group is to Targum Onkelos. Again using the Mann-Whitney U-test significant differences in rates were found for seven of the 22 *a priori* criteria: relative particles ($p=.008$, Targum Onkelos higher), *dy* ($p=.048$, Targum Onkelos lower), *yt* ($p=.008$, Targum Onkelos higher), *m*-prefixed Peal infinitives ($p=.016$, Targum Onkelos higher), total Peal infinitives ($p=.016$, Targum Onkelos higher), third person feminine plural verbs ($p=.008$, Targum Onkelos higher) and *-w* ending of third person masculine plural perfect forms ($p=.024$, Targum Onkelos higher). In terms of the number of criteria that are significantly different, the 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 group is closer to Targum Onkelos than to Targum Jonathan. Moreover, since total relative particles and *dy* are highly correlated in the Qumran scrolls as are *m*-prefixed Peal infinitives and total Peal infinitives, we would expect to find a significant difference for one of the pair if there is a significant difference for the other. This principle underpinned the exclusion of total relative particles and total Peal infinitives from the standardised rates clustering.

Hence, actual textual differences may apply to only five textual aspects. This proximity suggests that the so-called Babylonian Targums have a degree of communality greater than that found within the Qumran Aramaic corpus itself. That is to say Targum Onkelos, and to a lesser extent Targum Jonathan, are textually closer to 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 than some of the Aramaic scrolls found at Qumran. In view of this we might tentatively term 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 as an ‘Onkeline’ group. There is debate about the dating of other Targums, which might be later than either Onkelos or Jonathan, but comparison of the Qumran Aramaic scrolls with these might illuminate relationships further.

Targum Neofiti

Book	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
Genesis	11022	6308	4492	1704	18442	789	42757
Exodus	8947	4216	2711	1387	14290	439	31990
Leviticus	6025	2761	1978	732	10013	364	21873
Numbers	10320	3858	2707	1774	13878	363	32900
Deuteronomy	7631	4145	3406	1132	13033	463	29810

Table 8.5 Parts of speech count and total words for the Targum Neofiti by book.

The books of Targum Neofiti contain more words than the equivalent books of Targum Onkelos (Table 8.5). There is a wide range of expansion: Genesis has 23.5% more words, Exodus 13.1% more, Leviticus 8.3%, Numbers 20.8% and Deuteronomy 19.9%.

Criterion	Genesis	Exodus	Leviticus	Numbers	Deuteronomy
Determined nouns	121	130	145	131	130
Afel	12.8	11.1	12.9	7.6	12.5
Hafel	0.1	0.3	0	0.1	0.2
Itpeel	6.9	5.8	7.1	4.4	6.2
Hitpeel	0	0	0	0	0
<i>m</i> -prefixed Peal infinitives	7.7	6.1	4.5	5.6	10.1
Peal infinitives	7.7	6.1	4.5	5.6	10.2
<i>l</i> prefixed imperfects	0	0	0	0	0
<i>y</i> prefixed imperfects	9	19	49	20	26
3fp verbs	1.1	1.0	0.3	0.6	0.9
Perfect 3mp ending <i>w</i>	8.3	6.7	2.2	8.4	3.1
Perfect 3mp ending <i>w'</i>	0	0	0	0	0
Retained aleph before suffix	0.1	0.3	0	0	0
2ms pronominal suffix <i>kh</i>	0	0	0	0	0
3fs pronominal suffix <i>h'</i>	0	0	0	0	0
Relative particles	48	47	55	61	56
Relative particle <i>dy</i>	6.6	7.8	13.5	10.2	18.6
Direct object marker <i>yt</i>	31	39	41	26	37
<i>dnh</i> or <i>dn'</i> forms	0.1	0	0	0.2	0.1
Demonstrative pronoun ' <i>lh</i>	0	0	0	0	0
Demonstrative pronoun ' <i>lyn</i>	1.3	0.1	0.6	1.6	0.3
Demonstrative pronoun ' <i>ln</i>	0	0	0	0	0

Table 8.6 Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Neofiti .

4Q208 17 ↕ ↕
4Q209 18 ↕ ↕ ↕ ↕

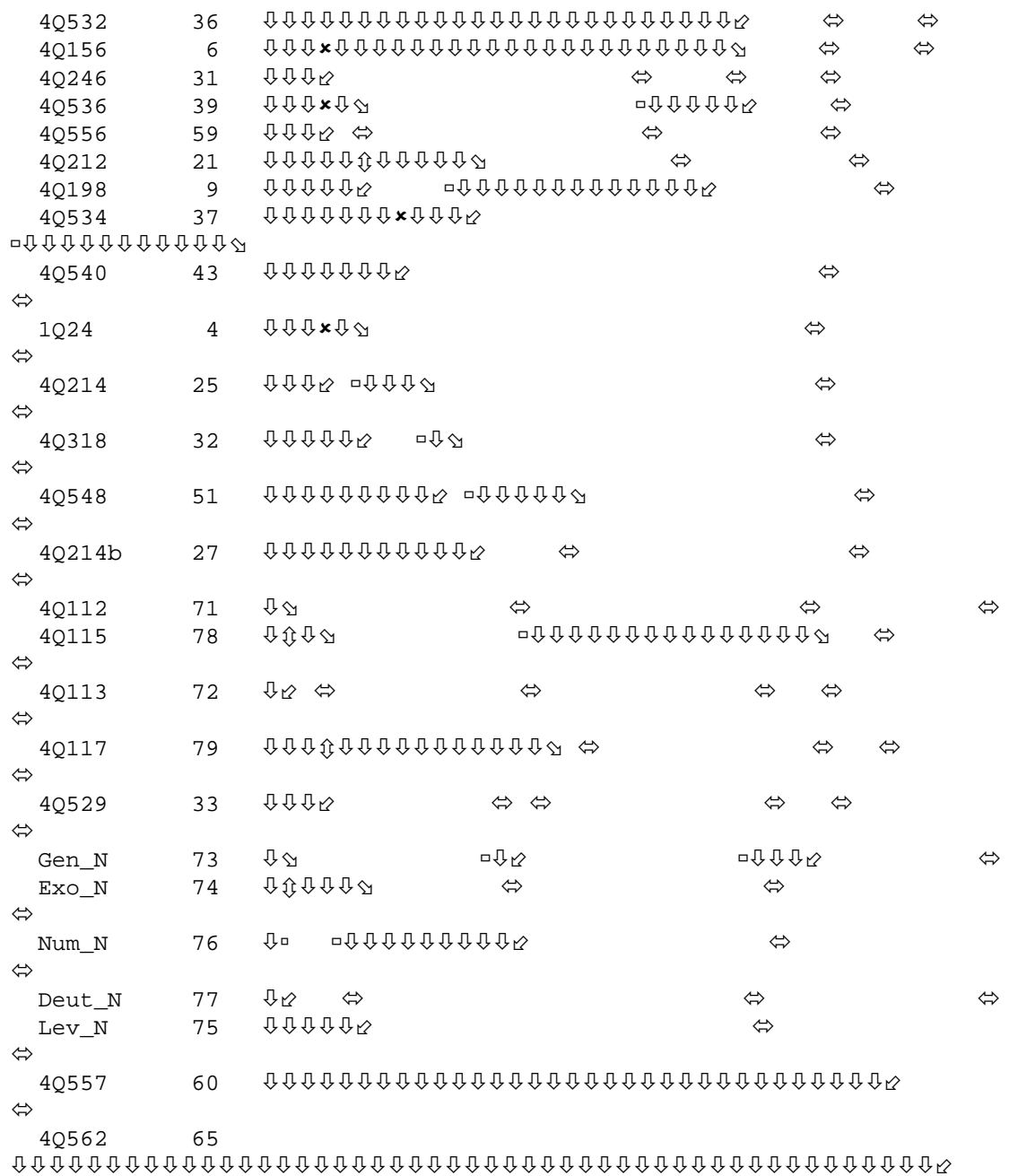


Figure 8.5 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the five books of Targum Neofiti (Gen_N, Exo_N, Lev_N, Num_N, Deut_N) using raw rates of the 22 *a priori* criteria.

The rates of the 22 *a priori* criteria are shown by book in Table 8.6. One clear difference with Targum Onkelos is the prevalence of *dy* relative particle forms. Figure 8.5 shows the hierarchical cluster dendrogram for Targum Neofiti books together with the Qumran Aramaic scrolls using raw rates of the 22 *a priori* criteria. Just as with the other Targums, Targum Neofiti books cluster together. Targum Neofiti books cluster in a similar way to Targum Onkelos being most closely related to the Biblical group of Qumran Aramaic scrolls, but then to 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548, which was previously termed the ‘Onkeline’ group. Given the extra length of Targum Neofiti compared with Targum Onkelos, this relationship with Targumic material persists even when text that is not directly related to a Hebrew *Vorlage* is included in the analysis. Comparing Targum Neofiti with the Qumran Aramaic Biblical group, significant differences are present for rates of Afels ($p=.032$), Hafels ($p=.016$), *dy* ($p=.016$), *yt* ($p=.016$), *m*-prefixed Peal infinitives ($p=.016$), total Peal infinitives ($p=.016$), *dnh* or *dn*’ ($p=.016$), Itpeels ($p=.016$) and Hitpeels ($p=.008$), that is nine of the 22 *a priori* criteria. This compares to the 13 criteria for which there were significant differences between the Biblical group and Targum Onkelos. Comparing Targum Neofiti with the Onkeline group, only four of the 22 *a priori* criteria show significant differences: total relative particles ($p=.008$), *yt* ($p=.008$), third person feminine verbs ($p=.008$) and *-w* ending third person masculine plural verb forms ($p=.024$). In view of this it is difficult to claim that 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548 are specifically Onkeline, but appear to represent a broader ‘Targumic’ subclassification of the Qumran Aramaic scrolls.

4Q208	17	↓×↓↗
4Q209	18	↓↗↔
4Q211	20	↓↗↔
4Q213b	24	↓↗↓□
4Q543	46	↓↗ □↓↗
2Q24	5	↓×↓□↔
4Q210	19	↓↗↔↔
1Q24	4	↓×↓□↔
4Q554	58	↓↗↔↔
4Q552	56	↓↗↓↗↔
4Q553	57	↓↗↔
4Q558	61	↓↗↓↗↔
4Q196	7	↓□↔↔
4Q213	22	↓□ □↓□
4Q197	8	↓↗↔ □↓↗
4Q544	47	↓↗↓□↔↔
4Q546	49	↓↗↓↗↔↔
4Q547	50	↓↗↓↗↓□↔
4Q538	41	↓×↓↗↓□↔
6Q8	68	↓↗↔↔
4Q530	34	↓↗↓×↓↗↔
4Q550c	55	↓↗↓↗↔
4Q204	13	↓×↓↗↔
4Q205	14	↓↗ □↓↗ □↓↗
4Q531	35	↓↗↓↗ □↓□↔
4Q198	9	↓↗↓↗↓↗↔↔
1Q21	2	↓×↓↗↔↔
11Q18	70	↓↗ □↓↗↓□ □↓↗
4Q318	32	↓↗↓↗↔↔↔
4Q529	33	↓↗↓↗↓×↓↗↔↔
4Q550	53	↓↗↓↗↓↗↔↔
4Q560	63	↓↗↓↗↓↗↓↗↓↗↓↗↓□↓↗
4Q545	48	↓↗↓↗↓↗↓↗↓↗↓↗↓□↔
4Q201	10	↓↗↓↗↓↗↓×↓↗↔↔
4Q542	45	↓↗↓↗↓↗↓↗↓↗↓↗↓↗↔
4Q202	11	↓↗↓↗↓↗↓↗↓↗↓↗↔
4Q212	21	↓↗↓↗↓×↓↗↓↗↔
4Q550a	54	↓↗↓↗↓↗↓ □↓↗↓↗↓↗↗
4Q535	38	↓↗↓↗↓↗↓↗↓↗↓↗↔↔
4Q213a	23	↓↗↓×↓↗↔↔
4Q539	42	↓↗↓↗↓ □↓↗↓↗↓↗↓↗↔
1Q20	1	↓↗↓↗↓↗↔
Exo_N	74	↓↗↔
Deut_N	77	↓□↔
Gen_N	73	↓↗↓↗↓↗↓↗↓↗↓↗↓↗↓↗↓↗↗
Num_N	76	↓□↔↔
Lev_N	75	↓↗↔↔
4Q156	6	↓↗↓↗↓×↓↗↔↔
4Q246	31	↓↗↓↗↓↗↓ □↓↗↓↗↓↗↔↔

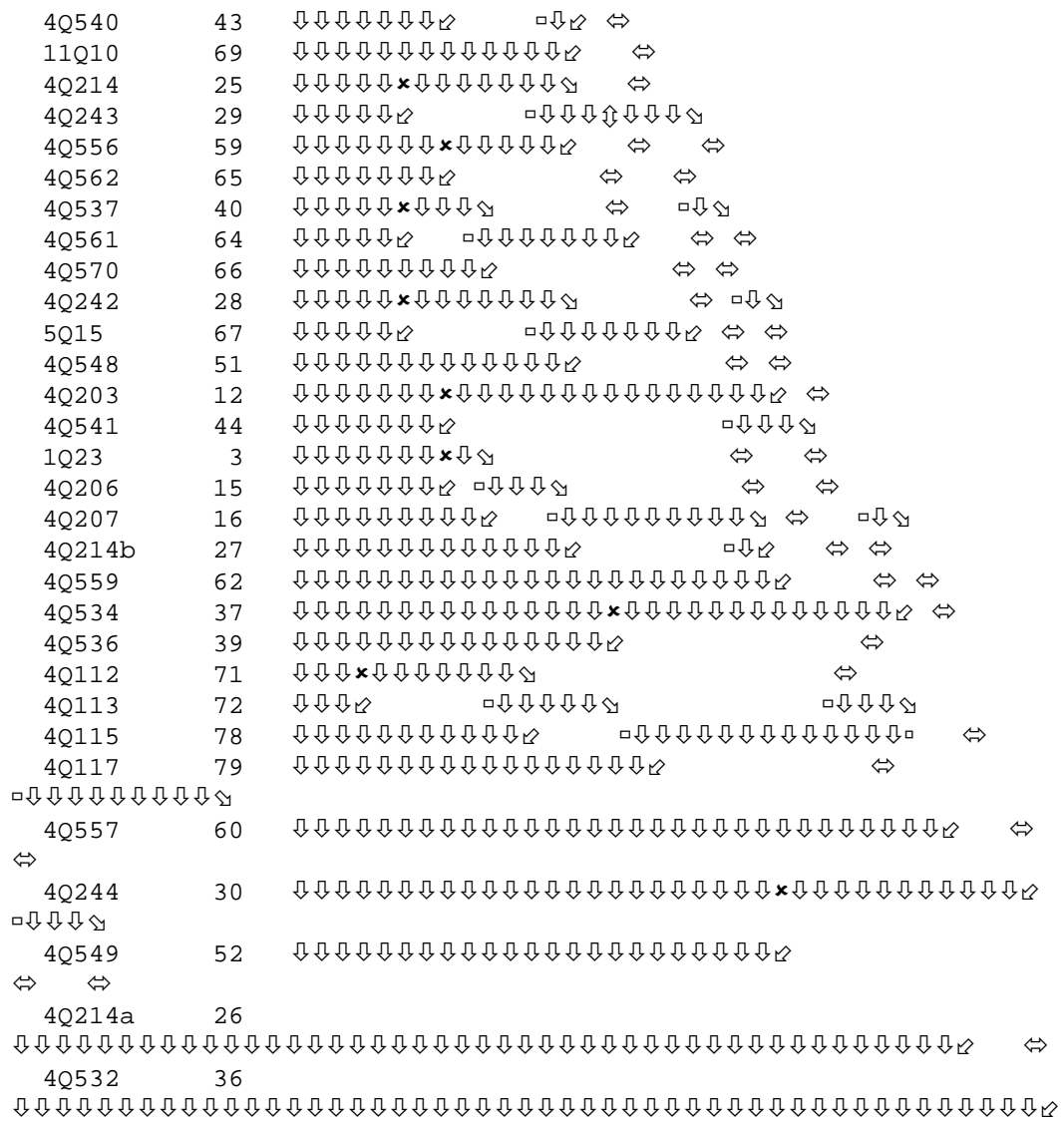


Figure 8.6 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with Targum Neofiti books using standardised rates of 20 *a priori* criteria.

Figure 8.6 shows the hierarchical clustering dendrogram for standardised rates using the 20 selected *a priori* criteria. Again, the agglomeration is far more stepwise than with the raw rates suggesting that raw rates are superior for detecting distinct clusters. Targum Neofiti books cluster tightly together, in a very similar manner to Targum Onkelos on using standardised rates (Figure 8.2). However, Targum Neofiti is less close textually to 11Q10 Qumran Aramaic Targum Job, than Targum Onkelos. The distance between Targum Neofiti and the Qumran ‘Biblical’ group is also considerable using standardised rates, a feature also noted for Targum Onkelos. These findings suggest that similarities between Targum Neofiti and the ‘Targumic’ Qumran Aramaic scrolls rely on a few relatively frequent criteria.

Targum Pseudo-Jonathan

Book	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
Genesis	11998	6944	4370	1882	19596	697	45487
Exodus	10255	4872	2967	1560	16499	445	36598
Leviticus	6768	3104	1989	893	11258	364	24376
Numbers	10873	4402	2540	1802	15445	374	35436
Deuteronomy	8301	4655	3550	1233	14149	400	32288

Table 8.7 Parts of speech count and total words for the Targum Pseudo-Jonathan by book.

Like Targum Neofiti, Targum Pseudo-Jonathan, sometimes known as Targum Yerushalmi, is thought to be Western (Palestinian) in character and also comprises the Pentateuch only. Table 8.7 shows counts of parts of speech count and total words for the Targum Pseudo-Jonathan by book. Targum Pseudo-Jonathan has more words than both Targum Onkelos and Targum Neofiti, though compared with Targum Neofiti not all parts of speech show an increase, notably pronominal suffixes and pronouns (compare Table 8.5). Rates per 1,000 words for the 22 *a priori* criteria are shown in Table 8.8. One possible indicator of the textual character of Targum Pseudo-Jonathan is that the *dy* rates for Targum Pseudo-Jonathan are lower than those of Targum Neofiti, but higher than those of Targum Onkelos. Figure 8.7 shows the hierarchical clustering dendrogram for the 74 Qumran Aramaic scrolls together with the five Targum Pseudo-Jonathan books according to raw rates of the 22 *a priori* criteria. The pattern is essentially the same as that for Targum Onkelos. The five Targum Pseudo-Jonathan books cluster together and are most closely related to the Qumran Biblical Aramaic group and the Qumran Targumic group of scrolls.

Formal statistical testing using the Mann-Whitney U-test reveals significant differences between Targum Psuedo-Jonathan and the Qumran Biblical Aramaic group on rates of *afels* ($p=.016$, Targum Pseudo-Jonathan higher), *hafels* ($p=.016$, Targum Pseudo-Jonathan lower), *dy* ($p=.016$, Targum Pseudo-Jonathan lower), *yt* ($p=.016$, Targum Pseudo-Jonathan higher), *m*-prefixed Peal infinitives ($p=.016$, Targum Pseudo-Jonathan higher), total Peal infinitives ($p=.016$, Targum Pseudo-Jonathan higher), and *hitpeels* ($p=.032$), that is on seven out of the 22 *a priori* criteria,

Criterion	Genesis	Exodus	Leviticus	Numbers	Deuteronomy
Determined nouns	121	137	152	131	135
Afel	13.1	11.5	15.0	9.1	13.9
Hafel	0.4	1.1	0.4	0.4	0.3
Itpeel	8.8	7.6	9.9	6.9	7.9
Hitpeel	0	0	0	0	0
<i>m</i> -prefixed Peal infinitives	9.5	7.7	6.4	7.1	12.1
Peal infinitives	9.5	7.7	6.4	7.1	12.1
<i>l</i> prefixed imperfects	0	0	0	0	0
<i>y</i> prefixed imperfects	9	17	48	21	25
3fp verbs	1.2	0.9	0.4	0.7	0.5
Perfect 3mp ending <i>w</i>	12.4	8.7	2.8	13.0	4.4
Perfect 3mp ending <i>w'</i>	0	0	0	0	0
Retained aleph before suffix	0.1	0.2	0	0.1	0.1
2ms pronominal suffix <i>kh</i>	0	0	0	0	0
3fs pronominal suffix <i>h'</i>	1.6	1.2	1.1	1.4	0.8
Relative particles	45	46	49	49	55
Relative particle <i>dy</i>	1.3	1.2	2.7	1.6	1.7
Direct object marker <i>yt</i>	24	30	31	21	22
<i>dnh</i> or <i>dn'</i> forms	0.1	0	0	0	0
Demonstrative pronoun ' <i>lh</i>	0	0	0	0	0
Demonstrative pronoun ' <i>lyn</i>	0	0	0	0.1	0
Demonstrative pronoun ' <i>ln</i>	0	0	0	0	0

Table 8.8 Rates per 1,000 words for the 22 criteria occurring in the five books of Targum Pseudo-Jonathan.

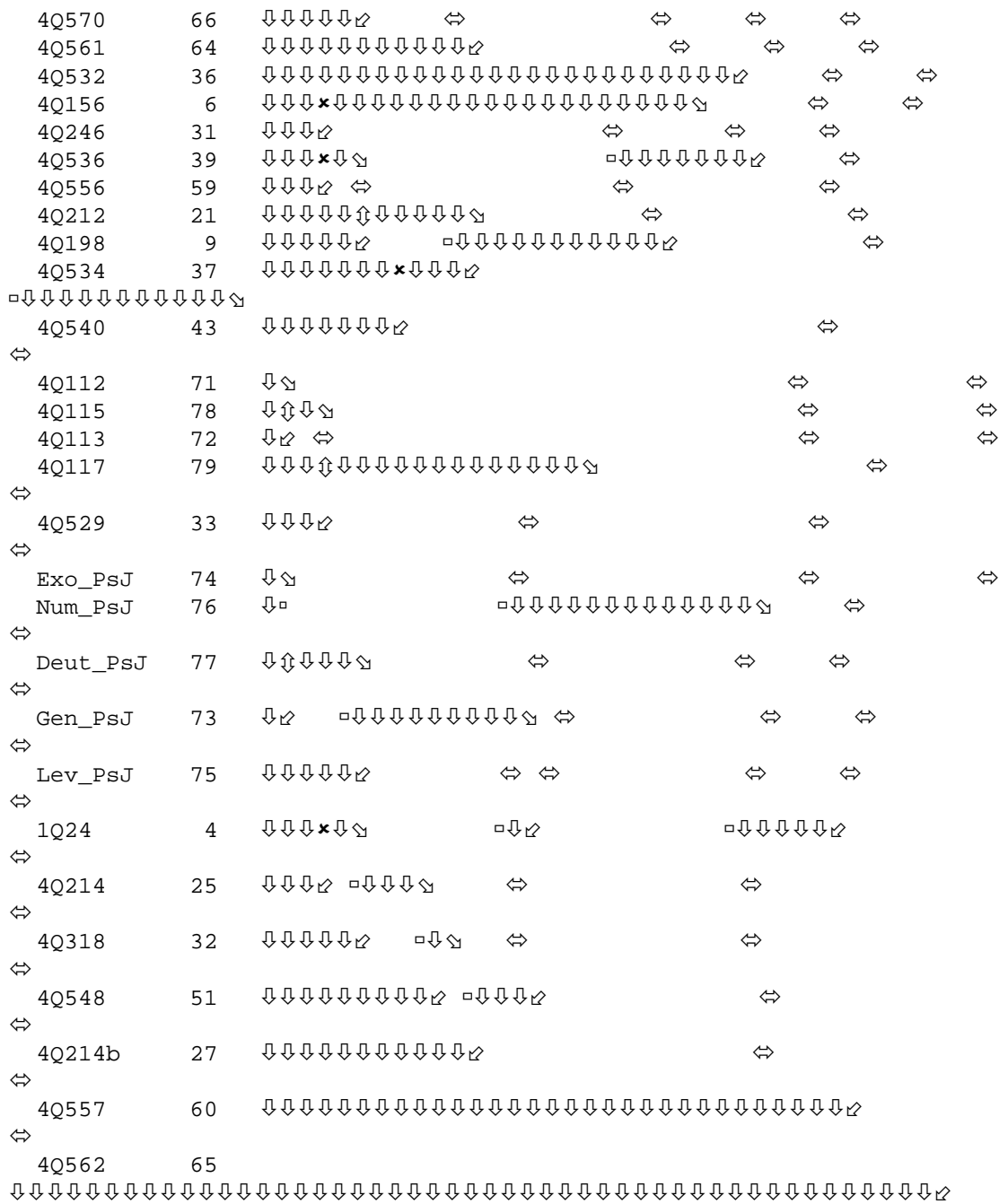


Figure 8.7 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with the five books of Targum Pseudo-Jonathan (Gen_PsJ, Exo_PsJ, Lev_PsJ, Num_PsJ, Deut_PsJ) using raw rates of the 22 *a priori* criteria.

the same number as for Targum Neofiti, but fewer than Targum Onkelos. Comparing Targum Psuedo-Jonathan with the Qumran Targumic group reveals significant differences for relative particles ($p=.008$), *yt* ($p=.008$), *m*-prefixed Peal infinitives ($p=.016$), total Peal infinitives ($p=.016$), third person singular feminine suffix *-h'* ($p=.008$), hitpeels ($p=.008$), third masculine plural perfect ending *-w* ($p=.024$), that is seven criteria compared with four criteria for Targum Neofiti.

Figure 8.8 shows the dendrogram for hierarchical clustering using the standardised rates of the 20 independent *a priori* criteria. The agglomeration is, like those for the other Targumic groups above, markedly incremental step-by-step which suggests that it may not be that useful. Again all Targum Pseudo-Jonathan books cluster together, a little more distant from the Qumran Biblical group than in the raw rates dendrogram, there is no clear relationship with the Qumran Targumic group nor with 11Q10.

4Q208	17	↓*↓↘
4Q209	18	↓↘ □↓↘
4Q213b	24	↓*↓□ ⇔
4Q543	46	↓↘ ⇔ ⇔
4Q211	20	↓↓↓↘ ⇔
2Q24	5	↓*↓↘ ⇔
4Q210	19	↓↘ ⇔ ⇔
1Q24	4	↓*↓□ ⇔
4Q554	58	↓↘ □↓↘↓↘
4Q552	56	↓↓↓□ ⇔ ⇔
4Q115	78	↓↓↓↘ ⇔ ⇔
1Q21	2	↓*↓↘ ⇔ ⇔
11Q18	70	↓↘ □↓↘ ⇔
4Q318	32	↓↓↓↘ ⇔
4Q538	41	↓*↓↓↓↘ ⇔
6Q8	68	↓↘ □↓□
4Q530	34	↓↓↓*↓↘ ⇔
4Q550c	55	↓↓↓↘ ⇔
4Q553	57	↓↘ ⇔
4Q558	61	↓↘↓↘ ⇔
4Q196	7	↓□ ⇔ ⇔
4Q213	22	↓□ □↓↘ ⇔
4Q197	8	↓↘ ⇔ ⇔ □↓↘
4Q544	47	↓↓↓□ ⇔ ⇔ ⇔
4Q546	49	↓↓↓↘ ⇔ ⇔ ⇔
4Q547	50	↓↓↓↓↓↘↓□ ⇔
4Q204	13	↓*↓↘ ⇔ ⇔ ⇔
4Q205	14	↓↘ □↓□ ⇔ □↓↘
4Q531	35	↓↓↓↘ ⇔ ⇔ ⇔ ⇔
4Q198	9	↓↓↓↓↓↘ ⇔ ⇔ ⇔
4Q529	33	↓↓↓↓↓*↓↘ ⇔ ⇔
4Q550	53	↓↓↓↓↓↘ ⇔ ⇔
4Q560	63	↓↓↓↓↓↓↓↓↓↘ ⇔
4Q545	48	↓↓↓↓↓↓↓↓↓↓↓□
4Q201	10	↓↓↓↓↓↓↓*↓↘ ⇔
4Q542	45	↓↓↓↓↓↓↓↘ □↓↘↓↘
4Q202	11	↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔
4Q212	21	↓↓↓↓↓*↓↓↓↘ ⇔ ⇔
4Q550a	54	↓↓↓↓↓↘ □↓↘ ⇔
4Q156	6	↓↓↓↓↓↘ ⇔ □↓↘
4Q246	31	↓↓↓↓↓↘↓↓↓↘ ⇔ ⇔
4Q540	43	↓↓↓↓↓↘ ⇔ ⇔
4Q213a	23	↓↓↓*↓↘ ⇔ ⇔
4Q539	42	↓↓↓↘ □↓↓↓↓↓↓↓↘ ⇔
1Q20	1	↓↓↓↓↓↘ ⇔
4Q214	25	↓↓↓↓↓*↓↓↓↓↓↓↓↓↘ □↓↘
4Q243	29	↓↓↓↓↓↘ □↓□ ⇔
4Q556	59	↓↓↓↓↓*↓↓↓↓↓↓↓↘ ⇔ ⇔
4Q562	65	↓↓↓↓↓↘ ⇔ □↓↘
11Q10	69	↓↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔
4Q537	40	↓↓↓*↓↓↓↓↓↘ ⇔ □↓↘

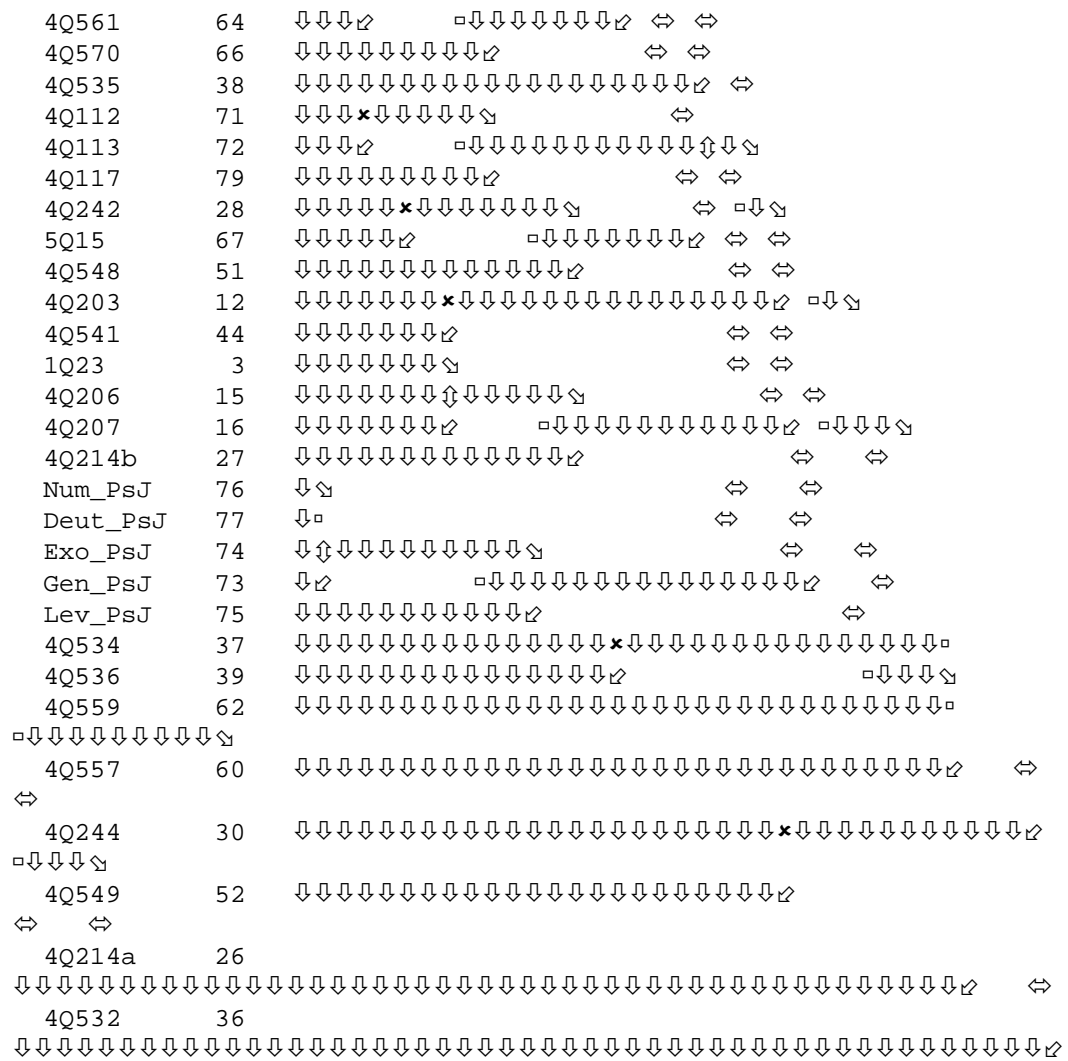


Figure 8.8 Hierarchical clustering dendrogram of the 74 Qumran Aramaic scrolls with Targum Pseudo-Jonathan books using standardised rates of 20 *a priori* criteria.

Qumran Biblical and Targumic Aramaic

Internal comparison indicated significant textual heterogeneity within the Qumran Aramaic corpus. The Targums considered thus far appear to relate closely to the Qumran Biblical Aramaic group, but in addition a further ‘Targumic’ group of scrolls has emerged during combined clustering of Qumranic and Targumic material. The only Qumran Aramaic Targum, 11Q10, appears to relate relatively poorly to the non-Qumranic Targums. The relative proximity to the Qumran Biblical Aramaic group raises the possibility of this group, or the related biblical texts presumably existing beyond Qumran, influencing the text type of the non-Qumran Targums. So-called scribal archaizing would be an example of this. Such archaizing may be difficult to detect. This is not made easier by possible later alterations of an archaized text. Since it is not an easy matter to decide between temporal and spatial influences on a text, one way forward is to consider ‘biblicising’ rather than ‘archaizing’. Table 7.4 reveals distinctive ‘biblical’ features compared with other Qumran text types. As noted above these are:

high determined state nouns rates,

low *afel* rates,

high *hafel* rates,

low *itpeel* rates,

high *hitpeel* rates,

high *y-* prefixed imperfects rates,

high relative particles represented by *dy* rates,

presence of *dnh* or *dn'* forms.

	Qumran Biblical group	Targum Onkelos	Targum Jonathan	Targum Neofiti	Targum Pseudo- Jonathan
Determined nouns ratio	0.49	0.48	0.45	0.48	0.48
Hafel/Afel	2.9	0.01	0.01	0.01	0.04
Hitpeel/Itpeel	10.3	0	0	0	0.002
y-prefixed imperfects ratio	0.16	0.19	0.16	0.19	0.16
<i>dy</i> proportion	1.0	0	0.004	0.21	0.03
<i>dn'</i> or <i>dnh</i>	Present	Absent	Present	Present	Present

Table 8.9 Ratios (presence/absence) of Biblical group specific features in the Targums.

These are therefore the features that would be expected to be artificially different from the rates predicted according to other textual features if a text had been ‘biblicised’.

Table 8.9 shows the proportion of determined state nouns compared with total nouns, hafel/afel ratio, hitpeel/itpeel ratio, proportion of y-prefixed imperfects for all verbs, *dy*/total relative particles ratio and whether *dnh* or *dn'* forms are present for the Qumran Biblical group and the Targums. Table 8.9 reveals little difference between text groups for determined nouns and y-prefixed imperfect ratios. There are large differences between the Qumran Biblical group and the Targums for hafel/afel ratio and hitpeel/itpeel ratios. This suggests that if ‘biblicising’ of these criteria has occurred, this has brought determined nouns and y-prefixed imperfect forms more into harmony with biblical texts than any scribal ‘correction’ of afel into hafel forms or itpeel into hitpeel

forms. It is a little difficult to conceive of scribes carefully altering the proportion of determined state nouns to harmonise with biblical texts whilst not altering afels to hafels, but such a process cannot be excluded. Also, ‘biblicising’ of other textual features that are not included in the 22 *a priori* criteria has not been examined and may well have occurred.

Despite these findings for overall rates, some portions of the Targums may have been ‘biblicised’ rather than others. Stephen Kaufman has hypothesized that portions of text near the beginning of the Pentateuch are likely to be affected more than those near the end.³ This can be considered for those ‘biblical’ textual criteria that differ significantly between the Qumran Biblical group and the Targums. First inspecting hafel/afel ratios from Genesis to Deuteronomy there is no obvious trend in Targum Onkelos, Targum Neofiti or Targum Pseudo-Jonathan. Since all hitpeel rates are close to zero, there is no trend for hitpeel/itpeel ratios in any of the Targums either. Similarly, *dy* rates are zero for all Targum Onkelos books, so no trend is present, whilst for Targum Neofiti *dy* rates are, if anything, higher in the later books than the earlier books. The *dy* rates are lower for Pseudo-Jonathan, but again no trend is discernible. There are so few *dnh* or *dn*’ forms in any of the Targums that any statement about trends is not possible, but there is nothing striking. Again, then, on these criteria there is no evidence to support any ‘biblicising’ of Targum texts.

³ E. M. Cook (2006).

Aramaic Documents from Ancient Egypt

	Verbs	Nouns	Adjectives	Particles	Pronouns	Suffixes
A1	9	31	1	23	0	5
A2	147	351	26	365	20	148
A3	160	371	52	405	15	150
A4	286	816	150	924	31	182
A5	38	90	26	105	6	21
A6	255	769	199	975	28	215
B1	25	75	10	71	0	23
B2	316	1468	296	1297	96	364
B3	496	1996	462	1818	140	550
B4	93	378	78	299	18	73
B5	85	226	51	286	19	92
B6	38	230	72	167	11	43
B7	40	117	14	140	6	37
B8	121	279	88	246	9	86
C0	15	30	6	47	5	18
C1	434	720	139	966	78	333
C2	218	329	144	491	26	125
C3	3	138	28	27	1	1
C37A	15	97	48	62	1	6
C37B	9	65	33	34	2	5
C37C	17	116	58	67	1	8
C37D	23	143	69	90	4	9
C37E	23	150	77	90	3	11
C37K	32	207	105	123	4	12
C37G	19	155	98	67	1	8
C37F	8	127	70	48	0	4
C37J	24	172	86	94	3	11
C37JV	7	56	27	33	0	3
C37FV	8	128	95	49	0	3
C37GV	9	128	77	72	0	4
C37KV	40	267	86	217	0	21
C37EV	20	177	57	126	0	6
C37DV	4	75	77	16	0	0
C37BV	0	1	1	2	0	0
C37AV	0	3	1	0	0	0

C37H	3	21	17	11	1	3
C37L	3	17	6	10	0	1
C38	28	158	54	140	15	4
C3A	64	1948	719	482	20	73
C4	0	277	3	8	0	5
D1	81	215	34	191	10	65
D2	30	161	38	111	7	29
D3	21	245	71	81	4	13
D4	12	46	13	30	1	8
D5	12	69	16	44	0	10
D6	25	79	17	77	1	41
D7	307	429	100	740	31	204
D8	7	326	128	68	1	11
D9	0	143	1	0	0	1
D10	0	44	0	0	0	0
D11	1	67	6	20	1	1
D12	0	28	4	4	0	0
D13	0	37	35	2	0	0
D14	0	14	2	5	0	0
D15	3	17	0	7	0	0
D16	0	1	0	2	0	1
D17	1	11	2	7	1	0
D18	1	59	0	10	0	0
D19	0	21	0	1	0	0
D20	10	55	9	26	0	2
D21	0	49	2	8	0	0
D22	11	167	49	72	1	4
D23	89	124	30	200	16	60
D24	2	72	14	26	1	2

Table 8.10 Parts of speech count and total words for the Egyptian Aramaic papyri.

As was the case with the Targums, where Qumran textual criteria would not necessarily be those chosen to classify them, so with the Ancient Egyptian documents, only here

Criterion	Mean	Median	Minimum, Maximum
Determined nouns	65.3	60	0, 182
Afel	1.1	0	0, 25.5
Hafel	4.1	0	0, 33.3
Itpeel	2.8	0	0, 17.9
Hitpeel	0	0	0, 0
<i>m</i> -prefixed Peal infinitives	1.6	0	0, 14.5
Peal infinitives	2.7	0	0, 19.9
<i>l</i> prefixed imperfects	0	0	0, 0
<i>y</i> prefixed imperfects	7.1	0	0, 41.6
3fp verbs	0	0	0, 0
Perfect 3mp ending <i>w</i>	7.7	0.3	0, 63.4
Perfect 3mp ending <i>w'</i>	0	0	0, 0
Retained aleph before suffix	1.5	0	0, 25.0
2ms pronominal suffix <i>kh</i>	0	0	0, 0
3fs pronominal suffix <i>h'</i>	0.1	0	0, 4.2
Relative particles	8.9	5.7	0, 45.5
Relative particle <i>dy</i>	0	0	0, 0
Direct object marker <i>yt</i>	0	0	0, 0
<i>dnh</i> or <i>dn'</i> forms	0.2	0	0, 7.0
Demonstrative pronoun <i>'lh</i>	0	0	0, 0
Demonstrative pronoun <i>'lyn</i>	0	0	0, 0
Demonstrative pronoun <i>'ln</i>	0	0	0, 0

Table 8.11 Mean, median and range of rates per 1,000 words for the 22 criteria occurring in the Aramaic Documents from Ancient Egypt.

criteria for classifying Aramaic at an earlier period than Qumran would be appropriate.⁴

Sixty-four documents from a variety of sources were available for analysis. Parts of

⁴ H. Gzella (2008) suggests some textual criteria that might be helpful classifying early Aramaic texts.

speech frequencies per document are shown in Table 8.10. Table 8.11 shows descriptive summaries for the 22 *a priori* criteria in this collection. The Table shows that 10 of the 22 *a priori* criteria that occur in at least ten percent of the Qumran Aramaic scrolls are entirely absent from the Aramaic Documents of Ancient Egypt. This compares with five criteria that were absent from Targum Jonathan books, six absent from Targum Onkelos, Targum Neofiti and Targum Pseudo-Jonathan. Thus, taking the Aramaic Documents of Ancient Egypt as a single corpus, Qumran Aramaic scrolls appear less close to these texts than to the Targums in terms of the 22 *a priori* criteria. However, neither the Qumran Aramaic scrolls nor the Aramaic Documents of Ancient Egypt can be considered homogeneous morphologically. Indeed, as noted in Materials Examined, the Comprehensive Aramaic Lexicon considers that at least three distinct dialects are to be found amongst the Aramaic Documents of Ancient Egypt. Figure 8.9 shows the two-dimensional Euclidean distance plot of the relationship between the 22 *a priori* criteria in the Aramaic Documents of Ancient Egypt. Young's S-stress for this model is 0.0021 indicating a very good fit. This is helped by the number of variables with zero value that cluster tightly together at the origin (0,0). Comparing Figure 8.9 to the equivalent model for the Qumran Aramaic scrolls (Figure 7.3) shows that determined state nouns are clearly separate in both, but that otherwise variables in the Aramaic Documents of Ancient Egypt are more homogeneously distributed. The inference is that, apart from determined state nouns, raw rates and standardised rates are likely to result in similar clustering of documents. In general, comparison of the distributions indicates that the morphological characteristics of the Aramaic Documents of Ancient Egypt on these 22 *a*

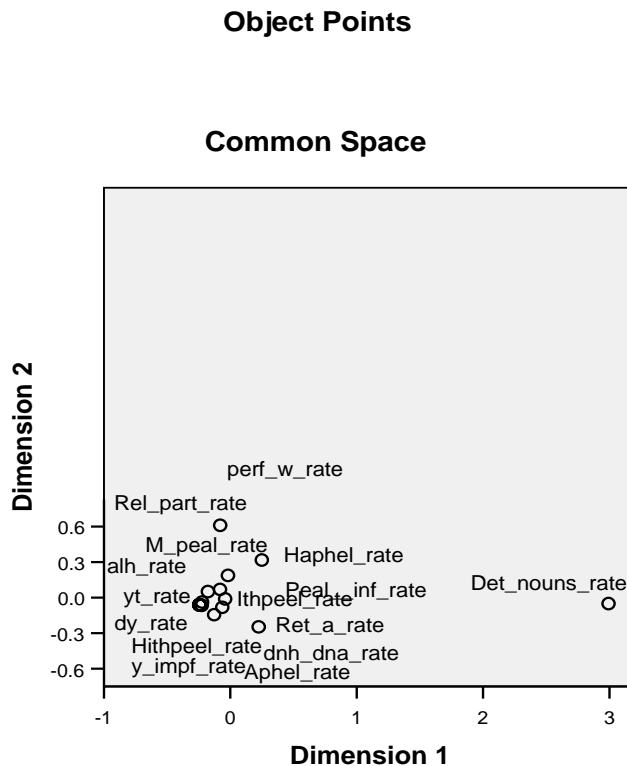


Figure 8.9 Two dimensional scaling using the Euclidean distance between the 22 *a priori* criteria for the 64 Aramaic Documents of Ancient Egypt.

priori criteria are somewhat different from those of Qumran Aramaic texts. Given the relatively large proportion of these *a priori* criteria with zero values, they are not ones that would have been chosen to cluster the Aramaic Documents of Ancient Egypt for their own sake; other criteria with greater discriminating power would be needed for this task. In other words, only 12 of the 22 *a priori* criteria are available for the hierarchical clustering algorithm.

Figure 8.10 shows the hierarchical clustering dendrogram for the 64 Aramaic Documents of Ancient Egypt using raw rates of the 22 *a priori* criteria. One document, D17, stands quite separate. This document consists of only 22 words, hence the likely reason for its failure to agglomerate easily: inadequate morphological data. Otherwise, the dendrogram shows that hierarchical clustering has been useful in identifying several low level clusters with fairly large distances between them. The first low level cluster comprises C3, C37AV, C37BV, C37DV, C4, D9, D12, D13, D16 and D19. Closely related to this cluster is a large group: B1, B6, C3A, C37A, C37B, C37D, C37E, C37F, C37G, C37H, C37J, C37K, C37FV, C37GV, C37JV, C38, D3, D8, D11, D14, D15, D18, D20, D22 and D24. A final low level cluster comprises A2, A3, A6, B2, B3, B4, B5, B7, B8, C0, C37L, D1, D2, D4, D5, D7 and D23. Another group, slightly less closely clustered together consists of A1, C2, C37EV and C37KV.

Turning now to the hierarchical clustering using standardised rates for the 22 *a priori* criteria shown in Figure 8.11, which shows a more gradually stepped dendrogram that also has several low level clusters, but also slightly more documents that do not agglomerate that well. As has been seen previously, standardised rate clustering does not perform so well. The first low levels cluster to appear is C3, C3A, C37AV, C37BV, C37DV, C4, D3, D8, D9, D10, D11, D12, D13, D14, D15, D16, D19 and D22. This is the same as the first low level cluster using raw rates with the addition of eight other

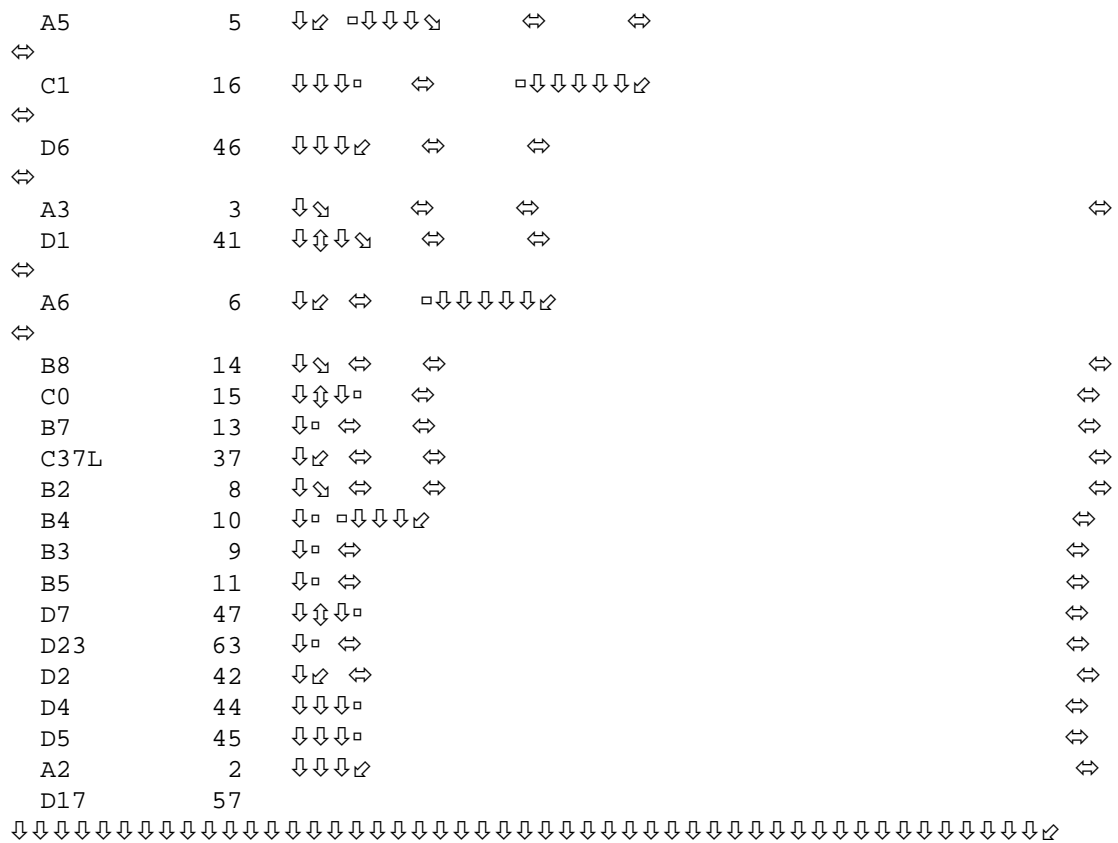


Figure 8.10 Hierarchical clustering dendrogram of the 64 Aramaic Documents of Ancient Egypt using raw rates of the 22 *a priori* criteria.

documents that are part of the closely related cluster in the raw rates dendrogram. The next cluster comprises B1, C0, C37B, C37C, C37D, C37E, C37F, C37G, C37H, C37J, C37K, C37L, C37FV, C37JV, D2, D20, D21 and D24. This is very similar to the second raw rate cluster and again is closely related to the first cluster. A third, smaller cluster contains A3, A4, A6, C1, D1 and D7. This bears some resemblance to the final raw rates low level cluster.

Notably, A1 and C2, identified as ‘non-Egyptian’ by the Comprehensive Aramaic Lexicon project cluster together in both raw and standardised rates clustering, whilst A6, also considered ‘non-Egyptian’ appears more closely related to the bulk of the documents. However, the criteria were not chosen to distinguish Egyptian and non-Egyptian provenance specifically, but were geared towards classifying the Qumran scrolls. It is possible that if some of the *a priori* criteria discarded because of low

D16	56	↓↘	
D19	59	↓□	
C37BV	34	↓□	
D12	52	↓□	
D13	53	↓□	
C37AV	35	↓□	
D10	50	↓□	
D9	49	↓□	
C4	40	↓□	
C3	18	↓□	
C37DV	33	↓□	
D11	51	↓□	
D14	54	↓□	
D15	55	↓□	
D8	48	↓↕↓↘	
D22	62	↓□ ↔	
C3A	39	↓□ ↔	
C37A	19	↓□ ↔	
D3	43	↓↗↘ ↔	
D18	58	↓↘ ↔	
D21	61	↓□ ↔	
C37FV	29	↓□ ↔	
D24	64	↓↕↓□	
C37F	26	↓□ ↔	
D20	60	↓↗↘ ↔	
C0	15	↓↘ ↔	
D2	42	↓□ ↔	
C37L	37	↓↕↓□	
B1	7	↓↗↘ □↓↓↓↓↓↘↘	
C37G	25	↓↘ ↔	↔
C37JV	28	↓□ ↔	↔
C37E	23	↓□ ↔	↔
C37J	27	↓↕↓□	↔
C37D	22	↓□ ↔	↔
C37K	24	↓□ ↔	↔
C37C	21	↓□ ↔	↔
C37B	20	↓↗↘ ↔	□↓↘
C37H	36	↓↓↓↗	↔ ↔

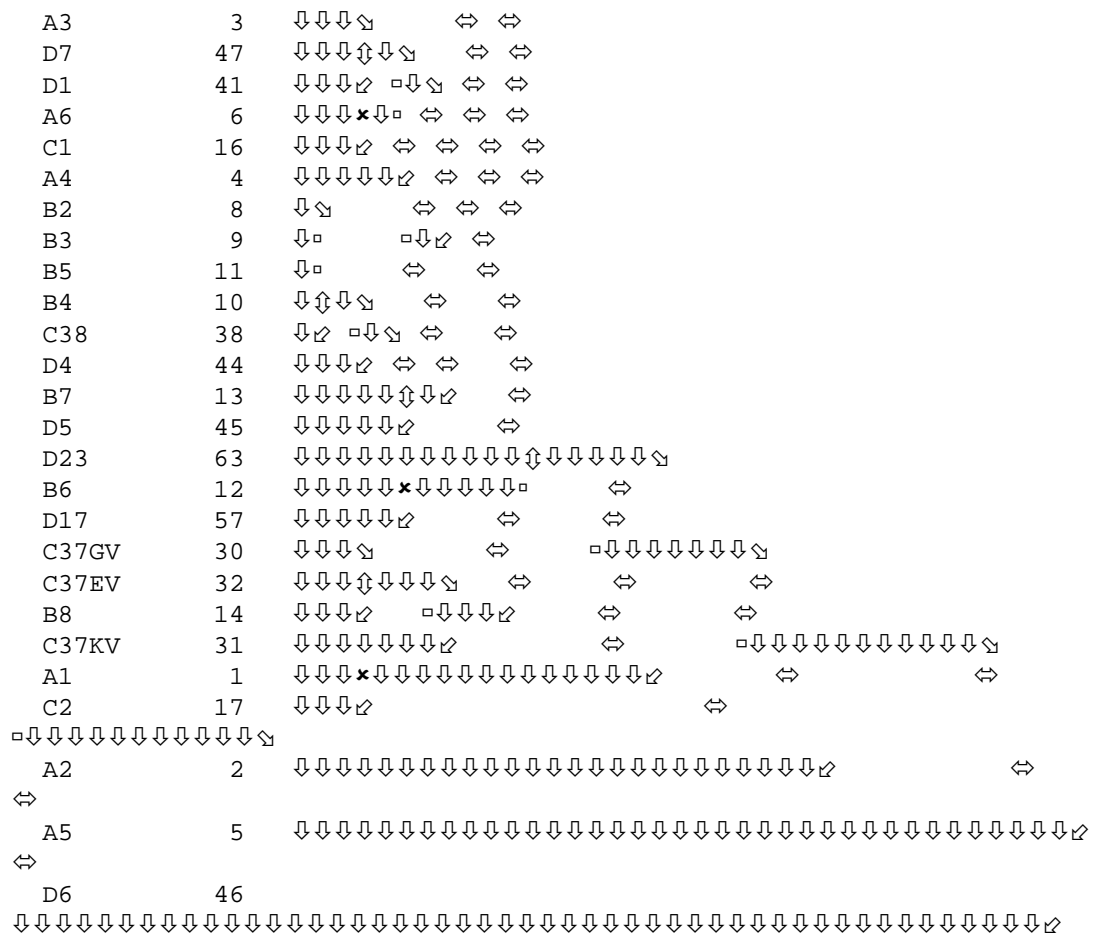


Figure 8.11 Hierarchical clustering dendrogram of the 64 Aramaic Documents of Ancient Egypt using standardised rates of the 22 *a priori* criteria.

frequency of occurrence in Qumran Aramaic scrolls had been used, these may have proved more discriminatory in the Egyptian Aramaic corpus. Nevertheless, since A1, A6 and C2 appear in distinct clusters from the bulk of Egyptian documents, the first two low level clusters on raw rates can be considered more representative of an ‘Egyptian’ Aramaic text type from the Imperial period. The reproducibility of these clusters, with a

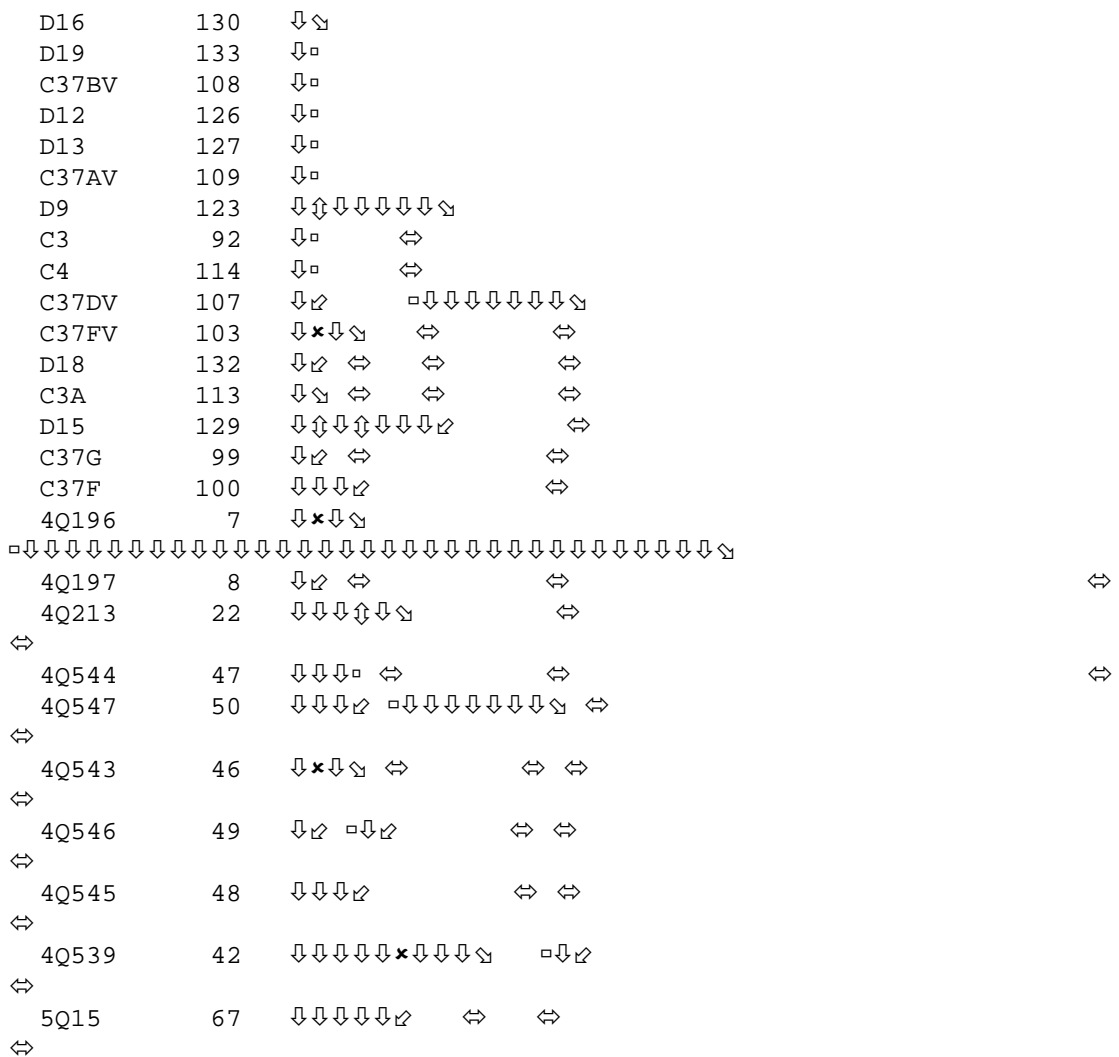
little interchange, on standardised rates indicates that this text type is likely to be fairly robust.

Selecting these first two clusters and comparing their members (n=34) with the other Ancient Egyptian Aramaic documents (n=30) shows significant differences on raw rates for *afel* (p=.003, members lower), *hafel* (p<.001, members lower), determined state nouns (p<.001, members lower), *m*-prefixed peal infinitives (p<.001, members lower), total peal infinitives (p<.001, members lower), *y*-prefixed imperfects (p<.001), *-w* ending perfects (p=.032, members lower) and retained alephs (p<.001, members lower), that is eight of the possible twelve non-zero criteria. Six of these criteria involve verbs, and comparing these two groups shows that the proportion of verbs in the cluster members is a mean of 3.7% compared with 10.3% in non-members (p<.001). This compares with a mean of 13.6% verbs for the Qumran Aramaic scrolls. Adjusting verbal forms for the total number of verbs in a document still showed significant differences for all the verbal criteria, with cluster member documents lower, except perfect *-w* ending perfects (p=.55). Hence, even adjusting for the type of text in terms of parts of speech composition, these verbal forms are significantly less frequent in the typical Ancient Egyptian Aramaic documents.

This division into core, typical Ancient Egyptian Aramaic documents and those less typical, of which some are considered to have dialects representative of a more Eastern type, can thus be used to provide two groups of external correlates for the Qumran Aramaic scrolls.

Aramaic Documents from Ancient Egypt and their relationship to Qumran Aramaic scrolls

Figure 8.12 shows the dendrogram produced by hierarchical clustering of the typical Ancient Egyptian Aramaic documents together with the Qumran Aramaic scrolls that represent different text types using raw rates. This clearly indicates that the texts fall into



	B6	86	↓↓↓↓x↓↓↓↓↓↓↓↓	⇔	
⇔					
	D20	134	↓↓↓↓⇨	⇔	⇔
⇔					
	4Q213a	23	↓↓↓↓x↓↓⇨	⇨↓↓↓↓⇨	
⇔					
	4Q213b	24	↓↓↓↓⇨	⇔	⇔
⇔					
	D11	125	↓⇨	⇔	⇔
	D14	128	↓□	⇔	⇔
	D8	122	↓□	⇔	⇔
	D22	136	↓□	⇔	⇔
	C37A	93	↓□	⇔	⇔
	D3	117	↓□	⇨↓⇨	⇔
	C37E	97	↓⇨↓⇨⇨	⇔	⇔
⇔					
	C37J	101	↓□	⇔	⇔
⇔					
	C37D	96	↓□	⇔	⇔
⇔					
	C37B	94	↓□	⇨↓□	⇔
	C37JV	102	↓□	⇔	⇔
⇔					
	D24	138	↓⇨	⇔	⇨↓⇨
⇔					
	C38	112	↓↓↓↓⇨	⇔	⇔
⇔					
	4Q208	17	↓⇨	⇔	⇔
	4Q209	18	↓⇨↓⇨⇨	⇔	⇔
⇔					
	B1	81	↓⇨	⇨↓⇨	⇔
	C37H	110	↓↓↓↓⇨	⇔	⇔
	C37GV	104	↓↓↓↓↓↓↓↓↓↓		
⇔					
	2Q24	5	↓↓↓↓⇨		⇔
	11Q18	70	↓↓↓↓□		⇔
	4Q210	19	↓↓↓↓⇨↓↓↓↓↓↓↓↓		
⇔					
	1Q21	2	↓↓↓↓□	⇔	⇔
	4Q554	58	↓↓↓↓⇨	⇔	⇔
	4Q553	57	↓↓↓↓⇨	⇨↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓	
⇔					
	4Q558	61	↓↓↓↓⇨↓⇨	⇔	⇔
⇔					
	4Q204	13	↓↓↓↓□	⇔	⇔
⇔					
	4Q531	35	↓↓↓↓⇨	⇨↓↓↓↓⇨	⇔
⇔					
	4Q552	56	↓↓↓↓↓↓↓□		⇔
	4Q205	14	↓↓↓↓↓↓⇨		⇔
⇨↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓					

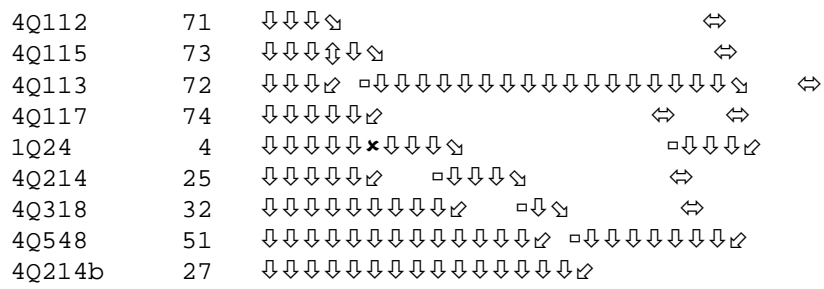


Figure 8.12 Hierarchical clustering dendrogram of the typical Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls of discrete text types using standardised rates of the 22 *a priori* criteria.

two main groups. One contains the typical Ancient Egyptian Aramaic documents together with some of the Qumran Aramaic scrolls. The other group consists of Qumran Aramaic scrolls only representing the Enochic-Four Kingdoms, New Jerusalem, Biblical and Targumic text types. Within the first group, a sub-group comprising only typical Ancient Egyptian Aramaic documents appears relatively separate. The Tobit-Visions of Amram text type also form a discrete cluster. This leaves a final sub-group with a mixture of typical Ancient Egyptian Aramaic documents and Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls. The members are:

Qumran

4Q208, 4Q209, 4Q213a, 4Q213b, 4Q539 and 5Q15;

Ancient Egyptian Aramaic documents

B1, B6, C37A, C37B, C37D, C37E, C37H, C37GV, C37JV, C38, D3, D8, D11, D14, D20, D22 and D24.

Figure 8.13 shows the dendrogram produced by hierarchical clustering of the non-core Ancient Egyptian Aramaic documents together with the Qumran Aramaic scrolls that represent different text types using raw rates. Unsurprisingly, the non-core Ancient Egyptian Aramaic documents are more dispersed throughout the dendrogram than the typical Ancient Egyptian Aramaic documents. 4Q208, 4Q209 and 4Q318 fall within a large group of Ancient Egyptian Aramaic documents. 4Q213a and 4Q213b cluster with C37C, C37K and D21. Otherwise relationships between Qumran scrolls and Ancient

B2	82	↓↘		
B4	84	↓□		
B3	83	↓↕↓↘		
B5	85	↓□ ↔		
D7	121	↓□ ↔		
D23	137	↓↘ ↓□ ↘		
D4	118	↓↓↓□ ↔		
D5	119	↓↓↓↘ ↓□ ↘		
4Q208	17	↓×↓↘ ↔ ↔		
4Q209	18	↓↘ ↓□ ↘ ↔		
A2	76	↓↓↓↘ ↓□ ↓↓↓↓↓↓↓↓↓↓↘		
A3	77	↓↘ ↔	↔	
D1	115	↓↕↓↘ ↔	↔	
A6	80	↓↘ ↓□ ↓↓↓↘	↔	
B8	88	↓×↓□	↔	
C0	89	↓↘ ↔	↔	
4Q318	32	↓×↓□	↓□ ↓↓↓↘	
C37L	111	↓↘ ↔	↔	↔
B7	87	↓×↓↘	↔	↔
D2	116	↓↘	↔	↔
A4	78	↓↓↓×↓↓↓↘	↔	↔
A5	79	↓↓↓↘ ↓□ ↓↓↓↘	↔	↔
C1	90	↓↓↓↓↓×↓↘ ↔	↔	↔
D6	120	↓↓↓↓↓↘ ↓□ ↓↓↓↓↓↘ ↓□ ↓↓↓↓↓↘		
1Q24	4	↓↓↓↓↓×↓↓↓↘ ↔	↔	↔
4Q214	25	↓↓↓↓↓↘ ↓□ ↘	↔	↔
4Q548	51	↓↓↓↓↓↓↓↓↓↘	↔	↔
4Q539	42	↓↓↓↓↓×↓↓↓↘	↔	↔
5Q15	67	↓↓↓↓↓↘ ↔	↔	↔
2Q24	5	↓×↓↘ ↔	↔	↔
11Q18	70	↓↘ ↔ ↓□ ↓↓↓↓↓↓↓↓↓↓↘		↔
4Q210	19	↓↓↓↕↓↓↓↓↓↓□	↔	
1Q21	2	↓↓↓□ ↔	↓□ ↓↓↓↘	
4Q554	58	↓↓↓↘ ↔	↔	↔
4Q553	57	↓×↓↘ ↔	↔	↔
4Q558	61	↓↘ ↔	↔	↔
4Q204	13	↓↓↓↕↓↘ ↔	↔	↔
4Q531	35	↓↓↓□ ↓□ ↓↓↓↘	↔	↔
4Q552	56	↓↓↓↘ ↔	↔	↔
4Q205	14	↓↓↓↓↓↘	↔	↔
C2	91	↓↓↓×↓↓↓↘	↔	↔
C37EV	106	↓↓↓↘ ↓□ ↘	↔	↔
A1	75	↓↓↓↓↓↓↓↓↓↘ ↓□ ↓↓↓↓↓↓↓↓↓↓↘		
↔		↓↓↓↓↓↓↓↓↓↘		
C37KV	105	↓↓↓↓↓↓↓↓↓↘		↔
↔				
4Q196	7	↓↘	↔	↔
4Q197	8	↓↕↓↘	↔	↔
4Q213	22	↓↘ ↔	↔	↔
4Q544	47	↓×↓□	↔	↔

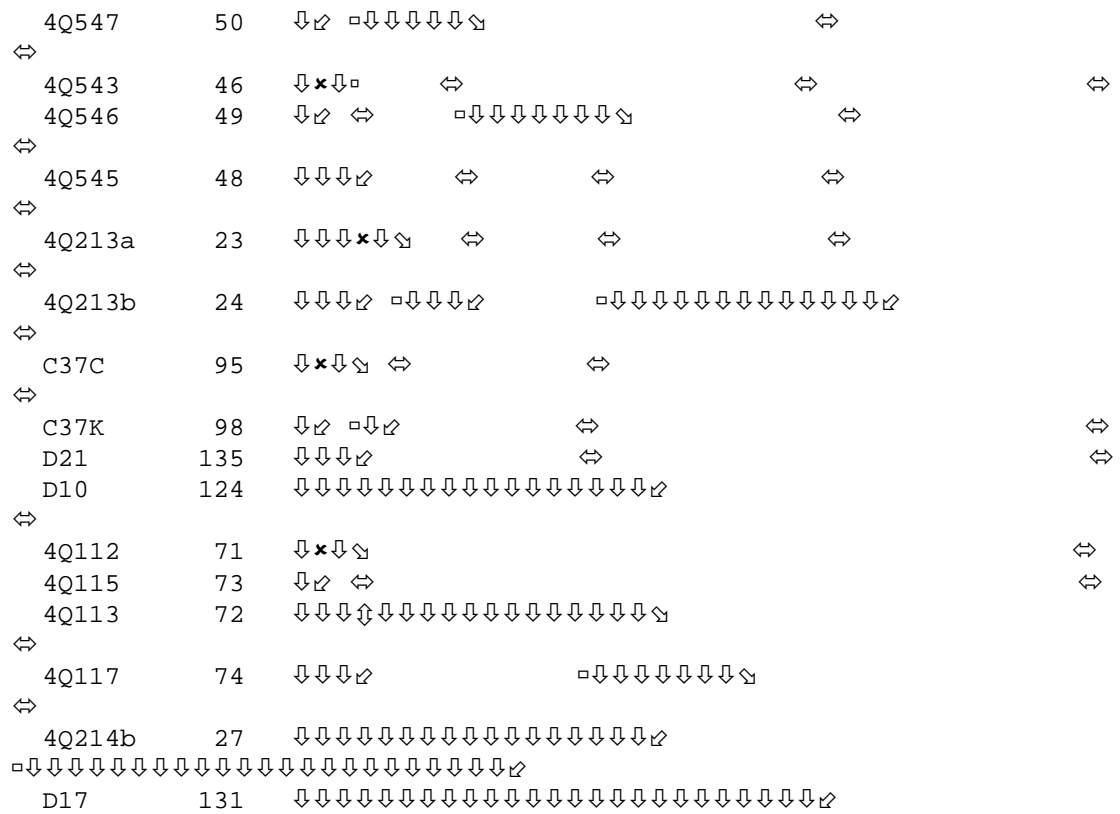


Figure 8.13 Hierarchical clustering dendrogram of the non-core Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls of discrete text types using standardised rates of the 22 *a priori* criteria.

Egyptian Aramaic documents are less close as indicated by the horizontal distances of the connecting lines. Taking Figures 8.12 and 8.13 together, there is a suggestion that 4Q208, 4Q209, 4Q213a, 4Q213b share similar textual features with the Ancient Egyptian Aramaic documents or, perhaps more likely, lack textual features that are common in other Qumran scrolls, just as the Ancient Egyptian Aramaic documents also lack these features. Comparing the typical Ancient Egyptian Aramaic documents with the Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls show significant differences for rates of Afels ($p < .001$, Egyptian Aramaic lower), *dy* ($p = .008$, Egyptian Aramaic lower) and determined state nouns ($p = .009$, Egyptian Aramaic lower) only. However, given the small sample size of the Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls ($n = 6$), Type 2 statistical error is very likely. To put these differences in context, whilst Afels are more common in the Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls, Hafels do not occur in this group whereas they are present, albeit infrequently, in the typical Ancient Egyptian Aramaic documents. This is consistent with the typical Ancient Egyptian Aramaic documents representing an older form of Aramaic. Similarly, *dy* is not found as a relative particle in the typical Ancient Egyptian Aramaic documents, the form is *zy*, whilst it is present in the Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls, though *d-* is the usual form in these, again consistent with a temporal shift. The higher determined state of nouns rate in the Enochic-Aramaic Levi/Apocryphon of Joseph B/New Jerusalem Qumran scrolls may reflect the increasing use of this state for semantically non-determined nouns in later Aramaic. The same criteria are significantly different between the Enochic-Aramaic Levi/Apocryphon of

Joseph B/New Jerusalem Qumran scrolls and the non-core Ancient Egyptian Aramaic documents. But in addition there are significant differences for rates of Hafels ($p=.005$, non-core Egyptian Aramaic higher), third person feminine single *-h'* suffix ($p=.040$, non-core Egyptian Aramaic lower) and *y*-prefixed imperfects ($p=.013$, non-core Egyptian Aramaic higher). The Hafel rate difference is expected from the dating of these text groups, but the 2ms *-kh* personal pronominal suffix rate difference is not, though Ancient Egyptian Aramaic documents are considered to be peculiar with this regard.⁵

As with the Targumic Qumran Aramaic text type, it is possible that a discrete text type exists within the non-aligned Qumran Aramaic scrolls that relates to the Ancient Egyptian Aramaic documents. Figure 8.14 shows the hierarchical clustering dendrogram of the typical Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the 22 *a priori* criteria. The typical Ancient Egyptian Aramaic documents cluster separately from the non-aligned Qumran Aramaic group, though there is a loose Qumran Aramaic cluster that is more closely related to them comprising 4Q538 (Testament of Judah), 4Q550c (proto-Esther), 4Q561 (Physiognomy/Horoscope), 4Q570 (Aramaic R) and 6Q8 (Giants). This is a mixed group in terms of content. The rates for the 22 *a priori* criteria are shown for each of these scrolls together with the mean rate for the typical Ancient Egyptian Aramaic documents in Table 8.12. The pattern of rates is heterogeneous, such

⁵ U. Scattner-Rieser (2004), p58.

D16	130	↓↘			
D19	133	↓□			
C37BV	108	↓□			
D12	126	↓□			
D13	127	↓□			
C37AV	109	↓□			
D9	123	↓↑↓↓↓↓↓↓↘			
C3	92	↓□	↔		
C4	114	↓□	□↓↘		
C37DV	107	↓↘	↔	↔	
4Q549	52	↓↓↓↓↓↓↓↓×↓↘	↔		
4Q559	62	↓↓↓↓↓↓↓↘	□↓↓↓↘		
B6	86	↓×↓↓↓↓↘	↔	↔	
D20	134	↓↘	↔	↔	↔
4Q211	20	↓×↓↓↓↓□	↔	↔	
C37GV	104	↓↘	□↓↓↓↘	↔	
C37FV	103	↓↘	↔	↔	
D18	132	↓□	↔	↔	
C3A	113	↓□	↔	↔	
D15	129	↓↑↓↓↘	↔	↔	
C37G	99	↓□	↔	↔	↔
C37F	100	↓↘	↔	↔	↔
B1	81	↓×↓□	↔	↔	
C37H	110	↓↘	□↓↘	↔	
D11	125	↓↘	↔	↔	
D14	128	↓□	↔	□↓↓↓↓↓↓↘	
D8	122	↓□	↔	↔	↔
D22	136	↓□	↔	↔	↔
C37A	93	↓□	↔	↔	↔
D3	117	↓□	↔	↔	↔
C37E	97	↓□	↔	↔	↔
C37J	101	↓□	↔	↔	↔
C37D	96	↓□	↔	↔	↔
C37B	94	↓↑↓↓↘	↔	↔	↔
C37JV	102	↓□	↔	↔	↔
D24	138	↓□	↔	↔	↔
C38	112	↓↘	↔	↔	↔
4Q538	41	↓×↓↓↘	↔	↔	↔
6Q8	68	↓↘	□↓↘	↔	↔
4Q550c	55	↓↓↓↓↘	□↓↓↓↘	↔	□↓↓↓↓↓↓↓↘
4Q570	66	↓↓↓↓↓↓↘	□↓↓↓↘	↔	↔
4Q561	64	↓↓↓↓↓↓↓↓↘	↔	↔	↔
4Q529	33	↓↓↓↓×↓↘	↔	↔	↔
4Q550	53	↓↓↓↓↘	□↓↓↓↓↓↓↓↘	↔	↔
4Q244	30	↓↓↓↓↓↓↘	↔	↔	↔
4Q206	15	↓↓↓↓×↓↓↓↓↓↓↓↘	↔	↔	↔
4Q207	16	↓↓↓↓↘	↔	□↓↓↓↓↓↓↓□	↔
1Q23	3	↓↓↓↓↓↓×↓↘	↔	↔	↔
4Q203	12	↓↓↓↓↓↓↘	↔	↔	↔

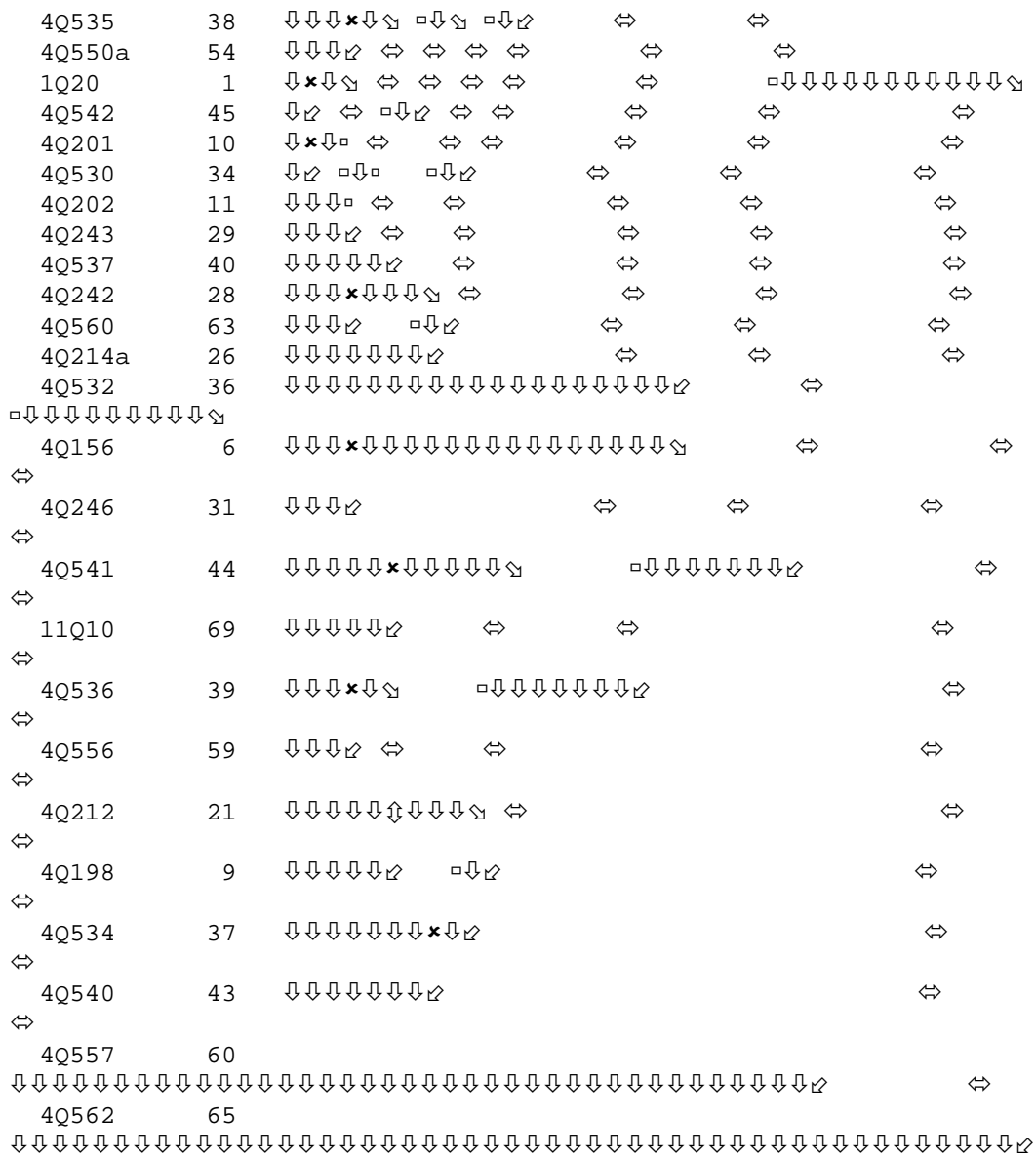


Figure 8.14 Hierarchical clustering dendrogram of the typical Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the 22 *a priori* criteria.

Variable	4Q538	4Q550c	4Q561	4Q570	6Q8	Ancient Egyptian
Determined nouns	14.7	32.3	9.1	34.1	28.9	38.7
Afel	29.4	8.1	9.1	0	19.2	0.2
Hafel	0	0	0	11.4	0	0.7
Itpeel	0	4.0	10.4	0	0	2.5
Hitpeel	0	0	0	0	0	0
<i>m</i> -prefixed Peal infinitives	0	4.0	18.2	11.4	0	1.3
Peal infinitives	7.4	4.0	18.2	11.4	0	1.3
<i>l</i> prefixed imperfects	0	0	10.4	22.7	0	0
<i>y</i> prefixed imperfects	14.7	16.1	52.1	11.4	9.6	2.2
3fp verbs	0	0	0	0	0	0
Perfect 3mp ending <i>w</i>	22.1	28.2	0	11.4	9.6	4.0
Perfect 3mp ending <i>w'</i>	0	0	0	0	0	0
Retained aleph before suffix	0	0	0	0	0	0.02
2ms pronominal suffix <i>kh</i>	0	4.0	0	0	0	0
3fs pronominal suffix <i>h'</i>	0	0	0	0	0	0
Relative particles	36.8	48.4	18.2	34.1	38.5	8.0
Relative particle <i>dy</i>	36.8	48.4	18.2	34.1	38.5	0
Direct object marker <i>yt</i>	0	4.0	0	0	0	0
<i>dnh</i> or <i>dn'</i> forms	0	0	0	11.4	0	0
Demonstrative pronoun ' <i>lh</i>	0	0	0	0	0	0
Demonstrative pronoun ' <i>lyn</i>	0	0	0	0	0	0
Demonstrative pronoun ' <i>ln</i>	0	0	0	0	0	0

Table 8.12 Rates per 1,000 words for the 22 criteria occurring in the Qumran `Aramaic scrolls most closely associated with the typical Ancient Egyptian documents together with mean rates for the typical Ancient Egyptian documents.

as *l*-prefixed imperfect forms, and similarities appear to depend on those criteria with zero rates. Given this, it is hard to argue for a discrete typical Ancient Egyptian Aramaic text type amongst the non-aligned Qumran Aramaic scrolls.

Figure 8.15 shows the hierarchical clustering dendrogram of the non-core Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the 22 *a priori* criteria. As might be expected, given that this is a clustering of two non-typical, heterogeneous groups, there is no neat division between Ancient Egyptian and Qumran Aramaic. The group of Qumran Aramaic scrolls that were most closely related to the typical Ancient Egyptian Aramaic documents cluster together again, but near the end of the dendrogram, indicating that they are more distantly related to the non-core Ancient Egyptian Documents. There is nothing in the dendrogram that is convincing in terms of identifying a major further Aramaic text type.

B2	82	↓↘
B4	84	↓□
B3	83	↓□
B5	85	↓□
D7	121	↓↕↓↘
D23	137	↓□ ⇔
D2	116	↓↘ ⇔
D4	118	↓↘↓↕↓↘
D5	119	↓↘↓□ ⇔
A2	76	↓↘↓↘ ↘↓↘
A3	77	↓↘ ⇔ ⇔
D1	115	↓↕↓↘ ⇔ ⇔
A6	80	↓↘ ↘↓↘ ⇔
B8	88	↓↘ ⇔ ↘↓↘
C0	89	↓↕↓↘ ⇔ ⇔
B7	87	↓□ ⇔ ⇔
C37L	111	↓↘ ⇔ ⇔
4Q214a	26	↓↘↓↘↓↘↓↘↓↘ ⇔
4Q242	28	↓↘↓↘*↓↘↓↘↓↘↓↘↓↘
4Q560	63	↓↘↓↘ ⇔ ⇔
A4	78	↓*↓↘ ⇔ ⇔
A5	79	↓↘ ↘↓↘↓↘↓↘ ⇔
C1	90	↓↘↓□ ⇔
D6	120	↓↘↓↘ ↘↓↘
C37C	95	↓*↓↘ ⇔ ⇔
C37K	98	↓↘ ↘↓↘ ⇔ ⇔
4Q211	20	↓↘↓↘ ↘↓↘↓↘ ⇔ ⇔
D21	135	↓↘↓↘↓↘ ⇔ ⇔ ⇔
1Q23	3	↓↘↓↘↓↘*↓↘ ↘↓↘↓↘ ⇔
4Q203	12	↓↘↓↘↓↘ ⇔ ⇔ ⇔
4Q535	38	↓↘↓↘*↓↘ ↘↓↘ ⇔
4Q550a	54	↓↘↓↘ ⇔ ⇔ ⇔
1Q20	1	↓*↓↘ ⇔ ⇔ ↘↓↘↓↘↓↘↓↘↓↘
4Q542	45	↓↘ ⇔ ↘↓↘ ⇔ ⇔
4Q201	10	↓*↓□ ⇔ ⇔ ⇔
4Q530	34	↓↘ ↘↓□ ⇔ ⇔ ⇔
4Q202	11	↓↘↓□ ⇔ ⇔ ⇔
4Q243	29	↓↘↓↘ ⇔ ⇔ ⇔
4Q537	40	↓↘↓↘↓↘ ⇔ ↘↓↘
4Q206	15	↓↘↓↘↓↘*↓↘↓↘ ⇔ ⇔ ⇔
4Q207	16	↓↘↓↘↓↘ ⇔ ⇔ ⇔ ⇔
C2	91	↓*↓↘↓↘ ↘↓↘↓↘↓↘ ⇔ ⇔
C37EV	106	↓↘ ⇔ ⇔ ⇔ ⇔
A1	75	↓↘↓↘↓↘↓↘↓↘↓↘ ⇔ ↘↓↘
C37KV	105	↓↘↓↘↓↘ ⇔ ⇔ ⇔
4Q557	60	↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘ ⇔ ⇔
4Q156	6	↓↘↓↘*↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘↓↘ ⇔ ⇔
4Q246	31	↓↘↓↘ ⇔ ⇔ ⇔
4Q541	44	↓↘↓↘↓↘↓↘*↓↘↓↘ ↘↓↘↓↘↓↘ ⇔
11Q10	69	↓↘↓↘↓↘↓↘↓↘ ⇔ ⇔ ⇔
4Q536	39	↓↘↓↘*↓↘ ↘↓↘↓↘↓↘↓↘↓↘ ↘↓↘

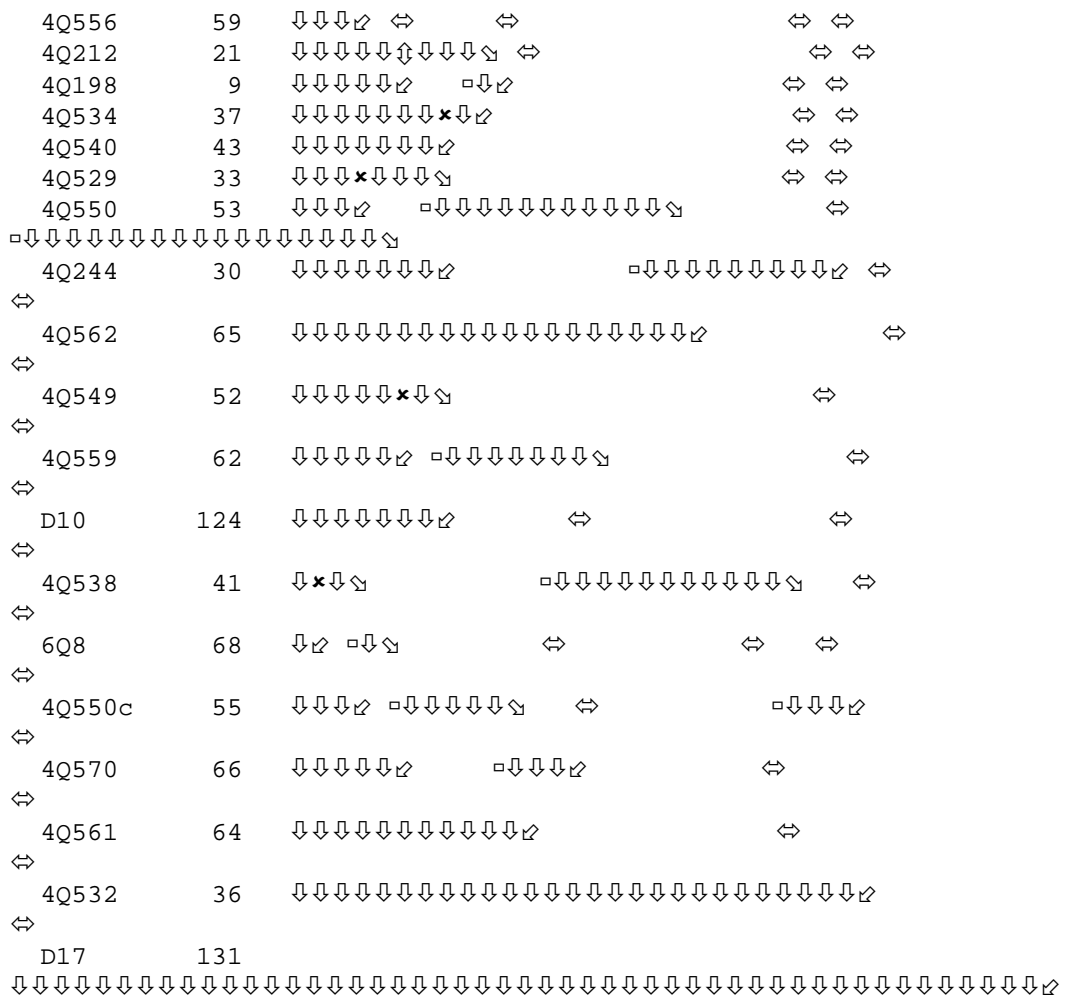


Figure 8.15 Hierarchical clustering dendrogram of the non-core Aramaic Documents of Ancient Egypt together with the Qumran Aramaic scrolls not aligned to any discrete text types using standardised rates of the *22 a priori* criteria.

Non-Qumran Judean Desert Aramaic texts

	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
WDSP1	82	23	26	29	19	12	191
WDSP2	78	25	17	27	15	8	170
WDSP3	72	20	13	15	12	4	136
WDSP4	67	22	19	24	21	5	158
WDSP5	82	23	22	26	30	10	193
WDSP6	64	23	27	33	17	9	173
WDSP7	98	25	24	35	22	8	212
WDSP8	75	18	13	20	13	3	142
WDSP9	102	25	27	28	29	7	218
WDSP10	49	10	3	15	7	0	84
WDSP15	91	26	17	40	8	3	185
WDSP18	10	5	1	2	4	1	23
Mur8	12	1	0	7	1	0	21
Mur10	26	0	0	7	0	0	33
Mur18	39	8	6	4	12	2	71
Mur19	76	29	18	18	58	9	208
Mur20	32	10	9	6	33	2	92
Mur21	53	21	30	4	52	7	167
Mur23	5	2	1	2	6	0	16
Mur25	35	6	7	9	13	2	72
Mur26	17	10	9	9	3	1	49
Mur27	10	4	7	0	15	1	37
Mur28	14	2	6	3	8	3	36
Mur31	9	1	0	4	3	1	18
Mur32	14	3	2	7	8	2	36
Mur33	7	1	0	1	2	1	12
Mur72	16	19	2	8	16	3	64
Sdeir2	36	7	8	7	18	1	77
XHevSe7	60	7	4	11	17	6	105
XHevSe8	142	29	30	26	35	6	268
XHevSe8a	102	10	22	18	17	5	174
XHevSe9	142	43	31	48	17	9	290
XHevSe9a	1	2	3	2	2	1	11
XHevSe10	12	1	1	3	2	0	19

XHevSe11	6	3	1	2	10	1	23
XHevSe12	20	2	6	7	16	0	51
XHevSe13	46	3	9	4	18	5	85
XHevSe21	39	16	6	14	22	3	100
XHevSe22	19	0	2	1	10	1	33
XHevSe23	7	7	5	2	9	2	32
XHevSe24	5	1	1	0	2	0	9
XHevSe24a	3	4	2	1	2	1	13
XHevSe25	6	0	0	1	4	0	11
XHevSe26	10	1	1	2	2	0	16
XHevSe31	2	1	2	0	2	0	7
XHevSe32	6	0	0	2	4	0	12
XHevSe33	2	0	0	2	0	0	4
XHevSe34	1	0	0	1	6	0	8
XHevSe40	0	0	0	0	0	0	0
XHevSe50	48	20	14	18	11	3	114
34Se3	5	0	2	5	8	0	20

Table 8.13 Absolute counts of parts of speech and total words for the non-Qumran Judean Desert Texts.

Finally, a miscellany of Aramaic texts has been discovered in the Judean Desert at locations other than Qumran. These texts are generally fragmentary and preserved on ostraca as well as papyrus. Fifty-one morphologically-tagged texts available from the Accordance Judean-T module preserved more than just one or two words; the absolute counts for the different parts of speech together with total word count for each text is shown in Table 8.13. Using the same criteria for adequate length for statistical analysis as applied to the Qumran Aramaic texts, 26 of the 51 texts have 50 words or more: 11 out of 12 from Wadi ed-Daliyeh, six out of 15 from Murabba'at, eight out of 20 from Nahal Hever, and the one from Wadi Sdeir. Rates per 1,000 words for the different parts of

	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronoun s
WDSP1	429	120	136	152	99	63
WDSP2	459	147	100	159	88	47
WDSP3	529	147	96	110	88	29
WDSP4	424	139	120	152	133	32
WDSP5	425	119	114	135	155	52
WDSP6	370	133	156	191	98	52
WDSP7	462	118	113	165	104	38
WDSP8	528	127	92	141	92	21
WDSP9	468	115	124	128	133	32
WDSP10	583	119	36	179	83	0
WDSP15	492	141	92	216	43	16
Mur18	549	113	85	56	169	28
Mur19	365	139	87	87	279	43
Mur20	348	109	98	65	359	22
Mur21	317	126	180	24	311	42
Mur25	486	83	97	125	181	28
Mur72	250	297	31	125	250	47
Sdeir2	468	91	104	91	234	13
XHev/Se7	571	67	38	105	162	57
XHev/Se8	530	108	112	97	131	22
XHev/Se8a	586	57	126	103	98	29
XHev/Se9	490	148	107	166	59	31
XHev/Se12	392	39	118	137	314	0
XHev/Se13	541	35	106	47	212	59
XHev/Se21	390	160	60	140	220	30
XHev/Se50	421	175	123	158	96	26

Table 8.14 Rates per 1,000 words for parts of speech for non-Qumran Judean Desert texts of adequate length for statistical analysis.

speech for these 26 texts are shown in Table 8.14. Excluding the single text from Wadi Sdeir, there was a significant difference between geographical locations in terms of the parts of speech composition ($F=2.66_{10,38}$, $p=.014$, partial $\eta^2=0.412$): means and 95% confidence intervals for the means for each part of speech by geographical site are shown in Table 8.15. Inspecting Table 8.15, rates for both pronouns and pronominal suffixes are relatively low and very similar across all sites; it is difficult to comment on the single Wadi Sdeir text, mean rates are presented to provide some general comparison. Nouns form the largest proportion of texts at all sites, but significantly less so at Murabba'at. Rates of verbs are significantly lower at Nahal Hever. Rates of adjectives and particles show the most discriminatory pattern between sites. Adjective rates are significantly higher at Wadi ed-Daliyeh compared with Nahal Hever, which in turn has significantly higher rates than Murabba'at. By contrast, particle rates are significantly lower at Wadi ed-Daliyeh compared with Nahal Hever, which in turn has significantly lower rates than Murabba'at. Since several of the morphological criteria identified for Qumran scrolls relate to pronouns and pronominal suffixes, it might be expected that, given the low rates across all sites of these parts of speech, that these criteria will not perform so well for the non-Qumran Judean Desert texts. Again, this is a further example where, had discriminatory criteria been focused on an external correlate text type, different criteria would be likely to be more appropriate.

	Site	Mean	Lower 95% confidence interval	Upper95% confidence interval
Nouns	Wadi ed-Daliyeh	470	420	520
	Murabba'at	386	318	454
	Nahal Hever	490	431	549
	Wadi Sdeir	468		
Verbs	Wadi ed-Daliyeh	130	99	160
	Murabba'at	144	103	186
	Nahal Hever	39	63	135
	Wadi Sdeir	91		
Pronominal suffixes	Wadi ed-Daliyeh	107	85	129
	Murabba'at	96	66	126
	Nahal Hever	99	72	125
	Wadi Sdeir	104		
Adjectives	Wadi ed-Daliyeh	157	135	179
	Murabba'at	80	50	110
	Nahal Hever	119	93	145
	Wadi Sdeir	91		
Particles	Wadi ed-Daliyeh	102	63	141
	Murabba'at	258	205	311
	Nahal Hever	161	116	207
	Wadi Sdeir	234		
Pronouns	Wadi ed-Daliyeh	35	24	45
	Murabba'at	35	21	49
	Nahal Hever	32	19	44
	Wadi Sdeir	13		

Table 8.15 Means and 95% confidence intervals of rates per 1,000 words for parts of speech by site in the non-Qumran Judean Desert texts.

Table 8.16 gives the rates per 1,000 words for the 22 morphological criteria identified as discriminatory for the Qumran Aramaic scrolls. As a first step, the place of Sdeir2 can be considered to ascertain whether it can be included with one of the other sites. Sdeir2 has no instances of seven of the 22 criteria that also do not occur in any of the texts at other sites: hitpeel stems, perfect 3mp ending -w', retained aleph before suffix, 2ms

Criterion	Wadi ed-Daliyeh	Murabba'at	Nahal Hever	Wadi Sdeir
Determined nouns	129	60	121	78
Afel	0	3.6	3.0	0
Hafel	23.4	2.6	0	0
Itpeel	2.6	0	0.4	0
Hitpeel	0	0	0	0
<i>m</i> -prefixed Peal infinitives	0	3.2	9.9	0
Peal infinitives	0	3.2	9.9	0
<i>l</i> prefixed imperfects	0	5.2	0	0
<i>y</i> prefixed imperfects	29	7.9	4.6	13
3fp verbs	0	3.8	0	0
Perfect 3mp ending <i>w</i>	15	0	0	0
Perfect 3mp ending <i>w'</i>	0	0	0	0
Retained aleph before suffix	0	0	0	0
2ms pronominal suffix <i>kh</i>	0	0	0	0
3fs pronominal suffix <i>h'</i>	0	0	0	0
Relative particles	41	53	43	39
Relative particle <i>dy</i>	0	38	38	39
Direct object marker <i>yt</i>	0	1.0	1.2	0
<i>dnh</i> or <i>dn'</i> forms	0.5	7.1	13	13
Demonstrative pronoun <i>'lh</i>	9.9	0	0	0
Demonstrative pronoun <i>'lyn</i>	0	0	0	0
Demonstrative pronoun <i>'ln</i>	0	0	0	0

Table 8.16 Rates per 1,000 words for the 22 morphological criteria occurring in the non-Qumran Judean Desert texts by site.

pronominal suffix *kh*, 3fs pronominal suffix *h'*, demonstrative pronoun *'lyn*, and demonstrative pronoun *'ln*. In addition, Sdeir2 has no instances of six of the criteria as does Wadi ed-Daliyeh: afel stems, *m*-prefixed peal infinitives, total peal infinitives, *l*-prefixed imperfects, third person plural feminine verbs, and direct object marker *yt*. Similarly, Sdeir2 has no instances of two of the criteria as does Murabba'at: itpeel stems, demonstrative pronoun *'lh*. Finally, Sdeir2 has no instances of four of the criteria as does Nahal Hever: hafel stems, *l*-prefixed imperfects, third person plural feminine verbs, and demonstrative pronoun *'lh*. This leaves five criteria which do occur in Sdeir2. Sdeir2 determined noun rate falls within the 95% confidence intervals of Murabba'at texts only. The *y*-prefixed imperfect verb rate falls within the 95% confidence intervals of both Murabba'at and Nahal Hever. Sdeir2 relative particle rate falls within the 95% confidence intervals of the means of all three other sites, but the *dy* relative particle form falls within the 95% confidence intervals of Murabba'at and Nahal Hever only. Sdeir2 *dnh* or *dn'* rate also falls within the 95% confidence intervals of Murabba'at and Nahal Hever only. Hence, in terms of positive occurrences, Sdeir2 resembles both Murabba'at and Nahal Hever, whilst for absence of criteria forms it is more like Wadi ed-Daliyeh texts. In summary, excluding the seven criteria which are uniformly absent in all texts, Sdeir2 shares six rates with Wadi ed-Daliyeh, seven with Murabba'at, and eight with Nahal Hever. Furthermore, of the seven criteria rates where Sdeir2 is zero and Nahal Hever's mean is positive, the Nahal Hever lower 95% confidence interval includes zero for three: afel stems, itpeel stems, and direct object marker *yt*. Hence, Sdeir2 is similar to Nahal Hever for 18 of the 22 criteria. In view of this, Sdeir2 can be considered along with Nahal Hever texts for practical purposes.

Criterion	Wadi ed-Daliyeh	South Judean Desert texts	P-value
Determined nouns	129 (102, 155)	94 (71, 117)	0.052
Afel	0	3.0 (0.6, 5.5)	0.11
Hafel	23 (19, 28)	1.0 (0, 4.6)	<0.001
Itpeel	2.6 (0.9, 4.2)	0.2 (0, 1.6)	0.037
<i>m</i> -prefixed Peal infinitives	0	6.5 (1.8, 11.3)	0.075
Peal infinitives	0	6.5 (1.8, 11.3)	0.075
<i>l</i> prefixed imperfects	0	2.1 (0, 4.5)	0.26
<i>y</i> prefixed imperfects	29 (17, 42)	6.5 (0, 17.6)	0.01
3fp verbs	0	1.5 (0, 3.2)	0.22
Perfect 3mp ending -w	15 (13, 18)	0	<0.001
Relative particles	41 (30, 53)	47 (37, 57)	0.46
Relative particle <i>dy</i>	0	38 (28, 48)	<0.001
Direct object marker <i>yt</i>	0	1.0 (0, 2.2)	0.24
<i>dnh</i> or <i>dn'</i> forms	0.5 (0, 5.8)	10 (6.0, 15)	0.007
Demonstrative pronoun <i>'lh</i>	9.9 (6.9, 12.9)	0	<0.001

Table 8.17 Mean (95% confidence intervals of the mean) rates per 1,000 words for the 15 of the 22 *a priori* morphological criteria occurring in the non-Qumran Judean Desert texts comparing Wadi ed-Daliyeh to the three other sites grouped together as south Judean Desert texts. P-values <0.05 indicate significant differences between text types. 95% confidence intervals are given for non-zero means with zero fixed as the lowest value for the lower confidence interval.

Table 8.16 shows some heterogeneity of rates for many of the criteria, but whether this is statistically significant remains to be tested. *A priori*, Murabba'at and Nahal Hever

texts would be expected to be morphologically closer than either with Wadi ed-Daliyeh texts. Comparing means with general linear models shows that of the 14 criteria where some instances occur, there are significant differences between Murabba'at and Nahal Hever rates for determined nouns ($p=.040$) only. However, once the differences in rates for all nouns were adjusted for, this difference is no longer significant ($p=.13$). Hence, Murabba'at and Nahal Hever texts can be considered homogeneous in terms of the 22 morphological criteria.

Combining Murabba'at and Nahal Hever texts with Sdeir2 allows for a more accurate estimate of any differences between these texts with the earlier texts from Wadi ed-Daliyeh. Table 8.17 gives means, 95% confidence intervals and p-value for difference for the Wadi ed-Daliyeh, which is well to the north of Qumran, and the southern sites text groups. There are significant differences in rates for seven of the 15 criteria considered and a statistical trend⁶ towards difference for a further three. There are no statistical differences between rates for five criteria. These findings support two morphologically distinct text types in the non-Qumran Judean Desert Aramaic texts: 1) earlier texts from Wadi ed-Daliyeh and 2) later texts from the Judean Desert south of Qumran. Earlier texts have more hafe! stems, *y*-prefixed imperfect verbs, third masculine plural perfect verb *-w* endings and demonstrative pronoun *'lh* forms, and fewer *dnh* or *dn'* forms, and *dy* forms of the relative particle (tends to be *zy* instead) than later texts. There are also more itpeel stems in the earlier texts.

⁶ A statistical trend is a technical term for a p-value between 0.05 and 0.1, that is between a 1 in 10 and 1 in 20 probability of happening by chance.

Non-Qumran Judean Desert Aramaic texts and their relationship to Qumran Aramaic scrolls

The previous section established that there are two morphological non-Qumran Judean Desert Aramaic text types: 1) from Wadi ed-Daliyeh to the north-west of Jericho is thought to originate from a Samaritan community around the end of the fourth century BCE; 2) texts from caves in valleys of water courses flowing into the Dead Sea south of Qumran are thought to date to the first and early second centuries CE. Some of these texts are letters from supporters of Simon Bar Kokhba. Many of these texts are of limited length so that around one third of the morphological criteria discriminating between Qumran Aramaic text types are not present in any text. This means that type 2 statistical error is likely: if more of these texts had survived, we could be more certain about rates of infrequently occurring forms. Nevertheless, the finding that seven of the a priori morphological criteria differed significantly between these dateable text groups and that a further three showed a statistical trend, supports the use of these criteria in investigating possible heterogeneity within the Aramaic corpus found at Qumran which is likely to have arisen between the time of Wadi ed-Daliyeh and Bar Kokhba.



Figure 8.16 Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with the Qumran Aramaic scrolls using standardised rates of the 22 *a priori* criteria.

Figure 8.16 shows the hierarchical cluster dendrogram using the *22 a priori* morphological criteria with all Qumran Aramaic scrolls and non-Qumran Judean Desert Aramaic scrolls of adequate length. First, it is notable that all Wadi ed-Daliyeh texts, except WDSP 10, cluster together separately from other texts. WDSP clusters towards the end of the agglomeration schedule and hence is rather separate from all the other texts. Closest to the Wadi ed-Daliyeh cluster are 1Q24, 4Q214, 4Q214b, 4Q548 and 8Hev/Se8a. The dendrogram with only aligned Qumran texts (Figure 8.17) confirms this relationship, but here XHev/Se8a is no longer part of this cluster. These scrolls have already been identified as a ‘Targumic’ text type. That this text type has affinities with texts from a Samaritan community at the end of the 4th century BCE as well as to the Targums is of note since it indicates that the features are unlikely to reflect solely diachronic influence. Inspection of the dendrogram for non-aligned texts shows that no further Qumran scrolls appear to relate to this group.

Figure 8.17 Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls aligned to a text type using standardised rates of the *22 a priori* criteria (over page).

Unlike Wadi ed-Daliyeh texts, the south Judean Desert texts do not cluster together. Three texts, Mur18, Sdeir2 and XHev/Se12 cluster at a low level with 1Q21, 2Q24, 4Q210, 4Q242, 4Q554, 5Q15 and 11Q18 (Figure 8.16). This is essentially a New Jerusalem text type except for the inclusion of 5Q15 which, though a New Jerusalem text, is in a different internal Qumran text type. Again this cluster is confirmed looking at aligned Qumran scrolls only (Figure 8.17), but appears separate from non-aligned scrolls (Figure 8.18). Two texts, XHev/Se9 and XHev/Se50 cluster at a slightly higher level with the Biblical group, 4Q112, 4Q113, 4Q115 and 4Q117. In the dendrogram with aligned scrolls only, this cluster is joined by XHev/Se8 and XHev/Se8a. These south Judean Desert texts do not cluster with any non-aligned Qumran scrolls (Figure 8.18). Finally, Mur21, Mur25 and XHev/Se8 cluster with 4Q244 (Pseudo-Daniel) and 4Q550 (Proto-Esther). These are joined by 4Q529 (Words of Michael) in the non-aligned dendrogram (Figure 8.18).

1Q24, 4Q214, 4Q214b and 4Q548 have already been text typed as Targumic.

Comparing rates for the 22 criteria variables for these four scrolls with those for the Wadi ed-Daliyeh texts shows significant differences for *afels* ($p=.009$), *hafels* ($p=.007$), relative particles ($p=.001$), *dy* relative particle form ($p=.001$), *'lh* ($p=.007$), and perfect third masculine plural *-w* ending ($p=.007$). Hence there are significant differences for six criteria compared with seven criteria for Targum Onkelos and ten for Targum Jonathan. Hence the label of 'Targumic' text type may need to be modified: this cluster of Qumran Aramaic scrolls bears similarities with both Targums and the Wadi ed-Daliyeh texts that originate some centuries before. The tendency for the

WDSP4	78	↓↘	
WDSP7	81	↓□	
WDSP8	82	↓↕↓↘	
WDSP9	83	↓□ □↓↘	
WDSP6	80	↓↕ ↕ ↕	
WDSP1	75	↓✕↓↕ □↓↓↓↓↓↓↓↓↓↓↓↘	
WDSP5	79	↓↕ ↕	↕
WDSP2	76	↓↘ ↕	↕
WDSP15	85	↓↕↓↓↓↕↕	↕
WDSP3	77	↓↕	↕
4Q206	15	↓↓↓↓✕↓↓↓↓↓↘	□↓↓↓↕
4Q207	16	↓↓↓↕	↕ ↕
1Q23	3	↓↓↓↓↓↓✕↓↘ ↕	↕ ↕
4Q203	12	↓↓↓↓↓↕ ↕ ↕	↕ ↕
4Q535	38	↓↓↓↓✕↓↘ ↕ ↕	↕ ↕
4Q550a	54	↓↓↓↕ ↕ ↕ □↓↓↓↓↓↓↓↕ ↕	
Mur18	86	↓↘ ↕ ↕ ↕	↕
XHevSe12	97	↓↕↓↓↓↕↓↕↓□ ↕	↕
Sdeir2	92	↓↕ ↕ ↕ ↕	↕
1Q20	1	↓↘ ↕ ↕ ↕	↕
4Q542	45	↓↕↓↕↘ ↕ ↕ ↕	↕
4Q201	10	↓□ ↕ ↕ ↕ □↓↕	↕
4Q530	34	↓↕ □↓□ ↕	□↓↓↓↕↘
4Q202	11	↓↓↓↓□ ↕ ↕	↕ ↕
4Q243	29	↓↓↓↕ ↕ ↕	↕ ↕
4Q537	40	↓↓↓↓↓↕ ↕ ↕	↕ ↕
4Q242	28	↓↓↓↓✕↓↓↓↓□	↕ ↕
4Q560	63	↓↓↓↕ ↕	↕ ↕
4Q211	20	↓↓↓↓✕↓↓↓↓□	↕ ↕
XHevSe7	93	↓↓↓↕ ↕	↕ ↕
4Q214a	26	↓↓↓↓↓↓↓↓↕↕	↕ ↕
4Q550	53	↓↓↓↘	↕ ↕
XHevSe8	94	↓↓↓↓↕↓↘	↕ ↕
4Q529	33	↓↓↓↕ □↓↘	↕ ↕
4Q244	30	↓↓↓↓↓↕ □↓↓↓↓↓↓↓↘ ↕ ↕	
Mur21	89	↓↓↓↓↓↓✕↓↕ □↓↓↓↓↓↕ ↕ ↕	
Mur25	90	↓↓↓↓↓↕ ↕	↕
4Q562	65	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘	□↓↓↓↓↓↓↓↓↓↓↓↕
4Q532	36	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓	↓↓↓↓□ ↕
4Q541	44	↓↓↓↓↓↓✕↓↓↓↓↓↘	↕ ↕
11Q10	69	↓↓↓↓↓↕ ↕	↕ ↕
4Q536	39	↓↓↓↘ □↓↓↓↓↓↓↓↓↓↓↘	↕ ↕
4Q556	59	↓↓↓↕↓↕ ↕	↕ ↕
4Q212	21	↓↓↓↕ □↓↓↓↘ ↕	↕ ↕
4Q198	9	↓↓↓↓↓↕ □↓↕	↕ ↕
4Q534	37	↓↓↓↓↓↓✕↓↓↓↕ □↓↓↓↕	↕
4Q540	43	↓↓↓↓↓↕ ↕	↕
4Q549	52	↓↓↓↓↓↓✕↓↓↓↓↓↓↓↓↘	↕
4Q559	62	↓↓↓↓↓↕ ↕	↕ ↕
4Q570	66	↓✕↓↓↓↕ ↕	↕ ↕
Mur19	87	↓↕ ↕ □↓↓↓↓↓↓↓↕	↕ ↕
4Q538	41	↓✕↓↘ □↓↓↓↕ ↕	↕ ↕
6Q8	68	↓↕ ↕ ↕ ↕	↕ ↕
4Q550c	55	↓↓↓↕↓↕ ↕ ↕	↕ ↕
XHevSe13	98	↓↓↓↕ □↓↓↓↕	↕ ↕
4Q561	64	↓↓↓↓↓↓↓↓↓↓↓□	↕
↕↓↓↓↓↓↓↓↘			
Mur20	88	↓↓↓↓↓↓↓↓↓↕	↕ ↕
↕			
4Q156	6	↓↓↓↓✕↓↓↓↓↓↓↓↘	↕ ↕
↕			

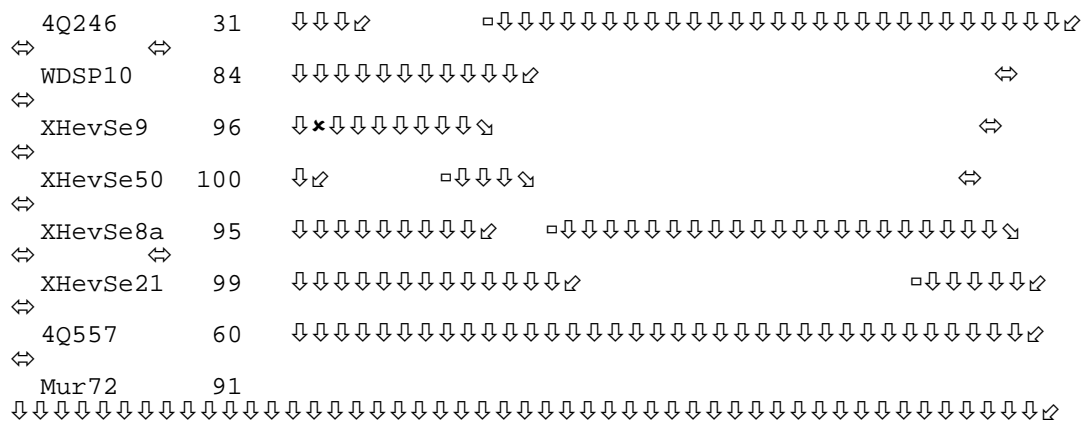


Figure 8.18 Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls not aligned to any text type using standardised rates of the 22 *a priori* criteria.

Targums to use archaic forms has already been noted, and this tendency may offer one explanation of this finding.

There is no clear southern Judean Desert Aramaic text type emerging with these morphological criteria. In other words, despite these southern Judean Desert texts being a relatively homogeneous group in themselves, when forced into a cluster analysis with Qumran texts, they do not remain as a tightly defined group, unlike the Wadi ed-Daliyeh texts. Just a few texts appear close to previously identified Qumran Aramaic text types. Mur18 is dated to 55CE, Sdeir2 and XShev/Se12 are undated. The three texts together comprise only 199 words and similarities may well be attributed to the relatively low variable mean rates of nearly all the criteria in the New Jerusalem text type. XHev/Se9 at 290 words and XHev/Se50 at 114 words are more substantial. Adding XHev/Se8 at 268 words and XHev/Se8a at 174 words provides four texts with over 800 words in total that cluster, albeit not tightly, with the Biblical Aramaic scrolls. All four texts are deeds of sale, one being dated to the days of Bar Kokhba. This suggests, along with the similarities between Targums and the Wadi ed-Daliyeh texts that the morphological criteria have only a limited diachronic component at best. XHev/Se8 also clusters with Mur21 and Mur25 to form a group with the non-aligned 4Q244, 4Q529 and 4Q550. Mur25 is dated to 133 CE and is, like XHev/Se8 dated in the days of Bar Kokhba, a deed of sale. Mur21 is an undated marriage contract. It is unclear whether this group represents a new, distinct text type or is an artifact of the clustering process. This is where confirmatory analyses are required to test whether having this extra text type helps to explain heterogeneity in Aramaic morphology found at Qumran. Nevertheless,

it is worthy of note that the Qumran texts relate closely to just a few of the Aramaic texts deposited not that far away in the Judean desert at a time not that long after the *terminus ad quem* of Qumran. By contrast, one Qumran text type appears fairly closely related to a presumed Samaritan text type of the fourth century BCE.

Babatha Archive Aramaic Texts

	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns	Total words
pYadin 7	650	155	144	45	726	41	1870
pYadin 8	64	4	4	7	29	6	119
pYadin 10	39	12	5	0	12	1	70
pYadin 17	39	12	5	0	12	1	70
pYadin 18	46	9	5	0	9	1	70
pYadin 20	31	10	2	0	7	2	52
pYadin 42	68	19	15	4	53	6	172
pYadin 47A	27	12	7	2	32	8	92
pYadin 47B	37	11	9	8	36	10	116

Table 8.18 Absolute counts of parts of speech and total words for the CAL Babatha Archive Texts with more than 50 words. (N.B. Total words also include number of determined nouns).

Finally, Babatha Archive texts in Aramaic, not included in the Accordance Judean-T module, are available from the Comprehensive Aramaic Lexicon project. The project includes these within the Palestinian Aramaic category. The texts were deposited in the Cave of Letters at Nahal Hever and are dated 124-132 CE. As noted in Chapter 5, some

texts contain duplicate sections. Nine texts contain at least 50 words. Frequencies of principle parts of speech are shown in Table 8.18; rates per 1,000 words (Table 8.19) are comparable to other Judean Desert non-Qumran Aramaic texts (Table 8.14).

	Nouns	Verbs	Pronominal suffixes	Adjectives	Particles	Pronouns
pYadin 7	348	83	77	24	388	22
pYadin 8	538	34	34	59	233	50
pYadin 10	557	171	71	0	171	14
pYadin 17	557	171	71	0	171	14
pYadin 18	657	129	71	0	129	14
pYadin 20	596	192	38	0	135	38
pYadin 42	395	110	87	23	308	35
pYadin 47A	293	130	76	22	348	87
pYadin 47B	319	95	78	69	310	85

Table 8.19 Rates for parts of speech per 1,000 words for the CAL Babatha Archive Texts with more than 50 words. (N.B. Total words also include number of determined nouns).

Table 8.20 gives the rates per 1,000 words for the 22 morphological criteria identified as discriminatory for the Qumran Aramaic scrolls. Comparing Table 8.20 with Table 8.16, the Babatha archive texts have lower rates of determined nouns than either Wadi ed-Daliyeh or South Judean texts, higher rates of Afels, Peal infinitives and direct object marker *yt*, but otherwise are closer in profile to the South Judean texts than those from Wadi ed-Daliyeh.

	pYd 7	pYd 8	pYd 10	pYd 17	pYd 18	pYd 20	pYd 42	pYd 47A	pYd 47B
Determined nouns	58	42	14	14	0	0	41	43	43
Afel	5.4	0	29	29	0	0	5.8	0	0
Hafel	0	0	0	0	14	0	0	0	0
Itpeel	4.3	0	0	0	0	0	0	0	0
Hitpeel	0	0	0	0	0	0	0	0	0
<i>m</i> -prefixed Peal infinitives	13	0	0	0	0	0	5.8	11	8.6
Peal infinitives	13	0	14	14	0	0	5.8	11	8.6
<i>l</i> prefixed imperfects	0	0	0	0	0	0	0	0	0
<i>y</i> prefixed imperfects	7.0	8.4	0	0	0	0	0	11	0
3fp verbs	0.5	0	0	0	0	0	0	0	0
Perfect 3mp ending <i>w</i>	0	0	0	0	0	0	0	0	0
Perfect 3mp ending <i>w'</i>	0.5	0	0	0	0	0	0	0	0
Retained aleph before suffix	0	0	0	0	0	0	0	0	0
2ms pronominal suffix <i>kh</i>	0	0	0	0	0	0	0	0	0
3fs pronominal suffix <i>h'</i>	0	0	0	0	0	0	0	0	0
Relative particles	43	0	14	14	14	38	23	54	43
Relative particle <i>dy</i>	47	8	14	14	14	77	29	33	43
Direct object marker <i>yt</i>	6.4	0	14	14	14	0	0	11	0
<i>dnh</i> or <i>dn'</i> forms	5.4	8.4	0	0	0	0	17	33	17
Demonstrative pronoun ' <i>lh</i>	0	8.4	0	0	0	0	0	0	0
Demonstrative pronoun ' <i>lyn</i>	1.1	0	0	0	0	0	0	0	0
Demonstrative pronoun ' <i>ln</i>	0	0	0	0	0	0	0	0	0

Table 8.20 Rates per 1,000 words for the 22 morphological criteria occurring in the Babatha archive Aramaic texts.

A multivariate analysis of variance comparing the Babatha archive Greek texts with the Legal texts found no overall significant difference ($F=0.43$, $p=.83$) with only rates for

determined nouns ($p=.006$) differing significantly on univariate comparisons. In view of this, the Babatha archive texts can be considered as a homogeneous text type. Figure 8.19 shows the non-standardised hierarchical cluster dendrogram for all Qumran Aramaic scrolls with 50 words or more together with the Babatha archive texts. The Babatha archive texts, unlike the South Judean texts, form a fairly tight cluster with pYd 10,17 and 18 closely related to Q549 and Q559 and pYd7, 8, 42, 47A and 47B related to 4Q196, 4Q197, 4Q213, 4Q538, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547, 4Q550c, 4Q570 and 6Q8. pYd20 appears quite separate.

The hierarchical clustering indicates that the Babatha archive texts are textually close to the Tobit, Visions of Amram text type, members being 4Q196, 4Q197, 4Q213, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547. Interestingly 4Q538 (Testament of Judah), 4Q550c (proto-Esther), 4Q570 (Aramaic R) and 6Q8 (Giants), all previously unaligned, are also included in this group. Multivariate analysis of variance finds the difference between the Tobit and Babatha groups to be just non-significant ($F=16.3$, $p=.059$). However, when pYd20 is excluded, so that the Babatha group are more homogeneous, the difference with the Tobit group becomes statistically significant ($F=36.7$, $p=.027$). Univariate statistics indicate that this difference is largely due to different rates of the object marker *yt* (Tobit lower), *m*-Peal infinitives (Tobit higher, though total Peal infinitives are similar), *dnh* or *dn'* (absent in Tobit group), feminine 3rd person plural verbs (Tobit higher despite Babatha archive texts relating to a woman), *l*-prefixed imperfects (absent in Babatha),

pYd10 103 ↕ ✖ ↕ ↕

itpeels (Tobit higher) and masculine 3rd person plural *-w* endings absent in Babatha). Apart from itpeels, the other differences would be consistent with the Tobit text type being older than the Babatha archive. Similarities between the two text types may indicate that the Babatha texts, which are of a legal nature, retain archaic features and/or that Babatha and Tobit text types share a similar geographical provenance.

Figure 8.20 shows the standardised hierarchical cluster dendrogram for all Qumran Aramaic scrolls with 50 words or more together with the Babatha archive texts. It is very similar to the non-standardised dendrogram. pYd7, 10, 17, 18, 20, 42 and 47B cluster closely together with texts from the Qumran Aramaic Tobit text type group; pYd8 and 47B do not cluster with the rest of the Babatha archive texts. However, other Qumran Aramaic text types, though not the Biblical one, are also interspersed. This probably reflects a loss of information when relatively short texts have variables standardised.

pYd10	77	↓×↓↓↓↓↓↓↓↓
pYd17	78	↓↗ □↓↓↓↓↓↓
pYd18	79	↓↓↓↓↓↓↓↓↓↗ ↔
4Q213a	23	↓↓↓↓×↓↗ ↔
4Q539	42	↓↓↓↓↗ □↓↓↓↓↓↓↓↓↓
1Q20	1	↓↓↓↓↓↗ ↔
4Q156	6	↓↓↓↓↓×↓↓↓↗ ↔
4Q246	31	↓↓↓↓↓↗ □↓↓↓↗ ↔
4Q212	21	↓↓↓↓↓×↓↗ ↔ ↔
4Q550a	54	↓↓↓↓↓↗ □↓↗ ↔
4Q540	43	↓↓↓↓↓↓↓↗ ↔ ↔
4Q529	33	↓↓↓↓↓×↓↓↓↗ ↔ ↔
4Q550	53	↓↓↓↓↓↗ □↓↗ ↔
pYd20	80	↓↓↓↓↓↓↓↓↓↗ ↔ ↔ ↔
pYd42	81	↓×↓↓↓↓↓↗ ↔ ↔ ↔
pYd47B	83	↓↗ □↓↗ ↔ ↔ ↔
4Q214	25	↓↓↓↓↓×↓↗ ↔ ↔ ↔ □↓↗
4Q243	29	↓↓↓↓↓↗ ↔ □↓ ↔ ↔
4Q208	17	↓×↓↗ ↔ ↔ ↔ ↔
4Q209	18	↓↗ □↓↗ ↔ ↔ ↔ ↔
4Q213b	24	↓×↓↗ ↔ ↔ ↔ □↓ ↔
4Q543	46	↓↗ ↔ ↔ ↔ ↔ ↔
4Q211	20	↓↓↓↓↓↓↓↗↗ ↔ ↔ ↔ ↔
1Q21	2	↓×↓↓↓↓ ↔ □↓↗ ↔ ↔
11Q18	70	↓↗ ↔ ↔ ↔ ↔ ↔
2Q24	5	↓×↓↗ ↔ ↔ ↔ ↔ ↔
4Q210	19	↓↗ ↔ ↔ ↔ ↔ ↔
1Q24	4	↓×↓↗↗↗ ↔ ↔ ↔ ↔
4Q554	58	↓↗ ↔ □↓ ↔ ↔
4Q552	56	↓↓↓↗ ↔ ↔ ↔ ↔
4Q318	32	↓↓↓↓↓↓↓↓ ↔ ↔ ↔
4Q560	63	↓↓↓↓↓↓↓↗ ↔ ↔ ↔
4Q542	45	↓↓↓×↓↓↓↗ ↔ ↔ ↔
4Q547	50	↓↓↓↗ ↔ ↔ ↔ ↔
4Q201	10	↓↓↓↓↓↓↓↓ ↔ ↔ ↔
4Q538	41	↓×↓↗ ↔ ↔ ↔ □↓↓↗
6Q8	68	↓↗ □↓↗ ↔ ↔ ↔ ↔
4Q550c	55	↓↓↓↗ □↓↗↗ ↔ ↔ ↔
4Q530	34	↓↓↓↓↓↗ ↔ ↔ ↔ ↔
4Q544	47	↓↓↓×↓↗ ↔ ↔ ↔ ↔
4Q546	49	↓↓↓↗ ↔ ↔ ↔ ↔
4Q553	57	↓×↓↗ □↓ ↔ ↔ ↔
4Q558	61	↓↗ ↔ ↔ ↔ ↔ ↔
4Q196	7	↓×↓ ↔ ↔ ↔ ↔ ↔
4Q213	22	↓↗ □↓↗ ↔ ↔ ↔ ↔
4Q197	8	↓↓↓ ↔ ↔ ↔ □↓↗
pYd7	75	↓↓↓↗ ↔ ↔ ↔ ↔ ↔
4Q204	13	↓↓↓↗ ↔ ↔ ↔ ↔ ↔
4Q205	14	↓↓↓↗↗ ↔ ↔ ↔ ↔ ↔
4Q531	35	↓↓↓↗ □↓↗ ↔ ↔ ↔ ↔
4Q198	9	↓↓↓↓↓↗ ↔ ↔ ↔ ↔ ↔
4Q545	48	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↗ ↔ ↔ ↔
11Q10	69	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↗ ↔ ↔ ↔
4Q537	40	↓↓↓↓↓×↓↗ ↔ ↔ ↔
4Q561	64	↓↓↓↓↓↗ □↓↓↓↓↓↓↓↓↓↗ ↔ ↔
4Q570	66	↓↓↓↓↓↓↓↗ ↔ ↔
4Q556	59	↓↓↓↓↓↓↓↓↓↓↓↓↓↓×↓↓↓↓↓↓↓↓↗ ↔
4Q562	65	↓↓↓↓↓↓↓↓↓↓↓↓↗ ↔
4Q535	38	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓
4Q242	28	↓↓↓↓↓×↓↓↓↓↓↓↓↓↓ □↓↗
5Q15	67	↓↓↓↓↓↗ □↓↓↓↓↓↓↓↓↓↓↓ ↔
4Q548	51	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↗ ↔ ↔

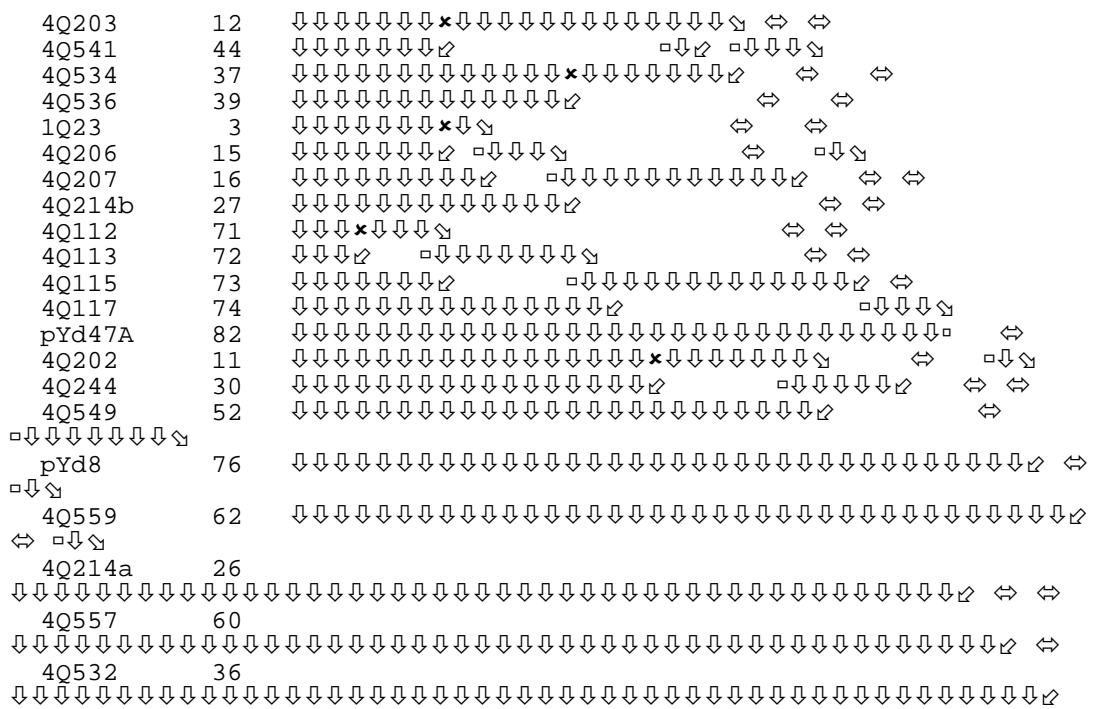


Figure 8.20 Hierarchical clustering dendrogram of the non-Qumran Judean Desert Aramaic Texts together with Qumran Aramaic scrolls not aligned to any text type using standardised rates of the 22 *a priori* criteria.

Summary of exploratory analyses of internal and external Aramaic text types

Exploratory analyses of morphological variables have suggested six distinct Qumran Aramaic text types plus a residual non-aligned Qumran Aramaic group. These can be provisionally described as:

Qumran Biblical

Qumran Targumic

Qumran Enochic-Aramaic Levi

Qumran Enochic-Four Kingdoms

Qumran New Jerusalem

Qumran Tobit-Visions of Amram

Qumran non-aligned

External text correlation identified only one of these groups, Targumic, beyond what was found by analyses of the Qumran texts on their own. The Babatha archive helped to confirm the integrity of the Tobit group. To examine how these different text types relate to each other, post hoc Scheffe's test⁷ was performed. This test is statistically

⁷ Scheffe's test seeks to establish homogeneous sets within a factor (here text type) in an analysis of variance. In essence establishing where the mean rate of a textual feature overlaps in terms of 95% confidence intervals for any set. Hence any text type can be a member of more than one set because its mean rate may overlap with both a lower mean rate of another text type and with a higher mean rate of yet another text type.

conservative with regards to detecting differences between groups. The post hoc tests identify the statistically significant text features that differentiate the text types. This is helpful to move from a description of the text types based on the kinds of texts they each encompass towards a description based on textual characteristics.

Aphels per 1000 words

Chosen	N	Subset	
		1	2
Scheffe ^{a,b} Wadi ed-Daliyeh	11	.0000	
Egyptian typical	34	.2118	
Egyptian atypical	30	2.0595	
South Judean Desert	15	3.0215	
Biblical	4	5.9454	5.9454
New Jerusalem	5	6.8747	6.8747
Babatha	9	7.5894	7.5894
Tobit-Visions of Amram	8	9.7456	9.7456
Qumran non-aligned	40	12.8999	12.8999
Enochic-Four Kingdoms	6	13.8007	13.8007
Targumic	5	15.9456	15.9456
Enochic-Aramaic Levi	6		20.1761
Sig.		.103	.244

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 59.192.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.21 Scheffe's test identifying homogeneous sets for Afel rates in the various text types. Sample sizes and mean rates per 1,000 words for each text type are displayed.

Univariate statistics showed that feminine 3rd plural verbs, masculine 2nd singular masculine suffix *-kh*, feminine 3rd singular suffix *-h'* and 3rd masculine plural perfect

ending *-w*' were all non-significant between all text types. Despite all other variables showing significant univariate statistically significant differences, Scheffe's test, which

Haphels per 1000 words

Chosen	N	Subset		
		1	2	3
Scheffe ^{a,b,c} Enochic-Aramaic Levi	6	.0000		
Enochic-Four Kingdoms	6	.0000		
Egyptian typical	34	.7451		
South Judean Desert	15	1.0417		
New Jerusalem	5	1.1765		
Tobit-Visions of Amram	8	1.3217		
Targumic	5	1.3793		
Babatha	9	1.5873		
Qumran non-aligned	40	3.0867		
Egyptian atypical	30	7.8540	7.8540	
Biblical	4		17.1372	17.1372
Wadi ed-Daliyeh	11			23.3734
Sig.		.723	.450	.930

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 32.133.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.22 Scheffe's test identifying homogeneous sets for Hafel rates in the various text types.

adjusts for multiple comparisons, found differences for afels (Table 8.21), hafels (Table 8.22), relative particles (Table 8.23), *dy* (Table 8.24), *yt* (Table 8.25), determined state nouns (Table 8.26), *dnh* or *dn'* (Table 8.27), *'lh* (Table 8.28), and hitpeels (Table 8.29) only. Of these, *'lh* and hitpeels select out only one text type as distinct each. For the other morphological criteria, note that text types can belong to more than one group; this

indicates that they are of a mixed type that might, for example, represent a diachronic or geographical transition.

Relative particles per 1000 words

Chosen	N	Subset		
		1	2	3
Scheffe ^{a,b} Egyptian typical	34	8.0078		
Egyptian atypical	30	9.8778	9.8778	
Enochic-Aramaic Levi	6	13.6712	13.6712	
Targumic	5	14.0764	14.0764	
Tobit-Visions of Amram	8	23.0174	23.0174	23.0174
Babatha	9	27.2007	27.2007	27.2007
Qumran non-aligned	40	31.4124	31.4124	31.4124
New Jerusalem	5	33.1109	33.1109	33.1109
Enochic-Four Kingdoms	6		40.8880	40.8880
Wadi ed-Daliyeh	11		41.4502	41.4502
South Judean Desert	15			47.1084
Biblical	4			49.9459
Sig.		.344	.061	.231

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 208.457.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.23 Scheffe's test identifying homogeneous sets for relative particle rates in the various text types.

dy per 1000 words

Chosen	N	Subset			
		1	2	3	4
Scheffe ^{a,b} Wadi ed-Daliyeh	11	.0000			
Egyptian typical	34	.0000			
Egyptian atypical	30	.0000			
Enochic-Aramaic Levi	6	10.2980	10.2980		
Targumic	5	14.0764	14.0764	14.0764	
Tobit-Visions of Amram	8	22.6119	22.6119	22.6119	22.6119
Qumran non-aligned	40		30.0023	30.0023	30.0023
Babatha	9		31.0544	31.0544	31.0544
New Jerusalem	5		32.5703	32.5703	32.5703
Enochic-Four Kingdoms	6		37.3895	37.3895	37.3895
South Judean Desert	15			37.8473	37.8473
Biblical	4				49.9459
Sig.		.250	.056	.179	.050

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 150.749.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.24 Scheffe's test identifying homogeneous sets for *dy* rates in the various text types.

yt per 1000 words

Chosen	N	Subset	
		1	2
Scheffe ^{a,b} Biblical	4	.0000	
Wadi ed-Daliyeh	11	.0000	
Targumic	5	.0000	
Enochic-Four Kingdoms	6	.0000	
Egyptian typical	34	.0000	
Egyptian atypical	30	.0000	
Tobit-Visions of Amram	8	.2834	
Enochic-Aramaic Levi	6	.4151	
New Jerusalem	5	.4415	.4415
South Judean Desert	15	1.0341	1.0341
Qumran non-aligned	40	1.2905	1.2905
Babatha	9		6.6826
Sig.		1.000	.051

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 7.864.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.25 Scheffe's test identifying homogeneous sets for yt rates in the various text types.

Determined state nouns per 1000 words

Chosen	N	Subset			
		1	2	3	4
Scheffe ^{a,b} Babatha	9	28.4618			
Tobit-Visions of Amram	8	36.6741	36.6741		
Egyptian typical	34	38.7280	38.7280		
Enochic-Aramaic Levi	6	66.0071	66.0071	66.0071	
Qumran non-aligned	40	70.3690	70.3690	70.3690	
Enochic-Four Kingdoms	6	76.4576	76.4576	76.4576	76.4576
South Judean Desert	15	93.8749	93.8749	93.8749	93.8749
Egyptian atypical	30	95.3846	95.3846	95.3846	95.3846
New Jerusalem	5		102.8456	102.8456	102.8456
Wadi ed-Daliyeh	11			128.6106	128.6106
Targumic	5				138.2184
Biblical	4				142.6708
Sig.		.054	.061	.105	.061

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 914.983.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.26 Scheffe's test identifying homogeneous sets for determined state nouns rates in the various text types.

dnh or dna per 1000 words

Chosen	N	Subset	
		1	2
Scheffe ^{a,b,c} Enochic-Aramaic Levi	6	.0000	
New Jerusalem	5	.0000	
Tobit-Visions of Amram	8	.0000	
Egyptian typical	34	.0000	
Enochic-Four Kingdoms	6	.2612	.2612
Egyptian atypical	30	.3738	.3738
Wadi ed-Daliyeh	11	.4760	.4760
Qumran non-aligned	40	1.8698	1.8698
Targumic	5	3.7248	3.7248
Biblical	4	5.9746	5.9746
Babatha	9	9.0048	9.0048
South Judean Desert	15		10.5430
Sig.		.201	.065

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 22.338.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.27 Scheffe's test identifying homogeneous sets for *dnh* or *dn'* rates in the various text types.

alh per 1000 words

Chosen	N	Subset	
		1	2
Scheffe ^{a,b,c} Targumic	5	.0000	
Enochic-Aramaic Levi	6	.0000	
New Jerusalem	5	.0000	
Tobit-Visions of Amram	8	.0000	
South Judean Desert	15	.0000	
Egyptian typical	34	.0000	
Egyptian atypical	30	.0000	
Enochic-Four Kingdoms	6	.0723	
Qumran non-aligned	40	.3236	
Babatha	9	.9337	
Biblical	4	1.6811	
Wadi ed-Daliyeh	11		9.9003
Sig.		.993	1.000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = 4.234.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.28 Scheffe's test identifying homogeneous sets for 'lh rates in the various text types.

hithpe'els per 1000 words

Chosen	N	Subset	
		1	2
Scheffe ^{a,b} Wadi ed-Daliyeh	11	.0000	
Targumic	5	.0000	
Enochic-Aramaic Levi	6	.0000	
Enochic-Four Kingdoms	6	.0000	
New Jerusalem	5	.0000	
Tobit-Visions of Amram	8	.0000	
South Judean Desert	15	.0000	
Babatha	9	.0000	
Egyptian typical	34	.0000	
Egyptian atypical	30	.0000	
Qumran non-aligned	40	.0706	
Biblical	4		7.3593
Sig.		1.000	1.000

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = .295.

a. Uses Harmonic Mean Sample Size = 8.192.

b. Alpha = .05.

Table 8.29 Scheffe's test identifying homogeneous sets for hitpeel rates in the various text types.

Afels and Hafels

These forms comprise an alternate pair for the so-called H-stem conjugation. Afels are divided into low and high rates groups. Wadi ed-Daliyeh, the typical and atypical Ancient Egyptian documents and the Non-Qumran South Judean texts belong solely to the low rate group; Enochic-Aramaic Levi is solely in the high rate group. All the other text types belong to both low and high rate groups. There are three hafel groups: a high rate group comprising Wadi ed-Daliyeh and Qumran Biblical texts; a medium rate group

comprising the Qumran Biblical and Egyptian atypical texts; and a low rate group comprising all texts except Wadi ed-Daliyeh and Qumran Biblical. These groupings cannot be explained purely by a diachronic shift from hafel to afel. Wadi ed-Daliyeh is dateable c.332 BCE, the Ancient Egyptian texts to the 5th century BCE and the South Judean texts to early 2nd century CE. Dates for the Biblical texts are disputed, depending on to what extent they are considered to represent original Babylonian and Persian Imperial documents.

Proportion of Aphels

Scheffe^{a,b,c}

Chosen	N	Subset				
		1	2	3	4	5
Wadi ed-Daliyeh	11	.0000				
Egyptian typical	34	.0330				
Egyptian atypical	30	.0948	.0948			
South Judean Desert	15	.2380	.2380	.2380		
Biblical	4	.2852	.2852	.2852	.2852	
Babatha	9	.4031	.4031	.4031	.4031	.4031
Targumic	5	.5312	.5312	.5312	.5312	.5312
Qumran non-aligned	40	.6357	.6357	.6357	.6357	.6357
New Jerusalem	5		.7057	.7057	.7057	.7057
Tobit-Visions of Amram	8			.7813	.7813	.7813
Enochic-Four Kingdoms	6				.9118	.9118
Enochic-Aramaic Levi	6					.9471
Sig.		.066	.098	.240	.077	.238

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = .086.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.30 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of afels within total afels and hafels plus one.

Assuming that Afel and Hafel represent potential alternatives, the proportion of afels and hafels to the total of afels and hafels plus one (to avoid division by zero) can provide an index of Afel/Hafel tendency for each text. This controls for genre, although there is little evidence from Tables 8.21 and 8.22 that absolute afel or hafel rates relate to genre. Table 8.30 shows the Scheffe's ranges' groupings for the proportion of Afels. This confirms the impression that this criterion is not a simple diachronic correlate. Similarly Table 8.31 shows the Scheffe's ranges' groupings for the proportion of Hafels. These are less frequent which probably explains the smaller number of homogeneous groups. Nevertheless, the results largely mirror the afel proportion findings. The overlap of groups indicates that Afel and Hafels are relatively limited as single criteria to discriminate between Qumran text types, although they are able to identify Qumran Biblical texts.

Proportion of Haphels

Scheffe^{a,b,c}

Chosen	N	Subset		
		1	2	3
Enochic-Aramaic Levi	6	.0000		
Enochic-Four Kingdoms	6	.0000		
Targumic	5	.0483	.0483	
South Judean Desert	15	.0627	.0627	
Tobit-Visions of Amram	8	.1035	.1035	
Babatha	9	.1038	.1038	
Egyptian typical	34	.1104	.1104	
Qumran non-aligned	40	.1569	.1569	
New Jerusalem	5	.1709	.1709	
Egyptian atypical	30	.5288	.5288	.5288
Biblical	4		.6699	.6699
Wadi ed-Daliyeh	11			.8738
Sig.		.263	.074	.882

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = .084.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.31 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of hafels within total afels and hafels plus one.

Relative particles and *dy*

Although not completely identical, Tables 8.23 and 8.24 indicate that rates for both *dy* and relative particles are highly correlated which may indicate these features are being influenced by genre. Against this is the separation of South Judean texts from the Babatha archive, and the separation of various Enochic text types. Nevertheless, *dy* is

dy_prop

Scheffe^{a,b,c}

Chosen	N	Subset	
		1	2
Wadi ed-Daliyeh	11	.0000	
Egyptian typical	34	.0000	
Egyptian atypical	30	.0000	
Enochic-Aramaic Levi	6	.5973	.5973
Targumic	5	.7505	.7505
South Judean Desert	15	.8152	.8152
Qumran non-aligned	40	.8859	.8859
Enochic-Four Kingdoms	6	.8914	.8914
Tobit-Visions of Amram	8	.9430	.9430
New Jerusalem	5	.9527	.9527
Biblical	4	.9801	.9801
Babatha	9		1.8871
Sig.		.395	.051

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = .337.

- a. Uses Harmonic Mean Sample Size = 8.192.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.
- c. Alpha = .05.

Table 8.32 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion *dy* to relative particles plus one.

one potential form of the relative particle and, uncommonly, can also act as a conjunction instead. Table 8.32 shows Scheffe's ranges' groupings for the proportion of *dy* to relative particles plus one (again to avoid division by zero). Again, this fails to support a purely diachronic explanation with the Qumran Biblical texts being at the other end of the spectrum from those from Wadi ed-Daliyeh or Ancient Egypt. The overlap indicates that this criterion is probably of limited use to discriminate text types taken on its own for Qumran texts.

The object marker *yt*

Table 8.25 shows that there are two homogeneous groups for *yt*: a high rate group comprising New Jerusalem, South Judean, Qumran non-aligned and Babatha archive texts and a low rate group comprising all texts except those of the Babatha archive. The concurrence of South Judean and Babatha archive texts in a high rate group might be explained diachronically, but may also relate to the specific nature of legal texts. As with *dy*, *yt* is not a promising candidate to discriminate between Qumran Aramaic text types taken on its own.

Determined state nouns

As expected for a variable which has relatively high rates in most texts, determined state nouns divide into four homogeneous groups (Table 8.26). The Qumran Biblical and Targumic text types have the highest rates, not belonging to any lower rate groups. The Babatha archive texts have low rates of determined state nouns. Moreover texts from Wadi ed-Daliyeh also have high rates. The profile from the raw rates runs against any presumed diachronic trend for an increased tendency towards the use of determined nouns over time. Although determined state nouns are able to separate two of the Qumran Aramaic text types from the others, the overlap between lower rate groups indicates that this criterion cannot be used on its own to discriminate adequately.

Table 8.33 shows the Scheffe's ranges' homogeneous groups for the proportion of determined nouns within total nouns. Again, the purpose is to correct for genre because some text types typically have a higher proportion of nouns. The correction has very little effect on the homogeneous groups with Babatha and typical Egyptian texts, from the opposite ends of the time window, both having low proportions of determined state

Proportion of determined nouns

Scheffe^{a,b,c}

Chosen	N	Subset			
		1	2	3	4
Babatha	9	.0759			
Egyptian typical	34	.0802			
Tobit-Visions of Amram	8	.1438	.1438		
South Judean Desert	15	.2071	.2071		
Enochic-Aramaic Levi	6	.2310	.2310		
Qumran non-aligned	40	.2632	.2632		
Egyptian atypical	30	.2738	.2738	.2738	
Wadi ed-Daliyeh	11	.2769	.2769	.2769	
Enochic-Four Kingdoms	6		.3470	.3470	
New Jerusalem	5		.3495	.3495	
Biblical	4			.4838	.4838
Targumic	5				.5705
Sig.		.092	.074	.060	.984

Means for groups in homogeneous subsets are displayed.

Based on Type III Sum of Squares

The error term is Mean Square(Error) = .009.

a. Uses Harmonic Mean Sample Size = 8.192.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = .05.

Table 8.33 Scheffe's post hoc ranges test indicating Aramaic text type that form homogeneous groups for proportion of determined state nouns within total nouns.

nouns. The typical Egyptian and Wadi ed-Daliyeh texts appear in the middle which would militate against any hypothesis that determined state flourished then waned over time.

Demonstrative pronoun *dnh* or *dn'*

Table 5 reveals that these forms do not occur in four of the text types: Qumran Enochic-Aramaic Levi, Qumran New Jerusalem, Qumran Tobit-Visions of Amram, and atypical Ancient Egyptian texts. Non-Qumran South Judean texts belong solely to the high rate group and Babatha archive texts also have high rates. The high rates in the most recent dateable texts would support a diachronic effect, but this would not explain how three of the Qumran text types have zero rates, significantly lower than the Qumran Biblical text type. The split indicates that this criterion is promising with regard to distinguishing between text types but, like the other criteria, not sufficient on its own.

Evidence for local dialects?

Having found little support for a purely diachronic explanation of different text type morphologies at Qumran in these exploratory analyses, an alternative model of persistent local dialects needs to be considered. Some of the external texts can be reliably attributed to specific regions: the typical Ancient Egyptian texts, the South Judean and Babatha archive texts, and the Wadi ed-Daliyeh texts representing a Samaritan provenance. It has already been noted that, although the South Judean and

Babatha archive texts are contemporaneous and were found in the same region, there are significant differences between them consistent with them representing two distinct dialects; it is possible that the scribes who wrote these documents came from different regions even though they were working in Judea. We cannot, therefore, label these text types as “Judean” or “Palestinian” Aramaic except in terms of where the texts were discovered and the subjects with which they are concerned.

As an initial step, those Qumran morphological criteria that differentiate these external texts can be reviewed from Tables 8.21 through Table 8.33, bearing in mind that morphological criteria specifically designed to discriminate between these texts may have been different from the ones chosen for the Qumran corpus. First, three features which are absent or at low rates in the typical Ancient Egyptian and Wadi ed-Daliyeh texts, but at relatively high rates in the South Judean and Babatha archive texts may represent diachronic shift: the particles *dy* and *yt*, and the demonstrative pronoun *dnh* or *dn'*. This leaves the presence of *'lh* as a specific feature of Wadi ed-Daliyeh texts which also have a relatively high rate of determined nouns, whilst Babatha archive texts have a low rate of determined state nouns. In addition Wadi ed-Daliyeh has a high proportion of *hafels* (Table 8.31) not seen in the older typical Egyptian texts.

Turning to the Qumran text types we can see that the Biblical text type also has a relatively high rate of *'lh*, determined nouns and *hafels*, features typical of Wadi ed-Daliyeh texts with a known Samaritan provenance. This would be suggestive of both text types sharing a similar “dialect”. Links between the Babatha archive and the

Qumran Tobit-Visions of Amram have already been noted above, even though these texts are likely to have been written more than two centuries apart. Similarly, a Qumran Targumic group has been identified, though the provenance of Targum Onkelos is uncertain. Neither typical Egyptian texts nor those from the South Judean Desert have distinguishing features amongst the chosen morphological criteria beyond those that might be expected due to diachronic shift. We would need to look beyond the 22 morphological criteria chosen for the clustering of Qumran texts if we wished to identify features of a distinct dialect. In the next chapter syntactic features are examined to see whether these support the different text types emerging on the basis of morphological criteria.

9. Description of syntactic features of Aramaic scrolls

The number of different syntactic features considered as discriminating between different Aramaic text types is far more limited than morphological features. As noted in Chapter 4, there is an overlap between morphology and syntax so that it is sometimes difficult to be categorical as to what is a morphological and what is a syntactic feature. To illustrate this, the form of the relative particle, *dy*, has already been considered as a morphological feature. In this Chapter its use relative to the use of a construct chain will be considered as a syntactic feature. This assumes that in those texts where *dy* and *d-* are both used as relative particles, they have the same syntactic value and the choice is simply one of morphological style, whereas the choice between using a relative particle or a construct chain is one of syntactic style.

Verbal stem	Biblical scrolls	Other scrolls
Afel	0	42
Hafel	19	11
Itpaal	0	10
Hitpaal	0	0
Itpeel	3	10
Hitpeel	1	0
Pael	3	44
Peal	8	238

Table 9.1 Counts of infinitives in the Qumran Aramaic scrolls for the common stem forms.

Of the syntactic features considered as *a priori* discriminators, several relate to the infinitive. Table 9.1 shows the numbers of infinitives in the Biblical scrolls (4Q112,

4Q113, 4Q115, 4Q117) and the other Aramaic Qumran scrolls. There are no infinitive absolute forms in Aramaic; hence syntactic forms that rely on infinitive absolutes in Hebrew, such as the emphatic use of an infinitive absolute before an infinitive construct, will be of no use in discriminating between text types. Indeed, apart from the peal infinitive, other forms are relatively uncommon and are likely to lack statistical power if they are distributed across a wide range of scrolls. Morphological features of the peal infinitive have already been included in the exploratory morphological analyses in Chapter 7.

Lemma	Frequency
<i>'bd</i>	20
<i>Hzh</i>	12
<i>'kl</i>	10
<i>nsb</i>	9
<i>qwm</i>	8
<i>'mr</i>	7
<i>bnh</i>	7
<i>twb</i>	7
<i>'lp</i>	6
<i>'th</i>	6
<i>Hwh</i>	6
<i>Ht'</i>	6
<i>qtl</i>	6
<i>hwk</i>	5
<i>qbr</i>	5

Table 9.2 Frequencies of the 15 of the 85 lemmas that occur at least five times as a peal infinitive in the non-biblical Qumran Aramaic scrolls.

The numbers are relevant to any consideration of verbal rection. One feature that relates to this is which preposition is associated with infinitives. Since different lemmas may be

expected to prefer different prepositions (e.g. even in broad terms, transitive and intransitive verbs), it is important to know if some lemmas occur frequently enough to have adequate statistical power to distinguish between text types. Table 9.2 shows the frequencies of lemmas for peal infinitive forms in the non-biblical Aramaic scrolls. Even the commonest lemma, ‘*bd*’, occurs only 20 times rendering preposition association statistically inadequate as a discriminator between text types.

Stem	1 st	2 nd	3 rd	4 th	5 th	6 th	>6 th
Afel	1	1	0	0	2	1	35
Hafel	3	0	1	2	2	0	21
Itpaal	0	1	1	0	0	0	8
Itpeel	0	0	0	1	0	0	12
Pael	0	3	1	1	3	3	32
Peal	1	12	12	11	11	7	177
Total	5	17	15	15	18	11	285

Table 9.3 Word position in sentence of infinitive verbs according to stem in the Aramaic Qumran scrolls.

The frequencies that infinitive verbs occupy different places in the sentence are shown in Table 9.3. The Table shows that for all stems infinitives tend to occur later in longer sentences, often this is in dependent clauses. The data may be adequately powered to investigate specific sentence positions if these are grouped together (e.g. first or second position), but the distribution suggests that categorising infinitives according to whether they occur in the main clause of the sentence or in subordinate clauses is more likely to be adequately powered for statistical analyses. These data need to be treated with a little caution for fragmentary texts where it is not always possible to decide whether an infinitive verb is within a subordinate clause because the words which might be expected

to indicate this are absent. In such cases a default of main clause was scored. The number of infinitive verbs in main clauses can also reflect texts with a relatively greater paratactic clause structure. However, a tendency to paratactic clauses can, itself, be considered a syntactic feature so that it can be thought of as contributing variance to the infinitive verb main clause counts.

	1	2	3	4	5	6	7	8	9	10+
1Q20	0	8	1	2	2	1	0	0	2	26
1Q21	0	1	1	1	1	2	0	0	0	0
1Q23	0	0	0	0	0	0	1	0	0	0
1Q24	0	0	0	0	0	0	0	0	0	0
2Q24	0	1	0	0	0	0	0	1	0	1
4Q112	0	1	0	5	0	0	0	0	0	20
4Q113	0	0	0	1	0	1	0	1	0	10
4Q115	0	1	0	0	0	0	1	0	0	1
4Q117	0	0	0	0	0	0	0	0	0	0
4Q156	0	0	0	0	0	0	0	0	0	0
4Q196	0	2	3	0	0	0	1	2	0	8
4Q197	0	1	0	0	1	0	0	0	0	3
4Q198	0	0	0	0	0	0	0	0	0	0
4Q201	0	4	2	0	0	0	0	2	1	4
4Q202	0	3	1	0	0	2	1	0	1	4
4Q203	0	1	0	0	1	0	0	0	0	3
4Q204	0	4	0	3	1	1	1	1	1	22
4Q205	0	0	0	1	1	0	0	0	1	1
4Q206	0	2	0	1	0	1	0	0	0	6
4Q207	0	0	0	1	0	0	0	0	0	0
4Q208	0	0	0	0	1	0	0	0	0	0
4Q209	0	0	0	0	0	0	0	0	0	0
4Q210	0	0	0	0	0	0	0	0	0	0
4Q211	0	0	0	0	0	0	0	0	1	0
4Q212	0	2	0	0	0	0	0	0	1	1
4Q213	0	0	0	0	1	0	0	0	0	4
4Q213a	0	0	0	0	0	0	0	0	0	1
4Q213b	0	0	0	0	0	0	0	0	0	5
4Q214	0	0	0	1	0	0	0	0	0	1
4Q214a	0	0	0	0	1	0	0	0	0	0
4Q214b	0	0	0	0	0	0	0	0	0	0
4Q242	0	0	0	0	0	0	0	0	0	2
4Q243	0	0	0	0	0	0	0	0	0	1
4Q244	0	0	0	0	0	0	0	0	0	0
4Q246	0	0	0	0	0	0	0	1	0	0
4Q318	0	0	0	0	0	6	0	0	0	1
4Q529	0	1	0	0	0	0	0	0	0	1

4Q530	0	2	0	2	0	0	0	1	1	5
4Q531	0	0	0	0	0	0	0	0	0	0
4Q532	0	0	0	0	0	0	0	0	0	3
4Q534	0	0	1	1	1	1	0	0	0	6
4Q535	0	0	0	1	0	0	1	0	0	0
4Q536	0	0	0	1	0	0	0	0	0	2
4Q537	0	0	0	1	0	0	0	1	0	9
4Q538	0	0	0	0	0	0	0	0	0	1
4Q539	0	0	0	0	0	0	0	0	0	0
4Q540	0	0	0	0	0	0	0	0	0	0
4Q541	0	0	0	0	0	1	1	0	0	5
4Q542	0	0	0	0	0	0	1	0	0	6
4Q543	0	0	1	0	0	0	0	1	0	3
4Q544	0	0	2	0	0	0	0	1	1	5
4Q545	0	0	0	0	0	0	0	0	1	4
4Q546	0	0	0	0	0	1	1	0	0	5
4Q547	0	2	1	0	0	0	0	1	0	9
4Q548	0	0	0	0	0	1	0	1	0	3
4Q549	0	0	0	0	0	0	1	0	0	0
4Q550	0	1	0	0	0	0	0	0	0	0
4Q550a	0	0	0	0	0	0	0	0	0	0
4Q550c	0	0	0	0	0	0	0	0	0	0
4Q552	0	0	0	0	1	0	0	0	0	2
4Q553	0	0	0	0	0	0	0	0	0	0
4Q554	0	0	0	0	0	0	0	0	0	0
4Q556	0	0	0	0	0	0	0	0	0	1
4Q557	0	0	0	0	0	0	0	0	0	0
4Q558	0	0	1	0	0	0	0	0	0	0
4Q559	0	0	0	0	0	0	0	0	0	0
4Q560	0	0	0	0	0	0	0	0	0	0
4Q561	1	0	0	1	0	0	0	0	0	6
4Q562	0	1	0	0	1	0	0	0	0	0
4Q570	0	0	0	0	0	1	0	0	0	1
5Q15	0	0	0	0	0	0	0	0	0	0
6Q8	0	0	0	0	0	1	0	0	0	1
11Q10	0	1	0	1	2	0	2	0	0	3
11Q18	0	1	0	0	0	0	1	0	1	7

Table 9.4 Word position in sentence of *hwh* in the Aramaic Qumran scrolls.

The position of forms of the verb *hwh* has also been suggested as a syntactic criterion that may discriminate between different text types. Table 9.4 shows the frequency per 1,000 words in the Aramaic Qumran scrolls. No forms occur in the initial position and counts are fairly sporadic in other positions. Occurrences in the tenth position and

beyond have been combined because of this. It is unclear whether these data will be adequately powered to discriminate between scrolls without further combinations of positions. Moreover, it is unclear, due to varying clause length and structure, how any such combinations should be selected.

A more general syntactical feature that involves verbs is the position of the verb in the clause. Verbal position is flexible in Aramaic, though often verbs occur in the middle of clauses after subjects and before objects. However, verbs in an initial position, or preceded only by a particle, do occur typically mirroring the usual Hebrew Verb-Subject-Object order.¹ By contrast, verbs can occupy the final position in a clause as is routine for Akkadian prose with subject and object preceding any adjuncts and finally the verb (excluding participles).² Comparison of frequencies of verbs occurring at the beginning or end of clauses within texts is therefore one method. Clause level data for the place of words are only available for Qumran texts in Accordance. Table 9.5 shows the rates per 1,000 words of verbs at the beginning or ends of clauses. In general verbs occur more frequently at the beginning of clauses than at the end. There is a fair degree of heterogeneity indicating that this syntactic feature may be valuable for textual classification statistically.

¹ B. T. Arnold & J.H. Choi (2003) p169ff. discusses Hebrew verbal position at some length.

² J. Huehnergard (2005), p19.

Scroll	Verbs at beginning of clause	Verbs at end of clause
1Q20	11.2	1.9
1Q21	0	1.3
1Q23	1.9	0
1Q24	0	0
2Q24	4.0	0
4Q112	2.4	8.0
4Q113	5.0	9.2
4Q115	10.1	6.3
4Q117	3.6	3.6
4Q156	36.1	10.3
4Q196	9.6	1.8
4Q197	14.8	1.8
4Q198	11.4	0
4Q201	16.1	5.1
4Q202	13.7	1.1
4Q203	3.4	1.7
4Q204	9.2	4.4
4Q205	11.5	4.9
4Q206	7.3	4.6
4Q207	0	0
4Q208	13.6	3.9
4Q209	32.7	8.8
4Q210	9.8	1.6
4Q211	5.9	17.7
4Q212	6.5	6.6
4Q213	11.7	5.0
4Q213a	3.0	0
4Q213b	15.6	0
4Q214	3.6	3.6
4Q214a	0	0
4Q214b	0	8.2
4Q242	0	0
4Q243	1.2	0
4Q244	0	0
4Q246	9.8	14.6
4Q318	0	2.2
4Q529	13.5	0
4Q530	8.9	2.2
4Q531	1.3	0
4Q532	0	4.9
4Q534	2.0	9.8
4Q535	0	0
4Q536	0	14.5
4Q537	3.1	0
4Q538	16.1	10.8
4Q539	13.9	0
4Q540	0	0
4Q541	2.2	3.3
4Q542	5.4	1.8

4Q543	1.9	3.9
4Q544	14.1	8.8
4Q545	5.6	2.8
4Q546	0	1.2
4Q547	9.3	1.6
4Q548	3.6	3.6
4Q549	6.4	0
4Q550	25.0	0
4Q550a	0	0
4Q550c	8.8	4.4
4Q552	0	3.3
4Q553	2.9	0
4Q554	3.9	1.0
4Q556	0	2.0
4Q557	0	0
4Q558	0	0
4Q559	0	0
4Q560	0	0
4Q561	2.4	0
4Q562	0	0
4Q570	0	0
5Q15	9.1	1.8
6Q8	0	2.2
11Q10	27.4	13.4
11Q18	0	0

Table 9.5 Rates per 1,000 words of verbs as first word in the clause or preceded only by a particle together with verbs as the final word in the clause in the Qumran Aramaic scrolls.

Scroll	Nouns in construct state	Nouns followed immediately by relative particle
1Q20	98.5	9.5
1Q21	64.7	0
1Q23	61.2	0
1Q24	89.3	0
2Q24	72.9	10.4
4Q112	98.0	21.1
4Q113	65.3	11.2
4Q115	54.2	10.0
4Q117	156.8	38.1
4Q156	121.0	6.4
4Q196	135.7	2.9
4Q197	98.3	1.2
4Q198	99.2	16.5
4Q201	110.7	10.2
4Q202	107.6	7.9
4Q203	94.2	10.9
4Q204	92.0	11.7
4Q205	48.5	7.2
4Q206	62.1	11.6
4Q207	65.4	0
4Q208	36.8	0
4Q209	54.7	0
4Q210	61.2	5.7
4Q211	33.3	0
4Q212	140.2	5.7
4Q213	108.8	2.3
4Q213a	168.3	0
4Q213b	119.7	0
4Q214	75.9	0
4Q214a	52.6	0
4Q214b	81.4	5.8
4Q242	72.4	6.6
4Q243	125.5	0
4Q244	86.4	0
4Q246	122.8	0
4Q318	2.6	0
4Q529	106.7	16.9
4Q530	87.7	4.4
4Q531	67.4	1.6
4Q532	24.7	0
4Q534	145.8	5.8
4Q535	56.3	0
4Q536	117.0	16.0
4Q537	127.1	5.7
4Q538	95.6	7.4
4Q539	138.9	0
4Q540	44.1	14.7

4Q541	126.6	0
4Q542	104.9	10.7
4Q543	130.8	2.0
4Q544	114.2	8.0
4Q545	144.9	6.0
4Q546	123.9	2.9
4Q547	90.9	7.6
4Q548	159.4	0
4Q549	136.8	0
4Q550	132.7	10.2
4Q550a	173.1	19.2
4Q550c	72.6	4.0
4Q552	55.3	16.6
4Q553	48.9	0
4Q554	91.6	16.6
4Q556	64.3	14.3
4Q557	166.7	0
4Q558	57.4	8.2
4Q559	129.4	0
4Q560	40.4	10.1
4Q561	113.6	0
4Q562	52.1	31.3
4Q570	45.5	0
5Q15	105.9	11.2
6Q8	57.7	0
11Q10	104.7	0.8
11Q18	106.8	8.1

Table 9.6 Rates per 1,000 words of nouns in the construct state together with nouns in determined or absolute state followed by a relative particle in the Qumran Aramaic scrolls.

Both nouns in the construct state and nouns immediately followed by a relative particle occur frequently enough in the Aramaic Qumran corpus to be suitable for statistical analysis without further data reduction. Table 9.6 presents the rates per 1,000 words for the Qumran Aramaic scrolls. Figure 9.1 shows the scattergram, labelled by scroll, between nouns in construct state and nouns followed by a relative particle rates. Although there are quite a few zero values for nouns followed by a relative particle, there is still a fair spread. Construct chains are the commoner syntactical choice in all

the scrolls. Three scrolls appear to be outliers: 4Q117 (Biblical Ezra), 4Q550a (Proto-Esther) and 4Q562 (Aramaic D). The mean number of nouns in the construct state is 93.3 (standard deviation 37.9) per 1,000 words and these data did not depart significantly from a normal distribution (Kolmogorov-Smirnov statistic 0.093, $p=.18$). The rate for nouns followed by a relative particle did depart significantly from a normal distribution ($p<.001$) with a mean rate of 6.4 and a median of 5.9 per 1,000 words. Hence there were on average around 15 occurrences of a construct chain for every occurrence of a noun followed by a relative particle.

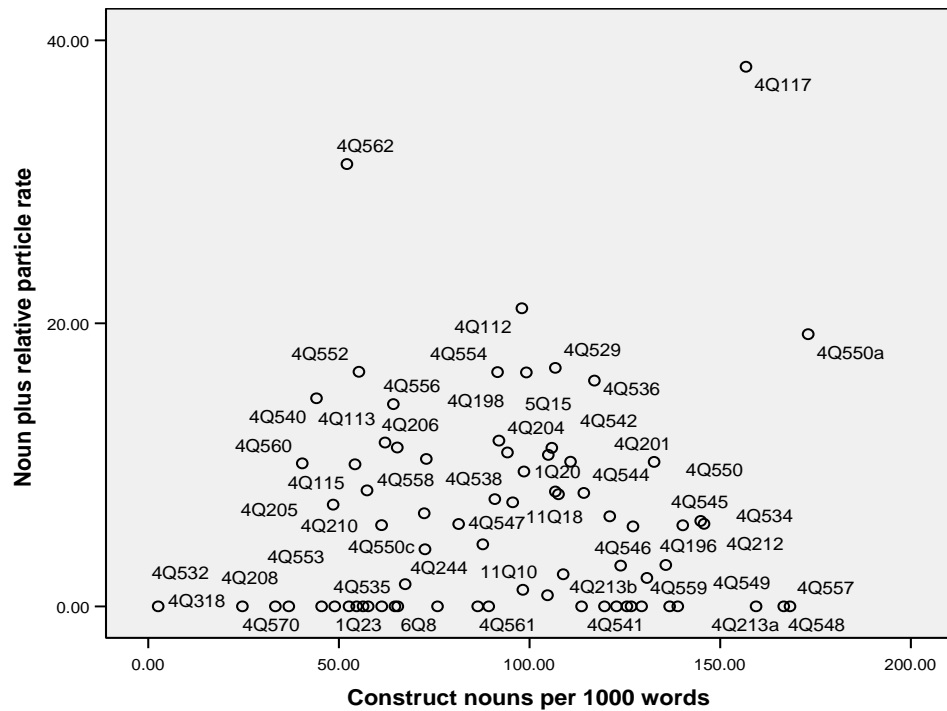


Figure 9.1 Scattergram of frequencies per 1,000 words of nouns in the construct state against nouns followed by a relative particle labelled by Qumran Aramaic scroll.

These syntactic features are also frequent in Aramaic texts chosen as external correlates. Table 9.7 shows frequencies of nouns followed immediately by a relative particle for the Pentateuch in Targum Onkelos, Targum Neofiti and Targum Pseudo-Jonathan and Table 9.8 the rates per 1,000 words for nouns in the construct state. The rates of nouns followed by a relative particle is fairly homogeneous across books and between Targums, with around twice the number than the average Aramaic Qumran scroll. By contrast,

	Onkelos	Neofiti	Pseudo-Jonathan
Genesis	16.8	17.7	19.5
Exodus	21.7	19.0	22.8
Leviticus	22.2	22.8	24.1
Numbers	24.7	26.7	26.6
Deuteronomy	24.4	22.6	26.6

Table 9.7 Rates per 1,000 words of nouns in determined or absolute state followed by a relative particle in Targums Onkelos, Neofiti and Pseudo-Jonathan.

rates of nouns in the construct state are on average around half of those in the Qumran corpus. Hence nouns followed by a relative particle is relatively more favoured as a syntactic construction and construct nouns relatively less favoured in the Targums compared with Aramaic Qumran scrolls.

	Onkelos	Neofiti	Pseudo-Jonathan
Genesis	36.0	22.7	36.8
Exodus	41.1	39.7	48.0
Leviticus	43.4	42.2	51.5
Numbers	55.8	59.5	65.1
Deuteronomy	29.4	42.1	43.4

Table 9.8 Rates per 1,000 words of nouns in the construct state in Targums Onkelos, Neofiti and Pseudo-Jonathan.

Several of the Ancient Egyptian Aramaic documents are quite short and this probably explains the considerable heterogeneity in rates of both nouns followed by a relative particle and nouns in the construct state (Table 9.9). Inspection of Table 9.9 shows that nouns followed by a relative particle are uncommon in this corpus whereas nouns in the construct state are present at rates comparable to the Targums though not as high as found at Qumran. The implication of these data is that there is a strong preference for construct chains over the use of relative particles compared to either Qumran Aramaic texts or the Targums. This indicates that this criterion may be useful for classifying the Aramaic Qumran scrolls as there are clear differences between distinct external correlate text types.

	Nouns in construct state	Nouns followed immediately by relative particle
A1	89.9	0
A2	27.8	0
A3	26.6	2.6
A4	46.0	7.5
A5	47.4	7.0
A6	23.7	1.6
B1	56.9	0
B2	67.3	1.3
B3	53.8	1.5
B4	67.5	0
B5	38.9	0
B6	68.3	1.8
B7	56.1	0
B8	19.9	0
C0	7.1	0
C1	31.7	0.4
C2	8.2	7.5
C3	71.0	0
C37A	54.4	0
C37B	39.3	0
C37C	52.0	0
C37D	47.2	0
C37E	49.3	0
C37K	44.6	0
C37G	64.5	0
C37F	78.7	0
C37J	51.8	0
C37JV	75.3	0
C37FV	74.3	0
C37GV	65.1	0
C37KV	44.3	0
C37EV	51.2	0
C37DV	52.2	0
C37BV	0	0
C37AV	133.3	0
C37H	23.8	0
C37L	71.4	0
C38	60.4	2.5
C3A	69.7	3.0
C4	170.1	0
D1	25.6	1.7
D2	37.4	0
D3	45.6	0
D4	21.2	0
D5	24.8	0
D6	3.3	0
D7	21.7	1.1
D8	50.6	3.7

D9	172.8	0
D10	0	0
D11	140.4	0
D12	236.2	0
D13	0	0
D14	90.9	0
D15	148.2	0
D16	0	0
D17	74.1	45.5
D18	223.7	14.3
D19	269.2	0
D20	102.8	9.8
D21	194.0	16.9
D22	83.1	3.3
D23	12.2	1.9
D24	145.3	8.5

Table 9.9 Rates per 1,000 words of nouns in the construct state together with nouns in determined or absolute state followed by a relative particle in the Ancient Egyptian documents.

Scroll	Passive participles	<i>t</i> -stem participles
1Q20	1.0	1.2
1Q21	0	0
1Q23	1.9	0
1Q24	4.6	0
2Q24	0	0
4Q112	4.0	1.4
4Q113	0.8	1.3
4Q115	3.8	0
4Q117	10.0	0
4Q156	0	0
4Q196	4.8	0
4Q197	2.8	0
4Q198	0	0
4Q201	0.7	5.2
4Q202	3.2	3.3
4Q203	3.4	0
4Q204	3.2	5.6
4Q205	3.3	9.8
4Q206	0.9	0.9
4Q207	0	8.9
4Q208	6.8	0
4Q209	11.1	0.3
4Q210	0	1.6
4Q211	5.9	17.7
4Q212	1.3	1.3
4Q213	1.7	1.7
4Q213a	0	3.0
4Q213b	0	0
4Q214	7.3	0
4Q214a	6.6	0
4Q214b	12.2	4.1
4Q242	15.5	0
4Q243	1.2	0
4Q244	0	0
4Q246	0	0
4Q318	0	0
4Q529	4.5	4.5
4Q530	3.3	3.3
4Q531	1.9	0.6
4Q532	4.9	0
4Q534	2.0	0
4Q535	0	15.9
4Q536	0	0
4Q537	4.7	1.6
4Q538	0	0
4Q539	0	0
4Q540	0	0
4Q541	3.3	3.3
4Q542	1.8	0

4Q543	0	2.0
4Q544	0	5.3
4Q545	0	1.4
4Q546	0	0
4Q547	0	3.2
4Q548	0	0
4Q549	0	0
4Q550	25.0	8.3
4Q550a	0	0
4Q550c	0	0
4Q552	0	0
4Q553	0	0
4Q554	2.9	0
4Q556	0	2.0
4Q557	0	0
4Q558	0.8	1.6
4Q559	0	0
4Q560	0	0
4Q561	4.8	0
4Q562	0	0
4Q570	0	0
5Q15	4.6	0
6Q8	0	0
11Q10	1.3	0.3
11Q18	2.1	0.5

Table 9.10 Rates per 1,000 words of all passive participles together with *t*-stem participles in the Qumran Aramaic scrolls.

Another syntactic choice that was suggested in Chapter 4 was that between the use of a passive participle and a *t*-stem verb, and for equivalence a *t*-stem participle. It is difficult to find an exact equivalence of a passive participle as this may act as a verbal adjective (unlike Akkadian, Aramaic does not have a separate form), infinitives can often carry the sense of a verbal participle, there is a wider range of passive stems in Aramaic (e.g. polel) and a passive sense can be intended by use of an impersonal third male plural active verb. Deciding on when each of these verbal forms could be an alternative to a

passive participle is a matter of opinion, hence here the analysis is restricted to participle forms. Although there may not always be a straight choice between a passive participle and a *t*-stem participle, some choice will contribute to the different proportions found in different texts and is thus open to statistical analysis. Table 9.10 presents the rates of passive participles and *t*-stem participles in the Aramaic Qumran corpus. The rates are relatively low but there appears to be adequate variance to allow statistical analysis.

	Onkelos		Neofiti		Pseudo-Jonathan	
	Passive participles	<i>t</i> -stem participles	Passive participles	<i>t</i> -stem participles	Passive participles	<i>t</i> -stem participles
Genesis	2.0	0.7	2.7	1.3	2.6	2.2
Exodus	3.3	0.8	4.6	1.4	6.3	2.6
Leviticus	2.7	1.2	2.6	2.7	4.1	3.9
Numbers	3.9	1.3	3.7	2.0	4.4	3.3
Deuteronomy	3.7	0.6	5.7	1.3	5.1	2.3

Table 9.11 Rates per 1,000 words of all passive participles together with *t*-stem participles in Targums Onkelos, Neofiti and Pseudo-Jonathan.

The rates for the Targums are shown in Table 9.11. Rates for *t*-stem participles in Targum Neofiti are about twice those in Targum Onkelos and those in Targum Pseudo-Jonathan about three times the Onkelos rates. Any trends for passive participles are less clear. Very few *t*-stem participles occur in the Ancient Egyptian documents whereas passive participle rates are relatively high (Table 9.12). Just as with construct chains versus relative particle use, these trends may be spatial and/or temporal in nature. Either

way, there appears enough differentiation between different external correlate texts to aid classification of the Qumran corpus.

	Passive participles	t-stem participles
A1	0	0
A2	3.5	0.9
A3	6.5	0
A4	6.7	0
A5	7.9	0
A6	5.1	1.2
B1	0	0
B2	2.0	0
B3	6.6	0
B4	7.9	0
B5	5.6	0
B6	1.5	0
B7	0	0
B8	9.9	0
C0	0	0
C1	1.6	0.6
C2	0.7	0
C3	0	0.2
C37A	36.2	0
C37B	28.1	0
C37C	42.2	0
C37D	36.0	0
C37E	39.4	0
C37K	32.3	0
C37G	29.1	0
C37F	21.0	0
C37J	34.6	0
C37JV	32.3	0
C37FV	20.0	0
C37GV	2.6	0
C37KV	0	0
C37EV	0	0
C37DV	4.4	0
C37BV	0	0

C37AV	0	0
C37H	23.8	0
C37L	53.6	0
C38	4.2	0
C3A	1.8	0
C4	0	0
D1	1.4	0
D2	3.3	0
D3	2.7	0
D4	0	0
D5	1.2	0
D6	2.2	0
D7	4.4	0.4
D8	1.4	0
D9	0	0
D10	0	0
D11	0	0
D12	0	0
D13	0	0
D14	0	0
D15	0	0
D16	0	0
D17	0	0
D18	0	0
D19	0	0
D20	0	0
D21	0	0
D22	0	0
D23	2.2	1.1
D24	0	0

Table 9.12 Rates per 1,000 words of all passive participles together with *t*-stem participles in the Ancient Egyptian documents.

Beyond choices of single syntactic features, alternative types of standard syntactic formulae may be worth examining. One example is that for vows, though these are not

as prominent in Aramaic literature beyond Qumran as they are in Hebrew.³ The usual root for vow beyond Qumran is *ndr*.⁴ However this only occurs once at Qumran in Q213b. However, three other verbs are found in the Aramaic Qumran scrolls that cover this semantic range: *ymh*, *š b'* and *Hrm*. *ymh* occurs twice in 1Q20, four times in 4Q201, once in 4Q202, once in 4Q204, once in 4Q530, once in 4Q531, and four times in 4Q560. *š b'* occurs once each in 4Q213a, 4Q543 and 4Q546. *Hrm* occurs three times in 4Q201 and once in 4Q202. The noun for oath is generally *mwmh* occurring once each in 1Q20, 4Q197, 4Q201, 4Q202 and 4Q530. The only other noun that occurs covering this semantic range is *š bw'h* occurring just once in 4Q221, one of the short fragments. Hence, just as in the non-Qumran Aramaic literature, vow formulae are too infrequent in Qumran Aramaic texts to be of statistical discriminatory value. It is unlikely that other formulaic phrases would serve any better.

In summary three main syntactic criteria emerge as candidates for statistical analyses that may be adequately powered to discriminate between different text types: 1) verbal position, with infinitives and the verb *hwh* possibly representing sub-categories, 2) use of construct state or relative particle *dy* and 3) use of passive participle or *t*-stem participle.

³ T. W. Cartledge (1992) p123ff.

⁴ T. W. Cartledge (1992) p127 notes that *nzr* is found in early Aramaic.

Correlations between syntactic features

	Mean	Median	Minimum	Maximum
Verbs at beginning of clause	6.4	3.6	0	36.1
Verbs at end of clause	3.2	1.8	0	17.7
Nouns in construct state	93.3	94.9	2.6	173.1
Nouns followed immediately by relative particle	6.4	5.7	0	38.1
Passive participles	2.5	0.9	0	25.0
<i>t</i> -stem participles	1.6	0	0	17.7

Table 9.13 Means, medians and ranges of syntactic variables rates per 1,000 words in the Qumran Aramaic corpus.

Table 9.13 provides summary descriptive values for rates per 1,000 words of verbal position (beginning or end of clause), construct state or relative particle *dy* and passive participle or *t*-stem participle in the Qumran Aramaic corpus. Table 9.14 shows Spearman (non-parametric) correlation coefficients between the three main syntactic features that emerge from descriptive data. The main feature is that the rate of verbs at the beginning of clauses correlates positively with the rates of verbs at the end of clauses, nouns in the construct state and *t*-stem participles. There are no other significant associations.

	Verbs at end of clause	Nouns in construct state	Nouns followed immediately by relative particle	Passive participles	<i>t</i> -stem participles
Verbs at beginning of clause	.39*	.24*	.12	.19	.24*
Verbs at end of clause		-.01	.13	.16	.19
Nouns in construct state			.03	-.08	.01
Nouns followed immediately by relative particle				.06	.13
Passive participles					.17

Table 9.14 Spearman correlation coefficients between syntactic variables in the Qumran Aramaic corpus. * indicates $p < .05$ (statistically significant association).

Figure 9.2 depicts a two-dimensional scaling of relationships between these syntactic variables; Young's S -stress=.00008 indicating good fit. In Figure 9.2 the rates of verbs at the end of clauses, *t*-stem active and passive participles are close with the other variables spread further apart. Nouns in construct state are quite separate on Dimension 1 which may relate to their far greater frequency in Qumran Aramaic texts compared to the other syntactic variables (Table 9.13). In view of this, cluster analysis is likely to be

more informative if standardised values of syntactic variables are used rather than absolute values, otherwise a single variable, nouns in a construct state, will far outweigh other clustering criteria.

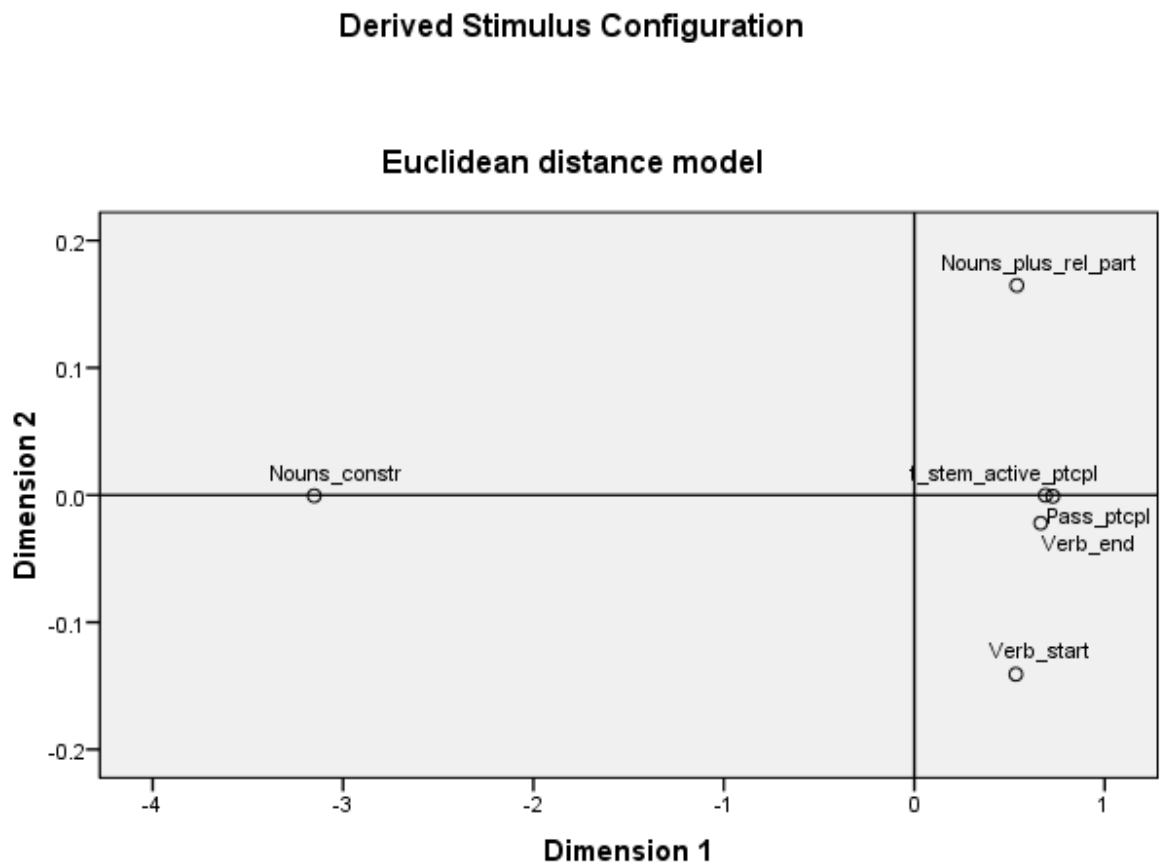
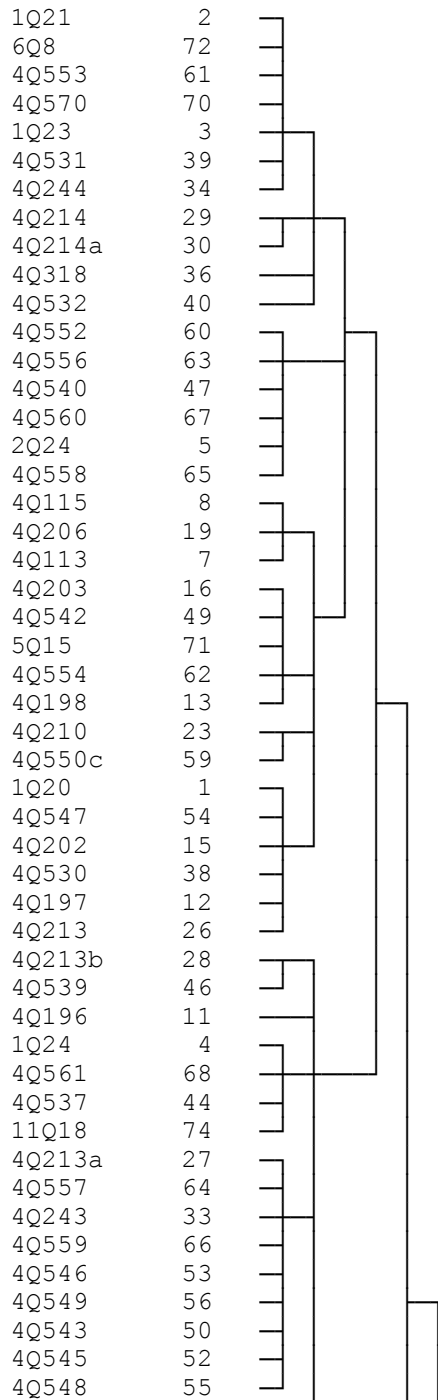


Figure 9.2 Two dimensional scaling of rates of syntactic features in the Qumran Aramaic scrolls.

Hierarchical clustering by syntactic features of the Qumran

Aramaic scrolls



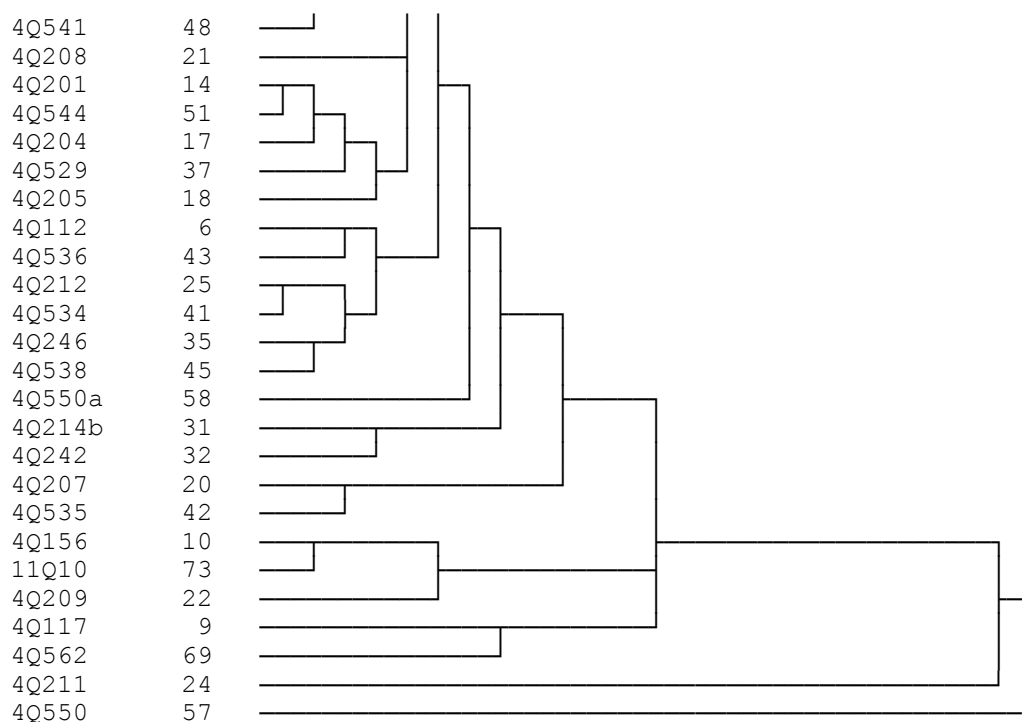


Figure 9.3 Hierarchical cluster dendrogram of the Qumran Aramaic scrolls using standardised values of the syntactic variables as clustering criteria.

Figure 9.3 shows the hierarchical cluster dendrogram of the Qumran Aramaic scrolls using standardised values of the syntactic variables as clustering criteria; standardised rates were chosen because of the large differences between the mean of the rate of nouns in the construct state compared with the mean rates of other syntactic features.

Inspection of the dendrogram shows clustering of fairly close proximity down to 4Q538.

Below 4Q538 scrolls agglomerate in ones or twos quite distant from each other, so are not really characterisable as low level clusters. Those that cluster well at a low level fall into the following groups:

- a) 1Q21 (Testament of Levi), 1Q23 (Enoch Giants), 4Q244 (Pseudo-Daniel), 4Q531 (Book of Giants), 4Q553 (Four Kingdoms), 4Q570 (Aramaic R), 6Q8 (Giants) plus, more loosely connected, 4Q214 (Aramaic Levi), 4Q214a (Aramaic Levi), 4Q318 (Brontologion), 4Q532 (Book of Giants)
- b) 2Q24 (New Jerusalem), 4Q540 (Apocryphon of Levi), 4Q552 (Four Kingdoms), 4Q556 (Book of Giants), 4Q558 (Vision), 4Q560 (Exorcism)
- c) 4Q113 (Daniel), 4Q115 (Daniel), 4Q206 (Enoch) plus 4Q198 (Tobit), 4Q203 (Enoch), 4Q542 (Testament of Qahat), 4Q554 (New Jerusalem), 5Q15 (New Jerusalem) plus 4Q210 (Enoch), 4Q550c (Proto-Esther) plus 1Q20 (Genesis Apocryphon), 4Q197 (Tobit), 4Q202 (Enoch), 4Q213 (Aramaic Levi), 4Q530 (Book of Giants), 4Q547 (Visions of Amram)
- d) 4Q213b (Aramaic Levi), 4Q539 (Apocryphon of Joseph B) plus 4Q196 (Tobit) plus 1Q24 (Enoch Giants), 4Q537 (Testament of Jacob), 4Q561 (Physiognomy/Horoscope), 11Q18 (New Jerusalem) plus 4Q213a (Aramaic Levi), 4Q243 (Pseudo-Daniel), 4Q543 (Visions of Amram), 4Q545 (Visions of Amram), 4Q546 (Visions of Amram), 4Q548 (Visions of Amram), 4Q549 (Work mentioning Hur and Miriam), 4Q557 (Vision) plus 4Q541 (Apocryphon of Levi)

These five low level clusters form a high level cluster (syntactic high level cluster A) together with two further high level clusters:

- B) 4Q201 (Enoch), 4Q204 (Enoch), 4Q205 (Enoch), 4Q529 (Words of Michael), 4Q544 (Visions of Amram)
- C) 4Q112 (Daniel), 4Q212 (Enoch), 4Q246 (Aramaic Apocalypse), 4Q534 (Noah), 4Q536 (Aramaic C), 4Q538 (Testament of Judah)

The remaining scrolls are not so closely connected in groups, though there are some interesting doublets. 4Q156 (Targum of Leviticus) and 11Q10 (Targum of Job) cluster together syntactically. 4Q207 (Enoch) and 4Q535 (Aramaic N) are also fairly close. The rest of the scrolls that fail to cluster are largely representatives of other groups (e.g. Enoch. Aramaic Levi) that do cluster with the exception of 4Q117, Biblical Esther.

The questions that follow the hierarchical clustering are how distinct are the different clusters syntactically? Is each low level cluster sufficiently distinct from the others to show statistically significant differences in rates of the various syntactical features or do they form a relatively homogeneous group? Is the real interest syntactically in the few scrolls that fail to cluster and are these really very different from the rest? An analysis of variance between cluster groups may enable us to answer these questions.

Analysis of variance by syntactic features between Qumran Aramaic scrolls clusters

Table 9.15 shows low level and high level cluster membership for the Qumran Aramaic scrolls according to syntactical membership by scroll numerical order as currently classified. Those scrolls that have nothing beside them in the Table do not cluster into any group. A multivariate general linear model with all six syntactical variables as dependent variables and low level cluster as the independent factor showed a significant difference between low level cluster groups ($F=12.5$, $p<.001$, partial $\eta^2=.640$). This means that cluster group explains 64% of the variance (variability) in the syntactic features' rates. Within the multivariate analysis, univariate statistics indicated that all syntactical features showed significant differences between groups except passive participles ($p=.27$) and *t*-stem participles ($p=.14$).

Scroll Low & High Level Clusters

1Q20	c	A
1Q21	a	A
1Q23	a	A
1Q24	d	A
2Q24	b	A
4Q112		C
4Q113	c	A
4Q115	c	A
4Q117		
4Q156		
4Q196	d	A
4Q197	c	A
4Q198	c	A
4Q201		B
4Q202	c	A
4Q203	c	A
4Q204		B
4Q205		B
4Q206	c	A
4Q207		
4Q208		
4Q209		
4Q210	c	A
4Q211		
4Q212		C
4Q213	c	A
4Q213a	d	A
4Q213b	d	A
4Q214	a	A
4Q214a	a	A
4Q214b		
4Q242		
4Q243	d	A
4Q244	a	A
4Q246		C
4Q318	a	A
4Q529		B
4Q530	c	A
4Q531	a	A
4Q532	a	A
4Q534		C
4Q535		
4Q536		C

4Q537	d	A
4Q538		C
4Q539	d	A
4Q540	b	A
4Q541	d	A
4Q542	c	A
4Q543	d	A
4Q544		B
4Q545	d	A
4Q546	d	A
4Q547	c	A
4Q548	d	A
4Q549	d	A
4Q550		
4Q550a		
4Q550c	c	A
4Q552	b	A
4Q553	a	A
4Q554	c	A
4Q556	b	A
4Q557	d	A
4Q558	b	A
4Q559		
4Q560	b	A
4Q561	d	A
4Q562		
4Q570	a	A
5Q15	c	A
6Q8	a	A
11Q10		
11Q18	d	A

Table 9.15 Low level and high level cluster membership according to syntactical features

Dependent Variable	Syntactical low level cluster group	Mean	N	95% Confidence Interval	
				Lower Bound	Upper Bound
Verbs at beginning of clause	a	0.9	11	-1.2	3.0
	b	0.7	6	-2.2	3.5
	c	9.0	16	7.2	10.8
	d	4.3	16	2.5	6.0
Verbs at end of clause	a	1.3	11	0.1	2.4
	b	0.9	6	-0.7	2.4
	c	2.9	16	1.9	3.8
	d	1.0	16	.09	2.0
Nouns in construct state	a	53.4	11	41.3	65.6
	b	55.7	6	39.3	72.2
	c	87.7	16	77.6	97.7
	d	132.1	16	122.1	142.2
Nouns followed immediately by relative particle	a	0.1	11	-1.8	2.1
	b	12.4	6	9.7	15.0
	c	8.8	16	7.2	10.5
	d	1.7	16	0.1	3.3
Passive participles	a	2.1	11	0.8	3.3
	b	0.1	6	-1.5	1.8
	c	1.9	16	0.9	2.9
	d	1.6	16	0.6	2.6
t-stem participles	a	0.06	11	-0.6	0.7
	b	0.6	6	-0.3	1.5
	c	1.0	16	0.5	1.6
	d	0.7	16	0.2	1.3

Table 9.16 Sample size (N), means and 95% confidence intervals of the means for each of the syntactical variables by low level cluster group in the Qumran Aramaic scrolls. Note negative values are 'theoretical'.

Table 9.16 shows means and 95% confidence intervals of the means for each of the syntactical variables by low level cluster group. Inspection of means and 95% confidence intervals allow identification of the key statistical differences between the low level cluster groups: where the mean of a group does not fall within the range of the 95% confidence intervals of another group, a statistically significant difference exists.

The following differences are apparent:

Verbs at the beginning of clauses – groups a and b are similar, groups c and d are distinct with group c having higher rates than group d, and group d having higher rates than groups a and b.

Verbs at the end of clauses – groups a, b and d are similar, group c has higher rates.

Nouns in construct state - groups a and b are similar, groups c and d are distinct with group c having higher rates than group d, and group d having higher rates than groups a and b.

Nouns followed immediately by relative particle – groups a and d are similar, groups b and c are distinct with group b having higher rates than group c, and group c having higher rates than groups a and d.

Passive participles – groups a, c and d are similar, group b rates are lower.

t-stem participles – groups c and d are similar and have higher rates than group a; group b is intermediate.

These comparisons allow the following group profiles:

Group a is generally similar to group b with low rates of most features except for nouns followed immediately by relative particle. Group a is next closest to group d and most distinct from group c.

Group b is, as noted, similar to group a, but has distinctively high rates of nouns followed immediately by relative particle and low rates of passive participles.

Group c has high rates of verbs at the beginning of clauses, verbs at the end of clauses and nouns in construct state. Group c has fairly high rates of nouns followed immediately by relative particle and *t-stem* participles.

Group d often has rates intermediate between group c and groups a and b, but has particularly high rates of nouns in the construct state.

The analysis of variance model has revealed significant heterogeneity within the high level cluster group A. Bearing this in mind, a similar multivariate model shows significant differences between high level cluster groups in terms of syntactical features

($F=17.0$, $p<.001$, partial $\eta^2=.657$). Again, this indicates that high level clusters explain 65.7% of total variance in rates of syntactical features. Univariate analyses showed that differences between groups on individual syntactical variables were significant except for nouns in construct state ($p=.18$) and passive participles ($p=.64$).

Dependent Variable	Syntactical high level cluster group	Mean	N	95% Confidence Interval	
				Lower Bound	Upper Bound
Verbs at beginning of clause	A	4.6	49	3.3	6.0
	B	12.9	5	8.6	17.21
	C	6.1	6	2.2	10.1
Verbs at end of clause	A	1.7	49	1.0	2.3
	B	4.6	5	2.6	6.7
	C	10.7	6	8.9	12.6
Nouns in construct state	A	90.6	49	80.3	100.8
	B	94.4	5	62.3	126.5
	C	119.9	6	90.6	149.2
Nouns followed immediately by relative particle	A	5.0	49	3.4	6.6
	B	10.8	5	5.7	15.9
	C	9.3	6	4.7	14.0
Passive participles	A	1.6	49	1.0	2.2
	B	2.34	5	0.6	4.1
	C	1.2	6	-0.4	2.9
t-stem participles	A	0.7	49	0.3	1.0
	B	6.1	5	5.0	7.1
	C	0.5	6	-0.5	1.4

Table 9.17 Sample size (N), means and 95% confidence intervals of the means for each of the syntactical variables by high level cluster group in the Qumran Aramaic scrolls. Note negative values are ‘theoretical’.

Table 9.17 shows means and 95% confidence intervals of the means for each of the syntactical variables by low level cluster group. Given that group A is heterogeneous, an initial comparison of groups B and C is appropriate. Group B has more verbs at the beginning of clauses than group C; by contrast group C has more verbs at the end of clauses than group B. There are no statistical differences between the two groups for nouns in construct state or nouns followed immediately by relative particle. Rates of passive participles also do not differ between the two groups, but group B has a higher rate of *t*-stem participles. Comparing groups B and C with rates in the lower level clusters of group A, we see that rates of verbs at the beginning of clauses are higher in Group B than all low level cluster groups, group C rates are higher than groups a, b and d but not group c. Both group B and C have higher rates of verbs at the end of clauses than any of the low level cluster groups in group A. For nouns in the construct state, Group C has higher rates than groups a, b and c; group B has higher rates than groups a and b. Groups B and C have similar rates of nouns followed immediately by relative particle to groups b and c, but higher than groups a and d. Apart from group B having higher rates than group b, there are no significant differences in passive participle rates between all the low and high level cluster groups B and C. Finally, group B has higher rates of *t*-stem participles than all group A low level cluster groups, whereas there are no significant differences between group C rates and the group A low level cluster groups.

The multivariate general linear models (multivariate analyses of variance) indicate that the hierarchical clustering procedure has identified clearly separable clusters both at low and higher levels. Moreover, this is not due to differences in rates across all syntactical

variables, such that some groups score higher on all measures compared to others, but there is a distinct patterning with some groups having lower rates of some features and higher rates of others compared to other groups. Such patterning suggests that the pairs of syntactical variables may represent stylistic choices. If this is the case then there should be an inverse relationship between these pairs. Such relationships might be hidden in overall statistics, such as bivariate correlation coefficients for all the Qumran Aramaic scrolls. They are also unlikely to be observed in cluster groups where the variance of rates is very low.

Relationships between syntactical variable pairs in syntactical cluster groups of the Qumran Aramaic scrolls

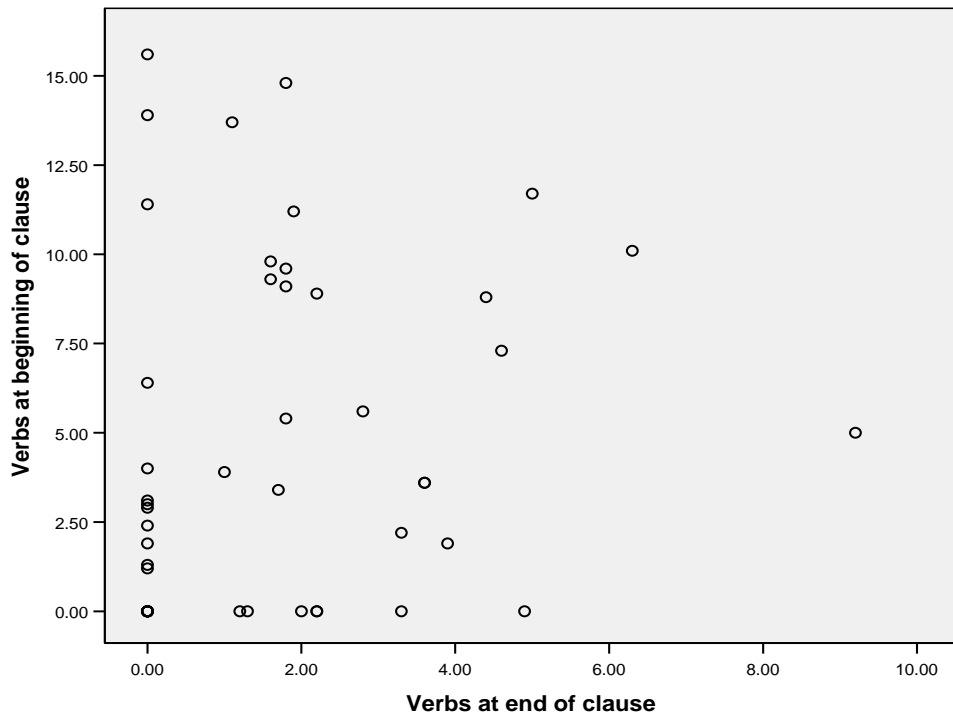


Figure 9.4 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group A of the Qumran Aramaic scrolls.

Figure 9.4 shows a scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group A. Of note are the significant differences between low level cluster groups for these two variables. Figure 9.4 shows that many scrolls in this cluster group have zero values for one or other of the variables and the distribution of the rest is far from elliptoid. We can see the effect of zero values more

clearly if we exclude low level cluster group c which has the highest mean rates for both features (Figure 9.5) where a possible inverse relationship is swamped by zero values.

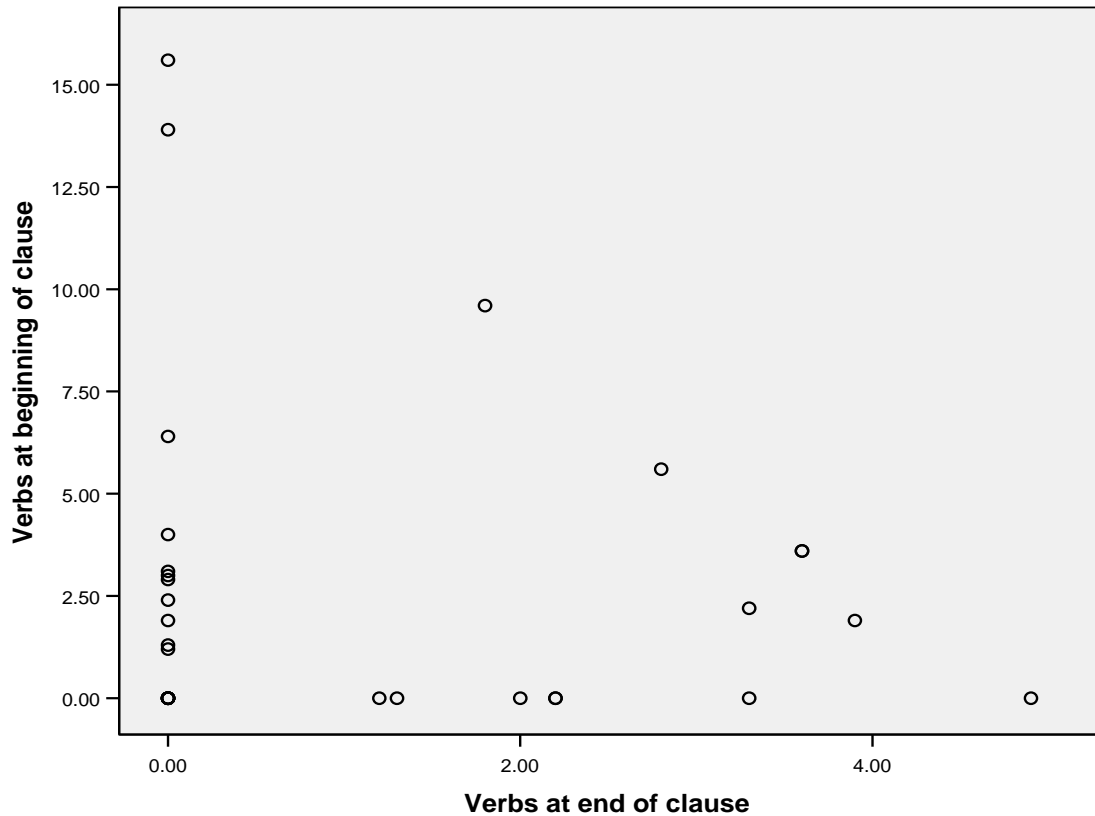


Figure 9.5 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group A, omitting low level cluster group c, of the Qumran Aramaic scrolls.

Figure 9.6 shows the same scattergram, including zero values, for high level cluster group B. There are only five data points and no clear relationship between variables is apparent. Similarly there is no clear relationship seen for high level group C (Figure

9.7). In summary, no clear inverse relationship emerges between verbs at the beginning versus verbs at the end of clauses from scatterplots, largely due to the abundance of zero values that introduce a 'floor' effect into any inverse relationship and this, in turn, probably reflects the brevity of texts under consideration.

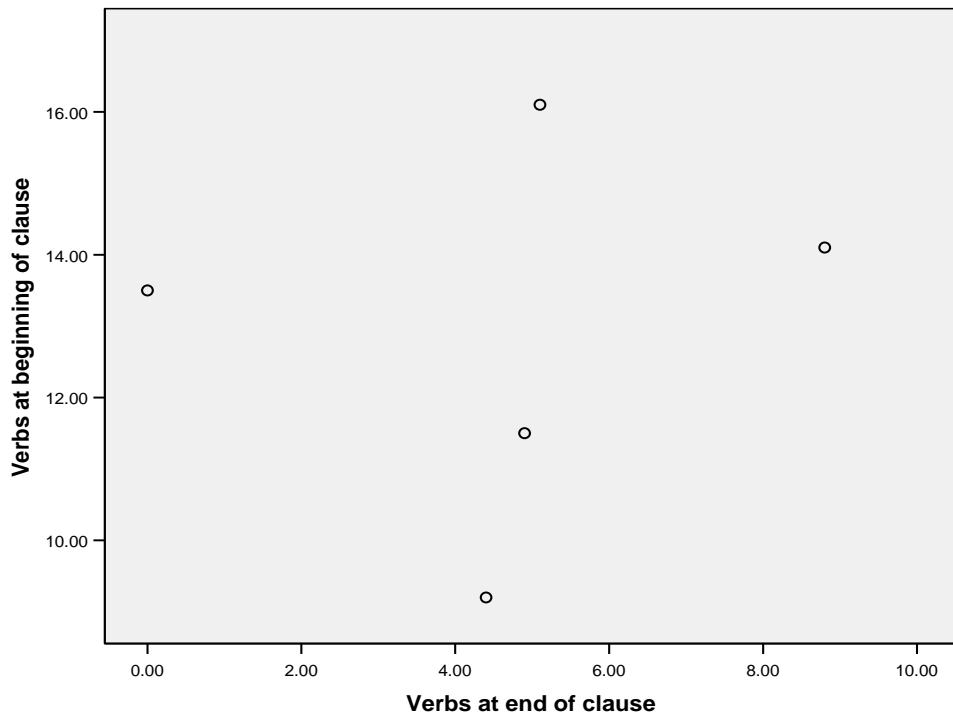


Figure 9.6 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group B of the Qumran Aramaic scrolls.

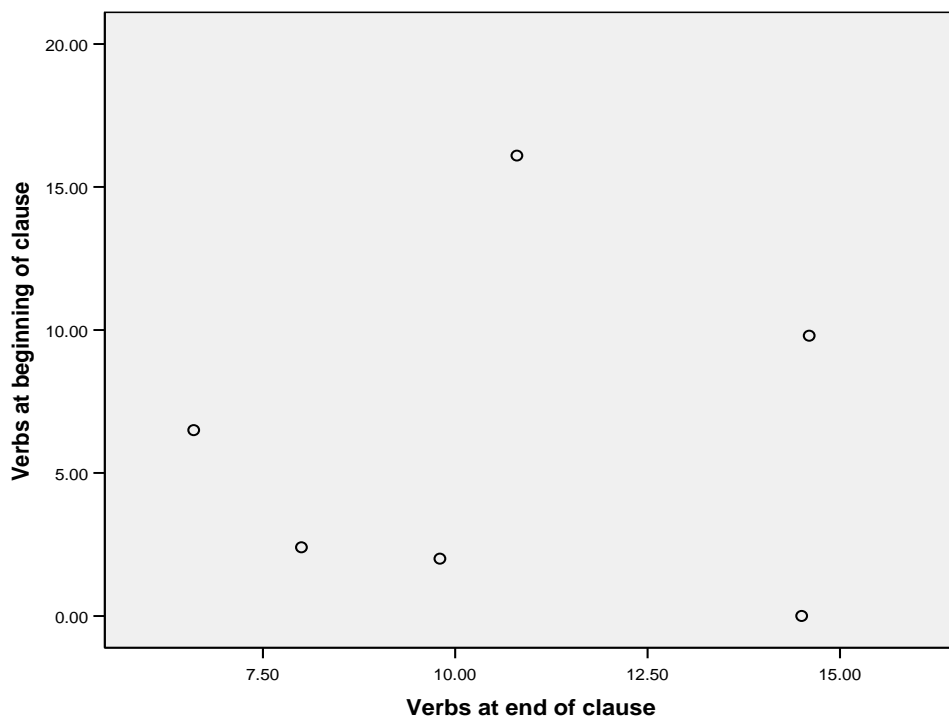


Figure 9.7 Scattergram of verbs at the beginning of clauses against verbs at the end of clauses in high level cluster group C of the Qumran Aramaic scrolls.

Figure 9.8 shows a scattergram of rates of nouns in the construct state versus rates of nouns followed immediately by a relative particle in high level cluster group A of the Qumran Aramaic scrolls. As shown by Table 9.6, there are a lot of zero values for nouns followed immediately by a relative particle. Figure 9.9 shows the same scattergram excluding these zero values. An inverse relationship is clearer; this is confirmed formally by Spearman $\rho = -0.44$, $p = .016$. As with verbal position, small sample size prevents adequate interpretation for high level cluster groups B and C.

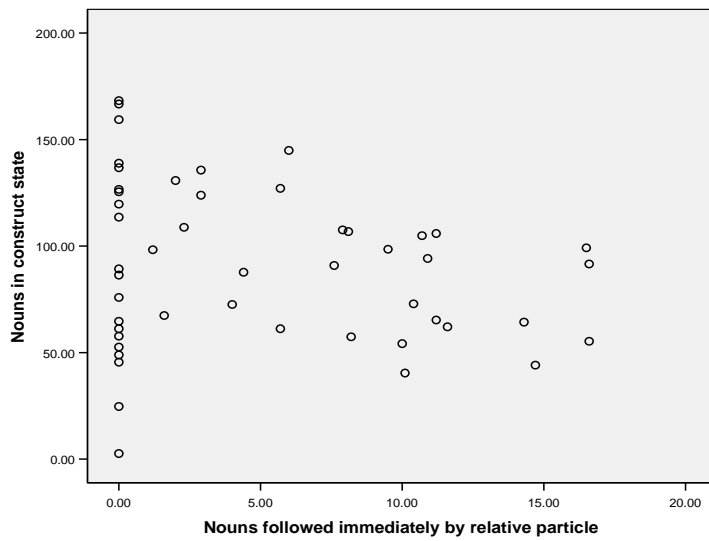


Figure 9.8 Scattergram of rates of nouns in the construct state versus rates of nouns followed immediately by a relative particle in high level cluster group A of the Qumran Aramaic scrolls.

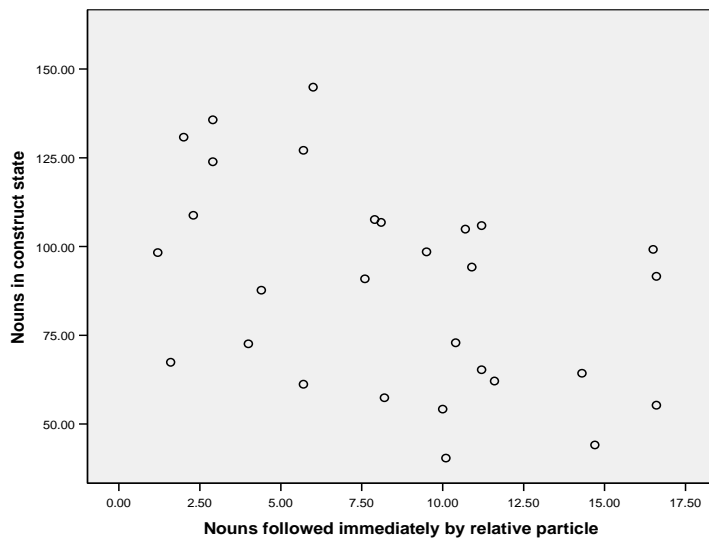


Figure 9.9 Scattergram of rates of nouns in the construct state versus rates of nouns followed immediately by a relative particle in high level cluster group A, excluding zero values, of the Qumran Aramaic scrolls.

Figure 9.10 shows a scattergram for passive participles versus *t*-stem participles in high level cluster group A of the Qumran Aramaic scrolls. There are very few non-zero values so that any inverse relationship would depend on a relatively small number of data points. Again, small sample size prevents adequate interpretation for high level cluster groups B and C.

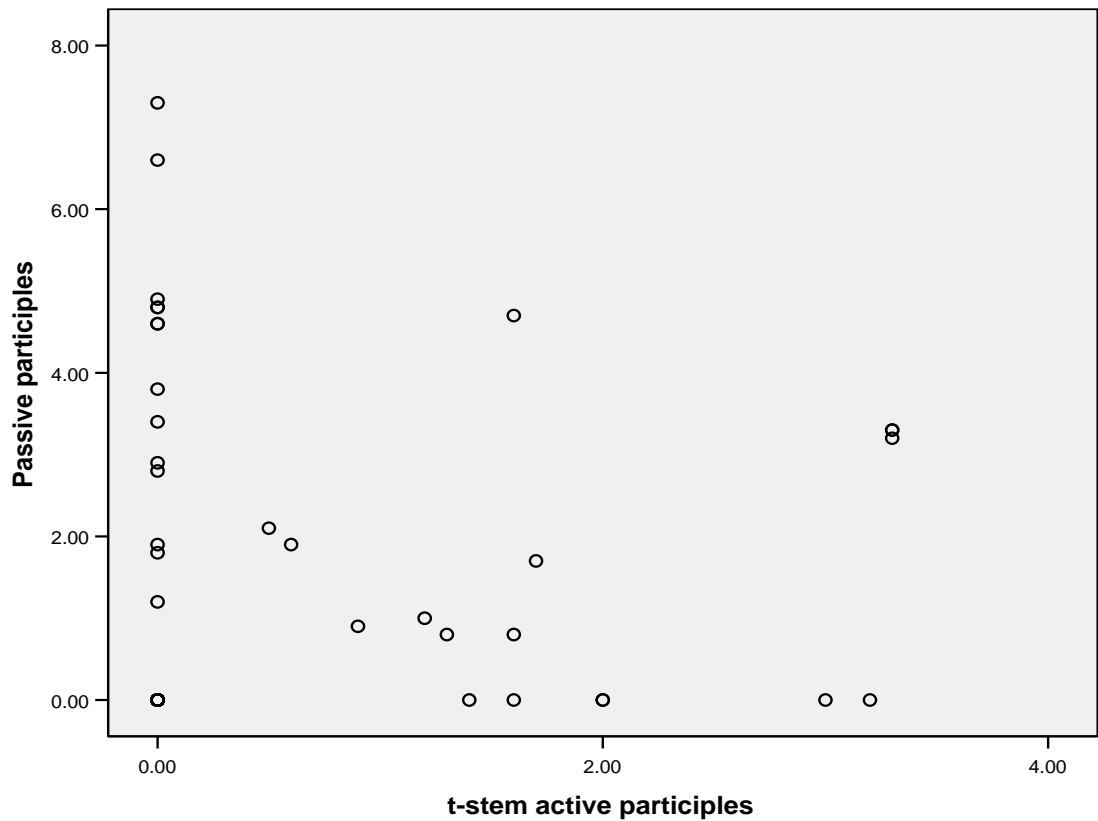


Figure 9.10 Scattergram for passive participles versus *t*-stem participles in high level cluster group A of the Qumran Aramaic scrolls.

In summary, there is evidence to support syntactical pair choice in the case of using nouns in construct state versus nouns followed immediately by a relative particle, but no

evidence for the other syntactical pairs where rates are lower. In view of the multitude of zero values, a simple present or absent classification may be more appropriate for verbal position and passive/*t*-stem participle choice; it is inappropriate for nouns in the construct state versus nouns followed immediately by a relative particle because there are no zero values for nouns in the construct state.

	Verbs at end of clauses		
Verbs at beginning of clauses	Absent	Present	Total
Absent	15	9	24
Present	14	36	50
Total	29	45	74

Table 9.18 Verbal position at beginning or end of clauses presence or absence counts for all the Qumran Aramaic scrolls.

Using a simple present/absent classification for a continuous variable naturally results in a loss of statistical power so examination of small cluster groups is inappropriate. Table 9.18 shows data on verbal position presence or absence for all the Qumran Aramaic scrolls. Inspecting the diagonals shows that the majority of counts fall in the absent-absent and present-present boxes indicating a positive relationship between these syntactical features (Fisher's exact test $p=.006$). This implies that if scrolls have verbs at the beginning of clauses they will also have them at the end which would be consistent with some scrolls having a higher verbal content.

Passive participles	<i>t</i> -stem participles		Total
	Absent	Present	
Absent	25	9	34
Present	18	22	40
Total	43	31	74

Table 9.19 Passive participle and *t*-stem participle presence or absence counts for the Qumran Aramaic scrolls.

Table 9.19 shows the data for passive participle and *t*-stem participle counts. As with verbal position there is a positive relationship between features (Fisher's exact test $p=.018$). Hence, there is no clear preference for first or last verbal position or for passive participles versus *t*-stem participles in the Qumran Aramaic corpus. It may represent a transitional phase in terms of these syntactical features, whilst in a subset of scrolls represented by high level cluster group A, a clear choice between using nouns in the construct state versus rates of nouns followed immediately by a relative particle is evident.

Comparison of syntactic features in the Qumran Aramaic scrolls with the Targums and Ancient Egyptian documents

As noted above, verbal position in clause is not searchable for either the Targums or the Ancient Egyptian documents in Accordance. Nevertheless, the preceding analyses indicate that the Qumran Aramaic corpus is heterogeneous in terms of the other syntactical features, so it is possible to consider whether external texts relate more or less closely to different low and high level clusters syntactically.

An initial inspection of means and 95% confidence intervals for construct state nouns in the Qumran Aramaic higher level cluster groups B and C (Table 9.17) compared with mean rates for Targums Onkelos, Neofiti and Pseudo-Jonathan (Table 9.8) shows that only the mean rate for the Book of Numbers in Targum Pseudo-Jonathan falls within any of the 95% confidence intervals (for higher level cluster group B). By contrast mean rates for many of the Ancient Egyptian documents (Table 9.9) fall within 95% confidence intervals for both higher level cluster groups B and C, and not a few are above the upper 95% confidence interval for mean rates. In terms of nouns in the construct state, then, higher level cluster groups B and C are closer to the Ancient Egyptian documents than to the Targums. A similar comparison can be made with the lower level cluster groups (Table 9.16). Both low level clusters c and d have lower 95% intervals above all the mean rates in the Targums. In contrast nearly all the mean rates in the Targums fall within the 95% confidence intervals of the mean rates for lower level clusters a and b of the Qumran Aramaic corpus. As expected many mean rates of many

of the Ancient Egyptian documents (n=23) are above the upper 95% confidence interval of lower level cluster group a, and nearly as many above that of lower level cluster group b. Hence, in terms of construct state nouns, lower level cluster groups a and b more closely resemble the Targums than the Ancient Egyptian documents.

A similar exercise to that for construct state nouns can be performed for nouns followed immediately by relative particle. Rates for all the Targums (Table 9.7) are above the upper 95% confidence intervals of the mean rates for high level cluster groups B and C. By contrast, only eight of the Ancient Egyptian documents have mean rates that are not below the lower 95% confidence interval of mean rates for high level clusters B and C (Table 9.9). Hence, by this syntactical criterion, high level cluster groups B and C hold an intermediate position between the Ancient Egyptian documents and the Targums, being slightly closer to the former than the latter. As noted above, lower level cluster groups b and c have similar mean rates of nouns followed immediately by relative particle to higher level cluster groups B and C and thus also are in an intermediate position. Lower level cluster groups a and d have mean rates in the same range as the vast majority of the Ancient Egyptian documents and therefore, in terms of nouns followed immediately by relative particle, are indistinguishable from them.

Passive participle rates show the opposite tendency: relatively high in many of the Ancient Egyptian documents (Table 9.12) and lower in the Targums (Table 9.11), though there is a subset of Ancient Egyptian documents with zero or near-zero rates. Mean rates of higher level clusters B and C are also relatively low, overlapping with the

majority of rates of the Targums. The upper 95% confidence intervals for all the lower level cluster groups falls below the mean rates of the majority of the Targums and closer to a subset of the Ancient Egyptian documents with similarly low rates.

Finally, high level cluster group B has far higher rates of *t*-stem participles than higher level cluster group C. The lower 95% confidence interval for high level cluster group B is above all the mean rates of the Targums (Table 9.11) and those of the Ancient Egyptian documents where only a few documents have non-zero rates. 95% confidence intervals for higher level cluster group C overlap with all of the Targum Onkelos means and two out of the five Targum Neofiti means, whilst all of the Targum Psuedo-Jonathan mean rates are higher. Higher level cluster mean rates overlap with all the Ancient Egyptian documents mean rates. The position is similar to higher level C cluster group for all the lower level cluster groups.

Table 9.20 summarises the relationship of the different high and low level cluster groups to the Ancient Egyptian documents and the Targums on these four syntactical variables. There is a spectrum of resemblance with the Qumran Aramaic cluster groups tending to fall between the Ancient Egyptian documents and the Targums. Lower level cluster group d is closest to the Ancient Egyptian documents, whilst lower level cluster group b is closest to the Targums.

Qumran Aramaic cluster	Ancient Egyptian documents rates	Intermediate rates	Targums rates	Higher rates
a	+	++	+	
b		+++	+	
c	+	+++		
d	++	++		
B	+	++		+
C	+	+++		

Table 9.20 Corresponding rates between Qumran Aramaic cluster groups and the Ancient Egyptian documents and the Targums. A plus (+) sign is inserted for each of the four syntactical variables (nouns in construct state, nouns followed immediately by relative particle, passive participles and t-stem participles) where rates coincide.

Comparison of syntactic features in the Qumran Aramaic scrolls with the Judean Desert Corpus and the Babatha archive

Most of the Aramaic Judean Desert Corpus texts are brief and limited in genre to receipts or contracts. Their syntactic range is therefore very narrow. Only one text, Mur19, features a verb at the start of a clause and none of the texts have a verb at the end of a clause. Table 9.21 shows rates per 1,000 words for the other four syntactical variables. It is striking that only one text, XHev/Se9 has a noun followed immediately by a relative particle and that no *t*-stem participles occur in any of the texts. With regard to nouns followed immediately by a relative particle, the Judean Desert Corpus is very similar to the Ancient Egyptian documents and quite distinct from the Targums. The situation is the same for *t*-stem participles. A fair number of the Qumran Aramaic scrolls have higher rates of both nouns followed immediately by a relative particle and *t*-stem participles than the essentially zero rates found in the Judean Desert Corpus.

As would be expected for short texts, there is a wide range of rates of nouns in a construct state in the Aramaic texts of the Judean Desert Corpus (Table 9.21). These are in general somewhat lower than those in the Qumran Aramaic scrolls (Table 9.6) and not clearly distinct from those in either the Ancient Egyptian documents or the Targums (Tables 9.8 and 9.9). It is of note that rates in earlier texts (WDSP) are similar to those in later texts (Mur) indicating a lack of evidence for any diachronic shift from nouns in a construct state to nouns followed immediately by a relative particle in this corpus.

	Nouns in construct state	Nouns followed immediately by relative particle	Passive participles	t-stem participles
WDSP1	32.49	0	0	0
WDSP2	42.47	0	3.86	0
WDSP3	70.35	0	20.1	0
WDSP4	31.87	0	3.98	0
WDSP5	37.93	0	0	0
WDSP6	20	0	0	0
WDSP7	37.97	0	9.49	0
WDSP8	61.14	0	4.37	0
WDSP9	57.75	0	3.04	0
WDSP10	35.18	0	5.03	0
WDSP15	30.03	0	3	0
WDSP18	11.76	0	0	0
Mur8	31.25	0	0	0
Mur10	53.33	0	0	0
Mur18	88.5	0	0	0
Mur19	80.77	0	0	0
Mur20	61.22	0	6.8	0
Mur21	30.38	0	0	0
Mur23	95.24	0	0	0
Mur25	20.59	0	0	0
Mur26	7.94	0	0	0
Mur27	18.18	0	0	0
Mur28	29.13	0	0	0
Mur31	23.26	0	0	0
Mur32	28.17	0	0	0
Mur33	103.45	0	0	0
Mur72	23.26	0	0	0
Sdeir2	113.21	0	9.43	0
XHevSe7	90.36	0	0	0
XHevSe8	42.4	0	7.31	0
XHevSe8a	60.15	0	7.52	0
XHevSe9	45.61	3.51	5.26	0
XHevSe9a	0	0	0	0
XHevSe10	88.24	0	0	0
XHevSe11	45.45	0	0	0
XHevSe12	89.29	0	0	0
XHevSe13	125	0	0	0
XHevSe21	4.61	0	0	0
XHevSe22	48.78	0	0	0
XHevSe23	0	0	0	0
XHevSe24	37.04	0	0	0
XHevSe24a	0	0	0	0
XHevSe25	46.88	0	0	0

XHevSe26	39.22	0	0	0
XHevSe31	0	0	0	0
XHevSe32	57.14	0	0	0
XHevSe33	0	0	0	0
XHevSe34	0	0	0	0
XHevSe40	0	0	0	0
XHevSe50	21.74	0	0	0
34Se3	12.5	0	0	0

Table 9.21 Rates per 1,000 words for syntactical variables in the Aramaic texts from the Judean Desert Corpus.

By contrast, there is a difference in rates between earlier Judean Desert Corpus Aramaic texts (WDSP) and later ones (Mur) for passive participle rates. These are closer to the high rates of many of the Ancient Egyptian documents for WDSP but mainly zero, and hence closer to the lower rates found in the Targums and some of the Qumran Aramaic text groups for Murabba'at texts. This would be consistent with a diachronic syntactic shift, but the heterogeneity apparent in this relatively uncommon feature makes it unwise to draw any strong inferences, especially given that WDSP texts likely reflect a Samaritan tradition. On the whole, though, Qumran Aramaic texts of all groups are more similar to later Murabba'at than earlier Wadi ed-Daliyeh texts.

Table 9.22 shows rates per 1,000 words of the six syntactic criteria in the Babatha Archive texts. Like texts in the rest of the non-Qumran Judean Desert corpus, all texts have a high rate of nouns in a construct state, but unlike those texts genitive constructs using the relative particle are present in five of the nine texts. Similarly, passive participles, although infrequent, were present in the archive, but so were t-stem

participles in contrast to the other non-Qumran Judean Desert texts. There were far higher rates of verbs at the start of clauses than at the end, though the latter were present in four of the nine texts. The mean rate of verbs at the beginning of clauses for the Babatha texts archive at 26 and median at 23 is higher than any upper 95%

Text	Verbs at start of clause	Verbs at end of clause	Nouns in construct state	Genitive construction with relative particle	Passive participles	<i>t</i> -stem participles
pYd7	23.0	0.5	93	1.1	2.7	1.6
pYd8	0	0	118	0	0	0
pYd10	86	0	343	14	0	43
pYd17	29	57	114	0	0	0
pYd18	0	43	157	0	0	0
pYd20	39	115	154	38	0	0
pYd42	12	0	122	5.8	5.8	0
pYd47A	33	0	109	11	0	0
pYd47B	17	0	78	0	0	0

Table 9.22 Rates per 1,000 words of the six syntactic criteria in the Babatha Archive.

confidence interval of any of the Qumran clusters. The mean (23) and median (0) of rates of verbs in the final position differ so markedly that making comparisons would be of questionable validity. Mean rate of nouns in a construct state at 143 (median 118) is in line with low level cluster c and high level cluster C. Mean rate of genitive construct with a relative particle at 7.8 (median 1.1) is consistent with Qumran low level clusters b and c, and high level clusters B and C. Mean rate of passive participles at 0.9 (median 0) is consistent with all Qumran low level and high level clusters except high level cluster A. Finally, mean rate of *t*-stem participles at 4.9 is above that for all Qumran clusters except high level cluster B, but the median of zero is consistent with most of the

clusters. Taken together Babatha archive texts most closely resemble Qumran low level c and high level C clusters, but there remain differences in rates of nouns in a construct state.

1Q21	11	↓↘
6Q8	81	↓□
4Q553	70	↓□
4Q570	79	↓□
1Q23	12	↓□
4Q531	48	↓□
4Q244	43	↓□
4Q318	45	↓□
4Q214	38	↓□
4Q214a	39	↓□
4Q532	49	↓□
4Q208	30	↓□
4Q540	56	↓□
4Q552	69	↓□
4Q556	72	↓□
4Q558	74	↓□
4Q560	76	↓□
4Q113	16	↓□
4Q206	28	↓□
2Q24	14	↓□
4Q115	17	↓↘↙↓↘
4Q203	25	↓□ ↔
4Q542	58	↓□ ↔
11Q18	83	↓□ ↔
5Q15	80	↓□ ↔
4Q112	15	↓□ ↔
4Q554	71	↓□ ↔
4Q198	22	↓□ ↔
4Q529	46	↓□ ↔
4Q197	21	↓□ ↔
4Q213	35	↓□ ↔
Yd7	1	↓□ ↔
Yd47B	9	↓□ ↔
4Q213b	37	↓□ ↔
4Q539	55	↓□ ↔
4Q246	44	↓□ ↔
4Q210	32	↓□ ↔
4Q550c	68	↓□ ↔
1Q20	10	↓□ ↔
4Q547	63	↓□ ↔
4Q538	54	↓□ ↔
4Q201	23	↓□ ↔
4Q544	60	↓□ ↔
4Q202	24	↓□ ↔
4Q530	47	↓□ ↔
4Q204	26	↓□ ↔

Babatha Archive texts which represent Aramaic sections from Greek letters, pYd10, 17, 18 and pYd20, appear distinct from most of those at Qumran. This, again, is reassuring with regard to the validity of clustering methods for classification. pYd47A clusters with 4Q156 and 11Q10, neither part of the low level or high level clusters previously identified. By contrast pYd7, 8, 42 and pYd47B cluster with a large group of Qumran Aramaic scrolls. These legal texts, therefore, can be considered close to Qumran as a text type syntactically, but do not pull out any distinct clusters of affiliated texts from within the Qumran corpus.

Summary of findings from syntactic data

The data indicate that it is incorrect to speak of the syntax of Qumran Aramaic: there is heterogeneity within the Qumran Aramaic corpus with discrete identifiable syntactical types. In general Qumran Aramaic texts lie somewhere intermediate between the Ancient Egyptian documents and the Targums syntactically and fairly close to the Babatha archive legal texts. The Aramaic texts of the Judean Desert Corpus are, in general, closer to the Ancient Egyptian documents than the Qumran Aramaic corpus, though there is evidence of diachronic shift within the Judean Desert Corpus itself with rates of passive participles in later Judean Desert Corpus texts being closer to those found at Qumran. Having noted these general features, two of the low-level clusters exhibit syntactic profiles closer to those found in the Targums than any of the other groups: low-level clusters a and b. These low-level clusters comprise:

1Q21 (Testament of Levi), 1Q23 (Enoch Giants), 2Q24 (New Jerusalem), 4Q214 (Aramaic Levi), 4Q214a (Aramaic Levi), 4Q244 (Pseudo-Daniel), 4Q318 (Brontologion), 4Q531 (Book of Giants), 4Q532 (Book of Giants), 4Q540 (Apocryphon of Levi), 4Q552 (Four Kingdoms), 4Q553 (Four Kingdoms), 4Q556 (Book of Giants), 4Q558 (Vision), 4Q560 (Exorcism), 4Q570 (Aramaic R), 6Q8 (Giants).

In Chapter 8 a Targumic text type cluster was identified from morphological criteria. This comprised 1Q24, 4Q214, 4Q214b, 4Q318 and 4Q548. Two of these five Targumic scrolls, 4Q214 and 4Q318, also appear in the Targumic syntactical text type clusters. Identification of the same texts as Targumic by two independent statistical methods suggests that the findings are robust.

Of all the Qumran Aramaic syntactical clusters, low-level cluster d has features most close to those found in the Ancient Egyptian documents. This group comprises:

1Q24 (Enoch Giants), 4Q196 (Tobit), 4Q213a (Aramaic Levi), 4Q213b (Aramaic Levi), 4Q243 (Pseudo-Daniel), 4Q537 (Testament of Jacob), 4Q539 (Apocryphon of Joseph B), 4Q541 (Apocryphon of Levi), 4Q543 (Visions of Amram), 4Q545 (Visions of Amram), 4Q546 (Visions of Amram), 4Q548 (Visions of Amram), 4Q549 (Work mentioning Hur and Miriam), 4Q557 (Vision), 4Q561 (Physiognomy/Horoscope), 11Q18 (New Jerusalem).

In Chapter 8 (Figure 8.28) the following Qumran Aramaic scrolls which belonged to a discrete morphological text type clustered with those from Ancient Egypt:

4Q196, 4Q197, 4Q208, 4Q209, 4Q213, 4Q213a, 4Q213b, 4Q539, 4Q543, 4Q544, 4Q545, 4Q546, 4Q547, 5Q15.

Of these fourteen scrolls, seven of them also occur in the syntactical cluster: chance would have expected only three or four to have been in common. The scrolls common to both morphological and syntactical clusters are 4Q196, 4Q213a, 4Q213b, 4Q539, 4Q543, 4Q545 and 4Q546. Thus, again, syntactical clustering findings corroborate those of morphological clustering. Moreover, scrolls that are classified similarly on both morphological and syntactic criteria provide suitable anchor points for confirmatory analyses and, hence, classification of other scrolls that are less clearly clustered. In particular, syntactic analyses support the utility of comparisons with non-Qumran Aramaic texts as an aid to characterization of the text types uncovered.

Bibliography

Aitken, J. K. Rhetoric and poetry in Greek Ecclesiastes. *Bull Int Org Septuagint & Cognate Studies* (2005) **38**, 55-77.

Alfaro, M. E., Zoller, S., Lutzoni, F. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evolution* (2003) **20**, 255-266.

Alexander, P.S. *Targum, Targumim*. In: The Anchor Bible Dictionary, ed. Freedman, D.N. New York: Doubleday, 6:320-31, 1992.

Alter, R. *The Art of Biblical Narrative*. New York: Basic Books, 1981.

Arnold, B. T. & Choi, J. H. *A Guide to Biblical Hebrew Syntax*. Cambridge: Cambridge University Press, 2003.

Baayen, R. A. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press, 2008.

Ben Tosef Tawil, H. *An Akkadian Lexical Companion for Biblical Hebrew*. Jersey City, NJ: Ktav Publishing Hopuse Inc., 2009.

Biber, D. *Variations across Speech and Writing*. Cambridge: Cambridge University Press, 1988.

Black, D. A. *Linguistics for students of New Testament Greek*. 2nd edition. Grand Rapids: Baker Books, 1995.

Blau, J. *A conservative view of the language of the Dead Sea scrolls*. In: *Diggers at the Well*. Muraoka, T & Elwolde J. F. (eds). Leiden: Brill, 2000.

Boccaccini, G. *Qumran and the Enoch groups: revisiting the Enochic-Essene hypothesis*. In: *The Bible and the Dead Sea Scrolls*. Volume 1. *Scripture and the Scrolls*. Charlesworth, J. H. (ed). Waco, Texas: Baylor University Press, 2006.

Boyarsky, A. & Góra, P. A comparative statistical dynamical analysis of Hebrew texts. *Discrete Dynamics in Nature and Society* (2000) **4**, 293-295.

Brock, S. P. Abraham and the ravens : a Syriac counterpart to *Jubilees* 11-12 and its implications. *Journal for the Study of Judaism in the Persian, Hellenistic and Roman Period* (1978) **9**, 135-152

Brooke, G. J. *E pluribus unum: Textual variety and definitive interpretation in the Qumran scrolls*. In: *The Dead Sea Scrolls in their Historical Context*. Lim, T. H. (ed). Edinburgh: T.&T. Clark, 2000.

Carmignac, Jean., Un aramaïsme biblique et qumrânien : l'infinitif placé après son complément d'objet, *Revue de Qumran* (1966) **5**, 503-520.

Cartledge, T. W. *Vows in the Hebrew Bible and the Ancient Near East*. Sheffield: JSOT Press, 1992.

Child, D. *The Essentials of Factor Analysis*. 3rd Edition. London: Continuum International Publishing Group, 2006.

Choi, J. J. The Aramaic of Daniel: Its date, place of composition and linguistic comparison with extra-biblical texts. PhD, Trinity Evangelical Divinity School, 1994.

Chum, B. K.-Y. A literary analysis of 1QH 10-17:36. Ph.D., University of Edinburgh, Edinburgh, 2000.

Conway, B. J. The Thanksgiving Hymns from Qumran: a statistical approach to the authorship question. Ph.D., King's College, University of London, London, 1999.

Cook, E. M. *Qumran Aramaic and Aramaic Dialectology*. In: Studies in Qumran Aramaic, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Cook, E. M. *The Aramaic of the Dead Sea Scrolls*. In: The Dead Sea Scrolls after Fifty Years. Flint, P. W. & VanderKam, J. C. (eds). Leiden: Brill, 1998.

Cook, E. M., "The 'Kaufman effect' in the Pseudo-Jonathan Targum," *Aramaic Studies* (2006), **4**, 123-32

Cook, J. A. The semantics of verbal pragmatics: clarifying the roles of *wayyiqtol* and *weqatal* in Biblical Hebrew prose. *Journal of Semitic Studies* (2004) **XLIX**, 247-273.

Coxon, Peter W., The distribution of synonyms in biblical Aramaic in the light of official Aramaic and the Aramaic of Qumran, *Revue de Qumran* (1978) **9**, 497-512.

Cross, Frank Moore, The development of the Jewish scripts. Doubleday: New York, 1961

Cross, F. M. & Saley, R. J. A Statistical Analysis of the Textual Character of 4QSamuel^A (4Q51). *Dead Sea Discoveries* (2006) **13**, 46-54.

Damsma, A. An analysis of the dialect and early Jewish mystical lore in a Targumic Tosefta to Ezekiel 1.1 (Ms Gaster 1478). *Aramaic Studies* (2008) **6**, 17-58.

Danove, P. L. *Linguistics and Exegesis in the Gospel of Mark*. Sheffield: Sheffield Academic Press Ltd., 2001.

Davies, P. Genesis: An Authorship Study in Computer-Assisted Statistical Linguistics by Y. T. Radday (Review). *Applied Statistics* (1986) **35**, 218.

Davila, J. R. (How) can we tell if a Greek Apocryphon or Pseudepigraphon has been translated from Hebrew or Aramaic? *Journal for the Study of the Pseudepigrapha* (2005) **15**, 3-61.

de Caën, V. Hebrew Linguistics and Biblical Criticism: a Minimalist Approach. *The Journal of Hebrew Scriptures* (2001) **3**, 1-32.

de Caën, V. On the distribution of major and minor pause in Tiberian Hebrew in the light of the variants of the second person independent pronouns. *Journal of Semitic Studies* (2005) **50**, 321-327.

Diez Merino, L. *The adverb in Qumran Aramaic*. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Dobson, J. H. *Learn Biblical Hebrew*. Dallas: SIL International, 1999.

Doudna, G. *Dating the Scrolls on the Basis of Radiocarbon Analysis*. In: *The Dead Sea Scrolls after Fifty Years*, edited by Flint Peter W., and VanderKam, James C., Vol.1. Leiden: Brill, 1998, 430-471.

Duffy, C. A. *Prayer*. In: *The Firebox*. O'Brien, S. (ed). London: MacMillan Publishers Ltd., 1998.

Epp, E. J. & Fee, G. D. *Studies in the Theory and Method of New Testament Textual Criticism*. Grand Rapids: Eerdmans, 1993.

Fassberg, S. E. *Hebraisms in the Aramaic documents from Qumran*. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Fassberg, S.E. *The Language of the Aramaic Documents from Qumran*. In: Meghillot, *Studies in the Dead Seas scrolls: II*, Bar-Asher, M. & DEimant, D. (eds). Haifa: University of Haifa, 2004.

Fassberg S.E. *The forms of 'son' and 'daughter' in Aramaic*. In: *Aramaic in its Historical and Linguistic Setting*. Wiesbaden: Harrassowitz Verlag, 2008.

Fassberg, S.E. *Salient features of the verbal system in the Aramaic Dead Sea Scrolls*. In: *Aramaica Qumranica*, Berthelot, K. & Stökl Ben Ezra D (eds). Leiden: Brill, 2010

Fee, G. D. *New Testament Exegesis*. 3rd edition. Louisville, Kentucky: Westminster John Knox Press, 2002.

Fitzmyer, J. A. *A Wandering Aramean: Collected Aramaic Essays*. Missoula, MT: Scholars Press, 1979.

Foucault, M. *The Order of Things*. London: Routledge Classics, 2002.

Fried, H. M. The Dead Sea Scrolls: a statistical analysis. *Qumran Chronicle* (2010) **18**, 123-135.

Gesenius, W. *Gesenius' Hebrew Grammar*, as expanded by Kautzsch, E. & Cowley, A.E.. 2nd English Edition. Oxford: Clarendon Press, 1010.

Ginsberg, H.L., "The Northwest Semitic Languages," Pages 102-124, 293. In: *The World History of the Jewish People, Vol 2: The Patriarchs*. Mazar, B. (ed). Tel-Aviv: New Brunswick, 1970

Golb, N. *Who wrote the Dead Sea scrolls?* New York: Scribner,1995.

Greenfield, J. C. & Sokoloff, M. *The Contribution of Qumran Aramaic to the Aramaic Vocabulary*. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Groom, S. *Linguistic Analysis of Biblical Hebrew*. Carlisle: Paternoster Press, 2003.

Güting, E. W. & Mealand, D. L. *Asyndeton in Paul: a text-critical and statistical enquiry into Pauline style*. Lampeter: Edwin Mellen Press, 1998.

Gzella, H. The heritage of Imperial Aramaic in Eastern Aramaic. *Aramaic Studies* (2008) **6**, 85-109.

Gzella, H. Dating the Aramaic Texts from Qumran: Possibilities and Limits. *Revue Qumran* (2009) **24**, 61-78.

Halliday, M. A. K. *Language as System and Language as Instance*. In: Data, Description, Discourse: Papers on the English Language in Honour of John McH. Sinclair. Hoey, M. (ed). London: Harper Collins, 1993.

Hasel, G. F. The book of Daniel and matters of language evidences (sic) relating to names, words, and the Aramaic language. *Andrews University Seminary Studies* (1981) **19**, 211-225.

Hempel, C. *The Place of the Book of Jubilees at Qumran and Beyond*. In: The Dead Sea Scrolls in their historical context. Lim, T. (ed.). Edinburgh: T.&T. Clark, 2000.

Herbert, E. D. & Tov, E. *The Bible as Book: the Hebrew bible and the Judaean Desert discoveries*. London: The British Library, 2002.

Herbert, W. N. *Ode to Scotty*. In: The Firebox. O'Brien, S. (ed). London, MacMillan Publishers Ltd., 1998.

Holmstedt, R. D. *The relative clause in Biblical Hebrew: a linguistic analysis*. Ph.D., The University of Wisconsin, Madison, 2002.

Holmstedt, R. Issues in the Linguistic analysis of a Dead language, with particular reference to Ancient Hebrew. *The Journal of Hebrew Scriptures* (2006) **6**, 2-21

Houk, C. B. Statistical analysis of Genesis sources. *JSOT* (2002) **27**, 75-105.

Houk, C. B. Multiple poets in Lamentations. *JSOT* (2005) **30**, 111-125.

Huehnergard, J. A Grammar of Akkadian. Second Edition. Winona Lake, Indiana, Eisenbrauns, 2005.

Joosten, J. *L'Araméen de Qumran entre l'Araméen d'empire et les Targumim: L'emploi de la préposition "Devant" pour exprimer le respect dû au roi et à Dieu.* In: Aramaica Qumranica, Berthelot, K. & Stökl Ben Ezra D (eds). Leiden: Brill, 2010

Kaddari, M. Z. *The existential verb hwh in Imperial Aramaic.* In: Arameans, Aramaic and the Aramaic Literary Tradition. Sokoloff, M. (ed). Ramat-Gan, Israel: Bar-Ilan Press, 1983.

Kelley, P H., Mynatt, D. S. & Crawford, T. G. *The Masorah of Biblia Hebraica Stuttgartensia.* Cambridge: Wm. B. Eerdmans Publishing Co., 1998.

Kenny, A. *A stylometric study of the New Testament.* Oxford: Clarendon Press, 1986.

Kesterson, J. C. *Tense usage and verbal syntax in selected Qumran documents.* PhD, The Catholic University of America, 1984

Koller, A. *Four dimensions of linguistic variation: Aramaic dialects in and around Qumran.* In: *The Dead Sea Scrolls in Context*, volume 1, Lange, A., Tov, E. and M. Weigold (eds). Leiden: Brill, 2011.

Kugel, J. How old is the Aramaic Levi Document? *Dead Sea Discoveries* (2007) **14**, 291-312.

Kutscher, E. Y. *Studies in Galilean Aramaic*. Sokoloff, M (trans). Ramat-Gan: Bar-Ilan University, 1976.

Larson, E. *The Relation between the Greek and Aramaic Texts of Enoch*. In: *The Dead Sea Scrolls: Fifty Years after their Discovery 1947-1997*, Schiffman, L. H., Tov, E. & VanderKam, J. C. (eds). Jerusalem: Israel Exploration Society, 2000.

Le Deaut, R. *Targum Du Pentateuque*, 5 vols. Translations of Tgs. Neof. And Ps-Jon. Paris: Editions du Cerf, 1978-1981.

Leonard, T. *Unrelated Incidents*. In: *The Firebox*. O'Brien, S. (ed). London, MacMillan Publishers Ltd., 1998.

Li, T. Non-active participles in the Aramaic of Daniel. *Aramaic Studies* (2008) **6**, 111-136.

Lim, T. H. *Holy Scripture in the Qumran Commentaries and Pauline Letters*. Oxford: Clarendon Press, 1997.

Louw, J. P. *Greek-English Lexicon of the New Testament: Based on Semantic Domains*. New York: American Bible Society, 1988.

Lowery, K. E. *Toward a discourse grammar of biblical Hebrew (cohesion, narrative, paragraph, computer-aided, syntax)*. Ph.D., University of California, Los Angeles, 1985.

Lüdemann, G. *Heretics: the other side of early Christianity*. London: SCM Press, 1996.

Malina, B. J. *Interfaith Dialogue: challenging the received view*. In: *Ancient Israel: the Old Testament in its Social Context*. Esler, P. F. (ed). London: SCM Press, 2005.

Masterson, K. J. *An Information-Theoretic approach to the analysis of Semitic texts*. Ph.D., University of California, Los Angeles, 1994.

McLay, R. T. *The Use of the Septuagint in New Testament Research*. Cambridge: Wm. B. Eerdmans Publishing Co., 2003.

McMahon, A. & McMahon, R. *Language Classification by Numbers*. Oxford: OUP, 2005.

McNamara, M. *Targum Neofiti, Genesis*. Collegeville, MN: Liturgical Press, 1992.

McNamara, M. *Targum Neofiti 1, Exodus*. Collegeville, MN: Liturgical Press, 1994.

McNamara, M. *Targum Neofiti 1, Leviticus*. Collegeville, MN: Liturgical Press, 1994.

McNamara, M. *Targum Neofiti 1, Numbers*. Collegeville, MN: Liturgical Press, 1995.

McNamara, M. *Targum Neofiti 1, Deuteronomy*. Collegeville, MN: Liturgical Press, 1997.

Menken, M. J. J. *Old Testament Quotations in the Gospel of John*. In: *New Testament Writers and the Old Testament*. Court, J. M. (ed). London: SPCK, 2002.

Metzger, B. *The Canon of the New Testament*. Oxford: OUP, 1987.

Metzger, B. M. & Ehrman, B. D. *The Text of the New Testament*. Oxford: OUP, 2005.

Miller, D. M., and Ken M. Penner, K.M.. *Jubilees*. In: *The Online Critical Pseudepigrapha*. Penner, K.M., Miller, D.M., & Scott, I.W. (eds). Atlanta: Society of Biblical Literature, 2006. Online: <http://www.purl.org/net/ocp/Jub.html>.

Morag, S. *On some concepts in the world of Qumran: polysemy and semantic development*. In: *Diggers at the Well*. Muraoka, T & Elwolde J. F. (eds). Leiden: Brill, 2000.

Morton, A. Q. & McLeman, J. *Christianity and the Computer*. London: Hodder & Stoughton Ltd., 1964.

Moyise, S. *The Old Testament in the New*. London: Continuum, 2001.

Muraoka, T. *On the morphosyntax of the infinitive in Targumic Aramaic*. In: *Arameans, Aramaic and the Aramaic Literary Tradition*. Sokoloff, M. (ed). Ramat-Gan, Israel: Bar-Ilan Press, 1983.

Muraoka, T. *The verbal rection in Qumran Aramaic*. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Muraoka, T. *Hebrew*. In: *Encyclopedia of the Dead Sea Scrolls*. Schiffman, L. H. & VanderKam, J. C. (eds). New York, Oxford University Press, 2000.

Muraoka, T. *A Grammar of Qumran Aramaic*. Leuven, Netherlands: Peeters, 2011.

Muraoka, T. & Rogland, M., The *waw* consecutive in Old Aramaic? A rejoinder to Victor Sasson. *Vetus Testamentum* (1998) **48**, 99-104.

Murray, R. *Symbols of Church and Kingdom*, revised edition. London: T&T Clark, 2006.

Muthén, L. K. & Muthén, B. O. *Mplus Statistical analysis with Latent variables User's Guide*. 6th Ed. Los Angeles: Muthén & Muthén, 2010.

Novum Testamentum Graece. 27th ed. Stuttgart: Deutsche Bibellgesellschaft, 2001.

Neumann, K. J. *The authenticity of the Pauline epistles in the light of stylostatistical analysis*. Atlanta: Scholars Press, 1990.

O'Donnell, M. B. *Designing a Register-Balanced Corpus*. In: *Diglossia and other Topics in New Testament Linguistics*. Porter, S. E. (ed). Sheffield: Sheffield Academic Press Ltd., 2000.

O'Donnell, M. B. *Corpus Linguistics and the Greek of the New Testament*. Sheffield: Sheffield Phoenix Press, 2005.

O'Keefe, R. A. Critical remarks on Houk's 'Statistical analysis of Genesis sources'. *JSOT* (2005) **29**, 409-437.

Parry, D. W. *Linguistic profile of the nonbiblical Qumran texts: a multidimensional approach*. In: *From 4QMMT to Resurrection*. Martinez, F. G., Steudel, A., & Tigchelaar, E (eds). Leiden: Brill, 2006.

Parunak, H. V. D. Linguistic Density Plots in Ezekiel. *The Computer Bible*, vol. 27. Wooster: Biblical Research Associates, 1984.

Penner, K.M. & Scott I.W. "1 Enoch." In *The Online Critical Pseudepigrapha*. Penner, K.M., Miller, D.M., & Scott, I.W. (eds). Atlanta: Society of Biblical Literature, 2006. Online: <http://ocp.tyndale.ca/1-ethiopic-apocalypse-of-enoch>.

Pollatschek, M. and Radday, Y. T. *Vocabulary richness and concentration in the Hebrew biblical literature*. *Ass. Lit. Ling. Comput. Bull.* (1981) **8**, 217-231.

Porter, S. E. & O'Donnell, M. B. *Semantics and patterns of argumentation in the Book of Romans: Definitions, proposals, data and experiments*. In: *Diglossia and other Topics in New Testament Linguistics*. Porter, S. E. (ed). Sheffield, Sheffield Academic Press Ltd., 2000.

Portnoy, S. & Peteren, D. Biblical Texts and Statistical Analysis: Zechariah and Beyond. *Journal of Biblical Literature* (1984) **103**, 11-21.

Puech, E. L'Araméen des manuscrits de la mer Morte. 1. Grammaire. Book review.. *Dead Sea Discoveries* (2011) **18**, 264-273.

Qimron, E. *The Hebrew of the Dead Sea scrolls*. Winona Lake, Indiana: Eisenbrauns, 1986.

Qimron, E. The pronominal suffix $\text{h}k\text{-}$. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Qimron, E. *The nature of DSS Hebrew and its relation to BH and MH*. In: *Diggers at the Well*. Muraoka, T & Elwolde J. F. (eds). Leiden: Brill, 2000.

Radday, Y. T. Two computerized statistical-linguistic texts concerning the unity of Isaiah. *Journal of Biblical Literature* (1970) **89**, 319-324.

Radday, Y. T. *The Unity of Isaiah in the Light of Statistical Linguistics*. Hildesheim: Gerstenberg, 1973.

Radday, Y. & Wickmann, D. The Unity of Zechariah Examined in the Light of Statistical Linguistics. *Zeitschrift für die alttestamentliche Wissenschaft* (1975) **87**, 30-55.

Radday, Y. T., Leb, G., Wickmann, D., & Talmon, S. Book of Judges examined by statistical linguistics. *Biblica* (1977) **58**, 469-499.

Radday, Y. T., *et al.* *Genesis: An authorship study in computer-assisted statistical linguistics*. Rome: Biblical Institute, 1985.

Rosenthal, F. *A Grammar of Biblical Aramaic*. (5th printing). Wiesbaden: Otto Harrassowitz, 1983.

Rowley, H. H. *The Aramaic of the Old Testament*. Oxford: Oxford University Press, 1929.

Rubin, Aaron. 2007. On the third person preformative *l-/n-* in Aramaic, and an Ethiopic parallel. *Ancient Near Eastern Studies* 44.1–28.

Schattner-Rieser, U. *Some observations on Qumran Aramaic: the 3rd fem. sing. pronominal suffix*. In: *The Dead Sea Scrolls: Fifty Years after their Discovery 1947-*

1997, Schiffman, L. H., Tov, E. & VanderKam, J. C. (eds). Jerusalem: Israel Exploration Society, 2000.

Schattner-Rieser, U. L'Araméen des manuscrits de la mer Morte. 1. Grammaire. Lausanne : Éditions du Zèbre, 2004.

Schattner-Rieser, U. *L'apport de la philologie Araméenne et l'interprétation des archaïsmes linguistiques pour la datation des textes Araméens de Qumran*. In: Aramaica Qumranica, Berthelot, K. & Stökl Ben Ezra D (eds). Leiden: Brill, 2010

Schniedewind, W. M. Qumran Hebrew as an antilanguage. *Journal of Biblical Literature* (1999) **118**, 235-252.

Schniedewind, W. M. *Linguistic ideology in Qumran Hebrew*. In: Diggers at the Well. Muraoka, T & Elwolde J. F. (eds). Leiden: Brill, 2000.

Schorch, S. Determination and the use of the definite article in the Samaritan and the Masoretic text of the Torah. *Journal of Semitic Studies* (2003) **XLVIII**, 287-320.

Semantics of Ancient Hebrew Database. *Lexeme index*.
http://www2.div.ed.ac.uk/research/sahd/lexeme_index.html.

Seow, C. L. *A Grammar for Biblical Hebrew*. Revised Edition. Nashville: Abingdon Press, 1995.

Sharp, C. J. *Prophecy and Ideology in Jeremiah*. London: T.&T. Clark, 2003.

Shepherd, M. B. *The verbal system of biblical Aramaic: A distributional approach*. PhD, Southeastern Baptist Theological Seminary, 2006.

The Shorter Oxford English Dictionary. 3rd ed. Onions, C. T. (ed). Oxford: Clarendon Press, 1973.

Sichel, H. S. On a distribution law for word frequencies. *J. Am Statist. Ass.* (1975) **70**, 542-547.

Sichel, H. S. Word frequency distributions and type-token characteristics. *Math. Scient.* (1986) **11**, 45-72.

Sokoloff, M. *Qumran Aramaic in Relation to the Aramaic Dialects*. In: The Dead Sea Scrolls: Fifty Years after their Discovery 1947-1997, Schiffman, L. H., Tov, E. & VanderKam, J. C. (eds). Jerusalem: Israel Exploration Society, 2000.

Steck, O. H. *Die Erste Jesajarolle Von Qumran (1QIsa)*. Stuttgart: Verlag Katholisches Bibelwerk, 1998.

Stefanovic, Z. *The Aramaic of Daniel in the Light of Old Aramaic*. Sheffield: JSOT Press, 1992.

The International HapMap Consortium. *A haplotype map of the human genome*. *Nature* (2005) **437**, 1299-1319.

Thomas, S. I. *The "Mysteries" of Qumran: Mystery, Secrecy, and Esotericism in the Dead Sea Scrolls*. Atlanta, GA: Society of Biblical Literature, 2009.

Tov, E. The orthography and language of the Hebrew scrolls found at Qumran and the origin of these scrolls. *Textus* (1986) **XIII**, 31-57.

Tov, E. *Textual Criticism of the Hebrew Bible*. 2nd revised ed. Minneapolis: Fortress Press, 2001a.

Tov, E. A categorized list of all the 'Biblical' texts found in the Judaean desert. *Dead Sea Discoveries* (2001b) 8, 67-84.

Tov, E. *The Biblical Texts from the Judaean Desert*. In: *The Bible as Book: the Hebrew bible and the Judaean Desert discoveries*. Herbert, E. D. & Tov, E (eds). London: The British Library, 2002.

Tov, E. *Scribal Practices and Approaches Reflected in the Texts found in the Judean Desert*. Brill, Boston MA, 2004.

Trebolle-Barrera, J. *Qumran evidence for a biblical standard text and for non-standard and parabiblical texts*. In: *The Dead Sea Scrolls in their Historical Context*. Lim, T. H. (ed). Edinburgh: T.&T. Clark, 2000.

Trocme, É. *The Childhood of Christianity*. London: SCM Press, 1997.

Ulrich, E. C. *The Qumran Biblical Scrolls – the scriptures of late Second Temple Judaism*. In: *The Dead Sea Scrolls in their Historical Context*. Lim, T. H. (ed). Edinburgh: T.&T. Clark, 2000.

Ulrich, E. *The text of Daniel in the Qumran scrolls*. In: *The Book of Daniel: Composition and reception*, Vol. 2, Collins, J.J. & Flint, P.W. (eds). Leiden: Brill, 2001.

van de Merwe, C. H. J. *Lexical Meaning in Biblical Hebrew and Cognitive Semantics: a Case Study*. *Biblica* (2006) 87, 85-95.

VanderKam, J. & Flint, P. *The Meaning of the Dead Sea Scrolls*. T&T Clark International, London, 2002.

Vegas Montaner, L. *Biblia del Mar Muerto: Profetas Menores*. In: Edición crítica según manuscritas hebreas procedentes del Mar Muerto. Madrid: Instituto 'Arias Montano'. 1980.

Vegas Montaner, L. Computer-assisted study of the relation between *IQpHab* and the ancient (mainly Greek) biblical versions. *Revue Qumran* (1989) 14, 307-323.

Vermes, G. *The Complete Dead Sea Scrolls in English*. Revised ed. London: Penguin Books, 2004.

Wevers, J. W. The Dead Sea Scrolls and the Septuagint. *Bull Int Org Septuagint & Cognate Studies* (2005) 38, 1-24.

Wise, M. O. *Accidents and accident: A scribal view of linguistic dating of the Aramaic scrolls from Qumran*. In: *Studies in Qumran Aramaic*, Muraoka, T. (ed). *Abr-Nahrain* (1992) **Supplement 1**, 1-21.

Wong, S. S. M. *A Classification of Semantic Case-Relations in the Pauline Epistles*. New York: Peter Lang Publishing, Inc., 1997.

Wright, B. G. A note on the statistical analysis of Septuagintal syntax. *Journal of Biblical Literature* (1985) **104**, 111-114.

Wright, B. J. *No Small Difference: Sirach's Relationship to Its Hebrew Parent Text*. Atlanta: Scholars Press, 1989.

Young, I. Late Biblical Hebrew and the Qumran Peshar Habakkuk. *Journal of Hebrew Scriptures* (2008) **8**, 25.

Young, I. & Rezetko, R. Linguistic dating of biblical texts, 2 vols. London-Oakville: Equinox, 2009.

Appendix 1

Glossary of statistical terms

Chi square value. This is a measure of difference used for data that are counted in different categories (e.g. ten oranges, thirteen pears etc).

Cluster analysis. This is a statistical method of describing similarities between texts. The technique consists of estimating distances between texts according to a basket of variables. There are several ways of calculating such distances depending on the types of variables involved (interval, ordinal, binary etc).

Correspondence analysis. This is statistically related to χ^2 testing in that it deals with variables that are counts of different categories rather than data that are normally distributed. Standard correspondence analysis uses χ^2 to estimate the distance (extent of difference) between cases (e.g. manuscripts).

Discriminant analysis. This statistical technique determines the variables (features) that best distinguish between two or more groups of cases. In practice it has many limitations in its use (e.g. class means and covariances are unknown and have to be estimated, limited statistical power).

Factor analysis. This is a statistical method that provides a simplified summary of related variables. It identifies common factors that these related variables share. For

example, different meals might share common factors of spiciness, sweetness, acidity etc. even though the specific ingredients will differ between them. Factor analysis can be orthogonal, where the factors are unrelated to each other, or oblique, where the extracted factors are related to each other to some degree.

Finite mixture modelling. Many statistical tests assume population homogeneity, that is the sample is taken from a single population (e.g. treating the books of the Hebrew bible as a single population. Mixture modelling, by contrast, allows for a mixture of populations and does not assume homogeneity (e.g. it would allow for different characteristics between Torah books and the Prophets). Random effects models are an example of a mixture (or mixed) model. Finite models are restricted to situations where there is a finite number of sub-populations.

Hypothesis. A proposed explanation for a phenomenon. For example, if we thought that some texts had more nouns in than others, the hypothesis would be that the first set of texts was more likely to have a higher proportion of nouns than the other group. The null hypothesis would be that there is no difference. Conventionally, we use statistical calculations to test whether the null hypothesis can be rejected. If the probability of there being no difference is less than 1 in 20, we may infer that there is a statistically significant difference between the proportion of nouns in each group of texts.

Latent class analysis. This is an example of a finite mixture model where the number of sub-populations (classes) is pre-specified. Models with different numbers of classes can

be compared to determine which fits the data best. Models in which all variables are at interval level are termed Latent Profile Analyses. One feature of Latent Class Analysis is the ability to fix class membership of a proportion of cases and then estimate the probability of the remaining cases belonging to each of the pre-specified classes.

Loading plot. In factor analysis a loading plot displays how each variable in a factor analysis ‘loads’ onto each extracted factor. A factor loading indicates how close a variable is to a factor (0 indicates no relationship, 1 indicates the variable and factor are identical).

Markov chain. This refers to a type of analysis where there is ordered dependence of data

Mean. Means are one type of ‘average’; they are probably the most common type of average used in statistical analyses. The mean number of letters per word is calculated by summing all the letters for all of the words and dividing this grand total by the number of words in the book.

Multidimensional scaling. This is a way to display differences or similarities between either variables or cases. The algorithm considers the distances (i.e. the extent of difference) between variables and estimates the optimal spacing between variables in multi-dimensional space. The number of dimensions (e.g. two or three) has to be pre-specified.

Multivariate analysis of variance. This form of analysis of variance is termed multivariate because there is more than one dependent variable. Effects of factors or covariates are tested for significant effects on the shared variance of the dependent variables. For example, the effect of text (factor) on word length, syllable number, percentage of nouns in the construct state etc. (dependent variables) adjusted for text length (covariate). There is no assumption that dependent variables are independent, an important consideration when investigating style. Multivariate analysis of variance can test not only for significant effects on the shared variance of dependent variables, but also if there are significant differential effects between dependent variables.

Normal (sometimes termed 'Gaussian') distribution. This is a symmetrical distribution about the mean value with frequencies of observed variable values diminishing and tailing off the further from the mean they are.

Ordinal variables. These contain data that are ranked (1st, 2nd, 3rd etc) rather than having absolute values. Data that have absolute values can, of course, be allocated a rank (highest through lowest) to transform them into ordinal data. A different set of statistical analyses are appropriate to ordinal data.

Pearson correlation coefficient. This correlation coefficient is used when variables have a normal distribution is named after the British statistician Karl Pearson. It supposes that two variables have a linear relationship such that $y=bx+a$ where y and x are the two

variables, b is the gradient of the line and a is the intercept. If the observed data are plotted and the best line estimated mathematically (i.e. a and b estimated from the observed values of x and y), the correlation coefficient expresses how close the observed data group around that line. It is thus a measure of the strength of the relationship.

Power. Statistical power is the percentage likelihood of detecting an effect of a given size in a specific sample size at a pre-specified level of significance (e.g. $p=0.05$) if it is truly there in the population being sampled.

Principal components analysis. This is a form of factor analysis which seeks to apportion shared variance between variables to one or more principal component. The data reduction is termed orthogonal because the components are calculated so as not to correlate with each other.

Random effects models. These assume that the sample is a random sample from a larger population. The means and distributions may therefore differ between groups.

Significance level. The probability value arbitrarily chosen to represent what is unlikely by chance. When the probability is below this level, we say that it is statistically significant.

Standard deviation. Formally the standard deviation is given by the equation

$$\sigma = \sqrt{\text{E}[(X - \mu)^2]}.$$

where σ is the standard deviation, X is the variable and μ is the mean value of X . In practice this means subtracting the mean from each value of X , squaring it and summing the total for each value of X , then finally taking the square root of the total sum. It is a measure of how much variation there is in the values of any observed variable and is often used to express the size of statistical effects (the amount of influence of one variable on another).

Stochastic processes. These are situations where small numbers of events occur at random. Because the number of events that can be observed are so small, special statistical techniques are required to test hypotheses, usually involving some form of simulation of the events over and over again.

Stress. This is a measure of how well a multidimensional scaling model fits the observed data. The closer to zero the stress value, the better the model fits the observed data.

Type 1 statistical error. The more hypotheses that are tested, the greater the chance of a statistically significant result (i.e. for a significance level $p=0.05$, testing 20 hypotheses should give one 'statistically significant' result by chance). The potential error in interpreting such multiple hypotheses testing as probably being different is termed Type 1 statistical error.

Type 2 statistical error. This occurs when the sample size is too small to provide a tight estimate of the mean of the population the sample is drawn from. Hence comparing the estimated mean to another is unlikely to show a difference because the margins of error for the respective means will overlap.

Variance. This is the square of the standard deviation. It is a measure of how much variation there is in the values of any observed variable

Appendix 2

Supplementary analyses of the Hebrew Bible

Statistical considerations of vocabulary richness for individual books of the Hebrew bible

This Appendix provides detailed statistical analyses that relate to Chapter 6, the Statistics chapter. These analyses are not essential to the thesis itself, but are illustrative. They draw on the Hebrew bible as a suitable corpus in a Semitic language as a ‘proof of concept’ of the statistical techniques that will be applied to the Aramaic texts at Qumran. The analyses also assist in determining the likely statistical power of these different techniques when applied to the Aramaic corpus.

Vocabulary richness of individual books of the Hebrew bible

Table 1 shows the number of different lemmas per book. Figure 1 plots the relationship between the percentage of different lemmas and the total word count for each book. With the caveat that shorter books may contain less narrative and also be later, with a greater Hebrew vocabulary to draw upon, longer books tend to have a less varied vocabulary. The relationship is exponential as demonstrated by \log_e transforming the total number of words per book (Figure 2).

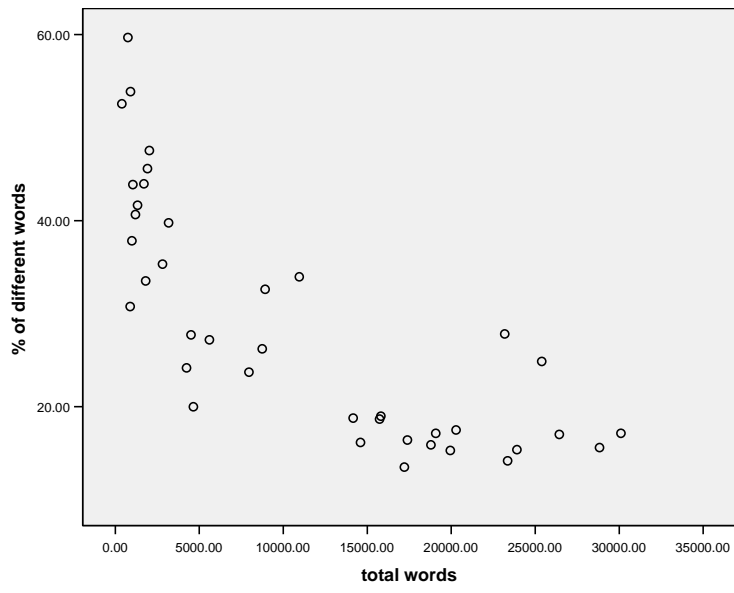


Figure 1 Percentage of different lemmas per total word count plotted against total word count for Hebrew bible books.

Book	Number of different lemmas/total words (%)	Book	Number of different lemmas/total words
Genesis	4,498/28,832 (15.6%)	Ecclesiastes	1,025/4,240 (24.2%)
Exodus	3,677/23,913 (15.4%)	Song of Songs	748/1,702 (43.9%)
Leviticus	2,324/17,208 (13.5%)	Isaiah	6,448/23,184 (27.8%)
Numbers	3,314/23,356 (14.2%)	Jeremiah	5,160/30,104 (17.1%)
Deuteronomy	3,549/20,284 (17.5%)	Lamentations	963/2,026 (47.5%)
Joshua	2,358/14,597 (16.2%)	Ezekiel	4,500/26,434 (17.0%)
Judges	2,659/14,161 (18.8%)	Daniel	2,293/8,745 (26.2%)
Ruth	605/1,805 (33.5%)	Hosea	1,262/3,174 (39.8%)
1Samuel	3,269/19,079 (17.1%)	Joel	552/1,325 (41.7%)
2Samuel	2,937/15,737 (18.7%)	Amos	994/2,814 (35.3%)
1Kings	2,988/18,791 (15.9%)	Obadiah	206/392 (52.6%)
2Kings	2,856/17,392 (16.4%)	Jonah	373/986 (37.8%)
1Chronicles	3,002/15,814 (19.0%)	Micah	873/1,915 (45.6%)
2Chronicles	3,050/19,945 (15.3%)	Nahum	447/749 (59.7%)
Ezra	1,524/5,606 (27.2%)	Habakkuk	488/906 (53.9%)
Nehemiah	1,887/7,958 (23.7%)	Zephaniah	458/1,044 (43.9%)
Esther	929/4,646 (20.0%)	Haggai	272/884 (30.8%)
Job	3,720/10,955 (34.0%)	Zechariah	1,250/4,509 (27.7%)
Psalms	6,313/25,388 (24.9%)	Malachi	485/1,193 (40.7%)
Proverbs	2,910/8,923 (32.6%)		

Table 1 Number of different/total words per Hebrew bible book.

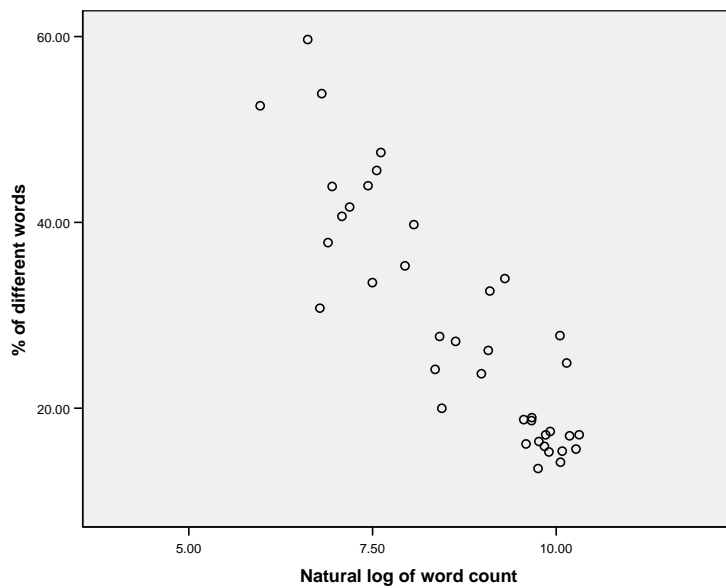


Figure 2 Percentage of different lemmas per total word count plotted against \log_e transformed total word count for Hebrew bible books.

Unsurprisingly, the total word count of any book accounts for 84.7% of the variance¹ in the number of different words in the book: the greater the total number of words, the more different words there are. The attenuation of the number of different words as books become longer can be described statistically by \log_e total word count accounting for 77.3% of the variance in the percentage of different words a book has: this time the relationship is a negative one. The standardised $\beta = -.883$ with 95% confidence intervals -

¹ Variance is a statistical measure of the variation of an observed variable, in this case the number of different words (lemmas) in a book. Such variation can be thought to comprise various components: measurement error (unlikely here), chance, and the effects of one or more causal factors (e.g. sex influences how height varies, mixed sex samples are likely to have a greater variation in height than single-sex samples).

.768 to -.998.² The inference that can be drawn is that, once the total word count is adjusted for, less than one quarter of variability in vocabulary between Hebrew bible books can be accounted for by other factors, such as genre, date of writing, authorship etc. At most less than half of variability may be accounted for, and at least, a vanishingly small amount.

Morphological characteristics of individual books of the Hebrew bible

The commonest individual morphological feature is *waw* which can be used as a conjunction in various ways. Arnold and Choi note that

Whether a clause is coordinate or subordinate is often interpretive because of the high frequency of the *waw consecutive* in verbal coordination.³

They take *w* and *'w* as the only coordinate conjunctions. The number of true conjunctions is moot in unpointed texts, as Arnold and Choi point out, because it forms part of the preterite prefix, so is not suitable for quantitative studies of a large corpus, though it may be useful for pointed texts such as the MT. Of the 321 occurrences of *'w* in the Hebrew bible, 137 occur in Leviticus, 44 in Numbers, 35 in Exodus and 34 in Deuteronomy, with single figures for other books. It is possible that effect sizes would

² The standardised β is the slope (gradient) of the line describing the relationship between two variables once the effects of different magnitudes of measurement units have been adjusted for (i.e. the standardised β will be the same for height versus weight whether height were measured in centimetres or inches).

³ B.T. Arnold & J.H. Choi (2003), p143.

be adequate, as indicated by the lexical power calculations for the direct object marker 't, but this may only apply to a limited number of texts.

Table 2 shows the frequencies of bound and independent pronouns for the entire Hebrew bible. The overall index of unbound to bound plus unbound suggested by Radday is 16.1%. It should be noted that this index does not represent a morphological choice (i.e. a choice between using an unbound or bound pronoun in any given clause) since unbound pronouns are often emphatic and are frequently implied by verb prefixes or suffixes. Hence, it could be argued that this index is more closely related to syntactic criteria. The numbers of pronouns suggest that there may be adequate power to consider these independently. Table 3 shows the number of independent pronouns by book in the Hebrew bible and Figure 3 plots these by total word count per book. Unlike the number of different lemmas, the relationship is remarkably linear even for short books with total word count accounting for 88.21% of variance in number of pronouns. The number of independent pronouns per book can be estimated as $-5.5 + (1.2\% \text{ of total words})$.

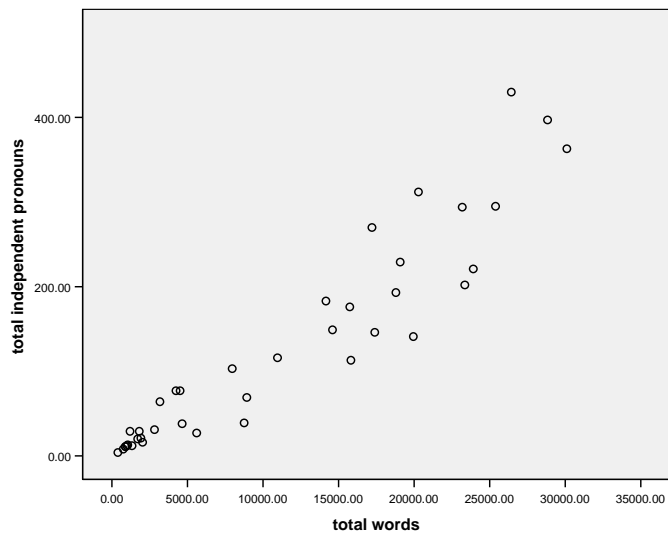


Figure 3 Number of independent pronouns by book word count.

Form	1 st person	2 nd person	3 rd person	None	Total
Noun	4,883	6,214	12,488	83,798	107,383
Perfect	375	256	554	13,429	14,614
Imperfect	376	390	871	14,229	15,866
Preterite	113	37	876	13,947	14,974
Waw- perfect	33	118	401	5,562	6,114
Imperative	256	1	136	3,861	4,254
Infinitive construct	260	306	912	5,087	6,565
Participle	219	134	355	7,460	8,335
Passive participle	3	3	59	926	991
Independent pronoun	1,359	1,092	2,490		4,941

Table 2 Frequencies of pronouns bound to nouns, verbs (by aspect) and independent in the Hebrew bible.

Book	Number of independent pronouns (%total words)	Book	Number of independent pronouns (%total words)
Genesis	397 (1.4%)	Ecclesiastes	77 (1.8%)
Exodus	221 (0.9%)	Song of Songs	20 (1.2%)
Leviticus	270 (1.6%)	Isaiah	294 (1.3%)
Numbers	202 (0.9%)	Jeremiah	363 (1.2%)
Deuteronomy	312 (1.5%)	Lamentations	16 (0.8%)
Joshua	149 (1.0%)	Ezekiel	430 (1.6%)
Judges	183 (1.3%)	Daniel	39 (0.4%)
Ruth	29 (1.6%)	Hosea	64 (2.0%)
1Samuel	229 (1.2%)	Joel	12 (0.9%)
2Samuel	176 (1.1%)	Amos	31 (1.1%)
1Kings	193 (1.0%)	Obadiah	4 (1.0%)
2Kings	146 (0.8%)	Jonah	12 (1.2%)
1Chronicles	113 (0.7%)	Micah	21 (1.1%)
2Chronicles	141 (0.7%)	Nahum	8 (1.1%)
Ezra	27 (0.5%)	Habakkuk	11 (1.2%)
Nehemiah	103 (1.3%)	Zephaniah	13 (1.2%)
Esther	38 (0.8%)	Haggai	11 (1.2%)
Job	116 (1.1%)	Zechariah	77 (1.7%)
Psalms	295 (1.2%)	Malachi	29 (2.4%)
Proverbs	69 (0.8%)		

Table 3 Number of independent pronouns (% of total words) per Hebrew bible book.

Power calculations show that even at $\alpha=.01$ there is almost 100% power to detect an effect size of 0.1%. Thus this criterion would be able to detect differences between most books of the Hebrew bible.

The definite article may be a useful stylistic discriminator if differences between the Masoretic text(s) and the Samaritan Pentateuch are a guide.⁴ Table 4 shows that frequencies are considerably greater than those of independent pronouns with total word count of book accounting for 86.0% of variance. Again the relationship between the number of definite articles and total word count per book is essentially linear (Figure 4).

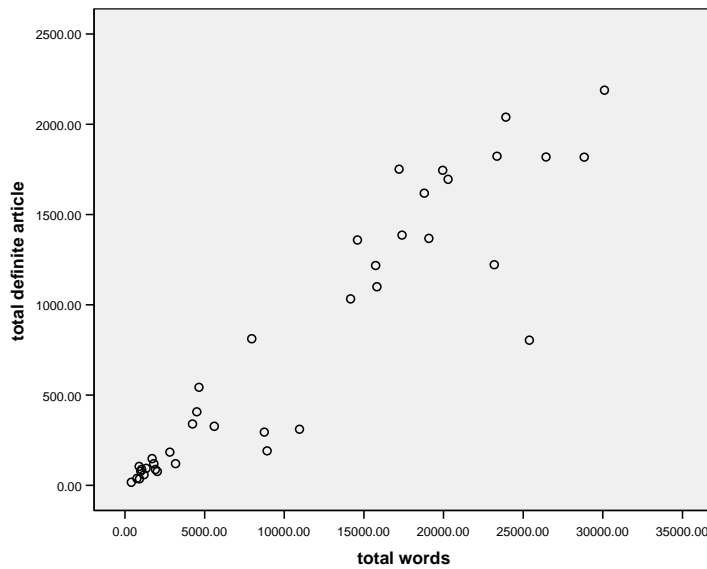


Figure 4 Total number of definite articles versus total word count.

⁴ S. Schorch (2003).

Hence, this criterion is also promising for texts of 1,000 words or more. Power calculations setting $\alpha=.01$ allowing 10% error variance provides 91.3% power to detect 1% difference for N=300, 84.4% for N=250 and 73.4% for N=200. For $\alpha=.05$, power is 89.5% for N=200, indicating that the criterion is suitable for relatively short fragments.

The ratio between common nouns in the construct and absolute state is normally distributed (Kolmogorov-Smirnov statistic =.103, $p>.20$,⁵ Figure 5) with a mean and median value of 1.07 construct to absolute state nouns and a standard deviation of .27. Using these values, power calculations show that at $\alpha=.01$, a difference of 0.1 between means comparing a text of 1,000 words with one of 100 words, there will be an 83.0% probability of detecting a real difference. That means that if there is a fragment of only 100 words from a larger text with a mean ratio of 1.17, there will still be over an 80% chance of detecting this using the fragment compared with a text which has the typical mean ratio of 1.07 for construct to absolute noun states. Thus a 100 word fragment taken at random from 29 of the 39 books of the Hebrew bible would be found to differ significantly from the mean for the entire bible eight out of every ten times

⁵ Kolmogorov-Smirnov statistic is a measure of how much a distribution of a variable differs from a normal distribution. If $p>0.05$ it can be taken that the distribution does not depart significantly from normality.

Book	Number of independent pronouns (%total words)	Book	Number of independent pronouns (%total words)
Genesis	1,818 (6.3%)	Ecclesiastes	390 (9.2%)
Exodus	2,039 (8.5%)	Song of Songs	148 (8.7%)
Leviticus	1,751 (10.2%)	Isaiah	1,222 (5.3%)
Numbers	1,823 (7.7%)	Jeremiah	2,189 (7.3%)
Deuteronomy	1,695 (8.4%)	Lamentations	77 (3.8%)
Joshua	1,359 (9.3%)	Ezekiel	1,819 (6.9%)
Judges	1,033 (7.3%)	Daniel	295 (3.4%)
Ruth	120 (6.6%)	Hosea	120 (3.8%)
1Samuel	1,368 (7.2%)	Joel	94 (7.1%)
2Samuel	1,218 (7.7%)	Amos	184 (6.5%)
1Kings	1,618 (8.6%)	Obadiah	17 (4.3%)
2Kings	1,386 (8.0%)	Jonah	78 (7.9%)
1Chronicles	1,100 (7.0%)	Micah	88 (4.6%)
2Chronicles	1,745 (8.7%)	Nahum	38 (5.1%)
Ezra	327 (5.8%)	Habakkuk	36 (4.0%)
Nehemiah	812 (10.2%)	Zephaniah	88 (8.4%)
Esther	543 (11.7%)	Haggai	105 (11.9%)
Job	311 (2.8%)	Zechariah	407 (9.0%)
Psalms	804 (3.2%)	Malachi	59 (4.9%)
Proverbs	191 (2.1%)		

Table 4 Frequency of the definite article (% of total words) per Hebrew bible book.

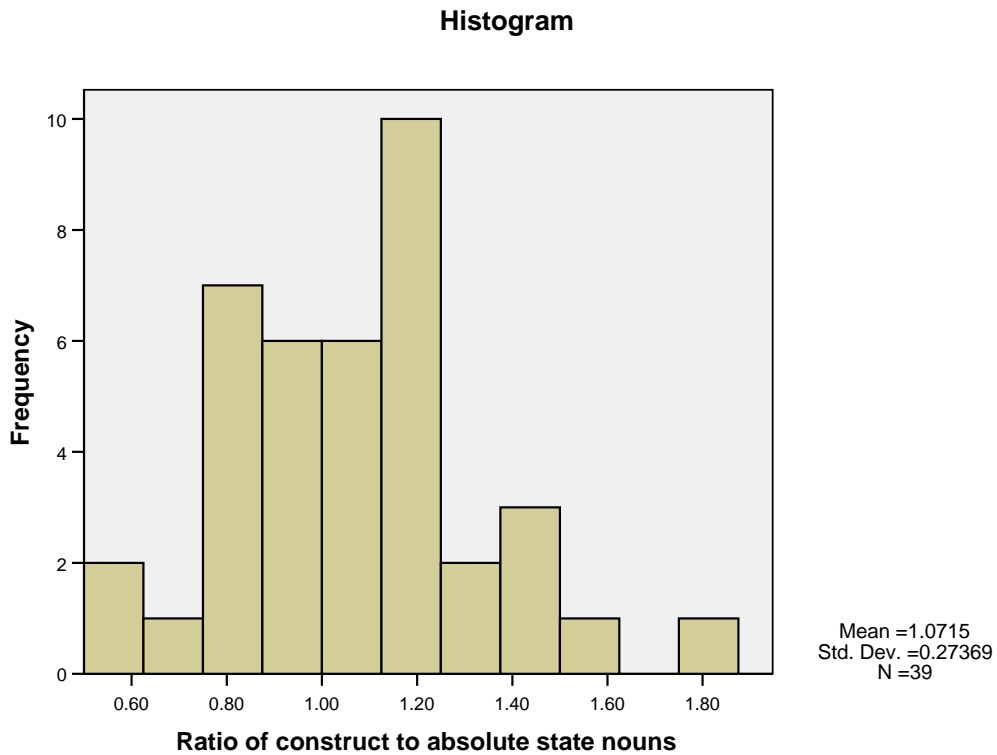


Figure 5 Frequencies of the number of Hebrew bible books with different ratios of construct to absolute state nouns.

Since using independent pronouns alone should have adequate power for most texts, Radday's suggested sums of pronoun types will also be adequately powered. So the last of Radday's morphological criteria to consider is verbs. Some initial data for the entire Hebrew bible have already been presented in Table 2. Adding in Infinite absolute forms (N=872), not shown in Table 2 because they do not take bound suffixes, there is a total of 51,568 finite forms and 21,017 non-finite forms. The magnitude of these numbers

indicates that, given the power calculations for nouns, texts with 1,000 words will be more than adequately powered to test verbal morphological criteria.

Exploratory factor analysis of individual books of the Hebrew bible

Variables entered into the analysis for each book:

- 1 Number of different lemmas per \log_e word count (vocabulary richness).
- 2 Ratio of construct to absolute state nouns.
- 3 Noun to verb ratio.
- 4 Unbound to bound pronoun ratio.
- 5 Independent pronoun percentage.
- 6 Definite article percentage.
- 7 Proper noun percentage.
- 8 Direct object marker percentage.
- 9 Pronouns bound to nouns or verbs percentage.
- 10 Persuasive verb forms (imperative, jussive, cohortative) percentage.
- 11 Preterite percentage.
- 12 *'šr* percentage.
- 13 *'m* percentage.
- 14 *ky* percentage.
- 15 All conjunctions percentage.
- 16 *hnh* percentage.
- 17 *whnh* percentage.

- 18 Passive verb forms percentage.
- 19 Non-finite to finite verb ratio.
- 20 Infinitive absolute followed directly by finite verb percentage.
- 21 Perfect verb followed by perfect plus *w* percentage x 10.
- 22 Perfect followed by preterite percentage x 10.

Of the 22 variables, five directly relate to nouns, three to pronouns, seven to verbs, six to particles, and one to both nouns and verbs. Thus there is a relatively even spread representing the different parts of speech.

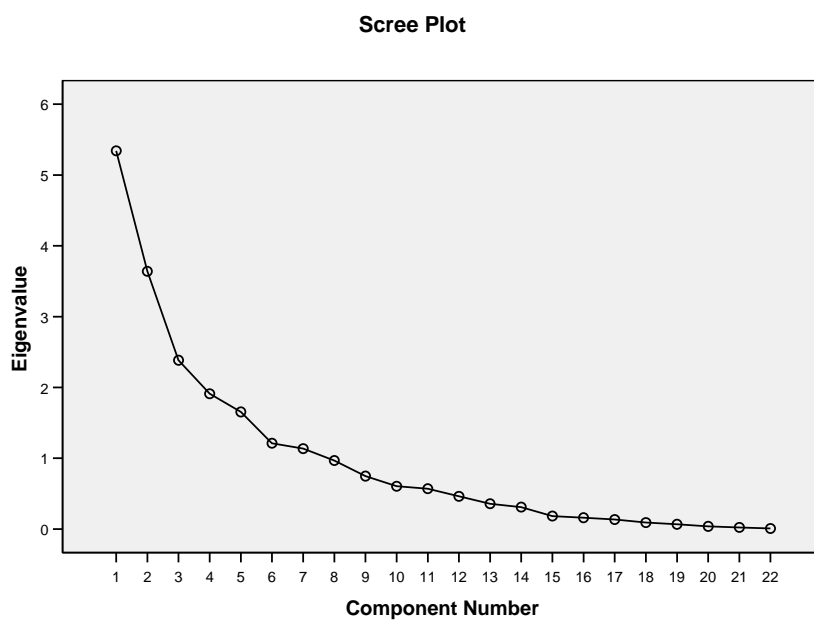


Figure 6 Scree plot of unrotated principal components of the rates of the 22 textual variables in the Hebrew bible.

Figure 6 shows the scree plot of the initial unrotated principal components analysis. Five factors stand out above the general scree slope. Together these five components explain 67.9% of variance of the 22 variables' values. The first component explains 24.3% of variance, the second 16.5%, the third 10.8%, the fourth 8.7% and the fifth 7.5%.

Table 5 shows the communalities, equivalent to the correlation coefficient between the first rotated component that accounts for around a quarter of variance in the textual variable scores, rotated using a varimax algorithm⁶ to align the components optimally with the contributing variables, and the 22 variables ranked in order. Those variables at the top or bottom of the Table are most informative of the component's dimension. The direction of association, positive or negative, relates to the magnitude of the correlated variables for each Hebrew bible book, but otherwise is arbitrary. The first component is thus a marker of determinate past action with qualifying, relative clauses, a

⁶ Principal components extraction provides what is termed an orthogonal solution because the components are derived to be uncorrelated with each other: in terms of vectors they can be considered to be at right angles with each other. Once the components have been extracted, this orthogonal reference frame can be rotated so that the rotated framework aligns as closely as possible to as many of the variables that went into the analysis as possible. Such a rotation is called 'varimax' because it maximises relationships between components and variables. This often makes it easier to label the components as the rotated components align more closely with specific variables.

Variable	Communalities with the first rotated component
Direct object marker %	.80
'šr %	.76
Preterite %	.76
All conjunctions %	.73
<i>whnh</i> %	.67
Definite article %	.49
Proper noun %	.43
Unbound to bound pronoun ratio	.41
Vocabulary richness	.26
Independent pronoun %	.24
Infinitive absolute followed directly by finite verb %	.17
<i>hnh</i> %	.02
Perfect verb followed by perfect plus <i>w</i> % x 10	-.003
Ratio of construct to absolute state nouns	-.01
Perfect followed by preterite % x 10	-.15
Non-finite to finite verb ratio	-.16
<i>ky</i> percentage	-.19
' <i>m</i> percentage	-.21
Noun to verb ratio	-.22
Passive verb forms %	-.26
Persuasive verb forms %	-.49
Pronouns bound to nouns or verbs %	-.56

Table 5 Communalities between stylistic variables and first rotated component.

predominance of transitive verbs and frequent use of proper nouns. The negative correlation coefficients are relatively smaller than the positive correlation coefficients, perhaps the only markers being those indicating a lack of personal pronouns bound to verbs or nouns and an absence of persuasive verbs. It is of interest that there is a negative correlation with perfect followed by preterite percentage x 10, suggesting that this dimension is not specifically related to temporal succession *per se*.⁷ The dimension is unrelated to the presence of exclamatory speech or immediacy of action and also unrelated to reflective or background passages. Associations with other textual variables fall somewhere in between.

Similarly Table 6 shows the correlation coefficients with the second rotated component that accounts for around one sixth of variance in the textual variable scores. The strongest positive correlation is with *ky* which has a wide semantic range. The next strongest positive correlation is with the use of the infinitive absolute followed by a finite verb, typical of emphatic statements. The strong negative correlation with noun to verb ratio indicates that this dimension has a strong verbal content, but the almost non-existent correlation with preterite verbal forms implies that this verbal dimension is not narrative in nature. Persuasive verbal forms are fairly strongly correlated. This is consistent with the strong negative correlation with proper nouns. Those nouns that are present are mostly in the absolute state indicating a lack of construct chains. There are

⁷ J. A. Cook (2004).

weak-to-moderate negative correlations with the determinate textual variables of the definite article and direct object marker. These associations, taken together, suggest that this second dimension has a rhetorical nature with emphatic and persuasive verbs linked to non-deterministic causal and evidential statements. A large proportion of the textual variables are essentially unrelated to this register.

Variable	Communalities with the second rotated component
<i>ky</i> %	.73
Infinitive absolute followed directly by finite verb %	.66
Persuasive verb forms %	.47
Independent pronoun %	.43
Pronouns bound to nouns or verbs %	.31
Perfect followed by preterite % x 10	.18
'š <i>r</i> %	.11
Unbound to bound pronoun ratio	.07
All conjunctions p%	.05
<i>hnh</i> %	.03
<i>whnh</i> %	.03
' <i>m</i> %	-.02
Preterite %	-.02
Perfect verb followed by perfect plus <i>w p</i> % x 10	-.03
Vocabulary richness	-.03
Passive verb forms %	-.06
Non-finite to finite verb ratio	-.17
Definite article %	-.22
Direct object marker %	-.25
Ratio of construct to absolute state nouns	-.68
Proper noun %	-.74
Noun to verb ratio	-.81

Table 6 Correlation coefficients between stylistic variables and second rotated component.

Table 7 shows the correlation coefficients with the third rotated component that accounts for just over one tenth of the variance in the textual variable scores. Strongest positive correlations are with *'m*, *hnh* and personal pronouns bound to nouns or verbs. Taken together, these associations indicate that this dimension is likely to relate to conditional, personal and exclamatory text. This would be consistent with a preponderance of finite over non-finite verbal forms. Like the second component or dimension, around half of the variables are unrelated to this dimension.

Table 8 shows the correlation coefficients with the fourth rotated component that accounts for just under one tenth of the variance in the textual variable scores. Hence we would expect correlation coefficients to be smaller on average. This is demonstrated by the large number of textual variables that correlate poorly with this component. Nevertheless, there is a very strong positive correlation with Perfect verb followed by perfect plus *w* percentage that Dobson considers typical of reflective and background passages. This dimension correlates with unbound pronouns, suggesting that, unlike the third dimension, it is not necessarily personal writing, and this would be consistent with a relative paucity of nouns in the construct state.

Variable	Communalities with the third rotated component
<i>'m</i> %	.61
<i>hnh</i> %	.60
Pronouns bound to nouns or verbs %	.59
Ratio of construct to absolute state nouns	.36
Passive verb forms %	.33
Vocabulary richness	.30
Perfect followed by preterite % x 10	.28
Persuasive verb forms %	.24
Independent pronoun %	.15
<i>ky</i> %	.11
Perfect verb followed by perfect plus <i>w</i> % x 10	.09
All conjunctions %	.07
<i>whnh</i> %	-.01
Direct object marker %	-.01
Infinitive absolute followed directly by finite verb %	-.03
Preterite %	-.03
Proper noun %	-.06
<i>'šr</i> %	-.18
Noun to verb ratio	-.37
Unbound to bound pronoun ratio	-.51
Definite article %	-.58
Non-finite to finite verb ratio	-.78

Table 7 Communalities between stylistic variables and third rotated component

Variable	Communalities with the fourth rotated component
Perfect verb followed by perfect plus <i>w %e</i> x 10	.87
Independent pronoun %	.69
Unbound to bound pronoun ratio	.65
<i>hnh</i> %	.43
<i>whnh</i> %	.24
All conjunctions %	.16
' <i>m</i> %	.08
Non-finite to finite verb ratio	.08
<i>ky</i> %	.06
Direct object marker %	.05
Definite article %	.02
Perfect followed by preterite % x 10	-.03
' <i>šr</i> %	-.09
Persuasive verb forms %	-.14
Noun to verb ratio	-.15
Proper noun %	-.16
Passive verb forms %	-.17
Pronouns bound to nouns or verbs %	-.20
Preterite %	-.24
Infinitive absolute followed directly by finite verb %	-.25
Vocabulary richness	-.28
Ratio of construct to absolute state nouns	-.36

Table 8 Communalities between stylistic variables and fourth rotated component.

Table 9 shows the correlation coefficients with the fifth rotated component that accounts for just under one twelfth of the variance in the textual variable scores. Again, this dimension is characterised by verbal forms: a strong positive correlation with passive forms and a strong negative correlation with perfect-preterite sequence and, to a lesser extent, also stand alone preterite forms. This dimension has therefore a relationship with narrative, but in a negative direction as the variables are counted. The preponderance of passive verbs may indicate a relationship with citation formulae such as 'it is written'. The relatively limited use of passive forms probably explains why its contribution to variance in all textual variable values is fairly low. Otherwise, it resembles the first dimension to a lesser degree with regard to a positive association with relative clauses and determinate content.

In summary, the Hebrew bible register determined by principal components analysis of the 22 textual variables counted on a book-by-book basis comprises five main dimensions:

- 1) Historical - determinate past action with subjects and objects;
- 2) Rhetorical – persuasive and evidential;
- 3) Personal exclamation with a conditional tendency;
- 4) Reflective – background text of a non-personal nature;
- 5) Passive – non-narrative text that refers to the effects of external agents

Variable	Communalities with the fifth rotated component
Passive verb forms %	.77
Definite article %	.44
'šr %	.27
Direct object marker %	.27
Independent pronoun %	.20
Pronouns bound to nouns or verbs %	.13
'm %	.11
Non-finite to finite verb ratio	.07
Noun to verb ratio	.06
Unbound to bound pronoun ratio	.05
Ratio of construct to absolute state nouns	.02
Vocabulary richness	.01
<i>whnh</i> %	-.02
Persuasive verb forms %	-.03
<i>hnh</i> %	-.13
Perfect verb followed by perfect plus <i>w</i> % x 10	-.16
<i>ky</i> %	-.19
Proper noun %	-.20
All conjunctions %	-.24
Infinitive absolute followed directly by finite verb %	-.30
Preterite %	-.34
Perfect followed by preterite % x 10	-.73

Table 9 Communalities between stylistic variables and fifth rotated component.

Having determined these dimensions, the next step is to examine which of the 22 variables contribute to them and, equally importantly, which do not. Variables that have a communality of $>.7$ or $<-.7$ will have at least 50% of the variance in their values accounted for by a dimension and are thus well represented. These variables are:

- 1 Direct object marker percentage (Dimension 1)
- 2 *’šr* percentage (Dimension 1)
- 3 Preterite percentage (Dimension 1)
- 4 All conjunctions percentage (Dimension 1)
- 5 *ky* percentage (Dimension 2)
- 6 Noun to verb ratio (Dimension 2)
- 7 Proper noun percentage (Dimension 2)
- 8 Non-finite to finite verb ratio (Dimension 3)
- 9 Perfect verb followed by perfect plus *w* percentage x 10 (Dimension 4)
- 10 Passive verb forms percentage (Dimension 5)
- 11 Perfect followed by preterite percentage x 10 (Dimension 5)

Variables that have no communalities $>.45$ or $<-.45$ with any of the variables have no more than 20% of their values explained by any one of the dimensions and are far less well represented. The only variable that meets this criterion is vocabulary richness. This suggests that vocabulary richness discriminates poorly between the different main dimensions. If it is of any value, it is measuring a different aspect of text from register. Since it is less affected by register it may be a useful measure for situations where

register may confound comparisons between texts. That is, it may be useful in comparing texts which have markedly different dimensions of register.

Hebrew bible books and their register dimensions

An important aim of factor analysis is to clarify relationships: seeing the wood for the trees. Having reduced the textual variables from 22 to five dimensions, it is easier to examine the textual style of the different books of the Hebrew bible, Daniel excepted. Without performing some sort of data reduction, such as principal components analysis, comparisons would have to be done variable by variable without a clear grasp of how the different textual variables might be related to each other.

Table 10 shows the principal components scores for Dimension 1 ranked in order of magnitude. Unsurprisingly, Genesis, Judges and 1Samuel score highly on this register dimension and Song of Songs, Lamentations and Proverbs score low. Interestingly, Leviticus scores higher than Exodus and Ezra, which might conventionally be considered an historical book, scores very low indeed. This dimension is the commonest in the Hebrew bible in terms of the stylistic variables considered; Genesis might therefore be considered as the paradigmatic book.

Book	Dimension 1 component score	Book	Dimension 1 component score
Genesis	1.62	1Chronicles	0.06
Judges	1.43	Ecclesiastes	-.03
1Samuel	1.38	Malachi	-.10
Ruth	1.37	Isaiah	-.30
2Kings	1.24	Amos	-.31
Leviticus	1.14	Hosea	-.40
2Samuel	1.06	Haggai	-.63
1Kings	1.06	Obadiah	-.71
Exodus	0.89	Job	-.72
Joshua	0.84	Zephaniah	-.79
Deuteronomy	0.79	Joel	-.91
Zechariah	0.61	Micah	-.93
Jonah	0.60	Habakkuk	-1.06
Ezekiel	0.52	Nahum	-1.28
2Chronicles	0.49	Psalms	-1.36
Numbers	0.37	Lamentations	-1.54
Jeremiah	0.31	Proverbs	-1.58
Esther	0.28	Ezra	-1.60
Nehemiah	0.11	Song of Songs	-1.91

Table 10 Books ranked by Dimension 1: Historical - determinate past action.

Book	Dimension 2 component score	Book	Dimension 2 component score
Jonah	1.86	Zephaniah	0.16
Ecclesiastes	1.71	Genesis	0.06
Joel	1.35	Ezekiel	0.02
Habakkuk	1.25	1Kings	-.18
Hosea	1.10	Judges	-.19
Ruth	0.90	2Samuel	-.21
Deuteronomy	0.86	Song of Songs	-.25
Psalms	0.66	Malachi	-.29
Lamentations	0.64	Esther	-.35
Job	0.61	Amos	-.44
Leviticus	0.53	2Kings	-.68
Isaiah	0.49	Haggai	-.77
Micah	0.46	2Chronicles	-.81
Nahum	0.33	Numbers	-.82
Proverbs	0.32	Joshua	-1.25
1Samuel	0.30	Nehemiah	-1.26
Exodus	0.26	Obadiah	-1.62
Zechariah	0.19	Ezra	-2.41
Jeremiah	0.19	1Chronicles	-2.75

Table 11 Books ranked by Dimension 2: Rhetorical – persuasive and evidential.

Table 11 shows the component scores for Dimension 2 ranked in order of magnitude. Rhetorical textual devices are common in Jonah, Ecclesiastes and some of the Minor Prophets, they are infrequent in 1Chronicles, Ezra and Nehemiah. This dimension is one that controls the flow of information: Table 11 provides some idea of the contribution of this dimension of evidence-related persuasion to different texts. This dimension appears to cut across genres more than Dimension 1 as both prose and poetry books score highly. Indeed, the lower scoring books are mainly prose texts.

Table 12 shows the component scores for Dimension 3 ranked in order of magnitude. High scoring books such as Job, Song of Songs and Psalms are unsurprising for this dimension which relates to personal exclamation and conditional statements. Comparing Dimensions 2 and 3 for the Minor Prophets is quite revealing in terms of the type of writing. Obadiah favours personal exclamation over persuasive evidence-related argument, the opposite is true for Joel and Jonah. It is also noteworthy that one 'Wisdom' book, Job, scores highly, whilst another, Ecclesiastes, scores low. Again Ezra scores very low on this dimension.

Book	Dimension 3 component score	Book	Dimension 3 component score
Obadiah	2.59	Micah	.10
Job	1.73	1Kings	.06
Malachi	1.11	Amos	-.01
Song of Songs	1.02	Deuteronomy	-.09
Genesis	1.02	Proverbs	-.12
Psalms	.86	1Chronicles	-.18
Isaiah	.72	Exodus	-.21
Ruth	.59	Joshua	-.29
Lamentations	.51	Leviticus	-.32
1Samuel	.47	Nehemiah	-.44
2Kings	.47	2Chronicles	-.46
Nahum	.41	Zechariah	-.50
Hosea	.37	Zephaniah	-.59
Ezekiel	.33	Joel	-1.07
Judges	.32	Esther	-1.35
Habakkuk	.29	Ecclesiastes	-1.52
2Samuel	.24	Jonah	-1.70
Jeremiah	.23	Ezra	-1.92
Numbers	.13	Haggai	-2.78

Table 12 Books ranked by Dimension 3: Personal exclamation with a conditional tendency.

Book	Dimension 4 component score	Book	Dimension 4 component score
Malachi	4.38	Proverbs	-.24
Haggai	1.90	Ruth	-.24
Zechariah	1.84	Micah	-.26
Amos	1.0	Song of Songs	-.30
Ecclesiastes	.97	Esther	-.34
Ezekiel	.66	Joshua	-.37
Hosea	.37	Zephaniah	-.42
Obadiah	.21	2Kings	-.47
Habakkuk	.19	Ezra	-.47
Jeremiah	.12	1Kings	-.54
Isaiah	.11	2Chronicles	-.59
Leviticus	.03	Joel	-.60
Job	.03	Jonah	-.73
Nahum	.02	Exodus	-.79
Judges	.00	1Chronicles	-.83
Nehemiah	-.11	Deuteronomy	-.83
Genesis	-.17	Numbers	-.84
2Samuel	-.21	Psalms	-.88
1Samuel	-.24	Lamentations	-1.38

Table 13 Books ranked by Dimension 4: Reflective – background text of a non-personal nature

Table 13 shows the component scores for Dimension 4 ranked in order of magnitude. Unlike the previous dimensions, Dimension 4 has a majority of variables with negative scores. There is not a great difference between the scores for most of the books, but Malachi stands out as containing a considerable amount of this background, reflective non-personal type of text. Other late Minor Prophets, Haggai and Zechariah also score highly. Psalms and Lamentations score low, indicating that background, non-personal text is less common in these books.

Finally, Table 14 shows the component scores for Dimension 5, passive – non-narrative text that refers to the effects of external agents, ranked in order of magnitude. This is the least common of the five dimensions, accounting for a relatively small proportion of variance in the 22 textual variable values, and produces some interesting comparisons. For example, Song of Songs and Deuteronomy have the highest scores. The former may reflect the passive role of the ‘Beloved’ and the latter the series of curses and blessings in chapters 28 and 29. In general books with narrative content score low as expected, though 1Kings scores unexpectedly highly with this regard.

Book	Dimension 5 component score	Book	Dimension 5 component score
Song of Songs	2.67	Psalms	-.02
Deuteronomy	1.63	Ruth	-.08
Leviticus	1.37	Zechariah	-.18
Obadiah	1.31	2Chronicles	-.27
Ecclesiastes	1.20	Genesis	-.30
Numbers	1.19	2Kings	-.38
Zephaniah	1.12	Proverbs	-.45
Esther	1.03	Job	-.54
Haggai	.65	Micah	-.55
Ezekiel	.53	2Samuel	-.78
Isaiah	.47	Malachi	-.86
Nehemiah	.47	1Samuel	-.90
1Kings	.41	Judges	-.92
Hosea	.31	Amos	-1.00
Joel	.16	Ezra	-1.04
Exodus	.08	1Chronicles	-1.13
Jeremiah	.00	Jonah	-1.33
Joshua	.00	Habakkuk	-1.56
Nahum	-.02	Lamentation	-2.30

Table 14 Books ranked by Dimension 5: Passive – non-narrative text that refers to the effects of external agents

Books contain more or less text reflecting these different dimensions. Since each dimension accounts for a different amount of variance in the textual variables examined, some weighting for this is appropriate when comparing books in terms of register. The weighting can most appropriately reflect the amount of variance explained by each component. Hence Dimension 1 is given around three times the weight of Dimensions 4 and 5.

Applying cluster analysis to the principal components scores to search for groups of similar Hebrew bible books

This exploratory principal components analysis has been performed as a ‘proof of concept’ for textual analysis before applying it to the Qumran Aramaic corpus. The textual character of the Hebrew bible has been thoroughly reviewed, though not statistically, by many scholars over the centuries. The present results indicate that principal components analysis can extract different textual registers which enables a preliminary inspection of which books of the Hebrew bible relate most closely to each other. The principal components scores can be used to produce a statistically more rigorous description of how the different books relate to each other. There are various options. If we knew that there were only two groups of texts of interest (e.g. for the New Testament letters, Pauline and non-Pauline), discriminant analysis might be applied.⁸

⁸ See Chapter 2 for a discussion of this statistical technique.

However, we do not know how many textual groups there might be. Indeed, we do not even know if, statistically, based on the five principal components that have been extracted, whether discrete groups exist at all. Instead, at this stage I will outline how cluster analysis can be used to search for discrete groups.

A hierarchical cluster analysis for the Hebrew bible books, excluding Daniel, is shown in Figure 7. Agglomeration distances suggest that the books may be considered in these small groups:

1. Genesis, Exodus, Leviticus, Deuteronomy, Judges, Ruth, 1Samuel, 2Samuel, 1Kings, 2Kings, Jeremiah, Ezekiel;
2. Numbers, Joshua, 2Chronicles, Nehemiah, Esther;
3. Amos, Zechariah;
4. Joel, Zephaniah;
5. Job, Isaiah, Hosea;
6. Psalms, Proverbs, Micah, Nahum, Habakkuk;
7. Lamentations;
8. Ecclesiastes;
9. Jonah;
10. Song of Songs;
11. Obadiah;
12. 1Chronicles;
13. Ezra;

14. Haggai;

15. Malachi.

Label	Num	+-----+-----+-----+-----+
Judges	7	↓↘
2Samuel	10	↓□
1Samuel	9	↓□
2Kings	12	↓↑↓↓↓↘
Genesis	1	↓↘ ⇔
Ruth	8	↓↓↓↓↓↑↓↓↓↘
Exodus	2	↓*↓↘ ⇔ ⇔
1Kings	11	↓↘ □↓↘ □↓↘
Jeremiah	24	↓*↓↘ ⇔ ⇔
Ezekiel	26	↓↘ ⇔ ⇔
Leviticus	3	↓*↓↓↓↓↓↓↓↓↘ ⇔
Deuteronomy	5	↓↘ □↓↓↓↘
Joshua	6	↓*↓↘ ⇔ ⇔
2Chronicles	14	↓↘ ⇔ ⇔ ⇔
Nehemiah	16	↓↓↓↓↓↑↓↘ ⇔ □↓↓↓↓↓↓↓↓↓↓↓↓↘
Numbers	4	↓↓↓↘ □↓↓↓↓↓↘ ⇔ ⇔
Esther	17	↓↓↓↓↓↘ ⇔ ⇔
Amos	30	↓↓↓↓↓↓↓*↓↓↓↓↓↓↓↓↘ ⇔
Zechariah	38	↓↓↓↓↓↓↓↘ ⇔
Joel	29	↓↓↓↓↓↓↓*↓↘ □↓↘
Zephaniah	36	↓↓↓↓↓↓↓↘ ⇔ ⇔ ⇔
Isaiah	23	↓*↓↓↓↘ □↓↓↓↓↓↘ ⇔ ⇔
Hosea	28	↓↘ □↓↘ ⇔ ⇔ ⇔ ⇔
Job	18	↓↓↓↓↓↘ ⇔ ⇔ ⇔ ⇔ ⇔
Micah	33	↓↘ □↓↘ ⇔ ⇔ □↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘
Nahum	34	↓↑↓↘ ⇔ □↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔
Proverbs	20	↓↘ □↓↘ ⇔ ⇔ ⇔ ⇔
Psalms	19	↓↓↓↘ □↓↘ ⇔ ⇔ ⇔ ⇔
Habakkuk	35	↓↓↓↓↓↘ ⇔ ⇔ ⇔ □↓↓↓↘
Lamentations	25	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔ ⇔
Ecclesiastes	21	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓*↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔
Jonah	32	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔ ⇔
⇔ Song of Songs	22	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓*↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘
⇔ Obadiah	31	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘ ⇔
⇔ 1Chronicles	13	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓*↓↓↓↓↓↓↓↓↓↓↓↓↘
⇔ Ezra	15	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘ □↓↓↓↓↓↓↓↓↓↓
⇔ Haggai	37	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘
⇔ Malachi	39	↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↘

Figure 7 Hierarchical cluster of Hebrew bible books using the five principal components derived from the 22 textual criteria.

If a greater degree of dissimilarity is allowed between books, groups 1-3 form a cluster, groups 4-7 a second cluster; Ecclesiastes and Jonah resemble each other; similarly, Song of Songs and Obadiah; 1Chronicles and Ezra; with Haggai and Malachi remaining quite distinct. Table 15 shows the mean principal component scores for each of these seven clusters.

Group	Dimension 1 Historical	Dimension 2 Rhetorical	Dimension 3 Personal exclamation	Dimension 4 Reflective background	Dimension 5 Passive
Gen etc	.80	-.15	.01	-.11	.10
Job etc	-.99	.67	.29	.12	-.31
Ecc & Jon	.28	1.8	-1.6	.12	-.06
Song & Obadiah	-1.3	-.93	1.8	-.04	1.99
1Chr & Ezra	-.77	-2.6	-1.1	-.65	-1.1
Haggai	-.63	-.77	-2.8	1.9	.65
Malachi	-.01	-.29	1.1	4.4	-.86

Table 15 Mean principal component scores for clusters of Hebrew bible books.

Gen etc = Genesis, Exodus, Leviticus, Deuteronomy, Judges, Ruth, 1Samuel, 2Samuel, 1Kings, 2Kings, Jeremiah, Ezekiel, Numbers, Joshua, 2Chronicles, Nehemiah, Esther, Amos, Zechariah.

Job etc = Job, Isaiah, Hosea, Psalms, Proverbs, Micah, Nahum,
Habakkuk, Lamentations.

The Genesis cluster, which comprises mostly books labeled as 'historical', scores highest on the historical dimension, with a fairly even balance across the other dimensions. Clearly, individual books within the group will deviate to some degree from the mean values. The Job cluster contrasts with the historical group, with relatively little in the way of historical text style. Instead this cluster is characterized by a rhetorical style, using persuasion and evidence. The other clusters are quite small, comprising one or two books only, and those clusters with two books have agglomeration distances indicating that the books differ substantially from each other. Ecclesiastes and Jonah are even more strongly rhetorical than the Job group, but more historical and less personally exclamatory. Song of Songs and Obadiah are characterized by a strong balance of personal exclamatory and passive dimensions, but have little historical or rhetorical text. 1Chronicles and Ezra are characterized by negative scores on all dimensions. Whatever text type they are characterized by, has failed to be captured by the five principal components. We can say that they are not typically historical, rhetorical, personal exclamatory, reflective or passive, but we cannot state what they are. This needs to be flagged up as an important outcome of this explanatory principal components analysis because there may be Qumran texts that also fail to be characterized by the register dimensions identified. Haggai and Malachi stand alone. Both contain reflective or background text, but Malachi exhibits a fair degree of personal exclamation whilst Haggai has very little of this dimension indeed.

Summary of exploratory principal components analysis and implications for the statistical analysis of Qumran texts

The first aim of this exploratory principal components analysis was to investigate whether textual variables identified from the literature would have enough shared variance to provide an informative set of principal components. This aim was achieved because around 70% of variance was explained by five components that stood out in the scree plot. There are two caveats. First, two books, 1Chronicles and Ezra, could not be classified in a positive way according to the five components or dimensions of register. If this were repeated for the Qumran corpus, we might expect around 5% of texts to lack a positive characterization. Secondly, one variable – vocabulary richness – appeared to contribute poorly to all of the components. This may have advantages for textual comparisons across texts with different register dimensions, but there may be a dimension that is strongly related to vocabulary richness which failed to be identified. If so, it accounts for only a small proportion of variance in the other 21 textual variables.

The second aim of the exploratory principal components analysis was to examine the validity of any identified dimensions. The construct validity is evident from the analyses themselves. Content validity depends on whether the dimensions relate to semantic categories recognized by scholars. This is not problematic for the commonest dimensions, historical and rhetorical, that typify the majority of books. There is more room for debate about the three less common dimensions that, although present to some

degree in many books, are typical of only a few. The personal exclamatory with a conditional tendency can be thought of as depicting the kind of texts associated with vows, oaths and promises. Song of Songs and Obadiah are typical of this kind of writing, but it is found in Malachi and, to a lesser degree, in the Job text cluster. This dimension, too, is recognized widely and was identified *a priori* when determining textual variables. The fourth dimension of reflective or background writing is also one that was determined *a priori*. It is present to some degree in the Job text cluster, but is more prominent in Haggai with its chapter one refrain of ‘Give careful thought to your ways’, but is most typified by the text of Malachi. The fifth and final dimension was not one identified *a priori*. I have characterized it as passive – non-narrative text that refers to the effects of external agents, noting that this can refer to a wide range of content. It accounts for the least variance of the textual variables of all the five components. It is not prominent in most books, but only in Haggai, Obadiah and Song of Songs. Two of these texts are quite short. There is thus some doubt over its content validity. It would not be surprising if a principal components analysis of the Qumran corpus found this component to be replaced by another.

The third aim, linked to the question of validity, was to assess whether register dimensions for Hebrew texts, as represented by the bible, were the same as those previously identified for English and Greek corpora. Biber’s dimensions, as discussed in Chapter 3, are:

- 1) informational vs involved production;

- 2) narrative vs non-narrative concerns;
- 3) explicit (situation-independent) vs situation-dependent reference;
- 4) overt expression of persuasion and
- 5) abstract vs non-abstract style.⁹

For the Hebrew bible, the Historical - determinate past action with subjects and objects – dimension appears to subsume Biber’s informational and narrative categories.

Interestingly, the grouping within the Genesis text cluster suggests that texts typified by the historical dimension can be sub-divided into more narrative or more informational, but that this may depend on the inter-play between dimensions. The second dimension for the Hebrew bible, Rhetorical – persuasive and evidential – appears to relate to Biber’s fourth dimension of overt expression of persuasion. Biber’s fifth dimension of abstract style may relate to the reflective or background dimension identified for the Hebrew bible, though an abstract style can encompass a wider range than covered by the Hebrew bible dimension. Perhaps this reflects a difference between corpora. A clear difference exists for the last categories to be considered. Situation independence does not appear as a clear dimension for the Hebrew bible texts. Instead a personal exclamatory with conditional tendencies dimension is more important. Finally, the passive – non-narrative text that refers to the effects of external agents – dimension does not align with any of Biber’s dimensions. In brief, there is some overlap between the dimensions identified for English and Greek corpora and those of the Hebrew bible, but

⁹ M. B. O’Donnell (2005), pp 82-83.

the shape of register of the Hebrew bible is also distinct in several ways. The implication for Qumran texts is that registers derived from substantially different corpora cannot be applied directly.

Cluster analysis

Cluster analysis has already been used to classify the books of the Hebrew bible when examining the validity of factor analysis. As noted, the five principal components explained around 70% of variance in the 22 textual variables. This means that 30% of information about the texts was not used in the cluster analysis procedure. For the classification Qumran texts cluster analysis is a statistical technique that will be worth exploring. Cluster analysis can be used to explore possible clusters of variables as well as clusters of cases. The additional 30% of information may be valuable when shorter texts are involved. Indeed it is notable that many of the texts that clustered poorly were relatively short. Use of the additional information available may have made a difference to this. It is important to know whether, if the text variable frequencies are similar in the Qumran corpus to the Hebrew bible, whether cluster analysis is likely to leave a significant proportion of small clusters that would be fairly uninformative. For example, there is little to be gained by having a cluster comprising of Haggai or Malachi on their own.

Clustering of textual variables

Some groupings of variables have been examined in the rankings for individual dimensions. However, as we have seen, groupings can change considerably when all dimensions are combined. Figure 8 shows the initial hierarchical clustering for all 22 variables using principal components. As indicated from the principal components analysis, vocabulary richness is quite distinct; the hierarchical cluster dendrogram of the 22 textual variables, Figure 8, supports this. To gain any useful information about the relationship of the other 21 variables, it needs to be excluded from further analyses. Of note is the finding that the other 21 variables are closely related, confirming the interdependence of these morphological textual variables. Vocabulary richness relates to all parts of speech, though it will reflect mainly nouns and verbs, so this inter-relationship would be less expected.

Figure 9 shows the cluster dendrogram for the 21 textual variables excluding vocabulary richness. Fifteen of the variables are closely related. The percentage of conjunctions is least closely associated with the other textual variables. The percentage of definite articles, proper nouns and bound pronouns comprise another group, though not as closely related as the main cluster. These relationships would be difficult to infer from the ranking of the 21 variables on each register dimension. The implication for the Qumran text analysis is that cluster analyses of morphological variables may illuminate different text groups. This will be examined once the text group clusters are established.

Clustering of Hebrew bible books

Hierarchical clustering using all 22 textual variables produces five clusters (Figure 10):

- 1 Exodus, Numbers, Deuteronomy, Judges, 1Samuel, 2Samuel, 1Kings, 2Kings, 1Chronicles, 2Chronicles, Job and Proverbs;
- 2 Leviticus, Joshua and Nehemiah;
- 3 Genesis, Jeremiah and Ezekiel;
- 4 Ruth, Ezra, Esther, Ecclesiastes, Song of Songs, Lamentations, Hosea, Joel, Amos, Obadiah, Jonah, Micah, Nahum, Habakkuk, Zephaniah, Haggai, Zechariah and Malachi;
- 5 Psalms and Isaiah.

The increased information compared with the clustering based on the five principal components has helped to group the Minor Prophets. On the other hand the so-called historical books have been split with Job and Proverbs included in what might be perceived as a Deuteronomistic group and Genesis clustered with Jeremiah and Ezekiel. The implication for Qumran text analysis and classification is that using all variables for clustering is likely to be helpful for shorter texts, but that a more focused approach, based on initial principal components analysis, may be more useful where there is more than adequate textual information.

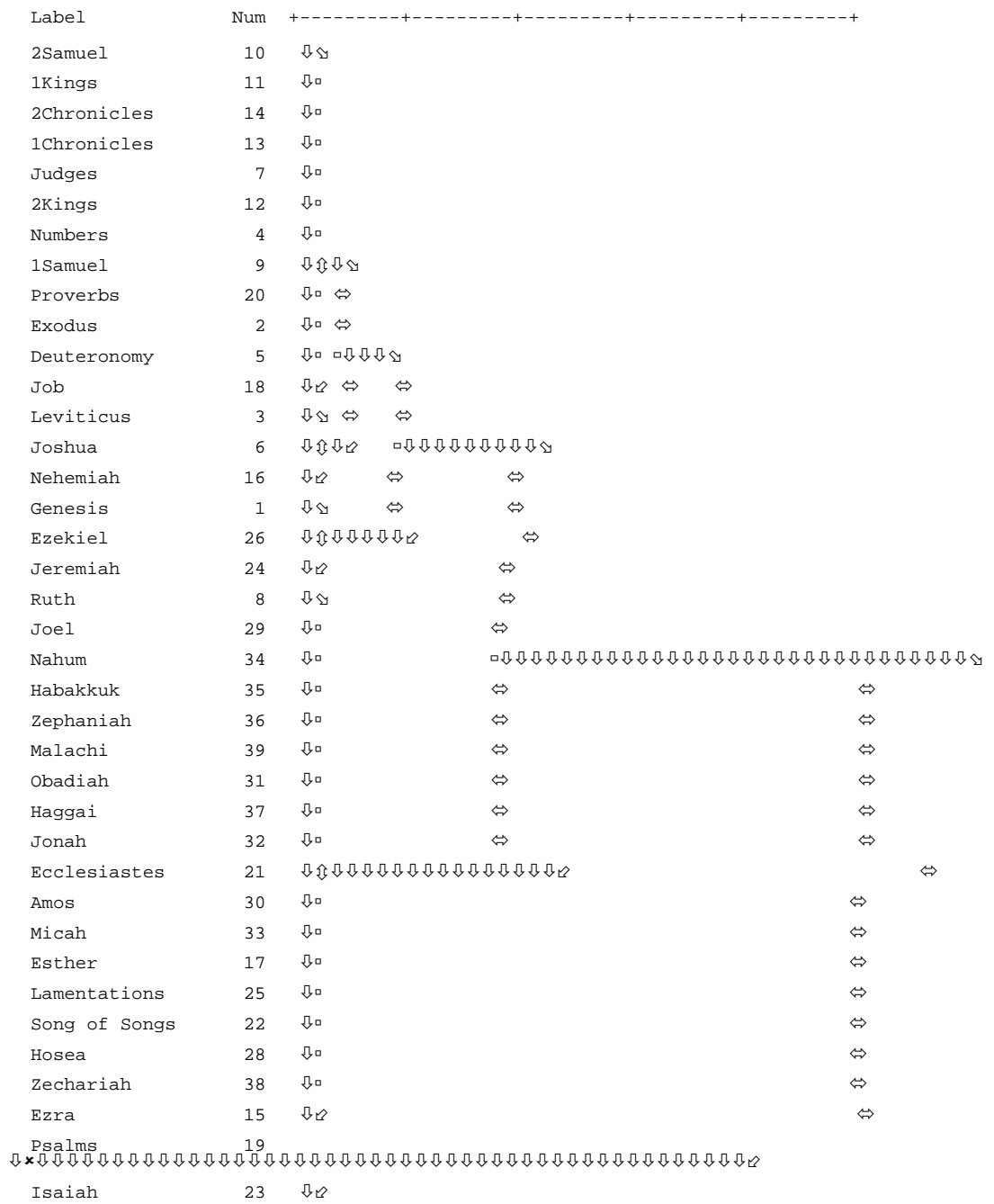


Figure 10 Hierarchical cluster analysis of the Hebrew bible books using raw scores of the 22 textual variables.

Clustering of textual variables within Hebrew bible book groups

The close association between many of the textual variables has been demonstrated above. In this final section on cluster analysis I will examine whether these relationships are similar in different text groups. An adequate number of texts are required for such comparisons. Hence, for exploratory purposes Deuteronomistic books common to both the clusters produced from the five dimensions and all the 22 textual variables (Exodus, Deuteronomy, Judges, 1Samuel, 2Samuel, 1Kings, 2Kings,) and the Minor Prophets are considered.

Using all 22 textual variables, clustering for both the Deuteronomistic and Minor Prophets groups reveals vocabulary richness to be distinct from all the other variables. Excluding vocabulary richness, the clustering pattern was very similar between the two book groups. For both the percentage of conjunctions was the final variable to be clustered. This is consistent with a paratactic text structure with clauses linked by conjunctions: the textual variables within clauses being more closely inter-related to each other than to the conjunctions between clauses. For the Deuteronomistic group, the percentage of definite articles and the percentage of proper nouns are closely associated as are the percentage of bound pronouns and the percentage of preterite verbs. These four variables then form a cluster which links with the remaining 15 textual variables. This indicates that in the Deuteronomistic group proper nouns, determinate nouns and preterite forms are associated with personal indicators, and may be considered markers of a narrative style. For the Minor Prophet group, the percentages of definite articles,

proper nouns and bound pronouns are also closely associated, the percentage of preterite and perfect followed by preterite verbs also are closely associated, linking to the other 15 variables before joining the definite articles, proper nouns and bound pronouns cluster. This suggests that narrative passages may be less closely linked to personal, determinate content than in the Deuteronomistic group. The implication for Qumran text classification is that clustering of variables within defined text groups may aid understanding of the text type and its description.

Summary of cluster analysis and implications for the statistical analysis of Qumran texts

Clusters analysis of the Hebrew bible books was useful for classifying shorter texts. It also helped elucidate text structure for different groups of books. It is thus a promising statistical method for Qumran texts. A limitation of hierarchical clustering is that it is relatively difficult to understand in terms of multi-dimensional space. Multi-dimensional scaling techniques may provide a more informative picture of relationships between variables. A brief exploration of this statistical technique may be informative about whether this is the case for the books of the Hebrew bible and thus should be attempted for the Qumran corpus.

Multi-dimensional scaling

Here a further statistical technique that can be useful in understanding the relationship between textual variables is multi-dimensional scaling. Multi-dimensional scaling, hierarchical cluster analysis and factor analysis are all related for ratio and interval data. Factor analysis provides measures of association and these can be displayed on loading plots for the different factors. Cluster analysis measures the distances between variables and produces an agglomeration schedule of clustering according to nearest distance between the remaining clusters. Multi-dimensional scaling, like cluster analysis, uses distances between variables as opposed to associations, though essentially these are two sides of the one coin, as it were. Multidimensional scaling can be undertaken with a preset number of dimensions for the solution. For the purposes of pictorial display, a two dimension solution is parsimonious, though it may not be the best fit for the data. Presetting a two dimension solution is akin to deciding on considering only two factors for a principal components analysis regardless of the components' eigenvalues or the scree plot.

Figure 11 shows the two dimensional plot of the relationship between the 22 textual variables using an alternating least-squares algorithm to fit squared Euclidean distances between variables. The plot shows vocabulary richness on its own on Dimension 1 with the other 21 variables lined up along Dimension 2. Using the same algorithm and excluding vocabulary richness produces the plot shown in Figure 12. The percentage of pronouns is most distant from the other variables, with the percentages of proper nouns

and bound pronouns next, about equidistant from the main cluster of variables, but in different directions on Dimension 2. For these Hebrew bible textual variables, two dimensional plots do not add a great deal to the information available from factor analysis and cluster analysis. The implication for the Qumran text analysis and classification is that the technique is probably best reserved for situations where the results of factor analysis or cluster analysis are unclear or conflicting.

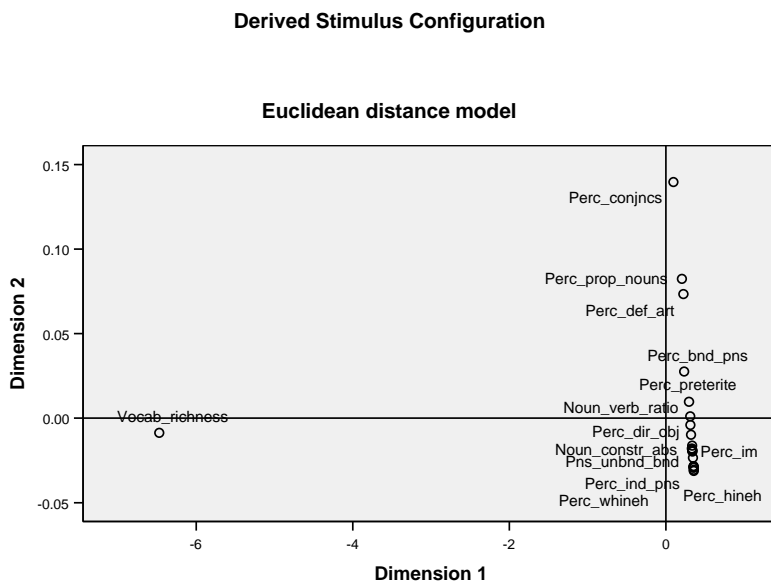


Figure 11 Two-dimensional solution by multi-dimensional scaling for the 22 textual variables (see Table 5).

Derived Stimulus Configuration

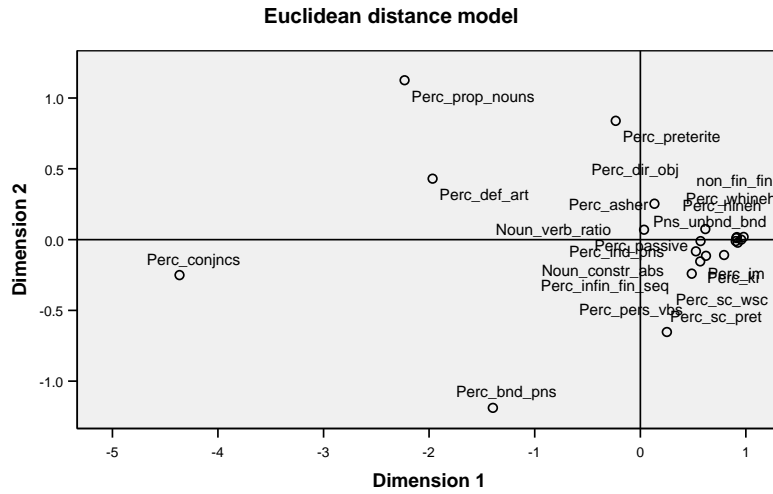


Figure 12 Two-dimensional solution by multi-dimensional scaling for the 21 textual variables excluding vocabulary richness