Key to the Kingdom: Variation as a Key to Understanding the Arabic Gospel Manuscripts

Phillip W. Stokes
University of Tennessee, Knoxville

(pstokes2@utk.edu)

Abstract

This paper takes a fresh approach to the study of the Arabic Gospel manuscripts. Although considerable success has been achieved at mapping out macro-families, there are still large lacunae in our knowledge of the Arabic Gospels as well as of the linguistic and scribal cultures that produced them. Arabic Gospel manuscripts notoriously vary at every level, and much of the variation is idiosyncratic. In previous work, this variation has by and large been considered background noise to be filtered out. In this paper, I study variation in the lexical, grammatical, and orthographic domains in the Gospel of Matthew as attested in twenty-two manuscripts belonging to multiple manuscript families. I use a Principal Component Analysis (PCA) to detect possible patterns in the variation. At each level, the variation is patterned in ways that contribute to our understanding of the manuscripts and their production. Most significantly, I argue that grammatical variation is not random, as previously assumed, and that several distinct grammatical traditions are detectable. I thus show that far from being an obstacle to the study of the Arabic Gospels, variation is in fact key to fully understanding them.

1. Introduction

Scholarly study of the Arabic Gospel manuscript tradition is by now well more than a century old.[1] The bulk of this scholarship has focused on two aspects of the manuscript tradition. First, much work has been done on translation efforts in the early Islamic period, with the goal of determining when the earliest translations were produced, and in what contexts. Second, scholars have devoted significant efforts to determining the *Vorlagen* from which the Arabic Gospels were translated, identifying the various translation styles and techniques used, and establishing families and subgroupings of manuscripts based on

1. For example, see I. Guidi, *Le Traduzioni degli Evangelii in arabo e in etiopico, Atti della reale accademia dei Lincei*, ser. 4, vol. 4 (Rome: Tipografia della reale accademia dei Lincei, 1888).

verbal similarities. Considerable progress has been made,[2] especially in the latter endeavor, such that Hikmat Kashouh in a recent monograph proposes a comprehensive classification of a few hundred manuscripts based on lexical similarities.[3]

Although the state of the field has advanced dramatically, there are still crucial lacunae in our understanding of both the relationships between manuscripts and families and especially the scribal practices attested across the corpus. The most immediate reason for these remaining lacunae is the fact that the corpus is characterized by a considerable degree of variation in all domains—lexical, grammatical, and orthographic. These overlapping arenas of variation have led scholars to describe the corpus as rather like a "jungle,"[4] or indeed a "massive forest, made up of many trees with intertwined branches."[5]

Numerous studies address the nature of the grammar and orthography attested in particular manuscripts.[6] However, these works have typically approached the manuscripts with the assumption that Classical Arabic was, originally at least, the target register, and that differences between Christian Arabic and Classical Arabic were due to mistakes. Consequently, instances of non-Classical features, both orthographic and grammatical, are cited from each manuscript, but with no systematic description of the distribution of most features in any one manuscript.[7] As a consequence, no systematic comparison of the patterns of orthographic or grammatical variation across manuscripts in the Arabic Gospels has yet been undertaken. The seminal work, which offers a synthesis of features from more than one manuscript, is Joshua Blau's three-volume grammar. In it, Blau offers general categories of what he considers hyper- and hypo-corrections,[8] with a handful of citations from numerous manuscripts as examples of these trends. However, Blau, too, refrains from

2. R. Vollandt, "The Status Quaestionis of Research on the Arabic Bible," in S*tudies in Semitic Linguistics and Manuscripts: A Liber Discipulorum in Honour of Professor Geoffrey Khan*, ed. N. Vidro, R. Vollandt, E.-M. Wagner, and J. Olszowy-Schlanger, 442–67 (Uppsala: Uppsala University Press, 2018).

3. H. Kashouh, *The Arabic Versions of the Gospels: The Manuscripts and Their Families* (Berlin: de Gruyter, 2012).

4. S. K. Samir, "La version arabe des évangiles d'al-Asʿad Ibn al-ʿAssāl," *Parole de l'Orient* 19 (1994): 441–551, at 444.

5. Kashouh, *Arabic Versions*, 2.

6. Among others, see J. Blau, "Über einige christlich-arabische Manuskripte aus dem 9. und 10. Jahrhundert," *Le Muséon* 75 (1962): 101–8; S. Arbache, "Une version arabe des évangiles: Langue, texte et lexique" (PhD diss., Université Michel de Montaigne Bordeaux III, 1994), and subsequently idem, *L'Évangile arabe selon saïnt Luc: Texte du VIII^e siècle, copié en 897, édition et traduction = al-Inǧīl al-ʿArabī bišārat al-qiddīs Lūqā* (Brussels: Éditions Safran, 2012); J. P. Monferrer-Sala, "Dos antiguas versiones neotestamentarias árabes surpalestinenses: Sin. ar. 72, Vat. ar. 13 y sus posibles Vorlagen respectivas greco-aljandrina y siriaca de la Pesitta," *Ciudad de Dios* 213, no. 2 (2000): 363–87.

7. This is the case even with otherwise quite meticulous works, such as Arbache's three-volume dissertation study of MS Sinai Arabic 72. For example, while Arbache provides statistics and data for his discussion of the verbal categories attested in the manuscript, his discussion of, e.g., nominal case is quite sparse, being limited to references to the "Middle Arabic" nature of the text and a few citations of non-Classical *tanwīn alif* use. He does not examine the syntactic contexts in which these citations occur or how frequently—if at all—Classical accusative marking occurs. See Arbache, "Une version arabe des évangiles," 123–24.

8. J. Blau, *A Grammar of Christian Arabic: Based Mainly on South-Palestinian Texts from the First Millennium*, 3 vols. (Louvain: Secrétariat du Corpus SCO, 1966–67).

undertaking either analysis of feature distribution within each manuscript or systematic comparison across manuscripts. Ultimately, by treating these features primarily as mistakes caused by the inability of Christian scribes to master aspects of Classical Arabic grammar not active in their dialects, previous studies have conceptualized non-Classical feature distribution (at least implicitly) as basically random within and across manuscripts.

Text-critical studies of the Arabic Gospel manuscripts have typically focused almost exclusively on lexical similarities to establish relationships between manuscripts.[9] Likely because of the nearly ubiquitous treatment of these texts in grammatical studies, neither grammatical nor orthographic patterns have received direct attention, and indeed in many cases they have been considered noise to be filtered out. For example, Kashouh, in discussing the role of linguistic variants in establishing textual relationships between manuscripts of a single family (his Family K), dismisses them as "very common grammatical mistakes" (that is, non-Classical Arabic variants) which "need not be taken into consideration."[10] Likewise, in the same section he argues that orthographic variation should be filtered out because "there seems to have been no standardized way of spelling some of the words," and such variants should thus "not be included among what we call *valuable* variants."[11] To be sure, no single standard existed by which all Gospel manuscripts were composed. Yet this fact has resulted in a rather extreme lack of attention to systematic documentation and comparison of variation in these domains within and across manuscripts. But the presence of variation does not entail the absence of meaningful patterns, and claims about grammatical and orthographic mistakes lacking value for text-critical purposes require testing. An approach that has not featured in text-critical studies of the Arabic Gospel manuscript tradition is the study of repeated lexical variation—how the same phrase is translated across multiple passages of the same manuscript and where that variation is replicated across manuscripts.

The goal of this paper is to explore nontraditional approaches and tools for conceptualizing and studying the variation attested in every domain of the Arabic Gospel manuscripts, as well as to illustrate how these might nuance, complement, and in some cases change our understanding of these manuscripts and the people who produced them. Specifically, I document and test whether and how lexical, grammatical, and orthographic variation form meaningful patterns across Arabic Gospel manuscripts by comparing instances of idiosyncratic variation in each domain. The idiosyncratic lexical variation measured here consists of different phrases used to translate the phrases "Kingdom of God" and "Kingdom of Heaven" in the Gospel of Matthew, both within and across manuscripts. For grammatical variation, I selected and compared fifty-eight grammatical variants from each manuscript, again taken from the Gospel of Matthew. Orthographic variation in this

9. In addition to Kashouh, *Arabic Versions*, see, among many others, G. Graf, *Geschichte der christlichen arabischen Literatur*, vol. 1, *Die Übersetzungen*, Studi e testi 118 (Rome: Biblioteca Apostolica Vaticana, 1944); C. Peters, "Proben eines bedeutsamen arabischen Evangelien-Textes," *Oriens Christianus* 11 (1936): 188–211; J. Valentin, "Les évangéliaires arabes de la bibliothèque du Monastère Ste-Catherine (Mont Sinai): Essai de classification d'après l'étude d'un chapitre (Matth. 28); Traducteurs, réviseurs, types textuels," *Le Museon* 116 (2003): 415–77; S. Griffith, *The Bible in Arabic: The Scriptures of the "People of the Book" in the Language of Islam* (Princeton, NJ: Princeton University Press, 2013).

10. Kashouh, *Arabic Versions*, 215.

11. Ibid.; the emphasis is Kashouh's.

study involves differences in the spelling of the word *malakūt*, "kingdom," as well as the orthographic means by which *tanwīn* is indicated. These patterns are mapped and compared across twenty-two manuscripts without coding for the family to which they are assigned in Kashouh's study. To analyze and visualize the results, I use a Principal Component Analysis (PCA), discussed further in Section 2.

I show that in each domain, the variants pattern in meaningful ways that correlate in many cases with manuscript family reconstructions established on the basis of lexical similarity. Idiosyncratic lexical patterns are replicated across manuscripts, sometimes with great faithfulness, confirming close textual relationships between several families. The results also offer insight into the nature of the grammatical and scribal practices attested across the corpus. Perhaps most significantly, my analysis of grammatical variation reveals the existence of some distinct grammatical traditions, which often correlate with text families. Finally, orthographic variation in one sense validates some of Kashouh's skepticism, insofar as a close study of the orthography of the noun *malakūt* in Family J^{b} does not result in a meaningful pattern within the family. However, there are nevertheless scribal patterns that connect certain families and that are suggestive of a connection between the circles that produced the manuscripts and distinct scribal subcultures and practices. I conclude that far from being an obstacle to understanding these manuscripts, the variation is in fact a defining feature and one that is significant both for our understanding of the linguistic and scribal background of Christian Arabic manuscripts and for the textual-critical study of the corpus. Documenting and studying such variation should consequently occupy a prime position in any investigation of the corpus.

2. Methodology, Manuscripts, and Manuscript Families

The scholarship on the comparison of Arabic Gospel manuscripts and the establishment of *Vorlagen* as well as genetic relationships between families based on those *Vorlagen* is long, and a full review is beyond the scope of the present paper.[12] I will take Kashouh's classification as a starting point and reference other proposals as relevant to the data presented below.[13] The manuscripts included in this study are listed in Table 1.[14]

12. In addition to Kashouh, *Arabic Versions*, some of the most significant are Guidi, *Le Traduzioni degli Evangelii*; G. Graf, *Catalogue de manuscrits arabes chrétiens conservés au Caire* (Rome: Biblioteca Apostolica Vaticana, 1934); idem, *Geschichte der christlichen arabischen Literatur*; A. Vööbus, *Early Versions of the New Testament: Manuscript Studies* (Stockholm: Estonian Theological Society, 1954); Blau, "Über einige christlich-arabische Manuskripte; idem, "Sind uns Reste arabischer Bibelübersetzungen aus vorislamischer Zeit erhalten geblieben?," *Le Muséon* 86 (1973): 67–72; A. S. Atiya, "Codex Arabicus (Sinai Arabic Ms. No. 514)," in *Homage to a Bookman: Essays on Manuscripts, Books and Printing Written for Hans P. Kraus on His 60th Birthday Oct. 12, 1967*, ed. H. Lehmann-Haupt, 75–85 (Berlin: Mann, 1967); S. H. Griffith, "The Gospel in Arabic: An Inquiry into Its Appearance in the First Abbasid Century," *Oriens Christianus* 67 (1983): 126–67; S. K. Samir, "La tradition arabe chrétienne: État de la question, problèms et besoins," in *Actes du premier congrès international d'études arabes chrétiennes*, ed. K. Samir, 21–120, Orientalia christiana analecta 218 (Rome: Pont. Institutum Studorum Orientalium, 1982); Arbache, "Une version arabe des évangiles"; Valentin, "Les évangéliaires arabes."

13. For a list of all the manuscripts included in Kashouh's study along with their proposed families, see Kashouh, *Arabic Versions*, 45–77.

14. All of the manuscripts with the exception of Sinai Arabic 106 and Sinai Arabic 84 were accessed via the Sinai Manuscripts Digital Library website, hosted by the UCLA Library and created in partnership with St.

Table 1. Manuscripts included in the study, with their dates and their families according to Kashouh.

Manuscript	Date	Family
Sinai Arabic 74	9th CE	A
Sinai Arabic 72	897 CE	A
Vatican Borg. Arabic 95	9th CE	A
Sinai Arabic 70	9th CE	C
Sinai Arabic 75	9th/10th CE	D
Vatican Borg. Arabic 13	8th/9th CE	H
Sinai Arabic 115	13th CE	J^a
Sinai Arabic 106	1056 CE	J^b
Sinai Arabic 69	1065 CE	J^b
Vatican Borg. Arabic 71	11th CE	J^b
Sinai Arabic 84	1262 CE	J^b
Sinai Arabic 82	1262 CE	J^b
Sinai Arabic 89	1285 CE	J^b
Sinai Arabic 90	1281 CE	J^b
Sinai Arabic 91	1289 CE	J^b
Sinai Arabic 80	1479 CE	J^b
Sinai Arabic 76	13th CE	J^c
Vatican Coptic 9	1204/5 CE	K
Sinai Arabic 112	1259 CE	K
Sinai Arabic 147	13th CE	K
Sinai Arabic 628	1336 CE	K
Sinai Arabic 68	14th CE	K

Catherine's Monastery of the Sinai, Egypt. For these manuscripts, the viewer provides folio, recto, and verso information. Sinai Arabic 106 and 84 are not available on this platform, and they were accessed through the website of the Library of Congress. For these manuscripts, folio, recto, and verso information is not given, so references are to the numbers of the images on which the cited examples are found.

Throughout the following sections, "Sinai Arabic" is abbreviated SAr.; Vatican Borg. Arabic is abbreviated Vat. Borg. Ar.; and Vatican Coptic is abbreviated Vat. Copt. The manuscripts are color-coded here according to family, and these colors recur in the charts and the PCA visualization below. Note that these proposed manuscript family designations are not coded in the analysis in any way.

Obviously, the present study does not pretend to be exhaustive; rather, the goal is to illustrate the value of such studies and to pave the way for the inclusion of more manuscripts and variants. The manuscripts included in this study were selected on the basis of two main criteria, namely online accessibility and manuscript family representation. The families represented in my sample include the three that are earliest historically (including the early families A and H) and the two that are most commonly attested in the medieval period, namely Family J (especially Family J^b, the "Melkite Version") and Family K (the "Alexandrian Vulgate"). Potential manuscripts were also evaluated for their completeness, with every effort made to compare variants from the same verses across manuscripts. For these reasons, three manuscripts (Sinai Arabic 54 and 71 from Family A, and Sinai Arabic 146 from Family J^a) were omitted, despite being accessible, because of their fragmentary nature.

In order to measure the variation in each domain that is the objective of this study, I created columns in a spreadsheet for each selected manuscript, and rows for the biblical citations in which each included variant occurred. The combination of dozens of manuscripts and dozens of features makes it virtually impossible to detect patterns by simply looking at the data in spreadsheet form; instead, a tool is required to analyze and visualize any patterns that might emerge. In order to detect such patterns as well as visualize them in helpful ways, I used a Principal Component Analysis (PCA). A PCA is particularly useful with the kind of high-dimensional data that result from comparison of many variants from nearly two dozen manuscripts. For an excellent lay introduction to PCA as well as an example of its use in an analysis of the Quran, see Behnam Sadeghi's excellent study.[15] For my purposes, the important benefit of this method is that with a data set as large as the one I use in this study, an attempt to find patterns of correlation in the data can reveal myriad ways in which aspects of the data pattern together.[16] A PCA is a tool (if

15. B. Sadeghi, "The Chronology of the Qurʾān: A Stylometric Research Program," *Arabica* 58 (2011): 210–99, at 247–52.

16. For the PCAs, I ran each in Python, with the specific PCA functions taken from scikit-learn. Empty cells in any manuscript for a particular feature resulted in the row in which the empty cell occurs being omitted from the PCA analysis. These data were nonetheless included in the charts in this study in order to provide as full a picture of the data as possible. Importantly, in many instances the tables that follow contain more than two features. For the sake of presenting the data in this article I have numbered these features 1, 2, 3, etc., but this is *not* how they were coded. If they had been, whatever feature happened to be coded 2 would have been treated as lying halfway between the one coded 1 and the one coded 3. This would, of course, be extremely problematic, since changing the order would result in a completely different distribution. Instead, all the data have been binarized. For example, let us say that the following three features were attested across all manuscripts:

ʾilay-hi "to him"
ʾilay-hī
ʾilay-hu

admittedly a rather crude one) for capturing the maximum amount of variance in the data and visualizing it plotted across two dimensions.[17] In the visualizations, each manuscript is color-coded according to its manuscript family (see Table 1 above); however, this family identification was not included in the data. Any clustering is therefore purely a result of the data distribution. The x axis corresponds to the first principal component, which is the one that covers the largest amount of variance in the data. The y axis corresponds to the second principal component, which covers the second-largest amount of variance in the data. Thus proximity along the x axis indicates similarity captured by the first principle component, whereas proximity along the y axis indicates similarity captured by the second principle component. Manuscripts that are close on both axes thus attest stronger relationships than those close on just one axis. We turn now to a discussion of the specific variants and analysis of the data.

3. "Kingdom of God" and "Kingdom of Heaven": Idiosyncratic Lexical Variation

As noted earlier (in footnote 9), most studies that have explored the relationships between Arabic Gospel manuscripts have focused on lexical similarities. In such studies, a passage or passages are collated and then compared with the same passages in other manuscripts. Identical or very similar lexical and stylistic usages, which can indicate a shared textual history, are used to group manuscripts. One weakness of this strategy, as Kashouh notes in a discussion of subgrouping manuscripts within Family K, is that a scribe's lexical choice in any particular instance might be independent of the same choice made by another scribe—that is, the similarity could be coincidental. Thus, in establishing secure relationships between manuscripts on the basis of shared lexical choices in particular places in a passage, the more cases, the better: "When the number of agreements rises, the level of probability grows."[18] Therefore, many examples from numerous passages are needed to establish a relatively high degree of confidence that the shared lexical choices indeed signify shared history.

Instead of simply labeling them 1, 2, and 3 in that order, which would result in *ʾilay-hī* being treated as a halfway form between *ʾilay-hi* and *ʾilay-hu*, each feature is split into three variables. So if a manuscript has *ʾilay-hi*, the data for that feature are interpreted as follows:

ʾilay-hi = 1, *ʾilay-hī* = 0, *ʾilay-hu* = 0

Whereas another manuscript, which has *ʾilay-hī*, is treated thus:

ʾilay-hi = 0, *ʾilay-hī* = 1, *ʾilay-hu* = 0

This way, the data—and the distribution of the PCA—are not affected by the numeric assignments given to the features. For the specific code used and a step-by-step reproduction of the process, see my GitHub page. The code, written by Hythem Sidky, can be found here. I also thank Marijn van Putten for his very valuable assistance in adapting the code to suit the analyses conducted in this study. Any errors in the PCAs and analyses of the data are mine alone.

17. For a similar use of PCA and its visualization in studying multiple features, see M. van Putten and H. Sidky, "Pronominal Variation in Arabic among the Grammarians, Quranic Reading Traditions, and Manuscripts," in "Formal Models in the History of Arabic Grammatical and Linguistic Tradition," ed. R. Villano, special issue, *Language and History* 65, no. 1 (forthcoming).

18. Kashouh, *Arabic Versions*, 215.

A related but distinct method, which does not suffer from the same limitation and can thus play a significant role in text-critical analysis of the manuscripts, is to investigate a particular word or phrase across a manuscript, compare how it is translated (or spelled; see Section 5 below) within a manuscript, and determine whether that same pattern is attested in other manuscripts. Good examples from the Gospels of phrases that occur frequently and whose distribution is frequently idiosyncratic are the phrases "Kingdom of God" and "Kingdom of heaven." Five variations in Arabic of the latter phrase are attested in the Gospel of Matthew in at least one (and usually more) of the twenty-two manuscripts included in my study (see Table 1); in total, the phrase occurs thirty-seven times in the Gospel of Matthew. In these variants, three words are used for "kingdom" (*mulk*, *malakūt*, and *mamlakah*), and both singular and plural forms of the word "heaven" (*samāʾ* and *samāwāt*) appear. The five variants are the following:

1. *mulk al-samāʾ*
2. *mulk al-samāwāt*
3. *malakūt al-samāʾ*
4. *malakūt al-samāwāt*
5. *mamlakat al-samāwāt* (*mamlakat al-samāʾ* not attested)

Some manuscripts, such as Vat. Borg. Ar. 13, SAr. 70, and SAr. 75, use a single phrase each for "Kingdom of God" and "Kingdom of heaven." However, in most of the manuscripts included in this study, more than one phrase is used, especially for "Kingdom of heaven." In some cases, the manuscripts use one and the same word for "kingdom" but render the word "heaven" variously in the singular and in the plural. This is the case in SAr. 112, 147, 68, and 628 and in Vat. Copt. 9. In the remaining manuscripts, both "kingdom" and "heaven" vary. This point is significant and bears repeating: the ways in which the same phrase is translated vary within a single manuscript. This is the kind of lexical variation studied here, not simply the use of one or another of these phrases to translate the phrase. If this lexical variation is idiosyncratic, dependent on the tastes of individual scribes, we would expect little meaningful clustering to occur. That is, the variation in one manuscript should not mirror that of another. On the other hand, meaningful clustering, especially in cases in which manuscripts contain multiple variants, is highly unlikely to be due to chance and is suggestive of a strong textual relationship. Table 2 shows the distribution of the five possible variants across the twenty-two manuscripts. In the table, 1 = *mulk al-samāʾ*, 2 = *mulk al-samāwāt*, 3 = *malakūt al-samāʾ*, 4 = *malakūt al-samāwāt*, and 5 = *mamlakat al-samāwāt*. Cells left empty indicate that another phrase is used (see Table 3 below). A dash (–) indicates that the manuscript does not contain any phrase in that location.

Table 2. "Kingdom of heaven" variants in the Gospel of Matthew across manuscripts.

Text	MS																					
	V13	74	72	V95	70	75	115	106	69	V71	84	82	89	90	91	80	76	112	147	68	628	V9
3:2	3	1	2	2	1	3	2	1	3	3	3	3	3	3	–	3	5	1	4	4	4	4
4:17	3	1	2	2	1	3	4	3	3	3	3	3	3	3	–	3	2	3	3	4	4	4
5:3	3	1	1	3	1	3	2	3	3	3	3	3	3	3	3	3	2	3	4	4	4	4
5:10	3	1	1	3	1	3	2	3	3	–	3	3	3	3	3	3	2	4	4	4	4	4
5:12	–	–	–	–	–	–	–	3	3	–	–	3	3	–	–	1	–	–	–	–	–	–
5:19a	3	2	2	4	1	3	4	3	3	–	3	3	3	3	3	3	2	4	4	4	4	4
5:19b	3	2	2	2	1	3	2	3	3	–	3	3	3	3	3	3	2	4	4	4	–	4
5:20	3	2	2	2	1	3	3	3	3	–	3	3	3	3	3	3	2	3	4	4	4	4
7:21	3	2	2	2	1	3	3	1	1	1	1	1	1	1	1	2	4	4	3	4	4	4
8:11	3	2	2	2	1	3	3	1	1	1	1	1	1	1	1	1	3	3	4	4	4	4
10:7	4	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	4	3	4	4	4
11:11	3	1	1	1	1	3	3	1	3	1	1	1	1	1	1	1	4	5	4	3	3	4
11:12	3	2	2	1	–	–	3	1	1	1	1	1	1	1	1	1	4	–	3	3	4	4
12:28	3								3										–	–		
13:11	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	4	2	4	4	4		4
13:24	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4
13:31	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4
13:33	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4
13:44	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
13:45	3	2	2	1	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
13:47	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
13:52	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	2		3	3	4	4
16:19	3	2	2	1	1	3	4	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
18:1	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	2	3	4	4	4	4

18:3	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	2		4	4	4	4
18:4	3	–	–	2	1	3	3	1	1	3	3	1	1	1	3	1	2	4	4	4	4	4
18:23	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
19:12	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4		4	4	3	4
19:14	3	2	2	2	1	3	3	3	3	3	3	3	3	4	3	3	4	3	–	–	4	4
19:23	3	2	2	2	1	3	3	1	1	1	1	1	1	1	1	3	4	3	3	3	4	4
19:24			4						3		3	3	3						3	3		
20:1	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4
21:31	3							3	3	3	3	3	3	3		3						
21:43								3		3	3	3		3								
22:2	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	4	4
23:13	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	3	3	4	4
25:1	3	2	2	2	1	3	3	3	3	3	3	3	3	3	3	3	4	3	4	4	4	4

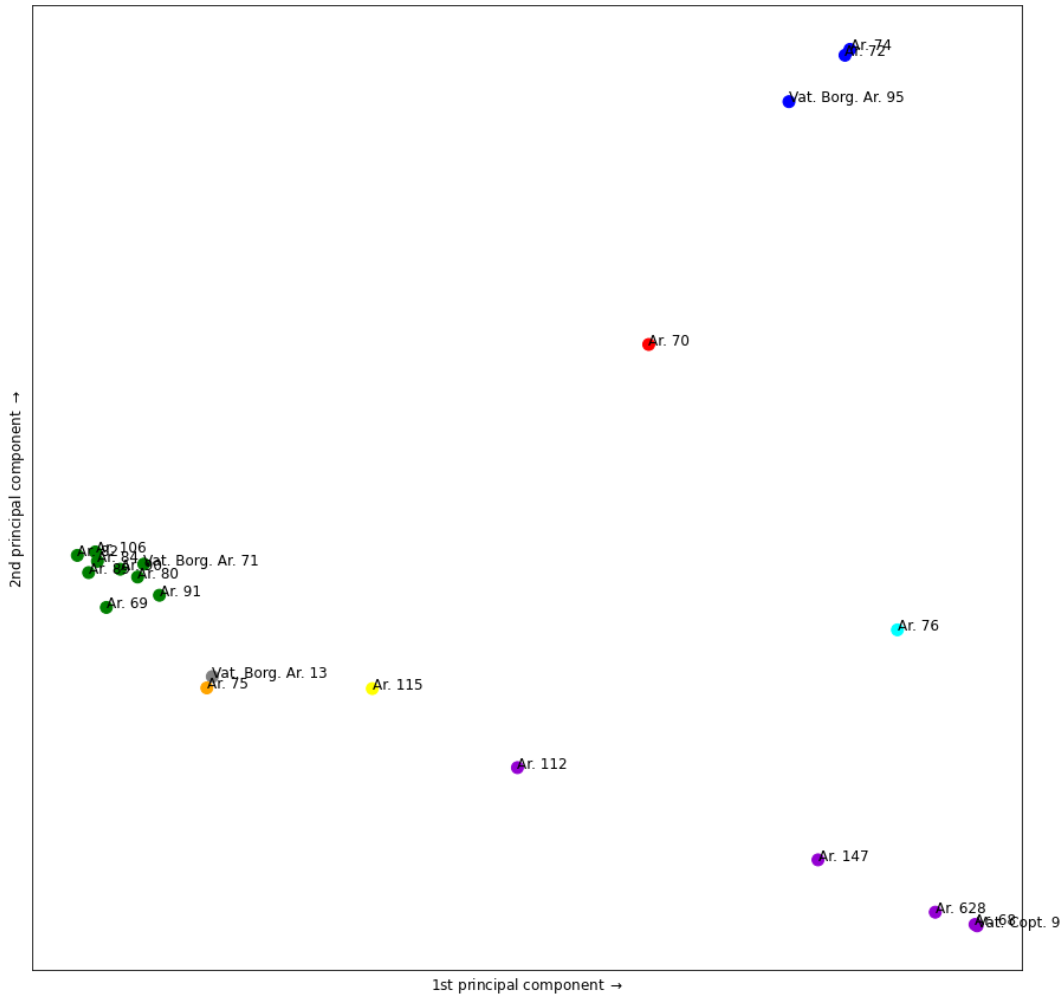
In five verses, Matthew 6:33, 12:28, 19:24, 21:31, and 21:43, the *Vorlagen* read "Kingdom of God." In the latter four verses, at least one of the manuscripts included here replaces this phrase with some version of "Kingdom of heaven." In Matthew 6:33, however, all manuscripts have a variant of "Kingdom of God." In Table 3, the variants of "Kingdom of God" across the manuscripts are tabulated: 1= *mulk allāh*, 2 = *malakūt allāh*. As in Table 2, empty cells indicate another phrase is used, and a dash marks the absence of any phrase in that location.

Table 3. "Kingdom of God" variants in the Gospel of Matthew across manuscripts.

Text	MS																					
	V13	74	72	V95	70	75	115	106	69	V71	84	82	89	90	91	80	76	112	147	68	628	V9
6:33	2	1	1	1	1	2	2	1	–	1	1	1	1	1	1	1	2	2	–	2	2	2
12:28		1	1	1	1	2	2	2		2	2	2	2	2	2	2	2	2	–	2	2	2
19:24	2	1		1	1	2	2	2		2				2	2	2	1	2			2	2
21:31		1	1	1	1	2	2								2		2	2	2	2	2	2
21:43	2	1	1	1	1	2	2		2				2		2	2	2	2	2	2	2	2

For this PCA, I concatenated both of these tables instead of running only one or the other or both separately. The results of the PCA of the data are displayed in Figure 1.

Fig. 1. Lexical variation in "Kingdom of heaven"/"Kingdom of God" across manuscripts.[19]



It is immediately clear that there are obvious clusters, which largely correspond to previously proposed manuscript families. As already noted, the variation characteristic of many of these families, especially Family J and Family A, is based on the replication across manuscripts of idiosyncratic patterns within manuscripts. There is no doubt that the clustering in these cases reflects the respective scribes' reliance on a shared exemplar, which, at least in the case of these phrases, the scribes followed quite closely. Although scribes generally felt fairly free to make lexical changes or updates to the exemplar, these phrases, which are perhaps a bit less salient than other, more stylistically or theologically

19. The first principal component accounts for 38.1% of the variation and the second principal component for 22.3%, so a total of 60.4% of the variation in the data is accounted for by the two principal components.

loaded terms, fly under the radar, as it were. They can thus serve as a highly meaningful indicator or marker of family identity when future manuscripts are sorted into these families.

First, it is important to note that, with a few exceptions (discussed below), relying solely on the variation in this phrase within and across manuscripts replicated the subgroupings from Kashouh's work. Indeed, several of the proposed manuscript families and subfamilies of which multiple manuscripts were included in my study emerged as strongly related, especially Family A (colored blue) as well as the large subfamily J^b (colored green). Family K (colored purple), a large and eclectic family, shows some clustering, especially between SAr. 68, Sar. 628, and Vat. Copt. 9; SAr. 147 is an outlier of this cluster and SAr. 112 a much further outlier. SAr. 115, the representative of Kashouh's family J^a (colored yellow), is much closer to J^b than it is to J^c, which is represented by SAr. 76 (colored aqua).

Second, families A, J^b, and K exhibit the most lexical variation in the translation of these phrases, which makes their results especially significant. The manuscripts of families A and J^b are especially closely clustered. Such strong clustering, defined by such idiosyncratic lexical variation, cannot be the result of chance. It therefore must be the result of textual relationships and dependence on a shared exemplar, which was often copied, at least in these cases, with great faithfulness. On the other hand, the close clustering of SAr. 75 (Family C, colored orange) and Vat. Borg. Ar. 13 (Family H, colored gray) is very possibly due to chance, insofar as they both used the same phrases throughout (*malakūt al-samāʾ* and *malakūt allāh*)—an overlap that is as likely to be due to independent preference for these phrases as it is to reflect a shared textual exemplar. Family K, as noted above, is a rather eclectic family, and thus we see close relationships between some of the family's manuscripts (e.g., Vat. Copt. 9 and SAr. 68 and 628) but not others (e.g., SAr. 112 and SAr. 147).

This approach has limits, of course, chief among them that it cannot always distinguish different families. The results above group families C and H together, but although a relationship between the two families is possible, it is far from certain, since it is also possible that the similarity represents an independent preference for the singular phrase used in both. In such cases, it is preferable to use other methods to supplement this kind of analysis.

I have argued that the study of replicated lexical idiosyncrasies can serve as a valuable tool for text-critical studies. It is important to note that the above comparison involved manuscripts that, according to most scholars, belong to separate language families. It is worth exploring briefly whether the same idiosyncrasies, insofar as they are not always replicated exactly, might be helpful for further grouping manuscripts from the same (sub) family. I tested this possibility on Family J, then further on Subfamily J^b, and finally on Family K. I chose these families because they were the families represented in my study with the largest number of manuscripts. Figure 2 illustrates the results of a PCA of Family J.[20] It recreates the preexisting subgroupings of the family quite neatly: the manuscripts of Family J^b cluster together, whereas those of J^a and J^c stand apart from them and from each other.

20. The first principal component captured 56% and the second principal component 18% of the variation.

Fig. 2. PCA of "Kingdom of God"/"Kingdom of heaven" variation within Family J

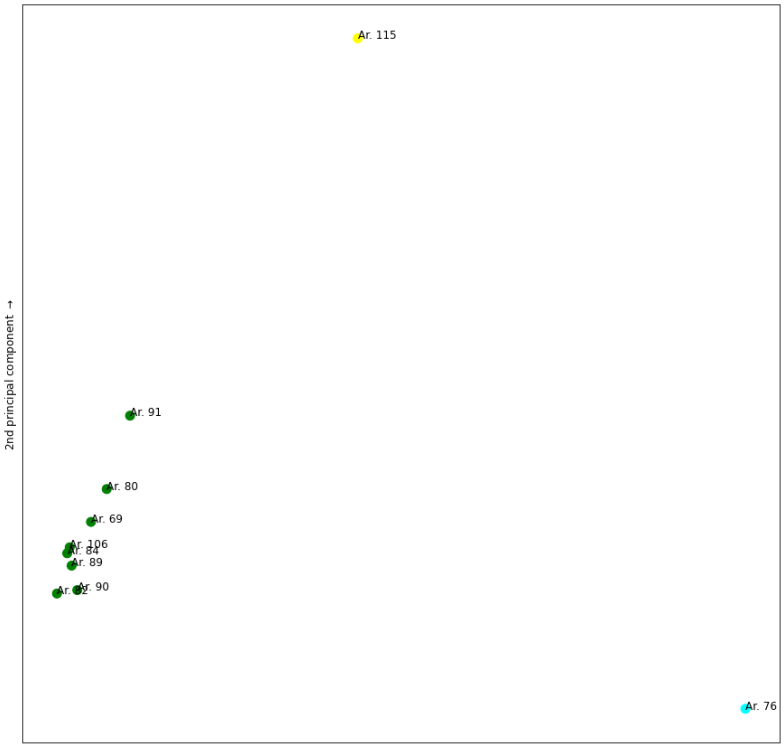
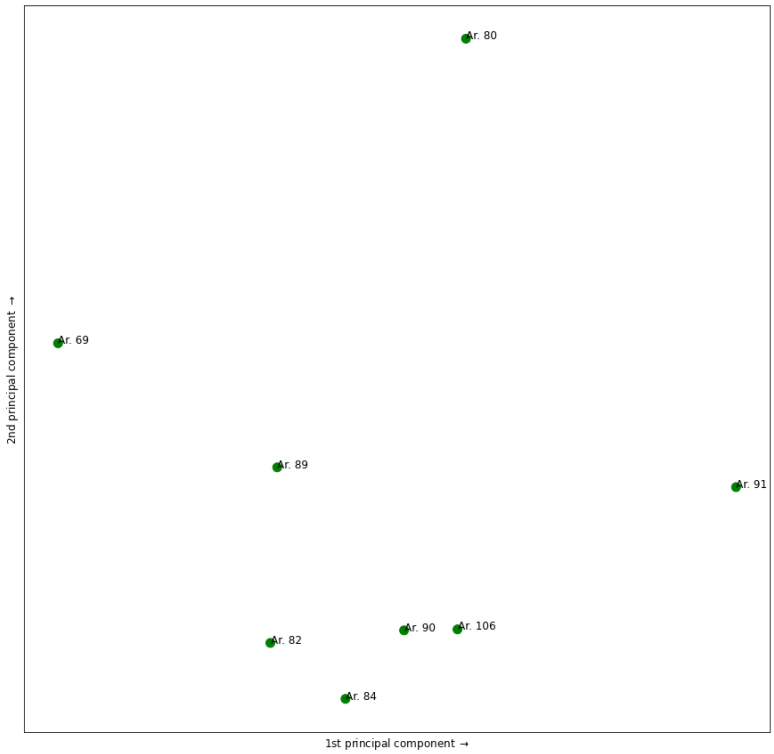
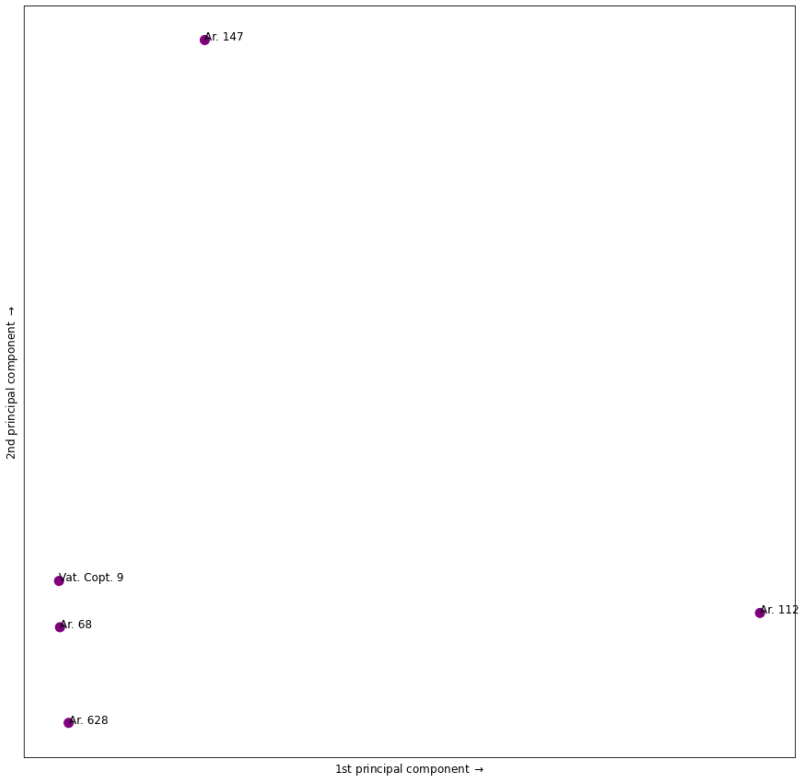


Fig. 3. PCA of "Kingdom of God"/"Kingdom of heaven" variation within Family J[b].



An analysis of just Family J[b] produced the results shown in Figure 3.[21] The manuscripts SAr. 82, SAr. 84, SAr. 90, and SAr. 106 cluster rather closely, with SAr. 89 an outlier of this group. SAr. 68, SAr. 91, and especially SAr. 80 are greater outliers. As will be seen in Sections 4 and 5 below, SAr. 69 and SAr. 80 diverge in other ways, too, from the norms of the group.

Fig. 4. PCA of "Kingdom of God"/"Kingdom of heaven" variation within Family K.



Finally, a study of Family K provided the results seen in Figure 4.[22] Three of the five manuscripts (SAr. 68, SAr. 628, and Vat. Copt. 9) cluster together; two others (SAr. 147 and SAr. 112) are outliers. As with the outliers in Family J[b], SAr. 147 and SAr. 112 are outliers from the other three in other ways as well (as discussed in Sections 4 and 5). It should be noted here that Family K is the largest of Kashouh's families, so more manuscripts would ideally be included to get a more nuanced picture of the family. However, these results—and those from Family J and Subfamily J[b]—suggest that this approach is potentially useful for grouping manuscripts within families as well as across them.

21. The first principal component captured 29% and the second principal component 24% of the variation. In order to capture the maximal amount of variation, I have omitted from this analysis Vat. Borg. Ar. 71, which is missing the first six instances of these phrases. As mentioned above, any empty cell in any column triggered the elimination of the entire row of variants from inclusion in the analysis.

22. The first principal component captured 62% and the second principal component 27% of the variation.

The present approach is thus most helpful when a particular word or phrase is repeated multiple times within or across the Gospels and when several words or phrases are used to translate it. When attested, idiosyncratic lexical variation can pattern meaningfully, as demonstrated here. The fact that these thirty-eight variants replicate to a large degree the findings of a study as detailed as Kashouh's is indicative of the potential usefulness of this approach.

4. Grammatical Variation within and across Manuscripts and Families

For all practical purposes, the modern study of the grammar of Christian Arabic corpora began with—and is still defined by—Blau's three-volume work on the subject.[23] Blau categorized these manuscripts as Middle Arabic, which he approached through the lens of Classical Arabic, focusing on those elements that are non-Classical. He provided no quantitative or frequency data, instead pulling examples of non-Classical features from various manuscripts. These are typically explained as either dialectalisms or hypercorrections.[24] It is important to note the implications of such a treatment: features are considered important only insofar as they indicate imperfect attempts to write Classical Arabic. No manuscript in Blau's grammar is treated in toto, and there is no effort to trace the distribution of the non-Classical features.[25]

Blau's framework was based on the belief that pre-Islamic Arabic had by and large been homogeneous, essentially identical to Classical Arabic, which led Blau to assume that Classical Arabic, and specifically a version of Classical Arabic equivalent to textbook descriptions like that of Wright[26] or Fischer,[27] should form the default grammar against which any corpus is analyzed.[28] Despite their differences in other areas, most scholars of

23. Blau, *Grammar of Christian Arabic*.

24. Hypercorrections are grammatically incorrect forms that result from a writer's attempt to use a prestigious form, usually one that differs from a dialectalism. For a discussion of these forms, see J. Blau, *On Pseudo-Corrections in Some Semitic Languages* (Jerusalem: Israel Academy of Sciences and Humanities, 1970); B. Hary, "Hypercorrection," in *Encyclopedia of Arabic Language and Linguistics*, ed. K. Versteegh, vol. 2, 275–79 (Leiden: Brill, 2007).

25. A good example of such impressionistic description is provided by Blau's discussion of features that supposedly distinguish Christian Arabic from other Middle Arabic varieties, such as Judeo-Arabic. In numerous places he claims that a feature, such as purported traces of living Aramaic (p. 55), is "not rare." But we are not told how frequent such features are, whether they are equally frequent in texts from a certain period, or whether they are equally frequent in translated and original texts. Blau's implicit claim is that they are found more frequently in Christian Arabic than in Judeo-Arabic, but he offers no quantitative comparisons between representatives of the two corpora.

26. W. Wright, *Arabic Grammar* (1896–98; repr., Mineola, NY: Dover, 2005).

27. W. Fischer, *A Grammar of Classical Arabic*, trans. Jonathan Rodgers (New Haven, CT: Yale University Press, 2002).

28. Blau, *Grammar of Christian Arabic*, 19–20; idem, "The Beginnings of the Arabic Diglossia: A Study of the Origins of Neoarabic," *Afroasiatic Linguistics*, no. 4 (1977): 1–28; idem, *A Handbook of Early Middle Arabic* (Jerusalem: Hebrew University of Jerusalem, 2002), 14–22.

Christian Arabic have adopted this framework, explicitly or implicitly,[29] assuming that since the manuscripts were written in Middle Arabic, any non-Classical features were unintentional, caused by imperfect knowledge. Insofar as the imperfections reflected an individual scribe's knowledge (or lack thereof) of Classical Arabic, the Middle Arabic features were bound to be random.[30]

The foregoing approach relies on a historical model of Arabic that posits that pre-Islamic Arabic was largely equivalent to Classical Arabic, that the Quran was composed in Classical Arabic (or perhaps a poetic koiné combining various Classical features), and that this variety of Arabic was the prestige variety in the early Islamic period.[31] However, these assumptions have all become extremely problematic in light of advances in the study of the early history of Arabic. For example, it has become clear that pre-Islamic Arabic was quite diverse.[32] Further, Classical Arabic, as described by the grammarians, is itself linguistically heterogeneous.[33] And evidence from other corpora, especially the Damascus Psalm Fragment, demonstrates that non-Classical Arabic varieties were prestigious in the

29. See, for example, Samir, "La tradition arabe chrétienne," 58; Kashouh, *Arabic Versions*, 6–8; M. L. Hjälm, *Christian Arabic Versions of Daniel: A Comparative Study of Early MSS and Translation Techniques in MSS Sinai Ar. 1 and 2* (Leiden: Brill, 2016), 371, n. 206; J. Grand'Henry, "Le moyen arabe dans les manuscrits de la version arabe du discours 40 de Grégoire de Nazianza," in *Moyen arabe et variétés mixtes de l'arabe à travers l'histoire: Actes du premier colloque international (Louvain-la-Neuve, 10–14 mai 2004)*, ed. J. Lentin and J. Grand'Henry, 181–91 (Louvain-la-Neuve: Peeters, 2008); P. Bengtsson, *Two Arabic Versions of the Book of Ruth: Text Edition and Language Studies* (Lund: Lund University Press, 1995), 85–94.

30. A reviewer of this article suggested that scholars in general would not assume that grammatical variation would necessarily be random. However, I do not see how one can draw another conclusion from the works and discussion outlined above. Indeed, if grammatical variation were seen to be patterned, it would be important to determine *how* it was patterned, which is a primary goal of this paper. No one to my knowledge has done this. It thus seems difficult to conclude anything other than that scholars have treated grammatical variation as implicitly idiosyncratic and random. This is exemplified in, e.g., Kashouh's introduction, in which he states, regarding "linguistic limitations" to the study of the Arabic Gospel manuscripts, that "in the earliest manuscripts, generally speaking, there seems to be no effort in producing a *linguistic homogenous* [*sic*] text. Moreover**,** there was no *systematic* linguistic rules, neither syntactical nor orthographical, that the scribes were following or even wanted to follow"; Kashouh, *Arabic Versions*, 7 (emphasis his).

31. For example, Kashouh makes this explicit when he claims that linguistic corrections were introduced in order to bring the language "close to the classical [*sic*] Arabic"; Kashouh, *Arabic Versions*, 7.

32. For example, Arabic inscriptions written in the Safaitic and Hismaic scripts attest to numerous non-Classical Arabic features, some archaic and others innovative; see A. Al-Jallad, *An Outline of the Grammar of the Safaitic Inscriptions* (Leiden: Brill, 2015); idem, "The Earliest Stages of Arabic and Its Linguistic Classification," in *The Routledge Handbook of Arabic Linguistics*, ed. E. Benmamoun and R. Bassiouney, 315–31 (New York: Routledge, 2018). Van Putten and I have argued that the language underlying the Quranic Consonantal Text typifies the dialects of the Hijaz and differs from Classical Arabic in numerous ways; M. van Putten and P. W. Stokes, "Case in the Qurʾānic Consonantal Text," *Wiener Zeitschrift für die Kunde des Morgenlandes* 108 (2018): 143–79.

33. The grammatical variation assembled in Sībawayh's *Kitāb* and al-Farrāʾ's *Luġāt al-Qurʾān* attest to this diversity, as does the Quran and its reading traditions; see, e.g., M. van Putten, "Arabe 334a: A Vocalized Kufic Quran in a Non-canonical Hijazi Reading," *Journal of Islamic Manuscripts* 10 (2019): 327–75; idem, *Quranic Arabic: From Its Hijazi Origins to Its Classical Reading Traditions* (Leiden: Brill, 2022), 47–98.

early Islamic Levant.[34] We can thus no longer assume that pre-Islamic Arabic was equivalent with Classical Arabic, that Classical Arabic itself—even as a target—consisted of a single set of features, or that there were no other prestigious varieties with which scribes might interact when composing a text or a translation.

Perhaps most significantly, Middle Arabic is only a negatively defined category, which "encompasses all the attested written layers of the language which can be defined as entirely belonging neither to Classical Arabic nor to colloquial Arabic."[35] Middle Arabic, insofar as it represents the mixing of various registers and/or varieties, is a sociolinguistic phenomenon that has likely always existed.[36] Authors of medieval Middle Arabic texts incorporated any number of registers and varieties when composing texts,[37] and they did so in patterned ways in various places and times.[38] Numerous systematic norms have been identified in different corpora, and in some cases these persisted over long periods of time.[39] And crucially, although features associated with what we think of as Classical Arabic were clearly part of the spectrum along which many Middle Arabic authors worked, they were not the only "high-register" forms from which authors might select.[40] In other words, the claim that authors of Middle Arabic were incompetent at producing Classical Arabic is, in most cases, unwarranted and untenable.[41] Thus, the fact that the Gospels are written in Middle Arabic simply indicates that they are not written in Classical Arabic; it does not entail that their authors were targeting Classical Arabic, nor that the variation in them is not patterned.

Another, related claim that serves as a reason for the relative lack of attention to the comparative grammar of the Gospels, in particular, is the scholarly belief that the Gospels' status as translations makes their grammar somehow different from "real" Arabic, however conceived. For example, Blau claims that "most of the ChA [= Christian Arabic] texts are translations from Greek and Syriac, sometimes (especially the translations of the Holy Writ)

34. A. Al-Jallad, *The Damascus Psalm Fragment: Middle Arabic and the Legacy of Old Ḥigāzī* (Chicago: Oriental Institute, 2020).

35. J. Lentin, "Middle Arabic," in *The Encyclopedia of Arabic Language and Linguistics*, ed. K. Versteegh, vol. 3, 215–24 (Leiden: Brill, 2008), at 216.

36. G. Khan, "Middle Arabic," in *Semitic Languages: An International Handbook*, ed. S. Weninger et al., 817–35 (Berlin: de Gruyter, 2011), at 817.

37. Lentin, "Middle Arabic," 217–18.

38. A. Bellem and G. R. Smith, "'Middle Arabic'? Morpho-syntactic Features of Clashing Grammars in a Thirteenth-Century Arabian Text," in *Languages of Southern Arabia: Papers from the Special Session of the Seminar for Arabian Studies Held on 27 July 2013*, supplement to *Proceedings of the Seminary for Arabian Studies* 44, ed. O. Elmaz and J. C. E. Watson, 9–17 (Oxford: Archaeopress, 2014).

39. J. Lentin and J. Grand'Henry, introduction to Lentin and Grand'Henry, *Moyen arabe et variétés mixtes*, xvii - xxiii, at xviii–xx.

40. P. W. Stokes, "In the Middle of What? A Fresh Analysis of the Language Attested in the Judaeo-Arabic Commentary on *Pirqê ʾĀvōṯ* (The Sayings of the Fathers), Middle Arabic, and Implications for the Study of Arabic Linguistic History," *Journal of Semitic Studies* 66, no. 2 (2021): 379–411.

41. J. den Heijer, "Introduction: Middle and Mixed Arabic, A New Trend in Arabic Studies," in *Middle Arabic and Mixed Arabic: Diachrony and Synchrony*, ed. L. Zack and A. Schippers, 1–26 (Leiden: Brill, 2012), at 11.

so awkward and literal that they *are hardly worthy of being called Arabic at all.*"[42] More recently, in his article on the *status quaestionis* of Arabic Bible research, Ronny Vollandt argues regarding the linguistic study of the texts that

> the linguistic aspect [of the Arabic Bible] is certainly important, and should not be neglected. However, *biblical translations often follow a grammar of their own*, which is governed by a wish to imitate the exalted source text and maintain a high degree of literalism in the translation. . . .The language used could often be described as a professional translation language. . . . This means that these biblical translations reflect only a rather specific register of Middle Arabic. *Thus the concentration on the linguistic aspects alone limits and undermines the historical significance of these translations.*[43]

Although it is undoubtedly the case that translators often imitated their source material, especially in lexical choice but also occasionally in aspects of syntax such as word order, these instances are far from sufficient to warrant such relegation of translations. Translations, even when intentionally wooden, were nothing if not intended to communicate the interpreted meanings of texts to their audience.[44] Labeling syntactic features as odd, apparently for no other reason than their divergence from the norms associated with Classical Arabic, is methodologically unsafe in light of the evidence for the diversity of pre- and early Islamic Arabic briefly reviewed above. Further, we still know far too little about too many aspects of the varieties of Arabic used by medieval Christians to feel confident judging how natural or unnatural any structure might have seemed to contemporary Arabic speakers.[45] In addition, although it is undeniable that genre-specific norms can present challenges for scholars and students of translations, whatever their source and target languages, this is no less the case with other genres. Indeed, any text of any genre will have norms and practices that must be taken into account in order to accurately deduce the linguistic reality evident in the text.[46]

---

42. Blau, *Grammar of Christian Arabic*, 20; emphasis mine.

43. R. Vollandt, "*Status Quaestionis*," 454; emphasis mine.

44. As Kashouh rightly notes, arguing that "the *meaning* which the *Sacred Text* conveys is the goal" of these translations; Kashouh, *Arabic Versions*, 7.

45. No lesser a text than the Quran contains numerous examples of oddities, in terms of both syntax and lexicon, to the degree that numerous works, from al-Farrāʾ's *Maʿānī al-Qurʾān* to Ibn Khalawayh's *al-Ḥuǧǧa*, were written to document linguistic variants of the Quran's consonantal text (*rasm*) and recitation traditions (*qirāʾāt*). And although debates have raged over the language behind the consonantal, no one, to my knowledge, has argued that the Quran is of limited or no linguistic significance for understanding the linguistic situation of the Hijaz in the seventh century CE.

46. This task is likely even more complicated in some cases. For example, Kootstra's study of the inflection of construct *ʾab*, "father," in the papyri from the first three Islamic centuries reveals that genre has some effect on the degree to which it is inflected as in Classical Arabic, but an even larger effect was found in contexts within documents that are more formulaic, such as lines of address. So in formulaic contexts, *ʾab* in construct is inflected more frequently according to Classical Arabic norms than it is in subsequent parts of letters that are less formulaic. See F. Kootstra, "A Quantitative Approach to Variation in Case Inflection in Arabic Documentary Papyri: The Case of *ʾab* in Construct," *Journal of American Oriental Society*, 2022. Therefore, genre, while meaningful, is always a challenge for understanding the language system and simultaneously far from the only

A final methodological issue with this perspective, and one that is often implicit, is its equation of lexicon (and to a lesser extent syntax) with "grammar." Though important, grammatical analysis goes far beyond lexicon and word order, and I see no reason translations cannot provide scholars with phonological and morphological data that are just as meaningful as what is found in other genres. Whether the registers attested in the Gospels were strictly those associated with scribal translations or whether they were more widespread, I take it as uncontroversial that it is important to determine what patterns exist and how they correlate with other domains, such as text type.

The present study therefore approaches the grammatical variation attested in the Arabic Gospel manuscripts without assuming any particular identity for the target register(s). Conversely, it does assume that the grammatical variants within and across manuscripts reflect intentional decisions made by the scribes. Even if there was no homogeneous target at which a scribe aimed,[47] the absence of a single grammatical standard can coexist with patterns in scribal decision-making just as a modern educated Arabic speaker who mixes colloquial and Modern Standard Arabic in various ways will produce output that falls along a spectrum of normed patterns and linguistic outputs.[48] I assume that scribes wrote what they intended, in ways that followed some norms or patterns, which it is our job to identify.

One possible methodological objection to my including multiple witnesses to the same family or subfamily is that given their close textual relationship, we might expect related manuscripts to replicate much of the same grammar and thus to be naturally closer to one another than to manuscripts from other families. However, the witnesses to a particular family or subfamily frequently attest to different grammatical patterns, especially in terms of case marking. The versions of Matthew 4:16 ("the people sitting in darkness have seen a great light") found in manuscripts in subfamily J^b illustrate the point (see Table 4).

In three of these eight manuscripts (SAr. 69, 84, and 80), both noun and adjective are marked accusative with *tanwīn* and *alif*, and in two others a "dialectal *tanwīn*" distribution is attested, with the noun marked (with *kasratān*) and the adjective unmarked in one (SAr. 90), and both noun and adjective marked with *tanwīn alif* but only the noun with *tanwīn* in another (Vat. Borg. Ar. 71).[49] In four of the eight, no *tanwīn* marking is attested. It is clear, then, that many grammatical categories, such as case and *tanwīn* marking, are variable even within manuscript subfamilies.

factor. In so-called original works, context and the nature of the document affect the nature of the grammar attested. Although translations present obvious challenges, these are not unique, nor are translations self-evidently farther from "real" Arabic. Indeed, the features detailed here are also common to other, non-translated texts.

47. Kashouh, in *Arabic Versions*, 7, argues that "there seems to be no effort in producing a *linguistic homogenous* text" (emphasis original). As I argue throughout this paper, although this is true, the common inference—that a lack of homogeneity equals a lack of meaningful patterns—is unwarranted and contradicted by the data.

48. G. Meiseles, "Educated Spoken Arabic and the Arabic Language Continuum," *Archivum Linguisticum* 11, no. 2 (1980): 118–43; R. Henkin, "Functional Codeswitching and Register in Educated Negev Arabic Interview Style," *Bulletin of the School of Oriental and African Studies* 79, no. 2 (2016): 279–304.

49. On the distribution of *tanwīn* in these and other manuscripts from the Arabic Gospel corpus, see P. W. Stokes, "Nominal Case in Christian Arabic Translations of the Gospels (9th–15th Centuries CE)," *Arabica*, forthcoming.

Table 4. Variation in case marking of "great light" in Matt. 4:16 in manuscripts of Family J[b].

MS/folio	Arabic text
SAr. 69, 8v	الشعب الجالس في الظلمه ابصروا ضوا عطيما [sic]
SAr. 80, 8r	الشعب الجالس في الظلمه لقد ابصر نورًا عظيمًا
SAr. 82, 11r	الشعب الجالس في الظلمه ابصر نور عظيم
SAr. 84, 16	الشعب السالك في الظلمه ابصر نورًا عظيمًا
SAr. 89, 5v	الشعب الجالس في الظلمه ابصر نور عظيم
SAr. 90, 12r	الشعب الجالس في الظلمه ابصر نورٍ عظيم
SAr. 106, 14	الشعب الجالس في الظلمه ابصر نور عظيم
Vatican Borg. Ar. 71, 6r	الشعب الجالس في الظلمه ابصر ضوًا عظيما

Another example is the treatment of the noun *ʾaḫ*, "brother," in Matthew 5:23. The noun *ʾaḫ* is one of five nouns (in Arabic called *al-ʾasmāʾ al-ḫamsah*, "the five nouns") that, when in construct, manifest a long case vowel; e.g., *ʾaḫūka* "your (nom) brother" / *ʾaḫīka* "your (gen) brother" / *ʾaḫāka* "your (acc) brother." In Matthew 5:23 ("If you bring your offering to the altar and remember that your brother has something against you"), *ʾaḫ* follows the particle *ʾan(na)*, which variously triggers nominative *ʾaḫūk* or accusative *ʾaḫāk* in various manuscripts. This variation occurs within a single manuscript family, as Table 5 shows for Family A.

Table 5. Variation in the inflection of *ʾaḫ* in Matt. 5:23 in manuscripts of Family A.

MS/folio	Arabic text
SAr. 74, 8r	وان قربت قربانك علي المذبح وذكرت ان اخوك غضبان عليك
SAr. 72, 6v	فان انت قربت قربانك على المذبح وذكرت هناك ان اخاك غضبان عليك
Vat. Borg. Ar. 95, 8v	وان قربت قربانك على المدبح وذكرت ثم ان اخوك غضبان عليك

In SAr. 74 and Vat. Borg. Ar. 95, the particle *ʾinna* triggers the nominative case, whereas in SAr. 72, it triggers the accusative.[50] Grammatical variation regularly crosses manuscript families. This variation includes differences that affect both consonantal representations and, in vocalized texts, vocalization markings. If grammatical variation were essentially random, as is often (implicitly) assumed, then we should expect to find either no meaningful clustering at all or clustering that does not correspond with manuscript families or subfamilies. As we will see, that is not the case.

50. The same variation between nominative and accusative following ان occurs in the early Arabic papyri; see Kootstra, "Quantitative Approach," §4.1.1 for data and analysis.

There is an almost unlimited number of variants that could be selected for such a study, and the following list should be expanded in future studies. I have chosen features that are attested in both unvocalized and vocalized manuscripts. These include a number of *tanwīn*-related features, which deserves methodological comment before proceeding. As I argue elsewhere, Christian scribes used the orthographic tools at their disposal in both unvocalized and vocalized manuscripts to mark *tanwīn* in certain syntactic contexts. Those contexts would, in Classical Arabic grammar, have variable case markings, depending on the context. I show that although Christian scribes indicated the presence of *tanwīn* in these contexts in some of the manuscripts included in his study, especially those belonging to Family J^{b}, there were apparently no underlying phonetic distinctions between cases in its realization. This, I suggest, parallels contemporary "dialectal tanwīn," where dialects attest a singular morpheme (usually /in/, /ən/, or /an/, but also /un/ in Yemen). This means that two manuscripts from the same family can mark the same word with *tanwīn*, but with two different orthographic means:

SAr. 82, 9r صوتٍ سمع في الرامه "A voice is heard in Ramtha"
but
SAr. 80, 6r صوتً سُمع في الرامه "A voice is heard in Ramtha"

This orthographic variation will be addressed further below in Section 5.

Whereas the vocalized manuscripts make use of *tanwīn* vocalizations as well as *tanwīn alif*, unvocalized manuscripts rely purely on the latter. Therefore, unless otherwise specified, the presence or absence of *tanwīn* will include any combination of orthographic means, whether *tanwīn alif* (Families A, C, D, and H) or various vocalizations (Families J and K). For this reason, I do not differentiate between *tanwīn* markings with different orthographic or vocalization signs. Since these manuscripts tend to mark the same roles, but not with the same frequency or in the same places, the goal here is to discover patterns of *tanwīn* marking and to reduce the degree to which orthographic variation might otherwise obscure grammatical patterning.[51] Finally, note that, in some instances, manuscripts use the *fatḥatān* vocalization without a *tanwīn alif*, even when orthographically eligible for the latter (e.g., SAr. 106 ابً instead of ابًا). To distinguish these two forms in transliteration, I have transliterated the former -an and the latter -AN, i.e., *ʾab-an* transliterates ابً and *ʾab-AN* transliterates ابًا. Otherwise, I have transliterated *tanwīn* simply as -N. The implications of these orthographic notes will be discussed in the context of other aspects of orthographic variation in Section 5 below.

Table 6 lists the grammatical variants included in this study. I have chosen features that, insofar as possible, are comparable across the manuscripts included. Another way of approaching a grammatical comparison would be to tally features for comparison. For example, we might choose a portion (or the whole) of each manuscript, tally up all occurrences of the dual on nominal forms, tagging for historically nominative and oblique cases, and then calculate the percentage of Classical and non-Classical Arabic forms for each. I have done this for triptotic case marking in both unvocalized and vocalized texts.[52] But although revelatory in one sense, such an approach obscures in another: the percentages

51. For a discussion of some patterns based on the vocalizations used, see Stokes, "Nominal Case," §3.

52. Stokes, "Nominal Case."

can vary depending on the frequency of a category's occurrence, and percentages are often similar across manuscripts in which the details are otherwise quite different. I have therefore chosen specific verses in which the manuscripts attest similar grammatical structures. This allows us to get a sense of how similar or different scribal grammatical choices were across the manuscripts.

Table 6. Selected grammatical variants from the Gospel of Matthew.

Biblical text	**Grammatical category**	**Variants attested**
25:36–44	KSW/KSY variation	1. fa-kasaytumūnī, 2. fa-kasawtumūnī
	KSW/KSY variation	1. fa-kasaynāk, 2. fa-kasawnāk, 3. ʾalbasnāk
	verbal negation	1. fa-lam taksūnī, 2. fa-mā kasay-tumūnī
	verbal negation	1. lam negation, 2. mā except for KSW/Y, 3. lam bookends
	adjectives nunated or not	1. nunated, 2. not nunated
14:21	men singular or plural	1. rağul, 2. riğāl
	men nunated or not	1. rağul-N, 2. rağul
	men before or after number	1. before, 2. after
	thousands singular or plural	1. ʾalf, 2. ʾālāf
15:38	men singular or plural	1. rağul, 2. riğāl
	men nunated or not	1. rağul-N, 2. rağul
	men before or after number	1. before, 2. after
	thousands singular or plural	1. ʾalf, 2. ʾālāf
16:9	five (loaves) masculine or feminine	1. ḫams ḫubz(āt), 2. ḫamsat ḫubz(āt)
	loaves singular or plural	1. ḫubz, 2. ḫubzāt
	five loaves definiteness pattern	1. al-ḫams ḫubz(āt), 2. al-ḫams al-ḫub-zāt, 3. ḫams al-ḫubzāt
	five (thousand men) masculine or feminine	1. ḫams thousands, 2. ḫamsat thou-sands
	thousand(s) singular or plural	1. ʾālāf, 2. ʾalf
	five (thousand men) definite or indefinite	1. ḫams(at) thousands, 2. al-ḫams(ah) thousands

16:9, cont.	seven (loaves) masculine or feminine	1. sabᶜ loaves, 2. sabᶜat loaves
	loaves singular or plural	1. ḫubzāt, 2. ḫubz
	seven loaves definiteness pattern	1. al-sabᶜ al-ḫubzāt, 2. al-sabᶜ ḫubzāt, 3. sabᶜ al-ḫubzāt
	four (thousand men) masculine or feminine	1. ʾarbaʿ thousands, 2. ʾarbaʿat thousands
	four (thousand men) definite or indefinite	1. ʾarbaʿ(at) thousands, 2. al-arbaᶜah thousands
	thousand(s) singular or plural	1. ʾālāf, 2. ʾalf
19:6	initial verb indicative or non-indicative	1. indicative, 2. non-indicative, 3. singular, 4. laysa
	predicate of verb nunated or not	1. unmarked, 2. fully marked, 3. dialectal *tanwīn*
	final predicate nunated or not	1. unmarked, 2. fully marked, 3. adjective only
17:17	"unbelieving generation" indefinite and nunated or definite	1. indefinite ǧīl-N multawiy-N, 2. definite
19:17	"No one is good but God" wording	1. laysa ʾaḥad, 2. mā ʾaḥad, 3. laysa ṣāliḥ, 4. wa-lā
	"No one is good but God" nunation pattern	1. ʾaḥad ṣāliḥ, 2. ʾaḥad-N ṣāliḥ, 3. ʾaḥad-N ṣāliḥ-N, 4. ṣāliḥ, 5. ʾaḥad-u ṣāliḥ, 6. ṣāliḥ-N
13:27	*zawān* nunated or not	1. zawān-N, 2. zawān, 3. definite
9:9	object of *naẓara* ʾilā nunated or not	1. fully marked, 2. dialectal *tanwīn*, 3. unmarked, 4. transitive verb used
17:14	subject *raǧul*/ʾinsān nunated or not	1. nunated, 2. not nunated
17:20	subject šayʾ nunated or not	1. šayʾ-N, 2. šayʾ
3:9	ʾab nominative or accusative	1. ʾab-N, 2. ʾabūnā, 3. ʾabānā
8:6	*ṭarīḥ* as predicate of ʾinna nunated or not	1. ṭarīh-N, 2. ṭarīḥ, 3. maṭrūḥ
	muḫallaᶜ nunated or not	1. muḫallaᶜ-N, 2. muḫallaᶜ

5:23	case of ʾaḫ	1. ʾn ʾaḫāk, 2. ʾn ʾaḫūk, 3. li-ʾaḫīk, 4. fī nafas ʾaḫīk
	predicate of ʾanna nunated or not	1. nunated, 2. not nunated
2:8, 3:3, 3:17, 17:5, 25:6, 27:46, 27:50	*ṣawt* nunated or not	1. marked, 2. unmarked
5:29–30	*ʾaʿḍāʾi-ka "your body parts" written with or without <Y>	1. ʾaʿḍāk, 2. ʾaʿḍāYk, 3. mafāṣil-ak, 4. ʾawṣāl-ak
19:8	**nisāʾi-kum* "your wives" written with or without <Y>	1. nisākm, 2. nisāYkm, 3. accusative
3:11	**ḥiḏāʾi-h* "his sandal" written with or without <Y>	1. ḥiḏāh, 2. ḥiḏāYh, 3. accusative, 4. no suffix
25:6	**li-liqāʾi-h* "to meet him" written with or without <Y>	1. li-liqāY-h, 2. li-liqā-h, 3. another root, 4. istiqbāl-h
21:42	"builders" nominative or oblique	1. bannāʾūn, 2. bannāʾīn
	"become the cornerstone" grammar	1. raʾs li-l-zāwiyah, 2. li-raʾs al-zāwiyah, 3. raʾs al-bunyān, 4. raʾs-N li-l-zāwiyah, 5. li-l-zāwiyah raʾs-A(N), 6. raʾs al-zāwiyah, 7. fī raʾs al-zāwiyah, 8. rukn al-banā
6:9–13	case of "Our Father"	1. ʾabūnā, 2. (yā) ʾabānā
	verbal form of "Hallowed be Thy name"	1. yataqaddas ism-ak, 2. yuqaddas ism-ak, 3. li-yataqaddas ism-ak, 4. taqaddas ism-ak
	verbal form of "Thy kingdom come"	1. yaʾtī mulk-ak, 2. taʾtī malakūt-ak, 3. li-taʾti malakūt-ak, 4. li-yaʾti malakūt-ak, 5. yaʾti malakūt-ak
	verbal form of "(Thy will) be done"	1. yakūn, 2. takūn, 3. (fa)li-takun, 4. li-takūn
	verbal form of "Give (us this day)"	1. ʾaʿṭīnā, 2. ʾaʿṭinā, 3. irzaqnā

While many of the features selected above are straightforward, several require some contextual discussion. The features from Matthew 25:36–44 concern the alternation between III-W and III-Y of the verb KSW/KSY "to clothe," which in some manuscripts is KSY in each verse but in others alternates (*kasaytamūnī* "you clothed me" but *kasawnāk* "we clothed you"). Further, the negation of specific verbs is in many manuscripts accomplished with

lam (Family A and Family K), whereas others use *mā* for the first and last in the series and lam in the remainder of instances (Family J[a]), and a third group uses *mā* (except in Matthew 25:43).

The next three sets of features all occur in three successive chapters of Matthew (14–16) in which Jesus first feeds five thousand men (not counting women and children; Matt. 14), then feeds four thousand (Matt. 15), and finally rebukes the disciples for not remembering both groups (Matt. 16). In each, the treatment (grammatical gender, number, and definiteness patterns) of the numbers (five, four, and thousands) and nouns (men and loaves of bread) vary across manuscripts.

In Matthew 17:17, Jesus addresses the "unbelieving generation," which is either grammatically indefinite or definite. In the former subset, the addressee is frequently nunated with *fatḥātān* and *tanwīn alif* (but once without *alif*):

SAr. 82, 42r يا جيلًا ملتويًا "O unbelieving generation"

but

Vat. Borg. Ar. 71, 30r يا جيلً ملتويا "O unbelieving generation"

This pattern is not the default one in Classical Arabic, but such a nunated form is attested in some poetic traditions, especially when the noun is followed by an adjective or another noun, as in *yā mūqid-an nār-an*, "O you who would kindle a fire!" and *ʾa rākib-an kamiyy-an*, "O you heroic horseman!"[53]

Finally, the word *ṣawt*, "a voice," occurs at least seven times in the Arabic versions of the Gospel of Matthew included in this study (Matt. 2:8, 3:3, 3:17, 17:5, 25:6, 27:46, 27:50). In many of these instances across the manuscripts the word is nunated, even in manuscripts that otherwise rarely indicate *tanwīn* orthographically.

As with the lexical data discussed in Section 3, I entered each variant into rows in a spreadsheet, with each column dedicated to one manuscript, as in Tables 2 and 3. In those tables as in Table 7, the abbreviated labels for the manuscripts that constitute the column heads are color-coded to denote manuscript families as in Table 1. The biblical citations in the left-hand column match those in Table 6 above. The numbers in the cells refer to the variant numbers in column 3 of Table 6. Finally, a dash indicates a lacuna in the manuscript, and an empty cell shows that a different phrase is used.

53. Fischer, *Grammar of Classical Arabic*, 96, n. 4.

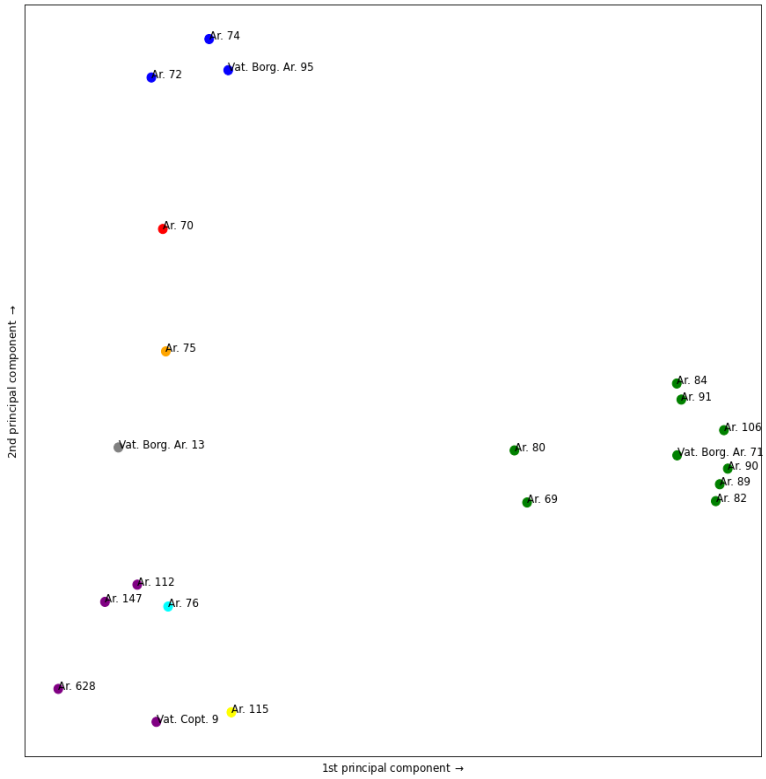
Table 7. Distribution of selected variants within and across manuscripts.

Text	MS																				
	74	V95	72	70	75	V13	115	106	V71	84	82	89	90	91	69	80	76	112	147	628	V9
25:36–44	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	–	2	2	2	2	2
	1	1	1	1	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	1	1	3	2	2	2	2	2	2	2	2	2	1	1	1	1	1
	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	4
14:21	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1	2	2	1
	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1	1	1
	1	1	1	1	1	2	2	1	1	1	1	1	1	1	2	1	2	2	2	2	2
	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	2	2	2	2	2	2
15:38	1	1	1	1	1	2	2	2	2	1	2	2	2	1	2	1	2	2	2	2	2
	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	–	1	1	1
	1	1	1	2	1	2	1	2	1	2	2	2	2	2	2	1	2	–	2	1	1
	1	1	1	1	1	1	–	1	1	1	1	1	1	1	1	1	1	–	1	1	1
16:9	1	1	1	2	2	2	1	1	1	2	1	1	1	1	1	2	2	2	1	1	1
	1	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	2	2	2	2	2	2	2	1	2	1	1	1	1	1	1	2	2	–	2	2	2
	1	1	1	2	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	3	1
	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	–	1	1	1
	1	1	1	1	1	2	2	2	1	1	2	2	2	1	2	1	1	–	2	1	2
	2	2	1	2	1	2	2	2	2	2	1	1	2	1	2	2	1	2	1	1	1
	2	2	1	1	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	3	2
	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1
	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	1	1	1	1	1	2	2	2	1	1	2	2	2	1	2	1	1	1	2	1	2

19:6	2	2	2	2	2	1	3	2	1	2	1	1	1	2	2	1	2	4	2	3	1
	1	3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	1	1	1	2	1	2	2	1	1	2	1	1	1	2	2	2	3	2	2	2	2
17:17	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2
19:17	1	1	3	3	1	3	2	1	1	1	1	1	1	1	1	1	2	2	4	2	2
	1	1	5	5	2	5	3	1	1	1	1	2	5	1	2	2	4	4	4	6	6
13:27	2	2	2	2	1	3	2	1	1	1	1	1	1	1	2	3	2	2	2	2	2
9:9	3	3	3	4	4	4	4	2	2	2	1	2	2	2	3	1	4	4	4	4	4
17:14	2	2	2	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1
17:20	1	1	1	2	2	2	2	1	–	1	1	1	1	1	2	1	1	2	2	1	1
3:9	1	1	1	1	4	4	1	1	3	3	3	3	3	–	4	3	1	1	4	4	4
8:6	2	2	2	2		2	2	2	2	2	2	2	2	2	1	2					
	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1
5:23	2	2	1	1	3	4	1	2	–	2	2	2	2	2	1	1	1	2	1	1	1
	2	2	2	1	3	2	2	2	–	1	1	1	2	2	2	1	1	1	1	1	1
2:8, 3:3, 3:17, 17:5, 25:6,	2	2	2	1	1	2	1	2	2	2	1	2	2	–	2	1	2	2	1	1	1
	2	2	2	2	1	2	2	1	1	2	1	2	2	–	2	1	2	1	2	1	1
	2	2	2	2	1	2	2	2	2	2	2	1	1	–	2	2	1	1	2	1	1
	2	2	2	2	1	2	1	2	2	2	2	2	2	2	2	1	2	2	2	1	1
	2	2	2	2	2	–	2	2	1	2	2	1	2	2	2	2	2	2	2	3	3
	2	2	2	2	2	2	2	2	1	2	1	2	2	2	2	1	2	2	2	1	1
	2	2	2	2	2	2	1	2	1	2	2	2	2	2	2	1	2	2	2	1	1
5:29–30	1	1	1	2	3	4	2	1	–	1	1	1	1	1	1	2	2	2	2	2	2
19:8	3	3	3	–	3	3	2	1	1	1	1	1	1	1	2	1	3	3	3	3	3
3:11	3	3	3	3	1	1	3	3	1	3	1	1	1	–	1		2	1	3	3	3
25:6	1	1	1	1		3	4	1	1	1	1	1	1	1	1	1	2	2	2	2	2

21:42	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
	1	1	1	3	2	8	4	5	6	5	5	5	5	5	5	5	7	4	6	6	6
6:9–13	1	1	1	1	1	2	2	2	–	2	2	2	2	2	2	2	2	2	–	2	2
	1	1	1	2	1	4	3	3	–	3	3	3	3	3	3	1	3	1	–	5	1
	1	1	2	1	2	2	3	3	–	3	3	3	3	3	3	3	4	2	–	2	2
	1	2	2	2	1	3	3	3	–	3	3	3	3	3	4	3	3	2	–	2	2
	1	2	2	1	2	4	2	2	–	2	2	2	2	2	2	1	2	2	–	2	2

Fig. 5. PCA of grammatical variation in Gospel manuscripts.[54]



A PCA of the data yields the results displayed in Figure 5. The figure shows clear clustering, especially among manuscripts of the same manuscript family. Since, as noted above, scribes felt comfortable changing or modifying the grammar of the manuscript from which they copied, the fact

54. The first principal component accounts for 29.2% and the second accounts for 18.1% of the variation in the data, for a total of 47.3%. This is rather low, but a low percentage is almost inevitable given the large amount and chaotic nature of the data.

that the manuscripts pattern together so consistently is significant—and even more so given that many of the features selected above are indicated via different orthographic means in textually related vocalized manuscripts. In other words, the scribes were clearly following similar grammatical paradigms, which they represented in a variety of orthographic ways, and thus were not simply copying what their exemplars attested.

Family A (marked in blue) forms a cluster, and within the family, SAr. 74 and Vat. Borg Ar. 95 are closer to each other than either is to SAr. 72. This supports previous scholars' claims that SAr. 72 represented a stylistic improvement over SAr. 74 and Vat. Borg. Ar. 95.[55] Of the other families attested, Family D (red) is closest to Family A, and indeed there are several places in which SAr. 70 shares non-Classical grammatical features word for word with members of Family A, such as marking *ʾaḥad* with *tanwīn alif* in Matthew 9.30:

SAr. 72, 11v وقال لهما انظرا ان لا يعلم احدا "And he said to them, 'See that no one knows'"

SAr. 70, 12r وقال انظر الا يعلم احدا "And he said, 'See that no one knows'"

It is, of course, possible that SAr. 75 represents "the culmination of the attempt on the part of a group of Palestinian Christians to achieve an Arabic version of the Gospel in the early Islamic period which could pass for Literary Arabic," as Griffith claims.[56] Alternatively, the results presented here suggest that SAr. 70 is another candidate for such a culmination, or at least belongs to a closely related grammatical tradition in which *tanwīn* on non-accusatives is limited to lexical items such as *ʾaḥad*.

The manuscripts of Family K (purple) also cluster together, and within Family K, SAr. 112 and 147 are particularly closely related to each other, while SAr. 628 and Vat. Copt. 9 form another close pair. This patterning is characterized, among other ways, by more consistent use of *tanwīn* vocalization signs in the latter two manuscripts.

Family J^b (green) once again forms a distinct cluster, which stands apart from the other families and from the representatives of J^a (yellow) and J^c (aqua). Seven of the nine J^b manuscripts cluster quite tightly; SAr. 69 and SAr. 80 are slight outliers. The latter two manuscripts are outliers in other respects, too. As we saw in Section 3, SAr. 69 stands apart from the other J^b manuscripts also in terms of its lexical and orthographic characteristics. SAr. 80 is the latest of the representatives of J^b included in this study and is grammatically closer to Family K, for example, than the other members of Family J^b are. This convergence with Family K makes sense in the light of Family K's increasing importance and dominance after the thirteenth century CE.[57] The representatives of J^a and J^c are much closer to each other than either group is to J^b, which supports Kashouh's claim of a closer relationship between the two on the basis of verbal and lexical similarities,[58] and they pattern quite closely with Family K. Again, although the scribes of J^a and J^c evidently relied on different exemplars than those of Family K did, they were clearly part of a scribal culture—with corresponding grammatical preferences—that was very similar to the culture behind Family K.

55. Kashouh, *Arabic Versions*, 93.

56. Griffith, "Gospel in Arabic," 155.

57. Kashouh, *Arabic Versions*, 205–6.

58. Ibid., 203.

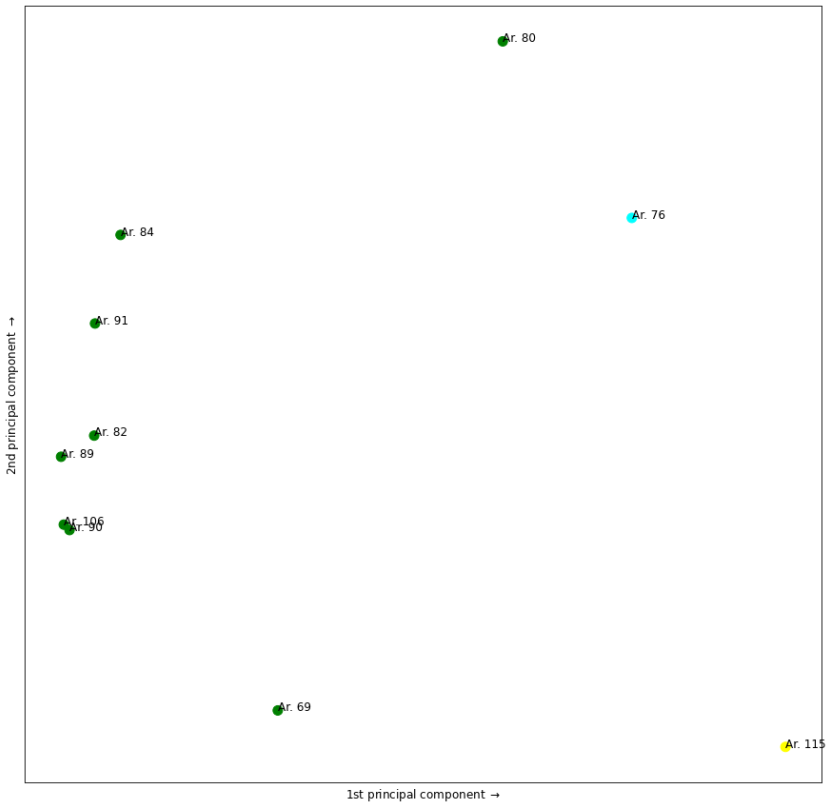
In general, we can discern three primary clusters:

1. Family A
2. Families K, J[a], and J[c]
3. Family J[b]

The latter is the outlier, as represented visually by the differences along the X and Y axes, which indicate a high degree of dissimilarity in both the first and the second principal components. Between 1 and 2 are several families that, so far, are represented by only a single manuscript each (at least, only one available to me at present).

In order to get a sense of how the manuscripts of each family pattern in relation to other manuscripts from the same (sub)family, I ran PCAs that included manuscripts from only one family. As in Section 3, I focused on the families represented by the most manuscripts: Family J, Family J[b], and Family K. The results of the PCA of the data from Family J are shown in Figure 6.[59] As in my analysis of the "Kingdom of God"/"Kingdom of heaven" data, SAr. 69 and SAr. 80 are clear outliers. Intriguingly, SAr. 80 is grammatically closest to SAr. 76. This proximity deserves closer study.[60]

Fig. 6. PCA of grammatical variation in Family J.

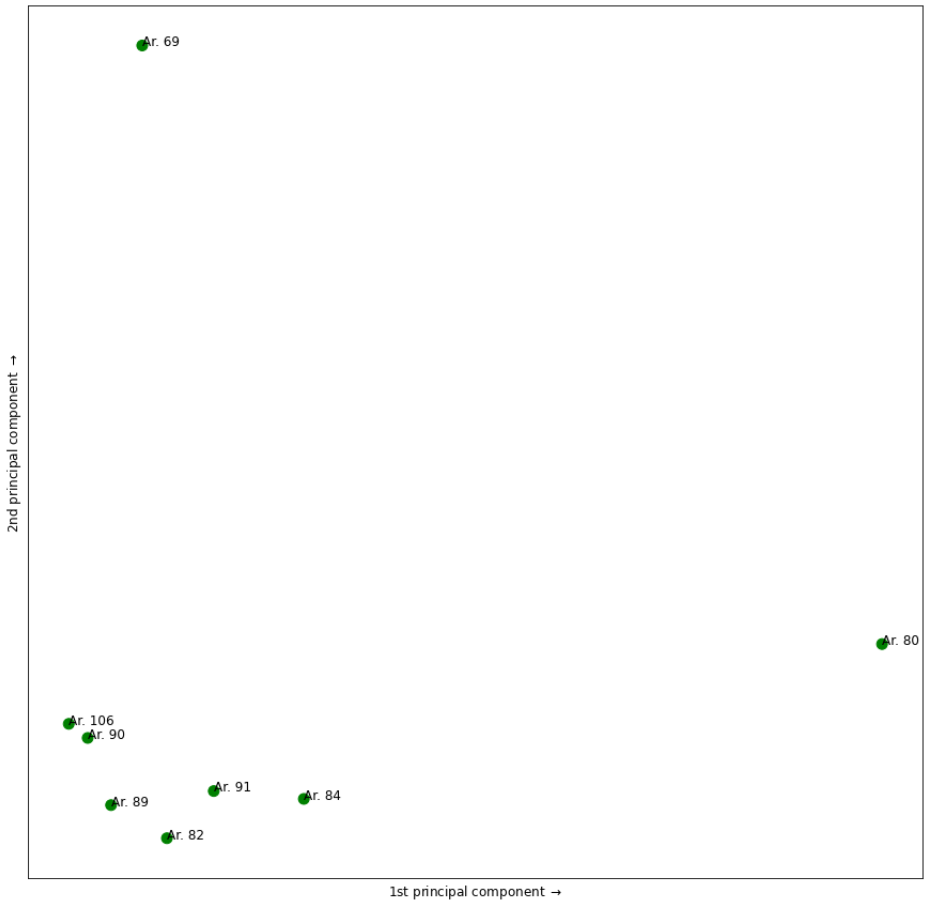


59. The first principal component captured 37% of the variation, and the second principal component captured 19%. Because Vat. Borg. Ar. 71 lacked several features, I omitted it from this analysis.

60. Impressionistically, this is unsurprising. Both, for example, utilize *tanwīn* vocalizations more frequently than other manuscripts in Family J[b] do.

When we include only manuscripts from Family J[b], we obtain the results shown in Figure 7. As can be seen, this does little to change the picture: most of the manuscripts are grammatically very similar, with SAr. 69 and SAr. 80 clear outliers.[61] Finally, Figure 8 illustrates the results of an examination of the manuscripts from Family K.[62] Again, there are two clear outliers, SAr. 147 and SAr. 112.

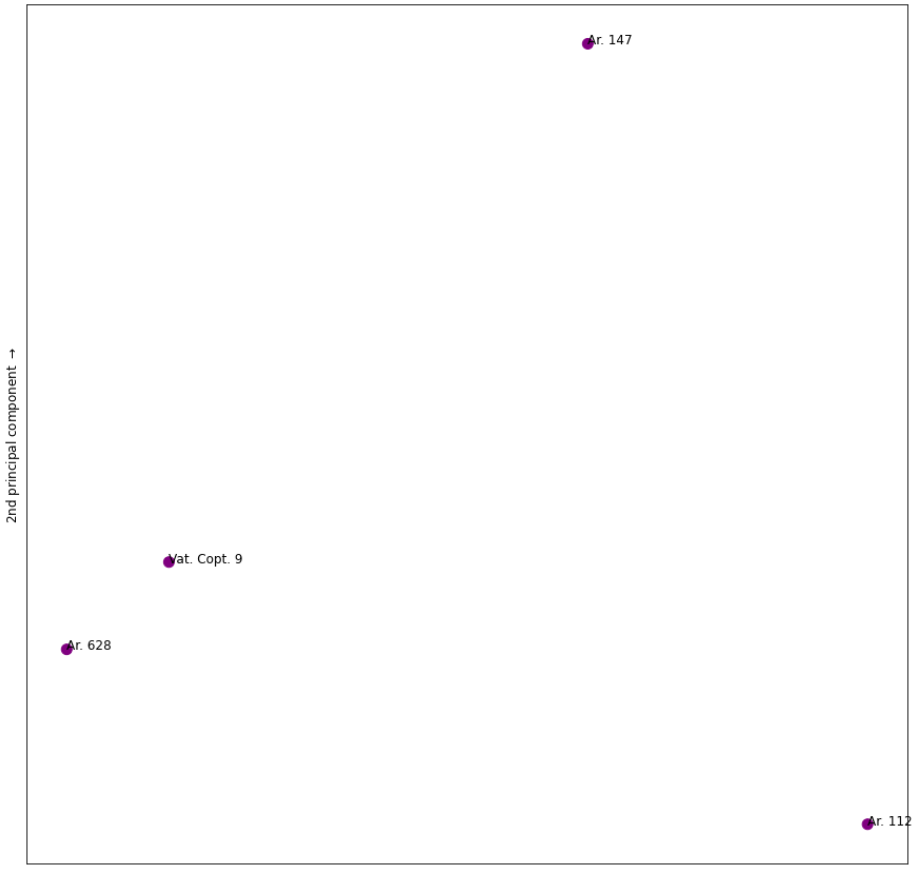
Fig. 7. PCA of grammatical variation in Family J.



61. For a discussion of SAr. 80 and its history, see both Kashouh, *Arabic Versions*, 185–94; and J. Valentin, "Des traces de la *vetus syra* des évangiles en traduction arabe? Étude critique des variantes significatives en Mc 5,1–20 dans le *Sinaï arabe 80*," in *Reading the Gospel of Mark in the Twenty-First Century: Method and Meaning*, ed. G. Van Oyen, 765–79 (Leuven: Peeters, 2019).

62. The first principal component captured 56% and the second principal component 29% of the variation.

Fig. 8. PCA of grammatical variation in Family K.



Like idiosyncratic lexical variation, grammatical variation, too, exhibits significant patterns. The different textual traditions are characterized by related but distinct grammatical traditions. As shown above, manuscripts do not replicate the grammar of the other members of their respective families or of the exemplar, which makes the strong clustering found here all the more meaningful: it shows that they participated in a shared grammatical tradition. It is striking that a systematic comparison of grammatical variants across manuscripts reveals both rather clear grammatical traditions and the fact that these generally align with text type. By focusing solely on grammatical variants, we get very close in many respects to Kashouh's subgroupings. And intriguingly, analyses of individual (sub) families result in subgroupings that are remarkably similar to those arising from studies of textual variation, such as that in Section 3. So although scribes do not seem to have replicated every grammatical choice made in their respective exemplars, idiosyncratic variants were nevertheless copied with sufficient frequency to leave distinct signatures in the manuscripts.

5. Orthographic Variation

In addition to idiosyncratic lexical and grammatical variation, repeated idiosyncratic orthographic variation can be a powerful tool for establishing textual and scribal relationships. In a recent article, van Putten documents the idiosyncratic variation in the spelling of the phrase *niʿmat allāh*, "the grace of God," which in early manuscripts of the Quran is spelled sometimes with *tāʾ marbūṭah* نعمه الله and sometimes with *tāʾ maftūḥah* نعمت الله, in about equal numbers.[63] He then compares the way those two spellings pattern across manuscripts and shows that the idiosyncrasies were carefully and faithfully reproduced by the scribes of the Quran. The fact that they were reproduced with such care can only mean that all of these manuscripts descended from a single written archetype. Consequently, instead of the variation's being random, at the whim of each scribe, the spelling is the same at each location in every manuscript. It is thus reasonable to study idiosyncratic orthographic variation in the Gospel manuscripts in order to determine whether the same kind of faithful replication occurred, and if it did not, whether any meaningful patterns appear at all. Such a study will also allow us to weigh Kashouh's contention that orthographic variants are not valuable for establishing relationships between manuscripts of the same family.[64]

The first variant under study is the spelling variation associated with malakūt, "kingdom," in the subset of manuscripts that use it for some (or all) of the "Kingdom of God"/"Kingdom of heaven" phrases investigated in Section 3 above. *Malakūt* is spelled in two ways depending on the manuscript—with either *tāʾ maftūḥah* ملكوت or *tāʾ marbūṭah* ملكوه/ملكوة—and the spelling can vary even in the same manuscript and sometimes on the same page (cf. the similar phenomenon in the Quranic niʿmat allāh in Q 16:71 بنعمه الله vs. Q 16:72 بنعمت الله).

The orthographic variation in the spelling of *malakūt* is part of a larger orthographic divide. Historically, some words ending in *ūt* exhibited the same orthographic variation; for example, *tābūt*, "chest, ark," was spelled variously as تابوت and as تابوه, with no consensus on what phonetic realization underlies the وه- spellings. Whatever the difference in pronunciation was historically, by the time the Arabic Gospels were composed, the fact that the form ملكوت was seen as a legitimate spelling strongly suggests that the pronunciation underlying both was /malakūt/, at least in construct.

In some manuscripts, the word *malakūt* is spelled either ملكوت or ملكوه / ملكوة, and the distribution of the two spellings is idiosyncratic. In others, only ملكوت is used. There are no manuscripts in which only ملكوه / ملكوة is attested. Table 8 shows the variation in the subset of manuscripts in which *malakūt* occurs. In the table, the spellings are indicated by a representative consonant: [ه] indicates a spelling with *tāʾ marbūṭah* ملكوه / ملكوة, and [ت] indicates a spelling with *tāʾ maftūḥah* ملكوت. As in Tables 2, 3, and 7, dash (–) indicates a lacuna in the manuscript, and an empty cell indicates that a different phrase (either *mulk* or *mamlakah*) is used instead.

63. M. van Putten, "'The Grace of God' as Evidence for a Written Uthmanic Archetype: The Importance of Shared Orthographic Idiosyncrasies," *Bulletin of the School of Oriental and African Studies* 82, no. 2 (2019): 271–88.

64. Kashouh, *Arabic Versions*, 215.

Table 8. Orthographic variation in manuscripts containing *malakūt*.

Text	MSS																	
	V13	75	115	84	106	82	89	91	90	80	V71	69	76	112	147	68	628	V9
3:2	ت	ت		ه	ه	ه	ه	_	ه	ت	ه	ه		ت	ت	ت	ت	ت
4:17	ت	ت	ت	ه	ه	ت	ه	_	ه	ت	ه	ه		ت	ت	ت	ت	ت
5:3	ت	ت		ه	ت	ه	ه	ه	ه	ت	ت	ت		ت	ت	ت	ت	ت
5:10	ت	ت	_	ه	ت	ه	ه	ه	ه	ت	_	ت		ت	ت	ت	ت	ت
5:12	_	_	_	_	ه	ه	ه	_	_		_	ه	_	_	_	_	_	_
5:19a	ت	ت	ت	ه	ت	ه	ه	ه	ه	ت	_	ت		ت	ت	ت	ت	ت
5:19b	ت	ت		ه	ت	ه	ه	ه	ه	ت	_	ت		ت	ت	ت	_	ت
5:20	ت	ت	ت	ه	ت	ه	ه	ه	ه	ت	_	ت		ت	ت	ت	ت	ت
6:33	ت	ت	ت									_	ت	ت	_	ت	ت	ت
7:21	ت	ت	ت										ت	ت	ت	ت	ت	ت
8:11	ت	ت	ت										ت	ت	ت	ت	ت	ت
10:7	ت	ت	ت	ه	ت	ت	ت	ه	ت	ت	ه	ت	ت	ت	ت	ت	ت	ت
11:11	ت	ت	ت									ت	ت	ت	ت	ت	ت	ت
11:12	ت	_	ت										ت	_	ت	ت	ت	ت
12:28	ت	ت	ت	ت	ه	ه	ه	ت	ه	ه	ت	ت	ت	ت	_	ت	ت	ت
13:11	ت	ت	ت	ه	ه	ه	ه	ه	ت	ه	ت	ت		ت	ت	ت	ت	ت
13:24	ت	ت	ت	ه	ه	ت	ت	ه	ت	ه	ه	ت	ت	ت	ت	ت	ت	ت
13:31	ت	ت	ت	ت	ت	ه	ه	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت
13:33	ت	ت	ت	ه	ت	ه	ه	ت	ت	ه	ه	ت	ت	ت	ت	ت	ت	ت
13:44	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ه	ت	ت	ت	ت	ت	ت	ت
13:45	ت	ت	ت	ه	ت	ت	ت	ه	ت	ه	ت	ت	ت	ت	ت	ت	ت	ت
13:47	ت	ت	ت	ت	ت	ت	ت	ه	ت	ه	ت	ت	ت	ت	ت	ت	ت	ت
13:52	ت	ت	ت	ت	ت	ت	ت	ت	ت	ت	ه	ت		ت	ت	ت	ت	ت
16:19	ت	ت	ت	ه	ه	ه	ه	ه	ه	ت	ه	ت	ت	ت	ت	ت	ت	ت
18:1	ت	ت	ت	ه	ه	ت	ت	ه	ه	ه	ت	ت		ت	ت	ت	ت	ت
18:3	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ت	ت		ت	ت	ت	ت	ت
18:4	ت	ت	ت	ه				ه			ت			ت	ت	ت	ت	ت
18:23	ت	ت	ت	ه	ه	ت	ت	ه	ت	ه	ه	ت	ت	ت	ت	ت	ت	ت
19:12	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ت	ت	ت	ت	ت	ت	ت	ت

19:14	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ت	ت	ه	ت	_	ت	ت	ت
19:23	ت	ت	ت							ت			ت	ت	ت	ت	ت	ت
19:24	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ت	ت		ت	ت	ت	ت	ت
20:1	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ه	ت	ت	ت	ت	ت	ت	ت
21:31	ت	ت	ت	ه	ه	ه	ه	ت	ه	ه	ت	ه	ت	ت	ت	ت	ت	ت
21:43	ت	ت	ت	ه	ه	ه	ه	ه	ه	ه	ت	ت	ت	ت	ت	ت	ت	ت
22:2	ت	ت	ت	ه	ه	ه	ه	ه	ت	ه	ت	ت	ت	ت	ت	ت	ت	ت
23:13	ت	ت	ت	ت	ه	ه	ه	ت	ت	ه	ت	ت	ه	ت	ت	ت	ت	ت
25:1	ت	ت	ت	ه	ه	ه	ه	ت	ه	ه	ه	ت	ه	ت	ت	ت	ت	ت

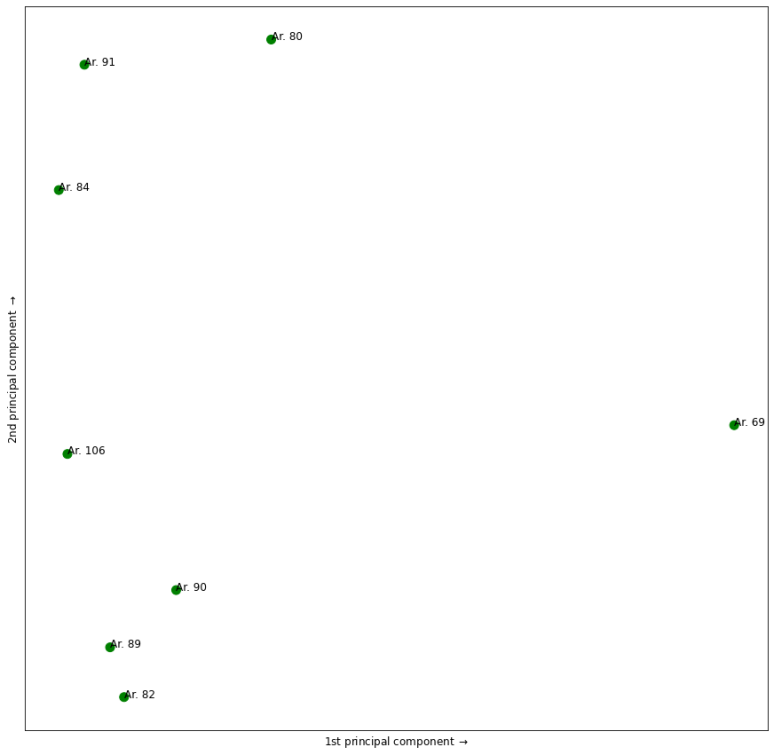
The most immediate observation arising from this examination is that Family J[b] patterns separately from all other families, with manuscripts regularly utilizing both ملكوه / ملكوة and ملكوت spellings of *malakūt*. At the same time, no two manuscripts in J[b] are identical in this respect, although SAr. 82 and 89 are differentiated only by the former's use of ملكوت in Matthew 4:17, where SAr. 89 uses ملكوة. This similarity is unsurprising, since the two manuscripts' colophons indicate that both are products of the same hand. The case of SAr. 69 is interesting, because it clearly belongs to J[b] in view of its idiosyncratic lexical and grammatical variation (Sections 3 and 4), yet it differs quite dramatically from the other J[b] manuscripts in the spelling of *malakūt*, patterning more closely with Family K in its preference for ملكوت over ملكوه / ملكوة (which occurs only four times in this manuscript, compared with at least nineteen instances in the other J[b] manuscripts). The sole J[c] manuscript included here, SAr. 76, is the only other manuscript in the study that uses ملكوة, and it does so only once (in Matt. 19:14). The third subgroup of Family J included here is Family J[a] (SAr. 115), and it attests only ملكوت spellings. Family K is consistent in its use of ملكوت spellings. There are, then, two main groups: Families J[b] and J[c], which use both spellings, and Families C, D, J[a], and K, which use only ملكوت.

The common use of the ملكوه / ملكوة spelling in Family J[b] and its concomitant absence (with one exception in SAr. 76) outside of this group demonstrate the distinct—and unique—nature of the scribal culture associated with Family J[b]. This observation correlates with the patterns of lexical and grammatical variation, where Families J[a] and J[c] again align with Family K (and in this case all others, too) against J[b]. At the same time, the spelling within the family is rather chaotic, with no manuscript replicating the pattern of another exactly. Thus we might ask whether the data do not in fact corroborate Kashouh's claim that orthographic variation is not useful for subgrouping manuscripts in the same family.[65] However, a PCA of the variation in these manuscripts turns up some visible clusters, shown in Figure 9.[66]

65. Ibid.

66. The first principal component captured 42% and the second principal component 22% of the variation.

Fig. 9. PCA of *malakūt* orthographic variation in Family J[b].



Several clusters are evident, including one consisting of SAr. 82, SAr. 89, and SAr. 90 and another with SAr. 84, SAr. 91, and SAr. 80. Another factor relevant to the interpretation of these visual results is that the x axis, which tracks the proximity or distance between any two manuscripts horizontally, plots the first principal component, which captures the greatest amount of variation between the manuscripts. In this case, all manuscripts except SAr. 69 are rather close to each other along the x axis; it is mainly the second principal component—tracked along the y axis with vertical space indicating the proximity or distance between any two manuscripts—that separates them. As we have seen in Sections 3 and 4, SAr. 69 is an outlier in the group. Although far from conclusive, this study of a single orthographic variant shows promise, and the topic warrants further study. It certainly casts doubt on Kashouh's claim that orthographic variants are not valuable for subgrouping manuscripts of the same family.[67]

The second area of orthographic variation, discussed briefly in Section 4, concerns the frequency and execution of *tanwīn* vocalizations. Once again, orthographic variation is the norm; no two manuscripts utilize the exact same combination of vocalization signs with the same frequency. However, here, too, there are patterns to the distribution. Four of the families studied here, J[a], J[b], J[c], and K, consist of manuscripts that are vocalized to one degree or another. The remarks below are therefore limited to manuscripts from these families. Further, I focus here on the means by which *tanwīn* is indicated orthographically; I make no

67. Kashouh, *Arabic Versions*, 215.

comments on morpho-syntax. Specifically, I do not consider the degree of overlap between Classical Arabic *tanwīn* and the *tanwīn* found in the manuscripts I study, nor am I grouping manuscripts on the basis of their treatment of any specific instance of tanwīn, since, as noted above, closely related manuscripts often diverge in particular instances (see the examples in Section 4). Instead, I focus on the combination of the three vocalizations (*ḍammatān* -ٌ / *kasratān* -ٍ / *fatḥatān* -ً) that each scribe utilized.[68] In this respect, the manuscripts in Families J^a, J^b, J^c, and K can be divided into three categories:

1. Use only *fatḥatān*/*tanwīn alif*
2. Use both *fatḥatān*/*tanwīn alif* and *kasratān*
3. Use all of the three vocalizations

Here, too, the data come from the Gospel of Matthew. Classifying each manuscript according to these three categories yields the groupings shown in Table 9:

Table 9. Orthographic categories of vocalized manuscripts according to *tanwīn*.

MS	Family	Category
SAr. 106	J^b	1
Vat. Borg. Ar. 71	J^b	1
SAr. 82	J^b	2
SAr. 84	J^b	2
SAr. 91	J^b	2
SAr. 69	J^b	2
SAr. 147	K	2
SAr. 89	J^b	3
SAr. 90	J^b	3
SAr. 80	J^b	3
SAr. 115	J^a	3
SAr. 76	J^c	3
SAr. 112	K	3
SAr. 628	K	3
SAr. 68	K	3
Vat. Copt. 9	K	3

68. For a discussion of the correlation between the orthography and (morpho-)syntactic categories in these manuscripts, see Stokes, "Nominal Case."

These category divisions do not fall neatly along manuscript family lines; however, we can still derive meaningful insights from this distribution. Only three of the nine manuscripts belonging to Family J^b included in this study utilize *ḍammatān*, and SAr. 89 uses it just once, while SAr. 90 attests it only three times. Within Family J^b, the use of only *fatḥatān/tanwīn alif* and *kasratān* is by far the dominant norm. By contrast, Families J^a, J^c, and K, with one counterexample in the latter family (SAr. 147), make use of all three forms. Interestingly, although the J^c manuscript uses *ḍammatān*, it is noticeably less frequent in this manuscript than it is in J^a and K manuscripts, occurring only nineteen times in SAr. 76's Gospel of Matthew. Still, orthographically Families J^a and J^c pattern much more closely with Family K than they do with Family J^b. Once again, although no two manuscripts are identical in terms of how and when they indicate tanwīn, there are clear patterns that point to a distinct scribal (sub)culture behind Family J^b.

6. Discussion

The data and results described above are significant in many ways for the study of the Christian Arabic Gospel manuscripts, as well as for the study of Christian and Middle Arabic more broadly. First, the data and analysis in Section 3 demonstrate the significance of lexical variation, especially idiosyncratic variation replicated across manuscripts, as a text-critical tool related to—but distinct from—simple lexical comparison in any one instance. Second, the results show that grammatical variation within and across manuscripts is not random. If grammatical variation were random, the result of idiosyncratic scribal behaviors, we should have found little meaningful clustering in the data. Instead, it appears that a text family was characterized by certain grammatical norms and patterns. Kashouh's statement that "there seems to be no effort in producing a *linguistic homogenous* text [*sic*]"[69] is thus technically true; however, the implication that "Arabic grammar and syntax were not the concern of the translators"[70] is contradicted by the patterns detected in this study. Although no two texts are grammatically identical, most pattern much more closely with related texts than they do with unrelated ones. This observation has potential text-critical implications that should be further explored. Third, and finally, the idiosyncratic orthographic variation points to different scribal cultures associated with the different text families, especially Family J^b. My findings generally corroborate Kashouh's intuition regarding the relative lack of importance of such variation for establishing relationships within manuscript families, but this topic requires much further work before such a principle should be broadly assumed.

These discoveries are meaningful and pave the way for future work to explore and define the nature of the patterns detected here. The clear distribution observed in the idiosyncratic lexical variation in the spelling of the phrases "Kingdom of God" and "Kingdom of heaven" in the Gospel of Matthew across twenty-two manuscripts provides conclusive evidence that the manuscripts classified as belonging to Family A, as well as those of Family J^b, were copied from single originals. The idiosyncrasies are copied too consistently for such replication to be attributable to chance. Relatedly, although scribes clearly felt free to alter and change

69. Kashouh, *Arabic Versions*, 7; emphasis in the original.

70. Ibid., 7.

lexical and grammatical aspects of their exemplars as style and preference dictated, they nevertheless faithfully copied many aspects of the exemplar. A focus on more banal lexical choices, and especially lexical variation, can help add nuance in ways that a focus on lexical similarities cannot.

The study of orthographic variation provides new insight into the scribal production of the Arabic Gospels. In Section 5, I demonstrated that the spelling of *malakūt* with a *tāʾ marbūṭah* was quite common in the scribal circles that produced the J^b manuscripts. That no two manuscripts replicate the same idiosyncratic distribution of those spellings confirms the absence of a norm and the freedom with which scribes alternated spellings, even disagreeing with the exemplar. Nevertheless, the fact that this spelling is found only in the manuscripts of J^b (with one exception in Family J^c) points to a distinct set of scribal practices. This finding correlates with the strong clustering of these manuscripts against the other families in terms of the lexical and grammatical variation studied here as well. Likewise, the orthographic variation that characterizes the representation of *tanwīn* supports the existence both of shared trends (notably writing *tanwīn* with *kasratān* or *fatḥatān* regardless of the context) and of different schools.

Perhaps most significant of all of my findings is the demonstration that the grammatical variation is in fact patterned. The present study provides one model for detecting patterns of grammatical variation in Middle Arabic corpora. When multiple versions of the same text or text tradition exist, comparing select variants across manuscripts can help identify where the main divisions fall. Of course, in-depth study of each manuscript within a group or family is also crucial. Every effort should be made to include as many variants as possible from as many grammatical categories as possible. When comparisons are made between texts that are not versions of the same original or related to one another by content, it is important to quantify grammatical trends by tagging all forms in a significant portion of each manuscript and making note of both form and context in order to facilitate comparison.

Since the goal of Section 4 was merely to determine whether the grammatical variation in the Arabic Gospel manuscripts included here attests to such patterns, attempting a typology of the various groups identified lies outside the scope of this paper. Nevertheless, some preliminary remarks can, I believe, further illustrate the significance of the present study.

Many of the features included here are morphological or morpho-syntactic in nature, including variation in III-W/Y roots, the use of *tanwīn*, nominal inflection of the "five nouns" in various contexts, and verbal mood. Of the vocalized manuscripts, those from Family J^b utilize *tanwīn* vocalizations least frequently, while J^a and J^c manuscripts as well as those belonging to Family K—especially SAr. 628 and Vat. Copt. 9—utilize them much more frequently. Families A and J^b share certain similarities, such as use of certain orthographic tools (*tanwīn alif* in Family A and the various *tanwīn* vocalizations in Family J^b) to mark a specific set of roles (e.g., nominal predicates, subjects of existential clauses, and words for either a single person or a group of people),[71] even when such marking is out of step with Classical Arabic grammatical and orthographic norms. These tools occur in Families

71. See Stokes, "Nominal Case."

J^a and K, too, but are less common than they are in A and K. The main exception is SAr. 80, which uses *tanwīn* vocalizations much more frequently than do other J^b manuscripts. This is likely a major reason SAr. 80 falls closer to Families J^a, J^c, and K than it does to the other manuscripts of Family J^b.

Other manifestations of case marking also attest to a spectrum. Family A regularly features the oblique in sound masculine plural nouns (cf. Matthew 21:42 above), whereas the other families, including J^b, display a more Classical distribution, with both nominative and oblique forms common. Words ending in *-*āʾ* (e.g., **ḥiḏāʾ*, "sandal," and **liqāʾ*, "meet"), which, when suffixed with a pronoun, will in Classical Arabic manifest a glide corresponding to the case vowel of the noun, appear with a Classical inflection in Families J^a, J^c, and K more frequently than they do in either A or J^b. Inflection of the "five nouns" (*al-ʾasmāʾ al-ḫamsah*) in the examples above is a mixed bag, with **ʾanna* triggering either the nominative or the accusative form in each of the families; however, in Families J^a, J^c, and K the accusative is more common than it is in Families A and J^b, for example.

At this juncture, it is worth stating explicitly that the non-Classical features that occur in each of the manuscripts studied here follow regular patterns and are thus clearly part of a register, or perhaps rather a spectrum of patterns, considered prestigious to one degree or another by each of the scribes. In some grammatical domains, such as the masculine plural, the register found in manuscripts belonging to Families J^a, J^c, J^b, and K is more similar to Classical Arabic than is the register visible in Family A. In terms of final glide inflection, for example, Families A and J^b tend to indicate nominal case with a glide less frequently than Families J^a, J^b, and K do. On the other hand, as noted above, case inflection of the "five nouns" varies. For example, the particle *ʾanna* triggered accusative and nominative inflection in members of Families A (SAr. 72) and J^b (SAr. 69 and 80). By contrast, although one manuscript of Family K (SAr. 112) shows nominative inflection in the same context, all others use the accusative, as do manuscripts of both J^a and J^c. Finally, orthographic representation of *tanwīn* in vocalized manuscripts corresponds closely with manuscript families, but all families, with the exception of Family H (Vat. Borg. Ar. 13), display the same patterns of non-Classical *tanwīn*.[72] Future work that examines the distribution of variants, such as historically nominative versus oblique dual and plural forms and the inflection of the five nouns and final-glide nouns in manuscripts from each family, will be crucial to understanding more precisely the patterns of variation.

7. Conclusion

This paper has investigated variation in three domains—lexical, grammatical, and orthographic—in order to determine the degree to which the variation attested in the domains of lexical patterns, grammar, and orthography across Arabic Gospel manuscripts is patterned in meaningful ways. In each domain, the variation is frequently idiosyncratic in any given manuscript. I have argued that it is precisely this idiosyncrasy that makes the patterns that emerge across manuscripts so compelling. Lexically, the idiosyncratic nature of the translation of the phrases "Kingdom of God" and "Kingdom of heaven" provides perhaps the strongest evidence yet that many of the manuscript families previously established reflect copies based on an exemplar that is, in this case at least, remarkably well preserved. Grammatical variation

72. For details of frequency and syntax, see Stokes, "Nominative Case."

patterns in clusters that by and large correlate with previously proposed manuscript families, and they thus offer strong support for the existence of distinct grammatical traditions in the various scribal circles that produced the manuscripts. Consequently, the various text types are characterized by distinct varieties (or registers). In turn, we should indeed think of multiple Christian Arabics, rather than a single "Christian Arabic." Orthographic variation is the most variable of the three types of variance investigated here, but even this variation corroborates the existence of distinct scribal cultures associated with major manuscript families. Most importantly, the present paper has demonstrated that the variation that characterizes the corpus, although often idiosyncratic, is neither patternless nor meaningless for understanding the corpus as a whole and the text-critical relationships between its manuscripts and manuscript families. Rather than background noise to be filtered out, variation turns out to be key to understanding the Arabic Gospels. Future work is needed to further map out these differences, as well as to expand the number of variants and the breadth of manuscripts and families included in the comparison.

Bibliography

Al-Jallad, A. *The Damascus Psalm Fragment: Middle Arabic and the Legacy of Old Ḥigāzī.* Chicago: Oriental Institute, 2020.

———. "The Earliest Stages of Arabic and Its Linguistic Classification." In *The Routledge Handbook of Arabic Linguistics*, edited by E. Benmamoun and R. Bassiouney, 315–31. New York: Routledge, 2018.

———. *An Outline of the Grammar of the Safaitic Inscriptions.* Leiden: Brill, 2015.

Arbache, S. *L'Évangile arabe selon saïnt Luc: Texte du VIII^e^ siècle, copié en 897, édition et traduction = al-Inǧīl al-ʿArabī bišārat al-qiddīs Lūqā.* Brussels: Éditions Safran, 2012.

———. "Une version arabe des évangiles: Langue, texte et lexique." PhD diss., Université Michel de Montaigne Bordeaux III, 1994.

Atiya, A. S. "Codex Arabicus (Sinai Arabic Ms. No. 514)." In *Homage to a Bookman: Essays on Manuscripts, Books and Printing Written for Hans P. Kraus on His 60th Birthday Oct. 12, 1967,* edited by H. Lehmann-Haupt, 75–85. Berlin: Mann, 1967.

Bellem, A., and G. R. Smith. "'Middle Arabic'? Morpho-syntactic Features of Clashing Grammars in a Thirteenth-Century Arabian Text." In *Languages of Southern Arabia: Papers from the Special Session of the Seminar for Arabian Studies Held on 27 July 2013, supplement to Proceedings of the Seminary for Arabian Studies* 44, edited by O. Elmaz and J. C. E. Watson, 9–17. Oxford: Archaeopress, 2014.

Bengtsson, P. *Two Arabic Versions of the Book of Ruth: Text Edition and Language Studies.* Lund: Lund University Press, 1995.

Blau, J. "The Beginnings of the Arabic Diglossia: A Study of the Origins of Neoarabic." *Afroasiatic Linguistics*, no. 4 (1977): 1–28.

———. *A Grammar of Christian Arabic: Based Mainly on South-Palestinian Texts from the First Millennium.* 3 vols. Louvain: Secrétariat du Corpus SCO, 1966–67.

———. *A Handbook of Early Middle Arabic*. Jerusalem: Hebrew University of Jerusalem, 2002.

———. *On Pseudo-Corrections in Some Semitic Languages.* Jerusalem: Israel Academy of Sciences and Humanities, 1970.

———. "Sind uns Reste arabischer Bibelübersetzungen aus vorislamischer Zeit erhalten geblieben?" *Le Muséon* 86 (1973): 67–72.

———. "Über einige christlich-arabische Manuskripte aus dem 9. und 10. Jahrhundert." *Le Muséon* 75 (1962): 101–8.

den Heijer, J. "Introduction: Middle and Mixed Arabic, A New Trend in Arabic Studies." In *Middle Arabic and Mixed Arabic: Diachrony and Synchrony*, edited by L. Zack and A. Schippers, 1–26. Leiden: Brill, 2012.

Fischer, W. *A Grammar of Classical Arabic.* Translated by Jonathan Rodgers. New Haven, CT: Yale University Press, 2002.

Graf, G. *Catalogue de manuscrits arabes chrétiens conservés au Caire.* Rome: Biblioteca Apostolica Vaticana, 1934.

———. *Geschichte der christlichen arabischen Literatur.* Vol. 1, *Die Übersetzungen.* Studi e testi 118. Rome: Biblioteca Apostolica Vaticana, 1944.

Grand'Henry, J. "Le moyen arabe dans les manuscrits de la version arabe du discours 40 de Grégoire de Nazianza." In *Moyen arabe et variétés mixtes de l'arabe à travers l'histoire: Actes du premier colloque international* (Louvain-la-Neuve, 10–14 mai 2004), edited by J. Lentin and J. Grand'Henry, 181–91. Louvain-la-Neuve: Peeters, 2008.

Griffith, S. *The Bible in Arabic: The Scriptures of the "People of the Book" in the Language of Islam.* Princeton, NJ: Princeton University Press, 2013.

———. "The Gospel in Arabic: An Inquiry into Its Appearance in the First Abbasid Century." *Oriens Christianus* 67 (1983): 126–67.

Guidi, I. *Le Traduzioni degli Evangelii in arabo e in etiopico. Atti della reale accademia dei Lincei*, ser. 4, vol. 4. Rome: Tipografia della reale accademia dei Lincei, 1888.

Hary, B. "Hypercorrection." In *Encyclopedia of Arabic Language and Linguistics*, edited by K. Versteegh, vol. 2, 275–79. Leiden: Brill, 2007.

Henkin, R. "Functional Codeswitching and Register in Educated Negev Arabic Interview Style." *Bulletin of the School of Oriental and African Studies* 79, no. 2 (2016): 279–304.

Hjälm, M. L. Christian *Arabic Versions of Daniel: A Comparative Study of Early MSS and Translation Techniques in MSS Sinai Ar. 1 and 2.* Leiden: Brill, 2016.

Kashouh, H. *The Arabic Versions of the Gospels: The Manuscripts and Their Families*. Berlin: de Gruyter, 2012.

Khan, G. "Middle Arabic." In *Semitic Languages: An International Handbook,* edited by S. Weninger et al., 817–35. Berlin: de Gruyter, 2011.

Kootstra, F. "A Quantitative Approach to Variation in Case Inflection in Arabic Documentary Papyri: The Case of ʾAb in Construct." *Journal of American Oriental Society* 142, no. 2 (2022): 387-414.

Lentin, J. "Middle Arabic." In *The Encyclopedia of Arabic Language and Linguistics*, edited by K. Versteegh, vol. 3, 215–24. Leiden: Brill, 2008.

Lentin, J., and J. Grand'Henry. *Introduction to Moyen arabe et variétés mixtes de l'arabe à travers l'histoire: Actes du premier colloque international* (Louvain-la-Neuve, 10–14 mai 2004), edited by J. Lentin and J. Grand'Henry, xvii - xxiii. Louvain-la-Neuve: Peeters, 2008.

Meiseles, G. "Educated Spoken Arabic and the Arabic Language Continuum." *Archivum Linguisticum* 11, no. 2 (1980): 118–43.

Monferrer-Sala, J. P. "Dos antiguas versiones neotestamentarias árabes surpalestinenses: Sin. ar. 72, Vat. ar. 13 y sus posibles Vorlagen respectivas greco-aljandrina y siriaca de la Pesitta." Ciudad de Dios 213, no. 2 (2000): 363–87.

Peters, C. "Proben eines bedeutsamen arabischen Evangelien-Textes." *Oriens Christianus* 11 (1936): 188–211.

Sadeghi, B. "The Chronology of the Qurʾān: A Stylometric Research Program." *Arabica* 58 (2011): 210–99.

Samir, S. K. "La tradition arabe chrétienne: État de la question, problèms et besoins." In *Actes du premier congrès international d'études arabes chrétiennes*, edited by K. Samir, 21–120. Orientalia christiana analecta 218. Rome: Pontificium Institutum Studiorum Orientalium, 1982.

———. "La version arabe des évangiles d'al-Asʿad Ibn al-ʿAssāl." *Parole de l'Orient* 19 (1994): 441–551.

Stokes, P. W. "A Fresh Analysis of the Origin and Diachronic Development of 'Dialectal Tanwīn' in Arabic." *Journal of the American Oriental Society* 104, no. 3 (2020): 637–64.

———. “In the Middle of What? A Fresh Analysis of the Language Attested in the Judaeo-Arabic Commentary on Pirqê ʾĀvōṯ (The Sayings of the Fathers), Middle Arabic, and Implications for the Study of Arabic Linguistic History.” *Journal of Semitic Studies* 66, no. 2 (2021): 379–411.

———. “Nominal Case in Christian Arabic Translations of the Gospels (9th–15th Centuries CE).” *Arabica*, forthcoming.

Valentin, J. “Les évangéliaires arabes de la bibliothèque du Monastère Ste-Catherine (Mont Sinai): Essai de classification d’après l’étude d’un chapitre (Matth. 28); Traducteurs, réviseurs, types textuels.” *Le Muséon* 116 (2003): 415–77.

———. “Des traces de la vetus syra des évangiles en traduction arabe? Étude critique des variantes significatives en Mc 5,1–20 dans le Sinaï arabe 80.” In *Reading the Gospel of Mark in the Twenty-First Century: Method and Meaning*, edited by G. Van Oyen, 765–79. Leuven: Peeters, 2019.

van Putten, M. “Arabe 334a: A Vocalized Kufic Quran in a Non-canonical Hijazi Reading.” *Journal of Islamic Manuscripts* 10 (2019): 327–75.

———. “‘The Grace of God’ as Evidence for a Written Uthmanic Archetype: The Importance of Shared Orthographic Idiosyncrasies.” *Bulletin of the School of Oriental and African Studies* 82, no. 2 (2019): 271–88.

———. *Quranic Arabic: From Its Hijazi Origins to Its Classical Reading Traditions*. Leiden: Brill, 2022.

van Putten, M., and H. Sidky. “Pronominal Variation in Arabic among the Grammarians, Quranic Reading Traditions, and Manuscripts.” In “Formal Models in the History of Arabic Grammatical and Linguistic Tradition,” edited by R. Villano, special issue, *Language and History* 65, no. 1 (forthcoming).

van Putten, M., and P. W. Stokes. “Case in the Qurʾānic Consonantal Text.” *Wiener Zeitschrift für die Kunde des Morgenlandes* 108 (2018): 143–79.

Vollandt, R. “The Status Quaestionis of Research on the Arabic Bible.” In *Studies in Semitic Linguistics and Manuscripts: A Liber Discipulorum in Honour of Professor Geoffrey Khan*, edited by N. Vidro, R. Vollandt, E.-M. Wagner, and J. Olszowy-Schlanger, 442–67. Uppsala: Uppsala University Press, 2018.

Vööbus, A. *Early Versions of the New Testament: Manuscript Studies*. Stockholm: Estonian Theological Society, 1954.

Wright, W. *Arabic Grammar*. 1896–98. Reprint, Mineola, NY: Dover, 2005.