Stylometric Perspectives on the Composition Debate of
Acts of Andrew

Sophie Robert-Hayek[1] ⬤, and Christian Houth Vrangbæk[2] ⬤

[1] Sorbonne University, STIH laboratory, Paris, France
[2] Department of Theology, Aarhus University, Aarhus, Denmark

Abstract

This paper engages with the ongoing scholarly debate surrounding the composition of the *Acts of Andrew*, particularly the question of whether the *Acts of Andrew and Matthew in the City of the Cannibals* (AAM) should be considered an original part of the text. While this issue has been extensively discussed, most prior studies have focused on selected excerpts rather than the complete corpus. Rather than offering definitive judgments or claiming methodological superiority, we apply computational methods to the full texts in order to revisit and enrich the debate. Through stylometric analysis, we reassess earlier claims in the literature, confirming, for instance, the overuse of periphrastic constructions and the prevalence of simple language in AAM, while offering nuanced support for others, such as the positioning of verbs and grammatical simplicity. We hope this study will demonstrate the potential of computational approaches to revisit humanities problems by treating a larger volume of data and systematizing previous scholarly claims made using a small data sample, due to the tediousness of manually counting word and grammatical occurrences.

Keywords: New Testament Apocrypha, Stylometry, Literary Composition

1 Introduction

Within the large, inhomogeneous text group known as the New Testament Apocrypha, we find a subgroup of Acts stories, i.e., narratives of the apostles' deeds, speeches, and deaths, among which the so-called apocryphal Acts of the Apostles [8]. These texts differ from the canonical acts and consist of stories of individual apostles, such as Paul, Peter, John, Andrew, or Thomas. Among literary works, the Acts of Andrew have come down to us in a complex textual state, leading to uncertainty regarding the relationship of the various textual materials to a possible original, or primitive, longer version of the Acts of Andrew [18]. The central concern in this paper is the relation of *Acts of Andrew and Matthew in the City of the Cannibals* (AAM) with the proper Acts of Andrew (AA) and possibly the AAM as an integral part of the AA.

The possible inclusion of AAM into an original larger Andrew cycle will have implications for the possible dating of the Andrew stories and it will affect the understanding of genre of apocryphal apostle stories to what extent they are more or less legendary in their style and content and to what extent they imitate canonical stories. All details of this huge discussion are outside the scope of this paper, in which we will revisit selected key points made in previous scholarship, however those implications provide the research context of the paper and the discussion of the matter.

The debate on the relation between AA and AAM began with Denis MacDonald, who proposed that the AAM, in accordance with Gregory of Tours' sixth-century epitome, was to be included in the AA [13]. This claim has been challenged by Hilhorst and Lalleman [5], David Warren [21],

and by Jean-Marc Prieur [16; 17]. We wish to take up this debate and, with the assistance of computational methods, reexamine the set of claims set forth on both sides of the debate. We regard this case as demonstrative for integrating computational methods into humanities scholarship of apocryphal literature by studying their style and content. Our findings contribute to the broader understanding of early Christian literature and demonstrate the potential of digital humanities approaches in resolving complex questions of textual criticism and composition in ancient religious texts.

The contributions of this paper to the state of the art are thus:

- The examination of three claims made regarding the relationship between the AAM and AA by leveraging stylometric approaches;

- More broadly, demonstrate the potential of computational approaches to contribute to well-known debates rooted in more traditional humanities scholarship.

The paper is structured as follows. Section 2 offers insights into the existing debates concerning the relationship between the Acts of Andrew (AA) and the Acts of Andrew and Matthew (AAM). Section 3 presents the selected texts used for the analysis, the performed pre-processing, and the sylometric features selected for discussing the existing claims regarding the relationship between the AAM and the AA. We then provide in section 3 the obtained results on the text, and provide in section 5 our conclusions regarding the relationship between the AAM and the AA, as well as insights regarding our further works on New Testament apocrypha.

2   The Relationship Between AA and AAM

2.1   Background of the Debate

The relationship between the Acts of Andrew (AA) and the Acts of Andrew and Matthew (AAM) has been debated since Flamion [4]. MacDonald argued that AA and AAM originally formed a single, continuous narrative, building on Lipsius' view that Gregory of Tours' epitome reflected the original Acts [9; 13], on the ground of shared literary and thematic connections between the texts, such as the motif of "Myrmidonia" and themes such as the sea, fishing, and miracles [10; 11; 12; 13]. Prieur, however, maintains that the two works are distinct, emphasizing differences in language, style, and the nature of miracles—AA being more complex and sober, AAM simpler and more prodigious [16]. Warren's stylometric analysis supports this, showing greater grammatical subordination and stylistic complexity in AA, and concluding the texts have different authors [21]. Zachariades-Holmberg also finds AA to be more literary and reminiscent of New Testament Greek, while AAM is simpler and less rich in vocabulary [22].
Hilhorst and Lalleman bridge the debate by noting both similarities and 11 significant differences in style and content, including themes, theological perspectives, and narrative motifs—arguing that AAM reflects a later, more Byzantine style [5].

In this paper, the investigation focuses on three quantitative arguments that advocate for a divergent stylistic approach between the AM and the AAM:

- The AAM has a simpler language than AA.

- There is a higher distribution of absolute genitives in AA than in AAM.

- The verb is more often positioned in the middle in AA than in AAM.

A summary of these different claims is available in table 3.

2.2 Research Question

Our paper is thus positioned within the field of computational stylometry, which applies statistical and computational tools to written texts in order to uncover underlying stylistic patterns, often for authorship attribution or compositional analysis in order to revisit established debates by moving from qualitative interpretations to quantitative, data-driven modeling of linguistic features such as function words or n-grams [3; 15; 19; 23]. While the previous quoted studies have made their stylometric claims based on selected passages due to the manual computation, our paper seeks to systematically test and generalize these findings in light of the ongoing debate about the composition of AA and AAM.

By systematically testing the different claims using quantitative metrics, we hope to **shed light on the possibility of the Acts of Andrew and Matthew being an integral part of the Acts of Andrew**, and highlight the potential of computational approaches to contribute to well-known debates rooted in humanities scholarship.

Stylometry and the question of authorship

3 Dataset and Methodology

3.1 Selected Texts and Pre-processing

Selected dataset The selected texts consist of 13 texts relating to what has generally been labeled as the Andrew cycle, taken from three text-critical editions: Prieur [**CCSA**], Bonnet [**Bonnet**], and Tischendorf [**Tischendorf**]. These editions include a variety of texts related to the stories of the Apostle Andrew. Even though some of the selected texts in the dataset exhibit minor differences, such as additional passions and martyrdoms, they have been included in our study to show the impact of selecting different critical texts on stylometric results, as well as to assert that our claims are not dependent on a selected edition. A summary of the selected texts is available in Table 2.

Abbreviated Title	Source	Chapters
MartAndrBonPrius	Bonnet	19
MartAndrBonAlt	Bonnet	11
MartAndrBonTert	Bollandiana	38
PAndr1	Bonnet/Tischendorf	15
PAndr2	Bonnet/Tischendorf	15
ActAndrBon	Bonnet	18
ActAndrPrieur	CCSA	50
MartAndrPrieurA	CCSA	13
MartAndrPrieurB	CCSA	19
AndrMattBon	Bonnet	33
AndrMattTisch	Tischendorf	33

Table 1: Texts used for the analysis. The titles of the individual texts follow the naming conventions used in the text-critical editions from which they have been retrieved [1; 2; 16; 20].

The editions of Tischendorf and Bonnet were retrieved from high-quality OCR scans via Google Books, which we then manually checked for any possible errors. Texts sourced from the *Corpus Christianorum Series Apocrypha* (abbreviated as CCSA in Table 2) have been copied manually, with prior consent. Since there is no dispute in scholarship that chapters 11-15 of AAM are a later addition [5; 13], we have extracted these chapters from AAM and excluded them from the analysis of AAM. However, we have kept the texts in the experiments so that we could validate that they are actually disjoint from the AAM.

Abbreviated Title	Source	Chapters
MartAndrBonPrius	Bonnet	19
MartAndrBonAlt	Bonnet	11
MartAndrBonTert	Bollandiana	38
PAndr1	Bonnet/Tischendorf	15
PAndr2	Bonnet/Tischendorf	15
ActAndrBon	Bonnet	18
ActAndrPrieur	CCSA	50
MartAndrPrieurA	CCSA	13
MartAndrPrieurB	CCSA	19
AndrMattBon	Bonnet	33
AndrMattTisch	Tischendorf	33

Table 2: Texts used for the analysis. The titles of the individual texts follow the naming conventions used in the text-critical editions from which they have been retrieved [1; 2; 16; 20].

Data processing

Data Parsing After the acquisition of the datasets and their manual cleaning to ensure high quality data, we perform automatic parsing using the pipeline for Ancient Greek GreCy [14], used through SpaCy's API [6]. We retrieve from this parsing the required information for computing the stylistic features: the lemmatized form of each word, the Part Of Speech (POS) of each word, as well as the morphological form of each available word. We then post-checked and corrected the resulting information partly via reading through the parsed files and partly via search queries to ensure that our statistics returned exact values, allowing us to compare our results with those from manual analysis. Afterwards we checked the parser again with random sampling.

3.2 Selected Stylometric Features

We divide our selected features into two sub-categories, classically used in stylometric analysis [19], and that have been used as arguments regarding the relationship between AAM and AA in traditional scholarship. A summary of the claims made by the different authors, along with our selected quantitative metrics for assessing these claims, is presented in Table 3.

Author	Claim	Metric
Prieur/Hilhorst & Lallemand [5; 16]	Simpler language in AAM than in AM	TTR and local TTR
Zachariades-Holmberg [22]	High distribution of abs genitive in AA than AAM	POS = Noun/Pron. + Participle: Morph = Gen.
Warren [21]	Verb in the middle position in AA than in AAM	Split sent. in 3, find finite verb position

Table 3: Scholars claims and selected metrics for validation

3.2.1 Vocabulary Richness Analysis

Vocabulary richness The first axis of our analysis assesses the *vocabulary richness* of the texts, in light of Prieur's observation that the language of the AA is much richer than that of the AAM, which he characterizes as employing a "simple" and "limited vocabulary" [17].
We verify this claim by relying on three widely used stylometric indicators: the *type-token ratio* (TTR), *sentence-level vocabulary richness*, and the *median sentence length* [7]:

1. *The type-token ratio (TTR)* is a widely used metric for lexical diversity. It is defined as the ratio of the number of distinct word forms (*types*), in our case in their lemmatized form, to the total number of word occurrences (*tokens*):

$$\text{TTR} = \frac{|\text{types}|}{|\text{tokens}|}$$

If Prieur's claim is right, then AAM should display a lower TTR.

2. *Sentence-level vocabulary richness* measures lexical diversity on a per-sentence basis as localized TTR, instead of being computed on the sentence as a whole as the standard TTR. For each sentence s_i, a localized TTR is calculated as:

$$\text{TTR}(s_i) = \frac{|\text{types}(s_i)|}{|\text{tokens}(s_i)|}$$

The mean of these values across all N sentences provides an aggregate indicator, computing the local variation in vocabulary use, and mitigating the bias linked to difference in sequence length.

3. *Median sentence length*: Median sentence length measure the median length of the sentences of the text, reflecting syntactic or rhetorical complexity, as longer sentences may reflect more elaborate structure.

Shared vocabulary

3.2.2 Sentence Construction Analysis

David H. Warren [21], Zachariades-Holmberg [22], and Hilhorst and Lalleman [5] argue for a lack of homogeneity between AAM and AM due to their different sentence structure. We leverage their different arguments by comparing the sentence structures in terms of subordination and the placement of verbs.

Grammatical subordination We rely on three measures to quantify grammatical subordination: the ratio of subordinating conjunctions, the ratio of subordinated verbs, and the use of absolute genitives.

1. *Ratio of subordinating conjunctions*: Subordinating conjunctions (e.g., ὅτι, ἵνα, ὡς …) introduce dependent clauses and are typically associated with more complex sentence structures. We compute their relative frequency as

$$\text{Sub. Conj. Ratio} = \frac{|\text{Subordinating Conjunctions}|}{|\text{All Conjunctions}|}$$

2. *Ratio of Subordinated Verbs*: This metric captures the extent of verbal subordination by calculating the proportion of finite verbs (filtered through the morphological tagging of SpaCy) that occur in subordinate clauses. Every finite verb embedded under subordinating conjunctions or relative pronouns is counted as subordinated.

$$\text{Sub. Verb Ratio} = \frac{|\text{Verbs in Subordinate Clauses}|}{|\text{All Finite Verbs}|}$$

3. *Frequency of genitive absolute*: The genitive absolute construction is composed of a noun or pronoun and a participle, both in the genitive case [22]. We compute its frequency per number of sentences:

$$\text{Genitive Absolute Rate} = \frac{|\text{Genitive Absolutes}|}{|\text{Sentences}|}$$

Verb positions

Following Warren's analysis, the verb position within the sentence serves as an indicator of the stylistic level in Ancient Greek prose. Skilled authors are typically associated with a preference for placing the (finite) verb in the middle of the sentence, rather than at the beginning or end. In contrast, less skilled or more colloquial styles tend to position the verb at the start of the sentence. While Warren examined seven chapters of AAM and ten chapters of AA, our computational approach allows us to treat a larger volume of data and consider the entirety of the book. To compute this score, sentences were divided into three equal parts, and we then identified in which part the finite verb occurred, defined as those tagged by SpaCy with a POS "AUX" or "VERB", as well as a VerbForm set to "Fin" in the morphological analysis. The number of verbs found in each part is then summed across each text and divided by the total number of verbs in the text to account for different lengths.

More formally, if we consider that a text consists of m sentences:

$$S^{(1)}, S^{(2)}, \ldots, S^{(m)}$$

with lengths $n_1, n_2, \ldots, n_m$, respectively.
Each sentence $S^{(j)} = [w_1^{(j)}, w_2^{(j)}, \ldots, w_{n_j}^{(j)}]$ is then divided into three equal parts (with rounding when n_j is not divisible by 3):

$$\text{Part 1:} \quad w_1^{(j)} \text{ to } w_{\lfloor n_j/3 \rfloor}^{(j)}$$

$$\text{Part 2:} \quad w_{\lfloor n_j/3 \rfloor + 1}^{(j)} \text{ to } w_{\lfloor 2n_j/3 \rfloor}^{(j)}$$

$$\text{Part 3:} \quad w_{\lfloor 2n_j/3 \rfloor + 1}^{(j)} \text{ to } w_{n_j}^{(j)}$$

We then define the set of finite verbs as:

$$FV = \{w \,|\, \texttt{pos}(w) \in \{\text{"AUX"}, \text{"VERB"}\} \text{ and } \texttt{VerbForm}(w) = \text{"Fin"}\}$$

If we denote $\{i_1, i_2, \ldots, i_K\}$ the sentence-relative positions (indices) of all finite verbs in the text, where $K = |FV|$ is the total number of finite verbs, then the finite verb position ratio for the text is computed as:

$$R_{\text{text}} = \frac{1}{K} \sum_{k=1}^{K} \frac{i_k}{n_{s(k)}}$$

where $s(k)$ denotes the index j of the sentence $S^{(j)}$ in which the k-th finite verb occurs, and $n_{s(k)}$ is the length of that sentence. This then leads to the average relative position of finite verbs within their respective sentences, that we multiply by 100 in order to obtain a percentage.

3.3   Implementation and Reproducibility

All code used in this study was written in Python, and the results are fully reproducible via the companion GitHub repository[1]. Due to copyright constraints, only the texts edited by Tischendorf and Bonnet are available on the GitHub repository. These datasets, to our knowledge, are the only freely available digital transcriptions of the Acts of Andrew and the Acts of Andrew and Matthew. We hope that these datasets will trigger further experimentation on these apocryphal texts.

---

[1] `https://github.com/Computing-Antiquity/acts_andrew_stylometry`

4 Results and discussion

We present the selected metrics for investigating and systematizing the stylometric claims in the following section. We summarize our results and their relationship to existing claims in table 4.2.1.

4.1 Vocabulary Analysis

Richness of vocabulary

The results regarding the values of the TTR, the Sentence Vocabulary Richness, and the Median sentences are displayed in Table 4. As claimed by Prieur, Hillhorst, and Lallemand, AAM presents a lower type-token ratio in the AAM texts, corroborating their claims. All of our results thus single out the AAM as being simpler in language. The AAM-editions, which are very close in their wording, have lower TTR-scores compared to the other texts.

	TTR	Sentence Vocabulary Richness	Median Sentence Length
AndrMattBon	0.33	0.88	11.00
AndrMattTisch1115	0.34	0.83	14.00
AndrMattTisch	0.34	0.83	17.00
AndrMattBon1115	0.34	0.87	10.00
MartAndrBonTert	0.40	0.86	16.00
PAndr1	0.41	0.90	12.00
MartAndrBonPrius	0.42	0.88	9.00
ActAndrPrieur	0.42	0.89	11.00
ActAndrBon	0.42	0.92	9.00
PAndr2	0.43	0.85	17.00
MartAndrPrieurB	0.46	0.88	11.00
MartAndrBonAlt	0.47	0.90	10.00
MartAndrPrieurA	0.48	0.89	11.00

Table 4: Type-Token Ratio, Sentence Vocabulary Richness, and Median Sentence Length for All Texts. The table is sorted by the TTR, and the metrics concerning the AAM are highlighted in grey.

In terms of vocabulary richness per sentence, we can observe a lesser difference between the different claims, with Tischendorf's version still presenting the lowest score.
The results of the median sentence length show a more complex image. Here, the *martyrium prius* and Bonnet's *acta* have the lowest median sentence length, whereas AAM from the Tischendorf version scores highest. These results should be correlated; for example, when AAM has a low TTR but a high median sentence length, it means that there are rather long sentences, but with few new terms, i.e., a simple and perhaps monotonous text.

4.2 Sentence Analyses

4.2.1 Sentence construction analysis

The first experiment examines the level of grammatical subordination through three measures: the ratio of subordinating conjunctions, the ratio of subordinated verbs, and the use of absolute genitives.
Grammatical subordination Table 4.1 presents the ratio of coordinating and subordinating conjunctions. The ratio between coordinating and subordinating conjunctions reveals that the

AAM texts, while exhibiting coordinating conjunction ratios within the 5–7% range, are not unique in this respect; several other texts display similar values. Notably, MartAndrBonTert has the highest ratio of coordinating conjunctions at 7.36. The texts that stand out most in these results are PAndr1 and PAndr2, both of which are characterized by relatively high ratios of subordinating conjunctions (2.17 and 2.06, respectively) and low ratios of coordinating conjunctions (4.23 and 3.89). This indicates a higher degree of *hypotaxis* and a lower degree of *parataxis* in these texts. Overall, the use of subordinating conjunctions does not distinguish the authors of the Acts and Martyrdoms of Andrew from those of the AAM texts, as their ratios are broadly comparable, invalidating Zachariades' claims regarding the higher rate of subordinating conjunctions.

	Ratio of subordinate conjunctions	Ratio of conjunctions
AndrMattTisch1115	2.27	6.54
AndrMattBon1115	2.15	6.80
PAndr1	2.17	4.23
PAndr2	2.06	3.89
AndrMattBon	2.00	6.82
AndrMattTisch	1.95	7.19
ActAndrBon	1.75	5.01
MartAndrBonAlt	1.66	5.55
ActAndrPrieur	1.65	5.32
MartAndrPrieurA	1.46	6.52
MartAndrPrieurB	1.21	6.51
MartAndrBonPrius	1.16	6.61
MartAndrBonTert	0.62	7.36

Table 5: Ratio of Coordinating Conjunctions and Subordinating Conjunctions. Data relative to the AAM is highlighted in gray

Subordinate and Non-Subordinate Verbs

Analysis of subordinate and main verb ratios reveals clear syntactic differences across the texts, as displayed in table 5.

	Subordinate Verbs Ratio (%)	Main Verbs Ratio (%)
PAndr1	28.95	28.57
PAndr2	25.93	24.20
ActAndrBon	24.80	22.93
ActAndrPrieur	23.63	28.00
AndrMattTisch1115	22.94	56.88
AndrMattBon1115	21.95	50.41
AndrMattBon	21.34	56.10
AndrMattTisch	20.46	70.10
MartAndrPrieurA	20.39	31.96
MartAndrBonAlt	17.86	27.86
MartAndrBonTert	15.85	37.72
MartAndrBonPrius	13.87	39.05
MartAndrPrieurB	13.33	35.87

Table 6: Ratios of Subordinate and Main Verbs per Text. Data relative to the AAM is highlighted in gray.

PAndr1, PAndr2, ActAndrBon, and ActAndrPrieur have the highest use of subordinate verbs and lower main verb ratios, indicating a preference for complex, hypotactic constructions. In contrast, martyrdom texts like MartAndrBonPrius, MartAndrPrieurB, and MartAndrBonTert show lower subordination and higher main verb ratios, reflecting a simpler, paratactic style. The AAM texts fall between, suggesting a moderate level of syntactic complexity.

These results must be seen together with the following, namely the ratio of absolute genitives which would give a more complete picture of grammatical subordination [21].

Absolute Genitives

The distribution of absolute genitives is displayed in Table 6 across the texts provides further insight into syntactic preferences and literary style. The results show that the later MartAndrBonTert stands out with the highest ratio of absolute genitives (6.70), followed by PAndr2 (3.70), MartAndrBonPrius (3.65), ActAndrPrieur (3.56), and MartAndrBonAlt (3.57). Most other texts, including the AAM versions, display lower ratios, typically between 1.8 and 3.3, however the distribution is does not indicate a clear overuse of absolute genitives in Prieur's versions which Zachariades-Holmberg used.

	Absolute genitive count	ratio genitive count
MartAndrBonTert	30	6.696
PAndr2	15	3.704
MartAndrBonPrius	10	3.650
MartAndrBonAlt	5	3.571
ActAndrPrieur	35	3.564
MartAndrPrieurB	10	3.175
PAndr1	7	2.632
AndrMattTisch	18	2.575
AndrMattBon	21	2.561
MartAndrPrieurA	9	2.479
ActAndrBon	8	2.133

Table 7: Number and ratio of detected absolute genitives.

The results regarding verb positions, available in Table 4.2.1, corroborate Warren's findings with some nuances.

	First part (%)	Middle part (%)	Last part (%)
AndrMattBon1115	46.81	32.62	20.57
AndrMattBon	45.60	33.58	20.83
AndrMattTisch	44.04	34.52	21.44
AndrMattTisch1115	44.28	35.88	19.85
MartAndrPrieurB	40.15	31.02	28.83
MartAndrBonPrius	39.69	28.79	31.52
MartAndrBonAlt	36.30	34.07	29.63
ActAndrPrieur	35.14	31.65	33.21
ActAndrBon	34.95	34.30	30.74
MartAndrPrieurA	34.30	32.37	33.33
MartAndrBonTert	33.17	30.92	35.91
PAndr2	32.62	30.24	37.14
PAndr1	31.19	33.90	34.92

Table 8: Normalized verb position distribution by sentence part. Results concerning the AAM texts are displayed as first and labeled in gray. Sorted according to First part position

The AAM texts (e.g., *AndrMattBon*, *AndrMattTisch*, and their variants) consistently show a higher proportion of verbs in the first part of the sentence (normalized values around 44–47%), compared to the middle (about 33–36%) and the last part (about 20–21%). Warren had found 49% , 40% and 11% for the first, middle and last position respectively for the AAM. For the AA, Warren found 23%, 43% and 34% respectively for the AA. However, Warren's investigation of AA was limited to ten chapters of Prieur's *Acta*, in which we also found a marked reduction in initial verb placement and an increase in medial and final positions. Warren's investigation of AAM was limited to chapters 1-7. In summary, the AAM texts' tendency to place verbs at the beginning of the sentence aligns to a large extent with Warren's findings, however, our results are more evenly distributed in the placement of the verb.

Ratio of other grammatical features
Ἐμί + participle (aor. or perf.)

Author	Claim	Confirmed	Confirmed w. Nuance	Unconfirmed
Warren	Verb in the Middle Position in AA		X	
Hilhorst & Lallemand	Grammatical Simplicity of AAM		X	
Hilhorst & Lallemand	Overuse of Periphrastic Construction in AAM	X		
Prieur/Hilhorst & Lallemand	Simple language in AAM	X		
Zachariades-Holmberg	High Distribution of Absolute Genitive in AA	X		

Table 9: Summary of results and relationship between scholarly claims

5   Conclusion and Further Works

Our results show that we on many points can corroborate the stylometric claims that have been conducted manually in previous scholarship, for example, concerning a lower vocabulary richness, a more paratactic sentence structure, and a high use of periphrastic constructions for the AAM, which would support the claim that AAM and AA do not belong to a common original. However, we could also detect important nuances: First, the placement of the verb is more evenly

distributed than previously shown; second, the use of hypotaxis was not in the low end for the AAM, for which reason we cannot deem it completely "simple" in this regard.

Our experiments so far have taken the texts in their present editions; however, proponents of the relationship between AAM and AA suggest a kinship based more on thematic lines that follow a now lost, primitive version of the text. In our future work, we will pursue such an investigation, attempting to analyze hidden literary layers with more topical and content-oriented analyses.

Acknowledgments

References

[1] Bonnet, M. *Acta Apostolorum Apocrypha*. Vol. 1-2. Hildesheim: G. Olms, 1972.

[2] Bonnet, M. "Martyrium sancti apostoli Andreae". In: *Analecta Bollandiana 13* (1894), pp. 354–372.

[3] Deneire, T. "Filelfo, Cicero and Epistolary Style: A Computational Study". In: *Francesco Filelfo, Man of Letters*. Brill, 2018, pp. 239–270.

[4] Flamion, J. *Les Actes Apocryphes de l'Apôtre André : Les Actes d'André et de Mathias, de Pierre et d'André et Les Textes Apparentés*. Bureaux du Recueil, 1911.

[5] Hilhorst, A. and Lalleman, P.J. "The Acts of Andrew and Matthias: Is It Part of the Original Acts of Andrew?" In: *The Apocryphal Acts of Andrew*. Vol. 5. Studies on the Apocryphal Acts of the Apostles. Peeters, 2000, pp. 1–14.

[6] Honnibal, M., Montani, I., Van Landeghem, S., and Boyd, Adriane. "spaCy: Industrial-strength Natural Language Processing in Python". In: *Online* (2020).

[7] Kettunen, K. "Can Type-Token Ratio be Used to Show Morphological Complexity of Languages?" In: *Journal of Quantitative Linguistics 21*, no. 3 (2014), pp. 223–245.

[8] Klauck, H. J. *The Apocryphal Acts of the Apostles: An Introduction*. Baylor University Press, 2008.

[9] Lipsius, R.A. *Die apokryphen Apostelgeschichten und Apostellegenden: Ein Beitrag zur altchristlichen Literaturgeschichte*. C.A. Schwetschke, 1883.

[10] MacDonald, D. R. *Christianizing Homer: The Odyssey, Plato, and the Acts of Andrew*. New York: Oxford University Press, 1994.

[11] MacDonald, D. R. "Response: [To: MacDonald, Dennis R., The Acts of Andrew and Matthias and the Acts of Andrew]". In: *The Apocryphal Acts of Apostles*, ed. by Dennis Ronald MacDonald. Vol. 38. Semeia. Decatur: Society of Biblical Literature, 1986, pp. 35–39.

[12] MacDonald, D. R. *The Acts of Andrew and the Acts of Andrew and Matthias in the City of the Cannibals*. Vol. 33. SBLTT. Atlanta: Scholars Press, 1990.

[13] MacDonald, Dennis Ronald. "The "Acts of Andrew and Matthias" and the "Acts of Andrew"". In: *The Apocryphal Acts of Apostles*, ed. by Dennis Ronald MacDonald. Vol. 38. Semeia. Decatur: Society of Biblical Literature, 1986, pp. 9–26.

[14] Myerston, J. and López, J. "greCy: Ancient Greek spaCy models for Natural Language Processing in Python". Version 1.0. URL: `https://github.com/jmyerston/greCy`.

[15] N., Kristoffer L., Perner, M. L., Larsen, C., Nielsen, J., and Laursen, Ditte. "Automated Compositional Change Detection in Saxo Grammaticus' Gesta Danorum". In: *HAL* (2019). URL: `https://hal.science/hal-02084682/document`.

[16] Prieur, J.-M. *Acta Andreae*. Vol. 6. Corpus Christianorum. Series Apocryphorum. Turnhout: Brepols, 1989.

[17] Prieur, J.M. "Response: [To: MacDonald, Dennis R., The Acts of Andrew and Matthias and the Acts of Andrew]". In: *The Apocryphal Acts of Apostles*, ed. by D. R. MacDonald. Vol. 38. Semeia. Decatur: Society of Biblical Literature, 1986, pp. 27–33.

[18] Roig Lanzillotta, L. "The Acts of Andrew". In: *The Oxford Encyclopedia of the Books of the Bible*, ed. by Michael D. Coogan. New York: Oxford University Press, 2011, pp. 34–39.

[19] Savoy, J. *Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling*. Springer Nature, 2020.

[20]   Tischendorf, C. *Acta apostolorum apocrypha. Ex triginta antiquis codicibus graecis vel nunc primum eruit vel secundum atque emendatius edidit Constantinus*. Leipzig: Avenarius et Mendelssohn, 1851.

[21]   Warren, D. H. "The Greek Language of the Apocryphal Acts of the Apostles: A Study in Style". In: *The Apocryphal Acts of the Apostles: Harvard Divinity School Studies*. Religions of the World. Cambridge, Mass.: Harvard University Press, 1999, pp. 101–124.

[22]   Zachariades-Holmberg, E. M. "Philological Aspects of the Apocryphal Acts of the Apostles". In: *The Apocryphal Acts of the Apostles*. Harvard Divinity School Studies. Religions of the World. Harvard University Press, 1999, pp. 125–142.

[23]   Zhang, O. R., Cohen, T., and McGill, S. "Did Gaius Julius Caesar Write De Bello Hispaniensi? A Computational Study of Latin Classics Authorship". In: *Human IT: Journal for Information Technology Studies as a Human Science 14*, no. 1 (2018), pp. 28–58.