

Scalable Syriac Paleography using Interactive Visualization

R. Jordan Crouser*
Smith College

Michael Penn
Stanford University

Nicholas R. Howe
Smith College

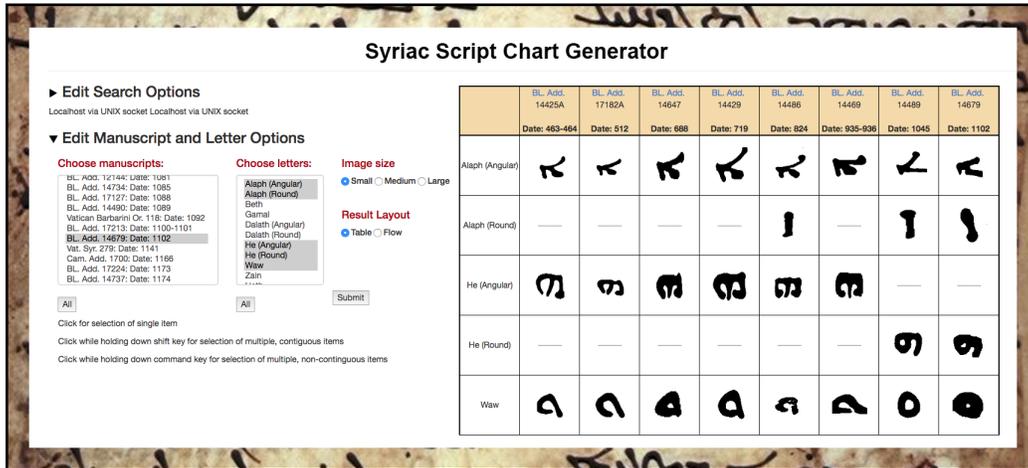


Figure 1: The Syriac Script Chart Generator enables historians and paleographers to generate script charts on demand for any subset of manuscripts drawn from the world’s largest repository of securely-dated Syriac manuscripts. Excerpts from manuscript images © British Library Board.

ABSTRACT

Syriac (a dialect of Aramaic) was the primary language spoken in the late ancient Middle East between the second and eighth centuries AD, and continues to be a language of Christian scholarship and liturgy up to the present day. There are approximately 20,000 known surviving Syriac manuscripts. Among early manuscripts, only around 10% include a scribal note that provides information regarding when, where, and by whom a given manuscript was written. For the remaining 90%, close examination of subtle differences in the handwritten script remains the primary tool for determining provenance. Prior to this study, scholars classified early Syriac manuscripts into two divergent script styles: *Estrangela* and *Serto*. In this paper, we present a case study of historians’ analysis of this collection of manuscripts supported by visual analytic tools. This approach uncovered major inaccuracies in this dichotomous model, resulting in profound disruption to the dominant understanding of the development of these texts.

Index Terms: H.5.2 [User Interfaces]: User Interfaces—Graphical user interfaces (GUI); H.5.m [Information Interfaces and Presentation]: Miscellaneous

1 INTRODUCTION

Paleography is the study and description of ancient and medieval manuscripts, documents, and systems of writing [1]. Since the field’s inception in the 17th century, this discipline has conducted a painstaking accounting regarding the various characters used at different periods by the scribes of different nations and languages. Capitalizing on the observation that no two scribes write in exactly the same way, early paleographers used subtle clues in letter forms

and preferred abbreviations to distinguish between scribes and to identify forgeries [13]. Paleography remained a limited field until the advent of photographic reproduction, at which point inscriptions and manuscripts could be made more readily accessible to scholars and asynchronous collaboration became possible.

A pivotal development in the understanding of the historical value of handwriting came in the 19th century, when scholars first connected the study of handwriting to the study of text transmission [7]. This advancement in the field allowed for better dating and geographical locating of a script, and also paved the way for future paleographical studies [2]. By 1912, the Latin written language could be classified into more than 25 different scripts [19], and in modern paleography metadata regarding the provenance of newly discovered manuscripts is uncovered by comparing them to repositories of securely dated and located manuscripts.

As part of an ongoing project on computer-assisted paleography entitled the Digital Paleography Project, a digital humanities team from Stanford University and Smith College applied computationally-supported handwriting analysis to extract a database of over 50,000 individual letters from the majority of ancient Syriac manuscripts in existence. The dataset includes 170 of the 182 known “securely dated” Syriac manuscripts written prior to 1100 AD, which include a scribal note detailing precisely when they were written. In addition to providing more fine-grained information about individual manuscripts, this project has led to new questions and new fields of inquiry when combined with traditional tools of historical and literary analysis.

1.1 Contributions

This paper presents an overview of a collaboration between visual analytics researchers and scholars of history and early Christianity. Specifically, we describe the application of several standard interactive visualization techniques to the world’s largest collection of securely dated Syriac manuscripts and individual letters. These visualizations illustrate that the actual development of Syriac script differs substantially from what the standard model predicts. An anal-

*e-mail: jrouser@smith.edu

ysis of the letter forms found in securely dated manuscripts using these visualizations enabled historians, for the first time, to accurately describe the chronological development of the Syriac script. This enables future scholars to better estimate the date and provenance of manuscripts that do not have a scribal note, and represents a profound breakthrough in the study of early Syriac texts.

2 BACKGROUND

Between the second and eighth centuries AD, the primary language spoken in the late ancient Middle East was a dialect of Aramaic called Syriac. Although gradually replaced by Arabic as a living language in the Middle East, Syriac continues to be a language of Christian scholarship and liturgy up to the present day. Syriac manuscripts have particular historical significance: because Syriac churches constituted the most widespread branch of ancient Christianity, spreading rapidly throughout the Middle East and into Asia, India, Tibet, and China, they developed their own theological, ecclesiastical, and monastic traditions. They were the first Christians to encounter Islam, and were a crucial link in the translation and preservation of the writings of Aristotle. Moreover, they formed an essential cultural bridge between Asia, the Middle East, and Europe.

As with the analysis of other manuscript traditions, the key to analyzing a Syriac manuscript is its handwriting. There are currently thought to be about 20,000 extant Syriac manuscripts. The majority of the earliest manuscripts are now housed in European libraries that provide images of individual pages. This has facilitated the rapid assembly of a database that includes handwriting samples from a substantial portion of the earliest extant Syriac texts. However, only about 10% of ancient manuscripts include specific information regarding their provenance. Despite their scholarly significance, many manuscripts lack the date and location information needed to interpret them within the larger historical context.

2.1 Estrangela and Serto

Introductory textbooks describe written Syriac as consisting of two divergent script styles. The first style (*Estrangela*, see Fig. 2-top) is attested in the earliest securely dated Syriac manuscript dated 411 AD and continues up to the present day. A second style (*Serto*, see Fig. 2-bottom) emerged in the eighth century and eventually became the dominant style of later manuscripts. Under this classification, six of the twenty-two Syriac letters (*alaph*, *he*, *waw*, *rish*, *dalath*, and *taw*) have distinct *Estrangela* and *Serto* forms. In many cases, the *Estrangela* form (abbreviated *E*) looks quite different from the *Serto* form (abbreviated *S*). For example, the *alaph* consists of three different strokes in the *E* form, but is reduced to a single vertical line in the *S* form. Other letters differ from being quite angular in the *E* form but much more rounded in the *S* form (Table 1).

According to this model, a manuscript will be either *Estrangela* script and thus consist only of *E* forms, or *Serto* script and thus consist only of *S* forms. This model asserts that prior to the eighth century, all manuscripts were written in the *Estrangela* script. Starting in the eighth century, some manuscripts continue to have only *E* forms, but these are quickly outnumbered by *Serto* manuscripts which have only *S* forms. This standard model is easy to explain, unambiguous, and dominates many prominent works in Syriac studies (e.g. [3, 8, 18]). Tables included in reference books, journal articles, text books on Syriac (e.g. [12, 14, 16, 17]) are virtually unanimous in how they define these two scripts. Additionally, the reader may note that William Henry Paine Hatch, *An Album of Dated Syriac Manuscripts* (Boston: The American Academy of Arts and Sciences, 1946), the only published album of securely dated Syriac manuscripts is organized into distinct *Estrangela* and *Serto* categories, and most manuscript catalogs use similar nomenclature. Indeed, this model works well for printed text and manuscripts written after the 13th century. It suffers from only one drawback: when applied to early Syriac manuscripts, **it is almost entirely wrong** [5].

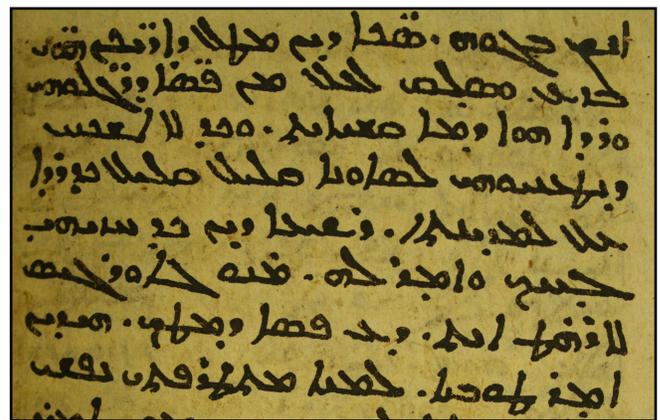
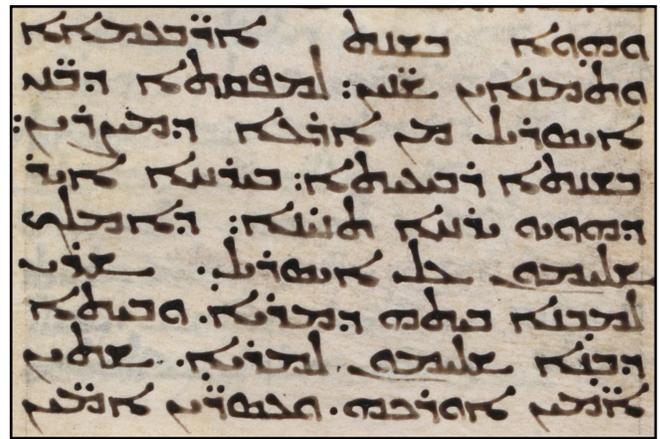


Figure 2: (top) A sample from *British Library Additional 14,430*, an example of an *Estrangela* manuscript written in the year 724 AD. (bottom) A sample from *British Library Additional 14,734*, an example of a *Serto* manuscript written in the year 1085 AD. Images © British Library Board.

Form	<i>alaph</i>	<i>dalath</i>	<i>he</i>	<i>rish</i>	<i>taw</i>	<i>waw</i>
<i>Estrangela</i>	Ⲁ	ⲁ	Ⲃ	ⲃ	Ⲅ	ⲅ
<i>Serto</i>	Ⲁ	ⲁ	Ⲃ	ⲃ	Ⲅ	ⲅ

Table 1: A sample of Syriac characters with markedly different *Estrangela* and *Serto* forms.

Historians have noted various inadequacies of this standard model [4, 6, 9, 15]. However, because these scholars did not have access to a large dataset of Syriac letters, their works were unable to detail or to quantify the extent to which this model failed, or if there were any patterns among the manuscripts it failed to properly classify. Falling short of providing a quantifiable explanation of the model’s weaknesses, this research has not yet influenced how the majority of students learn Syriac script or how the majority of specialists describe it. In the following sections, we present a brief case study detailing an ongoing collaboration in data science and the digital humanities the goal of which is to address this issue. We focus in particular on the design and use of several interactive visualization tools, which enabled historians to simultaneously examine all 170 securely dated manuscripts for the first time.

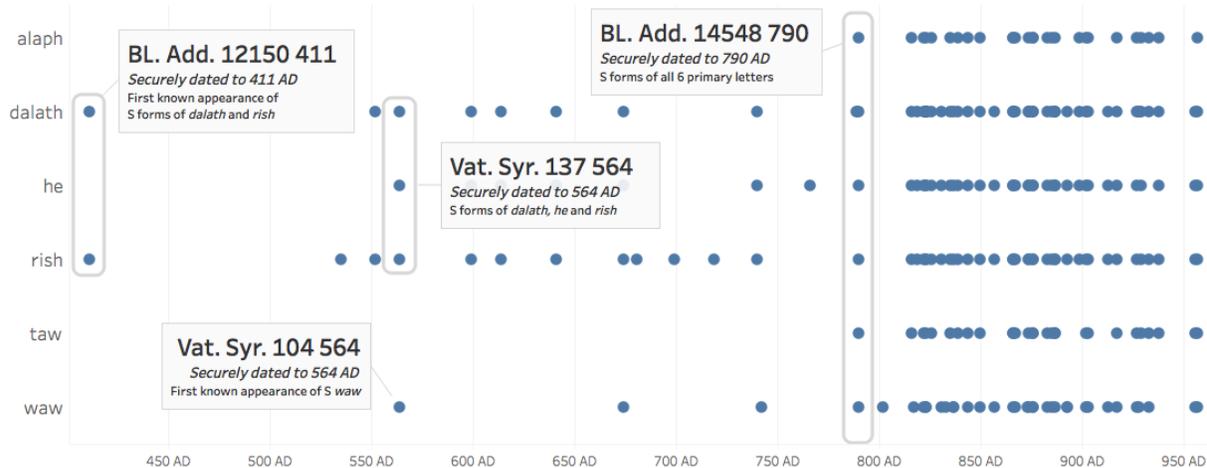


Figure 3: Through the use of an interactive timeline, historians were for the first time able to observe the 3-phased evolution of the *Serto* script: 411 AD (*dalath*, *rish*), 564 AD (*he*, *waw*), and 790 AD (*alaph*, *taw*). Each manuscript is represented as stacked dots along a timeline, indicating the presence of the *Serto* form of each letter on the *y*-axis.

3 SCRIPT CHARTS

One of the most common tasks a paleographer faces is comparing the letter forms of a manuscript of interest with a set of other manuscripts. When a paleographer compares manuscripts, they rarely do so at the level of complete manuscript pages. Instead, they want to compare at the level of individual letters. The most common tool for performing this fine-grained analysis is the **script chart**, a table which contains one or more examples of each letter form from a series of manuscripts (see Fig. 1 *right*). For Syriac, as for most ancient languages, script charts have historically been generated by hand. That is, a scholar will select a handful of manuscripts, trace a few examples of each letter form, and then publish the handwritten chart. Not only are such charts extremely time-intensive to generate, but they also provide only a very small, biased sample of the available data which cannot be filtered or manipulated.

3.1 The Syriac Script Chart Generator

This project re-conceptualizes the script chart so that it now shows images directly from actual manuscripts, driven by a nearly all-inclusive set of securely dated works. Using intuitive menus and radio buttons to select manuscripts and letters of interest, a scholar can for the first time construct script charts on demand to suit their particular needs (see Fig. 1). These interactively generated charts enable the reader to compare a given letter from the manuscript they are most interested in with the same letters in other manuscripts in order to find those that have the most similar form. At a broader level, the ability to generate script charts with a large number of securely dated manuscripts will enable a scholar to track how a given letter form changes over time. Discovering these subtle patterns not only assists in the analysis of a single manuscript, but also highlights larger trends that can help localize other manuscripts.

4 CASE STUDY: HISTORICAL MYTHBUSTING USING INTERACTIVE DATA VISUALIZATION

Interactive visualization can be a powerful tool for the exploration of literary collections [10, 11]. In this section, we include brief discussions of three primary discoveries directly enabled by the use of interactive visualization on this dataset. These questions arose based on discrepancies observed within the interactive script chart, and were further investigated using standard visualization techniques on the full dataset.

Myth 1: The *S* form first appeared in the 8th Century.

According to the standard model, the *S* form first appeared during the 8th century as a more cursive alternative to the earlier *E* forms. Charts in textbooks suggest that this emergence was a relatively sudden event, and these *S* forms eventually surpassed their predecessors in popularity. However, when we consider the complete collection of securely dated manuscripts together through the use of an interactive timeline (see Fig. 3), the alleged eighth-century birth of the *S* form becomes problematic. Although *British Library Additional 14548 790* was the earliest example of a manuscript that utilized exclusively *S* forms of all six primary letters (*alaph*, *dalath*, *rish*, *he*, *taw*, and *waw*), other manuscripts suggest that these letter forms developed separately and at different times, often well before the eighth century.

Consider for example the manuscript *British Library Additional 12,150*, which contains *S* forms of *dalath* and *rish* in 411 AD. *He* and *waw* were the next letters to appear in their *S* form among securely dated manuscripts, with an earliest appearance being in *Vatican Syr. 137*, securely dated to 564 AD. The *S he* is found in 5 of the 62 securely dated manuscripts produced before 700 AD. In total, 16% of securely dated manuscripts (10 out of 62) produced before 700 AD contain at least one *S* form of the letters *dalath*, *rish*, *he*, and *waw*. In contrast the *S* forms of *alaph* and *taw* in the main text of a securely dated manuscript does not occur until *British Library Additional 14,548*, two centuries after the first appearance of the *S* forms of *dalath*, *rish*, *he*, and *waw*. The *S alaph* and *taw* do not become commonly used until the mid-ninth century.

The standard model obscures a much more complicated history to Syriac scripts. Among the main text of securely dated manuscripts, there are three distinct points in which different *S* letter forms first appear: 411 AD (*dalath*, *rish*), 564 AD (*he*, and *waw*), and 790 AD (*alaph*, *taw*). Rather than an abrupt development of the *Serto* script in the eighth century, the letter forms that make up the standard model's definition of this script appear to have evolved over the span of nearly four centuries.

Myth 2: Manuscripts are written in either *E* or *S*.

According to the standard model, a given manuscript should have either *E* forms of all the letters or *S* forms of all the letters. That is, if the manuscript has an *E alaph*, so too should *rish*, *dalath*, *he*, and *taw* all appear in their *E* forms. Conversely, if a manuscript has an *S* form of *alaph*, so follows *rish*, *dalath*, *he*, etc. should all be *S*.

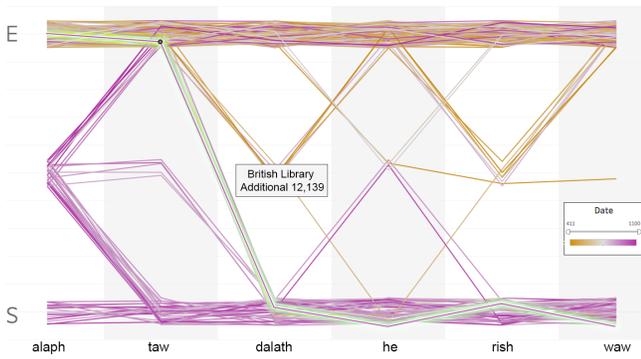


Figure 4: Parallel coordinates plot of all securely dated Syriac manuscripts available in the database. *S* forms of each letter appear along the bottom, and *E* forms appear along the top. Each line represents a single manuscript; position indicates which form(s) of each letter appear in the manuscript, and the line is color coded to indicate the manuscript's date (early to late). In this view, any diagonal line indicates a contradiction to the standard model.

In reality, *E* forms of some letters often coexist in the same manuscript alongside *S* forms of other letters. Consider, for example, *British Library Additional 12,139* highlighted in Figure 4, securely dated to 999 AD. This manuscript contains *E alaph* and *taw*, mixed with *S dalath*, *he*, and *rish*. Using a parallel coordinates visualization of the various letter forms appearing in all securely dated manuscripts in the collection (see Fig. 4), we observe that mixed-script manuscripts are far from uncommon. This pattern of *E alaph* and *taw*, *S dalath*, *rish*, and *he* is one of the most dominant patterns of deviation – 53% of securely dated manuscripts with a mix of *E* and *S* letter forms follow this pattern.

Myth 3: For each letter, a single manuscript contains either the *Estrangela* or *Serto* form.

According to the standard model, a scribe will use either an *E* or an *S* form for each primary letter (*alaph*, *dalath*, *he*, *rish*, or *taw*) in a given manuscript. Note that in the collection of securely dated manuscripts, all manuscripts are considered to be the work of a single scribe. If there is any evidence that more than one scribe contributed to a given manuscript, the manuscript is divided into parts such that each scribe's contribution is considered a separate manuscript. According to this model, if a scribe uses an *E alaph*, every other *alaph* in that manuscript should also take an *E* form. Similarly, if a scribe uses an *S he* in a given manuscript, every other *he* by that scribe in that manuscript should also appear in its *S* form.

In direct contradiction to this assertion, scholars noted when using the automated script chart generator that in many cases a single manuscript hand will contain both the *E* and *S* forms of the same letter, usually on the very same page, and at times even in the very same word. For example, Figure 5 shows a page from *British Library Additional 14,548* securely dated to 790 AD. These are not occasional occurrences: at least 20 of the 178 securely dated manuscript hands contain at least one Syriac letter in both its *E* and *S* form. This dual use becomes especially prevalent between the ninth and eleventh centuries (during which 18% of securely dated manuscripts show such overlap), although such violations of the standard model appear much earlier. Even the earliest securely dated Syriac manuscript, *British Library Additional 12,150* (written in the year 411) has both *E* and *S* forms of *dalath* and *rish*.

5 DISCUSSION AND BROADER IMPACT

Because most manuscripts lack verifiable dates of composition and their relationships with each other are indeterminate, researchers

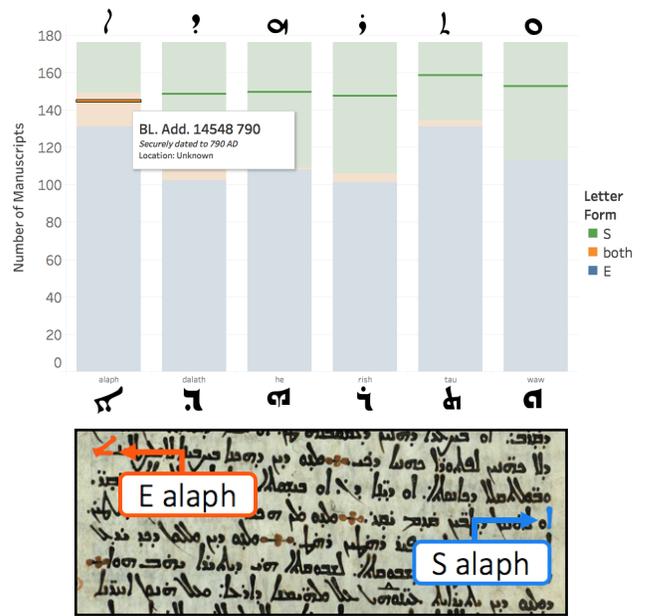


Figure 5: (top) A stacked bar chart showing the proportion of securely dated manuscripts exhibiting each letter form. For each manuscript, a bar shows whether the scribe used *E*, *S*, or both forms of each letter. Every bar in the orange *both* group contradicts the standard model for Syriac script. An excerpt from the highlighted manuscript *British Library Additional 14,548* (bottom) illustrates the appearance of both *E* and *S* forms of the letter *alaph* on the same page. Manuscript image © British Library Board: BL. Add. 14,548.

frequently use manuscripts solely as a means to reconstruct the earliest form of the individual texts they contain. This type of text criticism focuses on origins, but deemphasizes the transmission and reception history. In contrast, the data this project generates through the clustering of related manuscripts enables scholars to explore manuscripts as collections. In the early stages of this project, one historian was asked to think aloud about the kinds of yet unanswered questions scholars might ask if they had interactive access to the entire corpus of surviving manuscripts. Below is a transcript of a portion this thought exercise:

What if we discovered that most Syriac anti-Jewish writings were copied in the decades following the Islamic conquests? How would this change our understanding of the connections between Christian anti-Judaism and the rise of Islam? Would it strengthen the relationship between 7th century anti-Jewish policies in Byzantium and anti-Judaism further east?

Or what if one of the Syriac scribes who copied Aristotle was also responsible for transmitting anti-Muslim treatises? And what if we discovered that he was writing from Baghdad? Would this explain why Syriac Christians were so interested in classical logic?

The ability to ask (and in some cases, to begin answering) such questions has clear ramifications beyond the field of Syriac studies. It transforms our conceptualization of scribes from neutral, after-the-fact transmitters of ancient texts into *active historical agents*, individuals with values and agendas we now could effectively study.

REFERENCES

- [1] Part i: The auxiliary sciences. iii. paleography. *The Catholic Historical Review*, 2(3):367–372, 1916.
- [2] B. Bischoff. *Latin Palaeography: Antiquity and the Middle Ages*. Cambridge University Press, 1990.
- [3] S. P. Brock. *An introduction to Syriac studies*, volume 4. Gorgias Press LLC, 2006.
- [4] S. P. Brock and L. van Rompay. *Catalogue of the Syriac Manuscripts and Fragments in the Library of Deir Al-Surian, Wadi Al-Natrun (Egypt)*. Orientalia Lovaniensia analecta. Isd, 2014.
- [5] K. Bush*, M. Penn, , N. Howe, and S. Wu*. Challenging the estrangela / serto divide: Why the standard model of syriac scripts just doesn't work. *Hugoye: Journal of Syriac Studies*, 21(1), 2018.
- [6] F. B. Chatonnet. Some reflexions about the origin of the serto script. *The Harp*, 18:173–178, 2005.
- [7] L. Delisle. de l'emploi du signe abrégatif 9 à la fin des mots. *Bibliothèque de l'École des Chartes*, 67(67):591–592, 1906.
- [8] J. A. Fitzmyer. *A Wandering Aramean: Collected Aramaic Essays*, volume 25. Scholars Press, 1979.
- [9] J. F. Healey. The early history of the syriac script: a reassessment. *Journal of Semitic studies*, 45(1):55–67, 2000.
- [10] U. Hinrichs, S. Forlini, and B. Moynihan. Speculative practices: Utilizing infovis to explore untapped literary collections. *IEEE transactions on visualization and computer graphics*, 22(1):429–438, 2016.
- [11] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. pages 83–103, 2015.
- [12] G. A. Kiraz. *The New Syriac Primer: An Introduction to the Syriac Language with a CD*, volume 9. Gorgias Press LLC, 2007.
- [13] J. Mabillon. *De re Diplomatica*. Bibliopola, 1964.
- [14] T. Muraoka. *Classical Syriac: a basic grammar with a chrestomathy*, volume 19. Otto Harrassowitz Verlag, 2005.
- [15] A. Palmer. The syriac letter-forms of tur abdin and environs. *Oriens Christianus*, 73:68–89, 1989.
- [16] T. H. Robinson. *Paradigms and exercises in Syriac grammar*. Oxford University Press, 1962.
- [17] W. M. Thackston. *Introduction to Syriac*. Department of Near Eastern Languages and Civilizations, Harvard University, 1983.
- [18] W. M. Thackston. *Introduction to Syriac: An Elementary Grammar with Readings from Syriac Literature*. Ibex Publishers, 1999.
- [19] E. M. Thompson. *An introduction to Greek and Latin palaeography*. Cambridge University Press, 2013.