Schriften des Instituts für Dokumentologie und Editorik — Band 11

Kodikologie und Paläographie im digitalen Zeitalter 4

Codicology and Palaeography in the Digital Age 4

herausgegeben von | edited by

Hannah Busch, Franz Fischer, Patrick Sahle

unter Mitarbeit von | in collaboration with

Bernhard Assmann, Philipp Hegel, Celia Krause

2017
BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de/ abrufbar.

SPONSORED BY THE

Federal Ministry
of Education
and Research

Diese Publikation wurde im Rahmen des Projektes eCodicology (Förderkennzeichen 01UG1350A-C) mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) gefördert.
Publication realised within the project eCodicology (funding code 01UG1350A-C) with financial resources of the German Federal Ministry of Research and Education (BMBF).

2017

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-7448-3877-1
Einbandgestaltung: Julia Sorouri, basierend auf Vorarbeiten von Johanna Puhl und Katharina Weber; Coverbild nach einer Vorlage von Swati Chandna.
Satz: LuaTeX und Bernhard Assmann

Phenetic Approach to Script Evolution

Gábor Hosszú

Abstract

Computational palaeography, as a branch of applied computer science, investigates the evolution of graphemes, explores relationships between scripts, and provides support for deciphering ancient inscriptions, among others. The author applied methods often used to describe evolutionary processes in phylogenetics to analyse the development of scripts. Unlike in the clear evolution of phylogenetics, graphemes used to describe the evolution of scripts are sometimes indistinguishable from their glyph variants. Moreover, the historical background is at times incomplete. In order to reduce uncertainty, the author developed an exploratory data analysis method that combines phenetic analysis methods with a cladistic approach. The paper details the tests the author developed to explore the relationships among 66 different scripts with 186 different features. To extract data for analysis required determining the similarity groups of glyphs and orthographical rules in different scripts; the input is data from humanities-based palaeography. Creation of the similarity groups of the glyphs is based on minimizing the differences between the topological properties of the glyphs and individual decisions in order to avoid homoplasies, as well as the erroneous omission of slightly differing but otherwise related glyphs. For the second purpose, the layered grapheme model and the concept of characteristic transformations of related glyphs were used. Based on the extracted features of the scripts, various machine-learning methods were applied, including multidimensional scaling, k-means partitional clustering, and various hierarchical clustering methods. These algorithms produced similar results, represented in two- and three-dimensional scatter plots and phenograms, which visualize the relationship between the scripts. These results roughly concur with the results of humanities-based palaeography; however, new conclusions can be also derived, including the introduction of the concept of witness scripts, and glyph- and grapheme-level reticulations, which are used to describe the possible relationship of graphemes and scripts. The presented results demonstrate the usefulness of a developed modified phenetic method in exploring the similarities of scripts, and based on the results obtained, some improvements in modelling the distribution of certain historical scripts were also proposed.

Kodikologie und Paläographie im Digitalen Zeitalter 4 – Codicology and Palaeography in the Digital Age 4. Hrsg. Hannah Busch, Franz Fischer und Patrick Sahle, unter Mitarbeit von Bernhard Assmann, Philipp Hegel und Celia Krause. Schriften des Instituts für Dokumentologie und Editorik 11. Norderstedt: Books on Demand, 2017. 179–252.

Zusammenfassung

Computergestützte Paläographie als Zweig der angewandten Informatik untersucht unter anderem die Evolution von Graphemen, erforscht die Beziehungen zwischen Schriften und leistet Unterstützung bei der Entzifferung sehr alter Inschriften. Der Autor hat Methoden, die häufig für die Beschreibung evolutionärer Prozesse verwendet werden, angewandt, um die Entwicklung von Schriftsystemen zu untersuchen. Im Gegensatz zu der klaren Evolution in der Phylogenetik, sind Grapheme, die zur Beschreibung der Schriftevolution benutzt werden, manchmal nicht von ihren Glyph-Varianten zu unterscheiden. Zudem ist der historische Hintergrund zuweilen unvollständig. Um die Unsicherheiten zu reduzieren, hat der Autor eine explorative Methode der Datenanalyse entwickelt, die phänetische (numerisch taxonomische) Analysemethoden und einen kladistischen Ansatz kombiniert. Der Beitrag erläutert die Testreihen, die der Autor entwickelt hat, um die Beziehungen zwischen 66 verschiedenen Schriften mit 186 verschiedenen Merkmalen zu erforschen. Die Datenextraktion für die Analyse machte es notwendig, zunächst die Ähnlichkeitsgruppen von Glyphen und die orthographischen Regeln für verschiedene Schriften zu bestimmen; die Ausgangsdaten stammen also aus der traditionellen Paläographie. Die Bestimmung der Ähnlichkeitsgruppen basiert sowohl auf der Minimierung der Unterschiede zwischen den topologischen Eigenschaften der Glyphen und individuellen Entscheidungen zur Vermeidung von Homoplasien (zufälligen Ähnlichkeiten), als auch der falschen Aussonderung von nur leicht unterschiedlichen, ansonsten aber ähnlichen Glyphen. Für die zweite Aufgabe wurden das Graphem-Schichtenmodell und das Konzept der charakteristischen Transformationen verwandter Glyphen benutzt. Auf der Grundlage der bestimmten Merkmale wurden verschiedene Methoden des maschinellen Lernens wie multidimensionale Skalierung, k-Means Partitions-Clusteranalyse und verschiedene hierarchische Clusterverfahren angewandt. Diese Algorithmen haben zu ähnlichen Ergebnissen geführt, die in zwei- und dreidimensionalen Streudiagrammen und Phänogrammen (Kladogrammen) ausgedrückt werden und die Verhältnisse zwischen Schriften sichtbar machen. Die Ergebnisse stimmen grob mit den Resultaten der bisherigen paläographischen Forschung überein, allerdings können aus ihnen auch neue Erkenntnisse gezogen werden. Dazu gehören die Einführung des Konzepts der »Zeugenschriften« und Verbindungen auf der Glyph- und Graphemebene, die zur Beschreibung möglicher Beziehungen zwischen Graphemen und Schriften genutzt werden. Die hier vorgestellten Ergebnisse zeigen den Nutzen einer entwickelten phänetischen Methode für die Untersuchung von Schriftähnlichkeiten. Auf der Grundlage der erzielten Resultate werden außerdem Verbesserungsvorschläge für die Modellierung der Verbreitung und Verteilung einiger historischer Schriften gemacht.

1 Introduction

Computational palaeography, in other words *engineering in palaeography*, as a branch of applied computer science, deals with investigating the evolution of graphemes, exploring relationships between scripts, and providing support for deciphering ancient inscriptions, among others. Its main focus is using engineering methods to explore relationships found in the data of ancient inscriptions and other palaeographical (including epigraphic) information. Computational palaeography has an applied machine learning approach, and it extends the engineering modelling methods to any data of the written cultural heritage. The fields of computational palaeography are improving and tailoring phylogenetic algorithms for exploring relationships in palaeographical data and modelling the evolution of scripts and graphemes, including the spatial analysis of the various glyphs. The research efforts of the author and his colleagues cover a broad range of topics such as applying machine learning methods to explore similarities among scripts or orthographies (Hosszú 2014; Tóth et al. 2016), modelling graphemes in different abstraction levels (Pardede et al. 2016), reconstructing lineages of graphemes in various scripts (Hosszú 2015), investigating methods for testing the appropriateness of the reconstructed lineages, and developing algorithms for deciphering historical inscriptions (Tóth et al. 2015).

As opposed to computational palaeography, *digital palaeography* (Ciula 2005; 2009)—or in other words *computer-aided palaeography* (Stokes 2009)—is part of Digital Humanities, an interdisciplinary field of Palaeography, Computing, and Artificial Intelligence (Aussems and Brink 2010, 296). It is an extension of the type of palaeography found in the humanities using tools from computer science; their goals are similar (e.g. Aussems 2010). Humanities-based palaeography, with diplomatics and textual criticism, constitutes the main disciplines of philology. For simplicity, the term palaeography includes epigraphy in this article. Digital palaeography includes sub-fields such as *quantitative codicology* (Stokes 2015) and *quantitative palaeography*, and it entails the identification of scribes, reconstruction of fragmented texts with image analysis, digital representation of medieval scripts, digital description, imaging, recording, and reproduction of the manuscripts, image pre-processing for machine learning (e.g. feature extraction, pattern recognition, optical character recognition), textual analysis, physical analysis, storage in databases extending with semantic structures, digital presentation, and the teaching of palaeography (Ciula 2009; Fischer et al. 2010). Quantitative aspects can be measured by automated means and the results can be subjected to automated clustering techniques (Ciula 2005). Hierarchical clustering was used for creating the groups of the morphologically similar glyphs of a grapheme. A composite palaeographical classification method, including k-means clustering, was applied to match a particular document to a large set of palaeographical records (Wolf et al. 2011). Numerical tools were developed to automate the study of medieval writing samples in the context of the Graphem project, which is intended to explore, analyse, and categorize medieval scripts (Cloppet et al. 2011). It is noteworthy that the border between digital palaeography and computational palaeography is smooth, both of them use machine-learning tools, and they are related to analytical palaeography, which deals with the classification of glyphs and belongs to the palaeography of the humanities.

This paper details the investigations carried out to explore the relationships among 66 different scripts using clustering and factor analysis, where 186 different features of the examined scripts were involved in the phenetic analysis. As the input of the analysis, a data extraction step is necessary, which means determining the similarity features groups (SFGs) in different scripts; where the input was the result of humanities-type palaeography, and the criteria for constructing the SFGs include phenetic and cladistic considerations. Various machine-learning methods were applied, including multidimensional scaling, k-means partitional clustering, and different hierarchical clustering methods, and the different algorithms produced similar results; they are represented in two- and three-dimensional scatter plots and phenograms, where each point represents a single script, and the relative distance of these points represents the relationship between the scripts.

The paper is organized as follows: Section 1 gives background information, including a definition of the concepts and terminology of machine learning, comparison of phylogenetic approaches, details of phenetic tools, cluster validity techniques, and the terms and concepts of computational palaeography. Section 2 is dedicated to the newly developed exploratory data analysis method, including the general description of the algorithm. Section 3 presents the feature extraction with SFGs, section 4 evaluates the obtained results, and section 5 provides conclusions. A short Appendix presents some additional examples of the inscriptions written with the lesser-known scripts of the Eurasian Steppe.

2 Background

2.1 Computational palaeographical and machine learning terminology

A *writing system* is “a set of visible or tactile signs used to represent units of language in a systematic way” (Coulmas 1999). *Script* is the graphic form with orthographical rules of a writing system. A script has several versions, including the subset of the graphemes of the script belonging to various areal, cultural, temporal, stylistic, and typographical versions. An extinct script, for which only inscriptions have survived, is reconstructed from the surviving inscriptions, including their explored properties (e.g., orthographical rules).

Orthography is a certain set of the graphemes of a script and a set of rules for using a script in a particular language; e.g., some medieval orthographies of the Latin script include medieval German, medieval Italian, Middle English, Old French, Old Hungarian, Old Norse, etc. In computational palaeography, the term orthography means a specific set of graphemes with specific glyphs; e.g., the ê in French and medieval Italian, the ß in German, the ʒ in Old English, the ÿ in Dutch, French, medieval German, and Old Hungarian, ǫ in Old Norse and Old Hungarian orthographies, etc. All of these graphemes belong to the Latin script.

Taxon is a taxonomic unit, a set of objects classified into the same category in a formal taxonomic system. In biological evolution, taxa are usually species, and the entities of the species are called organisms. In computational palaeography, the taxa are the scripts, and the entities are the particular versions of a script (orthographies) used for each inscription. However, other approaches are also possible depending on the focus of the research – if the broad focus is on a particular orthography, it could be considered a taxon and variations of its graphemes would be the entities.

Grapheme is the smallest semantically distinguishing element in a script (Sukkarieh et al. 2012). A grapheme could be a letter, ideogram, logogram, ligature, numerical digit, diacritic, accent, phonogram, determinative, punctuation mark, syllabogram, etc. The grapheme is taken as an object with different features including its shape variations (called glyphs), transliteration values, sound values, age in which it was used, geographical distribution area, and the script to which it belongs.

Glyph refers to a unique shape of a grapheme that can be described by topological information. In the view of computational palaeography, the definition of a grapheme has the following conjunctive constituents: *(i)* different phonemes belong to the same grapheme if the sets of their possible phonetic values are identical or reasonably altering; *(ii)* the glyph variants of a certain grapheme must be visually very similar; *(iii)* any glyph variants of a certain grapheme must represent all phonetic values of that grapheme; and *(iv)* the usage of the grapheme is determined by the orthographical rules of a certain age in the history.

Inscription is a survived relic of one or more scripts independent of the writing materials (stone, wall, wood, ink and paper/papyrus/parchment, etc.), and physically it can be a fragment, a manuscript, a scroll, or a codex. In other words, the term inscription is used in the widest possible sense.

Symbols are the minimum individual units of the inscriptions from a visual perspective. In other words, inscriptions are composed of a sequence of symbols. Consequently, a symbol is the materialization of a particular glyph of a grapheme, and the grapheme is the abstraction of a symbol.

Feature (also called *character*) is a heritable trait (property) of a taxon. A feature can take one of more forms; these various forms are described by the feature states. It is noteworthy that in phylogenetics, the term "character" is much more frequently used than the term "feature"; however, in pattern recognition, the term "character" is used very similarly to the term "grapheme." In computational palaeography, the concepts of both phylogenetics and pattern recognition are used, which makes the term "character" ambiguous. Therefore, instead of the ambiguous term "character," the preferred terms are "feature" and "grapheme," respectively. A computational palaeographical feature is any property of scripts that can take one or more forms; these different forms are called states of the features. These features could be graphemic and orthographic. *Graphemic features* are represented using a binary variable having the two states "presence of a glyph variant of a grapheme" and "absence of a glyph variant of a grapheme". Similarly, *orthographic features* represent the presence or absence of various orthographic rules, e.g., directions or separator lines among rows of the inscriptions of certain scripts. In other words, the categorical variables are transformed into Boolean indicator variables (see below for details).

Object is the basic unit (data point) in machine learning methods, which is described with a vector of *variables* (in other words, *attributes*). If the machine learning method is used in phylogenetics, the object is the taxon, and the variable is the feature. Therefore, the object is usually the script, and the variable is the feature of the script, especially the existence of certain glyphs in the given script. In such a way, the *taxon-feature matrix* is composed of taxons in rows, and feature states in columns. If the feature is transformed into a Boolean indicator variable, the value of a feature state is 1 if it is present in a particular taxon, and 0 if it is absent.

Clade is a taxon and all of its descendant taxa (Hennig 1966). The taxa in a single clade share an evolutionary relationship. The taxa have features, and a taxon can be characterized by the feature states. *Apomorphy* is a derived feature state of a taxon; this feature state is known as *apomorphic*, and includes the types called autapomorphy, synapomorphy (homology), or homoplasy (analogy). *Autapomorphy* means a feature is present in an individual taxon, but not any of its ancestors. If there are descendants of a taxon that inherit this autapomorphic feature, then they create a clade, and this clade is characterized by this feature as apomorphy. This demonstrates that the properties apomorphy or autapomorphy are relative terms. *Synapomorphy* is a feature state shared by two or more taxa resulting from an innovation in their shortest common ancestor. Synapomorphy is a *homology,* meaning a similarity due to inheritance of a feature state from a common ancestor. *Homoplasy* (homoplastic feature state) is when two or more apomorphic feature states are identical; however, they originated from not a common ancestor, but rather by convergence or reversal. *Convergence* (parallel evolution) is when the same feature state presents in two unrelated taxa due to similar conditions. *Reversal* (back-mutation) is when a feature state reverts to an earlier state. *Plesiomorphy* is a feature state that taxa of a clade have retained from their ancestors; such feature state is *plesiomorphic* (ancestral). Considering a clade, the common feature states of its taxa may be plesiomorphic or synapomorphic.

Reticulate evolution happens in the case of *reticulate* events (Sneath 1975), such as hybridization (a new taxon is formed from two different taxons) or horizontal gene transfer (feature state transfer). In computational palaeography, *hybridization* means the combination of two scripts, e.g., the Early Cyrillic script is surely a hybrid, or combination, of the Greek and the Glagolitic scripts. *Feature state transfer* means transfer of a glyph or orthographical rule between contemporaneous scripts. It has two subcases: *(i) Glyph-level reticulation:* if a grapheme exists in a script, but an additional glyph for this grapheme is transferred (borrowed) from another script (loan glyph); or the grapheme did not exist in a script, and its glyphs are transferred from more than one grapheme (new grapheme with loan glyphs). See the comment on SFG-100 in table 10 for an example. *(ii) Grapheme-level reticulation:* if a grapheme with all of its glyphs in a script originated from a certain grapheme of another script (loan grapheme). For instance, the Latin script was developed from the Greek-origin Etruscan graphemes, and, later the graphemes Y and Z were directly adopted from the Greek. In this example, the graphemes were absent from the Latin script at the time of adoption.

Witness script is a script that has retained features from another script from a remote geographical region and/or a bygone era. For example, the Greek script retained features from an early form of the Phoenician script. Thus we can say that the Greek script bears witness to certain features existing in the early Phoenician script.

2.2 Phylogenetic approaches for computational palaeography

Phylogenetics aims to uncover the evolutionary relationships between taxa to obtain an understanding of their evolution. Phylogenetics in a wider sense has three areas: *phenetics* (numerical taxonomy), *cladistics,* and *phylogenetics* (in a narrow sense). The output of these methods is usually presented in tree-like branching diagrams (dendrogram, indexed tree) called *phylogenetic trees* (in a wider sense), or the *phenogram, cladogram,* and *phylogram* (in a narrow sense), respectively. A tree is a connected acyclic graph consisting of a set of vertices (nodes) and a set of edges (branches), each of which connects a pair of vertices. The differences between phenograms, cladograms, and phylograms are related to their underlying features: phenograms use phenotypic information, cladograms use hierarchical relationships among taxa based upon homologies (synapomorphies), and phylograms convey genealogical information. The phenetic relationships are usually multidimensional; therefore, different procedures can produce a variety of phenograms (Sokal et al. 1963). The cladogram is a synchronic representation of the evolution; it describes the relationships among the taxa of the same time. A phylogram is an estimation of genealogical relationships among a group of taxa (Kitching et al. 1998, 213); it represents evolutionary histories in which the main events are speciations (at the internal nodes of the tree) and descent with modification (along the edges of the tree).

The lengths of the branches of the tree have different meanings in the three approaches. In phenograms the length of the branch represents the similarity among the taxa. In cladograms, the length of the branch has no specific meaning. In phylograms, the length of the branch represents the amount of inferred evolutionary change: the longer the branch, the greater the variation between taxa. If in a tree, two scripts have a more recent common ancestor, then we expect these two scripts to have the most features in common, because they are the pair that has had the least opportunity to diverge. Using more than one feature provides a measure of the overall difference between them. It is assumed that the features in common are not convergent and have not evolved independently in the two branches by chance, either. The differences likely accumulate at a fairly steady rate, so that more differences mean that there is a less recent common ancestor.

The tree-based phylogenetic model is less suited for reticulate events, when the new taxon has more than one ancestor. In this case, *phylogenetic networks* better describe the evolution than phylogenetic trees. When scripts converge in the case of reticulate evolution, a network model is more appropriate with additional edges to reflect the dual parentage of a script. These edges could be bidirectional if both scripts borrow from one another. Change happens continually to scripts, but not usually at a constant rate, with its cumulative effect producing splits into orthographical variations and script families. Finally, there could be loss of any evidence of relatedness. Unlike biology, it cannot be assumed that scripts all have a common origin; relatedness must be established.

An analogue of the field of computational palaeography is the use of phylogeny for historical linguistics (Forster and Renfrew 2006). Methods of computational phylogenetics and cladistics can be used to define an optimal tree or network to represent a hypothesis about the linguistic evolution. Nakhleh et al. (2005) compared the following phylogenetic reconstruction methods on an Indo-European linguistic dataset: UPGMA, maximum parsimony, weighted and unweighted maximum compatibility, neighbour joining, and Gray-Atkinson algorithms. They found that UPGMA is inferior on these datasets, because they used data from different time depth. The other algorithms were not sensitive to this feature of the applied datasets. The maximum parsimony and the unweighted maximum compatibility methods returned similar dendrograms. Their dataset for the phylogenetic reconstruction for comparative historical linguistics contains lexical, phonological, and morphological features. Bar-

Issue	Species	Languages	Scripts
Reticulation	Possible (e.g., horizontal gene transfer)	Possible (e.g., loan words)	Possible (hybridization of scripts, glyph-level reticulation, grapheme-level reticulation)
Interrelation of features	Frequent (biological feature states usually affect each other)	Possible (shared cultural development)	Possible (e.g., influence of the writing techniques, or effect of geometric style)
Homoplasy	Possible (e.g., parallel evolution)	Rare, historical linguists can identify many of the borrowings; therefore, they can be screened out (Barbançon et al. 2013)	Frequent (e.g., presence of similar, but unrelated glyph variants in unrelated scripts)

Table 1: Comparison of biological, linguistic, and palaeographical feature evolutions

bançon et al. (2013) used Hamming distances in the UPGMA to define the distance matrix between the set of languages.

Another interesting application of the phylogenetic approach is the examination of the evolutionary relationships in software (Sampaio 2007). However, in the cases of modelling software evolution and biology, different methods are used to compare and classify the taxa.

Phylogenetic reconstruction methods originally designed for biological data could be used on palaeographical data for reconstruction of the phylogenies of script families (table 1). Analysis can be carried out on the features of scripts. Each feature being a Boolean indicator defines an equivalence relation on the set of scripts, such that two scripts are equivalent if they exhibit the same state for the same feature. In the case of identical feature states, the presumption could be that the shared state arose due to common inheritance. However, shared states can also arise due to homoplasies: borrowing (reticulate event) or random chance (autapomorphy).

A dendrogram is a common way to visualize the results of computational palaeographical analysis. There are different ways to represent this. One solution is when an internal node represents a palaeographical ancestor in a phylogenetic tree or network. Each taxon (usually script) is represented by a path (branch); the paths show the different states as the writing system evolves. There is only one path between every pair of vertices. Another solution for a phylogenetic tree is one where the taxa are represented by nodes and their evolutionary relationships are represented by branches.

Script classification is carried out based on a taxon-feature matrix, where the rows usually correspond to the various scripts being analysed and the columns correspond to different features by which each script may be described; however, altering approaches are also possible depending on the goal of the actual investigation. Skelton (2007) used phylogenetic systematics for orthographical variations of the Linear B script, where taxa represented scribal hands and phylogenetic features represented variants of the same Linear B glyph. The phylogenetic tree produced by running the data matrix using parsimony as the optimality criterion is consistent with and clarifies what is known or hypothesized about the history of Linear B. Skelton demonstrated the usability of phylogenetic analysis to reconstruct the evolution of writing systems.

Wheeler and Whiteley (2014) criticised the use of basing analyses on proto-languages in historical linguistics, and their arguments apply equally to proto-scripts in the humanities-based palaeography. In such palaeography, classification of scripts is based on the comparison of graphemes, glyphs, and orthographical rules to identify regular correspondence features. From such features a proto-script is reconstructed, and it is posited as the evolutionary ancestor of the observed scripts. A proto-script is regarded as a real script once used by a population in a particular time and place. Differentially shared patterns of change from the proto-script among descendant scripts are used to determine subgroups within the family. However, variation is compounded by the inherently sporadic data: there are no records for several extinct scripts, which might have served to falsify proposed proto-script reconstructions. The concept of proto-scripts is used in this paper as a method to identify group scripts having a common unknown ancestor; however, no analysis is based on such theoretical scripts.

Constructing phylogenies based on the surviving inscriptions has some difficulties. Namely, several scripts were originally used with perishable carriers (papyrus, wood, etc.). Even for those scripts that are well represented, only certain parts of the inscriptions survived, which limits the range of features that can be examined. The record is usually just too incomplete in both a spatial and a temporal sense to be of much use. One possible approach is to construct a phylogeny based on the characteristics of surviving inscriptions. However, the available fossil record (corpus of surviving inscriptions) is so fragmentary that the phylogeny of the vast majority of taxa is unknown. Phenetic classification is possible for all groups. By contrast, cladistic analysis, based on branching sequences, requires historical inferences about the direction of evolution in a group of taxa (Lindberg 2012). Phenetics attempts to classify taxa based on the concept of overall similarity, typically in morphology, without regard for their evolutionary relationships. Phenetic methods can be optimal when the distinctness of related taxa is important, and the data necessary for exploring the genetic relationships are missing. In phenetics, the more features on which the phenetic analysis is based, the better a given classification will be; every feature is of equal weight in creating natural taxa, and classifications are based on morphological similarity (Lindberg 2012). Subjectivity could be removed by examining as many features of the script as possible.

Decisions related to feature selection have the potential to impact a phylogenetic analysis, and these decisions also raise other issues, such as whether all features should be treated identically, or whether weighting schemes should be used to reflect the assumed reliability of the feature (Barbançon et al. 2013).

Consequently, in the case of limited information, it is futile to create an evolutionary tree, because there is no way to prove whether it is right or wrong. Instead, grouping taxa entirely on the basis of similarities is more efficient. As opposed to phylograms or cladograms, phenograms are only based on taxon similarities. In a phenogram, each branch point represents a step of increasing dissimilarity. In such case, the internal nodes of the graph do not represent ancestors but are introduced to represent the conflict between the different splits in the data analysis. The phenetic distance is the sum of the weights—represented as lengths—along the path between taxa. If discrete features are coded, the phenetic concept of homology is operationally identical to that used in cladistics. In phenetics, the homoplasy attending feature conflict is not reconciled (Wills 2001).

2.3 Machine-learning algorithms for phenetics

Phenetic analysis starts with the collection of raw measurement data on the chosen set of morphs, thus creating the taxon-feature matrix. Then a measure of dissimilarity is computed for each pair of taxa based on an appropriate metric. In the next step, a cluster analysis is performed to group taxa that are most similar. An index of average distance between each taxon could be calculated; then these distances are fitted into a hierarchical clustering pattern. It is difficult to decide which clustering algorithm should be used, and the methods do not all give the same answer. Therefore, following is an overview of some important algorithms used in the phenetic analysis, including the ordination, the clustering, the cluster validity indices, and the leaf ordering of the dendrograms to obtain a possible best cluster structure.

Clustering is an unsupervised learning (exploratory data analysis) method, which needs very little *a priori* knowledge. It is a useful technique for grouping data points such that points within a single group have similar characteristics, while points in different groups are dissimilar. Clustering is the task of categorizing objects having

	Features present in s_i	Features absent from s_i
Features present in s_j	f_{11} is the number of features present in both s_i and s_j	f_{01} is the number of features absent from s_i and present in s_j
Features absent from s_j	f_{10} is the number of features present in s_i and absent from s_j	f_{00} is the number of features absent from both s_i and s_j

Table 2: Parameters used in expressing the comparison of the features of scripts s_i and s_j

several attributes into different classes such that the objects belonging to the same class are similar, and those that are broken down into different classes are not. In the case of clustering, the problem is to group a given collection of unlabelled patterns into meaningful clusters. Labels are associated with clusters, but these category labels are data driven; that is, they are obtained solely from the data (Jain et al. 1999).

In clustering, the object is to place data points into the same cluster when they are similar enough according to some predefined metric. The predefined metric is one aspect that makes clustering a subjective process. In the case of computational palaeography, the features (variables) are the glyphs or orthographical rules, and the feature states (their values) are the presence or the absence of the glyphs or orthographical rules. Therefore, these variables are categorical with binary values. For comparing categorical data, the *Boolean* indicator variables are introduced. The formulae for the number of presence/absence feature states are written using the abbreviations in table 2.

The similarity of two objects (taxa, data points, in our case scripts) can be expressed by a metric. For categorical data, the *Jaccard* index (1) is widely applied, where M is the number of taxa. The Jaccard index is a statistic ordinarily applied to compare the similarity and diversity of the variables (features) of the examined objects, if the double absence (f_{00}) has no significance. This fits well with our dataset, since the clear majority of the features are glyphs, and the absence of a glyph in a script is not specific, since there are hundreds of glyphs that are absent from a certain script.

$$s_J(x_i, x_j) = \frac{f_{11}}{f_{11} + f_{10} + f_{01}}, i, j \in \{1, \dots, M\} \quad (1)$$

The Jaccard index is not a metric; however, it can be converted to a metric distance, shown in (2).

$$d_J(x_i, x_j) = 1 - s_J(x_i, x_j) = \frac{f_{01} + f_{10}}{f_{11} + f_{10} + f_{01}}, i, j \in \{1, \dots, M\} \quad (2)$$

The square root of Jaccard distance is an Euclidean metric (Gower and Legendre 1986), given as (3).

$$d_{SRJ}(x_i, x_j) = \sqrt{d_J(x_i, x_j)}, i, j \in \{1, \dots, M\} \quad (3)$$

Another approach is to examine the object-variable (taxon-feature) matrix using a geometric representation: the objects (taxa) are points in a space spanned by variables (features) as axes of a scatter plot. Since the number of variables (dimension) is very large, it is necessary to replace the original large number of dimensions by a few artificial axes so as to represent the data structure as efficiently and faithfully as possible. This method is called *ordination* (Goodall 1954). One kind of ordination is *multidimensional scaling* (MDS), which can produce a dimension-reduction of objects from their dissimilarities. Where in the original high-dimensional space the variables of the objects are Boolean indicators, in the reduced dimensional space resulting from MDS the variables of the objects are quantitative.

A clustering approach can be taken not only in the original high-dimensional space but also in the reduced dimensional space. The *Squared Euclidean* difference is widely applied as a measure between quantitative data. Let $x_i = \left[x_{i1}, x_{i2}, \dots, x_{iN}\right]$ and $x_j = \left[x_{j1}, x_{j2}, \dots, x_{jN}\right]$ be two data points in the N-dimensional space of the data points. The Squared Euclidean difference is given as (4).

$$d_{SE}\left(x_i, x_j\right) = \sum_{k=1}^{N} \left(x_{ik} - x_{jk}\right)^2, i, j \in \{1, \dots, M\} \tag{4}$$

Clustering can be broken down into the following main steps. *Definition of object proximity:* as measured by a distance function defined on pairs of objects. *Clustering:* can be hard (crisp) or fuzzy. In crisp clustering, one object can belong to one and only one cluster. In fuzzy clustering, each object belongs to each cluster but with a varying degree of membership. *Cluster validation:* uses a specific criterion of optimality (Jain and Dubes 1988; Jain et al. 1999).

Jain et al. (1999) defined several types of clustering algorithms. *Hierarchical clustering:* These algorithms create clusters recursively by merging smaller partitions into larger ones or splitting larger clusters into smaller ones. These produce a nested series of clusters based on similarity. *Partitional clustering:* decomposes data sets into a set of disjointed clusters. *Density-based clustering:* creates clusters based on density functions. Its main advantage is to create arbitrary shaped clusters. *Grid-based clustering:* quantises the search space into a finite number of cells.

The diameter of a cluster can be defined in a number of ways. *Single linkage* (nearest neighbour) deals with the area where the two clusters are closest to each other. It emphasizes cluster separation: elongated point clouds are recognized, but clusters connected by intermediate objects cannot be detected. It is a hierarchical algorithm that can deal with arbitrary shapes, potentially at the expense of simple clusters. However, this tendency may also produce clusters that are chained. *Complete linkage* (farthest neighbour) deals with the whole area of the clusters; it is sensitive to outliers, and a single point far from the centre can greatly modify the clustering. It emphasizes cluster cohesion; the separation of clusters is not influential (Podani 2000). It produces rightly bound or compact clusters (Baeza-Yates 1992). *UPGMA* (an unweighted pair group method of agglomeration, also called average linkage) merges in each iteration step the pair of clusters with the highest cohesion. In each grouping, the averages are calculated, and those groups with averages closest to each other are lumped together. It was developed for numerical taxonomy (Sokal and Michener 1958; Sokal and Sneath 1963). *WPGMA* (weighted pair group method, arithmetic average) uniformly weights all clusters independently of the number of their members. *Neighbour joining* (Saitou and Nei 1987) is based on the idea of parsimony; however, it does not attempt to obtain the shortest possible tree for a set of data. It operates on distance data, computes a transformation of the input matrix, and then computes the minimum distance of the pairs of objects. A weighted version of the method may also be used.

Ward's method minimizes the increase of the sum of squared deviations from the mean (Ward 1963). It optimizes the homogeneity of the clusters; it gives the most possibly homogenous clusters. In each step of the hierarchical clustering, Ward's method joins those two clusters where (5), the increase of the sum of squared deviations from the mean is minimal. In this case, M objects are clustered into a partition $C = \{C_1, \dots, C_K\}$ of clusters, C_l and C_t are two different clusters ($l \neq t$, $1 \leq l, t \leq K$, $C_l, C_t \in C$), $d_{Ward}\left(C_l, C_t\right)$ is the increase of the sum of squared deviations from the mean in the case of the fusion of C_l and C_t, and K is the actual number of clusters. Ward's method is appropriate for Euclidean distances, and it does not produce the clustering structure with the minimum error (Romesburg 2004, 129–135).

$$d_{Ward}(C_l, C_t) \coloneqq \sum_{x_i, x_j \in C_l \cup C_t} d^2(x_i, x_j) - \Big(\sum_{x_i, x_j \in C_l} d^2(x_i, x_j) + \sum_{x_i, x_j \in C_t} d^2(x_i, x_j) \Big) \quad (5)$$

Another kind of clustering, called partitional methods, decomposes data sets into a disjointed cluster set. Such an algorithm is the *k-means*. It runs quickly but tends toward clusters with non-convex shapes. The k-means process minimizes the error E in (6)

$$E = \sum_{i=1}^{K} \sum_{x \in C_i} d(x, \mu_i) \quad (6)$$

where μ_i is the center of cluster C_i, and K is the number of clusters. The number of iterations needed is unknown since standard k-means is not guaranteed to converge. Moreover, clustering produced by k-means is dependent on the starting points of the clusters.

Most clustering algorithms are very sensitive to their input parameters, and variations in the technique used can sometimes produce misleading results; verification through additional methods of dimensionality reduction analysis is essential, even though the ultimate objective of the research is classification (Podani 2000). Therefore, it is important to evaluate the result of the clustering process. Several clustering validity techniques and indices have been developed. The aim of cluster validity is to find the partitioning that best fits the underlying data. Two measurement criteria have been proposed for evaluating and selecting an optimal cluster structure (Berry and Linoff 1996): *(i) Compactness:* The member of each cluster should be as close to each other as possible. A common measure of compactness is the variance. *(ii) Separation:* The clusters should be widely separated. The basis of comparison is the validity index. A validity index can provide a measure of the quality of the clustering on different partitions of a data set. It helps to determine the appropriate number of clusters present in a data set.

Dunn index is a cluster validity measure introduced by Dunn (1974) that maximizes inter-cluster distances while minimizing intra-cluster distances; it is a ratio of between-cluster and within-cluster separations. In other words, it is the ratio of the smallest distance between objects not in the same cluster to the largest intra-cluster distance, defined as (7)

$$D_i = \frac{\underset{C_l, C_t \in C, l \neq t}{min} \left[\underset{x_i \in C_l, x_j \in C_t,}{min} d\left(x_i, x_j\right) \right]}{\underset{C_r \in C}{max} \left[\underset{x_i, x_j \in C_r,}{max} d\left(x_i, x_j\right) \right]} \tag{7}$$

where M objects are clustered into a partition $C = \{C_1, \dots, C_K\}$ of clusters, $d\left(x_i, x_j\right)$ is the distance between objects $x_i \in C_l, x_j \in C_t, l \neq t$, and $\underset{x_i \in C_l, x_j \in C_t,}{min} d\left(x_i, x_j\right)$ is an intercluster distance metric between clusters $C_l, C_t \in C, l \neq t$. High Dunn index means that the diameter of the clusters is small and the distance between clusters is large; therefore, the clusters are compact and separated. This measurement serves as a measure to find the right number of clusters in a data set, where the maximum value of the index represents the right partitioning given the index. Its disadvantage is that it is sensitive to noise, because the maximum cluster diameter can be large in a noisy environment.

Silhouette index is another approach to measure how similar a given object is to objects in its own cluster, as compared to objects in other clusters. Silhouette is higher when clusters are dense, well separated, or convex, and a zero value indicates overlapping clusters. It provides a graphical representation of how well each object lies within its cluster (Kaufman and Rousseeuw 1990). The S_i Silhouette index for the object x_i is defined as (8)

$$S_i = \frac{b_i - a_i}{max\left(a_i, b_i\right)} \tag{8}$$

where a_i is the average distance from the object x_i to the other objects in the same cluster, and b_i is the minimum average distance from the object x_i to objects in a different cluster, minimized over clusters. The range of Silhouette index is $[-1, 1]$. A high Silhouette value indicates that the object x_i is well matched to its own cluster, and poorly matched to neighbouring clusters. If most objects have a high Silhouette value, then the clustering solution is appropriate. The Silhouette index can be used with any distance metric.

Trees that result from cluster analysis are typically presented with their leaves in an undefined order. However, the distance of these leaves in the dendrogram could refine the cluster structure. Therefore, it is important to maximize the sum of the similarity of adjacent objects in the dendrogram. In the hierarchical clustering investigations, an optimal *leaf ordering* for hierarchical binary cluster tree (Bar-Joseph et al. 2001) was applied.

2.4 Notation and palaeographical sources

Different "runiform" or "Runic"-type scripts were used in largely different places in the Eurasian Steppe and in the Carpathian Basin; their surviving inscriptions are mainly from the 1st millennium AD. Their possible relationship has not been proved or widely accepted. Furthermore, a lot of inscriptions of the Eurasian Steppe have not been deciphered yet. However, many authors have previously demonstrated the similarities of the scripts used in some of these inscriptions (Nagy 1895; Sebestyén 1915, 143–160; Németh 1917–1920; Ligeti 1925; Kyzlasov 1994; Vasil'ev 1994; Vékony 1987b, among others).

Unlike in earlier attempts to decipher them, a single acknowledged scholar, G. Vékony (late Assoc. Prof. in the Eötvös Loránd University, Budapest, 1944–2004), provided a comprehensive decipherment for several of the inscriptions from the Carpathian Basin to Middle Asia. Therefore, his decipherment, including the determined sound values of the signs, was used in the phenetic analysis in this paper. Vékony published his results in several publications (1981; 1985a; 1985b; 1987a; 1987b; 1992a; 1992b; 1992c; 1993; 1996; 1999a; 1999b; 2004), mostly in Hungarian. Since 2008, the author of this article has systematically consulted with acknowledged scholars (linguists, archaeologists, historians), who validated and improved the readings of Vékony. The results of these collaborations are published in English (Hosszú 2012), and in Hungarian (Hosszú 2013; Hosszú and Zelliger 2013; 2014a; 2014b). It should be emphasized that computational palaeography uses the results of humanities-based palaeography. The author utilised these results as accurately as possible.

The very close similarities between some of the scripts of the Eurasian Steppe are demonstrated as an application of the phenetic method; however, there is no category name for these scripts. In the literature, mostly the terms "runiform" or "Runic" are used, which are largely inappropriate, since these scripts are fundamentally different from the Runic script and its various versions (older fuþark, Anglo-Saxon runes, younger/Danish fuþark, short-twig runes, etc.). In Hungarian scientific literature, these scripts are have, for the last century, usually been called "rovásírások" 'Rovash scripts' (e.g., Sebestyén 1909; 1915). Therefore, the author collectively calls these scripts "Rovash." It is noteworthy that modifying the name of a script based on the research results is not unknown. For instance, the Anatolian Hieroglyphic script was earlier denoted as Luwian hieroglyphic, and even earlier Hittite Hieroglyphic (Payne 2010, 2; Yakubovich 2015a, 5). Another example is the Cypro-Greek script, which name was proposed by Egetmeyer in 2010 to replace the traditionally used "Cypriot Syllabary."

Table 3 presents some abbreviations and symbols used throughout the paper.

/	The alternative hypotheses are separated by a slash.
/ /	Double slashes denote phonemic transcription (denoting phonemes), phonemic representation of grapheme.
?	Question mark denotes the non-consensual transcription or phonetic value.
[ ]	Square brackets denote phonetic transcription (denoting allophones) using IPA (International Phonetics Association) symbols. The square bracket denotes the optional texts, too.
< >	Angled brackets are used for denoting transliteration value. In transliteration, the case that a consonant used before or after a sound is denoted by writing the transliteration value of that sound in superscript, e.g. <$^{\ddot{W}}$k$^{\ddot{W}}$>.
<A>	Transliteration value of Rovash graphemes with /a, ä/ phonemes.
<W>	Transliteration value of Rovash graphemes with /o, u/ phonemes.
<Ẅ>	Transliteration value of Rovash graphemes with /ö, ü/ phonemes.
↔	A part of a sound continuum; e.g., /c ↔ ʦ/ is a part of the continuum [k] > [kj] > [c] > [cç] > [cɕ] > [ʨ] > [ʧ] > [ʦ] > [s], or /ɟ ↔ ʣ/ is a part of the continuum [gj] > [ɟ] > [ɟʝ] > [ɟʒ] > [ʥ] > [ʤ] > [ʣ] > [z] (based on Valério 2016, 217, 256, 259).
AH	Anatolian hieroglyphic (Luwian / Luvian / Anatolian Hieroglyphic / hieroglyphs / syllabary / syllabic) script
AGA	Anatolian-Greek alphabetic scripts
C	Consonant
CBR	Carpathian Basin Rovash (Nagyszentmiklós, Tisza) script
CGk	Cypro-Greek (Valério 2016) script
CM	Cypro-Minoan script
CT	Characteristic Transformation (in a topological layer of the layered grapheme model)
Cypro-Greek	Cypro-Greek syllabary (Cypriot syllabic, Cypro-Syllabic, Classical Cypriot syllabary, Linear C). The term *syllabaire chypro-grec* was introduced by Egetmeyer (2010) and supported by Valério (2014).
dextrograde	Left-to-right writing (direction of writing)
E. Cyrillic	Early Cyrillic script
I. Aramaic	Imperial Aramaic (*Reichsaramäisch,* Official/Standard Aramaic) script
Lin. A	Linear A script
Lin. B	Linear B script

Madhabic	Instead of the earlier Minaic, Macdonald (2000, 68) recommended use of Madhabic.
NE-Iberian	Northeastern Iberian (Levantine Iberian) script
Old Aramaic	*Altaramäish,* Ancient/Early Aramaic script
ONA	Oasis North Arabian is a script group, its members: Dumaitic, Taymanitic, Dadanitic, and Dispersed North Arabian (Macdonald 2004, 490).
P.-Campanian	Proto-Campanian (*Protocampano, Paleoitalico, Nucerino alphabet*) script
P.-Canaanite	Proto-Canaanite script
P.-Hebrew	Palaeo-Hebrew script
P.-Hispanic	Palaeo-Hispanic (Palaeohispanic) script family
P.-Sinaitic	Proto-Sinaitic script
P.-Umbrian	Palaeo-Umbrian script
Proto-Rovash	The supposed common ancestor of the Rovash scripts (TR, SR, CBR, SHR), as hypothesized by the author.
S. Picene	South Picene script
S. Semitic	South Semitic script family
SE-Iberian	Southeastern Iberian (Meridional Iberian) script
SFG	Similarity Features Group
SHR	Székely-Hungarian Rovash (Székely, *Sekler,* [Old] Hungarian) script
sinistrograde	Right-to-left writing (direction of writing)
SR	Steppean Rovash (Khazarian Rovash, Don-Kuban-South-Yenissei-Ačïqtaš-Isfar, East European Runic Script). Note that the meanings of these scripts (the sets of inscriptions classified to each of them) partly differ from each other.
stiktogram	Punctuation mark (Karnava 1999, 37).
SW	Southwestern (Southwest, South Lusitanian, Tartessian, Bastulo-Tartessian, Southern Portugal) script
syllabogram	A grapheme that represents a syllable.
Th.	Thamudic is the tentative name of Ancient North Arabian scripts, which differs from the well-defined ONA, Safaitic, or Hismaic scripts. The Thamudic scripts are the following: Th. B, Th. C, Th. D, and Southern Th. (Macdonald 2004, 492).
TR	Turkic Rovash (Orkhon-Yenissei-Talas, [East/Old] Turkic runiform / Runic / "Runic") script. In this paper, the term *Turkic Rovash* is used instead of the more common *Turkic runic,* since this script is grouped together with the other Rovash scripts to avoid confusion with the fundamentally different Runic script.
V	Vowel

Table 3: Abbreviations, alternative names, and symbols

In representing graphemes, if the phonemic transcription (e.g., /b/) is obviously based on the transcription value (e.g., <b>), the phonemic transcription is usually not denoted, to simplify the description. Note that in some sources, only the transliteration values are available. Moreover, there are several differences between the sound values of the same grapheme in different sources. Therefore, the presented computational palaeographical analysis can further be made more accurate depending on the new results of the palaeography of the humanities.

The graphemes are identified by the script name and the transliteration value (e.g., AH <kar>, CGk <ko>, CM <ko?>, Greek <α>, NE-Iberian <ga/ka>) or the script name and the grapheme name (e.g., AH *315, CM 15). If the grapheme name includes the abbreviation of the script name (e.g., CM 15), the script name can be omitted. Usually, one or more typical glyphs are also included in the grapheme identification (e.g.: AH [glyph], [glyph], [glyph] *315 <kar>; NE-Iberian Ʌ <ga/ka>).

In the case of sequencing graphemes, usually the graphemes are separated by a semicolon (;). However, the repeated identical script names, glyphs, or transliteration values are omitted in order to save space. In the case of an omission, the glyphs are separated by a comma (,), e.g., "Carian 𐊠, Lydian 𐤠 <a>" is written instead of "Carian 𐊠 <a>; Lydian 𐤠 <a>." Another example: "Taymanitic, Hasaitic 𐪚 <y>" is written instead of "Taymanitic 𐪚 <y>; Hasaitic 𐪚 <y>".

The use period of the examined scripts (table 4) are mostly estimations due to the inaccuracies of dating archaeological relics, and since if in a certain period a script was written on perishable materials, no relic survived. In table 4, the use periods of closely related scripts are given collectively. The grouping of scripts is based on historical and phenetic features, and not on proved genealogical relationships. If the sound value of a grapheme of any script has not been proved, a question mark (?) denotes this fact, and such grapheme is omitted from the numerical analysis.

The Lin. A and CM scripts are still undeciphered; therefore the author omitted all of their graphemes from the numerical analysis. However, in the case of several signs, there is a consensus about their probable sound values (Valério 2016); therefore, several Lin. A or CM graphemes were included in the SFGs for information purposes.

The sources of the palaeographical data of graphemes and scripts are generally not detailed in table 10 due to the very large number of used glyphs. The author used the glyphs and other palaeographical statements from the sources table 5.

3 Method

3.1 The concept of the developed method

In computational palaeography, the variability of a grapheme could easily result in identical glyphs of unrelated graphemes; therefore, identical feature states could appear without any genealogical relationship (homoplasy). Consequently and significantly, the identity of a computational palaeographical feature state is generally

Groups	Estimated period of use of scripts
Aegean	*CGk*: 11th–2nd c. BC; *CM*: 17th/16th–11th c. BC; *Lin. A*: 18th–14th c. BC; *Lin. B*: 15th–13th c. BC
AH	*AH:* 17th–7th c. BC
AGA	*Carian:* 7th–3rd c. BC, *Greek:* 8th c. BC –, *Lemnian:* 6th c. BC; *Lycian:* 5th–4th c. BC, *Lydian:* 8th–3rd c. BC, *Phrygian* (archaic period only): 8th–4th c. BC, *Sidetic:* 5th–2nd c. BC
Ancient Italic	*Camunic, Elymian, Etruscan, Faliscan, Gallo-Etruscan, Gallo-Greek, Gallo-Latin, Latin, Lepontic, Messapic, Oscan, P.-Umbrian, P.-Campanian, Raetic, S. Picene, Umbrian, Venetic:* 8th c. BC – 1st c. AD
Ancient Semitic & Canaanite	*P.-Sinaitic & P.-Canaanite:* 19th (?) – 11th (?) c. BC; *Old Aramaic:* 925–700 BC; *Phoenician, P.-Hebrew, Samaritan:* 15th c. BC – 2nd c. AD
Aramaic & Persian	*Arabic:* 6th c. AD –; *Hatran:* 1st c. BC – 3rd c. AD; *Hebrew:* 3rd c. BC –; *I. Aramaic:* 700–200 BC, *Middle Persian:* 3rd–7th c. AD; *Nabataean* 2nd c. BC – 4th c. AD; *Palmyrene* 1st c. BC – 3rd c. AD; *Parthian:* 2nd c. BC – 3rd c. AD; *Sogdian:* 3rd–13th c. AD; *Syriac:* 1st c. AD –
Libyco-Berber	*Libyco-Berber:* 8th/7th c. BC – 7th c. AD
P.-Hispanic	*Celtiberian, NE-Iberian, SE-Iberian, SW:* 8th–1st c. BC
Rovash	*CBR:* 6th–11th/12th c. AD; *SHR:* 10th c. AD –; *SR:* 8th–12th c. AD; *TR*: 7th–10th c. AD
Runic	*Runic:* 2nd c. BC –
S. Semitic	*Ancient North Arabian (Taymanitic, Dadanitic, Dumaitic, Dispersed ONA, Safaitic, Hismaitic, Th. B, Th. C, Th. D, Southern Th.):* 8th c. BC – 4th c. AD; *Ancient South Arabian (Sabaic, Madhabic, Hasaitic):* 11th c. BC – 6th c. AD; *Geʿez abjad:* 8th c. BC – 4th c. AD
Slavic	*E. Cyrillic:* 10th c. AD –; *Glagolitic:* 9th c. AD –

Table 4: The groups of scripts and the estimated use periods of scripts based on surviving inscriptions

less well determined than in biology (e.g., gene sequence). In order to obtain identical feature states after filtering out the homoplasies, the linguistic, historical and geographical circumstances must be taken into account along with the topological similarities of the glyphs (table 9).

It is difficult to directly determine a script's genealogy in part due to the long examined period (generally from 2nd millennium BC to 1st millennium AD), and during this time frame scripts may have influenced each other on multiple occasions. Moreover, presently unknown scripts and orthographies may have existed that may also have influenced the examined scripts. However, by narrowing the focus of study to individual graphemes, connections might be determined. A slightly similar concept

Adiego (2007a; 2007b; 2007c; 2007d; 2007e; 2015), Anders (2012), Bakkum (2009), Benkő (1996a; 1996b), Beyer (1998), Bordreuil (2005), Brixhe and Lejeune (1984), Colless (2010), Correia (1996), Cross (1989), Daniels and Bright (1996), Davies and Olivier (2012), Davis (2010), Doblhofer (1962), Erdal (1993), Eska (2008), Farrujia de la Rosa et al. (2010), Faulmann (1880), Ferrer i Jané (2005; 2013; 2014), Gabain (1941), Garbini (1979), Gibson (1975), Grimme (1923), Hampel (1884), Hawkins (1986; 2000; 2010), Healey (1990), Hempl (1899), Hesperia (2005), Hoffmann (1987; 2011), Hosszú (2012; 2013), Hosszú and Zelliger (2013; 2014a; 2014b), Jeffery (1961), Jensen (1969), Kairzhanov (2014), Kalinka (1901 *apud* Adiego 2015), Kara (1996), Karali (2007), Kenyon (1899), King (1992), Konkobaev et al. (2015), Kononov (1980), Krings (1995), Kyzlasov (1994), LBI, Lemaire and Sass (2013), Looijenga (1997), Macdonald (2004; 2005; 2015), Marchesini (2009; 2012; 2014), MacKenzie (1971), Masson (1976; 1978), Mees (2006), Melchert (2004; 2008a; 2008b, 2008c), Miller (1994), MNAMON, Morandi (1982; 2004), NLR, Nollé (2001), O'Connor (1996), Olivier and Vandenabeele (2007), Olivier (2007-2008), Payne (2010), PROEL, Rilly and de Voogt (2012), Rodríguez Ramos (2004), Rogers (1999), Rollston (2008), Róna-Tas (1987), Rosenthal et al. (1986-2011), Röllig (1995), Sándor (1991), Sass (1988), Sebestyén (1915), Sims-Williams and Grenet (2007), Skjærvø (1996), Sprengling (1931), Swiggers (1996), Swiggers and Jenniges (1996), Taylor (1883), Tekin (2003), Thelegdi (1994/1598), Thompson (1912), Thomsen (1893), Tzanavari and Christidis (1995), Urbanová (2003), Valério (2008; 2013; 2016), Vékony (1985a; 1987a; 1992a; 1999a; 1999b; 2004), Wallace (2007), Weeden (2014), Woodard (1997; 2014), Woudhuizen (1982–1983; 1984–1985a; 1984-1985b), Yakubovich (2015), Young (1969), Younger (2000; 2003–2012).

Table 5: The sources of the palaeographical data

was proposed by Bernal (1990), who traced "isographs" of each grapheme instead of whole scripts.

The comparison of the glyphs of different scripts is supported by Boisson's stability principle, i.e. graphemes representing a sound existing in the acceptor language are adopted with their original glyph and sound value (Boisson 1994, 225 *apud* Adiego 2007e, 2).

Macdonald (2015, 10–12, 28–29) differentiated between literate and non-literate societies. He only considered a society literate if the written word was essential to its day-to-day functions. It is not necessary for the majority in the society to be able to read and write, but if a society had a written script and members who could read and write, but the skill was not used in everyday life, such a society is considered non-literate. According to Macdonald, most nomadic societies were non-literate, or mostly non-literate. An inscription in an illiterate society does not serve practical reasons, such as the majority of the Safaitic inscriptions on the boulders scattered in the desert. The reader is less important, so reader requirements do not affect the development of the script. Typically, writing is continuous with no word-dividers. Occasionally decorative variants were created in particular inscriptions, but had no consequences on the stability of the script itself.

Macdonald is aware that many societies cannot be strictly confined to a single category (literate or non-literate) but are in transition from one to the other. Even the scripts used exclusively by nomadic societies show development (i.e. new glyph variants becoming widespread), meaning that the written word must have had readers who were able to select the new alternative of a glyph. Applied to the Rovash scripts (used in the Altai Mountains, the Eurasian Steppe, and the Carpathian Basin; the majority of the surviving inscriptions are read in Turkic and Hungarian; see some examples of Rovash inscriptions in the Appendix), they must have been used largely in non-literate societies, as widespread modifications are largely absent. Thus Macdonald's model gives the basis to compare Mediterranean glyphs from the 1st millennium BC with Rovash scripts, which are attested only after the 6th c. AD (table 4). The only extant Rovash script, Székely-Hungarian Rovash (SHR), was used in the relatively isolated community of the Székelys (living in the mountains of Transylvania) up to the 17th c. AD. The Székelys used the Hungarian orthography of the Latin script for day-to-day functions; SHR was used as an unofficial writing system, and knowledge of SHR was passed almost exclusively from father to son. Consequently, in the case of the Székely-Hungarian Rovash (SHR) script, inscriptions made up to the 17th c. AD can be used for the present analysis.

3.2 Conversion of the palaeographical data into similarity features groups

In the analysis, the objects (taxa) are the scripts, and the features of a script are their graphemes and orthographic rules. The main properties of graphemes are the glyphs, especially their shapes. Other features of the scripts are their orthographical properties. Since these features are categorical variables, they are transcoded to Boolean indicators with the value being 1 if the feature is present in a script and 0 otherwise. The input data are given by the matrix $X(S, F_x)$ where S is the vector of scripts (objects, taxa) and F_x is the vector of features (presence of glyphs or orthographic rules). In this matrix a total of 66 scripts have been recorded with over 186 features. An illustrative example of the $X(S, F_x)$ matrix is given below as equation (9), which is generated from the similarity features groups (SFGs), where the features are the presence of glyphs or orthographical rules in a script (table 6).

$$X(S, F_x) = \left(\begin{array}{cc|cc|cc} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right) \tag{9}$$

Automatic comparison of glyphs is an inherently difficult problem because *(i)* related glyphs may be realised vastly differently in inscriptions due to differing

	Celtiberian	Elymian	Linear B	NE-Iberian	Raetic	SE-Iberian	SR	SW	TR
⩚ <a> (SFG-10)	0	1	0	0	1	0	0	0	0
⨝, X <b, bo, bu> (SFG-123)	1	0	0	1	0	1	1	1	0
ᛍ, ᑕ, ᐸ <ge/ke, k^2> (SFG-94)	1	0	0	1	0	0	1	0	1
⋀ <ga/ka, g> (SFG-89)	1	0	0	1	0	1	0	0	0
Ħ, ⧗ <pa_3, p^u> (SFG-150)	0	0	1	0	0	0	0	1	0
Boustrophedon in some relics (SFG-182)	0	0	0	1	0	0	0	1	1

Table 6: An illustrative example of $X^T(S, F_x)$ the transpose of data matrix (abbreviations in table 3)

calligraphic requirements, and *(ii)* unrelated glyphs may take very similar shapes when written. To describe the written variations of a glyph, the multilayer grapheme model has been developed, where each grapheme's visual identity has been determined and it was further assumed that in a script at any given period the glyphs may only diverge in so far as the common visual identity remains intact (table 8).

To handle the visual differences between glyph variants representing one grapheme, a typical set of transformations has been developed (table 6) that can describe how the shape of one glyph variant transformed into another. Using these transformations, it is easier to decide if certain symbols of the inscriptions are glyph variants of each other or not; however, it has to be noted that this is not a sufficient condition. Even with taking into account all available data, deciding on the relationship between two symbols in the same script or even in different scripts is uncertain. The more palaeographical data is taken into account, the less the uncertainty. The known glyph variants of one or more scripts are collected into a similarity features group (SFG). As the available palaeographical data (shapes of symbols in inscriptions, age of the inscriptions, published set of inscriptions, sound values of the graphemes, orthographical properties of the scripts, etc.) increases, the SFGs are split or restructured to fit with the new data. The more palaeographical data are analysed, the easier it is to build more realistic SFGs, and as a consequence, the size of the $X(S, F_x)$ matrix is usually increasing.

Macdonald (2015, 18–22, 24–26, 34–35, 40) criticised *comparative palaeography,* in which grapheme chains were stated as sequences of development. Among others, he cited the theories of Lidzbarski (1902, 122) and Praetorius (1904, 717–718). Lidzbarski proposed the genealogical relationship between the Phoenician 𐤀 <'>, Dadanitic ᚺ <'>, and Safaitic ᛯ <'>; Praetorius improved upon Lidzbarski's theory and suggested a sequence between Phoenician 𐤀, 𐤀 <'>, Safaitic 𐪑 <'>, and Dadanitic 𐪑 <'>. Macdonald (2015, 34) criticised Lidzbarski's opinion that there was a tendency towards modifying the irregular shapes of the North Semitic graphemes into symmetrical glyphs in both the South Semitic and the Greek scripts. Macdonald (2015, 35, 41) claimed that there is no evidence for any progressive development of the known South Semitic glyphs; and he presented examples of homoplasies, when a glyph in one script can develop a form similar to that of its equivalent in a different script.

Based on Macdonald's arguments, the present research is restricted to the collection of SFGs, and not the complete genealogical sequences, in order to minimize possible errors. However, there was evolution in glyphs; consequently, applying the results of the type of palaeography found in the humanities and in phylogenetics, a genealogical model must be achieved, too. Moreover, in the case of certain SFGs, there are glyphs that are obviously relatives; however, their shapes are slightly different. For instance, the Rovash X <χ> (SFG-101) and X <k^5> are cognates; however, the characteristic transformation *shortening lines* (CT-8, in table 7) can be applied to transform one to the other. It is worth noting that the presence of a characteristic transformation between two presumably cognate glyphs is only a supposition, and more palaeographical data could falsify or support its presence.

The objective in these examinations is to use the actual realised glyphs in inscriptions and not idealised glyphs. However, the conventions of the literature, which forms the basis of the $X(S, F_x)$ matrix, differ widely—many scientific articles publish the inscription only using idealised glyphs, while others publish faithful drawings of the inscriptions.

3.3 The developed exploratory data analysis

In order to explore the similarities of historical scripts (objects, taxa), a composite phenetic analysis method was developed, presented as the flow chart in figure 1, where X is the taxon-feature data matrix, Y is the taxon-feature data matrix transformed into 2-dimensional or 3-dimensional space, F_{MDS} is the vector of transformed features of the taxa, C is the matrix of cluster configuration, Z is a tree of hierarchical clusters, I is the vector of the cluster identifiers, and K is the actual number of clusters.

One result of the present research is the multilayer model of the graphemes, which was developed for modelling the grapheme in computational palaeography. The developed grapheme model is composed of four logical layers from bottom to top, namely the Topology, Visual Identity, Phonetic, and Semantic Layers. In the Topology Layer, a single glyph is described by a complete set of geometrical attributes. The Visual Identity Layer focuses on determining the possible unique identity of a writing symbol based on the human visual perspective in identifying an object. In this layer, the various glyphs of a single grapheme share some topological attributes in common.

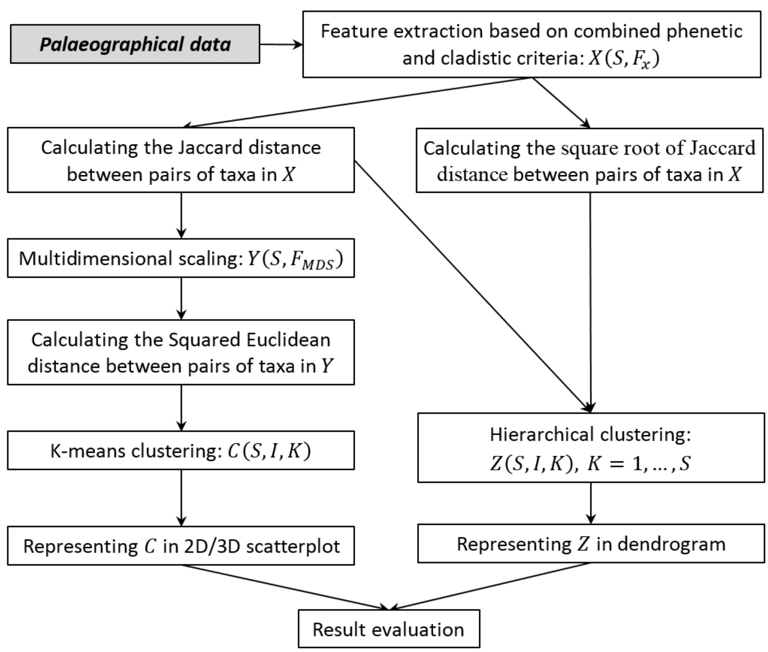


Figure 1: The flow chart of the developed method

The Phonetic Layer gives the sound values associated with the grapheme, and the Semantic Layer takes into account the context of the usage of the grapheme in the surviving and deciphered inscriptions (Pardede et al. 2016).

The Topology Layer of cognate, albeit slightly different glyphs can be transformed into each other by topological transformations called *characteristic transformations.* The characteristic transformation (CT) usually does not change the visual identity of the original glyph. Some examples of these transformations are listed in table 7, where there are references to SFGs in table 10.

Table 8 presents the four-layer grapheme model for the NE-Iberian <be> grapheme. This model helps to differentiate between the less important glyph variants and the significant altering graphemes. The CTs in table 7 are ideal geometric transformations in the topology layer of the grapheme model in table 8; however, the actual realization in the glyphs' evolution is unique in each case. It is noteworthy that on one hand, two glyphs are not necessarily relatives even if their shapes are identical or the difference can be covered by a CT; and on the other hand, the differences between cognate glyphs can usually be covered by CTs.

CT-1: *Bending or straightening*, e.g.: TR ⩓, ⌒, ⌂ <t^1> (SFG-29). Carian <, ⟩, C, Ɔ <d> (SFG-30). SHR ⨯, ⨳ <e> (SFG-68). NE-Iberian ⌶, ⨳ <ge/ke> (SFG-95). Libyco-Berber ○, □ <R> (SFG-160). Etruscan 8, ⊟ <fh> (SFG-166).

CT-2: *Closer-shape forming or vice versa,* e.g.: In the following two cases, Macdonald reconstructed the sequence of glyph development: Dadanitic ᚺ > ⩛ > ⩛ > ⩛ <'> (SFG-3), (Macdonald 2010, 12–14) and Dadanitic ᚺ > ⩛ > ⩛ > ⩛ <s^1> (SFG-142) (Macdonald 2010, 13–14). Other example: TR ⌒, ⌂, ⌂, ⌂, ⌂ <t^1> (SFG-29); TR ⋟, CBR ⟩, SHR Ɔ <W> (SFG-44).

CT-3: *Connecting,* e.g.: TR ⩛, ⩛ <g^1/$^1\gamma$> (SFG-103). CM ⩛ CM 56 (SFG-133), TR ⩛ <n^2>.

CT-4: *Cursivizing,* e.g.: Sabaic (early zabūr) 8, (middle zabūr) ⩛ < ṯ > (SFG-166).

CT-5: *Increasing or decreasing the number of repeating lines or curves,* e.g.: SHR ǂ, ⨍ <ḏ> (SFG-52). P.-Campanian ⩛, ⋙ <s> (SFG-58). SHR ≠, ⋕ <z> (SFG-59). P.-Sinaitic ⨳, ⨳ <ẖ> (SFG-68). S. Semitic ⊟, ⊟ <ḍ> (SFG-70). AH ⩛, ⩛ *315 <kar> (SFG-90). CBR ⩛, ⩛, ⩛ <z> (SFG-92). Runic ᛒ, ᛒ <b> (SFG-119). SW ⩛, ⩛, ⩛ <s> (SFG-168).

CT-6: *Line insertion or deletion,* e.g.: NE-Iberian H, H, H, H <o> (SFG-66); NE-Iberian ⩛, ⩛, ⩛ <be> (SFG-116).

CT-7: *Loop opening or vice versa,* e.g.: Dadanitic ⩛, ⩛ <ḏ> (SFG-55); NE-Iberian ⩛, ⩛ <be> (SFG-116).

CT-8: *Shortening of lines,* e.g.: SR ⧖ <k^5> /q/, SHR ⧖ <χ> /χ/ (SFG-101), the differences in the sound values are linguistically justifiable (Vékony 2004, 108–109). Sabaic zabūr (early) ⧖, (middle) ⩛ <z> (SFG-53).

CT-9: *Straight to curve or vice versa,* e.g.: SR ⧖ <k^5>, CBR 8 <q> (SFG-101). NE-Iberian ⩛, ⩛ <be> (SFG-116). Safaitic ⩛, ⩛, ⩛, ⩛, ⩛, ⩛, Hasaitic ⩛ <ṯ> (SFG-166) (Macdonald 2005, 82; 2015, 37). Greek (Corinth) ⩛, ⧖ <ε> (Swiggers 1996, 264). Greek (cursive, 6th c. AD) ß, ⩛, ⩛ <β> /b/ (Thompson 1912).

CT-10: *Turning ±90°,* e.g.: TR ⩛ <b^2> and ⩛ <m> (SFG-116). It is a typical Anatolian feature (Woudhuizen 1984–1985a, 92). Carian I, H <λ>. Carian Φ, ϴ <ś> (SFG-170). AH ◉, ◉, (Payne 2010, 14, 79), ◉ (Anders 2012) *412 <ru>.

CT-11: *Vertical mirroring,* e.g.: Old Aramaic 9, Greek ⩛ <b> (SFG-13); NE-Iberian ⩛, ⩛ <be> (SFG-116).

Table 7: Examples of characteristic transformations

4 Realization

The main goal of this paper is to demonstrate that the developed exploratory data analysis algorithm is applicable to processing palaeographical datasets and evaluating their statistical modelling. The realization of the method is presented below.

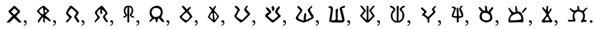
Layers	Example
Semantic	It was used in a semi-syllabic script in the period of 5^{th}–1^{st} c. BC in Northeastern Iberia (today Spain) and the Roman province Gallia Narbonensis (today France).
Phonetic	/be/
Visual Identity	A loop like shape with at least two legs up or down.
Topology	Glyph variants:  The applied topological transformations (citing CTs in table 7) are line insertion or deletion (CT-6), loop opening or vice versa (CT-7), straight to curve or vice versa (CT-9), and vertical mirroring (CT-11).

Table 8: Example of the four-layer grapheme model for the NE-Iberian <be> (SFG-116 in table 10)

C-1: The sound values of the graphemes are identical, or the difference is linguistically justified by acknowledged scholars.
C-2: Typologically the examined glyphs are identical or their difference is reasonable.
C-3: The historical and geographical facts prove or at least do not rule out the relationship between the scripts of the examined graphemes.
C-4: In the case of phonetically or topological differences, such SFG structure is chosen in which the supposed number of evolutionary changes is minimal.

Table 9: Criteria for constructing similarity features groups (SFGs)

4.1 Feature extraction based on similarity of glyphs and orthographical rules

In collecting the members of each SFG, the conjunctive criteria C-1 and C-2 (table 9) for the assumption of borrowing were considered. If both C-1 and C-2 are met, the appropriate glyphs are taken to be a member of a same SFG. This procedure is a phenetic analysis of the scripts. However, there are two problems: First, in the cases of several glyphs, based on the conditions above, a glyph could be classified into more than one SFG. Second, some glyphs that fulfill the criteria C-1 and C-2 could be homoplasies. Therefore, two further criteria were added in the analysis, C-3 and C-4 (table 9). The criterion C-3 helps to identify the homoplasies and to select such glyphs into separate SFGs. By using the criterion C-4, the developed combined method is governed by the *lex parsimoniae* (Ockham's razor) as is usual in cladistics.

When constructing the SFGs, the scientific literature has been taken into account, especially the dissertation of Valério (2016). This article thus primarily attempts to show the usefulness of phenetic modelling in developing the SFGs; phonetic and topological similarity is only used as a rough guide. In more specific palaeographical analyses, the method may be made more precise. While the SFGs used in this article are novel concepts and do not appear in the general literature in this form, similar approaches have been used, e.g., groups of presumably cognate graphemes in Valério (2016).

The explored SFGs are presented in table 10; the SFGs are sequence numbered. In each SFG, the topology of the glyphs or the orthographical rules is similar or identical. It is noteworthy that in the case of each grapheme, mainly those glyphs are listed that best fit in the appropriate SFG. The number of possible SFGs could be several hundred, but the set of SFGs is limited to those 186 that are the most significant. In the performed numerical analysis, only the presented SFGs are used. Naturally, more SFGs means more accurate results.

Analysing phonetic changes belongs to the palaeography of the humanities, and is outside the scope of computational palaeography. For example, SFG-91 is based on the combination of palaeographical and phonetic analysis (Valério 2016, 253–256). Similarly, SFG-92 is only a proposal, since it lacks palaeographical and linguistic evidence.

In each cell of table 10, first the members of the actual SFG are listed with their script names in italics. Following them are comments, which could contain further graphemes. However, the graphemes occurring only in comments are not included in the SFGs or any numerical analysis.

In general, the sources of individual glyphs in this article are not individually cited, since those may be found in the reference material (table 5). Furthermore, the name and age of the inscription in which a glyph is found is omitted for brevity, except in critical cases where these data are important for the analysis. In constructing the SFGs, in addition to the properties of the graphemes (glyphs, transliteration values, grapheme name, sound values), proposed relationships of different graphemes were obtained from palaeography publications from the humanities. Due to the large number of SFGs, in most cases, the author of the present paper could not detail all data from the scientific sources; therefore, publications used in constructing the SFGs are collected in table 5. Note that the goal of this paper is to present a developed phylogenetic procedure optimized for palaeographical data, and not to offer detailed palaeographical analyses (which is the task of humanities-type palaeography).

SFG-1: *P.-Sinaitic* 𓃾, 𓃾; 𓃾, *P.-Canaanite* 𐤀, 𐤀, *Phoenician* (Nora, ca. 900 BC) 𐤀, (Kilamuwa, ca. 825 BC) 𐤀, (Cyprus, ca. 880 BC) 𐤀, (Limassol, ca. 750 BC) 𐤀, 𐤀, 𐤀 <’>; *Phrygian* A, *Greek* 𐤀, Α, *Lemnian* 𐌀, *Carian* 𐊠, A, *Lydian* 𐤠, *Elymian* A, *S. Picene* A, *Etruscan* 𐌀, 𐌀, *Raetic* 𐌀, 𐌀, *Faliscan* 𐌀, *Venetic* 𐌀, *Messapic* A, *Lepontic* 𐌀, *P.-Umbrian* 𐌀, *Umbrian* 𐌀, 𐌀, *Oscan* 𐌀, *Latin* A, A, *Gallo-Greek* A, *SW* A, A, *E. Cyrillic* А, A <a>

SFG-2: *Greek (medieval cursive)* ✢ (Faulmann 1880, 171) <α>; *Glagolitic* Ⰰ *azъ* <a>

SFG-3: *Sabaic* 𐩱, *Dispersed ONA, Dumaitic, Taymanitic, Th. B, Th. D* 𐩱, *Dadanitic* 𐪑, 𐪑, 𐪑, 𐪑 <’> /ʔ/. *(i)* Glyph evolution CT-2.

SFG-4: *Phoenician* (al-Khader, 11^{th} c. BC) 𐤀, (Amurru, 11^{th} c. BC) 𐤀, (Aḥīrām,ca. 1000 BC) 𐤀, 𐤀, (Jeḥīmilk, ca. 950 BC) 𐤀, (Elībaal, ca. 900 BC) 𐤀, (Tekke, ca. 900 BC) 𐤀, *Hismaitic* 𐪑, 𐪑, 𐪑, *Safaitic* 𐪑, 𐪑, 𐪑, 𐪑 <’> /ʔ/; *SHR* 𐲀 <ë> /ä, e, ē/, 𐲀, 𐲀 <ö> /ö, ő/; *TR* 𐲀 <A> /a, ä/

SFG-5: *Phoenician* (*Punic,* Motya, mid-6^{th} c. BC) 𐤀, 𐤀 (Röllig 1995, 210–211) <’> /ʔ/ *SW* ⍑, ⍑, ⍑, (Espanca) ⍑ <o> /o/; *SE-Iberian* ⍑, ⍑, ⍑ (Rodríguez Ramos 2004, 99) <o> /o/. *(i)* Rodríguez Ramos proposed that the P.-Hispanic ⍑ <o> originated from the Phoenician <’> (2002, 192).

SFG-6: *Messapic* Α, Α, *Oscan* Α, *Gallo-Greek* Α <a>

SFG-7: *Greek* (not later than 5^{th} c. BC) Α, *Lycian* 𐊀, (TL 5) 𐊀, (TL 33) 𐊀, 𐊀 (Kalinka 1901 *apud* Adiego 2015, 20–21), *S. Picene* 𐌀, *SE-Iberian* 𐌀, *NE-Iberian* 𐌀, 𐌀, *Celtiberian* 𐌀, *SHR* 𐌀 <a>. *(i)* The similarity of the shapes of the AH 𐘀, 𐘀, 𐘀 *19 <á> and the 𐌀, 𐌀 <a> has not been clarified.

SFG-8: *NE-Iberian, Celtiberian* ᐅ <a>; *SHR* ᐊ <a>

SFG-9: *Carian* 𐊠, *SE-Iberian* 𐌀, *Elymian* Α, *Latin (epigraphic cursive)* Λ, (*cursive majuscule,* Pompeii) ⋋, *Raetic* Α, Α, *Lepontic* Α, 𐌀, *Gallo-Etruscan* 𐌀, 𐌀, *Camunic* Α, *Runic (older fuþark)* ᚨ; ᚨ <a>. *(i)* The Runic ᚨ could be an autapomorphy.

SFG-10: *Elymian* Α, *Raetic* Α, Α <a>

SFG-11: *Parthian* 𐭀 <’> /a, ā/; *Sogdian* 𐼰, 𐼰 <’> /a, ā, ə/; *Syriac* ܐ <’>

SFG-12: *AH* 𐘀, 𐘀, 𐘀 *19 <á>; *Sidetic* 𐤀, 𐤀, 𐤀, 𐤀, 𐤀 <a> /a/; *TR* 𐲀, 𐲀, 𐲀, (manuscripts) 𐲀 <A> /a, ä/. *(i)* The possible relationship between SFG-11 and SFG-12 is unclear. The TR 𐲀 <A> could belong to SFG-11 and not SFG-12. It is noteworthy that there are some interesting, but maybe unrelated, orthographical features as follows. (a) The TR 𐲀, 𐲀 <A> also used as word separator. (b) According to Younger, the Cretan Hieroglyphic ⌇ (Younger 2003-2012) is a phrase termination; however, Karnava (1999) handles this as a syllabogram and not a stiktogram. (c) The AH 𐘀 *450 <a> was also used as a word ending mark (Payne 2010, 81); however, it differs from the AH *19 <á> (SFG-12).

SFG-13: *P.-Canaanite* 𐤁 <b>; *Phoenician* (Byblos, 11^{th}–10^{th} c. BC) 𐤁 <b>; *P.-Hebrew* (late 8^{th} c. BC) 𐤁 <b>; *Old Aramaic* (Zinjîrlû, late 9^{th}–8^{th} c. BC) 𐤁, (8^{th} c. BC) 𐤁, (8^{th}–7^{th} c. BC) 𐤁 <b>; *SW* 𐤁; 𐤁, 𐤁, 𐤁 <p^e> /p/; *NE-Iberian* 𐤁, 𐤁 <bi>; *Celtiberian* 𐤁, 𐤁, 𐤁, 𐤁, 𐤁 <bi>; *Greek* 𐤁, 𐤁, 𐤁, (Naxos, 8^{th}–7^{th} c. BC) 𐤁, (Argos) 𐤁 <β>; *Lycian* 𐊂 <b>; *Parthian* 𐭁, 𐭁 <b> /b, u̯/; *Sogdian* 𐼱, 𐼱 <b> /b, β/; *Hatran* 𐣡 <b>; *Syriac* ܒ <b, ḇ>; *SR* 𐰉, 𐰉 <b^1> /b, β/ (Vékony 2004, 315); *TR* 𐲂, 𐲂, 𐲂, 𐲂, 𐲂, 𐲂, 𐲂, 𐲂, 𐲂 <b^1> /b/. *(i)* Glyph evolution: CT-11. *(ii)* For the possible evolution of the Greek 𐤁 <b> cf SFG-119.

SFG-14: *P.-Sinaitic* ⌙, *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) Γ, *Phoenician* 𐤂, 𐤂, 𐤂, Λ <g>; *Old Aramaic* 𐤂, Λ, *Greek* 𐤂, 𐤂, Λ, 〉, Γ, *Elymian* Γ <γ>; *P.-Umbrian* (Tolfa, ca. 530–525 BC) ∩ <c> (Urbanová 2003, 33; Bakkum 2009, 380); *Phrygian* Γ, ꓶ <g>; *Etruscan* 𐌂 <$c^{e, i}$> /k/; *Messapic* Γ, *Oscan* Γ, *Gallo-Greek* Γ, *I. Aramaic* Λ, 𐡂, *Parthian* 𐭂, *Middle Persian* 𐭢, *Hatran* 𐣢, *E. Cyrillic* Γ <g>

SFG-15: *Greek* <, ☾, *Elymian* < <γ>; *Lydian* 𐤢, (<g>; *Etruscan* 〉 <$c^{e, i}$> /k/; *Faliscan* (, < <c>; *S. Picene* < <c/g>; *Oscan* >,), C, G, G, *Camunic* <, 〉 <g>; *Umbrian* C <c/k> [k]; *Umbrian* *(late)* 𐌂 <g>; *Runic (older fuþark)* ᚲ, › <k>; *Latin (archaic)* C <c/g>; *Latin (classical)* G <g>; *NE-Iberian* ᐸ, (<ge/ke>

SFG-16: *I. Aramaic* 𐡂, *Hebrew* ג, ג, *Parthian* 𐭂, *Palmyrene* 𐡢, *Nabataean* 𐢄 <g>; *Sogdian* (earlier than 4^{th} c. AD) 𐼂 (Sims-Williams and Grenet 2007), (Ancient Letters) 𐼂 (Skjærvø 1996, 519), *(Manichean)* 𐫃, *(Christian)* ܓ <g, γ>; *Syriac* ܓ, ܓ, *Arabic* ج <ǧ>

SFG-17: *Greek* (*early minuscular*, 9^{th} c. AD) ϒ <γ> (Taylor 1883, 154); *Galgolitic* Ⰳ, Ⰳ <g>

SFG-18: *Lycian* 𐊄, 𐊄, (TL 5) 𐊄 (Kalinka 1901 *apud* Adiego 2015, 21) <g> /γ/; *TR* 𐲃, 𐲃, 𐲃, (manuscript) 𐲃, 𐲃 <ṅ> /η/

SFG-19: *Madhabic* (Dadan) 𐪕, 𐪕, *Sabaic, Hasaitic, Dispersed ONA, Taymanitic, Dadanitic* 𐩵, *Dumaitic* ᐅ, *Taymanitic, Th. B* 𐪕, *Safaitic* 𐪕, ᑭ (Macdonald 2015, 37), *Hismaitic* ᕮ (Macdonald 2005, 82), 𐪕, ᗱ, *Th. C, Th. D* ⏏, *Geʿez abjad* ደ, ፻ <d> /d/

SFG-20: *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) ᗜ (Cross 1989, 82) <d>; *Phoenician* 𐤃, 𐤃 (Sprengling 1931, 55), (Byblos, 11^{th}–10^{th} c. BC) 𐤃 <d>; *P.-Hebrew* (late 8^{th} c. BC) 𐤃 <d>; *Old Aramaic* (10^{th}–9^{th} c. BC) 𐡃, 𐡃, 𐡃, 𐡃, (8^{th} c. BC) 𐡃; (Deir ʾAllā, around 800 BC) 𐡃; (8^{th}–7^{th} c. BC) 𐡃 <d>; *Greek* ᐅ <δ>; *Phrygian* Δ (Adiego 2007e, 3) <d>; *Lycian* Δ (Adiego 2007e, 8) <d> /d/; *Faliscan* ᐅ, *Elymian* (5^{th} c. BC) Δ, *S. Picene* ᐅ; *Oscan* Δ, *Messapic* ᐅ, Δ, Δ <δ, d> [d]; *Gallo-Greek* Δ, Δ <d> /d, t/; *SW* Δ <t^u> /t/; *SE-Iberian* Δ <tu> /tu/; *NE-Iberian* Δ <du/tu> /du, tu/; *Celtiberian* (Botorrita, Spain) Δ (Eska 2008, 166–167), Δ, Δ <tu>

SFG-21: *Greek* D <δ>; *Etruscan* (Marsiliana d'Albegna, 8^{th} c. BC) ᗡ, ᗡ, (Veias, Caere, 7^{th} c. BC) D <d> /t/; *Latin, Faliscan, Elymian, Messapic, Umbrian, Oscan* D <d, δ> [d]; *(i)* The Runic (older fuþark) ᚦ, (Jutland, ca. AD 160–350) D, ᚦ (Looijenga 1997, 82–83) <Þ> may belongs to SFG-21.

SFG-22: *Greek (medieval cursive)* δ (Faulmann 1880, 171) <δ>; *E. Cyrillic* Д *dobro* <d>

SFG-23: *Greek (medieval cursive)* δ <δ>; *Glagolitic* Ⰴ <d>

SFG-24: *Lin. A* 𐘀 LA 01 <da>; *Lin. B* 𐀅 <da> /da/; *CM* ⊦, ⊦ CM 04 <ta?> (Valério 2016, 428); *CGk* (Common) 𐠫, (Paphian) ⊦ (Olivier 2008, 617–618), ⊢ <ta> /da, ta/; *Lydian* 𐤣, 𐤣 <d>. *(i)* Cf SFG-176. *(ii)* Cf Sidetic 7 <t>.

SFG-25: *Lin. A* 𐘅 (Valério 2013, 15–17) LA 05 <to?>; *Lin. B* 𐀵 <to> /to, t^ho/; *CM* ⊼ CM 13 / ⊼ CM 78 <to?> (Valério 2016, 111–112, 430); *CGk* (ICS 172, *early*) ⊼ (Valério 2016, 237), *(Common)* 𐠯; (Davis 2012, 38–61) *(Common)* 𐠯, *(Paphian)* ⊼; (Olivier 2008, 617–618) (*Paphian*, 6^{th} c. BC) ⊼, ⊼, ⊼ (Valério 2016, 228) <to> /do, to/; *CGk* 𐠰 (Valério 2016, 230), *(Common)* 𐠰 (Davis 2012, 38–61) 𐠰, (*Paphian, late*) 𐠰 (Olivier 2008, 617–618) (*Paphian*, 6^{th} c. BC) 𐠰, 𐠰 (Valério 2016, 228) <tu> /du, tu/. *(i)* The SFG-25 and SFG-26 are likely relatives. *(ii)* The SFG-25 and SFG-30 may be relatives.

SFG-26: *NE-Iberian* Ш, Ѱ, ᐯ <do/to>; *Celtiberian* (Botorrita, Spain) ᗯ, Ѱ <to>. *(i)* Cf SFG-25.

SFG-27: *SW* ⩓, ⩓ <t^o/$t^{o/u}$?> /t/; *SE-Iberian* Δ<tu> /tu/; *NE-Iberian* Δ, ⩓, Δ, Δ <du/tu>; *Celtiberian* Δ, Δ <tu>. *(i)* The SFG-27, SFG-28, and SFG-29 are likely relatives.

SFG-28: *SE-Iberian* ⩚ <tu> /tu/; *Celtiberian* ⩚, ⩚ <tu> /tu/. *(i)* Cf SFG-27 and SFG-29.

SFG-29: *Carian* 𐊼, 𐊼 <δ> /<u>md</u>/d/^{n}t/; *SW* ⩔ <t^o>; *Libyco-Berber* ⵡ; ⵟ <T_3>; *TR* ⩓, ⩔, ⌒, ⌒, ô, ⌒, ⌒, ⌒, ⌒, δ, ⌒ <t^1>. *(i)* The sound value /^{n}t/ of the Carian <δ> is supported by Kloekhorst (2008, 138–139). *(ii)* For the development of the TR glyphs, see CT-2 and CT-1. *(iii)* Cf SFG-27 and SFG-28.

SFG-30: *Greek* (Crete) Λ <δ>; *Phrygian* Λ <d>; *Libyco-Berber* ⊓, ⊏, ⊂, ‹ <D>; *Sidetic* ᑌ <d>;

Carian <, 〉, C, Ɔ <d> /d/; *Celtiberian* Λ <tu>; *CBR* > <d> /d/ (Table 15 in Appendix); *SR* >, ⊁ <d> /d, δ, ǰ/ (Vékony 2004, 243, 251, 267, 287, 294); *SR* ⋂ <d> /d, ǰ/ (Vékony 2004, 253, 264, 287, 294). *(i)* Glyph evolution: CT-1. *(ii)* Cf SFG-25.

SFG-31: *P.-Sinaitic* (Serabit al Hadim, early 15th c. BC) 𐤄, ꀀ, Ψ, Ψ; (Sinai 358) ⊀ (Colless 2010, 96) <h>; *AH* ⅄ (Hawkins 1986, 370–371) *451 <hur>; *Messapic* ⇡, Ұ, Υ <h?>; *SR* ⅄, ⅄ <h> /h/ (Vékony 2004, 287, 294); *CBR* ⅄, ⅄, ⩚ <χ> /χ/ (Vékony 2004, 151)

SFG-32: *Madhabic* (Dadan) Y, *Sabaic* Ψ, Y, *Hasaitic* Y, *Dumaitic* V, *Taymanitic* V, Y, *Dadanitic* ⅄, ⅄, Λ, *Th. D* Y, ≺, *Th. C* ≺, *Th. B* ⅄, Y, *Hismaitic* Y (Macdonald 2005, 82), ⅄, ⅄ (King 1992, Figure 1 between pages 5 and 6), Y, *Safaitic* ⅄, Ꮧ, ୮, ⅄; Y, Y, ⅄, *Geʿez abjad* V <h> /h/; *SR* Ꮧ, Y, ⅄, ⅄, Ψ <A> /a, ā, ä/ (Vékony 2004, 314); *CBR* Ꮧ, Y, Y, Y <A> /ȧ, a, ä, e/ (Table 15 in Appendix; Vékony 2004, 164, 185). *(i)* For the sound values see the comment in SFG-68. Moreover, in the Old Aramaic, the word-end <-'> and <-h> represented /-ā/. In the 10th–9th c., in the Old Aramaic the <-h> denoted /-ā/-t and /-ē/ (Segert 1978, 112–113). In the P.-Hebrew, the <h> denoted the word-ending /o/, /a/ or /e/ (Healey 1990, 35).

SFG-33: *Phoenician* (Inscription of King Kilamuwa, Zincirli, ca. 825 BC) ꟻ <h>; *NE-Iberian* ⋲, ⋲ <e> /e/; *SR* ∋, ∋ <e> /ä, e/ (Vékony 2004, 314); *SHR* ∋, ∋ <e> /ä, ē/. *(i)* Cf SFG-38. *(ii)* The NE-Iberian ⋲, ⋲ <e> could be a direct variant of the NE-Iberian ⊬ <e> (SFG-39) and not a direct descendant of the Phoenician ꟻ <h>.

SFG-34: *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) E, *Phoenician* (Byblos, 11th–10th c. BC) ꟻ, ꟻ, *P.-Hebrew* (end of 8th c. BC) ꟻ <h>; *Phrygian* ꟻ, E, E, *Greek* (Athens, 8th–7th c. BC) ꟻ, (Corinth) ꟻ, ꟻ, *Etruscan* ꟻ, ꟻ, *Faliscan* E, *S. Picene* E, *Lemnian* ꟻ, *Messapic* E, E, *Venetic* ꟻ, ꟻ, *Camunic* ꟻ, *Elymian* E, E, *Raetic* ꟻ, ꟻ, ꟻ, *Lepontic* ꟻ, ꟻ, *Gallo-Etruscan* E, ꟻ, *Oscan* ꟻ, *Umbrian* ꟻ, ꟻ, ꟻ, *Latin* (*archaic*) E <e, ε> /e/

SFG-35: *Greek* (8th–7th c. BC) Ǝ, Ǝ, Ǝ <ε> /e/, *Lycian* E <i> /i/; *Lydian* Ǝ, *P.-Campanian* (Sorrento) Ǝ, *S. Picene* E, Ǝ, *Oscan* Ǝ, Ǝ, E, *Elymian, Faliscan, Gallo-Greek, Umbrian, Latin (classical)* E, *Messapic* E, E <e, ε> /e/

SFG-36: *Phrygian* ꟻ, ꟻ (Young 1969, 262–268) <e> /e/; *Lydian* ꟻ, Ǝ, ☰ (Adiego 2007e, 7) <e> /e/ [e:]; *Camunic* ꟻ (Morandi 2004, 476) <e>; *NE-Iberian* ꟻ, ꟻ, ꟻ <e> /e/; *SR* ꟻ <e> /ä, e/ (Vékony 2004, 314). *(i)* Presumably, the SW ꟻ <h/H?> is also relative of the graphemes in SFG-36.

SFG-37: *Greek* (before 280 BC) ϵ, ϵ; (*minuscular,* 10th–11th c.) ϵ; *(medieval cursive)* ϵ; *Oscan* ∈ <ε, e>; *Messapic* ∈, *Umbrian* ꟻ, *Gallo-Greek* ∈, ∈, *Glagolitic* Ⰵ <e>, *E. Cyrillic* Э <e>

SFG-38: *Greek* (*cursive*, AD 701–718) ∉, ∉ <ε>; *Glagolitic* (Codex Zographensis, 10th–11th c.) Ⰵ, Ⰵ <e>. *(i)* The similarity between the Glagolitic Ⰵ, Ⰵ <e> and the SHR ∋ <e> (SFG-33) is maybe a homoplasy.

SFG-39: *Phoenician* (Sarepta, ca. 725 BC) ꟻ <h>; *Greek* ꟻ <ε>; *Lydian* ꟻ, F, ꟻ <e>; *NE-Iberian* ⊬, ⊬, ⊬, ꟻ <e> /e/; *Celtiberian* ⊬ <e>; *SHR* ꟻ <e> /ä, ē/

SFG-40: *P.-Umbrian* (Tolfa, ca. 530–525 BC) ꟽ <e> (Urbanová 2003, 33; Bakkum 2009, 380); *P.-Campanian* (Nuceria, second half of the 6th c. BC) ꟽ <e> (MNAMON)

SFG-41: *Phoenician* (ca. 900 BC) Y, Ꮧ, *Old Aramaic* (Zinjîrlû, late 9th–8th c. BC) Ꮧ, (8th c. BC) Ꮧ, (Deir 'Allā, ca. 800 BC) Ꮧ, (8th c. BC) Ꮧ, Ꮧ <w>; *SW* ⊬, ⊬, (Espanca) Ꮧ, *SE-Iberian* Ꮧ <u> /u/. *(i)* Cf SFG-44. *(ii)* Cf SFG-45.

SFG-42: *Greek* ꟻ, ꟻ <ϝ> /u̯/ [w]; *Phrygian* F, ꟻ <v> /w/; *Lycian* F <w> /w/; *Lydian* ꟻ <v> /v/; *Lemnian* ꟻ, ꟻ <v>; *Etruscan* ꟻ <v> /β, u̯/ [β]; *Raetic* ꟻ, ꟻ <v> /v/; *Messapic* F, F, *Venetic* ꟻ,

Lepontic ꟻ, ᖷ <v>; *Umbrian* ꟻ, *Oscan* ᖷ <v> [w]; *Umbrian, Oscan* F, *Latin* ᚠ, F <f> [f]; *Runic (older fuþark)* ᛉ, ᚠ, ᚡ <f> /f/; *CBR* ꟻ (Table 15 in Appendix; Hosszú and Zelliger 2014a, 186), *SR* ꟻ <β> /β, v/ (Vékony 2004, 314)

SFG-43: *Greek* (8th–7th c. BC) Ⴤ, Ⴤ, Y, Ⴤ, Y, Y, Ⴤ, Ⴤ, Y; Ⴤ <υ> /u, ū/; *Greek (classical)* Y <υ> /u, ü/; *Phrygian* ᚴ, *Sidetic* Y, *Lydian* Ⴤ, *Carian* Y, *Etruscan* (7th c. BC) ᚴ, *Elymian* Y, *Messapic* Y, *Oscan* Y, *Gallo-Greek* Y, Y, ᚴ, ᚴ, Ⴤ, Y, Y <u, υ> /u/; *Latin* (*archaic,* 4th–2nd c. BC) Y <v>

SFG-44: *I. Aramaic* (7th c. BC) ᚴ, ᚴ, ᚴ, ᚴ, ᚴ; (6th–4th. c. BC) ᚴ, ᚴ, ᚴ, ᚴ, (Aśoka, around 250 c. BC) ᚴ, *Parthian* (Nisa, 1st c. BC) ᚴ <w> /u̯, ō, ū/; *Hebrew* (1st c. BC) ו <w>; *Hatran* ᚴ <w>; *TR* 〉, ᚴ, ᚴ, ᚴ, *CBR* ᚴ (Vékony 2004, 151; Hosszú and Zelliger 2014a, 188), *SHR* Ͻ <W> /o, u/. *(i)* The possible ancestor of the Rovash 〉 glyph is attested in the Aramaic script in 7th–3rd c. BC. The more characteristic Rovash ᚴ, ᚴ glyphs are attested from a narrower period, the 7th c. BC. Consequently, Rovash most probably borrowed the Aramaic <w> in that time. The Rovash Ͻ was probably derived from ᚴ–like glyph by turning the short bars to obtain a closer shape (CT-2). *(ii)* Cf SFG-45.

SFG-45: *TR* ᚴ, ᚴ, ᚴ, *SHR* ᚴ, ⋧ <Ẅ> /ö, ü/. *(i)* Sebestyén (1915, 158) argued that the TR <Ẅ> is a descendant of the Greek (classical) Y <υ> /u, ü/ (SFG-43); *(ii)* In some Semitic scripts (e.g., Uyghur and Sogdian), the /ö, ü/ are represented by the ligature of the <y> and <w> (Erdal 2004, 42). Sims-Williams (1981, 359; 1989, 181; 1996, 313–314) demonstrated the Sogdian tradition of representing front rounded vowels *(ö, ü)* by the combination of <w> and <y>. Supposing the influence of the Sogdian script, the Rovash <Ẅ> could have been constructed of the Rovash ᚴ, ᚴ <i, y> (SFG-81) and 〉, ᚴ <W> (SFG-44) as follows: ᚴ < ᚴ + 〉; ᚴ < ᚴ + ᚴ; ᚴ < ᚴ + 〉; however, there is not direct evidence for this ligature-based evolution of the Rovash <Ẅ>. *(iii)* Erdal (2016) discovered use of the graphemes <o> for /ö/ and <u> for /ü/ as demonstrated in a Turkic text written with Brāhmī script in the IOL Toch 81 inscription (Maue 2008). According to the present author, a possible ancestor of the TR ᚴ, ᚴ, ᚴ <Ẅ> could be the Greek Ⴤ, Ⴤ <υ> /u, ū/, which could be used for representing /ö, ü/ as happened in the Brāhmī script; the glyph variants of the TR ᚴ, ᚴ, ᚴ <Ẅ> can be easily derived from the Greek Ⴤ, Ⴤ <υ>. In this solution, either the ligature forming or the /ü/ sound value of borrowed Greek Y <υ> have not to be assumed; therefore, based on *lex parsimoniae*, this lineage is the most probable. It is noteworthy that the glyphs of the SW ᚺ, ᚺ <u> (SFG-41) and the TR ᚴ, ᚴ <Ẅ> are very similar to each other. *(iv)* The Rovash ᚴ and ⋧ <Ẅ> are presumably variants of ᚴ <Ẅ>.

SFG-46: *P.-Sinaitic* ⊸, ♀ <w>; *AH* ᚴ *280 <wa/i_9>. *(i)* The existence of this SFG is very tentative.

SFG-47: *Sabaic, Dispersed ONA, Taymanitic, Th. B, Hasaitic* ♀, *Dumaitic, Dadanitic* ♀, ♀, *Hismaitic* ♀, ᑫ, ♀, *Safaitic* ♀, ᚴ <y> /y/ [ç]; *Carian* ᚴ <ý> [ɥ]; *TR* ʔ, ᑫ, 9 <y^1> /y/ [j]; *TR* ♀, 9, ᑭ, ᑫ <y^2> /y/ [j]

SFG-48: *P.-Umbrian* (Tolfa, ca. 530–525 BC) ↑ <f>; *Faliscan* ↑ <f>. *(i)* They may be relatives of the CGk (Paphian, 6th c. BC) ↑, ↑ (Valério 2016, 228), (Common) ↑ (Valério 2016, 230) <wo>. *(ii)* Cf SFG-49.

SFG-49: *NE-Iberian* ↑, ⩚, ⋒, ⁞, *Celtiberian* ↑ <u>; *S. Picene* ⩚ <ú> [uː]. *(i)* Cf SFG-48.

SFG-50: *P.-Sinaitic* ═, ═ <z/ḏ?>; *Phoenician* ═, ═, I <z> /d͡z/; *P.-Hebrew* ═ <z>; *Old Aramaic* I, Z, I <z>; *Greek* (Crete, 8th-7th c. BC) I <ζ> /ds, sd/, /zd/ or /dz/ [d͡ʒ/d͡z]; *Oscan* I <z> [z, t͡s, d͡z]; *Libyco-Berber* ᚺ, ⊢⊣, I <Z_1>; *Dadanitic* ∀, ∀, ⊢⊣, *Taymanitic, Th. D* ⊢⊣, *Th. C,* *Hismaitic* Ⲓ, ⊢⊣, *Th. B* T, *Safaitic* T, ⫰, *Geʿez abjad* H <z> /z/

SFG-51: *Greek* (Athens, 8th–7th c. BC) I <ζ> /ds, sd/, /zd/ or /dz/ [d͡ʒ/d͡z]; *Sidetic* \ <z>; *Libyco-Berber* —, I <Z_2>; *I. Aramaic* (Aśoka, around 250 c. BC) \ <z>; *Parthian* (Nisa, 1st c. BC) I <z> /z, ž/; *Hebrew* (Qumran, 1st c. BC) I <z>; *Hatran* (shortly before AD 240) I <z>; *Sogdian* (earlier than 4th c. AD) \ <z>; *Syriac* (1st c. AD) \ <z>

SFG-52: *Phoenician* (Limassol [Cyprus], ca. 750 BC) ‡ <z>; *Greek* ⱡ, ǂ <ζ> /ds, sd/, /zd/ or /dz/ [d͡ʒ/d͡z]; *Etruscan* ⱡ <z> /ts/; *Faliscan* ⱡ, *Raetic* ǂ, *Lepontic* ‡, ‡, *Venetic* ✗ <z>; *Umbrian* ǂ, ǂ, ⱡ <z> [t͡s]; *SHR* ⱡ, ǂ, ⨍, ⫽ <ď> /ɟ/. *(i)* After 10th c. AD, in the Hungarian language, there was a /d͡ʒʲ/ > /ɟ/ change. Presumably, the origin of the Rovash ⱡ, ǂ, ⫽ <ď> is the Greek ǂ <ζ>, and its ancestor is the P.-Sinaitic ═ <z/ḏ?> (SFG-50); however, the similarity between the Rovash ⫽ <ď> and the P.-Sinaitic ═ <z/ḏ?> is surely a homoplasy. *(ii)* Glyph evolution: CT-5.

SFG-53: *P.-Canaanite* ⧗, ⧖ <z>; *Madhabic* (JSMin 24) ⧗, *Sabaic (early musnad)* ⧗, *(early zabūr)* ⧗, *(middle musnad)* ⧗, *(middle zabūr)* ⧗ <z> (Macdonald 2015, 37, 39); *TR* (Yar Khoto grafitti no. 21) ⧗ (Erdal 1993, 91, 104–105); ⧗, ⧗, ⧗, ⧗, (Mendur-Sokkon IV) ⧗ (Konkobaev et al. 2015, 41) <z> (Kairzhanov 2014, 18). *(i)* Presumably, a ⧗ > ⧗, ⧗ shape transformation happened in TR, similarly to the ⧗ > ⧗ shape transformation (CT-8) in the Sabaic script. *(ii)* Cf SFG-59.

SFG-54: *CM* ⧗, ⧗, ⋊, ⧗, ⧗, ⧗ CM 107, ⧗, ⧗ CM 108 <za_2?/zo_2?/zi?>; *CGk* ⧗, ⧗, ⧗, ⧗, ⧗ <za?> /ɟja ↔ ʣa/? (based on Valério 2016, 227); *Carian*)(, ⧗, ⧗, (coins) ⧗, ⧗ <z> /sd/. *(i)* Cf SFG-59.

SFG-55: *Madhabic, Sabaic* H; *Hasaitic* H; *Dadanitic* ⧗, ⧗, ⧗, ⧗ (Macdonald 2010, 13–14), *Dumaitic* ⊟, *Taymanitic* H, *Th. B* ⧗, ⧗, ⧗ <ḏ> /ḏ/ [ð]. *(i)* Cf SFG-59. *(ii)* Glyph evolution: CT-7.

SFG-56: *Th. C* ⧗, *Hismaitic* ⧗, ⧗, ⧗, ⧗, *Safaitic* ⧗, ⧗, Ψ, Y, ⅄, Y <ḏ> /ð/. *(i)* Cf SFG-58.

SFG-57: *NE-Iberian* ⧗, ⧗, ⧗ <da/ta> /da, ta/, ⧗ <ta>; *Taymanitic* ⧗, ⧗, *Dadanitic* ⧗, ⧗, ⧗, ⧗, *Th. C* ⧗, *Hismaitic* ⧗, ⧗, H <ṯ> /θ/; *SHR* ⧗, ⧗ <ť> /ť/ (cf Hung. /t/ > /ť/); *SR* ⧗ (Vékony 2004, 253, 264, 315) <t> /t/. *(i)* Cf Lycian)(, ⧗, ⧗ <θ> /θ/ (Adiego 2007e, 8; Mechert 2008b, 49)

SFG-58: *P.-Campanian* (Nuceria) ⧗, (Sorrento) ⋙ <s>; *Camunic* ⧗, ⧗, ⧗ <z>; *Runic (older fuþark)* Y <z> /z/. *(i)* The graphemes in SFG-56 and SFG-58 are maybe relatives. *(ii)* Glyph evolution: CT-5.

SFG-59: *SHR* ⧗, ⊟, ⧗, ⧗, ⧗, ⧗, ⧗, ⧗, ⧗ Z <z> /z/; *SHR* ⧗, ⧗, ⧗, □ <č>; *SR* ⧗, ⧗ <č> /č/; *SR* ⧗, ⧗ <č> /s/ (Vékony 2004, 314). *(i)* The closest relative of the Rovash ⊟ <z> could be one of the following: CM ⧗ 108 <za_2?/zo_2?/zi?> (SFG-54), S. Semitic ⊟ <ḏ> (SFG-55) or S. Semitic ⧗ <z> (SFG-53). *(ii)* Glyph evolution: CT-5.

SFG-60: *Lin. B* ⧗ <zo> [t͡s, d͡z] (Valério 2016, 216); *AH* ⧗, ⧗, ⧗ *376 <zi> /tsi/; *Phrygian* ⧗, ↑, ⧗ <↑> /tʃ/ts/; *Lydian* ↑ <c> /z/dz/dʒ/ (Melchert 2004, 602–603; 2008b, 58) or /tˢ/ (Valério 2008, 130); *Carian* ↑ <τ> /tš/; *Sidetic* ↑ <ts>; *SHR* ↑ <c> /č, ts/. *(i)* Valério (2008, 130–131.) proposed (referring to Melchert 2004) the relation between the AH ⧗ <zi> and Phrygian ↑ /tʃ?/. According to Adiego (2004, 302–303) the Lydian ↑ <c> and the Carian ↑ <τ> are relatives.

SFG-61: *Madhabic* Ψ, *Sabaic* Ψ, Ψ, *Hasaitic* Ψ, *Dispersed ONA, Taymanitic* Y, *Dadanitic, Th. C* ⧗, *Hismaitic* ⧗, ⧗; ⧗, ⧗, *Th. B* Y, *Geʿez abjad* ⧗ <ḥ> /ħ/. *(i)* Cf SFG-62.

SFG-62: *Greek* (red) Y, ↓, Ψ, Ψ <kh> /kʰ/; *Etruscan* Y, Ψ <χ> [kʰ]; *Raetic* Y, V, ⅄ <χ> /ch/; *Lepontic* ↓, Ψ <χ>; *Venetic* Y <χ> [g]; *Camunic* Y, Y <χ> [g]; *Gallo-Etruscan* Y, ⧗ <χ>. *(i)* Cf SFG-61.

SFG-63: *P.-Sinaitic* ⊏⊐, 𐤇, ⌶, *P.-Canaanite* 𐤇, θ, 目, *Phoenician* 𐤇, ⊟, 𐤇, 𐤇, *P.-Hebrew* 𐤇, *Old Aramaic* (Zinjîrlû, late 9^{th}–8^{th} c. BC) 𐤇, 𐤇, (8^{th} c. BC) 𐤇 <ḥ>; *SW* 𐤇, ⊟ <h>; *Greek* ⊟ (Woodard 2014, 37), ⊟, ⊓, ⊟, 𐤇, ⊟ <η> /ē, h/; *Elymian* (5^{th} c. BC) 𐤇, ⊟ <H>; *Etruscan* ⊟, ⊟, *Raetic* 𐤇, 𐤇, 𐤇, 𐤇, *Faliscan* ⊟, ⊟, *Venetic, Oscan, Latin (archaic)* ⊟, *Messapic* ⊟, *Umbrian* ⊟, 𐤇 <h>; *Runic (Anglo-Saxon)* ᚻ, *(older fuþark)* (Oostum, The Netherlands) ᚺ (Looijenga 1997, 73) <h>. *(i)* Cf SFG-65.

SFG-64: *Greek* (Naxos, 8^{th}–7^{th} c. BC) ▯ <η>, *Carian* (Memphis, Sinuri, Stratonikeia) ▯ <e>

SFG-65: *Greek* (6^{th} c. BC) H, H <ε> /ē, h/; *Lemnian* (6^{th} c. BC) I <h>; *Elymian* (5^{th} c. BC) H, N <H>; *Carian* (Mylasa) H <e>, *Messapic* H <ē, h>, *Oscan* H <η, ē>, *Gallo-Greek* H, Ƕ, N <h>; *Runic (older fuþark)* ᚺ <h>; *Latin, Oscan, Umbrian* H <h>. *(i)* Cf SFG-63.

SFG-66: *NE-Iberian* H, H, 𐤇, 𐤇, ⊟ <o>; *Celtiberian* (Botorrita, Spain) H (Eska 2008, 166–167) <o>. *(i)* The SFG-66 and SFG-63 could be related if vowel value of SFG-66, since in the early age, the grapheme <h> was occasionally used to denote /o/ in the Old Aramaic and P.-Hebrew scripts (Healey 1990, 35). *(ii)* Glyph evolution: CT-6.

SFG-67: *Greek (medieval cursive)* ϰ, χ <kh>; *Glagolitic* ⱈ <x>

SFG-68: *P.-Sinaitic* ⌘, ⌘ <ḫ> /χ/; *SHR* ⨳, ⨳, ⨳ <h> /h/; *SHR* ⨳, ⨳ <e> /e/; *TR* ⨳ (Tekin 2003, 23); ⨳, ⨳, ⨳ (Kairzhanov 2014, 17) <e> /e/; *SR* ⨳ <e> /e/ (Vékony 2004, 287, 294). *(i)* Using the <h> or <ḥ> for representing /ē/ was specific for the Greek (similarly in Lydian, Lycian, Phrygian); however, in the Old Aramaic, the <-h> was also used for /-ē/ (Segert 1978, 113) in the 10^{th}–11^{th} c. BC. Therefore, the value /e/ of the Rovash grapheme ⨳, ⨳ <e> could originate from the Greek, Lydian, Lycian, Phrygian, or the Old Aramaic, but not from the later Aramaic. *(ii)* Glyph evolution: CT-1, CT-5. *(iii)* Cf SFG-32.

SFG-69: *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) ⊘ <ṭ>; *Phoenician* ⊕, ⊕, ϴ <ṭ/ṯ> /tˤ/; *Old Aramaic* ⊗, ⊘, ⊘ <ṭ/ṯ>; *Greek* ⊙, ⊖, ⊕, ⊗, (cursive) θ, ϑ, *Etruscan* ⊙, ⊕, *Lemnian* ⊗, *Messapic* ⦶, ⊖, ⊕, *Venetic* ◊, ⊙ <θ>; *Sidetic* ⊙, O <th> /θ/; *Oscan* ⊖ <f> [f]; *Gallo-Greek* ⊖, ⊖, θ, θ <θ> /t͡s/; *Safaitic, Th. D, Hismaitic* ◎ <ḍ> [ðˤ]; *SW* ⦶ <tⁱ> /d, t/; *SE-Iberian* ⦶ <ti>, *NE-Iberian* ⦶, ⊘, ⊖, ⊕, ⊗, *Celtiberian* ⊗ <de/te> /de, te/; *TR* ⊗ <dä/ed>; O, ⊙, ◡, ◡, ⊙, ⊙ <nd/nt>; *SHR* 0, θ, ⊙, ⊗, ⊕, ⊙ <f>. *(i)* In the AH and Aegean syllabaries, the /n/ before consonant was not written (Fischer 2001, 75). Maybe that is why the TR <nd/nt> represented /n/ plus consonant (cf SFG-87 and SFG-100). *(ii)* The Greek glyphs ⊗, ⊕ were typical even in the 7^{th} c. BC (McCarter 1975); however, in the 4^{th} c. BC these glyphs did not appear in the surviving Greek inscriptions (Thompson 1912, 144–145). Consequently, these glyphs were borrowed by Rovash before the 4^{th} c. BC. See comment *(ii)* in SFG-166.

SFG-70: *Madhabic* ⊟, ⊟, *Sabaic (early musnad)* ⊟, *(early zabūr)* ϑ, *(middle zabūr)* 𐤇, *Hasaitic* ⊟, *Dumaitic* H, *Th. C* ⌶, *Safaitic* H, H (Macdonald 2015, 30, 37), *Taymanitic* ⊟, ⊟, *Dadanitic* Ⅲ, *Th. B* H, ⌶, ⌶, ⌶ <ḍ> /ḍ/ [ðˤ]. *(i)* Glyph evolution: CT-5. *(ii)* Cf SFG-71.

SFG-71: *SW* H, H, 𐤇, (Espanca) 𐤇 <tᵉ> /d, t/; *SW* 目, ⋈ <tᵃ> /t/. *(i)* The graphemes in SFG-70 and SFG-71 could be indirect relatives.

SFG-72: *Greek* (*cursive,* 2^{nd} c. BC – 9^{th} c. AD) ϑ, (*minuscular,* 9^{th} c. AD) ϑ, (*late uncial,* 9^{th} c. AD) ϑ, (*minuscular,* 10^{th}–11^{th} c. AD) θ <θ>; *Glagolitic* ⱚ, ⱚ <f>; *E. Cyrillic* ѳ <f>. *(i)* See comment *(ii)* in SFG-166.

SFG-73: *Greek* ⟐ <θ>; *Lemnian* ⊞ <θ>; *SE-Iberian* ⦶ <ti>, *NE-Iberian* ⟐, ⟐, ⟐, ⟐, ⟐, *Celtiberian* (Botorrita, Spain) ⟐, *(Eastern)* ⟐ <de/te>

SFG-74: *Parthian* , <ṭ/ṯ> /tˤ/; *TR* , , , , <d¹> /d/. *(i)* In the Parthian and Middle Iranian languages, in intervocalic position a voicing occurred: /p/ > /b/, /t/ > /d/ and /k/ > /g/ (Skjærvø 1996, 519). This could be a reason why the Parthian <ṭ/ṯ> represented /d/ in the Rovash.

SFG-75: *Lin. A* (Valério 2013, 15–17), LA 57 <ja>; *Lin. B* , <ja>; *CM* , , CM 69, , , CM 71 <ja?>; *CGk* , , , <ja>; *TR* , , , , , <y¹> /y/ [j]; *SR* , <y> /y/ [j] (Vékony 2004, 315); *CBR* <y> /j, i̯/ (Table 15 in Appendix; Vékony 2004, 164); *SHR* , , , , , <ı́> /j/. *(i)* Cf Carian , , , , , , , , , , , , (Adiego 2007a, 209–210, 508) <i> and Lydian <y> /i/. The Lydian <y> /i/ is an unstressed allophone of [i] (Melchert 2008b, 59).

SFG-76: *P.-Umbrian* (Tolfa, ca. 530–525 BC) <í> (Urbanová 2003, 33; Bakkum 2009, 380); *S. Picene* , , , , <e> (MNAMON)

SFG-77: *Phoenician* , , , , *Old Aramaic* (8th c. BC) , *I. Aramaic* (7th c. BC) , (6th c. BC) , <y> /y/ [j]; *Lydian (archaic)* <i> (Woudhuizen 1984–1985a, 93). *(i)* Cf SFG-81.

SFG-78: *Lydian (archaic)* <i> /i/ (Woudhuizen 1984–1985a, 93); *SW* , (Espanca) , *SE-Iberian* , <i> /i/; *NE-Iberian* , , , , , , , , *Celtiberian* (Botorrita, Spain) , <i>

SFG-79: *Sidetic* (Adiego 2007e, 14), ; , (Woudhuizen 1984–1985b, 117) <i> /i/; *NE-Iberian* , <i> /i/

SFG-80: *Greek* , , , <ι> /i, ī/; *Lydian* (Woudhuizen 1984–1985a, 93) <i> /i/; *Phrygian* , <y> /j/; *Sidetic* <j>; *Libyco-Berber* , , , , , , , <Y/I>; *Runic (older fuþark, Anglo-Saxon)* <ï> /ij/

SFG-81: *I. Aramaic* (7th c. BC) , , (6th c. BC) , , , , (5th/4th c. BC) , *Hebrew* (Qumran, 1st c. BC) <y>; *TR* , , <i, y> /i, ï/; *SHR* , <j> /i, j/; *CBR* <i> /i/ (Vékony 2004, 164); *SR* , <i> /e, i, ï/ (Vékony 2004, 314). *(i)* Cf SFG-45. *(ii)* Cf SFG-77.

SFG-82: *Hatran* <y>; *Sogdian* <y> /y, ē, ī/, *Nabataean* <y>. *(i)* The similarity of the SFG-80 and SFG-82 is probably homoplasy due to the lack of known historical and geographical relationship.

SFG-83: *Palmyrene* (Palmyra, 2nd c. AD) <y>; *Middle Persian (Inscriptional)* , *(Psalter)* , *(Early Cursive Pahlavi)* , *(Book Pahlavi)* , , <y> /y, ĕ, ĭ, y̆/; *TR* <y²> /y/; *CBR* (Table 15 in Appendix), *SR* , , <i> /i, ï/ (Vékony 2004, 314).

SFG-84: *Greek (medieval uncial)* , (Faulmann 1880, 171), (*late uncial,* 9th c. AD) (Taylor 1883, 154) <ι>; *Glagolitic* , *iže* <i>; *E. Cyrillic* *iže* <i>

SFG-85: *Glagolitic* *jerь* <ь>; *E. Cyrillic* Ь *jerĭ* <ь>. *(i)* Cf *Glagolitic* <i> (SFG-84).

SFG-86: *Glagolitic* *ju* <j>; *E. Cyrillic* Ю *ju* <j>. *(i)* Cf *Greek* (*cursive,* AD 701–718) , , (Thompson 1912) <ι>.

SFG-87: *Lin. A* , , LA 77 <ka?>; *Lin. B* <ka> /ka, ga, kʰa/; *CM* (ENKO Atab 001, not later than 1525–1425 BC) CM0 09 (Valério 2016, 186) <ka?>; *S. Picene* <q>; *NE-Iberian* , , <gu/ku> /gu, ku/, , <gu>, , <ku>; *Celtiberian* <ku>; *Carian* , , <q> /q/; *Th. B* , *Th. C* , *Safaitic* <g>; *Runic (older fuþark)* <ŋ> /ŋ/; *TR* , , <ṅ> /ŋ/. *(i)* Cf Th. D <g?>. *(ii)* In the AH and Aegean syllabaries, the /n/ before consonant was not written (Fischer 2010, 75). Maybe that is why the Runic <ŋ> and the Rovash <ṅ> could represent a nasal sound. Cf SFG-69 and SFG-100.

SFG-88: *CM* , , , , , CM 25 <ka?>; *CGk* <ka> /ga, ka, kʰa/; *TR* , <k⁵/ʷkʷ> /q/.

(i) The possible relation of the Greek (red) Ψ, ↓ <kh> to SFG-88 has not been clarified.

SFG-89: *SE-Iberian* ʌ, A <ga>; *NE-Iberian* ʌ, ʌ, ʌ, A, A, A, A, A, F <ga/ka>; *Celtiberian* (Botorrita, Spain) ʌ, ʌ, ʌ, A, A, A <ka>; *SHR* ʌ, A, A <g>

SFG-90: *AH* (KARKAMIŠ A14b, 10^{th} c. BC) [glyph] (Hawkins 2000, 83–86), [glyph], [glyph] *315 <kar>; *CBR* B <k> (Vékony 2004, 165); *TR* B, B, B, 8, K, K <k^4, $^{w}k^{w}$>. *(i)* It is noteworthy that a relationship between the Rovash B <k> and Rovash 8 <q> (SFG-101) could exist if supposing the CT-9 characteristic transformation. However, the altering sound values of the Rovash B <k> and 8 <q> implies different origins. Cf CT-5.

SFG-91: *Lin. A* [glyph], [glyph], [glyph], [glyph], [glyph], [glyph] LA 74 <ze?> /ce, ɟe ↔ tse, dze/; *Lin. B* [glyph], [glyph] <ze>; *CM* [glyph], [glyph], [glyph], [glyph], [glyph] CM 112 <k/ze?> /ce, ɟe ↔ tse, dze/?; *CGk* [glyph], [glyph], [glyph], [glyph], [glyph] <ke> [k^je, g^je, $k^{hj}e$ ↔ ce, ɟe, c^he] (Valério 2016, 256). *(i)* The SFG-91 is based on Valério (2016, 253–256).

SFG-92: *Lin. A* [glyph], [glyph], [glyph], [glyph], [glyph], [glyph] LA 74 <ze?> /ce, ɟe ↔ tse, dze/; *CM* [glyph], [glyph], [glyph], [glyph], [glyph] CM 112 <k/ze?> /ce, ɟe ↔ tse, dze/ (Valério 2016, 256); *CBR* [glyph], [glyph], [glyph], [glyph], [glyph] <z> /z/ (Table 15 in Appendix; Vékony 2004, 164). *(i)* Both the similarity between the Lin. A [glyph] and the Rovash [glyph] glyphs and the sound relation between the Lin. A <ze?> and Rovash <z> are tentative proposals and need more evidence. *(ii)* Glyph evolution: CT-5.

SFG-93: *Phoenician* (Nora, ca. 900 BC) [glyph], (Kilamuwa, ca. 825 BC) [glyph], (Limassol, ca. 750 BC) [glyph]; *Old Aramaic* (10^{th}–5^{th}/4^{th} c. BC) [glyph] (Gibson 1975), (8^{th} c. BC) [glyph]; (Deir ʼAllā, around 800 BC) [glyph] <k> /k, χ/; *SW* (Espanca) [glyph], [glyph], K <k^e> /k/; *NE-Iberian* [glyph], [glyph] <ke>; *Greek* (8^{th}–7^{th} c. BC) ꓘ, ꓘ, ꓘ, H, ꓘ, ꓘ, ꓘ (Healey 1990, 37), *Phrygian* ꓘ, *Lemnian* ꓘ, *Lydian* ꓘ, *Elymian* K, K, K, *Raetic* ꓘ, ꓘ, ꓘ, ꓘ, *S. Picene* K, K, *Faliscan* K, *Lepontic* ꓘ, ꓘ, ꓘ, *Venetic* ꓘ, ꓘ, ꓘ, ꓘ, *Camunic* ꓘ, *Messapic* K, *Oscan* ꓘ, ꓘ, K, *Latin (archaic)* K <k, κ> /k/; *Lycian* K <k> /k</ (Melchert 2008a, 48) or /c/ (Adiego 2007e, 8); *Etruscan* ꓘ <k^a> /k/; *Umbrian* ꓘ, ꓘ, ꓘ, ꓘ, ꓘ, *Gallo-Etruscan* K, K, K, K, ꓘ, *Gallo-Greek* K, K <k> [g, k]

SFG-94: *Phoenician* (Amurru, 11^{th} c. BC) V, V, (Aḥīrām, Byblos, ca. 1000 BC) V, V, (Jeḥīmilk, Byblos, ca. 950 BC) V, (Elībaal, Byblos, ca. 900 BC) V <k>; *NE-Iberian* ‹, ‹, *Celtiberian* ‹ <ge/ke>; *SR* › (Vékony 2004, 315), *TR* [glyph], [glyph], [glyph], [glyph] <k^2> /k/

SFG-95: *NE-Iberian* X, X <ge/ke>; *TR* X <k^2> /k/. *(i)* Glyph evolution: CT-1. *(ii)* The similarity to the Runic (Anglo-Saxon) ᛝ <ŋ> /ŋ/ is presumably a homoplasy.

SFG-96: *NE-Iberian* [glyph] <ge/ke>; *TR* [glyph], [glyph], (Mendur-Sokkon IV) [glyph] (Konkobaev et al. 2015, 41) <g^2>

SFG-97: *Lin. A* [glyph], [glyph], [glyph], [glyph] LA 67 <ki?>; *Lin. B* [glyph]; [glyph] <ki> /gi, ki, k^hi/; *CM* [glyph], [glyph], [glyph], [glyph], [glyph], [glyph] CM 70 <ki?> (Valério 2016, 436, 442); *CGk* [glyph], [glyph], [glyph], [glyph], [glyph], [glyph], [glyph], [glyph] <ki> /gi, ki, k^hi/; *Carian* ▽, ∇, △, Y; <k> /k/; *CBR* A <k> /k/ (Hosszú 2013, 38–39); *SHR* A, △, A <k> /k/; *TR* ◁, ▷, ▽, (manuscript) ◀ <k^3/$^{ï}k^{ï}$> /q/

SFG-98: *Greek* (Crete, Sikinos, Phrygia) [glyph] (Jeffery 1961, 35–37, 39–40) <kh?>; *Sidetic* И <g>; *SE-Iberian* [glyph] <ki>; *NE-Iberian* [glyph], [glyph], [glyph], И, ꟿ, ꟿ, Z, [glyph], [glyph], [glyph], [glyph], [glyph], [glyph], [glyph] <gi/ki> /gi, ki/; *Celtiberian* (Botorrita, Spain) [glyph], [glyph] <ki/ci>; *SHR* [glyph], [glyph] <k> /k/; *CBR* [glyph] <k> /k/ (Vékony 2004, 164); *SHR* И <γ> /γ/; *TR* И, N, N, N, N, (manuscript) [glyph] <k^1> /q/; *TR* [glyph] <g^1> /γ/; *CBR* [glyph], И, N, И <γ> /γ/ (Vékony 2004, 164, 192; Hosszú and Zelliger 2014a, 186); *SR* [glyph], [glyph], [glyph], H, H, И, N <g> /g, γ/ (Vékony 2004, 315). *(i)* The Lycian [glyph], [glyph] <β/K> /k^w/? (Adiego 2007e, 8; Melchert 2008a, 48) is maybe also member of SFG-98.

SFG-99: *Lin. A* [glyph] LA 70 <ko?>; *Lin. B* [glyph], [glyph] <ko> /ko, go, k^ho/; *CM* [glyph] CM 21 <ko?>; *CGk* Λ, Λ, Λ, Π, Λ, [glyph] <ko> /go, ko, k^ho/; *Libyco-Berber* [glyph], Γ, ꓶ, ꓶ, Γ, L, ∩, Π, ∩, ∧ <G>; *SW* Λ <k^a> /k/;

SE-Iberian Λ <ka>; *NE-Iberian* Λ <ka>; *Celtiberian* Λ <ka/ca>

SFG-100: *CM* ◊, ◊ CM 15 <ko?> (Valério 2016, 430, 442); *AH* ◈ (Payne 2010, 14), (SÜDBURG) ◇, ◈, ◈ *423 <ku> /gu, ku/; *Lycian* ◊, ◊, ◇ <k>; *NE-Iberian* ◇, ◊, ◈, ◊, ⌗, ⍰ <gu/ku> /gu, ku/; *Celtiberian* (Botorrita, Spain) ◈, ◇ <ku/cu>; *S. Picene* ⏀, ◈, ◇, ◈ <q>; *Safaitic* ◊, ◊, ◊, ▭ (Macdonald 2015, 37) <g>; *Runic (older fuþark)* ◇, □ <ŋ> /ŋ/; *SHR* ◊, ◊, <k> /k/; *SR* ◊, ◊ <k^{1}> /q/ (Vékony 2004, 315); *TR* ◊, ◊, ◈ <ṅ> /ŋ/. *(i)* Cf SFG-87. Presumably, both CM ◈ CM0 09 <ka?> (SFG-87) and *CM* ◊, ◊ CM 15 <ko?> (SFG-100) are ancestors of the P.-Hispanic, Ancient Italic, and Rovash graphemes in SFG-87 and SFG-100; which is an example of glyph-level reticulation.

SFG-101: *CGk* (Paphian) ⚘, ⚘ <ko> /go, ko, k^{h}o/ *Carian* ⧗, ⧗?, (E.Me 30) ⚘ <γ> /g/, (coins) ⧗, (M33) 8, 8 <γ?> (Adiego 2007a, 483–509); *SW* ⋈, ⧖ <k^{o}> /k/, *SE-Iberian* ⋈ <go/ko> /go, ko/, *NE-Iberian* ⧗, ⧖, ⋈ <go/ko> /go, ko/, ⧗, ⧖, ⧖ <ko> /ko/; *Celtiberian* (Botorrita, Spain) ⧗, ⧗, ⧖ <ko/co>; *Th. C* ⊸, *Hismaitic* 8 <g>; *SHR* ⧖ <χ> /χ/; *CBR* 8 <q> /q/ (Vékony 2004, 165; Hosszú and Zelliger 2014a, 186), *SR* 8 <k^{1}> /q/; *SR* ⧖ <k^{5}, ^{w}k^{w}> /q/ (Vékony 1992a, 542; 2004, 315). *(i)* The Celtiberian ⧖ <ko> could be autapomorphy; however, it is more probable that this open shape also existed in other cognate scripts, cf Aegean ⚘ <ko>. *(ii)* The Rovash 8 <q> and S. Semitic 8 <g> are probably not homoplasies. *(iii)* Cf *Th. D* 8 <g?>. *(iv)* Cf CT-8 and CT-9.

SFG-102: *Dadanitic* ϑ, ϑ, ⅃, *Dispersed ONA* ⅃, *Taymanitic* ⅁, ⅁, <g>; *SR* ϑ, ϑ <g> /g, γ/ (Vékony 1992a, 542). *(i)* These may be the one-loop version of the ⧗ or ß shapes in SFG-101; since the probable relative S. Semitic 8 <g> surely has a relationship with the 8 shapes in SFG-101.

SFG-103: *Lin. A* ⾮, ⾮ LA 44 <ke?>; *Lin. B* ⾮ <ke> /ke, ge, k^{h}e/; *CM* ⾮, ⾮, ⾮, ⾮, ⾮, ⾮, ⾮ CM 110 <ke/u?>; *CGk* ⁑, ⁑, ⁑, ⁑, ⁑, ⁑, ⁑ <ku> /gu, ku, k^{h}u/; *Libyco-Berber* ÷, ⁞ <Q>; *Lycian* ⟩|⟨, ⟩|⟨, ⁑, ⁂ <q>; *TR* ⁑, ⁑, ⁑, ⁑, ⊤, ⁑, *SR* ⁑ <g^{1}/^{1}γ> /ġ, γ/ (Vékony 1992a, 542). *(i)* Glyph evolution: CT-3.

SFG-104: *Carian* X, + <k̂/χ (Simon 2008, 459–460)> /c?/k^{j}?/k^{w}?/; *Lydian* + <q> /k^{w}/? (Adiego 2007e, 7; Melchert 2008b, 57); *Runic (older fuþark)* X <g> [g, g, ɣ, j].

SFG-105: *AH* ⊲ *176; ⊲, ⊲, (BABYLON 1) ⊲ (Payne 2010, 121) *175 LINGUA <la>; *Carian* Δ, Δ <l>. *(i)* The relation of the AH *175 and the Carian <l> is uncertain, cf SFG-162.

SFG-106: *P.-Sinaitic* ⌐, ᒪ, ᓚ (Sprengling 1931, 55) <l>; *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) ϭ (Rollston 2008, 84) <l>

SFG-107: *P.-Sinaitic* ʔ <l>; *Madhabic, Sabaic, Dispersed ONA, Dumaitic, Dadanitic, Th. C, Th. B., Hasaitic* 1, *Taymanitic, Hismaitic* 1, I, *Th. D* ¬, *Safaitic* I, *Geʿez abjad* ᐱ <l>; *Greek* ˥, 1, ↾ <λ>; *Phrygian* ↾, *Lydian* 1, *Lemnian* ˥, *S. Picene* ↾, *Camunic* 1, *Messapic* ↾, *Raetic* 1, *Venetic* 1, *Runic (older fuþark, Anglo-Saxon, younger/Danish fuþark)* ↾, *SW* ↾, (Espanca) 1, *SE-Iberian* 1, *NE-Iberian* ↾, *Celtiberian* (Botorrita, Spain) ↾, ↾ <l>

SFG-108: *Phoenician* L, L, L, *P.-Hebrew* L, *Old Aramaic* (ca. 800 BC) L; (7th c. BC) L, L, L, *I. Aramaic* L, *Middle Persian* L, L, *Syriac* ܠ, *Arabic* ل <l>; *Sogdian* (Ancient Letters) L <δ> (Skjærvø 1996, 519); *Greek* V, ⊬ <λ>; *Faliscan* ⊬, *Camunic* ⊬, *Etruscan* ⅃, *Raetic* ⊬, ⅃, *Lepontic* ⅃, ✓, *Venetic* ⅃, *Oscan* ⅃, *Umbrian* ⅃, ✓, *Latin (archaic)* ⊬, *SR* L <l> (Vékony 2004, 315); *TR* ⅃, ⊬, ⊬, ✓, ∨ <l^{1}>

SFG-109: *Greek* (Ionia, Corinth) Λ, *Elymian* ∧, ∧, ∧ <λ>; *Oscan* ∧ <λ, l>; *Lycian* (TL 29) ▲, (TL

5) 𐊍 (Kalinka 1901 *apud* Adiego 2015, 14, 21), *NE-Iberian* Λ, *Messapic* Λ, Λ, *Gallo-Greek* Λ, Λ, ⅄, *SHR* Λ, *SR* Λ <l> (Vékony 1992a, 542). *(i)* The possible relationship of the Lycian 𐊍 <l> and the ᛘ CM 011, CGk Λ, and SHR Λ <l> in SFG-162 is unclear.

SFG-110: *Greek* (cursive, AD 824–830) λ <λ>; *Glagolitic* ⰾ <l>

SFG-111: *Greek* (cursive, AD 701–718) λ, λ, λ (Thompson 1912) <λ>; *E. Cyrillic* Л <l>

SFG-112: *P.-Sinaitic* 〰, 〰, ⌇, *P.-Canaanite* (mid-11th c. BC) ⌇, *Phoenician* ⌇, (Byblos, 11th–10th c. BC) ⌇, *Old Aramaic* (Tell Fekheriye, 9th c. BC) ⌇, *Greek* (Crete, 8th–7th c. BC) 𐌌, *SW* ⌇, ⌇, ⌇, ⌇, *P.-Umbrian* (Tolfa, ca. 530–525 BC) ⌇, ⌇, *SR* ⌇, ⌇ (Vékony 2004, 315), *CBR* ⌇ <m> (Table 15 in Appendix)

SFG-113: *Phoenician* (Kilamuva-stele, Sam'al, around 820 BC) 𐤌, 𐤌 (Lemaire and Sass 2013, 125), 𐤌, 𐤌, *Old Aramaic* (8th c. BC) 𐡌, (Deir 'Allā, around 800 BC) 𐡌, (10th–7th c. BC) 𐡌 (Gibson 1975), *Greek* 𐌌, (8th–7th c. BC) 𐌌, (8th–7th c. BC) 𐌌, (8th–7th c. BC) Μ <μ>, *Phrygian* 𐌌, *Lydian* 𐌌, *Lemnian* 𐌌, *Etruscan* (7th c. BC) 𐌌, *Raetic* Μ, 𐌌, 𐌌, *Lepontic* 𐌌, *Camunic* 𐌌, *Faliscan* 𐌌, *Venetic* 𐌌, *Latin (archaic)* 𐌌, *Carian* N, 𐊪 <m>

SFG-114: *P.-Hebrew* (8th c. BC) 𐤌, *Old Aramaic* (8th–3rd c. BC) 𐡌, *Phoenician* (Karatepe, ca. 700 BC) 𐤌, (Byblos, 5th–4th c. BC) 𐤌, *I. Aramaic* (7th–5th/4th c. BC) 𐡌, 𐡌, *Parthian* 𐭌, 𐭌, 𐭌, 𐭌, 𐭌, *Hatran* 𐣬 <m>

SFG-115: *I. Aramaic* 𐡌, *Hebrew* ם, *Middle Persian (inscriptional)* 𐭬, 𐭬, 𐭬, 𐭬, *(Psalter)* 𐮋, 𐮋, *(Book Pahlavi)* 𐮋, 𐮋, *Syriac* ܡ, ܡ, ܡ, *Sogdian* 𐼽, 𐼽 <m>

SFG-116: *CGk* (Paphian, 6th c. BC) 𐠊, 𐠊, 𐠊, 𐠊 (Valério 2016, 228, 278) <me>; *Libyco-Berber* ⵅ, ⵅ, ⵅ, ⵓ, ⵓ <P/F> /f/p?/; *NE-Iberian* Ω, ᛟ <ba>, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, Ш, Ψ, ᛟ, ᛟ, ᛟ, ᛟ, Ω <be>, *Celtiberian* ᛟ, ᛟ, Ω, ⌇, ᛟ, W <be/pe>; *TR* ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ <b^2> /b, u̯/; *CBR* ᛟ <b> /u̯/ (Table 15 in Appendix); *TR* ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ᛟ, ◇, (Mendur-Sokkon IV) ᛟ (Konkobaev et al. 2015, 41) <m>. *(i)* The TR ᛟ <b^2> ~ ᛟ, ᛟ, ᛟ <m> could be related to the Old Turkic onset [b] > /m/ change (Erdal 2004, 62, 74), cf SFG-118. The TR ᛟ <m> probably originated from the TR ᛟ <b^2> by turning the shape ᛟ with -90°, cf CT-10. *(ii)* The similarity between the TR ᛟ <m> and the Middle Persian 𐮋, 𐮋, 𐮋, 𐮋 <m> (SFG-115) is likely a homoplasy.

SFG-117: *TR* ᛟ (Gabain 1941) ᛟ, ᛟ, ᛟ (Kairzhanov 2014, 17) <m>; *SHR* ᛟ, ᛟ, ᛟ, (Constantinople, AD 1515) ᛟ <mb>. *(i)* The relationship between TR <m> and SHR <mb> was proposed by Németh (1934) and Vékony (2004). *(ii)* Presumably, the glyphs in SFG-117 are ornamented versions of the glyphs in SFG-116, cf TR ᛟ <m> (SFG-116). *(iii)* The glyphs in SFG-117 are maybe comparable to the *SW* (Espanca) ᛟ, ᛟ; ᛟ, (Stele of Mestras) ᛟ (Correia 1996, 105) <p^i?> (Valério 2008, 125–126).

SFG-118: *CGk* 𐠋, 𐠋 <mi>; *Lydian* (650–600 BC) ΔΔ <m> (Woudhuizen 1984–1985a, 98); *Madhabic, Sabaic* 𐩣, *Dispersed ONA* 𐩣, 𐩣, *Dumaitic, Taymanitic* 𐩣, *Dadanitic* 𐩣, 𐩣, *Th. B* 𐩣, 𐩣, *Hasaitic* 𐩣, 𐩣, *Geʿez abjad* መ, *SHR* 𐳘, 𐳘, 𐳘, 𐳘 <m>. *(i)* Cf AH ◇ (Hawkins 1986, 370–371), ᛟ (Payne 2010, 14) *419 <mí>.

SFG-119: *Phrygian* B, 8 <b>; *Greek* 8, 8, 8, 8 (Jeffery 1961, 23), (Crete, 8th–7th c. BC) 8, (Argos) B <β>; *Carian* ΔΔ, M, M, (coins) 8, 8 <p> /p/; *Lycian* B (Adiego 2007e, 8) <b>; *Lydian* 8 <b> /p/; *Etruscan* 8 <b> [p]; *P.-Campanian* (Nuceria) 8 <b>; *Elymian* B, B <β>; *Umbrian* 8, B, *Oscan* 8, 8, B, B, *Latin* B, *Messapic* B, B, *Gallo-Greek* B <b>; *Runic (older fuþark, Anglo-Saxon, younger/Danish fuþark)* ᛒ, *(older fuþark)* ᛒ, ᛒ, (Oostum, The Netherlands) ᛒ (Looijenga 1997, 6, 73) <b>; *TR* 𐰉 (Kairzhanov 2014, 17) <b^1>. *(i)* It is possible that the Greek 𐌁 <b> was not invented from the glyph 𐀁 (SFG-13), but it indirectly originates from the *CGk* 𒀭, 𒀭 <mi> (SFG-118). In this case, this is an example for glyph-level reticulation. *(ii)* The TR 𐰉 <b^1> is related to the graphemes in SFG-119, or the SHR 8 <m> (SFG-118) ~ TR 𐰉 <b^1> correspondence originates from the Old Turkic onset [b] > /m/ change (Erdal 2004, 62, 74), cf SFG-116. *(iii)* Valério (2016, 282–284) pointed out that languages without phonemic /m/ typically possess /b/ that—depending on its position—can be pronounced allophonically as [b], [m], or prenasalised [^{m}b]; the realization of the sound depends on if the following vowel is plain or nasalized. Valério also mentioned the possibility that the language of CM possessed a sound that varied between [b] and [m]. This opinion supports that the SFG-118 and SFG-119 could be relatives. *(iv)* Glyph evolution: CT-5. *(v)* Cf SFG-120.

SFG-120: *NE-Iberian* ᛘ <m>, V, ᚠ, T, Y, V, Y, ⅄ <ḿ/m̄>; *Celtiberian* ᛗ, ᚠ, ᛘ, ᛘ <ḿ>. *Sidetic* <, C <m>, *Umbrian* Λ <m>; *Libyco-Berber* ⊔, ∪, ∨ <M>. *(i)* Presumably, the graphemes in SFG-120 are related to the CGk 𒀭 <mi> in SFG-118.

SFG-121: *Madhabic, Sabaic, Dispersed ONA, Dumaitic, Taymanitic, Dadanitic, Th. B, Th. C, Th. D* Π, *Safaitic* ⊃, (,), >,], ⊐, ⊏, ∩ (Macdonald 2015, 31, 33, 37), *Hismaitic* ⊏, ⊐, ⊃, *Geʿez abjad* ⊓ <b>

SFG-122: *CM* 𒀭, 𒀭, 𒀭, 𒀭 CM 73 <mo?>; *CGk* ⦶; (Paphian, 6th c. BC) ▽, ◫, ▽, ◫, ⊟, ◻, (Paphian, late) ⦶, ⦶, ʮ, ◎ (Valério 2016, 278) <mo>; *SW* ▯ <p^o>; *SE-Iberian* ▯ <bo?>; *NE-Iberian* ⊟, ▯, ▥ <bu> /bu/; *Celtiberian* (Botorrita, Spain) D, ▯ <bu/pu>; *Libyco-Berber* ⊙, ⊡, ⦶, ⊕, ⊞ <B>; *TR* 𐰋, 𐰋, 𐰋, ⊙, O, *SR* O (Vékony 2004, 243, 251, 315) <m>. *(i)* Note that the shapes of the 𒀭 CM 73 <mo?> and Libyco-Berber ⊞ <B> are close to each other. *(ii)* The *AH* ⦶ *362 <má> maybe belongs to SFG-122.

SFG-123: *CM* 𒀭, 𒀭, 𒀭, 𒀭, 𒀭, 𒀭 CM 39/49 <mu?>; *CGk* 𒀭, 𒀭, 𒀭, 𒀭, 𒀭, 𒀭 <mu>; *SW* ⋈ <bu>; *SE-Iberian* ⋈, *NE-Iberian* ⋇, ⋇, X, X, X, X, X, X, ⋇, *Celtiberian* (Botorrita, Spain) ✱ (Eska 2008, 166–167), ⋇, X <bo>; *SHR* X, *CBR* X (Vékony 2004, 164), *SR* X <b> (Vékony 1992a, 542)

SFG-124: *P.-Canaanite* 𐤍, *Phoenician* 𐤍, *P.-Hebrew* 𐤍, *Old Aramaic* (8th c. BC) 𐡍, *I. Aramaic* (5th/4th c. BC) 𐡍, *Parthian* 𐭍, *CBR* (Nagyszentmiklós) 𐤍, 𐤍 <n> (Table 15 in Appendix)

SFG-125: *Madhabic* (Dadan) 𐪌, 𐪌, *Sabaic* 𐩬; *Dispersed ONA* 𐪌, *Dumaitic* 𐪌, 𐪌, *Taymanitic* 𐪌, 𐪌, *Dadanitic* 𐪌, *Th. D* 𐪌, *Hasaitic* 𐩬, 𐩬, *Geʿez abjad* 𐩬 <n>. *(i)* The difference between SFG-125, SFG-126, SFG-127, and SFG-128 is very small, and it is difficult to distinguish them.

SFG-126: *Greek* 𐌍, *Phrygian* 𐌍, *Lydian* 𐌍 (Adiego 2007e, 7; Melchert 2008b, 57), 𐌍 (Swiggers and Jenniges 1996, 283), *Etruscan* 𐌍, *Faliscan* ᚴ, *Lemnian* 𐌍, *Lepontic* 𐌍, *Raetic* 𐌍, ᛅ, *Venetic* 𐌍, *Camunic* 𐌍, *Messapic* N, *Gallo-Etruscan* ᚴ, N, 𐌍, *Latin (archaic)* ᚴ, *Oscan* ᚴ, *Elymian* ᚴ, *SW* ᚴ, 𐌍, *SE-Iberian* 𐌍, *NE-Iberian* ᚴ, *Celtiberian* ᚴ <n/ν>. *(i)* See comment *(i)* in SFG-125.

SFG-127: *Greek* И, *Dispersed ONA* N, *S. Picene* N, N, *Lycian* N, N, 𐌍, *Etruscan* 𐌍, *Messapic* N, *Elymian* N, N, N, *Raetic* N, *Faliscan* N, *Oscan* N, И, *Gallo-Greek* N, N, *Umbrian* N, N, *Lepontic* N, *NE-Iberian* N, *Celtiberian* N <n/ν>. *(i)* See comment *(i)* in SFG-125.

SFG-128: *Dispersed ONA* Ⲏ, *Etruscan* 𐌍, *Messapic* N, Ⲏ <n>; *Elymian* N, *Gallo-Greek* Ⲏ <n/ν>. *(i)* See comment *(i)* in SFG-125.

SFG-129: *Oscan* И, *Umbrian* И, *Greek* (*cursive,* 601–640) ν (Thompson 1912), *E. Cyrillic* H <n/ν>

SFG-130: *Greek* (*early minuscular,* 9th c. AD) ᚠ, ᚠ (Thompson 1912), (*cursive,* AD 701–718) ᚠ

<ν>; *Glagolitic* Ⱀ, ⱀ <n>

SFG-131: *Phoenician* 𐤍, /,), *Hebrew* נ, ן, *Parthian* 𐭍, ـ, ـ, ـ, |,), *Sogdian* (Ancient Letters, early 4th c. AD) 𐼌, *Middle Iranian (Psalter)* 𐮋, *(Book Pahlavi)* ו, *Hatran* 𐣭, *Palmyrene* 𐡮, *Nabataean* 𐢓, 𐢔, *Syriac* ܢ (individual), ܢ (ending), ܢ (beginning, middle), *Sogdian (Christian)* ܢ, ܢ <n>. *(i)* Cf SFG-135.

SFG-132: *AH* ⅽ, ⅽ, ⑊ *35 <na>; *SR* ⸂, ⸃, ⸅ (Vékony 2004, 251, 267, 294, 315) <n>. *(i)* Cf Sidetic ⸎ <n>.

SFG-133: *Lin. A* ⸸, Ⲓ, ⸸ LA 24 <ne?>; *Lin. B* 𐀚, 𐀚, 𐀚, 𐀚, 𐀚; 𐀚 <ne>; *CM* ⸸, ⸸, ⸸, ⸸ CM 02, ⸸, ⸸, ⸸, ⸸ CM 34, ⸸, ⸸ CM 56 <ne?>; *CGk* 𐠛, 𐠛, 𐠛, 𐠛, 𐠛, 𐠛, 𐠛 <ne>; *TR* 𐰤, 𐰤, 𐰤, 𐰤, 𐰤, 𐰤, 𐰤 <n^2> /n/. *(i)* Glyph evolution: CT-3.

SFG-134: *CM* (Ugarit) ⸸ (Valério 2016, 106–108) CM 02 <ne?>; *Carian* (Kaunos, Stratonikeia) Φ <ñ>

SFG-135: *AH* ⊃ (Hawkins 2010, 184, 188–189), ⊃ (Payne 2010, 119), ⊃ (Yakubovich 2015a, 12), ⊂ (Anders 2012), ⊂ (Payne 2010, 14), ⋐ (Payne 2010, 116) *411 <ni>; *TR* 𐰣, 𐰣 <n^1> /n/; *SHR*) <n>; *CBR*) <n> /n, ñ/ (Vékony 2004, 164); *SR*) (Vékony 1992a, 542) <n> /n/. *(i)* Probable homoplasies: glyphs in SFG-131, since their glyphs are similar to the glyphs in SFG-135; however, in the scripts in SFG-131, dextrograde writing is impossible, and dissimilarly, in SFG-135, the glyphs have two opposite versions, e.g., the Rovash), (<n, n^1>. Another difference, that all glyphs in SFG-135 are arched while certain glyph variants in SFG-131 are straightened: /, |. From this it follows that the arch ()) is only part of the visual identity of SFG-135, and not of SFG-131.

SFG-136: *CM* ⸸, ⸸; ⸸, Ψ, ⸸ CM 65/67/99/100 <ni?> (Valério 2016, 435–436, 442); *CGk* ⸸, ⸸, ⸸ <ni>; *Carian* V, Y <n>

SFG-137: *AH* ⸸, ⸸ *395 <nú>; *Lycian* Ξ, Ξ <ñ> /n̥/

SFG-138: *P.-Canaanite* ⸸, ⸸ <s>; *Phoenician* 𐤎, 𐤎, 𐤎, 𐤎, 𐤎, 𐤎, (Lachish letters, 6th c. BC) 𐤎; (Byblos, 5th–4th c. BC) 𐤎 <s> /t͡s/; *P.-Hebrew* 𐤎, 𐤎 <s>; *Old Aramaic* 𐡎, 𐡎, 𐡎, 𐡎 <s> /s/; *SW* ⸸ <s>, *SE-Iberian* ⸸, ⸸ <s>; *Libyco-Berber* ⵌ, H, ⵌ <S$_1$>; *Greek* Ξ, ⸸, Ξ <ξ> /ks/. *(i)* Other members of SFG-138 could be: Elymian Ξ <ξ?> and Lydian Ξ, Ξ <τ> /ts/? (Adiego 2007e, 7, Melchert 2008b, 57–58).

SFG-139: *I. Aramaic* Δ, □ (Faulmann 1880, 171), *Hebrew* ס, ס, *Nabataean* 𐢖, 𐢖 <s>

SFG-140: *Parthian* 𐭎, *Palmyrene* 𐡯, *Hatran* 𐣮, *Sogdian* (Ancient Letters) 𐼎 <s>. *(i)* The SFG-140 can be relative of the SFG-144 or the SFG-139.

SFG-141: *Lin. A* Y (Valério 2013, 15–17) LA 31 <sa?>; *Lin. B* 𐀭, 𐀭, 𐀭, 𐀭, 𐀭, 𐀭 <sa>; *CM* Y, (RASH Atab 004) Y (Valério 2013, 19–20) CM 82 <sa?>; *CGk (Paphian)* Y <sa>; *SR* Y, Ƭ <s> /s/ (Vékony 2004, 315); *TR* 𐰽, 𐰾 <s^1, š>; *TR* 𐰞 <l^1>, 𐰞, (manuscript) 𐰞 (Gabain 1941) <l^2>. *(i)* The TR 𐰽, 𐰾 <s^1, š> and 𐰞 <l^1> are relatives according to Róna-Tas (10). *(ii)* The close relationship of the Old Turkic /s/ and /š/ was discussed by Erdal (2004, 102).

SFG-142: *Madhabic* 𐩪; *Sabaic, Dispersed ONA, Dumaitic, Hasaitic, Taymanitic, Th. B, Th. C, Th. D* 𐩪, *Dadanitic* 𐪊, 𐪊, 𐪊, 𐪊 (Macdonald 2010, 13–14), *Hismaitic* ⸸, 𐩪, *Safaitic* ⸸, ⸸, ⸸, ⸸, ⸸ <s^1> [ʃ] (Macdonald 2004, 496, 499); *Geʿez abjad* ሰ <s^1/s>. *(i)* Glyph evolution: CT-2.

SFG-143: *Hismaitic* ⸸, λ, ⸸, ⸸, *Safaitic* ⸸, ⸸, Λ, ⸸ <s^1> [ʃ] (Macdonald 2004, 496, 499); *TR* 𐰲, 𐰲 <č> /č, ǰ/. *(i)* Cf the Old Turkic /š/ ~ /č/ (Erdal 2004, 103). *(ii)* SFG-142 and SFG-143 are certainly glyph variants of each other.

SFG-144: *CGk* V <sa>; *AH* ⌓, ⌒ *415 <sa>; *Safaitic* Λ, Λ, Λ, Λ (Macdonald 2005, 82) <s^1> [ʃ]; *SHR* Λ <š>; *TR* ^, ∧ <$š^1$, $š^2$>, ∩ <š>; *SR* Ð <š> (Vékony 2004, 315)

SFG-145: *Lin. A* ᛘ, ᛘ LA 09 <se?>; *Lin. B* ᛘ <se>; *CM* ᛘ CM 44 <se?> (Valério 2016, 433, 442); *CGk* ᛘ, Ψ, ᛘ, Ψ, ᛘ, ᛘ, ᛘ, Y, Y <se> /se/ (Valério 2016, 206); *TR* Y, Y, Y, ᛦ, ᛕ, (J) Y, Y, Y, Y, ᛋ, ᛘ <s^1, š> /s, š/; *SR* ᛋ <š> (Vékony 2004, 315); *TR* ¥ < s^1, $š^1$, s^2, $š^2$, r^2>; *SR* Y <š> (Vékony 2004, 315); *SHR* Y <ž> /š, ʒ/; *TR* Y, ⅄ <^{i}č^i, ^{i}ğ> /č, ǰ/. *(i)* Erdal (2004, 102) discussed the close relationship of the Old Turkic /s/ and /š/. *(ii)* Erdal (2004, 103) described several cases of the /š/ ~ /č/ alternation.

SFG-146: *Greek* (*early minuscular,* 9th c. AD) ꝏ <ω> /ō/; *Greek* (*cursive,* 3rd c. AD) ℓ, (*cursive,* AD 302–359) ϑ <o> /o, ō/; *Glagolitic* Ⰹ *onъ* <o>; *SHR* (Vargyas, 12th–13th c. AD) ⅂ <o> /o/

SFG-147: *E. Cyrillic* (9th–10th c. AD) Ѫ *onsъ (жcъ)* <ǫ>; *CBR* (Nagyszentmiklós) Ѫ <ǫ> /ǫ/ (Table 15 in Appendix)

SFG-148: *Lin. A* ǂ LA 03 <pa?>; *Lin. B* ‡ <pa> /ba, pa, pʰa/; *CM* ǂ CM 06 <pa?>; *CGk* ‡, ǂ, ǂ <pa> /ba, pa/; *SHR* ǂ, ǂ <p> /p/. *(i)* Cf SFG-149.

SFG-149: *SW* ᛈ <pᵃ> /p/; *SE-Iberian* ǂ <be?>; *SHR* ǂ, Ǝ <p> /p/. *(i)* The glyphs in SFG-149 are probably relatives of *SHR* ǂ, ǂ <p> in SFG-148; however, a relationship to SFG-153 or SFG-154 is also possible.

SFG-150: *Lin. A* ᚺ, ᚺ LA 56 <pa_2?>; *Lin. B* ᚺ <pa_3>; *CM* ● CM 72b <pa_2?>; *SW* ᚺ <pᵘ> /p/

SFG-151: *Phoenician* ʔ, *Old Aramaic* ʔ, ʔ, ʔ, ʔ, ʔ, *I. Aramaic* (7th c. BC) ʔ, ʔ, ʔ, ʔ, (6th c. BC) ʔ, (4th–3rd c. BC) ʔ, *Greek* ʔ, *Lycian* Γ, ᚪ, *Etruscan* ʔ, ʔ, *Umbrian* ʔ, ʔ, *Faliscan* ᚪ, P, *Raetic* ʔ, ᚪ <p>; *Lepontic* ʔ <P> /b, p/; *Gallo-Etruscan* ᚪ, ʔ <p> /b, p/; *SHR* Γ, *TR* ʔ, *SR* ʔ <p> /p/ (Vékony 2004, 315). *(i)* Cf SFG-152.

SFG-152: *Carian* Γ, Λ, Γ <b> /b/; *SE-Iberian* ʔ, *NE-Iberian* I, (, (<ba>; *Celtiberian* I, (<ba/pa>. *(i)* Cf SFG-151.

SFG-153: *Lin. A* ⅄, ⅄, ⅄, ⅄ LA 50 <pu?>; *Lin. B* ⅄, ⅄ <pu> /bu, pu, pʰu/; *CM* ⅄, ⅄, ⅄, ⅄ CM 41 <pu?>. *(i)* Cf SFG-149.

SFG-154: *Lin. A* ᛘ LA 29 <pu_2?>; *Lin. B* Ψ, Ψ, Ψ <pu_2>; *CM* ᛘ (?), Ш, Ш, Ψ, Ш, Ш CM 37 <pu?/so?> (Valério 2016, 432, 442); *CGk* Ψ, Ψ, Ш, Ψ, Ψ, Ψ <pu> /bu, pu/. *(i)* Cf SFG-149.

SFG-155: *NE-Iberian* ⋈ <ś>; *Etruscan* ⋈ <ś> [ʃ]; *Raetic* ⋈ <ś> /ś/; *Lepontic* ⋈, ⋈ <ś>; *Camunic* 8, ᛒ <ś>; *Gallo-Etruscan* (4th–2nd c. BC) ⋈, ⋈ <š>; *Libyco-Berber* ⧖, 8 <S_2/S_1/S>

SFG-156: *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) ᚦ, *Phoenician* (Byblos, 11th–10th c. BC) ʔ, *P.-Hebrew* ʔ, *Old Aramaic* (10th–5th/4th c. BC) ʔ (cf SFG-158), *SW* ᚦ, ʔ, *SE-Iberian* ʔ, *NE-Iberian* ʔ, ʔ, *Celtiberian* ʔ, *Greek* (8th–7th c. BC) ʔ, ʔ, *Phrygian* P, ʔ, *Lemnian* ʔ, *Lycian* P, *Lydian* ʔ, *Faliscan* P, *Etruscan* ʔ, *Messapic* P, P, *Venetic* ʔ, *S. Picene* P, P, *Elymian* P, P, *Raetic* ʔ, *Gallo-Etruscan* ʔ, *Latin (archaic)* P, ʔ, *Oscan* P, *Gallo-Greek* P, P, ϙ, P <r>; *Umbrian* ʔ <ḍ/ř>, *Glagolitic* Ⰼ, *E. Cyrillic* Р <r>

SFG-157: *Greek* (8th–7th c. BC) ◁ <ρ>; *Etruscan* ◁, *P.-Campanian* (Nuceria) ◁, *Raetic* ◁, ◁, ▷, *Lepontic* ◁, ◁, ◁, *Venetic* (6th–1st c. BC) ◁, ◁, *Camunic* ◁, *Oscan* (Etruscan-like, 4th–1st c. BC) ◁, ◁, *Umbrian* (Etruscan-like, 4th–1st c. BC) ◁, ◁, *Gallo-Etruscan* (4th–2nd c. BC) D, *SW* (Espanca) D, D, *NE-Iberian* ◁, ▷, ◁, D <r>

SFG-158: *Old Aramaic* (middle 7th c. BC) ʔ (cf SFG-156), *I. Aramaic* (middle 8th c. BC) ʔ, (7–5th/4th c. BC) ʔ, ʔ, ʔ, (Aśoka, ca. 250 BC) ˥, *Hatran* ʔ, ʔ, *Sogdian* ➤ <r>; *TR* ᛋ, ᛋ, ᛋ, ᚠ, ᛋ, ʔ, ʔ, ʔ, ᚠ <r^1>; *SR* H, h (Vékony 2004, 315), *SHR* H, H, H, H, N, И <r>. *(i)* The Rovash ᛋ glyph is attested to in Aramaic from 8^{th}–5^{th}/4^{th} c. BC; therefore, borrowing into the presumably common ancestor of the Rovash scripts called Proto-Rovash could happen in this period.

SFG-159: *Lin. A* 𐘧, 𐘧, 𐘧 LA 60 <ra?>; *Lin. B* 𐀨 <ra> /la, ra/; *CM* 𒾰 CM 87 <la?> (Valério 2016, 438, 442); *CGk* 𐠩, 𐠩, 𐠩, 𐠩 <la>; *Sidetic* 𐤮 (Adiego 2007e, 14) <l>. *(i)* Glyph evolution: CT-10.

SFG-160: *CM* 𒿂, 𒿂, 𒿂 CM 75 <ra>; *CGk* 𐠭, 𐠭, 𐠭, 𐠭, 𐠭 <ra>; *NE-Iberian* ○, ◇, ◈, ⚲, ⚲, ◇, ⚲, ⚲ <ŕ>; *Celtiberian* (Botorrita, Spain) ◈, ○, ◇, ◈, ⚲, ⚲, ⚲, ◇, Φ <r/ŕ>; *Libyco-Berber* ○, □ <R>; *CBR* ▯, ▯ (Table 15 in Appendix; Hosszú and Zelliger 2014a, 188), *SR* ▯ <r> /r/ (Vékony 2004, 154, 314). *(i)* NE-Iberian <ŕ> is a liquid (Valério 2008, 130) or a trill (Ferrer i Jané 2013, 448). *(ii)* Glyph evolution: CT-1.

SFG-161: *AH* 𔐀, 𔐀, ⟍, ⟋ *383 <ra/i>; *SHR* ⁄, ǀ <r>. *(i)* The similarity to the S. Semitic — <r> (SFG-164) is presumably a homoplasy.

SFG-162: *Lin. A* 𐘚 LA 27 <re?>; *Lin. B* 𐀩 <re> /le, re/; *CM* 𒾐 CM 011, 𒾐, 𒾐, 𒾐, 𒾐, 𒾐, 𒾐, 𒾐 CM 24 <le?>; *CGk* 𐠪, 𐠪 <le>; *SHR* ʌ, ʌ, (Constantinople, AD 1515) ʌ, ʌ, ʌ, ʌ, ʌ <l>. *(i)* The SHR ʌ, ʌ <l> and the Carian Δ, Δ <l> (SFG-105) are comparable. *(ii)* Cf SFG-105. *(iii)* Cf SFG-109. (*iv*) The relationship between the glyphs 𐘚 and 𒾐 are discussed by Valério (2016, 266).

SFG-163: *CM* 𒾓 CM 33 <re?>; *CGk* 𐠫, 𐠫, 𐠫, 𐠫, 𐠫, 𐠫 <re> /re/; *Carian* (Kildara, Sinuri, Stratonikeia) 𐊠, (Memphis, Kaunos) F, (Memphis, E.Me 14) F, (Memphis, E.Me 37) ⅂, (bronze lion, ca. 500 BC, E.xx 7, sinistrograde) ⅂, (Tralleis) Λ <r>; *Sidetic* ⅄, Λ <r>. *(i)* The Sidetic ⅄, Λ <r> may belong to SFG-164.

SFG-164: *Madhabic* (Dadan) 〉, 〉, *Sabaic* 〉, 〉, 〉 (Sprengling 1931, 55), *Hasaitic* 〉, *Dispersed ONA* 〉, 〈, 〉, *Dumaitic* 〉, *Taymanitic* 〉, 〉, ⊃, *Dadanitic* 〉, 〉, *Hismaitic* 〉, ⊃, ⊂ (Macdonald 2005, 82), *Th. B* 〉, 〈, *Safaitic* 〉, ɾ; ⊂, ɾ, ɾ, ⊏, ⊏, ɾ, ɾ (Macdonald 2015, 37), (*Th. D* –, *Th. C* ∩, – <r>; *SR* ∩ $<r^2>$ /r/ (Vékony 2004, 315). *(i)* Cf SFG-161. *(ii)* Cf Carian (Memphis) G <ŕ> /r^j?/.

SFG-165: *Lin. A* 𐘙, 𐘙, 𐘙 LA 26 <ru?>; *Lin. B* 𐀪 <ru> /lu, ru/; *CM* 𒾏, CM 010, 𒾏 CM 28 <lu?>; *Lydian* (550–500 c. BC) 𐤣, 𐤣, 𐤣 <λ>; *TR* ϒ $<r^1>$ /r/, ϒ, ϒ (manuscript) ϒ, ϒ, ϒ $<r^2>$ /r/

SFG-166: *Madhabic* 8, *Sabaic (early musnad)* 8, *(early zabūr)* 8, *(middle musnad)* 8 (Macdonald 2015, 39), *Th. B, Hasaitic* 8, *Safaitic* ℓ, ℓ, ℓ, ℓ, 8, 8 <ṯ> /θ/; *Lydian (archaic)* 8, *(classical)* 8 <f>; *Etruscan* 8, 日 <fh> /f/; *Oscan* 8, Ꮾ, 8, 8 <f> [f]; *Umbrian* 8 <f> [f]; *S. Picene* : <f> /f/. *(i)* Glyph evolution: CT-1, CT-4. *(ii)* Presumably, a /θ/ > /f/ change occurred. The relationship of the S. Semitic <ṯ> /θ/ and the Anatolian and Italic <f> is similar to the relationship of the graphemes with /θ/ or /f/ sound values in SFG-69 and SFG-72. *(iii)* Likely, the Phrygian 8, 8 /b/p^h/ also member of SFG-166. (*iv*) The S. Picene I <t> (SFG-167) is maybe related to the S. Semitic 8 <ṯ> /θ/.

SFG-167: *Raetic* ፕ, ↾ <t'>; *S. Picene* ↾, ǀ, I, I <t>. *(i)* The Carian 𐊭, 𐊭, P <t> /t/ could also be a member of SFG-167. *(ii)* Cf SFG-166.

SFG-168: *P.-Canaanite* (Izbet Sartah, ca. 1100 BC) ʑ, ⅀ <ś/š>; *Greek* ⸗, ⸗, *Elymian* 𐌔, ⸗, *Messapic* 𐌔, ⸗, ⸗, *Oscan* 𐌔, Σ <σ>; *Lemnian* 𐌔, 𐌔 <ś>; *Lydian* 𐤭, ⸗, 𐤭 <ś> /s/; *Libyco-Berber* ⪢ $<S_3/Ṡ>$; *Etruscan* ⸗, ⸗, 𐌔 <s/ś>; *Phrygian* ⸗, Σ, Σ, 𐌔, 𐌔, *SW* ⸗, ⸗, ⸗, *NE-Iberian* ⸗, ⸗, 𐌔, *Lycian* ʃ, *P.-Umbrian* (Tolfa, ca. 530–525 BC) ⸗, *S. Picene* 𐌔, ⸗, *Faliscan* ⸗, ⸗, *Gallo-Etruscan* ⸗, *Camunic* 𐌔, *Latin (archaic)* 𐌔, *Lepontic* 𐌔, ⸗, ⸗, *Raetic* 𐌔, *Umbrian* 𐌔, *Venetic* 𐌔, 𐌔, *Runic (older fuþark)* ᛊ,

⦚, ⦚, ⦚, ⦚ <s>; *Madhabic, Sabaic, Hasaitic, Dispersed ONA, Dadanitic, Taymanitic* ❯ <s^2> [ɬ]; *Geʿez abjad* ʍ <s^2/ś>; *TR* ⦚, ⦚, ⦚, ⦚, ⦚ <nč>. *(i)* Cf Old Turkic /š/ ~ /č/ change (Erdal 2004, 103). *(ii)* Glyph evolution: CT-5. *(iii)* Cf SFG-169.

SFG-169: *AH* I *380 <sa$_8$>; *Hismaitic* \, I <s^2> [ɬ]; *Sidetic* I <ś> (Adiego 2007e, 14); *SHR* I, \, *(cursive)* ⟍, *CBR* I (Table 15 in Appendix), *SR* I <s> /s/ (Vékony 1992a, 542), *TR* I, ⟍ <s^1, s^2, š, š^1, š^2> /s, š/. *(i)* The sound values of the Hismaitic <s^2> [ɬ] and of the Rovash <s> /s/ are not identical; however, this is the situation in SFG-168, too. Moreover, the sound value of the Sidetic <ś> has not been clarified. Therefore, the closer known relative of the Rovash is the AH I *380 <sa$_8$>.

SFG-170: *AH* (KARKAMIŠ A31) O (Hawkins 2000, 141), ◎ (Payne 2010, 14), ◎ (Anders 2012), ⊙ (Hawkins 1986, 371) *402 SCUTELLA <sa$_4$>, Δ (Payne 2010, 14), △ (Weeden 2014, 88) *370 <su>, Ⅲ *104 <sà> /s/ [ʃ, s], (MALATYA 6) ◊ <us> (Hawkins 2000, 33); *Carian* Φ, Φ, Θ, ⦸, Θ <ś> /ç?/ (Adiego 2007a, 32, 250; Adiego 2007e, 10) (CT-10); *TR* ◻, ⌑, ⌑, ◇ <š^1>, ◊, ◻ <š^2>; *CBR* ◊, Φ (Vékony 2004, 164; Hosszú and Zelliger 2014a, 186, 188), *SHR* Φ <š>

SFG-171: *AH* II⊢ (Hawkins 1986, 370–371) *389 <tara/i>; *Libyco-Berber* IIII, ΠΠ, ᚎ, ···· (LBI), IIII (Farrujia de la Rosa et al. 2010, 33) <Ṭ/T$_1$>; *Madhabic, Sabaic (early musnad)* Ⅲ, *(early zabūr)* Ⅲ, *(middle zabūr)* ⦚, *(late zabūr)* ⦚ (Macdonald 2015, 39), *Dispersed ONA* ᗷ, *Taymanitic* ⊠, *Dadanitic* ΠΊ, Ɔ, Ǝ, *Th. D* III, *Hismaic* Ɛ (Macdonald 2005, 82), Ǝ, Ǝ, Ǝ (King 1992, Figure 1 between pages 5 and 6), *Th. B* Ǝ, ΙΊ, Ɔ, *Safaitic* ᚎ (Macdonald 2015, 30, 37), *Hasaitic* Ⅲ <ṭ> /t^ʕ/; *Geʿez abjad* ጠ <ṭ>

SFG-172: *Lin. A* ∩ (Valério 2013, 15–17) Λ, Λ LA 37 <ti>; *Lin. B* ⋒, Λ, ⋒ <ti> /ti, t^{h}i/; *CM* ⋒ CM 23 <ti> (Valério 2016, 430–431, 442); *CGk* ↑ (Woudhuizen 1984–1985b, 120) *(Common)* ↑, (*Paphian*, 6th c. BC) ↑ (Olivier 2008, 617–618) <ti> /di, ti/; *Runic (older fuþark, Anglo-Saxon, younger/Danish fuþark)* ↑ <t> /d, t/

SFG-173: *AH* ⊲, ⊲, ⊲, ⊲, ⊲ (Payne 2010, 6, 14, 79, 81) *90 PES <ti>; *CBR* ◊, ◊, ◊ <t> (Vékony 2004, 164; Hosszú and Zelliger 2014a, 188). *(i)* Glyph evolution: CT-10.

SFG-174: *P.-Sinaitic* +, ⨯, *P.-Canaanite* ✕, +, *Phoenician* +, ✕, *P.-Hebrew* ✕, *Old Aramaic* ✗, ✗ <t>; *Madhabic, Sabaic* ✕; *Hasaitic* ✕, *Dispersed ONA, Dumaitic, Taymanitic, Dadanitic, Hismaitic, Th. B, Safaitic, Geʿez abjad* ✕, + <t>; *SW* X <t^a>; *SE-Iberian* +, X <ta>; *NE-Iberian* X <ta> /da, ta/; *Celtiberian* ✕ <ta>; *Etruscan* ⊹ <θ> /t^h/; *Raetic* X, X, ↑, +, ↑ <t>; *Venetic* ⊹, X <t> [d]; *Camunic* X <t>; *Umbrian* ⊹ <t> /t, d/; *Lepontic* X, + <T> /d, t/; *Gallo-Etruscan* X <t> /d, t/; *Gallo-Greek* +, ✕ <t> /d, t/; *SHR* +, + <d>; *TR* +, +, ✕ <d^2>

SFG-175: *Safaitic* ⊹, ⊹, ✕ (Macdonald 2015, 31, 37) <t>; *SHR* X <t́>; *CBR* X, (Vékony 2004, 192, 197, 198), *SR* X <t> (Vékony 2004, 315)

SFG-176: *Greek* (ca. 700 BC) ⅄, Y <t> /t/; *Faliscan* ⅄ <t>; *Umbrian* ⅄ <t> [d, t]; *SHR* ⅄ <t> /d, t/. *(i)* Cf SFG-24.

SFG-177: *Greek* ↑, *Etruscan, Elymian* ↑, ↑, *Faliscan* ↑, <t, τ> /t/; *Umbrian* ↑ <t> [d, t]

SFG-178: *Greek* T, T, *Faliscan* T, *Latin (archaic), Messapic* T, *Oscan* T, <t, τ> /t/; *Venetic* T <t> /d/

SFG-179: *Greek, Phrygian, Lydian, Lycian, Lemnian, P.-Campanian, P.-Umbrian, Messapic, Elymian, Etruscan, Oscan, Latin (archaic), S. Picene, E. Cyrillic* T, *Messapic* T <t, τ> /t/; *Gallo-Greek* T, T, ⊤, ⊤, T, T, ⊤ (MNAMON) <t> /d, t/

SFG-180: *Greek* (*late uncial,* 9th c. AD) T (Taylor 1883, 154) <τ> /t/; *Glagolitic* (Preslav, ca. AD

893) ᚁ, ᛏ (NLR) *tvrdo (tverdo)* <t>

SFG-181: *I. Aramaic* ᚦ, *Parthian* ᚴ, *Hatran* ᛇ, *Sogdian* ᛘ <t>; *TR* ᚻ, ᚺ, ᛞ, ᚻ <t^2>; *SR* ᛣ <t^2> (Vékony 1992a, 542)

SFG-182: Writing ductus is boustrophedon in a part of the inscriptions: *AH, CM* (Valério 2016, 179–180, 182, 193), *Latin (archaic), Libyco-Berber, Greek, Hasaitic, Hismaitic, Lemnian, Lepontic, Messapic, NE-Iberian, P.-Sinaitic, S. Picene, Sabaic, Safaitic, SW, Taymanitic, Umbrian, Venetic, TR*

SFG-183: Writing ductus is spiral or circle in a part of the inscriptions: *Etruscan, Latin (archaic), Libyco-Berber, Safaitic, Th. B, Venetic, TR*

SFG-184: Writing versus is bottom-up in a part of the inscriptions: *Libyco-Berber, Safaitic, Th. B*

SFG-185: No word divider in any inscriptions: *Elymian, Hasaitic, Safaitic, Th. B, Th. C, Th. D*

SFG-186: *AH* ǂ, ǂ, ǂ *216a FINES (ends) *ARHA* <arha>; *SR* ▤, ▥, ▦, ▧ separator, end-mark. *(i)* The existence of this SFG is very tentative.

Table 10: Similarity features groups (SFGs)

4.2 Results of the phenetic analysis

The appropriateness of these clusterings strongly depends on the data structure to be clustered. Since the investigated scripts were developed based on a kind of evolution, some branches of the scripts remained close to each other during their evolutionary history. Therefore, the single linkage clustering method is not efficient, since it cannot distinguish clusters with elements close to each other. Moreover, there are outlier members of the script branches, so complete linkage clustering is also not optimal. Certain scripts had several descendants (e.g. Aramaic script), while others remained singular (e.g. Libyco-Berber script). Consequently, the numbers of elements in the clusters largely vary. The UPGMA gives weights to each cluster according to the number of elements of the cluster in each step. Sneath and Sokal demonstrated that the UPGMA would favour clusters more similar in size (fig. 2). Conversely, WPGMA is appropriate when there is a reason *a priori* to eliminate size differences between the resulting clusters. The middle diagram in figure 2 presents the phenogram of the scripts calculated by using WPGMA.

The clearest result is obtained from the Ward method (fig. 2), since it is optimised for homogeneity and filters out the feature similarities that are shared between largely unrelated scripts due to long-term coexistence and cultural interactions. In case of the Ward's method, the square root of Jaccard distance (3) as an Euclidean metric was used.

It is noteworthy that the higher-level joins of the clusters in the dendrograms in figure 2 are analytically uninteresting, since these higher-level joins represent very large dissimilarities in the hierarchical cluster structure. The cluster structure was

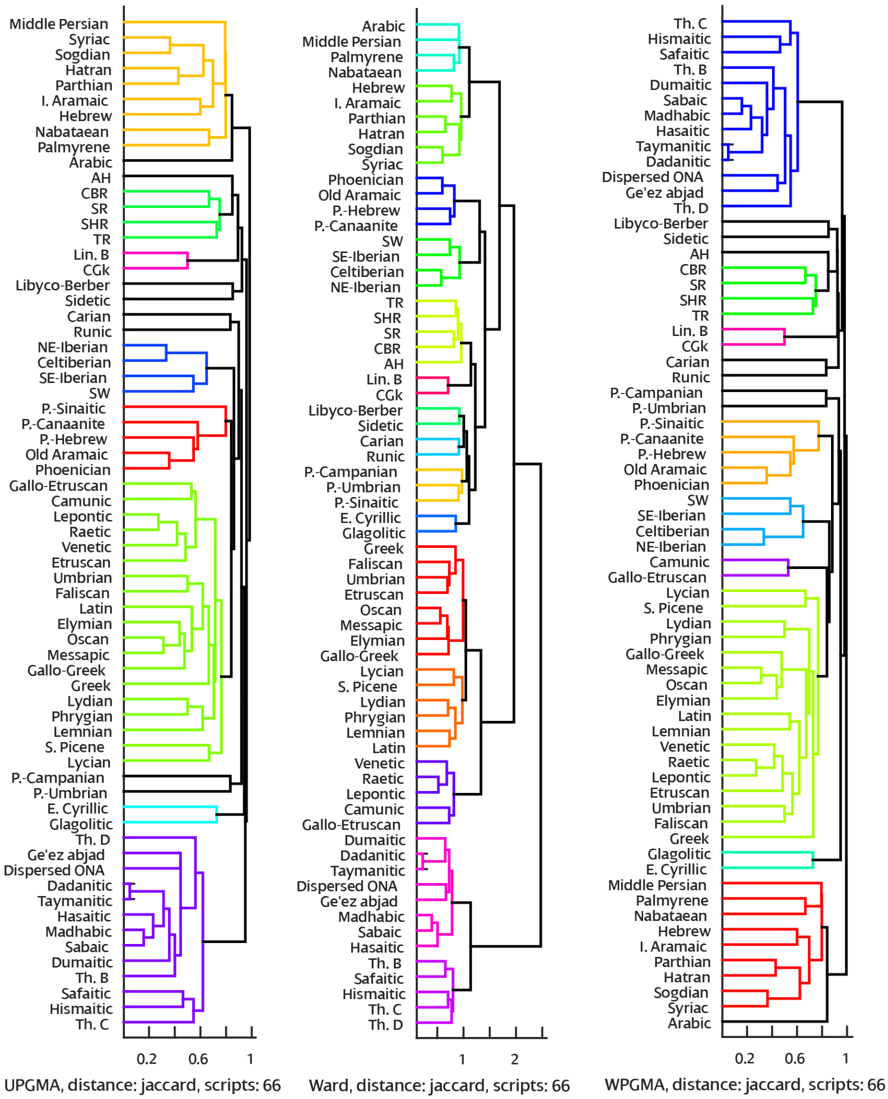


Figure 2: UPGMA (left), WPGMA (middle), and Ward (right) results ($M = 66$ scripts, $N = 186$ features)

further refined using leaf ordering methods, which placed leaves next to each other on the dendrogram that are in different clusters but still share some similarity.

Examining the results of the phenetic analysis in figure 2, these mostly medieval, Greek-derivative script Slavic scripts (Glagolitic and E. Cyrillic) were not grouped close to the Greek script. The probable reason for this is that Greek has a large number of glyphs, while the Slavic scripts have much fewer, and thus the calculated distance between them is relatively large. Furthermore, Greek has a large number of other relatives that are unrelated to the Slavic scripts. It can further be observed that the results of all three clustering methods largely agree, differing only in details supporting the stability of the phenetic methods.

An important feature of the k-means clustering is that the mean value of the clusters must be calculated. Consequently, it cannot be used in the case of categorical attributes. Since the features of the scripts can be described with categorical variables, this variable space has to be transformed into a quantitative variable space. For this purpose, multidimensional scaling (MDS) was applied, which transforms the 186-dimensional data points (scripts) to 2- or 3-dimensional data points ($n_{MDS} = 2$ or $n_{MDS} = 3$, respectively). Figure 3 presents the results for the 2-dimensional variable space.

The data points remained representative of the scripts; however, their two quantitative variables (the coordinates in figure 3) are abstract values without interpretable meaning. Then, the k-means clustering was performed on the 2- and 3-dimensional variable space of the MDS output using Squared Euclidean distance; see (4). In the k-means clustering algorithm, the Squared Euclidean distance was used; therefore, each centroid is the mean of the objects in that cluster. The resulting scatter plot in the case of 3-dimensional scaling and $K = 6$ clusters is presented in figure 4; the computation was carried out with the use of MATLAB. The cluster structure was validated by the Dunn index (7), which was 0.7 in the presented case.

The quality of the clusters in figure 4 is measured by the Silhouette index for each cluster, based on (8); see figure 5.

5 Evaluation

5.1 Some observations about the possible origin of some Western Mediterranean scripts

The P.-Hispanic scripts are descendants of Phoenician script. However the phenetic results (SFGs in table 10) present several P.-Hispanic graphemes as being unrelated to the Phoenician. Instead, they are similar to various Aegean, AH, and AGA scripts. The following data support the possibility of transmitting the literacy of the Cypriots to Iberia.

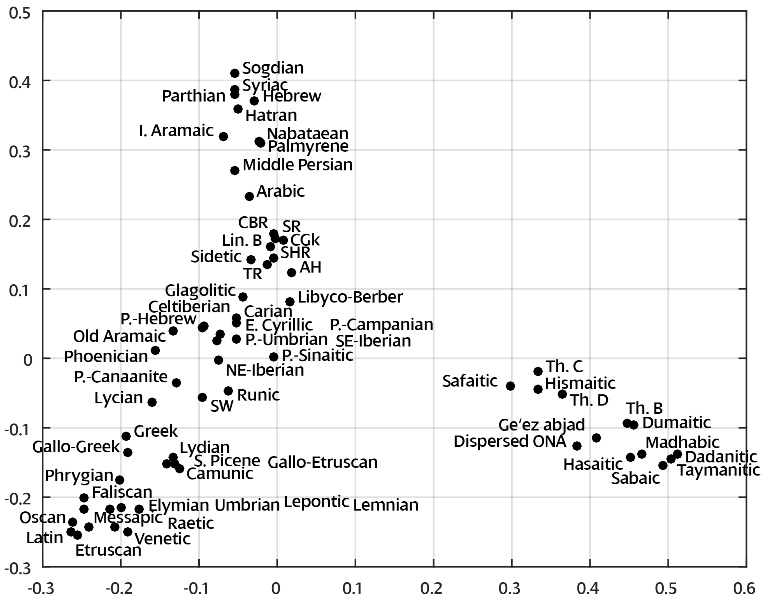


Figure 3: Multidimensional Scaling (MDS), $m = 66$, $n_{MDS} = 2$

According to Botto, between the late 10th and the early 9th c. BC, the Phoenicians used their strongest bond, with the Cypriot element, to penetrate the southern Tyrrhenian Italian and Sardinian markets. Moreover, in the late Bronze Age connections existed between southern Iberia and Sardinia. There was an alliance between the main Phoenician and Cypriot coastal cities. In the 11th–10th c. BC between Cyprus and Sardinia, the relationships became vital. After the fall of Mycenaean power, the Cypriots played a significant role in trade between Levante and the western part of the Mediterranean (Botto 2016).

Another important fact is that the P.-Hispanic scripts are syllabic for the plosives and alphabetic (monophonemic) for the rest of the consonants and the vowels. Moreover, the syllabic graphemes for the plosives do not mark a voicing contract. This is the reason why they are called semi-syllabaries. Especially interesting is the so-called *principle of redundancy*, which means that in one of the P.-Hispanic scripts, the SW, each syllabic grapheme is accompanied by a redundant grapheme representing the vowel of the syllabic grapheme (Valério 2008, 112; 2014, 440). A possibly related fact is

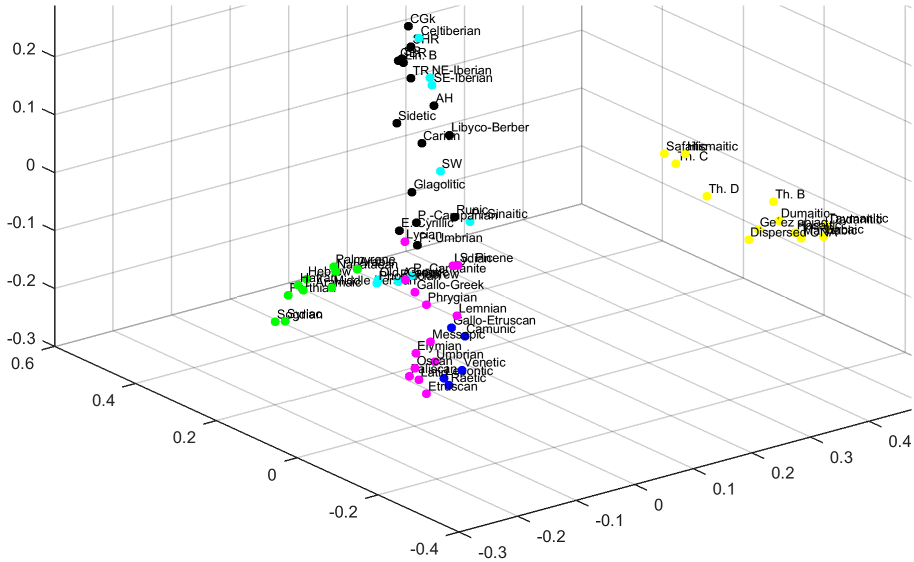


Figure 4: K-means mapped to MDS, $m = 66$, $n_{MDS} = 3$, $K = 6$

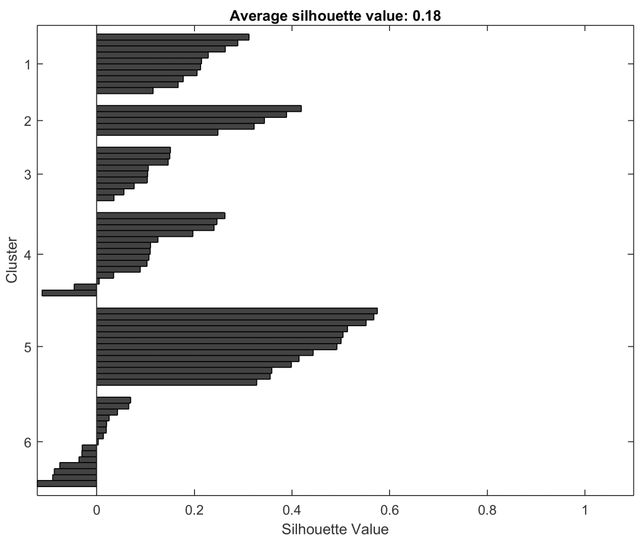


Figure 5: Silhouette values of k-means clusters

that in the Assyrian cuneiform and the AH scripts, word-ending long consonants are represented with the <CV> + <V> grapheme combination, e.g. <ki>+<i> represented /kī/ (Segert 1978, 111–112). Consequently, it is not ruled out that the P.-Hispanic scripts were affected by the AH. This conforms to Valério's supposition—citing Craig Melchert—that the AH ↾ *376 <zi> at least indirectly affected the SW script (Valério 2008, 130–131).

Considering the Ancient Italic scripts, the northern version of the Etruscan script probably originated from Lydia, and the southern version could be from Cilicia (Woudhuizen 1982–1983, 98). Woudhuizen claimed that in Pithecussae there was the presence of Lydians in the 8th c., who disappeared in the 7th c. BC. According to Woudhuizen, Lydian and eastern Greek merchants founded Pithecussae in the early 8th c. and Cumae in the late 8th c. According to Szabó (2015, 352), the Etruscan territories around Bologna were affected by northern Balkan and Hallstatt archaeological features, and oppositely, the southern Etruscan areas were influenced by Anatolian features. These archaeological data are used only in support of the results presented here, and no archaeological conclusions are drawn. However, a possible consequence of these (and several others, not cited) archaeological data could be that the culture in Italy was heterogeneous, which could lead to the preservation of ancient glyphs in the orthographies. Considering the dendrogram obtained by the Ward method in figure 2, it is interesting to note the strong relationship between the Runic, some Ancient Italic, and AGA scripts. It could imply that the spread of writing knowledge in Italy happened in multiple waves. In such case, the Runic maybe preserved an early layer of literacy in Italy. This approach is not contradictive to the model of Looijenga (1997, 55–56), who demonstrated that the Runic originated by adaptation of some kind of Northern Italic local script in Romanized regions along the Rhine.

To summarize, besides the Phoenician, there was another source of the P.-Hispanic, Libyco-Berber, Ancient Italic, and Runic scripts that could be rooted in the eastern Mediterranean.

5.2 An approximative model for the origin of the Rovash scripts

Based on the SFGs in table 10, there are several graphemes that are very similar to the Rovash graphemes. The possible donors or close indirect relatives of the Rovash graphemes are summarized in table 11 and table 13 with SFG references to table 10. The graphemes of scripts that might have been earliest affected by the Rovash script are listed in table 11. Considering the very early age of use of the Lin. A, Lin. B, and CM scripts (table 4), if similar glyphs exist in different Aegean scripts, only the latest occurrence is denoted in table 11, usually in the CGk script. However, in some cases, the most similar glyph occurred in earlier scripts than the CGk. The Rovash ᚿ, ✕ <d, d^2> grapheme was left out from table 11, since it is not possible to choose the most probable source due to the large number of candidates in SFG-174.

Group	Script	Probable donor or indirect relative of the Rovash glyph	Sum
	CM	⚙ CM0 09 <ka?> (SFG-87); ⋇ CM 112 <k/ze?> (SFG-92, or ≩ LA 74 <ze?>); CM ♈, ♈ 70 <ki?> (SFG-97, cf Lin. B 𐀓, 𐀓 <ki>, CGk 𐠌, 𐠌 <ki>); CM ▢ 75 <ra> (SFG-160, cf CGk 𐠣 <ra>); CM ⋒ 24 <le?> (SFG-162, cf CGk 𐠐 <le>)	16
Aegean	CGk	𐠼 <za?> (SFG-54); 𐠅 <ja> (SFG-75); 𐠊 <ka> (SFG-88); 𐠕 <me> (SFG-116); 𐠖 <mi> (SFG-118); 𐠗 <mo> (SFG-122); 𐠘 <mu> (SFG-123); 𐠚 <ne> (SFG-133); 𐠨 <sa> (SFG-141); 𐠩, 𐠩, 𐠩 <se> (SFG-145); ‡ <pa> (SFG-148, SFG-149)	
Anatolian hieroglyphic	AH	⅄ *451 <hur> (SFG-31); ↑, ↑, ↑ *376 <zi> (SFG-60); ☞ *315 <kar> (SFG-90); ◇ *423 <ku> (SFG-100, or CM ◊ 15 <ko?>); ⋐ *35 <na> (SFG-132); ᐳ *411 <ni> (SFG-135); 𝒪, ⁄ *383 <ra/i> (SFG-161); ǀ, ▯ *380 <sa_8> (SFG-169); O *402 <sa_4>, ◊ <us> (SFG-170); ⊲ <ti> (SFG-173); ⫼ *216a FINES (SFG-186)	11
Anatolian-Greek alphabetic (AGA)	Carian	⩓ <δ> (SFG-29); ⧗ <γ> (SFG-101); 〉 <d> (SFG-30)	
	Greek	◀ <α> (SFG-7); ∂ <β> (SFG-13); ꟼ <ε> (SFG-33); ⅎ <ϝ> (SFG-42); ⵜ <ζ> (SFG-52); ⊙, ϴ, ⊕, ⊗ <θ> (SFG-69, or Old Aramaic ⊗, ⦸, ⊘ <ṭ/ṯ>); ˪ <l> (SFG-108); Λ <λ> (SFG-109); ⲋ, ϟ <σ> (SFG-168); Y <τ> (SFG-176)	17
	Lycian	Y <g> (SFG-18);)(<q> (SFG-103, or CGk 𐠎 <ku>)	
	Sidetic	ㄣ <a> (SFG-12, cf AH ʕ *19 <á>); И <g> (SFG-98, or Greek ⟆ <kh?>)	
Ancient Semitic & Canaanite	Proto-Sinaitic	Ψ, 𐤄 <h> (SFG-31); 𐤇, 𐤇, 𐤇 <ḫ/ḥ> (SFG-68)	
	Phoenician	𐤄 <h> (SFG-33); 𐤌 <m> (SFG-112, or Old Aramaic 𐡌 <m>); 𐤍 <n> (SFG-124)	6
	Old Aramaic	𐡓 <r> (SFG-158)	
S. Semitic	S. Semitic	𐩱 <’> (SFG-4, cf Phoenician 𐤀 <’>); Y, Y, Ψ, 𐩠 <h> (SFG-32); 𐩺, 𐩺 <y> (SFG-47); 𐩻, 𐩻 <ṯ> (SFG-57); 𐩴 <g> (SFG-102); 𐩪 <s^1> (SFG-143, cf CGk 𐠨 <sa> in SFG-141); Λ <s^1> (SFG-144, cf *AH* ෧ *415 <sa>, CGk V <sa>); 𐩧 <r> (SFG-164, cf CGk 𐠤 <re> and Sidetic Λ <r> in SFG-163); 𐩩 <t> (SFG-175)	9

Table 11: Presumably direct donors or close indirect relatives of the Rovash graphemes

The SFGs suggest that the common ancestor of the Rovash scripts had to have been developed in Anatolia, after the distribution of the Semitic consonantal scripts (Proto-Sinaitic, Phoenician, Old Aramaic), likely before the end of the syllabaries, which originated from the AH and the Aegean scripts, and surely before the 3rd c. BC, when the Greek script became dominant in Anatolia.

It is known that the Turkic Rovash (TR) originated from the nomad region of the Altai Mountains (Vasil'ev 1994, 328). From this it follows that a presumably common ancestor of the Rovash scripts (Proto-Rovash) must have reached the Altai Mountains beforehand. According to Marsadolov (2000a, 247–250; 2000b, 51), during the 6th – 4th c. BC the Pazyryk *(Pazîrîk, Пазырык)* culture ruled the Altai region, and the descendants of the Cimmerians, who settled there after being expelled from Anatolia, may have participated in the growth of the Pazyryk culture, which was also influenced by the Chinese and Achaemenid Persian empire. In the first half of the 6th c. BC, numerous innovations appeared in the Altai region which, according to Marsadolov, can be linked to the arrival of nomadic tribes from Anatolia at the end of the 7th c. BC or the beginning of the 6th c. BC. Presumably the nomadic tribes from Gordion or the surrounding region settled the most fertile valleys, Tuekta and Bashadar, assuming rule over the local Pazyryk population. The 4000 km distance between Anatolia and the Altai region was not insurmountable, in part due to the existence of trade routes. The nomads could have crossed this distance in as little as one to two years.

The Altai region later became part of the Yüeh-chih *(Yuèzhī)* Empire, and the sites at Pazyryk should be related to the Yüeh-chih (Enoki et al. 1994). According to Harmatta, the Yüeh-chih is known as Tochari in Greek and Latin sources. Between 203 and 177/176 BC, the Hsiung-nu *(Xiongnu)* defeated the Yüeh-chih, who migrated to the west (Harmatta 1994).

In the necessary timeframe (7th–6th c. BC), there is knowledge of only a single ethnic group, the Cimmerians, who could have taken literacy from Anatolia to the East. The Cimmerians seized Phrygia from King Midas in the first half of the 7th c. BC. During the same period, Caria fell to the Lydians (Adiego 2007b, 758). For generations the Cimmerians lived around Gordion (the late Phrygian capital), making two attempts to capture Lydia during 650–640 BC. Eventually the Lydian ruler Alyattes expelled the Cimmerians from Anatolia in the late 7th – early 6th c. BC (Marsadolov 2000a, 249).

If the Cimmerians borrowed the later Rovash graphemes, the S. Semitic scripts could not be ancestors of these graphemes, since they had no known contact. More probably, both the Rovash and S. Semitic scripts originated from a common region; see also table 12. Based on the SFGs, it seems very likely that there are numerous S. Semitic graphemes of *non-Phoenician origin.* Moreover, these non-Phoenician S. Semitic glyphs appear in other scripts of the first half of the 1st millennium BC. Table 12 presents each presumably non-Phoenician S. Semitic grapheme and the occurrence

S. Semitic	Ancient Italic	Aegean	AH	AGA	Libyco-Berber	P.-Hispanic	Rovash	Runic
ꓫ <’> (SFG-3)	0	0	0	0	0	0	1	0
ʏ <’> (SFG-32)	0	0	0	0	0	0	1	0
ʔ <y> (SFG-47)	0	0	0	1	0	0	1	0
⧗ <z> (SFG-53)	0	0	0	0	0	0	1	0
Ħ <ḍ> (SFG-70 and SFG-71)	0	0	0	0	0	1	0	0
O <g> (SFG-87)	1	1	0	1	0	1	1	1
◊, □ <g> (SFG-100)	1	1	1	1	0	1	1	1
ȣ <g> (SFG-101)	0	1	0	1	0	1	1	0
◊ <g> (SFG-102)	0	0	0	0	0	0	1	0
⌉ <l> (SFG-107)	1	0	0	1	0	1	0	1
ȝ <m> (SFG-118)	0	1	0	1	0	0	1	0
И <n> (SFG-127)	1	0	0	1	0	1	0	0
ɦ <n> (SFG-128)	1	0	0	0	0	0	0	0
λ <s^1> (SFG-143)	0	1	0	0	0	0	1	0
Λ <s^1> (SFG-144)	0	1	1	0	0	0	1	0
❯, ∩, – <r> (SFG-164)	0	0	0	0	0	0	1	0
ȣ, ʃ <ṯ>, ȣ, 8 <f> (SFG-166)	1	0	0	1	0	0	0	0
ʒ <s^2> (SFG-168)	1	0	0	1	1	1	1	1
I <s^2> (SFG-169)	0	0	1	0	0	0	1	0
ϯ, ϯ, ✕ <t> (SFG-175)	0	0	0	0	0	0	1	0
boustrophedon (SFG-182)	1	1	1	1	1	1	1	1
spiral or circle (SFG-183)	1	0	0	0	1	0	1	0
bottom-up (SFG-184)	0	0	0	0	1	0	0	0
Summary	9	7	4	10	4	8	17	5

Table 12: Occurrence of cognates of non-Phoenician S. Semitic graphemes

of their counter pairs in other scripts based on the SFGs in table 10. For this study only, SFG-70 and SFG-71 were hesitantly unified.

The resulting numbers of cognate graphemes in table 12 cannot be evaluated quantitatively, since the populations of each group of the examined scripts are largely different. Nevertheless, it can be observed that scripts other than Rovash had a significantly weaker relationship with the S. Semitic scripts. Consequently, the region where graphemes were transferred to a supposed ancestor of the Rovash scripts was presumably not farther from the region from which the S. Semitic scripts borrowed certain graphemes than was the region that lent graphemes to other examined scripts. Based on the known historical data, the S. Semitic groups did not reach any region in Anatolia except a part of the Neo-Hittite states (Syria and Southeastern Anatolia). Therefore, the region lending graphemes to Rovash scripts could not be far from the Neo-Hittite states.

According to Macdonald, two forms of the Sabaic script (a kind of S. Semitic) are the formal musnad and the informal, cursive zabūr. Several zabūr relics have been carbon dated and found that the oldest one was from the period 1150-901 BC with a confidence of 2σ (94%) (Macdonald 2009, Addenda and Corrigenda, 10). Consequently, the common ancestor of the S. Semitic and Rovash scripts could not have developed later than the 11^{th} c. Since the start of the CGk is about the 11^{th} c. BC (Valério 2016, 237), this may justify that the most similar Rovash graphemes may have come from the CM script, which was still used in the 11^{th} c. BC (Valério 2016, 27), rather than the CGk (table 11).

Lehmann claims that in the 12^{th}-11^{th} c. BC, both Syria and Cilicia were affected by the Aegean culture in part due to the Aegean settlers in the costal regions and also due to the Aegeans' trade with Syria, Lebanon, Cyprus, and Cilicia at the end of the Bronze Age (Lehmann 2013, 265, 325, 328). According to Yakubovich, the Cilician leaders were of Greek-speaking Aegean descent in the Early Iron Age (Yakubovich 2015b, 35–36, 38, 40–41). Thus, the Cypriot scripts could have affected Cilician literacy.

Que (Assyrian name; its Luwian form was Hiyawa) situated on the Cilician plain was one of the Neo-Hittite states (Yakubovich 2015b, 49). Greek pottery from the 12^{th}–11^{th} c. BC is found in large quantities in the Cilician plain. The Greek settlers in Pamphylia succeeded in establishing their linguistic dominance in this region. Cilicia represents the only region where Luwians and Greeks may have coexisted. A neighbour of the Greeks in Southwestern Anatolia was the Carians (Yakubovich 2008, 200). The main official language of Que was not Luwian, even though Luwian was historically spoken by the bulk of its population. The socially dominant language was Greek, and the attested written language is Phoenician.

According to Yakubovich, the Phoenician language was emblematic of the rulers of Que, who claimed Greek descent, and the Luwian language was used by the indigenous population of Que from before the collapse of the Hattusa empire. Yakubovich claims that the adoption of Phoenician as a language of written expression by the Greek colonists in Cilicia happened at the point when the Linear B script had been forgotten and represented the first step toward the creation of the Greek script. Furthermore, the Greek script originated from Cilicia in the late 9^{th} century BC. In Que, no Semitic personal names are attested to in these inscriptions in connection with local individuals. Valério (2008, 116) claims that the Phoenician script was used for recording Luwian personal names. Swiggers (1996, 266–267) stated that the Cilicians

Group	Invoked graphemes of the donors	Sum
Aramaic & Persian	I. Aramaic 𐡅 <w> (SFG-44); I. Aramaic 𐡉, 𐡉 <y> (SFG-81); I. Aramaic 𐡐 <p> (SFG-151); I. Aramaic 𐡕 <t> (SFG-181); Sogdian orthographical rule <w>+<y> for representing /ö, ü/ (uncertain, see comments in SFG-45); Middle Persian 𐭉 <y> (SFG-83, or Palmyrene 𐡩 <y>);	6
Slavic	Glagolitic: Ⱁ <o> (SFG-146); E. Cyrillic: Ѫ <ǫ> (SFG-147)	2

Table 13: Sporadic influence on the Rovash scripts

could have adopted the Phoenician script but only used it for the inscriptions in the Phoenician languages. Thus it is proven that the Phoenician script was present in Cilicia.

Although no local script relics in the Greek language have been found for the relevant place and period, several new findings and methodological advances made since the year 2000 have strengthened the case for a Greek existence in Early Iron Age Cilicia (Yakubovich 2015b, 49). The Early Iron Age assemblages excavated in Cilicia match those of the northern Levant in attesting to the presence of materials connected with the Aegeans. From the period between the 12th to mid-8th c. BC, no AH inscription was found in Cilicia. Cilicia is the only region of south-central Anatolia and northern Syria in which a Neo-Hittite tradition begins in the very late 8th century BC without any earlier trace of a post-Hittite tradition (d'Alfonso and Payne 2016). Consequently the AH script did not dominate in Cilicia.

Based on the above data and geographical factors (the Cimmerians were neighboured by Cilicia), it is likely that the Rovash graphemes originated from the region around Cilicia. Similarly, non-Phoenician S. Semitic glyphs in table 12 may also have originated from the Cilicia region.

In the period 700–200 BC, the I. Aramaic, from 1st c. BC to 7th c. AD, the Late Aramaic, and the Middle Iranian in Central Asia, and around 10th c. AD, the Slavic scripts in the Carpathian Basin affected the Rovash scripts by the graphemes listed in table 13. It is noteworthy that some of the I. Aramaic graphemes (e.g., 𐡅 <w>, SFG-44) could have been borrowed in the earliest time, maybe even in Anatolia.

The Rovash ᐳ, ᚬ, ᚭ, ᚿ <W> (SFG-44) may have originated from the Old Aramaic 𐤅, 𐤅 <w> or the I. Aramaic 𐡅, 𐡅 <w>, and surely not from the Old Aramaic 𐤅, 𐤅, 𐤅, 𐤅, 𐤅 <w> (SFG-41). In the case of the Rovash <i, y>, the typical glyphs are ᚫ, ᛚ, ᚴ (SFG-81), which are unrelated to the Phoenician 𐤆, Old Aramaic 𐤆, and I. Aramaic (7th c. BC) 𐡆, (6th c. BC) 𐡆, 𐡆 <y> (SFG-77), but more probably related to the Old Aramaic (7th c. BC) 𐡉, 𐡉, (6th c. BC) 𐡉, 𐡉, 𐡉, 𐡉 <y> (SFG-81). Consequently, the adaptation of

TR grapheme	Supposed meaning in Old Turkic	SFG with some example glyphs
𐰉, 𐰊 <b[2]>	*äb, äβ* 'tent, house'	SFG-116, e.g., P.-Hispanic 𐰉 <be>
𐰓 <d[2]>	*ed* 'property, livestock'	SFG-174, e.g., P.-Hispanic X <ta> /da, ta/, Ancient Italic X, + <t, T> /d, t/
𐰍, 𐰎, 𐰏 <g[1]>	*aγ* 'net'	SFG-103, e.g., Lycian)(, ↄ(, ⚹ <q>
𐰖, O, 𐰗 <y[1]>	*ay* 'moon'	SFG-75, e.g., CGk O, 0, ϙ <ja>
𐰶, 𐰷 <k[3]/[ï]k[ï]>	*ïq* 'spindle'	SFG-97, e.g., Carian ▽, ∇, △ <k>
↓, ↑ <k[5], [w]k[w]>	*oq* 'arrow'	SFG-88, e.g., CGk ⇞ <ka> /ga, ka, k^ha/
Y <l[2]>	*el* 'hand'	SFG-141, e.g., CGk Y <sa> (see comments in SFG-141)
𐰤, 𐰥 <n[2]>	*en* 'declivity'	SFG-133, e.g., CM 𐰥, 𐰥 34 <ne?>; CGk 𐰤, 𐰥 <ne>
ϒ <r[2]>	*er* 'man'	SFG-165, e.g., Lin. B ϒ <ru> /lu, ru/, Lydian ϒ, ϒ, ϒ <λ>
𐱃, 𐱄, ⌒, ⌒, ⌒, ⌒, ⌒, ⌒, ⌒, ⌒, ⌒ <t[1]>	*at* 'horse'	SFG-29, e.g., Carian ⩓, ⌃ <δ> /md/d/^{n}t/, SW ⩔ <t[o]>

Table 14: The relationship of the TR graphemes that were traditionally supposed to be ideograms

the Aramaic <y> may have happened in the 7th– 6th c. It is noteworthy that there is more cursive Rovash ↄ <i, y> (SFG-83), which had to have been adapted in the period 1st c. BC – 7th c. AD. Due to historical reasons, this adaptation had to have happened in Middle Asia. In the case of the Rovash Ч, Ƴ, N, H <r>, the ancestor is surely the Aramaic 𐡓 <r> (SFG-158). The strictly geometric forms of the Rovash glyphs point to an early adaptation; however, cf Turkic Rovash ˥, ˥ <r[1]> and I. Aramaic (Aśoka, ca. 250 BC) ˥ <r>.

5.3 The question of the TR ideograms

Several authors have hypothesized that some of the TR graphemes originate from ideograms (pictograph, tamgha). The history of this direction of research is summarized in Róna-Tas (1987, 8). However, similar counterpairs of the TR graphemes in question (Róna-Tas 1987, 9) can be found in the SFGs of table 10, as is demonstrated in table 14. The listed example glyphs in the last column of table 14 are usually not direct relatives of the appropriate TR graphemes; however, they show the probable relationships of the TR graphemes in question. Using the *lex parsimoniae,* it is unnecessary to assume they have an ideogrammatic origin.

5.4 Syllabic traces in the Rovash scripts

There are traces of syllabary in the Turkic Rovash; namely, Kyzlasov (1994, 131) explored that the TR is partly a syllabary. Kyzlasov claimed that the ancestor of the TR (he called: проторуническое слоговое письмо) goes back not to alphabetic systems but to the ancient, probably Semitic, syllabaries of an unknown (not West Semitic) origin. He claimed that the ancestor script developed by eliminating a part of the earlier used presumed syllabic graphemes, and the surviving Orkhon and Yenisei inscriptions demonstrate the final stage of this process. According to Kyzlasov, the ancestor of the TR was not invented but borrowed. He further supposed that among many such systems the ancient Turkic "linguists" wisely chose the alphabetical system best suited to the Turkic language. The outward similarity of the symbols of the various Euro Asiatic and Asiatic Turkic inscriptions can be explained by their basis, the ancient Semitic scripts of Central Asia. Each of these versions of writing systems used for TR inscriptions was formed under different conditions and on a different basis. In the reconstruction by Kyzlasov, most of the consonants are denoted by two different kinds of graphemes, depending on the vowel in the syllable of the consonant (velar or palatal sound values). A consonant is called *velar* if it is used near back vowels, and it is called *palatal* if it is used near front vowels. Consonants are harmonized with the vowels of their syllables. Graphemes that represent consonants next to back and front vowels are transliterated by adding a superscript 1 or 2, respectively, to the transliteration value of the consonant, e.g. b^1 and b^2.

As Erdal (2004, 39) pointed out, synharmonism (vowel harmony) and the presence of the front rounded vowels *ö* and *ü*, both are equally untypical of Semitic, Caucasian, East Asian, and Early Indo-European. The TR script distinguishes front and back harmony in rounded vowels and also in consonants; there are, for example, sets of very different-looking graphemes for front *b* and back *b,* front *y* and back *y,* etc. The palatal consonant *y* is sometimes used in the Old Turkic language beside front vowels. Semitic scripts distinguish only between velar and uvular /k/ (*'k'* and *'q'*) and /g/ (often noted *g* and *γ* respectively), a distinction which has been used for expressing synharmonism in Turkic languages. A further specific feature of TR is the preponderance of closed syllables as opposed to open ones. For example, unlike Semitic and Indo-European scripts, the grapheme for a consonant *t* implies not a following vowel, but a preceding vowel. Moreover, all coda vowels are written out as separate features in the TR, again unlike the Semitic and Indic scripts (Erdal 2004, 39–40). Possibly related to Erdal's observation is that in the earliest SHR relics, the consonant grapheme names begin with a vowel (to ease pronunciation), different from the usual European practice where the vowel is placed after the consonant.

Following are known synharmonism of consonants in the TR: /b/, /p/, /d/, /t/, /g, γ/, /k, q/, /t/, /l/, /n/ /r/, /s/, /j/, and /ïq/, /oq/, /üq/. Moreover, as Kyzlasov claimed, certain graphemes could have been used for syllables /ït/, /ïš/, /ïs/, /id/ed/, /ič/eč/, /im/em/. His reconstruction supports the possibility that the common ancestor of the Rovash scripts originated from at least partly syllabaries. However, no known Rovash script is a syllabary. Even synharmonism exists only in the TR. Sporadically and not consequently, the SR also applied synharmonism in the case of some consonants, as Vékony (2004) demonstrated. In the CBR and SHR there are some consonants with multiple graphemes: In the CBR, for the /k/ and /t/, and in the SHR, for the /č/, /k/, /š/, /r/, and /t/, there are multiple graphemes; the reason for this has not been clarified. In the surviving CBR and SHR relics, usually there is no synharmonism. However, in a very few SHR relics, the differentiation of the <k> graphemes near front and back vowels can be detected. Moreover, in the Constantinople inscription, the grapheme 𐲔 <k> seems to represent also the syllable /aːk/ besides the consonant /k/ (table 17 and comments).

Consequently, the Rovash scripts may have preserved traces of an ancient syllabary, but there is no evidence for an ancient syllabary as the common origin of the Rovash scripts (Proto-Rovash). However, taking into account the fact that, according to the phenetic analysis, several graphemes of the semi-syllabic P.-Hispanic and the Rovash scripts (see SFGs in table 10) are markedly similar, it can be supposed that the Proto-Rovash could have had some syllabic property.

5.5 Witness scripts as a consequence of the centre-periphery effect

In the 3rd–2nd millennia BC, the centre of script development was in the Middle East. Presumably, the North-West Semitic and the S. Semitic writing traditions separated in the 2nd millennium BC (Macdonald 2015, 32). In the 1st millennium BC, it gradually diverged into the Aramaic world (east) and the Anatolian-Greek world, and later (classical) Italy. Using Macdonald's model for literate and non-literate societies, in these areas the societies were literate; therefore, these places can be considered central. Conversely, in the nomadic or partly nomadic Arabian Peninsula, Hispania, Northern Africa, and the Eurasian Steppe from the eastern Altai Mountains to the western Carpathian Basin, the societies can be modelled predominantly as non-literate; therefore, they are considered peripheral. Theoretically, the peripheries could preserve glyphs that were already forgotten in the centre in favour of the later developments. A centre–periphery (core–periphery) model can be used for the spatial distribution of certain glyphs.

The extracted SFGs (table 10) clearly show the significant similarities in several glyphs and orthographical rules in the Ancient Italic, Libyco-Berber, P. Hispanic, Rovash, Runic, and S. Semitic scripts. Taking into account some historical facts, these scripts probably originated from Levantine or the Anatolian coast. All of these scripts left Anatolia not later than the 7th c. BC; therefore, they could have preserved a certain state of the grapheme evolution in Anatolia. These groups of scripts can be qualified as witnesses of the graphemes used in Anatolia and the surrounding regions in the first half of the 1^{st} millennium BC.

It is noteworthy that the property of being a witness script is a relative quality, since a certain script could be witness of the development of another script, which could be witness of another. For example, the AGA scripts also witness the age of their development; they testify a mainly alphabetic environment from the early age of the Greek script. The beginning of the AGA scripts is about the 8^{th} c. BC (table 4), based on the earliest dated inscriptions. The accurate development of these scripts remains unknown, however, they did preserve even earlier graphemes, such as the Lycian ◊ <k> (SFG-100).

6 Conclusions

The paper presented a new composite data analysis method to explore the similarities between scripts. Computational palaeography concentrates on the topological relationships of each grapheme. The premise is that the glyphs of the graphemes are relatively stable during the development of the writing, and the changes can usually be described by well-defined rules. During this, the linguistic, historical, geographical, and archaeological circumstances are taken into account as accurately as possible.

The developed method starts with searching for sets of possible cognate glyphs. It utilizes the determined typical *characteristic transformations* of the topology of the glyphs, which can be observed on the evolution of the graphemes. The characteristic transformation usually does not change the visual identity of the original glyph. The topological and the visual identity layers belong to the layered grapheme model, which was developed for modelling the grapheme in computational palaeography. The developed data analysis method selects orthographical rules and sets of possible cognate glyphs from the phonetically similar graphemes by minimizing the necessary topological transformations between glyphs. In such way, the similarity features groups are constructed. Then various machine-learning methods are applied to obtain a phenetic model for the investigated scripts based on the similarity group of features. In this stage, the multidimensional scaling and various clustering algorithms were applied. The obtained results give an overall picture about the phenetic relationships of the examined scripts. In order to filter out the possible homoplasies, a cladistic approach was also used, in a limited fashion.

Some special concepts were elaborated and introduced in the computational palaeography in order to apply the phylogenetic methods for palaeography. Beside the existing term of characteristic transformation, the concept of the witness script is also introduced. A script is taken as a *witness script* for a certain area and time period if the continued evolution of the script happened in isolation. Further new concepts are the glyph- and grapheme-level reticulations as reticulate events. A *glyph-level reticulation* occurs if part of the glyphs of a grapheme is borrowed from another script, and a *grapheme-level reticulation* exists if all glyphs of a grapheme are borrowed in a certain evolutionary event.

The results show the usability of the phenetic approach combined with cladistic elements in exploring the similarities of scripts. The present study concentrated on the phenetic analysis of Mediterranean-origin scripts; but the presented method could be extended to other writing systems. The main goal was to prove the usability of the combined exploratory data analysis method; however, during the evaluation of the resulting phenetic model, some approximative consequences can be derived about the relationships of the examined scripts as follows. *(i)* Some groups of witness scripts are identified which attest the state of the grapheme evolution in the first centuries of the Iron Age in the Mediterranean. These are the S. Semitic, the P.-Hispanic, the Ancient Italic, the Libyco-Berber, the Runic, and the Rovash. *(ii)* The origin of these witness scripts is at least partly connected to south Anatolia. *(iii)* The probable source of the Rovash graphemes was approximately determined as the region of Cilicia before the 6^{th} c. BC.

The developed method for script analysis might be used for further applications. Changing the focus of the research, it is possible that the basic taxonomical unit (taxon) is not the script, but a version of the script (e.g., grapheme set of the medieval English orthography), or a certain writing style, typography, and so on. The introduced approach may give support to palaeographers in exploring the relationships among scripts and deciphering ancient inscriptions. The present method can be highly automatized; therefore, it could be scaled to library-wide databases.

7 Appendix: Examples of Rovash inscriptions

7.1 A quadrilingual CBR inscription of the Golden Treasure of Nagyszentmiklós

The Golden Treasure of Nagyszentmiklós is a tableware collection of 23 gold pieces found in Nagyszentmiklós, Hungary (currently Sânnicolau Mare, Romania) on 3 July 1799. The treasure is unique in the region; the total weight is 10 kg. Its style cannot be connected to any great cultural center; most probably, it is a local product, made in the 7^{th}–8^{th} c. AD (Bálint 2010); however, the majority of the inscriptions could have been carved later. The names of the beverages to fill the jugs and the names of the foods to be served on the plates were carved onto the bottom of the pieces in CBR script (Vékony 2004). That is why the Rovash texts are mainly names of drinks and food.

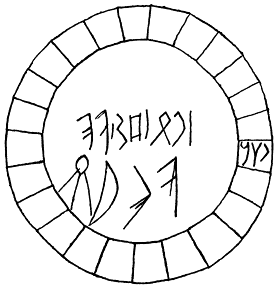


Figure 6: Drawing of the quadrilingual Nagyszentmiklós inscription (Hampel 1884)

First row in the middle	[illegible] /sïu̯ *s^{o}r$^{\ddot{i}}$m/ *(right part in Ogur),* /β^{i}z^{i}/ *(left part in Hungarian)* 'water [and] beer/wine' *(right part),* 'water' *(left part)*
Second row in the middle	[illegible] /v^{o}d^{o}jɔ̃/ *(in Slavic)* 'with water'
Text in the sidewall frame	[illegible] /dan$^{(u)}$/ *(in Alan)* 'water'

Table 15: The transliteration, transcription, and translation of the quadrilingual Nagyszentmiklós inscription

One piece of the treasure, jug No. 6, is a unique relic, since the inscriptions on its underside (fig. 6) are in different languages having similar meaning (table 15). In the first row there is a short vertical bar, which separates the Ogur (Turkic) text (right) and the Hungarian text (left). The detailed palaeographical analysis of this inscription, including the alternative readings, is published in Hosszú and Zelliger (2014a).

7.2 The Vargyas SHR inscription

A Székely-Hungarian Rovash inscription was found in Vargyas (Romanian: Vârghiș, Székelyland, Romania) on a stone in a church in 1994 (Benkő 1994, 487–489). For linguistic reasons the text could not have been created later than the 12th–13th c. AD (Zelliger 2016). The drawing of the inscription is presented in figure 7. The writing of the inscription is sinistrograde; see table 16. The sentence in the inscription was cited from the Gospel of John (Ioh 19:26): "*Woman, behold your son.*" Zelliger claimed that

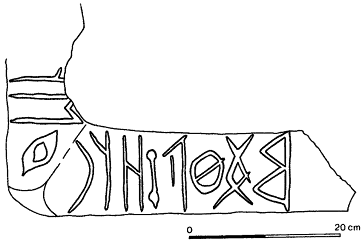


Figure 7: Drawing of the Vargyas inscription (Benkő 1996a, 79; 1996b, 31–33)

Transliteration with Rovash graphemes	ΥΜ◊Ͻ Υ ΝꞮ1θ ⨳ᐊ
Transcription with phonetic symbols	/^{i}mē fioγ t^{e} n^{ä}küd/
Translation to English	'[Woman,] behold your Son'

Table 16: The transliteration, transcription, and translation of the Vargyas inscription

the SHR text of the Vargyas Inscription resembles a Greek translation of the Bible. The detailed palaeographical analysis of this inscription, including the alternative readings, is published both in Zelliger (2016) and in Hosszú (2013).

7.3 The Constantinople SHR inscription

In 1515 in Constantinople (Istanbul), Barnabas Bélay, the ambassador of the Hungarian King Vladislaus II (1490–1516), found he had to wait for two years for his admittance to the Sultan Selim I (1512–1520), and during this time, a Hungarian person named Thomas Kidei Székely wrote this SHR inscription on the wall of the Ambassadors' House. Between 1553 and 1555, the numismatist and epigraphist Hans Dernschwam (1494–1568 or 1569) discovered and copied it (fig. 8); later the building was destroyed in an accidental fire (Babinger 1914, Sebestyén 1915). The writing of the inscription is dextrograde; see table 17. The detailed palaeographical analysis of this inscription, including the alternative readings, is published in Zelliger and Hosszú (2014).

The inscription contains several ligatures, e.g., the symbol Ħ (first row) is maybe the ligature of the graphemes *F <e> (SFG-33) and *H <r> (SFG-158), the sound value of the ligature being /er/. The glyph *F is presumably the mirrored version of the SHR Ꞁ <e>, which is attested in the Dálnok and the Rugonfalva inscriptions (Hosszú and Zelliger 2013).

In the Constantinople inscription, the ᛏ <k> is used consequently in the syllables containing /a, ā/ vowels. Therefore, it is possible that the sound value of ᛏ <k> was /ak, aːk/ in the orthography of the Constantinople inscription. This is supported

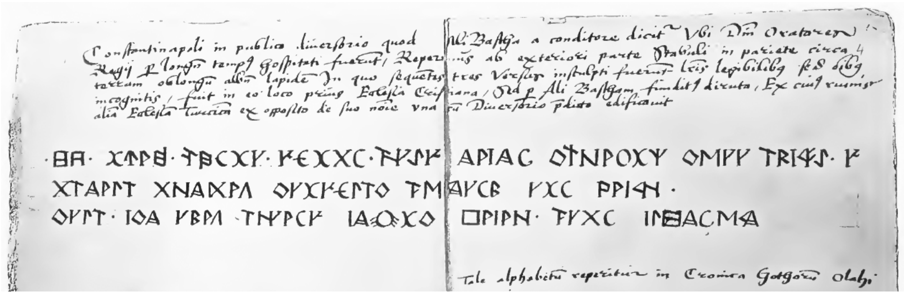


Figure 8: Copy of the original mural inscription in Constantinople (Sebestyén 1915)

First row	[illegible] /$^{\text{ä}}$zer e$\widehat{\text{ts}}^{\text{ts}}$āz tiz$^{\text{e}}$net $^{\text{ä}}$st$^{\text{ä}}$ndēb$^{\text{ä}}$n īrtāk/īrt$^{\text{ā}}$k $^{\text{ä}}$st lāslō kirāʎērt k$^{\text{e}}$v$^{\text{ä}}$t$^{\text{ä}}$t jārɔtt$^{\text{ɔ}}$k $^{\text{i}}$tt/ ‘It was written in 1515; delegate of King Vladislaus was sent here.’
Second row	[illegible] /bīlɔji b$^{\text{ɔ}}$rlɔbāš k$^{\text{ä}}$t äst$^{\text{ä}}$nd$^{\text{ä}}$jik it vɔlt; n$^{\text{ä}}$m tēn $\widehat{\text{tʃ}}$āsār/ ‘Barnabas Bélay waited here for two years; the emperor did not do [anything for them].’
Third row	[illegible] /k$^{\text{i}}$d$^{\text{ä}}$ji s$^{\text{ē}}$k$^{\text{e}}$l t$^{\text{ɔ}}$māš īrtān $^{\text{ä}}$st, s$^{\text{e}}$l$^{\text{i}}$mbēk $\widehat{\text{tʃ}}$āsār id$^{\text{ä}}$ tēn sāz lōvɔl/ ‘Thomas Kidei Székely wrote here, Emperor Selim housed here with one hundred horses.’

Table 17: The transliteration, transcription, and translation of the Constantinople inscription

by the fact that in the word [illegible] /īrtāk/īrt$^{\text{ā}}$k/ ‘written’ the /ā/ is not written with an individual grapheme; however, the long vowels were generally written even in the early Rovash inscriptions (Zelliger and Hosszú 2014).

Bibliography

Adiego Lajara, Ignasi-Xavier. [Ignacio-Javier Adiego Lajara.] “Los alfabetos epicóricos anhelénicos de Asia Menor.” In Bádenas de la Peña, Pedro et al. (eds.). *Lenguas en contacto: el testimonio escrito.* Madrid: Consejo Superior de Investigaciones Científicas, 2004. 299–320.

Adiego Lajara, Ignasi-Xavier. [2007a.] *The Carian language.* Leiden: Koninklijke Brill, 2007.

Adiego Lajara, Ignasi-Xavier. [2007b.] “Greek and Carian.” In Christidis, Anastassios-Fivos (ed.). *A History of Ancient Greek. From the Beginnings to Late Antiquity.* Cambridge, New York (NY): Cambridge University Press, 2007. 758–762.

Adiego Lajara, Ignasi-Xavier. [2007c.] "Greek and Lycian." In Christidis, Anastassios-Fivos (ed.). *A History of Ancient Greek. From the Beginnings to Late Antiquity.* Cambridge, New York (NY): Cambridge University Press, 2007. 763–767.

Adiego Lajara, Ignasi-Xavier. [2007d.] "Greek and Lydian." In Christidis, Anastassios-Fivos (ed.). *A History of Ancient Greek. From the Beginnings to Late Antiquity.* Cambridge, New York (NY): Cambridge University Press, 2007. 768–772.

Adiego Lajara, Ignasi-Xavier. [2007e.] *The Spread of Alphabetic Writing among the Non-Greek Peoples of Anatolia.* Unpublished, 2007.

Adiego Lajara, Ignasi-Xavier. "Lycian nasalized preterites revisited." *Indogermanische Forschungen* 120 (2015). 1–30.

d'Alfonso, Lorenzo and Annick Payne. "The Paleography of Anatolian Hieroglyphic Stone Inscriptions." *Journal of Cuneiform Studies* 68 (2016). 107–127.

Anders, Gunter. *Luwhitta/B. Luwische Hieroglyphen Fonts.* 2012. <http://www.hethport.uni-wuerzburg.de/luwglyph>.

Aussems, Johannes F. A. "Christine de Pizan and the scribal fingerprint—A quantitative approach to manuscript studies." Utrecht: unpublished Research Master's Thesis, 2006. <http://dspace.library.uu.nl/handle/1874/12537>.

Aussems, Mark, and Axel Brink. "Digital Palaeography." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßen (eds.). *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age (KPDZ 1).* Norderstedt: Books on Demand, 2009. 293-308.

Babinger, Franz. "Eine neuentdeckte ungarische Kerbinschrift aus Konstantinopel vom Jahre 1515." *Ungarische Rundschau* III (1914). 41–52.

Baeza-Yates, Ricardo A. "Introduction to data structures and algorithms related to information retrieval." In Frakes, William B., and Ricardo A. Baeza-Yates (eds.). *Information Retrieval: Data Structures and Algorithms.* Upper Saddle River (NJ): Prentice-Hall, Inc, 1992. 13-27.

Bakkum, Gabriël C. L. M. *The Latin Dialect of the Ager Faliscus. 150 Years of Scholarship. Part I.* Amsterdam: University of Amsterdam, 2009.

Bálint, Csanád. *Der Schatz von Nagyszentmiklós: archäologische Studien zur frühmittelalterlichen Metallgefäßkunst des Orients, Byzanz' und Steppe.* Budapest: Balassi Kiadó, 2010.

Bar-Joseph, Ziv, David K. Gifford, and Tommi S. Jaakkola. "Fast optimal leaf ordering for hierarchical clustering." *Bioinformatics* 17 (2001). S22–S29.

Barbançon, François et al. "An experimental study comparing linguistic phylogenetic reconstruction methods." *Diachronica* 30 (2013). 143-170.

Benkő, Elek. "Középkori rovásfelirat Vargyasról." [Medieval Rovash inscription of Vargyas.] *Magyar Nyelv* 90 (1994). 487-489.

Benkő Elek. [1996a.] "A székely rovásírás korai emlékei." [The early relics of the Székely Rovash script.] *Magyar Nyelv* 92 (1996). 75–80.

Benkő Elek. [1996b.] "A székely rovásírás. A legújabb kutatások." [The Székely Rovash script. The latest research.] *História* 18 (1996). 31–33.

Bernal, Martin. *Cadmean Letters. The Transmission of the Alphabet to the Aegean and Further West before 1400 B.C.* Winona Lake (IN): Eisenbrauns, 1990.

Berry, Michael J. A. and Gordon Linoff. *Data Mining Techniques for Marketing. Sales and*

Customer Support. New York (NY): John Wiley & Sons, Inc, 1996.

Beyer, Klaus. *Die aramäischen Inschriften aus Assur, Hatra und dem übrigen Ostmesopotamien: (datiert 44. v.Chr. bis 238 n.Chr.).* Göttingen: Vandenhoeck & Ruprecht, 1998.

Boisson, Claude. "Conséquences phonétiques de certaines hypotheses de déchiffrement du carien." In Gianotta, Maria Eliana et al. (eds.). *La decifrazione del Cario.* Rome: Consiglio nazionale delle ricerche, 1994. 207–232.

Bordreuil, Pierre. "Migraines d'Épigraphiste." In Bienkowski, Piotr, Christopher Mee, and Elizabeth Slater (eds.). *Writing and Ancient Near Eastern Society. Papers in Honour of Alan R. Millard.* New York (NY), London: T&T Clark International, 2005. 15–28.

Botto, Massimo. "The Phoenicians in the central-west Mediterranean and Atlantic between 'precolonization' and the 'first colonization'." In Donnellan, Lieve, Valentino Nizzo, and Gert-Jan Burgers (eds.). *Contexts of Early Colonization.* Rome: Palombi Editori, 2016. 289–309.

Brixhe, Claude and Michel Lejeune. *Corpus des inscriptions paléo-phrygiennes.* Paris: Éditions Recherche sur les civilisations, 1984.

Christidis, Anastassios-Fivos (ed.). *A History of Ancient Greek. From the Beginnings to Late Antiquity.* Cambridge (NY): Cambridge University Press, 2007.

Ciula, Arianna. "Digital palaeography: using the digital representation of medieval script to support palaeographic analysis." *Digital Medievalist* 1 (2005). DOI: 10.16995/dm.4.

Ciula, Arianna. "The Palaeographical Method Under the Light of a Digital Approarch." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßen (eds.). *Kodikologie und Paläographie im digitalen Zeitalter – Codicology and Palaeography in the Digital Age (KPDZ 1).* Norderstedt: Books on Demand, 2009. 219–235.

Cloppet, Florence et al. "New Tools for Exploring, Analysing and Categorising Medieval Scripts." *Digital Medievalist* 7 (2011). <http://www.digitalmedievalist.org/journal/7/cloppet>.

Colless, Brian E. "Proto-alphabetic inscriptions from the Wadi Arabah." *Antiguo Oriente* 8 (2010). 75–96.

Correia, Virgílio Hipólito. *A epigrafia da Idade do Ferro do Sudoeste da Península Ibérica.* Porto: Etnos, 1996.

Coulmas, Florian. *The Blackwell Encyclopedia of Writing Systems.* Oxford: Blackwell, 1999.

Cross, Frank M. "The Invention and Development of the Alphabet." In Senner, Wayne M. (ed.). *The Origins of Writing.* Lincoln (NE): University of Nebraska Press, 1989. 77–90.

Daniels, Peter T. and William Bright (eds.). *The World's Writing Systems.* New York (NY), Oxford: Oxford University Press, 1996.

Davies, Anna Morpurgo, and Jean-Pierre Olivier. "Syllabic scripts and languages in the second and first millennia BC." *British School at Athens Studies* 20 (2012). 105–118.

Davis, Brent. "Introduction to the Aegean pre-alphabetic scripts." *Kubaba* 1 (2010). 38–61.

Doblhofer, Ernst. *Jelek és csodák.* [Signs and wonders.] Budapest: Gondolat, 1962.

Dunn, Joe C. "Well separated clusters and optimal fuzzy partitions." *Journal of Cybernetics* 4 (1974). 95–104.

Egetmeyer, Markus. *Le dialecte grec ancien de Chypre.* Berlin; New York (NY): de Gruyter, 2010.

Enoki, Kazuo, Gennadij A. Koshelenko, and Z. Haidary. "The Yüeh-chih and their Migrations." In Harmatta, János (ed.). *History of Civilizations of Central Asia: The development of sedentary* *and nomadic civilizations: 700 B.C. to A.D. 250. Volume II.* Paris: UNESCO Publishing, 1994. 165–183.

Erdal, Marcel. "The runic graffiti at Yar Khoto." *Türk Dilleri Araştırmaları* 3 (1993). 87–108.

Erdal, Marcel. *A Grammar of Old Turkic.* Leiden: Koninklijek Brill, 2004.

Erdal, Marcel. *Personal e-mail communication.* 7 Nov. 2016.

Eska, Joseph F. "Continental Celtic." In Woodard, Roger D. (ed.). *The Ancient Languages of Europe.* Cambridge: Cambridge University Press, 2008. 165–188.

Farrujia de la Rosa, A. José et al. "The Libyco-Berber and Latino-Canarian Scripts and the Colonization of the Canary Islands." *African Archaeological Review* 27 (2010). 13–41.

Faulmann, Carl. *Das Buch der Schrift. Enthaltend die Schriftzeichen und Alphabete aller Zeiten und aller Völker des Erdkreises.* Wien: Druck und Verlag der Kaiserlichen Hof- und Staatsdruckerei, 1880.

Ferrer i Jané, Joan. "Novetats sobre el sistema dual de diferenciació gràfica de les oclusives sordes i sonores." *Acta Palaeohispanica* 5 (2005). 957–982.

Ferrer i Jané, Joan. "Els sistemes duals de les escriptures ibèriques." *Palaeohispanica* 13 (2013). 445–459.

Ferrer i Jané, Joan. "Ibèric *kutu* i els abecedaris Ibèrics [Iberian kutu and the Iberian Abecedaries.]" *Veleia* 31 (2014). 227–259.

Fischer, Franz, Christiane Fritze, and Georg Vogeler (eds.). *Kodikologie und Paläographie im Digitalen Zeitalter 2 – Codicology and Palaeography in the Digital Age 2 (KPDZ 2).* Norderstedt: Books on Demand, 2010.

Fischer, Steven R. *A History of Writing.* London: Reaktion Books, 2001.

Forster, Peter and Colin Renfrew (eds.). *Phylogenetic Methods and the Prehistory of Languages.* Cambridge: McDonald Institute for Archaeological Research, 2006.

Gabain, Annemarie von. *Alttürkische Grammatik mit Bibliographie, Lesestücken und Wörterverzeichnis auch Neutürkisch. Mit vier Schrifttafeln und sieben Schriftproben.* Leipzig: Harrassowitz, 1941.

Garbini, Giovanni. *Storia e problemi dell'epigrafia semitica.* Naples: Istituto Orientale, 1979.

Gibson, John C. L. *Textbook of Syrian Semitic Inscriptions.* Vol. 2. Oxford: Oxford University Press, 1975.

Goodall, David W. "Objective methods for the classification of vegetation III. An essay in the use of factor analysis." *Australian Journal of Botany* 2 (1954). 304–324.

Gower, John C. and Pierre Legendre. "Metric and Euclidean Properties of Dissimilarity Coefficients." *Journal of Classification* 3 (1986). 5–48.

Grimme, Hubert. *Althebräische Inschriften vom Sinai: Alphabet, Textliches, Sprachliches mit Folgerungen.* Hannover: Heinz Lafaire, 1923.

Hampel, József. "A nagyszentmiklósi kincs. Tanulmány a népvándorláskori művészetről." [The treasure of Nagyszentmiklós. A study of the art of the Great Migrations.] *Archeológiai Értesítő* 4 (1884). 1–166.

Harmatta, János. "Conclusion." In Harmatta, János (ed.). *History of Civilizations of Central Asia: The development of sedentary and nomadic civilizations: 700 B.C. to A.D. 250.* Volume II. Paris: UNESCO Publishing, 1994. 476–483.

Hawkins, John David. "Writing in Anatolia: Imported and Indigenous Systems." *World Archae-*

ology 17 (1986). 363–376.

Hawkins, John David. *Corpus of Hieroglyphic Luwian Inscriptions. Vol. I: Inscriptions of the Iron Age.* Berlin; New York (NY): de Gruyter, 2000.

Hawkins, John David. "A Unique Hieroglyphic Luwian Document?" In Cohen, Yoran, Amir Gilan, and Jared L. Miller (eds.). *Pax Hethitica – Studies on the Hittites and their Neighbours in Honour of Itamar Singer.* Wiesbaden: Harrassowitz, 2010. 183–190.

Healey, John F. *The Early Alphabet.* Berkeley, Los Angeles (CA): University of California Press, 1990.

Hempl, George. "The Origin of the Latin Letters G and Z." *Transactions and Proceedings of the American Philological Association* 30 (1899). 24–41.

Hennig, Willi. *Phylogenetic systematics.* Urbana (IL): University of Illinois Press, 1966.

Hesperia: *Banco de datos de lenguas paleohispánicas.* <http://hesperia.ucm.es>.

Hosszú, Gábor and Erzsébet Zelliger. "Rovásfelirat a megfejtés útján. Módszertani kérdések." [On the deciphering of Rovash inscriptions. Methodological questions.] *E-nyelv Magazin* 2 (2013). <http://e-nyelvmagazin.hu/rovasfelirat-a-megfejtes-utjan-modszertani-kerdesek>.

Hosszú, Gábor and Erzsébet Zelliger. [2014a.] "Többnyelvű feliratok a Nagyszentmiklósi aranykincsen." [Multilingual inscriptions on the golden treasure of Nagyszentmiklós.] *Magyar Nyelv* 110 (2014). 177–195. <http://www.c3.hu/~magyarnyelv/14-2/HosszuZelliger14-2.pdf>.

Hosszú, Gábor and Erzsébet Zelliger. [2014b.] "A Bodrog-alsóbűi rovásemlék számítógépes írástörténeti kapcsolatai és egy olvasati kísérlete." [Computational palaeographical relations of the Bodrog-Alsóbű Rovash relic and a reading attempt.] *Magyar Nyelv* 110 (2014). 417–431. <http://www.c3.hu/~magyarnyelv/14-4/HG_ZE_2014-4.pdf>.

Hosszú, Gábor. *Heritage of Scribes. The Relation of Rovas Scripts to Eurasian Writing Systems.* Budapest: Rovas Foundation, 2012.

Hosszú, Gábor. *Rovásatlasz.* [Rovash Atlas.] Budapest: Milani, 2013.

Hosszú, Gábor. "Mathematical Statistical Examinations on Script Relics." In Bhatnagar, Vishal (ed.). *Data Mining and Analysis in the Engineering Field.* Hershey (PA): Information Science Reference, 2014. 142–158.

Hosszú, Gábor. "A Novel Computerized Paleographical Method for Determining the Evolution of Graphemes." In Khosrow-Pour, Mehdi (ed.). *Encyclopedia of Information Science and Technology.* Hershey (PA): Information Science Reference, 2015. 2017–2031.

Jain, Anil K. and Richard C. Dubes. *Algorithms for Clustering Data.* Upper Saddle River (NJ): Prentice-Hall, 1988.

Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. "Data Clustering: A Review." *ACM Computing Surveys* 31 (1999). 264-323.

Jeffery, Lilian H. *The local scripts of archaic Greece.* Oxford: Clarendon Press, 1961.

Jensen, Hans. *Sign, Symbol, and Script: An Account of Man's Efforts to Write.* Transl. George Unwin. New York (NY): Putman, 1969.

Kairzhanov, Abai K. [Каиржанов, Абай Каиржанович.] Генезис тюркского рунического письма (на материале тамги-знаков евразийских кочевников). *Мова* 22 (2014). 12–24.

Kalinka, Ernst. *Tituli Asiae Minoris.* Vol. 1: *Tituli Lyciae lingua Lycia conscripti.* Wien: Hölder,

1901.
Kara, György. “Aramaic Scripts for Altaic Languages.” In Daniels, Peter T., and William Bright (eds.). *The World’s Writing Systems.* New York (NY), Oxford: Oxford University Press, 1996. 536–558.
Karali, Maria. “Writing systems.” In Christidis, Anastassios-Fivos (ed.). *A History of Ancient Greek. From the Beginning to Late Antiquity.* Cambridge (NY): Cambridge University Press, 2007. 197–207.
Karnava, Artemis. *The Cretan Hieroglyphic script of the second millennium BC: description, analysis, function and decipherment perspectives.* Unpublished doctoral dissertation. Bruxelles: Université libre de Bruxelles, 1999.
Kaufman, Leonard and Peter J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis.* New York (NY): John Wiley & Sons, 1990.
Kenyon, Frederic G. *The Palaeography of Greek Papyri.* Oxford: Clarendon Press, 1899.
King, Geraldine. *A Report on the Work of the Dhofar Epigraphic Project.* <http://krc2.orient.ox.ac.uk/aalc/images/documents/mcam/dep_report_1992.pdf>.
Kitching, Ian J. et al. *Cladistics: The Theory and Practice of Parsimony Analysis.* Oxford: Oxford University Press, 1998.
Kloekhorst, Alwin. “Studies in Lycian and Carian phonology and morphology.” *Kadmos* 47 (2008). 117–146.
Konkobaev, Kadyraly, Nurdin Useev, and Negizbek Shabdanaliev [Кадыралы Конкобаев, Нурдин Усеев, Негизбек Шабданалиев]. *Атлас древнетюркских письменных памятников Республики Алтай.* Astana: Ғылым, 2015.
Kononov, Andrej Nikolaevič. *Grammatika jazyka tjurkskix runičeskix pamjatnikov.* Leningrad: Nauka, 1980.
Krings, Véronique (ed.). *La civilisation phénicienne et punique: manuel de recherche.* Leiden, New York (NY), Köln: E.J. Brill, 1995.
Kyzlasov, Igor L. (Кызласов, Игорь Леонидович). *Рунические письменности евразийских степей*, Москва: Издательская фирма «Восточная литература» РАН, 1994.
LBI: *Libyco-Berber Inscriptions Online Database. © LBI-Projekt / Institutum Canarium* <http://www.institutum-canarium.org>.
Lemaire, André and Benjamin Sass. “The Mortuary Stele with Sam’alian Inscription from Ördekburnu near Zincirli.” *BASOR* 369 (2013). 57–136.
Lidzbarski, Mark. “Der Ursprung der nord-und südsemitischen Schrift.” In Lidzbarski, Mark (ed.). *Ephemeris für semitische Epigraphik. Vol. I. 1900–1902.* Giessen: Ricker, 1902. 109–136.
Ligeti, Lajos. “A magyar rovásírás egy ismeretlen betűje.” [An unknown letter of the Hungarian Rovash script.] *Magyar Nyelv* XXI (1925). 50–52.
Lindberg, David R. *Principals of Phylogenetic Systematics: Phenetics. Integrative Biology 200A Principles of Phylogenetics: Systematics.* University of Berkeley, 2012. <http://ib.berkeley.edu/courses/ib200a/lect/ib200a_lect09_Lindberg_phenetics.pdf>.
Looijenga, Tineke. *Runes Around the North Sea and On the Continent AD 150–700; Texts & Contexts.* Groningen: SSG Uitgeverij, 1997.
Macdonald, Michael C. A. “Reflections on the linguistic map of pre-Islamic Arabia.” *Arabian*

archaeology and epigraphy 11 (2000). 28–79.

Macdonald, Michael C. A. "Ancient North Arabian." In Woodard, Roger D. (ed.). *The Cambridge Encyclopedia of the World's Ancient Languages.* Cambridge: Cambridge University Press, 2004. 488–533.

Macdonald, Michael C. A. "Literacy in an Oral Environment." In Bienkowski, Piotr, Christopher Mee, and Elizabeth Slater (eds.). *Writing and Ancient Near Eastern Society. Papers in Honour of Alan R. Millard.* New York (NY), London: T&T Clark International, 2005. 45–114.

Macdonald, Michael C. A. *Literacy and Identity in Pre-Islamic Arabia.* Farnham, Burlington: Ashgate Publishing, 2009.

Macdonald, Michael C. A. "Ancient Arabia and the written word." In Macdonald, Michael C.A. (ed.). *The development of Arabic as a written language.* (Supplement to the Proceedings of the Seminar for Arabian Studies 40). Oxford: Archaeopress, 2010. 5–28.

Macdonald, Michael C. A. "On the uses of writing in ancient Arabia and the role of palaeography in studying them." *Arabian Epigraphic Notes* 1 (2015). 1–50.

MacKenzie, David Neil. *A Concise Pahlavi Dictionary.* London: Oxford University Press, 1971.

Marchesini, Simona. *Le lingue frammentarie dell'Italia antica: manuale per lo studio delle preromane.* Milano: Hoepli, 2009.

Marchesini, Simona. "The Elymian language." In Tribulato, Olga (ed.). *Language and Linguistic Contact in Ancient Sicily.* Cambridge: Cambridge University Press, 2012. 95–114.

Marchesini, Simona. "Über die Rätische Inschrift aus Pfatten/Vadena im Tiroler Landesmuseum Ferdinandeum, Innsbruck." In Meighörner, Wolfgang (ed.). *Wissenschaftliches Jahrbuch der Tiroler Landesmuseen 2014.* Innsbruck, Wien, Bozen: StudienVerlag, 2014. 202–217.

Marsadolov, Leonid. [2000a.] "The Cimmerian Traditions of the Gordion Tumuli (Phrygia): Found in the Altai Barrows (Bashadar, Pazyryk)." In Davis-Kimball, Jeannine et al. (eds.). *Kurgans, Ritual Sites, and Settlements: Eurasian Bronze and Iron Age.* Oxford: Archaeopress, 2000. 247–258.

Marsadolov, Leonid. [2000b.] "The Nomads of Kazakhstan, the Altai, and Tuva." In Aruz, Joan et al. (eds.). *The Golden Deer of Eurasia. Scythian and Sarmatian Treasures from the Russian Steppes.* New York (NY): Yale University Press, 2000. 49–53.

Masson, Olivier. "Un lion de bronze de provenance égyptienne avec inscription carienne." *Kadmos* 15 (1976). 82–83.

Masson, Olivier. *Carian Inscriptions from North Saqqâra and Buhen.* London: Egypt Exploration Society, 1978.

MATLAB: *MATLAB Release R2015a.* Natick (MA): The MathWorks, Inc., 2015.

Maue, Dieter. "Three Languages on One Leaf: On IOL Toch 81 with Special Regard to the Turkic Part." *Bulletin of the School of Oriental and African Studies, University of London* 71 (2008). 59–73.

McCarter, P. Kyle, Jr. *The antiquity of the Greek alphabet and the early Phoenician scripts.* Missoula (MT): Scholars Press, 1975.

Mees, Bernard. "Runes in the First Century." In Stoklund, Marie et al. (eds.). *Runes and their Secrets. Studies in runology.* Copenhagen: Museum Tusculanum Press, 2006. 201–232.

Melchert, H. Craig. "Lydian." In Woodard, Roger D. (ed.). *The Cambridge Encyclopedia of the World's Ancient Languages.* Cambridge: Cambridge University Press, 2004. 601–607.

Melchert, H. Craig. [2008a.] "Lycian." In Woodard, Roger D. (ed.). *The Ancient Languages of Asia Minor.* Cambridge: Cambridge University Press, 2008. 46–55.

Melchert, H. Craig. [2008b.] "Lydian." In Woodard, Roger D. (ed.). *The Ancient Languages of Asia Minor.* Cambridge: Cambridge University Press, 2008. 56–63.

Melchert, H. Craig. [2008c.] "Carian." In Woodard, Roger D. (ed.). *The Ancient Languages of Asia Minor.* Cambridge: Cambridge University Press, 2008. 64–68.

Miller, D. Gary. *Ancient Scripts and Phonologial Knowledge.* Amsterdam, Philadelphia (PA): John Benjamins Publishing, 1994.

MNAMON: *Antiche Scritture del Mediterraneo. Guida critica alle risorse elettroniche.* <http://lila.sns.it/mnamon>.

Morandi, Alessandro. *Epigrafia italica.* Roma: L'Erma di Bretschneider, 1982.

Morandi, Alessandro. "Epigrafia e lingua dei Celti d'Italia." In Agostinetti, Paola P., and Alessandro Morandi (eds.). *Celti d'Italia.* Vol. II. Roma: Spazio Tre, 2004.

Nagy, Géza. "A székely írás eredete." [The origin of the Székely script.] *Ethnographia* VI (1895). 269–276.

Nakhleh, Luay et al. "A comparison of phylogenetic reconstruction methods on an IE dataset." *Transactions of the Philological Society* 3 (2005). 171–192.

Németh, Gyula. "A régi magyar írás eredete." [The origin of the Ancient Hungarian script.] *Nyelvtudományi Közlemények* XLV (1917–1920). 31–44.

Németh, Gyula. *A magyar rovásírás.* [The Hungarian Rovash script.] Budapest: Hungarian Academy of Sciences, 1934.

NLR: *Российская национальная библиотека.* [National Library of Russia.] <http://expositions.nlr.ru/slav_culture>.

Nollé, Johannes. *Side im Altertum. Geschichte und Zeugnisse.* Band II. Bonn: Dr. Rudolf Habelt, 2001.

O'Connor, Michael P. "Epigraphic Semitic Scripts." In Daniels, Peter T. and William Bright (eds.). *The World's Writing Systems.* New York (NY), Oxford: Oxford University Press, 1996. 88–107.

Olivier, Jean-Pierre and Frieda Vandenabeele. *Édition holistique des textes chypro-minoens (HoChyMin).* Pisa; Rome: Fabrizio Serra Editore, 2007.

Olivier, Jean-Pierre. "Les syllabaires chypriotes des deuxième et premier millénaires avant notre ère: État des questions." In Sacconi, Anna et al. (eds.). *Colloguium Romanum. Atti del XII colloquio internazionale di Micenologia, Roma 20–25 febbraio 2006.* Pisa; Rome: Fabrizio Serra, 2008. 605–619.

Pardede, Raymond E. I. et al. "Four-Layer Grapheme Model for Computational Paleography." *Journal of Information Technology Research* 9 (2016). 64–82.

Payne, Annick. *Hieroglyphic Luwian. An Introduction with Original Texts.* Wiesbaden: Harrasowitz, 2010.

Podani, János. *Introduction to the Exploration of Multivariate Biological Data.* Leiden: Backhuys Publishers, 2000.

Praetorius, Franz. "Bemerkungen zum südsemitischen Alphabet." *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 58 (1904). 715–726.

PROEL: *Promotora Española de Lingüística.* <http://www.proel.org>.

Rehbein, Malte, Patrick Sahle, and Torsten Schaßan (eds.) *Kodikologie und Paläographie im digitalen Zeitalter 1 – Codicology and Palaeography in the Digital Age 1 (KPDZ 1).* Norderstedt: Books on Demand, 2009. URN: urn:nbn:de:hbz:38-29393.

Rilly, Claude and Alex de Voogt. *The Meroitic Language and Writing System.* Cambridge: Cambridge University Press, 2012.

Rodríguez Ramos, Jesús. "El origen de la escritura sudlusitano-tartesia y la formación de alfabetos a partir de alefatos." *Rivista di Studi Fenici.* 30 (2002). 187–216.

Rodríguez Ramos, Jesús. *Análisis de epigrafía íbera.* Vitoria-Gasteiz: Universidad del País Vasco, 2004.

Rogers, Henry. "Sociolinguistic factors in borrowed writing systems." *Toronto Working Papers in Linguistics* 17 (1999). 247–262.

Rollston, Christopher A. "The Phoenician Script of the Tel Zayit Abecedary and Putative Evidence for Israelite Literacy." In Tappy, Ron E. and P. Kyle McCarter Jr. (eds.). *Literate Culture and Tenth-Century Canaan. The Tel Zayit Abecedary in Context.* Winona Lake (IN): Eisenbrauns, 2008. 61–96.

Romesburg, H. Charles. *Cluster Analysis for Researchers.* Raleigh: Lulu.com, 2004.

Róna-Tas, András. "On the development and origin of the East Turkic «runic» script." *Acta Orientalia Academiae Scientiarum Hungaricae* XLI (1987). 7–14.

Rosenthal, Franz, Jonas C. Greenfield, and Shaul Shaked. "Aramaic." *Encyclopædia Iranica* II (3) (1986, updated in 2011). 250–261. <http://www.iranicaonline.org/articles/aramaic-#pt2>.

Röllig, Wolfgang. "L'Alphabet." In Krings, Véronique (ed.). *La civilisation phénicienne et punique: Manuel de recherche.* Leiden, New York (NY), Köln: E.J. Brill, 1995. 193–214.

Saitou, Naruya, and Masatoshi Nei. "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular Biology and Evolution* 4 (1987). 406–425.

Sampaio, Alberto. "Software Phenetics, Phylogeny and Evolution." *Proceedings of the 3rd International IEEE Workshop on Software Evolvability* 2007. 60–66.

Sándor, Klára. *A Bolognai Rovásemlék.* [The Bologna Rovash relic.] Szeged: József Attila Tudományegyetem, Magyar Őstörténeti Kutatócsoport, 1991.

Sass, Benjamin. *The Genesis of the Alphabet and Its Development in the Second Millennium B.C.* Wiesbaden: Harrassowitz, 1988.

Sebestyén, Gyula. *Rovás és rovásírás* [Rovash and Rovash writing.] Budapest: Magyar Néprajzi Társaság, 1909.

Sebestyén, Gyula. *A magyar rovásírás hiteles emlékei.* [The credible relics of the Hungarian Rovash script.] Budapest: Hungarian Academy of Sciences, 1915.

Segert, Stanislav. "Vowel Letters in Early Aramaic." *Journal of Near Eastern Studies* 37 (1978). 111–114.

Simon, Zsolt. "Critica. Ignacio J. Adiego: The Carian Language. With an Appendix by Koray Konuk [Handbuch der Orientalistik 86]. Leiden; Boston (MA): Brill, 2007. xiv + 518 pages with 2 maps & 4 plates." *Acta Antiqua Academiae Scientiarum Hungaricae* 48 (2008). 457–463.

Sims-Williams, Nicholas. "The Sogdian Sound-System and the Origins of the Uyghur Script." *Journal Asiatique* (1981). 347–360.

Sims-Williams, Nicholas. "Sogdian." In Schmitt, Rüdiger (ed.). *Compendium Linguarum Irani-*

carum. Wiesbaden: Dr. Ludwig Reichert Verlag, 1989. 173-192.
Sims-Williams, Nicholas. "The Sogdian manuscripts in Brāhmī script as evidence for Sogdian phonology." In Emmerick, Ronald E. et al. (eds.). *Turfan, Khotan und Dunhuang. Vorträge der Tagung Annemarie von Gabain und die Turfanforschung, Berlin, 9.–12. 12. 1994.* Berlin: Akademie Verlag, 1996. 307–315.
Sims-Williams, Nicholas and Franz Grenet. "The Sogdian Inscriptions of Kultobe." *Shygys* 2006/1 (2007). 95–111.
Skelton, Christina. "Phylogenetic Systematics to Reconstruct the History of the Linear B Script." *Archaeometry* 50 (2007). 158–176.
Skjærvø, Prods Oktor. "Aramaic Scripts for Iranian Languages." In Daniels, Peter T. and William Bright (eds.). *The World's Writing Systems.* New York (NY), Oxford: Oxford University Press, 1996. 515–535.
Sneath, Peter H. A. and Robert Reuven Sokal. *Numerical Taxonomy. The Principles and Practice of Numerical Classification.* San Francisco (CA): Freeman, 1973.
Sneath, Peter H. A. "Cladistic Representation of Reticulate Evolution." *Systematic Zoology* 24 (1975). 360–368.
Sokal, Robert Reuven and Charles D. Michener. "A statistical method for evaluating systematic relationships." *The University of Kansas Science Bulletin* XXXVIII (1958). 1409–1438.
Sokal Robert Reuven and Peter H. A. Sneath. *Principles of Numerical Taxonomy.* San Francisco (CA), London: Freeman, 1963.
Sokal, Robert Reuven et al. "Numerical Taxonomy: Some Points of View." *Systematic Zoology* 14 (1965). 237–243.
Sprengling, Martin. *The Alphabet. Its Rise and Development from the Sinai Inscriptions.* Chicago (IL): University of Chicago Press, 1931.
Stokes, Peter Anthony. "Computer-Aided Palaeography, Present and Future." In Rehbein, Malte, Patrick Sahle, and Torsten Schaßen (eds.). *Kodikologie und Paläographie im digitalen Zeitalter 1 – Codicology and Palaeography in the Digital Age 1 (KPDZ 1).* Norderstedt: Books on Demand, 2009. 309–338.
Stokes, Peter Anthony. "Digital Approaches to Paleography and Book History: Some Challenges, Present and Future." *Frontiers in Digital Humanities* 2 (2015). 1–3.
Sukkarieh, Jana Z., Matthias von Davier, and Kentaro Yamamoto. *From Biology to Education: Scoring and Clustering Multilingual Text Sequences and Other Sequential Tasks.* Princeton (NJ): Educational Testing Service, ETS Research Report No. RR-12–25, Dec. 2012.
Swiggers, Pierre. "Transmission of the Phoenician Script to the West." In Daniels, Peter T. and William Bright (eds.). *The World's Writing Systems.* New York (NY), Oxford: Oxford University Press, 1996. 261–270.
Swiggers, Pierre and Wolfgang Jenniges. "The Anatolian Alphabets." In Daniels, Peter T. and William Bright (eds.). *The World's Writing Systems.* New York (NY); Oxford: Oxford University Press, 1996. 281–287.
Szabó, Géza. "Keleti mítoszelemek nyugati megjelenése a regölyi astralagos leletek tükrében." [Occurance of eastern mythological elements in the West considering the astralag finds of Regöly.] In Csabai, Zoltán et al. (eds.). *Európé égisze alatt. Ünnepi tanulmányok Fekete Mária hatvanötödik születésnapjára kollégáitól, barátaitól és tanítványaitól.* [Under the

perspective of Európé. Celebrating studies for the 65th birthday of Mária Fekete from her collegues, friends and disciples.] Budapest: L'Harmattan, 2015. 321–364.

Taylor, Isaac. *The Alphabet. An Account of the Origin and Development of Letters. Vol. II. Aryan Alphabets.* London: Kegan Paul, Trench, & Co., 1883.

Tekin, Talat. *Orhon Türkçesi Grameri.* Istanbul: Mehmet Ölmez, 2003.

Thelegdi, Ioannis. *Rudimenta, Priscae hunnorum linguae brevibus quaestionibus ac responcionibus comprehensa opera et studio. (1598).* Budapest: Ars Libri, 1994.

Thompson, Edward M. *An Introduction to Greek and Latin Palaeography.* Oxford: Clarendon Press, 1912.

Thomsen, Vilhelm. *Inscriptions de l'Orkhon déchiffrées.* Helsingfors: La société de literature Finnoise, 1893.

Tóth, Loránd L., Raymond Pardede, and Gábor Hosszú. "Novel Algorithmic Approach to Deciphering Rovash Inscriptions." In Khosrow-Pour, Mehdi (ed.). *Encyclopedia of Information Science and Technology.* Hershey (PA): Information Science Reference, 2015. 7222–7233.

Tóth, Loránd L. et al. "Application of the Cluster Analysis in Computational Paleography." In Samui, Pijush (ed.). *Handbook of Research on Advanced Computational Techniques for Simulation-Based Engineering.* Hershey (PA): Engineering Science Reference, 2016. 525–543.

Tzanavari, Katerina and Anastasios-Phoebus Christidis. "A Carian Graffito from the Lebet Table, Thessaloniki." *Kadmos* 34 (1995). 13–17.

Urbanová, Daniela. "Oština, umberština a jihopikénština – sabellské jazyky." *Sborník prací Filozofické fakulty brněnské univerzity. N, Řada klasická* 52 (2003). 5–37.

Valério, Miguel. "Origin and development of the Paleohispanic scripts: the orthography and phonology of the Southwestern alphabet." *Revista Portuguesa de Arqueologia* 11 (2008). 107–138.

Valério, Miguel. "Cypro-Minoan Tablet RASH Atab 004 as Akkadian Text and its Role in the Decipherment of the Script." In Martín, Sergio C. et al. (eds.). *Mediterráneos: An Interdisciplinary Approach to the Cultures of the Mediterranean Sea.* Newcastle upon Tyne: Cambridge Scholars Publishing, 2013. 5–28.

Valério, Miguel. "«Cypro-Greek» Syllabary: An Unambiguous Solution to a Terminological Problem." *AION Linguistica* 3 (2014). 259–270.

Valério, Miguel. *Investigating the Signs and Sounds of Cypro-Minoan.* Doctoral dissertation, Barcelona: University of Barcelona, 2016.

Vasil'ev, Dmitrij D. "Versuch zur Lösung der Kerbinschrift aus der Umgebund von Kalocsa im Spiegel der eurasischen Parallelen." *Folia Archaeologica* XLIII (1994). 181–191.

Vasil'ev, Dmitrij D. "The Eurasian Areal Aspect of Old Turkic Written Culture." *Acta Orientalia Academiae Scientiarum Hungaricae* 58 (2005). 323–330.

Vékony, Gábor. "Das nordwestliche Transdanubien im 9. Jahrhundert und die 'Uungariorum marcha'." *Savaria* 15 (1981). 215–229.

Vékony, Gábor. [1985a.] "Késő népvándorláskori rovásfeliratok." [Rovash inscriptions from the Late Period of the Great Migrations.] *Életünk* XXII (1985). 71–84.

Vékony, Gábor. [1985b.] "A szarvasi felirat és ami körülötte van." [The Szarvas inscription and its circumstances.] *Életünk* XXII (1985). 1133–1145.

Vékony, Gábor. [1987a.] *Későnépvándorláskori rovásfeliratok a Kárpát-medencében.* [Rovash inscriptions in the Carpathian Basin from the Late Period of the Great Migrations.] Szombathely: Életünk Szerkesztősége – Magyar Írók Szövetsége Nyugat-Magyarországi Csoportja, 1987.

Vékony, Gábor. [1987b.] "Spätvölkerwanderungszeitliche Kerbinschriften im Karpatenbecken." *Acta Archaeologica Academiae Scientiarum Hungaricae* 39 (1987). 211–256.

Vékony, Gábor. [1992a.] "A halomi honfoglalás kori tegezfelirat." [Quiver inscription of Halom from the age of the Magyars' Land Acquisition.] *Életünk* XXX (1992). 542–546.

Vékony, Gábor. [1992b.] "Varázsszöveg a halomi honfoglalás kori temetőből." [Magic text from the cemetery of Halom from the age of the Magyars' Land Acquisition.] In Sándor, Klára (ed.). *Rovásírás a Kárpát-medencében.* [Rovash scripting in the Carpathian Basin.] Szeged: József Attila University, 1992. 41–50.

Vékony, Gábor [1992c.] "A Bolognai Rovásemlék." [The Rovash relic of Bologna.] *Magyar Könyvszemle* 108 (1992). 288–289.

Vékony, Gábor. "A székely rovásírás." [The Székely Rovash script.] *Napjaink* VII (1993). 39.

Vékony, Gábor. "The Peoples of the Period of the Great Migrations in the Carpathian Basin." *Specimina Nova* 12 (1996). 01–04.

Vékony, Gábor. [1999a.] "A székely írás legrégibb emléke Bodrog–Alsóbű vaskohászati műhelyéből." [The oldest relic of the Székely script from Bodrog-Alsóbű iron smelting yard.] In Gömöri, János (ed.). *Traditions and Innovations in the Early Medieval Iron Production*, Sopron – Somogyfajsz, 30th May 1997 – 1st June 1997. Dunaújváros, Veszprém, 1999. 226–229.

Vékony, Gábor. [1999b.] "10. századi székely felirat a Somogy megyei Bodrog határában." [Székely inscription from the 10th c. in the outskirts of Bodrog of Somogy County.] *História* 8 (1999). 30–31.

Vékony, Gábor. *A székely írás emlékei, kapcsolatai, története.* [Relics, relationships and history of the Székely script] Budapest: Nap Kiadó, 2004.

Wallace, Rex E. *The Sabellic Languages of Ancient Italy.* Muenchen: LINCOM, 2007.

Ward, Joe H. Jr. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (1963). 236–244.

Weeden, Mark. "Anatolian Hieroglyphs: Logogram vs. Ideogram." In Grodin, Shai (ed.). *Visualizing Knowledge and Creating Meaning in Ancient Writing Systems.* Berlin: PeWe-Verlag, 2014. 81–100.

Wheeler, Ward C., and Peter M. Whiteley. "Historical linguistics as a sequence optimization problem: the evolution and biogeography of Uto-Aztecan languages." *Cladistics* (2014). 1–13.

Wills, Matthew A. "Morphological disparity: A primer." In Adrain, Jonatahn M., Gregory D. Edgecombe, and Bruce S. Lieberman (eds.). *Fossils, Phylogeny, and Form: An Analytical Approach.* New York (NY): Plenum, 2001. 55–144.

Wolf, Lior et al. "Computerized Paleography: Tools for Historical Manuscripts." *18th IEEE International Conference on Image Processing (ICIP).* Brussels, 2011. 3545–3548.

Woodard, Roger D. *Greek Writing from Knossos to Homer.* New York (NY), Oxford: Oxford University Press, 1997.

Woodard, Roger D. *The Textualization of the Greek Alphabet.* New York (NY): Cambridge University Press, 2014.

Woudhuizen, Fred C. "Etruscan Origins: The Epigraphic Evidence." *Talanta* XIV–XV (1982–1983). 91–117.

Woudhuizen, Fred C. [1984–1985a.] "Lydian: Separated from Luwian by Three Signs." *Talanta* XVI–XVII (1984–1985). 91–113.

Woudhuizen, Fred C. [1984–1985b.] "Origins of the Sidetic script." *Talanta* XVI–XVII (1984–1985). 115–126.

Yakubovich, Ilya S. *Sociolinguistics of the Luvian Language.* PhD. Dissertation. Chicago (IL): University of Chicago Press, 2008.

Yakubovich, Ilya S. [2015a.] "The Luwian Language." *Oxford Handbooks Online* (21 Oct. 2015). DOI: 10.1093/oxfordhb/9780199935345.013.18.

Yakubovich, Ilya S. [2015b.] "Phoenician and Luwian in Early Iron Age Cilicia." *Anatolian Studies* 65 (2015). 35–53.

Young, Rodney S. "Old Phrygian Inscriptions from Gordion: Toward a History of the Phrygian Alphabet. (Plates 67–74)." *Hesperia: The Journal of the American School of Classical Studies at Athens* 38 (1969). 252–296.

Younger, John. "Linear A Texts in phonetic transcription & Commentary." (30 November 2000, last update: 26 October 2016). <http://people.ku.edu/~jyounger/LinearA/>.

Younger, John. "Cretan Hieroglyphic Grids: (very tentative suggestions)." (29 July 2003, last update: 21 April 2012). <http://people.ku.edu/~jyounger/Hiero/Hgrids.html>.

Zelliger, Erzsébet. "Nyelvemlék? Művelődéstörténeti emlék? Gondolatok az ÓMS és a vargyasi rovásírásos felirat kapcsán." [Language relic? Cultural historical relic? Thoughts on the Old Hungarian Lament of Mary and the Vargyas Rovash inscription.] In Juhász, Dezső (ed.). *Kerekasztal körül. Huszonöt nyelvészeti tanulmány.* Budapest: Eötvös Loránd University, 2016. 92–98.

Zelliger, Erzsébet, and Gábor Hosszú. "A Konstantinápolyi székely-magyar rovásfelirat számítógépes paleográfiai elemzése." [The computational palaeographical analysis of the Constantinople Székely-Hungarian Rovash inscription.] *Alkalmazott Nyelvtudomány* XIV (2014). 89–124. <https://www.academia.edu/11537828/>.