

Chapter 8

Mathematical Statistical Examinations on Script Relics

Gábor Hosszú

Budapest University of Technology and Economics, Hungary

ABSTRACT

This chapter presents statistical evaluations of script relics. Its concept is exploiting mathematical statistical methods to extract hidden correlations among different script relics. Examining the genealogy of the graphemes of scripts is necessary for exploring the evolution of the writing systems, reading undeciphered inscriptions, and deciphering undeciphered scripts. The chapter focuses on the cluster analysis as one of the most popular mathematical statistical methods. The chapter presents the application of the clustering in the classification of Rovash (pronounced “rove-ash,” an alternative spelling: Rovas) relics. The various Rovash scripts were used by nations in the Eurasian Steppe and in the Carpathian Basin. The specialty of the Rovash paleography is that the Rovash script family shows a vital evolution during the last centuries; therefore, it is ideal to test the models of the evolution of the glyphs. The most important Rovash script is the Szekely-Hungarian Rovash. Cluster analysis algorithms are applied for determining the common sets among the significant Szekely-Hungarian Rovash alphabets. The determined Rovash relic ties prove the usefulness of the clustering methods in the Rovash paleography.

INTRODUCTION

The computational paleography belongs to the applied computer science and deals with the use of mathematical methods in exploring the meaning and the ties of various old inscriptions. The chapter focuses on the cluster analysis as one of the most popular mathematical statistical method. The chapter presents the fundamental concepts of the computational paleography, several earlier approaches to applying the mathematical tools for

exploring the ties of the old orthographies, and the basics of the clustering methods and applied metrics. The clustering is applied to an ancient script, the Szekely-Hungarian Rovash (pronounced “rove-ash,” an alternative spelling: Rovas). The chapter presents the application of the clustering in the classification of Rovash relics. Cluster analysis algorithms are used for determining the ties among the significant Szekely-Hungarian Rovash alphabets.

DOI: 10.4018/978-1-4666-6086-1.ch008

BACKGROUND

Mathematical statistical tools as the cluster analysis are useful in various scientific fields, including the computational paleography. The chapter deals with computational paleography that means modeling and analyzing methods to explore links among graphemes, deciphering ancient inscriptions, representing various glyphs in digital form by using mathematical algorithms and software tools. The applied mathematical statistical tools may accelerate the research time and provide more accurate results through the automatic process for exploring the genealogical ties among the various branches of a script.

The *writing system* is a symbolic representation of a language described in terms of linguistic units (Malatesha & Aaron, 2006). The *script* is a writing system, some examples of the scripts are the following: Aramaic, Arabic, Armenian, Brahmi, Carpathian Basin Rovash, Chinese, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Hebrew, Kannada, Kharoshthi, Latin, Runic, Szekely-Hungarian Rovash, Telugu, Umbrian, Venetic, etc.

The *grapheme* is a minimally distinctive unit in a writing system. The *symbol* is a distinct unit of an inscription, the inscriptions are composed of a series of symbols. *Grapheme* is the abstraction of a symbol. Graphemes can be letters, ligatures, numerical digits, or punctuation marks. The grapheme has the following properties: (i) the script belonging into, (ii) its glyphs, (iii) its sound values, (iv) periods of use. The *glyph* is the shape of the grapheme with topological information. One grapheme has usually more glyphs. The symbol in an inscription is an implementation of a glyph.

The *orthography* is the visual representation of a language, which uses graphemes belonging to a certain script, and it is determined by the features of a language. A script is used for different orthographies. For instance, the Latin script is used for several orthographies, including the French, German, English, Hungarian,

etc. orthographies. The orthography of a spoken language changes periodically (Rogers, 1999). The changes can occur as changing set of graphemes, which encompasses the shape transformation of a glyph. A cause of writing system alteration is the establishment of more advanced writing media or instruments. The advanced writing technology introduces new writing technique, which impacts the glyphs of a grapheme.

The *script family* is a group of scripts, which are closely related to each other. The Rovash (pronounced “rove-ash”) script family includes the Proto-Rovash, the Early Steppean Rovash, the Carpathian Basin Rovash, the Steppean Rovash, and the Szekely-Hungarian Rovash scripts (Hosszú, 2013a, 2013b). The history and the genealogical ties of the Rovash scripts have been the subject of heavy research efforts from the past until today (Hosszú, 2012). It is a difficult task to determine the accurate genealogy and timelines of the Rovash scripts (Hosszú, 2012). The various Rovash scripts gradually differentiated after the geographical isolation of their users. Carpathian Basin Rovash was in use in the Carpathian Basin by Hungarians mainly, the others were used by nations and tribes of the Eurasian Steppe up to the 10th/13th centuries. Most of the Rovash scripts became extinct in the Medieval Times; however, the Szekely-Hungarian Rovash continued to be used since the 9th century throughout history until the present by the Szekelys—an organized border guard subgroup of the Hungarians.

The studies related to graphemes are challenging topics for paleographers and archaeologists, including deciphering undeciphered glyph discovered through excavation, reading patterns in glyphs transformation, etc. The mathematical statistical methods can be used in the computational paleography in different fields. Due to the significant need to decipher the inscriptions of the various archaeological finds, serious research efforts are carried out worldwide. A part of them deals with modeling the graphemes (Pardede et al., 2012). Quantitative aspects can be measured by

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/mathematical-statistical-examinations-on-script-relics/109980

Related Content

Ant Programming Algorithms for Classification

Juan Luis Olmo, José Raúl Romero and Sebastián Ventura (2014). *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining* (pp. 107-128).

www.irma-international.org/chapter/ant-programming-algorithms-for-classification/110456

Big Data – Small World: Materializing Digital Information for Discourse and Cognition

Ian Gwilt (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 288-301).

www.irma-international.org/chapter/big-data--small-world/150170

An Efficient Pruning and Filtering Strategy to Mine Partial Periodic Patterns from a Sequence of Event Sets

Kung-Jiuan Yang, Tzung-Pei Hong, Yuh-Min Chen and Guo-Cheng Lan (2014). *International Journal of Data Warehousing and Mining* (pp. 18-38).

www.irma-international.org/article/an-efficient-pruning-and-filtering-strategy-to-mine-partial-periodic-patterns-from-a-sequence-of-event-sets/110384

XML Tree Classification on Evolving Data Streams

Albert Bifet and Ricard Gavaldà (2012). *XML Data Mining: Models, Methods, and Applications* (pp. 199-218).

www.irma-international.org/chapter/xml-tree-classification-evolving-data/60910

An Approach for Retrieving Faster Query Results From Data Warehouse Using Synonymous Materialized Queries

Sonali Ashish Chakraborty and Jyotika Doshi (2021). *International Journal of Data Warehousing and Mining* (pp. 85-105).

www.irma-international.org/article/an-approach-for-retrieving-faster-query-results-from-data-warehouse-using-synonymous-materialized-queries/276766