



ASZPK 2012

I. Alkalmazott Számítógépes Paleográfiai Konferencia

**Budapesti Műszaki és Gazdaságtudományi Egyetem
Budapest, 2012. december 1.**



ASZPK 2012

Konferenciakiadvány

ASZPK 2012

**I. Alkalmazott Számítógépes
Paleográfiai Konferencia**

**Budapesti Műszaki és Gazdaságtudományi Egyetem
Budapest, 2012. december 1.**

Szerkesztő: Dr. habil. Hosszú Gábor, a műszaki tudomány kandidátusa, okl. villamosmérnök, okl. jogász, egyetemi docens a Budapesti Műszaki és Gazdaságtudományi Egyetem Elektronikus Eszközök Tanszékén

Lektorálta: Demeczky Jenő, okl. villamosmérnök, általános és alkalmazott nyelvészet szakos okl. nyelvész, IBM nemzetközi fordítási terminológus, IBM magyar terminológus, IBM Translation Services Center közép- és kelet-európai terminológus, International Business Machines Corporation Magyarországi Kft.

Dr. Zelliger Erzsébet, a nyelvészeti tudományok kandidátusa, dialektológus, nyelvtörténész, nyugalmazott egyetemi docens az Eötvös Loránd Tudományegyetem Magyar Nyelvtörténeti, Szociolingvisztikai és Dialektológiai Tanszékén

Kiadó: Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki És Informatikai Kar

Címlap Rumi Tamás

Copyright © Dr. Hosszú Gábor és Budapesti Műszaki és Gazdaságtudományi Egyetem,
Villamosmérnöki és Informatikai Kar. Minden jog fenntartva, 2012.

ISBN: ISBN 978-963-421-785-5
ISBN 978-963-421-786-2 (online)

Tartalom

Előszó	6
Vorwort	7
Foreword	8
Előadások	9
Írásrendszerek, hangrendszerek (<i>Zelliger Erzsébet</i>)	10
A grafémák leszarmazási elemzése (<i>Hosszú Gábor</i>)	18
Az andronovói bronzkori civilizáció és a rovásírás keletkezése (<i>Banai Miklós, Lukács Béla</i>)	37
A nagyszentmiklósi kincs korai keltezésének kérdései (<i>Obrusánszky Borbála</i>)	55
A román nyelv tanúságtétele: adalékok Közép-Európa népvándorlás kori történetéhez (<i>Hölbling Tamás</i>)	70
Gyökrovás (<i>Sípos László</i>)	76
A latin betűs írás hatása a székely-magyar rovásra (<i>Rumi Tamás</i>)	91
Szövegek rovásra alakításának lehetőségei (<i>Kliha Gergely</i>)	105
Grafémák kanonikus összetevőkre bontása (<i>Tóth Loránd Lehel, Raymond Pardede, Hosszú Gábor</i>)	112
Zusammenfassungen	127
Schriftsystemen, Lautsystemen (<i>Erzsébet Zelliger</i>)	127
Abstammungsanalyse der Grapheme (<i>Gábor Hosszú</i>)	127
Die Andronowo-Kultur (Bronzezeit) und die Gestaltwerdung der Rowaschchrift (<i>Miklós Banai, Béla Lukács</i>)	127
Fragen um die frühen Datierung des Goldschatzes von Nagyszentmiklós (<i>Borbála Obrusánszky</i>)	128
Zeugnis der rumänischen Sprache: Beiträge zur Geschichte Mitteleuropas in der Völkerwanderungszeit (<i>Tamás Hölbling</i>)	128
Wurzelrowasch in der Schrift (<i>László Sípos</i>)	128
Die Schrift mit lateinischen Buchstaben und ihr Einfluss auf die szekler-ungarischen Rowaschchrift (<i>Tamás Rumi</i>)	129
Verfahren der Transkription von Texten mit Rowaschchrift (<i>Gergely Kliha</i>)	129
Zerlegung der Graphemen zur kanonischen Komponenten (<i>Loránd Lehel Tóth, Raymond Pardede, Gábor Hosszú</i>)	129
Abstracts	131
Writing systems – Phonetic structures (<i>Erzsébet Zelliger</i>)	131
Genealogical analysis of graphemes (<i>Gábor Hosszú</i>)	131
The Andronovo Bronze Age civilization and the formation of the Rovash script (<i>Miklós Banai, Béla Lukács</i>)	132
Problems of the early date of the Golden Treasure of Nagyszentmiklós (<i>Borbála Obrusánszky</i>)	132
Testimony of the Romanian language: Additional aspects to the history of the Migration Period in Central Europe (<i>Tamás Hölbling</i>)	132
Root Rovas scripting (<i>László Sípos</i>)	133
Influence of the Latin script on the Székely-Hungarian Rovas (<i>Tamás Rumi</i>)	133
Methods of texts transcription to Rovash (<i>Gergely Kliha</i>)	133
Decomposition of graphemes to canonical components (<i>Loránd Lehel Tóth, Raymond Pardede, Gábor Hosszú</i>)	134
Életrajzok	135

Előszó

Az **I. Alkalmazott Számítógépes Paleográfiai Konferencia** 2012. december 1-jén a Budapesti Műszaki és Gazdaságtudományi Egyetem Villamosmérnöki és Informatikai Karán került megrendezésre.

A konferencia szervezőbizottsága: Dr. Zelliger Erzsébet, a nyelvészeti tudományok kandidátusa, dialektológus, nyelvtörténész, nyugalmazott egyetemi docens, Demeczky Jenő okl. villamosmérnök, okl. nyelvész, IBM nemzetközi és magyar fordítási terminológus, valamint Dr. Hosszú Gábor, a műszaki tudomány kandidátusa, okl. villamosmérnök, okl. jogász, egyetemi docens.

A konferencia meghirdetett témái a következők voltak: számítógépes paleográfia, magyar és eurázsiai nyelvészet, történelem, régészet, néprajz, ismeretlen írásemlékek algoritmizált megfejtése, jelenkori székely-magyar rovás-helyesírás és a rovás helye a digitális kommunikációban.

A konferencia alkalmat nyújtott a különböző tudományterületek képviselőinek, hogy megosszák a módszertanuk és megközelítésük szempontjából azokat az elképzeléseiket, elméletüket, kutatási eredményeiket, amelyek hasznosak lehetnek más tudományágak művelőinek is.

A kiadványban a szerzők által a konferenciát követően benyújtott tanulmányok szerepelnek. Az egyes tanulmányok tartalmáért a szerzők teljes felelősséget vállalnak.

A munka szakmai tartalma kapcsolódik a „**Új tehetséggondozó programok és kutatások a Műegyetem tudományos műhelyeiben**” c. projekt szakmai célkitűzéseinek megvalósításához. A projekt megvalósítását a **TÁMOP – 4.2.2.B-10/1--2010-0009** program támogatja.

Szervezőbizottság

Vorwort

Die **1. Konferenz für Angewandte Computer-Paläographie** fand am 1. Dezember 2012 an der Fakultät für Elektrotechnik und Informatik an der Technische und Wirtschaftswissenschaftliche Universität Budapest statt.

Das Organisationskomitee der Konferenz: Dr. Erzsébet Zelliger (Sprachwissenschaftlerin in ung. Dialektologie und Sprachgeschichte, Univ.-Dozentin), Jenő Demeczky (MSc in Elektrotechnik, MSc in Linguistics, IBM Internationaler und ungarischer Übersetzungsterminologe) und Dr. Gábor Hosszú (MSc in Electrical Ingenieurwesen, MSc in Law, Associate Professor).

Die Themen der Konferenz waren Computer-Paläographie, ungarische und eurasische Linguistik, Geschichte, Archäologie, Ethnographie, algorithmische Entzifferung von unbekanntem Schriftrelikte, die heutige szekler-ungarische Rowaschorthographie und die Lage der Rowasch in der digitalen Kommunikation.

Die Konferenz bot den Vertretern aus verschiedenen Disziplinen Gelegenheit, ihre Ideen, Theorien und Forschungsergebnisse hinsichtlich ihrer Methodik und ihrer Ansätze zu teilen, die für andere Disziplinen nützlich sein können.

Diese Publikation enthält die von den Autoren nach der Konferenz vorgelegten Studien. Die Autoren übernehmen die Verantwortung für den Inhalt jeder Studie.

Der berufliche Inhalt der Arbeit bezieht sich auf die „**Neuen Talentprogramme und Forschungen an den akademischen Werkstatt der Universität für Technologie**“. Die beruflichen Ziele des Projekts sind die Umsetzung des Projekts, die von **TÁMOP – 4.2.2.B-10/1--2010-0009** unterstützt wird.

Organisationskomitee

Foreword

The **1st Applied Computational Palaeography Conference** was held on December 1, 2012 at the Faculty of Electrical Engineering and Informatics at the Budapest University of Technology and Economics.

The Organizing Committee of the conference: Dr. Erzsébet Zelliger (linguist, Associate Professor), Jenő Demeczky (MSc in Electrical Engineering, MSc in Linguistics, IBM international and Hungarian translation terminologist), and Dr. Gábor Hosszú (MSc in Electrical Engineering, MSc in Law, Associate Professor).

The topics of the conference were computational palaeography, Hungarian and Eurasian linguistics, history, archaeology, ethnography, algorithmic deciphering of script relics without reading, present-day Székely-Hungarian Rovash orthography and location of the Rovash in digital communication.

The conference provided an opportunity for representatives from different disciplines to share their ideas, theories and research findings with regard to their methodology and approaches that may be useful to other disciplines.

This publication includes the studies submitted by the authors after the conference. The authors take full responsibility for the content of each study.

The professional content of the work is related to the “**New Talent Programs and Research at the Academic Workshops of the University of Technology.**” Project's professional goals. The implementation of the project is supported by **TÁMOP – 4.2.2.B-10/1--2010-0009**.

Organizing Committee

Előadások

Grafémák kanonikus összetevőkre bontása

TÓTH LORÁND LEHEL, RAYMOND PARDEDE, HOSSZÚ GÁBOR

Budapesti Műszaki és Gazdaságtudományi Egyetem

Elektronikus Eszközök Tanszéke

e-mail: tothl@eet.bme.hu

Kivonat: A cikk bemutat egy új eljárást egy írás különböző írásváltozatai rokonsági viszonyainak feltárására. Ennek során a vizsgált írásváltozatot reprezentáló írásemlék grafémaalakjait kétdimenziós részalakzatokra, úgynevezett kanonikus alakokra bontották fel. Kanonikus alak lehet például egy körszerű hurok, ferde szakasz, függőleges szakasz vagy kereszteződés. A vizsgált írásemlékben szereplő grafémaalakok kanonikus felbontásában szereplő együtthatókat összegezve és normálva egy ábécélenyomatnak nevezett adatsor jön létre, amely az ujjlennyomathoz hasonlóan a vizsgált írásváltozat grafémaalakjai együttesére jellemző. Az ábécélenyomatokat klaszteranalízis segítségével összehasonlító megállapításokat lehet tenni az írásemlékek által reprezentált írásváltozatok hasonlósági viszonyaira.

Kulcsszavak: grafémaanalízis, feliratok azonosítása, számítógépes írásmodellezés, számítógépes paleográfia, klaszteranalízis.

1. Bevezetés, irodalmi áttekintés

A különböző írásokkal készült történelmi írásemlékek szerzőjének vagy a létrejövetelük körülményeinek meghatározása gyakran nehézséget okoz a kutatók számára. Ennek oka az írást hordozó anyag (fa, kő, tégl, papír stb.) romlásától eltekintve elsősorban az, hogy az írásokban használatos betűk alakjai a történelem során időben változtak (Dian 2009).

Számos algoritmus ismert a szövegbányászat, illetve számítógépes paleográfia területén, amelyek az írásfelismerésre, a jelalak-felismerésre, és a nehezen olvasható szövegek jelentésazonosítására szolgálnak (Tikk 2007; Wolf és mtsai. 2011). A nemzetközi szakirodalomban található cikkek, folyóiratok, konferenciák sokszínűsége alátámasztja e téma hasznosságát, aktualitását, mint például az Izadi és társai által kidolgozott eljárás a régi perzsa szövegek azonosítására (Izadi és mtsai. 2008), a Jón által írt algoritmus, amely ősi egyiptomi hieroglifák megfejtésére irányul neurális hálózatok alkalmazásával (Kristjánsson 2007), valamint az ugyancsak neurális hálózat kifejlesztésével foglalkozó másik kutatócsoport által kifejlesztett, a kézzel írt arab szövegek gépi felismerésére szolgáló algoritmus (Khatatneh és mtsai. 2006). Egy hasonló kutatási témát közöltek Le Cun és mtsai. (1990) a kézírással írt szövegek azonosítására, míg Mutalib és mtsai. (2007) olyan módszert fejlesztettek ki, amely képes a kézzel írt „t” betű topológiai vizsgálatából a betűt leíró személy személyiségi elemzésére. Indiában, a kannada nyelvet beszélők között régóta használatos saját írás emlékei korának automatikus meghatározásával foglalkoznak Kashyap és mtsai. (2003).

Acharya és mtsai. (2009) neurális hálózatok alkalmazásával kísérik meg a különböző történelmi korokban használatos betűalakokra jellemző azonosítók felismerését egy tetszőleges szövegben. Khan értekezésében (2000) egy alakzati modellt állít fel, amelyet graféma- és szófelismerésre használ. Ennek lényege az, hogy azzal kísérli meg az optikai karakterfelismerés

(OCR) során szükséges tanulási szakasz idejét csökkenteni úgy, hogy a grafémák ideális alakjának a mindennapokban szokásos transzformációit veszi számításba.

Grafémának (más szóval karakternek) nevezzük az írás alapegységét. A graféma különböző típusú lehet: betű, írásjel, számjegy és ékezet. A graféma mint objektum attribútumai: *írás (script)*, amin belül használták, *grafémaalak (glyph)*, *grafémanév (betűnév)*, *átbetűzési érték (transzliteráció)*, *hangérték (transzkripció)*, *használati időszak*, *használati terület* stb. Grafémaalakból – amely egy graféma alakzati tulajdonságait tartalmazó kép – és hangértékből – amely egy adott nyelven egy adott korban a grafémához tartozó hangot jelöli – több is lehet. Grafémanév a graféma rendszerint hagyományos elnevezése, pl. a görög *alfa*.

Egy adott írás eltérő változataival készült írásemlékek stílusa, grafémaalakjai különböznek egymástól attól függően, hogy az egyes írásemlékek milyen korban és földrajzi területen jöttek létre. Az eltérések adódhatnak abból is, hogy az egyes írásemlékeket eltérő kézírással és írástudással rendelkező emberek hozták létre. Így az írásemlékek készítéséhez alkalmazott grafémaalakok topológiai tulajdonságaira jellemző stílusjegyek írásemlékről írásemlékre változnak.

Jelen cikkben olyan kutatás első eredményeiről számolunk be, amely gyakran ismeretlen korban és körülmények között készült feliratok hasonlósági viszonyainak beazonosításához kíván a kutatóknak segítséget nyújtani azzal, hogy egy minősítő algoritmussal meghatározza egy adott felirathoz alkalmazott írásváltozat lehetséges rokonsági viszonyait, és a különböző feliratokat hasonlóságuk alapján matematikai statisztikai eszközök felhasználásával osztályozza.

Az olvasattal rendelkező feliratok rokonsági fokának meghatározása eltér az optikai karakterfelismerés (OCR) problematikájától, ahol egy ismert írással vagy írásváltozattal, de azon belül egyedi grafémaalakokkal készült felirat elemzéséből kell visszakövetkeztetni a felirat jelentésére. Esetünkben viszont egy írás grafémaalakjainak adott feliratban megfigyelhető megjelenéséből kell meghatározni azt, hogy az adott feliratban szereplő grafémaalakok az adott írás mely változatához köthetők. A kidolgozott eljárás az egyes grafémák topológiai felbontására, a vizsgált feliraton azonosítható stílusjegyekre és azok kiértékelésére kifejlesztett algoritmusra épül, továbbá felhasználja a vizsgált írás ismert körülményekkel rendelkező írásemléke grafémaalakjainak (ábécéjének) kanonikus összetevőiből alkotott kanonikus lenyomatait, röviden abécélenyomatait.

A kidolgozott módszert a rovás íráscsaládba tartozó (Hosszú 2012) székely-magyar rovás (szmr.) írásváltozatait képviselő feliratok, grafémagyűjtemények stílusa rokonsági viszonyainak feltárásához alkalmaztuk. Az *1. táblázat* a vizsgált írás különböző változatait képviselő írásemlékekben (Nikolsburgi ábécé [Jakubovich 1935:2–11], Bél Mátyás-féle ábécé [Bél 1718 *apud* Sebestyén 1904: 403–408], Patakfalvi Sámuel-féle felirat [Hosszú 2010b]) található néhány grafémaalakot mutat be. A részletekből az is kitűnik, hogy amíg a Nikolsburgi és Patakfalvi Sámuel-féle írásemlékek grafémaalakjai egyenes szakaszokból épülnek fel, addig Bél Mátyás ábécéjét jellemzően íves alakzati elemek alkotják. Léteznek grafémák, amelyek a legtöbb írásemlékben hasonló topológiai tulajdonságokat mutatnak, ugyanakkor ezek ellentéte is előfordul, amikor az adott hangot jelölő grafémaalak egymástól eltérő alakzati elemekből épül fel (Hosszú 2010a).

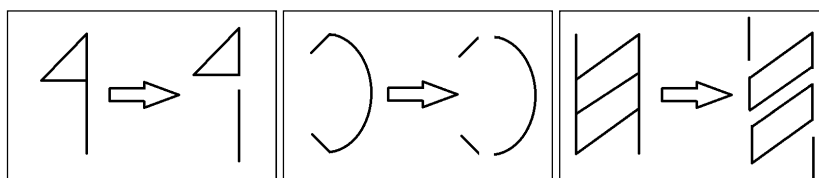
1. táblázat: Kiválasztott szmr. írásemlékekben használt néhány grafémaalak és átbetűzési értéke

Írásemlék	<a>		<c>	<cs>	<d>	<e>	<é>	<f>
Nikolsburgi ábécé	ḡ	ḥ	↑	ḡ	ḡ	ḡ		ḡ
Bél Mátyás-féle ábécé	ḡ	ḥ	ḡ	ḡ	ḡ	ḡ	ḡ	ḡ
Patakfalvi Sámuel-féle biblia	ḡ	ḥ	↑		ḡ	ḡ	–	ḡ

2. A kifejlesztett eljárás

2.1. Grafémák kanonikus felbontása









Az általunk bevezetett *kanonikus alak* alatt a grafémák kétdimenziós topológiai részalakzataiból, úgynevezett *kanonikus összetevőiből* vagy röviden *primitívjeiből* való felépítését értjük. Jelen kutatásunkban a következő primitíveket különböztetjük meg: háromszögszerű alakzat, körszerű hurok, négyszögszerű alakzat, függőleges szakasz, vízszintes szakasz, ferde szakasz, ív, keresztezés, pont. Így összesen $z = 9$ primitívet alkalmazunk.



1. ábra: A szmr. \sqsubset $\langle a \rangle$, \smile $\langle o \rangle$ és \mathbb{N} $\langle z \rangle$ grafémaalakok kanonikus felbontása

A grafémák kanonikus alakra bontását az 1. ábra szemlélteti, ahol az \sqsubset $\langle a \rangle$ grafémát háromszögszerű alakzatra és függőleges szakaszra, az \smile $\langle o \rangle$ grafémát ívre és ferde szakaszokra, míg a \mathbb{N} $\langle z \rangle$ grafémát négyszögszerű alakzatokra és merőleges szakaszokra bontottuk. A kanonikus alakokat úgy választottuk meg, hogy azok páronként függetlenek, azaz diszjunktak legyenek. Egy adott írásváltozat grafémáihoz tartozó $z = 9$ elemű graféma-primitíveket *kanonikus alakoknak* nevezzük. A z számú elemből álló kanonikus alakokat megfeleltethetjük egy-egy pontnak a graféma-primitívek z -dimenziós terében. Példaként a Bél Mátyás-féle szmr. ábécé néhány grafémájának kanonikus alakját bemutatja a 2. táblázat.

2. táblázat: A Bél Mátyás-féle szmr. ábécé néhány grafémájának kanonikus felbontása (részlet)

Kanonikus alak (graféma-primitív)	$\langle a \rangle$	$\langle b \rangle$	$\langle c \rangle$	$\langle cs \rangle$	$\langle d \rangle$	$\langle e \rangle$	$\langle é \rangle$	$\langle f \rangle$
								
Háromszögszerű alakzat	0	0	0	0	0	0	0	0
Körszerű hurok	1	0	0	0	0	0	0	1
Négyszögszerű alakzat	0	0	0	1	0	0	0	0
Függőleges szakasz	0	0	0	1	1	0	0	0
Vízszintes szakasz	0	0	0	1	1	0	0	0
Ferde szakasz	0	0	0	0	0	0	1	0
Ív	1	1	1	0	0	1	1	0
Keresztezés	0	1	0	0	1	1	0	1
Pont	0	0	0	0	0	0	0	0

A 2. táblázat soraiba kerülnek a kanonikus alakok típusai, az oszlopaiba pedig a vizsgált írásemlék grafémaalakjai. Az analízis során grafémánként a kanonikus alakoknak (graféma-primitíveknek) csupán a létezését jelöljük, és nem vesszük figyelembe egy adott grafémában az egyes kanonikus alakok előfordulási számát. Ebből adódik, hogy az egyes mezőkben csak 1 vagy 0 szerepel attól függően, hogy az adott graféma grafémaalakjának kanonikus felbontásában szerepel-e vagy nem egy adott kanonikus alak (graféma-primitív). Megjegyzendő, hogy egy korábbi cikkünkben (Tóth és mtsai. 2010) bemutattunk egy másik, általunk kidolgozott jelentésazonosító

eljárást, ahol a grafémák topológiai tulajdonságait vizsgáltuk, és figyelembe vettük a topológiai tulajdonságok előfordulásainak számát is.

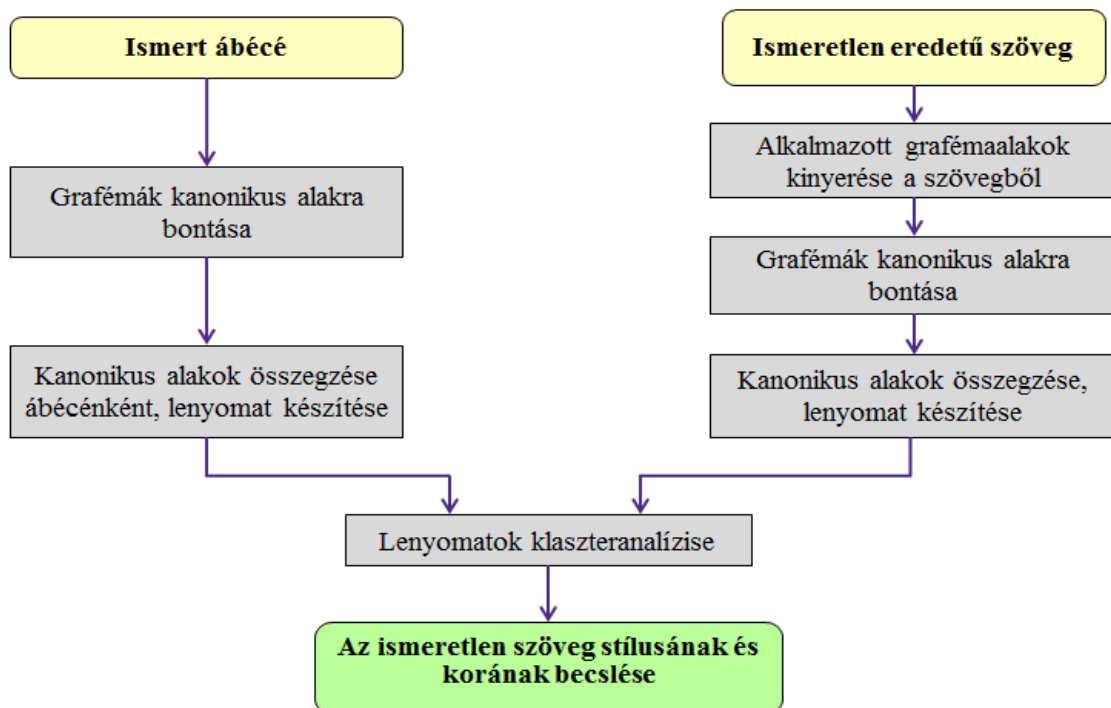
2.1. Az ábécék kanonikus lenyomata

Egy írásváltozatot képviselő, m számú grafémából álló grafémagyűjteményben, röviden ábécében szereplő grafémák grafémaalakjainak kanonikus felbontása egy $z = 9$ sorból és m oszlopból álló Φ mátrixot alkot. Ennek a Φ mátrixnak egy eleme $\alpha_{i,j}$, amely az ábécé j -dik grafémája kanonikus alakjának mint vektornak az i -dik eleme. Ennek graféma primitívénkénti összegzéséből és normálásából (1) szerint létrehozható egy mérőszámnak tekinthető vektor, ami egy írásváltozatot képviselő **ábécé** $\Phi = (\phi_1, \dots, \phi_j, \dots, \phi_z)$ **kanonikus lenyomata**, röviden **ábécélenyomat** (*alphabet printout*). Az ábécélenyomat így a 2. táblázathoz hasonló mátrix sorainak összegzéséből tevődik össze.

$$\phi_i = \frac{\sum_{j=1}^m \alpha_{i,j}}{m}, \quad i = 1, 2, \dots, z \quad (1)$$

2.3. A kifejlesztett algoritmus leírása

Az általunk kifejlesztett algoritmus célja az olvasattal rendelkező, de ismeretlen keletkezési körülményekkel rendelkező írásemlékekben használt írásváltozat rokonsági kapcsolatainak felderítése. Így nem foglalkozunk az egyes grafémák pl. bitkép alapján való felismerése, amellyel számos közlemény foglalkozik (Doermann & Jaeger 2008). A kifejlesztett és *kanonikus alaktár*nak nevezett graféma primitív-készlettel az eddig elemzett írásemlékekben előforduló grafémaalakokat teljes egészében le lehet írni. Így a jelenleg $z = 9$ elemből álló kanonikus alaktárat a jövőben csak abban az esetben kell bővíteni újabb graféma primitívekkel, ha egy vizsgált írásemlékben olyan grafémaalakot találunk, amely a jelenleg használt graféma primitívekkel nem bontható fel. Az algoritmus tömbvázlatát a 2. ábra szemlélteti.



2. ábra: Az algoritmus tömbvázlata

Az algoritmusnak két bemenete van. Az egyik bemenet rendre megkapja egy adott írásváltozattal készült írásemlékből származó grafémaalakokat, a másik bemenetre pedig egy ismeretlen írásváltozatú írásemlékből kinyert grafémaalakok kerülnek. Az algoritmus lépései: (i) Meghatározzuk az általunk definiált kanonikus alakokat grafémánként; ez intuitív módon történik, a már korábban definiált kanonikus alakok felhasználásával. (ii) A kanonikus alakokból úgynevezett lenyomatot készítünk, amely a vizsgált írás egy adott változatában szereplő jellegzetes grafémaalakokra jellemző, akár egy ujjlenyomat. (iii) Az ismeretlen eredetű feliratban található jelekhez rendelhető grafémaalakokból egy ábécét készítünk. Megjegyzendő, hogy a jelen vizsgálatban a feliratnak van olvasata, a jeleit hozzá lehet rendelni egy adott írás grafémáihoz, csak a feliratnak csak az eredete ismeretlen. (iv) Meghatározzuk az általunk definiált kanonikus alakokat grafémánként. (v) A kanonikus alakokból lenyomatot készítünk (kanonikus lenyomat), ami az ismeretlen eredetű felíratra jellemző. (vi) Az ábécék kanonikus lenyomatait (ábécélenyomatokat) adatelemzési eszközökkel vizsgáljuk, így lehetővé válik az egyes ábécék által reprezentált vizsgált írásváltozatok hasonlósági viszonyainak analízise.

2.4. Az ábécélenyomatok klaszteranalízise

A klaszteranalízis arra a problémára keresi a megoldást, hogy hogyan rendezhetjük megfigyeléseinket – azok hasonlósága, illetve különbözősége alapján – valamilyen struktúrába úgy, hogy ezzel egy csoportosítást hajtsunk végre. Segít az új minták azonosításában, csoportosításában és osztályozásában. Vizsgálataink során csak kemény klaszterezéssel (*hard/crisp clustering*) foglalkozunk, vagyis mindegyik objektumot pontosan egy klaszterbe sorolunk (Romesburg 2004).

Az objektumok hasonlóságának, illetve eltérésének mértéke az objektumok páronkénti különbözősége (*dissimilarity*). A klaszterező algoritmusok a vizsgált objektumokat valamilyen különbözőségi mérték alkalmazásával csoportosítják, illetve különítik el. Ezért fontos az alkalmazáshoz megfelelő különbözőségi mérték kiválasztása, mert az eltérő különbözőségi mértékek használata ugyanazon adathalmaz különböző csoportosításaihoz vezethet (Khan 2000; Kovács 2008). A különbözőség bizonyos feltételek teljesülése esetén matematikai értelemben vett *metrikának* tekinthető. Jelöljük $d(x_i, x_j)$ -vel az x_i és x_j objektum különbözőségét. Egy $d(x_i, x_j)$ különbözőségi függvényt akkor tekintünk távolságnak, ha teljesülnek rá a (2)-beli feltételek, amelyek a metrikus tér általános tulajdonságait írják le, és az ezeket kielégítő $d(x_i, x_j)$ függvényt metrikus függvénynek vagy röviden metrikának is nevezik (Füstös 2009).

1. $\forall x_i, x_j: d(x_i, x_j) = d(x_j, x_i)$ (szimmetria),
2. $\forall x_i: d(x_i, x_i) = 0$,
3. $\forall x_i, x_j: 0 \leq d(x_i, x_j) < +\infty$ (nem-negativitás),
4. $\forall x_i, x_j, x_k: d(x_i, x_k) \leq d(x_i, x_j) + d(x_j, x_k)$ (háromszög-egyenlőtlenség).

Ha a (2)-beli feltételek közül csak a háromszög-egyenlőtlenségnek nevezett utolsó nem teljesül, akkor a $d(x_i, x_j)$ különbözőségi függvényt *szemimetrikának* hívják. A leggyakrabban alkalmazott különbözőségek egyike az euklideszi távolság, amely metrika, és másika az euklideszi különbözőség négyzete, amely pedig szemimetrika.

Célunk egy optimális különbözőségi mérték kiválasztása volt, amely alkalmazásával egy adott írás változatait reprezentáló ábécék olyan csoportosítása érhető el, amelynek eredménye a legközelebb áll a múltban ténylegesen lezajlott fejlődésükhöz. Az ábécélenyomatok klaszteranalízisét a MATLAB R2009A keretszoftver alkalmazásával végeztük el (MathWorks 2006). Az összevonó típusú, hierarchikus klaszterezési eljárás általános menete a következő: (1. lépés) Kiindulás q számú egyelemű csoportból (klaszterből), lényegkiemelés. (2. lépés) A hasonlósági (vagy távolság-) mátrix maximális (minimális) elemének, vagyis a két leghasonlóbb klaszternek a megkeresése. (3. lépés) A két leghasonlóbb klaszter összevonása, ezzel a klaszterek számának eggyel csökkentése. Az új klaszter többtől mért távolságának (hasonlóságának) újraszámítása. (4. lépés) A 2. és 3. lépést

$(q - 1)$ -szer elvégezve minden objektum egy közös klaszterbe kerül.

A lényegkiemelés az objektumok, azaz esetünkben a vizsgált grafémakészletek (ábécék) leképezését jelenti egy jól kezelhető kódba (vektorba), amelyek a grafémák kanonikus alakjaiból elkészített ábécélenyomatok. Tekintsünk egy $(q \times z)$ dimenziós X adatmátrixot. Ennek sorait a vizsgált q számú ábécének az (1) szerint számított ábécélenyomatai alkotják. Az adatmátrix $z = 9$ számú oszlopa az ábécélenyomatok egyes elemei. Így összesen q különböző objektumot, vagyis ábécélenyomatot (x_1, x_2, \dots, x_q) hasonlítunk össze. Ezen objektumok között értelmezhető a $d(x_i, x_j)$ különbözőségi függvény.

Az összevonó eljárások eredménye a klaszterek hierarchikus elrendezését tükröző kétdimenziós ábrán is megjeleníthető, amelyet dendrogramnak nevezünk, ha nem engedünk meg átfedő klasztereket (Füstös 2009). A dendrogram vízszintes tengelyén az egyedek sorszámainak, függőleges tengelyén pedig a klaszterek összevonásának szintjeit tüntetjük fel. A függőleges tengelyen hasonlósági vagy távolságmértékek találhatók a bemeneti adatoknak megfelelően (MathWorks 2007).

A különbözőségi mértékek használatát a korrelációs metrika példáján szeretnénk részletezni. Ehhez az egyes ábécék lenyomatvektorát független valószínűségi változónak tekintjük, amelynek felvehető értékei az egyes grafémaprimitívek előfordulási aránya, amint azt a 3. táblázat bemutatja. Tekintsük az (x, y) változópárt mint két valószínűségi változót. Ha ezek Φ ábécélenyomatok, akkor a Φ egyes ϕ_i elemeit (1) szerint számíthatjuk. Így az egyes valószínűségi változóknak $z = 9$ különböző elemét ismerjük, ugyanis ezek az ábécélenyomatok egyes elemei. Az (x, y) valószínűségi változópár függőségét a kovarianciájukkal mérjük, lásd (3).

$$\text{cov}(x, y) = M[(x - M(x))(y - M(y))], \quad (3)$$

ahol $M(x)$, ill. $M(y)$ az x , ill. y valószínűségi változók várható értéke. Az x és y valószínűségi változók függőségét a $\rho_{x,y}$ Pearson-féle korrelációs együttható adja meg, amely a (4) szerint számítható (Rodgers & Nicewander 1988).

$$\rho_{x,y} = \frac{\text{cov}(x, y)}{D(x)D(y)}, \quad (4)$$

ahol $D(x)$, ill. $D(y)$ az x , ill. y valószínűségi változók szórása.

A korrelációs számítással azt vizsgáljuk, hogy az elemzésbe vont metrikus változók között mennyire szoros a kapcsolat. Két metrikus (x, y) változó közötti kapcsolat vizsgálatánál pontdiagramot készíthetünk az (x, y) változópár alapján. A pontdiagram alapján megállapíthatjuk a változópár közötti kapcsolat típusát, amely lehet lineáris vagy nemlineáris. Lineáris kapcsolat esetén a pontok egy képzeletbeli egyenes, nemlineáris kapcsolat esetén egy szabályos görbe körül szóródnak. Lineáris kapcsolatnál a pontoknak a képzeletbeli egyenes körüli szóródásából következtethetünk arra, hogy mennyire szoros a kapcsolat a két változó között. Az egyenes meredekségéből pedig következtethetünk a kapcsolat irányára, amely pozitív vagy negatív lehet. A pozitív irányú kapcsolat azt jelenti, hogy ha az egyik változó értéke nő, akkor a másik változó értéke is nő és fordítva. A kapcsolat erősségének jellemzésére lineáris kapcsolat esetén az úgynevezett *lineáris korrelációs együtthatót* használjuk. A lineáris korrelációs együttható értéke a $[-1, 1]$ tartományba esik. Előjele megadja a két változó közötti kapcsolat irányát, míg abszolút értéke a kapcsolat erősségét. A nullához közeli érték gyenge, az egyhez közeli érték erős kapcsolatot jelent (Kovács 2008). Mivel esetünkben az egyes ábécélenyomatoknak, mint valószínűségi változónak csak $z = 9$ különböző értékét ismerjük, ezért a $\rho_{x,y}$ Pearson-féle korrelációs együttható helyett az r_z Pearson-féle mintakorrelációs együttható (*sample Pearson correlation coefficient*) számítható, lásd (5) (Deza & Deza 2009).

$$r_z = \frac{\sum_{i=1}^z (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^z (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^z (y_i - \bar{y})^2}}, \quad (5)$$

ahol \bar{x} , ill. \bar{y} az x , ill y valószínűségi változók tapasztalati várható értéke.

3. Eredmények

A vizsgálatba $q = 12$ írásváltozatot vontunk be, amelyek közül a legtöbbet a XV–XVIII. sz.-ből származó történelmi szmr. írásemlékek képviselnek (Sebestyén 1904: 403–408; Sebestyén 1909: 245; Sebestyén 1915: 84–85, 112, 124–126; Hosszú 2012: 194–197, 223–224, 232–233). Ez alól kivétel a történelmi előzményeken alapuló, a nemzetközi karakterkódolásra előterjesztett szmr. grafémakészlet (Demeczky és mtsai. 2012), a XX. sz.-i Magyar Adorján-féle szmr. ábécé (Magyar 1970: 14), a szmr.-hoz sok tekintetben hasonló steppei rovás írásemlékekből származó ábécé (Vékony 1987; Vékony 2004; Hosszú 2013: 42–44), és az ismeretlen eredetű, talán XVIII. sz.-i, a szmr.-tól eltérő, ismeretlen íráshoz tartozó Gellei-féle ábécé (Sebestyén 1915: 128). Egydimenziós vektor formájában elkészítettük a vizsgált írásváltozatokban található grafémaalakok kanonikus lenyomatait, vagyis az ábécélenyomatokat, amelyeket bemutat a 3. táblázat. Ennek oszlopaiban a kanonikus alakok (grafémaprimitívek) típusai, soraiban pedig a vizsgált írásemlékek és az azokhoz tartozó ábécélenyomatok találhatók.

3. táblázat: A vizsgált írásváltozatok kanonikus lenyomatai (ábécélenyomatok)

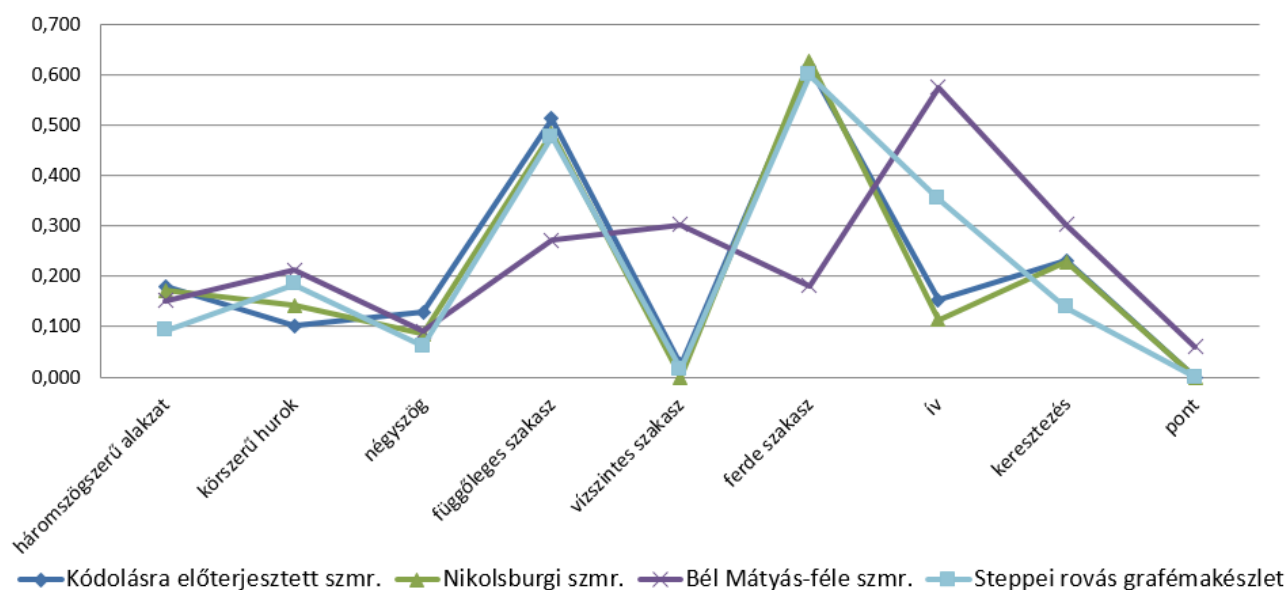
Sor-szám	Ábécé	háromszög-szerű alakzat	körszerű hurok	négy-szög	függőleges szakasz	vízszintes szakasz	ferde szakasz	ív	keresztelés	pont
1	Kódolásra előterjesztett szmr.	0,179	0,103	0,128	0,513	0,026	0,615	0,154	0,231	0,000
2	Patakfalvi Sámuel-féle szmr.	0,243	0,135	0,081	0,432	0,054	0,595	0,216	0,189	0,000
3	Nikolsburgi szmr.	0,171	0,143	0,086	0,486	0,000	0,629	0,114	0,229	0,000
4	Bél Mátyás-féle szmr.	0,152	0,212	0,091	0,273	0,303	0,182	0,576	0,303	0,061
5	Magyar Adorján-féle szmr.	0,200	0,086	0,114	0,457	0,000	0,571	0,143	0,314	0,000
6	Rettegi István-féle szmr.	0,091	0,182	0,061	0,212	0,182	0,697	0,242	0,303	0,000
7	Kájoni János-féle régi szmr.	0,174	0,174	0,087	0,435	0,087	0,304	0,261	0,348	0,043
8	Gyulafehérvári szmr.	0,182	0,212	0,091	0,455	0,121	0,424	0,212	0,273	0,030
9	Harsányi János-féle szmr.	0,182	0,242	0,091	0,242	0,091	0,303	0,667	0,364	0,000
10	Gönczi György-féle szmr.	0,156	0,156	0,125	0,344	0,156	0,281	0,406	0,344	0,063
11	Steppei rovás grafémakészlet	0,092	0,185	0,062	0,477	0,015	0,600	0,354	0,138	0,000
12	Gellei-féle ismeretlen írású	0,194	0,639	0,056	0,583	0,500	0,278	0,722	0,528	0,000

A vizsgált $q = 12$ számú ábécélenyomatra kiszámított korrelációs együtthatókból alkotott korrelációs mátrix a 4. táblázatban látható. Ennek főátlójában az ábécélenyomatok önmagukkal vett korrelációját kifejező 1-esek állnak. Sötét háttérrel kiemeltük a 0,9-nél nagyobb korrelációs együttható értékek.

4. táblázat: Az ábécélenyomatok korrelációs mátrixa

Sorszám	Ábécé	1	2	3	4	5	6	7	8	9	10	11	12
1	Kódolásra előterjesztett szmr.	1,000	0,973	0,992	0,049	0,982	0,777	0,796	0,936	0,291	0,571	0,905	0,147
2	Patakfalvi Sámuel-féle szmr.		1,000	0,972	0,137	0,952	0,816	0,767	0,923	0,388	0,589	0,930	0,204
3	Nikolsburgi szmr.			1,000	0,007	0,978	0,798	0,783	0,937	0,270	0,531	0,899	0,150
4	Bél Mátyás-féle szmr.				1,000	0,058	0,234	0,443	0,269	0,868	0,794	0,337	0,836
5	Magyar Adorján-féle szmr.					1,000	0,781	0,837	0,927	0,332	0,612	0,855	0,145
6	Rettegi István-féle szmr.						1,000	0,570	0,739	0,440	0,544	0,791	0,268
7	Kájoni János-féle régi szmr.							1,000	0,921	0,594	0,866	0,774	0,568
8	Gyulafehérvári szmr.								1,000	0,439	0,714	0,898	0,461
9	Harsányi János-féle szmr.									1,000	0,873	0,554	0,706
10	Gönczi György-féle szmr.										1,000	0,698	0,708
11	Steppei rovás grafémakészlet											1,000	0,382
12	Gellei-féle ismeretlen írású												1,000

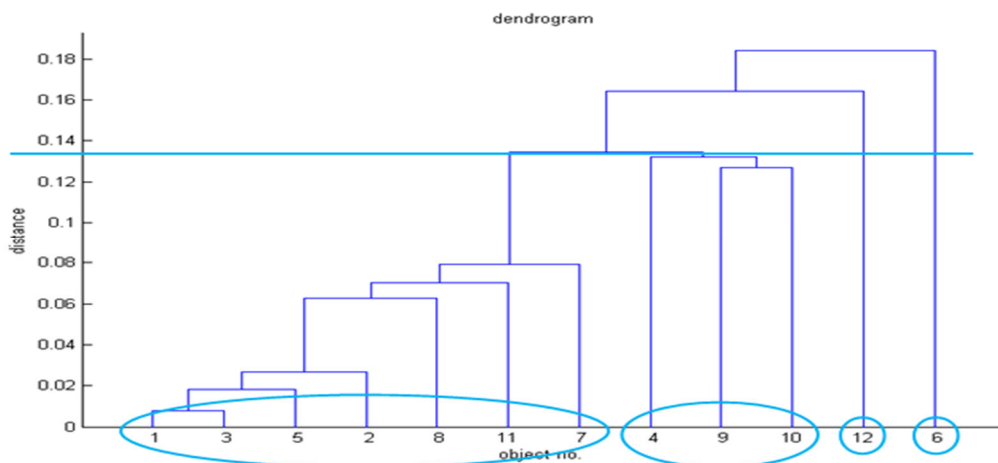
Megállapítható például, hogy a Nikolsburgi ábécé lenyomata erős korrelációban áll a Patakfalvi-féle íráslel grafémaalakjaiból alkotott ábécé lenyomatával. A nemzetközi kódolásra előterjesztett szmr. ábécé nagyrészt egyenes vonalvezetésű kanonikus alakokból tevődik össze, ezért nem meglepő, hogy lenyomata erősen korrelál a Patakfalvi-féle íráslel és a Nikolsburgi ábécé lenyomataival. Ugyanakkor Bél Mátyás szmr. ábécéje, amely nagyrészt íves kanonikus alakokból tevődik össze, kisebb korrelációt mutat az előbbiekkal.



4. ábra: A kódolásra összeállított, a Nikolsburgi és a Bél Mátyás-féle szmr. ábécé, valamint steppei rovás íráslelkek alapján összeállított steppei rovás grafémakészlet lenyomatának grafémaprivitívénkénti összevetése

Amint az a 4. ábrán is látható, a kódolásra előterjesztett szmr., a Nikolsburgi szmr. és a steppei rovás ábécék lenyomatára illesztett görbék többnyire követik egymást, szemben a Bél Mátyás-féle szmr. ábécé lenyomatával. Ez egybevág azzal, hogy a Bél Mátyás-féle ábécé lenyomatával számolt korrelációs együttható a másik három ábécé lenyomatával nullához közeli eredményt adott, vagyis alig korrelálnak egymással. Ugyanakkor észrevehető a korrelációs metrika lenyomatokra való alkalmazhatóságának problémája, mivel csak az adatsorok emelkedését és csökkenését vizsgálja, nem ad információt az adatok méretéről. Ugyanez a gond látható a korrelációs metrika alkalmazásával végzett hierarchikus klaszterelemzésből is, amelynek eredményét bemutatja az 5. ábra. Ezen látható

dendrogram vízszintes tengelyén lévő sorszámok a 4. táblázat első oszlopa szerint az egyes vizsgált ábécélenyomatokat jelentik. A hierarchikus klaszterelemzés eredményéből egy konkrét klaszterstruktúrát úgy lehet megállapítani, hogy kijelölünk egy elfogadható csoporton belüli távolságot, amit a dendrogram függőleges tengelyén mérünk. A függőleges tengelyen kijelölve egy távolságértéket, ezen keresztül a vízszintes tengellyel párhuzamosan húzott egyenes a dendrogramban kijelöli az aktuális távolságértékhez tartozó klasztereket. Az ugyanazon klaszterbe sorolt ábécélenyomatok sorszámait egybe karikáztuk. Az 5. ábra dendrogramján látható, hogy egy klaszterbe került a Nikolsburgi szmr. ábécé a steppei rovás grafémakészlettel, holott a lenyomatuk számértékben jelentősen eltér egymástól.



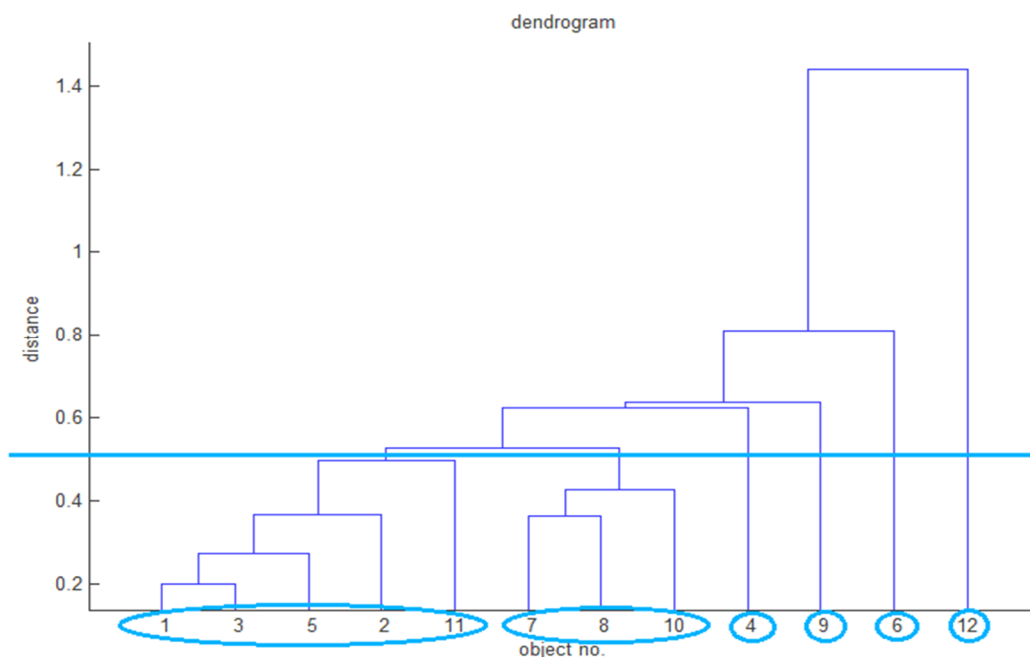
5. ábra: A korrelációs metrikával történt számítás eredménye (MATLAB program kimenet)

A korrelációs metrika helyett alkalmazható a *Minkowski-metrika* (6), illetve ennek speciális esetei (Deza & Deza 2009).

$$d_{xy} = \sqrt[p]{\sum_{j=1}^n |x_j - y_j|^p} \quad (6)$$

A Minkowski-metrika elnevezése $p = 1$ esetén *Manhattan-metrika* (más néven *taxi-* vagy *cityblock-metrika*), ahogy az a (7)-ből látható. A Manhattan-metrikával számított dendrogramot bemutatja a 6. ábra.

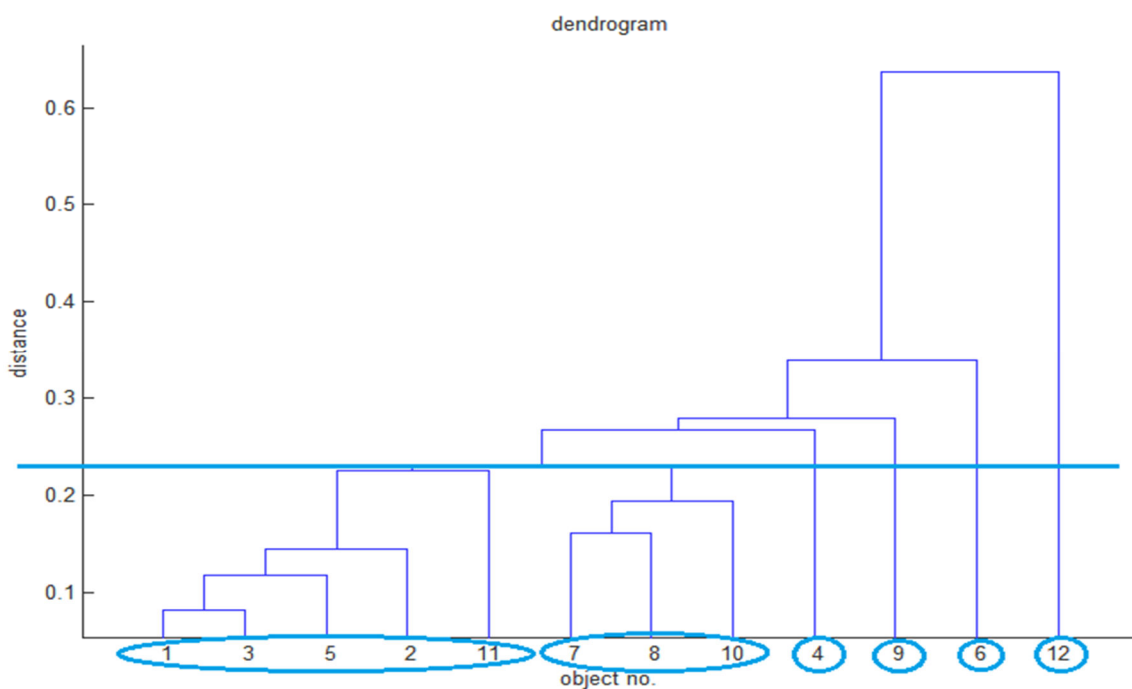
$$d_{xy} = \sum_{j=1}^n |x_j - y_j| \quad (7)$$



6. ábra: A Manhattan-metrikával történt számítás eredménye (MATLAB program kimenet)

A Minkowski-metrika elnevezése $p = 2$ esetén *euklideszi metrika*, amelynek képlete a (8). Az euklideszi metrikával számított dendrogram a 7. ábrán látható. A 6. ábra és a 7. ábra összehasonlításából látható, hogy a Manhattan- és az euklideszi metrika lényegében azonos klaszterezésre vezetett.

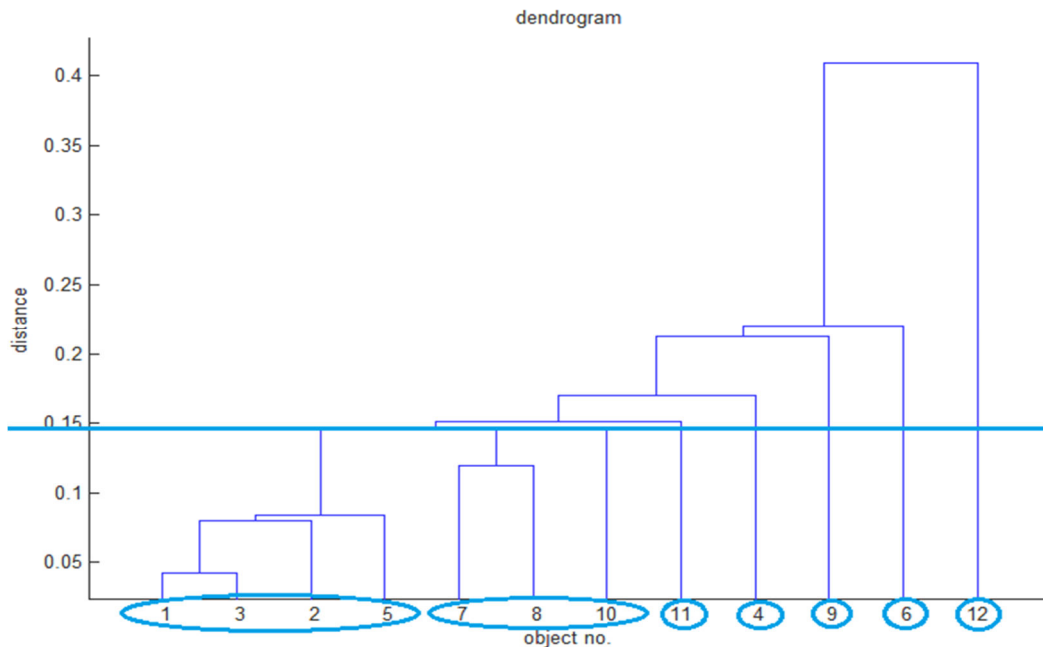
$$d_{xy} = \sqrt{\sum_{j=1}^n |x_j - y_j|^2} \quad (8)$$



7. ábra: Az euklideszi metrikával történt számítás eredménye (MATLAB program kimenet)

A Minkowski-metrika elnevezése $p = \infty$ esetén *Csebisev-metrika*, ahogy az a (9)-ből látható. A Csebisev-metrikával számított dendrogram a 8. ábrán látható. A Csebisev-metrikával készült klaszterezés az eddiekhez képest sokkal jobban elkülönítette a steppei rovás grafémakészletet a topológiailag hasonlóan egyenes vonalaktól álló, de eltérő íráshoz tartozó szmr. grafémakészletektől.

$$d_{xy} = \max_j \{|x_j - y_j|\} \quad (9)$$

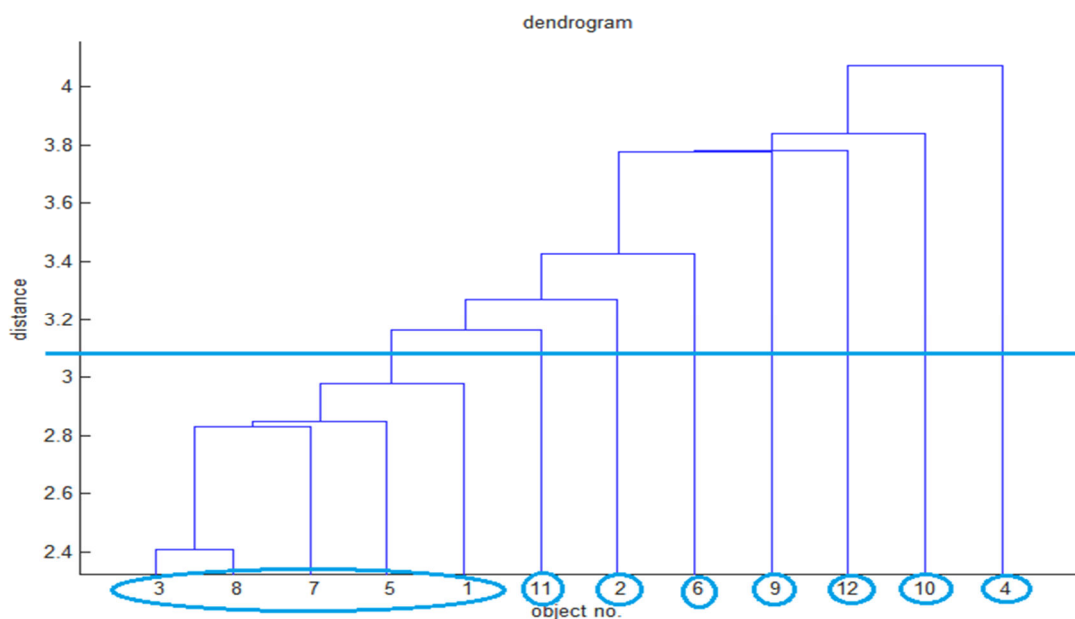


8. ábra: A Csebisev-metrikával történt számítás eredménye (MATLAB program kimenet)

Ha az az x és y változók korreláltak, akkor érdemes lehet a (10) szerint definiált *Mahalanobis-metrikát* alkalmazni.

$$d_{xy} = \sqrt{(x - y)^T COV^{-1}(x - y)}, \quad (10)$$

ahol az x és y változók vektorok, és a COV egy kovariancia mátrix. Ha a kovariancia mátrix egységmátrix, vagyis x és y teljesen korrelálatlan, akkor (10) képlet a (8) szerinti euklideszi metrikával azonos. A Mahalanobis-metrikával végzett hierarchikus klaszterelemzés eredményét bemutatja a 9. ábra.



9. ábra: A Mahalanobis-metrikával történt számítás eredménye (MATLAB program kimenet)

A különböző metrikával végzett klaszterelemzések összevetéséből látható, hogy a metrika megválasztása jelentősen befolyásolja az eredményt. A klaszterezés csak felhívja a figyelmet az adatstruktúrában rejlő lehetséges összefüggésre, de nem ad bizonyítékot azok meglétére. A legtöbb vizsgálat elkülönítette a különösen kalligrafikus írásváltozatokat, ilyenek pl. a 6-os Rettegi István-féle szmr., a 9-es Harsányi János-féle szmr. és a 12-es Gellei-féle – ismeretlen íráshoz tartozó – ábécé. Ezzel szemben az egyenes vonalakat tartalmazó grafémaalakokat tartalmazó ábécék általában közös csoportba kerültek.

Az alkalmazott metrikák használatát és a vizsgált írásemlékeket összehasonlítva megállapítottuk, hogy az ábécélenyomatok elkészítése, és azok klaszteranalízise együtt lehetőséget adnak ismeretlen eredetű vagy stílusú írásváltozatok kategorizálására, beazonosítására.

4. Összefoglalás, következtetések

A cikkben bemutatásra került egy eljárás a különböző írásváltozatokat reprezentáló, olvasattal rendelkező feliratok vagy grafémagyűjtemények (ábécék) hasonlósági viszonyainak feltárására. Az eljárás az egyes feliratok grafémáit azok topológiai tulajdonságait kifejező kanonikus alakokra bontja, majd az így kapott kanonikus alakokat összegezve és normálva létrehozza az adott írásemlékre jellemző mérőszámnak tekinthető vektort, az ún. ábécélenyomatot. Így az írásváltozatok összehasonlítása átalakítható az ábécélenyomatok hasonlósági elemzésére. A kifejlesztett eljárás gyakorlati haszna abból adódik, hogy számos történelmi felirat létezik, amelynek jelalakjai ugyan világosan azonosíthatók, viszont a felirat készítője és a készítés körülményei ismeretlenek. A cikkben bemutatott vizsgálatok azt igazolják, hogy az írásváltozatok jellemzésére kifejlesztett kanonikus alakokra felbontás és az ezekből kialakított összegzett jellemző vektor, az ún. ábécélenyomat alkalmas lehet a különböző feliratok hasonlósági viszonyainak feltárására. Így ha ismert feliratok (írásemlékek) ábécélenyomatait egy ismeretlen keletkezésű felirat ábécélenyomatával vetjük össze, az utóbbi stílusbeli rokonsági viszonyaira becslést lehet adni. A kidolgozott módszerünk illeszthető egy fejlesztés alatti jelentésazonosító szoftverbe, amelynek célja a nehezen olvasható feliratok, szövegek megfejtése (Tóth és mtsai. 2010).

5. Köszönetnyilvánítás

A munka szakmai tartalma kapcsolódik a „Új tehetséggondozó programok és kutatások a Műegyetem tudományos műhelyeiben” c. projekt szakmai célkitűzéseinek megvalósításához. A projekt megvalósítását a TÁMOP – 4.2.2.B-10/1--2010-0009 program támogatja.

6. Irodalomjegyzék

- Acharya és mtsai. 2009: Acharya U. Dinesh; Reddy, N. V. Subba & Makkithaya, Krishnamoorthi (2009): Multilevel classifiers in recognition of handwritten Kannada numerals. In: *Proceedings of World Academy of Science, Engineering and Technology (WASET)*, Vol. 32, Aug 2009. 278–283.
- Bél 1718: Bél Mátyás (1718): *De vetere literatura hunno-scythica exercitatio*, Lipcse.
- Demeczky és mtsai. 2012: J. Demeczky, G. Hosszú, T. Rumi, L. Sípos & E. Zelliger (2012): *Revised proposal for encoding the Rovas in the UCS*. Individual Contribution for consideration by UTC and ISO/IEC JTC1/SC2/WG2, October 14, 2012, Universal Multiple-Octet Coded Character Set. *ISO/IEC JTC1/SC2/WG2 N4367*, <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4367.pdf>; UTC: L2/12-331.
- Deza & Deza 2009: M. M. Deza & E. Deza (2009): *Encyclopedia of Distances*, Berlin, Heidelberg: Springer.
- Dian 2009: Dian Szabolcs (2009): Jelentésazonosító szoftver fejlesztése, *Tudományos Diákköri Konferencia*, Budapest: Budapesti Műszaki és Gazdaságtudományi Egyetem, Villamosmérnöki és Informatikai Kar.
- Doermann & Jaeger 2008: Doermann, David & Jaeger, Stefan (2008): *Arabic and Chinese Handwriting Recognition*, Springer Berlin Heidelberg, SACH 2006. Lecture Notes in Computer Science 4768. Berlin, Heidelberg: Springer.
- Füstös 2009: Füstös László (2009): *A sokváltozós adatelemzés módszerei*. Módszertani Füzetek. Budapest: MTA Szociológiai Kutatóintézete, Társadalomtudományi elemzések Akadémiai Műhelye (TEAM).
- Hosszú 2010a: Hosszú Gábor (2010): Az informatika írástörténeti alkalmazásai. (Meghívott plenáris előadás) In: Cserny László (szerk.), *IKT 2010 Informatika Korszerű Technikai Konferencia (2010. március 5–6.)*, Dunaújváros: Dunaújvárosi Főiskola Informatikai Intézet. 5–21.
- Hosszú 2010b: Hosszú Gábor (2010): *A székelly jog megjelenése egy rovásemlékben*. Diplomamunka, Budapest: Pázmány Péter Katolikus Egyetem, Jog- és Államtudományi Kar.
- Hosszú 2012: Hosszú, Gábor (2012): *Heritage of Scribes. The Relation of Rovas Scripts to Eurasian Writing Systems*. Second, Extended Edition. Budapest: Rovas Foundation. <http://books.google.hu/books?id=TyK8azCqC34C&pg=PA1>.
- Hosszú 2013: Hosszú Gábor (2013): *Rovásatlasz*. Budapest: Milani Kft.

- Izadi és mtsai. 2008: Izadi, S.; Sadri, J.; Solimanpour, F. & Suen, C. Y. (2008): A Review on Persian Script and Recognition Techniques. In: D. Doermann & S. Jaeger (Eds.), *Arabic and Chinese Handwriting Recognition. SACH 2006*. Lecture Notes in Computer Science 4768. Berlin, Heidelberg: Springer. 22–35.
- Jakubovich 1935: Jakubovich Emil (1935): A székely rovásírás legrégebbi ábécéi. *Magyar Nyelv XXXI* (1–2): 1–17.
- Kristjánsson 2007: Kristjánsson, Jón Orri (2007): *Glyph identification using neural network techniques: HORUS project*. University of Akureyri. <https://skemman.is/handle/1946/1015>, letöltve: 2009. november 20.
- Kashyap és mtsai. 2003: Kashyap, K. Harish; Bansilal; Koushik, P. Arun (2003): Hybrid neural network architecture for age identification of ancient Kannada scripts. In: *Proceedings of the 2003 IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. 5, 25–28, May 2003. V-661–664.
- Khan 2000: Khan, Nadeem Ahmad: Thesis (2000): *A Shape Analysis Model with Application to Character and Word Recognition*. Eindhoven: Technische Universiteit Eindhoven, Proefschrift.
- Khatatneh és mtsai. 2006: Khatatneh, Khalaf; El Emary, Ibrahim M. M. & Rifai, Basem A. (2006): Probabilistic Artificial Neural Network For Recognizing the Arabic Hand Written Characters, *Journal of Computer Science* 2 (12): 879–884. www.scipub.org/fulltext/jcs/jcs212879-884.pdf, letöltve: 2009. november 20.
- Kovács 2008: Kovács Péter (2008): *Elméleti összefoglalók*. Szeged: Szegedi Tudományegyetem Gazdaságtudományi Kar, Statisztikai és Demográfiai Tanszék. <http://www.eco.u-szeged.hu/oktatas/marketing-szak-ma/matstat-elmeleti>, letöltve: 2013. január 6.
- Le Cun és mtsai. 1990: Le Cun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W. & Jackel, L. D. (1990): Handwritten Digit Recognition with a Back-Propagation Network. In: David S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2*. San Francisco (CA): Morgan Kaufmann. 396–404.
- Magyar 1970: Magyar Adorján (1970): *Ős magyar rovásírás*. Warren (OH): A Fáklya.
- MathWorks 2006: The MathWorks, Inc.: MATLAB, *The Language of Technical Computing Version 7*, September 2006. www.mathworks.com.
- MathWorks 2007: The MathWorks, Inc.: *Cluster Analysis, Statistics Toolbox User's Guide*, Revised for Version 6.0, 2007. 929–946.
- Mutalib és mtsai. 2007: Mutalib, Sofianita & Rahman, Shuzlina Abdul; Yusoff, Marina & Mohamed, Azlinah (2007): Personality Analysis Based On Letter 't' Using Back Propagation Neural Network. In: *Proceedings of the International Conference on Electrical Engineering and Informatics*, Institut Teknologi Bandung, Indonesia, June 17–19, 2007. 496–499.
- Pardede és mtsai. 2012: Pardede, Raymond Eliza Ivan; Tóth, Loránd Lehel; Hosszú, Gábor & Kovács, Ferenc (2012): Glyph Identification Based on Topological Analysis. In: *Proceedings of the PhD Workshops at BME*. Budapest, March 9, 2012. 99–103.
- Rodgers & Nicewander 1988: Rodgers, J. L. & Nicewander, W. A. (1988): Thirteen ways to look at the correlation coefficient. *The American Statistician* 42 (1): 59–66. (February 1988).

- Romesburg 2004: H. C. Romesburg (2004): *Cluster Analysis for Researchers*. Raleigh (NC): Lulu.com.
- Sebestyén 1904: Sebestyén Gyula (1904): Rovás és rovásírás. (Nyolczadik közlemény) *Ethnographia* XV (8–9): 337–422.
- Sebestyén 1909: Sebestyén Gyula (1909): *Rovás és rovásírás*. Budapest: Magyar Néprajzi Társaság.
- Sebestyén 1915: Sebestyén Gyula (1915): *A magyar rovásírás hiteles emlékei*. Budapest: Magyar Tudományok Akadémia.
- Tikk 2007: Tikk Domonkos (2007): *Szövegbányászat*. Szerk.: Gerner József. Budapest: TypoTeX.
- Tóth és mtsai. 2010: Tóth L., Hosszú G., Dian Sz., Pardede R. & Kovács F. (2010): Jelentésazonosító eljárás a 16–18. századi székely-magyar rovás emlékek értelmezésére. In: Dr. Cserny László (szerk.), *IKT2010, Informatika Korszerű Technikai Konferencia*, 2010. március 5–6., Dunaújvárosi Főiskola, Dunaújváros. Dunaújvárosi Főiskola Informatikai Intézet. 5–21.
- Vékony 1987: Vékony Gábor (1987): *Későnépvándorláskori rovásfeliratok a Kárpát-medencében*. Szombathely: Életünk Szerkesztősége – Magyar Írók Szövetsége Nyugat-Magyarországi Csoportja.
- Vékony 2004: Vékony Gábor (2004): *A székely írás emlékei, kapcsolatai, története*. Budapest: Nap Kiadó.
- Wolf és mtsai. 2011: Wolf, Lior; Potikha, Liza; Dershowitz, Nachum; Shweka, Roni & Choueka, Yaacov (2011): Computerized Paleography: Tools for Historical Manuscripts. *18th IEEE International Conference on Image Processing (ICIP)*. Brussels (Belgium). 3545–3548.

Zusammenfassungen

Schriftsystemen, Lautsystemen

Abstrakt: Die Anfänge der ungarischen Schriftlichkeit von lateinischen Buchstaben weisen einige Kontaktpunkte zweierlei Sprachen auf. Diese Prozesse behaupten eine Art von Kontrastivität. Anfangs als die ungarische Sprache mit lateinischen Buchstaben geschrieben wurde, ist aufgefallen, dass das Ungarische auf dem Gebiet sowohl der Vokalen als auch der Konsonanten eine Überzahl gegenüber der Lateinischen aufweist. Daraus folgte eine Relation in der Schrift, wie ein Buchstabe vs. mehrere Phonemen bzw. mehrere Buchstaben vs. eine Phonem. In der vorhandenen Artikel handelt es sich von Bezeichnung einiger Phonemen in zwei Sprachdenkmäler mit Rowaschchrift (von Bodrog-Alsóbü in Komitat Somogy, und Vargyas in Siebenbürgen, Rumänien), zwei Sprachdenkmäler mit griechischen Buchstaben (Konstantin De administrando imperio, und die Gründungsurkunde des griechischen Nonnenkloster in Veszprémvölgy), arabisch-persische Quellen, und die Gründungsurkunde der Abtei in Tihany. Die dargestellten Phonemen sind: [e:], [β], [γ] und [dʒʲ].

Schlüsselwörter: Lautschrift, Linguistik, Phonemen Überzahl, Rowaschchrift.

Abstammungsanalyse der Grapheme

Abstrakt: Die dargelegten Forschungsergebnisse zeigen die Erkundung der Herkunftsbeziehungen der Grapheme. Es handelt sich um Verwandtschaftsbeziehungen zwischen zwei Grapheme, wenn diese über dieselbe Lautwert verfügen, oder die Lautwertunterschiede durch entsprechende linguistische bzw. paläographische Gründe zu erklären sind. Es gibt ferner immerhin eine irgendwelche Ähnlichkeit unter Graphemformen (glyphs) der Grapheme die miteinander in einer Abstammungsbeziehung stehen. Die Ähnlichkeit kann ein Zusammenfall oder eine Modifikation sein. Es ist typisch, dass die Graphemmodifikation zu Folge einer topologischen Transformation ist. Der Autor hat die bei den Graphementwicklungen erwähnenswerten topologischen Transformationen auf dem Grund der untersuchten Schriftentwicklungsprozesse festgelegt. Der Autor hat einen Algorithmus, welcher für den Aufbau zur Abstammungskette ausgearbeitet ist, zur Klärung der Herkunft jener Schriftzeichen, die schon entzifferten Schriftdenkmalen vorhanden sind, verwendet. Die Summierung der Abstammungsbeziehungen der einzelnen Grapheme hat die Entwicklungsmodellierung der von den untersuchten Graphemen gestalteten Schrift ermöglicht.

Schlüsselwörter: Computer-Paläographie, Graphem, Maschinelles Lernen, Modellierung.

Die Andronowo-Kultur (Bronzezeit) und die Gestaltwerdung der Rowaschchrift

Abstrakt: Durch unseren Vortrag wollen wir auf die Andronowo-Kultur, bzw. auf deren Beziehung zur Rowaschchrift aufmerksam machen. Nach der Zeittafel handelt es sich um die Herausbildung der Schrift, die Bilderschrift, Silbenschrift, Buchstabenschrift und eine Eigenheit der Bronzezeit, nämlich wie das Zinn als Rohstoff dem Verbraucher zur Verfügung stehen konnte. Danach erörtern wir die Quantität der Vokalen in den verschiedenen indoeuropäischen und ugrischen

Sprachen. Es folgt dann die Untersuchung der Inschriften von Orkhon. Im Spiegel der akustischen Messungen erörtern wir das etruskische Alphabet und die alttürkische Sprache. Schließlich sprechen wir über die Rowasch vor dem Zeitalter von Kagan Bilge. Zum Schluss wird noch über die Rowasch und Runenschrift diskutiert.

Schlüsselwörter: Andronowo-Kultur, alttürkische Schrift, Paläographie, Rowaschschrift.

Fragen um die frühen Datierung des Goldschatzes von Nagyszentmiklós

Abstrakt: Der Artikel will auf Fakten über den Ursprung des goldenen Schatzes von Nagyszentmiklós aufmerksam machen, welche Fakten von der wissenschaftlichen Forschung nicht oder nur marginal behandelt sind. Der Autor wirft die Möglichkeit der frühen skythisch-hunnische Entstehung des Schatzfundes auf, stellt die diesbezügliche Theorien, die Funde aus der Antike mit Rücksicht auf den Fundort vor. Diese Kenntnisse weisen darauf, welche Reiche Machtzentren in der Fundortnähe zustande bringen konnten. Ferner wird der Herkunft der Schriften des Goldschatzes, bzw. die Beziehung zwischen dem Christentum und der Steppevölker erwähnt.

Schlüsselwörter: Goldschatz von Nagyszentmiklós, Skythen, Hunnen, Rowaschschrift.

Zeugnis der rumänischen Sprache: Beiträge zur Geschichte Mitteleuropas in der Völkerwanderungszeit

Abstrakt: Es gibt eine Möglichkeit um die sprachliche Zugehörigkeit der Hunnen durch eine neue, noch nicht verwendete Methode zu untersuchen. Die Forscher haben bisher den Umstand, dass das letzte Kapitel der Hunnengeschichte zeitlich und örtlich mit dem Anfang der rumänischen Geschichte übereinstimmt, ausser Acht gelassen. Von der lateinischen Sprachgemeinschaft ausgeschiedene protorumänische Gruppen lebten am selben Gebieten, wohin sich die Überreste der Hunnen laut Forscher der hunnischen Geschichte flüchteten. Es heißt, die beiden Volksgruppen sollten miteinander in Sprachkontakt geraten. Die rumänische Sprache aufweist die sprachlichen Spuren aller Völker, die auf dem Balkan nach der Zeit von Herodot herumgekommen sind; z. B. thrakische, lateinische, griechische, slawische, ungarische, albanische usw. Elemente. Es ist der Umstand zu erklären, daß die Rumänische über keine türkischen Lehnwörter vor dem 12. Jahrhundert verfügt, obwohl die für türkischsprachig gehaltenen Hunnen jahrhundertlang im engen Sprachkontakt zu den Rumänen leben sollten.

Schlüsselwörter: Geschichte, Moesia, Hunnen, hunnische Sprache, gemeintürkische Sprachen, rumänische Sprache.

Wurzelrovasch in der Schrift

Abstrakt: Nach der Auffassung des Autors ist die wurzelbasierte Architektur der ungarischen Sprache durch ein visuelles System besser zu verstehen. Dieses System macht nicht nur die grundlegenden Bausteine der Sprache sichtbar, sondern stellt auch ihre Hierarchie dar. Die Seklerungarischen Rowaschschrift bietet eine Lösung für die Darstellung des Wurzelsystems in der ungarischen Sprache durch die Zusammendrängungsregeln in ihrer Rechtschreibung. Die technischen

und sprachwissenschaftlichen Bedingungen zum Zustandebringen der Wurzelrowasch werden diskutiert. Durch Beispiele begründet der Autor die Anwendbarkeit der Wurzelrowaschschrift für die Darstellung der ungarischen Sprache hinsichtlich der phonetischen und kognitiven Aspekte.

Schlüsselwörter: Computer-Schriftmodellierung, Szekler-ungarischer Rowaschschrift.

Die Schrift mit lateinischen Buchstaben und ihr Einfluss auf die szekler-ungarischen Rowaschschrift

Abstrakt: Die neuen Buchstaben, welche durch die Graphemenmodifikationen der szekler-ungarischen Rowaschschrift entstanden sind, stellen eine Parallelbildung mit der Transformationserscheinungen des ungarischen ABC lateinischen Ursprungs. Man kann die Wirkung der Lautbezeichnung des Ungarischen durch lateinischen Buchstaben auf die szekler-ungarischen Rowaschschrift als einen Konvergenzprozess betrachten, welche eine historische Notwendigkeit war. Durch diese Tatsache wird unter anderem bestätigt, daß die beiden Schreibsysteme gleichzeitig miteinander existierten. Der Konvergenzprozess hat sich bis zu unseren Tagen nicht abgeschlossen, da die beiden Schreibsysteme ihren Synchronzustand entsprechend dem Computerzeitalter bloß nach der Kodifizierung der szekler-ungarischen Rowasch erreichen. Aus diesem Grund sind einige historische Aufgaben, wie die Definierung der Transliterationsregeln, die Ausarbeitung der Graphemenmuster für die Unicode, welche nicht mehr weiter zu verschieben sind. Es gibt nur eine einzige Schreibart unter den Europäischen lebendigen Schreibsystemen, nämlich die szekler-ungarische Rowaschschrift, welche bisher nicht unifiziert wurde. Diese Tatsache ist seitens der Verbraucher eventuell eine marginale Frage (z.B. das Fehlen der Rowaschbuchstaben für x, y, q, w), aber in der Informationsumgebung verhindern diese die Vollbringung der Datenbasisoperationen, die genaue und eindeutige Entsprechungen unter den Corpora, womit die Rowasch endgültig aus der modernen Anwendung ausgerissen und zu einem Anzeichen vom beschränkten Gebrauch sein würde.

Schlüsselwörter: Schriftart, Rowasch Paläographie.

Verfahren der Transkription von Texten mit Rowaschschrift

Abstrakt: Dieser Artikel fasst die Art der Transliteration ungarischer Texte vom lateinischen Alphabet zur Rowaschschrift zusammen und legt die Vor- und Nachteile der Methoden dar. Die Umsetzung der regelbasierten Transkription und ihre aktuellen Anwendungen werden eingehend erörtert.

Schlüsselwörter: Rowaschschrift, Transkription, Schriftdarstellung.

Zerlegung der Graphemen zur kanonischen Komponenten

Abstrakt: Der Artikel präsentierte ein neues Verfahren zur Aufdeckung verwandtschaftliche Verhältnisse von Schriftreliquien, die verschiedene Versionen einer Schrift darstellen. Dabei wurden die Glyphen einer untersuchten Schrift in zweidimensionalen Teilformationen, sogenannten kanonischen Formen zerlegen. Die kanonische Form ist z. B. eine kreisförmige Schleife, oder ein schräger Abschnitt, ein vertikaler Abschnitt oder eine Kreuzung. Die Koeffizienten in der kanonischen Auflösung der grafischen Formen, die in den untersuchten Schriftreliquien enthalten sind, werden durch ein mehrdimensionales Maß (Alphabetenabdruck) zusammengefasst und

normalisiert, das wie der Fingerabdruck für eine Gruppe von Glyphen der untersuchten Schriftvarianten charakteristisch ist. Vergleicht man die Alphabetenabdrucke mit der Clusteranalyse, kann man feststellen, ob die Schriftvarianten Ähnlichkeiten aufweisen.

Schlüsselwörter: Cluster-Analyse, Computer-Paläographie, Computer-Schriftmodellierung, Graphem-Analyse, Identifikation der Inschriften.

Abstracts

Writing systems – Phonetic structures

Abstract: The beginnings of the Latin letter writing of the Hungarian language show one of the contact points between these two languages. When a writing system is created, a certain kind of contrastivity can be observed. During the development of the Latin letter phonetic transcription of the Hungarian language, the problem of excess phonemes had to be solved both for vowels and consonants. Initially, these problems created one letter versus many sounds, and many letters versus one sound relationships. The study analyses two literary records of the Rovash Script, one from Bodrog-Alsóbű (Somogy shire), and another one from Vargyas (Transylvania), also two literary remains of the Greek language: De administrando imperio by Constantine Porphyrogenitus, and the Deed of Gift for the Nuns of Veszprém Valley, furthermore the Arabic-Persian sources, and the Deed of Foundation of Tihany in order to study various attempts to solve the designation problem of the following sounds: [e:], [β], [γ] and [dʒⁱ].

Keywords: excess phonemes, linguistics, phonetic transcription, Rovash script.

Genealogical analysis of graphemes

Abstract: The research results described in this article deal with the exploration of the genealogical relationships of graphemes. Two graphemes are relatives if the voices they designate are the same, or their differences are justified by proper linguistic or palaeographic reasons. Furthermore, graphemes being in genealogical relationship are always similar to each other in a certain extent. This similarity may be complete coincidence, or a certain kind of modification. A variation of a grapheme is typically a result of topological transformation. Based on the development of writings, the possible topological transformations for the development of graphemes were determined. The author applied the genealogical chain construction algorithm to clarify the origin of various symbols of already deciphered writing monuments. Aggregating the genealogical relationships created for individual graphemes made possible to model the development of the writing systems that consist of those individual graphemes.

Keywords: computational palaeography, grapheme, machine learning, modelling.

The Andronovo Bronze Age civilization and the formation of the Rovash script

Abstract: This presentation wishes to draw the attention of professionals to the Andronovo culture, and the relationship between that civilization and the Rovash script. After a short chronology, a brief summary follows on the evolution of writing, image writing, syllabic writing and letter writing, then a special feature of the Bronze Age is discussed, namely the restricted availability of tin as a raw material. Then the relative frequency of vowels in Indo-European and Ugric languages is discussed, followed by an examination of the Orkhon inscriptions. The Etruscan alphabet and the Old Turkic language are also discussed using the results of acoustic measurements. Finally, the Rovash script of the age before Bilge Khagan is discussed, along with the comparison of the Rovash scripts and the Runic scripts.

Keywords: Andronovo culture, Old Turkic script, palaeography, Rovash script.

Problems of the early date of the Golden Treasure of Nagyszentmiklós

Abstract: The article wants to draw the attention to facts about the Golden Treasure of Nagyszentmiklós, which are not or only marginally addressed in scientific research. The author suggests that the treasure might be of early steppe Scythian-Hun origin, introduces relevant theories of that aspect, and the available ancient artefacts, without losing sight of local historical examination of the finding place of the treasure. The author describes those empires that had power centres nearby, discusses the origin of inscriptions of the treasure, and the relationship of steppe peoples to Christianity.

Keywords: Golden Treasure of Nagyszentmiklós, Scythians, Huns, Rovash script.

Testimony of the Romanian language: Additional aspects to the history of the Migration Period in Central Europe

Abstract: It seems to be possible to have a new method for taking a position in the age-old debate over the Huns' linguistic affiliation. Researchers so far have not paid attention to the following fact: the place and time of the last chapter in the history of the Huns covers the place and time of the first chapter of Romanian history. The ancestors of the Romanians with their newly developed Latin dialect dwelled in the same territory, where the remains of the Hun Empire's population fled, according to researchers of the Hun history. All peoples who dwelled in the Balkans since the time of Herodotus, had influenced the Romanian language: Thracian, Latin, Greek, Slavic, Hungarian, Albanian, etc. It's waiting to be explained by scientists that, no Turkic loanwords can be dated before the 12th century, while two peoples, the Huns and the Avars, both are assumed to belong to the Turkic-speaking tribes, had dwelled in the neighbourhood of the Romanians' ancestors for centuries, or even mingled with them.

Keywords: history, Moesia, Hunnish, Hunnish language, Common Turkic, Romanian.

Root Rovas scripting

Abstract: The article presents a visual system, which helps to understand the root-based architecture of the Hungarian language, makes clear the basic building blocks of the language, and displays their hierarchy. In the Székely-Hungarian Rovas, the root-based rovas orthography built on the compression rules applied to rovas ligature – the root rovas – is a solution to offer an accurate and systematic presentation of the root system of the Hungarian language. The technical and linguistic preconditions of creating root rovas are discussed. The author gives examples of the suitability of root rovas scripting for the optimal phonetic and semantic representation of the Hungarian language.

Keywords: computational script modelling, Székely-Hungarian Rovas.

Influence of the Latin script on the Székely-Hungarian Rovas

Abstract: The new letters, which are modifications of the Székely-Hungarian Rovas graphemes, show a parallelism with the transformation phenomena of the Hungarian alphabet of Latin origin. One can consider the effect of the phonography of Hungarian by Latin letters on the Székely-Hungarian Rovas script as a process of convergence, which was historically unavoidable. Their convergence is another evidence of their co-existence through Hungarian history. The convergence process is still not completed, as the two writing systems in the computer age reach their synchronous state only after the codification of the Székely-Hungarian Rovas. For this reason, the historic tasks of defining the rules of transliteration, and creating the standard glyphs of graphemes for Unicode cannot be deferred any more. There is only one of the living European writing systems, namely the Székely-Hungarian Rovas font, which is not standardized yet. Some seemingly marginal issues for the everyday user, such as the lack of rovas graphemes for x, y, q, w, fundamentally prevent the accomplishment of database operations, and the accurate and unequivocal correspondence among corpora, which might separate the Rovas script from the modern application era, and degrade it to a local phenomenon of restricted use.

Keywords: font design, rovas palaeography.

Methods of texts transcription to Rovash

Abstract: This article summarizes the pros and cons of the methods for transliterating Hungarian Latin letter-based texts to Rovash script. The rule-based transliteration approach and its current applications are shown in more detail.

Keywords: Rovash script, transcription, script presentation.

Decomposition of graphemes to canonical components

Abstract: This article introduces a new procedure for exploring the style related relationships of script relics representing different versions of a script. In doing so, the glyphs of an investigated script were divided into two-dimensional sub-formations, so-called canonical forms. Examples of canonical shape include a circular loop, an oblique section, a vertical section, or an intersection. The coefficients in the canonical resolution of the graphical forms included in the examined script relics are summed up and normalized by a multidimensional measure called an alphabet printout, which, like the fingerprint, is characteristic of a set of glyphs of the examined script variant. Comparing the alphabet printouts with cluster analysis, findings can be made about the similarities of the script variants represented by the script relics.

Keywords: computational palaeography, computational script modelling, cluster analysis, grapheme analysis, script identification.

Életrajzok

Banai Miklós

Sz.: 1953, Kisújszállás. 1977. ELTE TTK fizikus. A fizikai tudományok egyetemi doktora 1981. Munkahelyek: KFKI, ELTE TTK Elméleti Fizika Tanszék, MultiRáció Kft. Kutatási területek: komplex rendszerek matematikai modellezése, információs technológia. Releváns publikációk: Banai M., Lukács B.: Közgazdasági példák egzaktul megoldható variációs problémákra, Matematikai Lapok, 34. évf., 4. sz., 1991; M. Banai, B. Lukács: Attempts at Closing Up By Long Range Regulators. Technological Lag and Intellectual Background: Problems of Transition in East Central Europe, ed. J. Kovács, Dartmouth, Aldershot, 1995. Könyv: Banai Miklós, Lukács Béla: A Kárpát-medence egysége, Helikon Kiadó, Bp. 2010.

Demeczky Jenő

Sz.: 1949. febr. 24., Budapest. Diploma: 1972. BME Villamosmérnöki Kar híradástechnika szak. 2005. ELTE BTK általános és alkalmazott nyelvészet. Mh.: 1972–1976. Pannónia Filmstúdió; 1977–1986. MMG Automatika Művek; 1987–1997 Omikron Kft.; 1997- IBM Magyarország Kft. Szakterület: terminológia a műszaki fordításban.

Hosszú Gábor

Sz.: 1963, Budapest. Okl. villamosmérnök (BME), okl. jogász (PPKE), a műszaki tudomány kandidátusa, egyetemi docens, BME Elektronikus Eszközök Tanszék. Kutatási területek: internetes kommunikáció, digitális rendszerek modellezése, számítógépes paleográfia. Fontosabb publikációk: Internetes médiakommunikáció (2001), Current Multicast Technology (2005), Az internetes kommunikáció informatikai alapjai (2005), A székely jog megjelenése egy rovásleírásban (2010), Heritage of Scribes. The Relation of Rovas Scripts to Eurasian Writing Systems (2012), VHDL alapú rendszertervezés (2012), Rovásatlasz (2013).

Hölbling Tamás

Sz.: 1976, Pécs. PTE BTK történész (2001). Kutatási területek: korai magyar történelem, népvándorlás kora. Fontosabb publikációk: A honfoglalás forráskritikája I. A magyar kútfők (2010), A honfoglalás forráskritikája II. A külföldi kútfők (2010).

Kliha Gergely

Sz.: 1984. ápr. 18., Budapest. Programtervező matematikus (MSc), Eötvös Loránd Tudományegyetem, Budapest, rovásátíró szoftver-fejlesztő. Szakterület: programozás, rovás számítógépes megjelenítése. Firefox, Chrome bővítmény teljes weboldalak rovásra való átíráshoz, (2010-től, <http://rovasmag.hu/>). Munkahely: Insiron Informatika Kft.

Lukács Béla

Sz.: 1947, Budapest. 1970. ELTE TTK, fizikus. A fizikai tudományok akadémiai doktora 1988. Munkahely: MTA KFKI & jogutódai, nyugdíjas. Kutatási területe: Elméleti fizika, csillagászat, társadalomtörténet. Releváns publikációk: B. Lukács & L. Végső: The Chronology of the „Sumerian King List”. *Altorient. Forsch.* **2**, 25 (1975); Banai M., Lukács B.: Közgazdasági példák egzaktul megoldható variációs problémákra, Matematikai Lapok, 34. évf., 4. sz., 1991; M. Banai, B. Lukács:

Attempts at Closing Up By Long Range Regulators. Technological Lag and Intellectual Background: Problems of Transition in East Central Europe, ed. J. Kovács, Dartmouth, Aldershot, 1995. Könyv: Banai Miklós, Lukács Béla: A Kárpát-medence egysége, Helikon Kiadó, Bp. 2010.

Obrusánszky Borbála

Sz.: 1972, Ózd. Történész-keletkutató, néprajzos. PhD filológia (1999), Ulánbátor, Mongol Állami Egyetem. Honosítva: Debreceni Egyetem (2006). Journal of Eurasian Studies, főszerkesztő-helyettes. Kutatási területek: Belső-Ázsia története, a magyar őstörténet keleti vonatkozásai, keleti kereszténység. Legfontosabb publikációk: Keleti kereszténység (2012) Hunok a Selyemúton (2008), A mongol népek története (2005), A vérszerződés keleti párhuzamai (2004).

Raymond Pardede

Sz.: 1976. április 28., Jakarta. Okl. mérnök informatikus (BME, 2002), doktorandusz, BME Elektronikus Eszközök Tanszék. Kutatási területek: távközlés, informatikai rendszerek, internetes kommunikáció, többesadás, számítógépes paleográfia. Fontosabb publikációk: Thematic-Based Group Communication (2008), Jelentésazonosító eljárás a 16–18. századi székely-magyar rováslelkek értelmezésére (2010), Glyph Identification Based on Topological Analysis (2012), Grafémaanalízis kanonikus összetevők alapján (2012).

Rumi Tamás

Sz.: 1971. Okl. építészmérnök (1998, BME Építészmérnöki Kar), Master of Business Administration (2001, BME). Reklám és marketing területen vállalkozó, Rovás Infó szakmai portál főszerkesztő. Kutatási területek: rovás története, rovás tipográfia, rovás jelenkori alkalmazási lehetőségei. Fontosabb publikációk: Rovás alapismeretek (2010); Rovás kártya 1 (2010), Élő rovás – Nemzeti írásunk az egységesítés útján (2008; 2010); A rovás szabványosítása (2013); A magyar nyelvtörténet és a rovás paleográfia találkozása (2013).

Sípos László

Sz.: 1970. Mérnök-MBA (Budapesti Műszaki Egyetem, Tokyoi Waseda Egyetem és a Hawaii Állami Egyetem). Rováskutató, rovásos tartalomfejlesztő, a Rovás Alapítvány elnöke, a Rovás Infó szerkesztőségének tagja, munkáját a nemzetközi project management területén – több iparágban és a kultúrában – végzi. A rovásműveltséggel gyermekkorától foglalkozik, funkcionális rováshasználó. Jelenleg a rovásalkalmazások fejlesztési területein aktív – különös tekintettel a korszerű rováshasználatra, tartalomfejlesztésre és technológiai fejlesztésekre. Egyik kutatás-fejlesztési területe a gyökrovás. Fejlesztések: Kárpát-medencei egységes székely-magyar rovás helynévtábla rendszer; Latin-rovás átíró és szövegszerkesztő rendszer. Fontosabb publikációk: Rovás alapismeretek (2010); Rovás kártya 1 (2010), Élő rovás – Nemzeti írásunk az egységesítés útján (2008; 2010); A rovás szabványosítása (2013); A magyar nyelvtörténet és a rovás paleográfia találkozása (2013).

Tóth Loránd Lehel

Sz.: 1983. Okl. villamosmérnök (BME), doktorjelölt, BME Elektronikus Eszközök Tanszék. Kutatási területek: digitális rendszerek tervezése, számítógépes paleográfia. Fontosabb publikációk: Novel Applications of the Peer-to-Peer Communication Methodology (2009), Jelentésazonosító eljárás a 16–18. századi székely-magyar rováslelkek értelmezésére (2011), Pattern Identification for Computerized Paleography (2012), Grafémaanalízis kanonikus összetevők alapján (2012), Novel Algorithmic Approach to Deciphering Rovas Inscriptions (2013), Jelentésazonosító algoritmus

rovásfeliratokhoz (2013).

Zelliger Erzsébet

Sz.: 1945, Somorja. ELTE BTK magyar nyelv és irodalom szakos tanár, magyar nyelvtörténet speciális képzés 1969. Bölcsészdoktor 1975, CSc 1990. Mh.: 1969–2010 ELTE BTK Magyar Nyelvtörténeti, Dialektológiai, Szociolingvisztikai Tanszék: egyetemi docens; 1993–1999 PPKE BTK Magyar Nyelvi Tanszék; 2005–2006 Universität Wien Finnisch-Ugrisches Seminar; 2010 óta nyugdíjban. Kutatási területek: magyar nyelvtörténet, dialektológia, (történeti) szociolingvisztika. Fontosabb publikációk: A kupuszinai nyelvjárás igeragozási rendszere (1977), A magyar igetövek története a XVI. század közepéig (1990), A Tihanyi Alapítólevél (2005, 2011), Az első magyar nyelvű népének és művelődéstörténeti háttere. A Zsigmond-kori Húsvéti Népének keletkezésének körülményei, nyelvi kérdései (2006), A magyar nyelv használata a felső-ausztriai magyar diaszpóra körében (2012).