

ISO/IEC JTC1/SC2/WG2 N4274

2012-05-25

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации
ህጻህጻህጻህ ስለህጻህጻህጻህ ስለህጻህጻህጻህ

Doc Type: Working Group Document

Title: Contemporary Rovas in the word processing

Source: Tamás Somfai

Status: Individual Contribution

Action: For consideration by UTC and ISO/IEC JTC1/SC2/WG2

Irattípus: Munkacsoport dokumentum

Cím: Korszerű Rovás a szövegszerkesztésben

Forrás: Somfai Tamás

Státusz: Egyéni előterjesztés

Tárgy: Az UTC és ISO/IEC JTC1/SC2/WG2 figyelmébe

This document contains the comments of Tamás Somfai (Applied Computer Science Engineer) on the recent issues of the word processing of Szekely-Hungarian Rovas texts. Please, send any response to the author (tomas.lung@gmail.com).

Ez az irat Somfai Tamás (alkalmazott számítástechnikai mérnök) megjegyzéseit tartalmazza a székely-magyar rovás szövegszerkesztésének jelenlegi kérdéseivel kapcsolatban. Válaszok a tomas.lung@gmail.com címre küldhetők

Summary

Based on the practical experience in the development and use of the – until now – sole Hungarian Latin-based to Szekely-Hungarian Rovas transliteration and word processing system, I collected the practical requirements of the contemporary Szekely-Hungarian Rovas orthography that must be reflected in the encoding process.

Összefoglaló

A mai napig egyedülálló latin-betűs magyar írásról székely-magyar rovásra átfordító és rovásos szövegszerkesztő rendszer fejlesztésének és használatának tapasztalataira alapozva összeállítottam a korszerű székely-magyar rovás gyakorlati használatának követelményeit, melyeket a kódolási folyamatnak teljesíteni kell.

About the author

The author holds BSc in Applied Computer Science, and graduated as well at the Naganuma School (Japan) Open University (UK). The author has rich multinational-multicultural working experiences (IT, engineering, content development, governmental service) and language skills (Hungarian, English, Japanese, German) - recently working as application project engineer at leading automotive company in Hungary.

The achievements of the author in linguistic IT development are ranging from KM-based dictionary development (acknowledged by the scholarship of The Naganuma School, Japan) to multilingual character DB processing innovations and the Hungarian word-stem system DB development.

The author is the IT leader of the comprehensive Rovas transliteration and word processing system development that made possible the Rovas book publishing started in 2009. As participant of the Rovas IT workshop, Rovas Typography Workshop and active member of the Rovas user community, the author fluently reads and writes Székely-Hungarian Rovas.

Contribution text

1. Transliteration issues

Due to the traditions of the Latin-based old Hungarian orthography (used from the 10th century to the 19th century), the phonetic accuracy of the recent Latin-based orthography is lower than that of the Rovas orthography. This may cause contradictions in the transliteration methods of different Rovas users. Therefore, the integrative aspect in the cross-transliteration of large text corpuses - from Latin-to-Rovas, and from Rovas-to-Latin - do requires the possibility of the "one-to-one" clear connection, ensuring the seamless reverse transliteration to the original form of the texts.

A szerző

A szerző alkalmazott számítástechnikai mérnök, illetve a Naganuma School (Japán), és az Open University (Anglia) végzett hallgatója. A szerző gazdag nemzetközi-multikulturális munkatapasztalattal (IT, mérnöki tervezés, tartalomfejlesztés, önkormányzati szolgálat) és nyelvismerettel (magyar, angol, japán, német) rendelkezik. Jelenleg alkalmazásfejlesztő mérnökként dolgozik Magyarország egyik vezető autóiipari vállalatánál.

A szerző eredményei a nyelvészeti informatikában a TM-alapú szótárfejlesztéstől (a Naganuma School – Japán - ösztöndíjával elismerve) a többnyelvű karakter-adatbáziskezelési újításokig és a magyar szógyök-rendszer adatbázisának fejlesztéséig terjednek.

A szerző informatikai vezetője annak a teljes rovás-átíró és szövegkezelő rendszer fejlesztésnek, mely megteremtette a rovásos könyvkiadást lehetőségét 2009-ben. A rovás IT-műhely és a tipográfiai műhely tagjaként, illetve a rovóközösség aktív tagjaként a szerző folyékonyan ír és olvas székely-magyar rovással.

Az előterjesztés szövege

1. Átírási (transzliterációs) kérdések

A régi latin-betűs magyar írás (X. sz –XIX. sz) hagyományainak következtében a jelenkori latin-betűs írás fonetikai megjelenítési képessége a rovásnál szerényebb, ami ellentmondásokat okozhat a különböző rováshasználók átírási módszerei között. Emiatt a nagyobb mennyiségű szövegek – latin betűről rovásra és rovásról latin betűre - átírásának egybefoglaló szemlélete megköveteli a tisztán egy-az-egybeni megfeleltetés lehetőségét, biztosítva az információvesztéstől mentes visszaírást az eredeti szövegre.

1.1. The contemporary Hungarian Latin-based orthography as well may contain non-phonemic letters of 1 character or even of 2–3 combined characters from Hungarian Latin-based orthographies of older times, or in loanwords from foreign languages.

- combined: CZ, OE, SS, TH, TS, FFY, SCH, etc.
- single: X, Y, Q, W

Many of these letter combinations following Old Hungarian Latin-based orthography are typically used in traditional (family) names having additional meanings, representing identity or belonging to a community. For example:

- SCH: German-origin names
- SZKY, CZKY: Polish-origin names
- Y, FFY endings: nobility marking

Responding to the needs of the users, these combinations may need to be kept letter by letter in the Rovas orthography as well. Examples:

1.1. A korszerű latin-betűs magyar írás is tartalmaz olyan nem fonéma betűket, melyek 1, vagy akár 2–3 jelből is állnak. Ezek a betűk a korábbi idők magyar helyesírásaiból származnak, vagy idegen nyelvekből átvett szavakban fordulnak elő.

- összetett betűk: CZ, OE, SS, TH, TS, FFY, SCH, stb.
- egytagú betűk: X, Y, Q, W

Ezek a betűkapcsolatok többnyire régi magyar latin-betűs helyesírás követnek, jellegzetesen a (család)nevekben, ahol ennek többletjelentése van, önazonosságot, vagy közösséghez való tartozást mutat. Például:

- SCH: német eredetű nevek
- SZKY, CZKY: lengyel eredetű nevek
- Y, FFY végződés: nemesi származás

A felhasználói igényeknek megfelelően, ezeknek a kombinációknak a betű szerinti megtartása szükséges lehet a rovásos írásmódban is. Példák:

Ybl vs. Íbl

Xantus Xavér vs. Kszantus Kszavér

* *Note, that this issue is not a new need of the users. The usage of these characters have an old history. Examples:*

- X, Y, Q since 1629 (see Bonyhai Moga Mihály's alphabet)
- SCH since 1655 (in Portsalmi's handwritten book)
- Q W (also used by boy scouts from 1933)

** *These characters are not obligatory to use but the possibility of their application has to be provided with individual code points.*

* *Megjegyzés: Ez az igény nem új keletű, e jelek használatának hosszú története van. Például:*

- X, Y, Q 1629 óta (Bonyhai Moga Mihály ABC-je)
- SCH 1655 óta (Portsalmi kéziratos könyve)
- Q W (1933 óta a cserkészek is használják)

** *E jelek a használata nem kötelező, de alkalmazásuk lehetőségének biztosítása céljából önálló kódpontokat igényelnek.*

1.2. The usage of the Szekely-Hungarian Rovas characters **DZ** and **DZS** is a practical standard in the contemporary Rovas texts. Furthermore, the Hungarian linguistics has a very clear statement about these letters - that can be found even in the Hungarian Grammar textbooks as early as in the elementary school: DZ and DZS represent Hungarian phonemes. In the Latin-based Hungarian orthography the **DZ** represents a *voiced alveolar affricate* and the **DZS** represents a *voiced postalveolar affricate*.

Therefore, the principally phonetic Szekely-Hungarian Rovas orthography do use these Rovas characters, thus they have to be encoded individually.

As the **D+Z** and **D+ZS** letters are not equal to digraph **DZ** and trigraph **DZS** respectively, omitting the individual Szekely-Hungarian Rovas **DZ** and **DZS** characters will cause inconsistency of the clear phonetic interpretation of the words written with Szekely-Hungarian Rovas. Examples:

DZS vs D+ZS and DZ vs D+Z:

- lúdzsír (IPA: lu:d ʒi:r) dzs=d+zs vs. lándzsa (IPA: la:ndʒa) dzs=dzs
- kádzománc (IPA: ka:d zoma:nc) dz=d+z vs. bodza (IPA: bodza) dz=dz

1.2. A **DZ** és **DZS** nevű rovásjelek használata a korszerű székely-magyar rovás szövegekben gyakorlati szabvány. Sőt, ezen két betű tekintetében a magyar nyelvészet álláspontja is teljesen világos – mely egyébként már az általános iskolai magyar nyelvtankönyvekben is megtalálható: a **DZ** és **DZS** magyar fonémát jelenít meg. A latin-betűs magyar írásban a **DZ** zöngés *dentalveoláris spiráns* hangot, a **DZS** zöngés *alveoláris spiráns* hangot jelöl.

Ezért az alapvetően hangjelölő székely-magyar rovás használja ezeket hangjelölő betűket, így önálló kódolásuk is szükséges.

Mivel a **D+Z** és a **D+ZS** betűsorok nem azonosak a kettősbetű **DZ**-vel és a hármásbetű **DZS**-vel, a székely-magyar rovás önálló **DZ** és **DZS** jeleinek elhagyása következetlenséget okoz a székely magyar rovással írott szavak hangértékének világos értelmezésében.

DZS kontra D+ZS és DZ kontra D+Z:

- lúdzsír (IPA: lu:d ʒi:r) dzs=d+zs kontra lándzsa (IPA: la:ndʒa) dzs=dzs
- kádzománc (IPA: ka:d zoma:nc) dz=d+z kontra. bodza (IPA: bodza) dz=dz



* Note: The earliest known **DZ**, **DZS** Rovas characters appeared in Verpeléti's alphabet in 1935.

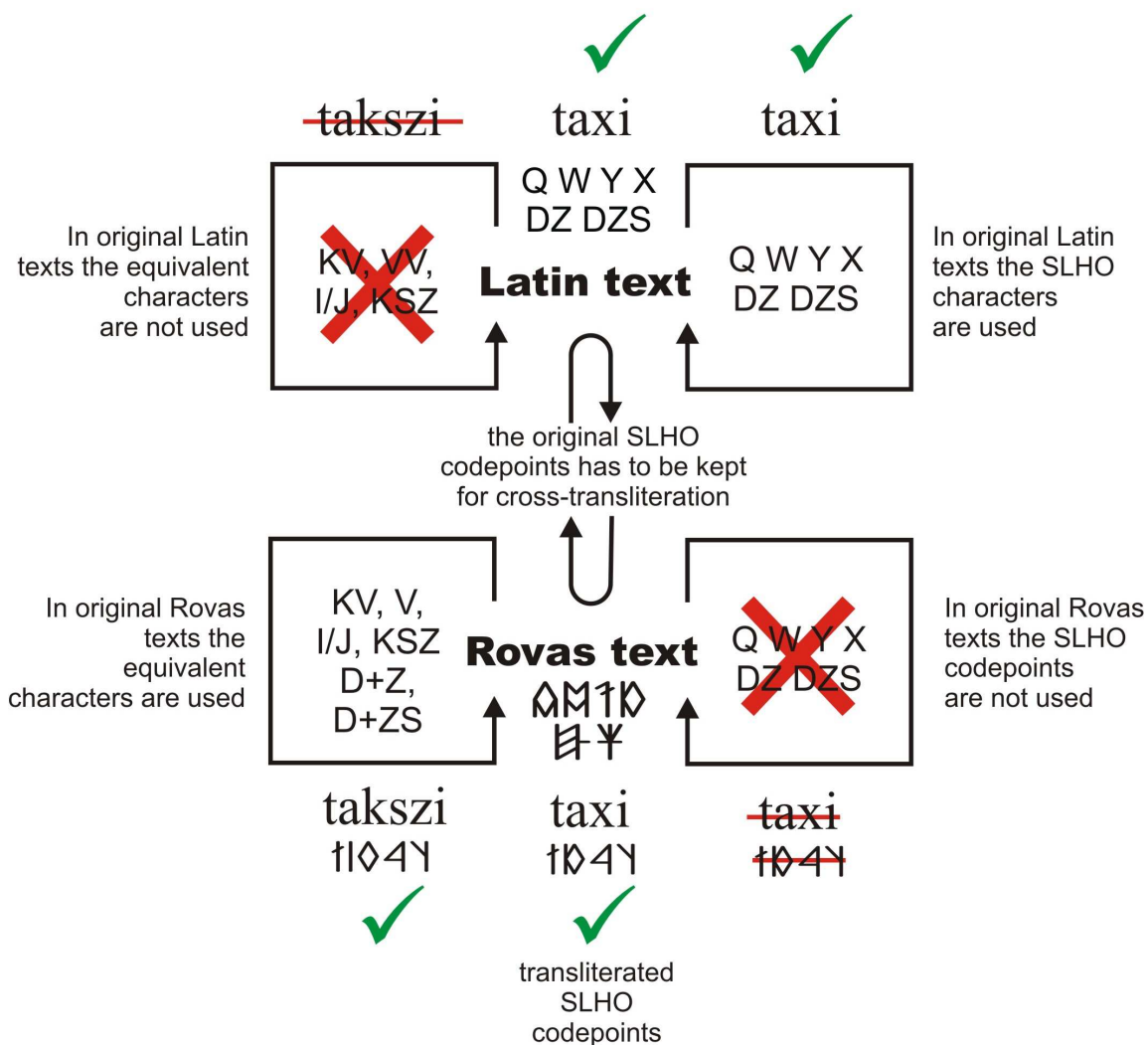
Megjegyzés: A legkorábbi ismert **DZ** és **DZS** rovás jelek Verpeléti ABC-jében jelentek meg, 1935-ben.

1.3. In the contemporary digital Rovas applications, the use of "rare" or historical characters are requested by the Rovas content developers. Therefore, all these characters - as listed in the Rovas proposal of the Hungarian Standards Institution - do need individual code points. Few examples:

1.3. A korszerű digitális rovásalkalmazásokban az ún. történelmi és „ritka” jelekre is igényt tartanak a tartalomfejlesztők. Ezért ezek a karakterek is – amint a Magyar Szabványügyi Testület javaslata is felsorolja – egyéni kódpontokat igényelnek. Néhány példa:

- **CLOSED E** is a close-mid front unrounded vowel. This letter is practically used in the word processing development for the Szeged-dialect. *Note: this phoneme is widely used in the Hungarian dialects, except the North-East (literal), Mezőség and Northern Csángó dialects.*
- **OPEN V** represents an extinct consonant of the the Hungarian language - namely the voiced bilabial fricative *v*, which was in use up to the 14th century.
- **CLOSE OE UE, CLOSE OEE UEE.** The character CLOSE OE UE can represent a short close-mid front rounded vowel or a short close front rounded vowel. Similarly, the CLOSE OEE UEE may represent a long close-mid front rounded vowel or a long close front rounded vowel - depending on the actual custom of the Rovas user.
- **ZÁRT E:** középső nyelvállású palatális illabiális magánhangzó. Ennek betűje már gyakorlatilag is használatban van a szegedi tájszólás szövegszerkesztőjének fejlesztésében. *Megjegyzés: ez a fonéma a magyar tájszólások többségében jelen van, kivéve az észak-keleti (irodalmi), a mezőszegi, és az észak-csángó tájszólást.*
- **NYÍLT V** a magyar nyelv egyik, mára már kihalt mássalhangzója – zöngés bilabiális spiráns – melyet a XIV. századig használtak.
- **Zárt Ö Ü, Zárt Ó Ú:** A ZÁRT Ö Ü jelölhet akár középső, akár felső nyelvállású rövid palatális labiális magánhangzót. Hasonlóan a ZÁRT Ó Ú jelölhet akár középső, akár felső nyelvállású hosszú palatális labiális magánhangzót – a felhasználó szokásaitól függően.

Usage of Special Letters of the Hungarian Orthography (SLHO) in contemporary Latin and Rovas texts



2. Contemporary Rovas in the modern electronic media

Recently, there are growing needs in the digital media for Rovas-written digital contents, ranging from words, titles or shorter texts to larger texts (letters, articles, books). Now, the vast majority of the Szekely-Hungarian Rovas texts are transliterated from Latin-based texts. Besides the static transliteration (printed materials) the demand for real-time transliteration is rapidly emerging. To answer the wide user needs, the full contemporary character set has to be encoded individually.

2. Korszerű rovás a modern elektronikus médiában

Jelenleg a digitális médiában növekedő igény van a rovással készült tartalmak iránt a szavaktól és címeiktől kezdve a rövidebb és hosszabb szövegekig (levelek, cikkek, könyvek). Pillanatnyilag a székely-magyar rovásos szövegek túlnyomó többsége a latin-betűs szövegről kerül átírásra. A statikus átírások (nyomtatott anyagok) mellett a valós idejű átírásra való szükséglet megjelent. Ezen széleskörű felhasználói igények kielégítéséhez a teljes korszerű betűkészlet egyéni kódolására van szükség.

3. Implementation of Rovas text processing in software products

As the majority of the contemporary Szekely-Hungarian Rovas alphanumeric digital content is created by transliteration, the data synchronization is very important for content creators and owners. To provide consistency and unambiguous use - for example in searchable databases or internet search engines - the possibility to use the full Rovas character set (including DZ, DZS, Q, X, Y W) is inevitable. Therefore, individual code points for all the contemporary Rovas characters are needed.

The latest book where Rovas and Latin based alphanumeric database was used: Postal codes of Inner-Hungary (ISBN 978-963-88952-3-3). An example from the book:

- DZ as an individual consonant (dz): Meggyesbodzás

3. Rovásos szövegfeldolgozás megvalósítása a szoftverekben

Mivel a korszerű alfanumerikus székely-magyar rovásos tartalmak többsége átírással keletkezett, az adatok összehangolása nagyon fontos a tartalmak fejlesztőinek és tulajdonosainak. A következetesség és egyértelműség biztosítása végett – kereshető adatbázisok, internetes keresés – a teljes rovás jelkészlet használatának lehetősége (beleértve a DZ, DZS, Q, X, Y W betűket) elengedhetetlen. Ezért szükséges önálló kódpont a korszerű rovás jelkészlet minden betűjének számára.

A legújabb könyv, amelybe rovás és latin-betűs adatbázis került: Postai irányítószámok Belső-Magyarország (ISBN 978-963-88952-3-3). Egy példa a könyvből:

- DZ mint egy önálló mássalhangzó (dz): Meggyesbodzás

H3H33 9176 Mecser
 A9H33X3333 5663 Medgyesbodzás
 4H33333333 5666 Medgyesegyháza

Usage of the traditional **non-phoneme or composite letters** of contemporary Latin-based Hungarian Orthography in Rovas transcription:

- X, W, Y in street names: Textilgyári u., Wlassics Gyula u., Xántus u., Ybl Miklós tér

Hagyományos **nem-fonéma és összetett betűk** használata a korszerű latin-betűs és rovásos átírásban:

- X, W, Y utcanevekben: Textilgyári u., Wlassics Gyula u., Xántus u., Ybl Miklós tér



Irányító szám	Város, község, lakott hely neve
9324	Bogyoszló
4114	Bojt
8741	Bókaháza
2855	Bokod
3066	Bokor
3016	Boldog
7937	Boldogasszonyfa
3884	Boldogkőújfalú
3885	Boldogkővára
3794	Boldva
7517	Bolhás
7586	Bolhó
7754	Bóly
8992	Boncodfőle
7150	Bonyhád

1182	Wlassics Gyula u.
1148	Xántus u.
1013	Ybl Miklós tér
1033	Zab u.
1141	Zablja u.
1101	Zách u.
1186	Zádor u.
1155	Zafir u.
1182	Zágon u.
1107	Zágrábi köz
1107	Zágrábi u.
1224	Zagyva u.
1031	Záhony u.
1112	Zajzon u.
1222	Zakariás József u.
1152	Zala György u.
1192	Zalaegerszegi u.

4. Standardization

The goal of the encoding shall be to provide the technical possibility of using the Rovas character set in the information technology for the widest range of users. I carefully compared the various proposals of the different contributors based on the following practical aspects:

- digital reproduction of historical Rovas findings
- digital use of the contemporary Rovas scripts
- cross-transliteration between the Latin-based and Rovas orthographies
- consistency and reproduction of historical proper names
- Rovas-specific DB processing, wordprocessing and spelling solutions

Bakonyi's approach, e.g.:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3566.pdf>

It is based on the outdated state of the paleography and Rovas usage in the beginning of the 20th century. No information about the technical issues of the contemporary Rovas usage.

Everson-Szelp' submissions (practically identical to the approach of the WG2 ad-hoc meeting in Helsinki in 2011 e.g.:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4110.pdf>

It is a static approach based on arbitrarily selected Rovas findings. The proposal disregards the contemporary Rovas usage and the needs of the vast majority of the Rovas user community of today. Thus, it does not contain all the necessary Rovas characters, hindering the seamless digital utilization. Furthermore, it uses the ambiguous script name "Old Hungarian" and erroneous character names.

4. Szabványosítás

A szabványosítás fő célja rovás jelhasználat technikai feltételeinek biztosítása a legszélesebb értelemben vett felhasználói kör számára. Az alábbi nézőpontok figyelembevételével összehasonlítottam a különböző javaslattevők előterjesztéseit:

- történeti rovásemlékek digitális újraelőállítás
- a korszerű rovás szövegek digitális használata
- átírás a latin-betűs írás és a rovás írásbeliség között
- következetesség és alakhűség a történelmi tulajdonnevek átírásában
- rovás-specifikus adatbázis-kezelés, szövegszerkesztés és helyesírás-ellenőrzés

Bakonyi-féle megközelítés

pl.: <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3566.pdf>

A XX. század elejének idejétmúlt rováshasználatán és írástörténeti ismeretein alapul. Nincs benne információ a technikai kérdésekkel, illetve a korszerű rováshasználattal kapcsolatban.

Everson-Szelp beadványokkal (gyakorlatilag azonos, A WG2 ad-hoc megbeszéléssel – Helsinki 2011 pl.:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4110.pdf>

Statikus megközelítés, önkényesen választott rovásemlékek alapján. A beadvány nem vesz tudomást a korszerű rováshasználatról és a mai rovótársadalom többségének igényeiről. Így a javaslat nem tartalmazza az összes szükséges rovásbetűt, megakadályozva a hibamentes digitális használatot. Ráadásul félreérthető írásnevet és hibás betűneveket tartalmaz.

Prof. Róna-Tas' approach:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4232.pdf>

Remarkable conclusion is that a "general acceptance of a modern variant" is needed and "a final solution can be found with the help of the Hungarian Academy of Sciences."

Szondi's approach:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4267.pdf>

Pamphlet of the local lifestyle association "Természetesen (Naturally)" with no traceable professional credit in the encoding process, nor in the Rovas script research. Without any relevant technical details, the sole point of the contribution is the preclusion of 4 practically used characters on ideological base (Q, X, Y, and W as "foreign" letter), and 2 official phonemes (DZ, DZS) due to lack of elementary knowledge of Hungarian grammar.

Approach of the Irish National Body:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4196.pdf>

It is a contribution without clear intent and relevant technical details. Its sole point is to provide a basic Rovas font, with typographically primitive and erroneous character glyphs, with missing characters, and with erroneous script and character names.

Approach of the Hungarian Standard Institution

(Dr. Hosszú), e.g.:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4183.pdf>

This approach fully conforms with the current user needs and fulfills the requirement of the contemporary digital Rovas usage (DB- and word processing, cross-transliteration, consistency, etc.).

Consequently, I support the proposal of the Hungarian Standards Body and its legitimate role of coordination between the groups of the Szekely-Hungarian Rovas user community in the encoding process.

This contribution is officially sent to the Hungarian Standards Body as well.

Prof. Róna-Tas megközelítése:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4232.pdf>

Figyelemre méltó következtetés a „korszerű (rovás)változat elfogadásának” szüksége, illetve hogy „a végső megoldás megtalálása a Magyar Tudományos Akadémia segítségével lehetséges”.

Szondi megközelítése

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4267.pdf>

Egy helyi életmód egyesület (Természetesen) pamfletje, a szabványosítás és a rováskutatás nyomon követhető szakmai referenciája nélkül. Értékelhető technikai javaslat híján az egyetlen mondanivalója 4 gyakorlatban használt rovásjel kizárásának igénye ideológiai alapon (Q, X, Y, W mint „idegen” betűk), illetve 2 hivatalos magyar fonéma (DZ, DZS) kizárása a magyar nyelvtan alapvető ismereteinek hiánya következtében.

Az Ír Nemzeti Testület megközelítése:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4196.pdf>

Világos cél és vonatkozó technikai részletek nélküli előterjesztés. A javaslat egyetlen eleme a rovás alap-betűcsalád, mely tipográfiaiailag kezdetleges és hibás betűalakokból áll, hiányzó karakterekkel, hibás íráselnevezéssel és betűnevekkel.

A Magyar Szabványügyi Testület (Dr. Hosszú)

megközelítése, pl.:

<http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4183.pdf>

Ez a megközelítés teljesen kielégíti a jelenlegi rováshasználati igényeket és teljesíti a digitális rováshasználatához (adatbázis-kezelés és szövegszerkesztés, átírástechnológia, következetesség, etc.) támasztott feltételeket.

Következésképpen én a Magyar Szabványügyi Testület javaslatát és a szabványosítást érintően a székely-magyar rovás felhasználói csoportok közötti egyeztetésben annak legitimitását támogatom.

Ezt az előterjesztést hivatalosan megküldöm a Magyar Szabványügyi Testület számára is.