

DR. MARCEL COURTHIADE
 ODELENJE RROMSKIH STUDIJA – INALCO (PARIS)
 KOMESAR ZA JEZIK I JEZIČNA PRAVA
 MEĐUNARODNA RROMSKA UNIJA



Od "glave vola" (naime od slova alef – א) do upravljačkog programa "EuroUniv"
 [poslao je doprinos, jer nije bio prisutan zbog zdravstvenih razloga]

Povjest pismenosti raznih jezika dobro je dokumentirana kod naroda koji su ostavljali pisani trag, ali vrlo malo znamo o jezicima za koje pisani trag ne postoji. Upravo oni predstavljaju veliku većinu drevnih govora. Gotovo sve ove jezike i kulture u određenoj mjeri poznajemo samo kroz prizmu drugih naroda, dakle onih koji su o njim pisali.

Naravno, svoj doprinos dala je i arheologija, nudeći relevantne informacije, ali one nisu toliko izrazite kao pisani tekst. Bez obzira, upitna može biti relevantnost pisanih tekstova, koji nerijetko mogu prikazivati više laži nego istine, posebice kada govore o drugom narodu.

Stav o pisanju u narodu, nekad i danas, nije bio i nadalje nije isti. Danas imamo, prije svega židovsko-kršćansko-muslimanski svijet, u kojem se pisanje smatra običnim, poželjnim i štovanim djelom. Opće je poznato kako su se razvijale fenička i hebrejska abeceda, na osnovi kasnijeg egipatskog demotskog pisma, te kako se tijekom vremena izrodila latinica i ćirilica u Europi. Manje je poznata činjenica da je drugi razvoj iste osnove stvarao slovo indijskih jezika, kroz aramsku abecedu.

Aramski jezik bio je jezik praperzijskih dvorova, te je na taj način došao na svjero-zapadni dio indijskog subkontinenta. Kasnije se, kroz oblik brahmi, širio cijelom Indijom. Prvi pisani dokumenti bili su natpisi cara Devanāpija Piadasi, poznatijeg kao Ašoka (carevina 269 – 232 g. prije Krista), a koji je urezao svoje «dekrete» širom Indije, na stijenama i kolonama u lokalnim varijantama govornih jezika, kako bi svi ljudi znali koje su najvažnije osnove njegovog zakona.

U sjevero-zapadnoj Indiji (danas Pakistan) natpisi su bili urezani aramskom i grčkom abecedom (pošto je i grčki bio službeni jezik), te dalje na jug i istok u brahmi abecedi. To je dakle, prvo svjedočanstvo pisanih tekstova u Indiji.

Međutim, kako se radi o pažljivo formiranim slovima na velikom teritoriju, sa međusobno velikim udaljenostima, možemo zaključiti, da ovo nije bio prvi pokušaj pisanja, već da je postojala izvjesna tradicija pisanja. Drugu informaciju nalazimo u traktatu «Arthašastra» dvorskog savjetnika i teoretičara politike Čanakje (poznatog i kao Kautilja – 370-283 g. prije Krista). Govori o politici, piše da «arhive imaju izrazitu vrijednost za vladara». To bi značilo da su Indijci pisali i pohranjivali u arhive, ali su one vjerojatno bile nesigurne, na nestabilnim držačima, primjerice na velikim osušenim listovima, koje su kukci vrlo brzo i jednostavno uništili. Stoga su, po nekoliko puta, prepisivali sadržaje iz arhiva, dok je njihov sadržaj imao značaj za društvo i vladara. Kasnije su prestali prepisivati starije arhive, te prednost dali

novima. Možemo zato vjerovati zašto pisana riječ nikada nije posjedovala opsjenu u Indiji. Ovaj tehnički faktor Indijci su maštovito izrazili, objašnjavajući kako pismenost pripada djelu nerealnog svijeta, u suprotnosti sa izgovorenim i slušanim riječi, koja sama ulazi u sustav značenja. Ova koncepcija izražava suprotan položaj s jedne strane brahmana, čija je riječ «prava, istinita i vjerna» kao i Veda, koje oni znaju napamet, a s druge strane pisara, kojeg sanskrska književnost predstavlja kao lažljivca i lopova. Drevni Indijci pisanje su smatrali nevrijednom i tužnom neophodnošću, prozaičnim instrumentom činovnika i trgovaca, bez ikakve plemenitosti, nasuprot ugledu, koji je pisanje imalo kod semitskih naroda. Pravi intelektualac nije bio onaj koji čita i piše, već onaj koji napamet zna tisuće i tisuće stihova sa njihovim komentarima. Pisanje su uspoređivali sa štakama koje služe slabom i hromom umu. Zato se može razumijeti nedostatak umjetnosti kaligrafije kod Indijaca, dok se ona u punom cvatu razvijala kod drugih civilizacija, koje su koristile trajnije držače pisane riječi (papirus, terra-cota, zidove hrama – hramovi su se kasno pojavili u indijskoj civilizaciji, i dr.)

Neki drugi narodi imali su još radikalniji odnos prema pisanju, primjerice Gali, koji su se plašili pisma, smatrajući da se u njemu krije opasna magija s obzirom da živ čovjek može izreći ono što je mrtav čovjek napisao. Vidjeli su to u nadgrobnim natpisima naroda Italije, Rimljana i drugih, što im se činilo nesigurno s obzirom da svaki čovjek može ukrasti nešto napisano i pobjeći sa ovom tajnom koju je netko drugi povjerio.

Riječ «pisati» u indo-europskim jezicima

Sama riječ za pojam "pisati" pokazuje, da je on došao kasno u indo-europske kulture: slavenski korijen u glagolu "pisati" potječe od staroslovenskog < **пѣсати, пишѡ**, naime "ukrasiti", što je još vidljivo u poljskoj riječi *pisanka* "jajca Uskrasa" i ima srodstvo sa sanskritskim पित्ति [piṣṭi] "ukrasiti", sa litvanskim *piešiù* "crtati uglom" i sa latinskim *pingere* "ukrasiti, vesti" (odakle *pictum* i *pictura*, *pictogram* itd.). Sa druge strane je korijen latinskog glagola *scribere* "pisati" vezan sa pojmom "grepsti" sa varijantama *gr*, *kr*, *skr*, *wr* itd. – sve od praindoeuropskog **gerbh-*, koji daju osim *grebem* grčko γράφω "pišem", γράμμα "slovo", γραμμή "linija" (pa *grafologija*, *gramatika* itd.), staro-germ. *krattan* "grepsti", germ. *kratzen* "ogrepsti", *kerben* "zarezati", *Krebs* "rak", engl. *scratch* "ogrepsti" itd... Njemački je preuzeo sa latinskog oblik *scribere* da bi stvorio *schreiben* "pisati" (kroz *scriban* i *schriben*), dok se staro-engl. *gr+i+t~/wr+i+t~* razvijao u *writanan* "ogrepsti, pisati" i dalje u široko poznati oblik *to write* "pisati".

Ovdje nije mjesto za opisivanje različitih riječi koje izražavaju pojam "čitati", ali treba napomenuti, da i sistem dotičnog rječnika također karakteriziraju neologija i improvizacija (lat. *legere* "vezati", eng. *to read* < *ridan* "savjetovati, pričati", gr. διαβάζω "staviti kroz", alb. *këndon*, *knon* "pjevati", mađ. *olvasni* "brojiti" itd.).

Kad su govornici rromskog jezika pristigli u Evropu, kao što smo rekli, bez posebnog štovanja prema simboličnoj vrijednosti čina pisanja, adaptirali su se, kao uvijek, lokalnim uvjetima i preuzeli su često riječ dominantnog jezika kraja gdje su živjeli, kao što su na primjer u Hrvatskoj stvorili **pišinel** "pisati" od *pisati*. Ceokupna zbirka lokalnih riječi za pojmove "pisanje" i "čitanje" može biti predmet osobnog člana i ovdje bi će dovoljno da se pripomne riječi zajedničkog rromskog jezika: **lekhel** ili **lekhavel** "piše" i **drabavel** ili **drabarel** "čita".

Ipak, pod utjecajem drugih naroda, Rromi su često usvojili lokalnu riječ i stvorili oblike kao *skril*, *iskril* (rum. *a scrie*), *irinel* (mađ. *írni*), **pišinel**, **pišil** (sl. *pisati*), **škroinèla** (alb. *shkruan*), **ğrafinèla** (gr. γράφω), **jazinèla** (tk. *yazmak*), **činel** "rezati" etc... odnosno **čitinel**, **čitil** (sl. *čitati*, rum. *a citi*), **ginel** (bukv. "brojiti", cf. *olvasni* v. gore), **điavanèla** (gr. διαβάζω), **ledzonèla** (alb. *lexon*), **gilabel** (bukv. "pjeva", cf. alb. *këndon*, *knon* "1. pjevati; 2. čitati" itd... Međutim bilo bi greška smatrati da je **drabavel** neologizam, čim ovu riječ nalazimo već sa značenjem "čitati" u rromskoj gramatici XVIII. stoljeća iz Švedske.

Kao što je rečeno u tekstu, do danas, svi narodi čak i unutar Europe, nemaju isti odnos prema pisanju, bez obzira na njihovu kulturnu povjest. Općenito govoreći, na zapadu postoji stroži i

precizniji odnos prema uporabi pravopisnih pravila nego na jugo-istoku Europe. U starim tekstovima primjećujemo da nije uvijek bilo tako. Primjerice do XVI. ili XVII. stoljeća postoji više pisanja jedne te iste riječi u razmaku od nekoliko redova. Nesigurnost dakle predstavlja arhaični aspekt pisanja. Njemačko-skandinavska rigoroznost pravopisa širila se u susjedne slavenske zemlje na istok i dalje u Rusiju već u XVIII. stoljeću, te kasnije na cijeli Sovjetski Savez. Do Mađarske (uključujući Transilvaniju) i Hrvatske proširio se preko austrijske tradicije. Ipak, Balkan je ostao izvan ovog utjecaja i nepismenost je tamo vladala trajno, tako da se narodi (osim malobrojne elite) nisu navikli, da marljivo pišu prema određenim pravilima. Nije u pitanju drevnost pismenosti, nego širina narodne osnove ove pismenosti, jer na primjer grci, koji imaju pravadnu književnost, ipak često malo mare za ispravnost pisanja – istina je da imaju težak sistem, gde se isti zvuk mora pisati jednim od nekoliko načina, ovisno od etimologije, koju često ljudi više ne znaju. Uglavnom možemo reći da, dok je na neki način određena sramota vezana sa pravopisnim greškama na zapadu i sjeveru Evrope, ne nalazimo takvu emocionalnu reakciju na jugu Balkana i većina ljudi bez ikakve kočnice piše «kao što govori» – ili barem tako misli.

Inače rječeno, u svim zemljama tradicionalna elita piše marljivo ali procent ove elite zavisi od povjesnih uvjeta dane zemlje. U zemljama, gdje je sistem obrazovanja, često silom, bio djelotvoran, brojnija je elita – primjerice u Bugarskoj pod njemačkim a posle sovjetskim uticajem, dok tamo gdje se manje poštovalo školovanje, ljudi nisu tako osjetljivi na pravopis. Naravno, u sredinama, gdje uopće nije postojao odgoj na maternjem jeziku, najmanje imamo izraženu brigu oko pravopisa, kao kod Rroma, Arumuna, Pomaka i drugih manjina izvan matične zemlje, jer im nije dostupno obrazovanje na materinjem jeziku (Makedonci u Grčkoj, Rumunji u južnom dijelu Srbije itd.). Najčešće ove grupe nemaju ni novine, ni ostale tiskovine ili knjige na materinjem jeziku, pa nemaju odakle crpsti emulaciju za pravilno pisanje.

Četiri osnovna elementa svijesti o ispravnom pisanju

U stavu svakog naroda prema pisanju općenito i posebice o ispravnom pisanju, možemo razlikovati četiri osnovna elementa:

1. Unutarnja tradicija: hebrejski jezik na primjer ima jaku tradiciju pisanja, bilo gdje u Europi, što se može zapaziti već od ranog odgoja djece, pa do njihovog običaja da se piše i hebrejska varijanta lokalnog jezika hebrejskim pismom – što daje židovsko-njemački ili jidiš, židovsko-španski ili džudezmo, židovsko-arapski itd...

2. Područje boravka nekog naroda : do Židova je slična situacija sjevernih jezika Europe, gdje se školsko obrazovanje razvilo rano, intenzivno i široko. Čak i narodi bez države, kao što su Sami u Skandinaviji, redovito čitaju šta se o njim piše, vode rasprave preko novine i stvorili su već, nedugo nakon drugog svjetskog rata, abecedu vlastitog jezika. Na žalost, stvorili su faktički 13 različitih dijalektalnih abeceda, što je dovelo do raspada pisanog jezika. Danas gotovo da i nije više u uporabi, ne zbog nedostatka abecede ili interesa, već zato što su previše male grupe koristile svaku abecedu umjesto da imaju zajedničko pismo. Taj sistem, dakle, nije bio niti praktičan niti ekonomski održiv. Sa druge strane imamo narode, kojim ne drže toliko do pisanja, kao na Balkanu, u Turskoj i dr.... Naravno ovo je slučaj većine Rroma – ali ne svih, i ne samo slučaj Rroma, kako bismo mogli vjerovati sudeći po zapisima raznih "stručnjaka".

3. Pristunost ili odsutnost školovanja u maternjem jeziku, dostupnost ili nedostupnost novina i ostalog pisanog materijala na materinjem jeziku, itd...

4. Jednostavnost ili kompliciranost pisemnog sistema, u odnosu na govorni jezik : jasno je da logičan usklađen sistem pisanja, kao hrvatski ili španjolski ima više uvjeta za masovno pravilno korištenje nego sistem tipa grčkog, francuskog, čak i bugarskog (sami bugari često miješaju "e" i "я" zbog šopskog karaktera govora prijestolnice Sofije) ili rumunjskog jezika.

Ako pogledamo situaciju nekoliko jezika u svjetlu ovih elemenata, možemo zapaziti specifičan položaj romskog jezika:

	unutr. tradicija	Područje boravka	školovanje	sistem pisanja	total +
hrvatski	++	++	++	++++	10
romski	0	0/+	0	++++	4/5
grčki	++	0	++	0	4
albanski	0	0	++	+++	5
bugarski	++	0	++++	+++	9
makedonski	+	0	++	++++	7
rumunski	+	0	++	+++	6
mađarski	+++	++	+++	+++	11
poljski	+++	+++	++++	+++	8
francuski	+++	++	++++	0	8
engleski	++	++	++++	0	7
samski	0	+++	+	0	4
španski	++	+	++++	++++	11

Broj pluseva izražava naklonost datog elementa prema poštivanju pravilnog pravopisa : 0 ima loš utjecaj, ++++ ima vrlo pozitivan utjecaj. Total izražava vjerojatnost da budu pisane velike količine tekstova pravilno, to znači da preživi jezik ispravno napisan.

Prijelaz na mehaničko pisanje (pisaće mašine, računala, internet)

Nisu svi jezici istovjetno postupili u odnosu sa mehaničko pisanje. Na primjer, za vrijeme konferencije o pisanju albanskog jezika u Monastiru u g. 1912, sudionici su izabrali kao slova sa dijakritičkim znakovima samo ona, koja su već postojala u slaganju tiskanica i na pisačim mašinama, odnosno *ç* i *ë* – iako stalno ponavljanje slova *ë* u ovom jeziku čini pisanje vrlo zamornim svaki put, prvo *ç* a posle *e*. Naprotiv Rumunji nisu imali ovu dilemu i od samog početka su stvorili iz olova slova *ă*, *ș* i *ț*, dodatno do već postojećih *â* i *î*. Tako i Turci nisu vidjeli problem, da stvore za vrijeme prelaza s arapske abecede na latinski, nova slova *ı*, *ğ* i *ş*. Danas, sistem računala dozvoljava, da se bilo koji jezik može pisati odgovarajućom abecedom (čak i samski u 13 raznih abeced!) na računalu ili internetu. Nije uvijek bilo tako : u početku demokratizacije informatike, većina jezika nije imala svojih dijakritičkih znakova na raspolaganju, pa su tada Jugoslaveni sve pisali bez dijakritičkih znakova, to znači na isti način *celo* i *čelo*, *sis* i *šiša*, ili *zelje* i *želje*... Mađari, koji nisu imali u računalu *ő* i *ű* zamijenili su ih pomoću tilda nad samoglasnikom : *õ* i *ű*. Kasnije su ipak producenti informatike uveli sva slova, uključujući mađarska *ő* i *ű*, sve suglasnike tipa *š*, *č*, *ž*, *š*, *ć*, *ž* ili *ž* itd.... Ipak mnogo ljudi je i dalje pisalo bez dijakritičkih znakova, iako su oni bili dostupni, čak lako dostupni. Kad se forum na internetu otvara, primijetimo da veliki broj cibernauta piše iz područja bivše Jugoslavije bez znakova.

Problem je još gori sa jezicima, koji imaju drugu abecedu, kao grčki, makedonski ili bugarski. Iako postoji cjelovit sistem elektronskog pisanja grčkim slovima, još uvijek veliki broj Grka piše aproksimativnom latinicom, koju zovu *φραγκοχιώτικα*, pa faktički postoje uglavno dvije vrste ovog pisma, "fonetski" i tada se tekst čita više manje kao što bi zvučio na grčkom, ili "pandanski" i tada za svako grčko slovo imamo latinski pandan ili broj sličnog oblika (8 za *θ*, 3 za *ξ*), kao na primjer:

grčki	φραγκοχιώτικα		grčki	φραγκοχιώτικα	
	fonetski	pandanski		fonetski	pandanski
α	a	a	ξ	ks, x	3
ο	o	o	θ	th	8, 9
ω	o	w	β	v	b
υ [i]	i	y, u	σ, ζ	s	c, s
ου	u	oy, ou	υ [v, f]	v, f	y
αι	e	ai	δ	d	d

η	i	h		μπ	b	mp
---	---	---	--	----	---	----

Primjeri¹ sa riječima :

grčki	φραγκοχιώτικα			grčki	φραγκοχιώτικα	
	fonetski	pandanski			fonetski	pandanski
ξέρουν ²	kserun	3eroun		φωνή	foni	fwnh
θέλω	thelo	8elw		συμφωνώ	simfono	sumfwnw
μπαίνει	beni	mpainei		μπορεί	bori	mporei
αξίζει	aksizi, axizi	a3izei		ευχαρίστουμε	efxaristume	eyxaristoyme
συνέργω	sinergo	cynergo		εντύπωση	entiposi	entipwsi
στην ζωή	stin zoi	cthn zwh		χωρίς	xoris	xwric

Tako da ista rečenica može zavisno od pisara izgledati kao slijedeće:

grčki ερωτεύτηκε αυτή και μας ξεφούρνισε ωραία τραγούδια με την γλυκή της φωνή
fonetski eroteftike afti ke mas ksefurnise oreá tragudia me tingliki tis foni
pandanski erwteythke ayth kai mas 3efoynise wraia tragoydia me thn glyki ths fvnh
prijevod zaljubila se i poslužila nam lijepe pjesme svojim slatkim glasom

Prilična je razlika među dvjema transkripcijama, ali faktički ih većina cibernauta miješa u velikoj meri. Naprimjer česta transkripcija *entipwsi* nije niti fonetska (treba bi bilo *entiposi*), niti pandanska (treba bi bilo tada *entypvsh*), već mješavina obje. Opća je tendencija da se sve više piše pandanski ali nerijetko nalazimo transkripcije fonetskog tipa, prije svega kad su kratke.

Isto važi za bugarski i makedonski jezik, ali bez tolikog kolebanja kao na grčkom.

Primjeri sa bugarskim :

improlatinica	bugarski	prijevod
vsi4ko vurvi/varvi/vyrvi	всичко върви	sve u redu
ako jelaesh, bašta mi	ако желаеш, баща ми	ako želiš, oče moj
naI-grozniq ³ 4ovek v bulgariq	най грозния[т] човек в България	najružniji čovjek u Bugarskoj
priqteli, zabavlqvaite se	приятели, забавляйте се	prijатели, zabavljajte se
trqbva da izgledja 4ovek	трябва да изглежда човек	treba izgledati čovjek
ne me obarkwaite we h0raa	не ме обърквайте в хора	nemojte me mješati
slan4eviyat	слънчевият човек	sunčani čovjek
sladak grqh	сладък грях	Slatki grijeh
napyln0 otka4en i smeбен	напълно откачен и смешен	potpuno otkaćen i smiješan

Vidimo da su se pojavile konvencije za dopunjenje slova koja nedostaju: *a, e, i, o, b, v, s* itd. imaju latinsko slovo, dok su naprimjer za *ъ, ѝ, я, ч, ш* odnosno *щ* bugari izabrali *a* (ili *u, y*), *I* (ili *y*), *q* (ili *ya*), *4* (ili *ch*), *6* (ili *sh, ch*) odnosno *6t* (ili *sht*)... Nesigurnost zapažamo prije svega pri transkripciji tvrdog znaka *ъ*, koji se u improlatinici piše *a, u* ili *y*: *bilgariq, bulgariq, balgariq* (България – obično bez majuskule). I slovo *ю* predstavlja problem pa zato se riječ *кучек* možemo naći u slijedećim pisanjima : *kuchek, kiuchek, ku4ek, kiu4ek, kio4ek, cucek, kucek* itd. (mn. *qki kiu4eci* «jaci čučeci»).

U slučaju makedonskog jezika, ima manje nesigurnosti, jer je praksa obočnih alfabeti u srpsko-hrvatskom jeziku za vrijeme Jugoslavije dovela do lakog prelaza od jednog do drugog i opuštenje znakova ne smatraju makedonski cibernauti za veliku grešku – pa zato nalazimo *bidejki, prodolzete, pesnichka, ke najdam*. Rijetko nalazimo bugarske brojke u mjesto slova, ali ipak postoje : *de4ki zakon* «dječji zakon»... Nalazimo i dijalektalni izgovor, kao na primjer

¹ Svi su primjeri preuzeli od komentara pod snimanjima Youtube'a.

² Prijevod : znaju, hoću, ulazi, vredi, surađujem, u životu; desna kolona : glas, slažem se, može, radujemo se, utisak, bez.

³ Pravilno bi trebalo dopisati -т на kraju.

u poruci : *ne se znaí koj prv ke pukne od ubaite momi* (prav. *ne se znae koj prv ke pukne od ybavute momu*). Rečenica *Q sakam da te galam* predstavlja istovremeno srpsku zamjenicu «ja» i bugarsko pisanje ovog zvuka uz pomoć slova *Q* (za *Я*) pa nalazimo često kod cibernauta i zamjenicu *as* «ja», koja nije ni makedonska (pravilno : *jac*) ni bugarska (*az*).

Dosta iznenadno, albanski podnosi isti tretman, iako taj jezik ima latinsku abecedu i samo dva posebna znaka, kao što smo već gore uveli, naime: *ë* i *ç*. Možemo u @-forumima i u komentarima youtube'a čitati poruke tipa :

youtube *mandi ost sun larg pi muharemi numer 1 ost muharemi per kong dashnive*
 albanski *Mandi është shumë larg prej Muharemi. Numër 1 është Muharemi për këngë dashnive* (knjiž. *dashurie*)

prijevod *Mandi je vrlo daleko od Muharema. Muharem je broj 1 za ljubavne pjesme.*

Taj nemar nije vidljiv samo kod polupismenih cibernauta (u ovom slučaju kod Albanca, koji je bio školovan van albanske škole, vjerojatno u inozemstvu, ili sasvim nepismen), već je možemo naći i na nekim sasvim «oficijalnim sajtovima» Albanije ili Rumunjske. Većina stranica rumunjskih ministarstava, pa čak i dugo vrijeme predsjednika republike, nisu koristili slova *t, ș* ili *ă*, nego su dugo pisali *t, s* ili *a...* Od 2005. g., otprilike su ove stranice pravilno, čak i ponekad precizno i točno napisane. Nisu u pitanju ovdje Rromi iz mahale nego činovnici visokog ranga, koji tako malo brinu za vlastiti jezik, da su ga oni dugo vremena osakatili.

Osim problema dijakritičkih znakova, koji je uglavnom riješen za internet stranice, iako nije za većinu foruma, imamo u južnim oblastima Srbije problem brkanja zvukova *č* i *dž* sa jedne strane i *ć* odnosno *đ* sa druge strane, jer tamo ljudi u izgovoru ne razlikuju «reči» i «reći»⁴, i dakle pišu na primjer **govoreći* umjesto *govoreći* (isto kao Albanci na Kosovu ne razlikuju *ç* od *q* i *xh* od *gj* pa pišu primjerice *qmim* «cijena» umjesto pravilnog oblika *çmim* itd...). Zato je izuzetno važno i potrebno školovanje na vlastitom jeziku – ne možemo ozbiljno očekivati pisanje bez grešaka od ljudi, koji nikad nisu prošli kroz barem jedan tečaj o maternjem jeziku. Istina je da u nekim uređajima (SMS, outlook itd.) nije lako pisati datom abecedom i zato je odjel romistike u Bukureštu zamislilo pojednostavljen sistem pisanja za takve situacije (ali ne za rromski uopće) : *sx, c, ch, zx* umjesto *ś, ć, čh* i *ž*, pa onda *8* i *3* umjesto *θ* i *z*, pisanje bez infleksa (*a, e, i, o, u* umjesto *ă, ë, ĩ, ő, ũ*) ili sa «y» kad je neobhodno (*ya, ye, yi, yo, yu*) itd... Primjerice, tri sledeće rečenice, uzete slučajno u zbirci, predstavljaju sledeće frekvencije promene (između 1/6 i 1/21):

Normalno pisanje	Pojednostavljeno pisanje za SMS
Gelem gelem, lungone dromençar, maladilem baxtale Rromençar. Aj Rroma!len katar tumen aven E caxrençar, bokhale çhavorrençar?	Gelem gelem, lungone dromencar, maladilem baxtale Rromencar. Aj Rroma!len katar tumen aven E caxrencar, bokhale chavorrencar? [5 promene za 107 slova = 1/21]
Dilino phenel so zanel – godăver zanel so phenel	Dilino phenel so 3anel – godyaver 3anel so phenel [3 prom. za 40 sl. = 1/13]
O maçho lel te khàndel e șeresθar	O macho lel te khandel e sxeres8ar [4 prom. za 27 sl. = 1/7]
O laçho manus thulöl panëshθar, o biužo manus nasvalöl avginesθar	O lacho manusx thulol panyes8ar, biuzo manus nasvalöl avgines8ar [9 prom. za 55 sl. = 1/6]

⁴ Poznata je Vukova formula « пиши као што говориш », ali manje je poznato, da ju je on faktički preuzeo od Save Mrkalja, koji je sam u 1810 g. citirao bezautorsku hrvatsku gramatiku iz 1779. g. «kak se izgovaraju réchi, tak se y pišzati imaju». Uprkos toga, Vuk nikad nije dopustio mješanje ч i һ niti u i һ, jer je dosta kritično napisao « простота покварено говори » i ljudi koji se na Vuka pouzdaju na osnovi prve formule (kao bajagi da se može kako god pisati jezik) moraju ipak biti poštjeni i uzimati u obzir i drugu vukovu formulu.

Nije neophodno označiti *ž* prema *z*, jer je *z* okolo 10 puta češće nego *ž* u tekstu. Obratno, *ć* je nekoliko puta češće nego *c* i kontekst u maltene svim slučajevima uklanja kolebanje – pogotovo kad se tiče slova *ćh*, jer *ch* uopće ne postoji. Za par *s* i *ś*, čim je manja razlika (*s* se pojavlja 3-4 puta češće nego *ś*), dobro je označiti razliku grupom *sx*. Što se tiče samoglasnicima sa infleksom, podsjećamo da niz dijalekata (prije svega superdijalekt bez mutacije, odnosno gurbetski, džambaski, čergarski itd...) samo retko imaju jotizaciju i pisanje na SMS i Outlook može slediti njihov primjer, dok znakova *c*, *3* i *8* dobro mogu zamjestiti *ç*, *з* i *θ* u kratkim porukama.

Ukratko, važno ja zapamtiti:

1. da se taj sistem odnosi samo na nedovoljne sisteme kao SMS ili Outlook
2. da se ovo ne odnosi na direktno pisanje govornih oblika rromskog, nego da je pojednostavljenje izvadeno iz već standardnog pisanog jezika :

lokalne varijante	→→	zajednički jezik, zajedničko pisanje	→→	pojednostavljeno pisanke za SMS ili Outlook
-------------------	----	--------------------------------------	----	---

Ukratko, mnogo toga ovisi o tradiciji, marljivosti ili nemarljivosti prema jezika, od poštovanja njegove vrijednosti, od osjećaja koliko je bitno pisati pravilno da bi svaki čitatelj lako, brzo i bez zabune razumio značenje teksta. Nije ovo dakle nikako rromska osobina, nego više balkanska – uključujući grčki usprkos prдавne književnosti na visoko ocjeenjenom jeziku, zbog zanemarivanja za vrijeme osmanske vladavine.

Čime danas raspolažemo kod računala ?

Faktički još prije nekoliko godina nije bilo tako lako pisati rromski na računalima. Postojalo je poluimprovizirano rješenje, da se ubace slova pomoću naredbe "Insert", pa posle "Symbols" i tada se pojavljuje na ekranu okvir podijeljen na kvadratiće, svaki sa jednim slovom unutra.

Za svako specifično rromsko slovo, naime (ć Ć ś Ś ź Ż з З θ ä Ä ë Ě ĭ Ĩ ő Ő ů Ű ç Ç), trebalo je tada po redu kliknuti na odgovarajući kvadrat, gde se pojavljivalo dotično slovo i poslije toga kliknuti na znak "Short cut". Na ekranu se pojavljivao tada drugi prozor sa naslovom "Customize keyboard". U njemu se registrirala izabrana prečica za dotično slovo pritiskom na taster prema slijedećem popisu:

za ć Alt istovremeno sa tasterom "c"
za Ć Alt i Shift istovremeno sa tasterom "c"
za ś Alt istovremeno sa tasterom "s"
za Ś Alt i Shift istovremeno sa tasterom "s"
za ź Alt istovremeno sa tasterom "z"
za Ź Alt i Shift istovremeno sa tasterom "z"
za з Alt istovremeno sa tasterom "y"
za З Alt i Shift istovremeno sa tasterom "y"
za θ Alt istovremeno sa tasterom "8"
za ä Alt istovremeno sa tasterom "a"
za Ä Alt i Shift istovremeno sa tasterom "a"
za ë Alt istovremeno sa tasterom "e"
za Ě Alt i Shift istovremeno sa tasterom "e"
za ĭ Alt istovremeno sa tasterom "i"
za Ĩ Alt i Shift istovremeno sa tasterom "i"
za ő Alt istovremeno sa tasterom "o"
za Ő Alt i Shift istovremeno sa tasterom "o"
za ů Alt istovremeno sa tasterom "u"
za Ű Alt i Shift istovremeno sa tasterom "u"

za ç Alt istovremeno sa tasterom "w"

za Ç Alt i Shift istovremeno sa tasterom "w"

Kada je već odabrana prečica dotičnog slova, klikne se na "Ascribe" i "Close". Ovo se ponavlja za svako specifično rromsko slovo (cela procedura traje za svako slovo nekoliko sekunda) i posle toga se svako od njih piše pritiskanjem osnovnog tastera istovremeno sa Alt (odnosno Alt i Shift za velika slova).

Taj sistem, koji još uvijek vrijedi, omogućio je pisanje rromskom abecedom, ali ipak je ukazao i na tri smetnje:

1. dugo nisu sva rromska slova bila dostupna u glavnim tipovima slova, kao Times New Roman. Ipak moguće je bilo pisati rromski u Arial Unicode MS, Lucida Sans Unicode ili naravno Microsoft Sans Serif. Sada je već od godine 2009 Microsoft uveo cjelovitu rromsku abecedu također u Times New Roman, što znači da je ova smetnja nestala.

2. druga prepreka je bila, činjenica da korisnik mora pripremiti svoje računalo gore opisanom operacijom. Nije ni ovo bila velika prepreka, jer je programacija ostala poslije toga u računalu, dok se ne izbriše.

3. najveći se problem veže sa činjenicom, da je bila ova programacija operativna samo u Wordu i zato ograničena *Microsoft Outlooku* (nije operativna na primjer sa *Outlook Expressom*, u formularima ili internet-tražilacama tipa *Google, Internet Explorer* itd.).

Najnovija razvojna kretanja rromske tehnologije

Evo stižemo do cilja ovoga referata: unatrag nekoliko tjedana, već postoji drajver (upravljački program) zvani EuroUniv, koji dozvoljava pisanje ne samo rromskom abecedom, već i abecedama svih europskih jezika, koji se služe latinicom (ima ih preko 40 što ukupno predstavlja oko 154 slova) i u latinskoj transkripciji jezika sjeverne Indije (još desetak drugih slova tipa ṭṭḍḍṛṛṇṇṁṁ). Sve tri gore navedene prepreke nisu više aktualne, jer se može pisati u fontovima Times, Arial, Lucida itd... naime u bilo kojem fontu *Unicode*: prethodna programacija više nije potrebna, jer se upravljački program može dobiti kao *download* tipa *patcha*, uvjesti u računalo i odmah je operativan – i to ne samo kao ranije u *Wordu* nego i u svim sistemima, uključujući na primjer i programe filmskih podnaslova.

Svatko će za sada dobiti po jedan CD, gdje je upravljački program graviran ali tehnički detalji još su tajna (pod zaštitnim brojem Nr. INPI 403.723-281.210). Ovaj upravljački program ima poseban značaj u današnjoj Europi, jer:

a. njime se može pisati pedesetak europskih jezika⁵ sa njihovim varijantama, uz pomoć samo jedne tastature, naime engleske, i jedne jedine "*linguistic and regional options*", bez promjene za svaki prijelaz s jezika na jezik; koncepcija upravljačkog programa je bila zamišljena na takav način, da slova i znakovi odgovaraju u 96% slučajeva onim što je gravirano na fizičkim tasterima;

b. u jednom tekstu može se pisati na više jezika, što je korisno u publikacijama o jezikoslovlju, posebno usporednom jezikoslovlju, o povjesti ili zemljopisu, jer se mora citirati veliki broj riječi i osobnih imena. Ipak taj upravljački program nije namijenjen za ljude, koji pišu "samo" u dva tri jezika bez čestog prijelaza od jednog na drugi, jer u njihovom slučaju ostaje bolji sistem prijelaza sa jedne jezične opcije na drugu uz promjene tastature (jer kad ostaju iste tipke, ovo ipak znatno smeta brznoj uporabi), nego baš za korisnike, kojima su

⁵ Naime albanski, arumunski/armînj, azerski, bretonski, češki, danski, estonski, euskerski (euskera batua), faeroenski, finski, francuski, galicijski, galski, hrvatski, irski, islanski, kašubski, katalanski, korsikanski, kurmanji (kurdski latinica), latinski (sa skandiranjem), letski, litvanski, luksemburski, lužički (obe varijante), mađarski, maltanski, nederlandski, njemački (sve varijante), norveški (sve tri varijante), okcitanški (sve varijante i grafije), poljski, portugalski, retoromanski, rromski (sve varijante), rumunski (i para-rumunski jezici), samski (odnosno laponski, sápmigiele – sve 13 varijante), sardinski, slovački, slovenački, španjolski, srpski (latinica), švedski, taljanski, turski (uklj. karaimski, tatarski & gagauski), transliteracija indo-arijanskih jezika i transliteracija pīnyīn'a.

potrebne brojne abecede u kratkom vremenu, kao jezikoslovci, povjesničari, zemljopisci i slični.

c. nedavno su počele skandinavske zemlje primjenjivati u praksi pravilan, doslovan pravopis, to znači doslovno pisanje svih osobnih imena (ljudi i njihovih gradova, sela ili mjesta rođenja, venčanja i smrti) u dokumentima osobnih podataka, u svrhu poštovanja identiteta svakoga ali također smanjenja rizika krivotvorenja dokumenata prilikom propuštanja ovog ili onog dijakritičnog znaka.

Principi našeg upravljačkog programa su sledeći :

1. apsolutan prioritet imaju slova rromskog jezika;
2. relativan prioritet imaju slova jezika sa najvećim rromskim stanovništvom, jer kad u svojoj zemlji rrom piše tekst, mora stalno uvoditi vlastita imena ili citate na većinskom jeziku, znači u prvom redu rumunjski (2.000.000 rroma) pa i drugi po redu;
3. slijedeći prioritet imaju u okviru svake abecede najčešće upotrebljavana slova, uzevši u obzir istovremeno slova, koja su u upotrebi u više jezika (kao *š, â, ç*) i tada imaju prednost u odnosu na ona slova, koja su ograničena na manji broj jezika (kao *ȷ, ʀ* ili *ć*).

Naravno, raspored slova na tipkama odgovora mnemotehničkim principima, da bi čovjek lako, brzo i dugotrajno zapamtio kombinaciju tipaka za pisanje datog slova. Zato je izuzetno važno da predstavimo ovdje, na Svjetski dan rromskog jezika u Zagrebu, jer imamo obavezu da otvorimo rromskom jeziku što više vrata, kao što su ciberneta, IT, tiskanje, filmski podnaslovi, I-pad, smartphone itd... Ne smijemo se zadovoljiti kukanjem oko sudbine jezika ili deklaracijama namjere, već je neophodno da ga uključujemo aktivno u sve dijelove sutrašnjeg života, ne samo zbog tehničkih razloga nego i u svrhu povratka jeziku priznanja i prestiža, koje nikad ne bi trebao gubiti. Zagreb je s tog gledišta važan grad s obzirom na inicijativu Udruge za promicanje obrazovanja Roma u Republici Hrvatskoj «Kali Sara» i također na prisutnost grupe inženjera, koji vode odličan projekat « rjecnik.net » i koji su suglasni da surađuju sa nama da bi rromski jezik bio uključen u sastav europskih jezika na internetu.

Da zaključimo, možemo još naglasiti, da je današnja situacija pisanja u IT slijedeća :

1. dugo nisu institucije, čak i glavna državna tijela, koristile sve mogućnosti pravilnog pisanja vlastitog jezika na internetu;
2. ipak, unazad 5-6 godina, one uglavno pišu tekstove kako treba;
3. pretežno prosječni cibernetai ipak ne koriste sve tehnološke mogućnosti i pišu bez znakova, mada ponekad sa brojčanim pandanom, kao bugari i grci;
4. takvo pisanje otežava znatno čitanje i dakle ograničava do najnižijeg nivoa razmjenu misli i osećaja, obično do nivoa divljenja bez nijansa ili psovki – što naravno ne zahtijeva bogati jezik niti mudru glavu. Ima se utisak, da cilj nije prikaz informacije ili osećaja nego samo ispražnjenje agresivnosti. Vidljivo je ovo također u engleskom jeziku, isto kao u onim jezicima, u kojim navodno znakovi predstavljaju poteškoće... Ja bih ovo – s oprostjenjem, ilustrirao slijedećim primjerom, koji sam baš jučer našao : « *macadonia is grrek beacause the greeks fukc the girls of macdonia.....so macedonias is bastard of the grEeks* » [nisam promijenio nijedno slovo]⁶. Engleska stilistika, pravopis riječi, dostojanstvo autora, njegova politična doličnost i ljudska umjerenost su više manje na istom (žalosnom) nivou. Možemo zaključiti, da takav tekst ne zaslužuje niti bolji pravopis, niti više pažnje, niti čak dug život. Autor se htio opustiti i nije mario da li će netko čitati.
5. divljačko pisanje ugrožava jezike, koje već imaju standardni oblik – ili barem ugrožavalo bi ih, kad ne bi imali potporu tradicije, akademije, obrazovanja u školi i van nje, bogatog tiskovnog materijala, medija itd. Dakle tvrdnje, da se može standard rromskog jezika stvoriti sam od sebe putem divljačkog pisanja na internetu (kad nemamo potporu tradicije, akademije,

⁶ Mogli bismo tisuće takvih primjera uvesti, što je nemoguće, ali i sledeći dosta je zanimljiv i ne zahtijeva valjda komentar : « *makedonija je vecna a bugari se makedonski zigani...*»

obrazovanja na rromskom, dovoljno tiskovnog materijala itd.) potiče ili od krajnje gluposti ili od sistematske strategije uništavanja našeg jezika.

6. najverojatnije će cibernauti dalje pisati još dugo psovke i druga priopćenja nižeg nivoa bez brižljivosti prema pravopisu, pa ne predstavlja ovo posebnu štetu.

7. kao što su institucije polako prihvatile da poštuju pravopis svog jezika, uvjeren sam da će sada, čim je i rromski dostupan uz pomoć upravljačkog programa EuroUniv, i naši cibernauti koristiti taj sistem, kad budu htjeli stvarno priopćiti nešto umno drugim Rromima. Potreban je samo način proširenja dostupnosti upravljačkog programa i njegovog uputstva.

DALJE DAJEMO ODLOMAK ZVANIČNOG OPISA UPRAVLJAČKOG PROGRAMA EUROUNIV

REQUIREMENTS FOR SUPPORT OF WRITTEN RROMANI LANGUAGE

(and of an all-European Latin script operative system)

IN DATA PROCESSING SYSTEMS

SO KAMLŎL PES KERDINO KAJ TE ŠAJ LEKHAVDŎL E ĆHIB RROMANI

(thaj vi sa e avera evroputne ćhibă latinikane patranenća)

AND-E ELEKTRONIKANE SISTÈME

This document states the minimum requirements for support of use of written Rromani in IT. In addition to Rromani character sets and keyboard layout requirements for all the forms of Rromani in use world-wide and for all kinds of platforms e. g. Microsoft "Window", Apple "Macintosh", Unix systems etc. it covers also practically all European language scripts using Latin letters. It is intended as information for industry, and as a basis for possible work in word processing standardization. It can also be used as a reference document in purchasing.

A first draft was sent out for comment in February 2010, and the document was eventually adopted, with changes resulting from the comments received, by the Board of the above mentioned Rromani structures.

Endorsing Entities:

International Rromani Union

Address: Warszawska 43 PL-15062 BIAŁYSTOK

International Rromani Federation

Address: Suite 464 Boardwalk, WILDWOOD New Jersey 08260

Fundación Instituto de Cultura Gitana (Ministerio de Cultura)

Address: C/ San Marcos, 39-1ºA SP-28004 MADRID

Ref. 01/350-S2.

Foreword : Purpose of the present document

This document is intended to serve as an authoritative source of information for the IT industry and also as a basis for work in formal written standardization, as regards the IT requirements of the Rromani language, as well as all other European languages written in Latin script, made available through typing on a single common key-board and a single "Regional and linguistic option" (namely U.K. English). It complements and explains the EuroUniv driver.

Further, it is intended for use as a reference document in purchasing situations. It has therefore been written in standard-type style and structure.

Problems of Rromani-language use in IT

Data processing of text in Rromani language has always been viewed as problematical, since they require a few letters that are generally not used in the majority languages of the countries

where Rromani population lives, i.e. practically all countries of Europe, Northern and Southern America as well as Australia and some areas of Asia (mainly Turkey). This includes countries where other alphabets than Latin are in use, mainly Serbia, Russia, Ukraine, Belarus, Bulgaria, Macedonia and Greece. Yet, Rromani is written in Latin script in these countries as well. For the time being the main companies in computer industry have not yet developed appropriate solutions to meet the need of writing in Rromani due to the fact that Rromani-language computer users are geographically very much scattered all over the world and they have addressed only lately these companies.

Some partial solutions, like sets of fonts, have been produced and circulated by local companies like Piard Ltd. in the early 90ies, for both "Windows" and "Macintosh" computer environments. These solutions have been made available as "freeware" by Universities and Rromani NGOs, and have been quite important in facilitating information interchange in written Rromani, including publication of many books. However, such solutions are perceived as temporary, since they do not cover all varieties of IT activity, especially since the development of web exchanges. Only when the main computer companies decide to include full support of the Rromani language in their product offerings will the problem become finally solved.

Moreover, the present proposal encompasses, in addition to Rromani script, writing support for all other European languages, provided they are written in Latin script. This is a major requirement, owing to the fact that everywhere the Rroms use daily various local languages (mainly for proper names and quotations) in addition to their Rromani mother tongue. Accordingly a suitable key-board must meet this requirement also, in addition to the use of the Rromani special letters. The necessity of such a keyboard encompassing through a single regional and linguistic option all European languages using a Latin script has also been expressed by European and national authorities in order to grant a more appropriate rendering of the names related to the civil status of the persons living in a given area.

Languages covered

Albanian, Arumanian/Armînj, Azeri, Bosnian SH, Breton, Catalan, Corsican, Croatian, Czech, Danish, Dutch, English, Esperanto, Estonian, Euskera (batua), Faeroese, Finnish, French, Galician, German (all varieties), Hungarian, Icelandic, Irish, Italian, Kaszubian, Kurmanji (lâtin), Latin (with scansion), Latvian, Letzбургisch, Lithuanian, Maltese, Norwegian (all three varieties), Occitan (all varieties and spellings), pînyîn and rômaji transcriptions (incl. revised Hepburn and kunrei-siki), Polish, Portuguese, Romantsch, Rumanian (& para-Rumanian languages), Rromani (all varieties), Same (or Sâpmigiele, all 13 varieties), Sardinian, Serbian (Latinica), Slovak, Slovenian, Sorbian (Upper & Lower), Spanish, Swedish, Transliteration of Indo-Aryan languages, Turkish (incl. Karaim, Tatar and Gagauz), Welsh and basic (school level) IPA transcription.

Apart from the world-wide dispersion of the Rromani-language market, a factor that has delayed industry development of Rromani language support is that technical documentation on the IT needs of the Rromani language, in a form that can be readily treated by Industry, has been lacking.

The determination of filling in this lack is the primary reason for the development of the present document.

Standardization activities

The present solution relies on some preliminary researches and discussions which have been carried out by LWP of Stockholm and presented to the CDSG expert meeting in Berlin (4 April 2004) as "A Possible CEN/CENELEC Standard For Rromani IT Requirements". At this stage only the Rromani letters were taken into consideration. Later on, after consultations in Stockholm and Uppsala (7-9 January 2007) with the majority worldwide of Rromani experts in Rromani linguistics participating in the Stockholm-Uppsala linguistic conference, it was presented under an

improved form to the second CDFG expert meeting in Luxemburg (16 March 2007) as “CEN Standard for Rromani Written Language IT Requirement”. The participation of INALCO to the European project "*Restore the European Dimension of Rromani language and culture*" (EACEA 2009-2163/001-001 CU7MULT7) since 1-st September 2009 made it crucial to produce for the project itself an experimental keyboard driver on the basis of the British keyboard, with appropriate character representation (“Coding”) allowing the use of existing and available Unicode fonts. It was then decided by the INALCO group to include into the driver not only Rromani but also the alphabets of all five languages in Latin script of the participating countries (French, Hungarian, Polish, Rumanian and Slovak) and later on those of all other European Latin script-based alphabets. The final result was achieved with success in summer 2010 with the production of an experimental executive driver meeting all the needs defined by the INALCO group. The driver and its description were deposited in Paris in Dec. 2010 under number Nr. INPI 403.723-281.210. It was concluded that a descriptive document needs anyway to be produced, for presentation the European standardization CEN, and for circulation among entities likely to use this keyboard executive driver, since ultimate responsibility for declaring the needs of such an all-European Latin script system and of its support in IT rests with such entities, among others the specialized Rromani bodies and users at European level (first of all those involved in the above mentioned European project), keeping definitely in mind the world-wide perspective.

SUPPORT FOR WRITTEN RROMANI LANGUAGE IN IT

1 Scope

The present document defines the minimum IT requirements applicable to the alphabets of the Rromani language written in the Latin script.

The present requirements relate to the present-day common use of the Rromani language as well as to letters used in the countries where Rroms live and need to write proper names and majority language quotations. In spite of initial fears, it came into view during the development of the project that it is indeed possible not only to integrate all European alphabets to the requirements described in the present document, but also to do so in such a mnemonic way which grants an easy access to all letters, without any specific training. Outside Europe, the situation is even less complex, since Rromani computer users are based in countries of English, Spanish and Portuguese majority language, where all local proper names have their equivalent in the spelling of one of these languages.

The present document is intended primarily for the IT industry, and also as a basis for possible future formal standardization. Its contents and general layout therefore follow common practices for IT standards.

The present document is also intended to be immediately usable as a base reference in purchasing. Since it however specifies only minimum requirements, references need generally be made also to formal standards and/or other material.

It shall be particularly noted that, while several alternatives for satisfying the IT support requirements of the Rromani language are specified in this document, it contains no recommendations for selection of alternatives in purchasing. The user of the document must decide the selection for each situation, considering what local and/or higher-level IT policies apply.

The requirements are applicable to three common computer environments, namely ISO/IEC 8859-compatible (e.g. Unix systems), Microsoft "Windows", and Apple "Macintosh", for both 8-bit and multi-octet coding schemes.

Acknowledgment: "Macintosh" is a trademark of Apple Computer, Inc. "Windows" is a trademark of Microsoft Corporation. These trademarks are registered in all countries where Rromani computer users live and are likely to use written Rromani.

Requirements on character sets, character coding, keyboard functional layouts and fonts are specified. The present document is intended to be useable in a wide range of purchasing situations, from acquisition of complete systems to individual system components, like e.g.

keyboards. Therefore, the applicability of the requirements varies with the situation, and must be considered if the document is referred to in purchasing.

The requirements relate to data processing computer systems. They are not intended to be generally applicable to communications systems (e.g. mobile telephony text messaging or cash dispensers).

2 Implementation levels

This document defines two levels of implementation:

- Basic level satisfying the requirements of the Rromani language alphabet as defined in the Warsaw decision of April 1990 and confirmed in the Stockholm-Uppsala declaration of January 2007 (see Annex C page 24); and
- Extended-alphabet level satisfying the requirements of Rromani texts as written in all particular local scripts and those of all local majority languages alphabets for proper names and quotations in the languages of the relevant countries.

This division into two levels is intended to facilitate implementations, and thereby to secure a wider industry support, for cases where only basic-level may be necessary. It shall not be interpreted as a "ranking" among varieties of the Rromani language, which are all of equal importance.

3 Requirements

3.1 Scope of requirements

In this section are stated the requirements that a computer system shall comply with to be in accordance with the present document.

It must be noticed, however, that the various components of a computer system may have differing degrees of language support. For instance, the operating system may be capable of handling a larger character set than can be input directly through the keyboard.

The requirements in the present document may consequently be inadequate to ensure sufficient total functionality for specific situations. Therefore, when referring to the present document in purchasing it will in general be necessary to also specify additional requirements.

3.2 Character set

The character repertoire to be handled by operating systems shall contain as a minimum:

a) The characters of registration no. 6 of the ISO "International register of coded character sets to be used with escape sequences", ISO-IR 6; and additionally

b) the specific Rromani letters àÀ äÄ éÉ çÇ èÈ ěĚ ìÌ įŎ ðÐ óÓ ůÛůŮ źŻ ʒŽ and ʈʈ ; and

c) for the extended-alphabet level, additionally the letters áÁ äÄ âÂ ãÇ ċĈ đĎ đĎ đĎ (with dot beneath) éÉ ěĚ ħĦ íĪ îĬ ĩĬ ĺŁ łŁ ņŃ ñŇ ñŇ óÓ ôÔ öÖ őŐ řŘ řŘ řŘ (with dot beneath) șȘ șȘ țȚ țȚ úÚ üÜ ůŮ źŻ źŻ (this mean letters of Rumanian, Hungarian, Slovak, Polish and French).

d) the ultimate version encompasses all other European languages alphabets in Latin script have been included (this means Estonian, Latvian, Lithuanian, Azeri, Maltese, Welsh etc...).

Note:

The coding scheme of ISO-IR 6 is identical to the scheme defined in ISO/IEC 646:1991 as the "International Reference Version" (IRV). It is also coding-wise identical to ASCII, "American Standard Code for Information Interchange".

3.4 Keyboard layouts

Three different user situations with need for Rromani letters in data processing exist:

- a strictly Rromani language covering all Europe,
- a local "Majority-language" situation and

- a European wide situation of all languages written in Latin alphabet (including all minority as well as majority languages and transcription of Rromani in foreign alphabets). This document therefore defines keyboard layouts for all three situations, focusing upon the Rromani keyboard. The Rromani keyboard layouts are designed especially for ease-of-use primarily in European common Rromani.

BASIC INFORMATION ABOUT RROMANI AS A WRITTEN LANGUAGE

In descriptive linguistic literature uncertainty about the nature of Rromani has prevailed until the end of the 20th century. It was viewed sometimes as a collection of different (mutually related or not) languages, more often as one language with a number of dialects, more or less remote between them, and in some descriptions as a unique language with several "varieties" or "registers". These doubts have been maintained by the lack of global analysis of the problems but by now it has been ascertained that the real nature of Rromani is a composite of these different views: there is one Rromani language, divided in two super-dialects, each of them presenting a series of local varieties characterized more by partial and comparatively recent oblivion of the genuine linguistic material than by any development into sharply isolated dialects or languages. These two super-dialects cover the heritage of some 90% of the Rromani population all around the world. In addition, there is an important branch (around 7%) of the Rromani population who lost the regular use of Rromani as a home language and have retained some Rromani basic vocabulary (a few dozen of words), which they use in the local language of their country, mainly Spanish and English, leading to the creation of some kind of mixed idiom. The last 3% consist of strongly modified local varieties of Rromani under the influence of local languages, to such an extent that communication between their speakers and speakers of Rromani is very difficult. The two last groups have sometimes been referred to as the "Atlantic branch".

The Rromani population (including the Atlantic branch which represents approx. 10% of the whole population) is evaluated world wide between 12 and 15 millions of persons, among whom – according to estimations – 50 to 70% (depending on the method of estimation) can speak Rromani under one form or another. The number is increasing thanks not only to demographic factors but also to the care of European institutions and some governments to protect and promote the Rroms' linguistic heritage.

Among the existing local varieties, differences affect mainly the lexical level (due to both an increasing oblivion of Rromani vocabulary, changes in the needs of how to name new realities and disorderly integration of local foreign lexical elements) but only very limitedly other linguistic levels: morphology, syntax, and phonemics, due to the fact that beyond some conspicuous local discrepancies in pronunciation, the common phonemic underlying structure is well preserved from one variety to another and this common phonemic structure can be expressed quite properly by a common graphic system.

Unfortunately the lack of communication in the past when Europe was divided by countries and political regimes did not allow the emergence of a common form of Rromani, at least until 1971 when the first Rromani World Congress in London laid the pedestal of a common Rromani language based in the key principles expressed in the London declaration that "no dialect is better than any other" (principle of identity respect) and that "a common alphabet has to be elaborated to allow easy communication between users of the various varieties" (principle of linguistic convergence in written). These principles were developed by the Linguistic commission of the International Rromani Union and led to the adoption by the fourth Rromani Congress in Warsaw in 1990 of the "Alphabet of the common Rromani language", based on a polylectal principle recognizing the underlying phonemic structure common to all forms of Rromani spoken all over the world and therefore suitable to all of them, while respecting their specificities and ensuring written communication between users

of these various forms. The adoption of this alphabet, embodied in the so-called Warsaw decision in 21 points, has been confirmed in July 2000 at the 5th Congress in Prague (Czech Republic) and in September 2004 at the 6th Congress in Chieti (Italy). It has been in addition endorsed in January 2007, after seventeen years of practice, by the Stockholm-Uppsala conference referred to above.

However one may still encounter Rromani written in the alphabet of the local majority languages, often due to the lack of appropriate electronic device for data base processing, but also for some very specific purpose like the local circulation of political electoral advertising or portions of the Bible in Rromani translation among people unfamiliar to written Rromani. There is also a quantity of early texts in Rromani written in majority language alphabets which may have to be quoted for some reason or another – not to mention proper names in majority language.

The alphabet of the common Rromani language is based on the Latin alphabet, with a few additions, as most of other spellings are also. In several countries, mainly Russia, Bulgaria and more seldom Macedonia, some users (mainly non-Rroms) write Rromani in Cyrillic script, while attempts to write Rromani in Greek alphabet has been exceptional. Both Cyrillic and Greek alphabets are disregarded in the present document since they are beyond its scope. In addition to the letters of the basic Latin alphabet, the following Rromani letters are needed: from Latin extended A:

- vowels with a grave accent: **à è ì ò ù**
- consonants with an acute accent: **ć ś ź**
- consonant with a cedilla: **ç**
- a preytoticized vowel: **ě**

from C1 Controls and Latin-1 Supplement:

- the Spanish inverted question mark **¿**

from Latin extended B

- further preytoticized vowels with a caron: **ǎ ě ő ů**
- a specific consonants: **ʒ** (ezh); note that the Stockholm-Uppsala declaration leaved the option of using **y** (missing in the alphabet of the common Rromani alphabet) instead of **ʒ** when the letter is unavailable – such a practice can be observed among some Rroms living in Latin America.

from Greek and Coptic:

- and Greek **θ** (theta).

All these letters are represented in the fonts Arial Unicode MS, Microsoft Sans Serif and Lucida Sans Unicode.

In addition to the letters listed above, a number of "odd" letters, be they Latin based like (**z**, **ṛ**, **ṛ**) or not (like **𑂔** or **𑂕**) can be found in various texts of historical character. These are however not covered by this document, which is intended to be applicable only to today's mainstream common use of the languages. Nevertheless, two consonants broadly used in historical context ought to be included into the extended-alphabet scheme – these are **ḍ** and **ṛ** with a dot beneath. The solution of a floating dot beneath (ASCII 0323) would be quite satisfactory and allow an easy transcription of Indian words.

Recommendations exist at the European level supporting the rights of Rromani-speaking individuals to communicate between them and progressively also with governmental authorities and judicial bodies in their own language, even if these recommendations have been so far very incompletely implemented. One of the reasons is the lack of specific support for Rromani in computer software and even of IT-oriented specification of language requirements for this purpose.

For complete data processing coverage of the Rromani language, only basic level implementation (namely Basic Latin letters plus the eighteen letters [sixteen Rromani letters

and historical **ḍ** and **ṛ** with a dot beneath] mentioned above and the Spanish inverted question mark) is necessary (see clause 2 "Implementation levels").

For complete coverage of the local majority languages, extended-alphabet implementation is however necessary. Decisions on what level that should be required in a purchasing situation must therefore be considered very carefully.

DOCUMENT BACKGROUND

A.1 Status of this document

In the 7th to 9th January 2007 conference "A language without borders... an International Rromani Conference" organised in Stockholm and Uppsala by the Roma Cultural Centre in Stockholm⁷ and bringing together the majority worldwide of Rromani experts in Rromani linguistics (called hereafter representatives of the Rromani specialized bodies), it was decided that "the Conference gives a mandate to one group to examine with IT specialists how to develop as soon as possible a Rromani system in all electronic devices and on the Internet" (see Annex C page 24). Therefore, a draft for this document was developed during the same year by the *ad hoc* group under the authority of the chairs of Rromani studies of the Universities of Paris (France), Bucharest (Rumania) and Veliko Tŕrnovo (Bulgaria) and in co-operation with LWP Consulting, Stockholm.

The text has been circulated for comments and eventually amended according to the remarks. The *ad hoc* group decided to adopt it with the proposed changes.

A further step was achieved in September 2009 when two of the above mentioned chairs of Rromani studies (those of the Universities of Paris [INALCO] and Bucharest, while the one in Veliko Tŕrnovo was closed down by administrative decision) were involved in the European project "*Restore the European Dimension of Rromani language and culture*" (EACEA 2009-2163/001-001 CU7MULT7) and decided to extend the linguistic scope of the keyboard driver beyond the Rromani language and encompass all European Latin-based scripts if possible – this undertaking resulted possible and an experimental driver, called "European Universal" was therefore produced.

It is important to emphasize that the present document is of informative and orientational nature, and has in itself no binding status (see Foreword and clause 1 "Scope"). It can however be used as a reference in purchasing, in which case additional requirements may need to be specified; see A.6 (page 10) "Examples of use of document in purchasing" below.

A.2 Coverage of languages

It shall be noted that the document's definition of two implementation levels – "basic" and "extended alphabet" means that one has to distinguish between a kernel for use in writing Rromani (all varieties) and an additional system intended to write all the various languages of European countries where Rroms live, be it to write proper names, quotations or entire texts. The system proposed encompasses not only the common spelling of Rromani, a medium allowing to write all the forms of spoken Rromani in the respect of their specificities, but also all other spellings developed so far locally for this language, provided they are based on Latin script.

All varieties of the Rromani language are of equal importance and accordingly it was indispensable to develop a system able to cover all of them.

The documents treats only the aspect of language writing, so that the description of prerequisites for any device aimed at voice processing is beyond its scope.

A.3 Policies for character coding

⁷ The conference was organised in partnership with Södertörn University College, the Centre for Multiethnic Research/Uppsala University, Romano Paso and the International Romanes Institute in Hamburg.

This document does not declare any data-processing policies, preferences or recommendations. Such declarations would be outside its scope, which is to specify what alternatives exist for the IT support of the Rromani language in so far all European languages in Latin script are included.

As regards coding, only fundamental schemes are covered. Transformation formats, like e.g. UTF-8, are in this connection considered application-oriented, and are therefore not dealt with in the text.

The industry direction in character coding is towards multi-octet schemes, and that direction should ultimately provide the solution to today's multi-lingual problems in IT. At the moment, however, vital areas in data processing may still be dependent on 8-bit coding schemes. It is therefore desirable that policies for "migration" from today's 8-bit situation to multi-octet coding are declared by authorities and other organisations. This document however does not, and cannot, declare such policies.

A.4 Document limitation

This document, as stated in its "Scope" clause, specifies only minimum IT requirements for support of Rromani and a series of European majority languages, for present-day common use. The reason for this "minimal" approach is that data processing in these languages may vary to some extent in detail requirements, and flexibility for individual purchasing situations should therefore be provided. In the specific case, stricter requirements may naturally need to be prescribed in addition to those of this document; see "Examples of use of document in purchasing" below.

A.5 Keyboard considerations

For text-writing in Rromani, the user requires a keyboard layout with convenient access to all the letters of the language under discussion. True enough, in an application where only occasional input of some Rromani words in a national language database is needed and the use restricted to Word texts, usual national keyboards of the respective country will be acceptable, provided that the Rromani letters can also be input from them through the commands "Insert", "Symbols" and the selection and assignment of the Rromani letters to specific keys of the keyboard (using "Shortcut", "Assign" and "Enter" commands). Several fonts provide users with Rromani letters: mainly **Arial Unicode MS**, **Microsoft Sans Serif**, **Lucida Sans Unicode** and since 2009 also **Times New Roman**.

For more systematic and broader language applications it is preferable however that the Rromani letters are available "readymade". It may be also acceptable that Rromani letters are input by typing "dead-key" diacritics followed by base letters. For instance, immediate access to a complete pre-composed letter **ă** could be needed in some cases, but typing caron followed by **a** could be acceptable in others. This document therefore specifies *alternative requirements*, for the user of it to choose from.

Further, one aim of the document is to provide industry with a certain freedom in detail implementations; taking into account the danger that, if specifications are too strict, they could hinder some software development, and/or possibly favour certain vendors at the expense of others.

A.6 Examples of use of document in purchasing

This clause contains some examples of how the document can be used as a reference in purchasing, although these examples do not state requirements, and are not intended as recommendations. As a matter of fact, after the examination of the various keyboard used for European languages, it was concluded that the best configuration can be based on the British keyboard. Nevertheless, local conditions and existing policies, local or higher-level, e.g. as

regards the choice of coding systems, will decide the exact use of the document in the specific case.

The following hypothetical purchasing situation is exemplified:

System for a British English-language application, with input needed of Rromani and other Latin script-based European languages. Unicode/10646 coding has been decided on for the system.

Possible extension to other keyboards is in preparation and a common introductory text could be used for these situations with the following indications:

“This Request for Proposal relates to the acquisition of a system for X, intended to be used for Y and consisting of the following components:

[Specification of hardware/software]

The system shall conform to the requirements stated in the present document and also to the requirements stated in the following documents, as applicable:

[Ergonomics, safety, etc.]”

Basically, the present document applies for:

- a) System-specific text for British English language system supporting Rromani.
- b) Character set: Shall contain basic and extended-alphabet level repertoire and the additional characters of ISO/IEC 8859-1 as stated in the document.
- c) Character coding: Multiple-octet coding is required.
- d) Keyboard layout: Majority-language basic level British English layout is required.
- e) Rromani letters and letters of other European languages in Latin script may be made available pre-composed or as diacritic-letter sequences. Character positioning not specified in the document should agree with current commercially-available British English keyboards, but deviations can be accepted. Key marking with the Rromani letters is desirable and will be the object of a specific exploration, although it should not be required at least during the initial stage of the project.
- f) Fonts: The complete repertoires for the fonts proposed shall be specified in the proposal. The specification may take the form of references to publicly available or included proprietary documents.
- g) Special attention will be paid to the Rromani vowels **ǎ**, **ȳ**, **ǒ** and **ǔ** with caron in order to avoid two technical problems (it will be noticed that these problems do not exist with **ě**):
 - problems in segmentation and hyphenation, since they are treated by most soft wares as Chinese pīnyīn characters and therefore they do not stick to the preceding and following letters, breaking the continuity of the word when occurring at the end of a line,
 - problem of font switching, since in many cases, if used through specific assignment of a key through the “Insert a symbol” method, they are produced in a different font (often Simsun or similar) than the one in which the text is being written.

1. Le manus save kidisajle and-o Švédó and-o Romano Kulturálo Céntró line angl-i jakh le probléme pal-i čhib romani thaj laqí standardizácia. Von areste k-i konklúzia so adés o pučimos si maj but socio-politiko desar strikt texniko. I standardizácia si te kerdól maj dur pđal-i konvergencia le variantonqi pal-el bare linie e štare bare variantonqe kaj si and-i čhib, dindor pativ savorrenqe, sar vorbin pen maškar-el Rroma anda sa i lúmia (o sundal), aj nisar ulavde maškar-el thema thaj le politikane gránice. Si amari propozicia te nimaj phenel pes "dialéko" and-i romani lingvistika (i vórba "dialéko" si la adés jekh baro emocionálo pharimos) thaj te phenas dakana "variánta".
2. Vázdas amen akate speciálo grúpe pal-el pučimáta sar gráfia, informácia, pedagógia thaj leksikografia thaj te keras kedimáta sarkon beré vi duje Rromenčar andar-i Amérika (North thaj Latin America), ke si amaro pláno globálo pe sa i lúmia. Si te avel vi jekh grúpe kaver pal-o copyright (©, le autoronqe xakaja) kana si and-o pučimos le purane tókste andar-i romani tradicia ta kreácia, te šaj den len avri le manus kana trebul lenqe, bi te pokinen love vaš-el autoronqe xakaja, kana keren kadí buti vaš-o phiravimos aj pinzarimos lenqe, a na te keren šešto.
3. Trebul te keras maj but buti pal-i Biblia thaj kavera spirituálo tókste pe čhib romani, but serioz thaj p-o maj uprutno nivéli. I lavutik (dikcionáro) so sas kerdini k-i lingvistikani grúpa te kerdól vi and-i nordutni Evrópa, e thanesqe gazikane čhibenča thaj Kaalenqe vorbenčar andar-o Finitiko Them.
4. Si te dikhas, te kidas, te zuráras sa i buti so sas kerdini ži adés pal-i romani gráfia, lindor sar báza la Varšavaqi khetani gráfia, thaj te keras xurdikani analiza kaj te arakhas sa so šaj te las kaj te lacharav aj te zoráras maj dur o jekhimos la čhibáqo kaj te dikáras palvača i identiteta sa laqe variantonqi, so si ekh barvalimos baro amaro thaj pe aver rig te šaj haláren pen so maj leznes le Rrom andar sa le Thema, bi te xolanón pe-l vórbe, so naj anda lenqi variánta.
5. Maj dur si te keras buti sa pal-i poliektálo metodológia, sar sas phendino kotar-o 1-to Kongrésó k-o London and-o b. 1971, 4-to Kongrésó k-i Varšava and-o b. 1990 thaj vi sar palem mothode amare specialiste anda penqe prezentácie and-o Stockholm.
6. I Konferénca del mandáto jekhe grupaqe, kaj te kerel sváto manusenča andar-i IT kaj te ovel so maj sig jekh romani sistéma anda sa le elektronikane aparátúre thaj vi p-o internet.
7. Provizor šaj te das avri, le manusenqe so si lenqe phares te xramon and-i romani khetani ortografía p-o komputéro, jekh alternativó gráfia vaš-o internet, kaj "z" si te ovel la "y" sar variánta (sar paralel si a thaj o, g thaj g thaj kavera), thaj le kavera grafeme te šaj oven bi diakritikonqo, žikaj te ovel p-i piáca i nevi tastatúra thaj o soft (kon kamel te haznil i khetani gráfia šaj p-o Microsoft Outlook).
8. Mangas te zan 4-5 manus andar amare grúpe (mothodine maj opre) k-o bešimos pal-o kurikulum so kerdóla k-o Raj Joseph Sheils (Politika le čhibenqi and-o Sombeš la Evropaqo and-o Strasbourq) and-o čhon Maj 2007, kaj te ažutis anda lenqe butá.
9. Si amen godi kaj te kidas le étar katédre andar-i romológia (k-o Bukurešto, k-o Greenwich, k-o Paris thaj k-o Veliko Tárnovo) kaj te den vast k-o Romano Kulturálo Céntró and-o Stockholm, kaj te del avri jekh but zorali prográma vaš-i didaktika aj i pedagógia (inkluziv jekh @-lavustik) vaš-i temi generácia thaj le Rrom so siklón vi anda-l barine berés. Kadí buti si te kerdól anda jekh but pašutni kooperácia e universitatača, savi si te ovel la mandáto te siklárel i romani čhib and-o Švédó.
10. Pal-o avindimos la romana čhibáqo, si amen paša kadí deklarácia jekh propozicia le Forumesqe kaj te dikhel la thaj te lingrel la maj dur ži k-i Parlamentáro Asambléa and-o Strasbourq.

Chhon Gelpies - Gies Romani Kuvon, Hungary, Holland
Dr. Trojane Petrov
Laut Hancock
LONGE BEARD
Jasna K. H. H. H.
GHEORGHE STILIA
Dorcas Jhon
Kecce Jhon
Carlos M. Nieto
Dreel Vankon
Hristo Kyuchy xpv / 9.01.07 /
Hristo Kyuchy
Gregor D. Kuch

A CONCISE BIBLIOGRAPHY:

Requirements for Support of Sami Languages in data processing / Gáibádus sámegiela sáhttit geavahit dihtoris / Krav om støtte for samiske språk innen databehandling / Vaatimus tietojenkäsittelyn saamenkielisestä tuesta. s.l. – s. d.

I alfabèta e standardone rromane éhibáqiri; Decìzia "i rromani alfabèta" / Decision "The Rromani Alphabet"; Rezolùcia N. 7/1990 – In: Informaciaqo Lil e Rromane Uniaqoro n° 1-2 (Žulaj-Avgùsto) 1991. Warszawa.

Comparisons of Standardized Character Sets for Europe (Statskontoret & LWP Consulting). Stockholm, 2000:2.

Name-writing in Swedish Authorities' electronic administration / Namnskrivning i svenska myndigheters elektroniska förvaltning. Stockholm, 2004

CEN Standard for Rromani IT Equipment, by K. I. Larsson In: CDSG Meeting. Berlin, 2005.

Character repertoires, ordering and assignments to the 12-key telephone keypad (for European languages and other languages used in Europe). Etsi, 2006.

Presentation of the situation for Rromani in ICT in Sweden and Europe by K. I. Larsson (LWP Consulting) [document CDFG-N258] In: Cultural Diversity Focus Group. Berlin, 2007.

Unicode website : www.unicode.org