

dr. **Marcel COURTHIADE**

Оделење ромских студија – INALCO (Париз)

Комесар за језик и језичка права међународне ромске Уније (IRU)

Од "главе вола" (наиме од слова *alef* – א) до управљачког програма «EuroUniv»

[послао је допринос преко e-mail, јер није био присутан због здравствених разлога]

Питање писања ромског језика није изолирана проблематика. Треба га сматрати кроз историју писања различитих других језика на свету.

Историја писмености разних језика добро је документирана код народа који су остављали писани траг, али врло мало знамо о језицима за које никакав писани траг не постоји. Управо они представљају велику већину древних говора. Готово све ове језике и културе у одређеној мери познајемо само кроз призму других народа, дакле оних који су о њима писали.

Наравно, свој допринос дала је и археологија, нудећи релевантне информације, али оне нису толико изразите као писани текст. Без обзира, упитна може бити релевантност писаних текстова, који неретко могу приказивати више лажи него истине, посебно када говоре о другом народу.

Став о писању у народу, некад и данас, није био и надаље није исти. Данас имамо, пре свега жидовско-кршћанско-муслимански свет, у којем се писање сматра обичним, пожељним и поштованим делом. Опште је познато како су се развијале феничка и еврејска абецеда, на основи каснијег египатског демотског писма, те како се током времена изродила латиница и ћирилица у Еуроци. Мање је позната чињеница да је други развој исте основе стварао слово индијских језика, кроз араамску абецеду.

Араамски језик био је језик праперзијских дворова, те је на тај начин дошао на северо-западни дио индијског субконтинента. Касније се, кроз облик брахми, ширио целом Индијом. Први писани документи били су натписи цара Devanāria Piadasia, познатијег као Ашока (царевина 269 – 232 г. пре Христуса), који је урезао своје «декрете» широм Индије, на стенама и колонама у локалним варијантама говорних језика, како би сви људи знали које су најважније основе његовог закона.

У северо-западној Индији (данас Пакистан) натписи су били урезани араамском и грчком абецедом (пошто је и грчки био службени језик), те даље на југ и исток у брахми абецеди. То је дакле, прво сведочанство писаних текстова у Индији.

Међутим, како се ради о пажљиво формираним словима на великом територију, са међусобно великим удаљеностима, можемо закључити, да ово није био први покушај писања, већ да је постојала извесна традиција писања. Другу информацију налазимо у трактату «Артхашастра» дворског саветника и теоретичара политике Чанакије (познатог и као Каутиља – 370-283 г. пре Христуса). Говорећи о политици, он пише да «архиве имају изразиту вредност за владара». То би значило да су Индијци писали и похрањивали записе у архиве, али су оне вероватно биле несигурне, на нестабилним држачима, на пример на великим осушеним листовима, које су кукци врло брзо и једноставно уништили. Стога су, по неколико пута, преписивали садржаје из архива, док је њихов садржај имао значај за друштво и владара. Касније су престали преписивати старије архиве, те предност дали новима. Можемо зато веровати зашто писана реч никада није поседовала велику вредност у Индији. Овај технички фактор Индијци су маштовито изразили, објашњавајући како писменост припада делу нереалног света, у супротности са изговореном и слушаном речи, која сама улази у сустав значења. Ова концепција изражава супротан положај с једне стране брахмана, чија је реч «права, истинита и верна» као и Веда, које они знају напамет, а с друге стране писара, којег санскртска књижевност представља као лажљивца и лопова. Древни Индијци писање су сматрали невредном и тужном неопходношћу, прозаичним

инструментом чиновника и трговаца, без икакве племенитости, насупрот угледу, који је писање имало код семитских народа. Прави интелектуалац није био онај који чита и пише, већ онај који напамет зна хиљаде и хиљаде стихова са њиховим коментарима. Писање су успоређивали са штакама које служе слабом и хромом уму. Зато се може разумети недостатак уметности калиграфије код Индијаца, док се она у пуном цвату развијала код других цивилизација, које су користиле трајније држаче писане речи (папирус, terra-cota, зидови храма – храмови су се касно појавили у индијској цивилизацији, и др.)

Неки други народи имали су још радикалнији однос према писању, на пример Гали, који су се плашили писма, сматрајући да се у њему крије опасна магија с обзиром да жив човек може изрећи оно што је мртав човек написао. Видели су то у надгробним натписима народа Италије, Римљана и других, што им се чинило несигурно с обзиром да сваки човек може украсти нешто написано и побећи са овом тајном коју је нетко други поверио.

Реч «писати» у индо-европским језицима

Сама реч за појам "писати" показује, да је он дошао касно у индо-европске културе: славенски корен у глаголу "писати" потиче од старословенског < **пѣсати, пишж**, наиме "украсити", што је још видљиво у пољској речи *pisanka* "јајца Ускрса" и има сродство са санскритским पित्ति [piṣati] "украсити", са литвинским *piešiù* "цртати углом" и са латинским *pingere* "украсити, вести" (одакле *pictum* и *pictura, pictogram* итд.). Са друге стране је корен латинског глагола *scribere* "писати" везан са појмом "грести" са варијантама *gr, kr, skr, wr* итд. – све од праиндоевропског **gerbh~*, који дају осим *grebem* грчко γράφω "пишем", γράμμα "слово", γραμμή "линија" (па *grafologija, gramatika* итд.), старо-герм. *krattan* "грести", герм. *kratzen* "огрести", *kerben* "зарезати", *Krebs* "рак", енгл. *scratch* "огрести" итд... Немачки је преузео са латинског облик *scribere* да би створио *schreiben* "писати" (кроз *scriban* и *schriben*), док се старо-енгл. *gr+i+t~ / wr+i+t~* развијао у *writanan* "огрести, писати" и даље у широко познати облик *to write* "pisati".

Овде није место за описивање различитих речи које изражавају појам "читати", али треба напоменути, да и систем дотичног речника такођер карактеризирају неологија и импровизација (лат. *legere* "везати", енгл. *to read* < *ridan* "саветовати, причати", гр. διαβάζω "ставити кроз", алб. *këndon, knon* "певати", мађ. *olvasni* "бројити" итд.).

Кад су говорници ромског језика пристигли у Европу, као што смо рекли, без посебног штовања према симболичној вредности чина писања, адаптирали су се, као увек, локалним условима и преузели су често реч доминантног језика краја где су живели, као што су на пример у Србији створили **pišinel** "писати" од *pisati*. Цеокупна збирка локалних речи за појмове "писање" и "читање" може бити предмет особног члана и овде ће бити довољно да се припомену речи заједничког ромског језика: **lekhel** или **lekhavel** "пише" и **drabavel** или **drabarel** "чита".

Ипак, под утицајем других народа, Роми су често усвојили локалну реч и створили облике као *skril, iskril* (рум. *a scrie*), *irinel* (мађ. *írni*), **pišinel, pišil** (сл. *pisati*), *škroinèla* (алб. *shkruan*), **ğrafinèla** (гр. γράφω), **jazinèla** (тк. *yazmak*), **čhinel** "резати" етц... односно **čitinel, čitil** (сл. *čitati*, рум. *a citi*), **ginel** (букв. "бројити", cf. *olvasni* в. горе), **điavanèla** (гр. διαβάζω), **ledzonèla** (алб. *lexon*), **gilabel** (букв. "пева", cf. алб. *këndon, knon* "1. певати; 2. читати) итд... Међутим било би погрешно сматрати да је **drabavel** неологизам, чим ову реч налазимо већ са значењем "читати" у ромској граматици XVIII. века из Шведске.

Као што је речено у тексту, до данас, сви народи чак и унутар Европе, немају исти однос према писању, без обзира на њихову културну историју. Уопштено говорећи, на западу постоји строжи и прецизнији однос према употреби правописних правила него

на југо-истоку Европе. У старим текстовима примећујемо да није увек било тако. На пример до XVI. или XVII. века постоји више писања једне те исте речи у размаку од неколико редова. Несигурност дакле представља архаични аспект писања. Немачко-скандинавска ригорозност правописа ширила се у суседне славенске земље на исток и даље у Русију већ у XVIII. веку, те касније на цели Совјетски Савез. До Мађарске (укључујући Трансилванију) и Хрватске проширио се преко аустријске традиције. Ипак, Балкан је остао изван овог утицаја и неписменост је тамо владала трајно, тако да се народи (осим малобројне елите) нису навикли, да марљиво пишу према одређеним правилима. Није у питању древност писмености, него ширина народне основе ове писмености, јер на пример грци, који имају прадавну књижевност, ипак често мало маре за исправност писања – истина је да имају тежак систем, где се исти звук мора писати једним од неколико начина, овисно од етимологије, коју често људи више не знају. Углавном можемо рећи да, док је на неки начин одређена срамота везана са правописним грешкама на западу и северу Европе, не налазимо такву емоционалну реакцију на југу Балкана и већина људи без икаквих кочница пише «као што говори» – или барем тако мисли. Још дан данас, огромна већина омладине Балкана, кад пише на рачунару – пре свега интернетом, користи искључиво „notepad“ или „text dokument“, који не садржи ниједан диакритични знак, па не знају уопште да постоји на свету, у употреби *Word* у више верзија, помоћу којег би могли писати сопствени језик на правилан начин, не осакаћен. Наравно, не можемо подцењивати чињеницу, да је нотепад бесплатан и унапред упрограмиран у рачунар, док је *Word* релативно скуп и треба га купити и инсталирати посебно.

Уопштено говорећи, у свим земљама традиционална елита пише марљиво али проценат ове елите зависи од историчних услова сваке земље посебно. У земљама, где је систем образовања, често силом, био успешан, бројнија је елита – на пример у Бугарској под немачким а после советским утицајем; док тамо где се мање поштовало школовање, људи нису тако осетљиви на правопис.

Наравно, у срединама, где уопште није постојао одгој на матерњем језику, најмање имамо изражену бригу око правописа, као код Рома, Арумуна, Помака и других мањина изван матичне земље, јер им није доступно образовање на матерњем језику (Македонци у Грчкој, Румуњи у јужном делу Србије итд.). Најчешће ове групе немају ни новине, ни остале штампане материјале или књиге на матерњем језику, па немају одакле црпсти емулацију за правилно писање.

Четири основна елемента свести о исправном писању

У ставу сваког народа према писању општенито и посебице о исправном писању, можемо разликовати четири основна елемента:

1. Унутарња традиција: еврејски језик на пример има јаку традицију писања, било где у Европи, што се може запазити већ од раног одгоја деце, па до њиховог обичаја да се пише и еврејска варијанта локалног језика еврејским писмом – што даје жидовско-немачки односно јидиш, жидовско-шпански односно джудезмо, жидовско-арапски итд...

2. Подручје боравка неког народа : до жидова је слична ситуација северних језика Европе, где се школско образовање развило рано, интензивно и широко. Чак и народи без државе, као што су Сами у Скандинавији, редовито читају шта се о њима пише, воде расправе преко новина и створили су већ, недуго након другог светског рата, абечеду властитог језика. На жалост, створили су фактички 13 различитих дијалекталних абечеда, што је довело до распада њиховог писаног језика па и чак самског језика уопште. Данас готово да и није више у употреби, не због недостатка абечеде или интереса, већ зато што су превише мале групе користиле сваку абечеду посебно, уместо да имају заједничко писмо. Тај систем, дакле, није био нити практичан

нити економски одржив. Са друге стране имамо народе, који не држе толико до писања, као на Балкану, у Турској и др.... Наравно ово је случај већине Рома – али не свих, и не само случај Рома, како бисмо могли веровати судећи по записима разних "стручњака".

3. Приступност или одсутност школовања на матерњем језику, доступност или недоступност новина и осталог писаног материјала на матерњем језику, итд...

4. Једноставност или комплицираност писменог система, у односу на говорни језик : јасно је да логичан усклађен систем писања, као хрватски или шпањолски, који имају више услова за масовно правилно коришћење него систем типа грчког, француског, чак и бугарског (сами бугари често мешају "е" и "ј" због шопског карактера говора у престоници Софија) или румунског језика.

Ако погледамо ситуацију неколико језика у светлу ових елемената, можемо запазити специфичан положај ромског језика:

	унутр. традиција	Подручје боравка	школовање	систем писања	тотал +
српски	+	++	++	++++	9
ромски	0	0/+	0	++++	4/5
грчки	++	0	++	0	4
албански	0	0	++	+++	5
бугарски	++	0	++++	+++	9
македонски	+	0	++	++++	7
румунски	+	0	++	+++	6
мађарски	+++	++	+++	+++	11
полски	+++	+++	++++	+++	8
француски	+++	++	++++	0	8
енглески	++	++	++++	0	7
самски	0	+++	+	0	4
шпански	++	+	++++	++++	11

Број плусева изражава наклоност датог елемента према поштовању правилног правописа : 0 има лош утицај, ++++ има врло позитиван утицај. Тотал изражава вероватност да буду писане велике количине текстова правилно, то значи да може преживети исправно написан језик .

Прелаз на механичко писање (писаће машине, рачунар, интернет)

Нису сви језици истоветно поступили у односу на механичко писање. На пример, за време конференције о писању албанског језика у Монастиру у г. 1912, судионици су изабрали као слова са дијакритичким знаковима само она, која су већ постојала у слагању тисканица и на писаћим машинама, односно *ç* и *ë* – иако стално понављање слова *ë* у овом језику чини писање врло заморним сваки пут, прво "а" после *e*. Напротив Румуни нису имали ову дилему и од самог почетка су створили из олова слова *ă*, *ș* и *ț*, додатно до већ постојећих *â* и *î*. Тако и Турци нису видели проблем, да створе за време прелаз са арапске абецеде на латински, нова слова *ı*, *ğ* и *ş*.

Данас, систем рачунара дозвољава, да се било који језик може писати одговарајућом абецедом (чак и самски у 13 разних абецеда!) на рачунару или интернету. Није увек било тако : у почетку демократизације информатике, већина језика није имала својих дијакритичких знакова на располагању, па су тада Југословени све писали без дијакритичких знакова, то значи на исти начин *celo* и *čelo*, *sis* и *šiša*, или *zelje* и *želje*... Мађари, који нису у почетку имали у рачунару *ő* и *ű* заменили су их помоћу тилда над самогласником : *õ* и *ü*. Касније су ипак продуценти информатике увели сва слова, укључујући мађарска *ő* и *ű*, све сугласнике типа *š*, *č*, *ž*, *ś*, *ć*, *ź* или *ż* итд.... Ипак много људи је и даље писало без дијакритичких знакова, иако су они били доступни, чак

велома лако доступни. Кад се форум на интернету отвара, примећујемо да велики број цибернаута пише из подручја бивше Југославије без знакова.

Проблем је још гори са језицима, који имају другу абeцeду, као грчки, македонски или бугарски. Иако постоји целовит систем електронског писања грчким словима, још увек велики број Грка пише апроксимативном латиницом, коју зову *φραγκοχιώτικα* [франгохиотика], па фактички постоје углавном две врсте овог писма, "фонетски" и тада се текст чита више мање као што би звучало на грчком, или "пандански" и тада за свако грчко слово имамо латински пандан или број сличног облика (8 за θ, 3 за ξ), као на пример:

прави грчки	φραγκοχιώτικα		прави грчки	φραγκοχιώτικα	
	фонетски	пандански		фонетски	пандански
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	ou, ou	υ [v, f]	v, f	y
αι	e	ai	δ	d	d
η	i	h	μπ	b	mp

Примери¹ са речима :

прави грчки	φραγκοχιώτικα		прави грчки	φραγκοχιώτικα	
	фонетски	пандански		фонетски	пандански
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

Тако да иста реченица, може, зависно од писара, изгледати на следећи начин:

грчки ερωτέυτηκε αύτη και μας ξεφούρνισε ωραιά τραγούδια με την γλυκή της φωνή

фонетски eroteftike afti ke mas ksefurnise orea tragudia me tingliki tis foni

пандански erwteythke ayth kai mas 3efoynise wraia tragoydia me thn glyki ths fvnh

превод заљубила се и послужила нам лепе песме својим слатким гласом

Прилична је разлика међу двама транскрипцијама, али фактички их већина цибернаута меша у великој мери. На пример честа транскрипција *entipwsi* није нити фонетска (треба би било *entiposi*), нити панданска (треба би било тада *entypvsh*), већ мешавина обе. Општа је тенденција да се све више пише пандански али неретко налазимо транскрипције фонетског типа, пре свега кад су кратке.

Исто важи за бугарски и македонски језик, али без толиког колебања као на грчком.

Примери са бугарским :

импролатиница	бугарски	превод
vsi4ko vurvi/varvi/vyrvi	всичко вврви	sve u redu
ako jelaesh, babta mi	ако желаеш, баца ми	ako želiš, oče moj
naI-grozniq ³ 4ovek v bulgariq	нај грознија[т] човек в Българија	najružniji čovek u Bugarskoj

¹ Сви су примери преузели од коментара под снимањима Youtube'a.

² Превод : знају, хоћу, улази, вреди, сурађујем, у животу; десна колона : глас, слажем се, може, радујемо се, утисак, без.

prijeteli, zabavljajte se	приятели, забавляйте се	prijatelji, zabavljajte se
trqбva da izglejda 4ovek	трябва да изглежда човек	treba izgledati čovek
ne me obarkwaite we h0raa	не ме обърквайте в хора	nemojte me mešati
slan4eviyat	слънчевият човек	sunčani čovek
sladak grqh	сладък грях	Slatki greh
napyлно отка4en i smeбен	напълно откачен и смешен	potpuno otkачen i smešan

Видимо да су се појавиле конвенције за допуњење слова која недостају: *a, e, i, o, b, v, s* итд. имају латинско слово, док су на пример за *ъ, ѝ, я, ч, ш* односно *щ* бугари изабрали *a* (или *u, y*), *I* (или *y*), *q* (или *ya*), *4* (или *ch*), *б* (или *sh, ch*) односно *bt* (или *sht*)... Несигурност запажамо пре свега при транскрипцији тврдог знака *ъ*, који се у импрлатиници пише *a, u* или *y*: *bilgariq, bulgariq, balgariq* (България – обично без мајускуле). И слово *ю* представља проблем па зато се реч *кључек* можемо наћи у следећим писањима : *kuchek, kiuchek, ku4ek, kiu4ek, kio4ek, cucek, kucek* итд. (мн. *qki kiu4eci* «јаци чучеци»).

У случају македонског језика, има мање несигурности, јер је пракса побочних алфавета у српско-хрватском језику за време бивше Југославије довела до лаког прелаза од једног до другог и опуштање знакова не сматрају македонски цибернаути за велику грешку – па зато налазимо *bidejki, prodolzete, pesnichka, ke najdam*. Ретко налазимо бугарске бројке у место слова, али ипак постоје : *de4ki zakon* «деџи закон»... Налазимо и дијалектални изговор, као на пример у поруци : *ne se znai koj prv ke pukne od ubaite tomi* (прав. *не се знае кој прв ќе пукне од убавите моме*). Реченица *Q sakam da te galam* представља истовремено српску заменицу «ја» и бугарско писање овог звука уз помоћ слова *Q* (за *Я*) па налазимо често код цибернаута и заменицу *as* «ја», која није ни македонска (правилно : *јас*) ни бугарска (*аз*).

Доста изненадно, албански подноси исти третман, иако тај језик има латинску абецеду и само два посебна знака, као што смо већ горе увели, наиме: *ë* и *ç*. Можемо у @-форумима и у комен-тарима уоутубе'а читати поруке типа :

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

превод Манди је врло далеко од Мухарема. Мухарем је број 1 за љубавне песме. Тај немар није видљив само код полуписмених цибернаута (у овом случају код Албанца, који је био школован ван албанске школе, вероватно у иностранству, или сасвим неписмен), већ је можемо наћи и на неким сасвим «официјелним сајтовима» Албаније или Румунске. Већина страница румунских министарстава, па чак и дуго време председника републике, нису користили слова *ț, ș* или *ă*, него су дуго писали *t, s* или *a...* Од 2005. г., отприлике су ове странице правилно, чак и понекад прецизно и тачно написане. Нису у питању овде Роми из махале него чиновници високог ранга, који тако мало брину за властити језик, да су га и они дуго времена сакатили.

Осим проблема дијакритичких знакова, који је углавном решен за интернет странице, иако није за већину форума, имамо у јужним областима Србије проблем бркања звукова *č* и *dž* са једне стране и *ć* односно *đ* са друге стране, јер тамо људи у изговору не разликују «речи» и «рећи»⁴, и дакле пишу, на пример **govoreći* уместо *govoreći* (исто као што Албанци на Косову не разликују *ç* од *q* и *xh* од *gj* па пишу на пример *qmit*

³ Правилно би требало дописати -т на крају.

⁴ Позната је Вукова формула «пиши као што говориш», али мање је познато, да ју је он фактички преузео од Саве Мркаља, који је сам у 1810 г. цитирао безауторску хрватску граматику из 1779. г. «*как ње изговарају гѣчи, так ње у рјѣзати имају*». Упркос тога, Вук никад није допустио мешање *ч* и *ћ* нити *џ* и *ђ*, јер је доста критично написао «простота покварено говори» и људи који се на Вука уздају на основи прве формуле (каои да се наводно може како год писати језик) морају ипак бити поштени и узимати у обзир и другу вукову формулу.

«цена» уместо правилног облика *çmim* итд...). Зато је изузетно важно и потребно школовање на властитом језику – не можемо озбиљно очекивати писање без грешака од људи, који никад нису прошли кроз барем један озбиљнији курс на матерњем језику. Истина је да у неким уређајима (СМС, оутлоок итд.) није лако писати датом абecedом и зато је одељење ромистике у Букурешту замислило поједностављен систем писања за такве ситуације (али не за ромски уопште) : *sx, c, ch, zx* уместо *ś, ć, čh* и *ž*, па онда *8* и *3* уместо *θ* и *з*, писање без инфлекса (*a, e, i, o, u* уместо *ă, ă, ı, ő, ŷ*) или са «у» кад је неопходно (*ya, ye, yi, yo, yu*) итд... На пример, три следеће реченице, узете случајно у збирци, представљају следеће фреквенције промене (између 1/6 и 1/21):

Нормално писање	Поједностављено писање за СМС
Gelem gelem, lungone dromençar, maladilem baxtale Rromençar. Aj Rromallen katar tumen aven E saxrençar, bokhale čhavorrençar?	Gelem gelem, lungone dromencar, maladilem baxtale Rromencar. Aj Rroma!len katar tumen aven E saxrencar, bokhale chavorrencar? [5 промене за 107 слова = 1/21]
Dilino phenel so zanel – godăver zanel so phenel	Dilino phenel so Zanel – godyaver Zanel so phenel [3 пром. за 40 сл. = 1/13]
O mačho lel te khàndel e šeresθar	O macho lel te khandel e sxeres8ar [4 пром. за 27 сл. = 1/7]
O lačho manuś thulöl panėsθar, o biužo manuś nasvalöl avginesθar	O lacho manusx thulol panyes8ar, biuzo manuś nasvalöl avgines8ar [9 пром. за 55 сл. = 1/6]

Није неопходно означити *ž* према *z*, јер је *z* око 10 пута чешће него *ž* у тексту. Обратно, *ć* је неколико пута чешће него *c* и контекст у малтене свим случајевима уклања колебање – поготово кад се тиче слова *čh*, јер *ch* уопште не постоји. За пар *s* и *ś*, чим је мања разлика (*s* се појавља 3-4 пута чешће него *ś*), добро је означити разлику групом *sx*. Што се тиче самогласника са инфлексом, подсећамо да низ дијалеката (пре свега Е-супердијалект без мутације, односно гурбетски, джамбаски, чергарски итд...) само ретко имају јотизацију и писање на СМС и Outlook може следити њихов пример, док знакова *c, 3* и *8* добро могу заменити *ç, з* и *θ* у кратким порукама.

Укратко, важно је запамтити:

1. да се тај систем односи само на недовољне системе као СМС или Оутлоок
2. да се ово не односи на директно писање говорних облика ромског, него да је поједностављење извадено из већ стандардног писаног језика :

локалне варијанте	→→	заједнићи језик, заједничко писање	→→	поједностављено писање за СМС или Оутлоок
-------------------	----	------------------------------------	----	---

Укратко, много тога овиси о традицији, марљивости или немару према језику, од поштовања његове вредности, од разумевања колико је битно писати правилно да би сваки читатељ лако, брзо и без забуне разумео значење текста. Није ово дакле никако ромска особина, него превасходно балканска – укључујући грчки упркос прадавне књижевности на високо оцењеном језику, због његовог занемаривања за време османске владавине.

Чиме данас располажемо код рачунара ?

Фактички још пре неколико година није било тако лако писати ромски на рачунарима. Постојало је полуимпровизирано решење, да се убаце слова помоћу наредбе "Insert", па

после "Symbols" и тада се појављује на екрану оквир подељен на квадратиће, сваки са једним словом унутра.

За свако специфично ромско слово, наиме (ć Ć ś Ś ź Ż ź 3 θ ă Ă ę Ę ĩ Ĭ ǫ Ő ŷ Ŭ ę Ć), требало је тада по реду кликнути на одговарајући квадрат, где се појављивало дотично слово и после тога кликнути на знак "Short cut". На екрану се појављивао тада други прозор са насловом "Customize keyboard". У њему се регистривала изабрана пречица за дотично слово притиском на тастер према следећем попису:

за ć Alt истовремено са тастером "c"
за Ć Alt и Shift истовремено са тастером "c"
за ś Alt истовремено са тастером "s"
за Ś Alt и Shift истовремено са тастером "s"
за ź Alt истовремено са тастером "z"
за Ż Alt и Shift истовремено са тастером "z"
за 3 Alt истовремено са тастером "y"
за 3 Alt и Shift истовремено са тастером "y"
за θ Alt истовремено са тастером "8"
за ă Alt истовремено са тастером "a"
за Ă Alt и Shift истовремено са тастером "a"
за ę Alt истовремено са тастером "e"
за Ę Alt и Shift истовремено са тастером "e"
за ĩ Alt истовремено са тастером "i"
за Ĭ Alt и Shift истовремено са тастером "i"
за ǫ Alt истовремено са тастером "o"
за Ő Alt и Shift истовремено са тастером "o"
за ŷ Alt истовремено са тастером "u"
за Ŭ Alt и Shift истовремено са тастером "u"
за ę Alt истовремено са тастером "w"
за Ć Alt и Shift истовремено са тастером "w"

Када је већ одабрана пречица дотичног слова, кликне се на "Ascribe" и "Close". Ово се понавља за свако специфично ромско слово (цела процедура траје за свако слово неколико секунди) и после тога се свако од њих пише притискањем основног тастера истовремено са Alt (односно Alt и Shift за велика слова).

Тај систем, који још увек вреди, омогућио је писање ромском абecedом, али ипак је указао и на три сметње:

1. Дуго нису сва ромска слова била доступна у главним типовима слова, као Times New Roman. Ипак могуће је било писати ромски у Arial Unicode MS, Lucida Sans Unicode, Calibri или наравно Microsoft Sans Serif. Сада је већ од године 2009 Microsoft увео целовиту ромску абечеду такођер у Times New Roman, што значи да је ова сметња нестала.
2. Друга препрека је била чињеница да корисник мора припремити свој рачунар горе описаном операцијом. Није ни ово била велика препрека, јер је програмирање остајало после тога у рачунару, док се не избрише.
3. Највећи се проблем веже са чињеницом, да је ово програмирање било оперативно само у Wordy и зато ограничено *Microsoft Outlooku* (није оперативна на пример са *Outlook Expressom*, у формуларима или интернет-потрагама типа *Google*, *Internet Explorer* итд.).

Најновија развојна кретања ромске технологије

Ево стижемо до циља овога реферата: од пре неколико седмица, већ постоји драјвер (управљачки програм) звани EuroUniv, који дозвољава писање не само ромском абecedом, већ и абечедама свих европских језика, који се служе латиницом (има их

преко 50, што укупно представља око 156 слова) и у латинској транскрипцији језика северне Индије (још десетак других слова типа tTdDrRnNnM). Све три горе наведене препреке нису више актуалне, јер се може писати у фонтовима Times, Arial, Lucida итд... наике у било којем фонту *Unicode* : претходно програмирање више није потребно, јер се управљачки програм може добити као *download* типа *patcha*, увести у рачунар и одмах је оперативан – и то не само као раније у *Wordy* него и у свим системима, укључујући на пример и програме филмских поднаслова.

Свако ће за сада добити по један CD, где је управљачки програм гравирани али технички детаљи још су тајна (под заштитним бројем Nr. INPI 403.723-281.210). Овај управљачки програм има посебан значај у данашњој Европи, јер:

а.) -ниме се може писати на педесетак европских језика⁵ са њиховим варијантама, уз помоћ само једне тастатуре, наике енглеске, и једне једине "*linguistic and regional options*", без промене за сваки прелаз с језика на језик; концепција управљачког програма је била замишљена на такав начин, да слова и знакови одговарају у 96% случајева онима који су гравирани на физичким тастерима;

б.) -у једном тексту може се писати на више језика, што је корисно у публикацијама о језикознању, посебно упоредном језикознању, о историји или географији, јер се мора цитирати велики број речи и личних имена. Ипак тај управљачки програм није намењен за оне, који пишу "само" на два до три језика, без честог прелаза са једног на други, јер у њиховом случају остаје бољи систем прелаза са једне језичке опције на другу уз промене тастатуре (јер кад остају исте типке, ово ипак знатно смета брзој употреби), него баш за кориснике, којима су потребне бројне абецеде у кратком времену, као језикозналци, историчари, географи и слични.

ц.) - недавно су почеле скандинавске земље примењивати у пракси правилан, дослован правопис, то значи дословно писање свих личних имена (људи и њихових градова, села или места рођења, венчања и смрти) у документима особних података, у сврху поштовања идентитета свакога али такођер смањења ризика кривотворења докумената приликом пропуштања овог или оног дијакритичког знака.

Принципи нашег управљачког програма су следећи :

1. апсолутан приоритет имају слова ромског језика;
2. релативан приоритет имају слова језика са најбројнијим ромским становништвом, јер кад у својој земљи ром пише текст, мора стално уводити властита имена или цитате на већинском језику, значи у првом реду румуњски (2.000.000 рома) па и други по реду;
3. следећи приоритет имају у оквиру сваке абецеде најчешће употребљавана слова, узевши у обзир истовремено слова, која су у употреби у више језика (као *ž, â, ç*) и тада имају предност у односу на она слова, која су ограничена на мањи број језика (као рецимо *ц, ř* или *č*).

Наравно, распоред слова на типкама одговора мнемотехничким принципима, да би човек лако, брзо и дуготрајно запамтио комбинацију типки за писање одговарајућег слова. Зато је изузетно важно да представимо овде ово решење за што ширу публику, јер имамо обавезу да отворимо ромском језику што више врата, као што су циберсфера, ИТ, штампање, филмски поднаслови, И-пад, смартпхоне итд... Не смемо се задовољити ламентирањем око судбине језика или декларацијама намере, већ је неопходно да га

⁵ Наике албански, арумунски, азерски, бретонски, чешки, дански, естонски, еускерски (еускера батуга), фæроенски, фински, француски, галицијски, галски, хрватски, ирски, ислански, кашубски, каталански, корсикански, курмањи (курдска латиница), латински (са скандирањем), летски, литвански, луксембурски, лужички (обе варијанте), мађарски, малтански, недерландски, немачки (све варијанте), норвешки (све три варијанте), окцитански (све варијанте и графије), пољски, португалски, реторомански, ромски (све варијанте), румунски (и пара-румунски језици), самски (односно лапонски, сапмигиеле – све 13 варијанте), сардински, словачки, словеначки, шпањолски, српски (латиница), шведски, таљански, турски (укљ. караимски, татарски & гагауски), транслитерација индо-аријанских језика и транслитерација рп̄n̄ȳn̄'a.

укључујемо активно у све делове сутрашњег живота, не само због техничких разлога него и у сврху повратка језику признања и престижа, које никад више не би требало, нити смело губити. Радујем се да постоји на терену бивше Југославије, а конкретно у Загребу, група инжењера, који воде одличан пројекат « gjesnik.net » и који су сагласни да сарађују са нама да би ромски језик био укључен у састав европских језика на интернету.

Да закључимо, можемо још нагласити, да је данашња ситуација писања у ИТ следећа :

1. дуго нису институције, чак и главна државна тела, користиле све могућности правилног писања властитог језика на интернету;
2. ипак, уназад 5-6 година, оне углавном пишу текстове како треба;
3. претежно просечни цибернаути ипак не користе све технолошке могућности и пишу без знакова, мада понекад са бројчаним панданом, као бугари и грци; значајан је факт у томе, да је за правилно писање језика, неопходно посебно купити прилично скуп програм *Word* (или слични), па га увести у рачунар, док у њему већ постоји програм *notepad*, бесплатан, мада не дозвољава правилно писање свих слова : некако се подразумева да се већина корисника задовољава тиме, јер њихови текстови никад нису веома суптилно израђени.
4. такво неправилно писање знатно отежава читање и дакле ограничава до најнижијег нивоа размену мисли и осећаја, обично до нивоа дивљења без нијанса или псовки – што наравно не захтева богати језик нити мудру главу. Има се утисак, да циљ није приказ информације или осећаја него само пуко пражњење агресивности. Видљиво је ово такођер у енглеском језику, исто као у оним језицима, у којима наводно знакови представљају потешкоће... Ја бих ово – с опрштењем, илустрирао следећим примером, који сам баш јучер нашао : « *macadonia is grrek beacause the greeks fukc the girls of macdonia.....so macedonias is bastard of the grEeks* » [нисам променио ниједно слово]⁶. Енглеска стилистика, правопис речи, достојанство аутора, његова политичка доличност и људска умереност су више мање на истом (жалосном) нивоу. Можемо закључити, да такав текст не заслужује нити бољи правопис, нити више пажње, нити чак дуг живот. Аутор је хтео дати себи на вољу и није марио да ли ће нетко прочитати његово опуштање.
5. дивље писање угрожава језике, које већ имају стандардни облик – или барем угрожавало би их, кад не би имали потпору традиције, академије, образовања у школи и ван ње, богатог штампаног материјала, медија итд. Дакле тврдње, да се може стандард ромског језика створити сам од себе путем дивљег писања на интернету (кад немамо потпору традиције, академије, образовања на ромском, довољно штампаног материјала итд.) потиче или од крајње глупости неодговорности или од систематске стратегије уништавања нашег језика.
6. највероватније ће цибернаути и даље још дуго писати псовке и друга приопштења нижег нивоа без брижљивости према правопису, па не представља ово посебну штету.
7. као што су институције временом полако прихватиле да поштују правопис свог језика, уверен сам да ће сада, чим је и ромски доступан уз помоћ управљачког програма EuroUniv, наши цибернаути такође користити тај систем, и када буду имали воље приопштити стварно нешто умно другим Ромима. Потребан је само начин проширења доступности управљачког програма и његовог упутства.

⁶ Могли бисмо хиљаде таквих примера навести, што је немогуће, али и наведени је доста илустративан и не захтева додатни коментар : « *makedonija je vecna a bugari se makedonski zigani...*»

REQUIREMENTS FOR SUPPORT OF WRITTEN RROMANI LANGUAGE
(and of an all-European Latin script operative system)
IN DATA PROCESSING SYSTEMS

SO KAMLÖL PES KERDINO KAJ TE ŠAJ LEKHAVDÖL E ĆHIB RROMANI
(thaj vi sa e avera evroputne ĉhibă latinikane patranenĉa)
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, 39-1ª 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), pînyîin and rômaji transcriptions (incl. revised Hepburn and kunrei-siki), Polish, Portuguese, Romantsch, Rumanian (& para-Rumanian languages), Rromani (all varieties), Same (or Sápmigiele, 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:

- 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
- the specific Rromani letters àÀ äÄ éÉ ċĈ èÈ ěĚ ìÌ įĪ òÒ óŎ śŚ ùÙüŮ žŽ žŹ and 00 ; and
- for the extended-alphabet level, additionally the letters áÁ äÄ âÂ ãÃ ąĄ ċĈ đĎ đĎ đĎ (with dot beneath) éÉ ěĚ ħĦ íĪ ĩĬ ĺŁ ĺŁ ŋŊ ñŊ ñŊ ōÓ ôÔ öÖ őŐ řŘ řŘ řŘ (with dot beneath) șȘ șȘ țȚ țȚ úŮ ůŮ ůŮ žŽ žŽ (this mean letters of Rumanian, Hungarian, Slovak, Polish and French).
- 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: ě ĭ ō 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, r, r̃) or not (like ɓ or □) 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 ɖ and ɣ 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 ɖ and ɣ 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 document 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

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.

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.

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**, **ĩ**, **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 Roma anda sa i lúmia (o sundal), aj nisar ulavde maškar-el thema thaj le politikane gránice. Si amari propozicia te nimaj phenel pes "dialéktó" and-i romani lingvistika (i vórba "dialéktó" 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 leksikográfia 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 Konferenča 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 aparatúre thaj vi p-o internet.
7. Provizor šaj te das avri, le manusenqe so si lenqe phares te xramon and-i romani khetani ortográfia 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 Sombes la Evropaqo and-o Strasbourq) and-o čhon Maj 2007, kaj te ažutis anda lenqe butá.
9. Si amen godi kaj te kidas le štar katédre andar-i rromológia (k-o Bukurešto, k-o Greenwich, k-o Paris thaj k-o Veliko Tárnovó) 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 Bernard
Jasodko L. Harari
GHEORGHE STILIAH
Dobrošlav Štefan
Marek Covicinovic
Carlos M. Nieto
Dreel Vankov
Hristo Kyuchukov / 9.01.07 /
Hristo Kyuchukov
Gregor D. Kuick

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 éhibäqiri; 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