

Милош Ранчић

Норма кодних распореда и сродна питања

2001

САДРЖАЈ

<u>НОРМА КОДНИХ РАСПОРЕДА И СРОДНА ПИТАЊА</u>	1
<u>САДРЖАЈ</u>	1
<u>О РАДУ</u>	2
<u>ПИСМА</u>	2
<i>Кратак преглед токова</i>	2
<i>Минимум стилских обележја српске ћирилице</i>	3
<i>О изгледу слова у данашњој масовној употреби</i>	4
<i>Скуп слова углавном стабилне граfiје: б, в, з, ђ, ћ, (љ), (њ), (р), (ч)</i>	4
<i>Скуп стилски повезаних слова нестабилне системске граfiје: д, ж, к, л, љ, м, у</i>	4
<i>Неусаглашена ћирилска граfiја: мала курзивна слова „г“, „п“ и „т“</i>	5
<i>О посебним стилским недоследностима</i>	5
<i>Неке напомене о сетовима слова</i>	6
<u>О КОДНИМ РАСПОРЕДИМА И ТАСТАТУРАМА</u>	6
<i>Преглед постојећих кодних распореда</i>	6
<i>Кодни распореди</i>	6
<i>Српски језик и кодне стране</i>	7
<i>Проблеми пресловљавања и њихова решења</i>	8
<i>Проблеми функционалне и стилске граfiје српских азбука</i>	8
<i>Тастатуре</i>	10
<i>Оперативни системи и кориснички програми за унос и обраду текста</i>	11
<i>Микрософтови Прозори</i>	11
<i>Линукс</i>	11
<u>НАШЕ ПОТРЕБЕ И СКУПОВИ ЗНАКОВА</u>	11
<i>Основна ћирилица и латиница</i>	12
<i>Ћирилица</i>	12
<i>Латиница</i>	12
<i>Основни незбућни знаци (бројеви, реченични и правописни знаци, акценатски знаци итд.)</i>	13
<i>Акцентована слова</i>	13
<i>Ћириличка</i>	13
<i>Латиничка</i>	13
<i>Скупови научних, културних и осталих знакова за друштвену употребу</i>	13
<i>Ћирилице у културној употреби нашег народа</i>	13
<i>Савремене словенске ћирилице</i>	14
<i>Старосрпска, старословенска и црквенословенска ћирилица</i>	14
<i>Латинице у културној употреби нашег народа</i>	14
<i>Савремене европске латинице</i>	14
<i>Даничићева латиница</i>	14
<i>Грчко писмо</i>	14
<i>Основни скуп знакова</i>	14
<i>Проширени скуп знакова</i>	14
<i>Глагољица</i>	14
<i>Јеврејско писмо</i>	15
<i>О осталим писмима</i>	15
<u>ПРИЛАГОЂАВАЊЕ ПИСМА</u>	15
<u>РЕШЕЊА</u>	16
<i>Кодни распореди</i>	17
<i>Тастатуре</i>	17
<i>Оперативни системи и кориснички програми за унос и обраду текста</i>	17
<i>Преводи</i>	17
<u>ЛИТЕРАТУРА</u>	18

О РАДУ

Лингвистичкој јавности је неопходно да се подробније упозна са проблематиком кодних распореда, графије и сродним питањима. Ово су граничне области науке о језику, али данас, у времену све веће везаности за електронску технологију, од пресудне важности за несметану и ваљану комуникацију.

У том смислу сам претходну верзију овог рада приложио Одбору за стандардизацију српског језика. Ипак, у питању је интерна публикација која излази једном годишње, па не може пратити брз развој дешавања у овим областима нити да отвори динамичку дискусију, неопходну за стварање ваљане норме. Зато сам рад приредио за објављивање у *Лингвистичким актуелностима*, с тим да је због величине подељен на два дела.

Први део чине поглавља „Писма“ и „О кодним распоредима и тастатурама“, где су укратко описани наши проблеми у светлу међународних токова. У другом делу се налазе поглавља „Наше потребе и скупови знакова“, „Прилагођавање писма“ и „Решења“; где су дати предлози за решења ових питања.

Сваки од ових делова чини врло кратак опис своје области, тако да ћу се потрудити да у скорије време приложим јавности и детаљније прегледе заједно са графичким и формалним описима кодних распореда и знакова у нашој информационој и културној употреби.

ПИСМА

Кратак преглед токова

Почетком коришћења писма један народ улази у своју историју, у писану прошлост. До тада, тај се народ налази у праисторији.

По ономе што је несумњиво дошло до нас, почетак српске историје везује се за стварање сталних насеобина на балканској обали Јадранског мора¹ и латиничко писмо. Почетак историје, дакле, јединствен је за Србе и Хрвате.² У каснијим вековима средиште српског народа прво је примило глагољицу, а потом, до данас, за своје основно писмо прихватило ћирилицу. Од краја 19. века и Ђуре Даничића, слабије, а од 1945. године све јаче, Срби користе и југословенску латиницу.³ Поред те латинице, Срби и Хрвати су током 19. и 20. века користили тзв. Даничићеву латиницу, коју одликују монограми за слова „љ“, „њ“ и „џ“.

Услед разноликих политичких околности, данас је стање писама које користе Срби у превирању.

Колико зарад идеје о стварању југословенске нације,⁴ толико и због политичког утицаја Запада (пре свега Ватикана) и привлачности западне културе насупрот (из разних разлога произведене) одбојности према средишту словенске културе у Русији — ћирилица је само на папиру основно српско писмо (Република Србија има Закон о службеној употреби језика и писма). Напротив, данас је преовлађујуће писмо Срба наведена југословенска латиница.⁵

Ствар је сложена и кад је типографски изглед у питању. Услед привредног расула, па тиме и немогућности финансирања чак ни битних културних пројеката, у времену наглог развоја рачунарске технологије, савремена припрема за штампу ослања се искључиво на понуђена типографска решења. У масовној употреби понуђена решења су прве половине деведесетих искључиво стизала од америчке фирме Микрософт, а и то тек преко посредника, не у изворном облику. С друге стране, управо те године

¹ Ово време можемо апроксимовати на шести или седми век после Христа, а простор на данашњу црногорску обалу, Дубровник и Далмацију. За разлику од овог „несумњивог“, постоје индиције да су на просторима Подунавља и северне Италије Словени познавали писмо и око хиљаду година пре Христа, па и раније.

² Јединствен је стога што су племена која су населила тај простор етнички била српска, док се готово сви потомци тих људи данас осећају Хрватима.

³ Колико се језик лингвистички може називати „српскохрватским“ или „српско–хрватским“, толико се латиница не може, јер је у целости и део словеначке културе, иако словеначки језик не користи сва слова те латинице.

⁴ Идеја о стварању југословенске нације на језику се путем политичких притисака остваривала неколиким тежњама: (1) лексичким уједначавањем (познати пример је „тачка и зарез“ насупрот хрватској „точка и зарезу“ а српској „тачки и запети“); (2) писање еквицом; (3) уједначавање изговора фонема „ч“ са „ћ“ и „џ“ са „ђ“; и (4) коришћења искључиво латиничног писма.

⁵ Српски језик је званичан у три државна ентитета (не рачунајући СР Југославију као заједничку државу Србије и Црне Горе). Иако је на званичним местима ћирилица често и једино писмо, због потребе за очувањем националног идентитета, само се у Републици Српској ћирилица знатно више користи од латинице. У Црној Гори и Србији ствар је готово обратна.

одликује готово истоветно привредно расуло и у Русији, па оне пролазе у ваљаном типографском изгледу латинице (коју и није било потребно дорађивати графички) и катастрофалном изгледу ћирилице. Такво стање довело је до оправданог става многих аутора изван лингвистичких и типографских токова — да им је боље да дела објављују латиницом, јер знају да ће им књиге и типографски ваљано изгледати.

У другој половини деведесетих година привредно стање у Русији се сређује, па се руска варијанта ћирилице типографски устаљује. Нажалост, овај се напредак претвара у увођење додатног хаоса у нашу типографију. Не само да се масовно појављују руске варијанте курзивних малих слова „г“, „п“ и „т“, већ се стварају стилска одступања међу малим словима у нормалу и курзиву „л“ и „љ“, односно „л“, „љ“ и „м“. Питање стилске недоследности слова „к“ и „ж“ у нормалу су решена (о чему ће бити речи нешто ниже), али се унела додатна забуна у то шта је сада „правилније“. Низови слова у масовној употреби с најправилнијим типографским решењима имају, ипак, једну ману: поседују „плеонастично“ мало курзивно „г“.

С краја деведесетих година почели су јасније да се артикулишу захтеви за нормирањем српске ћирилице. Четири су кључна момента у томе: (1) књига Митра Пешикана *Наша азбука и њене норме* из 1993. године, (2) оснивање Одбора за стандардизацију српског језика при САНУ 1997. године, (3) приступање решавању ових проблема у Друштву информатичара Србије у сарадњи с Одбором, што се током 1998. године остварило у виду предлога који је почетком марта 1999. године објавио Савезни завод за стандардизацију и (4) оснивањем Радне групе за српско писмо на Интернету децембра 1999. године (<http://www.srpsko-pismo.org/>), која електронским путем окупља стручњаке из информатике и лингвистике у циљу омогућавања несметане ћириличне електронске комуникације међу говорницима српског језика.⁶

Данас, на почетку 2001. године, стање је боље само уколико што постоје сви неопходни предуслови за приступање озбиљном раду у овој области: (1) артикулисао се одређени број стручних људи који су спремни да раде на овом питању, (2) постоји нормативна окосница, пре свега у делу Митра Пешикана, (3) постоји неопходна правна подлога у виду Одбора за стандардизацију и (4) мада је рано о томе са сигурношћу говорити — и економско и политичко стање у држави се поправљају. Насупрот томе, човек без типографских склоности данас још увек нема могућност да пише ћирилицом правилно (осим ако му се дело не штампа каквом старијом технологијом). Човек који није довољно упућен у рачунаре не може увек ни да прочита неки ћирилички текст на рачунару, нити може бити уверен да ће његова писанија бити ваљано прочитана.⁷

Минимум стилских обележја српске ћирилице

До стварања једног сложенијег и формалнијег рада о минимуму функционалних и стилских обележја писама у информационој и културној употреби Срба, на овом месту ћу покушати да дам преглед најбитнијих спорних особености српске ћирилице.

Најважнија тековина структуралне лингвистике — појам о систему — мора се применити и на писмо. Како је писмо у току свог вишемиленијског развоја добило статус не само изведеног него и самосвојног средства општења — и оно је добило свој стандард, своја функционална и стилска обележја.

Писмо не може потпуно економично да преноси све видове говора, као што и говор мора да се поштапа називима из писма (нпр.: „То је под наводницима било добро.“). Као што писмо не може да пренесе насмејан или уплакан говор а да не каже експлицитно да се неко смеје или да плаче, тако ни говор не може да пренесе курзив или масна слова. То говори да се мора размишљати о посебним функционалним (лингвистичким) обележјима писма.

Две су основне парадигме функционалних обележја писма. Једна је парадигма типова сетова писама, где један сет повезује одређен стил линија и других особености слова и знакова. Тако се разликују типови „Времена“ („Times“), „Хелветика“, „Куријер“ итд. Друга је парадигма унутартипска. Слова једног сета имају своје варијанте према величини, дебљини, укошености и повезаности. По **величини**, слова се класификују према броју тачака. У традиционалној штампи постојали су сетови слова неколико величина. Данас, захваљујући векторским форматима сетова слова (фонтова) могуће је штампати готово ма којом величином: мале величине су ограничене квалитетом излазног уређаја

⁶ Први облик овог рада је писан средином децембра 2000. године. У односу на то време, овоме бих додао и стварање Српског рачунарског друштва „Уликс“ које је у току. Друштво има за циљ стварање конкретних програмских решења и превода у сврху прилагођавања рачунара српским потребама, а окупља лингвисте, програмере и остале непосредно заинтересоване за рад на тим питањима. Рад је усмерен пре свега према прилагођавању слободног софтвера српском језику.

⁷ Како по занимању нисам типограф, а рачунаре доживљавам као оруђе које служи мени (а не ја њему), па тиме тежим ка што бржем писању, први примерак овог рада (а сигурно и прву десетину примерака, бар) штампам типографски руском ћирилицом. Истина је да ће исти рад, готово електронски неизмењена облика, бити могуће штампати правилно, али данас такво решење не постоји.

(штампача, екрана, штампарских машина), а велике величином хартије. Уобичајена величина текста у штампи јесте десет тачака (том величином су исписани и ови редови). По дебљини слова могу бити танка, средња и дебела (с могућим међустањима), по **укошености** неукошена и укошена (са могућим међустањима и по **повезаности** „штампана“⁸, „полуписана“ и „писана“⁹). Ове се четири парадигме могу укрштати, мада је плеонастично мешати укошен с полуписаним и писаним стилем.

О изгледу слова у данашњој масовној употреби

Нису сва слова српске ћирилице у данашњој масовној употреби подједнако (не)стабилне граfiје. Она која у потпуности следе латинички или грчки изглед нису проблематична. Она која следе у потпуности руску ћирилицу углавном су стабилна мада често системски лоше уклопљена, док највећа одступања показују посебно наша слова, па имала она нарочите граfiјске појединости или не.

Слова се по степену стабилности могу поделити на она са стабилном граfiјом и она с мање или више нестабилном граfiјом.

Услед тога што су слова по својој граfiји мање сложена, што имају упориште у изгледу латиничких слова и/или грчких слова, или зато што у штампи нису често употребљавана, може се утврдити да су следећа слова стабилне граfiје: (1) сва писана (букварска) слова; (2) у свим осталим облицима слова: а, е, и, ј, о, р, с, х, ц, ч, ш; (3) у нормалу стабилна су слова: г, м, н, п, т, ф, ц.

Другу групу чине слова углавном стабилне граfiје која повезује полукружна линија, тзв. трбушчић: б, в, з, ђ, ћ, (љ), (њ), (р).

Трећу групу чине слова која углавном имају упориште у руској ћирилици. Она чине скуп стилски повезаних слова нестабилне (системске) граfiје: д, ж, к, л, љ, м, у.

Четврту групу чине мала курзивна и букварска слова „г“, „п“ и „т“, која немају системског упоришта ни у латиници ни у руској ћирилици.

Пету групу чине посебне стилске недоследности, које углавном нису глобалне природе, а чине их слова из претходних група. У питању су мала слова у нормалу и курзиву „б“, „д“, „ж“, „у“ и „ц“.

Скуп слова углавном стабилне граfiје: б, в, з, ђ, ћ, (љ), (њ), (р), (ч)

У групу углавном стабилних слова која чине стилски скуп налазе се: (1) велико слово „б“ у нормалу, (2) велико, мало и мало полуписано слово „в“, (3) све три преостале варијанте слова „з“, (4) слова „ђ“, (5) слова „ћ“ и (6) слова „ч“. Њих одликују већа доња и мања горња полукружна линија. Заједно са њима, придружени су им чланови и слова „љ“ и „њ“ у свим облицима и верзално слово „р“ у нормалу.

Већа полукружна црта истоветна је код великих слова „б“ (доле), „в“ (доле), „з“ (доле), „љ“ (доле), „њ“ (доле) и „р“ (горе). Код слова „ћ“ је у једној варијанти иста црта, а у другој следи отвореност (и делимичну сличност линије) слова „ћ“, а доњу црту повија тек испод основне линије. Слова „ћ“ и „ч“ не прате у потпуности ову линију, па ипак спадају у ову врсту слова. Изузев за слова „б“ и „р“ исто важи и за мала слова. Наравно, велике полукружне црте мање су у мањих него у великих слова.

Слова „в“ и „з“¹⁰ у сва три облика имају једнаке мале горње полукружне црте.

Скуп стилски повезаних слова нестабилне системске граfiје: д, ж, к, л, љ, м, у

Изузев слова „м“ што је битно само у курзиву, и то само у односу на мала курзивна слова „л“ и „љ“, ово је група наших најспорнијих слова, која ни у старијој типографији нису увек задовољавајуће решена.

⁸ Под наводницима су „штампана“ зато што је свако одштампано слово штампано, док се у обичном говору појам „штампаног“ супротставља појму „писаног“ текста према повезаности слова. Називи су: нормал за „штампана“, курзив за „полуписана“ и букварска за „писана“. Ипак, кроз рад ћу користити називе из обичног говора.

⁹ Посебан стилски облик унутартипске парадигме представљају само варијанте повезаности. Иако један типограф мора посебно водити рачуна и о томе каква ће бити масна, односно укошена слова, то се чини искључиво у сврху бољег изгледа одређеног сета. Не рачунајући тај рад, удебљање и укошење представљају једноставну математичку функцију на нормално слово. С друге стране, полуписана (курзивна) и писана (букварска) слова имају различите линије, па су готово независни сетови унутар једног сета слова. Треба, такође, знати да су велика полуписана слова само укошена нормална, док су мала некад изгледом ближа нормалу а некад букварском стилу.

¹⁰ Ћирилично слово „з“ поседује проблем истоветности са бројем „з“. Ово би свакако граfiјски требало решити, а можда је најбоље тако што би се за слово „з“ прихватило решење са равном цртом горе.

Главна особеност ове групе слова (изузев слова „м“) јесте капљица коју треба или да имају сва слова у нормалу или ниједно.¹¹ При томе, непостојање капљице повлачи и другачији изглед слова „д“, „л“ и „љ“, који је у том случају заснован на грчком слову ламбда („Λ“), за разлику од уобичајеног заснивања изгледа на ћириличком слову „П“.

Неправилно је, дакле, мешати ова два начина, као што се то на један начин чини стварањем сет слова са капљицама на словима „к“, „л“, „љ“ и „у“, а не и на словима „ж“ и „к“, односно на други начин готово супротно.

(Не)постојање капљице повлачи ланчано, преко одређеног изгледа, и изглед малих полуписаних слова „л“, „љ“ и „м“. Тако, у случају да капљица постоји (па је изглед слова „д“, „л“ и „љ“ заснован на слову „п“), мала курзивна слова „л“, „љ“ и „м“ имају незаобљене прелазе међу својим косим линијама, што се може остварити у две варијанте. У случају да капљице не постоје (па су слова „д“, „л“ и „љ“ заснована на изгледу грчке ламбде), прелази међу косим линијама су заобљени.

Неправилно је обратном везивати нормал и курзив, као и мешање две варијанте у случају када капљица постоји.

Неусаглашена ћирилска графика: мала курзивна слова „г“, „п“ и „т“.

На слици 1 приказане су варијанте малих курзивних слова „г“, „п“ и „т“. Укратко, постоје три врсте ових слова и једна неправилна варијанта.

Прва варијанта је руска (дакле, она што следи латинична курзивна мала слова „н“ и „м“, док је мало курзивно „г“ специфично). Мала полуписана слова „п“ и „т“, у односу на иста та писана, букварска руска слова, иста су као мала полуписана латинична слова „н“ и „м“.

Друга варијанта је предлог Митра Пешикана за уједначен изглед ових слова у ћирилицама. Тако би писана слова изгледала као и до сада, док би штампарски курзив био уједначен. Битно је уочити да та слова јесу заснована на грчким словима пи („π“) и тау („τ“), али да ћирилично „п“ и „т“ нису идентична овим грчким словима. Јер, особеност грчких малих слова јесте да имају функцију писаних слова (грчка азбука не познаје букварски, писани стил), па су више заобљена него што је курзив.

Претпоследња је наша традиционална варијанта у односу на писана слова. Битно је приметити разлику између овог малог курзивног „г“ и наредног. Наше мало полуписано „г“ има за основу латиничко курзивно „и“, односно малу грчку јоту („ι“).

О посебним стилским недоследностима

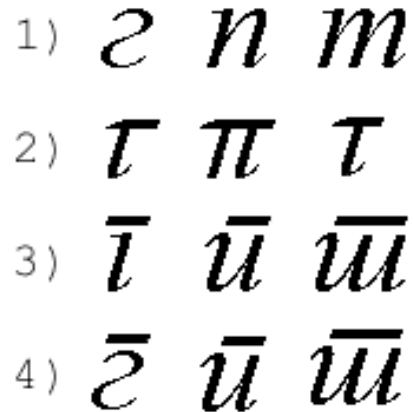
Ове се недоследности углавном јављају услед неразумевања разлике између полуписане и писане варијанте.

Испрва, писало се искључиво великим словима. Потом су створена, ради бржег писања, мала, док су велика остала за сврху посебног наглашавања (почетно слово реченице, личног имена, имена уопште у немачком језику нпр., наслова итд.). Пошто је економија писања захтевала природнији и повезанији облик, настали су брзописи, тј. писана слова (односно, букварски стил).

Претходно је „природна“ историја стварања писаног стила. Полуписани стил није настао из потребе за бржим писањем, већ као посебан стилски облик за истицање текста. Тако он има црте писане варијанте, али и доследност традицији и систему штамpane.

У том смислу треба гледати и двојне варијанте неких малих штампаних и полуписаних слова: „б“, „д“, „ж“, „у“ и „ц“.

Варијанта малог (и штампаног и полуписаног) слова „б“ са горњом кривом цртом која почиње из средишта круга (тј. са десне стране) настала је под утицајем писаног стила. Правилна је варијанта у којој та линија почиње са леве стране. Слично важи и за остала наведена мала слова. Мало слово „д“ се везује



Слика 1: Мала курзивна слова „г“, „п“ и „т“: 1) руска варијанта, 2) варијанта коју је предложио Митар Пешикан, 3) традиционална варијанта и 4) варијанта са „плеонастичним“ словом „г“, која се често среће у нашим сетовима слова. Засновано на сету слова Times New Roman, произвођач Монотип корпорација.

¹¹ Слово „д“ нема изражену капљицу, али у нормалу потпуно прати слово „л“, па чак има и криву црту која капљицу наговештава. Мало курзивно „д“ другачијег је постанка (у вези је с малом грчком делтом [„δ“]), па не спада у ову групу.

и у нормалу и у курзиву за грчку малу делту („δ“), а не за свој писани облик и мало латинично „г“. Слово „у“ има своје посебне облике о којима је раније било речи, а не везује се за свој писани облик, тј. латиничко ипсилон. Слово „ц“ се, такође, не везује за свој писани облик. Слово „ж“ се једино у букварском, писаном стилу не везује за слово „к“. У свим осталим облицима слово „ж“ изгледа као једно обично слово „к“ и једно пресликано у огледалу.

Неке напомене о сетовима слова

Сетови слова на рачунару зову се фонтовима. Унутрашње особености фонта зависе од оперативног система и програма који их користи. Ипак, по одређеним функционалним (значи, не уопште графичким) особинама фонтови се могу груписати. Прву осу чини опозиција тачкасти—векторски, а другу непропорционални—пропорционални.

Примери за тачкасте фонтове јесу они које користе старији оперативни системи и режим рада под терминалом (DOS, Линукс конзола, конзола Микрософтових Прозора итд.), као и системски фонтови окружења Икс прозор под различитим оперативним системима који личе на Јуникс. Основна њихова особина јесте да су састављени од одређеног броја тачака и да се не дају повећавати или смањивати.

Векторски фонтови су они који се могу математичком функцијом увећавати без губитка на изгледу слова. За разлику од тачкастих, они су састављени од линија, па је могуће такву математичку функцију извести. Примери за те фонтове су „тип 1“ произвођача Адоб, „тачни тип“ произвођача Микрософт и Метафонт пакет Доналда Кнута. Проблем са прве две врсте је у томе што не поштују типографске детаље,¹² док Метафонт пакет иде искључиво уз TeX програм који није широко прихваћен (за разлику од Микрософтове Речи, коју готово сваки власник рачунара има).

И тачкасти и векторски фонтови могу бити пропорционални или непропорционални. Ова особина односи се на ширину слова. Код непропорционалних фонтова слова су увек исте ширине, док се код пропорционалних ширина природно мења (тако је, нпр., слово „ж“ шире од слова „к“). Непропорционални су тачкасти фонтови за терминални режим рада, док се они које користи Икс прозор деле на пропорционалне (Times, Helvetica) и непропорционалне (Fixed). Непропорционални векторски фонтови су нпр. Courier New Монотип корпорације и Lucida Console произвођача Bigelow & Holmes Inc., док су пропорционални нпр. Times New Roman, Arial (Монотип корпорација) и Verdana (Микрософт).

О КОДНИМ РАСПОРЕДИМА И ТАСТАТУРАМА

Један друштвоњак се може наћи у поприличној збрци кад дође у додир с принципима представљања знакова и писама на рачунару. Још када покуша да то некако формално сложи, настаје прави хаос.

Отприлике, треба разликовати три појма: (1) **писмо**, у филолошком значењу те речи, (2) **знак (код) с кодним распоредом**, што представља одређени функционални скуп договорених знакова, без обзира на изглед и функцију датих знакова „ван рачунара“ и (3) **изглед знакова**, тј. **изглед писама (фонг)**, што представља конкретну реализацију неког кодног распореда и његову везу с писмом у основном значењу.

Преглед постојећих кодних распореда

Кодни распореди

Кодни распоред је по својој бити ствар стручног и политичког договора одређене групе људи. Тиме представља **норму** по којој људи могу општити путем рачунара.

Технички најважнија разлика између кодних распореда јесте у колико бита неки од њих стаје (тј. колико знакова има у одређеном кодном распореду, гледано путем експонента броја два). Тако су први кодни распореди били седмобитни, што значи да су садржали 128 знакова (од којих нису сви видљиви). Данашњи кодни распореди махом су осмобитни (значи, 256 знакова), док полако у употребу улазе и шеснаестобитни, односно кодни распореди са више од шеснаест бита, тј. вишебајтни (један бајт је осам бита).

¹² Ширина повлаке, дуге црте („—“), на малим величинама слова (око десет тачака) износи две ширине броја нула („0“), док се увећавањем слова сужава да би на великим (око педесет тачака и више) износила ширину једне нуле. „тип 1“ и „тачни тип“ не поштују ову особеност.

Историја савремених кодних распореда почиње 1968. године, када је Национални институт за стандарде Сједињених Америчких Држава (ANSI) донео *привремени* кодни стандард назван Амерички стандардни код за размену информација (ASCII, /'aski:/, /аски/). Тај је стандард, ето, остао привремен до данас. Њега је 1972. године ISO прогласио званичним међународним стандардом (ISO 646), а осамдесетих година су почели су се појављивати национални стандарди засновани на ASCII кодном распореду, који место ређих ASCII знакова поседују специфичне знаке датих националних писама. Међу двадесетак седмобитних националних кодних стандарда налазе се и три наша: JUS I.B1.001 (српска, хрватска и словеначка латиница), JUS I.B1.002 (српска ћирилица) и JUS I.B1.003 (македонска ћирилица). Прва два стандарда су и данас у масовној употреби у нашој земљи, а познатија су као YUSCII /јуски/ (оба) или YUSCII за латинички, а SRPSCII /српски/ за ћирилички.

Треба рећи да данас првих седам бита у свим кодним распоредима представља управо ASCII кодна страна. Једини шире прихваћени кодни стандард који не поштује ASCII јесте IBM-ов EBCDIC (Проширени бинарно кодирани децимални информациони код), који је осмобитног карактера.

Следеће раздобље у развоју кодних распореда представљају осмобитни кодни распореди. Захваљујући спорости међународних организација, а из потребе за проширивањем скупа знакова, почетком осамдесетих година створен је прави хаос у кодним распоредима. Микрософт и IBM су направили свој први стандард: CP 437, који је по садржају знакова налик данас познатијим страницама CP 850, CP 1250 и ISO 8859–1. Следиле су: латиничка за централну и источну Европу (CP 852), једна ћириличка (CP 855), друга (CP 866), па трећа (CP 878), при чему је само прва код нас употребљива; па за грчки (CP 737), португалски (CP 860), турски (CP 857)... Уз неке измене, ове странице су, потом, прилагођене Микрософтовим Прозорима; осмице и седмице на почетку заменила је дванаестица... Ћирилична страна, која се и данас највише користи у свету, јесте CP 1251.

Средином осамдесетих година Европско удружење произвођача рачунара (ECMA) направило је скуп осмобитних кодних страна данас познатих као ISO 8859 (иза чега следи цртица, па број од 0 до 15), што је озваничила ISO. Тако је словенска ћириличка страна добила број 5, а латиничка за средњу и источну Европу број 2.

На основу стандарда ISO 8859, организација Јуникод конзорцијум 1991. године озваничила је пројекат стварања једног кодног распореда који би садржао слова/знакове свих писама и свих стручних скупова знакова. Исте године ISO је озваничила овај кодни стандард као ISO/IEC 10646 (10000 + 646, званични број ASCII стандарда). За такав пројекат је испрва одређено 16 бита, тј. два бајта (65536 знакова), али се данас ради о три, односно четири бајта. Овај стандард, иако по својој природи решава све проблеме многојезичног света, није у потпуности прихваћен — што због своје гломазности, што због потребе да се постојећи програми из сржи промене.

Српски језик и кодне стране

У односу на српски језик, стање би се могло свести на следеће:

Највећи проблем српског језика у овом питању јесте коришћење седмобитних JUS I.B1.001 и JUS I.B1.002 у корисничким програмима. IBM-ов EBCDIC кодни стандард данас готово нико не користи, сваки други подразумева ASCII страну у првих седам бита (мада, и да је ко још користи, у овом случају не би било од неке помоћи). Коришћење фонтова у овим кодним распоредима представља солидан проблем за све могуће надоградње.

Осмобитни кодни распореди решавају само најелементарније потребе наше културе. Ови кодни распореди не задовољавају ни минималне штампарске потребе: није могуће написати наше акценатске знаке. За ове потребе се користе посебни фонтови, какве је направио Владо Ђукановић за електронско уношење Речника САНУ.

Док се примена Јуникод стандарда не омасови, текстове на српском језику требало би писати и чувати у ћириличком облику, а у зависности од потребе преводити али не у латинички. Јер, латиница није, за разлику од ћирилице, једнозначна: слова 'љ', 'њ' и 'џ' могу бити исто тако и комбинације слова 'л'+'ј', 'н'+'ј' и 'д'+'ж'. Јуникод теоријски подржава диграмске карактере, али их у пракси једино фонт Lucida Sans Unicode садржи.

Поред тога, транслитератори могу успешно да преводити текст са енглеским и српским само ако је текст написан ћирилицом. Ако се покуша превести наша латиница уз какву другу латиницу, и туђа ће се латиница превести у ћирилицу.

Под Микрософтовим Прозорима могуће је користити ваљано само CP 1250 и CP 1251 кодни распоред. Под Линуксом је могуће користити и ове, али је подршка за ISO 8859–2 и ISO 8859–5 кодне распореди боља.

Коришћењем транслитератора могућ је успешан и ефикасан рад на српском језику помоћу рачунара. За сада не постоји јединствено решење и мора се прибегавати различитим помоћним решењима. Ове странице на Растку и имају за циљ да вас у та решења упуте.

Проблеми пресловљавања и њихова решења

Једна од значајних предности рачунара у односу на традиционалну штампу и писаћу машину, посебно везана за двоазбучни српски језик, јесте могућност транслитерације, пресловљавања, из једног писма у друго. За потребу једноазбучне штампе више није неопходно прекуцавање текста, већ једноставно извршење програма,¹³ а понекад и још једноставнија замена фоната.¹⁴

Ово је, ипак, решење за ограничену количину текста. Проблеми који се јављају у процесу пресловљавања сложени су до те мере, да их рачунари у процесу са високим степеном аутоматизације могу само делимично решити.

Проблеми функционалне и стилске графиције српских азбука

Ови се проблеми могу поделити на четири групе (преко оних описаних у првом поглављу): (1) проблем неједнозначности традиционалне¹⁵ латинице; (2) проблем општости латиничког писма; (3) проблем односа верзала и капитале; и (4) проблем дијакритика у Даничићевој латиници.

1) Проблем с највећим последицама по ваљаност пресловљавања представља неједнозначност традиционалне латинице. Комбинације латиничких слова „л“ + „ј“, „н“ + „ј“ и „д“ + „ж“ не морају увек одговарати фонемама „љ“, „њ“ и „џ“.

Ствар је најједноставнија, мада не и потпуно решива, када је у питању комбинација „л“ + „ј“. И старије и ијекавско јотовање дало је на споју ова два гласника палатал „љ“. Једино у посебним случајевима транскрипције нпр. словеначких имена (као што је, нпр., име града Љубљане /лјублјана/) или ретких личних имена, ова се комбинација јавља двофонемски. Опет, на већим количинама текста није упутно вршити потпуно пресловљавање.

И поред тога што је за српски језик неприродан спој гласова „н“ и „ј“, речи латинског порекла као што су „конјункција“, „конјугација“ и сл. у хиперкоректном изговору и Правопису изричито се захтева изговор/испис у два гласа, односно два слова — па био испис латиницом или ћирилицом. (Свакако, и овај је спој у оба јотовања дао палатал, па изворно српске речи са двофонемским обликом не постоје; чиме број проблематичних речи остаје на минимуму.) У случају да је текст поетског карактера углавном није потребно водити рачуна о томе налазе ли се на том месту две фонеме или не. У случају да се ради чак и о новинском језику — неопходно је.

Најнезгоднија комбинација, али, на срећу, не одвећ честа, јесте комбинација слова „д“ и „ж“. Главна тешкоћа се састоји у томе што је комбинација фонема „д“ и „ж“ знатно чешћа од фонеме „џ“ у српском језику. Фонема „џ“ се изворно у српском језику јавља искључиво као алофонска реализација фонема „џ“ („убеник“, „отаџбина“)¹⁶, док је своје место нашла захваљујући данас готово изумрлим турцизмима, а потпору у систему (као што је могућа нпр. звучне варијанте фонема „џ“ и „х“) и именима енглеског порекла (која, опет, системски нису значајна, за разлику од појаве дифтонга¹⁷ под утицајем енглеског језика, такође). Сваки се текст мора проверити за овај случај.

¹³ У ову сврху могу се, на различите начине, употребити програми „unistd“ (који се може наћи на Мрежној адреси <http://www.srpsko-pismo.org/>) и „YUCC“ (који се може наћи на Мрежној адреси http://members.xoom.com/_XOOM/graft_art/freetools/msoffice/yucc.html). Програм „YUCC“ користи се из Микрософтове Речи. Много већи избор и програмабилност доноси програм „unistd“ Владе Кешеља (његова електронска адреса је vkeselj@uwaterloo.ca), с тим што не може да ради са листовима Микрософтове речи, већ мења глобално листове написане у обичном текстуалном облику (са наставком „.txt“) или у хипертекст језику за означавање (са наставком „.html“). Ипак, програм „unistd“ се и даље развија. Треба поменути и да Српско рачунарско друштво „Уликс“, у сарадњи са пројектом Растко (<http://www.rastko.org.yu/>), развија сложенији програм Методије, за употребу у окружењу Микрософтове Речи, који поседује елементе претраге по језицима за означавање.

¹⁴ Ћирилички кодни распоред JUS I.B1.003-S (тј. ISO IR-146) и латинички JUS I.B1.002 (тј. ISO IR-141) готово су у потпуности транслитерарни са становишта српског језика. Тако је могуће, уз извесне преправке, само заменити фонат у једном кодном распореду за фонат у другом и добити прелаз из једног писма у друго.

¹⁵ Назив „традиционална латиница“ супротстављам називу „Даничићева латиница“.

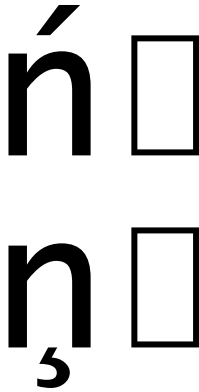
¹⁶ У том смислу њен се легитимитет своди на легитимитет офонемљавања алофонске реализације фонеме „џ“ („џ“, „с“); ако се потоња транскрибује историјски са „з“, односно фонетски са „џ“, то би исто и за фонему „џ“ могло да важи (са „ж“, односно „џ“).

¹⁷ У фонолошки систем српског језика лагано су се увукли дифтонзи под утицајем латинског, грчког и енглеског језика, што тренутно стање чини сличним оним када су из словенског, односно српског нестајали назални самогласници, тј. глас „јат“. Латински и грчки језик су утицали готово неприметно путем уношења већег броја речи са „ј“ у крајњем неслоготворном делу слога, углавном уз самогласник „а“ — чинећи дифтонг /ај/. (Историјски, српски језик познаје фонему „ј“ искључиво као почетну у слогу, иза које следи самогласник, што доказује и првобитна графика са „прејотованим“ самогласницима и без самосталног знака.

На срећу, учесталост ових скупова и фонема обрнуто је сразмерна учесталости двофонемских вредности: „љ“ је најчешће (историјски, чешће је „н“ било назализатор претходног самогласника него што је „л“ прелазило у „у“, односно „о“), док је учесталост латиничког споја „д“ + „ж“ тек нешто већа од релативно ретке појаве изведеница са префиксима „пред-“, „под-“, „над-“ и сл. и речи која почиње фонемом „ж“.¹⁸

2) Проблем општости латиничког писма¹⁹ показује се, на другачији начин, истолико сложен проблем колико претходни. Ма којим кодним распоредом и ма којом варијантом јатинице (једнозначном Даничићевом или неједнозначном традиционалном) 22 знака од 27, односно 30 наших — представљају општи латинички фонд, па ће рачунар пресловити, поред текста на српском, и текст написан на неком другом језику.

С формалног становишта, унутар једноставног текстуалног записа, овај је проблем нерешив. Једина могућност у овом случају (за једноставни текстуални запис) јесте да се текст на српском пише ћирилицом, а да се за енглески језик користи латиница. Тиме је једнозначно решено шта је написано на српском а шта није. А за потребе латиничне штампе могуће је пресловити ћирилицу у латиницу (в. слику 21, пресловљен латинички текст).



Слика 2: Даничићева и летонска варијанта

3) Проблем односа верзала и капитале јесте један ситнији и решив проблем пресловљавања ћирилице у латиницу, као и самог латиничког писања.

Неопходно је разликовати два појма: (1) верзал, велико слово и (2) капиталу, насловна велика слова. Углавном, ови проблеми није видни, па их није ни неопходно разликовати, осим ако дотични текст није намењен (формалној) лингвистичкој обради.

Проблем се јавља код латиничких диграма („lj“, „lj“ и „dž“). Једини правилни облик верзала ових диграма јесте онај у коме је прво слово велико, а друго мало. Само у случају када је у питању насловни текст састављен од великих слова, правилно је да и друго слово буде велико.

Када је пресловљавање у питању, потребно је свако од датих ћириличких слова пресловити у верзал (не у капиталу). С друге стране, како латиничке, тако и ћириличне наслове треба писати уобичајено, с великим словима само где им је место по Правопису, док насловна велика слова треба посебно одредити у језику за означавање, односно у програму за унос и обраду текста.

4) Може се учинити да је решење у томе да с традиционалне латинице пређемо на Даничићеву. Али, и она поседује један стилски проблем: изнад слова „њ“ није могуће поставити акценат²⁰ (в. слику 2) који ће стилски прихватљиво изгледати.

Решење које захтева измену графиције могло би да буде преузимање варијанте из летонске азбуке (слика 3), чиме би се скуп нама потребних знакова ограничио на тзв. централно и источноевропски латинички скуп подржан у свим скуповима осмобитних кодних распореда (за разлику од досадашње потребе за скупо знакова „Latin Extended B“ из Јуникода).

Постојање два слова за фонему „и“ друге је природе, идентичне постојању два слова за фонему „о“: покушају потпуне транспарентности грчке и словенске азбуке, односно заблуде о транспарентности фонолошких система ова два тадашња језика.

Иза тога следиле су још драстичније појаве: одустајање од монофтонгне транскрипције енглеског дифтонга /ei/ („тролејбус“ према „Шекспир“) и форсирање хиперкоректне суфиксације придева са /ни/ дифтонгом („индијски“ према Белићевом „индиски“). Колико су махом експресивни дифтонзи /ai/ и /oi/ природни српском језику (за експресивну употребу) због велике разлике у артикулацији суседних фонема (па је питање постоје ли ту дифтонзи или је у питању спој самогласника и сугласника), толико приближена артикулација ствара прави дифтонг. (А приметан је и природни отпор говорника у честом монофтонгом изговору.) Такво је стање довело до данас све чешћег дифтонга /ou/ (најпознатији пример је реч „Виндоуз“). Јавно, путем овакве транскрипције, и скривено, путем изворног исписа енглеских речи, дифтонзим се даје употребни легитимитет, а тиме се непосредно руши фонолошки систем српског језика.

¹⁸ Ствари се изузетно компликују не само за пресловљавање у ћирилицу, већ и за свођење текста на стандардну латиницу, у случају да се за фонему /њ/ користи скуп „д“ + „ј“. Фонема /њ/ није одвећ честа (мада знатно чешћа од фонема /ц/ и /ф/), али је скуп „д“ + „ј“ готово једнако чест ако не и чешћи. Преко тога, постоје и нерешиве хомографије: „дјетић“ и „ђетић“ нпр., где постоји и секундарна семантичка разлика.

¹⁹ Проблем општости латиничког писма могао би, не без ироније, да се сведе на констатацију „и енглески језик се пише латиницом!“. И ћирилицом пишу други народи другим језицима, али, уместо задржавања оригинала ми (све ређе) транскрибујемо, односно (све чешће) узимамо енглеску транскрипцију (или транслитерацију).

²⁰ Фонема (гласник) „њ“ припада скупу дистрибутивно слоготворних фонема. За разлику од слоготворног „р“, које је историјски слоготворно, па и слоготворних „л“ и „н“, који су ушли у наш језик путем туђица; гласници „љ“ и „њ“ потенцијално су слоготворни. Оказионализми „шкљ“ и „жњ“ су пример за тај „увозни“ потенцијал.

Тастатуре

Проблем тастатура раслојава се на две групе.

Прву чини изглед саме тастатуре као уређаја. У ту сврху је неопходно наметнути увозницима и могућим произвођачима правна решења по којима је обавезно на нашем тржишту пласирати тастатуре које су у складу с стандардом.²¹



Слика 3: Летонски облик као могућа замена за Даничићев.

Увођење писаћих

машина у нашу земљу догодило се у време везаности наше културе за немачку (аустројску, пре свега). Што због тога што је готово све до краја 19. века центар српске културе био у Војводини, што због везаности самог Вука Караџића за Беч — основна техничка решења преузета су од Немаца.

Тако се места слова „з“ и „ипсилон“ на тастатурама мењају, као и одређени број „виших“ знакова у горњем реду тастатуре. Свакако, ту је и 30, односно 27 наших знакова наспрам 26 енглеских.

	наводници:	полунаводници:
а)	"текст"	'текст'
б)	“текст”	‘текст’
в)	„текст”	,текст’
г)	„текст”	,текст’
д)	»текст«	>текст<
ђ)	«текст»	<текст>

Слика 4: Различити типови наводника и полунаводника: 1) наводници: (а) једнообразни, (б) облик са горњим наводницима, (в) први наш облик са изводником, (г) други наш облик са обрнутим изводником, (д) облик веће–мање, (ђ) облик мање–веће; 2) полунаводници: (а) једнообразни, (б) облик са горњим полунаводницима, (в) први наш облик са полуизводником, (г) други наш облик са обрнутим полуизводником, (д) облик веће–мање, (ђ) облик мање–веће.

Типичан пример за нашу традиционалну (званичну) тастатуру јесте IBM-ов уређај, односно Микрософтово програмско решење за Прозоре 95, 98, 2000, NT 4 и Миленијум. Проблем са овом тастатуром јесте — што то није српска тастатура, већ општа ћириличка тастатура СФРЈ. Притиском на тастере „“ и „”“, односно „“ и „”“ добијају се македонска слова „ѓ“ и „ќ“, док на месту ипсилона на латиничкој, на ћириличкој постоји македонско „дз“ („s“). Како више не живимо у једној земљи, нема ни потребе да подржавамо македонски језик (који има своју засебну тастатуру).

Друго решење, које сам направио према предлогу Митра Пешикана и применио под

оперативним системом Линукс, на месту латиничког ипсилона доноси ћирилично „ђ“, док „ж“ смешта на место са старијих писаћих машина.

Код Микрософтових тастатура постоји још и проблем наводника.

Постоји пет типова наводника (слика 4) с којима се данас срећемо и толико типова полунаводника. По три су заснована на изгледу запета а по два на изгледу знакова веће и мање. Правилни облици наших наводника јесу под (в) и (г). Једино у случају да је скучен број знакова (као што је случај у седмобитним и осмобитним кодним распоредима), допуштено је користити једнообразне (а) наводнике.

У складу са стилским уклапањем и читљивошћу, за наводнике се препоручује²² варијанта под (г), док се за полунаводнике препоручује једнообразни облик (а). Нечитљивост у случају да и полунаводници имају облик под (г) може се приметити приликом спајања полунаводника с наводницима.

Кад је унос текста на рачунару у питању, срећемо се са три облика наводника (а, б и д) и два полунаводника (а и б), при чему је само један облик полунаводника правилан (а), па се може користити и

²¹ Пример за овакво решење јесте понашање компаније IBM, која у потпуности поштује потписан уговор с нашом Владом, па и за све претходне године на наше тржиште доноси тастатуре које су у складу с нашом нормом.

²² У питању је усмена препорука Миљка Ковачевића. Овај проблем је код нас званично и даље отворен, па се очекује дефинитивно нормативно решење.

у штампи. То не значи да знакови нама потребни не постоје у кодним распоредима, већ значи само да нису повезани на тастатурама како ваља.

Оперативни системи и кориснички програми за унос и обраду текста

Неопходно је да наша држава донесе минимум регулативе у области коришћења рачунара за унос и обраду текста, као и осталих облика података. У овом делу даћу кратак преглед данашњег стања.

Сведено на један општи ниво, у посао прилагођавања српском језику оперативних система и програма за унос и обраду текста мора се укључити већи број релевантних стручњака из области лингвистике, рачунарства и типографије.

Микрософтови Прозори

Овај оперативни систем задовољава минимум потреба српског језика. Ипак, постоје сложени проблеми на системском нивоу. Није могуће (на системском нивоу) користити одједном нашу ћирилицу и латиницу. Често ни датум (тј. име дана и месеца) није могуће прочитати ако системски нисте одредили да користите искључиво ћирилицу као додатно писмо.

Од програма, тек пакет Микрософтов Уред 2000 углавном доноси могућност несметаног ћириличног писања кроз све његове програме. До тада је једино програм из раније верзије тог пакета (Микрософтов Уред 97) Микрософтова Реч 97 то омогућавала.

Програми које није направио Микрософт углавном преводе Јуникод у осомбитне кодне распореде, па често писање ћирилице у њима представља солидан проблем. Испис написане ћирилице мањи је проблем за интерпретацију.

Ова област је, ипак, до те мере сложена, да захтева посебну студију с темом: Који програми подржавају и на који начин правилно писање српским писмима.

Линукс

За разлику од Микрософтових Прозора, Линукс, као и остали оперативни системи који личе на Јуникс, по својој концепцији нису били окренути што једноставнијем коришћењу, већ што квалитетнијем. Такав је приступ довео до тога да корисник не мора видети да пише ћирилицом, али ће и поред тога имати изузетно квалитетан излаз на штампачу, односно у слици која се програмски генерише (и коју није могуће мењати текстуално).

Тек се од половине 1999. године Линукс почео знатније окретати кориснику. Данас, у оквиру онога што се ради у незваничном пројекту Српског Линукса, готово да се у потпуности овладало могућношћу да обичан корисник може користити наша писма (пре свега ћирилицу) на задовољавајући начин.

Ипак, још доста тога потребно је да се на системском нивоу Линукса уради не би ли вишејезични, па тако и српски ћирилички и латинички испис био у потпуности доступан крајњем кориснику.

Програми под оперативним системом Линукс на различите начине решавају проблеме примене различитих језика. Углавном се тежи да то буде по званичним стандардима (ISO и Јуникод), али програмска примена, захваљујући отворености Линукса, толико је разноврсна, да је неопходно да тим стручњака решава сваки проблем појединачно. У будућности, када Јуникод буде био у потпуности примењен и на Линуксу (што ће се догодити можда и у току наредне године), неће више тако компликовано бити стварање подршке за одређени језик.

Треба издвојити програм TeX, који већ пар деценија у потпуности подржава све наше језичке потребе. Једини проблем је у томе што програм, чак и у својим облицима усмереним кориснику (LaTeX и L^AX), није ни приближно тако удобан за коришћење, као што је то нпр. Микрософтова Реч.

И ова област, дакако, захтева једну сложенију студију.

НАШЕ ПОТРЕБЕ И СКУПОВИ ЗНАКОВА

Наше азбучне и знаковне потребе могу се поделити на следеће скупове (по важности):

- 1) Основна ћирилица и латиница.
 - 1.1.) Ћирилица.
 - 1.2.) Латиница.

- 2) Основни неазбучни знаци (бројеви, реченични и правописни знаци, акценатски знаци итд.).
- 3) Акцентована слова.
 - 3.1.) Ћириличка.
 - 3.2.) Латиничка.
- 4) Скупови научних, културних и осталих знакова за друштвену употребу.
- 5) Ћирилице у културној употреби нашег народа.
 - 5.1.) Савремене словенске ћирилице.
 - 5.2.) Старосрпска и старословенска ћирилица.
- 6) Латинице у културној употреби нашег народа.
 - 6.1.) Савремене европске латинице.
 - 6.2.) Историјски коришћене латинице (као што је Даничићева латиница).
- 7) Грчко писмо.
 - 7.1.) Основни скуп знакова.
 - 7.2.) Проширени скуп знакова.
- 8) Глагољица.
 - 8.1.) Обла.
 - 8.2.) Углата.
- 9) Јеврејско писмо.

Следи детаљан опис односа наших потреба према постојећим кодним распоредима.

Основна ћирилица и латиница

У ову групу спадају основна слова наших азбука. Конкретно везано за представљање слова на рачунару, наша латиница подразумева и једнокарактерне диграме за слова „љ“, „њ“ и „џ“.

Ћирилица

Основна слова наше ћирилице подржана су у већини кодних распореда.

Седмобитни кодни распоред који подржава нашу ћирилицу јесте JUS I.B1.003–S, који се као интернационални стандард зове и ISO IR–146. Постоје још два седмобитна ћириличка кодна распореда, који су некад били у широј употреби, али они подржавају само руску ћирилицу. У питању су кодни распореди KOI и KOI 7.

Особеност осмобитних кодних стандарда јесте стварање већег броја различитих који истом служе. Тако постоји већи број осмобитних кодних распореда који подржавају нашу ћирилицу: ISO 8859–5, CP 855, CP 870, CP 1251, KOI 8 European, KOI 8 Unified и Mac Cyrillic. Руски ћирилички осмобитни кодни распореди који не подржавају српску ћирилицу јесу: CP 866, KOI 8a, KOI 8b, као и низ KOI 8 кодних распореда којима се служе различити народи са простора бившег Совјетског Савеза.

Наша основна ћирилица спада у основни скуп ћириличких знакова Јуникода.

Латиница

С латиницом ствари стоје мало сложеније. Разлог за то је њена неједнозначност. Доскора смо морали себе прилагођавати рачунарима. Морали смо писати онако како нам уске могућности наметнутих америчких стандарда допуштају. Сада то више није потребно; напротив, рачунари се прилагођавају људима. У том смислу немамо ни националног ни научног права да одустанемо од једнокарактерних латиничких диграма за „љ“, „њ“ и „џ“ као основног дела нашег латиничног писма. Ситни проблеми који се јављају приликом коришћења латинице са овим знаковима много су мањи од оних који се јављају приликом коришћења тог писма без њих.

У овом смислу једино Јуникод подржава нашу латиницу. Остали, седмобитни и осмобитни кодни распореди не подржавају је у потпуности, већ се за диграме морају користити два карактера.

Седмобитни кодни распоред који делимично покрива наше латиничке потребе јесте JUS I.B1.002, односно ISO IR–141.

Осмобитни кодни распореди који делимично покривају наше латиничке потребе јесу: ISO 8859–2, CP 852, CP 880, CP 1250 итд.

Основни неизбучни знаци (бројеви, реченични и правописни знаци, акценатски знаци итд.)

У ову групу знакова иду они неизбучни знаци који се налазе у Јуникод скуповима: C0 Controls and Basic Latin, C1 Controls and Latin-1 Supplement, Diacritical Marks, General Punctations, Currency Symbols и Number Forms. Сви нама потребни знакови (осим краткосилазног акцената и можда дугосилазног) представљају општи скуп знакова европске цивилизације. Зато је потребно само следити оно што међународне институције усвоје и реализују кроз Јуникод.

Осим најосновнијих знакова, као што су бројеви и основни реченични и правописни знаци, ове знакове није могуће користити у седмобитним и осмобитним кодним распоредима. Зато се користе вишекарактерне замене (рецимо, пишу се три тачке уместо знака за три тачке, пишу се велика латиничка слова уместо римских бројева, користи се један облик црте уместо бар три различите по ширини итд.), што није добро ни стилски ни формално, а прихватљиво је једино ако није могуће друкчије.

Акцентована слова

Типографска правила поседују нешто већи стандард него функционална лингвистичка (и осталих наука). Као што је раније речено, осим програмског пакета TeX, ниједан други не омогућава „паметне“ измене изгледа слова. Ово је посебно важно код акценатских знакова.

Тако постоје два типична примера: први је тај што је акценат изнад великих слова и на већој висини и сам по себи већи од акценатског знака над малим словом. Други се тиче акцентованог малог латиничког слова „и“, које нема тачку, већ акценат стилски тачку надомешћује.

Треба напоменути да постоји одређени број знакова који надомешћује латинички скуп акцентованих знакова (углавном су у питању самогласници с краткоузлазним и дугоузлазним акцентом, као и генитивска дужина над словом „а“), али да то не задовољава наше потребе.

У нашем језику постоји пет самогласника и шест дистрибутивно слоготворних фонема. Поред тога, за потребе транскрипције наших дијалеката неопходни су и акцентовани полугласници, односно знак „шва“. За слово „а“ је потребно имати и облик с генитивском дужином.

Ћириличка

За ћирилицу је, дакле, потребно:

$$2 * (6 * [a + 5 * ([e, и, о, у, њ, њ, р, л, љ, м, н, њ])) = 132 \text{ додатна знака.}$$

Латиничка

За латиницу је, пошто постоји по три облика слова „љ“ и „њ“, потребно је нешто више додатних слова:

$$3 * (5 * ([l, j, n])) + 2 * (6 * [a] + 5 * ([e, i, o, u, x, y, r, l, m, n])) = 142 \text{ додатна знака.}$$

(Латиничка слова „х“ и „у“ узета су место тврдог и меког полугласника.)

Скупови научних, културних и осталих знакова за друштвену употребу

У ову додатну групу спадају знакови из следећих Јуникод скупова: IPA Extensions, Spacing Modifier Letters, Superscripts and Subscripts, Combining Marks and Symbols, Letterlike Symbols, Arrows, Mathematical Operators, Miscellaneous Technical, Control Pictures, Optical Character Recognition, Enclosed Alphanumerics, Box Drawing, Block Elements, Geometric Shapes, Miscellaneous Symbols, Dingbats, Braille Patterns, Alphabetic Presentation Forms и Small Form Variants. У Јуникоду 3.1 постоје још и следећи посебни скупови симбола који би могли бити од интереса за нашу норму: (1) византијски музички симболи, (2) савремени музички симболи и (3) додатни математички симболи.

Ови додаци су посебни за одређене науке, културе итд. У том смислу, ако су некој нашој науци потребни додатни знаци, онда то Јуникоду треба и пријавити.

Ћирилице у културној употреби нашег народа

Две су основне групе ћирилица у културној употреби нашег народа, поред српске. То су савремене словенске ћирилице и старословенска, старосрпска и црквенословенска ћирилица. Остале ћирилице, оне којим пишу разнолики несловенски народи на простору бившег Совјетског Савеза и

Монголи, треба третирати истоветно осталим писмима која нису у непосредном додиру са нашом културом, па их тако ваља преузимати искључиво из међународних стандарда.

Савремене словенске ћирилице

Особитни ћирилички кодни стандарди који покривају нашу ћирилицу (ISO 8859–5, CP 855, CP 870, CP 1251, KOI 8 European, KOI 8 Unified и Mac Cyrillic), покривају и све остале словенске ћирилице. Једини проблем чини ново украјинско слово „Г“, које у кодним распоредима насталим осамдесетих година и раније не постоји.

Јуникод ове ћирилице у потпуности покрива (у скупу знакова Cyrillic), док их седмобитни кодни распоред JUS I.B1.003–S не покрива.

Старосрпска, старословенска и црквенословенска ћирилица

У оквиру Јуникод скупа знакова Cyrillic постоји минимална старословенска ћирилица. Она, ипак, не покрива не само све типографске захтеве (посебна слова с дијакритикама), већ не поседује ни нека посебна варијантна слова. Зато је неопходно начинити студију о словима и знацима који су за старословенски, старосрпски и остале црквенословенске језике потребни. То би ваљало учинити у сарадњи с типографима који су већ имали искуства у раду с фонтовима.

Латинице у културној употреби нашег народа

И кад су латинице у питању, потребно је ограничити се на европске језике пре свега.

Савремене европске латинице

Њих покривају углавном Јуникод скупови C0 Controls and Basic Latin, C1 Controls and Latin–1 Supplement, Latin Extended A и Latin Extended B.

Различити седмобитни и особитни кодни распореди, као што су ISO 8859–1, CP 850, CP 1252 итд., покривају ова слова, али о њима не треба пуно размишљати јер директно онемогућавају писање ма којом нашом азбуком.

Даничићева латиница

Даничићеву латиницу покривају Јуникод скупови Latin Extended A и Latin Extended B.

Грчко писмо

Грчко писмо је део наше културе и науке. Стога треба и њега уврстити у скуп потребних знакова.

Основни скуп знакова

Основни скуп грчких знакова покривају, поред особитних кодних распореда ISO 8859–7, CP 737 и CP 1253, Јуникод скуп Greek and Coptic.

Проширени скуп знакова

Проширени скуп грчких знакова сличан је скупу знакова који би требали да садржати наша акцентована слова. Ту се налазе грчка слова за самогласнике и сугласник „р“ с акцентима и спиритусима. Постоји, као и наш скуп акцентованих слова, због типографских потреба.

Глагољица

У званичном документу Јуникода „Unicode Technical Report #3“ из 1992. године детаљно су описане потребе глагољице. Ипак, до данас (Јуникод верзија 3.1) глагољица у Јуникоду није покривена. Како је глагољица део наше негдашње културе, потребно је и њу покрити. И то обе њене варијанте: облугу и углату.

Проблем с Јуникодом је тај што се задржава, као и када је стара ћирилица у питању, само на основном скупу знакова (од четрдесет њих у овом случају), док не води рачуна о одређеним варијантама слова, нити о словима са дијакритикама.

Ниједан седмобитни ни осмобитни кодни распоред, као ни Јуникод, не покривају глагољицу.

Јеврејско писмо

Део наше хришћанске културе представља и јеврејско писмо. Као што је део рачунарских наука енглески језик и његово писмо, тако је део теологије јеврејски језик и његово писмо.

Јеврејско је писмо покривено у следећим осмобитним кодним распоредима: ISO 8859–8, CP 862 и CP 1255.

Јуникод јеврејско писмо покрива у кодној страни Hebrew.

О осталим писмима

Од осталих писама, једино је потребно размишљати о арапском, као делу наше културе за време турског раздобља и култури нашег муслиманског становништва.

Остала писма треба једноставно преузимати из међународних и националних стандарда других земаља. Наравно, у Јуникоду искључиво.

ПРИЛАГОЂАВАЊЕ ПИСМА

Своје особености један народ треба да чува, али је исто тако важно знати одредити приоритете. У томе смо као народ кроз историју често грешили.

Највиши приоритет који један овакав рад себи може поставити јесте српски језик сам. А кад му каква његова особеност у одређеном тренутку прети уништавањем, боље је те особености се одрећи, него се одрећи читавог језика.

Моја је жеља да српски језик не постане музејски, хир појединаца — већ да смело иде савременим токовима, без комплекса мање вредности, без страха од новог; управо како је ишао српски народ у временима када је српска култура била привлачна не само околним, већ и далеким народима. Зато је неопходно бити отворен за промене у одређеним структурама размишљања.

Пре свега, потребно је омогућити сваком на свету да се користи српским језиком независно од писма и знакова којим се служи. Наши су се људи током претходних година тога и досетили, па су створени различити начини за писање српским језиком помоћу енглеске, односно руске азбуке (слика 5).²³

²³ Имена прве три од ових српских услужних азбука узете су са Интернета. За „дуал“ је занимљиво и то да је исти предложио и Митар Пешикан у књизи *Наша азбука и њене норме* као једнозначан испис наше латинице у случају да нема наших слова. Усавршавањем ове идеје добијена су два крајња предлога: танко јер у ћирилицу и ипсилон у латиници створили би потпуно транспарентан азбучни систем српског језика, а тиме и могућност да се на **сваком** рачунару на свету српским језиком може писати (нисам до сада чуо за рачунар који не дозвољава писање енглеском или руском азбуком [можда неки стари јапански рачунари?]). Штавише, пресловљавање са ћирилицу у латиницу и обратно не би био никакав проблем.

врста слово	дуал	qwux	интернетица++	'y' варијанта	варијанта 'танко јер'
ж	zx	x	"z ("z)	zy	зъ
ш	sx	w	"s ("s)	sy	съ
ц	dx	dx (d\ x)	d"z (d\ "z, d"\z)	gy	гъ
ч	cx	y	"c	ky	къ
љ	ly	lj (l\ j)	lj (l\ j)	ly	лъ
њ	ny	nj (n\ j)	nj (n\ j)	ny	њъ
ђ	dy	dq	dj (d\ j)	dy	дъ
ћ	cy	q	'c ('\ c)	ty	тъ

Слика 5: Варијанте писања српског језика енглеском и руском азбуком. Приказана су само слова која се не могу написати енглеском азбуком (нешто је мање слова која се не могу написати руском азбуком [„ж“, „ш“ и „ч“ припадају основном скупу ћириличних слова]).

Данас више не морамо водити рачуна о политичким условљавањима зарад заједничког језика, тако да имамо одрешене руке за ваљано регулисање норме. А норма се мора пронаћи између два унеколико супротна процеса: између слеђења савремених потреба и њиховог регулисања. Значи, норма пре свега мора да регулише, али не сме прописивати нешто што је далеко од стварности. Понекад је неопходно начинити већи рез, као што је то и Вук Караџић учинио.

Српским језиком се у основи пише помоћу два писма, две азбуке.²⁴ Ово представља велику предност и не треба је се никако лишавати било на који начин. Коришћење два писма омогућава сваком нашем човеку блискост и источне и западне културе на једном суштинском нивоу.

Ипак, појаве које су механичке природе треба уједначавати. Тако распоред тастера на писаћој машини, односно тастатури рачунара треба да буде исти и за латиницу и за ћирилицу. Потребно је да и знаци „љ“, „њ“ и „џ“ на латиничној тастатури постоје на месту где постоје и на ћириличкој. Односно, треба пронаћи можда неки други начин по коме би, опет, изглед тастатура био уједначен.

Ћирилица има проблем малих курзивних слова „г“, „п“ и „т“. Да ли је то проблем српске ћирилице или ћирилице уопште, мање је важно. Неопходно је уједначити писање ових слова. Да ли ће се уједначити према руском облику или према, нпр., предлогу Митра Пешикана — мањи је проблем.

Потребно је, зарад формалних потреба, изменити, односно прецизирати одређена правописна правила. Правопис данас мора ући и у сферу припреме за штампу и уноса текста у електронски облик. Тако је потребно стриктно одредити коришћење капитале и верзала; потребно је дефинисати тачно на ком месту иде цртица, на ком црта, а на ком повлака. И то у складу са што јаснијим формалним разазнавањем функције датих знакова. Потребно је несрпске унете речи у српски текст означити и у латиничком тексту (тј., треба користити цртицу иза страног имена у основном, нескраћеном облику). Ово је проблем који треба тим стручњацима, пре свега из формалне лингвистике, да реши.

Такође, у складу са савременим токовима, добро би било размишљати о раслојавању графије и ортографије. Већ дуже време видна су превирања око наше језичке норме. Ова превирања, како их ја видим, нису одраз личне супротстављености, већ супротстављености процеса у самом српском језику. На једној страни покушава се истрајати у либералној традицији наше језичке норме, с друге се покушавају задржати одређени традиционални елементи у норми.

Решење је, можда, управо раслојавање и графије и ортографије. Садашња графија би се, по томе, везала за фонетско–фонолошку ортографију, а једна измењена графија, која би омогућавала писање српског језика напоредо само руским ћириличком, односно енглеским латиничком азбуком везала би се за једну морфофонолошку ортографију. Деца би у школама прво учила фонетско–фонолошку графију и ортографију, а потом и морфофонолошку. (У овом би се случају чак могло извести писање српским језиком и нпр. грчким писмом.)

РЕШЕЊА

Укратко, решења би се могла свести на следећа:

²⁴ Треба разликовати два појма: писмо, које је једна општија културна појава; и азбуку, која представља скуп знакова једног писма у служби комуникације једног народа, односно језика.

Кодни распореди

Седмобитни кодни распореди потпуно су застарели. Осмобитни не задовољавају наше потребе, али су реалност, па их треба разматрати. С Јуникодом наша држава мора успоставити контакт и приложити наше потребе. Ако Јуникод не жели наше потребе да задовољи, треба створити посебан шеснаестобитни кодни распоред који ће их задовољити.

Тастатуре

Потребно је уједначити ћириличку и латиничку тастатуру, са могућношћу да се историјским писмима пише уз незнатне измене на тастатури. Дobar пример за то је предлог Митра Пешикана. Треба посебно водити рачуна и о нашим дијакритичким знацима и њиховим местима на тастатури.

Оперативни системи и кориснички програми за унос и обраду текста

На овом месту треба посебно водити рачуна о могућностима за развој и квалитетну штампу и обраду података.

Треба овај проблем раслојити на четири нивоа:

- 1) Први ниво представљају крајњи корисници. Њих није потребно ничим условљавати за личну употребу, док би се у другим случајевима морали прилагођавати. У ту сврху добри су и програми каква је Микрософтова Реч.
- 2) Други ниво представљају штампарије и издавачи. Они би морали у тачно одређеном формату (TeX или SGML) достављати електронски облик издања Народној библиотеци (односно, другој надлежној институцији) и то у одређеним фонтовима које је држава откупила и прогласила јавним.
- 3) Трећи ниво представљају наше културне институције, које би морале имати унифициран оперативни систем (оперативни систем Линукс врло је добро решење, јер омогућава много већу слободу стварања и коришћења програма, за разлику од многих осталих, међу којима су и Микрософтови Прозори) под којим раде њихови рачунари, као и идентичне програме итд.
- 4) Четврти ниво престављале би остале државне институције, које би морале уважавати минимум претходног, а користиле би све у зависности од својих потреба.

Преводи

Неопходно је и започети систематично превођење рачунарске литературе и програма; одредити правилне преводе; начинити законску регулативу која обесхрабрује неправилну употребу итд. Потребно је у све то укључити што већи број људи, што Интернет омогућава.

У том смислу требало би направити на различитим посећеним местима на Интернету (какав је Интернет чвор пројекта „Растко“) форме за слободан унос предлога за преводе страних термина. После одређеног раздобља (месец дана, два месеца, пола године, година) ови би се предлози прикупљали, а групе стручњака би вршила одабир (или би на основу ових предлога дошла каква нова идеја). Тиме би се омасовио утицај на стандардни језик, а постојала би и одређена врста легитимитета за новоуведену реч. Спремност за овакве подухвате постоји, једино Одбор још треба да их озваничи.

ЛИТЕРАТУРА

- 1) Света Јовановић; *UNIX*; Институт за нуклеарне науке "Винча", Београд, 1996.
- 2) Цветана Крстев; *Један прилаз информатичком моделирању текста и алгоритми његове трансформације*, докторска дисертација; Универзитет у Београду, Математички факултет; Београд; 1997.
- 3) Митар Пешикан; *Наша азбука и њене норме*; Вукова задужбина, НИП Политика, Завод за уџбенике и наставна средства; Београд; 1993.
- 4) Митар Пешикан, Мато Пижурица, Јован Јерковић; *Правопис српскога језика са речником*; Матица Српска; Нови Сад; 1993.
- 5) Предраг Пипер; *Увод у славистику I*; Завод за уџбенике и наставна средства; Београд; 1998.
- 6) Donovan Rebbeci; *Font HOWTO*; LDP.
- 7) Ellen Siever, the Staff of O'Reilly & Associates; *Linux in a Nutshell*; O'Reilly, 2nd Edition, 1999.
- 8) Радоје Симић, Бранислав Остојић, Божо Ђорић, Живојин Станојчић, Милош Ковачевић; *Правопис српскога језика са речником*; Унирекс; 1993.
- 9) Живојин Станојчић, Љубомир Поповић; *Грамматика српскога језика*; Завод за уџбенике и наставна средства; Београд; прво издање 1992., више издања.
- 10) Михаило Стевановић; *Савремени српскохрватски језик I и II*; Научна књига; 1969.
- 11) *Unicode Technical Report #3*, Unicode Inc., 1992.
- 12) Bruno Haible; *Unicode HOWTO*; LDP; v0.12, 19 October 1999.