



ПРЕДЛОГ ЗА РЕГИСТРАЦИЈУ СТАРОСЛОВЕНСКОГ ЌИРИЛИЧКОГ ПИСМА У УНИКОДУ



Усвојен на међународној конференцији одржаној у Српској академији наука и уметности,
15–17. октобра 2007. године, у организацији Одељења језика и књижевности САНУ
и Института за српски језик САНУ, са темом:

СТАНДАРДИЗАЦИЈА СТАРОСЛОВЕНСКОГ ЌИРИЛИЧКОГ ПИСМА И ЊЕГОВА РЕГИСТРАЦИЈА У УНИКОДУ

Учесници конференције

Хајнц Миклас, Јоханес Рајнхарт (Аустрија); Румјан Лазов, Анисава Милтенова (Бугарска); Стана Јанкоска, Виолета
Мартиновска (Македонија); Елена Л. Алексејева, Виктор А. Баранов, Анатолиј А. Турилов (Русија); Душица Грбић,
Јасмина Грковић–Мејдор, Гордана Јовановић, Зоран Костић, Катарина Мано–Зиси, Предраг Миодраг, Слободан
Павловић, Виктор Савић, Гојко Суботић, Гордана Томовић, Ђорђе Трифуновић, Бранкица Чигоја, Новка Шокица,
Ирена Шпадијер (Србија); монах Јосиф (Хиландар); Вацлав Чермак (Чешка); Јелица Стојановић (Црна Гора)

пример писања). Остатак слова треба сматрати глифовима који припадају делу Уникода за приватну употребу. На тај начин би и најсложенији старословенски текст био Уникод–текст, потпуно компатибилан са свим фонтовима који поштују стандард (Уникод).

- б) Сва надметнута слова са титлом и без титле (писана изнад слова или између слова) за мала и велика слова у складу са решењем No2 (види Композитна слова).
- в) Сви надредни знаци за мала и велика слова и знаци интерпункције у складу са решењем No2 (види Композитна слова)
- г) Сви бројеви за мала и велика слова (види на пример www.unicode.org – U10140-Ancient Greek Numbers)

2. Редослед Уникод–кодова треба да буде у складу са азбуком, усвојеној на Београдској конференцији, због сортирања.

Имајући у виду принципе, није могуће извршити додавање на оно што је до сада регистровано, а то је и закључак Београдске конференције. Потребно је извршити ново регистровање из два разлога. Први разлог је тај што у садашњој регистрацији нема места за убацивање великог броја нових кодова. Други разлог је тај што, чак и да има места, постојећи ћирилични кодови би морали потпуно да се преуреде како би се омогућило сортирање. Чињеница је да се сада не може мењати редослед јер је ћирилична кодна страна већ у широкој употреби.

Регистрација старословенског писма одвојено и независно од садашње ситуације како је регистрована ћирилица омогућава да се старословенско писмо региструје у потпуности, у складу са оба принципа, и тада можемо имати савремену и старословенску ћирилицу у истом фонту.

Према нашем предлогу, имајући у виду принципе за регистровање писма, примере писања и принципе израде композитних слова, потребно је регистровати 1.303 карактера (види прилог). Глифови (у жутом пољима) треба да се кодирају у делу Уникода који је за приватну употребу (Private Use Area), о чему треба такође направити интерни стандард како би старословенски фонтови били компатибилни.

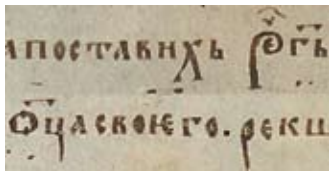
Преглед карактера за регистровање

Кураентна слова	стандард	глифови
Слова	96	246
Надметнута слова без титле (писана изнад слова)	96	43
Надметнута слова без титле (писана између два слова)	96	43
Надметнута слова са титлом (писана изнад слова)	96	43
Надметнута слова са титлом (писана између два слова)	96	43
Дијакритички знаци и титле (писана изнад или између слова)	49	37
Бројеви и њихови симболи	99	61
Интерпункција и симболи	26	43
Верзална слова		
Слова	96	124
Надметнута слова без титле (писана изнад слова)	96	43
Надметнута слова без титле (писана између два слова)	96	43
Надметнута слова са титлом (писана изнад слова)	96	43
Надметнута слова са титлом (писана између два слова)	96	43
Дијакритички знаци и титле (писана изнад или између слова)	49	37
Бројеви и њихови симболи	99	61
Интерпункција и симболи	21	60
Тотални број јединствених карактера	1.303	1.013

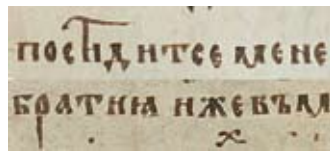
Лигатуре за кураент и верзал (види на пример www.unicode.org – UFB00 Ligatures for Latin, Armenian and Hebrew Ligatures) за сада остају у делу Уникода који је за приватну употребу (Private Use Area).

Лигатуре за кураент	85	7
Лигатуре за верзал	111	26

Пример писања:



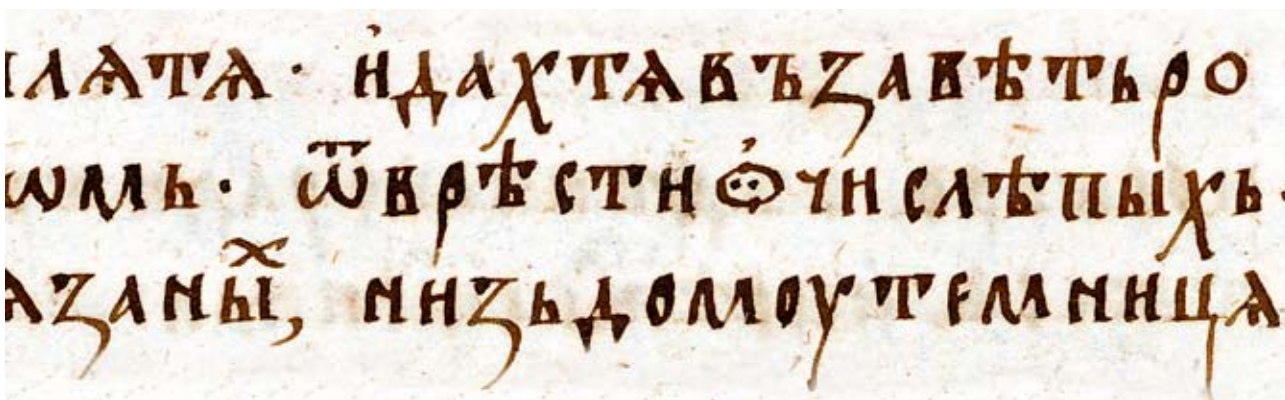
Лигатура „ре“ **Ɑ**, слова „ре“ **ре**.



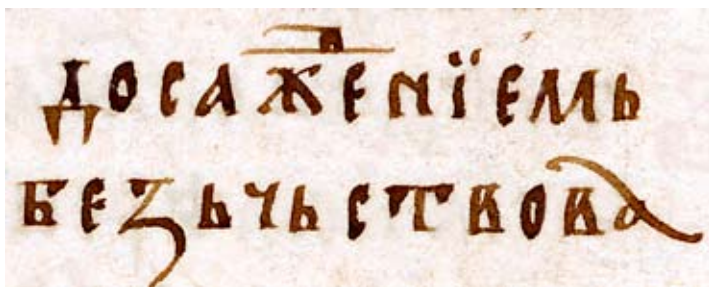
Лигатура „ти“ **Ɱ**, слова „ти“ **ТИ**.



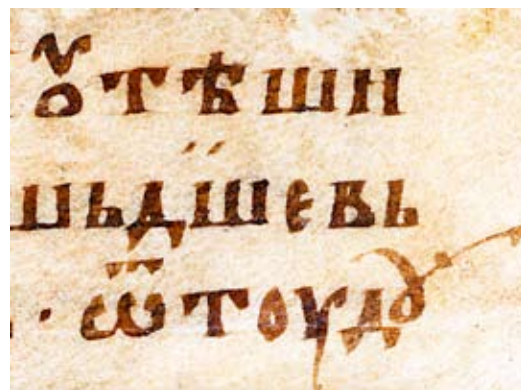
Слово „он широко 2“ **Ɱ**, „он широко“ **О**, „он“ **Ɱ**.



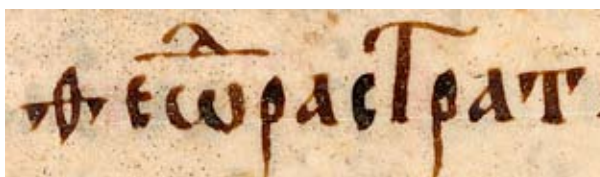
Слово „он широко“ **О**, „он двооко“ **Ɱ**, „он“ **Ɱ**.



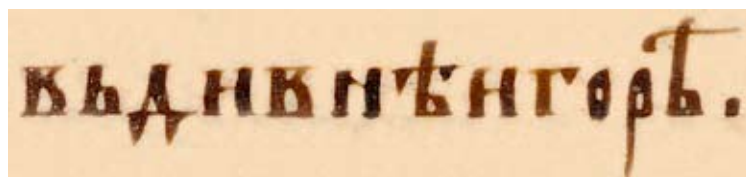
Слово „аз“ **А**, „алфа 5“ **λ**.



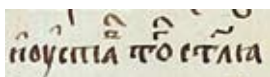
Слово „ук“ **Ѡ**, „оник“ **Ɱ**, „ук два 2“ **Ѡ**.



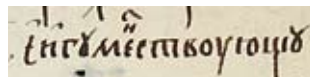
Слово „тврдо високо“ **Т**, „тврдо“ **Т**.



Слово „јат“ **Ѣ**, „јат високи“ **Ѣ**.



Слово „тврдо 2“ **Ш**, „тврдо 3“ **П**.



Слово „јота“ **Ѡ**, „иже“ **Н**.

Композитна слова

Да би се направило слово са дијакритичким знаком изнад њега (композитни карактер) постоје две могућности. Показаћемо то на примеру конструкције курентног слова *udieresis* “ü” и верзалног *Udieresis* “Û”.

1. У Уникоду имамо дефинисано слово “u” са кодом (0075), дијакритички знак дијерезис “¨” са кодом (00A8) и композитно слово удијерезис “ü” са кодом (00FC). Дијакритички знак нема неки одређен положај, већ га типограф у композитном слову тачно подешава у односу на ширину и висину слова.

За верзално композитно слово Удијерезис “Û” са кодом (00DC), користимо слово “U” са кодом (0055) и исти дијакритички знак дијерезис “¨” са кодом (00A8) као и у случају курентног слова. Типограф га поново тачно подешава за нову ширину и висину.

Видимо да у оба случаја није битан положај дијакритичког знака у свом пољу (00A8) већ да тај положај одређујемо у композитном слову.

Сва остала слова са дијакритичким знацима се конструишу на исти начин. Када куцате, куцате директно “ü” у једном потезу. Последица овог приступа је да за сва слова са дијерезисом имате два кода, односно два места у фонту (слово и композитно слово) и само један код (место) за сам дијерезис. То значи да вам за 12 слова (ä, ë, ï, ö, ü, ÿ, Ä, Ë, Ì, Ö, Û, ÿ) треба 25 Уникод кодова, односно места у фонту.

u	¨	ü	U	¨	Û
0075	00A8	00FC	0055	00A8	00DC

2. Други приступ је упрошћен и са типографског становишта лошији. Као и у горњем случају имамо слово “u” са кодом (0075) и дијакритички знак дијерезис “¨” са кодом (00A8), али немамо композитно слово. Уместо тога када правимо дијерезис позиционирамо га тачно за висину курентног слова, стављамо да му је ширина нула и померамо га улево како би дошао изнад слова. Када куцамо, куцамо прво “u” па онда дијерезис које ће се појавити тачно изнад “u” и добијемо “ü”.

За верзал имамо слово “U” са кодом (0055) али сада морамо узети други дијакритички знак Дијерезис “¨” са кодом (F6CB). Дијерезис сада подешавамо за висину верзалног слова, стављамо ширину нула и померамо га улево како би дошао изнад “U”. Када куцамо, куцамо прво “U” па онда дијерезис које ће се појавити тачно изнад “U” и добијемо “Û”.

У оба случаја, ако смо добро подесили положаје дијакритичких знакова, добићемо добар резултат. Тако за истих 12 слова (ä, ë, ï, ö, ü, ÿ, Ä, Ë, Ì, Ö, Û, ÿ) треба 14 кодова, односно места у фонту, 12 за слова и 2 за дијакритичке знакове, што је значајно мање него у првом случају (25).

У чему је проблем оваквог приступа. Како смо позиционирали положај дијерезиса за ширину слова “u” а постоји само један дијерезис, то ће положај дијерезиса за слово “i” које је уже бити погрешан.

u	+	¨		=	ü		U	+	¨		=	Û	али за idieresis	ı	+	¨		=	ï	
0075		00A8					0055		F6CB						0131		00A8			

Да закључимо:

- Решење No1 је типографски оправдано и сви романски језици као и савремена ћирилица су урађени на тај начин. У случају старословенског писма то би, због многобројних надметнутих слова и надредних знакова, захтевало огroman број кодних места (више десетина хиљада).
- Решење No2 је типографски мање оправдано али захтева знатно мање кодних места.
- Решење No3 не постоји.