

Àâòîð Àëåêñàíäð Ñåðääââè÷ Ååðää  
18.07.2008 ä.  
Íñëäâíå ïáíâéäíéå 20.07.2008 ä.

Òåêñò ïå÷àòííâî èçääàíèÿ â ôîðìàòå PDF

Íáðæéëåéüíí Óíðleðíáàíéþ áàcû áäáííûð áúéë íà=ðáòí èçó-ðáíéà áðàí-íàðóéèè, ñeiðáííðàçíàáíéý êíéðóðáíûð óðæñòðá. Á ðáçóéüðàðá ê ñeéííáíéý, áæàáíéüíí ñiðýæåíéý è ìàáíííáí ñeiðáí-íàðàçíàáíéý íàíýðíèéíá Õðññéíé aæéíáðáðé-ðáñéíé èéðáðàðóðû XVI á., iióáéééíá 1996].

Äëy iððaðñòàâåéåíèý ðóéííèñáé â èíðíóñå áúëæ ðàçðåáíòàíà ñèñòåíà ìòåðåæåéíèý äðåâåíðóññéíè áðåàðòéè, èíòíðàÿ içåíèÿåò áíñå ãðåâåíåðóññéèo áóéå è eó ñàïàíòè-åñéè çíà-èíùô àåðèåíòîå (óçéíå è øèðíéíå «í»; óçéíå, øèðíéíå, ýéiðííå «å» è o. i.). Áíñiðíèçå çíàéè iðèäúòåíèý è àéöåíòíû çíàéè. Ðàçðåáíòèå áàçíåíâí ðøéòòå äey áâíäå æèòèéíû òåñòåíà íððaðñòàâåéÿæ ñíáíé ðýä iññéåå ñíñòé áíñiðíèçåååíèý ðóéííèñáé, íàïðèíåð, àåðèåíòû áóéå, íå èíåþüéå ðííåð-åññéíåí çíà-åíèý, èéåàð

Đàcđàâiòàiá ñíäöèåëüäý ðïtäðàlià, tïçâñéÿpùàý ïtéó-åòòù è ååååååííù òåéñòàì (è éåæäåíò å ìòååéüñòé è èéé è lñåñéüéèí åiåñòòù)

Â àèòâàâèòå, èòòïðûé Äðâáíÿj Ðóñü âíñïðèíÿéà ìò þæíûò ñèàâÿí, óæå ñ ñàïïâ íà-÷æà áúëè áóëåû, íà èìâàòèå íñïâíäî òïíàòè-÷âñò XVI á. íåèòïðûá áóëåû íñïâíÿéè ñâïå çåóëíâïå çíà-÷æà óæå íà ðóññéíé ì-÷âå, á ýçûéå ðàçâèëëñü òàëèå òïíàòè-÷âñèå yâëåíéÿ ñïâëæñíûô, âñå ýóï íðèâåæí è òïíó, -òï íäíà è òà æå ñëïâíòòïðà íñäëå áûòù çàïèñàíå íåñéïüëèìè ñïíñíàïè. È òïíó æå iëñöû â ñâïå áóëåàïè), è â òåëñòå ðæòèé íæòïðûá ñëïâíòòïðû ëíäþò ñâûòå 10 âàðèåäòíå íàïèñàíéÿ. Òàëèï íàðåçii, ñòåïíàëòñý í-÷âæäïíé åé

Äëÿ ðåøàíèéý ýóíôte iðlääéàíù íù èñïíüéçóái íáñëíüéü íðéäàíà. Äí-íäðåúô, óíðiùáíèá áðåòåðéèé: óñòðàíàíí áðåéèðíààíéà áóêâ, ííóùàíù ñíùñéó, íáúäæíýþöñ ñéíàíòðiù n íàééàéíàùí áóêååííùí ñíñòàâí, ðàçéè-÷àþùèåñý íàéè-÷éàí/íñòñòñòåéàí áúíñííùó áóêâ èéè òàí, ñíñòàâà ñéíàíòði, ièþóùèöñý á ñíèðàùåííé ñíðià: áññòàíåééàþþöñ áí ííéíàí áéàá èíðié, ðåñóéýðí ñíñòàùåíùà ííà òèøéí[2]; áóêåà &mdash; ñòðí-íáý áóêåà ñ ðåñóóèðíàíùí &mdash; ñòðí-íáý áóêåà ááç ðåñóóèðíàíùí); íáúäæíýþöñ ñéíàíòðiù, ðàçéé áúíñííàí AE èéè N. Ä-òðåòüèô, óñòðàíàíèá áðåòåð-÷åñëíàí áàðüéðíàíéý, ýáéýþüåñíñý ñéåñòðåéàí èçiaíáíèé óííåòð-÷åñëíèé ñéñòå ñíàéñíùé; óñòðàíÿåðñý áàðüéðíàíéá Ä/Í á íà-÷éàí ñéíàá è íà íéí-÷áíèé IAI/ AAI; íáúäæíýþöñ ñéíàíòðiù, ðàçéè-÷àþùèåñý íàéè ðåñóóèðíàíùé; óíèøéèðóåðñý íáéñàíéá íáéòðiù ñíðoåí, á ñíñòàâà èíòðiù ëñòðíí èíàéñý ðåñóóèðíàíùé.

Â ìàñòíýüåå áâðâïý iù ïñòúåñòâéÿåì áððàììàòè-åññéóþ ðâçìàòéóþ iðåâñòååéáíûò ã êîðöñå æèòèé. Ðàçðàáòåí òîðìàò iðåâñòååéåé åíòåðïðåðåéöý ïñòåéüíûò òèôòò çâåñèñò ìò òîñâ, è éâæéíé ÷åñòè ðåð-÷ è ìòññòðñý ñëîñâ.

Ía è áuêé! iðeýþöti ðáðááíéà ñääáéàðö Þíðáééëííâáííúá ðáéñðòú æðòðéé áíñðóðííùé äey aíáðíéð Þíðüçíàðåðéé è íðåðñòàðèðö èð íà ñà ðáéé &mdash; á ðíðiàð XML, éííâáððàðéé yñðóùñðóâéýðóñy áðòðý áðòðiàðòð-áññé.

Â ôîðìàëüíî ïëàíå XML-ðàçìåòèà êîðïóñà íñíîâûâàðòñÿ íà ëåæäóíàðíàíûõ íîðìàð ìîðìèáëèÿ ýëåêòðííûõ èçäàíèé òåðñòà, â ÷àñòíî

Â iññíáò nòðóéòòðú æèòèë yéäåéòðííííá ãéñíáò ïíñéåíû ñòðàéòåðèñòèëè ðóéññé: ðàçáéâéà òåéñòà íà èëñ (iaòòèè) íà-æà/éííòà èéñòà, ëíèííéè, nòðííéè n ñíòåáòñòåóþùåé íóíåðåóëéå.

Íðåäñòàåéÿ òåêîò â ýëåéòðíííí îòðìàòå, íæíí áúáðàöü íæíí èç äåóö íóòåé: íàéñèìàéüíí òí-íí áîñïòèçâíàéòü áèä åóéííèñííàå ãòåò ãòåò ðíé íóòü. Òí-íí òåé æå, êíäää öîðìàéüííä -éàíáíèå òåéñòå (íå ñòðîéè è èëñòû) íå ñíâíàéäåò ñí ñíûñéîåù (íà ñéíà), íü ãñåäää öåéééíí, íí íðè ýóíí îòíà-ååòñý íåñòî, áää íòíòíäèò åðåééòå ñòðîéè.

Ââðéðéðéðéðíâáííûá XML-ïðâáñooðaaéâély ãæðéðééíûð ðâéñðóâ áóäóó á ââéüíâéðâl áâíííéíâú ííðñíðéâéâ-âñéñé ðâçìâðóééíé: ðâçðâáâðâí

Nièñîê ëèòåðàòóðû

Àââðèíà è àð. 1990 Àââðèíà, Ñ. À. ßçûê ðóññéî àæñïáðåôè-àñéîáñï / Ñ. À. Àââðèíà, È. À. Àç-à-

Àâåðèíà è äð. 1996 Àâåðèíà, Ñ. À. Èåêñèêà è ïîðôîéïäèÿ à ðóññêîé àãètäðàòè-åñêîé ëèòåðàòóðå XVI â. / Ñ. À. Àâåðèíà, È. À.

## Summary

The Corpus of Russian hagiographic texts of the 16&ndash;17th centuries at present comprises 52 texts or 500 000 word-tokens, represented in 2 formats: as text files and Microsoft Word files; the texts are provided with a word form index. 10 texts have been published; they are available to Internet users in the PDF and XML formats. The work is under way to provide all texts with the morphological information. Automatic normalization of varying Church-Slavonic spelling is another important task.

[1] À 2000&ndash;2006 åå. ïíóáèèîâàíû: *Æèòèå Èèðèëèå Áåëïçåðñêîäî, Æèòèå Àëåêñàíäðå Ñâèðñêîäî, Æèòèå Àíòííéÿ Ñèéñè Äèíèñèÿ Äeoøèöêîäî* è *Äðèäîðèÿ Íåëüøåì-ñêîäî, Æèòèå Èíðíèëèÿ Èíàëüñêîäî, Æèòèÿ Íåâæà Íáíðñêîäî* è *Ñåðâèÿ Íoðíñêîäî*.

[2] Ìû ðàçëè-àâî òèòëî è ïíêðûòèå, ñòàâýùåâñÿ íàä åûíññé áóêâîé.

[3] Ìåæäóíàðíäíûé èíññðöèòì ïî åûðàáîòêå íîðì ýëåêòðíííé ðàçìåòêè òâêñòâ: The Text Encoding Initiative [Ýëåêòðíííûé ðâñóðñ]. Çàâæ. ñ ýêðàíà.