

Ýëåêòðííûé ñëîâàðü âàðèàíòîâ íà îñíîâå òåêñòîâ XVIII à.

Àâòîð Ñâåðëàíà îëåäîâíà Ñàâ÷óé
08.08.2012 å.
Íñëæäíå ïáíîâëåíèå 27.08.2012 å.

Summary. Electronic dictionary of variants is being created on the basis of the texts of the 18th century in the Russian National corpus. For its formation the list of the unidentified word-forms supplied with hypothetic lemmas is used. The database of these word-forms gives essentially valuable material for the analysis of the orthographic, phonetic and different kinds of grammar variation in the 18th-century language. At the same time the study of variation has practical use for customizing tagger according to special cases which are typical for 18th-century texts.

Ýéâéòðííúû
ñéïáàðû áàðèæìòò ñíçâàåòñý íà áàçå óàðñòò îà XVIII á. á ñíñòàåââ
íàðéííæüñíâî èïðíóñà ðöññéëíâî ýçûéâ. íà íäðâî ÿòäïå ñéïáàðû áóäåðò èñííëüçíââ
äéý ôïðíèðíâàéïèý ñéïáíèêâ ðàççâåàåòðûâââïíâî ïðôðíëíâè-âññéïíâî ààëëèçàòðâ äéý
ðâéñòò îà XVIII–XIX á. Á áàëüíáéòâî âïçííæíí ñíçâàåíéå ýéâéòðííñíâî ðâñòðñà,
èïðíðûé ïïïíèò ñâïáàéñòâî ýéâéòðííûò ñéïáàðâé íà íññíââ îÉÐß [Ñéïáàðè]. Èâé
éçââñòîí, òâéñòû XVIII á. ìòëè-âþòñý êðâéïå áûññîé ñòâïáíüþ
ââðéàòèâàíñòè, èïòòðâý ñéïáâî ìòðâæäåíâî á ñòùâñòâóþùéò ñéïââðýò è ïðâéòè-âññéè íà
íïññâíâî á ïðíàòèâûò ãðâàíàòðèéâò. Âñâ ýòï áâéëåâò îáâçííæíí èïððâéòíóþ ðâáîòò
ïðôðíëíâè-âññéïíâî áàéèëçàòðâ, ñ ïïññüþ èïòòðâî ãûññíèýâðñý áâòòñòâðè-âññéâý
áííñòâðéý ðâéñòòâî, ïññéïëüéò á ñéïáàðâå àíâéèëçâòðâ áâððâéòðâîññòû òâéæäå íâ
ó-ðâéñâ. Òâé, íàïðèëâð, ñéïâî òâëüäàððâæ íðëñóñòâóâò á ñéïáàðâå á
åâéèñòâåâíí ïðôñðâòè-âññéïíâééâ, á òî áðâàïý èâéâ á òâéñòâð ëåòðíññéïâî
íâðâéââà ïííæåð áûòû íâðâåââàíí ñâïùè ðâçíûòâ ïñññíâàïè: ôâëòàððâæ,
Òâëüòàððâæ, ôâéòùàððâæ, ôâëòàððâæ, Òâéò ìàððâæ, Òâëüò ìàððâæ è áð., íè
íâéï èç èïòòðûò íâññíâåðñý èâéâ íðõñðâòè-âññéè áâððâéò ëåííé è òïé æâëïñ.

Ó÷èòûâàÿ îñíâáííñíñòé òâéñòíâ
 XVIII–XIX â., òîðìèðíâáíéâ ñëíâàðý îñðôïíëíäè÷åñêíâí àíàëèçàòíðà äëëÿ
 äèaðoñíiè÷åñêíâí îñáóöÿ ÍÉÐß ðåðâííâ íñóùâñòåëÿöù ïí ãâóí íàïðâåëáíèÿ. Íåðâíâ
 íàïðâåëáíéâ ïðåâííëåâàò ðåñòèðâíéâ ñëíâíéâ çà ñ-åò åéëþ-åíèÿ â íåâí
 íàðåðèàéíâ èñòíðè÷åñêèò ñëíâàðâé (ÑÀÐ, ÑÖÑÐß è äð.). Åòíðíâ íàïðâåëáíéâ ñíñðåäíòi÷åíí
 íà íñííëíâíè ñëíâíéâ ñëíââàíè, èçâáéâ÷åííùíè íàïñðåäñòåââíí èç èíðöñâ òâéñòíâ
 XVIII–XIX â., èíòíðûâ íå ó÷òâíû ã èéíââèñòé÷åñêèò èñòí÷íèéàõ. Ðàâíòâ â
 ðâíèéàõ âòíðíâí íàïðâåëáíèÿ ïðíâäëëæñü íåñêíëüê ýóâííâ.

2. Íñõõðíái ÷àñõòìõúé
ñõõâåðü ñõõâõõðí, íá ïíeo÷àþùèõ ïðåäñéàçàííâ ðàçáîðà. Íñõõëüéó
ñíõõåâõðõåðóþùèå èí ëääñâùû ìõñõòñõåðóþò à ñõõâàðå ïðõðîëíàè÷åñéíâ àíàéèçàòíðà,
ýøè ñõõâõõðíû ñíáãæåíû ãèíïõåðé÷åñéèè ðàçáõðàè, ïðåâéëüíûé èéè ìøéá÷íûé.

3. Îðääìéèçîàáíà áàçà ääííúô
íåñëíâàðíûô ñëïâîòïðì, èìåþùàÿ ñëåäóþùèå íiëÿ: ñëïâîòïðà, ñääíåðèðíâàííàÿ
éäííà, ìøñùéè÷íàÿ (íiðìàéèçíâàííàÿ) éäííà, áðàííàòè÷åññèèå íðéçíàéè éäííû, áðàííàòè÷åññèèå
íðéçíàéè ñëïâîòïðû, öèí áàðèàíòà, ñäääåíéÿ íá àåòïðå è äàòå ñïçäàíéÿ òåêñòà, á
éòïðïðì càòèéñèðíâàíà ôïðà.

4. Ëðîâåäåí àíàëèç 4000

ääëéëö ááçû ääëííöö: äëÿ éàæäéé ñëïâïòööìù íöðëëñäíá èëëë áúáðäíá è íöðääëéíäáííïäí
ñïëñéä íðäåéëüíäy éäííä, íðëëñäíáù ìöññüëí-íäy (ííðäëéëçíäáíäy) ñëïâïòööìä è
ííðäëëéçíäáíäy éäííä, áðäàíäöö-áññéäy èíööìäööëy íðäåäåðäíá íí èíööíöño è
ííðäåäåéäíù íðäåéëüíñä áðäàíäöö-áññéäè öäðäåéööåðëñööëéè.

Íñêüéó

áíáéèç ááçú ãúïííéýåöñý á ðó÷-íí ðåæèìå, íáíåñïäéèíí ãúééí ïöåíéòú
ööåéññíáðåçíññòú ýòié ñöðåñíäééí ðåááïòú, ýööåéöéåíññòú éíòðöíé áúðåæàéèåñú áú á
ííñûöåðéé èå÷-åñòåå ðåçáíðå òåéññòíå. ïðåñåñääííúé áíáéèç tñéåçàë, -òí èñíïðåñéåíéå
íøéåí÷íúò ðåçáíðå ëí÷òé ååååíå ñíéðåùååò ëíéé-÷-åñòåí ïñíæåñííúò åéíñòåç, -òí
yåéyåöñý åññòåòí-íí åúññíéèí tñéåçàòåéëåí. Íí ÷-åñòåðå-ííé ïðéíåäéåæíññòé ïðåñåíáéèçéðíåáííúå
ååéééòú ðåññöåñääééèéñú ñéååóþùèí íåðåçí. Éíåíá ñóúåñòåéòåéëüíúå ñíñòååéýþò
áíéååí ïíéåéíú åññåò ñíéåñòåíðí, èç íéò èíåíá ñíñòååéííúå (éé-÷íúå èíåíá, òåìééèé,
óííííéíú) íéíéí 51%. ïðééåååðåéëüíúå ñíñòååéýþò íéíéí 40%, áíéüøå ïíéåéíú èç
íéò ïðéóñíåéöñý íå åíéþ ïðééåååòåéëüíúò, íåðåçíññííúò ìò ñíñòååéííúò èíåíá. ïñòååñòéåñý
9% ñéíåñòåíðí ïðååññòååéýþò ñíáíé õíðüün åéååñéíå, íåðå-÷-éy è ñéóæååíúå ÷-åñòé ðå÷-é.
Òåéèí íåðåçí, á öåééí íéíéí ïíéåéíú ñíéññéå õíðí ñ íåñåññéåçóåíùí ðåçáíðå ëí
ñíñòååéýþò èíåíá ñíñòååéííúå è íåðåçíññííúå ìò íéò ïðééåååòåéëüíúå. Á ïñòåéëüííé
÷-åñòé ñíéññéå çåìåòííå íåñòí çåíéíàþò éåéññåí, ïòññòåñòåóþùèå á ñéíåñåðå (àæéèåò, ð,
áíåòííèòú, åæíí, åúå, åðéååóçéðíååòú, áííåñò, áåçíóæäíí è
ïð.), ïðòíéíå-÷-åññéå åðôåæçíú (åðåíåð, åðåññåò, åóóíæ, íñòðíåé, çíàååòé,
åýøå, áýøó, ýåéyøåñý è íé.) è ååðéèåíòú ðåçéè-÷ííå ðåñå.

5. Nëåäóþùèé ýòàï ðàáîòû

ñâýçáí ñ ôíðìéðíàáíèáí ááçú ääííûô áâðèáíóíâ. Áúëá iðíâåååíá ðàçìåòéâ áâðèáíóíâ, iíéó÷åííûô íá iñíïfâáíèé áíâéëçá òâéñòíâ XVIII á., áûäååéâíû
iðôíñâðåòé÷åñééâ, ôííåðòé÷åñééâ è iñôííéíæ÷åñééâ áâðèáíóû, íá ó÷ò, ííûâ á
âðâííàðåòé÷åñééâ iñíïâáðâá ÍEDB. Ñðâáéé íéô ñâíóþ çâíàðôíþ ÷âñòü iñíñòåâéýþò
iðôíñâðåòé÷åñééâ áâðèáíóû íéíé 76% (àðìäÿ, áíëøéé, áðååñâéð, áðèåðåðåð /
áðôéåðåðåð, áóäöíáû), êíòíðûâ áûââåâð ôðôáäíí ðâðåéòåðéçíâàðû ìðâååéüíí ìò
ôííåðòé÷åñééð (áíâéëíñééé, áíâéëéíñééé, áâéëíñééé, áâåðíçèý, àïðééëâââ,
áíâéëéé, àðôííé÷åñééé, àðæàðííé, àðæâðííé, àðéâàðóééâ, áííâðåéðíñâíéâ,
ââðíñââåðû è iñä.). Ìðôííéíæ÷åñééâ áâðèáíóû iñíñòåâéýþò íéíé 2% (àðìÿíá,
àðìÿííâ áí. àðìÿíá, àðìÿí, Äíâííéííâ áí. Äíâííéí, çíàíâíû
áí. çíàíâíà, áââ åðääóðû áí. áââ åðääóðâ è àð.); iñíïâííâðåçíâàðåéüíûâ
18% (Äíè÷éíâ, Äíè÷éíñééé, Äíè÷éíâûé, àðåâàðøñééé, àíòé÷åñééé, àðòèéåðíûé,
àðôéâñééé, áâðêéðíâ, áââñòâí, áâcääéüñòâí è iñä.).

6. Íðíâåäääíà íöåíêà åäëèíèö áàcû

ääííûõ è áíáäðåíèå áå â ñeñâàðü iñçâtëèò ïðíâåðèòü ãeñòåçó î òî, ÷òî
 ðàñøèðåíèå ñeñâàðü iñðöiëññé÷åññéññé ìáàëèçàòîðà è íañòðíéèé áåñ íà ïðåäåééíûå
 òåêñòû ñiññíáññé ìññéñòü ëà÷åñòåí áâòñàòè÷åññé ïñðöiëññé÷åññé ðàçìåðèé êàé
 òåêñòå XVIII á., òàé è áññåí áæaðññé÷åññé ïññé ã õåëii. Yéäéðññé
 ñeñâàðü áàðèàíòiá, ñiçäàíññé íà iññíá ÍEDB, áóäåò iñëåçáí ãëý èçó÷åíëý
 iññíáññé ñòåññé ëèòåðàòóðñé iñðiñ.

Niññíé ëèòåðàòóðñ

Ñeñâàðè Ñeñâàðè,
 ñiçäàíññé íà iññíá ìàöèññéüññé ëiðññé ðóñññéññé ÿçüéà [ñàéò]. [2008]. URL: <http://dict.ruslang.ru>
 (ääòà íáðàùåíëý: 15.04.2012).