

# Езико-лингвистичний обробник

Актуальна дата: 04.08.2012 р.

Інші дати: 27.08.2012 р.

## Summary. Current

linguistic processors are not suitable for analysis of pre-reform Russian texts due to numerous graphical, morphological and lexical differences from the modern language. We have developed a lemmatizer which can properly analyze pre-reform texts and has a facility for flexible adaptation to other spelling systems.

### 1.

Інформація про текст.

#### Новість та орфографія

Це новий текст, який не має аналогів в нашому зібранні. Орфографія відрізняється від сучасної російської мови завдяки багатьом графичним, морфологічним та лексичним відмінностям. Ми розробили лемматизатор, який може коректно аналізувати предреформальні тексти та має можливість гнучкої адаптації до інших написань систем.

#### 2. Інформація про текст.

Текст є новим та неподібним у зібранні. Орфографія відрізняється від сучасної російської мови завдяки багатьом графичним, морфологічним та лексичним відмінностям.

#### 3. Інформація про текст.

Текст є новим та неподібним у зібранні.

#### 4. Інформація про текст.

Текст є новим та неподібним у зібранні. Орфографія відрізняється від сучасної російської мови завдяки багатьом графичним, морфологічним та лексичним відмінностям.

#### 5. Інформація про текст.

Текст є новим та неподібним у зібранні. Орфографія відрізняється від сучасної російської мови завдяки багатьом графичним, морфологічним та лексичним відмінностям.

#### 6. Інформація про текст.

Текст є новим та неподібним у зібранні.

#### 7. Інформація про текст.

Текст є новим та неподібним у зібранні.

#### 8. Інформація про текст.

Текст є новим та неподібним у зібранні.

#### 9. Інформація про текст.

Текст є новим та неподібним у зібранні. Орфографія відрізняється від сучасної російської мови завдяки багатьом графичним, морфологічним та лексичним відмінностям.

Ãðàììàòè÷åñèé ñëîâàðü &mdash; nïëññé ëåêñåì ñ  
ïðèièñàíííé èíóïðìàöèåé í ñëîâèçìáíàíèé, èíóïðàÿ âéëþ÷àåò, êàé ièíèíóí,  
ñéäåóþùèå iàðàìåòðû:

- 1) ññíñâà ñ óêàçàíèåì  
÷åðåäåíàíèé;
- 2) iññòÿíúå iðèçíàéé  
ëåêñåíû (-àñòü ðå÷è, ðíä, iäóøååëåíñòü...);
- 3) èíä ñëîâèçìáíèåéüííä  
òèíä (iàðàäéäíû).

Ñëîâèçìáíèåéüíû òèí (iàðàäéäíà) &mdash;  
íááíð öéåéñéé, iáùéé äëÿ íâéíòíðíí ññòåñòå ëåêñåí, èíòíðûé çàäàåò  
ññòåñòå ñòåéå íææåó ãðàììàòè÷åñèéè çíà÷åíèé è ññòåñòå ñòåóþùèé èí ñòåéå.

## 2.2. Iðèíøeíû.

Ëåììàòèçàòð íå iðèåýçáí  
æåñòéí è íííéðåòííþ ýçûéò è á iðèíøeíä íææåò áúòü iàñòðííí áàëíé ýçûé  
öéåéòéåíííí òèíä iðè íæéé÷èé ññòåñòå ñòåóþùåé ãðàììàòè÷åñèé ññòåñòå è íðàéé  
òíéåíèçàöèé. Ány èííéðåòíí-ýçûéíàÿ èíóïðàöèÿ (éåííû, iññíáû, iàðàäéäíû,  
öéåéñèé, ãðàììàû) íå çàøèòå á èíä iðíäðàíû, à áúíáñåí áí áíåøíèå ðàáééöü.  
Äëÿ çàíèñè ñëîâèçìáíèåéüííé èíóïðàøèé áúéå ðàçðàáíòåíà ññòøèåéüíàÿ iñòàöèÿ,  
èíòíðóþ éåäéí ÷èòàòü è iðàéèòü áðó÷íóþ.

Ëåììàòèçàòð áéëþ÷ååò  
íàðàíéçí, iñçâíéþþùéé áíó àäääíøèðíâàòüñý è ðàçíûí iðôíâðàòè÷åñèé ñèñòåíàí. Á  
iðôíâññå áíàééçà èññòåíû ñàéñò iðâíáðàçóåòñý áí áíóøðåííå (íðòíàééçíåííå)  
iðâñòååéåíèå, èíòíðíà ýäéÿåòñý ááñòðåééòéé íò ñðåéüíííí íàéñåíèÿ è íææåò  
ññòåñòå ñòåéå ñòåéå ñåñéüé è íðôíâðàòè÷åñèé ñòåéåíòåí. Ýòò iàðàíéçí iñçâíéÿåò  
àíàééçèòðíâàòü ñàéñòü ñ iéàâàþùåé iðôíâðàòèåé, áäå íí÷òé ëþáíá ñéíâí íææåò èíåòü  
íáñèéüéí ñòåéåíòåí íàéñåíèÿ (é/ú, ô/s, úé/ú, ú/&rsquo;/íóñòí).

## 2.3. Àéäíðèòí.

Àéäíðèòí ðàáíòû  
ëåììàòèçàòð ññòåñòèò èç ñéäåóþùèõ ýòàííâ:

- 1) Òíéåíèçàöèÿ &mdash; iðíäðàíà  
ðàçáéåååò ñàéñò íà òíéåíû è áúäåéÿåò ñëîâà.
- 2) Íðíàééçàöèÿ &mdash; iðíäðàíà  
íàðåååíæò ñëîâí áí áíóøðåííå iðâñòååéåíèå, à èññòíàÿ ñëîâà ññòðàíÿåòñý  
äëÿ áúäå÷é.
- 3) Íðíäðàíà ññòøàåòñý á  
òéééå ðàçáéòü ñëîâí íà iññíâó è óéåéñèþ, à çàòåí äëÿ êàæäíííí áàðèåíòå ðàçáéåíèÿ  
íðíâåðÿåò äíííéøåéüíû áññéâèÿ:

&mdash; iññíâà ñí÷åòååòñý ñ  
ääííé óéåéñèåé (éíååò íóæíóþ iàðàäéäíò);

&mdash; iñíñáà èìååò ïðàâèëüíûé  
âàðèàíò (iðè íàëè÷è ÷åðåäñàíèý);

&mdash; iñíñáà ñíñáñòèìà ñ  
ëéññáñé ï ðåññòðó;

&mdash; ãðàññáñòèíû õëåññéè  
ñíñáñòèíû ñ ãðàññáñàíè ëåññáñû.

Äëý óñéñðåíèý àíàëëçà  
èññíñëüçóþòñý çàðåññá ïñòðññíû òàáëëöû õëåññéè, iñíñá è íàðåäñé.

4) Iðè ìòñóòñòâèè âàðèàíòíâ  
ïðñññññíà ïñòðññòðó ãëíñòðññé ãññéèé ðàçáñ ï ìàëëññé ñ ñóùññòðññóþùè  
ññéññé. Iðè ýòí ñññíñëüçóþòñý ñòðññòðé ÷åññéàý òàáëëöà õàðåäñòðññû ëíñññá ññéññé  
è ñò ãíçññæñû ðàçáñññ, à ãëíñòðññû ñññòðóþòñý ã ïðýäñé õáññññéý ãñññýòíññòðé.

#### 2.4. Äíçññæññòðé è íáëññòðé ïðèíññé.

Ðàçðñññòðññé íàíè  
ëíñññæñæñòðé ñòðññòðññé ññéññé:

1) ï ñòðññòðññé ñòðññòðññé  
ñòðññòðññé ññéññé:

2) ï ñòðññòðññé ñòðññòðññé  
ñòðññòðññé ññéññé:

3) ï ñòðññòðññé ñòðññòðññé  
ñòðññòðññé ññéññé:

Íðññññññíà ïññæñòðññé ññéññé:  
ññéññé:

1) ïññññññé ññéññé  
ññéññé:

2) ÷àññòðé ññéññé  
ññéññé:

3) ññéññé  
ññéññé:

Â íàñòñññé  
ëíñññæñæñòðé:

&mdash; Ññçññæñæñé  
ññéññé:

&mdash; Nîçäàíèå êíðíóñà òåêñòâî  
è ñëîâàðý ýçûéà È.Í. Áàòþøêîâà;

&mdash; lîäáíòâåéà òåêñòâî â  
ñòâðíé îðôíâðåðèè äëÿ ýëåêòðííûõ áèáéèòåé;

è íåéòòðûõ äðóåèõ.

3. Èèíâèñòè÷åñèàÿ lîäåéü.

3.0. Åååååíèå.

ïðè nîçäàíèè èåíìàòèçàòîðà  
ïù ìòòåéèååéèñû ìò äðàìàòè÷åñéîé lîäåéè ñîâðåìåíííâ ðóññéíâ ÿçûéà,  
çàòèèñèðîâåííé à ñëîâàðå Çàëèçíÿéà. Äàéåå, iû îðîàíæçèðîâåéè lîñíâíûõ  
îñíâåííñòë ýçûéà 18&mdash;19 ååéà íà iñíâåå ãðàìàòè÷åñéèõ lîèñàíèé è ðåàéüíûõ  
òåêñòâî. Ýòè iñíâåííñòë lîæíí óñëîâíí ðàçäåéèòü íà òðè ãðóíû:

1) îðôíâðåðèÿ

(äññéòåéüíûå áóêâû, äðóåèå ïðàâèëå íàëñàíèÿ);

2) ñëîâîèçìåíèå

(óñòàðåâðèå ôîðíû);

3) èåéñèéà (óñòàðåâðèå

ñëîâà).

3.1. Îðôíâðåðèÿ (óïðíùåííûé âàðèàíò).

lîñéíèüé ñîâðåìåííàÿ

îðôíâðåðèÿ à iñíâííí ÿäéÿåòñÿ óïðíùåíèà ñòàðíé, àïçíèéååò èääÿ, ÷òî ñòàðóþ  
îðôíâðåðèþ lîæíí àíæçèðîâåòü ïòòåì îðèååååíèÿ è ñîâðåìåííé. Âîò iñíâíûõ  
ïðàâèéà ïðåíåðàçíâàíèÿ:

1) çàìåíèòü óñòàðåâðèå

áóêâû íà ñîâðåìåííûå (³=>è, c=>å, s=>ô, u=>è);

2) ìòñå÷ü êíñå÷íûé -ú;

3) çàìåíèòü áåç-/â)ç-/ëç-/íèç-/ðàç-/ðíç-/÷(å)ðåç- => -c íåðåå ãëóôèìè (NB);

4) çàìåíèòü êíñå÷íûå -àãí/ÿãí(ñÿ) => -iãí/âãí, -ûÿ/ëÿ(ñÿ) => ûå/èå (NB).

íû îðîâåðèéè ýòíò iåòâî íà

áéüøíí îðîóñå òåêñòâî è lîéó÷èéè íåíèíðéå ðåçóéüòâò. Áíéüøàÿ ÷àñòü ñëîâ áûëà  
ðàçíâàíà îðàâèéüíí, íåñíçíâííûè lîñòàéèñû óñòàðåâðèå ôîðíû è èåéñâíû.

íäíàéî iåòâî óïðíùåíèÿ,

öiòý iïçâíèÿåò áûñòðí iïéó÷èòü ðåçóéüòâò, äàåò ñëèøèíí iíñâí øóià. Âî-iåðåûõ,  
íåðåíè÷åñéàÿ çàìåíà áóêâ áåç ó÷åòà îðôíèíæ÷åñéîé ñòðóéòòðû ñëîâà èíñâå  
ðåàòååòåò íåâåðíí (îðàâèéà 3 è 4). Âî-âòðûõ, ñîåðèâàþòñÿ ñëîâà, êîòîðûå  
ðàçèé÷àþòñÿ áóêâàíè c&mdash;å (ñëççú&mdash;ñëçú, ñçéú&mdash;ñëéú, ñâcäcí³å&mdash;ñâåååí³å, Âcíà&mdash;âåíà, iñçå  
Â-ôðåòüèõ, íåâîçíæíí lîéé÷èòü lîéé÷íûå íàëñàíèÿ, êîòîðûå iðè íåðååíåå à  
ñîâðåìåííûé áéà ñîâíàäàþò ñ îðàâèéüíûè. Í-ååèäíí, ÷òî äëÿ lîéííòåíííâ àíàéèçà

ñòàðíé ìðôðàðàòè è íáíáðíæì ñëîâàðü è ãðàììàòè÷åñêíà ëíèñàíéå, ãää òì÷íí  
óéàçáíú ìåñòà óíòðåáéåíèý áóéå (íðåæäå áñãåäí, c è s) á ñíñòàâå êîðíåé, ñóôôèññâ è ôéåêñèé, à òàéæå  
ó÷òåíú ñóñòàðåâøèå òíðìú.

### 3.2. Ñëîâîèçìåíåíèå.

Åñëè åçyöü çà íñííâó ìäåëü  
ñeiâîèçìåíåíèý ñíâðåìåííííñ ðóññéíñí ýçûêà, òí äëý àíàëèçà òåéñòâ á ñòàðíé  
íðôðàðàòè, ìíèíí ãðàòè÷åñêèò ìòëë÷éé, íáíáðíæìí ãíáàâèòü á ãðàììàòè÷åñêóþ  
íäåëü òíðìú, ìòñóòñòâóþùèå á ñíâðåìåííííñ ýçûêå èëéé íá ó÷òåííûå á ñëîâàðå  
Çàéèçíýéå:

1) Åäúåéòèâíûå ôéåêñèé (-àãí/-ÿãí,  
-ûÿ/-<sup>3</sup>ÿ).

2) Óñâ÷åííûå òíðìú  
ïðèëåâàðåéüíûõ (éðàñíà/o/û/ó), êîòíðûå òíðìàëüíí ñíââàäàþò ñ êðàòêèìè  
òíðìàìè, íäíáêè ìàþò è äðóâèå íäåäåæè (éðàñíó).

3) Íñíáûå òíðìú ìåñòíèåíèé  
(âÿ, ííc, íäícõú).

4) Ôâîðèòåéüíûé íäåäåæ 3-âí  
ñééíåíèý íá -<sup>3</sup>þ (íèéñò<sup>3</sup>þ,  
ííñù<sup>3</sup>þ).

5) Ñðàâíèòåéüíàÿ ñòåíáíû á -cé (ñëëüícé)  
é -ÿå (ñëëüíÿå, ñéíöÿå).

6) Åàðèàíò ÷àñòèöû -ñÿ ìñëå ãëàñíûõ (åàëþñ, åàëëëëàñ),  
êîòíðûé óíòðåáéÿåñý òàéæå á ñíâðåìåííííñ ýçûêå (á íåéíòíðûõ èäèíëåéòåõ).

7) Åååíðè÷àñòèÿ  
ñíâðåðåííííñ áèàà ìò íñíáû ïðåçáíñà (íðèéäÿ, óâèäÿ, åçäðííçäÿñü),  
êîòíðûå áííéíà óíòðåáéüíû á ñíâðåìåííííñ ýçûêå.

8) Åëàâíèüíûå ôéåêñèé -òè è -øè (ðíäèøè, ðíäèòè), êîòíðûå  
ðàðàéòåðíû áëý 18-âíí åâéèà è, åèäèíí, åíèæíû òðàéèòíâàòüñý êàé  
óðåéíâíñëåâÿèçìû.

Å ãðàììàòè÷åñêèò òàáëèöåò  
íâéíòíðûå èç ýòèò òíðì ìàþò ñíâðèàéüíû á ìíåðòû (óñòàðåâøàÿ,  
óðåéíâíñëåâÿèçì), ÷òíáû èõ ííæíí áúëè ìòëë÷èòü ìò íáùèò è ñíâðåìåííûõ òíðì).

### 3.3. Íðôðàðàòè÷åñêàÿ åàðèàòèâíñòü.

Óàðàéòåðííé ñíâðåìåííñòüþ  
ýçûêå 18&mdash;19 åâéèà íí ñðàâíåíèþ ñ ñíâðåìåííû ýâëÿåñý çíà÷èòåéüíàÿ ëåéñè÷åñêàÿ è  
íðôðàðàòè÷åñêàÿ åàðèàòèâíñòü. Òàì, ãää á ñíâðåìåííííñ ýçûêå çàôéñèðíâàí ìäéí  
åàðèàíò ñéíâà, á ñòàðûõ òâéñòåò ìíæíí áñòðåòèò ìññéíëüéí åçàéííçäíáÿíûõ  
åàðèàíòíâ. Íðôðàðàòè÷åñêèå íðàâèéà íáííííéðàòíí íáÿëëèñü á òâ÷âíéå 18&mdash;20 åâéíâ,  
à á ñíâðåìåííû áðåâíÿ ííâéà òâéñòû áúëè íáðâèçäàíû á ÷-àñòè÷íí ìäåðíéçéðíâàíííé  
íðôðàðàòè. Ííyòííò á ðâæëüíû òâéñòåò ìâæþâåðñý íðè÷óäéèå ñíâðåìåííû ðàçíûõ  
íðôðàðàòè, ñ êîòíðûõ á ñíâðåìåííû ñíâðåìåííû ñíâðåìåííû ñíâðåìåííû ñíâðåìåííû.

Íññáàñúå òî÷éè  
âàðèàòéâíñòè â ñòàðíé íðôíñåðàôèè &mdash; òå æå, ÷òî è â ñíñåðåìåííé:

- 1) Íáíçíà÷åíèå ìýäéíñòè äëÿ  
øèïýùèö (iå÷ü&mdash;iå÷, èääøü&mdash;èääø, íí÷üííé&mdash;íí÷íé);
- 2) Íáíçíà÷åíèå ,/í ññëå øèïýùèö è ö (÷, ðíúé&mdash;÷íðíúé, äüÿ÷, è&mdash;äüÿ÷íé,  
ø, è&mdash;øíé, èèöí&mdash;èèöå, çàéöíâ&mdash;çàéöåå);
- 3) ãëóõèå/çâííèå ñíñëæñíúå  
â íïçèöèè íáéòðåèèçàöèè (âíçñèÿöü&mdash;âíññèÿöü, âíñòîê&mdash;âíçòîê,  
ääðçèèé&mdash;ääðçñèèé&mdash;ääðññèèé&mdash;ääðñèèé);
- 4) óíñòðåáëåíèå íðñèñíúö  
áóéå (Áíäéèéñééé, Áíäéè÷àíå, Ñëàâåñíñéèé);
- 5)  
ñëèòííå/ðàçäåëüííå/ääôèñííå íàíèñàíèå (êòî íèáóäü, íâéèæíííó, íí ðóññêè, äí  
òéå, áøòðéó, òî-âñòü, êòî-áû&mdash;êòîíá, íâáîéñý, äíéäíëü);
- 6) íòðåæåíèå ôííåòè÷åñéèö  
éíëåáàíéé, íäíðèìåð è/ü (äûòàíéå&mdash;äûòàíüå, ñ÷àñòèå&mdash;ñ÷àñòüå,  
âàðèàíò&mdash;âàðüýíò, èòàëèÿíñéé&mdash;èòàëüýíñéé).

Äëÿ àíàëèçà íðôíñåðåòè÷åñéèé  
âàðèàòéâíñòè íáò óíéâåðñàëüííå íàòíäà, íí íæíí íðåäéíæèòü ðåøåíèÿ äëÿ  
éíéðåòíúö ñëó÷àåå:

- 1) Ííðíàëèçàöèÿ &mdash;  
íðåíáðàçíàíèå íàíèñàíèå â ííðíàëèçíàííóþ ôíðíó, êíòíðàÿ äàéåå åíàëèçèðóåòñý  
íí ñòàíáðòííé äðàííàòè÷åñéèé ííäåéé.
- 2) Đàñøèðåíèå  
ääðííàòè÷åñéèé ííäåéé &mdash; áéëþ÷åíèå äíííéíèòåëüíúö ôíðí â íàðåäéäííü.
- 3) Đàñøèðåíèå ëåéñè÷åñéèé  
ííäåéé &mdash; ñíåöèåëüíàÿ íòàöèÿ äëÿ çàëèñè ëåéñè÷åñéèé âàðèàòéâíñòè  
("ãèíåðåðåôåíü") èëè áéëþ÷åíèå äíííéíèòåëüíúö âàðèàíòíå â ñëíâàðü.
- 4) Ííäéòèéåöèÿ àëäíðèòíà  
äëÿ àíàëèçà ñëèòíúö íàíèñàíéé.
4. Íáíðååëåíèå ääëüíåéøåé ðàáíòü.

Äëÿ àíàëèçà òåéñòíå 18&mdash;19  
âåéå íáíåöíäèííñíçäåíèå ííñíñåðííé ñëíâàðííé áàçû, íðåæåå åñååííå, äðàííàòè÷åñéíäíñíçäåíèå  
éíòíðúé ííñíñåðííé åñå íññíñåðííé ñëíâàðííé ñëíâàðííé åñå åñååííå, ôíðíû, ôàðåéòåðííå äëÿ  
ýçûéå ääíííäííå íáðéíäà. Å èà÷åñòåå äéåéñè÷åñéèé áàçû äëÿ òåéñòíå ñëíâàðý íóæíí  
éññííèçíàäòü ñòàòèñòè÷åñéèå ääíííå, ííëó÷åííúå èç ðåàëüíúö òåéñòíå, à òåéæå  
ñëåäåðþùèå äåéñèéåðåòè÷åñéèå èñòíí-íééé:

&mdash; Ñëíâàðü Åéàäåíèè

Điññèéñêé (1789&mdash;1794);

&mdash; Ñëiâàðü  
öåðéíâíñëàâýíñéíâ è ðóññéíâí ýçûêà (1847);

&mdash; Íieíûé ðóññéèé  
íðôíâðàòè÷åñéèé ñëiâàðü (1898);

&mdash; Ñëiâàðü ðóññéíâí ýçûêà 18  
âåéà;

è äðóæèå ñëiâàðè.