

Íðíáæðiú íðåæñòàâæðíèý áóêâ â æðàøðííè÷åñéíí êíðíóñå èëè ìåòàêíðíóñå

Àâòìð Àëåâñàíð Âèêòîðàè Êâàëáíè
26.06.2008 á.
Íñéäåíå íáàëåíèå 16.07.2008 á.

Óâcèñû à ôîðìàòå RTF (57.87 kB 2008-07-13 21:44:21) Óâcèñû à ôîðìàòå PDF (121.06 kB 2008-07-13 21:43:45)

Էնիշանական պատճենները հայության մասին տեսական գիտական հարցերի վեհական առաջատար է և առաջարկություն է առ այսօն աշխարհում:

Â òi æå âðåìâý, ñ ìâîíîé ñòîðñíû, ñïçäåþþöñy êîðïóñâ, ñïáèðåþþùèå
ðåçëè-÷íûå ií âðåìâíàé è ìåñòó ìàïèñàíèÿ òåðñòû á ñâï, i ïñíáîí öäéëåâíî ðåçðåçå è íå
íåðåìâ-÷åíûå iäííé ðòðåæöèåé èéé iäíèì âðåìâíåí, òi áñòù yâéýþþùèåñy äèåðñíè-÷åñèè, è
èéé iðñòð iñåðøáíûè. N áðóåíè ñòîðñíû, ñòîðø ãíiðñ í ñïçäåíèé ìåðåéíðñíâ –
ñèñòåí, iïçâíèýþþùèò ìåííåðåìâíí ðåáíòåðòû ñ ååéèéòåíè ðòðåíáíéý iïçåâéñèíí
ðåçðåàíðåííûò êîðïóñâ. Íåá ýòe iðñâéàíàðèéé ðòðåáóþò ó÷, òà áññâò ðòðåáâàíèé è
ñèñòåíàí iðñâéñòåâééíèÿ ièñüìåííñòè (ÑII) ìòåéëüíûò ðòðåæöèé.

Íí ñóììèðíâàíèå òàêèò ÷àñòíûô
ííæåéàíèé ñóùâñòâåííí íàíýåò ñèòóàöèþ.

Âi-ïâðâûõ, âñèè â ìòäâëüíîé òðàäëëöèè
ðíëü êàæäíâí çíàëà ñìðíøí èçó÷åíà è àëôâàâèò íðâëñòàâëåí èçâåñòíûì ÷èñëèí
çíàëíâ, òi òâðíè÷åñêíâ ðåøáíèå äëÿ âñâé èñòîðè÷åñêíè ñïâîéóíñòè òðàäëëè
ëèøáíí ãïçíñæíñòè ïðâðåðûñü íà ñâí, åëâåâíâí èíòóéòéâíâí ñïñàíèå –
èñòîðè÷åñêè ñïèæèâøééñü àëôâàâèò. îèàçûñàâòñü, ÷òi, ñ ëàííè ñòîðííû, íèëíàäà
íåëüçý áûòü óââðåííû ã àïñòàòíííé ëëíòå íà ñïæñòâà ó÷ò, ííûõ çíàëíâ, à ñ äðóäíé ñòîðííû, ñàíí ñïñòâåòñòâèå çíàëà è
íçíà÷åñâíí íàÿðûñü, íðè÷åí cä ñ÷, ò íâåëò ñòîðíí.

Âî-âòîðûõ, íáíàðóæèåàåòñÿ, ÷òî ðàçíûâ îïæåéèàíèÿ íåÿâî
îïåðäçóïåâåþò ðàçíûâ óàðñíèëäè ðàáîðû, î ðåðäíèÿ, óää÷ûâ äëÿ íäíîé
òåðñíèëäè, îäóó êàçàöüñÿ íåöäíàùè äëÿ äðóáîé.

Â èòîâå, øòóðíóàìàÿ ñääïäíý ðàçíûè
èññéåâàìàòåðéÿè çàâàà-à ñïçääàíéÿ óíéâåðñàëüííâ iðåäñòàâéåíéÿ ñèìâíéíâ
íéàçüâåàòñv íá ïñòàâéåííé âî âñåð ãàæíûò àñíäéòà.

1. Íáðiðåäääë, ííñòü iðåääiaðòà.
Íáò ääëííáí iðåäñòàåäääíèý ääæäå í íåðå÷íá óðåäæööéé, êòîðûå äíææíà
íáññóåæäåàòü íèñêííàÿ ñèñòðàìà. Äèàëüacíí iðåääëííååíéò íí iðåääëñòàåäääíèh éèññéëëè-åññêèñ

ieññüìåííñòåé â ýòí ìòíñòåíè èøðëîé. Ía íäíí tïëþñá iù áéëæì íðääéëåéäíèý, èíåþùéå â áéëæó ðåçíüå èçâíäú öäðéëåíññëåáýíñéíåí ýçûéà (ðåññíðíñòåíýåíüå, â íðéíøëå, è ía ãëëåíéè-åñéíå íëñüí). Ía äðóäí – tïëäåððæéó áñåö áéëäå íëñüìåííñòé ía ééðééèé-åñéíé íññíåå, áéëþ-áý èñéñöññòåíåíññíçääííüå óæå â XX áåéåå àéôååéòú íáñéàáýíñéèó ýçûéå íaðíäíå NNND.

2. Íáíðåáâæé, ííñòú áäýöåðëüíñòúííáí
éííòåéñòà. Çääá-à ñòåâèòñý áíá éííéðåòííé äèñòëíëëíú ðåàáíòù èññëåäíâàòåëÿ.
Íáíáéí áëý éàæäéíé òåðííéíé ìëàòòíðíù óæå ñàïà òåðíé-âñéàÿ ííñòåíâéà
âííðñíà í NIII áûñëÿëèò íí-ðäçííó. Éðííà òííáí, éåæäàÿ äèñòëíëëíà ðåàáíòù
åééþ-àåò á ñääý ðäçíúå ñòíðííù áçæèííäéñòåëÿ -åéíâåéà è -åéíâåéà, -åéíâåéà è
ìàøéíú, áíóòðéíàøéííúò íðíòéíëíà, äëý éíòíðùò ðäçíúå NIII íí-ðäçííó óåíáíù èëëë
åáçðåçéè-íú. Òàéèí íáðäçíí, ñàïà ííñòåíâéà çàëä-è íà NIII äëý ííðíöñà ðòåáóåò íáçíðå
âñåé íáíðíæíé áäýöåðëüíñòé ñ éíðíöñii èéé íðé ååí íáííéíæé, ðòåíáíéè è
íóáééæàòéè, òåé è íðé ðåðåéèé íà ååí íáòåðéèåéò íáó-íúò çàëä-.

Éítíðiáòéèéè, áæäý íáíðóðåáééåííññóù çáääà÷è,
íá ááðóðóñý ðáðøàðöü ááç íáíðiáééííé éítíðiáòéè ìò óèëéíéíáíá, íí ñiaiè íá ðáðøàþþóñý
ñðåàáééðü çáääà÷ó ðéëéíéíáíá íá eò ýçûñé – á ðáðíðiáð óíííéíáéé, èñòñðé÷-ññééíé è
ñðåááééðåéüííé ééíááééññóðéèé. Óéëéíéíáé, éáðóñéá íí íðóðe íðíðåáééæáíéý íðåááéüíúð óáííéíáíéé,
íá áññáááá áæäýo íðíáéáíó á öäéíí è ñéááí íðåááñóááéýþò áíçííñéíúð ñíáééð òáðíé-áññééò
ðáðøáíéé áéáé á ìóííðáíéé íéàðóðíði, íá éíòñðóñ ñðóðíýòñý éíðíðiñà, òáé è á íáéáñòé
ððéðóðíáüñ ðáðøáíéé áéý éííðåðóííé íéàðóðíði. Á ðáçóðéüòàð áíùáý íðíáéáíà íá
íáúááéíýþùáé Níí ííáíáýáðóñý -áññóðííé íðíáéáíéý ñíçäááíéý óíéááðññéüííé ððéðóðíáíé
éíàéðíáéé á ðáìéáà ñòáíááðòá Óíééíá. Á ðáçóðéüòàð áðøáíéá ñòáíááéòñý íá ðíéüéí
íáéí ííéáçíúí áéý éíðíðiñíá íá öäéíí ðýäá íéàðóðíði, íí è íáññááíáíúí ìò ðáíááá
óðåááðæá, ííñ ðáðøáíéé ýóíé -áññóðííé çáääà÷é.

Ñëäâîàòåëüíí, ðåøâíéå íðîáëåìû äíëæíí
èòöe îí íðîáëåìû ñïíå áå íñóðøåíéý èéíñâèñòàíè è íðîáðàííèñòàíè. Å öåëíí
ííëíí áíáðèòú î áäoo àñïäóòå áå ðàññííòðåíéý.

À. Áíáøíá íæfíæfíâúâ çíàèè ðàçíûô
 òðääëëöèé ííäóó áûñíëíyòù ðàçíûâ ôóíëöèè, à áíáøíá ðàçíûâ – èääíòè=íóþ ôóíëöèþ.
 lïäðèìâð, íäíá è òà æå (â èíòöèòëâîíí ìíèìàíèè) áóéâà i íæåò â ðàçíûô òðàëëëöèýò èìåòü â ñååá 0, 1 èëè 2 òî÷êè; íi ïðè ýòíí
 óéðæíñéàÿ ï yäéyåðñý äðóáíé áóéâíé. Íòñþäà ñéååóâð, ÷òí ãððåè=âñéàÿ
 èääíòè=íñòü – íäíñòàðòð=ííá ìñííàíéà äëÿ ìòíæäñòâéíëý áóéâ ðàçíûô òðàëëëöèé,
 êâè è ãððåòð=âñéâíá ðàçëè=ëà – äëÿ èô ðàçëè=áíèý. Íðæòð=âñéè ýòí íçíà=âåâò, ÷òí
 Ñíi íæåò ðàçëè=àòü íæñíëüñí ñòúíñòâé, èòíðûâ ïðè áéçòàëëçàòëè â èííéðåòííé
 òðääëëöèé ííäóó áûñäëýåâòü íæfíæfíâ.

Á. Đàc cù - ái èå ñèl ài èå ñâ, ía ñâo ñèl ài èå ñâ
òi èü êåi ài èü êå - òi ñâ ñíi èò òå ñâ ñèl ài èå ñâ.
âi èü èå ñâ ñíi èò òå ñâ ñèl ài èå ñâ èò ñâ ñíi èå ñâ

Â. Íi â ðàçíûô ðòðàæöèÿô òàêàÿ
äïïíéèòåðæüïíñòû ê íäííé è ôíé æä áóêâå âðàôè÷âñêè ðàçíàÿ. Ðàññìàððèâàÿ íäíó
òðàæöèÿþ, iû íäééë ìïüçïâàðöñÿ âïñòàòî÷íû ëðèòåðèåì ðàçëè÷âíèÿ áóêâ:
âñêèé âðåðåðåùå âñòðå÷âþþñÿ â íäííò ðàæñòå â ðàçíîé ðîëè, òî èõ íóæíí ðàçëè÷àòü.
Ýòîò êðèòåðèé ñàï iñ ñåáá ðañïëüâ÷àò, è iðîòèâîðå÷èå, íí iðèìåíàíèå åäí î áóêâàì
èç ðàçíûô ðòðàæöèéë ïðñòò îòëðåàåò ðàçëè÷ëÿ: ñåðåñêàÿ “; ñîñëàñíí
êðèòåðèþ, íå ìòëè÷âðöñÿ ìò ýéòòñêîé “; ñåðåñêàÿ “; çíà÷èò, èõ âðåôè÷âñêîå ðàçëè÷ëå
ñòàïíâèòñÿ âïíðñò îí ìï÷âðêà. Îñâåéäíí, òàééïé éíäíè÷íû ê ãûñïà íå óñòðîò òåð
ïðæðèéñâ, èïòòðû ñïòòðò ñåðåñêèéë ñåðåñêàÿ “; ñåðåñêàÿ “; Arial Unicode ñåðåñêèéë òåðñò ñ êííìåíðàðèÿìè íà ýéòòñ

Ííáiáíúá ìðíòèáíðá-èý, áíçíéèáþþùéá
éíáíí á hâýçé ní níáðáééàl òððæééòéé, òððåáóþò óðíáàíáíòàéüííáí ðàçãíáíðá í òíí,
éâééàá níóúíñòé áíéæíú ðàçéè-àðöñý è éâééíáí áíéæíí áúðò èò nííòíðåíéá ñ
âéäéíúé ãððåòé-åññéééé ýéáíáíòàíé, ÷òíáú èñéííàÿ Níí áúðá áíóððáííá íáðíòèáíðá-èáíé
é óñòíé-éáíé è áíáááéáíéþ ííáúñ óððæéöéé. Yóí òððåáóáò óùáðáéüííáí
óíðíöééðíáíéý íðéíöéíá. ßçúé áéý òâéíáí ðàçãíáíðá óáíáíí ñòíðéòú íá íñííáá
èñííéüçóáíúñ ó ðýáá íáííðíáíúñ êíðíöñíá óððáíñééòåðåöéííúñ íðåáñòàáééáíéé, á êíòíðúñ
ññáé íáíçíá-áíéý ííóé-áþþ, íáíðíáíð, óâééá “;ñíóúíñòé”;, éâé óâáðáíéá è “çáãéèáíñòü”;
áóéáú, è êíðíðúñ ííçáíéýþò éâááéí áâáíáéòú ííáúá íáíçíá-áíéý éâé áéý
óíííéíáé-åññééó è áðáíáàðé-åññééó áñíáéóíá, òàé è íðíñóí áéý áððåòé-åññééó ðàçéè-èé.

2. Ñ äðóáîé ñòðííû, iïäáîòâåéå
iâæäèñòðíèéèíàðííå ðàçäåíâîðà ñíñòðíèò à iïñòðíâíè è ñèñòåííí ñòðóéòðèðíâåíè è
ääýöåäüñòðííé ñòðåíû ðàáíòû ñ òåéñòåíè ëîðíóñíâ. Ýòí iïíæåò óâéëåòü ðàçéè÷èå à
àéðòåäüñíñòð è ñíäåðæàíèò ìòäåëüñû ðòðåáâåíè è ÑÍ: à) íà ðàçíû ÿðäåò
ðàáíòû ñ êîðíóñíí, á) à êîðíóñàô íà ðàçíû òåðíè÷åññéò ïëàòòðíàô, á) äëÿ
ðàçíû çäåäà÷, ðåðàäåíû íà iàòåðèåëåò êîðíóñà.

Â ðâçóöüòàòå
ñòàïíâèòöny ÿñíí, -òî çàäà-÷ àääåéâàòííé âèçóàëëçàöè è ðâéñòà – íå
ååéèïñòååáíìàÿ è àäæå íå ñàïàÿ åâæíàÿ çàäà-÷, äëÿ ëòòðíé ñïçäà, òñý èïðíöñ.
Íàïðèïåð, äëÿ çàäà-÷ ëïòåéñòííà ïïñèñèà è ñòðååéòåéüùííà ñïññòååéàëëÿ
ôòååíäíòíà åíëåå öäiaíùè ìèäçüåàþþöny õïðíû ððäiaííéÿ ðâéñòà, ñïññíàíùå
åéåéî öääðæéåàòò íå òïéüéí íåíåöïäèíùå äëÿ åèçóåëëçàöè è åòåðëè, íí è ðàçíüå
åñïåéòò íåíåù, íííàí ñïñåðæäíèÿ. Äëÿ çàäà-, ñâyçäííûò ñ èçó-åíéàë ýçûéà è
åðååéèëë èïñéðåðííé ðòðååéëè, íåíåöïò, ñòàïíâèòöny åâæíùí ìòðåçèòü åñâå åâòåðëè,
ïðé-÷, í àäæå åíéåå ïïðíáíí, -÷í íå ýòò ñïññíàíí ëþáíà ðòèñèðííàíííà
éïñéðåíí-÷ííà ðåòåíéà. Ýòò ðäçääéåíéà ïïæåò áúòö yóooåéòéåíí ìðíååååíí à ñèòååéëè
íåòåéïòíñà, êíñää íåéåíéå ïïðíáíay éíòòðååéëÿ í ðâéñòà – ñòðåéòöny è
íåðåååòååååååñòå ïïðíååååðåíé íåðåé-÷ííà íåííòåååíí ëïðíöñà, à ïðíöååååðå

Ðáçóöüòàòòí î lâzäæéñööí ëëéíåðííáí
î áññööðäåíéý äïéæíá áûòù èïððåéòíäý îññòàííâéà çàääà÷è í ðàçëë÷åíéè ñèlâñëíâ íðë
éò ððåíáíéè å êïðöñå, íá ðåøäíéå èïlõíðé, íá ñâýçäííá ñ ääòåéëýle
ðåâéëçàööè, lñäéëé áú ïíèðåðüñý âññå ðåðííñëíäé÷åññéëå íëéàðòíðíû. Áûëí áú lñéåçíí,
åññéé íá íâlâñöäéíí, ñíâlâñöíí áûðåáíòàöü òàéöþ îòðíö ïðåññòàäéëíéý ñâñäííé
éïðíðåööè í ððåäëëöéý, èïðíðåý áú, ñ lññíé ñòðíðíû, áâéðäéà á ñâðå
ðåéöñéå äïñòðéæåíéý ýçüñéïçíáíéý lñí áññåð íâlâñöäéíñû ãëý ñíçäáíéý NÍÍ áñlâéòàð
éâæäíé ðòðåäéööè, à ñ áððåáíé ñòðíðíû, ñéóðæéèå áú ïññíâíé äéý ðåðíé÷åññéíâí
çàäääíéý IT-ñíâöéàëëñòàí á èâæäíé èç ñòùñâñðåóþùëö íëéàðòíðí. lññòðíé íâðåð÷åíü
íâlâñöäéíñû èçíâðåæåíéé ñèlâñëíâí áú, íá lñæåò ñéóðæéòü òàééíé ïðíðíé, ýâëëýñü ëëøü
÷åññíû, íðëçâññíû ìò íá, ðåçóöüòàòíí.

Problems of representing letters in a diachronic corpus or metacorpus

Alexander V. Kovalenin

A. P. Ershov Institute of
Informatics Systems of the Siberian Branch of the Russian Academy of Science,
Novosibirsk, Russia

To devise a character representation system for texts in a diachronic corpus is a more intricate problem than to represent characters of a certain written tradition. Both philological and technical difficulties appear. Proposed solutions often ignore this difference and implicitly have in mind a certain operational environment. Consideration of these issues is important for discussing the correct target setting and proper allocation of philologists' and IT specialists' responsibilities.