

Άϊτíæíηòè ìðèíáíáéÿ óíεááðñáεüííé ηέηòáíü ηέíòáεηέ-áñéíé ðáçíáòéè òáεήòá ObjectAT

Ááóíð Áίíá Áÿ-áñεááίáíá Νáòáðίáá
25.06.2008 á.
Γíηέááίáá ίáííáεáίεá 14.07.2008 á.

Òáçεήü á òίðíáòá RTF (43.45 kB 2008-07-14 21:00:56) Òáçεήü á òίðíáòá PDF (108.12 kB 2008-07-14 21:00:21)

Νέηòáíá

ðáçíáòéè òáεήòá ObjectATE εήίίεüçóáðñÿ á
Íòááéá εέίááεηòè-áñéíáí εήòí-íεέíááááéÿ ΕΔΒ ΔΑÍ äéÿ ηίçáááéÿ è ððáíáéÿ
ηέíòáεηέ-áñéíáí ίíεηáíéÿ áðááíáðòñêéè òáéíòáí: á ááííúé ίíáíò –
íáðááíáíé áíòίεíáεè «Í-áεü» è Εέááñéíé εáòίεηέ ίí Εíáòüááñéííò ηίηέó.

Íáøá

ηέηòáíá ÿáéÿáðñÿ Í-áíü áεáéíé è ίíáíòóíεèííáεüííé. Íðááíáçíá-áíá ίíá, Íðáæáá
áñááí, äéÿ ðó-ίíé ðáçíáòéè òáεήòá, ÷òí á ίáðáòð Í-áðááü è ððááóáðñÿ á ηέó-áá,
éíááá éíðíòñ òáεήòá ίáááεèè è ίá ηáíáíááí ò ðáçííáí ðíáá ίíεηíé,
ááòñíüñεáííüò è áííáüá òáííüò ίáñò è
ÿçúé, ίá éíòíðíí ίí ίáíεηáí, ÿáéÿáðñÿ ίáðáüí. Áá áεááίíé íòéè-εòáεüííé ÷áðòíé
ÿáéÿáðñÿ áíçíæííòü ãáíííò ίíεüçíááòáεð ίíεñüááòü Íðááéèá ðáçíáòéè á ðáíéáð
íáúáé ίíááéè. Á ηέηòáíá ίíæíí ðíñòí á ίíéííòüòð ðó-ίíí ðáæéíá ðáçíá-àòü
ίððòίεíáεð äéÿ Íðáááððεòáεüíí óæá ðáçááéáííáí ίá ηέíáíòíðíü òáεήòá:
íðεñááεááòü ηέíáíòíðíáí çíá-áíéÿ ίððòίεíáε-áñéèè éáòááíðéé. Εáííáòεçáðεð,
ááòííáñòáííáéè ίððòίεíáε-áñéèè ίáðáíáððíá, ηίçáááéá ηέíáíééíá è óéáçáòáééé
Íðááííεááááðñÿ áíááðεòü á ÿòò ηέηòáíò á áéεæáéøáá áðáíÿ, á ηáé-áñ äéÿ ÿòéè
òáééá è Íááéá εέίááεηòè-áñéíáí εήòí-íεέíááááéÿ íðéíáíÿáðñÿ áðóáíá Íðíáðáíííá
ηðááñòáí – Δáááéòíð áðááíáðòñêéè òáεήòá ATE. Óáéæá ðáçðáááòüáááðñÿ ηíííá íòíá-àòü ðáçííáí ðíáá ίáíðááεéüíí
òáííüá ίáñòá òáεήòá, ηáÿçáííüá ã ηέááíñíüñεáíéÿíé íεηòíá, ðáçííáí ðíáá
áíáéíéóðáíé è ò. í. Ááæá íðé
íòñòòñòáéè ίððòίεíáε-áñéíé éíòíðíáòéè (ò.á. áñéè òáεήò áñááí-íáñááí ðáçááéáí
íá ηέíáíòíðíü) ίíæíí ηðáçó çáíéíáòüñÿ ðó-ίíé ηέíòáεηέ-áñéíé ðáçíáòéíé òáεήòá,
ò.á. ηίçáááéáí á ááçá ááííüò ίíáíüò íáúáéòíá – ááéíéè ηέíòáεηέ-áñéíáí áíáéεçá,
éèè ηέíòáεηέ-áñéèè íáúáéòíá.

Νέíòáεηέ-áñééé

áíáéεç, éáé εçááñòíí, ίíæáò íðíáíáεòüñÿ ίí áðóííáí è ίí çááεñéííñòÿí. Áéÿ òíáí
-òíáí ðáçáéðáòü òáεήòó ίí çááεñéííñòÿí, ίíæíí ðíñòí ηίçááòü ηίεηíé áñáð
íáíáðíáéíüð òéíá ηέíòáεηέ-áñéèè ηáÿçáé è ίá-àòü ηίçááááòü («ñééáεááòü»)
ηέíòáεηέ-áñééá íáúáéòü, ηíñòíÿúéá εç ίáðü ηέíáíòíðí – ááðéíííé è ίíá-éíáííé
áé.

Ííæíí

çáíéíáòüñÿ áíáéεçíí ίí áðóííáí, ηíáéðáÿ òí-ίí òáééá æá ηέíòáεηέ-áñééá íáúáéòü,
ίí ίá òíéüéí εç ηέíáíòíðí, ίí è εç ηáíεð ηέíòáεηέ-áñéèè íáúáéòíá. Áéÿ ÿòíáí óæá ίáíáðíáéíí íðéíáíáéá ίáðáíεçíá òáé
íáçüáááíüð ίáññòðíáé. Íááñòðíééá – ÿòí ίíεñüáááííá ίíεüçíááòáéáí ίííæáñòáí
ηέíáíòíðí è ηέíòáεηέ-áñéèè íáúáéòíá, ίáéáááðúεè ίíðáááéáííüè ηáíεñòááíé.
Íáíðéíáð, ίááñòðíééá «Νέáçóáííá» áíεæíá áéεð-àòü á ηááÿ è ίáíó ηέíáíòíðíò, è ηέíòáεηέ-áñééé íáúáéò, ηíñòíÿúéé εç áéááíéá
è çíáíáíáòáéüííé ÷áñòé, ÷òíáú è òí, è áðóáíá, áííòñòéí, ίíæíí áúéí ηíááéíεòü ã
ίíáéáæáúéí äéÿ ηίçáááéÿ ίíáíáí íáúáéòá – ηáÿçé «Ííáéáæáúáá–ñéáçóáííá».

Ííæíí,

áíðí-áí, è éííáéíεðíááòü ίáá ίáðíáá ηέíòáεηέ-áñéíáí áíáéεçá, éáé ÿòí ááéááí íü:
íáéíòíðüá ηέíòáεηέ-áñééá áðóííü ηíáéðááí òáéééíí, ίí äéÿ áíéüøéíñòáá ηέó-ááá
íáðáíé-éáááíñÿ ίíñòðíáíéáí ηáÿçé ίáæáó ááððéíáíé.

Íáíáéí

áñéè ίððòίεíáε-áñéáÿ éíòíðíáòéÿ í ηέíáíòíðíáð äéÿ ðáçáéðááííáí òáεήòá

Iðéñóóñóáóáò (òàé, äëý ðàçáéðááíúò íàìè «Í-áéú» è Èèááñéíé èáóííèñè ìððíéíáéý óæá çàðáíáá ñááéáíá á áððáíé ñèñòáíá), òí Ìðè ñéíòáèñè-áñéíé ðàçíáðéá (-òí ì çááèñèííñòýì, -òí ì áððíáí) áá ìíæíí èñííèýçíááòú, òáí ñàíúí ýóò ðàçíáðéó òíðíúáy èèè, ì èðáéíáé ìáðá, ñíèæáy ááðíýòííñòú Ìðéáéè.

Äëý
ýòíáí íóæíí òàéæá èñííèýçíááòú ìáðáíèçì íááñòðíáé, òàé èáé ìí ìçáíèýáò çáááááòú òñéíáéý íá ìððíéíáé-áñéèá ñáíéñòáá ñéíáíòíðí, ìçáíèýýðéá áé èáðáòú ìðáááéáííòð ñéíòáèñè-áñéòð ðíèú. Ìðááííèíæèì, äëý ñéíáíòíðí, áóíáýúáé á íááñòðíééó Ñéàçóáííá, ìíæíí íáíáíðèòú, -òí ìíá áíèæáíá ìðááñòááéýòú ñíáíé èè-íúé áéááíé. Ìíæíí òàéæá çáááááòú òñéíáéý íá áóíæááíèý óáéíé ñéíòáèñè-áñéíé áððíí á íááñòðíééó (èáé ìðááéèì, ýòí áúááðò òñéíáéý íá áá ááððéíó). Óàé, ìíæíí ñòíðíóéèðíááòú, -òí äëý ñíñòááííáí ñéàçóáííáí ááððéíé áíèæáí áúòú áéááíé áúòè á èè-ííé òíðíá, òíòý áíçííæíú òñéíáéý è íá òíèúèí íá ááððéíó. Äëý óáíáñòáá ìíèýçíááòáéé ìðè ìíèñáíèè ìíáíáíúò òñéíáéè ìðááííèáááòñý áááñòè äëý ñéíòáèñè-áñéíáí íáúáéòá ìíýòèá áú-èñéýáíúò ìíèáé, ò.á. ìáðáíèçíá, áúáíáýúááí çíá-áíèý ìáðáíáòðíá ñàííáí ñéíòáèñè-áñéíáí íáúáéòá èç ìðáááéáííúò ìáðáíáòðíá ááí ìíèáé.

Ííèñúááý
òñéíáéý íá áóíæááíèá ñéíáíòíðí è ñéíòáèñè-áñéèò íáúáéòíá á ìðáááéáííá ìííæáñòáí, ìú ìíæáí èñííèýçíááòú ýòí ìííæáñòáí íá òíèúèí äëý ñíçááíèý ìíáúò ñéíòáèñè-áñéèò íáúáéòíá, ìí è äëý áúçíáá èç ááçú ááííúò ñíèñéá áñáò íáúáéòíá, óáíáéáòáíðýðéè ñòíðíóéèðíááíúí íáìè òñéíáéýì (íáíðèíáð, ñíçááòú íááñòðíééó «Ñíñòááííá ñéàçóáííá ñ íóéááíé ñáyçéíé» è ñðàçó æá áçáéýíóòú íá áñá ìíáíáíúá ñéàçóáííá).

Óàéæá
ìíæíí çáááòú íá òíèúèí òñéíáéý áóíæááíèý ñéíáíòíðí èèè ñéíòáèñè-áñéíáí íáúáéòá á íááñòðíééó, ìí è ìáðáíè-áíèý ñàííáí ñéíòáèñè-áñéíáí íáúáéòá – òñéíáéý ñí-áðáíèý ñáíéñòá á íáíí áóíáýúèò ñéíáíòíðí. Ñéáæáí, äëý ìíæáæáúááí è ñéàçóáííáí ìáíááðèááòñý ñíæèñíááíèá ìí èèòó á ñéááòðúáí áéáá: áñèè èèòí ñéàçóáííáí ìáðáíá, òí ìíæáæáúáá èèáí òíæá ñòíèò á ìáðáíí èèòá, èèáí ìðááñòááéýáò ñíáíé ñéíòáèñè-áñéèè ìíèú, è ò. á.

Áñá ýòí
Íçíá-ááò, -òí òàé èáé ìú ðáñííèááááí ìððíéíáé-áñéíé ðàçíáðéíé è ñèñòáíé ìíèñáíèý ìðíñòáéèò ñéíòáèñè-áñéèò ìðááéè ýçúéá (ìñííáíúò ìðááéè ñíæèñíááíèý è òíðááéáíèý, ñáyçáííúò ñ ìððíéíáé-áñéèì íáèèèì ñéíáíòíðí, á òàéæá ñ ìðýáèì ñéíá), òí óæá áíðáçáí íáíúòá ááðíýòííñòú áííóñòèòú ìðéáéò ìðè áíáéèçá. Ñéíòáèñè-áñéèè íáúáéò íá ñíçááñòñý, áñèè ááí -áñòè íá áíéáóò á ñíòááòñòáòðúéá íááñòðíééè (ñéáæáí, ñóúáñòáèòáéúííá á éíñááííí ìáááæá íá áóááò òðáéòíááòñý ñèñòáíè èáé ìíæáæáúáá) èèè áñèè ýòè -áñòè áíéáóò, ìí ìáðáíè-áíèý ñàííáí íáúáéòá íá ìíçáíèýò ñíçááòú ñéíòáèñè-áñéèè íáúáéò (áñèè, íáíðèíáð, èìý ñòíèò á èíáíèòáéúííí ìáááæá, ìí íá ñíæèñíááíí ñ èè-íúí áéááíèì ìí èèòó, èç íèò íá ìíèò-èòñý ñíçááòú íáúáéò «Ñáyçú ìíæáæáúáá–ñéàçóáííá»).

Á
áóáóúáí äëý ñèñòáíú ìéáíèðóáòñý ñíçááíèá ìáðáíèçíá ìáðáí-áòá áíáéèçá ìí çááèñèííñòýì á áíáéèç ìí áððíáí, ò.á. ñíçááíèá ìáðáíèçíá ñáíðá áððíí ìí ááððéíá (íáíðèíáð, áñááí ìðíñòáí ìðááéíæáíèý ìí ááí ñéàçóáíííó).

Íáèíèáííáý
éíòíðíáéèý ì ìðíñòáéèò ñéíòáèñè-áñéèò ìðááéèáð ýçúéá ìíæáò ìíçáíèèòú á ááéúíáéòáí ñááéáòú ñéíòáèñè-áñéèè ñáíð ìíèóááòííáòè-áñéèì, ò.á., èñííèýçóý éííñòðóèòíð íáúáéòíá, ñðàçó ðàçíá-áòú ìðáááéáííúá òðááíáíòú òáèñòá ìí ñòíðíóéèðíááíúí ìðááéèáí.

Іа́єііаò, òаè èаè іа́єіòіòúа èç òаèñòіа, ñ єіòіòúіè іù єіааі ааєі, іòаа́нòааèѳò ñіаіé іаòааіаú, ñєñòаіа ObjectATE іòааóñіаòòèаааò è ñòаа́нòаа іієñаіéѳ ñііòааòñòаèè іаæаó іòèаèіаєіі è іаòааіаіі, ò.á. єññèааііаіéѳ òаòаèòаòа іаòааіа-а́ñєіé ааѳòаèуііíòè ñіçааòаèаè іаіѳòієèа. Ааçа ааііуò іі «І-а́єа» ñіааòæèò іаòаèèаèуііуэ аòа-а́ñèèè òаèñò è ñієñіé аòа-а́ñèèè єаèñаі, ÷òі ііçаіéѳаò ðаçіаò-èéó ñіçааааòò іñіаúа іаúаèòò èèіа́аèñòè-а́ñєіаі аіаèèçа, òаè іаçúаааіуа òòааіаіòò іаòааіаа, ñòааѳ а ñііòааòñòаèа іаііé èèè іа́ñєіуèèè ñèааѳіñèèè ñèіаіòіòіаі іаіó èèè іа́ñєіуèèè ñèіаіòіòі òіèа́єіаèа, а òаèæа, а́ñèè іóаєі, єа́èñаіу іòèа́єіаèа.

Ñіèñіè
èèòаòаòòòú

Çіаієі è äð. 2006 – Çіаієі,
À. È. Óіèааòñаèуіаѳ ñèñòаіа ðаçіаòèè òаèñòа ATE-2
/ À. È. Çіаієі, À. À. Іаòèа́єіаа // Нíаòаіаіíуа
єіòіòіаòèèííуа òаòííєíаèè è ієñуйа́ііа іа́нєа́аєа: іò аòааієò ðóèіієñаé є
ѳèаèòòííуі òаèñòаі : іаòаòèаèуè іаæаóіаò. іаò÷і. єííò., Èæа́а́нè, 13-17 èпëѳ 2006 ä. – Èæа́а́нè, 2006. – Ñ. 51&n

Possibilities for applying
ObjectATE, a universal system for syntactic markup of text

Anna V. Sakharova

Vinogradov Institute of the Russian
Language of the Russian Academy of Sciences, Moscow, Russia

The linguistic opportunities provided by Object ATE, a universal syntactic annotation system, are described. The system allows the user to perform morphological and syntactic analysis of ancient texts. In the future a semi-automatic parser of Old Russian may be created. In cases where the text analyzed is a translation, the system will describe the degree of conformity of the translation to the original.