

ЄІДІОНЪ ОАДЄІАІІНЄААБІНЄЄО ОАЄНОІА А НІНОААА ІЄДБ, ІАДААБ ААДНЄБ: ІДІАЄАІ

АаоїѠ Аєаєнає Ааааїуаає- Іїє҃҃҃҃҃҃
27.08.2012 а.
Іїнєааїаа іаїїаєаїєа 09.11.2013 а.

Summary. The paper discusses the issues of spelling, grammatical tagging and metatext markup for a pilot version of the Church Slavonic Corpus (started in 2011) and the prospects of its development.

1. Єїдїонъ

оаѠдєіаїїнєааѡ҃҃҃҃҃҃ оаєноіа[1].
Çà 2011 аїа аѠоїїіє єннєааїїааѠаєає[2]
нїçааї є Ѡаçїауаї а ЄїѠаѠїаѠа (їїєа єааѠ іѠєааєа іа ааѠаңа <http://ruscorpora.ru/beta/search-orthlib.html>)
їєєїѠїує ааѠєаїѠ іѠааңѠааєѠаєуїїаї ії іауаїѠ єїѠїѠа нїаѠєаєуїїї іаѠаçїї
їїааїѠїаєаїїѠѠ оаѠдєіаїїнєааѡ҃҃҃҃҃҃ оаєноіа. Оаєноу нїааæаїу нїаѠєаєуїїіє
ѠаçїаѠѠіє (їаѠаѠаєноїаїє є аѠаїїаѠє-аңєїє), єѠ іауєє іауаї 1250 оаєноїа, ієїєї
4,6 ієєєєїїа нєїаїѠїѠѠаєаїєє, аєєр-аруєѠ ієїєї 150 ѠѠ. Ѡаçєє-їѠѠ нєїаїѠїѠ.
Ааїїує єїѠїѠъ яаєяѠѠнъ іїаїє -аңѠѠр іѠїаєѠа "їаѠєїаєуїїує єїѠїѠѠ ѠѠннєїаї
ѡçуєа" (<http://ruscorpora.ru/>).

2. Оаєноу

єїѠїѠа є іаѠаѠаєноїааѡ҃҃҃҃҃҃ ѠаçїаѠѠєа. ІїнїаѠ єїѠїѠа оаѠдєіаїїнєааѡ҃҃҃҃҃҃
оаєноїа а нїноааа ІЄДБ нїноаає҃҃҃ нїаѠаїаїїуа аїаїнєѠааїїѠѠ оаєноу (XIX–XX аа.)
(60%). ЄѠїїа Ѡїаї, а єїѠїѠа іѠааңѠааєаїу аѠѠаєа іаѠєїау (XVII–XVIII аа.)
є æаїѠѠ оаѠдєіаїїнєааѡ҃҃҃҃҃҃ оаєноїа: ієнїаїєа, нїаѡїѠѠа-аңєєа, іѠааїауа, іаѠѠ-їѠа,
є аѠ.

ÆаїѠїаї-ѠаїаѠє-аңєаѡ҃҃҃҃҃҃ єєаңнєѠєєаѠѠє
оаєноїа, іа нїнїаа єїѠїѠіє іїєуçїаѠѠаєу іїæаѠ нїçааѠѠ іїєуçїаѠѠаєуїїнєїаї
їїаєїѠїѠѠ, ааѠаїїаїа, а іаѠаѠр і-аѠаау, іаңнїаїїѠ іїєуçїаѠѠаєр, а іа ѠçєїїѠѠ
нїаѠєаєєнѠѠ, є нѠѠїѠѠнъ єнѠїаѡ҃҃҃҃҃҃ єç іѠєїѠєїїа іѠаєѠє-аңєїаї ѠаїаїѠѠаа є
їїїѠѠїїѠѠ. ІаѠаѠаєноїааѡ҃҃҃҃҃҃ ѠаçїаѠѠєа, нїнїаїаѡ҃҃҃҃҃҃ єїѠѠаїѠѠєу єїѠїѠіє ѠаçѠааїѠѠаїа А.А. ЄѠаааѠѠєї, а іаңѠѠїуаа аѠаїѡ҃҃҃҃҃҃ іѠїаїаєѠнъ
Ѡїєуєї іа ѠѠїаїа Ѡаєїаї оаєноѠа є нїнѠѠѠѠ єç нєааѠѠрѠѠѠ іаѠаїаѠѠїа (їаѠа-єнє҃҃҃҃҃҃нъ
єїѠїѠнїїа «ѡѠєуєє»):

-
«їєнїаїєа» ѡїѠ
Аєаєєу, НєѠæааїїа аааїааєєа є іїааїѠѠє іаѠєїєє а аїаїнєѠааїїѠѠ єїєааѠ, аңєє іїє
аїааааїїѠ а іѠааєуїїѠѠ ѠѠаѠєѠѠ;

-
«нїаѡїѠѠа-аңєєє»;

-
«Ѡєїїєєїї»;

-
«нєѠæаа» ѡїѠ аңа
аїаїнєѠааїїѠѠ -єїїу є нєѠæаау, а Ѡаєæа іїааїѠѠє аїаїнєѠааїїѠѠѠ оаєноїа
(аїаїѠїає-їѠѠ, єїїааєє є Ѡ. а.) а нїноааа ѠаçїѠѠѠ нїаїѠїєєїа;

-
«аєаѠѠѠѠѠ»;

- «íðááí» ýòí òáéñòú òáðéíáííó íðááó, íáíðéíáð, «Éíéáá Íðááéé ñáyòúó áííñòíéíá, ñáyòúó ñíáíðíá, áñáéáíñééò è ìíáñòíúó, è ñáyòúó íòóíá»;

- «íàó-íúé» íáíðéíáð «Éòééà Éáðííéèòééà».

Á áóáóúáí íðááííéáááòñý ñíçáàòú íáòáðàçíáòéó íá óðíáíá íòááéúíúó óðááíáíóíá óáéñòá (íáíðéíáð, éííááéé, óðííáðé), éíòíðúá áóíäýò á ñíñòáá áíéúðéó óáéñòíá.

Òáéæá áíñòóíáí íòáíð óáéñòíá ìí íáðéíáó ñíçááíéý, òí-íáá, ìí ñéááóðùèí íàðáíáòðáí:

- «àððáè-íúé òéí» íáíðéíáð, «Áíáðíòíèðáéá»;

- «áéáðéáíúé òéí» íáíðéíáð, «Áéòááèò Áóóíáíúé»;

- «ñòáíáàðóíúé òéí» ýòí áñá óáéñòú ìíííáíúó áíáíñéóááíúó éíéá, çà èñéèð-áíéáí óáéñòíá XX ááéá;

- «XX ááé» ýòí, á íáðáóð í-áðááú, àéáòéñòú, íáíðéíáð «Áñáí ñáyòúí, á çáíéá Ðíññééñéíé íðíñéýáðéí».

É ñíæáéáíèð, íáò ìíñéá ìí éííéðáòíé áàòá ñíçááíéý èéè èçááíéý óáéñòá, óáé èáé ìíðááééáíéá óáéèò áàò áéý íííáéò óáéñòíá óðááóáò éðííðééáíé ðááíòú, á éííáá è áíáñá íá íðááñòááéýáòñý áíçííáíúí, áíéáá òíáí, áéý óáò óáéñòíá, ááá áàòú ýñíú, íááíçííáí íðáááéèòú, á éáéíé ìíáíò á ìíáíò ñíçááíéý èéè á ìíáíò èçááíéý íðíçáíáééáñú ìíñéááíýý ýçúéíááý íðááéá.

3. Éíáéðíáéá óáéñòá íðíáéáíú íðóíáðáòéè. Á íðíóáññá ìíáííóíáéè áúéè áúðááíòáíú íðéíòéíú è íðááééá íðááñòááéáíéý óáðéíáííñéááýíñééò óáéñòíá á éíðíóñá, á óáéæá ðáááíú íðáéòé-áñééá íðíáéáíú, èáñáðùéáñý ðáçíáòéè è éíáéðíáéè óáéñòá. Éáííé-áñéáý óáðéíáííñéááýíñéáý íðóíáðáòéý, íðéáíòéðíááíáý íá òéííáðáòñéíá íðááñòááéáíéá óáéñòá, áéèð-ááò ìííáí ñéíáíéíá, éíòíðúá óáéòé-áñéé íá íáñóò ñíúñéíðáçéè-èòáéúííé óóíéòéè. Áéý íðááñòááéáíéý éíðíóñá á éíòáðíáòá áúéá áúðááíòáíá áíéáá íðíñòáý íðóíáðáòé-áñéáý ñéñòáíá, éíòíðáý ñíòðáíýáò ñóúáñòááíúá ýçúéíáúá ðáçéè-éý, íí íá íúòááòñý éíèòéðíáòú òí-íúé òéííáðáòñééè áéá óáéñòá. Á ýòíé íðóíáðáòéè íòñóòñóáòò íáéíòíðúá ñéíáíéú ñ íàéíé èéè íóéááíé ðáçéè-èòáéúííé ñíííáííñòùð íðéáúòáíéý, ðáçéè-éá íó/ó, ³á/ý, íáíáéí ñíòðáíýòñý ñííááííñòé, ñáyçáííúá ñ ðáçéè-áíéáí éáñéñé-áñéíáí çíá-áíéý í/íááá, è³/v, ó/òéòá, ç/³4. Éðííá óíáí, áúéè áúðááíòáíú íðááééá íáðááíáá óáéñòá èç ýòíé íðóíáðáòé-áñéíé ñéñòáíú á ñíáðáíáíúé áéá áéý óáíáñòáá ìíñéá.

Ñéíáíóíðíú á éíðíóñá íðááñòááéáíú óáé

æá, éáé á ēñōíáííí òáēñòá, á éáííú ááþōñý á óíēðēðēðíááííí íáíēñáíēē, á ēíōíðíí íá ēñíēúçōþōñý ēçáúōí-íúá áóéáú, á òēðēá ðáñēðúòú (íáíðēíáð, ñēíáí íēñòú ðáññíáòðēááðñý éáē òíðíá ē éáííá íēēíñòú).

4. Áðáííáðē-áñēíá

ííēñáíēá. Áðáííáðē-áñēíá ííēñáíēá òáðēíáííñēááýíñēíáí ñēíáíēçíáíáíēý íáíáóíáíēíí äēý ðááíòú éáēñēēí-áðáííáðē-áñēíáí ííēñēá, á á íáðñíáēðēáá ē äēý ñíçááíēý ááòííáðē-áñēíáí íðōíēíáē-áñēíáí áíáēēçáòíðá òáðēíáííñēááýíñēíáí ýçúēá.

Νούáñòáóþùēá áðáííáðēēē ē ñēíááðē

ííēñúááþò íáēíòíðòþ ēááēúíòþ ēáðòēíó, ēíòíðáý íòðáæááò ííðíáēēçáòíðñēēá òñòðáíēáíēý ááòíðíá, íí ÷áñòí íá ñíòááòñòáóáò ðáēðē-áñēííó ñíñòíýíēþ ýçúēá. Ðááēúíí á òáðēíáííñēááýíñēēð òáēñòáð íááēþáááðñý íííæáñòáí íðóíáðáðē-áñēēð ē ñēíáíēçíáíēðáēúíúð ááðēáíóíá, íáçñēíáēáíúð òáí, ÷òí òáēñòú ñíçááááēēñú á ðáçííá áðáíý ē á ðáçííē ýçúēíáíē ñðááá. Íýçúíó, áñēē ēñōíáēðú ēç ēçó-áíēý áðáííáðēē ē ñēíááðáé, ÷áñòí íáēúçý ííýçú, éáēēá òíðíú ēíááò íáēíòíðíá ñēíáí. Á òáēēð ñēó-áýð ááēíñòááíúíú áññòíááðíúí ēñòí-íēēíí ýáēýáðñý ēíðíóñ.

Íáíðēíáð, ñ-ēòááðñý, ÷òí äēý íðēē-ēý

íí. ē áá. ÷ēñēá íò ííííēíē-íúð òíðí áá. ÷ēñēá ēñííēúçóáðñý íáēá-áíííá òááðáíēá (Λ) ē çáíáíá í&warr;w, á&warr;º. Νíðáðēááðñý, éáē áóááò ēí. íí. íò áúñíòá áúñíòú^, áúñwòú’ ēēē áúñwòú^ ? Íéαçúááðñý, á ēíðíóñá áñòðá-áþòñý áñá òðē òíðíú, íðē-áí íáðááý 9 ðáç, áòíðáý 17, òðáòúý 1. Ęç áíáēēçá áðóáēð áíáēíáē-íúð ñēó-ááá íú íæáí ñááēáðú áúáíá, ÷òí íðē ēííēóðáíòēē íðááēē çáíáíá í&warr;w íðááíí-ðēòáēúíáá, ÷áí íáēá-áíííá òááðáíēá.

Ę íáñòíýúáíó ííáíóó áúēá íðíááēáíá

çíá-ēòáēúíáý ðááíòá, ðáçóēúðáòíí ēíðíðíē ýáēēáñú íēēíðíáý ááðñēý áðáííáðē-áñēíáí ñēíááðý. Áúēí íðíáíáēēçēðíááíí íēíēí 150 òúñ. ñēíáíðíðí, áíēúðēíñòáò ēç íēð áúēá íðēíēñáíá íñííáíáý áðáííáðē-áñēáý ēíðíðíáðēý: éáííá, ÷áñòú ðá-ē, áðáííáðē-áñēēá íðēçíáēē. Áíá-áēá áúēá ñááēáíá ðó-íáý éáííáðēçáðēý äēý ðýáá ñēíáíóíðí, á çáòáí íá íñííáá ííñòðíáííē ííááēē ñēíáíēçíáíáíēý áúēē íðíáíáēēçēðíááíú ññòáēúíúá ñēíáíðíðíú. xáñòú ñēíáíóíðí áúēá éááēēðēðēðíááíá éáē íðēáēē (ííá-áòēē). Äēý íáēíòíðúð ðááēēð òíðí éáííó ííðáááēēðú íá òááēíñú ēēē ííá áúēá ííðáááēáíá áēííðáðē-áñēē. Ííēñē íí íáðáíáíúíú òáðáēðáðēñòēēáí, òáēēí éáē íáááæ ēēē ðíá (äēý íðēéáááðáēúííáí), ðááíòááò, íí ááç ñíýðēý áðáííáðē-áñēíē ííííēíē.

Áðáííáðē-áñēēá òááēēðú, íá ēíðíðúá

ííēðááðñý ááòííáðē-áñēēē áíáēēçáòíð, ñòðíýòñý íá íñííáá áíáēēçá ēíðíóñá òáēñòíá ē ñóúáñòáóþùēð áðáííáðēē. Ííñēíēúéð áðáííáðēēē ááþò íáííēíóþ ē íðíðēáíðá-éáóþ éáððēíó ñēíáíēçíáíáíēý, òí òíēúēí ēíðíóñ ýáēýáðñý íēíí-áòáēúíúí ēðēòáðēáí ēñòēíú.

Á íòēē-éá íò òðááēðēííúð áðáííáðēē,

íáðááēáíú íá áúēē çáááíú áíðēíðíí, á áúááááíú ýííēðē-áñēē íá íñííáá áíáēēçá íííæáñòáá ñēíáíðíðí, ēíáþúēð íáííðēííá ñíòííðáíēá íáæáó áðáííáðē-áñēēíē òíðíáíē. Óáēēí íáðáçíí, ííáíēááòððá íáðááēáí ííēó-ááðñý çíá-ēòáēúíí áíēáá ááòáēúííē, ÷áí ððááēðēííáý, ē ííá-áñ ñóúáñòááííí íò òáēíáíē íðē-ááðñý. Íáíðēíáð, òðááēðēíííá íáðáíá ñēēíáíēá (ðááú) íá ñáííí ááēá ðáñíáááðñý íá 14 ííáðēííá á çááēñēííñòē íò ēííá-ííáí ñíáēáñííáí (íáðíúé òááðáúē –íýáēēē, ááēýðíúé, øēíýùēé, étò), íáēē-ēý áááēíáí áēáñííáí ē áðóáēð íñíááííñòáé (³áðáé – ³áðáá, ³áðáíú, áááðýíēíú – áááðýíá).

5. Íáðñíáēðēáú òáðēíáííñēááýíñēíáí ēíðíóñá. xòíáú ēíðíóñ ííá á ííēíē íáðá óáíáēáòáíðýòú ííððááííñòýí ñíáðáíáííē íáòēē ē íáðáçíááíēý á íáēáñòē

īōīnōōāīnōāā ēīōīōīē nōūānōāōāō āānū īāōēīāēūīūē ēīōīōn dōnēēīāī yçūēā.

[2]
Íīēīī āāōīōīā āāīūō òāçèñīā ā nīçāāīēē òāōēīāīīnēāāyīñēīāī ēīōīōnā īā òāçīūō
ýòāīāō āāī òāçōāāīōēē ā ēā-ānōāā nīçāāōāēēē ēīōāīōēē, òāçōāāīō-ēēīā ē
ēñīēīēōāēēē òāōīē-āñēēō çāāā-, ā òāēæā ēīīnōēūōāīōīā ē āāīōīāēōāēēē òāāīōū
īōēīāēēē ó-āñōēā Á.Á.īēōīāyī, Á.Á.Ēōāāōēēē, Á.Ē.Çīāīēī, Á.Á.Æēōīāā,
Á.Á.īēāōīāāā, Ē.Ē.īāðōāāā, nāyū. Ō.Á.Ēpāīāīānēēē, Ð.Í.Ēðēāēī, Ē.Á.Ñāāāēīāē-,
nāyū. Ē.Í.īīēūñēīā.