

Корпус церковнославянских текстов в составе НКРЯ

Тексты.

Современные богослужebные тексты
(библия, служба, типикон, святоотеческие, церковное право)

1250 текстов

Объем = 4.6 миллиона словоупотреблений,
150 тыс. различных словоформ

Источник

Библиотека святоотеческой литературы (<http://orthlib.ru>)

Проблемы:

- нет поиска
- нет лемматизации
- кодировка NIP

Метатекст.

- 1) Заголовок текста (+название книги).
- 2) Количество слов.
- 3) Жанр текста:
 - а) библия: ветхий завет, новый завет, служба (17%)
 - б) служба: акафист, ирмологий, минея, октоих, служебник, требник, триодь, часослов, прочее (64%)
 - в) типикон (5%)
 - г) святоотеческий (10%)
 - д) право (1%)
 - е) научный (0.5%)
- 4) Период создания текста:
 - а) стандарт (XIX-XX = 86%)
 - б) архаика (XVII-XVIII — Пролог, Добротолубие = 2%)
 - в) 20 век (3%)
 - г) гибридный (8%)
- 5) Перевод или оригинал.

Издания

Книга	Год издания	Слов	%%
Библия	1900 (1993)	650 253	14.36%
Евангелие	1984	88 246	1.95%
Апостол	1989	105 236	2.32%
Ирмологий	1913 (1995)	46 641	1.03%
Канонник	1986	1 284	0.03%
Минея	1893 (1996–97)	1 429 651	31.57%
Минея общая	2002	102 493	2.26%
Минея праздничная	1914 (1993)	188 819	4.17%
Минея (из разных источников)	–	13 923	0.31%
Молитвы читаемые на молебнах	1915 (1993)	146 598	3.24%
Молитвы (из разных источников)	–	15 249	0.34%
Октоих	1981	254 655	5.62%
Псалтирь следованная (Часть I)	1978	71 440	1.58%
Служебник	1896	65 385	1.44%
Требник	1906	108 325	2.39%
Триодь постная	1992	217 295	4.80%
Триодь цветная	1992	131 010	2.89%

Часослов	1980	29 434	0.65%
Типикон	1900	246 784	5.45%
Алфавит духовный святителя Димитрия Ростовского	1837 (1997)	38 098	0.84%
Добротолубие, или слова и главизны священного трезвения	2000	294 909	6.51%
Книга Правил святых апостолов, святых соборов и святых отец	1893 (1992)	75 513	1.67%
Ифика иерополитика, или Философия нравоучительная	1764	39 937	0.88%
Пролог (старообр.)	1643 (1915)	120 009	2.65%
Акафисты (из разных источников)	—	47 597	1.05%

Кодировка.

НІР

_o='ч~е на'шъ, и='же _e='си` на нб~сјь'хъ, да ст~и'тся и='мя твое` : да прии'детъ цр\ствіе твое` : да бу'детъ во'ля твоя` , ја='кѡ на нб~си` , и= на земли` : хлјь'бъ на'шъ насу'щный да'ждь на'мъ дне'сь: и= w='ста'ви на'мъ до'лги на'шя, ја='кѡ и= мы` w='ставля'емъ должникѡ'мъ на'шымъ: и= не введи` на'съ въ напа'сть, но и=зба'ви на'съ w\т лука'вагѡ: ја='кѡ твое` _e='сть цр\ствіе и= си'ла и= сла'ва во вѣ'ки. А=ми'нь.

Unicode

о́че на́шь, и́же еси́ на нбсѣхъ, да стѣ́тса і́ма
твоѣ́: да прѣ́идеть црѣ́ствіе твоѣ́: да бѣ́детъ во́ла
твоѡ́, я́кѡ на нбси́, и на земли́: хлѣ́бъ на́шь
насу́щный да́ждь на́мъ дне́сь: и вста́ви на́мъ
до́лги на́ша, я́кѡ и мы́ вставля́емъ должникѡ́мъ
на́шимъ: и не введи́ насъ́ въ напáсть, но изба́ви
насъ́ ѿ лука́ваго: я́кѡ твоѣ́ е́сть црѣ́ствіе и си́ла и
сла́ва во вѣ́ки. Ами́нь.

Орфография.

0) Каноническая

- типографски точный вид (придыхания, оу/8, іа/а, букво-титла)
- многие символы не несут смысловоразличительной функции

1) Точная

- точный вид словоформы (є/ѡ широкое, титла, ударение).

2) Упрощенная

- лексические оппозиции в лемме (о/ѡ, и/і/ѡ, ф/ѳ, з/ѕ, ударение).

3) Модернизированная

- только современные буквы.

Проблема

Символы Unicode 5.2 (8, іа, букво-титла, паерок) – отсутствуют в шрифтах, не отображаются.

Символ	Каноническая (типографская)	Точная (словоформы)	Упрощенная (леммы)	Модернизи- рованная
придыхание	(есть)	(нет)	(нет)	(нет)
букво-титла	мл ^с ть, блг ^д ть	мл ^с ть, блг ^д ть	милость, благодать	млсть/ милость
паерок	без ^з язычный	без ^ъ язычный	безъязычный	
титла	м ^т и, аг ^г ль, бл ^г дть	м ^т и, аг ^г ль, бл ^г дть	мати, ангель, благодать	мати, ангель благодать
оу- / -8-	оучити, нав ^ч ити	учити, научити	у	у
іа- / -а-	іати, вб ^ъ ати	яти, вб ^ъ яти	я	я
ѣ	сѣль	ѣ	ѣ	е
є- / -є-	єзеро, кон ^ѣ мь	є (єзеро)	е (єзеро)	е
о- / -о-	о ^т ецъ, пра ^о тець	о (о ^т ецъ)	о (отець)	о
w во флекс.	раб ⁺ wмь	раб ⁺ wмь	(нет)	о
w в корне	ід ^w ль, ік ^w на	w	w	о
w в префиксе	wб ^р азую, wб ^р азъ /об ^р азъ	w	w	о
̄w	̄wдати, ̄wати	̄wдати, ̄wяти	wтдати, wт ^ъ яти	от-
і+согл.	ід ^w ль, м ^і ръ, в ^і на	і	і	и
і+глас.	знаніе	і	і	и

и	миръ	и	и	и
v v̄	мvро, єvсv̄гнiй	v	v	и / в
ѳ	анѳраѣъ	ѳ	ѳ	ф
ѣ	анѳраѣъ	ѣ	ѣ	кс
ψ	апокалv̄ψiсъ	ψ	ψ	пс
s	сло, smiа	s	s	з

Словоформы – реальный вид (езеро, мл^сть)

Леммы – нормализованный вид (езеро, милость).

Орфографические колебания:

1) i/i/v — ассvрійскiй/ассирiискiй;

2) o/w — вvссонъ/вvссонь, образъ/wобразъ—wобразовати, воwобразити;

3) геминаты — вис(с)арiшнъ, бжѣствен(н)ый, иноплемен(н)икъ, молитвен(н)икъ, камен(н)одѣлатель;

4) ь/ноль — довол(ь)ный, начал(ь)никъ, жител(ь)ствовати;

5) паерок/ноль — без(ъ)пристрастiе, без(ъ)wобразный/без(ъ)образный, возъму/возьму/возму.

Грамматическое описание.

Грамматические категории.

Часть речи:

S=существ. A=прилагат. ADV=наречие V=глагол

PREP=предлог PART=частица INTJ=междометие

S-PRO=мест.-сущ.(кто) A-PRO=мест.-прил.(уже)

Число:

sg=ед. pl=мн. du=двойственное

Падеж:

nom=им gen=род acc=вин dat=дат loc=мест ins=твор voc=зват

Степень сравнения:

comp=сравнительная (*сильнѣй, сильнѣйшій*), нет super

Время:

praes=наст. fut=буд. praet=прош. (только причастия),

imperf=имперфект

aor=аорист

perf=перфект(ное причастие) (*ходилъ*)

Грамматические таблицы (парадигмы).

Парадиг- ма	N1t	N1t*	N1j	N1k	N1x	N1k*
Пример	раб+ъ	осе*л+ъ, со*н+ъ	кон+ъ, цар+ъ	отро(к ц ч) +ъ	ду(х с ш)+ ъ	свит(к ок ц ч)+ъ
sg,nom	ъ	2ъ	ь	ъ	ъ	2ъ
sg,acc	=nom/gen	=nom/gen	=nom/gen	=nom/gen	=nom/gen	=nom/gen
sg,gen	а	а	я	а	а	а
sg,dat	у	у	ю	у	у	у
sg,loc	ѣ	ѣ	и	2ѣ	2ѣ	3ѣ
sg,ins	омъ	омъ	емъ	омъ	омъ	омъ
sg,voc	е	е	ю	3е	3е	4е
pl,nom/voc	и	и	и/іе	2ы	2и	3ы
pl,acc	ы/=gen	ы/=gen	и/=gen	и/=gen	и/=gen	и/=gen
pl,gen	ѡвѣ/ѣ^	ѡвѣ/ѣ^	ей	ѡвѣ	ѡвѣ	ѡвѣ
pl,dat	ѡмѣ	ѡмѣ	ѣмѣ	ѡмѣ	ѡмѣ	ѡмѣ
pl,loc	ѣхѣ	ѣхѣ	ехѣ	2ѣхѣ	2ѣхѣ	3ѣхѣ
pl,ins	ы	ы	и/ьми	и	и	и
du,nom/acc	а^	а^	я^	а^	а^	а^
du,gen/loc	у^	у^	ю^	у^	у^	у^
du,dat/ins	ома	ома	ема	ома	ома	ома

Парадигма	N1s	N1sj	N1c*	N1a	N1i	N1e
Пример	муж+ъ	врач+ь	от(ц ец ч)+ ъ	кра+й	агапі+й	іере+й
sg,nom	ъ	ь	2ъ	й	й	й
sg,acc	=nom/gen	=nom/gen	=nom/gen	=nom/gen	=nom/gen	=nom/gen
sg,gen	а	а	а	я	а	а
sg,dat	у	у	у	ю	ю	ю
sg,loc	и	и	ѣ!	и	и	и
sg,ins	емъ	емъ	емъ	емъ	емъ	емъ
sg,voc	у	у	3е	ю	ю	ю/е
pl,nom/voc	и/іе!	и/іе!	ы	и^	и^	є
pl,acc	ы/=gen	и!/=gen	ы!/=gen	и^/=gen	и^/=gen	и^/=gen
pl,gen	ей	ей	євъ/ъ^	євъ	євъ	євъ
pl,dat	ємъ	ємъ	ємъ	ємъ	ємъ	ємъ/ѡмъ
pl,loc	ахъ	ахъ	ѣхъ!	ехъ	ехъ	ехъ
pl,ins	ы	и^!	ы^!	и^	и^	и^
du,nom/acc	а^	а^	а^	я^	я^	а^
du,gen/loc	у^	у^	у^	ю^	ю^	ю^
du,dat/ins	ема	ема	ема	ема	ема	ема/ома

Проблемы.

Грамматики описывают идеальную картину.

Много орфографических и словоизменительных вариантов.

Единственным достоверным источником является корпус.

Вопросы.

1) pl/du vs sg = облеченное ударение (^) или замена o→w, e→є.

высо́ты — вы́соты̂ (9), вы́сѡты̂ (17) или вы́сѡты̂ (1)

ми́лости — ми́лѡсти (93), ми́лѡсти (2)

єле́ни — єле́ни (8)

2) pl, nom = и (кони), е (іерее), іе (царіе)

дѣ́латели (36), дѣ́лателе (13), дѣ́лателіе (18)

3) церковнославянские vs. русские формы:

зако́ны (15), зако́нами (1), зако́нами (9)

зако́нѡмъ (59), зако́намъ (4)

зако́нѣхъ (15), зако́нахъ (1)

зако́нѡвъ (71), зако́нѣ (10)

Грамматический словарь (1).

лексема	часть	словоформа	граммемы
благій	А	благь	brev,sg,m,nom/acc
благій	А	блаже	brev,sg,m,voc
благій	А	блази	brev,pl,m,nom
благій	А	блзѣ	brev,sg,f,dat/loc brev,sg,m/n,loc
благій	А	блзѣй	plen,sg,f,dat/loc
благій	А	блзѣмь	plen,sg,m/n,loc
благій	А	благѣ	brev,pl,ins brev,pl,m,acc brev,pl,f,nom/acc
благій	А	благѣ	brev,sg,f,nom brev,sg,m/n,gen
благій	А	благѣго	plen,sg,m,acc
благій	А	благѣгѣ	plen,sg,m/n,gen
благій	А	благѣя	plen,sg,f,nom
благій	А	благѣя	plen,pl,n,nom/acc plen,du,m,nom/acc
благій	А	блѣгѣя	plen,sg,f,gen
благій	А	блѣгѣя	plen,pl,m,acc plen,pl,f,nom/acc
благій	А	блѣжайшая	comp,plen,sg,f,nom
благій	А	блѣжайшии	comp,plen,pl,m,nom comp,plen,du,n/f,nom/acc
благій	А	блѣже	brev,sg,m,voc
благій	А	блѣзи	plen,pl,m,nom
благій	А	блѣзѣ	brev,sg,f,dat/loc brev,sg,m/n,loc

Грамматический словарь (2).

благій А (4840)

хороший; добрый

благі (24), благшымъ (1), благъ (183), блаже (30), блази (13), блазѣ (32), блазѣй (3), блазѣмъ (1), благи (17), блага (22), благаго (12), благагw (84), благая (53), благаѣ (240), благими (45), благимъ (56), благихъ (301), благимъ (45), благое (176), благой (7), благому (37), благомъ (12), благою (12), благою (38), благіе (3), благіи (6), благій (83), благія (21), благія (18), блажайшая (2), блажайшее (1), блажайшій (1), блазіи (14), блазѣмъ (1), блга (68), блга (9), блгаго (52), блгагw (56), блгая (310), блгаѣ (168), блги (10), блгими (49), блгимъ (234), блгихъ (394), блгимъ (21), блго (227), блгое (188), блгой (9), блгому (48), блгомъ (9), блгою (10), блгу (16), блгую (46), блгъ (483), блги (3), блгий (279), блгия (13), блгия (20), блжайшая (1), блжайшіи (1), блже (517), блзіи (1), блзѣ (4)

благово́лити V,ipf,intr (567)

доброжелательствовать, благодетельствовать, изъяснить милость

благово́ли (66), благово́ли (2), благово́ливша (1),
благово́лившаго (9), благово́лившагw (2), благово́ливше (1),
благово́лившему (8), благово́лившу (1), благово́лившіи (1),
благово́ливъ (4), благово́ливый (15), благово́лила (5), благово́лиль
(27), благово́лимъ (4), благово́лите (1), благово́лити (1),
благово́литъ (17), благово́лихомъ (5), благово́лихъ (9),
благово́лиша (6), благово́лиши (3), благово́лю (3), благово́ляй (1),
благово́лять (5), благово́ляща (2), блѣгово́ли (127), блѣгово́ли (8),
блѣгово́лившаго (13), блѣгово́лившагw (2), блѣгово́лившая (1),
блѣгово́лившему (5), блѣгово́ливъ (3), блѣгово́ливый (42), блѣгово́лила
(10), блѣгово́лиль (72), блѣгово́лити (4), блѣгово́литъ (24), блѣгово́лихъ
(15), блѣгово́лиша (4), блѣгово́лиши (34), блѣгово́лю (3), блѣгово́лять
(1)

Перспективы

- 1) Грамматический словарь ЦС языка
(по материалам корпуса)
- 2) Грамматические таблицы
(по материалам корпуса)
- 3) Краткие толкования для непонятных слов
(СЦРЯ, САР, словарь 11—18 века)

Ссылки

1) Корпус

<http://ruscorpora.ru/search-orthlib.html>

2) Тексты

<http://feb-web.ru/febupd/slavonic/>

3) Грамматический словарь (бета)

<http://feb-web.ru/febupd/slavonic/dic/>