

Case study:

Проект по церковно-славянским текстам

О проекте

Учебный проект студентов магистратуры "Цифровые методы в гуманитарных науках" НИУ ВШЭ.

Целью проекта является создание алгоритма выравнивания богослужебных текстов сложной структуры. На первом этапе работы используются тексты Цветной Триоди на греческом и церковнославянском языках.

Участники:

- Шадрин Лев
- Дрожжина Анастасия

О проекте

Куратор проекта:

- игумен Пантелеимон (Королев),
настоятель Данилова монастыря в Переславле-Залесском.



<https://liturcorpora.ru/>

Работа над проектом началась в декабре 2021 года в рамках научно-исследовательского проектного семинара программы “Цифровые методы в гуманитарных науках”. Консультативную поддержку проекту оказывает Анастасия Александровна Бонч-Осмоловская.



Команда

игумен Пантелеимон (Королёв),
Александр Кравецкий, Анастасия Бонч-Осмоловская, студенты ВШЭ.



Новости

Известия о развитии проекта и новости о смежных событиях публикуются в телеграм-канале [@liturcorpora](#).



Код и данные

Программный код на Django и базу данных SQLite вы можете [скачать с GitHub](#).



Принять участие

Если вы желаете внести свою лепту в проект, пишите игумену Пантелеимону:
pantlmn@gmail.com.

Эксперименты по применению методов компьютерной лингвистики к корпусу богослужебных текстов Православной Церкви.



Словари

Леммы (словарь А. Е. Полякова)



Тексты

Церковнославянские тексты в гражданской графике

Библия 1751 года в гражданской графике

Бытие		
Исход		
Левит		
Числа		
Второзаконие	Глава 1	
Иисус Навин	Глава 2	
Судьи	Глава 3	¹ В начале сотвори Бог небо и землю.
Руфь	Глава 4	² Земля же бе невидима и неустроена, и тма верху бездны, и Дух Божий ношашеся верху воды.
1-я Царств	Глава 5	³ И рече Бог: да будет свет. И бысть свет.
2-я Царств	Глава 6	⁴ И виде Бог свет, яко добро, и разлучи Бог между светом и между тмою.

Леммы (словарь А. Е. Полякова)

Найти слово...

ИСКАТЬ

Найдено лемм: 32902

а

а́бїе

а́вва

а́велевъ

а́вель

а́ггелъ

Леммы (словарь А. Е. Полякова)

áгница

Грамматические характеристики: S,f,anim

Встречающиеся формы:

áгница (sg,nom)

áгницами (pl,ins)

áгнице (sg,voc)

áгницу (sg,acc)

áгниць (pl,gen/acc)

áгницы (sg,gen)

Другие возможные формы (NЗс):

áгниць* (sg,dat | sg,loc)

áгницею* (sg,ins)

áгницамъ* (pl,dat)

áгницахъ* (pl,loc)

áгницама* (du,dat/ins)

Поляков А. Е.

Грамматический словарь церковнославянского языка (по материалам корпуса).

Ресурс создан при поддержке РФФИ, проект No 17-04-12064-ОГН «Разработка модулей НКРЯ для автоматической разметки и словарной поддержки старорусских и церковнославянских текстов».

Грамматический словарь церковнославянского языка создан на базе корпуса церковнославянских текстов, который является составной частью проекта «Национальный корпус русского языка» (<http://ruscorpora.ru/>).

Церковнославянский корпус включает около 1250 текстов, которые охватывают все основные типы и жанры церковнославянской литературы (богослужебные, святоотеческие, писание, типикон, церковное право). Корпус имеет объем около 4.6 миллиона словоупотреблений и включает около 150 тыс. различных словоформ, которые группируются примерно в 30 тыс. лексем.

О проекте

В настоящий момент в репозитории представлены:

- [тексты](#) Цветной Триоди
- результаты работы с [lingtrain-aligner](#), в том числе [скрипт](#)
- [лемматизированные тексты](#) службы Вознесения
- [код](#) лемматизации с Udpipes

Специфика богослужебных текстов

Наиболее известным параллельным корпусом всех времен, безусловно, является Библия. Будучи классическим примером строгого текста, перевод которого не терпит вольных интерпретаций, порядковых перестановок компонентов или добавления художественной составляющей, Библия представляет собой эталонный пример выровненного корпуса текстов. Чуть сложнее дело обстоит художественными текстами, в которых переводчик обладает уже большей свободой. Выравнивание семантически многоуровневых текстов является предельно сложной и ресурсоемкой задачей, требующей поиска специфического подхода.

Специфика богослужебных текстов

- многослойность → тексты песнопений сопровождаются служебными комментариями по организации и проведению службы
- тексты песнопений на различных языках могут иметь отличный друг от друга порядок следований
- молитвы могут приводиться полно или сокращенно
- возможно расхождение и в темпоральной организации

Μногослойность

Ἦχος α΄.

ΤΩ ΣΑΒΒΑΤΩ ΕΣΠΕΡΑΣ, ΕΝ ΤΩ ΜΙΚΡΩ Ε-
ΣΠΕΡΙΝΩ

Στίχ. Ἀπὸ φυλακῆς πρωίας μέχρι νυκτός, ἀπὸ
φυλακῆς πρωίας, ἐλπισάτω Ἰσραὴλ ἐπὶ τὸν Κύριον.

ГЛАСЬ А

Въ сѣббѣѣ вѣчера, на малѣй вечерни,

на Гдѣи воззвахъ, поставимъ стѣхѣвъ дѣ: и по-
ѣмъ стѣхиры воскрѣны ѡсмогласника гѣ, повторѣ-
юще аѣ-й стѣхъ, гласъ аѣ. Творѣнїе прѣбнаго ѡтца
нашего іωάνна дамаскина

Стѣхъ: Ѡ стражи ѹтреннїа до нѣщи, Ѡ стра-
жи ѹтреннїа, да ѹповаѣтъ иѣль на гдѣа.

Для кого этот проект

Выровненные тексты в первую очередь могут быть необходимы церковнослужителям. Потенциальными пользователями также являются филологи и лингвисты, студенты и преподаватели высших учебных заведений Русской православной церкви.

DHonorare – проект «Богослужебный интерлинеар»

Проект «Богослужебный интерлинеар» (<https://dhonorare.ru/>) является результатом многолетней работы сотрудников Санкт-Петербургской Духовной Академии, специалистов других институций, а также волонтеров, принимавших активное участие в оцифровке церковнославянского словаря протоиерея Григория Дьяченко.

Цели проекта «Богослужебный интерлинеар» предельно созвучны основным задачам настоящего исследования, а во многом превосходят их.

DHonorare – проект «Богослужебный интерлинеар»

Как отмечают члены команды «Богослужебного интерлинеара», существует множество онлайн-ресурсов, посвящённых текстам Библии на разных языках. Однако, «до сегодняшнего дня таких проектов, посвящённых богослужебным текстам, не создано».

DHonorare – проект «Богослужебный интерлинеар»

The Greek text	Трѣбникъ. В 2 частяхъ. — М.: Синодальная типографія, 1906	Требник: молитвослов для совершения священнодействий православной церкви. На русском языке / В переводе свящ. Василия Адаменко. — М., 2010
¹ Εὐχὴ εἰς γυναῖκα λεχὼ , τῇ πρώτῃ ἡμέρᾳ τῆς γεννήσεως τοῦ παιδίου αὐτῆς	¹ МѢТВЫ ВЪ ПЕРВЫЙ ДЕНЬ , ПО ВНЕГДА РОДИТИ ЖЕНѢ ѠТРОЧА .	¹ Молитвы в первый день после рождения младенца
		² Молитвы : 1 .
³ Τοῦ Κυρίου δεηθῶμεν .	³ ГДѢ помѡлимсѧ .	³ Господу помолимся .

DHonorare – проект «Богослужебный интерлинеар»

<https://dhonorare.ru/dict/dyachenko/>

Frequencies wordforms

Change subcorpus:

Hide stop words

Show stop words

Only stop words

Greek 38144			Church Slavonic 687367			Russian 39601			English 2746		
№	word	count / ipm	№	word	count / ipm	№	word	count / ipm	№	word	count / ipm
1.	Θεὸς	223 / 5846	1.	рече	3856 / 5610	1.	Господи	197 / 4975	1.	king	26 / 9468
2.	Κύριε	193 / 5060	2.	гдѣ	3803 / 5533	2.	священник	127 / 3207	2.	men	12 / 4370
3.	Κυρίου	176 / 4614	3.	земли	1936 / 2817	3.	Бог	116 / 2929	3.	captives	11 / 4006
4.	δεηθῶμεν	135 / 3539	4.	бгѣ	1853 / 2696	4.	ныне	112 / 2828	4.	God	11 / 4006
5.	Τερεὺς	123 / 3225	5.	сынове	1419 / 2064	5.	имя	107 / 2702	5.	Jews	11 / 4006

DHonorare – проект «Богослужебный интерлинеар»

Благодаря словарю в тексте реализована пословная подсветка:

[https://dhonorare.ru/texts/trebnik/molitvy-v-pervyy-den-posle-rozhdeniya-mladent
sa?ch56=on&ch57=on&ch60=on&view=rows](https://dhonorare.ru/texts/trebnik/molitvy-v-pervyy-den-posle-rozhdeniya-mladent
sa?ch56=on&ch57=on&ch60=on&view=rows)

Цветная Триодь

Цветная Триодь включает в себя богослужебные последования пасхального цикла, то есть круг подвижных праздников от самой Пасхи до Недели Всех святых, даты которых меняются в зависимости от дня празднования Пасхи.

Название книги связано с тем, что каноны, содержащиеся в ней, по большей части состоят из трех песен. Такой канон и называется трипеснец или *τριώδιον* («триодион») по-гречески.

Предварительная обработка текста

Греческий текст Цветной Триоди выгужен с сайта Греческой православной архиепископии Америки. Церковнославянский текст Цветной Триоди выгужен с сайта проекта «Азбука веры».

Предварительная обработка текстов на этапе экспериментов осуществлялась при помощи текстового редактора Sublime Text, а также отдельными скриптами с базовыми NLP-элементами (нормализация регистра, токенизация, очистка пунктуации).

Работа с текстами: οχιστκα

123 Μετὰ τὴν α' Στιχολογίαν, Κάθισμα τῆς Πεντηκοστῆς
124 Ἦχος δ'
125 Κατεπλάγη Ἰωσήφ **ΤΟ ΑΚΟΥΤΕ**
126 Τὴν μεθέορτον πιστοί, καὶ τελευταίαν ἐορτὴν, ἐορτάσωμεν
Παρακλήτου εὐθύς, κατέβη ἐπὶ γῆς, ὥσπερ ἐν εἶδει γλωσσῶ
τόν κόσμον, ἐφώτισε. (Δίς)
127 Μετὰ τὴν β' Στιχολογίαν, Κάθισμα τῆς Πεντηκοστῆς
128 Ἦχος δ'
129 Κατεπλάγη Ἰωσήφ **ΤΟ ΑΚΟΥΤΕ**
130 Ἡ τοῦ Πνεύματος πηγὴ, ἐπιδημοῦσα τοῖς ἐν γῇ, εἰς πυρφόρ
δροσῶδες τὸ πῦρ, φωτίζουσα αὐτούς, καὶ ὑετίζουσα φλόξ,
κόσμον ἐφώτισε. (Δίς)
131 Κανὼν α', ᾠδὴ α', τῆς Πεντηκοστῆς
132 Ἦχος βαρὺς
133 Ὁ Εἰρμὸς **ΤΟ ΑΚΟΥΤΕ**
134 "Πόντῳ ἐκάλυψε Φαραῶ σὺν ἄρμασιν, ὁ συντρίβων πολέμους
135 Ἔργῳ, ὡς πάλαι τοῖς Μαθηταῖς ἐπηγγείλω, τὸ Παράκλητον Π
136 Νόμῳ τὸ πάλαι προκηρυχθὲν καὶ Προφήταις, ἐπληρώθη, τοῦ
137 Κανὼν β', ᾠδὴ α', τῆς Πεντηκοστῆς
138 Ἦχος δ'

Μετὰ τὴν α' Στιχολογίαν, Κάθισμα

Ἦχος δ'

Κατεπλάγη Ἰωσήφ **ΤΟ ΑΚΟΥΤΕ**

Τὴν μεθέορτον πιστοί, καὶ τελευταίαν ἐορτὴν, ἐορτάσωμεν φαιδρῶς· αὕτη
ὥσπερ ἐν εἶδει γλωσσῶν, καὶ Μαθητὰς ἐφώτισε, καὶ τούτους οὐρανομύστα

Μετὰ τὴν β' Στιχολογίαν, Κάθισμα

Ἦχος δ'

Κατεπλάγη Ἰωσήφ **ΤΟ ΑΚΟΥΤΕ**

Ἡ τοῦ Πνεύματος πηγὴ, ἐπιδημοῦσα τοῖς ἐν γῇ, εἰς πυρφόρους ποταμούς,
καὶ ὑετίζουσα φλόξ, δι' ὧν ἡμεῖς ἐλάβομεν τὴν χάριν, διὰ πυρός τε καὶ ὕδα

Μετὰ τὸν Πολυέλεον, Κάθισμα

Ἦχος πλ. δ'

Τὸ προσταχθὲν μυστικῶς **ΤΟ ΑΚΟΥΤΕ**

Μετὰ τὴν Ἑγερσιν Χριστέ, τὴν ἐκ τοῦ τάφου, καὶ τὴν πρὸς ὕψους οὐρανοῦ,
κιθάρα μουσουργικὴ, πᾶσιν ἐτραυολόγησαν, τῷ θεῷ πλήκτρῳ μυστικῶς, Σ

Предварительная обработка текста

4588 Ὁ οἶκος
4589 Αὐτὸς μόνος ὑπάρχεις ἀθάνατος, ὁ ποι
ὁ πλάσας με καὶ εἰπὼν μοι· Ὅτι γῆ εἶ
4590 Σ·Υ·Ν·Α·Ξ·Α·Ρ·Ι·Ο·Ν, Πεντηκοσταρίου
4591 Τῇ αὐτῇ ἡμέρᾳ, μνείαν πάντων τῶν ἀπ'
4592 Στίχοι

ΣΥΝΑΞΑΡΙΟΝ, Πεντηκοσταρίου

Τῇ αὐτῇ ἡμέρᾳ, μνείαν πάντων τῶν ἀπ' αἰῶνος

Στίχοι

Ἀμνημόνησον παισμάτων νεκροῖς, Λόγε,
Τὰ χρηστὰ νεκρὰ σπλάγχνα σου μὴ δεικνύων.

Предварительная обработка текста

- Особенностью обрабатываемых текстов являются диакритические знаки, встречающиеся как в греческом, так и в церковнославянском корпусе
- Лемматизация. В контексте задачи параллельного выравнивания текстов, лемматизация может быть необходима для последующей обработки текста при помощи элайнеров или кода, использующего словарь переводных эквивалентов. В других ситуациях, например, при использовании векторных моделей подобных Lingtrain Aligner, лемматизация оказывается не столь значимой.

Немного о проблеме лемматизации

В условиях отсутствия готового модуля, позволяющего провести лемматизацию церковнославянского текста, а также учитывая описанную специфику Цветной Триоди, было решено попробовать лемматизировать Триодь как при опоре на модели современного русского языка (Spasy, PyMorphy2, UDPipe), так и используя древние языки (UDPipe).

Наилучшие результаты показал UDPipe – обучаемый пайплайн для сегментации предложений, токенизации, морфологического парсинга и лемматизации, основанный на синтаксических компонентах грамматики зависимостей.

Алгоритм работы UDPipe можно описать так – сначала текст разделяется на предложения и далее на слова (этот этап выполняется посредством однослойной двухсторонней нейронной сети GRU), после чего происходит тегирование. Генерируются гипотезы о части речи и морфологических параметрах словоформы. Лучший вариант отбирается при помощи перцептрона. По результатам POS-тегирования происходит лемматизация – подбор оптимальной начальной формы.

Для русского языка – современного и древнего – представлено 6 моделей. В таблице 1 изложены описания данных моделей

	old_russian-rnc-ud-2.6-200830		old_russian-torot-ud-2.6-200830		russian-syntagrus-ud-2.6-200830		russian-gsd-ud-2.6-200830		russian-taiga-ud-2.6-200830	
1	В	въ	В	въ	в	в	В	в	в	в
2	среду	сръда	среду	сръда	среду	среда	среду	среда	среду	среда
3	вечера	вечерь	вечера.	вечера.	вечера	вечер	вечера	вечер	вечера	вечер
4	Празднуем	празднити	Празднуем	празднити	Празднуем	праздновать	Празднуем	праздновать	Празднуем	праздновать
5	Вознесение	вознесение	Вознесение	вознесение	Вознесение	Вознесение	Вознесение	вознесение	Вознесение	вознесение
6	Господа	господь	Господа	господь	Господа	Господь	Господа	господин	Господа	господин
7	Бога	Богъ	Бога	богъ	Бога	Бог	Бога	бог	Бога	Бог
8	и	и	и	и	и	и	и	и	и	и
9	Спаса	спасъ	Спаса	спасъ	Спаса	Спас	Спаса	спас	Спаса	спас
10	нашего	нашъ	нашего	нашъ	нашего	наш	нашего	наш	нашего	наш
11	Иисуса	Иисусъ	Иисуса	иисусъ	Иисуса	Иисус	Иисуса	Иисус	Иисуса	Иисус
12	Христа	Христъ	Христа.	христа.	Христа	Христос	Христа	Христос	Христа	Христос
13	По	по	По	по	По	по	По	по	По	по
14	начале	начало	начале	начало	начале	начало	начале	начало	начале	начало
15	глаголет	глагоити	глаголет	глагоити	глаголет	глаголет	глаголет	глаголет	глаголет	глаголет
16	чтец	чтець	чтец	чтець	чтец	чтец	чтец	чтец	чтец	чтец
17	На	на	На	на	На	на	На	на	На	на
18	малеи	малый	малеи	малый	малеи	малея	малеи	малея	малеи	малея
19	вечерни	вечерьнъ	вечерни.	вечерьня	вечерни	вечерни	вечерни.	вечерни.	вечерни	вечерьня
20	Господи	господь	Господи	господь	Господи	Господь	Господи	Господь	Господи	Господь
21	воззвах	воззвати	воззвах:	воззвати	воззвах	воззва	воззвах	воззва	воззвах	воззва

Код лемматизации расположен в открытом доступе на Github: https://github.com/Drozhzhinastya/GSPC/blob/main/texts/lemmatization/udpipe_lemmas_vozneseniye.ipynb

CLTK:

ὁ κύριος ἀναλαμβάνω εἰς οὐρανός ἵνα πέμπω ὁ παράκλητος ὁ κόσμῳ, ὁ οὐρανός ἐτοιμάζω ὁ θρόνος αὐτός νεφέλῃ ὁ ἐπίβασις αὐτός ἄγγελος θαυμάζω ἄνθρωπος ὁράω ὑπεράνω αὐτός ὁ πατήρ ἐκδέχομαι ὅς ἐν κόλπος ἔχω συναῖδιον ὁ πνεῦμα ὁ ἅγιος κελεύω πᾶς ὁ ἄγγελος αὐτός αἰρώ πύλῃ ὁ ἄρχων ἡμεῖς πᾶς ὁ ἔθνος κροτέω χεῖρ ὅτι ἀναβαίνω χριστός ὅπου εἰμί ὁ πρότερος

UDPipe - Perseus

ὁ Κύριος ἀαλαμβάνω εἰς οὐρανός, ἵνα πέμπω ὁ Παράκλητος ὁ κόσμος , ὁ οὐρανός ἐτοιμάζω ὁ θρόνος αὐτός , νεφέλα ὁ ἐπίβασις αὐτός , Ἄγγελοι θαυμάζω , ἄνθρωπος ὁράω ὑπεράνω αὐτός , ὁ Πατήρ ἐκδέχομαι , ὅς ἐν κόλπος ἔχω συναῖδιος ὁ Πνεῦμα ὁ ἅγιος κελεύω πᾶς ὁ Ἀγγέλοί αὐτός · Ἄρατε πύλῃ ὁ ἄρχων ἐγώ , πᾶς ὁ ἔθνος κροτάω χεῖρ ὅτι ἀνέβη Χριστός , ὅπου εἰμί ὁ πρότερος

LF-aligner

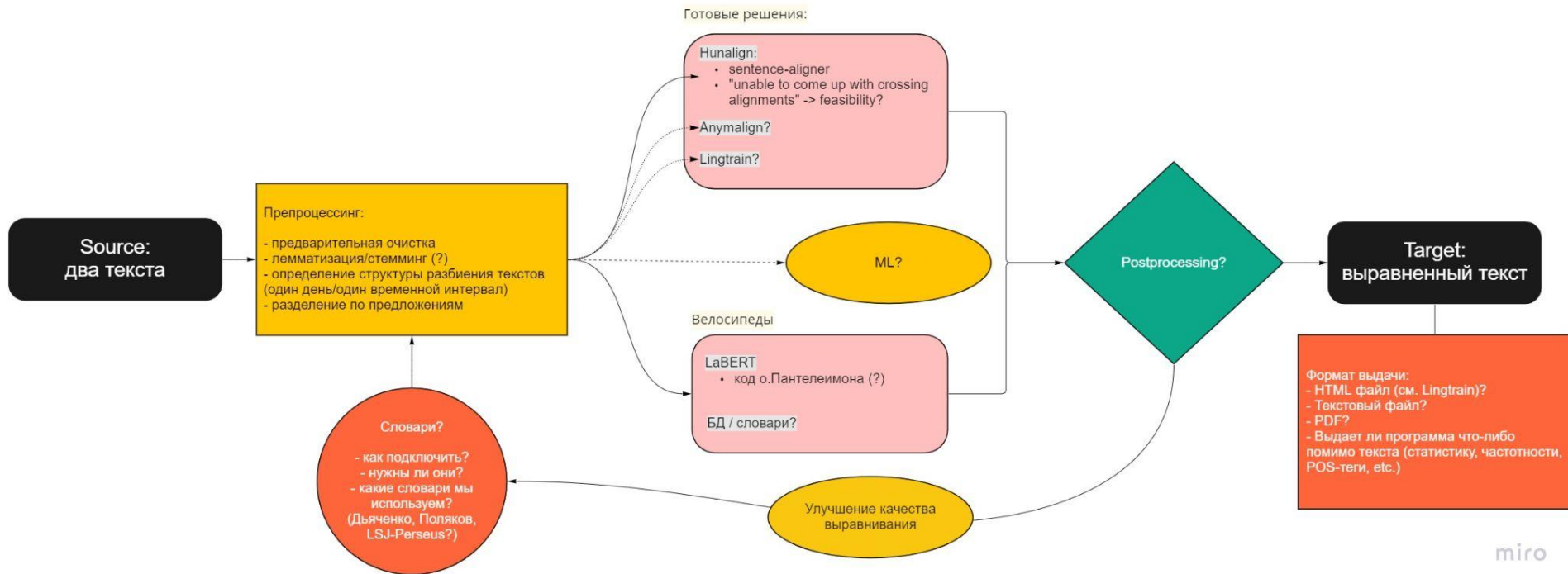
Следующая пара инструментов – *Hunalign* и графический интерфейс *LF-aligner* – требуют более сложной предварительной работы с текстом. В основе алгоритма *Hunalign* – выравнивание лемматизированных текстов по предложениям. При условии наличия словаря переводных эквивалентов, *Hunalign* использует его, комбинируя словарные данные с информацией о длине предложения. При отсутствии словаря он сначала обращается к длине предложения, а затем на основе этого выравнивания автоматически строит словарь. После этого *Hunalign* выравнивает текст повторно, используя словарные данные.

Lingtrain-Aligner

Среди основных аспектов, повлиявших на низкий результат работы элайнера, стоит отметить следующие:

- Несоответствие количества предложений церковнославянского фрагмента древнегреческому: объем последнего больше;
- Текст молитв в церковнославянском может быть представлен только инципитом, в греческом же всегда приводятся полные формы, что усложняет выравнивание по предложениям;
- Различия в богослужебных указаниях

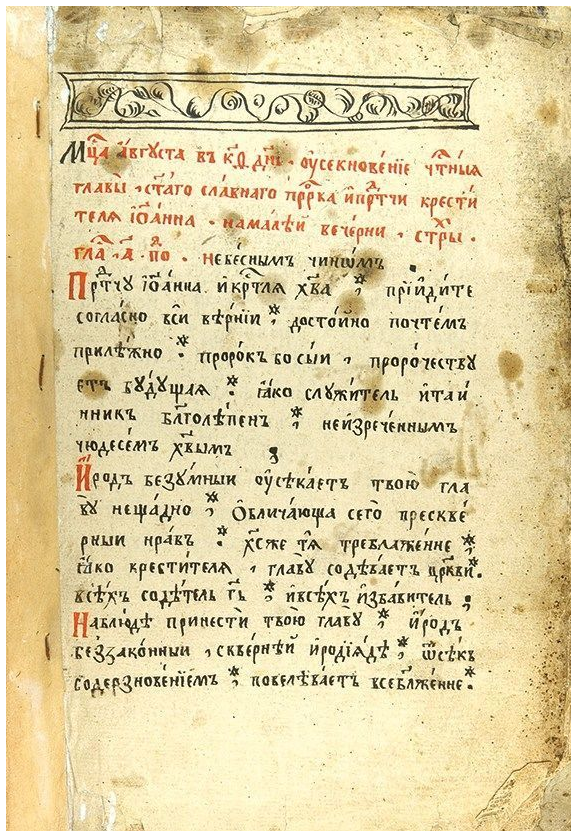
Результаты на текущий момент



Направления дальнейшей работы

- все же необходимо учитывать сложную структуру текста
- Lingtrain демонстрирует результаты лучшие, чем LF-aligner. Однако, это готовый пайплайн, механизмы которого не так легко регулировать, как при написании собственного решения
- возможно, нужна модель с дообучением (на той же Библии или Октоихе)

Киноварь и двусложная структура богослужебных текстов



КАНОН, ГЛАС 1.

И начинает предстоятель канон, творение господина Иоанна Дамаскина. Глас 1. Ирмос: Воскресения день: Ирмосы на 4: и тропари на 12, с припевы: Христос воскресе из мертвых. И паки последи кийждо лик ирмос. Последи же на сходе Катавасия, ирмос тойже: Воскресения день: И по нем Христос воскресе: весь трижды. Творит же начало канона на куюждо песнь всегда предстоятель, десной или левой стране случившейся начати. И кадит в начале канона святые иконы и облики, и братию по чину. И по коейждо песни бывает ектения малая вне олтаря, якоже рехом, во святой сей день. Возглас внутрь олтаря от иерея. По 1-й песни поет десная страна. По 3-й же поет левая. Сие поем и по прочих песнях.

Киноварь в проекте:

- разметка оригинальных текстов с azbyka.ru
 - отдельные классы в HTML-тегах
- также присутствует и в греч. текстах
- разные уровни текстов в структуре БД
 - киноварь: служебные указания
 - чернила: тексты ГИМНОВ

Νεφέλην σε φωτὸς αἰδίου, Παρθένε, ὁ Προφήτης ὠνόμασεν· ἐν σοὶ γὰρ ὡς ὑετὸς ἐπὶ πόκον, καταβάς ὁ Λόγος τοῦ Πατρός, καὶ ἐκ σοῦ ἀνατείλας, τὸν κόσμον ἐφώτισε, τὴν πλάνην κατήργησε, Χριστὸς ὁ Θεὸς ἡμῶν. Αὐτὸν ἱκετεύουσα ἐκτενῶς, Παναγία δεόμεθα, μὴ παύσῃ ὑπὲρ ἡμῶν, τῶν ἀληθῆ Θεοτόκον ὁμολογούντων σε.

Ἀπολυτίκιον. Ἦχος α΄.

ΤΩ ΣΑΒΒΑΤΩ ΕΣΠΕΡΑΣ, ΕΝ ΤΩ ΜΕΓΑΛΩ ΕΣΠΕΡΙΝΩ

Στιχηρὰ Ἀναστάσιμα τῆς Ὀκτωήχου. Ἦχος α΄.

Τὰς ἐσπερινὰς ἡμῶν εὐχάς, πρόσδεξαι Ἁγίε Κύριε, καὶ παράσχου ἡμῖν ἄφεσιν ἁμαρτιῶν, ὅτι μόνος εἶ ὁ δείξας ἐν κόσμῳ τὴν ἀνάστασιν.

Облакъ та свѣта присносѣщнаго дѣо, прѣрокъ именова: изъ тебѣ бо ꙗ́кѡ дождь на рѣно сшѣдъ слѣво ѡчее, и изъ тебѣ возсі́авый, міръ просвѣти, прѣлестъ ѱпраздни хрѣтѣсь бгѣ нашѣ. того мо́лаши прилѣжнѡ престѣа, мо́лимсѧ, не престѣй ѡ́ насѣ, иже ѣстиннѣю бгѣмъ исповѣдающихъ та.

Таже: Нынѣ ѡпѣщаеши: Трисѣ́ое. по О́че нашѣ: тропѣрь воскрѣный. Сла́ва, и нынѣ, бгѣ-ро́дичень ѣгѡ. Ектенѣа ма́лаа, и ѡпѣсть.

Въ сѣббѡтѣ на велицѣй вечерни,

по предначинáтельномъ ѱалмѣ, ѡ́бычное стѣхослѡ́вѣ ѱалтіра. На Гдѣ́и воззвахѣ, поста́вимъ стѣхѡвъ ї: и поѣмъ стѣхиры воскрѣны ѡ́смогласника г, и а́нато́ліевы д, и мине́и г, или ѕ. Аще прѣ́зднѣемый стѣйи: Сла́ва, мине́и: И нынѣ, бгѣро́дичень а́-и, гласа.

Стѣхиры воскрѣны ѡ́смогласника, гласъ а́

Стѣхъ: Изведи изъ темни́цы дѣшѣ́ мою, испо́вѣдатисѧ ѣмени твоѣмъ.

Вече́рнѣа на́ша млтвы прѣими стѣйи гдѣ́и, и по́даждь на́мъ ѡ́ставленіе грѣхѡ́въ, ꙗ́кѡ є́динъ є́си ꙗ́влѣй въ міръ воскрѣніе.

→ Работать со смысловыми компонентами текста отдельно

	red_text_b	black_text	red_text_e	color
0	## Служба в Светлое Христово Воскресение			temp
1	## Во Святую и Великую Неделю Пасхи			temp
2	<Об часе утреннем параклисиарх взем благословение у настоятеля, исходит и ударяет в великое, и клеплет довольно. И вшед во храм, вжигает свещи вся, и кандила: устрояет же сосуды два со углием горящим, и влагает в них фимиама много благовоннаго, и поставляет один сосуд среди церкви, другой же во святом олтарѣ, яко да исполнится церковь вся благовония. Также настоятель вшед во святой олтарь со иереи и диаконы, облачаты в весь светлейший сан. И раздаёт свещи братии, и вземлет Честный Крест: кадило же вземлет диакон. Иерей же святое Еванглие, ин иерей образ Воскресения Христова: и ставятся лицом на запад. И затворят врата церковная, яже к западу. Исходит настоятель со иереи в притвор северными дверьми, предыдущим пред ним диаконом со двумя свечами, и оба лица поюще стихиру, глас 6:>			red
3	<Также ударяют во вся кампаны и тяжкая, и клеплют довольно. И вшедше в притвор, станут со Еванглием, и со образом, лицом на запад, якоже указася прежде. Также настоятель вземлет у диакона кадило в правую руку, Крест же в левую, и кадит образы, и клиросы, и братию по обычаю. Диакону же предносящу пред ним свещу горящую. Братия же вси стоят держаще свещи свои, со вниманием молящиеся в себе, и благодаряще нас ради потрудившаго и воскресшаго Христа Бога нашего. По окончании же каждения, приходит настоятель пред великия врата церкви, и покадит диакона предстоящего ему со свечью. Тогда взем кадило диакон от рук настоятеля, и покадит самого настоятеля. И паки восприим настоятель кадило, став пред церковными дверьми, зря на восток, и назнаменает великия врата церкви, [затворённая суща], кадилом крестообразно, трижды, держа в левой руке Честный Крест, и светильником стоящим со обою страну.>			red
4		Воскресение Твое Христѣ Спасе, Ангели поют на небесех, и нас на земли сподоби чистым сердцем Тебѣ славити.		black
5	<И велегласно возгласит.>			red
6	<И нам отвечающим.>			red
7		Слава Святей, и Единосущней, и Животворящей, и Неразделимой Троице, всегда, ныне и присно, и во веки веков.		black
8	<Начинает настоятель, с прочими служители, велегласно настоящий тропарь, во глас 5:>			red
9		Аминь.		black
10	<И мы тойже поём подобно, со сладкопением. Поётся же сей тропарь трижды от настоятеля, и трижды от нас.>			red

Размеченные тексты

https://docs.google.com/spreadsheets/d/166jSDbbLF_5jlAmuJGrhm5BLTBdYADH_UH-75GzhmRil/edit?usp=sharing

https://docs.google.com/spreadsheets/d/1T6yhl53d-F8F3fQ3tsV6itnvPgxsWVPoS2NslwT_9tU/edit?usp=sharing

Чуть подробнее о механизмах внимания и нейросетях

<https://habr.com/ru/post/486158/> → Визуализируя нейронный машинный перевод (seq2seq модели с механизмом внимания)

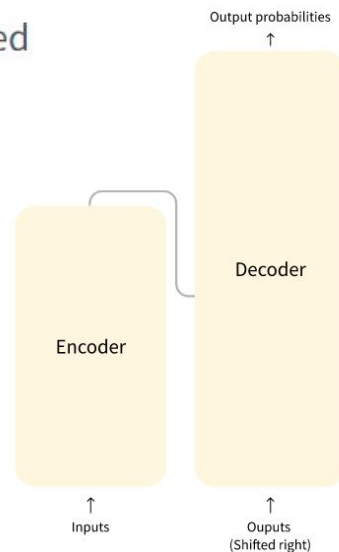
<https://habr.com/ru/post/486358/> → Transformer в картинках

The model is primarily composed of two blocks:

- **Encoder (left):** The encoder receives an input and builds a representation of it (its features). This means that the model is optimized to acquire understanding from the input.
- **Decoder (right):** The decoder uses the encoder's representation (features) along with other inputs to generate a target sequence. This means that the model is optimized for generating outputs.

- [CTRL](#)
- [GPT](#)
- [GPT-2](#)
- [Transformer XL](#)

- [ALBERT](#)
- [BERT](#)
- [DistilBERT](#)
- [ELECTRA](#)
- [RoBERTa](#)



Model	Examples	Tasks
Encoder	ALBERT, BERT, DistilBERT, ELECTRA, RoBERTa	Sentence classification, named entity recognition, extractive question answering
Decoder	CTRL, GPT, GPT-2, Transformer XL	Text generation
Encoder-decoder	BART, T5, Marian, mBART	Summarization, translation, generative question answering

The model can be used in this state, but it will output gibberish; it needs to be trained first.

Note that this model is primarily aimed at being fine-tuned on tasks that use the whole sentence (potentially masked) to make decisions, such as sequence classification, token classification or question answering. For tasks such as text generation you should look at model like GPT2.

<https://huggingface.co/bert-base-cased> (pretrained model)

Код для доработки

https://colab.research.google.com/drive/1Bgtst3Ys1qoQ4vMfroFRwapIfDwVZO6P#scrollTo=_KR8tjE4_sNJ

Не забываем про дневник :)

https://docs.google.com/forms/d/1PUilZN_ScyAyzGMcCFZ6SdbO5_r6OhNzM-IVxZezOMo/edit