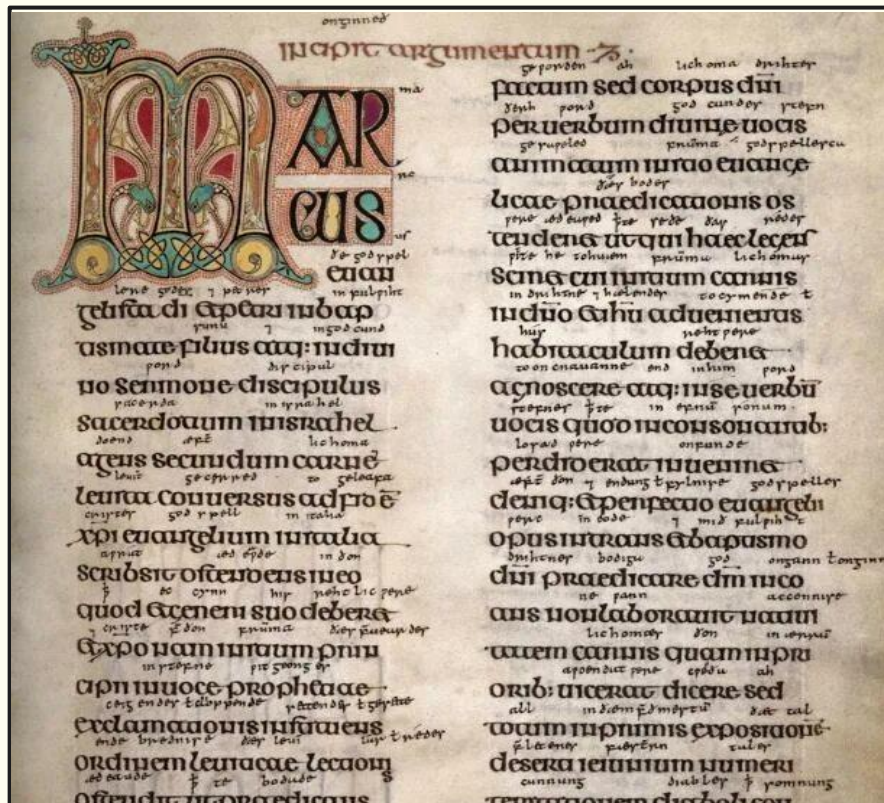


Anna
Rogova

Exploratory Corpus
Analysis of Old
English Gospels
(Lindisfarne &
Rushworth)



Lindisfarne Gospels & Rushworth Gospels

Два перевода Евангелия на староанглийский язык. Оба текста представляют собой построчные глоссы (пометки) к латинскому оригиналу.

Глобальная цель исследования — выяснить степень сходства между двумя переводами и предположить, пользовались ли писцы: общим (ныне утраченным) источником, или один опирался на работу другого.

Этапы работы

3

1.

Создание
параллельного
корпуса

2.

Токенизация и
удаление стоп-
слов

3.

Частотный анализ

4.

Векторизация и
метрики сходства

5.

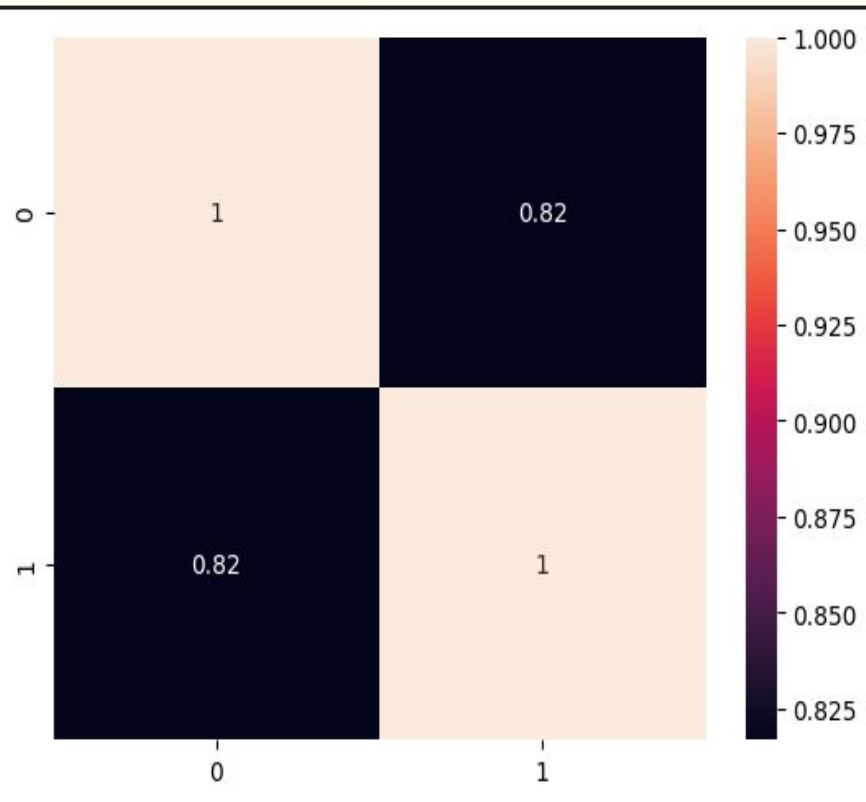
Поиск словоформ
и построение
интерактивного
поиска

Формирование корпуса

- Загрузка параллельных текстов Линдисфарнского и Рашвортского переводов
- Очистка строк от меток стихов ([123 (1.2)])
- Сопоставление строк по общему латинскому фрагменту
- Построение таблицы с колонками: LI, Common Latin, RU
- Выделение отдельных колонок в .txt файлы
- Подготовка файлов для анализа: li_column.txt и ru_column.txt

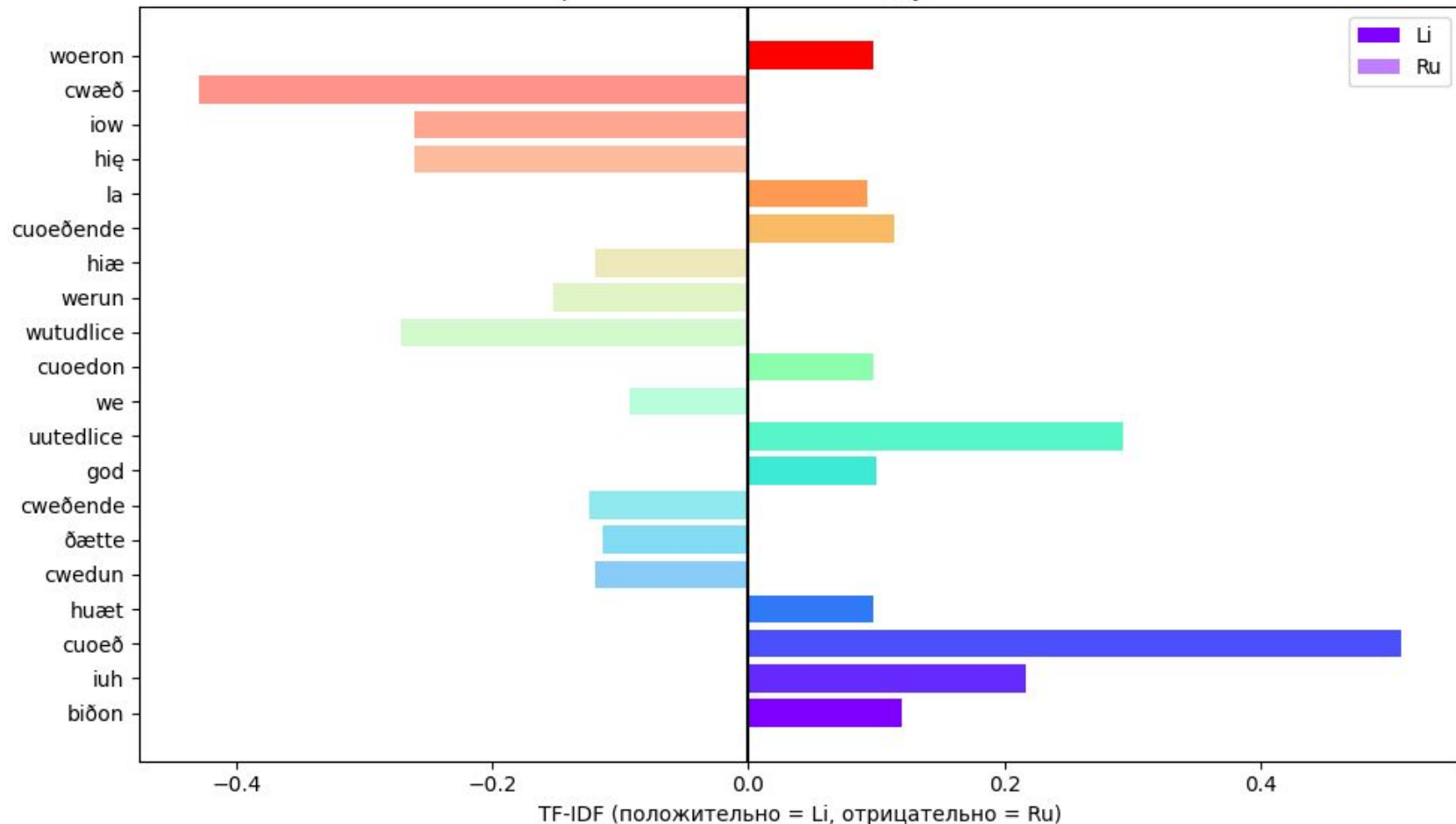
Векторизация и метрики сходства

6



- Создание **мешка слов (Bag of Words)** для каждого текста
- Расчёт **косинусного сходства** между текстами
Применение **TF-IDF** для оценки значимости слов
- Сравнение TF-IDF-лемм между переводами

сравнение TF-IDF лемм в двух текстах



Поиск словоформ и контекстов

8

- Разработка функции поиска словоформ по регулярному выражению
- Извлечение предложений с искомым словом
- Выравнивание найденных контекстов в параллельной таблице
- Сравнение употребления конкретных слов в разных переводах
- Ввод формы через `input()` для живого поиска по корпусу

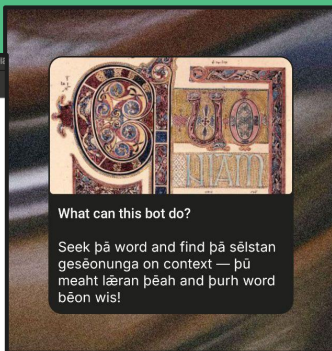
	Context Lindisfarne	Form Lindisfarne	Form Rushworth	Context Rushworth
0	ðy eade him of scip recone togægnes arn him of byrgennum † of bendum mon in gast unclæne.	gast	gaste	ðy eode to him of scipe togægnes arn him of byrgennum † of bendum monn in gaste unclænnum.
1	spirite inmundē ab homine cwæð forðon to him gaa ðu gaast unclæne from ðæm menn.	gaast	gast	i spiritus inmundē ab homine cwæð forðon to him gaa gast unclæne from ðæm menn.
2	epraecabantur eum spiritus dicentes mitte nos in porcos ut in eos introeamus & bedon hine ða gaastas cwæðendo send usic in ðæm bergum þætte in him † in ðæm we ingeonga † gæ.	gaastas	gastas	ipraecabantur eum spiritus dicentes mitte nos in porcos ut in eos introeamus & bedun hine gastas cwæðende send usih in ða bergas þætte we in hia ingonge † ingæ.
3	esus et exeuntes spiritus inmundi introierunt in porcos et magno impetu grex praecipitatus est in mare ad duo milia et suffocati sunt in mare & gelefde him † ðæm recone se hælend & miððy færende weron ða gastas unclæno inæodon in ðæm bergum & mið micle hræs † ongeong worn todrifen wæs in sæ to twæm ðusendum & underdrencdo wæron in sæ.	gastas	gastas	hesus et exeuntes spiritus inmundi introierunt in porcos et magno impetu grex praecipitatus est in mare ad duo milia et suffocati sunt in mari & gilefde him sona ðe hælend & mið ðy færende werun gastas unclæne inæodon in ða bergas & micle ræse † ongonge worn todrifen wæs on sæ to twæm ðusendum & adrencte weren on sæ.
4	illis potestatem spiritum inmundorum & efnegeceigde tuoefto & ongann hia sende tuoge & gesalde him † ðæm mæht gasta unclænra.	gasta	gasta	eis potestatem spirituum inmundorum & efne gicegde twelfe & ongan hia senda twoege & gisalde him mæhte gasta unclænra.

Создание телеграм бота

new Zoom Add Category Print Table Insert Table Chart Text Shape Media

+ Results

Gospel	Verse	Context L1	Form L1	Form Ru	Context Ru
Mk	1:2	[300 (1:2)] <i>icut scriptum est in esaiā propheta ecce mitto angelum meum ante faciem tuam qui praeparabit uiam tuam suae auertit is in esaiā donec uigilo henuc engel min beforan cruxione dñe seße «fongaraw» wage dñe.</i>		henu	[300 (1:2)] <i>icut scriptum est in esaiā propheta ecce mitto angelum meum ante faciem tuam qui praeparabit uiam tuam suae auertit is in esaiā donec uigilo henuc ic sende engel min beforan cruxione dñe seße fongaraw wæg dñe.</i>
Mk	12:44	[50300 (12:44)] <i>omnes enim ex eo quod abundabat illis miserrant haec uere de paenuria sua emissa quae habuit misit totum suum uile forbon of ðon þarte gemonglæde! (growe him sendon ðios outelice of hendu! unspeod hie alle ðæde hæfde sende al gebregneisse hie.</i>	hendu	hendum	[50300 (12:44)] <i>omnis enim ex eo quod habundabat illis miserrant haec uere de paenuria sua emissa quae habuit misit totum suum uile forbon of him þarte gemonglæde him sendon ðas wutlice of hendum! unspeodum hie alle ðæde hio hæfde sende ale gebregneisse hie.</i>
Mk	8:2	[28800 (8:2)] <i>miserror super turbam quia ecce iam triduo sustinent me nec habent quod manducet ic milia offer ðreat forbon heno gee i seßlice ðio doger ge i abides mec ne habbas ða þarte ða gætte.</i>	heno		[28700 (8:2)] <i>miserror super turbam quia ecce iam triduo sustinent me nec habent quod manducet ic milia offer ðreat forbon henuc ge ðio doger gebides mec ne habbas hē gætte hē etc.</i>
Lk	10:16	[48700 (10:16)] <i>Qui uos audit me spernit qui autem me spernit spernit eum qui me misit seße iuh geheneð mec geheneð & seße iuh teled i geheneð mec hene seße uætelice mec henes geheneð ðone seße mec sende.</i>	henes	heneð	[48400 (10:16)] <i>Qui uos audit me spernit et qui uos spernit me spernit et qui autem me spernit spernit eum qui me misit seße iowah gheres mec gheres & seße iowah teled i heneð mec teled i heneð & seße wutlice mec heneð heneð i teled hie seße mec sende.</i>
		[49000 (10:18)] <i>ecce desit uobis</i>			[48700 (10:18)] <i>ecce desit uobis</i>



- Пользователь вводит слово или его форму (например: **wif**, **hus**)
- Бот ищет все совпадающие словоформы в корпусе
- Результаты поиска выводятся в таблицу, содержащую:
 - название Евангелия и стих
 - найденную словоформу
 - контекст (предложение)
- Бот отправляет пользователю Excel-файл (.xlsx) с таблицей совпадений



Q&A

?