Федеральное государственное автономное образовательное учреждение

высшего образования

«Национальный исследовательский университет

«Высшая школа экономики»

###### Факультет гуманитарных наук

Левченко Мария Владимировна

**Славянские словари XVII века в межкультурном контексте**

Выпускная квалификационная работа – магистерская диссертация

по направлению подготовки «45.04.03 Фундаментальная и прикладная лингвистика»

###### Образовательная программа

###### «Компьютерные методы в гуманитарных науках»

|  |
| --- |
| Научный руководитель |
| канд. филол. наук |
| О.Н. Ляшевская |

Рецензент

канд. истор. наук

Н.Г. Поврозник

Москва, 2022

Оглавление

[Введение 3](#_Toc105378318)

[Епифаний Славинецкий и его «Лексикон греко-славяно-латинский» 5](#_Toc105378319)

[Электронное представление словаря Епифания в формате TEI 7](#_Toc105378320)

[RDF: моделирование для всемирной сети 10](#_Toc105378321)

[OntoLex Lexicon Model for Ontologies (lemon) 11](#_Toc105378322)

[Мировой опыт конвертирования словарей в машиночитаемый формат 13](#_Toc105378323)

[Конвертация Словаря Епифания в RDF 18](#_Toc105378324)

[Методология 18](#_Toc105378325)

[Правила URI 19](#_Toc105378326)

[Схема RDF 21](#_Toc105378327)

[Валидация через SPARQL 28](#_Toc105378328)

[Связанная работа: межкультурный анализ 30](#_Toc105378329)

[Заключение 36](#_Toc105378330)

# Введение

С развитием компьютерных методов обработки данных многие разделы гуманитарных наук получили новые инструменты анализа, что послужило толчком к созданию новых направлений и развитию существующих. Не исключением стала и лексикография, раздел лингвистики, направленный на изучение и составление словарей. Помимо работы по простому сохранению культурного наследия (digital conservation), ведется работа по конвертации отсканированного физического словаря в машиночитаемый формат (retro-digitization). Регламентированная структура машинного языка позволяет не только облегчить чтение и навигацию по словарям, но и открывает новые, не доступные при близком чтении возможности для анализа не только отдельного словаря, но связи нескольких словарей и текстов. Таким образом создается «Semantic Web», связывающая тексты и понятия в рамках локального проекта или в Интернете.

Наша работа продолжает исследование Левченко 2021, направленное на разработку подходов к приведению в машиночитаемый формат Славяно-греко-латинского словаря, составленного монахом Чудовского монастыря и переводчиком Епифанием Славинецким во второй половине XVII века.

Работа проводится на материалах, собранных германо-российским проектом «Греко-славяно-латинский словарь Епифания Славинецкого в восточнославянском и европейском языковом пространстве XVI – начала XVIII в.» (руководители проекта - И. А. Подтергера[[1]](#footnote-1), М. О. Новак[[2]](#footnote-2)). Этот проект направлен на изучение культурных и языковых связей в восточной Европе раннего Нового времени на материале ранее не изученных и не изданных текстах так называемого Чудовского круга, оказавшего значительное влияние на развитие русского языка во второй половине XVII века, и особенное внимание уделяется трудам Епифания Славинецкого, одного из ведущих московских книжников. Проект также ставит целью изучить взаимное влияние греческой, латинской, славянской и польской языковых традиций на уровне лексики, ведь восточнославянская лексикография раннего Нового времени и сегодня остается малоизученной областью языковой культуры допетровского времени. Вводя эти тексты в научный оборот, исследователи надеются окончательно опровергнуть гипотезу о «грекофильстве» чудовского круга и раскрыть роль латинских и польских источников[[3]](#footnote-3).

В рамках проекта планируется объединить оцифрованные с помощью программы автоматического распознавания текста Transkribus[[4]](#footnote-4) словари и манускрипты в единую базу данных с возможностью представления в форматах TEI и RDF. В дальнейшем планируется связать словари между собой на уровне лемм, возможна также имплементация в исторический подкорпус НКРЯ[[5]](#footnote-5). Полученная в результате проекта база данных позволит исследовать тенденции развития письменного церковнославянского языка в контексте взаимодействия с другими европейскими языками, в том числе проследить трансформацию языка от русинской (староукраинской, старобелорусской) и церковно-славянской лексики к русской.

В нашей курсовой работе мы выполнили часть задач по введению словаря Епифания в научный оборот, представив данные в формате TEI XML. Задача дипломной работы – продолжить начатый процесс и перейти к этапу презентации словаря в формате RDF (Resource Description Framework) для будущего формирования сети связанных данных проекта.

Цель работы – используя существующие наработки в области цифровой лексикографии, разработать модель для репрезентации Славяно-греко-латинского словаря Епифания Славинецкого в формате RDF и специализации Turtle.

Для этого потребуется выполнить следующие задачи:

1. Провести краткий обзор проделанной работы по трансформации словаря в TEI
2. Изучить структуру модели RDF
3. Исследовать мировой опыт трансформации словарей в формат RDF
4. Разработать RDF- схему для словаря Епифания
5. Описать фрагмент словаря на языке Turtle и протестировать через запросы SPARQL
6. Провести попытку предварительного сравнительного анализа ГСЛ и Польско-латинско-греческого тезауруса Григория Кнапского
7. Опубликовать Python-библиотеку по трансформации .csv файла словаря в открытом доступе в репозитории Python Package Index (PyPI)

# Епифаний Славинецкий и его «Лексикон греко-славяно-латинский»

Епифаний Славинецкий, постриженик Киево-Печерской лавры, в 1649 году был приглашен царем Алексеем Михайловичем в Москву для участия в книжных реформах, подготовки нового издания Библии и организации школы для обучения детей греческому языку[[6]](#footnote-6).

Хотя Епифаний создал большое количество переводческих и филологических трудов, многие его работы до сих пор не опубликованы и не изучены. Среди прочих, известен его перевод лемм толкового словаря латинского языка Амвросия Калепина, а также «Филологический лексикон» - объяснения терминов Священного Писания на греческом языке с латинскими соответствиями и вписанными позднее кириллическими[[7]](#footnote-7).

По предположениям исследователей, «Лексикон греко-славяно-латинский» был написан в 1660е-1680е годы (между 1664 и 1680 годами, как считает С. Н. Брайловский[[8]](#footnote-8)) и включает от 75 до 85 тысяч греческих лемм с их переводами на церковнославянский и латинский языки[[9]](#footnote-9). Автограф рукописи находится в Государственном историческом музее[[10]](#footnote-10).

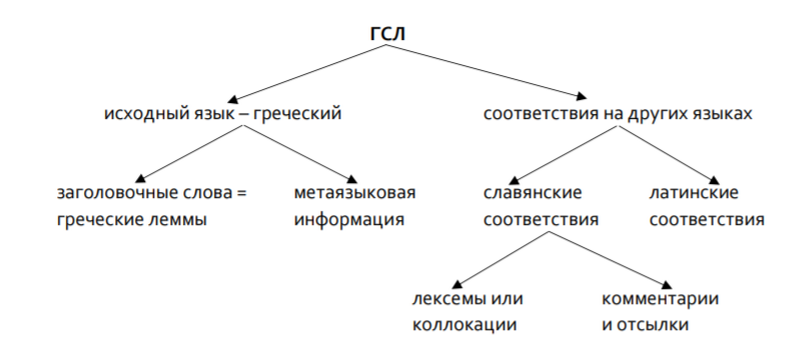
Словарь Епифания включает славянскую лексику других словарей, входя в предшествующую ему лексикографическую традицию. По мнению Т.В. Пентковской, «В переводах Максима Грека и особенно Епифания Славинецкого и Евфимия Чудовского действует принцип, характерный для построения словарной статьи в лексиконах: привести как можно больше лексических эквивалентов для одного понятия,»[[11]](#footnote-11) - большой объем словаря может говорить об использовании Епифанием множества как славянских, так и латинско-польских источников.

При составлении перевода Калепина Епифаний, знавший польский, пользовался и польско-латинскими источниками, например, польско-латинско-греческим тезаурусом Григория Кнапского – гипотеза О. Б. Страховой о том, что Епифаний пользовался им и при составлении словаря, требует своего исследования[[12]](#footnote-12), ведь известно, что польско-латинские источники активно использовались монахами Чудовского круга[[13]](#footnote-13). Мы предпримем попытку связать словари Славинецкого и Кнапского в формате RDF и найти структурные параллели в рамках нашей работы.

# Электронное представление словаря Епифания в формате TEI

Изображение выглядит как текст

Автоматически созданное описаниеПри конвертации исходного файла словаря в формате таблицы .csv мы опирались на руководство проекта Text Encoding Initiative, в частности раздел *Dictionaries*[[14]](#footnote-14) и спецификации TEI Lex-0[[15]](#footnote-15), а также на русскоязычное учебное пособие[[16]](#footnote-16) В. П. Захарова.

Структура словарной статьи достаточно четкая и последовательная: греческая лемма, грамматическая информация (родительный падеж для существительных и прилагательный и указание рода, форма аориста и перфекта для глаголов), церковнославянские и латинские переводы. Порой Епифания добавляет семантический комментарий, включающий смысловую трактовку или отсылку к синониму.

В нашей разметке мы по возможности придерживались данной схемы, Общий порядок дерева XML для словарной статьи таков: греческое слово, его форма и грамматические признаки, затем в тэге <sense>, в который входит славянская форма, и затем снова тэг <sense> внутри славянской формы, в который входит латинская форма. Все три языковых уровня мы старались приблизить по структуре и оформлению, чтобы в дальнейшем можно было инверсировать словарь и принимать любой язык за исходный.

Фрагмент отсканированной копии словаря и схема предоставлены И. А. Подтергерой

Работа кода на языке python устроена следующим образом: программа при помощи пакета *pandas* обходит каждую строчку и в соответствии с заданными условиями задает тэги.

<entry xml:id="Βαθέως" xml:lang="grc">

<form type="hyperlemma">

<orth>βαθύς</orth>

</form>

<form type="lemma">

<orth>Βαθέως</orth>

</form>

<sense xml:id="Βαθέως.1">

<cit type="translationEquivalent" xml:lang="chu">

<form xml:id="глꙋбокѡ [corr. < -бꙋк-]">

<orth>глꙋбокѡ [corr. < -бꙋк-]</orth>

</form>

<sense xml:id="глꙋбокѡ [corr. < -бꙋк-].1">

<cit type="translationEquivalent" xml:lang="lat">

<form type="hyperlemma" xml:id="profundere">

<orth>profundere</orth>

</form>

<form type="lemma" xml:id="profunde·">

<orth>profunde·</orth>

</form>

</cit>

</sense>

</cit>

</sense>

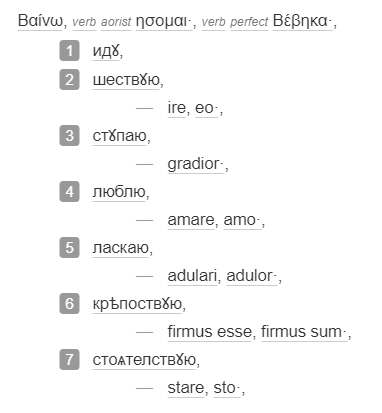
</entry>

Слева представлена разметка статьи для слова *Βαθέως*, у которого по одному переводу на два других языка. Тэги словарного описания (<entry> <sense> или <cit>) обладают атрибутами xml:id или xml:lang для облегчения навигации по документу.

Данные из столбца с метаинформацией выражены тэгом <form type="inflected">. В соответствии с рекомендациями TEI Lex-0 мы обобщили грамматические тэги <pos>, <number>, <case> и <gen> в атрибуты одного тэга <gram>. Кроме прочего наш выбор обусловлен тем, что в тексте словарной статьи нет прямого указания на число, падеж или часть речи: все эти характеристики вложены в артикль (ἡ). Мы бы хотели кодировать как текст лишь то, что напрямую есть в статье.

Для тестирования созданного XML файла мы выложили его на платформу для публикации словарей Lexonomy[[17]](#footnote-17), обеспечивающую поиск и навигацию по вхождениям.

Ручная настройка словаря и возможность загрузки собственной схемы DTD (Document Type Definition) позволили откалибровать внешний вид словаря, скрыв ненужные элементы и выделив наиболее важные. На этапе написания курсовой работы еще не было возможности способа сделать ссылки на другие вхождения в случае указания на них в словаре («*тожде*»). Lexonomy постоянно развивается[[18]](#footnote-18), и недавно появилась возможность связать два словаря и, в теории, вхождения внутри одного словаря – однако пока наша попытка настроить ссылки в словаре Епифания не увенчались успехом.

В целом, Lexonomy оказался очень удобным инструментом в публикации словаря на данном этапе. Ниже представлены некоторые примеры статей; [ссылка](https://www.lexonomy.eu/gsldict) на страницу словаря.

Изображение выглядит как текст

Автоматически созданное описание

В рамках дипломной работы мы доработали скрипт python, чтобы создать собственную библиотеку и выложить ее в открытый доступ. Библиотеку можно найти по адресу <https://pypi.org/project/xmlexicon/>, а инструкцию по установке и использованию – на главной [странице](https://github.com/wildmary/TEI-RDF-encoded-Dictionary#python-package) нашего проекта на GitHub. Программа принимает .csv документ, обрабатывает и записывает результат в .xml файл:



Завершив трансформацию словаря в TEI XML, мы приступили к следующему шагу – переводу из XML в RDF для ввода словаря в Semantic Web.

# RDF: моделирование для всемирной сети

RDF, Resource Description Framework – это набор спецификаций, разработанный группой World Wide Web Consortium[[19]](#footnote-19) для формализованного описания связанных данных, или ресурсов. Понятие *ресурс* крайне широко: это может быть и физический документ, и абстрактное понятие, и словарь – главное, чтобы этот ресурс можно было идентифицировать как уникальный набор данных.

Чтобы обеспечить уникальность каждого ресурса, ему присваивается идентификатор – URI (Uniform Resource Identifier). URI имеет определенный формат: для описания веб документов используется URL, для ссылок на другие, физические или абстрактные объекты автором формируется свой словарь и алгоритм построения идентификатора.

Модель данных RDF предполагает описание данных при помощи трехчастных утверждений, или триплетов (*triplets*), следующих структуре Субъект – Предикат – Объект. Субъект представляет собой сам ресурс, который требуется описать, например, Греко-Славяно-Латинский словарь. Он имеет уникальное имя в формате URI. Объект содержит в себе свойство субъекта, которое мы хотим ему присвоить: например, тот факт, что словарь написан от руки и является манускриптом. Формат объекта не ограничен строкой: он может быть и ссылкой на другой объект (URI автора), и датой (создания словаря) и числом (страниц). Наконец, предикат служит связующим звеном между субъектом и объектом, описывая характер их отношений. В итоге получается трехчастное высказывание, к примеру *«ГСЛ словарь –> создан автором –> Епифаний Славинецкий», «ГСЛ словарь –> имеет формат –> Манускрипт».*

Во избежание ошибок при трактовке и для простоты использования предикатам также присваиваются URI, которые имеют формально описание и набор сценариев использования. В таком случае предикаты становятся свойствами, которые описываются разработчиками и собираются в словари, или онтологии (мы будем использовать термин «онтология», чтобы избежать путаницы со словарем Епифания). Онтология, таким образом, представляет собой набор классов и свойств, который описывает модель определенных данных. Некоторые онтологии имеют широкую специализацию (например, RDF Schema[[20]](#footnote-20), которая описывает саму модель данных RDF). Другие более узко направленны, например модуль для описания лексики OntoLex Lemon[[21]](#footnote-21). Доступность, открытость и универсальность (в своей сфере) такого набора свойств расширяет диапазон использования и способствует большей совместимости данных. Так, используя одни термины для описания разных текстов и словарей XVII-XVIII веков, мы сможем анализировать их в их совокупности, устанавливать логические связи и определять взаимные влияния.

Для реализации модели RDF используются разные форматы, такие как RDF/XML, JSON-LD, N3 и связанный с ним по структуре Turtle. Turtle, или Terse RDF Triple Language, является одной из наиболее популярных спецификаций из-за ее удобства для восприятия человеком. Простота структуры и лёгкость при ручном написании и редактировании стали обоснованием выбора Turtle в качестве инструмента для нашей работы.

## OntoLex Lexicon Model for Ontologies (lemon)

Модель для описания лексиконов была разработана созданной в 2011 году международной группой Ontolex[[22]](#footnote-22) как дополнение к основным онтологиям, таким как RDF(S) и OWL, чтобы восполнить пробелы при описания естественного языка и онтологических концептов. Инструментами, предложенными в *lemon*, можно построить детализованную репрезентацию лексикона от лингвистического описания до установления переводных пар слов. В настоящее время эту модель можно назвать стандартом для представления лексической и лингвистической информации в формате Linked Data: с ее помощью были построены открытая база „*LiLa*: Linking Latin“[[23]](#footnote-23) и портал DBnary[[24]](#footnote-24), на котором представлена лексическая информация с Викисловаря в связанном RDF формате, а также ведется работа проектом K Dictionaries[[25]](#footnote-25) по переводу различных двуязычных словарей машиночитаемый формат.

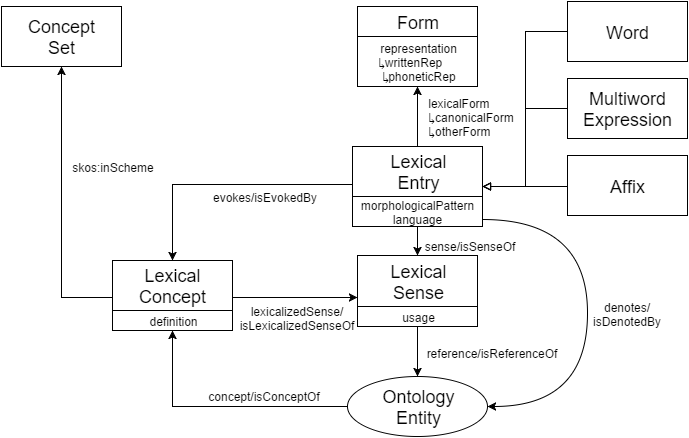
Ниже представлена схема основы *lemon*[[26]](#footnote-26), которая содержит базовые элементы, необходимые для описания слова, в нашем случае - словарного вхождения в словаре Епифания.

Рис.

1. Основа lemon

Модель строится на основе лексической единицы *Lexical Entry*. Ей может быть как слово, так и часть слова, так и фраза, случае словаря Епифания – лемма. У лексической единицы есть грамматически связанные формы набор смыслов, относящийся ко всем этим формам. В дополнение к основе доступны модули, описывающие синтаксис и рошлогоднюю попыткуику, варианты форм, переводы и метаданные, а также позволяющие разделить фразовое вхождение на словарные единицы – некоторыми из них мы воспользовались.

Онтологию *lemon* можно использовать частично. На данном этапе мы решили воздержаться от включения в модель ГСЛ лексических концептов и ссылок на онтологии ввиду недостаточности данных: эти элементы описывают семантику, которая в переводном словаре практически отсутствует.

При работе с онтологией lemon используются и другие модели. Ведется постоянная работа по расширению основной онтологии и дополнению ее новыми модулями. Так, в 2019 году на конференции eLex группой OntoLex был представлен[[27]](#footnote-27) лексикографический модуль *lexicog*[[28]](#footnote-28). Кроме того, мы использовали морфологический модуль *morph*[[29]](#footnote-29), который на момент написания работы еще не был опубликован и находится на стадии разработки[[30]](#footnote-30).

Онтология *lexinfo*[[31]](#footnote-31) использовалась при описании морфосинтаксических свойств: числа, рода, части речи, времени. Библиографическая онтология *bibo*[[32]](#footnote-32) предоставил инструменты для более детального описания физического словаря Епифания: форму, статус публикации.

## Мировой опыт конвертирования словарей в машиночитаемый формат

Следующим шагом после преобразования в электронный формат стали попытки связать словари в «Semantic Web» в виде связанных данных (Linked data). Публикация словарей в едином облаке в Интернете облегчает к ним доступ со стороны пользователей и исследователей с одной стороны, а с другой стороны, подход связанных данных позволяет отойти от традиционной жесткой структуры печатного словаря к более гибкой форме графа, где все связи четко и однозначно определены, превращая словарь в сеть независимых лексических элементов[[33]](#footnote-33).

Еще до завершения работы над онтологией lemon предпринимались попытки преобразования словарей в RDF, в результате которых были выявлены и предложены новые свойства (*properties*), которые позднее были включены в *lemon*. Например, в процессе разработки ресурса DBnary[[34]](#footnote-34) возникла необходимость обозначить перевод с одного языка на другой, в результате чего был предложен класс dbnary:Translation, включающий такие свойства как dbnary:isTranslationOf, которые сейчас вошли в состав модуля *vartrans*.

Развитие lemon идет параллельно с работой над проектом «The Global series of K Dictionaries», в составе которого исследователи и разработчики, также входящие в группу OntoLex. Проект ставит своей целью создание «межъязыковой, многоуровневой мозаики лексикографических ресурсов, которые эволюционируют в рамках единой системы и обладают одинаковой внутренней и внешней структурой словарного вхождения, которая может адаптироваться под нужды различных языков»[[35]](#footnote-35). Исходный формат словарей – XML. Чтобы расширить возможности перевода, связи двух языков и привязки к внешним ресурсам, анализа и преобразования данных на межсловарном уровне, возникла потребность перевода имеющихся данных в формат Linked Data с помощью OntoLex *lemon*.

Первая попытка была проведена в 2014 году, когда была начата работа над трансформацией немецкоязычного тезауруса. Были проведены параллели между элементами XML и классами RDF и обнаружены пробелы в существующей на тот момент версии *lemon*, а именно способы описания фразовых вхождений, семантических связей (синонимии, например). В дальнейшем предложенные авторами свойства и классы были включены в текущую версию *lemon*.

В 2016 году была проведена большая работа по моделированию в формате RDF мультиязычного испанского словаря, содержащего переводы лемм, определений и примеров на голландский и английский языки, а также варианты произношения и грамматическую информацию. На этом этапе снова были предложены дополнения к существующей модели lemon во избежание потери данных, содержащихся в XML-версии, например, kd:TranslationExampleCluster[[36]](#footnote-36).

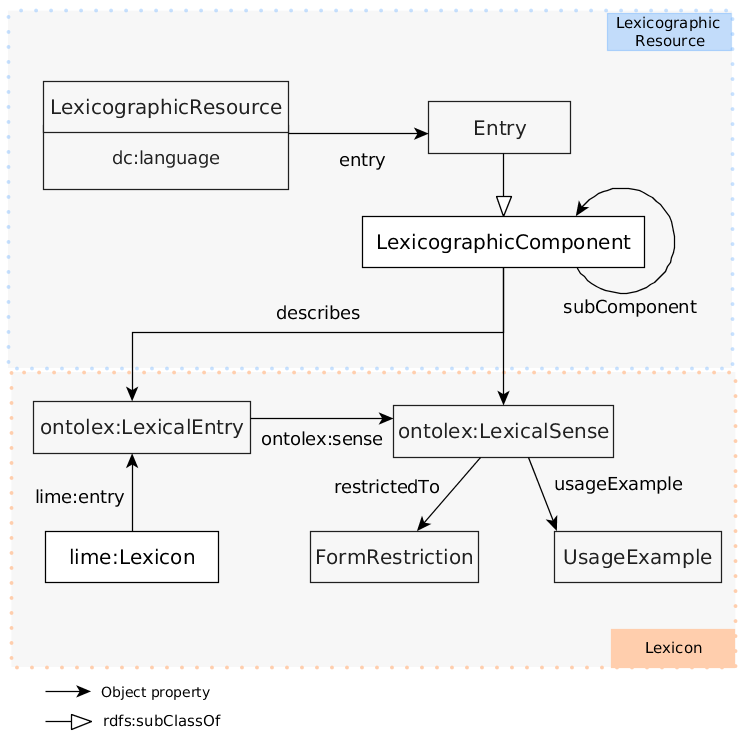
Наконец, в 2019 году, после проведенной работы над дополнением модели *lemon*, лексикографический модуль *lexicog* был опробован на словарях K Dictionaries. Новый модуль направлен на формальное представление словарей и структуры словарных статей, отделяя структурную словарную метаинформацию от собственно словарного описания (см Рис. 2). Так, вводится класс lexicog:LexicographicResource, который представляет список словарных статей в самом словаре (в случае словаря Епифания – только древнегреческие слова), в отличие от класса lime:Lexicon, который представляет собой все слова определенного языка (древнегреческого, церковнославянского или латыни). lexicog:Entry и lexicog:LexicographicComponent описывают конкретные словарные статьи. Таким образом, в класс lexicog:Entry войдут лишь древнегреческие слова, а в класс ontolex:LexicalEntry – слова всех трех языковых корпусов, представленных через lime:Lexicon.

Рис. 2: Схема лексикографического модуля[[37]](#footnote-37)

Нам также интересна работа других членов группы Ontolex по преобразованию билингвальных словарей с платформы Apertium, занимающейся машинным переводом. В результате часть словарей была включена в облако Linguistic Linked Open Data[[38]](#footnote-38). В качестве примера в статье[[39]](#footnote-39), вышедшей в 2017 году, описывается преобразование переводов с английского на испанский и с испанского на португальский. Этот случай особенно полезен к изучению, так как структура работы схожа со случаем словаря Епифания, который представляет собой набор переводов с древнегреческого на церковнославянский и с церковнославянского на латынь. Авторы использовали концепты Translation и TranslationSet, которые позднее войдут в модуль *lemon-vartrans*. Translation служит мостом между двумя случаями ontolex:LexicalSense на разных языках. Список всех переводов, или пар смыслов, содержится в TranslationSet.

Несмотря на то что случай конвертации в RDF трех словарей, описывающих диалекты арабского языка в Каире, Дамаске и Тунисе, не так близок нам по подходу, все же представляет собой любопытный пример работы со сложно структурированными словарными статьями. Словари состоят из списка лемм, указания на корень (в арабском языке состоящий из согласных, например, «bwb»)[[40]](#footnote-40) и переводы на несколько европейских языков. Опираясь на модель lemon OntoLex, авторы не пошли путем фиксации переводов через свойства Translation, предложив записывать эквиваленты на других языках как новые объекты ontolex:LexicalSense со свойствами rdfs:label (перечисляющим порой несколько вариантов перевода через запятую) и ontolex:language. При этом был создан отдельный субъект ontolex:LexicalSense, который, как мы предполагаем, позднее станет ontolex:LexicalConcept и представляет собой один концепт (например, ontolex:LexicalSense\_Door[[41]](#footnote-41)), который связан со всеми словарными вхождениями, у которых есть перевод на один из европейских языков (rdfs:label). Таким образом, три словаря были связаны, используя этот «смысло-ориентированный» подход.

Если же обратиться к опыту привязки не словаря к словарю, а словаря к тексту, то работа[[42]](#footnote-42) по публикации медицинского трактата со ссылками на онлайн-словарь может послужить примером и опорной точкой при будущем анализе славянских переводов. Целью проекта было привязать к словам (леммам) написанного на среднефранцузском языке трактата вхождения из этимологического словаря французского языка, чтобы облегчить понимание текста. Исходные данные в формате LATEX были сначала переведены в формат TEI XML с интеграцией определений из словаря, затем в структуру XML были добавлены атрибуты RDFa – структуры, позволяющей реализовывать RDF внутри HTML или XHTML-разметки. Авторам статьи было необходимо расширить словарь TEI, чтобы обеспечить связь с внешними ресурсами: были добавлены атрибуты @about, @property и @resource к классу att.global.linking[[43]](#footnote-43). Затем была произведена конвертация в формат HTML5+RDFa и публикация на сайте проекта[[44]](#footnote-44). Учитывая лёгкость конвертации в Turtle, появляется возможность анализа текста при помощи SPARQL-запросов и привязки его к другим словарям при необходимости.

Данный опыт может пригодиться при разработке базы данных, включающей трехъязычный словаря Епифания и переводы чудовских книжников с польского и латинского языков (книга Иова в пере­во­де Моисея Чудовского и Библия Пискатора по экземпляру 1643 г. в переводе Евфи­мия Чудовского) и перевод-редактура «Андриант», выполненный с греческого Евфимием Чу­довским. Интегрировав эти тексты в единую систему при помощи RDF, можно будет не только публиковать переводы со ссылками на различные словари, но и прово­дить сравнительный анализ лексики и делать выводы о рецепции словаря Епифания в пе­ре­водческой практике второй половины XVII в., а также будущем ответить на вопрос о степени вме­шательства Евфимия Чудовского в переводы Епифания Славинецкого, разнице между переводами с греческого и польского/латинского и в целом исследовать тенденции развития церковнославянского и русского языка в контакте с европейскими языками.

# Конвертация Словаря Епифания в RDF

## Методология

Помимо рекомендаций, предоставленных в описаниях моделей и онтологий, и практических примеров, описанных в вышеупомянутых статьях, мы обращались к руководствам по генерации многоязычных[[45]](#footnote-45) и двуязычных[[46]](#footnote-46) словарей в формате LD, опубликованных общественной группой Консорциума Всемирной паутины (W3C Community Group).

Процесс преобразования словаря Епифания в формат RDF можно разбить на несколько шагов:

1. Анализ исходных данных – анализ структуры (DTD) словарного вхождения в формате TEI XML с точки зрения иерархии и связей между лексическими единицами.
2. Разработка стратегии по именованию субъектов – дизайн URI для самого словаря, словарных вхождений и других элементов схемы.
3. Проектирование модели (схемы) RDF – формирование схемы, описывающей ГСЛ словарь без потери данных, на основе онтологии lemon с дополнением из модуля lexicog и других онтологий. Схема должна быть гибкой в плане адаптации под другие славянские словари, с которыми нам придется работать в будущем.
4. Ручной перевод из XML в RDF – трансформация образцов словарных вхождений в формат Turtle вручную и параллельная доработка схемы.

Так как детальное описание TEI модели уже было представлено выше, остановимся на главных компонентах, которые было важно перенести в новый формат, а также отсечем данные, которые не представляют ценности для анализа на данном этапе. В схеме RDF должны присутствовать словарные вхождения (лемма, гиперлемма), репрезентация их форм и грамматическое описание, а также семантическое описание: переводы и «комментарии» Епифания, либо указывающие на дополнительный смысл *(«Таже имѧ конѧ Ахіллева»),* либо отсылающие к синонимам *(«тожде єже Βαλάντιον»).* Структурные элементы бумажного словаря, такие как указания страниц и разворотов, напротив, из будущей схемы исключены.

## Правила URI

Унифицированные идентификаторы ресурсов, или URI, создавались c ориентацией на повторное использование и связь элементов, при этом представляя уникальный субъект. С одной стороны, необходимо исключить коллизии в Интернете и в рабочей среде исследователя, с другой, обеспечить возможность связывать различные словари через использование одинаковых URI.

В формировании URI для словаря Епифания мы придерживались рекомендаций программы ISA (сейчас - Semantic Interoperability Community) Европейской комиссии. Общая структура URI выглядит следующим образом: *http://{domain}/{type}/{concept}/{reference}[[47]](#footnote-47)* , где домен – это комбинация хоста и раздела, к которому относится текущая работа. Так как словарь в формате RDF еще не опубликован на веб-сервере, домен и другие элементы можно легко изменить – пока же мы используем www.slav.uni-heidelberg.de. Тип представляет предназначение объекта: следуя примеру проекта трансформации словарей *Apertium*[[48]](#footnote-48), мы выбрали тип «*id*», который описывает реальные объекты. Концепт представляет конкретный объект, в нашем случае – ГСЛ, который пока зашифрован как *«Slavineckij».* *Reference* будет отсылать к конкретным субъектам схемы RDF. Таким образом, базовый @prefix в файле, содержащем описания словаря, будет выглядеть как https://www.slav.uni-heidelberg.de/id/Slavineckij/, а полный URI лексикографического ресурса «Лексикон греко-славяно-латинский» - как https://www.slav.uni-heidelberg.de/id/Slavineckij/Epifanij\_dict.

URI для субъектов схемы, в основном, строятся вокруг элемента xml\_id. На лексикографическом уровне к xml\_id спереди добавляются «entry» или «lcomp\_» для представления словарного вхождения и лексикографического компонента соответственно (например, :entry\_βαλιός и :lcomp\_βαλιός). На уровне лексикона элементы добавляются после xml\_id, кроме обозначения ontolex:LexicalEntry, которое равно xml\_id. В таблице ниже представлены правила и примеры формирования URI:

|  |  |  |
| --- | --- | --- |
| Элемент | URI класса | URI элемента |
| Словарное вхождение | lexicog:Entry | entry\_+xml\_id *(напр. :entry\_βαλιός)* |
| Лексикографический компонент | lexicog: LexicographicComponent | lcomp\_+xml\_id *(:lcomp\_βαλιός)* |
| Слово | ontolex:LexicalEntry | xml\_id *(:βαλιός)* |
| Основная форма | ontolex:Form | xml\_id+\_form *(:βαλιός\_form)* |
| Форма при склонении | ontolex:Form | xml\_id+\_infl *(:βαλιός\_infl)* |
| Окончание при склонении | morph:AffixMorph | suffix\_+[окончание] *(:suffix\_ου)* |
| Определение | ontolex:LexicalSense | xml\_id+\_sense *(:βαλιός\_sense)* |
| Лексикон данного языка | lime:Lexicon | *:lexiconGRC, :lexiconCHU, :lexiconLAT* |
| Набор переводов (пар значений в 2х языках) | vartrans:TranslationSet | *:tranSetGRC-CHU* |
| Конкретный случай перевода (пара значений) | vartrans:Translation | sense1+"-"+sense2 *(:βαλιός\_sense-пестръ\_sense)* |
| Связь смыслов (синонимия) | vartrans:SenseRelation | Synonym\_+ sense1+"-"+sense2 *(Synonym\_βαλίος\_sense-βαλιός\_sense)* |

В словаре Епифания есть два вида связи двух смыслов: перевод и ссылка на другое словарное вхождение на том же древнегреческом языке. Второй случай, выраженный графически как «*тожде*» (если отсылает к предыдущему вхождению) или *«тожде еже …»,* мы интерпретировали как случай синонимии.

## Схема RDF

Как уже упоминалось выше, мы опирались на основу модели OntoLex *lemon*. Другие модели и онтологии, использовавшиеся при составлении схемы, будут упомянуты при описании соответствующих элементов.

Ниже представлена окончательная версия сконструированной схемы RDF. Элементы, окрашенные зеленым, составляют «лексикографическую» часть графа, описанную модулем *lexicog*. Красным мы обозначили наиболее важные элементы: словарное вхождение и смысл. Два желтых элемента представляют собой «наборы» данных: lime:Lexicon включает все слова конкретного языка, а vartrans:TranslationSet собрал в себе пары смыслов из двух языков. Оранжевым мы выделили остальные описательные элементы схемы: формы, окончание, связующие звенья между двумя смыслами и семантический комментарий.

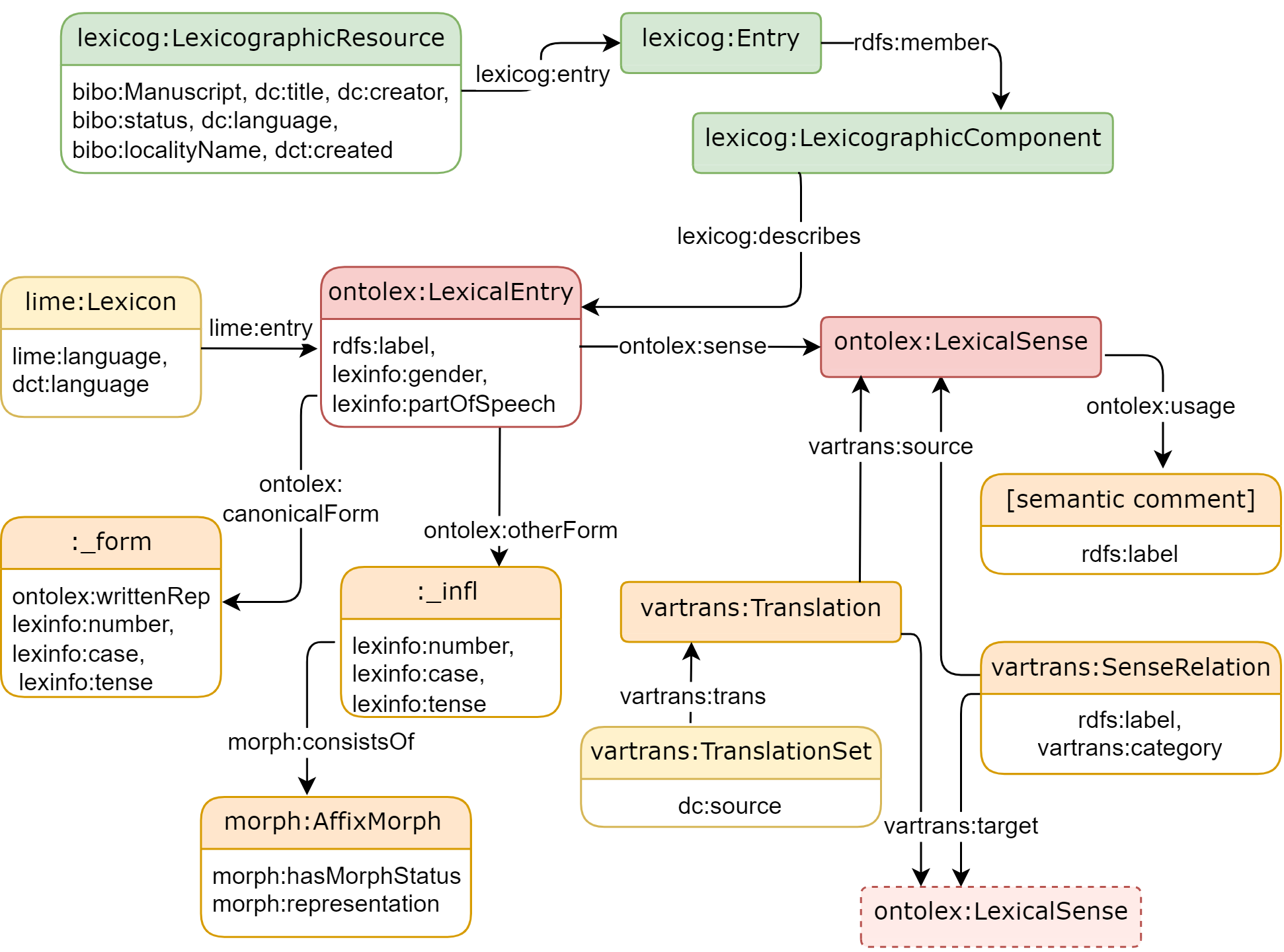


Рис. 3. RDF-граф для описания словаря Епифания

Элемент lexicog:Lexicographicresource описывает манускрипт словаря Епифания и является аналогом элементов <fileDesc> и <profileDesc> из файла XML, показанных на Рисунке 4. Так как онтология lemon не обладает библиографическим модулем на момент написания работы, мы обратились к библиографической онтологии BIBO, разработанной в 2008 году для описания библиографических элементов, таких как документов и цитат. С 2018 года онтология поддерживается и располагается на сайте The Dublin Core™ Metadata Initiative[[49]](#footnote-49). Из этой онтологии были взяты ресурсы:

* bibo:Manuscript – присвоение словарю класса «Манускрипт» для описания физического формата
* bibo:localityName - место создания документа. Этот ресурс отсылает к объекту «Москва» в DBpedia
* bibo:status + bibo:unpublished – статус публикации



Рис. 4.

В дополнение к этим атрибутам были использованы элементы онтологий FOAF, Dublin Core (& terms), RDFs. Не всю информацию, однако, удалось зафиксировать, используя доступные инструменты: так, место хранения манускрипта с точностью до репозитория и ID в онтологии не описаны. Так как эта метаинформация на самом манускрипте не указана и для анализа не важна, мы позволили себе опустить ее в текущей схеме. Ниже можно видеть репрезентацию в Turtle RDF:

@prefix gsl: <https://www.slav.uni-heidelberg.de/id/Slavineckij/> .

@prefix lexicog: <http://www.w3.org/ns/lemon/lexicog#> .

@prefix bibo: <http://purl.org/ontology/bibo/> .

@prefix dct: <http://purl.org/dc/terms/> .

@prefix dbr: <http://dbpedia.org/resource/> .

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix dc: <http://purl.org/dc/elements/1.1/> .

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

gsl:Epifanij\_dict a lexicog:LexicographicResource, bibo:Manuscript ;

dc:title "Лексикон греко-славяно-латинский"@ru ;

dc:creator gsl:Epifanij\_Slavineckij ;

bibo:status bibo:unpublished ;

dc:language "grc", "chu", "lat" ;

bibo:localityName dbr:Moscow ;

dct:created "XVII век" ;

lexicog:entry gsl:entry\_βαλιός, gsl:entry\_βαλίος .

gsl:Epifanij\_Slavineckij a foaf:Person ; rdfs:label "Епифаний Славинецкий"@ru .

dbr:Moscow rdfs:label "Москва"@ru .

В словарь Епифания через свойство lexicog:entry включены все словарные вхождения (древнегреческие): в нашем примере это gsl:entry\_βαλιός и gsl:entry\_βαλίος, идущие в словаре подряд. Так как словарные статьи в словаре Епифания содержат только одну древнегреческую лемму, то lexicog:entry будет связан лишь с одним :LexicographicComponent, который, в свою очередь, описывает одну лексическую единицу ontolex:LexicalEntry:

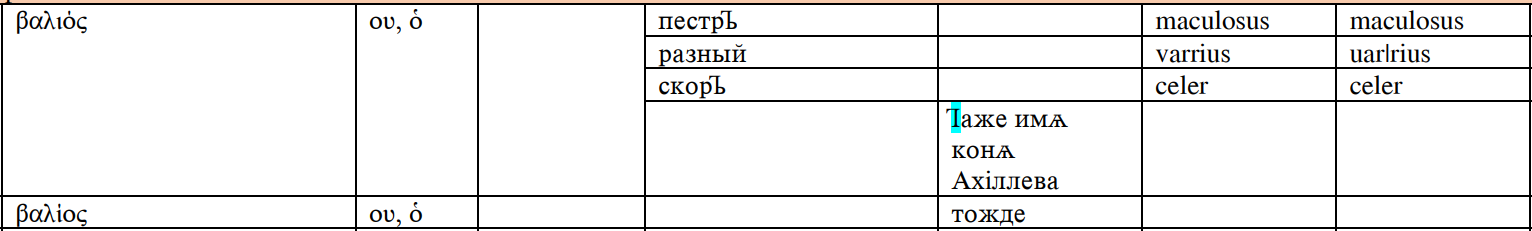
gsl:entry\_βαλιός a lexicog:Entry ; lexicog:subComponent gsl:lcomp\_βαλιός .

gsl:lcomp\_βαλιός a lexicog:LexicographicComponent ;

lexicog:describes gsl:βαλιός .

Вышеописанная структура в формате XML явно не указана ввиду иерархической природы языка – при переводе в плоскую структуру RDF нам показалось важным не потерять эту информацию.

На Рисунке 5 для наглядности представлена исходная таблица .csv, из которой был получен XML-файл; два вхождения gsl:entry\_βαλιός и gsl:entry\_βαλίος послужили образцом для конвертации в Turtle RDF, так как они включают в себя все возможные элементы схемы: грамматическое описание, переводы, синонимия и комментарий.

Рис. 5.

Итак, как мы определили, древнегреческие ontolex:LexicalEntry связаны с физическим словарем-манускриптом. Все слова, к тому же, связаны с лексиконом на соответствующем языке, который в будущем будет включать слова не только из ГСЛ, но и из других словарей, включенных в «Semantic Web». Эта информация также не имеет прямого эквивалента в формате XML. Пример ниже описывает древнегреческий лексикон: согласно рекомендациям[[50]](#footnote-50) W3C, мы включили также ссылки на внешние ресурсы для описания языка.

gsl:lexiconGRC a lime:Lexicon ; lime:language "grc" ;

dct:language <http://id.loc.gov/vocabulary/iso639-2/grc>, <http://lexvo.org/id/iso639-3/grc> ;

lime:entry gsl:βαλιός, gsl:βαλίος .

В таком же формате описаны лексиконы церковнославянского языка и латыни. Ниже представлены вышеописанные элементы RDF в виде схемы[[51]](#footnote-51):

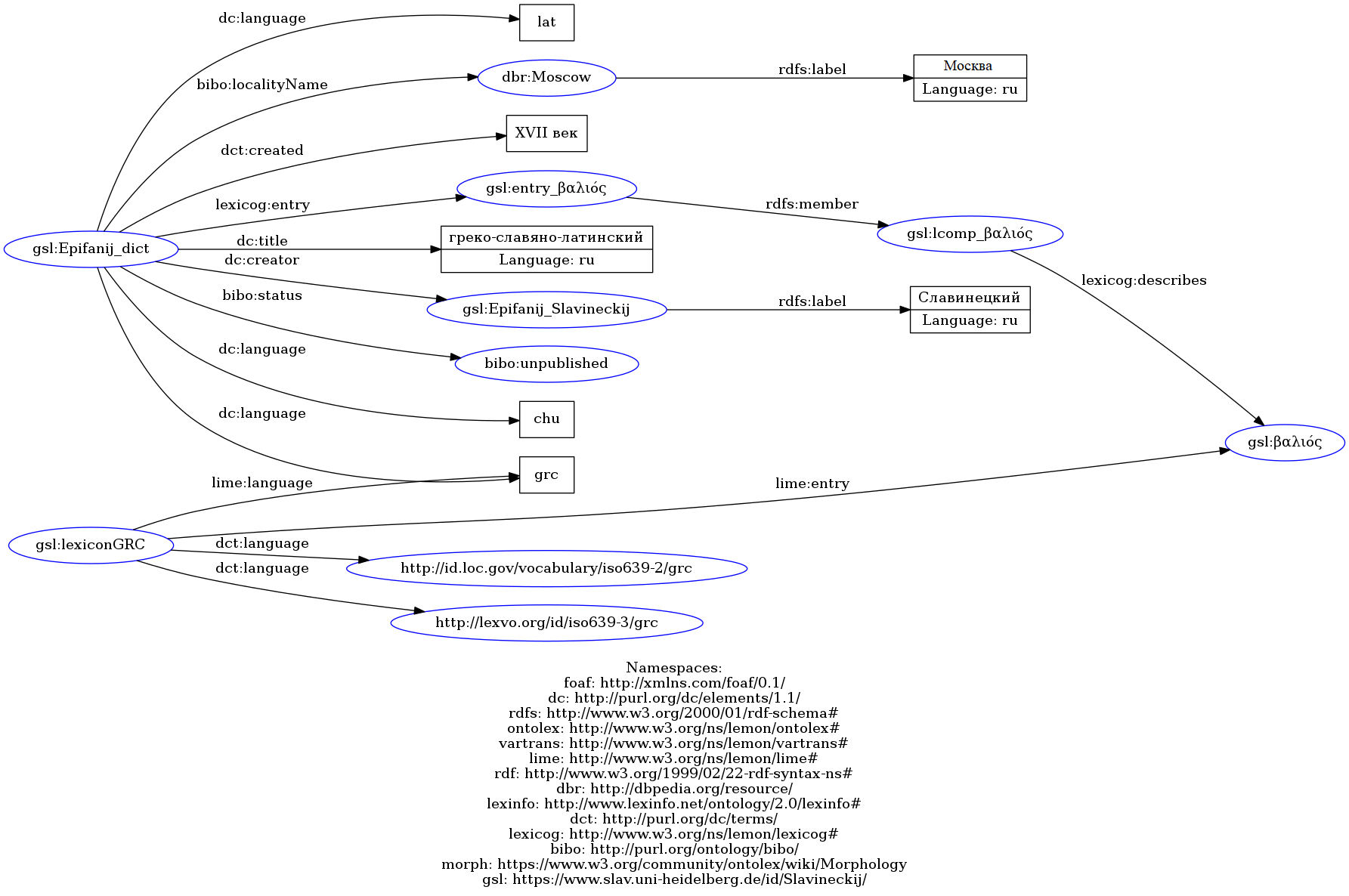


Рис. 6. Описание словаря, лексикона и связь со словом βαλιός

Трансформация грамматического описания слова из XML в RDF была проще с точки зрения поиска эквивалентов в TEI и сохранения исходной структуры двух форм. Первая форма – лемма – стала «канонической» формой; вариант склонения был записан как «другая» форма. Каждая форма имеет изменяемые грамматические атрибуты: число, падеж, время. Часть речи и род связаны напрямую с :LexicalEntry.

При описании окончания при склонении слова мы столкнулись с еще одним пробелом в модели *lemon*, которая пока не поддерживает описание морфем (в XML их можно описать с помощью атрибута extent="part"). Мы обратились к еще не опубликованному и верифицированному W3C морфологическому модулю, который на данный момент разрабатывается группой OntoLex. Мы опирались на отчет[[52]](#footnote-52), представленный на конференции eLex в 2019 году, а также на wiki-страницу[[53]](#footnote-53) разработки модуля. Согласно предложенной схеме модуля, морфемы (morph:AffixMorph) привязаны к форме (в нашем случае - gsl:βαλιός\_infl). Модуль позволяет выразить тип аффикса (слово- или формообразующий), однако разделение на определенные позиционные категории пока не предусмотрено. Ниже представлена часть описания слова *βαλιός* в разных форматах:

gsl:βαλιός a ontolex:LexicalEntry, ontolex:Word ;

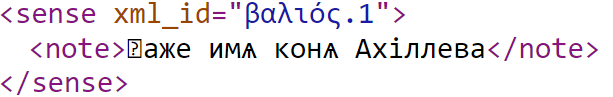
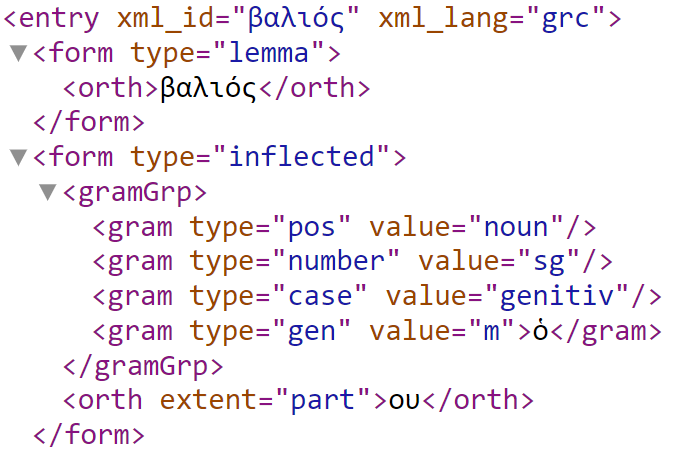


Рис. 7.

ontolex:sense gsl:βαλιός\_sense ;

rdfs:label "βαλιός"@grc ;

lexinfo:gender lexinfo:masculine ;

lexinfo:partOfSpeech lexinfo:Noun ;

ontolex:canonicalForm gsl:βαλιός\_form ;

ontolex:otherForm gsl:βαλιός\_infl .

gsl:βαλιός\_form ontolex:writtenRep "βαλιός"@grc ;

lexinfo:number lexinfo:singular ;

lexinfo:case "nominative"@en .

gsl:βαλιός\_infl lexinfo:number lexinfo:singular ;

lexinfo:case "genitive"@en ;

morph:consistsOf gsl:suffix\_ου .

gsl:suffix\_ου a morph:AffixMorph ;

morph:hasMorphStatus morph:inflectional ;

morph:representation "ου"@grc .

gsl:βαλιός\_sense a ontolex:LexicalSense ;

ontolex:usage [rdf:value "Таже имѧ конѧ Ахіллева"@chu] .

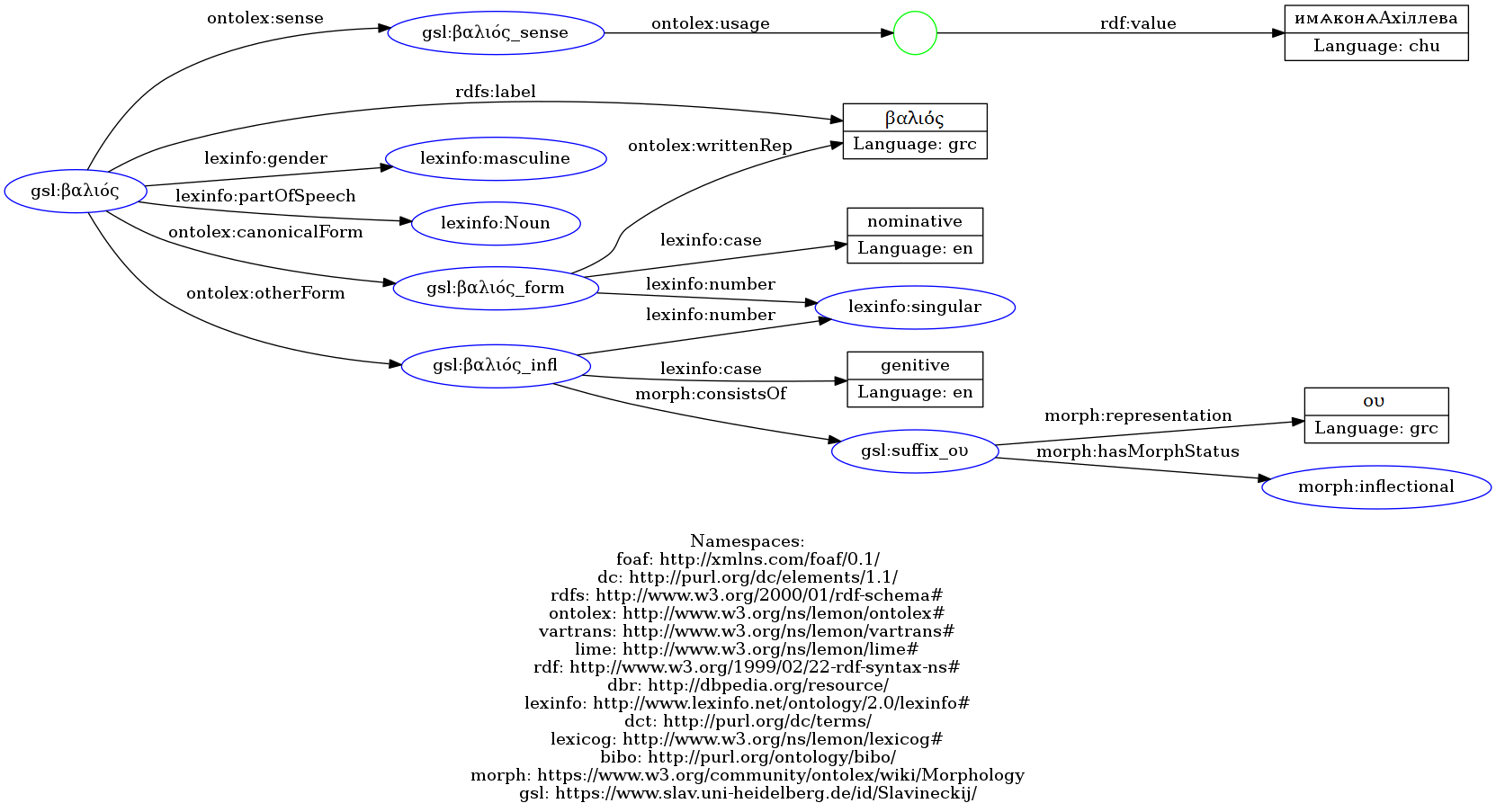
Рис. 8. Грамматическое описание и семантический комментарий к слову βαλιός

Схема на Рисунке 8 также включает в себя семантический комментарий, который Епифаний иногда давал словарным вхождениям: в данном случае слово *βαλιός* используется и как имя коня Ахиллеса – Балия.

Как видно по Рисунку 5, у слова *βαλιός* три славянских эквивалента, у которых в свою очередь по одному переводу на латынь. Слова на церковнославянском и на латыни не попадают в класс lexicog:Entry, так как они не являются словарными вхождениями, но они представлены классом ontolex:LexicalEntry наравне с древнегреческими словами. Каждый субъект :LexicalEntry обладает смыслом ontolex:LexicalSense, который может выражаться через перевод, синонимию или комментарий.

В случае перевода слова (Рис. 9) *βαλιός* его смысл :βαλιός\_sense связан с тремя смыслами церковнославянских слов :пестръ\_sense, :разный\_sense и :скоръ\_sense для соответствующих субъектов :LexicalEntry из :lexiconCHU. Эта связь направленная: элемент vartrans:Translation, работающий посредником между двумя смыслами, имеет свойства vartrans:source (исходный язык – древнегреческий) и vartrans:target (целевой язык – церковнославянский. Набор таких пар смыслов хранится в наборе vartrans:TranslationSet. По той же схеме смоделированы переводы с церковнославянского на латынь – таким образом, субъекты :βαλιός и :maculosus связаны лишь посредственно, через :пестръ. Ниже представлено описание первого значения *βαλιός* на церковнославянском:

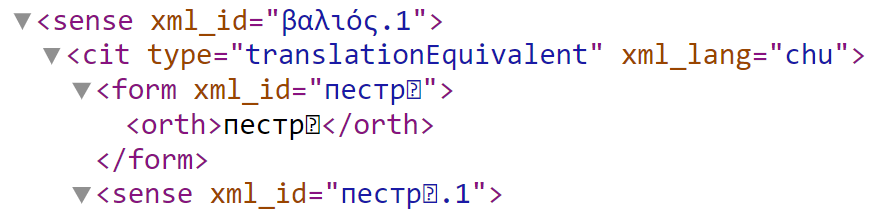


Рис. 9. Перевод

gsl:пестръ a ontolex:LexicalEntry, ontolex:Word ; ontolex:sense gsl:пестръ\_sense .

gsl:tranSetGRC-CHU a vartrans:TranslationSet ;

vartrans:trans gsl:βαλιός\_sense-пестръ\_sense .

gsl:βαλιός\_sense-пестръ\_sense a vartrans:Translation ;

vartrans:source gsl:βαλιός\_sense ;

vartrans:target gsl:пестръ\_sense .

Если же семантический комментарий указывает на тождественность (Рис. 10), два словарных вхождения так же связаны через: source и :target посредством атрибута vartrans:SenseRelation, которому присвоена категория lexinfo:synonym. Ниже введено второе словарное вхождение из Рисунка 5 – *βαλίος*, которое обладает теми же грамматическими и семантическими свойствами, что и предшествующее ему слово *βαλιός*.

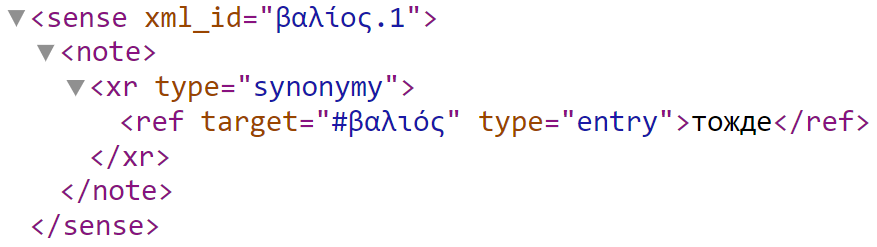


Рис 10. Синонимия

gsl:entry\_βαλίος a lexicog:Entry ; rdfs:member gsl:lcomp\_βαλίος .

gsl:lcomp\_βαλίος a lexicog:LexicographicComponent ;

lexicog:describes gsl:βαλίος .

gsl:βαλίος a ontolex:LexicalEntry, ontolex:Word ; ontolex:sense gsl:βαλίος\_sense .

gsl: Synonym\_βαλίος\_sense-βαλιός\_sense a vartrans:SenseRelation ;

vartrans:source gsl:βαλίος\_sense ;

vartrans:target gsl:βαλιός\_sense ;

vartrans:category lexinfo:synonym ;

rdfs:label "тожде"@chu .

Рисунок 11 демонстрирует схему смысловых связей между словами *βαλίος*, *пестръ* и *maculosus*:

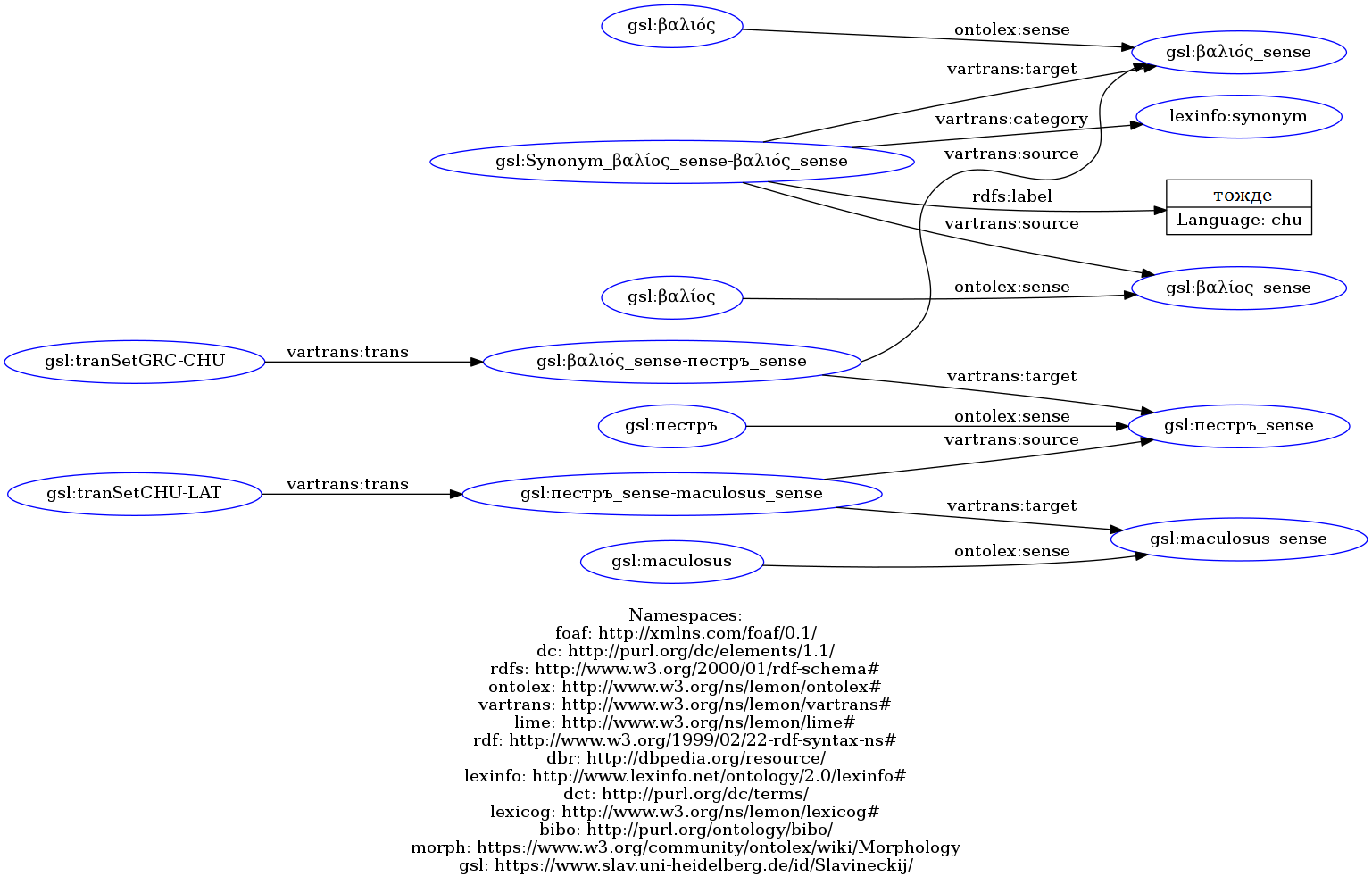


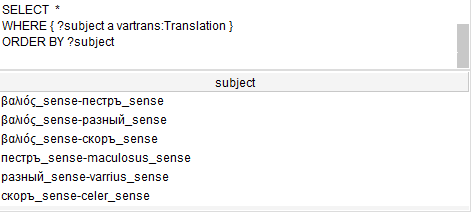
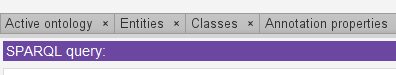
Рис. 11. Переводы и синонимия

Полная [схема](https://raw.githubusercontent.com/wildmary/TEI-RDF-encoded-Dictionary/main/rdf_entry_representation.png) репрезентации двух словарных вхождений *βαλίος* и *βαλιός* доступна в [репозитории](https://github.com/wildmary/TEI-RDF-encoded-Dictionary) проекта на GitHub.

## Валидация через SPARQL

Следующий шаг после разработки схемы RDF– проверка логики – был выполнен с помощью запросов SPARQL[[54]](#footnote-54). SPARQL Protocol and RDF Query Language – язык запросов, созданный для работы с данными в формате RDF и рекомендованный W3C. Запрос строится на той же структуре *«субъект-предикат-объект»*, что позволяет нам при тестировании проверить все элементы схемы.

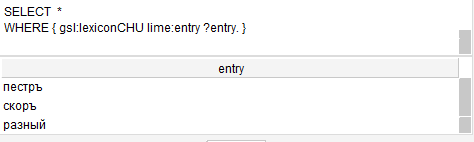
Так, Запрос 1 выводит список всех случаев перевода. Результат содержит шесть объектов vatrtans:Translation: три перевода с древнегреческого на церковнославянский и три – с церковнославянского на латынь.



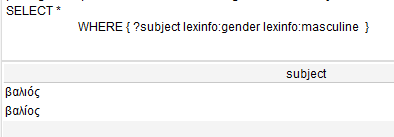
Запрос 1.

Запрос 2 возвращает все слова, входящие в церковнославянский лексикон: сейчас это три варианта перевода слова *βαλίος*.

Запрос 3 показывает все слова мужского рода: *βαλίος* и *βαλιός*.

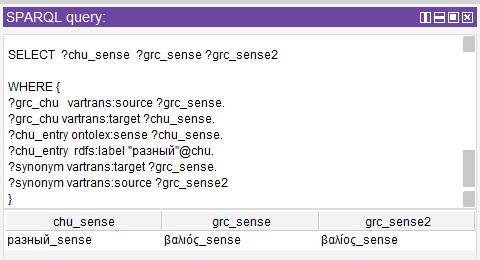


Запрос 2.



Запрос 3.

Более сложные запросы позволяют проследить связь между отдаленными точками графа.



Запрос 4.

В четвертом примере мы попробовали посмотреть, какие древнегреческие эквиваленты есть у слова «разный», при этом допустив не только прямой перевод, но и синонимы к нему. Таким образом, результат показал оба древнегреческих слова.

Наконец, пятый запрос дает переводы слова βαλιός на латынь через посредничество славянских эквивалентов.



Запрос 5.

Сравнивая исходные данные и результаты SPARQL-запросов, можно проверять качество трансформации из одного формата в другой при будущей машинной обработке словаря.

# Перспективы: межкультурный анализ

История создания Греко-Славяно-Латинского словаря остается до конца не изученной: так как словарь превышает по количеству греческих лемм большинство современных ему словарей[[55]](#footnote-55) и пока не найден такой словарь, который можно было бы считать основным источником, - исследователи пытаются найти влияния из нескольких известных Епифанию словарей и манускриптов[[56]](#footnote-56), с которыми он сталкивался в своей переводческой практике.

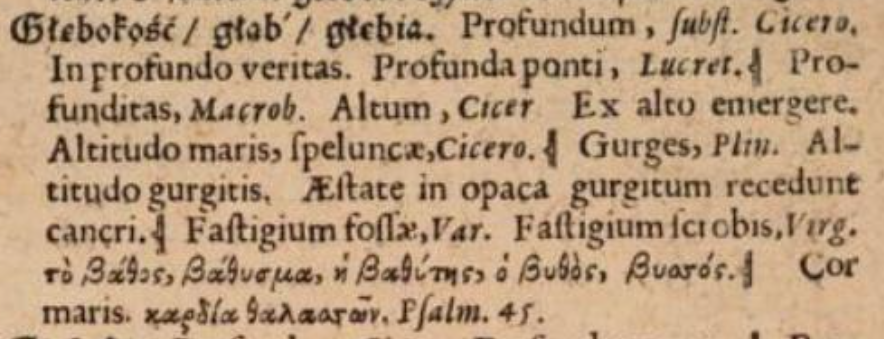
Среди таких мультиязычных словарей особый интерес представляет Польско-латинско-греческий тезаурус Григория Кнапского. Епифаний знал польский язык и пользовался словарем Кнапского при переводе словаря Калепина[[57]](#footnote-57). Хотя польский словарь достаточно мал, чтобы служить основным источником греческих слов, можно попытаться установить степень влияния его на формирование Епифанием словарных статей и церковнославянских и латинских эквивалентов.

Словарь Кнапского на данный момент доступен только в формате PDF/Djvu с достаточно плохим качеством распознавания. В рамках большого проекта планируется оцифровка и перевод в машиночитаемый формат. Пока же мы попробовали вручную сравнить словари Епифания и Кнапского на примере трех польских лемм *Głeboko,* *Głebokość и Głeboki*[[58]](#footnote-58) и их эквивалентов в ГСЛ.

Немного о структуре[[59]](#footnote-59) словаря Кнапского: хотя четкой структуры словарной статьи нет, постоянными элементами являются польские леммы в алфавитном порядке, включение синонимов и использование примеров из литературы. За польским словом следует значение на польском, латинские и греческие синонимы. Грамматическое описание встречается редко.

Так как на данном этапе полная трансформация трех образцов словарных статей не требуется, мы позволили себе описать в Turtle лишь переводы, придерживаясь предложенной ранее схемы RDF. Ниже представлено описание польского лексикона и польской леммы *Głebokość* и ее однословных переводов (vartrans:Translation), которые мы смогли опознать (описания польских синонимов, а также греческих и латинских лемм здесь мы опустили – они содержат лишь свойства ontolex:sense и rdfs:label).

Рис. 12. Фрагмент словаря Кнапского – слово Głebokość.



@prefix knap: <http://hint.org.pl/hid=A2076/>.

gsl:lexiconPOL a lime:Lexicon ;

lime:language "pl" ;

dct:language <http://id.loc.gov/vocabulary/iso639-1/pl>, <http://lexvo.org/id/iso639-3/pol> ;

lime:entry knap:Głeboko, knap:Głebokość,knap:Głeboki .

knap:Głebokość a ontolex:LexicalEntry, ontolex:Word ;

ontolex:sense knap:Głebokość\_sense ;

rdfs:label "Głebokość"@pl .

gsl:tranSetPOL-LAT a vartrans:TranslationSet ;

vartrans:trans knap:Głebokość\_sense-Profundum\_sense,

knap:Głebokość\_sense-Profunditas\_sense,

knap:Głebokość\_sense-Altum\_sense,

knap:Głebokość\_sense-Gurges\_sense,

knap:Głebokość\_sense-Fastigium\_sense.

gsl:tranSetPOL-GRC a vartrans:TranslationSet ;

vartrans:trans knap:Głebokość\_sense-Βάθος\_sense,

knap:Głebokość\_sense-Βάθυσμα\_sense,

knap:Głebokość\_sense-Βαθύτης\_sense.

Быстрый поиск по словарю Епифания показывает, что он содержит все три греческие леммы из словаря Кнапского, которые есть в обработанной части словаря (две леммы на *Βυ* пока не доступны в машиночитаемом формате):

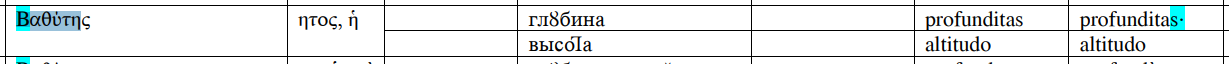
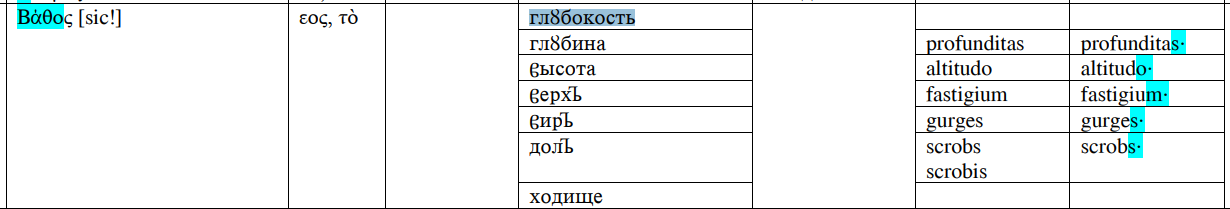


Рис. 13. Греческие эквиваленты слову Głebokość из ГСЛ

Можно заметить параллели между двумя словарями и закономерности в оформлении Епифанием переводов: хотя количество церковнославянских эквивалентов различается, часть из них повторяется от леммы к лемме. Так, «*глꙋбина*» и «*высота*» встречается во всех трех статьях, «*виръ*» - в двух. То же можно проследить и латинских переводах: четыре из пяти латинских переводов из ГСЛ есть в словаре Кнапского (*altitudo* в составе двусложного примера).

Порядок латинских переводов хотя и не точно одинаков, но достаточно схож: 2-4-6-5. Сейчас в модуле *lexicog* предложена возможность выразить порядок слов в статье только в рамках исходного языка в случае вложенных статей и подсмыслов. Это может быть важно при кодировании толковых словарей, где есть четкое разделение на смыслы внутри леммы. Однако регламента по описанию порядка переводов на данный момент отсутствует: есть возможность использовать класс rdf:Seq и свойства rdf:\_1, rdf:\_2, и т.д. или конструировать собственные элементы онтологии, что, хотя и решит проблему на локальном уровне проекта, может усложнить привязывание словарей к другим членам «Semantic Web». В анализе влияний между словарями порядок переводов может сыграть немалую роль, и потерять эти данные при трансформации в RDF не хотелось бы. Решение этой проблемы может стать одним из направлений будущей доработки модуля lexicog.

Интересно отсутствие перевода у слова «*глꙋбокость*» - опираясь на близость слова к польскому эквиваленту и порядок латинских переводов у Кнапского, было бы логичным провести связь со словом *Profundum*. Этот выбор Епифания может служить контраргументом в разговоре о влиянии польского словаря. Стоит заметить, однако, что слово *Profundum* в доступном нам фрагменте ГСЛ не встречается вовсе.

Похожие закономерности можно проследить и при сравнении двух других польских лемм с их эквивалентами из ГСЛ: у слова Głeboko оба греческих перевода есть в ГСЛ (Рис. 14), у слова *Głeboki* есть один перевод как минимум (Рис. 15).

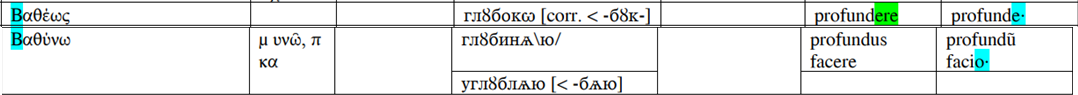
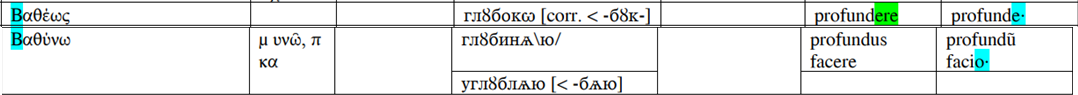
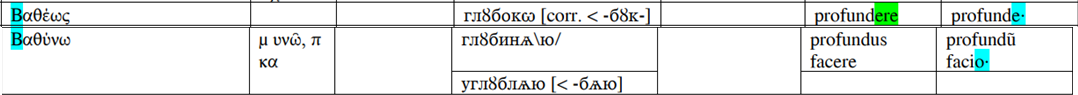
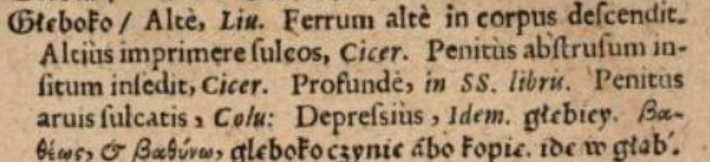
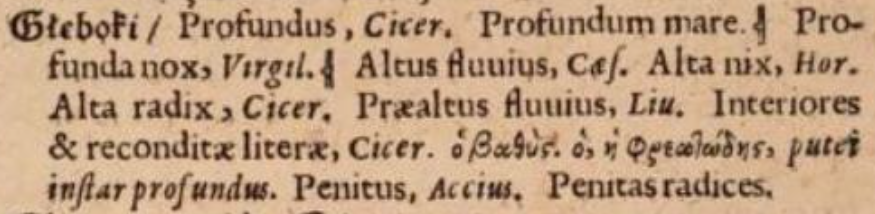
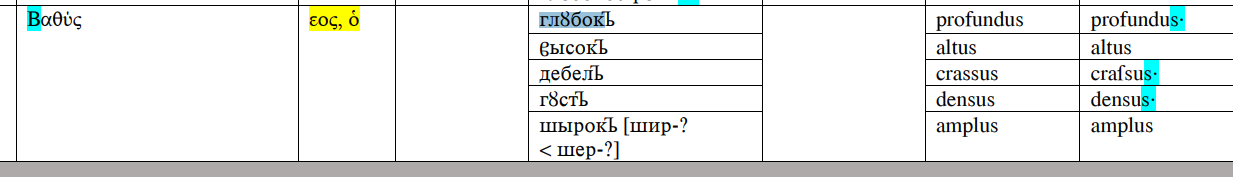
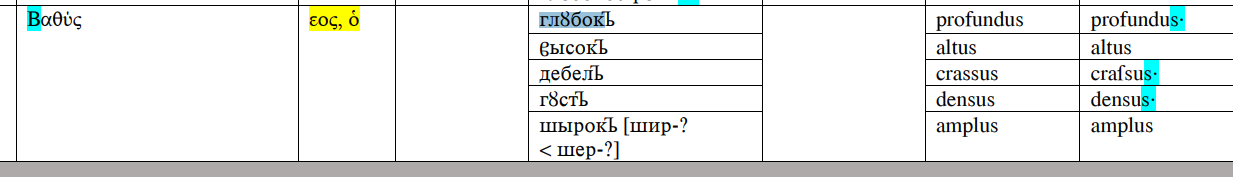
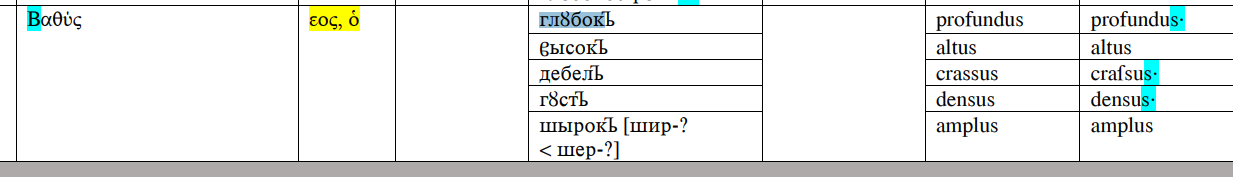


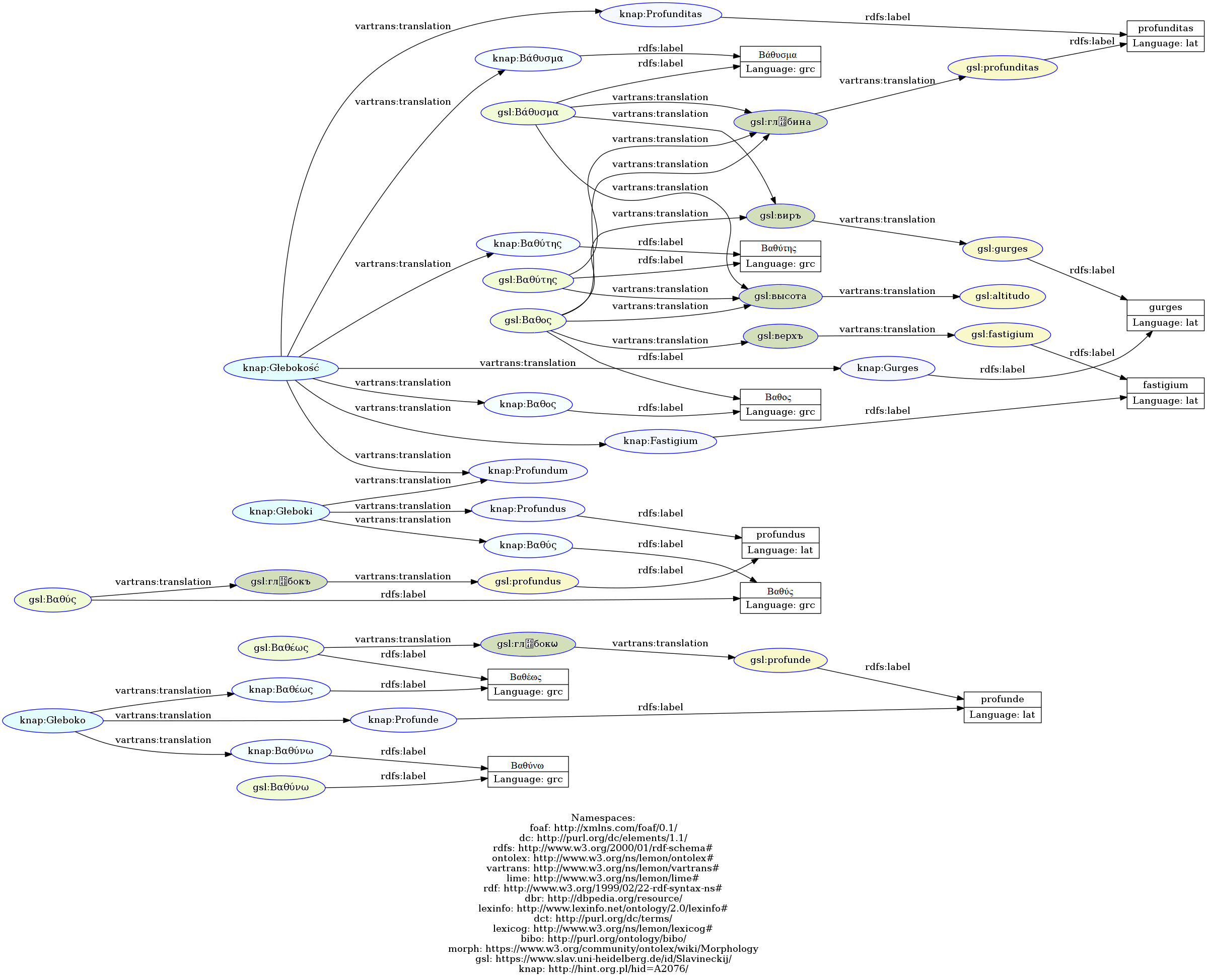
Рис. 14. Слово Głeboko и греческий эквивалент из ГСЛ

Рис. 15. Слово Głeboki и греческий эквивалент из ГСЛ



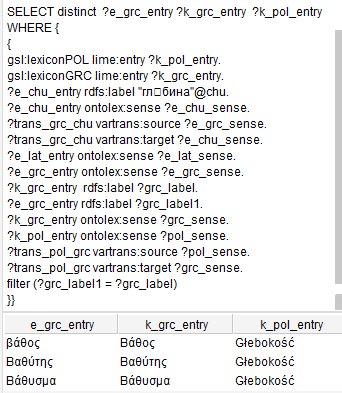
Лучше всего взаимосвязь двух словарей и четырех языков видно на графе. На Рисунке 16 представлена сеть лемм, связанных с польским словом *Głeboko* (см. также Рис. 14). Для наглядности мы упростили граф, который более не отображает RDF- модель в полной мере, но позволяет четко проследить путь от словаря Кнапского (голубые оттенки) к словарю Славинецкого (зеленые) через свойство rdfs:label. С одной стороны, словари разделяют греческие леммы *Βαθέως* и *Βαθύνω*, с другой – латинский перевод *Profunde*.

Рис. 16. Упрощенный RDF граф связи ПЛГ и ГСЛ

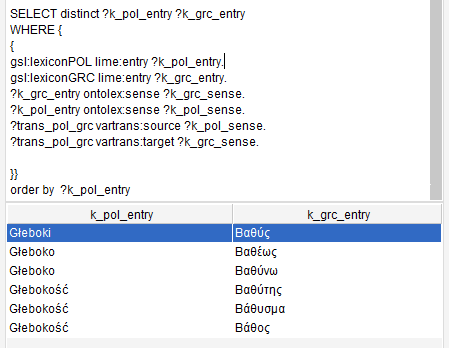


Имея в распряжении два словаря, размеченных языком Turtle, можно проводить анализ с помощью SPARQL-запросов. Так как сейчас у размечены нас лишь три леммы из словаря Кнапского, мы ограничимся тестированием и простыми сравнениями.

Запрос 6 выдает греческие переводы в рамках словаря Кнапского: один для *Głeboki*, два для *Głeboko* и три для *Głebokość* (во избежание ошибок при написании мы кодировали только те слова, которые есть в ГСЛ, однако в самом запросе этого фильтра нет).



Запрос 7.



Запрос 6.

Запрос 7 выдает конкретный случай перевода церковнославянского слова «*глꙋбина*», которое является эквивалентов трем древнегреческим леммам у Епифания, которые, в свою очередь, служат переводами польского слова *Głebokość*.

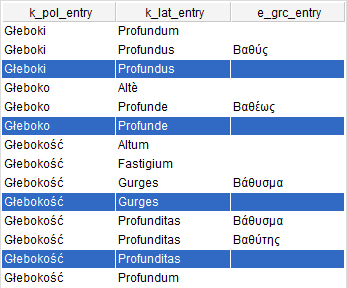
Запрос 8 выводит список польских слов, их переводов на латынь у Кнапского и связи с Епифанием через rdfs:label латинских слов. Если опустить выделенные синим строки-дубликаты, результат показывает, как польские слова разделяют с греческими латинские эквиваленты: порой у одного греческого слова два латинских эквивалента (*Βάθυσμα -> Profunditas, Gurges*), порой – у одного латинского слова два древнегреческих эквивалента (*Profunditas* -> *Βάθυσμα*, *Βαθύτης*).

select distinct ?k\_pol\_entry ?k\_lat\_entry ?e\_grc\_entry

WHERE {

{gsl:lexiconPOL lime:entry ?k\_pol\_entry.

Запрос 8.



?k\_pol\_entry ontolex:sense ?pol\_sense.

gsl:lexiconLAT lime:entry ?k\_lat\_entry.

?lcomp lexicog:describes ?e\_grc\_entry.

?e\_grc\_entry ontolex:sense ?e\_grc\_sense.

?trans\_grc\_chu vartrans:source ?e\_grc\_sense.

?trans\_grc\_chu vartrans:target ?e\_chu\_sense.

?trans\_chu\_lat vartrans:source ?e\_chu\_sense.

?trans\_chu\_lat vartrans:target ?e\_lat\_sense.

?e\_lat\_entry ontolex:sense ?e\_lat\_sense.

?k\_lat\_entry rdfs:label ?lat\_label1.

?e\_lat\_entry rdfs:label ?lat\_label.

?trans\_pol\_lat vartrans:source ?pol\_sense.

?trans\_pol\_lat vartrans:target ?lat\_sense.

?k\_lat\_entry ontolex:sense ?lat\_sense.

filter (?lat\_label1 = ?lat\_label)}

union

{gsl:lexiconPOL lime:entry ?k\_pol\_entry.

gsl:lexiconLAT lime:entry ?k\_lat\_entry.

?k\_pol\_entry ontolex:sense ?pol\_sense.

?trans\_pol\_grc vartrans:source ?pol\_sense.

?trans\_pol\_grc vartrans:target ?lat\_sense.

?k\_lat\_entry ontolex:sense ?lat\_sense.}}

order by ?k\_pol\_entry ?k\_lat\_entry

Итак, на примере этих графов и SPARQL-запросов можно видеть параллели между двумя словарями. Однако на основе такого пробного анализа можно лишь рассуждать о корреляции – чтобы доказать каузацию, потребуется больший объем словарных данных и отдельное исследование с использованием продвинутых SPARQL-запросов и анализа графов (например, в Gephi).

# Заключение

В рамках данного проекта была продолжена работа по представлению Греко-Славяно-Латинского словаря Епифания в цифровом формате.

Python-скрипт по трансформации словаря из формата .csv был оформлен в качестве библиотеки и выложен в открытый доступ.

Мы разработали модель репрезентации словаря в формате RDF, развиваясь в направлении, предложенном разработчиками онтологии OntoLex *lemon* на собственном опыте. Часть словаря была конвертирована в текстовую сериализацию RDF Turtle, что позволило протестировать предложенную логику и структуру при помощи SPARQL-запросов, а также визуализировать связи в виде графа. Успешное тестирование показало гибкость модели *lemon* и существующих (*lexicog*) и разрабатываемых (*morph*) модулей: словарь может быть конвертирован в формат RDF без потери данных.

Так как целью конвертации словарей в RDF является не только обеспечение связи между словами внутри словаря, но и создание единой сети документов и понятий, объединяющей разные ресурсы в рамках Linked Data, была предпринята попытка связать словарь Епифания с одним из возможных источников – Польско-Латино-греческим тезаурусом польского филолога Г. Кнапского. Тезаурус был также конвертирован в Turtle по предложенной нами модели. Установленная через графическое представление (rdfs:label) связь с ГСЛ была продемонстрирована при помощи визуализации и через SPARQL. В процессе были выявлены некоторые пробелы в существующих онтологиях (порядок переводов).

Кроме привязки словаря к словарю, очень перспективна для исследования связка словарь-текст (манускрипт). Дальнейшая работа может быть направлена на сравнение ГСЛ с манускриптами, из которых Епифаний мог выбирать греческие леммы для своего словаря, а также с переводными текстами современников и последователей монаха, которые могли пользоваться его словарем в своей работе.

Таким образом, конвертация текстового наследия в машиночитаемый формат не только позволяет сохранить данные, но и открывает широкий горизонт возможностей для будущих исследований: интертекстуальных, межъязыковых, межкультурных.

*Репозиторий проекта:* [*https://github.com/wildmary/TEI-RDF-encoded-Dictionary*](https://github.com/wildmary/TEI-RDF-encoded-Dictionary)

*Страница библиотеки на Pypi:* [*https://pypi.org/project/xmlexicon/*](https://pypi.org/project/xmlexicon/)

*Словарь на сайте Lexonomy:* [*https://www.lexonomy.eu/#/gsldict*](https://www.lexonomy.eu/#/gsldict)

**Литература и ресурсы**

1. Брайловский С. Н. Филологические труды Епифания (Славинецкого) //Русский филологический вестник. 1890. Т. 24. №2. С. 241-250. Цит. по: Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 42.
2. Левченко М.В. Электронное представление славянских словарей 17 века. Курсовая работа. Москва, 2021.
3. Литература и культура Древней Руси: Словарь-справочник / О.М. Анисимова, В.В. Кусков, М.П. Одесский, П.В. Пятнов. Под ред. В.В. Кускова. М.: Высш. шк., 1994. С. 38.
4. Николаев, С. И.: Словарь Г. Кнапского в России (Библиографические ра¬зы¬скания) // XVIII век. Вып. 20. Л., 1996. С. 160–161.
5. Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 41.
6. Пентковская, Т. В.: Новый Завет в переводе книжного круга Епифания Славинецкого и польская переводческая традиция XVI в.: Перевод аргументов к Апостолу // Русский язык в научном освещении № 1 (31), 2016. С. 184–229.
7. Фонкич, Б. Л.: Греческие рукописи и документы в России в XIV – начале XVIII в. (= Россия и Христианский Восток. Библиотека. Вып. 4). М., 2003. С. 281.
8. Электронный обменный формат для словарей проекта TEI (Text Encoding Initiative): Учебное пособие. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 80 с.
9. 9 Dictionaries // P5: Guidelines for Electronic Text Encoding and Interchange: https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html
10. Bosque-Gil, J., Lonke, D., Gracia, J., Kernerman, I. 2019. Validating the OntoLex-lemon lexicography module with K Dictionaries' multilingual data. Electronic Lexicography in the 21st Century. Proceedings of the Elex 2019 Conference. Pp. 726–746 (URL: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex\_2019\_41.pdf, дата обращения 15.04.2022)
11. Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., and Aguado-de-Cea, G. 2016. Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case. Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology. Pp. 65-72. (URL: http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-GLOBALEX\_Proceedings-v2.pdf, дата обращения 16.04.2022)
12. Declerck, T. and Mörth, K. 2016. “Towards a Sense-based Access to Related Online Lexical Resources”. In: Proceedings of the 17th EURALEX International Congress. Ed. by George Meladze Tinatin Margalitadze. Tbilisi, Georgia: Ivane Javakhishvili Tbilisi University Press. Pp. 660–667. (URL: https://www.dfki.de/fileadmin/user\_upload/import/8732\_Proceedings\_EURALEX2016\_Declerck\_Moerth\_104\_Final.pdf, дата обращения 26.04.2022)
13. G. Knapiusz (Knapski; Gregorius Cnapius), Thesaurus polono-latino-graecus seu promptuarium lingua Latinae et Graece [...], F. Caesario, Cracoviae [Kraków] 1643. Т. 1. С. 161. (URL: https://books.google.pl/books?id=1AQ-AAAAcAAJ&printsec=frontcover&hl=ru#v=snippet&q=profundum&f=false, дата обращения 30.05.2022). См. также http://hint.org.pl/hid=A2076
14. Gracia, Jorge & Villegas, Marta & Gomez-Perez, Asuncion & Bel, Nuria. 2017. The Apertium bilingual dictionaries on the web of data. Semantic Web. 9. Pp. 1-10. (URL: http://www.semantic-web-journal.net/system/files/swj1419.pdf)
15. Klimek, D., McCrae, J.P., Ionov, M., Tauber, J.K., Chiarcos, C., Bosque-Gil, J. and Buitelaar, P.. 2019. Challenges for the Representations for Morphology in Ontology Lexicons. Proceedings of Sixth Biennial Conference on Electronic Lexicography, eLex 2019. (URL: https://elex.link/elex2019/wp-content/uploads/2019/09/eLex\_2019\_33.pdf, дата обращения 19.05.2022)
16. Mambrini, F., Passarotti, M. Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin. In Proceedings of the 2020 Globalex Workshop on Linked Lexicography, pages 20–28, Marseille, France. European Language Resources Association.
17. McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The OntoLex-Lemon Model: development and applications. In Proceedings of eLex 2017, pages 587–597. (URL: (http://john.mccr.ae/papers/mccrae2017ontolex.pdf , дата обращения 02.05.2022)
18. Rambousek, A., Jakubíček M., Kosem, I. 2021. New developments in Lexonomy. In Proceedings of eLex 2021. (URL: https://www.sketchengine.eu/wp-content/uploads/2021-New-developments-in-Lexonomy.pdf, дата обращения 15.04.2022)
19. Sérasset, G. 2015. DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF. Semantic Web. 6. 355-361. Pp. 355-361. (URL: https://www.researchgate.net/publication/277903702\_DBnary\_Wiktionary\_as\_a\_Lemon-Based\_Multilingual\_Lexical\_Resource\_in\_RDF, дата обращения 15.04.2022)
20. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). Pp. 269-285.
21. Tittel, S., & Sabel, H. B. 2018. Using RDFa to Link Text and Dictionary Data for Medieval French. Proceedings of the 6th Workshop on Linked Data in Linguistics (LDL-2018), Miyazaki, Japan. (URL: https://www.deaf-page.de/pdf/SabineTittel\_rdfa.pdf, дата обращения 26.04.2022)
22. Toma Tasovac, Laurent Romary et al. 2018. TEI Lex-0: A baseline encoding for lexicographic data.. DARIAH Working Group on Lexical Resources: https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html
23. НКРЯ, исторический церковнославянский подкорпус: https://ruscorpora.ru/new/search-orthlib.html
24. Bibo: https://www.dublincore.org/specifications/bibo/bibo/
25. ELEXIS : https://elex.is
26. K Dictionaries: https://lexicala.com/k-dictionaries/
27. Lexinfo: https://lexinfo.net
28. Lexonomy: https://www.lexonomy.eu
29. pandas: https://pandas.pydata.org
30. Protégé: https://protege.stanford.edu
31. RDF converter: https://issemantic.net/rdf-converter
32. Rdf Grapher: https://www.ldf.fi/service/rdf-grapher
33. Text Encoding Initiative: https://tei-c.org
34. Transkribus: https://readcoop.eu/transkribus
35. W3C: https://www.w3.org/

1. Ирина Александровна Подтергера , канд. филол. наук: <https://www.slav.uni-heidelberg.de/personal/ipodtergera.html> [↑](#footnote-ref-1)
2. Мария Олеговна Новак, доктор филол. наук: <http://www.ruslang.ru/node/1736> [↑](#footnote-ref-2)
3. Выдержка из нашей курсовой работы Левченко М.В. Электронное представление славянских словарей 17 века. 2021. [↑](#footnote-ref-3)
4. Transkribus: <https://readcoop.eu/transkribus> [↑](#footnote-ref-4)
5. НКРЯ, исторический церковнославянский подкорпус: <https://ruscorpora.ru/new/search-orthlib.html> [↑](#footnote-ref-5)
6. Литература и культура Древней Руси: Словарь-справочник / О.М. Анисимова, В.В. Кусков, М.П. Одесский, П.В. Пятнов. Под ред. В.В. Кускова. М.: Высш. шк., 1994. С. 38. [↑](#footnote-ref-6)
7. Фонкич, Б. Л.: Греческие рукописи и документы в России в XIV – начале XVIII в. (= Россия и Христианский Восток. Библиотека. Вып. 4). М., 2003. С. 281. [↑](#footnote-ref-7)
8. Брайловский С. Н. Филологические труды Епифания (Славинецкого) //Русский филологический вестник. 1890. Т. 24. №2. С. 241-250. Цит. по: Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 42. [↑](#footnote-ref-8)
9. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 271. [↑](#footnote-ref-9)
10. ГИМ, Москва, Син. гр. №383 (по каталогу архим. Владимира №488). [↑](#footnote-ref-10)
11. Пентковская Т.В. Адаптирующие глоссы в поздних церковнославянских переводах с греческого // Вестник Московского университета. Серия 9. Филология. 2016. №1. С. 41. [↑](#footnote-ref-11)
12. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 275. [↑](#footnote-ref-12)
13. Пентковская, Т. В.: Новый Завет в переводе книжного круга Епифания Славинецкого и польская переводческая традиция XVI в.: Перевод аргументов к Апостолу // Русский язык в научном освещении № 1 (31), 2016. С.  184–229. [↑](#footnote-ref-13)
14. 9 Dictionaries // P5: Guidelines for Electronic Text Encoding and Interchange: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DI.html> [↑](#footnote-ref-14)
15. Toma Tasovac, Laurent Romary et al. 2018. TEI Lex-0: A baseline encoding for lexicographic data.. DARIAH Working Group on Lexical Resources: <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html> [↑](#footnote-ref-15)
16. Электронный обменный формат для словарей проекта TEI (Text Encoding Initiative): Учебное пособие. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 80 с. [↑](#footnote-ref-16)
17. Lexonomy: <https://www.lexonomy.eu> [↑](#footnote-ref-17)
18. Rambousek, A., Jakubíček M., Kosem, I. 2021. New developments in Lexonomy. In Proceedings of eLex 2021. (URL: <https://www.sketchengine.eu/wp-content/uploads/2021-New-developments-in-Lexonomy.pdf>, дата обращения 15.04.2022) [↑](#footnote-ref-18)
19. <https://www.w3.org/RDF/> [↑](#footnote-ref-19)
20. <https://www.w3.org/TR/rdf-schema/> [↑](#footnote-ref-20)
21. <https://www.w3.org/2016/05/ontolex/> [↑](#footnote-ref-21)
22. McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The OntoLex-Lemon Model: development and applications. In Proceedings of eLex 2017. P. 589. (URL: (<http://john.mccr.ae/papers/mccrae2017ontolex.pdf> , дата обращения 02.05.2022) [↑](#footnote-ref-22)
23. Mambrini, F., Passarotti, M. [Representing Etymology in the LiLa Knowledge Base of Linguistic Resources for Latin](https://aclanthology.org/2020.globalex-1.3). In *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, pages 20–28, Marseille, France. European Language Resources Association. [↑](#footnote-ref-23)
24. Serasset, G. 2012. Dbnary: Wiktionary as an LMF based Multilingual RDF network. In: Proceedings of the 8th Language Resources and Evaluation Conference, Istanbul. (URL: <https://www.researchgate.net/publication/225081670_Dbnary_Wiktionary_as_a_LMF_based_Multilingual_RDF_network/citations>, дата обращения 24.04.2022) [↑](#footnote-ref-24)
25. <https://lexicala.com/k-dictionaries/> [↑](#footnote-ref-25)
26. <https://www.w3.org/2016/05/ontolex/#core> [↑](#footnote-ref-26)
27. **Bosque-Gil, J., Lonke, D., Gracia, J., and Kernerman, I. 2019.** Validating the OntoLexlemon Lexicography Module with K Dictionaries’ Multilingual Data. [Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference: 726-746.](https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_41.pdf) [↑](#footnote-ref-27)
28. <https://www.w3.org/2019/09/lexicog/> [↑](#footnote-ref-28)
29. <https://www.w3.org/community/ontolex/wiki/Morphology#Update_Telco_06.04.2022> [↑](#footnote-ref-29)
30. Klimek, В., McCrae, J.P., **Bosque-Gil, J.,** Ionov, M., Tauber, J.K., Chiarcos, C. 2019. Challenges for the Representation of Morphology in Ontology Lexicons. (URL: <https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_33.pdf>, дата обращения 24.04.2022) [↑](#footnote-ref-30)
31. <https://lexinfo.net> [↑](#footnote-ref-31)
32. <https://www.dublincore.org/specifications/bibo/bibo/> [↑](#footnote-ref-32)
33. McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. 2017. The OntoLex-Lemon Model: development and applications. In Proceedings of eLex 2017. P. 599. [↑](#footnote-ref-33)
34. Sérasset, G. 2015. DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF. Semantic Web. 6. 355-361. P. 358. (URL: <https://www.researchgate.net/publication/277903702_DBnary_Wiktionary_as_a_Lemon-Based_Multilingual_Lexical_Resource_in_RDF>, дата обращения 15.04.2022) [↑](#footnote-ref-34)
35. Bosque-Gil, J., Lonke, D., Gracia, J., Kernerman, I. 2019. Validating the OntoLex-lemon lexicography module with K Dictionaries' multilingual data. Electronic Lexicography in the 21st Century. Proceedings of the Elex 2019 Conference. P.728. (URL: <https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_41.pdf>, дата обращения 15.04.2022) [↑](#footnote-ref-35)
36. **Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., and Aguado-de-Cea, G. 2016.** Modelling multilingual lexicographic resources for the Web of Data: The K Dictionaries case. Proceedings of GLOBALEX 2016: Lexicographic Resources for Human Language Technology. P. 68. (URL: <http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-GLOBALEX_Proceedings-v2.pdf>, дата обращения 16.04.2022) [↑](#footnote-ref-36)
37. <https://www.w3.org/2019/09/lexicog/#lexicography-module-lexicog> [↑](#footnote-ref-37)
38. <http://linguistic-lod.org/llod-cloud> [↑](#footnote-ref-38)
39. Gracia, Jorge & Villegas, Marta & Gomez-Perez, Asuncion & Bel, Nuria. 2017. The Apertium bilingual dictionaries on the web of data. Semantic Web. 9. Pp. 1-10. (URL: <http://www.semantic-web-journal.net/system/files/swj1419.pdf>) [↑](#footnote-ref-39)
40. Declerck, T. and Mörth, K. 2016. “Towards a Sense-based Access to Related Online Lexical Resources”. In: Proceedings of the 17th EURALEX International Congress. Ed. by George Meladze Tinatin Margalitadze. Tbilisi, Georgia: Ivane Javakhishvili Tbilisi University Press. P. 662.(URL: <https://www.dfki.de/fileadmin/user_upload/import/8732_Proceedings_EURALEX2016_Declerck_Moerth_104_Final.pdf>, дата обращения 26.04.2022) [↑](#footnote-ref-40)
41. Там же. С. 666. [↑](#footnote-ref-41)
42. Tittel, S., & Sabel, H. B. 2018. Using RDFa to Link Text and Dictionary Data for Medieval French. Proceedings of the 6th Workshop on Linked Data in Linguistics (LDL-2018), Miyazaki, Japan. (URL: <https://www.deaf-page.de/pdf/SabineTittel_rdfa.pdf>, дата обращения 26.04.2022) [↑](#footnote-ref-42)
43. Там же. C. 4. [↑](#footnote-ref-43)
44. <http://www.deaf-page.de/guichaulmTel/edition.html> [↑](#footnote-ref-44)
45. <https://www.w3.org/2015/09/bpmlod-reports/multilingual-dictionaries/> [↑](#footnote-ref-45)
46. <https://www.w3.org/2015/09/bpmlod-reports/bilingual-dictionaries/> [↑](#footnote-ref-46)
47. <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/document/10-rules-persistent-uris> [↑](#footnote-ref-47)
48. Gracia, Jorge & Villegas, Marta & Gomez-Perez, Asuncion & Bel, Nuria. 2017. The Apertium bilingual dictionaries on the web of data. Semantic Web. 9. P. 4. (URL: <http://www.semantic-web-journal.net/system/files/swj1419.pdf>) [↑](#footnote-ref-48)
49. <https://www.dublincore.org/specifications/bibo/bibo/> [↑](#footnote-ref-49)
50. <https://www.w3.org/2016/05/ontolex/#lexicon-and-lexicon-metadata> [↑](#footnote-ref-50)
51. Этот и последующие графы сгенерированы при помощи ресурса RDF Grapher (<https://www.ldf.fi/service/rdf-grapher>) [↑](#footnote-ref-51)
52. Klimek, D., McCrae, J.P., Ionov, M., Tauber, J.K., Chiarcos, C., Bosque-Gil, J. and Buitelaar, P.. 2019. Challenges for the Representations for Morphology in Ontology Lexicons. Proceedings of Sixth Biennial Conference on Electronic Lexicography, eLex 2019. (URL: <https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_33.pdf>, дата обращения 19.05.2022) [↑](#footnote-ref-52)
53. <https://www.w3.org/community/ontolex/wiki/Morphology> [↑](#footnote-ref-53)
54. Мы использовали редактор онтологий Protégé для работы с запросами в SPARQL: <https://protege.stanford.edu> [↑](#footnote-ref-54)
55. Strakhov, O. B.: Jepyfanij Slavynecʹkyj’s Greek-Slavic-Latin Lexicon: The History, Contents, and Principles Underlying the Composition of Its Greek Portion (Preliminary Re-marks) // Harvard Ukrainian Studies 28.¼ (2006). P 271. [↑](#footnote-ref-55)
56. Там же. С. 282. [↑](#footnote-ref-56)
57. Николаев, С. И.: Словарь Г. Кнапского в России (Библиографические ра­зы­скания) // *XVIII век*. Вып. 20. Л., 1996. С. 160–161. [↑](#footnote-ref-57)
58. G. Knapiusz (Knapski; Gregorius Cnapius), Thesaurus polono-latino-graecus seu promptuarium lingua Latinae et Graece [...], F. Caesario, Cracoviae [Kraków] 1643. Т. 1. С. 161. (URL: <https://books.google.pl/books?id=1AQ-AAAAcAAJ&printsec=frontcover&hl=ru#v=snippet&q=profundum&f=false>, дата обращения 30.05.2022). См. также <http://hint.org.pl/hid=A2076> [↑](#footnote-ref-58)
59. <http://www.leksykografia.uw.edu.pl/slowniki/39/thesaurus-polono-latino-graecus-krakow-1621> [↑](#footnote-ref-59)