

Параллельный корпус как грамматическая база данных и Новый Завет как параллельный корпус (предисловие)

В. А. Плунгян

Институт языкознания РАН (Москва, Россия); plungian@gmail.com;

ORCID: 0000-0002-2393-1399

Аннотация. В настоящем предисловии (которое является одновременно теоретическим введением к последующим статьям) дается общая характеристика проекта подготовки базы данных типологически релевантных грамматических контекстов на основе параллельного корпуса переводов Нового Завета. Также приводится краткий обзор состава и содержания публикуемых далее статей.

Ключевые слова: базы данных, грамматические категории, Новый Завет, параллельные корпуса.

The parallel corpus as a grammar database and the New Testament as a parallel corpus (Preface)

Vladimir A. Plungian

Institute of Linguistics, Russian Academy of Sciences (Moscow, Russia);

plungian@gmail.com; ORCID: 0000-0002-2393-1399

Abstract. The present foreword, which is also a theoretical introduction to the subsequent contributions, provides a general overview of the project aimed at a database of cross-linguistically relevant grammatical contexts based on a parallel corpus of New Testament translations. In the final part, the reader will find a brief summary of the structure and content of the papers included in this issue.

Keywords: databases, grammatical categories, New Testament, parallel corpora.

... читаю посланья Апостоловъ я,
слова Псалмопѣвца читаю.

А. Ахматова (1915)

1. Введение

Собранные ниже статьи составляют 9-й выпуск серии «Исследования по теории грамматики» и отражают предварительные результаты работы над типологическим проектом, конечной целью которого является создание базы данных универсальных элементов грамматического набора на основе параллельного корпуса переводов Нового Завета. Замысел такого проекта активно обсуждался большинством из участников сборника в ходе нескольких семестровых семинаров (проводившихся в разном составе на ОТиПЛ МГУ и в ШЛ НИУ ВШЭ в течение 2020–2022 гг.), а также в рамках конференции «К построению базы данных универсальных грамматических контекстов на материале параллельного корпуса переводов Нового Завета», прошедшей в Москве 27–28 августа 2021 г. (в конференции также приняли участие сотрудники Института перевода Библии).

В качестве введения к публикуемым статьям мы хотели бы несколько подробнее прокомментировать сам этот замысел, а также охарактеризовать общую структуру и содержание статей, вошедших в настоящий выпуск.

Начнем с главной задачи, на решении которой в той или иной степени сосредоточены все публикуемые материалы: ее можно кратко обозначить как использование корпусных инструментов в качестве эвристики в исследованиях по типологии грамматических категорий. Эта задача имеет два важных аспекта: использование параллельных корпусов в грамматической типологии и использование переводов Нового Завета в качестве параллельного корпуса для лингвистических исследований. Рассмотрим их поочередно.

2. Параллельные корпуса в лингвистической типологии

То, что параллельные корпуса могут быть эффективным средством решения задач, связанных с типологическим описанием языков, было, вообще говоря, очевидно задолго до возникновения современных электронных ресурсов. Собственно, любое сопоставительное исследование так или иначе опиралось на переводы одного текста на различные языки, и тех или иных «предшественников» корпусных методов можно найти очень много. Пожалуй, стоит выделить два параметра, при учете которых такие исследования несколько более заметно сближаются с современным представлением о формате использования корпусных методов.

Во-первых, это использование не просто нескольких языков, как это принято в рамках стандартной «сопоставительной лингвистики», а опора на значительное количество типологически разнородных языков (и, соответственно, отбор таких текстов, которые существуют в переводах на как можно большее число естественных языков). Для типологии — особенно такой типологии, которая стремится к построению сбалансированных выборок и максимально полному охвату языкового разнообразия человечества, — именно этот тип данных подходит в наибольшей степени. Понятно, что подобная постановка задачи сильно сужает набор текстов, пригодных для исследования, что противопоставляет стратегию использования корпусов в лингвистической типологии и в традиционном контрастивном анализе, для которого годится любая языковая пара, состоящая из оригинала и перевода. В современной типологии такой тип корпусных данных обычно называется *massively parallel corpora* [Wälchli, Sölling 2013; Östling 2016]. Использование подобных инструментов сегодня становится всё более частым, хотя пока еще далеко не доминирует в типологическом арсенале; но стоит заметить, что само обсуждение таких возможностей получает совершенно новый импульс именно с появлением современных электронных ресурсов, т. е. приблизительно с начала XXI века [Johansson 2007; Aijmer 2008; Doval, Sánchez Nieto 2019].

Вторая важная особенность использования параллельных корпусов в соответствии с принципами функциональной типологии (важная в том числе и для той задачи, которая решается в статьях настоящего выпуска) — это ориентация на поиск универсальных семантических контекстов, по возможности не зависящих от конкретно-языковых особенностей. Как известно, современная функциональная типология в целом стремится опираться на универсальную семантическую субстанцию, исследуя закономерности выражения «одних и тех же» смыслов средствами различных конкретно-языковых лексических и грамматических систем (программа, намеченная еще в [Wierzbicka 1988]; ср. также [Мельчук 1998] и [Wiltshko 2014]). В грамматической типологии это дает инструменты типа «универсального грамматического набора» и правил «совмещения» (или «кластеризации») универсальных элементов в составе значений полисемичных конкретно-языковых грамматических показателей (подробнее см. [Плунгян 2011: 53–68]; ср. также [Frajzyngier, Mettouchi 2015]); в лексической типологии — правила «совмещения» (или «колексификации») универсальных лексических значений и/или универсальных фреймов ([Koch 2001; Рахилина, Плунгян 2007; François 2008; Кортјевскаја-Тамм et al. 2015; Östling 2016] и др.).

Соответственно, параллельный корпус в такой оптике понимается прежде всего как инструмент обнаружения универсальных контекстов (или фреймов), по-разному выражаемых в разных языках. Его наиболее распространенный конкурент в типологических исследованиях — типологическая **анкета**, в которой те же универсальные контексты задаются исследователем искусственно, обычно с использованием распространенного языка-посредника. Анкета может состоять из коротких предложений-стимулов, предназначенных для перевода на исследуемый язык, или более пространных фрагментов текста; исследуемое явление может быть эксплицитно обозначено средствами языка-стимула или оставлено без перевода — в последнем случае предполагается, что контекст однозначно указывает на нужный смысл [Dahl 1985, 2000]; реже в полевой лингвистике используются невербальные стимулы — изображения или видеоматериалы.

В истории типологических исследований грамматических категорий анкета по понятным причинам предшествовала параллельным корпусам (важную роль в грамматической типологии сыграли анкеты Э. Даля для изучения видовременных систем, разработанные в начале 1980-х гг., а также анкеты, предложенные в исследованиях Петербургской типологической школы с конца 1960-х гг. преимущественно для описания залоговых категорий глагола и позднее распространенные на некоторые значения из аспектуально-таксисной и модальной сферы и ряд других). К достоинствам анкеты относятся возможность точно задать нужный исследователю контекст и возможность непосредственно воздействовать на носителя языка; связанные с этим недостатки анкеты — опасность получить искаженные данные, не отражающие естественно-языковое поведение носителей. Кроме того, анкета хорошо работает в том случае, когда исследователь изначально хотя бы в общих чертах представляет себе то явление, которое он хочет обнаружить; как инструмент эвристического исследования при отсутствии априорных представлений о грамматических особенностях изучаемого языка она не годится.

Напротив, корпусные данные отражают способы вербализации нужного исследователю смысла, созданные говорящими независимо от вмешательства исследователя, и это повышает надежность и достоверность такого материала. Корпусные данные существенно лучше отражают естественное функционирование языковых единиц, хотя следует помнить, что и они отражают его не полностью. Как известно, основные проблемы, связанные с использованием параллельных корпусов, заключаются в качестве присутствующих в них переводов; в целом переводной текст не эквивалентен спонтанно порождаемому тексту, а находится под (более или менее) сильным влиянием текста на языке-оригинале. Поэтому оптимальной является ситуация, когда исследователь может сочетать все доступные методы, используя преимущества каждого из них и нивелируя недостатки [Wälchli 2007; Dahl 2007].

Таким образом, желательная ситуация использования параллельного корпуса для типологических исследований грамматики (будем ниже говорить именно о типологии грамматических значений) предполагает, что это корпус:

- многоязычный (и, по возможности, содержащий типологически сбалансированную выборку языков);
- включающий переводы оригинального текста высокого качества (возможно, несколько переводов одного и того же текста — так называемый **поливариантный** корпус, ср. [Aijmer 2008; Сичинава 2015; Нестеренко 2019] и др.);
- включающий элементы **базы данных** универсальных грамматических значений: в корпусе имеется дополнительная разметка контекстов, в которых целесообразно ожидать появления показателей тех или иных грамматических значений (если в данном языке соответствующее значение грамматикализовано).

Последний пункт наименее тривиален: фактически он означает, что для эффективного использования параллельного корпуса в исследованиях по грамматической типологии необходима специальная и достаточно трудоемкая разметка этого корпуса, т. е. предварительный анализ и обработка всех входящих в него текстов, с указанием, например, прототипических контекстов для выражения различных значений детерминации, дейксиса, числа, основных семантических ролей аргументов, аспекта, времени, модальности, эвиденциальности, отрицания, вопроса и т. д. и т. п. В полном объеме такой разметки не только не было до сих пор сделано, но и задачи такой, насколько нам известно, по отношению к существующим параллельным корпусам не ставилось. Именно поэтому начальным стимулом для наших исследований было стремление получить хотя бы предварительный ответ на вопрос, насколько такая разметка вообще возможна, и реалистично ли построить базу данных с контекстами хотя бы для некоторых участков универсального грамматического набора. (Скажем сразу, немного забегаая вперед: мы считаем, что на этот вопрос был получен положительный ответ, и публикуемые статьи отчасти служат этому подтверждением.)

Несмотря на отмеченные выше недостатки и ограничения, эффективность использования параллельных корпусов была осознана исследователями достаточно давно, по крайней мере еще

в докомпьютерную эпоху. Не претендуя на исчерпывающий обзор, отметим здесь два ранних опыта такого рода, принадлежащих к различным лингвистическим школам, в которых предлагались идеи, весьма близкие представленным в настоящем выпуске.

Один из самых ранних известных нам опытов продуктивного использования параллельного корпуса для типологического исследования семантики грамматической категории — это работа по «функциональной типологии перфекта» воронежского лингвиста Вячеслава Борисовича Кашкина (1955–2014), написанная под руководством Ю. С. Маслова ([Кашкин 1991]; в дальнейшем В. Б. Кашкин развивал этот опыт и в ряде других исследований) и посвященная сопоставительному анализу категории перфекта в европейских языках¹. В качестве источника материала были взяты прозаические тексты русских классиков в переводе на основные европейские языки (романские и германские); задача тем более примечательная, что в русском языке грамматической категории перфекта, как известно, нет. Исходной единицей анализа служила так называемая «перфектная ситуация», т. е. контекст, семантически способствующий употреблению перфекта в тех языках, где он является грамматикализованным. Тем самым эта работа оказалась существенно смещена от традиционного контрастивного анализа («от формы к форме») к типологическому исследованию способов выражения некоторого универсального значения («от смысла к формам»); именно в этом аспекте следует понимать слова «функциональная типология» в названии работы. Учитывая, что исследование проводилось на материале параллельного корпуса художественных текстов, есть все основания считать его одним из первых примеров применения параллельных корпусов как инструмента работы с универсальными грамматическими элементами для решения типологических задач.

¹ Близкие по постановке задачи, но более масштабные исследования были выполнены современными типологами лишь примерно через четверть века [Сичинава 2016; Dahl 2014, 2022] и проводились уже на материале «настоящих» электронных параллельных корпусов; следует отметить, что они подтвердили справедливость многих наблюдений, сделанных в ранней работе В. Б. Кашкина.

Другая, гораздо более известная ранняя попытка использования корпусных инструментов в типологии *avant la lettre* — это монография [Haspelmath 1997]. Узкая задача этой работы — проверка утверждения об универсальном характере пространственно-временной полисемии; для решения этой задачи автор выделяет элементарные наборы универсальных пространственных и временных значений и исследует способы выражения этих значений в большом количестве языков мира с помощью специально подобранных контекстов, взятых из параллельных переводов одного исходного текста. Крайне важно, что в качестве основного материала для такого сравнения автору служат тексты Нового Завета — и в этом он оказывается прямым предшественником настоящего проекта. Выбор именно этих текстов, в принципе, достаточно очевиден для масштабной типологической задачи (хотя, как мы увидим ниже, встречались и другие решения), но к подробному обсуждению мотивов такого выбора мы обратимся в следующем разделе.

Конечно, с появлением электронных корпусов (в том числе и представительных коллекций параллельных текстов) количество типологических исследований с их применением начинает расти, но число работ, удовлетворяющих всем приведенным выше критериям (многоязычные параллельные корпуса с элементами типологической разметки) по-прежнему остается невелико² — как и не наблюдается большого числа попыток создать сами такие параллельные корпуса, с которыми типологу было бы удобно работать. До сих пор в этом направлении имеется всего несколько значимых опытов (обзор их см., например, в [Aijmer 2008; Āermák, Rosen 2012] и [Нестеренко 2019]: ведь далеко не всякое сопоставительное исследование двух или нескольких языков, даже выполненное с использованием параллельного корпуса, оказывается значимым вкладом в грамматическую типологию как таковую. В этом отношении этапной является публикация [Cysouw, Wälchli (eds.) 2007]: этот специальный выпуск журнала *Language Typology and Universals* целиком посвящен

² Из серьезных попыток такого рода стоит упомянуть, например, работы [Wälchli 2010] или [Bogaards 2022]; о других исследованиях см. также ниже.

проблеме использования параллельных корпусов именно для масштабных теоретических исследований функционирования грамматических категорий в языках мира. Существенно, что целый ряд статей в этом выпуске посвящен использованию параллельных корпусов на основе переводов Библии, и мы коснемся этих работ ниже. По-видимому, появление данной публикации и можно считать начальной точкой возникновения корпусной грамматической типологии как отдельной дисциплины со своими задачами и своим специфическим набором инструментов и методов.

3. Новый Завет как параллельный корпус

Естественный вопрос, возникающий непосредственно вслед за осознанием необходимости использования параллельных корпусов для исследования типологии грамматических значений — это вопрос о выборе наиболее подходящей для этой цели коллекции параллельных текстов. Возможности выбора, однако, здесь совсем невелики: типологам требуется такой текст, который был бы удовлетворительным образом переведен на очень большое число разнородных языков, в том числе и на такие, по которым затруднительно найти какие-либо иные данные. Таких текстов практически нет. Наиболее переводимые в мире тексты — такие, как тексты известных политиков или популярные художественные произведения европейских авторов (будь то Агата Кристи, Жюль Верн, Антуан де Сент-Экзюпери, Джоан Роулинг и т. п.) — переводятся всё же в основном на крупные языки, имеющие развитую издательскую инфраструктуру; их почти не переводят на малые языки с ограниченными сферами применения и тем более не переводят по понятным причинам на бесписьменные языки; между тем именно языки последних двух групп в наибольшей степени интересуют типологов. Единственный текст, который хотя бы немного приближается по своим свойствам к этим критериям — это, как легко видеть, текст Библии (и даже, если быть более точным, то именно Нового Завета).

Действительно, Новый Завет не имеет равных по количеству языков, на которые он был переведен, далеко опережая все остальные произведения: по данным Всемирного альянса им. Уиклифа (<https://www.wycliffe.net/resources/scripture-access-statistics>), к 2019 году Новый Завет был переведен на 2246 языков полностью и еще на 1138 — частично (итого 3384 языка, что составляет примерно половину от имеющихся в настоящее время на Земле живых языков³); более того — перевод Библии осознается как насущная миссионерская задача во многих христианских деноминациях, и эта деятельность активно продолжается (в том числе силами таких организаций, как Институт перевода Библии, Летний лингвистический институт, и рядом других). Переводы Библии свободно распространяются и доступны в интернете; несмотря на то, что в настоящее время не существует удобного специализированного ресурса по использованию параллельных переводов Нового Завета именно для типологических исследований языков⁴, использовать эти тексты для пилотных исследований достаточно легко — тем более легко, что эти тексты изначально предстают как уже «выровненные» (Новый Завет разделен на главы, главы — на пронумерованные стихи, и переводчики обычно сохраняют эту структуру). Греческая лексика Нового Завета хорошо изучена, и для работы с лексическими единицами существует так называемый конкорданс Стронга, облегчающий многие виды поиска [Cysouw et al. 2007]. Грамматические значения (тем более элементы универсального грамматического набора) пока таким образом

³ В отличие от подавляющего большинства других текстов, Библия стала переводиться уже в начале I тысячелетия н. э., и имеются ее переводы на мертвые языки — в том числе и на такие, для которых это сегодня практически единственный или основной письменный источник, дающий возможность их изучения (готский, старославянский, коптский).

⁴ Необходимость создания специального ресурса на основе переводов Библии для лингвистических исследований давно и широко обсуждалась (ср. [de Vries 2007; Christodouloupoulos, Steedman 2015] и др.), но, насколько нам известно, в открытом доступе на сегодняшний день такого инструмента нет; ряд коллективов (в том числе группа исследователей из НИУ ВШЭ совместно с ИПБ) ведут работу по его созданию.

не размечены, и это задача для будущих исследований. Именно подступы к ее решению и возможные варианты такой разметки и обсуждаются в статьях данного выпуска.

Идея использования переводов Библии для изучения языков сама по себе далеко не новая: как кажется, одной из первых попыток такого рода в европейской науке был *Synopsis universae philologiae* Готфрида Хензеля (1741), в котором на основе переводов «Отче наш» на все известные в то время языки была составлена языковая карта мира и приводились (соответствующие уровню знаний XVIII века) рассуждения о классификации языков человечества. В современной типологии, помимо уже указанной работы М. Хаспельмата, можно отметить такие исследования, как [Барентсен 2008; Wälchli, Cysouw 2012; Wälchli, Sölling 2013; Dahl 2014, 2022] и др. Отдельно следует отметить проект по созданию корпуса PROIEL [Eckhoff, Naug 2009; Naug et al. 2009], в котором предлагается тщательно выполненная морфосинтаксическая разметка древних переводов Нового Завета (об их важности для лингвистических исследований уже говорилось выше).

Разумеется, типологи активно используют и другие виды параллельных корпусов: при ограничении рамок исследования, например, только европейскими языками возможности здесь существенно расширяются. Есть опыты использования популярных художественных произведений (помимо упомянутой в самом начале работы В. Б. Кашкина, ср. например, [van der Auwera et al. 2005] или [Stolz 2007]; коллекции параллельных текстов А. Барентсена и Т. Штольца в этой области наиболее известны); существуют типологические работы, написанные на материале таких относительно больших и в то же время легко доступных коллекций, как корпус документов Европарламента [de Swart et al. 2022] или субтитры к фильмам [Levshina 2015]. Всё это популярные в корпусной лингвистике форматы, но для массовых типологических исследований по перечисленным выше причинам Новый Завет в переводах на современные языки остается практически единственным вариантом. Заметим, что это и в принципе хорошо подходящий для сравнительных исследований грамматики текст: он достаточно большой по объему и разнообразный по содержанию,

в нем встречаются как образцы диалогов (в которых взаимодействуют люди разного возраста, пола и социального статуса), так и классические нарративные фрагменты с чередованием основной линии повествования и комментариев; в тексте описываются как наблюдаемые говорящими события, так и те, о которых им известно не из личного опыта: существовавшие в прошлом, возможные в будущем, а также воображаемые или ирреальные. Всё это может быть важно для описания базовых грамматических категорий языков мира.

4. О настоящем выпуске

Как уже было сказано, в настоящем выпуске собраны статьи, нацеленные на решение следующих задач: на основе некоторой базовой коллекции текстов параллельных переводов Нового Завета (о ее структуре см. ниже) выявить прототипические контексты, в которых наиболее вероятно появление показателей, выражающих определенные элементы универсального грамматического набора. Иными словами, если исследователь захочет выяснить, выражается ли в изучаемом им языке *L* грамматическое значение *M*, то ему будет предложен набор контекстов Нового Завета, где появление данного значения в переводах на *L* высоко вероятно. В идеале тексты Нового Завета могли бы быть размечены таким образом максимально полно, с учетом всех поддающихся контекстной фиксации универсальных грамматических элементов⁵. Но эта задача, разумеется, существует только как очень далекий конечный этап в построении будущей базы

⁵ Очевидным образом, не все существующие в языках мира грамматические категории имеет смысл обнаруживать таким образом; так, вряд ли возможно предложить прототипические контексты для описания согласовательных категорий или, например, категорий, связанных с выражением залоговых отношений. Наилучшим образом фиксируются так называемые «ингерентные» (или «семантически наполненные») категории, модифицирующие прототипические имена и глаголы (число, детерминация, посессивность, аспект, отрицание и т. п.).

данных, а в представленных ниже статьях лишь обсуждаются возможные подходы к ее решению и, что важно, предлагаются наборы диагностических контекстов для отдельных универсальных элементов. Как правило, в статьях тестируются как эвристические процедуры обнаружения определенного значения, так и предварительные результаты классификации полученных контекстов.

Чтобы читатель лучше представлял себе, что мы имеем в виду, когда говорим о прототипических контекстах для обнаружения универсальных грамматических значений, приведем один достаточно простой пример. Возьмем отрывок из Лк. 2:41 в древнегреческом оригинале и для наглядности в латинском, испанском, английском (New King James Version) и русском Синодальном переводах:

ДРЕВНЕГРЕЧЕСКИЙ

- (1) а. Καὶ ἐπορεύοντο οἱ γονεῖς αὐτοῦ κατ' ἔτος εἰς Ἱερουσαλὴμ τῇ ἐορτῇ τοῦ πάσχα.

Kai eporeuonto hoi goneis autou
и идти.IPF.3PL DEF.M.PL родитель.NOM.PL 3SG.M.GEN
kat' etos eis Ierusalēm tē
каждый год.ACC.SG к Иерусалим.ACC.SG DEF.F.SG.DAT
heortē tou pascha.
праздник.DAT.SG DEF.N.SG.GEN Пасха.GEN.SG

ЛАТИНСКИЙ

- б. Et ibant parentes eius per omnes
и идти.IPF.3PL родитель.PL TOT.GEN.SG через весь.M.PL.ACC
annos in Ierusalem in die
год.PL.ACC в Иерусалим.ACC в день.ABL.SG
festo Paschae.
праздник.ABL.SG Пасха.GEN.SG

ИСПАНСКИЙ

- в. Iban sus padres todos los años
идти.IPF.3PL POSS.3PL родитель.PL весь.M.PL DEF.M.PL год.PL
a Jerusalén en la fiesta de la Pascua.
в Иерусалим в DEF.F.SG праздник GEN DEF.F.SG Пасха

АНГЛИЙСКИЙ

- г. His parents went to Jerusalem every year at the Feast of the Passover.

‘Каждый год родители Его ходили в Иерусалим на праздник Пасхи’. [Лк. 2:41]

Этот отрывок представляет собой простое предложение с единственной предикативной вершиной, а также субъектной и тремя обстоятельственными синтагмами. Обстоятельственные группы реализуют распространенные типы временных (‘период, включающий регулярно совершаемое повторяющееся действие’) и пространственных значений (‘цель движения’); группа субъекта представляет собой участника из двух лиц (с выраженным эксплицитно посессивным отношением родства). Наконец, предикативная вершина описывает регулярно повторяющееся в прошлом событие. Всё это очень яркие и однозначные контексты для поиска таких граммем, как, например, двойственное число, неотчуждаемая принадлежность, цель движения, хабитуальность в прошлом и т. п. Как можно видеть, в приведенных примерах переводов некоторые из этих значений действительно грамматикализованы: в древнегреческом и романских языках глагол употреблен в форме имперфекта (для которого хабитуальность — одно из основных значений, как и для английского Past Simple), и т. п.

Именно такие «прозрачные» и однозначные контексты предлагалось обнаружить участникам проекта для выбранных ими категорий. Для исследования была создана несложная инфраструктура, состоящая из пронумерованных стихов и глав Нового Завета (главным образом использовались четыре Евангелия, Послания в корпус не включались, Откровение и Деяния привлекались лишь избирательно). Тексты привлекались в древнегреческом оригинале и в переводе на перечисленные выше четыре языка (латинский, русский, испанский, английский); это была так называемая базовая выборка. Ее состав продиктован скорее прагматическими соображениями: это наиболее распространенные переводы (с которых в новейшее время делались переводы на многие другие языки); при этом представленные языки по своим морфосинтаксическим свойствам достаточно

разнообразны и обеспечивают возможность с неплохим приближением очертить круг релевантных контекстов для более масштабного типологического исследования (в некоторых случаях контуры такого исследования намечаются уже в самой статье).

Выделенные контексты анализировались и подвергались классификации (например, исследователь, изучающий хабитуальность, мог выделить контексты для разных типов хабитуальных ситуаций, в каких-то языках формально противопоставленных). К обязательному набору базовых языков мог добавляться какой-то язык по выбору исследователя, где интересующее его явление специально представлено или выражено каким-то более сложным образом (например, для изучения контекстов двойственного числа естественным решением было бы добавление словенского). Но основным итогом публикуемых исследований считались именно списки диагностических контекстов с их классификацией.

На наш взгляд, представленные в этом выпуске статьи, при всей их разнородности и предварительном характере полученных результатов, убедительно демонстрируют плодотворность такого подхода к параллельным текстам; их можно рассматривать как своего рода задел для создания полноценной базы данных для типологического исследования грамматики.

Как можно заметить, в подборке примерно в равном количестве присутствуют работы, посвященные именным и глагольным категориям. Из именных категорий анализируются контексты для странственной роли латива (К. В. Филатов) и ситуаций контактной локализации (Д. А. Парамонова), двойственного числа (Д. Д. Врубель и Л. И. Паско), показателей универсальной квантификации (И. А. Хомченкова), темпоральных обстоятельств (П. С. Плешак), значений компаратива и экватива (Ю. В. Сеницына). Из глагольных категорий анализируются контексты для экспериментивного значения (Ю. Н. Кузнецова и А. Д. Подгорная), императива (В. А. Орлов), континуатива (А. Б. Панова), плюсквамперфекта (Д. В. Сичинава), сопутствующего движения (Н. М. Стойнова), цитатива (М. Б. Коношенко). Несколько особняком стоят работы, анализирующие контексты для экзистенциальных предикатов (Э. К. Кожевникова) и причинных

конструкций (М. Л. Федотов, О. В. Кузнецова), привлекающие как именной, так и глагольный материал. В большинстве статей данные базовых языков используются в полном объеме, но у ряда авторов предпочтение отдается собственной выборке (как в статьях Д. А. Парамоновой, П. С. Плешак, Д. В. Сичиной, А. Б. Пановой, Н. М. Стойновой, М. Л. Федотова и О. В. Кузнецовой, М. Б. Коношенко). Разнообразие представленного материала как раз и призвано убедить читателя в возможности осуществления намеченного нами проекта в более полном объеме. Можно надеяться, что такая работа — дело ближайшего будущего.

Список условных сокращений

3 — 3-е лицо; AVL — аблатив; ACC — аккузатив; DAT — датив; DEF — определенность; F — женский род; GEN — генитив; IPF — имперфект; M — мужской род; N — средний род; NOM — номинатив; PL — множественное число; POSS — посесивность; SG — единственное число.

Литература

- Барентсен 2008 — А. Барентсен. О конструкциях при глаголах восприятия в различных европейских языках (на основе переводов Нового Завета) // E. de Haard, W. Honselaar, J. Stelleman (eds.). *Literature and Beyond: Festschrift for Willem G. Weststeijn on the Occasion of his 65th Birthday*. Vol. 1. Amsterdam: Pegasus, 2008. P. 103–134.
- Кашкин 1991 — В. Б. Кашкин. Функциональная типология перфекта. Воронеж: Изд-во ВГУ, 1991.
- Мельчук 1998 — И. А. Мельчук. Курс общей морфологии. Т. II. Часть вторая: Морфологические значения / Пер. с франц. М.; Вена: Языки русской культуры; Wiener Slawistischer Almanach, 1998.
- Нестеренко 2019 — Л. В. Нестеренко. Мультиязычные параллельные корпуса: новый источник данных для типологических исследований, перспективы использования и проблемы // *Вопросы языкознания*. 2019. № 2. С. 111–125. DOI: 10.31857/S0373658X0004308-7.
- Плунгян 2011 — В. А. Плунгян. Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира. М.: РГГУ, 2011.

- Рахилина, Плунгян 2007 — Е. В. Рахилина, В. А. Плунгян. О лексико-семантической типологии // Т. А. Майсак, Е. В. Рахилина (ред.). Глаголы движения в воде: лексическая типология. М.: Индрик, 2007. С. 9–26.
- Сичинава 2015 — Д. В. Сичинава. Параллельные тексты в составе Национального корпуса русского языка: новые направления развития и результаты // Труды Института русского языка им. В. В. Виноградова. 2015. Вып. 6. № 3. С. 194–235.
- Сичинава 2016 — Д. В. Сичинава. Европейский перфект сквозь призму параллельного корпуса // Acta Linguistica Petropolitana. 2016. Т. XII. Ч. 2. С. 85–114.
- Aijmer 2008 — K. Aijmer. Parallel and comparable corpora // A. Lüdeling, M. Kytö (eds.). Corpus Linguistics: An International Handbook. Vol. I. Berlin: De Gruyter Mouton, 2008. P. 275–291.
- Bogaards 2022 — M. Bogaards. The discovery of aspect: A heuristic parallel corpus study of ingressive, continuative and resumptive viewpoint aspect // Languages. 2022. Vol. 7. № 3. P. 158. DOI: 10.3390/languages7030158.
- Čermák, Rosen 2012 — F. Čermák, A. Rosen. The case of InterCorp, a multilingual parallel corpus // International Journal of Corpus Linguistics. 2012. Vol. 13. № 3. P. 411–427. DOI: 10.1075/ijcl.17.3.05cer.
- Christodouloupoulos, Steedman 2015 — Ch. Christodouloupoulos, M. Steedman. A massively parallel corpus: the Bible in 100 languages // Language Resources and Evaluation. 2015. Vol. 49. № 2. P. 375–395. DOI: 10.1007/s10579-014-9287-y.
- Cysouw et al. 2007 — M. Cysouw, Ch. Biemann, M. Ongyerth. Using Strong's Numbers in the Bible to test an automatic alignment of parallel texts // STUF — Language Typology and Universals. 2007. Vol. 60. № 2. P. 158–171. DOI: 10.1524/stuf.2007.60.2.158.
- Cysouw, Wälchli (eds.) 2007 — M. Cysouw, B. Wälchli (eds.). Parallel Texts: Using Translational Equivalents in Linguistic Typology. Theme issue in STUF — Language Typology and Universals. 2007. Vol. 60. № 2.
- Dahl 1985 — Ö. Dahl. Tense and Aspect Systems. Oxford: Blackwell, 1985.
- Dahl 2007 — Ö. Dahl. From questionnaires to parallel corpora in typology // STUF — Language Typology and Universals. 2007. Vol. 60. № 2. P. 172–181. DOI: 10.1524/stuf.2007.60.2.172.
- Dahl 2014 — Ö. Dahl. The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus // B. Szmrecsanyi, B. Wälchli (eds.). Aggregating Dialectology and Typology: Linguistic Variation in Text and Speech, within and across Languages. Berlin: De Gruyter Mouton, 2014. P. 268–289. DOI: 10.1515/9783110317558.268.
- Dahl 2022 — Ö. Dahl. Perfects Across Languages // Annual Review of Linguistics. 2022. Vol. 8. № 1. P. 279–297. DOI: 10.1146/annurev-linguistics-031120-123428.

- Dahl (ed.) 2000 — Ö. Dahl. (ed.). *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter, 2000. DOI: 10.1080/03325860120702.
- de Swart et al. 2022 — H. de Swart, J. Tellings, B. Wälchli. *Not ... until* across European languages: A parallel corpus study // *Languages*. 2022. Vol. 7. № 1. P. 56. DOI: 10.3390/languages7010056.
- de Vries 2007 — L. de Vries. Some remarks on the use of Bible translations as parallel texts in linguistic research // *STUF — Language Typology and Universals*. 2007. Vol. 60. № 2. P. 148–157. DOI: 10.1524/stuf.2007.60.2.148.
- Doval, Sánchez Nieto (eds.) 2019 — I. Doval, M. T. Sánchez Nieto (eds.). *Parallel corpora for contrastive and translation studies: new resources and applications*. Amsterdam: John Benjamins, 2019. DOI: doi.org/10.1075/scl.90.
- Eckhoff, Haug 2009 — H. M. Eckhoff, D. T. Haug. Aligning Syntax in Early New Testament Texts: the PROIEL Corpus // B. Hansen, J. Grković-Major (eds.). *Diachronic Slavonic Syntax: Gradual Changes in Focus*. München: Sagner, 2009. P. 41–52.
- Frajzyngier, Mettouchi 2015 — Z. Frajzyngier, A. Mettouchi. Functional domains and cross-linguistic comparability // A. Mettouchi, M. Vanhove, D. Caubet (eds.). *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Amsterdam: John Benjamins, 2015. P. 257–279. DOI: 10.1075/scl.68.08fra.
- François 2018 — A. François. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages // M. Vanhove (ed.). *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*. Amsterdam: John Benjamins, 2008. P. 163–215. DOI: 10.1075/slcs.106.09fra.
- Haspelmath 1997 — M. Haspelmath. *From space to time: temporal adverbials in the world's languages*. München: LINCOM Europa, 1997.
- Haug et al. 2009 — D. T. Haug, M. L. Jøhndal, H. M. Eckhoff, E. Welø, M. J. B. Hertzberg, A. Mũth. Computational and linguistic issues in designing a syntactically annotated parallel corpus of Indo-European languages // *Traitement Automatique des Langues*. 2009. Vol. 50. № 2. P. 17–45.
- Johansson 2007 — S. Johansson. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins, 2007. DOI: 10.1075/scl.26.
- Koch 2001 — P. Koch. Lexical typology from a cognitive and linguistic point of view // M. Haspelmath, E. König, W. Oesterreicher, W. Raible (eds.). *Language Typology and Language Universals*. Vol. 2. Berlin: Walter de Gruyter, 2001. P. 1143–1175. DOI: 10.15496/publikation-19017.
- Koptjevskaja-Tamm et al. 2015 — M. Koptjevskaja-Tamm, E. Rakhilina, M. Vanhove. The semantics of lexical typology // N. Riemer (ed.). *The Routledge Handbook of Semantics*. London: Routledge, 2015. P. 434–454. DOI: 10.4324/9781315685533.

- Levshina 2015 — N. Levshina. European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles // *Folia Linguistica*. 2015. Vol. 49. № 2. P. 487–520. DOI: 10.1515/flin-2015-0017.
- Östling 2016 — R. Östling. Studying colexification through massively parallel corpora // P. Juvonen, M. Koptjevskaja-Tamm (eds.). *The Lexical Typology of Semantic Shifts*. Berlin: De Gruyter, 2016. P. 157–176. DOI: 10.1515/9783110377675-006.
- Stolz 2007 — T. Stolz. *Harry Potter meets Le petit prince* — On the usefulness of parallel corpora in crosslinguistic investigations // *STUF — Language Typology and Universals*. 2007. Vol. 60. № 2. P. 100–117. DOI: 10.1524/stuf.2007.60.2.100.
- van der Auwera et al. 2005 — J. van der Auwera, E. Schalley, J. Nuyts. Epistemic possibility in a Slavonic parallel corpus — a pilot study // B. Hansen, P. Karlik (eds.). *Modality in Slavonic languages: new perspectives*. München: Sagner, 2005. P. 201–218.
- Wälchli 2007 — B. Wälchli. Advantages and disadvantages of using parallel texts in typological investigations // *STUF — Language Typology and Universals*. 2007. Vol. 60. № 2. P. 118–134. DOI: 10.1524/stuf.2007.60.2.118.
- Wälchli 2010 — B. Wälchli. Similarity Semantics and Building Probabilistic Semantic Maps from Parallel Texts // *Linguistic Discovery*. 2010. Vol. 8. № 1. P. 331–371. DOI: 10.1349/PS1.1537-0852.A.356.
- Wälchli, Cysouw 2012 — B. Wälchli, M. Cysouw. Lexical typology through similarity semantics: Toward a semantic map of motion verbs // *Linguistics*. 2012. Vol. 50. № 3. P. 671–710. DOI: 10.1515/ling-2012-0021.
- Wälchli, Sölling 2013 — B. Wälchli, A. Sölling. The encoding of motion events: Building typology bottom-up from text data in many languages // J. Goschler, A. Stefanowitsch (eds.). *Variation and Change in the Encoding of Motion Events*. Amsterdam: John Benjamins, 2013. P. 77–113. DOI: 10.1075/hcp.41.04w228l.
- Wierzbicka 1988 — A. Wierzbicka. *The semantics of grammar*. Amsterdam: John Benjamins, 1988.
- Wiltschko 2014 — M. Wiltschko. *The universal structure of categories: towards a formal typology*. Cambridge: Cambridge University Press, 2014. DOI: 10.1017/CBO9781139833899.

References

- Aijmer 2008 — K. Aijmer. Parallel and comparable corpora. A. Lüdeling, M. Kytö (eds.). *Corpus Linguistics: An International Handbook. Vol. I*. Berlin: De Gruyter Mouton, 2008. P. 275–291.

- Barentsen 2008 — A. Barentsen. O konstruksiyakh pri glagolakh vospriyatiya v razlichnykh yevropeyskikh yazykakh (na osnove perevodov Novogo Zaveta) [On constructions with verbs of perception in various European languages (based on translations of the New Testament)]. E. de Haard, W. Honselaar, J. Stelleman (eds.). *Literature and Beyond: Festschrift for Willem G. Weststeijn on the Occasion of his 65th Birthday. Vol. 1*. Amsterdam: Pegasus, 2008. P. 103–134.
- Bogaards 2022 — M. Bogaards. The discovery of aspect: A heuristic parallel corpus study of ingressive, continuative and resumptive viewpoint aspect. *Languages*. 2022. Vol. 7. No. 3. P. 158. DOI: 10.3390/languages7030158.
- Čermák, Rosen 2012 — F. Čermák, A. Rosen. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*. 2012. Vol. 13. No. 3. P. 411–427. DOI: 10.1075/ijcl.17.3.05cer.
- Christodouloupoulos, Steedman 2015 — Ch. Christodouloupoulos, M. Steedman. A massively parallel corpus: the Bible in 100 languages. *Language Resources and Evaluation*. 2015. Vol. 49. No. 2. P. 375–395. DOI: 10.1007/s10579-014-9287-y.
- Cysouw et al. 2007 — M. Cysouw, Ch. Biemann, M. Ongyerth. Using Strong’s Numbers in the Bible to test an automatic alignment of parallel texts. *STUF — Language Typology and Universals*. 2007. Vol. 60. No. 2. P. 158–171. DOI: 10.1524/stuf.2007.60.2.158.
- Cysouw, Wälchli (eds.) 2007 — M. Cysouw, B. Wälchli (eds.). *Parallel Texts: Using Translational Equivalents in Linguistic Typology*. Theme issue in STUF — Language Typology and Universals. 2007. Vol. 60. No. 2.
- Dahl 1985 — Ö. Dahl. *Tense and Aspect Systems*. Oxford: Blackwell, 1985.
- Dahl 2007 — Ö. Dahl. From questionnaires to parallel corpora in typology. *STUF — Language Typology and Universals*. 2007. Vol. 60. No. 2. P. 172–181. DOI: 10.1524/stuf.2007.60.2.172.
- Dahl 2014 — Ö. Dahl. The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus. B. Szmrecsanyi, B. Wälchli (eds.). *Aggregating Dialectology and Typology: Linguistic Variation in Text and Speech, within and across Languages*. Berlin: De Gruyter Mouton, 2014. P. 268–289. DOI: 10.1515/9783110317558.268.
- Dahl 2022 — Ö. Dahl. Perfects Across Languages. *Annual Review of Linguistics*. 2022. Vol. 8. No. 1. P. 279–297. DOI: 10.1146/annurev-linguistics-031120-123428.
- Dahl (ed.) 2000 — Ö. Dahl. (ed.). *Tense and Aspect in the Languages of Europe*. Berlin: Mouton de Gruyter, 2000. DOI: 10.1080/03325860120702.
- de Swart et al. 2022 — H. de Swart, J. Tellings, B. Wälchli. *Not ... until* across European languages: A parallel corpus study. *Languages*. 2022. Vol. 7. No. 1. P. 56. DOI: 10.3390/languages7010056.

- de Vries 2007 — L. de Vries. Some remarks on the use of Bible translations as parallel texts in linguistic research. *STUF — Language Typology and Universals*. 2007. Vol. 60. No. 2. P. 148–157. DOI: 10.1524/stuf.2007.60.2.148.
- Doval, Sánchez Nieto (eds.) 2019 — I. Doval, M. T. Sánchez Nieto (eds.). *Parallel corpora for contrastive and translation studies: new resources and applications*. Amsterdam: John Benjamins, 2019. DOI: doi.org/10.1075/scl.90.
- Eckhoff, Haug 2009 — H. M. Eckhoff, D. T. Haug. Aligning Syntax in Early New Testament Texts: the PROIEL Corpus. B. Hansen, J. Grković-Major (eds.). *Diachronic Slavonic Syntax: Gradual Changes in Focus*. München: Sagner, 2009. P. 41–52.
- Frajzyngier, Mettouchi 2015 — Z. Frajzyngier, A. Mettouchi. Functional domains and cross-linguistic comparability. A. Mettouchi, M. Vanhove, D. Caubet (eds.). *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. Amsterdam: John Benjamins, 2015. P. 257–279. DOI: 10.1075/scl.68.08fra.
- François 2018 — A. François. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. M. Vanhove (ed.). *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*. Amsterdam: John Benjamins, 2008. P. 163–215. DOI: 10.1075/sles.106.09fra.
- Haspelmath 1997 — M. Haspelmath. *From space to time: temporal adverbials in the world's languages*. München: LINCOM Europa, 1997.
- Haug et al. 2009 — D. T. Haug, M. L. Jøhndal, H. M. Eckhoff, E. Welo, M. J. B. Hertenzenberg, A. Muth. Computational and linguistic issues in designing a syntactically annotated parallel corpus of Indo-European languages. *Traitement Automatique des Langues*. 2009. Vol. 50. No. 2. P. 17–45.
- Johansson 2007 — S. Johansson. *Seeing through Multilingual Corpora: On the use of corpora in contrastive studies*. Amsterdam: John Benjamins, 2007. DOI: 10.1075/scl.26.
- Kashkin 1991 — V. B. Kashkin. *Funktsionalnaya tipologiya perfekta* [Functional typology of the perfect]. Voronezh: Voronezh State University Publishing House, 1991.
- Koch 2001 — P. Koch. Lexical typology from a cognitive and linguistic point of view. M. Haspelmath, E. König, W. Oesterreicher, W. Raible (eds.). *Language Typology and Language Universals. Vol. 2*. Berlin: Walter de Gruyter, 2001. P. 1143–1175. DOI: 10.15496/publikation-19017.
- Koptjevskaja-Tamm et al. 2015 — M. Koptjevskaja-Tamm, E. Rakhilina, M. Vanhove. The semantics of lexical typology. N. Riemer (ed.). *The Routledge Handbook of Semantics*. London: Routledge, 2015. P. 434–454. DOI: 10.4324/9781315685533.
- Levshina 2015 — N. Levshina. European analytic causatives as a comparative concept: Evidence from a parallel corpus of film subtitles. *Folia Linguistica*. 2015. Vol. 49. No. 2. P. 487–520. DOI: 10.1515/flin-2015-0017.

- Melchuk 1998 — I. A. Melchuk. *Kurs obshchey morfologii. Tom II. Chast vtoraya: Morfologicheskkiye znacheniya* [A course of general morphology. Vol. II. Pt. 2: Morphological meanings]/Translated from French. Moscow; Vienna: Yazyki russkoy kultury; Wiyener Slawistischer Almanach, 1998.
- Nesterenko 2019 — L. V. Nesterenko. Multiyazychnyye parallelnyye korpusa: novyy istochnik dannykh dlya tipologicheskikh issledovaniy, perspektivy ispolzovaniya i problem [Multilingual parallel corpora: an alternative source of language data for typological studies, applying perspectives and problems]. *Voprosy Jazykoznanija*. 2019. No. 2. P. 111–125. DOI: 10.31857/S0373658X0004308-7.
- Östling 2016 — R. Östling. Studying colexification through massively parallel corpora. P. Juvonen, M. Koptjevskaja-Tamm (eds.). *The Lexical Typology of Semantic Shifts*. Berlin: De Gruyter, 2016. P. 157–176. DOI: 10.1515/9783110377675-006.
- Plungian 2011 — V. A. Plungian. *Vvedeniye v grammaticheskuyu semantiku: grammaticheskkiye znacheniya i grammaticheskkiye sistemy yazykov mira* [Introduction to grammatical semantics: grammatical meanings and grammatical systems of languages of the world]. Moscow: Russian State University for the Humanities Press, 2011.
- Rakhilina, Plungian 2007 — E. V. Rakhilina, V. A. Plungian. O leksiko-semanticheskoy tipologii [On lexical-semantic typology]. T. A. Maisak, E. V. Rakhilina (eds.). *Glagoly dvizheniya v vode: leksicheskaya tipologiya* [Verbs of motion in water: lexical typology]. Moscow: Indrik, 2007. P. 9–26.
- Sitchinava 2015 — D. V. Sitchinava. Parallelnyye teksty v sostave Natsionalnogo korpusa russkogo yazyka: novyye napravleniya razvitiya i rezultaty [Parallel texts within the Russian national corpus: new directions and results]. *Trudy Instituta Russkogo Iazyka imeni V. V. Vinogradova*. 2015. Vol. 6. No. 3. P. 194–235.
- Sitchinava 2016 — D. V. Sitchinava. Yevropeyskiy perfekt skvoz prizmu parallelnogo korpusa [European Perfect viewed from a parallel corpus]. *Acta Linguistica Petropolitana*. 2016. Vol. XII. Pt. 2. P. 85–114.
- Stolz 2007 — T. Stolz. *Harry Potter* meets *Le petit prince* — On the usefulness of parallel corpora in crosslinguistic investigations. *STUF — Language Typology and Universals*. 2007. Vol. 60. No. 2. P. 100–117. DOI: 10.1524/stuf.2007.60.2.100.
- van der Auwera et al. 2005 — J. van der Auwera, E. Schalley, J. Nuyts. Epistemic possibility in a Slavonic parallel corpus — a pilot study. B. Hansen, P. Karlík (eds.). *Modality in Slavonic languages: new perspectives*. München: Sagner, 2005. P. 201–218.
- Wälchli 2007 — B. Wälchli. Advantages and disadvantages of using parallel texts in typological investigations. *STUF — Language Typology and Universals*. 2007. Vol. 60. No. 2. P. 118–134. DOI: 10.1524/stuf.2007.60.2.118.

- Wälchli 2010 — B. Wälchli. Similarity Semantics and Building Probabilistic Semantic Maps from Parallel Texts. *Linguistic Discovery*. 2010. Vol. 8. No. 1. P. 331–371. DOI: 10.1349/PS1.1537-0852.A.356.
- Wälchli, Cysouw 2012 — B. Wälchli, M. Cysouw. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics*. 2012. Vol. 50. No. 3. P. 671–710. DOI: 10.1515/ling-2012-0021.
- Wälchli, Sölling 2013 — B. Wälchli, A. Sölling. The encoding of motion events: Building typology bottom-up from text data in many languages. J. Goschler, A. Stefanowitsch (eds.). *Variation and Change in the Encoding of Motion Events*. Amsterdam: John Benjamins, 2013. P. 77–113. DOI: 10.1075/hcp.41.04w2281.
- Wierzbicka 1988 — A. Wierzbicka. *The semantics of grammar*. Amsterdam: John Benjamins, 1988.
- Wiltschko 2014 — M. Wiltschko. *The universal structure of categories: towards a formal typology*. Cambridge: Cambridge University Press, 2014. DOI: 10.1017/CBO9781139833899.

Получено / received 04.04.2022

Принято / accepted 30.09.2022

