

**Конференция «К построению базы данных
универсальных грамматических контекстов
на материале параллельного корпуса
переводов Нового Завета»**

**Conference on “Building a database of universal
grammatical contexts based on a parallel corpus
of New Testament translations”**

DOI: 10.37892/2313-5816-2021-2-229-238

Конференция с таким довольно длинным названием (но вполне точно передающим ее тематику) прошла в Москве 27–28 августа 2021 г. Ее организаторами выступили Отделение теоретической и прикладной лингвистики Московского университета и Проблемная группа по теории грамматики Института языкознания РАН. Большинство представленных докладов в той или иной степени являлись результатом работы семинара по использованию параллельных корпусов в лингвистической типологии, проходившего на ОТиПЛ МГУ в 2020–2021 гг. (под руководством В. А. Плунгяна и при участии И. В. Макаrchука, Н. В. Перковой и К. В. Филатова). В конференции также приняли участие сотрудники Института перевода Библии. Всего за два дня работы конференции было представлено 15 докладов.

Проблематика конференции находится на пересечении сразу трех (на первый взгляд, достаточно далеких друг от друга) областей: типологии грамматических категорий в языках мира, корпусов параллельных текстов и текстологии Нового Завета. Тем не менее все эти три области связаны. Типологические исследования языков мира сегодня всё чаще делаются с использованием параллельных корпусов — этот инструмент зарекомендовал себя как более

объективный и надежный по целому ряду параметров по сравнению с традиционным анкетированием, предполагающим прямое воздействие исследователя на носителя языка. Для масштабных типологических исследований (а не для сравнения двух или нескольких языков в рамках более банального «контрастивного» анализа) особенно подходят так называемые многоязычные параллельные корпуса (англ. *massively parallel corpora*), в которых исходный текст представлен в переводах на большое количество типологически разнородных языков (много десятков или даже сотен). Применение таких корпусов для типологических исследований грамматики и лексики — относительно новое направление в лингвистике, но уже успевшее привлечь внимание многих ведущих типологов и принести интересные результаты: ср. [Cysouw, Wälchli 2007; Dahl 2007; Aijmer 2008; Čermák, Rosen 2012; Östling 2016; Сичинава 2016; Нестеренко 2019] и др.

Разумеется, один из главных вопросов при таком формате исследования — какой именно текст может быть использован в многоязычном параллельном корпусе. Очевидным образом, тексты, существующие в переводе на многие сотни языков, не столь многочисленны; ещё меньше текстов, жанр и содержание которых делает их пригодными для масштабных типологических сопоставлений. И наконец, текстов, которые при этом переводились бы не только на крупные и культурно значимые языки, но и на малые (и даже бесписьменные) языки, практически не существует — а ведь именно такие языки и являются предметом интереса типологов в первую очередь. Точнее, такой текст более или менее единственный и это, как нетрудно понять, и есть текст Нового Завета: он не имеет равных ни по числу переводов (считается, что полностью он переведен более чем на 1,5 тыс. языков), ни по типологическому разнообразию языков перевода.

Уникальная «переводоемкость» Нового Завета, разумеется, была осознана и оценена лингвистами достаточно дав-

но, но значение этого факта особенно возросло с появлением в обиходе лингвистов электронных корпусов (и корпусной лингвистики как самостоятельной дисциплины со своим методом и набором инструментов). Неудивительно, что и типологи, и специалисты по корпусной лингвистике не раз обращали внимание на то, что существующие переводы Нового Завета могут рассматриваться как один и старейших, самых объемных и разнообразных параллельных корпусов, с помощью которого можно решать очень многие исследовательские задачи (ср., например, [Haspelmath 1997; de Vries 2007; Wälchli, Cysouw 2012]). В настоящее время известно даже несколько специальных проектов, посвященных поиску оптимального представления переводов Нового Завета для лингвистических исследований (хотя ни один из них пока полностью не завершен).

Но если преимущества использования Нового Завета как уникального параллельного корпуса для типологических исследований языков хорошо осознаются лингвистическим сообществом (и опыты такого рода имеются, ср., помимо указанных выше, также, например, работы [Барентсен 2008, 2015; Wälchli, Sölling 2013; Dahl 2014; Dahl, Wälchli 2016]), то пока еще нельзя сказать, что все возможности этого инструмента использованы в полной мере. В частности, все известные на сегодняшний день примеры использования параллельных корпусов в грамматической типологии сводятся к тому, что исследователи выбирают некоторое интересующее их грамматическое явление и пытаются обнаружить его в различных переводных версиях определенного текста. Иначе говоря, параллельный корпус остается при таком подходе лишь пассивным источником материала.

Идея, лежащая в основе докладов, представленных на настоящей конференции, другая. Во всех этих работах ставится принципиально иная задача: сделать параллельный корпус (в данном случае, корпус переводов Нового Завета) не просто используемым *ad hoc* источником, а активным

инструментом исследования грамматической семантики. Для этого требуется прежде всего превратить его в *базу данных универсальных грамматических контекстов*, где будет указана локализация важнейших семантических элементов из Универсального грамматического набора. Тем самым, исследователь будет заранее знать, в каких фрагментах текста наиболее целесообразно искать, например, грамматические показатели хабитуальности, проспектива, двойственного числа или слабой определенности. Именно эта идея и отражена в названии конференции.

Опыт работы семинара ОТиПЛ МГУ показал, что такая цель достижима — и даже первые исследования в этом направлении приносят нетривиальные результаты. Соответственно, в представленных на конференции докладах обсуждались как общие проблемы построения и использования такой базы данных, так и задачи поиска и классификации релевантных контекстов для конкретных семантических элементов.

Конференция открылась вводным докладом В. А. Плунгяна «Параллельный корпус как грамматическая база данных и Новый Завет как параллельный корпус», обозначившим теоретические рамки конференции и охарактеризовавшим основные направления ее работы. В последующих докладах обсуждались как именные, так и глагольные категории с точки зрения возможностей нахождения прототипических контекстов для их диагностики в Новом Завете. Как правило, участниками проекта использовались для первоначального поиска пять «базовых» языков (послужившие основой для большинства существующих переводов): древнегреческий, латинский, церковнославянский (и русский Синодальный перевод), английский (преимущественно King James New Version) и испанский. При всей ограниченности этого списка, в нем представлены достаточно разнородные языки, использующие отнюдь не тождественные грамматические стратегии. Дополнительно участники проекта могли в отдельных случаях использовать какие-то другие, более редкие

языки с целью проверить встречаемость определенных грамматических явлений именно в них (см. ниже список докладов).

Глагольные категории, исследованные по этой схеме, рассматривались в следующих докладах:

- А. Б. Панова (НИУ ВШЭ) «К описанию релевантных контекстов для континуатива (Евангелие от Луки)»;
- К. В. Кашинцева (НИУ ВШЭ) «Отрицательная фазовая полярность: дисконтинуатив»;
- А. А. Васяткина (МГУ) «К описанию релевантных контекстов для контрфактивности»;
- А. В. Алхазова (МГУ) «К описанию релевантных контекстов для хабитуального аспекта»;
- Ю. Н. Кузнецова, А. Д. Подгорная (МГУ) «К описанию релевантных контекстов для экспериентивности»;
- Д. В. Сичинава (ИРЯ РАН / НИУ ВШЭ) «Плюсквамперфект и предбудущее в Новом Завете»;
- К. Т. Гадилия (Институт перевода Библии) «Юссив, гортатив и императив в переводах Книги Бытия (1:3, 6, 9, 11, 20, 22, 24, 28) на грузинский, осетинский и кумыкский» (данный доклад по материалу и постановке задачи, как можно видеть, несколько отличался от остальных, но был встречен участниками конференции с большим интересом).

Исследованию неглагольных категорий были посвящены следующие доклады:

- Д. Д. Врубель, Л. И. Паско (МГУ) «К описанию релевантных контекстов для двойственного числа»;
- К. А. Флоринский (НКРЯ / ИРЯ РАН) «Инклюзив (на материале языков кечуа и чоль)»;
- Д. А. Парамонова (МГУ) «Контексты “точечных” контактных локализаций в уральских и алтайских языках»;
- К. В. Филатов (СПбГУ) «Контексты латива»;
- Т. Е. Фейгина (МГУ) «К описанию релевантных контекстов для пространственного и временного дейксиса»;

- И. А. Хомченкова (ИРЯ РАН / МГУ) «К описанию релевантных контекстов для универсальной квантификации»;
- Ю. В. Сеницына (МГУ) «К описанию релевантных контекстов для компаративных и эквативных конструкций».

Можно отметить значительное разнообразие исследованных категорий: в той или иной степени были затронуты аспектуальность, темпоральность, модальность, число и квантификация, лицо, дейксис, пространственные и другие падежные роли. Во всех случаях был представлен убедительный материал, позволяющий диагностировать соответствующий грамматический элемент в произвольном языке с помощью набора предложенных контекстов.

Разумеется, в ходе очень живого и заинтересованного обсуждения всех докладов было выявлено множество проблем, релевантных для дальнейшего исследования. Проект построения базы данных универсальных грамматических контекстов на материале параллельного корпуса переводов Нового Завета будет продолжаться, впереди публикации, а также новые семинары и конференции. Участники проекта заинтересованы в сотрудничестве и приглашают всех желающих внести посильный вклад в это исследование.

Литература

Барентсен А. 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 I. Amsterdam, 2008, 103–134.

Барентсен А. Наблюдения о встречаемости плюсквамперфекта в переводах евангелий на славянские языки // Л. Попович, Д. Војводић, М. Номачи (ред.). *У простору лингвистичке славистике: Зборник научних радова поводом 65 година живота академика Предрага Пипера*. Белград, 2015, 135–160.

Нестеренко Л. В. Мультиязычные параллельные корпуса: новый источник данных для типологических исследований, перспективы использования и проблемы // *Вопросы языкознания*, 2019, 2: 111–125.

Сичинава Д. В. Европейский перфект сквозь призму параллельного корпуса // Т. А. Майсак, В. А. Плунгян, Кс. П. Семёнова (ред.). *Исследования по теории грамматики*, вып. 7: *Перфект и смежные категории*. Санкт-Петербург, 2016 (ALP XII: 2), 85–114.

Aijmer K. Parallel and comparable corpora // A. Lüdeling, M. Kytö (eds.). *Corpus Linguistics. An International Handbook*. Vol. I. Berlin, 2008, 275–291.

Čermák F., Rosen A. The case of InterCorp, a multilingual parallel corpus // *International Journal of Corpus Linguistics*, 2012, 13 (3): 411–427.

Cysouw M., Wälchli B. (eds.). Parallel texts: Using translational equivalents in linguistic typology // *Theme issue in STUF — Language Typology and Universals*, 2007, 60 (2).

Dahl Ö. From questionnaires to parallel corpora in typology // *STUF — Language Typology and Universals*, 2007, 60 (2): 172–181.

Dahl Ö. 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, 2014, 268–289.

Dahl Ö., Wälchli B. Perfects and iamitives: two gram types in one grammatical space // *Letras de Hoje*, 2016, 51 (3): 325–348.

de Vries L. Some remarks on the use of Bible translations as parallel texts in linguistic research // *STUF — Language Typology and Universals*, 2007, 60 (2): 148–157.

Haspelmath M. *From space to time: Temporal adverbials in the world's languages*. München, 1997.

Östling R. Studying colexification through massively parallel corpora // P. Juvonen, M. Koptjevskaja-Tamm (eds.). *The lexical typology of semantic shifts*. Berlin, 2016, 157–176.

Wälchli B., Cysouw M. Lexical typology through similarity semantics: Toward a semantic map of motion verbs // *Linguistics*, 2012, 50 (3): 671–710.

Wälchli B., Sölling A. 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, 2013, 77–113.

References

Barentsen A. Nablyudeniya o vstrechaemosti plyuskvamperfekta v perevodakh evangelii na slavyanskije yazyki [Observations on the occurrence of pluperfect in translations of the Gospels into Slavic languages] // L. Popovich, D. Vojvodiћ, M. Nomachi (red.). *U prostoru lingvistichke slavistike: Zbornik nauchnikh radova povodom 65 godina zhivota akademika Predraga Pipera*. Belgrad, 2015, 135–160. (In Russ.)

Barentsen A. O konstruktsiyakh pri glagolakh vospriyatiya v razlichnykh evropeiskikh yazykakh (na osnove perevodov Novogo Zaveta) [About constructions with verbs of perception in various European languages (based on the New Testament translations)] // E. de Haard, W. Honselaar, J. Stelleman (eds.). *Literature and Beyond. Festschrift for Willem G. Weststeijn on the Occasion of his 65th Birthday*. Vol I. Amsterdam, 2008, 103–134. (In Russ.)

Čermák F., Rosen A. The case of InterCorp, a multilingual parallel corpus // *International Journal of Corpus Linguistics*, 2012, 13 (3): 411–427.

Cysouw M., Wälchli B. (eds.). Parallel texts: Using translational equivalents in linguistic typology // *Theme issue in STUF — Language Typology and Universals*, 2007, 60 (2).

Dahl Ö. From questionnaires to parallel corpora in typology // *STUF – Language Typology and Universals*, 2007, 60 (2): 172–181.

Dahl Ö. 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, 2014, 268–289.

Dahl Ö., Wälchli B. Perfects and iamitives: two gram types in one grammatical space // *Letras de Hoje*, 2016, 51 (3): 325–348.

de Vries L. Some remarks on the use of Bible translations as parallel texts in linguistic research // *STUF — Language Typology and Universals*, 2007, 60 (2): 148–157.

Haspelmath M. *From space to time: Temporal adverbials in the world's languages*. München, 1997.

Nesterenko L. V. Mul'tiyazychnye parallel'nye korpusa: novyi istochnik dannykh dlya tipologicheskikh issledovaniy, perspektivy ispol'zovaniya i problemy [Multilingual Parallel Corporuses: A New Data Source for Typological Research, Prospects for Use and Problems] // *Voprosy yazykoznaviya*, 2019, 2: 111–125. (In Russ.)

Östling R. Studying colexification through massively parallel corpora // P. Juvonen, M. Koptjevskaja-Tamm (eds.). *The lexical typology of semantic shifts*. Berlin, 2016, 157–176.

Sichinava D. V. Evropeiskii perfekt skvoz' prizmu parallel'nogo korpusa [European perfect through the prism of a parallel corpus] // T. A. Maisak, V. A. Plungyan, Ks. P. Semenova (red.). *Issledovaniya po teorii grammatiki, vyp. 7: Perfekt i smezhnye kategorii*. St.-Petersburg, 2016 (ALP XII: 2), 85–114. (In Russ.)

Wälchli B., Cysouw M. Lexical typology through similarity semantics: Toward a semantic map of motion verbs // *Linguistics*, 2012, 50 (3): 671–710.

Wälchli B., Sölling A. 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, 2013, 77–113.

Владимир Александрович Плунгян
Институт языкознания РАН
МГУ им. Ломоносова
Москва, Россия

Vladimir Alexandrovich Plungian
Institute of Linguistics, RAS
Lomonosov Moscow State University
Moscow, Russia
plungian@gmail.com