

Д. В. Сичинава
D. V. Sitchinava

ПОЛИВАРИАНТНЫЕ ПАРАЛЛЕЛЬНЫЕ ТЕКСТЫ В СОСТАВЕ НКРЯ¹

MULTIPLE TRANSLATION VARIANTS IN A PARALLEL CORPUS WITHIN THE RNC

Аннотация. В докладе представлен проект параллельного русско-французского корпуса в рамках НКРЯ, где каждому оригинальному тексту соответствует несколько вариантов перевода. На базе данного корпуса строится одновременно база данных соответствия видо-временных глагольных форм в русском и французском языках. В докладе указаны типы разметки несоответствий и неточных соответствий оригинала и перевода, а также параметры указанной базы данных.

Abstract. The paper presents a project of a parallel Russian-French corpus within the RNC with multiple translation variants of each Russian text. A database of cross-linguistical equivalences within the TAM zone based on the corpus is being built. In our paper we give the basic types of marking up the translation discrepancies as well as some properties of this database.

1. Поливариантный корпус: общее

Параллельные корпуса в составе Национального корпуса русского языка² до недавнего времени включали в себя только по

¹ Работа выполнена при поддержке программы Президиума РАН № 36-П «Корпусная лингвистика» и грантов РФФИ № 12-06-33038 «Контрастивные корпусные исследования русских и французских глагольных категорий в поливариантных параллельных текстах» и № 13-06-00403 «Контрастивное корпусное исследование специфических черт семантической системы русского языка».

² См. о них: *Добровольский Д. О., Кретов А. А., Шаров С. А.* Корпус параллельных текстов. Архитектура и возможности использования // Национальный корпус русского языка 2003–2005. Результаты и перспективы. М., 2005. С. 263–296; *Сичинава Д. В.*

одному переводу данного текста на данный язык. Единственным исключением был многоязычный корпус, содержащий 9 текстов, для некоторых из которых давались несколько альтернативных переводов (в частности, несколько переводов «Алисы в стране чудес» и «Винни-Пуха» на польский и русский языки). Со второй половины 2012 года команда НКРЯ совместно с исследователями Института проблем информатики РАН и французскими лингвистами (университет Paris–13, Институт восточных языков INALCO) разрабатывает поливариантный параллельный русско-французский корпус³. Одновременно на его материале строится база данных поливариантных соответствий (полиэквиваленций) глагольных аспектуальных и временных форм в русском и французском языках.

Поливариантные параллельные корпуса – сравнительно слабо развитое направление в рамках корпусной лингвистики, и тем меньше существует собственно сопоставительных грамматических исследований, выполненных на их базе. Существует поливариантный русско-немецкий корпус на материале выполненных в разное время переводов романов Достоевского, созданный в Австрийской академии наук в Вене⁴, но он не находится в открытом доступе. Он используется прежде всего для исследования лексики и эволюции переводческих

Комплексное использование одноязычного и параллельного корпуса в грамматических исследованиях // Труды международной конференции «Корпусная лингвистика-2011», Спб., 2011. С. 316–321; *Sitchinava D.* Parallel Corpora within the Russian National Corpus // *Prace Filologiczne*, LXIII, 2012. S. 271–278.

³ См. подробнее: *Loiseau S., Sitchinava D., Zalizniak A., Zatsman I.* Information technologies for creating the database of Equivalent verbal forms in a Russian-French multivariant parallel corpus // *Информатика и ее применения*, 2013 (в печати).

⁴ *Добровольский Д.О.* Корпус параллельных текстов в исследовании культурно-специфичной лексики // *Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы*. СПб., 2009. С. 383–401.

техник. По нескольку вариантов перевода ряда текстов на один и тот же язык (например, «Алиса в стране чудес» – шесть переводов на русский и четыре на польский) находится в ряде многоязычных параллельных корпусов, например, в Амстердамском корпусе параллельных текстов (ASPAC) и Регенсбургско-Бернском параллельном корпусе.

Эти корпуса использовались для контрастивных исследований, в частности, вида в славянских языках⁵. Однако рассмотрение нескольких вариантов перевода на один язык занимало в этих исследованиях подчинённую позицию: сравнивались, прежде всего, переводы на разные языки, специальной задачи исследования полиэквиваленций не ставилось, а отмечались лишь отдельные расхождения в некоторых местах. Аналогичных сопоставительных корпусных исследований фрагментов грамматики на базе параллельного корпуса ранее не создавалось, а поливариантный русско-французский корпус создаётся впервые.

Для поливариантного корпуса выбираются тексты, существующие не менее, чем в двух переводах на французский язык, причём эти переводы должны быть созданы, как правило, не ранее середины XX века. Более ранние французские переводы русской литературы, особенно выполненные в XIX веке, устарели в языковом отношении, а главное, содержат много ошибок и сокращений исходного текста.

Для многих произведений русской литературы (Н. В. Гоголь, И. А. Гончаров, Л. Н. Толстой) существует несколько новых переводов, выполненных на высоком уровне. Они демонстрируют как различные подходы к художественному переводу, так и

⁵ v. Waldenfels R. Compiling a parallel corpus of slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beitrage der Europäischen Slavistischen Linguistik (POLYSLAV) 9. Munchen, 2006, S. 123–138 (доступно по адресу <http://www-nw.uni-regensburg.de/%7E.war05297.slavistik.sprachlit.uni-regensburg.de/pub/WaldenfelsParallelCorpora2006.pdf>).

различный выбор языковых выражений, в том числе грамматических.

Таблица 1. Пример выравнивания оригинального текста с двумя переводами («Обломов» И. А. Гончарова)

<p>Цвет лица у Ильи Ильича не был ни румяный, ни смуглый, ни положительно бледный, а безразличный или казался таким, может быть, потому, что Обломов как-то обрюзг не по летам: от недостатка ли движения или воздуха, а может быть, того и другого.</p>	<p>Le teint d'Iliia Ilitch n'était ni rose, ni mat, pas véritablement pâle non plus; neutre plutôt, ou du moins ayant cette apparence, et cela peut-être parce qu'Oblomov était prématurément un peu flasque, faute d'exercice, ou faute d'air, ou faute d'exercice et faute d'air à la fois.</p>	<p>Le teint d'Iliia Ilitch n'était ni rose, ni hâlé, ni carrément pâle, mais indifférent ou, du moins, il le paraissait. Peut-être parce que la chair d'Oblomov était prématurément flasque: faute d'exercice ou manque d'air, peut-être l'un et l'autre.</p>
--	---	---

2. Типология несоответствий перевода

Разработана типология несоответствий выравнивания нескольких переводов на французский язык одного текста на русском языке. Эта типология включает следующие категории признаков для описания несоответствий для каждого варианта перевода:

1) Категория «Добавления/замены/пропуски», признаки которой полностью совпадают с признаками в программе выравнивания «Евклид», используемой при формировании моновариантных параллельных корпусов. В программе «Евклид» используются три признака: «значимое добавление», «значимая замена» и «значимый пропуск». Сохранение этих трех признаков в новой программе обеспечивает преемственность между

программой выравнивания «Евклид» и новой программой выравнивания.

2) Категория «Единица несоответствия» предназначена для описания несоответствия по следующей шкале признаков: «слово», «синтагма», «целое предложение». Эти признаки позволяют дополнить признаки первой категории, указав, например, что именно пропущено (слово, синтагма или предложение).

3) Категория «Уровень несоответствия». С помощью признаков этой категории можно будет указать уровень наблюдаемого несоответствия: «семантический», «стилистический», «грамматический».

4) Категория «Семантические аспекты добавлений / замен / пропусков». С помощью признаков этой категории можно будет указать, какую семантическую функцию несут изменения, вносимые переводчиком при переводе оригинального текста на французский язык. В эту категорию включены следующие признаки:

– «упрощение»: переводчик упрощает оригинальный текст, использует более простые выражения, опуская некоторые детали (пример: «Он позволял это себе потому, что чувствовал в себе силу всегда, когда ему понадобится, опять выделить одно служебное и откинуть человеческое» => «Il le faisait seulement parce qu'il se sentait de force à rétablir à n'importe quel moment les barrières fatidiques»);

– «усложнение»: переводчик использует более изысканные выражения или вносит дополнительные детали, отсутствующие в оригинале, дополняя описываемую ситуацию (пример: «В промежутки он курил, пил чай, беседовал немножко о политике, немножко об общих делах...» => «...pendant les suspensions d'audience, il savait se délasser et passer agréablement le temps à fumer une cigarette, prendre le thé, bavarder à bâtons rompus...»);

– «разъяснение»: переводчик описывает ситуацию более подробно, чем в оригинале, стремясь сделать ее более

простой и понятной для понимания (пример: «Дверцу – тонкую цинковую пластинку – отвел в сторону, слез, пугливо поглядел на окна, потрогал простыню». => «Il ôta la mince plaque de zinc qui la fermait, posa celle-ci sur l'étagère, descendit de sa chaise, et alla à la fenetre vérifier d'un air inquiet que le drap et le plaid n'avaient pas bougé»).

5) Категория «Трудности перевода». С помощью признаков этой категории можно будет помечать несоответствия, возникающие из-за сложности переводимых отрезков текста. В эту категорию включены следующие признаки:

– «сложность интерпретации оригинала»: смысл оригинального фрагмента непонятен или его сложно интерпретировать носителям языка;

– «культуроспецифическая лексика»: смысл оригинального фрагмента ясен, но он включает в себя понятия, которых нет в языке перевода.

6) Категория «Языки оригинала». С помощью признаков этой категории можно будет помечать в оригинале фрагменты текста на языках, отличных от языка оригинала. В эту категорию включены следующие признаки: «язык отмеченного фрагмента оригинала отличается от языка текста всего оригинала и от языка перевода», «язык фрагмента в оригинале совпадает с языком перевода», «диалект языка оригинала».

3. Построение базы данных эквивалентных глагольных форм

В базе данных эквивалентных глагольных форм объектом анализа являются «соответствия», т.е. пары функционально эквивалентных форм русского и французского языков. На материале русско-французского параллельного корпуса речь может идти о соответствиях двух типов. Это:

1) «модели перевода» – множество переводов $\{F_n \dots F_{n+m}\}$ для русской формы R (например, НСВ переводится как present, imparfait или в известных ограничительных контекстах passé simple);

2) «стимулы перевода» – множество «стимулов» {Rn...Rn+m}, «реакцией» на которые является французская форма F (например, *passé antérieur* может появиться для русских слов «немедленно» или «внезапно»).

Соответствия первого типа позволяют получить новые сведения об устройстве русского языка, второго – об устройстве французского. Поскольку объектом нашего интереса является русский язык, в рамках данного проекта рассматриваются лишь соответствия в направлении «русский => французский», т. е. «модели перевода».

Второе ограничение состоит в том, что в создаваемую базу данных функционально эквивалентных лексико-грамматических глагольных форм русского и французского языков включаются только те глагольные формы русского языка, которые содержат спрягаемый глагол – *verbum finitum* (т.е. исключаются слова категории состояния, перифразы с глаголом «быть» и некоторые другие типы предикативных единиц).

Когда ведение базы будет отработано на тестовом материале, она будет расширена. Последующее ее расширение возможно также в направлении включения данных, полученных из параллельного француско-русского корпуса; они позволят ответить на вопрос «На какие французские «стимулы» появляется в переводе в качестве «реакции» данная русская форма?».

В рамках данного проекта рассматривается кластер глагольных категорий, который в типологической литературе получил обозначение TAM (Tense-Aspect-Mood⁶; ср. также типологические исследования языков Европы в рамках комплексного проекта EUROTYPE).

⁶ *Bybee J. L., Perkins R., Pagliuca W.* The evolution of grammar: Tense, aspect and modality in the languages of the world. Chicago, 1994; *Comrie B.* Aspect. An introduction to the study of verbal aspect and related problems. Cambridge, 1976; *Comrie B.* 1985. Tense. Cambridge, 1985; *Dahl Ö.* Tense and aspect systems. Oxford, 1985; *Dahl Ö.* (ed.) Tense and aspect in the languages of Europe. Berlin – New York, 2000.

Вводится понятие «лексико-грамматической формы», которое представляет собой комбинацию значений трех категорий для русского языка (Tense, Aspect, Mood) и двух (Tense, Mood) – для французского. Понятие «лексико-грамматической формы» (ЛГФ) близко по своему содержанию к понятию «конструкции» – в том значении, которое придается этому термину в Грамматике конструкций⁷, т.е. это может быть как набор значений грамматических и семантических категорий, задающий класс удовлетворяющих данным требованиям реальных форм (т.е. «грамматическая конструкция» в традиционном понимании), так и единичная лексема или словоформа. Понятие лексико-грамматической формы представляется наиболее адекватным инструментом контрастивного анализа поливариантного параллельного корпуса.

⁷ *Goldberg A. Constructions: A Construction Grammar Approach to Argument structure. Chicago, 1995. Goldberg A. Constructions at Work. The nature of generalization in grammar. Oxford, 2006.*