

Divergent translation of connectives in human and machine translations

Inkova O.

Institute of Informatics Problems, FRC
CSC RAS / Moscow, Russia
University of Geneva, Geneva,
Switzerland
Olga.Inkova@unige.ch

Nuriev V.

Institute of Informatics Problems, FRC
CSC RAS / Moscow, Russia
nurieff.v@gmail.com

Abstract

The paper is focused on divergent ways of conveying discourse relations in translation. For data collection, we used the supracorpora database of connectives storing parallel texts from the Russian-French subcorpus of the Russian National Corpus. These data show what logical-semantic relations tend to be translated using divergent ways, i.e. other than connectives (exclusion in its various gradations, propositional concomitance and substitution, the share of divergent translations ranging from 30% to 50%). Also, such data help define what causes divergent ways of translation to be used. The causes may be as follows: (a) the lack of an adequate equivalent of a given connective in the target language; (b) differences in the syntactic structure of the source and target languages; (c) usage differences; (d) contextually determined use of divergent translation. If there is a prototypical indicator of logical-semantic relations (i.e. connective) in the source text, it also occurs in translation in more than 90% of cases. The data on human translations are then compared with those on machine translations, which shows that the machine translation system also tends to keep a connective if there is one in the source text (it occurs in almost 98% of cases). However, there are cases where the machine translation system has difficulties processing a multiword connective (failing to perceive it as a whole) or a polyfunctional unit (failing to tell a connective from a non-connective) and thus uses divergent ways to translate it. Some causes of divergently translating connectives are likely to be the same for human and machine translations. These are differences in the syntactic structure of languages and usage differences. Further research of divergent means of conveying discourse relations will allow to draw a sharper border-line between explicitly expressed and implicit discourse relations. The data collected from annotated corpora (both monolingual and multilingual and parallel) will help determine what the divergent ways of expressing logical-semantic relations are and how frequently they are used. The research results can be used both in automatic text processing and automatic text generation. Also, the data on divergent translations of discourse relations can serve to improve the machine translation quality.

Keywords: semantics; discourse relations; connective; divergent translation; parallel corpora; quantitative analysis; machine translation

DOI: 10.28995/2075-7182-2021-20-339-348

Дивергентный перевод коннекторов в авторских и машинных переводах

Инькова О.Ю.

ИПИ ФИЦ ИУ РАН / Москва, Россия
Женевский университет / Женева,
Швейцария
Olga.Inkova@unige.ch

Нуриев В.А.

ИПИ ФИЦ ИУ РАН / Москва, Россия
nurieff.v@gmail.com

Ключевые слова: семантика; дискурсивные отношения; коннектор; дивергентный перевод; параллельные корпуса; количественный анализ; машинный перевод

1 Введение

Проблема дивергентных средств выражения логико-семантических отношений (ЛСО) встает особенно остро при аннотировании корпусов текстов с точки зрения их связности. Как известно, дискурсивные отношения могут выражаться как коннекторами, так и другими языковыми средствами, получившими название «альтернативные лексикализации» [Prasad et al. 2010]. В дальнейшем была предложена классификация языковых средств, способных выражать дискурсивные (в данной терминологии) отношения [Das, Taboada 2013]; [Taboada, Das 2013]; [Das, Taboada 2014]. Выделяются лексические, морфологические (временные формы), семантические (синонимия, антонимия и др.), синтаксические (различные виды придаточных предложений и др.) и графические средства, которые могут сочетаться.

- (1) – Слава Богу, слава Богу, – заговорила она, – теперь всё готово. *Только* немножко вытянуть ноги. [Л. Н. Толстой. Анна Каренина (1873-1877)]
 – Dieu merci, me voilà prête. Il *ne me reste qu'*à étendre un peu les jambes. [Tr. H. Mongault]

В (1) *только*, выступающее в функции коннектора, устанавливает между положениями вещей «теперь почти все готово» и «(нужно) немножко вытянуть ноги» ЛСО исключения (всё готово, за исключением того, что нужно вытянуть ноги). Французский переводчик, вместо коннектора исключения, например, *sauf que*, использует два языковых средства: грамматическое (ограничительное отрицание *ne... que*, часто выступающее эквивалентом *только*) и лексическое (глагол *rester* ‘оставаться’, т.е. буквально ‘остаётся вытянуть ноги’).

При работе с параллельными корпусами для таких случаев используется термин «дивергентный перевод», предложенный в [Johansson 2007] и уточненный для описания средств выражения дискурсивных отношений в [Инькова 2019а]. Дивергентным считается такой перевод, когда показатель ЛСО – коннектор – переведен языковым средством, не принадлежащим к этому функциональному классу, и, наоборот, когда такое языковое средство переведено коннектором¹. Ср. (2), где в варианте а. выбрано конгруэнтное средство перевода коннектора *a*, выражающего сопоставительные ЛСО (французский коннектор *tandis que* ‘тогда как’), а в варианте б. – дивергентное: местоименный повтор подлежащего, создающий сопоставление действий двух субъектов, Обломова и Алексеева.

- (2) Обломов задумался, *a* Алексей барабанил пальцами по столу, у которого сидел.
 [И. А. Гончаров. Обломов (1848-1859)]
 а. Oblomov se replongea dans ses rêveries, *tandis qu'*Alexeïev tambourinait sur la table devant laquelle il était assis. [Tr. A. Adamov]
 б. Oblomov devint pensif, Alexéev, *lui*, pianotait sur la table. [Tr. L. Jurgenson]

В настоящем исследовании на основе информации, полученной в Надкорпусной базе данных коннекторов (НБД)², мы покажем, для каких ЛСО наиболее характерны дивергентные средства перевода и проанализируем причины их использования (раздел 2). Затем мы сравним данные, полученные для переводов, выполненных профессиональными переводчиками, с данными, полученными для машинного перевода (раздел 3). Особый интерес представляет тот факт, что в обоих случаях переводился один и тот же русский контекст.

¹ Уточним, что на основании этого определения случаи, когда коннектор переведен коннектором, выражающим другое ЛСО, считаются конгруэнтными переводами.

² Подробнее об устройстве и возможностях использования НБД для сопоставительного количественного анализа коннекторов см. [Inkova 2021]; представительный фрагмент НБД доступен на сайте <http://a179.frccsc.ru/RFH41002/main.aspx>.

2 Дивергентные средства выражения логико-семантических отношений в авторских переводах

Материалом для исследования послужили 11 252 двуязычные аннотации в направлении перевода русский-французский³. Из них аннотаций с пометой «Дивергентное межъязыковое соответствие» (Dvrg) – 903, т. е. 8,1%. Эта цифра показательна сама по себе: она говорит о том, что при наличии показателя ЛСО переводчики стараются его переводить соответствующим показателем ЛСО. Это обстоятельство важно и при анализе машинных переводов и оценки их качества.

В НБД все виды дивергентного перевода распределены по четырем кластерам: Прочие (лексические), Прочие (грамматические), Пунктуация, Отрицание. В кластер Прочие (лексические) входят, как говорит его название, лексические средства передачи ЛСО. Например, сравнительные ЛСО часто передаются предикатами, выражающими сравнение, как в (3), где сравнительный коннектор *так||как* заменен в переводе на предикат *être semblable à* ‘быть похожим’.

- (3) Она пела так чисто, так правильно и вместе *так... так... как* поют все девицы, когда их просят спеть в обществе; без увлечения. [И. А. Гончаров. Обломов (1848-1859)]
 Son chant, si pur, si juste, était à la fois si... *semblable au* chant de toutes les autres jeunes filles quand on les prie de chanter en société: sans aucune émotion. [Tr. L. Jurgenson]

В кластер Прочие (грамматические) входят синтаксические конструкции, глагольные формы деепричастий и причастий и др. Так, в (4) временное ЛСО, выражаемое *когда||то*, передано деепричастием настоящего времени, также устанавливающим одновременность действий, описанных деепричастием (*en regardant bien* ‘рассмотрев’) и предикатом *s’aperçut* ‘заметили’.

- (4) Но *когда* разглядели хорошенько Катерину Ивановну, *то* увидели, что она вовсе не разбилась о камень. [Ф. М. Достоевский. Преступление и наказание (1866)]
 Mais *en regardant bien* Catherine Ivanovna, on s’aperçut qu’elle ne s’était nullement blessée contre une pierre. [Tr. É. Guertik]

В кластер Пунктуация входят семантически насыщенные знаки препинания: тире и двоеточие. Они совместимы с ограниченным набором ЛСО, а значит могут считаться средством их выражения. Двоеточие, например, часто передает причинные отношения, как в русском, так и во французском языках, и его переводчик использует вместо коннектора *потому что* в (5):

- (5) А вы, матушка, и времени даром не теряйте, закажите ему теперь же сосновый гроб, *потому что* дубовый будет для него дорог. [Н. В. Гоголь. Шинель (1842)]
 Allons, ma bonne dame, ne perdez pas votre temps inutilement; allez vite commander un cercueil de sapin: le chêne serait trop cher pour lui. [Tr. H. Mongault]

В кластер Отрицание изначально заносились коннекторы, которые включают в свой состав отрицание (*не то чтобы||а, не только||но* и др.), а затем, по мере наполнения НБД и фиксации таких случаев, также отрицательные частицы (*ne|pas, ne|que*) или языковые единицы, в состав которых они входят (*ne pas empêcher* ‘не мешать’, *ne|que gérondif présent* ‘не_деепричастие настоящего времени’). Дивергентными средствами перевода ЛСО, входящими в кластер Отрицание, считаются только последние. При этом кластеры Прочие (лексические), Прочие (грамматические), с одной стороны, и Отрицание, с другой, являются пересекающимися, т.е. языковые единицы, приписанные к двум кластерам Прочие, могут быть одновременно приписаны и к кластеру Отрицание. Например, *je n’irais pas jusqu’à dire* ‘букв. я не дойду до того, чтобы утверждать’, передающее ЛСО замещения, выраженное в оригинале коннектором *не то чтобы*, приписано к кластерам Прочие (лексические) и Отрицание.

- (6) *Не то чтобы* меня выбросили из ресторана. Я выполз сам, окутанный драпировочной тканью. [Сергей Довлатов. Заповедник (1983)]

³ Это направление перевода выбрано как располагающее наиболее представительным массивом аннотаций в НБД, но описанный выше подход к дивергентному переводу может быть применен и для анализа большего количества языков перевода. См. примеры такого анализа в [Кобозева, Инькова 2018] и [Инькова 2018].

Je n'irais pas jusqu'à dire qu'on m'a jeté à la porte du restaurant. J'ai rampé dehors, drapé dans le store. [Tr. Ch. Zeytounian-Beloüs]

В таблице 1 приводятся данные о распределении дивергентных средств перевода ЛСО по четырем кластерам. Они показывают, что более 55% приходится на лексические средства.

Кластер	Всего МЭ	Для данного кластера	В %
Прочие (грамм.)	903	219	24,2%
Прочие (лекс.)	903	520	57,6%
Пунктуация	903	28	3,1%
Отрицание	903	136	15%

Таблица 1: Распределение дивергентных переводов коннекторов по кластерам

Интересен тот факт, что некоторые дивергентные средства перевода ЛСО являются специализированными, т.е. выражают только одно ЛСО, другие, в первую очередь, грамматические, обслуживают сразу несколько ЛСО. Так, наибольшим разнообразием передаваемых ЛСО отличается определительное придаточное, которое использовано как функциональный эквивалент показателей 15 ЛСО⁴: аддитивные иллокутивные (9 аннотаций); переформулирование и спецификация (по 5 аннотаций); сопоставительные, «вопреки ожидаемому», пропозициональные причинные, сравнительные, пропозициональное сопутствование (по 3 аннотации); условные, уступительные, иллокутивное сопутствование и иллокутивные причинные (по 2 аннотации); аддитивные пропозициональные, соединительные и противопоставление (единичные употребления). Ср. (7), где оба переводчика передают ЛСО, выражаемое *потому что*, относительным придаточным, в варианте а. в сочетании с местоименным повтором.

- (7) Коллежский ассессор был в этом сведущ *потому, что* был послан несколько раз на следствие еще в Кавказской области. [Н. В. Гоголь. Нос (1832-1833)]
 а. L'assesseur de collège était savant en la matière, *lui qui avait été chargé d'instruire maintes affaires criminelles lors de son séjour au Caucase.* [Tr. B. De Schloezer]
 б. L'assesseur, *qui avait procédé au Caucase à plus d'une enquête*, s'entendait en ces matières. [Tr. H. Mongault]

Далее по степени полисемичности следуют деепричастия настоящего времени, которые, в том числе в сочетании с местоимением *tout* 'всё', подчеркивающим одновременность действий, могут быть использованы для выражения 11 ЛСО: пропозициональное сопутствование (16 аннотаций); временные (11 аннотаций); соединительные (10 аннотаций); уступительные (8 аннотаций); «вопреки ожидаемому» (7 аннотаций); иллокутивное сопутствование (4 аннотации); условные и сопоставительные (по 2 аннотации); аддитивные пропозициональные и иллокутивные, противительно-уступительные (единичные употребления). Ср. (4) для временных отношений.

Среди лексических дивергентных средств в наибольшем спектре ЛСО задействовано наречие *même* 'даже', привносящее градационный оттенок. Оно зафиксировано для передачи 9 ЛСО: единственности (5 аннотаций); противительно-уступительные иллокутивные, уступительные, аддитивные пропозициональные (по 2 аннотации); аддитивные иллокутивные, аналогия, коррекция, пропозициональная альтернатива, спецификация (единичные употребления). Ср. (8) для противительно-уступительных иллокутивных ЛСО в обоих вариантах перевода.

- (8) Предположение, *хотя* легкое, шуточное, что она может быть несчастлива, неожиданно вызвало ее на откровенность. [И. А. Гончаров. Обломов (1848-1859)]
 а. La supposition, *même* faite à la légère et en plaisantant, qu'elle pouvait être malheureuse, la poussa soudain à faire des aveux. [Trad. L. Jurgenson]

⁴ В НБД и в работе используется классификация, разработанная О.Ю. Иньковой [Инькова 2019b], исходящая из базовой семантической операции, лежащей в основе того или иного ЛСО и различающая семантические уровни, на которых они могут быть установлены: пропозициональный, уровень высказывания (иллокутивный) и метаязыковой.

б. Cette hypothèse, *même* formulée à la légère, par plaisanterie, qu'elle pût être malheureuse, la poussait à avoir confiance. [Trad. A. Adamov]

В таблице 2 мы приводим распределение дивергентных переводов по ЛСО. В ней представлены только те ЛСО, для которых в НБД сформировано более 50 аннотаций⁵.

ЛСО	Всего	Диверг.	%
исключение	277	127	45,8%
исключение из рассмотрения	69	26	37,7%
пропозициональное сопутствование	126	42	33,3%
замещение	220	73	33,2%
аналогия	102	18	17,6%
тождество	65	9	13,8%
коррекция	277	31	11,2%
спецификация	616	59	9,6%
возмездительное противопоставление	131	12	9,2%
пропозициональная причина	363	33	9,1%
переформулирование	777	69	8,9%
сравнительные	307	27	8,8%
сопоставительные	373	31	8,3%
аддитивные иллокутивные	519	43	8,3%
временные	644	53	8,2%
соединительные	344	27	7,8%
иллокутивная причина	221	16	7,2%
иллокутивное сопутствование	211	15	7,1%
следствие	57	4	7,0%
условные	623	38	6,1%
уступительные	608	37	6,1%
неединственности	398	24	6,0%
противительные- уступительные	74	4	5,4%
аддитивные пропозициональные	272	12	4,4%
экстенсивная генерализация	103	4	3,9%
несоответствие	130	5	3,8%
“вопреки ожидаемому”	1373	46	3,3%
“вопреки ожидаемому” иллокутивные	543	19	3,3%
временные метаязыковые	126	4	3,2%
отрицательная альтернатива	190	5	2,6%
пропозициональная альтернатива	577	15	2,6%
уступительные иллокутивные	76	2	2,6%
контраст	54	1	1,8%

Таблица 2: Распределение дивергентных переводов коннекторов по ЛСО

⁵ По сравнению с данными из [Инькова 2019а], увеличение количества аннотаций в НБД и более детальная классификация ЛСО позволяет дать более полную панораму ЛСО и более точное представление о возможностях их перевода дивергентными средствами.

Лидируют в этом списке ЛСО исключения (почти 50% дивергентных переводов), исключения из рассмотрения, пропозициональное сопутствование и замещение, для которых доля дивергентных переводов превышает 33%. На противоположном полюсе находится ЛСО контраста с менее чем 2% дивергентных переводов. Для основной массы ЛСО доля дивергентных переводов находится в зоне 2-10%.

Поскольку, как мы видели, при наличии показателя ЛСО переводчики стараются его перевести соответствующим показателем ЛСО, то важно проанализировать причины, по которым переводчик выбирает дивергентный способ передачи ЛСО. Анализ данных позволил выявить следующие причины:

- *Отсутствие точного эквивалента коннектора в языке перевода.* Это наиболее очевидная причина выбора дивергентного перевода. Здесь, однако, следует различать два случая.
 - В обоих языках есть дивергентные средства выражения данного ЛСО, в языке перевода дивергентное средство является единственно возможными. Это случай ЛСО пропозиционального сопутствования: в обоих языках оно может быть выражено деепричастием настоящего времени, но в русском для него есть также коннектор *при этом*.
 - В языке перевода нет коннектора, являющегося точным эквивалентом, и используются близкие по семантике коннекторы или дивергентные средства; ср. в (2) варианты перевода а. и б.
- *Различия в синтаксической структуре сопоставляемых языков.* Этот случай касается, прежде всего, отношения исключения, выражаемого *кроме*. В том случае, когда множество, из которого исключается вводимый им элемент, задано отрицательным местоимением, во французском языке вместо предлога *sauf* 'кроме' используется семантически пустой союз *que* 'что'.

(9) Они все против, а мне *никто, кроме* тебя, не нужен. [Светлана Алексиевич. Время секунд хэнд]

Иls sont contre. Mais moi, je n'ai besoin de *personne d'autre que* toi! [Tr. S. Benech]

Различиями в синтаксической структуре языков объясняется и высокая доля дивергентных средств при переводе ЛСО аналогии (17,6%), поскольку соотносительные конструкции, при помощи которых оно выражается, как и сравнительные ЛСО, менее частотны во французском языке и в целом представляют значительные трудности для перевода; подробнее см. [Inkova 2014]. Ср.(10), где при переводе сравнительных отношений на французский язык невозможно использовать коррелятивную структуру.

(10) она делает это с такой обидной снисходительностью, *так* тихо, *как* делают только с детьми или с совершенными дураками. [И. А. Гончаров. Обломов (1848-1859)]

elle le faisait avec une condescendance si blessante, avec *cette douceur qu'*on a avec les enfants ou les idiots complets. [Tr. L. Jurgenson]

- *Различия в структуре коннекторов.* Появление дивергентного средства перевода может быть связано с тем, что переведена одна из частей двухместного коннектора, как правило, первая и содержащая отрицание, или один из элементов многоэлементного коннектора, как правило, конкретизатор, в терминологии русской грамматики, т.е. семантически более насыщенный компонент. Ср. (11), где семантика двухкомпонентного коннектора коррекции *ne||a* передана при помощи отрицания (при наличии во французском языке двухкомпонентного коннектора *ne|pas||mais*).

(11) он никакой *не* интурист, *а* шпион. [М. А. Булгаков. Мастер и Маргарита (1929-1940)]

Ce n'est *pas* du tout un touriste. C'est un espion. [Tr. Cl. Ligny]

- *Различия в узусе.* К таким случаям относятся коннекторы замещения по предпочтению, которые существуют в обоих языках, но во французском языке семантика этого ЛСО чаще

передается глаголами предпочтения (*préférer, aimer mieux* ‘предпочитать’, безличная форма *valoir mieux* ‘стоит лучше’ и др.); см. (12).

- (12) *Скорее* соглашусь умереть, – сказал я в бешенстве, – *нежели* уступить ее Швабрину!
[А. С. Пушкин. Капитанская дочка]
– *J’aimerais mieux mourir, dis-je avec fureur, que de la céder à Chvabrine.* [Tr. L. Viardot]

- *Контекстуально обусловленное использование дивергентных средств.* Использование дивергентного перевода может быть связано с особенностями конкретного контекста, где по тем или иным причинам невозможно употребление коннектора, что часто приводит к значительному изменению по сравнению с текстом оригинала.

Подчеркнем, что во всех перечисленных случаях, кроме последнего, дивергентное средство перевода является одновременно и «альтернативной лексикализацией» ЛСО в языке перевода. Поэтому изучение этих явлений представляет интерес не только для сопоставительного изучения языков, но и для изучения состава языковых средств, которые могут использоваться для выражения ЛСО.

3 Дивергентные средства выражения ЛСО в машинных переводах

Материалом для исследования дивергентных переводов, выполненных машинным переводчиком⁶, послужили 7631 двуязычная аннотация в направлении перевода русский-французский. Из них с пометой «Дивергентное межъязыковое соответствие» (Dvrg) – 180, т. е. 2,4%, что указывает на корреляцию с выявленной выше тенденцией, характерной для выполненных человеком переводов: наличие показателя ЛСО в исходном тексте порождает реализацию соответствующего показателя ЛСО и в переводе.

В НБД, как и в случае «человеческого» перевода, виды дивергентного машинного перевода распределены по четырем основным кластерам: Прочие (лексические), Прочие (грамматические), Пунктуация, Отрицание. Однако наряду с этими четырьмя классами для квалификации машинного перевода также использовался кластер Прочие без спецификации (лексические/грамматические). Сюда заносились переводные варианты, реализованные в контекстах, машинный перевод которых в целом содержит многочисленные ошибки, иногда не позволяя вообще установить какие-либо параллели с исходным фрагментом текста, см. (13), где *да еще*, показатель аддитивных пропозициональных ЛСО, передается утвердительным словом *oui* (‘да’).

- (13) Из всех присутствующих я узнал только музыковеда Лазарева, *да еще* фарцовщика Беллугу. [Сергей Довлатов. Заповедник (1983)]
– *De toutes les personnes présentes, j’ai appris que musicologue Lazarev, oui, фарцовщика Беллугу.* [Яндекс.Переводчик (14.11.2016, 16:38)]

Кластер	Всего МЭ	Для данного кластера	В %
Прочие (грамм.)	180	49	27,2%
Прочие (лекс.)	180	58	32,2%
Прочие	180	68	37,8%
Пунктуация	180	0	0%
Отрицание	180	5	2,8%

Таблица 3: Распределение дивергентных машинных переводов коннекторов по кластерам

Данные, приведенные в таблице 3, свидетельствуют о том, что системы машинного перевода, в отличие от человека-переводчика, вообще не используют пунктуационные возможности

⁶ В НБД хранятся аннотации машинных переводов, выполненных двумя системами – «Яндекс.Переводчик» (<https://translate.yandex.ru/>) и «Google Переводчик» (<https://translate.google.com/>).

дивергентного перевода⁷. По всей видимости, это связано, в первую очередь, со структурой тренировочных материалов, задействованных для обучения систем автоматического перевода, где количество таких переводных примеров (преимущественно заимствованных из художественных переводов) оказывается нерелевантным и не принимается в расчет.

ЛСО	Всего	Дивергент.	%
исключение из рассмотрения	12	6	50,00 %
переформулирование	267	52	19,48%
аналогия	419	29	6,92%
аддитивные иллокутивные	246	13	5,28%
причина	323	17	5,26%
пропозициональное сопутствование	124	6	4,84%
временные	532	21	3,95%
спецификация	274	8	2,92%
иллокутивное сопутствование	233	4	1,72%
замещение	161	2	1,24%
“вопреки ожидаемому”	517	6	1,16%
аддитивные пропозициональные	495	5	1,01%
сопоставительные	102	1	0,98%
экстенсиональная генерализация	288	2	0,69%
коррекция	773	4	0,52%
уступительные	387	2	0,52%
соединительные	239	1	0,42%
неединственности	1375	1	0,07%

Таблица 4: Распределение дивергентных машинных переводов коннекторов по ЛСО

В таблице 4 приводится распределение дивергентных машинных переводов по ЛСО. Панорама ЛСО здесь отличается от той, что представлена в таблице 2. Список ЛСО значительно меньше, и лидируют в нем ЛСО исключения из рассмотрения (доля дивергентных переводов 50%) и переформулирования (доля дивергентных переводов превышает 19%). На противоположном полюсе находится ЛСО неединственности с 0,07% дивергентных переводов. Для основной массы ЛСО доля дивергентных переводов находится в зоне 0,42-7%⁸.

Анализ машинных переводов позволяет выявить три группы случаев, когда реализуется дивергентный вариант.

- Система машинного перевода не распознает неоднословный коннектор как единую языковую единицу.

(14) *Как письмо прочел, так и пошел...* [Ф. М. Достоевский. Преступление и наказание (1866)]

En lisant la lettre, je suis allé... [Google Translate (03.09.2019, 15:06)]

В (14) при переводе контекста, содержащего показатель ЛСО аналогии – коннектор *как||так и*, вводится герундиальный оборот *en lisant*. Вариант, зафиксированный в машинном переводе, показывает, что, во-первых, синкретичное смысловое содержание, о котором сигнализирует коннектор (аналогия + временные ЛСО), интерпретируется в пользу временного толкования, и, во-вторых, временные отношения передаются ошибочно (предшествование заменяется одновременностью). Система машинного перевода не воспринимает двухкомпонентный коннектор как

⁷ Во всяком случае, пока таких переводных вариантов зафиксировано нами не было.

⁸ Очевидно, что представленная в таблице 4 панорама ЛСО будет изменяться по мере дальнейшего наполнения НБД. При актуальной структуре данных наблюдается дисбаланс в соотношении аннотированных машинных переводов для контекстов, где зафиксированы показатели тех или иных ЛСО.

единый комплекс, каждая часть которого участвует в смыслообразовании, реагируя только на один компонент. Перевод ЛСО аналогии представляет определенные трудности и для автоматического переводчика в направлении русский-французский.

- Система машинного перевода не различает разные употребления полифункциональных языковых единиц (как в (13))⁹;
- Случаи, когда использование дивергентных средств может объясняться так же, как и в переводе, выполненном человеком. Оно мотивировано различиями в синтаксической структуре языков и/или узуальными предпочтениями, которые отражаются в тренировочных данных, применяемых для обучения системы машинного перевода¹⁰.

- (15) *Так как* я знал, что заботы матушки о моих занятиях ограничатся этими немногими словами, то я и не почел нужным возражать ей... [И. С. Тургенев. Первая любовь (1860)]
Sachant que les préoccupations de ma mère au sujet de mes études se limiteraient à ces quelques mots, je n'ai pas jugé nécessaire de s'y opposer... [Google Translate (03.10.2019, 21:56)]

В (15) причинный коннектор *так как* переводится причастной конструкцией *sachant que*, что является вполне естественным для французского языка. В русском языке также возможно употребление деепричастия в данном контексте (ср. *Зная, что...*), вместе с тем при редактировании русского художественного текста, например, бытует установка по возможности избегать причастных и деепричастных оборотов.

4 Заключительные замечания

Дальнейшая разработка понятий «альтернативная лексикализация» и «дивергентные средства выражения дискурсивных отношений» позволит, на наш взгляд, провести более четкую границу между эксплицитно выраженными и имплицитными дискурсивными отношениями (см., например, [Martin 1992]; [Renkema 2004]; [Taboada 2009], где к имплицитным относятся все дискурсивные отношения, не выраженные коннекторами), а также между коннекторами, прототипическими показателями дискурсивных отношений, и другими средствами их выражения (ср. в этой связи довольно широкую трактовку понятия коннектор в [Toldova et al. 2018]). Данные, полученные в аннотированных корпусах, как одноязычных, так и многоязычных и параллельных, помогут, в свою очередь, составить более четкое представление о том, какие дискурсивные отношения могут быть выражены языковыми средствами, не принадлежащими к классу коннекторов, а какие – нет, каковы причины появления дивергентных средств выражения дискурсивных отношений в тексте и какова их частотность. Полученные результаты могут быть использованы как при автоматической обработке и генерации текста, так и – в первую очередь, данные о дивергентных переводах дискурсивных отношений – для улучшения качества машинного перевода. В отношении последнего было бы интересно на примере отдельно взятой системы машинного перевода проследить во времени, как изменяется доля дивергентных переводных вариантов по мере наращивания эпох обучения, проявляется ли тенденция к уменьшению этой доли или, наоборот, к ее увеличению.

References

- [1] Das D., Taboada M. (2013) Explicit and Implicit Coherence Relations: A Corpus Study. Proceedings of the 2013 annual conference of the Canadian Linguistic Association, available at: http://homes.chass.utoronto.ca/~cla-acl/actes2013/Das_and_Taboada-2013.pdf.
- [2] Das D., Taboada M. (2014) RST Signalling Corpus Annotation Manual, available at: https://www.sfu.ca/~mtaboada/docs/publications/RST_Signalling_Corpus_Annotation_Manual.pdf.
- [3] Inkova O. Yu. (2018) *Voobshche* [In general], O. Inkova (ed.), Semantics of connectives: a contrastive study [Semantika konnektorov: kontrastivnoe issledovanie]. Moscow: TORUS PRESS. Pp. 80–128.

⁹ После внедрения технологий нейронного машинного перевода частотность таких случаев существенно снизилась.

¹⁰ В статье зафиксированы результаты наблюдения за использованием дивергентных средств выражения ЛСО в машинных переводах, но при этом если и называются возможные причины реализации дивергентных вариантов, то только в качестве предположения. Как правило, на реализацию такого варианта в машинном (как статистическом, так и нейронном) переводе влияет не один, а целый ряд факторов.

- [4] Inkova O. Yu. (2019a) Annotirovanie parallel'nykh tekstov: ponyatie "divergentnyi perevod" [Annotation of parallel texts: the concept of divergent translation]. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2019". Moscow, May 29–June 1, 2019. <http://www.dialog-21.ru/media/4600/inkovaoyu-019.pdf>.
- [5] Inkova O. Yu. (2019b) Logical-semantic relations: classification problems [Logiko-semanticheskie otnosheniya: problemy klassifikatsii], O. Inkova, E. Manzotti, Text coherence: mereological logical-semantic relations [Svyaznost' teksta: mereologicheskie logiko-semanticheskie otnosheniya]. Moscow: Izdatel'skii Dom YaSK. Pp. 11–98.
- [6] Inkova O. (2021) La sémantique des connecteurs: méthodes quantitatives d'analyse [Semantics of connectives: quantitative methods of analysis]. Bern / Berlin: Peter Lang. 276 p.
- [7] Johansson S. (2007) Seeing through multilingual corpora: On the use of corpora in contrastive studies. Amsterdam / Philadelphia: John Benjamins. 377 p.
- [8] Kobozeva I.M., Inkova O. Yu. (2018) *Kak i ego dvukhmestnye ekvivalenty* [*Kak 'As' and its two-word equivalents*], O. Inkova (ed.), Semantics of connectives: a contrastive study [Semantika konnektorov: kontrastivnoe issledovanie]. Moscow: TORUS PRESS. Pp. 168–239.
- [9] Martin J. R. (1992) English Text: System and Structure. Amsterdam / Philadelphia: John Benjamins. 620 p.
- [10] Prasad R., Joshi A., Webber B. (2010) Realization of Discourse Relations by Other Means: Alternative Lexicalizations. Proceedings of the 23rd International Conference on Computational Linguistics (Beijing, China – August 23–27, 2010): Posters. Pp. 1023–1031.
- [11] Renkema J. (2004) Introduction to Discourse Studies. Amsterdam / Philadelphia: John Benjamins. 363 p.
- [12] Taboada M. (2009) Implicit and explicit coherence relations. Discourse, of Course. An overview of research in discourse studies. J. Renkema (ed.). Amsterdam / Philadelphia: John Benjamins. Pp. 127–140.
- [13] Taboada M., Das D. (2013) Annotation upon annotation: Adding signalling information to a corpus of discourse relations. Dialogue and Discourse, Vol. 4, No. 2. Pp. 249–281.
- [14] Toldova S., Pisarevskaya D., Kobozeva M., Vasilyeva M. (2018) The cues for rhetorical relations in Russian: "cause–effec" relation in Russian rhetorical structure treebank. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2018". Moscow, May 30–June 2, 2018. <http://www.dialog-21.ru/media/4338/toldovas.pdf>.