

ПРОБЛЕМА ЭКВИВАЛЕНТНОСТИ ЯЗЫКОВЫХ СТРУКТУР ПРИ ПЕРЕВОДЕ И СЕМАНТИЧЕСКОМ ВЫРАВНИВАНИИ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ

THE LANGUAGE STRUCTURES EQUIVALENCE PROBLEM IN TRANSLATION AND SEMANTIC ALIGNMENT OF PARALLEL TEXTS

Е. Б. Козеренко (kozerenko@mail.ru)

Институт проблем информатики РАН

Проблема эквивалентности языковых структур в исходном тексте и тексте перевода рассматривается с точки зрения а) выработки методик перевода ряда языковых реалий, характерных для научно-технических текстов, и б) создания корректных алгоритмов семантического выравнивания параллельных текстов на уровне значимых компонентов. Исследования проведены на материале русскоязычной и англоязычной научной периодики. Особое внимание уделяется переводу безличных и неопределенно-личных конструкций русского языка на английский язык, оборотов с неличными формами глаголов английского языка на русский язык, и других структур, наиболее частотных для научных текстов. Важное место в исследованиях занимает проблема перевода метафор англоязычного научного текста.

1. Введение

Современный этап исследований и разработок в области машинного перевода и обработки естественного языка требует применения новых методов представления лингвистической реальности, отражающих сложный характер естественных языков и, вместе с тем, сочетающих возможности уже существующих подходов. Формирование корпусов параллельных текстов дает возможность использовать машинное обучение в дополнение к традиционным методам на основе эвристических правил, составленных человеком-экспертом. Однако, для того, чтобы было возможно автоматически извлекать правила из параллельных текстов, необходимо разработать достаточно тонкие методы сопоставления лингвистических объектов, при этом главным основанием должны быть их смысловые характеристики.

До сих пор семантическое выравнивание параллельных текстов рассматривалось, в основном, с точки зрения построения лексико-семантических моделей, в нашем же случае речь идет о семантике структурных компонентов, т.е. способов конфигурирования языковых объектов (в семиотической лингвистике они получили название *структурных знаков* [1]). Эти исследования мотивированы тем обстоятельством, что применяемые методы автоматического выравнивания параллельных текстов пословно или по предложениям приводят к возникновению большого числа шумов, которые возможно устранить только вручную на этапе постредактирования.

В чем же причина возникновения подобных шумов? Прежде всего, в корпусах параллельных текстов собраны уже готовые переводы, и заранее невозможно предсказать, каким именно способом будет выражено некоторое значение в реальном тексте. Качественно выполненный перевод неизбежно содержит значительное число переводческих трансформаций, которые необходимо не только регулярным образом описать, но и формализовать для последующей алгоритмизации и введения в программный код разрабатываемых систем.

В данной работе рассматривается ряд конструкций, типичных для текстов научной и деловой прозы, которые приводят к трудностям при русско-английском и англо-русском переводе и автоматическом сопоставлении параллельных текстов. Какие же это конструкции? Прежде всего, это безличные и неопределенно-личные предложения, часто употребляемые в научной прозе на русском языке; экзистенциальные конструкции с формальным подлежащим английского языка; обороты с неличными формами глаголов в текстах на английском языке, особенно контринтуитивные для носителей русского языка абсолютные причастные обороты в английском языке. Рассматриваются наиболее типичные структурные трансформации при их переводе на русский язык (например, неличные конструкции в клаузы; нефинитная форма глагола -> финитная форма, и т.п.). Отдельное внимание уделяется проблемам сегментации предложений при переводе и некоторым наиболее типичным способам перевода метафор в научном тексте.

Суть метода, который применяется в наших исследованиях, состоит в интегральном подходе к представлению языковых структур и объектов в виде системы функционально-семантических полей - для многоязычной ситуации мы их назвали поля функционально-семантического переноса (ПФП), [2] - включающих

синонимичные языковые средства, которые выражают одни и те же типы значений. В настоящее время наши усилия сосредоточены на развитии системы категоризации лингвистических объектов (компонентов высказывания).

2. Общие характеристики текста научной прозы

Очевидно, что можно говорить только об относительной эквивалентности исходного текста и текста перевода. В нашей работе мы исходим из представления, что главное в любом переводе — это передача смысловой информации текста [3,4]. Все остальные ее виды и характеристики, функциональные, стилистические, эмоциональные, не могут быть переданы без воспроизведения смысловой информации, так как все остальное содержание компонентов сообщения наслаивается на смысловую информацию. Степень относительной эквивалентности научных текстов зависит от их типов и видов. Замечено, что чем более формализован научный текст, а это происходит, прежде всего, в естественных науках, тем ближе его перевод оригиналу. Переводы некоторых работ по математике, химии или биологии, состоящих из стереотипных фраз, вводящих соответствующие формулы, оказываются практически тождественными оригиналу.

Однако переводы работ по гуманитарным наукам — истории, социологии, философии — содержат весьма значительное число конструкций, специфических для исходного языка и претерпевающих весьма значительные трансформации при передаче смысла средствами другого языка. При переводе научных трудов главное — передать мысль, ее логику, суть научной доктрины, последовательность рассуждения. Для этого нередко переводчик меняет синтаксический строй фраз исходного текста, снижает эмоциональную тональность, если она есть в оригинале.

Уровень относительной эквивалентности переводов научной литературы обуславливается некоторыми грамматическими трансформациями, логическими и терминологическими уточнениями и разъяснениями, которые зависят от характера научного труда и прагматических требований к переводу. В научно-технической речи используются те же самые синтаксические структуры и морфологические формы, как и в других функциональных стилях, однако ряд грамматических явлений отмечается в данном стиле чаще, чем в других, некоторые явления, напротив, встречаются в нем сравнительно редко.

3. Перевод безличных глаголов и предложений; проявление субъекта

Характерным для русскоязычного научного текста является употребление безличных и неопределенно-личных предложений. Эти конструкции не имеют полных структурных аналогов в английском языке и, в основном, передаются предложениями, в которых появляется субъект предложения в позиции первой именной группы. При этом особо следует выделить предложения с формальным подлежащим «it» и конструкции с «there to be», которые в некотором смысле можно уподобить безличным конструкциям русского языка, с той лишь разницей, что в русском языке некоторое загадочное «оно» подразумевается, а в английском языке «оно» — «it» присутствует явным образом. Что же может скрываться за формально выраженным субъектом?

По мнению Степанова Ю. С. [5], в безличном предложении типа русского *Морозит* субъектом будет нечто, определяемое словом «погода». Локус-предложения, или «оно»-предложения, в которых субъект обозначает — в широком смысле — место (или время) в объективном мире: *Сегодня холодно* и др. К подобной трактовке субъекта склоняются исследователи в области теоретической грамматики русского языка Шмелев Д. Н. [6] и Золотова Г. А. [7].

Если рассмотреть безличные предложения русского языка и их возможные переводы на английский язык, т. е. двуязычную ситуацию, то можно заметить «проявление» субъекта. При этом семантические характеристики этих субъектов могут быть различными, но в значительной степени они определяются безличными глаголами. В грамматике [8-10] выделяют следующие самые общие семантические группы безличных глаголов.

1. Безличные глаголы, которые обозначают природные явления.

- (1) *Морозит. It freezes. The day is cold. It is frosty.*
- (2) *Вечерело. Dusk was falling. It was getting dark.*
- (3) *Похолодало. It got colder.*
- (4) *На террасе дуло. There was draught on the terrace.*

2. Безличные глаголы, которые обозначают состояние человека.

- (5) *Легко дышится в лесу. It is easy to breath in the wood.*
- (6) *Его лихорадило. He was feverish.*
- (7) *Его знобило. He was shivering.*
- (8) *Ему нездоровилось. He felt unwell.*
- (9) *Мне больно. It hurts me.*
- (10) *Ей грустно. She is sad.*
- (11) *Ему было тревожно. He felt uneasy.*
- (12) *Ему не спалось. He could not sleep.*

Рассмотрим варианты перевода, полученные с помощью двух современных систем МП:

(13) * *To it it was not slept;*

(14) * *Didn't be able to sleep to him.*

Эти переводы удивительным образом совпадают с типичными переводами людей, изучающих английский язык на начальных стадиях – имеются в виду попытки сделать грамматические кальки безличных конструкций русского языка.

(15) *В коммуникативной грамматике [11] эти модели (6-12) предложений рассматриваются как имеющие субъект, который выражен формой дательного или винительного падежа.*

Рассмотрим два варианта перевода данной фразы, сделанные носителем английского языка.

(16) *Communicative Grammar treats these sentence models as having a subject expressed by the forms of the Dative or Accusative cases.*

(17) *In Communicative Grammar these sentence models are treated as sentences having a subject expressed by the forms of the Dative or Accusative cases.*

Интересно, что с точки зрения методики синхронного перевода более предпочтительным является вариант (16) [4,12], т.е. переводческая трансформация

NP (в косвенном падеже) -> NP (субъект английского предложения).

Модели предложений, которые в традиционной грамматике классифицируются как безличные, характеризуются в коммуникативной грамматике как инволютивные. Признак инволютивности является определяющим для объединения в одну группу предложений, в которых сентенциальный предикатор обозначает действие (или состояние), совершаемое (или имеющее место) помимо воли субъекта этого предложения.

3. Конструкции с бытийными глаголами, выражающие наличие/отсутствие чего-либо, соотношенное с некоторым локусом.

(18) *Всего этого в российской истории было всегда в избытке. Russian history has always had a lot of it.*

4. Безличные глаголы, которые обозначают долженствование.

(19) *Вам следует оставаться дома. You should stay at home.*

(20) *Тебе стоит принять это предложение. You ought to accept this proposal.*

Последний случай является наиболее сложным с точки зрения выдвижения гипотез о возможном заполнении субъектной роли. Здесь необходимо обратиться к результатам в области построения глубинно-семантических представлений.

Инвариантное представление семантики модального предиката долженствования дается в работах И.М. Кобозевой и Н.И. Лауфер [14,15]. Лингвистические экспликации значений модальных операторов приводятся в работе А.А. Зализняк и Е.В. Падучевой [16].

В приводимых ниже примерах русско-английского перевода используются лексические корреляты модальных конструкций.

(21) *Исходя из требований жизни, надо по-новому взглянуть на некоторые теоретические представления и концепции.*

Life prompts us to take a new look at some theoretical ideas and concepts.

(22) *К поиску ответов на решение многих проблем приходилось возвращаться неоднократно, и всякий раз открывались все новые их грани, рождались новые, зачастую исключаящие друг друга варианты.*

Again and again we returned to the same problems, and every time we saw new aspects of them and came up with new often mutually exclusive solutions.

4. Транскатегориальные соответствия

Отсутствие полного совпадения между английскими и русскими языковыми конструкциями в научно-технических текстах можно обнаружить при изучении сравнительной частоты употребления в них отдельных частей речи, что важно для построения систем перевода, использующих машинное обучение [13]. Для научного изложения в целом характерен признак номинативности, т.е. более широкое использование существительных, чем в других функциональных стилях. При этом сопоставительный анализ переводов показывает, что в русском языке эта тенденция выражена более четко, и при переводе английские глаголы нередко заменяются существительными.

(23) *The engine is the source of power that makes the wheels go round and the car move.*

Двигатель служит источником энергии для вращения колес и движения автомобиля.

(24) *The fuel system is designed to store liquid gasoline and to deliver it to the engine cylinders in the form of vapor mixed with air.*

Система питания предназначается для заправки жидким топливом и подачи его в цилиндры в виде смеси паров бензина с воздухом.

(25) *A similar approach has marked the EU's efforts to expand the current club of 15 countries to embrace former communist countries further east.*

Точно таким же подходом характеризуются усиления ЕС по расширению нынешнего клуба 15 стран дальше на восток путем присоединения к нему бывших коммунистических стран.

(26) *Новизна задач, а вместе с нею и трудности их на этом не ограничиваются.*

It is not only this that makes their tasks novel and difficult.

5. Структурные расширения при переводе

При переводе могут происходить расширения конструкций. Например, развертывание свернутого предиката в клаузу:

(27) *Все это приходилось учитывать в ходе подготовки концепции.*

We had to take this into account as we worked out the basic concept.

Экспликация при переводе именных конструкций:

(28). *low-pressure producers - производители полиэтилена методом низкого давления.*

6. Членение высказываний

При появлении в исходном высказывании сразу нескольких предметных ситуаций переводной эквивалент подвергается членению: производится сегментация длинных предложений на два и более. В английской фразе в роли подлежащего чаще всего выступает тот же субъект, выраженный не именем, а личным местоимением:

(29) *Народы, широкие круги общественности горячо желают изменения положения дел к лучшему, хотя и учиться сотрудничать.*

Nations and the public at large ardently wish to see improvements. They want to learn to cooperate.

(30) *Оттого-то у нас до сих пор подозрительно относятся к иронии, что видят в ней нарушительницу серьезного взгляда на литературу как общественного просветителя.*

This is why our critics are still rather suspicious of irony. They believe it violates the serious view of literature as a social enlightener.

Препозитивные причастные и обособленные обстоятельственные обороты регулярно сегментируются: такие обороты желательно передавать самостоятельным отдельным предложением.

(31) *Никуда не уйти от ставшего условием выживания и прогресса требования находить баланс интересов в международных отношениях.*

There is no getting away from the necessity to balance interests on the international scene. Their balance is vital to survival and progress.

Рассмотрим пример перевода высказывания с обособленным обстоятельственным оборотом.

(32) *Попытки управлять творчеством, тем более явно не чистыми руками, морально и социально разрушительны.*

Attempts to streamline creative quest by fiat are morally and socially destructive. This is all the more so if such attempts are made by far from impeccable people.

7. Компрессия при переводе

При переводе может происходить структурное сжатие языковых конструкций, например, словосочетание в русском языке -> слово в английском языке.

(33) *Встреча на высшем уровне в Вашингтоне выявила существенные перемены в общественном мнении.*

The summit in Washington threw light on changes in public opinion.

Объединение высказываний при переводе с русского языка на английский иллюстрирует тенденцию английского языка к экономному использованию языковых средств при описании предметных ситуаций.

(34) *Ответ на этот вопрос имеет принципиальное значение. Он важен не только для выяснения места и роли профсоюзов в современном обществе на Западе.*

The importance of this fundamental question goes beyond the aim of elucidating the place and role of the unions in the West today.

Для научно-технических материалов английского языка характерно преобладание простых предложений, которые, по приблизительным подсчетам, составляют в среднем около 53% общего числа предложений в тексте. Это явление несвойственно научному стилю в русском языке, где сложные предложения используются очень широко. В связи с этим в англо-русских технических переводах часто используется прием объединения

предложений, в результате чего двум или более простым предложениям английского оригинала соответствует одно сложное предложение в русском переводе.

(35) This condition, however, changes at certain critical energies of the electrons. At these critical energies the gas atoms do absorb energy, and a sudden drop in the electron current is simultaneously observed.

Однако это условие нарушается при некоторых критических энергиях электронов, когда атомы газа поглощают энергию, и одновременно наблюдается внезапное падение электронного тока.

8. Абсолютные причастные обороты

Абсолютный причастный оборот английского языка фактически является самостоятельным предложением с собственным «подлежащим», однако, глагол в нем стоит в неличной форме, т.е. не является сказуемым в английском предложении. При переводе на русский язык неличная форма глагола трансформируется в личную и становится сказуемым в придаточном предложении. Если оборот стоит в начале предложения, то он обычно присоединяется к главному подчинительными союзами «так как», «хотя», «если», «после того, как» и другими.

(36) The other conditions being equal, the acceleration will be the same.

Если все остальные условия равны, ускорение будет одним и тем же.

(37) Water being cooled, the rate of the reaction was low.

Когда воду охлаждали, скорость реакции была низкой.

Если оборот стоит в конце предложения, то при переводе он обычно присоединяется к главному предложению сочинительными союзами «причем», «а», «и».

(38) The solid production was doubled with one-third of the produce shipped to the works.

Производство кислоты было удвоено, причем одна треть всей продукции отправлялась на завод.

9. Перераспределение семантических компонентов

Причиной перераспределения семантических компонентов являются ограничения, налагаемые нормами языка перевода на сочетаемость слов и словообразование. Этот метод состоит в отнесении семантических компонентов, принадлежащих исходному языковому объекту (слову, словосочетанию) к другим языковым объектам, что приводит к изменению конфигурации признаков, характеризующих предметную ситуацию [4.].

(39) Правительству было поручено глубоко изучить предложения и замечания, высказанные депутатами.

The Congress asked the government to make an in-depth study of the deputies' proposals and opinions.

(40) Вспомним, какой высокой степенью нетерпимости отличались порой дискуссии в русской литературе 19 века.

Let us remember the degree of intolerance which characterized literary debate in Russia in the 19-th century.

10. Стилистическая правка текста при переводе, проблема перевода метафор

Для научно-технического стиля как в английском, так и в русском языке характерно стремление к четкости и строгости изложения, точному употреблению терминов, отказу от косвенных, описательных обозначений объектов, широкому использованию стереотипов специальной лексики. Более детальный анализ показывает, что строгость в употреблении терминов и привычных формулировок, в целом, более свойственна русскому научно-техническому стилю, чем английскому. Поэтому при переводе на русский язык переводчик нередко осуществляет «стилистическую правку» оригинала, вводит вместо описания точный термин, заменяет авторский оборот более привычным штампом.

Сравним, например, следующий перевод с его оригиналом:

(41) It was discovered, however, that X-rays scattered by atoms exhibited not only the frequency V_0 of the incident X-rays but also a new frequency V_1 not present in the original X-rays.

Однако было обнаружено, что рентгеновское излучение, рассеянное на атомах, содержит не только частоту V_0 падающего излучения, но также новую частоту V_1 , которой не было в спектре первоначального рентгеновского излучения.

(42) The time between a research idea and its translation into product is diminishing quite rapidly.

Время между возникновением идеи при исследованиях и ее осуществлением в виде промышленной разработки весьма быстро сокращается.

По нашим данным, основными способами перевода метафор английского научного текста на русский язык являются калька (например, «дружественный» интерфейс), транслитерация, подходящий по смыслу образ и экспликация. Следует отметить, что метафора в англоязычной научной прозе является очень важным когнитивным инструментом.

По мнению американского философа и логика М. Блэка, метафора работает, проецируя на то, что мы должны понять, множество ассоциативных связей, соответствующих комплексу представлений об образце, с помощью которого человек осваивает неизвестное. Авторами многих метафор в генетике являются крупнейшие

биологи (А. Вейсман, Н. Вавилов, К. Уоддингтон и другие), например, *mitotic spindle* - митотическое веретено, *quantum species* квантовое видообразование, *biological clock* биологические часы. Метафорам в генетике посвящена работа [17].

Лакофф и Джонсон [18] подробно описывают процесс структурирования метафор и выделяют три главные области «концептуальных структур», из которых черпают «корневые» метафоры. Первая - это область «физического», т.е. структура, определяющая понимание предметов и идей как «объектов, существующих независимо от нас». Вторую область составляет культура, третью - собственно интеллектуальная деятельность. Эти области и ограничивают наши возможности описания мира. Выбрав какое-то понятие, принадлежащее одной из этих концептуальных структур, и сопоставляя его с понятием, входящим в другую структуру, мы связываем различные области и "структурируем одну в терминах другой". Примеры переводов некоторых метафор из научных и научно-политических текстов:

(43) *Big Bang Theory* - теория большого взрыва;

(44) *axis of evil* – ось зла;

(45) *about the old bear* - по поводу возможной российской угрозы.

Хотя стилю научной прозы характерна строгость изложения, в английских научных текстах нередко встречаются эмоциональные эпитеты, образные и фигуральные выражения, риторические вопросы и тому подобные стилистические приемы, оживляющие повествование и более свойственные разговорному стилю или художественной речи. Такая свобода стиля меньше свойственна научно-техническим материалам на русском языке. Сопоставительный анализ переводов показывает, что переводчики регулярно осуществляют стилистическую адаптацию переводимого текста, опуская эмоционально-стилистические элементы оригинала, которые кажутся им неуместными в «серьезном» научном изложении. Такие, например, оценочные эпитеты, как *dramatic*, *successful*, *excellent*, и т.п., нередко оказываются избыточными в русском переводе:

(46) *The spectral lines provide one dramatic example of the discreteness in nature.*

Спектральные линии являются примером дискретности в природе.

(47) *Having successfully obtained the expression for the experimental results for black body radiation...*

Получив формулу, соответствующую экспериментальным результатам для излучения абсолютно черного тела.

(48) *These conclusions, however, raised other uncomfortable questions.*

Эти заключения, однако, вызвали ряд вопросов.

(49) *Modern technology is growing at a very rapid rate, and new devices are appearing on the horizon much more frequently.*

Современная техника развивается настолько быстро, что новые типы приборов появляются значительно чаще, чем это было раньше..

11. Заключение

Наши исследования сосредоточены на максимально полном выявлении всех возможных типов структурных знаков, передающих сходные значения, и их детальной категоризации. Категории понимаются как обобщения языкового и ментального планов. Развиваемые языковые представления основаны на функционально-семантических характеристиках, близких современным направлениям исследований в функциональной грамматике [19,20].

Наша задача – выявить все типы структурных знаков – и выстроить их в единую систему семантико-синтаксических представлений для лингвистического процессора, снабдив языковые правила статистическими данными, которые могут быть получены при исследовании корпусов параллельных текстов.

Список литературы

1. Шаумян С.К. Семантическая Лингвистика как Объяснительная Наука. // Труды Международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии», М.: "Наука", 2005, с. 507-513.
2. Kozerenko, E.V. *Cognitive Approach to Language Structure Segmentation for Machine Translation Algorithms // Proceedings of the International Conference on Machine Learning, Models, Technologies and Applications, June, 23-26, 2003, Las Vegas, USA.// CSREA Press, pp. 49-55, 2003.*
3. Комиссаров В.Н. *Лингвистика перевода.* 1980.
4. Бреус Е.В. *Основы теории и практики перевода с русского языка на английский.* М.: Изд-во УРАО, 1998.
5. Степанов Ю.С. *Имена, предикаты, предложения (семиологическая грамматика).* М.: Эдиториал УРСС, 2004.
6. Шмелев Д.Н. *Синтаксическая членимость высказывания в современном русском языке.* М., 1976.
7. Золотова Г.А. *К вопросу об объекте синтаксических исследований.* – Изв. АН СССР. Сер. Литер. И языка, 1979, т. 38, № 1.
8. Валгина Н.С. *Синтаксис современного русского языка.* М.: Высшая школа, 1991.
9. Шахматов А.А. *Синтаксис русского языка.* М.: Эдиториал УРСС, 2001.
10. Пулькина И.М., Захава-Некрасова Е.Б. *Русский язык. Практическая грамматика с упражнениями.* М.: Русский язык, 2004.
11. Золотова Г.А., Ониненко Н.К., Сидорова М.Ю. *Коммуникативная грамматика русского языка.* М., 2004.

12. Visson, L. *From Russian Into English: An Introduction to Simultaneous Interpretation*. Ann Arbor, Michigan: Ardis, 1991.
13. Knight, K. *Automating knowledge acquisition for machine translation*. *AI Magazine* 18, pp.81-96, 1997.
14. Кобозева И.М., Лауфер Н.И. *Семантика модальных предикатов долженствования // Логический анализ языка. Культурные концепты*. М., 1991.
15. Кобозева И. М. *Лингвистическая семантика*. М.: Эдиториал УРСС, 2000.
16. Зализняк Анна А., Падучева Е.В. *Предикаты пропозициональной установки в модальном контексте // Логический анализ языка. Проблемы интенциональных и прагматических контекстов*. М., 1989. С. 92-115.
17. Седов А. Е. *Метафоры в генетике // Вестник Российской Академии Наук*. том 70, № 6, с. 526-534, 2000.
18. Lakoff G., Johnson M. *Metaphors We Live By*. Chicago: The University of Chicago Press, 1980.
19. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E. *Longman Grammar of Spoken and Written English*. Pearson Education Limited, 1999.
20. Шелякин М. А. *Справочник по русской грамматике*. М.: Русский язык, 2005.