

МОДЕЛИРОВАНИЕ ТРАНСФОРМАЦИЙ НОМИНАТИВНЫХ СТРУКТУР ПРИ РЕШЕНИИ ЗАДАЧ ФРАНЦУЗСКО-РУССКОГО МАШИННОГО ПЕРЕВОДА

MODELLING OF TRANSFORMATIONS OF NOMINATIVE STRUCTURES for the DECISION OF PROBLEMS OF FRENCH-RUSSIAN MACHINE TRANSLATION

И.В. Галина (irn_gl@mail.ru)

ИПИ РАН

В работе рассматриваются вопросы построения отдельных функционально-семантических моделей при трансформациях номинативных структур в рамках решения задач французско-русского (и обратного - русско-французского) машинного перевода (МП).

Анализ структур и блок многовариантных формальных логико-семантических правил разрабатываются с учетом функционального сходства и синтаксической полисемии для номинативных конструкций на материале фокусной выборки параллельных текстов на русском и французском языках. Задача переноса значений решается на базе анализа когнитивных структур. Построение моделей ведётся в рамках проекта по созданию многоязычного лингвистического процессора, реализуемого на основе функционально-семантического подхода.

1. Введение

Разработка систем машинного (иначе: автоматического) перевода, несмотря на давнюю историю вопроса, до сих пор является одной из актуальных и главных проблем в области прикладной лингвистики и информационных технологий. Анализ современного состояния МП и обзор систем МП проводился неоднократно, в частности, один из самых полных - в [2,3]; анализ существующих теорий МП и работающих зарубежных систем автоматического перевода был недавно представлен в [10].

В отечественной лингвистике имеется ряд хорошо известных – как теоретических, так и практических – работ в этой области, начиная с модели И.А. Мельчука «Смысл-Текст», нашедшей практическое воплощение в созданной под руководством Ю.Д. Апресяна системе французско-русского перевода ЭТАП-1 [1]; сейчас реализуется уже новая - англо-русская - версия системы ЭТАП. Среди систем МП по интересующему нас французско-русскому направлению перевода можно отметить систему ФРАП, разрабатывавшуюся ВЦП совместно с лабораторией машинного перевода МПНИИЯ им М. Тореза в 1976-1986 годах под руководством Н.Н.Леонтьевой (наследницей этой системы является АОР) [12]; а также чрезвычайно успешный коммерческий проект PROMPT, обеспечивающий прямой и обратный перевод с шести европейских языков [www.translate.ru]. Однако ни одна из этих систем МП не является совершенной, каждая имеет свои недостатки и, разумеется, не обеспечивает «идеального» перевода, что оправдывает новые попытки создания других систем МП, базирующихся на иных принципах.

В данной статье рассматриваются вопросы построения функционально-семантических моделей при трансформациях некоторых номинативных структур в рамках решения задач французско-русского (и обратного: русско-французского) машинного перевода.

Задачей нашей работы было исследование когнитивных функций линеаризации некоторых фразовых (номинативных) структур французского языка, что позволяет в дальнейшем представить их как систему формальных правил, встраиваемых в уже существующий лингвистический процессор (пока работающий в виде англо-русской версии перевода – [8, 9, 10]). Результатом является начальный этап представления моделей номинативных структур.

Анализ и построение этих моделей ведётся в рамках проекта по созданию многоязычного лингвистического процессора для машинного перевода [10], реализуемого на основе функционально-семантического подхода. В настоящее время к экспериментальной версии данного процессора постепенно подключается новый модуль, обеспечивающий французско-русский перевод.

Функционально-семантический подход, в частности, подход к исследованию языковых явлений на основе функционально-семантических полей, был предложен в работах А.В. Бондарко [4], в которых рассматривались примеры анализа отдельных функционально-семантических полей на материале русского языка. Подход, используемый при разработке указанного лингвистического процессора "заключается в построении

унифицированной модели функциональных значений синтаксисом, учитывающих сдвиги значений, производимых трансформациями" [8]. (Заметим, что термин "синтаксема" предложен Золотовой, определяющей её как минимальную синтаксическую единицу, являющуюся результатом взаимодействия трех составляющих: категориально-семантической – значения, синтаксической – функции, и морфологической – формы; три этих признака мотивируют друг друга, их "взаимопроникновение" и создаёт синтаксему как языковую единицу [5, 6, 7].)

Применяемый нами подход базируется на сопоставлении глубинных семантико-синтаксических структур входного и выходного языков; основой оценки степени совпадения языковых структур языков служит принцип движения от семантического значения к языковой форме и от формы к значению; процесс переноса значений (трансфера) происходит через функционально-категориальные значения языковых единиц. "Построение функционально-семантических представлений – плодотворная основа для создания лингвистического процессора. На основе выделяемых значений происходит формирование единого семантического пространства, где средства выражения сходных значений в разных языках рассматриваются как синонимические средства" [9].

2. Особенности использования функционально-семантического подхода при исследовании переводимых языковых структур

При разработке данного лингвистического процессора (на основе англо-русского и обратного трансфера) Е.Б. Козеренко было предложено понятие полей функционального переноса (ПФП), явившихся базисом сегментации языковых структур для решения задач машинного перевода. Основная идея такого поля состоит в принятии гипотезы о том, что в основе грамматических структур лежат структуры когнитивные (ментальные фреймы); функционально-семантическое поле отражает взаимодействие элементов разных языковых уровней. "Переносимость фразовых структур обусловлена выбором языковых единиц исходного и целевого языка, принадлежащих к одним и тем же полям функционального переноса (ПФП), не взирая на различия или сходство их традиционных категориальных значений" [10].

Глубокий и подробный анализ семантико-синтаксического уровня русского языка был выполнен в работах [4-7]. Однако при решении задач, связанных с разработкой систем машинного перевода, приходится ограничивать тонкость семантических описаний и стараться сжимать различия, огрубляя анализ исследуемых языковых явлений, так как основной проблемой при создании систем автоматического перевода становится поиск решений, направленных на получение описаний, позволяющих всё же осуществить приемлемый перевод. Это – ещё одна причина, с которой связано признание функционально синонимичными конструкциями языковые структуры русского и французского языков, демонстрирующие явные различия на уровне поверхностного синтаксиса.

Процесс линеаризации языковых структур управляется когнитивными механизмами, что хорошо показано в работе [11], однако когнитивный аспект линеаризации пока наименее исследован.

Особенности в структурных различиях французского и русского языков обусловлены рядом причин, в числе которых важное значение имеют различия в порядке развёртывания линейной цепочки. На это в своём классическом труде, посвящённом структурному синтаксису, обращал внимание ещё Л. Теньер [13], вводя понятие языков *центробежного* и *центростремительного* типа. Согласно Теньеру, в языках центробежного типа господствует центробежный, или нисходящий, порядок: в них подчинённый элемент следует за подчиняющим (*cheval blanc*), речевая цепочка строится *сверху вниз*, то есть удаляясь от центрального узла; напротив, в центростремительных языках доминирует центростремительный, или восходящий, порядок (*белая лошадь*): в таких языках подчинённый элемент предшествует подчиняющему, в них мы строим речевую цепочку, двигаясь *снизу вверх*, то есть приближаясь к центральному узлу. Французский язык относится к умеренно центробежным, а русский – к умеренно центростремительным языкам, что определяет отличия в порядке линеаризации французской и русской речевой цепочек. "Направление словорасположения становится особенно важным, когда речь идёт о переводе с центробежного языка на центростремительный или обратно. В этом случае действует следующее правило: *необходимо инвертировать порядок следования слов, связанных вертикальной связью, и сохранить порядок следования слов, связанных горизонтальной связью*" [13, с. 33-34].

Отметим, что в языках без склонения (к которым относится французский) словорасположение всегда играет важную роль, так как в них обычно субъект и объект (то, что традиционная грамматика называет подлежащим и прямым дополнением, а Теньер первым и вторым актантами) различаются не морфологически, а по их месту в линейной цепочке, что обуславливает специфическую разницу форм выражения актанта в русском и французском языках.

Рассуждая о проблемах перевода, Теньер отмечает, что далеко не всегда различные языки обращаются к идентичным структурам для выражения идей, точно соответствующих друг другу в семантическом плане, наоборот: "в ряде случаев при переводе с одного языка на другой приходится заменять одну структуру другой"; при этом, "опыт показывает, что поверхностные переводы в таком случае не только неуклюжи, но часто даже и неточны" [13, с. 297].

Такого рода структурные преобразования названы им *метатаксистом*, а переводы, содержащие метатаксист - *глубинными переводами*; эти структурные изменения могут варьировать от простой замены грамматической категории до полной трансформации структуры с изменением в центральном узле.

3. Описание отдельных функционально синонимичных именных структур французского и русского языков как основа для составления правил переноса значений при машинном переводе

Предметом нашего анализа являются номинативные структуры французского и русского языков и их возможные трансформации, возникающие при переносе значения из исходного языка; при этом, особый интерес представляют случаи транскатегориального соответствия различных предложно-падежных и некоторых адъективных конструкций.

Материалом служит фокусная выборка параллельных текстов на русском и французском языках (научный дискурс, язык деловой прозы, финансово-экономическая тематика, трудовое законодательство). Основное внимание уделялось конструкциям словосочетаний, управляемым субстантивом: беспредложным и предложным конструкциям во французском языке, падежным и предложно-падежным конструкциям в русском.

На данном этапе исследования нас интересуют два ПФП [10]: *ПФП номинации и относительности*, куда входят языковые структуры, выполняющие функции номинации (включая сентенциальные единицы) и *ПФП атрибутивности*, куда (кроме прилагательных и причастий, составляющих ядро этого поля) также входят номинативные модификаторы – единицы и структуры.

В результате анализа имеющегося лингвистического материала (рассматривались именные синтаксемы) оказалось возможным выделить следующие некоторые функционально-семантические конструкции во французском и в русском языках, которые считаются функционально синонимичными при трансфере с входного на выходной язык. (На начальном этапе анализировались только повествовательные предложения; вопросы и побудительные конструкции исключались.)

В данной работе приводятся описания базовых именных структур французского и русского языков, которые служат исходным материалом при построении правил трансфера для лингвистического процессора, применяемого в машинном переводе.

1. *Субъект действия* при активном предикате (первый актант по Теньеру) в общем случае регулярно выражается в русском языке через именительный падеж, а во французском находится в препозиции к предикатному глаголу.

N + Verbe (Pred)

Les aéroports offrent... – аэропорты предоставляют...

L'entreprise assure la promotion de ses produits par divers moyens – предприятие осуществляет продвижение своей продукции на рынок различными средствами

2. *Объект действия* при активном предикате (второй актант по Теньеру или прямое дополнение в терминах традиционной грамматики) в русском языке обычно выражен винительным падежом (в случае неодушевлённого существительного), генитивом (случае одушевлённого существительного), либо (реже) дательным падежом; во французском языке конструкции со значением объекта действия регулярно соответствует постпозиция субстантива по отношению к глаголу.

Verbe + N

L'entreprise assure la promotion – предприятие осуществляет продвижение...

Cela aide le client – это помогает покупателю ...

... ce qui doit être aider notre *entreprise* – ... то, что должно помочь нашему *предприятию*.

3. *Генитиву* в приименной беспредложной позиции в русском языке может быть дан универсальный перевод на французский язык через конструкцию с предлогом de. То же самое относится к часто встречающимся в деловой лексике и в научном дискурсе генитивным цепочкам.

N1 + de +N2

l'objet de la demande – предмет заявления (просьбы); *le représentant du fournisseur* – представитель поставщика; *description des marchandises* – опись товаров; *L'Institut national de la statistique et des études économiques* – Государственный институт статистики и экономических исследований; *les travaux du Groupement des Cartes Bancaires* – разработки Объединения кредитных карточек.

Это верно для всех значений данной синтаксемы, приведённых в Синтаксическом словаре Золотовой [5], а именно:

родительный количественный (стая птиц) – *un stock des marchandises (запас товаров)*;

родительный квалитативный (человек доброй души) – *un homme de haute taille (человек высокого роста), un produit de masse (массовая продукция)*;

родительный посессивный (дом отца) – *un entrepôt de l'entreprise (склад предприятия)*;

родительный агентивный (приезд гостя) – *une activité de l'entreprise (деятельность предприятия)*;

родительный носителя признака (зелень трав) – *la pureté de l'air (чистота воздуха)*.

4. *Посессивные конструкции* представляют для нас особый интерес. Можно выделить некоторые регулярные способы выражения посессива в рассматриваемых языках:

а) самый простой случай: словосочетание "притяжательное прилагательное + существительное" во французском языке, которому соответствует словосочетание "притяжательное местоимение + существительное" в русском (эти конструкции можно считать синтаксически сходными)

ADJ_Poss +N (fr) = PRON_Poss +N (rus)

nos documents – наши документы;

б) именная конструкция с предлогом *de* (встречается наиболее часто), соответствующая в русском языке родительному посессивному или словосочетанию "прилагательное + существительное"

$N1 + de + N2$ (fr) = $N1 + N2$ (Gen), ADJ +N

les guichets de banque – кассы банка, банковские кассы

с) именная конструкция с предлогом *à* (встречается гораздо реже), соответствующая в русском языке родительному посессивному или словосочетанию "прилагательное + существительное"

$N1 + à + N2$ (fr) = $N1 + N2$ (Gen), ADJ +N

la fille à ma tante – дочь моей тётки, тёткина дочь

5. *Атрибутивные определения*. Были выделены некоторые *регулярные способы выражения атрибутивного определения* (характеристики):

а) обычное адъективное определение (прилагательное) во французском языке: при трансфере на русский язык возможен пословный перевод

$N + ADJ$ (чаще), $ADJ + N$ (реже)

des caractéristiques diverses – различные характеристики, *les petites entreprises* – малые предприятия

Позиция прилагательного во французском языке (в общем случае: постпозиция) зависит от специфики словоупотребления лексемы и – реже – от специфики контекста (например, логическое ударение). Иногда от места определения-прилагательного в линейной цепочке меняется значение: *un homme brave* – храбрый человек, *un brave homme* – славный человек.

б) наиболее частый случай: французской конструкции "существительное + предлог *de* + существительное" в русском языке соответствует "прилагательное + существительное" = "существительное + существительное в посессивном генитиве"

$N1 + de + N2 = ADJ + N, N1 + N2$ (Gen)

moyens de paiement – платёжные средства, средства платежа

les guichets de banque – банковские кассы (принадлежность), кассы банка

в) атрибутивная характеристика во французском языке может быть дана и через конструкцию "существительное + предлог *de* + инфинитив", центральным узлом которой является именно субстантив; в русском языке обычно соответствует сочетанию "прилагательное + существительное" или "существительное + существительное в генитиве". Таким образом, при переводе часто необходимо совершать категориальные преобразования.

$N + de + infinitif = ADJ+N, N1+N2$ (Gen), $N+ infinitif$

sa façon de parler – речевая манера, манера речи, манера говорить(выражаться, изъясняться)

г) характеристика предназначения (для чего): именная конструкция с предлогом *à* во французском языке часто соответствует в русском языке именной конструкции с предлогом "для" или существительному с адъективным определением, хотя, разумеется, данная фразовая структура не может считаться регулярным способом выражения указанного значения. (Другие значения, выражаемые этой конструкцией будут обрабатываться иными правилами или заносится в отдельный Словарь идиом, как, например, "*fer à cheval*" – подкова.)

$N1 + à + N2$

un verre à liqueur – бокал для ликёра, ликёрный бокал

д) характеристика предназначения во французском часто выражается и через конструкцию "существительное + предлог *à* + инфинитив", соответствующую в русском языке сочетанию "прилагательное + существительное" или "существительное + существительное в генитиве"

$N + à + infinitif$

machine à coudre – швейная машина

une machine à plier les documents – машина для фальцевания, фальцевальный станок (в полиграфии)

В данном случае при переводе приходится производить категориальные изменения: французский инфинитив регулярно меняется на прилагательное или субстантив с предлогом.

е) Заметим, что во французском языке субстантив в роли неадъективного определения чаще всего вводится с помощью предлога *de*; но для того, чтобы существительное выступало в роли определения присутствие предлога *de* не обязательно: французская конструкция $N1 + de + N2$ регулярно используется для выражения *атрибутивных неадъективных определений* с разными значениями, но не является единственно возможной. В частности, допускается употребление беспредложных конструкций, в которых определяющее существительное следует за определяемым субстантивом (это касается словосочетаний, включающих собственные имена, обычно они означают "в честь кого", "имени кого" назван атрибулируемый объект). Такие определения передаются на русском языке адъективным определением или генитивом посессива.

$N1 + N2$ *ATR* (fr)

La Tour Eiffel – Эйфелева башня; *l'affaire Dreyfus* – дело Дрейфуса.

Данные именные структуры, управляющим узлом которых является субстантив, выделяются и анализируются как *структурные знаки* в смысле С.К. Шаумяна [14], считающего, что фундаментальная форма структурных значений заключается в синтаксических функциях, а морфологические функции в языках с

морфологией являются отображением синтаксических функций; таким образом, система значений с точки зрения коммуникативной функции языка инвариантна по отношению к знакам, которые признаются переменными (принцип переноса значений). Именно поэтому субъектно-объектные отношения могут выражаться либо морфологическими знаками, либо порядком слов.

Настоящее описание функционально сходных (но часто – морфологически различных) номинативных структур во французском и русском языках делалось для решения задач компьютерного моделирования по заданию конкретных правил переноса значений в лингвистическом процессоре, предназначенного для машинного перевода. Это описание, разумеется, не полно, и дальнейшая работа будет связана с анализом оставшихся предложных и предложно-падежных именных структур русского языка и соответствующих им функционально синонимичных структур на французском языке, с уточнением и развитием уже существующего материала, а также с разработкой формализованных представлений этих структур, пригодных для функционирования лингвистического процессора.

4. Заключение

В настоящее время для существующей экспериментальной версии процессора на базе английского и русского языков разрабатывается модуль французско-русского (и обратного) перевода. Сейчас мы находимся на начальном этапе разработки.

Механизм установления функционально-семантического тождества номинативных (и других) конструкций будет уточняться по мере накопления лингвистического материала для анализа. Ведутся работы по созданию блока семантико-синтаксических правил с учетом синтаксической неоднозначности для номинативных конструкций на материале фокусной выборки параллельных текстов на русском и французском языках (язык деловой прозы, финансово-экономическая тематика, трудовое законодательство). Задача межъязыкового переноса значений для русского и французского языков решается на базе анализа когнитивных структур, выделенных при создании англо-русского лингвистического процессора [10].

В дальнейшем планируется сосредоточиться на проблемах установления детальных семантико-синтаксических соответствий различных именных и предложных конструкций в русском и французском языках в ситуациях изменения категориальных значений при переводе.

Список литературы

1. Апресян Ю.Д. и др. Лингвистическое обеспечение системы французско-русского автоматического перевода ЭТАП-1. 1. Общая характеристика системы. // Теория и модели знаний. (Теория и практика создания систем искусственного интеллекта.) Труды по искусственному интеллекту. Ученые записки Тартуского государственного университета. Тарту, 1985, с. 20-39.
2. Бакулов А.Д., Леонтьева Н.Н.. Теоретические аспекты машинного перевода // Искусственный интеллект. Кн.1. Системы общения и экспертные системы. - М., 1990. - С. 201 -216.
3. Бакулов А.Д., Леонтьева Н.Н., Шалыпина З.М.. Отечественные системы машинного перевода // Искусственный интеллект. Кн.1. Системы общения и экспертные системы. - с. 248-260.
4. Бондарко А.В. Принципы функциональной грамматики и вопросы аспектологии. // М.: "Эдиториал УРСС", 2001.
5. Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. // М.: "Наука", 1988.
6. Золотова Г.А. Коммуникативные аспекты русского синтаксиса. // М., УРСС, 2001.
7. Золотова Г.А. Значение, функция и форма в языке. // Труды Международной конференции Диалог'2004 "Компьютерная лингвистика и интеллектуальные технологии", М.: "Наука", 2004
8. Козеренко Е.Б. Функциональная семантика в компьютерных решениях. // Труды Международного семинара Диалог'2002 «Компьютерная лингвистика и интеллектуальные технологии», М.: "Наука", 2002, с. 218-226.
9. Козеренко Е.Б. Моделирование языковых структур со значением условия и причины для англо-русского лингвистического процессора. // Труды Международной конференции Диалог'2003 «Компьютерная лингвистика и интеллектуальные технологии», М.: "Наука", 2003
10. Козеренко Е.Б. Моделирование переноса функциональных значений для англо-русского машинного перевода. // Труды Международной конференции Диалог'2004 "Компьютерная лингвистика и интеллектуальные технологии", М.: "Наука", 2004.
11. Лауфер Н.И. Линеаризация компонентов сочинительных конструкций. // в сб. "Моделирование языковой деятельности в интеллектуальных системах". Под ред. Кибрик А.Е., Нариньяни А.С. М: "Наука", 1987, с. 167-176.
12. Леонтьева Н.Н., Никогосов С.Л.. Система ФРАП как информационная система // Актуальные вопросы практической реализации систем автоматического перевода. - М., Изд-во МГУ, 1982. - с. 134-166.
13. Теньер Л. Основы структурного синтаксиса. // М. "Прогресс", 1988.
14. Шаумян С.К. Семиотическая Лингвистика как Объяснительная Наука. // Труды Международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии», М.: "Наука", 2005, с. 507-513.