

Т.Н. Медведева

**ФОРМАЛЬНЫЕ МОДЕЛИ
В ЛИНГВИСТИКЕ**

Учебное пособие

Часть 2

Саратовский государственный университет имени Н.Г. Чернышевского

УДК 81'33 (075.8)

ББК 81.1-923

М 42

Медведева Т.Н.

М 42 Формальные модели в лингвистике: учебное пособие: часть 2 / Т.Н.

Медведева. – Саратов, 2014. – 50 с.

Учебное пособие предназначено для студентов-бакалавров IV курса специальности (направления) 45.03.03 «Фундаментальная и прикладная лингвистика».

Рекомендовано к изданию кафедрой теории, истории языка и прикладной лингвистики Саратовского государственного университета им. Н.Г. Чернышевского.

УДК 81'33 (075.8)

ББК 81.1-923

Работа издана в авторской редакции

© Медведева Т.Н., 2014

Тема 5. Компьютерное моделирование пословицы

Моделируя пословицы при помощи компьютерной программы, В.С. Баевский¹ начинает с уровня наблюдения и стремится создать модель поверхностной структуры пословицы, причем предельно простую, и реализует ее на компьютере. При этом предполагается, что если поверхностная структура понята и смоделирована правильно, то и на уровне глубинном полученный текст будет соответствовать интуитивным представлениям носителей языка о пословице. Таким образом, критерием правильности признается языковая и паремиологическая компетенция носителей языка.

Задача состоит не в том, чтобы создавать новые пословицы для последующего их использования носителями языка. Но если удастся на основе модели поверхностного уровня пословицы создать тексты, которые носители языка признают подобными пословичным, это с большой вероятностью будет свидетельствовать о том, что структура пословицы понята и воссоздана правильно.

Для изучения берется один класс пословиц: *Как аукнется, так и откликнется; Что посеешь, то и пожнешь; Кого люблю, того казню; С кем поведешься, от того и наберешься; Где пьют, там и льют* и т.п. Этот класс пословиц весьма показателен для всего жанра. Пословице искони присуща четырехчастная структура, причем четыре наименьших сегмента сгруппированы по два, причем эти две половины противостоят одна другой. Первая половина пословицы называется головой (*кого люблю*), вторая – хвостом (*того казню*).

5.1. Структура пословицы изучаемого класса

1) **Тематика** отражает общественные отношения преимущественно в этических категориях. Такова же в значительной степени тематика всех пословиц вообще.

2) **Образность**. В пословице часто возникает образ, основанный на остранинии: хвост противоречит голове, их объединение, обычно острое, неожиданное, вызывает специфическую для искусства эстетическую реакцию «противочувствия» – катарсис. Слыша слова «кого люблю», мы, естественно, ожидаем чего-то наподобие «того и награждаю», «того и хвалю»; между тем, мы слышим нечто прямо противоположное: «того и казню». В первом случае мы видим лишь плоский трюизм, во втором – глубокую мысль, метафору, выражение нелицеприятности, требовательности, а, быть может, и более глубоких свойств человеческой психики. Вот почему в изучаемый класс не включаются пословицы типа *Кто ленив, тот и сонлив*: в них нет противоречия между головой и хвостом.

¹ Баевский В.С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. – М.: Языки славянской культуры, 2001.

3) **Логико-семантические отношения.** Голова и хвост пословицы находятся в отношении, близком к каузальной зависимости: $G \rightarrow X$. Оба глагола (обозначим их V_1 и V_2) состоят в отношениях, близких к антонимичности. В действительности это не совсем так. Если *аукнется – откликнется, посеешь – пожнешь, люблю – казню, пьют – льют* представляют более или менее ярко выраженные антонимические отношения, то в пословице *С кем поведешься, от того и наберешься* антонимичность глаголов весьма завуалирована. Но в модель пословицы вводится условие антонимичности глаголов, поскольку именно при этом условии максимально раскрывается специфика пословицы (оппозитивность – одно из основных свойств пословицы).

4) **Лексика.** V_1 и V_2 должны принадлежать к исконно русской лексике. Авторы ввели в память компьютера словарь иностранных слов (глаголы), и с ними сравнивали каждый V_1 и V_2 .

5) **Грамматика.** Пословица изучаемого класса представляет собой сложноподчиненное предложение с придаточным подлежащим, дополнительным, сравнительным, места или следствия (по классификации академической грамматики). Кроме V_1 и V_2 предложение содержит относительное и соотносительное слова r_1 и r_2 . Таким образом, на морфологическом уровне текст пословицы представляет собой последовательность $r_1; V_1; r_2; V_2$.

В образовании пословиц могут участвовать следующие пары относительных и соотносительных слов: *как – так и; где – там и; сколько – столько и; куда – оттуда и; кто – тот и; что – то и.*

В пословицах рассматриваемого класса глаголы стоят в изъявительном наклонении преимущественно в настоящем времени. Во множественном числе глаголы берутся во всех трех лицах, в единственном же – только во 2-м и 3-м.

Кроме того, для построения отмеченных предложений оказались важны направленность / ненаправленность обозначаемого глаголом действия и переходность / непереходность (с уточнением в случае переходности, переходит действие на одушевленный / неодушевленный предмет). Так, *идти* обозначает направленное действие, *сидеть* – ненаправленное, *баловать* – действие, переходящее на одушевленный предмет, *белить* – на неодушевленный, *бранить* – и на одушевленный и на неодушевленный.

б) **Фонологический уровень.** Пословица рассматривается как явление стихотворной речи, хотя в сборниках и в устном бытовании встречается много текстов аритмичных и безрифменных. Типична для пословицы стихотворная форма, так что иногда говорят даже о пословичном стихе, видя его отличительные признаки в том, что он состоит из двух колонов с интонационным переломом типа антикаденции между ними, с синтаксическим и семантическим параллелизмом колонов при тенденции к равноклаузульности. В каждой колонне кроме сильного, синтагматического ударения, возможно и более слабое побочное. Изучаемый класс пословиц ярко представляет все эти особенности, приписываемые пословичному стиху вообще.

Почти все эти признаки осуществляются сами собой как простое следствие функционирования семантической и синтаксической структур. Един-

ственный фактор, который должен быть учтен особо, – равноклаузульность. Поэтому необходимо ввести особое правило: V_1 и V_2 должны быть одинаково окситонами, парокситонами или пропарокситонами. Выполнение этого правила автоматически ведет к появлению глагольной рифмы, наиболее характерной для народного стиха вообще.

5.2. Алгоритм порождения пословицы

В память компьютера вводятся четыре списка:

А. Глаголы, взятые подряд из словаря (в эксперименте были взяты все глаголы, начинающиеся буквой *б*, и случайным образом выбрано 20 глаголов).

В. Глаголы, взятые подряд из словаря иностранных слов (в эксперименте – на те же буквы, что и в списке А).

С. Глаголы, антонимичные глаголам списка А, взятые из словаря антонимов. Каждый из них при кодировании снабжался пометой, указывавшей, к какому из глаголов списка А он относится.

Д. Пары относительных-соотносительных слов (см. выше).

Глаголы кодируются не в инфинитиве, а в форме 3 л. мн.ч. Эта форма в качестве исходной позволяет избегать ритмичных текстов при глаголах с подвижным ударением. Например, *белить* и *чернить* по инфинитиву подходят на роль V_1 и V_2 , но ни одного текста с равноклаузульными колонами они образовать не могут. Так как мы сравниваем их в формах *белят* – *чернят*, они сразу же отбрасываются.

Алгоритм работает следующим образом.

1. Первый по порядку глагол из списка А сравнивается с каждым поочередно глаголом из списка В.

1а. Если глагол из списка А есть в списке В, то он отбрасывается, после чего совершается переход к следующему глаголу из списка А.

1б. Если глагола из списка А нет в списке В, то совершается переход к операции 2.

2. Глагол из списка А сравнивается со списком С.

2а. Если к глаголу из списка А нет антонима в списке С, то он отбрасывается, после чего совершается переход к следующему глаголу из списка А и выполняется операция 1-2.

2б. Если к глаголу из списка А есть антоним в списке С, то совершается переход к операции 3.

3. Глагол из списка А сравнивается с антонимом из списка С по месту ударения.

3а. Если ударение в глаголе из списка А и в антониме стоит на разных от конца слова местах, антоним из списка С отбрасывается и совершается переход к операции 2 (поскольку к одному глаголу из списка А может быть несколько антонимов в списке С).

3б. Если место ударения, начиная от конца слова, в глаголе из списка А и в антониме из списка С совпадает, то они становятся V_1 и V_2 и совершается переход к операции 4.

4. Из списка D берется первая пара относительных-соотносительных слов *Как – так и* и организуется последовательность: *Как V_1 , так и V_2* (не забудем, что глаголы стоят в 3 л. мн.ч.).

5. Полученный текст выводится на печать.

6. В последовательности *Как V_1 , так и V_2* глаголы берутся в 1 л. мн.ч. (см. Примечание).

7. Как (5).

8. Как (6), но 2 л. мн.ч.

9. Как (5).

10. Как (6), но 2 л. ед.ч.

11. Как (5).

12. Как (4), но со следующей парой из списка D *Где – там и*.

13-19. Как (5-11).

20. Как (4), но со следующей парой *Сколько – столько и*.

21-27. Как (5-11).

28. Проверка: обозначают ли и V_1 и V_2 направленное действие, т.е. снабжены ли они при кодировании соответствующими индексами.

28а. Если не обозначают, то переход к операции 38.

28б. Если обозначают, то переход к операции 29.

29. Из списка D берется следующая пара *Куда – оттуда и* и далее как (4).

30-36. Как (5-11).

37. Проверка: обозначают ли и V_1 и V_2 действие одушевленного предмета, т.е. снабжены ли они при кодировании соответствующими индексами.

37а. Если не обозначают, то переход к операции 41.

37б. Если обозначают, то переход к операции 38.

38. Из списка D берется следующая пара *Кто – тот и* и далее как (4).

39-45. Как (5-11).

39. В тексте, организованном в операции 38, флексия глаголов меняются на флексии 3 л. ед.ч.

40. Полученный текст выводится на печать.

41. Проверка: обозначают ли и V_1 и V_2 переходность.

41а. Если не обозначают, то переход к операции 50.

41б. Если обозначают, то переход к операции 42.

42. Проверка: переходит ли действие на неодушевленный предмет.

42а. Если не переходит, то переход к операции 50.

42б. Если переходит, то из списка D берется пара *Что – то и* и далее как (4).

43-49. Как (5-11).

50-70. В текстах, организованных в результате операций 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 29, 31, 33, 35, 39, 42б, 44, 48, голова и хвост меняются местами.

71-91. Полученные тексты выводятся на печать.

92. Переход к операции 2, так как к одному глаголу из списка А может быть несколько антонимов в списке С.

92а. Если в списке С есть новый антоним к глаголу из списка А, то переход к (3-92).

92б. Если в списке С нет новых антонимов к глаголу из списка А, то переход к (1).

93. Если в списке А глаголы исчерпаны, алгоритм перестает работать.

Примечание. При организации текстов вместо *-ут, -ют* ставится: 1 л. мн.ч. *-ем*; 2 л. мн.ч. *-ете*; 2 л. ед.ч. *-ишь*. Вместо *-ат, -ят* ставится: 1 л. мн.ч. *-ите*; 2 л. ед.ч. *-ишь*.

5.3. Обсуждение результатов и выводы

Т.А. Самойлова по алгоритму составила программу, которая была реализована на компьютере. Сперва алгоритм и программа были несовершенны и формировали тексты вроде *Куда иду, туда и сижу*. Постепенно было создано несколько редакций алгоритмов и программ, формировавших только грамматически отмеченные тексты. Часть из них приближалась (по интуитивной оценке) к пословицам, например:

Как барствуют, так и холопствуют

Как барствуем, так и холопствуем

Как барствуете, так и холопствуете

Как барствуешь, так и холопствуешь

Где барствуют, там и холопствуют

Где барствуем, там и холопствуем

Где барствуете, там и холопствуете

Где барствуешь, там и холопствуешь

Сколько барствуют, столько и холопствуют

Сколько барствуем, столько и холопствуем

Сколько барствуете, столько и холопствуете

Сколько барствуешь, столько и холопствуешь

Сколько барствует, столько и холопствует

Еще 13 текстов образованы путем перестановки головы и хвоста:

Как холопствуют, так и барствуют и т.д.

Могут быть упомянуты тексты *Как благословляют, так и проклинают* (и вся парадигма), *Как берегают, так и расточают* (и вся парадигма), *Как взбираются, так и спускаются* (и вся парадигма).

В каждой парадигме обычно выделяются один-два текста, которые в роли пословицы представляются особенно выразительными: *Кто барствует, тот и холопствует*; *Как благословляют, так и проклинают*; *Где берегают, там и расточают*.

Итак, результаты показали достаточно полное понимание структуры пословицы изучаемого класса. Были смоделированы фонологический, морфологический, синтаксический, лексический уровни, а в результате получены тексты, имеющие некоторое эстетическое значение, содержащие анализ определенных логических отношений и моральных императивов. Эти особенности сближают полученные тексты с пословицами как жанром фольклора.

Вопросы и задания к Теме 5

1. Что В.С. Баевский признаёт критерием правильности пословиц, создаваемых посредством компьютерного моделирования?
2. Какой класс пословиц берётся в качестве образца для моделирования?
3. Какова структура пословиц исследуемого класса?
4. Опишите тематику и степень образности изучаемых пословиц.
5. Каковы логико-семантические отношения между частями пословицы?
6. В чём специфика лексики и грамматики пословиц данного класса?
7. Каковы фонологические особенности данных пословиц?
8. Назовите и охарактеризуйте четыре списка лексем, вводимых в память компьютера для моделирования пословиц.
9. Опишите работу алгоритма порождения пословицы.
10. В какой форме кодифицируются глаголы и почему?
11. Что показали результаты моделирования пословиц?

Литература к Теме 5

Баевский В.С. Лингвистические, математические, семиотические и компьютерные модели в истории и теории литературы. – М.: Языки славянской культуры, 2001.

Тема 6. Общие сведения о системе ЭТАП-2

ЭТАП-2 (Электро-Технический Автоматический Перевод, вторая очередь) – система качественного автоматического перевода электротехнических текстов с английского языка на русский². Основные лингвистические компоненты системы ЭТАП-2 могут быть использованы в широком спектре информационных систем, имеющих дело с переработкой текстов на естественных языках, в частности в диалоговых системах, предусматривающих общение с базами данных и базами знаний на относительно свободном естественном языке.

Панорама существующих в настоящее время систем автоматического перевода (АП) чрезвычайно разнообразна. Их можно классифицировать по самым различным признакам, из которых наиболее важными являются следующие:

1. Поколение системы АП. Различаются *четыре поколения* в зависимости от структуры алгоритма и от типа формального представления предложения, на уровне которого происходит переход от одного языка к другому.

В системах *первого поколения* алгоритм не разделяется на отдельные блоки анализа и синтеза: синтез данного фрагмента предложения протекает одновременно с его анализом в процессе однократного просмотра фразы слева направо. Лингвистическая информация не отделена и принципиально неотделима от алгоритма, т.е. задается чисто процедурно, поэтому в системах первого поколения не используется в явном виде никакое формальное лингвистическое представление предложения.

В системах *второго поколения* а) алгоритм анализа отделен и независим от алгоритма синтеза; б) лингвистическое обеспечение задается декларативно, т.е. независимо от алгоритмов; в) собственно перевод осуществляется на уровне синтаксической структуры предложения, явно выделяемой алгоритмом в процессе обработки фразы.

Системы *третьего и четвертого поколений*, в настоящее время находящиеся в стадии разработки или только проектирования, наследуют первые два свойства систем второго поколения. По замыслу, в системах *третьего поколения* перевод должен осуществляться на уровне полной семантической структуры предложения, а в системах *четвертого поколения* – на уровне «энциклопедического представления» – семантической структуры, слитой с определенным фрагментом представления знаний.

2. Уровень разработки. Среди систем, доведенных до машинной реализации, различаются *промышленно действующие* и *экспериментальные* системы. *Промышленно действующие* системы располагают словарями объемом от 12000-15000 до сотен тысяч слов в зависимости от числа обслужива-

² Лингвистическое обеспечение системы ЭТАП-2 / Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др. – М.: Наука, 1989.

емых предметных областей и количества рабочих языков системы. Что касается грамматик, то они могут быть более или менее полными в зависимости от поколения системы. В промышленно действующих системах второго поколения морфология и синтаксис должны быть весьма полными.

3. Тип перевода. Различаются переводы, выполняемые: а) машиной без вмешательства человека; б) машиной с привлечением человека в качестве пред-, интер- или постредактора; в) человеком с помощью машины (автоматические словари и т.п.).

4. Назначение перевода. Имеются два полярных типа систем: для *оперативного*, но, может быть, поверхностного информирования научных и инженерных работников о новых научных идеях и технических решениях и для *точного* перевода технической документации.

5. Число рабочих языков системы. Различаются *двуязычные* и *многоязычные* системы АП.

6. Тип рабочих языков системы. Различаются системы, работающие с естественным языком в его полном объеме или с весьма ограниченным фрагментом естественного языка, с типологически далекими или типологически близкими языками.

Перечисленные типы различий независимы друг от друга лишь с логической точки зрения. Фактически между ними обнаруживается значительная *степень корреляции*. Так, подавляющее большинство промышленно действующих систем АП – системы первого или «полуторного» поколений; подавляющее большинство систем АП второго поколения работают пока только в экспериментальном режиме. Промышленно действующие системы АП дают перевод низкого качества и поэтому нуждаются в серьезном вмешательстве человека на стадиях пред- и пост-, а иногда и интерредактирования. Экспериментальные системы второго поколения дают переводы значительно более высокого качества, что существенно сокращает, а иногда и сводит на нет долю участия человека в процессе перевода. Тем не менее в ситуациях, когда необходимо гарантировать высокую точность информации (таково, например, положение с переводом технической документации на экспортируемые изделия), самым надежным остается человеческий перевод, а наиболее перспективной его формой – перевод с использованием многоаспектных автоматических словарей.

Рассматривая ЭТАП-2 с перечисленных точек зрения, мы можем охарактеризовать его как экспериментальную двуязычную систему АП второго с половиной поколения, работающую без какого-либо вмешательства человека, имеющую дело с типологически «среднеудаленными» естественными языками и предназначенную для информирования научных и инженерных работников. ЭТАП-2 использует полную морфологию и достаточно полный синтаксис рабочих языков; в частности, правил синтаксического анализа, по видимому, достаточно для обработки любых научно-технических текстов на английском языке. Объемы комбинаторных словарей английского и русского

языков составляли 5748 и 6473 словарных статьи соответственно, причем «рабочая» часть словарей, прошедшая экспериментальную проверку, составляла около 4000 единиц в каждом. В связи с ограниченностью словаря правила перевода в собственном смысле не могут претендовать на полноту.

В чисто лингвистическом аспекте ЭТАП-2 отличается от других систем второго поколения следующими свойствами:

1. *Полная независимость описания входного и выходного языков.* Описание каждого из языков представляет собой редуцированный, но достаточно принципиальный вариант лингвистической модели «Смысл \Leftrightarrow Текст». Каждая из двух моделей дает интегральное (единое) описание морфологии, синтаксиса и словаря соответствующего языка. Принцип интегральности описания означает, что морфология, синтаксис и словарь полностью согласованы друг с другом по типам помещаемой в них лингвистической информации и что эта лингвистическая информация во всех трех компонентах записывается абсолютно единообразно, т.е. на одних и тех же формальных языках.

2. *Декларативность задания лингвистических знаний*, т.е. их полная независимость от алгоритма. Декларативность задания лингвистической информации имеет два достоинства.

Во-первых, она делает непосредственно доступной ту лингвистическую модель, которая лежит в основе системы АП. Ее легко охватить в целом, оценить достоинства и недостатки, сравнить с другими аналогичными моделями. Ее не надо реконструировать, извлекая по кусочкам из алгоритмических процедур. В этом отношении декларативно заданная лингвистическая модель очень похожа на обычное лингвистическое описание. В сущности, это те же грамматика и словарь, но только записанные на формальном языке.

Во-вторых, декларативно заданную лингвистическую модель легко корректировать в ходе машинных экспериментов при условии, что система одновременно с переводом каждой фразы выдает подробный протокол его получения.

3. *Стандартизованность форматов описания рабочих языков.* Оба рабочих языка системы ЭТАП-2 – английский и русский – описываются по единым достаточно общим схемам. Это открывает возможность включать в систему новые рабочие языки без изменения форматов описания и алгоритмов анализа, преобразования и синтеза текстов.

4. *Неориентированность лингвистического обеспечения* (за исключением узко терминологического словаря) на какую-либо одну предметную область. Это свойство системы является прямым следствием полноты лингвистических моделей рабочих языков.

Морфология и синтаксис рабочих языков рассчитаны на переработку самых различных научно-технических текстов, т.е. учитывают весьма широкий спектр форм и конструкций, встречающихся в них.

Что касается *словарей*, то в них около половины всей лексики составляют термины и терминологические слова, более или менее специфичные для данной предметной области. Однако остается несколько тысяч слов (и

связанных с ними правил), которые являются словами общего распространения, т.е. не ориентированы на узкую предметную область. Словарные статьи, включенные в словарь, достаточно представительны и принципиальны, т.е. обеспечивают анализ и перевод данного слова не в одном-двух реально встретившихся текстах, а в широком спектре релевантных для электротехники контекстов, как они представлены в специальных терминологических словарях.

Относительная лингвистическая полнота и принципиальность лингвистического обеспечения системы ЭТАП-2 открывают возможность использования системы или отдельных ее компонентов для переработки текстов самых разных предметных областей.

Указанные свойства лингвистического обеспечения системы ЭТАП-2 имеют и некоторые *теоретические импликации*.

Прежде всего можно заметить, что качество перевода тем выше, чем полнее и принципиальнее модель языка, составляющая лингвистический компонент системы АП (воздействие лингвистики на АП: достаточно полная модель языка как гарант высокого качества АП).

Кроме этой зависимости качества АП от полноты и принципиальности лингвистических моделей, существует и обратная зависимость качества лингвистических моделей от результатов функционирования систем АП. Автоматический перевод оказывается экспериментальным полигоном для лингвистики, полигоном, где она может проверять свой научный аппарат и оценивать основательность своих научных построений. Если в системе АП эксплуатируются серьезные модели естественных языков, то такая система начинает оказывать глубокое стимулирующее влияние на лингвистику. В ходе машинных экспериментов уточняются уже известные лингвистические правила и формулируются новые, осознается необходимость новых лингвистических понятий, вызываются к жизни целые теоретические концепции. Короче говоря, создаются предпосылки для формирования новой научной дисциплины – экспериментальной лингвистики, которая в качестве инструмента доказательства использует естественно-научную методологию поиска научной истины (обратное воздействие АП на лингвистику: высококачественный АП как естественная основа для постоянного совершенствования лингвистических моделей).

6.1. Качество переводов

Самой естественной, общепонятной и наглядной характеристикой любой системы АП являются, конечно, не ее технические параметры, а выполняемые ею переводы. ЭТАП-2 переводил оригинальные английские тексты двух типов: 1) связный текст (статья, реферат и т.п.), 2) заголовки патентов сводной базы данных INPADOK. Перевод в обоих случаях осуществлялся пофразно, без какого-либо вмешательства человека в процесс перевода, причем каждая фраза переводилась единственным образом. Иными словами, морфологическая, лексическая и синтаксическая омонимия и полисемия раз-

решались (быть может, не всегда правильно) до такой степени, что число морфологических, лексических и синтаксических альтернатив в любой точке и в любом отрезке предложения сокращалось до одной. Удовлетворительный перевод получался примерно в половине случаев. Неудовлетворительный перевод в остальных случаях объяснялся чаще всего отсутствием какой-то информации в словаре или (реже) недосмотрами в конкретных правилах. В большинстве таких случаев была возможна коррекция словарной информации или правил, не разрушающая дееспособности системы в целом.

В целом авторы считают возможным оценить русские тексты, построенные ЭТАПом-2, как машинные переводы достаточно хорошего качества, в принципе сравнимые с человеческими.

6.2. Идеология и архитектура ЭТАПа-2

По своей идеологии и архитектуре ЭТАП-2 является продолжением и развитием системы перевода с французского языка на русский ЭТАП-1.

Назовем *четыре важнейших отличия* ЭТАПа-2 от ЭТАПа-1.

1. Прежде всего перед ЭТАПом-2 была поставлена *новая цель* – более высокий уровень разработки, чем чисто экспериментальный, с ориентацией на возможности дальнейшего практического использования. Отсюда вытекают два новых требования к системе: а) перевод должен быть быстрым; б) система ЭТАП-2 должна быть изготовлена в виде подсистемы, которую можно было бы подсоединить к любой информационной системе, где возникает необходимость в переводе с английского языка.

2. Новая цель определила и *новую архитектуру* лингвистического обеспечения системы. Для ускорения работы системы все правила были разбиты на три типа – общие, частные (трафаретные) и словарные. *Общие правила*, составляющие не более одной четверти всех правил, участвуют в обработке любой фразы. *Частные и словарные правила* участвуют в обработке лишь тех фраз, словесный материал которых может потребовать их применения. Они активизируются словарными статьями соответствующих слов, в которых они либо упоминаются по имени (*трафаретные правила*), либо приводятся целиком (*словарные правила*). Таким образом, в ЭТАПе-2 была реализована идея оптимальной самонастройки системы на обработку текущей фразы. В результате время переработки фразы существенно сократилось, потому что теперь в ее анализе и переводе принимают участие не все правила системы, а лишь те, в которых может возникнуть необходимость.

3. ЭТАП-2 работает с *новым входным языком* – английским. Основная трудность, связанная со сменой языка, состоит в следующем.

Английский язык, из-за слабой формальной маркировки семантических и синтаксических отношений между словами во фразе, с точки зрения автоматической обработки по крайней мере на порядок сложнее французского и в особенности русского.

Действительно, части речи в нем почти не имеют характерных парадигм, и одно орфографическое слово сплошь и рядом бывает существитель-

ным, прилагательным и глаголом, а иногда даже наречием и предлогом; ср. ROUND – ‘круг’ (существительное), ‘круглый’ (прилагательное), ‘округлять’ (глагол), ‘кругом’ (наречие) и ‘вокруг’ (предлог).

Окончания разных частей речи омонимичны, ср. ROUNDS – ‘круги’ или ‘округляет’. Омонимичны и грамматические суффиксы, ср. ROUNDING – ‘округляющий’ или ‘округляя’, ROUNDED – ‘округлил’ или ‘округленный’. Грамматические суффиксы омонимичны словообразовательным, ср. ROUNDING – ‘округляющий’ или ‘округление’. Прилагательные и причастия не имеют никаких окончаний, и поэтому их согласование с существительными никак не выражено. Например: electric device – ‘электрическое устройство’ и electric devices – ‘электрические устройства’.

Подчинительные союзы и союзные слова могут опускаться. Например: He claimed he was ill => ‘Он утверждал, что был болен’, буквально ‘Он утверждал он был болен’, или The man I saw yesterday => ‘Человек, которого я видел вчера’, буквально ‘Человек я видел вчера’. Тем самым оказываются невыраженными подчинительные отношения.

В связи с этим в системе ЭТАП-2 введены правила реинтерпретации синтаксических структур, широко использующие знание лексического состава фразы.

4. ЭТАП-2 рассчитан на перевод не только текстов рефератов или статей, но и совершенно *нового жанра* текстов – заголовков патентов. Совершенно очевидно, что переводы, выполняемые системой, должны быть не только семантически и стилистически приемлемыми, но и технически грамотными. Это значит, что они должны быть настроены на терминологические эталоны, принятые в электротехнике, в частности, в патентоведении.

В лингвистическом обеспечении ЭТАПа-2 можно выделить следующие крупные блоки (выход каждого предшествующего блока является входом для предыдущего):

- (1) – морфологический анализ исходного английского предложения;
- (2) – синтаксический анализ английского предложения;
- (3) – нормализация синтаксической структуры английского предложения;
- (4) – преобразование нормализованной английской структуры в нормализованную русскую;
- (5) – развертывание нормализованной структуры в синтаксическую структуру будущего русского предложения;
- (6) – синтаксический синтез русского предложения;
- (7) – морфологический синтез русского предложения.

Все семь блоков переработки фразы требуют обращения к словарям. На этапах (1) и (7) происходит обращение к морфологическим словарям системы – английскому и русскому соответственно. До начала этапа (2) происходит однократное обращение к комбинаторному словарю (КС) английского языка. Для каждой лексемы предложения (точнее, для множества омонимичных

лексем в каждой позиции) из КС извлекается вся лексикографическая информация, необходимая для синтаксического анализа, нормализации и перевода английского предложения. На этапах (4)-(6) происходит обращение к русскому КС, откуда извлекается лексикографическая информация, необходимая для «доводки» русской синтаксической структуры и для ее морфологизации.

Английский КС включает следующие типы информации о лексеме: 1) ее номер в КС; 2) часть речи; 3) тривиальный перевод на русский язык; 4) синтаксические признаки; 5) семантические признаки (дескрипторы); 6) модель управления лексемы, а именно сведения о количестве ее (семантических) актантов, о способах их выражения и о требованиях к синтаксическим или семантическим свойствам ее потенциальных актантных зависимых; 7) имена трафаретных правил, необходимых для переработки фраз с данной лексемой, с указанием значений параметризованных переменных, если они есть в данном правиле; 8) словарные правила, необходимые для переработки фраз с данной лексемой.

Русский КС включает аналогичные типы информации, за исключением перевода и семантических признаков лексемы.

Вопросы и задания к Теме 6

1. Как расшифровывается аббревиатура ЭТАП-2?
2. Какие существуют поколения систем автоматического перевода?
3. Какие различают уровни разработки систем АП?
4. Опишите классификацию систем АП по типу перевода.
5. Какие имеются системы АП в соответствии с назначением перевода?
6. Какие различаются системы АП по числу рабочих языков системы?
7. Какие типы рабочих языков системы встречаются в системах АП?
8. Опишите степень корреляции между типами систем АП.
9. Охарактеризуйте систему ЭТАП-2 в соответствии с известными типами классификации систем АП.
10. Чем отличается система ЭТАП-2 от других систем АП второго поколения в лингвистическом аспекте?
11. Каковы отношения между входным и выходным языками системы ЭТАП-2?
12. Что подразумевается под декларативностью задания лингвистических знаний?
13. Объясните принцип стандартизованности форматов описания рабочих языков системы ЭТАП-2.
14. Следствием какого свойства является неориентированность лингвистического обеспечения системы ЭТАП-2?
15. Каков состав словарей рабочих языков ЭТАП-2?
16. Какие теоретические импликации можно отметить у лингвистического обеспечения ЭТАП-2?
17. Назовите типы оригинальных английских текстов, переводимых системой ЭТАП-2.
18. Назовите важнейшие отличия системы ЭТАП-2 от системы ЭТАП-1.
19. Дайте определение словарным и трафаретным правилам.
20. В чем сложность синтаксических отношений в английском языке по сравнению с французским?
21. Опишите правила реинтерпретации синтаксических структур.
22. Обоснуйте необходимость настройки переводов, выполняемых ЭТАП-2, на терминологические эталоны, принятые в патентоведении.
23. Какие крупные блоки можно выделить в лингвистическом обеспечении ЭТАПа-2?
24. Какие типы информации о лексеме включает английский комбинаторный словарь?
25. Каково отличие русского комбинаторного словаря от английского?

Литература к Теме 6

Лингвистическое обеспечение системы ЭТАП-2 / Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др. – М.: Наука, 1989.

Тема 7. Лингвистическая модель анализа текста

7.1. Морфологический анализ

7.1.1. Назначение этапа морфологического анализа.

Условные обозначения морфологических объектов

На вход морфологического компонента системы ЭТАП-2 поступает английское предложение. *Задача морфологического анализа* состоит в построении морфологической структуры этого предложения.

Морфологический анализ работает *пословно*. Это означает, что словоформы входного предложения обрабатываются по отдельности, без обращения к какому-либо контексту. Для каждой словоформы строится совокупность всех ее морфологических структур (МС); в частном случае, если словоформа неомонимична, эта совокупность содержит единственную МС. Под *МС словоформы* понимается имя соответствующей лексемы с полной грамматической характеристикой.

Грамматические характеристики и названия частей речи в системе представляются простыми, по возможности приближенными к традиционным. Ниже приводится полный список условных обозначений.

Части речи:

S – существительное;

V – глагол;

A – прилагательное;

ADV – наречие;

PR – предлог;

CONJ – союз;

NUM – числительное;

ART – артикль;

PART – частица;

COM – композит (первая часть сложных слов типа lexico-semantic ‘лексико-семантический’).

Число:

SG – единственное:

а) для существительных: book, author, author’s;

б) для глаголов: has, breaks;

PL – множественное:

а) для существительных: books, children, authors’;

б) для глагола be: are, were.

Лицо (глаголов):

1-P – первое лицо (только для глагола be): am, was;

3-P – третье лицо: has, breaks, was.

Степени сравнения (прилагательных и наречий)
COMP – сравнительная: better, more, less(er), greater;
SUP – превосходная: best, most, least, greatest.

Падежи:
NOM – именительный (для личных местоимений): I, you, he, they;
OBJ – объектный (для личных местоимений): me, him, her;
POSS – притяжательный: John's, children's, brother's, one's.

Время (глаголов):
PRS – настоящее;
PST – прошедшее.

Репрезентации (неличные формы глагола):
INF – инфинитив (только для глагола be): be;
ING – причастие I;
PP – причастие II;
MF – основная форма глагола.

Этап морфологического анализа разбивается на три блока: 1) предморфологического анализа, 2) выделение безусловных оборотов и 3) собственно морфологического анализа.

7.1.2. Предморфологический анализ

Начальным элементарным объектом морфологического анализа является *словоформа входного текста*, которая с алгоритмической точки зрения определяется как графически цельнооформленная последовательность буквенных символов и определенных знаков, ограниченная с двух сторон пробелами. Поэтому до начала морфологического разбора необходимо произвести целый ряд мелких операций по *расклейке словоформ*, чтобы привести их к каноническому виду. Эту задачу и выполняет блок предморфологической обработки. Объекты, с которыми приходится иметь дело этому блоку, подразделяются на три категории:

1) записанная слитно словоформа cannot, которую необходимо просто расклеить, т.е. переписать в виде: can not;

2) словоформы, пишущиеся через дефис в композитной конструкции, например: high-frequency modulations 'высокочастотные колебания', shoulder-fired rocket 'ракетный снаряд, запускаемый с плеча', three-legged table 'трехногий стол', input-output 'ввод-вывод'. В принципе они должны быть разъединены и переписаны в виде high – frequency, shoulder – fired. Однако поскольку в английском языке дефис ставится не только между разными словами, но и внутри одного слова (non-alignment 'неприсоединение', re-interpretation 'реинтерпретация', great-grandfather 'прадед' и т.д.), до расклей-

ки необходимо убедиться в том, что обрабатываемая словоформа с дефисом не включена в словарь;

3) словоформы, записанные через апостроф:

а) личное, указательное или вопросительное местоимение + личная форма глагола: I'm, they've, we're, that'll, who's, he'd;

б) личная форма глагола + отрицательная частица not: isn't, wasn't, aren't, weren't, doesn't, haven't, mustn't, mightn't, couldn't, shouldn't, wouldn't.

Сложность обработки этих словоформ связана с двумя обстоятельствами. С одной стороны, их необходимо отличать от слов типа o'clock 'часов' (в обозначении времени, ср. 5 o'clock '5 часов'), ha'penny 'полпенса' и т.п., у которых апостроф входит в состав слова. С другой стороны, их необходимо отличать от форм притяжательного падежа типа John's, master's, authors'. Даже когда эта проблема решена, остается еще неоднозначность словоформ, оканчивающихся на 'd и на 's и представляющих собой редуцированные формы соответственно had/would и is/has.

Таким образом, при обработке слов с апострофом после целого ряда проверок от них отделяется их конечная часть, а именно: 'm, 're, 've, 'll, n't, 'd, 's. Между начальной и конечной частями оставляется пробел.

7.1.3. Выделение безусловных оборотов

По сравнению с блоком предморфологического анализа блок выделения безусловных оборотов выполняет прямо противоположную функцию. Если задача первого из них заключается в расклейке слов, оформленных в тексте в виде единой словоформы, то *второй собирает несколько изолированных словоформ в единое целое.*

В английском, как и в любом другом, языке существует довольно большое количество устойчивых словосочетаний, или оборотов, которые в силу традиции пишутся раздельно, а по существу представляют собой одно слово. Между элементами безусловного оборота не могут вставляться посторонние слова, а сами эти элементы непереставимы и являются грамматически неизменяемыми формами, ср.: A FEW 'несколько', AS IF 'как если бы', BECAUSE OF 'из-за', FOR INSTANCE 'например', IN ACCORDANCE WITH 'в соответствии с', IN GENERAL 'вообще', ONE ANOTHER 'друг друга', PREVIOUS TO 'до, перед' и т.д. Такими оборотами при синтаксическом анализе удобно оперировать как цельными объектами.

Выделение безусловных оборотов производится следующим образом. Имеется множество словосочетаний, которые должны описываться как цельные лексические единицы. Оно организовано в виде упорядоченного по алфавиту списка кортежей словоформ словаря безусловных оборотов. Кроме того, все начальные элементы оборотов (первая словоформа оборота) получают в морфологическом словаре особую помету, которая при работе морфологического анализа служит сигналом к включению блока выделения безусловных оборотов. При появлении словоформы с данной пометой в тексте анализируемой фразы эта словоформа и ее ближайший правый контекст сли-

чаются со словарем оборотов. В случае, если последовательность словоформ в тексте совпадает с одним из кортежей в словаре, она опознается как оборот и пробелы между отдельными словоформами заполняются вспомогательными знаками. В дальнейшем безусловный оборот интерпретируется уже как единый объект, который включается в комбинаторный словарь на тех же основаниях, что и обычная лексема, и снабжается информацией тех же типов.

7.1.4. Морфологический анализ в собственном смысле

Блок морфологического анализа состоит из словаря основ, списка окончаний, списка грамматических характеристик (приписываемых основе и/или окончаниям) и вспомогательных средств компактной записи морфологической информации (стандартных парадигм, масок, форматов).

7.1.4.1. Словарь основ

Словарь основ представляет собой центральную часть аппарата морфологического анализа. По своему лексическому составу он является отраслевым: помимо общеупотребительной лексики, в него вошли термины по электротехнике из частотного словаря, полученного в результате обработки нескольких тысяч рефератов по электротехнике на английском языке.

Словарь основ в системе ЭТАП-2 содержит около 12 тыс. словарных статей. Словарная статья состоит из заголовка и одной или нескольких подстатей. Заголовок представляет собой основу, в частном случае совпадающую со словоформой, например CERTAINLY 'конечно'. Каждая подстатья включает в себя следующие три зоны:

1) номер подстатьи (NAR) – пятизначное число, являющееся номером соответствующей единицы комбинаторного словаря;

2) перевод (TRS) – русский эквивалент данной лексемы, записанный в своей словарной форме (т.е. в именительном падеже единственного числа в случае существительного, в инфинитиве несовершенного вида в случае глагола и т.д.). В зоне перевода может быть записано несколько русских эквивалентов, соответствующих различным значениям данной лексемы. В частном случае зона перевода может оставаться незаполненной (например, в словарной статье артикля THE или других служебных слов, не имеющих соответствия в русском языке). Информация, записанная в этой зоне, используется только при работе системы в режиме пословного перевода;

3) зона морфологической информации, состоящая из строк, в которых приводятся морфологические характеристики основы (СТН) и записываются сочетающиеся с ней окончания (LZ) с их морфологическими характеристиками.

7.1.4.2. Список окончаний

Список окончаний представляет собой полный перечень словоизменяемых окончаний английского языка. В общем случае они не совпадают с окончаниями, выделяемыми традиционной грамматикой. Это объясняется, в частности, тем, что в нашем варианте морфологии не описывается чередование гласных на стыке морфем. Поэтому для словоформ *happy* ‘счастливый’, *happier* ‘счастливее’, *happiest* ‘счастливейший’ дается следующий морфологический разбор: основа: *happ-*, окончание: *-y*, *-ier*, *-iest*.

Приведем *полный алфавитный список окончаний*, действующий в системе ЭТАП-2: *-a*; *-e*; *-ed*; *-en*; *-er*; *-es*; *-est*; *-ied*; *-ier*; *-iest*; *-ing*; *-is*; *-on*; *-s*; *-s'*; *-s*; *-t*; *-um*; *-y*; *-ying*; *-ies*.

7.1.4.3. Процедура морфологического анализа

На вход собственно морфологического анализа подается фраза, прошедшая предморфологический анализ и блок выделения безусловных оборотов. Конечный сегмент каждой словоформы фразы по очереди сличается со всем списком окончаний. Каждый раз, когда конечная часть словоформы совпадает с одним из окончаний списка, она отделяется, а начальная часть словоформы объявляется гипотетической основой. Например, словоформа *carries* получит следующие гипотетические интерпретации:

1) окончание: нулевое;

гипотетическая основа: *carries-*;

2) окончание: *-s* (например, окончание множественного числа обычного существительного: *boy* – *boys* ‘мальчик – мальчики’, или окончание 3-го лица настоящего времени глагола: *put* – *puts* ‘класть – кладет’);

гипотетическая основа: *carie-*;

3) окончание: *-es* (как, например, окончание множественного числа существительного: *bus* – *buses* ‘автобус – автобусы’, или окончание 3-го лица настоящего времени глагола: *push* – *pushes* ‘толкать – толкает’);

гипотетическая основа: *car-*;

4) окончание: *-ies* (как, например, окончание множественного числа существительного: *berry* – *berries* ‘ягода – ягоды’, или окончание 3-го лица настоящего времени глагола: *carry* – *carries* ‘нести – несет’);

гипотетическая основа: *car-*.

После этого происходит обращение к словарю основ, в котором отыскиваются словарные статьи с заголовками (основами), совпадающими с выделенными гипотетическими основами. Так, в разбираемом примере мы обнаружим в словаре словарные статьи с основой *caries*, представляющей лексемы *CARIES* ‘кариес’, и с основой *car*, представляющей лексемы *CAR* ‘автомобиль’, *CARE1* ‘забота’ и *CARE2* ‘заботиться’, но не обнаружим словарных статей с основами *carie* и *car-* (второй и третий варианты анализа).

Отобранные статьи прочитываются, чтобы установить, зафиксированы ли в их строках (приписанных им стандартных парадигмах или форматах) такие окончания, которые были выделены на предыдущем шаге. В результате этого просмотра полностью отсеется четвертый вариант анализа: ни у существительного CAR, ни у существительного CARE1 или глагола CARE2 нет словоформы caries. Тем самым остается только первый, правильный, вариант: основа caries + окончание # (ему приписывается характеристика SG). Эта MC (CARIES_{SG}) и будет результатом морфологического анализа.

7.1.5. Обработка текстовых элементов, не получивших морфологической интерпретации

После морфологического анализа предложения могут оказаться необработанными отдельные объекты текста, которые, хотя и удовлетворяют алгоритмическому представлению о словоформе как о последовательности символов «от пробела до пробела», но не имеют соответствия в морфологическом словаре. Эти объекты могут быть разной природы: во-первых, это реальные английские слова, не включенные в словарь в силу неполноты последнего; во-вторых, имена собственные, введение которых в словарь представляет собой особую проблему, решаемую всякий раз индивидуально; в-третьих, иноязычные вставки; в-четвертых, объекты, содержащие неалфавитные символы (различного рода формулы, цифры и т.п.); наконец, в-пятых, слова, попросту содержащие орфографические ошибки. Все такие объекты, за исключением цифровых, алгоритм морфологического анализа в неизменном виде подает на следующий этап анализа, приписывая им особую помету NID («неидентифицированная словоформа»). Таким образом, формула вида PL-1 получит «морфологическое» представление PL-1_{NID}.

В блоке синтаксического анализа имеется целый ряд синтагм, специально ориентированных на обработку слов с пометой NID, которые позволяют строить «аварийные» структуры для предложений, содержащих такие слова.

Обработка объектов, состоящих из цифр (5; 0,8; 6.000 и т.д.), производится следующим образом. Алгоритм морфологического анализа квалифицирует их как (количественные) числительные и приписывает помету части речи NUM. Тем самым в дальнейшем, на этапе синтаксического анализа, эти объекты смогут обрабатываться теми же синтагмами, что и (простые) числительные, записанные словами: FIVE ‘пять’, THOUSAND 1 ‘тысяча’ и т.д. при переходе к русской синтаксической структуре (на этапе TRADUCT) эти объекты получают синтаксические признаки, характерные для различных классов русских числительных, в зависимости от конкретного наполнения: так, цифровые последовательности 2; 0,8 получают признак «мал», цифровые последовательности 5; 6.000 получают признак «больш» и т.д., что обеспечит выработку правильных грамматических характеристик в выходном русском предложении (2 листа vs. 6.000 листов).

7.2. Синтаксический анализ

Блок синтаксического анализа является центральным и одновременно самым сложным компонентом системы ЭТАП-2. Он состоит из двух неравноценных в алгоритмическом отношении частей: небольшого вспомогательного блока предсинтаксического анализа, предназначенного для частичного снятия лексико-грамматической омонимии МС предложения по линейному контексту, и основного блока синтаксического анализа в собственном смысле, работающего с продуктом, получаемым на выходе предсинтаксического блока.

7.2.1. Предсинтаксический анализ

7.2.1.1. Назначение этапа предсинтаксического анализа

В каждом предложении, как правило, имеется значительное количество неоднозначных (омонимичных или полисемичных) словоформ. Для разрешения любых типов омонимии (грамматической, лексико-грамматической, лексической, синтаксической) в общем случае требуются весьма мощные языковые средства вплоть до семантических и даже энциклопедических. Однако в некоторых ситуациях для снятия грамматической и лексико-грамматической омонимии словоформ достаточно обратиться к их ближайшему контексту. Например, для правильной интерпретации слова *work* как инфинитива (а не как существительного) во фразе *they must work hard* 'Они должны много работать' достаточно убедиться в том, что непосредственно слева от нее находится модальный глагол. Таких ситуаций насчитывается довольно много, и представляется естественным ввести специальные правила для их обработки, которые действовали бы после морфологического анализа предложения, но до этапа синтаксического анализа. Это правило удобно организовать в отдельный вспомогательный анализирующий компонент системы – блок предсинтаксического анализа. С помощью правил этого блока еще до начала синтаксического анализа удастся устранить из переводимого предложения большое число «паразитических» омонимов, которые в противном случае породили бы множество заведомо ложных синтаксических гипотез.

7.2.1.2. Обзор правил предсинтаксического анализа

1. Разрешение грамматической омонимии «личный глагол в прошедшем времени – причастие II» в пользу причастия.
2. Разрешение лексико-грамматической омонимии «глагол в третьем лице единственного числа настоящего времени – существительное во множественном числе».
3. Разрешение омонимии «основная форма глагола – существительное/прилагательное».
4. Разрешение омонимии «причастие I – существительное» в пользу существительного.

5. Разрешение омонимии «наречие с признаком ADPHRAS – предлог». (Признак ADPHRAS приписывается наречиям, которые способны выступать в составе фразового глагола: give up ‘сдаваться’, let in ‘впускать’, come to ‘приходить в себя’. Большинство из этих наречий омонимичны предлогам.)

6. Разрешение лексической омонимии.

7.2.2. Синтаксический анализ (в собственном смысле)

7.2.2.1. Назначение этапа синтаксического анализа

Блок синтаксического анализа (СинтА) получает на входе редуцированную в результате работы предсинтаксического анализа МС и строит на ее основе синтаксическую структуру (СинтС) данного предложения.

Из всех известных *способов формального представления синтаксической структуры предложения* – деревьев зависимостей, систем составляющих, различных комбинаций этих способов – в системах ЭТАП выбран первый, который лучше других соответствует синтаксическому строю английского и русского языков.

Таким образом, *СинтС предложения* – это размеченное и расположенное *дерево зависимостей*, в узлах которого стоят МС словоформ, а связывающие их стрелки помечены именами синтаксических отношений, специфичных для данного языка. В СинтС оказывается полностью снятой неоднозначность словоформ: каждой словоформе предложения, даже первоначально омонимичной, соответствует ровно один узел в синтаксическом дереве.

Основными средствами синтаксического анализа являются лингвистические правила двух типов – *синтагмы* и *правила предпочтения*.

7.2.2.3. Общая схема алгоритма синтаксического анализа.

Два типа синтаксических правил – синтагмы и правила предпочтения

Синтагмы являются главным типом синтаксических правил. С формально-алгоритмической точки зрения синтагма представляет собой правило, которое соединяет МС двух словоформ предложения в гипотетическое бинарное поддерево. Каждая синтагма содержит набор условий, которые должны быть удовлетворены для того, чтобы некоторые две словоформы могли быть связаны данным синтаксическим отношением (СинтО). Этот набор условий подразделяется на две совокупности: 1) *линейные условия*, в которых фиксируется информация о главном и зависимом членах данного СинтО, их абсолютном и относительном размещении в предложении, наличии или отсутствии в предложении тех или иных третьих слов с определенными свойствами, знаках препинания и пр. и 2) *древесные условия*, задающие разного рода ограничения на сочетаемость разных бинарных поддеревьев в пределах одной правильной СинтС.

Если на очередном шаге алгоритма не происходит ни сокращения потенциального множества гипотез, ни пополнения подмножества окончательных связей, а граф потенциальных гипотез все еще не является деревом,

вступают в действие синтаксические правила второго типа – *правила предпочтения*.

Эти правила отдают предпочтение одним синтаксическим гипотезам перед другими на основании простых вероятностных соображений. Например, из двух претендентов на роль подлежащего предпочтается тот, который расположен ближе к сказуемому; актантные связи предпочтаются атрибутивным и сирконстантным; более тесные сирконстантные связи предпочтаются менее тесным (например, ограничительная связь предпочтается обстоятельственной).

Исчерпав материал правил предпочтения, алгоритм еще раз возвращается ко всему комплексу фильтрующих процедур.

Если и после этого возврата остаются лишние гипотезы, начинается действие *блок перебора альтернатив*. Он поочередно объявляет оставшиеся неокончательные гипотетические связи окончательными и вновь приводит в действие фильтрующие механизмы. При этом работа начинается с тех неокончательных гипотез, которые связывают слова, близко расположенные друг к другу, поскольку именно такие гипотезы с наибольшей вероятностью оказываются истинными. Если на каком-то шаге алгоритм приходит к противоречию (т.е. принятое допущение привело к нарушению какого-либо из свойств правильной СинтС), блок перебора альтернатив отказывается от этого допущения и приступает к разбору следующей альтернативы.

Алгоритм СинтА работает до тех пор, пока не будет построена первая СинтС, удовлетворяющая всем сформулированным требованиям: процедура перебора альтернатив устроена таким образом, что правильная СинтС чаще всего и получается первой.

В некоторых случаях в результате работы СинтА не удается построить СинтС предложения: граф потенциальных гипотез после применения всех фильтрующих механизмов оказывается несвязным. Причиной этого обстоятельства может быть неполнота или противоречивость формального описания синтаксиса, не замеченная ранее и проявившаяся в текущей фразе. В подобных ситуациях ЭТАП-2 переходит на аварийный режим работы – пословный перевод.

7.3. Глубинно-синтаксический анализ

7.3.1. Общие положения. Нормализованная структура

Глубинно-синтаксический анализ должен приводить к построению глубинной структуры предложения, приближенной к его логическому прототипу. В идеале в глубинной структуре должны быть проинтерпретированы все синтаксические отношения естественного языка, сняты чисто формальные языковые элементы (например, пустые предлоги), восстановлены языковые элементы, материально не представленные в предложении (опущенные актанты, нулевые лексемы и т.п.), и отмечены кореферентные связи между именными группами. Тем самым в глубинной структуре в значительной мере

снимается морфологическое и синтаксическое своеобразие конкретного естественного языка.

В данной системе, однако, мы не ставим перед собой в полном объеме задачу получения такой «идеальной» структуры. Глубинно-синтаксический анализ мы рассматриваем лишь как движение в сторону глубинной структуры. Для целей перевода часто оказывается достаточным пройти лишь часть пути, отделяющего собственно синтаксическую структуру предложения от его глубинной структуры. Та точка, где можно остановиться, определяется мерой типологического расхождения между входным и выходным языками. Если в каких-то отношениях выходной язык так же отличается от логического прототипа, как и входной, то нет смысла проводить соответствующие глубинные преобразования. Например, в большинстве случаев русский язык проводит сочинительное сокращение точно так же, как английский, и поэтому восстановление сокращенных элементов сочинительной конструкции при переводе с английского на русский оказывается излишним.

Возьмем в качестве примера следующее английское предложение:

It has turned out that the system they are planning to develop should be modelled after and adapted to the European system so that it can meet all the specifications required. ('Оказалось, что система, которую они планируют разработать, должна быть смоделирована по образцу европейской системы и приспособлена к ней так, чтобы она могла удовлетворять всем требуемым техническим условиям'.)

В синтаксической структуре этого предложения проводятся следующие глубинные преобразования.

1. Эксплетивное *it*, являющееся чисто формальным подлежащим предложения, устраняется и заменяется в этой синтаксической позиции реальным первым актантом предиката **TURN OUT** 'оказываться' – придаточным предложением, вводимым союзом **THAT**1.

2. Аналитические формы *has turned, are planning, be modelled* стягиваются в единые узлы: **TURN**_{prs, perf, sg, 3-p}, **PLAN**_{2prs, progr, pl}, **MODEL**_{2inf, pass}.

3. Устраняются артикли, а соответствующим существительным приписываются характеристики *def* или *indef*, например *the system* => **SYSTEM**_{def}.

4. В релятивную конструкцию *the system they are planning* вводится опущенное по правилам английского синтаксиса союзное слово **WHICH**: *the system they are planning* => *the system which they are planning*.

5. Устраняется приинфинитивная частица **TO**2, а у формы *develop* характеристика *mf* заменяется на *inf*.

6. В сочинительной конструкции некоторые характеристики передаются по цепочке всем однородным членам; в результате словоформа *adapted* приобретает следующий вид: **ADAPT**_{inf, pass}.

7. В специфичной для английского языка сочинительной конструкции с висячим предлогом у предлога восстанавливается опущенное зависимое, ко-

торое получает помету *antec* (антецедент); у конечного же элемента сочинительной цепочки реальное зависимое заменяется на фиктивную лексику (ФИКТ-ЛЕКС). В результате конструкция *modelled after and adapted to the European system* преобразуется в конструкцию *modelled after the European system_{antec} and adapted to ФИКТ-ЛЕКС*. Это позволяет заменить впоследствии лексику ФИКТ-ЛЕКС личным местоимением, соответствующим по роду и числу его антецеденту – в нашем случае русскому слову СИСТЕМА.

8. Устраняются пустые предлоги, а словам, зависящим от этих предлогов, приписывается вспомогательная характеристика *prep*.

9. В контексте модального глагола *can* глагольная словоформа *meet* с характеристикой *mf*, по существу скрывающей целый ряд омонимичных словоформ, интерпретируется как инфинитив.

В результате всех этих преобразований получается *нормализованная* структура предложения (считать ее глубинной структурой нельзя: здесь, в частности, проинтерпретированы далеко не все синтаксические отношения, восстановлены далеко не все актаны.)

7.3.2. Реинтерпретация СинтС

Центральная задача этапа *нормализации* СинтС состоит в снятии морфолого-синтаксической специфики английского языка. Поэтому с содержательной точки зрения *правила нормализации* можно рассматривать как перечень основных морфологических и синтаксических явлений, составляющих специфику английского языка по сравнению с русским, и способов снятия этой специфики.

7.3.3. Обзор правил нормализации

7.3.3.1. Стяжение аналитических форм глагола

Поскольку система глагольных форм в английском языке резко отличается от русской (в частности, гораздо более высокой степенью аналитичности), мы сочли целесообразным на этапе нормализации стягивать все аналитические формы глаголов в единые узлы; при этом грамматические значения, выражаемые вспомогательными формами, представляются в виде грамматических помет при полнозначном глаголе.

1. Стяжение форм перфекта. Как в личных, так и в неличных формах вспомогательный глагол HAVE перерабатывается в характеристику *perf* (*perfect*).

2. Стяжение форм прогрессива. Вспомогательный глагол BE перерабатывается в характеристику *progr* (*progressive*).

3. Стяжение форм будущего времени. Вспомогательные глаголы WILL и SHALL перерабатываются в характеристику *fut* (*future*).

4. Устранение вспомогательного глагола DO.

а) В вопросительных и отрицательных предложениях DO просто стирается, а полнозначный глагол ставится на его место и принимает грамматические характеристики DO;

б) В эмфатических предложениях вспомогательный глагол DO перерабатывается в характеристику *emph* (*emphasis*).

5. Стяжение форм пассива. Вспомогательный глагол BE перерабатывается в характеристику *pass* (*passive*).

7.3.3.2. Интерпретация морфологических характеристик

1. Интерпретация характеристики *mf*. Глагольная характеристика *mf* (*main form*) в разных условиях может соответствовать нескольким более конкретным характеристикам:

а) *mf* перерабатывается в характеристику *prs* (*present*), если при глагольной форме с характеристикой *mf* имеется подлежащее (предикативное зависимое);

б) *mf* перерабатывается в характеристику *inf* (*infinitive*), если словоформа с характеристикой *mf* зависит от частицы TO, а также модальных и других глаголов, управляющих инфинитивом без частицы TO (*can, must, make, help, see, hear* и т.п.);

в) *mf* перерабатывается в характеристику *sbj* (*subjunctive*) в контексте глаголов, требующих сослагательного наклонения;

г) *mf* перерабатывается в характеристику SØ. Характеристика SØ приписывается таким английским глагольным словоформам, которые при переводе на русский язык должны подвергнуться номинализации, в частности превратиться в отглагольное существительное, или (реже) в придаточное предложение, вводимое конструкцией *to, что*.

Характеристика SØ заменяет *mf* в композитной конструкции типа *read circuit* 'схема считывания'. Помимо *mf*, с помощью SØ интерпретируется характеристика *ing*.

2. Интерпретация характеристики *ing*. Уже указывалось, что *ing*-овые формы глаголов омонимичны, т.е. выполняют в предложении различные синтаксические функции. После того как формы прогрессива, содержащие причастия на *-ing*, стянуты в один узел, по существу, остаются три основных употребления *ing*-овых форм: в субстантивной функции (*his coming*), в функции определения (*the boy reading a book*) и в функции обстоятельства (*we rushed on, the dog running behind*):

а) в субстантивном употреблении (в контексте предлогов, детерминативов и т.д.) характеристика *ing* перерабатывается в SØ;

б) в определительных и обстоятельственных употреблениях характеристика *ing* перерабатывается в характеристику *part* (*active participle*).

В дальнейшем характеристика *part* преобразуется либо в русское действительное причастие, либо в деепричастие, либо в личный глагол (абсолютный оборот).

3. Передача морфологических характеристик по сочинительной цепочке. В случае, если первый из членов сочинительной конструкции (т.е. вершина этой конструкции) в результате применения каких-либо правил нормализации получает новые морфологические характеристики, эти характеристики передаются от вершины всем однородным членам; при этом ненужные старые характеристики стираются.

7.3.3.3. Опущение служебных слов

На этапе нормализации СинтС из нее устраняются два класса служебных слов. Первый класс составляют сильноуправляемые пустые предлоги, для которых соответствие в русском переводе отыскивается вне словарных статей этих предлогов, по модели управления русского эквивалента их синтаксического хозяина; ср. *depend on* 'зависеть от'. Ко второму классу относятся такие служебные лексемы, для которых словесного соответствия в русском переводе не находится вовсе (артикли, эксплетивные элементы и т.д.). В этот же класс входят вспомогательные глаголы и приинфинитивная частица *TO*, устранение которых из СинтС описано в предыдущих пунктах, поскольку оно сопровождается порождением новых грамматических характеристик.

1. Опущение пустых предлогов:

а) Опущение предлогов в актантных конструкциях. Сильноуправляемые предлоги, замещающие какую-либо валентность своего хозяина, устраняются из СинтС, если это не влечет за собой утраты какой-либо (семантической) информации, иными словами, если либо соответствующая валентность управляющего слова вообще не выражается другими средствами (ср.: *depend on*, *approve of*, *listen to*), либо средства ее выражения синонимичны (*conform to* и *conform with*, *arrive at* и *arrive in* и т.д.; тот факт, что разные предлоги имеют в некоторых подобных случаях разную сочетаемость, не имеет значения). При опущении пустого предлога слово, зависящее от него в СинтС, переподчиняется синтаксическому хозяину предлога.

Правило опущения пустых предлогов применяется и в тех случаях, когда в модели управления хозяина, помимо этого предлога, имеются и полнозначные предлоги.

В случае же, если глагол *look* управляет предлогом *after* или *for* и тем самым требует особых лексических единиц при переводе (*look after* 'наблюдать', *look for* 'искать'), предлог не опускается.

По ряду соображений технического характера на этапе нормализации не опускаются предлоги, замещающие валентности фразовых глаголов (*do away with*, *go in for*, *walk out on* и т.д.). Такие словосочетания целиком обрабатываются на этапе перевода.

Не опускаются также висячие предлоги (в конструкциях типа *This is the man I spoke about* 'вот человек, о котором я говорил'; *I have nobody to speak to* 'мне не с кем разговаривать' и т.д.), поскольку в противном случае одновременно с предлогом мы бы устранили имеющуюся в СинтС информацию о синтаксическом отношении, связывающем предлог с его хозяином.

б) Опускание предлогов в атрибутивных конструкциях. Если атрибутивным зависимым какого-либо слова является предлог OF (и только этот предлог), то он опускается.

В подобных случаях в блоке синтаксического синтеза у русского атрибутивного зависимого вырабатывается родительный падеж.

в) Опускание предлогов в сочинительных конструкциях. Пустые предлоги опускаются и тогда, когда они не непосредственно зависят от управляющего слова по одному из актантных или по атрибутивному СинтО, а входят в состав сочинительной цепочки.

2. Опускание артиклей. Все артикли в СинтС предложения опускаются, а их синтаксические хозяева приобретают пометы *def* или *indef*.

3. Опускание эксплетивных элементов. Местоимения IT и THERE, образующие вершину эксплетивной конструкции, опускаются, а их синтаксические зависимые переподчиняются сказуемому. Тем самым эксплетивная конструкция, которая не имеет соответствия в русском языке, преобразуется в предикативную.

4. Опускание подлежащего IT во вводном обороте.

5. Опускание прилагательных и наречий в тавтологических глагольных конструкциях. Имеются в виду обороты типа *kill dead, wipe clean, fall down*, где смысл прилагательного или наречия, по существу, вкладывается в смысл глагола и не влияет на перевод последнего.

7.3.3.4. Введение дополнительных узлов

Дополнительные узлы вводятся в СинтС для прояснения некоторых сложных конструкций с незаполненными синтаксическими позициями.

1. Введение WHICH в релятивную конструкцию.

2. Развертывание сочинительных цепочек:

а) В английском языке сочинительное сокращение, как правило, производится так же, как в русском. В некоторых случаях, однако, английский язык идет дальше русского, допуская при сочинении наличие ненасыщенных предлогов: *flowers behind and in front of the house* 'цветы за домом и перед ним'. Поскольку в русском языке такое невозможно, на этапе нормализации подобные конструкции разворачиваются, а именно: именная группа перевешивается к первому (ненасыщенному) предлогу и получает помету *antes*, а вместо этой именной группы в СинтС возникает временный элемент (ФИКТ-ЛЕКС), который позднее будет заменен русским личным местоимением, грамматические характеристики которого переносятся от лексемы с пометой *antes*, к этому моменту уже переведенной на русский язык;

б) К описанному случаю примыкает сочинительная конструкция типа *turn on and off* 'включать и выключать', где, по существу, сочиняются два фазовых глагола *turn on* и *turn off*.

7.3.3.5. Изменение порядка слов

Изменение порядка слов на этапе нормализации СинтС производится только в случаях значительного расхождения между английским и русским языками. В частности, порядок слов меняется в непроективных конструкциях с придаточным определительным (*a new method is discussed which allows*), а также в конструкциях с одиночным постпозитивным определением (*elements measured*).

7.3.3.6. Глубинизация синтаксических отношений

В глубинной структуре предложения все синтаксические отношения должны быть интерпретированы. При переводе с английского языка на русский, однако, необходимости в такой интерпретации обычно не возникает, и на этапе нормализации она систематически не проводится. Исключение составляют некоторые композитные конструкции, не имеющие прямых параллелей в русском языке, ср.: context-sensitive ‘чувствительный к контексту’, acid-resistant ‘устойчивый к кислоте, кислотостойкий’ и т.п. В таких ситуациях оказывается удобным интерпретировать композитное поверхностно-синтаксическое отношение (ПСО) путем актантного.

7.3.3.7. Правила реинтерпретации синтаксической структуры

Правила реинтерпретации служат для того, чтобы «подправлять» СинтС в тех случаях, когда из множества СинтС, потенциально допустимых для данного предложения, синтаксический анализ выбрал неправильную структуру. Дефекты структуры, которые устраняются правилами реинтерпретации, бывают двух основных видов: *неправильное разбиение словосочетания на составляющие* и *неправильный выбор вершины словосочетания*.

1. Неправильное разбиение на составляющие. В принятом нами способе представления синтаксической структуры предложения никакие средства эксплицитной передачи информации о составляющих (например, скобки) не применяются, поскольку в большинстве случаев для целей перевода такая информация оказывается излишней. Это происходит, с одной стороны, благодаря тому, что структуры составляющих в английском и русском языках достаточно близки, а с другой – благодаря тому, что имплицитно информация о составляющих удерживается в английской СинтС непосредственно отношениями зависимости и сохраняющейся информацией о порядке слов. Например, словосочетания, состоящие из двух адъективов и существительного, вида A^1A^2S практически всегда имеют в английском языке *одну и ту же структуру составляющих*: более далекий от существительного адъектив относится к сочетанию существительного с адъективом, стоящим рядом: $A^1(A^2S)$.

Такая же закономерность имеется и в русском языке. Благодаря этому достаточно сохранить при переводе таких сочетаний английский порядок слов, чтобы правильная структура составляющих оказалась автоматически перенесенной на русскую СинтС.

Однако бывают случаи, когда английские словосочетания допускают неоднозначное членение на составляющие и эти различия существенны для перевода. Рассмотрим характерные типы таких словосочетаний.

а) Реинтерпретация сочетаний типа $A(S^1S^2)$ в $(AS^1)S^2$. Сочетания адъектива с двумя существительными AS^1S^2 весьма близки к сочетаниям типа A^1A^2S . Так, в словосочетании *German staff cars* возможно двойное выделение составляющих, приводящее к различиям при переводе, ср.: *(German staff) cars* ‘машины немецкого штаба’ или *German (staff cars)* ‘немецкие штабные машины’.

Легко видеть, однако, что для передачи этих различий совсем необязательно вводить составляющие в структуру. Ту же информацию можно передать и конфигурацией отношений зависимости. В первом случае прилагательное *German* ‘немецкий’ связано с существительным *staff* ‘штаб’, а во втором – со словосочетанием *staff cars* ‘штабные машины’. Поскольку внешние связи сочетания в целом осуществляются через вершину этого сочетания, то второе осмысление вполне адекватно задается структурой, в которой прилагательное *German* присоединено к вершине сочетания *staff cars* – существительному *cars*.

Таким образом, задача установления структуры составляющих для сочетаний вида AS^1S^2 предстает как задача определения того, с каким существительным – S^1 или S^2 – синтаксически связан адъектив A .

б) Реинтерпретация сочетаний типа $(S^1S^2)S^3$ в $S^1(S^2S^3)$. Например: *government health service* ‘служба правительственного здоровья’ => ‘правительственная служба здравоохранения’; *zero insertion force* ‘сила нулевой вставки’ => ‘нулевая сила вставки’.

в) Реинтерпретация сочинительных конструкций типа S^1 and (S^2S^3) в $(S^1$ and $S^2)S^3$: *window and glass cleaner* ‘окно и моющее средство для стекла’ => ‘моющее средство для окна и стекла’.

2. Неправильный выбор вершины словосочетания. Вторая группа правил реинтерпретации обеспечивает перестройку СинтС в тех случаях, когда при синтаксическом анализе была неправильно выбрана вершина некоторого синтаксически омонимичного словосочетания. Этим правилам приходится иметь дело с ситуацией, несколько отличающейся от той, на которую настроены правила первой группы. Работа синтаксического анализа протекает таким образом, что словосочетания, допускающие неоднозначное разбиение на составляющие, всегда оказываются проанализированными единообразно.

Словосочетания, для которых предназначены правила второй группы, синтаксически анализируются иначе. Для них нельзя заранее предсказать, какая из омонимичных структур будет выбрана. Поэтому приходится для каждого типа словосочетаний, допускающих по два анализа, иметь пару про-

тивоположно направленных правил, преобразующих один вариант структуры в другой в зависимости от ее лексического наполнения.

Имеется три типа таких словосочетаний и соответственно три пары правил:

а) правила, обрабатывающие сочетания типа $V_{ing}S$: regulating winding 'регулирование обмотки' => 'регулирующая обмотка'; reducing temperature 'уменьшающаяся температура' => 'уменьшение температуры'.

В этом правиле учитывается то обстоятельство, что если в подобном сочетании при существительном имеется артикль или предлог, то возможна только комплетивная конструкция, а определительная является неправильной: regulating the winding, regulating of the winding. В этих случаях правило, естественно, не применяется.

б) правила, обрабатывающие сочетания типа $S^1V_{ing}S^2$: signal generating circuit 'сигнал, порождающий цепь' => 'цепь, порождающая сигнал'; circuit regulating current 'ток, регулирующий цепь' => 'цепь, регулирующая ток';

в) правила, обрабатывающие сочетания типа $S^1S^2S^3$: holder guide assembly 'блок держателя направляющей' => 'блок направляющей держателя'.

Вопросы и задания к Теме 7

1. В чём состоит задача морфологического анализа?
2. В чём суть пословного принципа работы морфологического анализа?
3. Каковы условные обозначения частей речи в система ЭТАП-2?
4. Какие блоки включает этап морфологического анализа?
5. Какую задачу выполняет блок предморфологической обработки?
6. Какие английские словоформы подвергаются процедуре расклейки?
7. Какова функция блока выделения безусловных оборотов?
8. Каковы признаки безусловных оборотов?
9. Опишите алгоритм выделения безусловных оборотов
10. Опишите состав блока морфологического анализа.
11. Что является центральной частью аппарата морфологического анализа?
12. Какова структура словарной статьи словаря основ в системе ЭТАП-2?
13. Какие характеристики имеет зона морфологической информации?
14. Каковы отличия списка окончаний английского комбинаторного словаря от окончаний, выделяемых традиционной грамматикой?
15. Опишите процедуру морфологического анализа системы ЭТАП-2.
16. Перечислите разновидности текстовых элементов, не получающих морфологическую интерпретацию после первичной обработки.
17. Из каких частей состоит блок синтаксического анализа?
18. Охарактеризуйте назначение этапа предсинтаксического анализа.
19. Перечислите правила предсинтаксического анализа.
20. Какой способ формального представления синтаксической структуры предложения наилучшим образом соответствует синтаксическому строю английского и русского языков?
21. Что собой представляет дерево зависимостей синтаксической структуры предложения?
22. Какие правила являются основными средствами синтаксического анализа?
23. Охарактеризуйте синтагмы как главный тип синтаксических правил.
24. В чём состоят особенности линейных и древесных условий?
25. Какова роль правил предпочтения?
26. Каковы функции блока перебора альтернатив?
27. В каких ситуациях ЭТАП-2 переходит на аварийный режим работы – пословный перевод?
28. Какова цель глубинно-синтаксического анализа?
29. Приведите примеры глубинных преобразований на материале английского языка.
30. Дайте определение нормализованной структуры предложения.
31. В чём состоит центральная задача нормализации синтаксической структуры?

32. Каковы основания для стяжения всех аналитических форм глаголов в единые узлы на этапе нормализации?
33. Опишите принципы стяжения форм перфекта, прогрессива, будущего времени, пассива и устранения вспомогательного глагола DO.
34. Охарактеризуйте принципы интерпретации характеристик *mf*, *ing* и передачи морфологических характеристик по сочинительной цепочке.
35. Какие классы служебных слов устраняются из синтаксической структуры на этапе ее нормализации?
36. Приведите примеры опущения пустых предлогов, артиклей, эксплетивных элементов, подлежащего IT во вводном обороте, а также прилагательных и наречий в тавтологических глагольных конструкциях.
37. Для чего в синтаксическую структуру вводятся дополнительные узлы?
38. Приведите примеры введения WHICH в релятивную конструкцию и примеры развертывания сочинительных цепочек.
39. В каких случаях на этапе нормализации синтаксической структуры производится изменение порядка слов?
40. Охарактеризуйте принцип глубинизации синтаксических отношений.
41. Для чего служат правила реинтерпретации синтаксической структуры?
42. Опишите два основных типа правил устранения дефектов структуры.
43. Каким образом исправляется дефект неправильного разбиения на составляющие? Приведите примеры.
44. Каким образом исправляется дефект неправильного выбора вершины словосочетания? Приведите примеры.

Литература к Теме 7

Лингвистическое обеспечение системы ЭТАП-2 / Ю.Д. Апресян, И.М. Богуславский, Л.Л. Иомдин и др. – М.: Наука, 1989.

Тема 8. Лингвистическое обеспечение собственно перевода

8.1. Назначение этапа перевода

Задача этапа перевода состоит в преобразовании английской нормализованной структуры в русскую. Она складывается из трех подзадач: *перевода лексики, перевода морфологических характеристик слов и перевода синтаксических отношений*. Эти подзадачи в определенной мере независимы друг от друга, хотя иногда они могут решаться в пределах одного правила.

Обратимся к каноническому примеру, чтобы сразу же создать представление о характере преобразований, осуществляемых на этапе перевода. Напомним каноническую фразу:

It has turned out that the system they are planning to develop should be modelled after and adapted to the European system so that it can meet all the specifications required. ('Оказалось, что система, которую они планируют разработать, должна быть смоделирована по образцу европейской системы и приспособлена к ней так, чтобы она могла удовлетворять всем требуемым техническим условиям'.)

На вход этапа TRUDUCT поступает английская нормализованная структура (см. разд. 2.3.1), которая подвергается следующим преобразованиям:

1) перевод морфологических характеристик, ср.: SYSTEM_{sg} => SYSTEM_{ед}; SPECIFICATION_{pl} => SPECIFICATION_{мн}; CAN_{prs} => CAN_{непрош,несов}; TURN_{perf,prs} => TURN_{прош,сов}; PLAN2_{progr,prs} => PLAN2_{непрош,несов}; MODEL2_{inf,pass} => MODEL2_{инф,страд};

2) тривиальный перевод лексики. Такой перевод не требует проверки каких-либо условий и состоит в замене одной английской лексемы одной русской: THAT1 => ЧТО1; SYSTEM => СИСТЕМА; PLAN2 => ПЛАНИРОВАТЬ; WHICH => КОТОРЫЙ; IT => ЭТО и т.п.;

3) нетривиальный перевод лексики. Такой перевод предполагает отсутствие пословных соответствий между английскими и русскими словами и/или те или иные проверки, ср.:

а) TURN OUT => ОКАЗЫВАТЬСЯ; SPECIFICATION => ТЕХНИЧЕСКОЕ УСЛОВИЕ; SHOULD => БЫТЬ ДОЛЖЕН;

б) MEET => УДОВЛЕТВОРЯТЬ (при условии, что MEET подчиняет по первому комплетивному отношению слово типа SPECIFICATION, CONDITION 'условие', DEMAND 'требование' и т.п.; в противном случае MEET переводится иначе: I met him => Я встретил его); REQUIRE => ТРЕБОВАТЬ_{прич, непрош, страд, несов} (при условии, что слово REQUIRE стоит в форме пассивного причастия и не имеет агентивного зависимого; в противном случае REQUIRE переводится иначе: required by the customer => потребованный покупателем); AFTER1 'по образцу' (при условии, что слово AFTER1 подчинено глаголу MODEL2 по второму комплетивному отношению; в против-

ном случае AFTER1 переводится иначе: He came after dinner => *Он пришел после обеда*);

4) перевод синтаксических отношений: they plan, European system.

После всех этих преобразований получается русская нормализованная структура.

8.2. Тривиальные и нетривиальные правила перевода

Как можно было видеть из разобранных примеров, правила перевода различаются по глубине того межъязыкового расхождения, которое они преодолевают. В соответствии с этим будем делить правила на *тривиальные* (соответствующие минимальному расхождению и сводящиеся, по существу, к переименованию) и *нетривиальные* (компенсирующие достаточно заметное расхождение).

Мера нетривиальности правила определяется двумя параметрами: сложностью контекстных условий, подлежащих проверке, и объемом производимого правилом преобразования. Чем меньше число условий надо проверить и чем проще проводимое преобразование, тем выше тривиальность правила. Предельный случай тривиальности – правило перевода лексики, не требующее проверки каких-либо контекстных условий и сводящееся к замене одной английской лексемы одной русской – той, которая записана в ее словарной статье в зоне TRANS. Столь же тривиальные правила имеются и в области перевода морфологических характеристик (*inf* => инф, *comp* => срав, *sup* => прев), (*predic* => предик, *agent* => агент, *prepos* => предл и т.п.).

Следующий шаг усложнения проверяемых условий заключается в исследовании контекста переводимой единицы, т.е. свойств тех элементов структуры, которые лежат вне обрабатываемого правилом узла. Такая проверка может фиксировать сколь угодно сложный древесный контекст переводимой единицы и учитывать любые морфологические и словарные свойства входящих в этот контекст единиц. Примеры:

1. Слово PAY, тривиальным переводом которого является слово ПЛАТИТЬ, переводится словом ОБРАЩАТЬ, если его первым дополнением оказывается слово ATTENTION ‘внимание’.

2. Многие глаголы со значением покрытия поверхности предмета имеют по два значения, являющихся конверсивами друг друга, например: *spread butter on the bread* ‘намазывать масло на хлеб’ – *spread the bread with butter* ‘намазывать хлеб маслом’. Если соответствующий русский глагол не обладает тем же свойством, то выбор перевода такого английского глагола зависит от того, каковы его первое и второе дополнения: *to sprinkle the floor with sand* ‘посыпать пол песком’ – *to sprinkle sand on the floor* ‘рассыпать песок по полу’; *to sow wheat* ‘сеять пшеницу’ – *to sow a field with wheat* ‘засевать поле пшеницей’.

3. Прилагательное HARD1, имеющее тривиальный перевод ТВЕРДЫЙ, переводится словом ТВЕРДЫЙ в двух контекстах: в случае, если это слово подчиняет инфинитив в функции атрибута (*a promise hard to fulfil* => *обеща-*

ние, которое трудно выполнить), и в эксплетивном контексте (it is hard to fulfil this promise => это обещание трудно выполнить).

4. Сочетание high quality обычно переводится как высокое качество; если же его вершина (QUALITY) либо подчиняется другому существительному по композитному отношению, либо входит в сочинительную цепочку, вершина которой является композитным или определительным зависимым существительного, то это словосочетание переводится словом ВЫСОКОКАЧЕСТВЕННЫЙ; ср.: high quality translation => высококачественный перевод; cheap and high quality translation => дешевый и высококачественный перевод.

Если в тривиальном правиле осуществляется лишь замена одной английской лексемы одной русской, то нетривиальное правило перевода может отличаться от него по следующим признакам:

1. Морфологические характеристики русских слов не наследуются из английского языка, а вырабатываются самим данным правилом; ср.: SUPERIOUR => ПРЕВОСХОДИТЬ_{прич, непрош, несов} (т.е. *превосходящий*); CLAD => ПОКРЫВАТЬ_{прич, прош, сов, страд} (т.е. *покрытый*).

2. Переводу подвергается не одна английская лексема, а две или больше, ср.: dust exhausting fan => пылесос; heat resistance => жаростойкость.

3. Переводным эквивалентом служит не одна русская лексема, а больше, ср.: preheat => *предварительно нагревать*; earth fault => *замыкание на землю*; moat guarded Josephson device => *устройство с использованием эффекта Джозефсона*.

8.3. Обзор правил перевода

Перейдем к рассмотрению *нетривиальных* правил.

8.3.1. Перевод морфологических характеристик

Для перевода морфологических характеристик, помимо правил, о которых уже говорилось, используются еще два правила.

1. Перевод видо-временных характеристик у глаголов пассивного залога. Это правило дает шесть разных переводов, иллюстрируемых следующими примерами: а) is given => *дан*; б) is being given => *дается*; в) was being given => *давался*; г) was <has been, had been> given => *был дан*; д) will be given => *будет дан*; е) would be given => *был бы дан*.

2. Перевод активных причастий. Английские формы на -ING нормализуются двумя способами. В одних условиях они приобретают характеристику SØ и в дальнейшем превращаются в отглагольные существительные или их аналоги, а в других условиях приобретают характеристику раст (активное причастие), которая переводится, в зависимости от синтаксического контекста, формой причастия или деепричастия, ср.: the approaching train => *приближающийся поезд*; Having said so, he left the room => *Сказав это, он вышел из комнаты*.

8.3.2. Перевод синтаксических конструкций

Синтаксические конструкции, требующие для своего перевода нетривиальных правил, могут быть в большей или меньшей степени ограничены лексически. Перечислим основные правила перевода, начиная с тех, в которых лексическая связанность конструкции минимальна.

8.3.2.1. Правила перевода, не зависящие от лексического заполнения конструкции

1. Абсолютно-предикативная конструкция с глаголом в вершине.

Если такая конструкция содержит глагол в неперфектной форме и находится в постпозиции к главному предложению, то она переводится предложением, присоединяемым к главному союзом И: We went to the river, the dog running in front => *мы пошли к реке, и пес бежал впереди.*

Если же глагол в этой конструкции стоит в перфекте или вся конструкция предшествует главному предложению, в переводе появляется союз ПОСКОЛЬКУ: We decided to agree, John having spoken to the president => *мы решили согласиться, поскольку Джон поговорил с президентом.*

2. Композитная конструкция. Основная разновидность композитных конструкций – конструкция, образованная двумя существительными, – один из наиболее крепких орешков для перевода. Это обусловлено тем, что в таких конструкциях одновременно представлены чрезвычайно скудные формальные средства и весьма разнообразные семантические отношения между элементами конструкции. Следует считать не исключением, а правилом ситуацию, когда две композитные конструкции, использующие одни и те же формальные средства и отличающиеся только лексическим наполнением, переводятся совершенно по-разному, например: creep rate ‘удельная ползучесть’ – tall rate ‘тариф на пользование’. Поэтому мы пошли по такому пути. Прежде всего были выделены случаи тривиального перевода композитной конструкции – те случаи, когда для адекватного перевода достаточно перенести зависимый член в постпозицию к главному и приписать ему родительный падеж (сохраняя единственное число – типичное свойство композитной конструкции): field theory => *теория поля*. Все остальные варианты перевода композитной конструкции мы постарались ограничить теми или иными контекстными условиями.

8.3.2.2. Правила перевода, зависящие от лексического заполнения конструкции

Имеются в виду конструкции, для перевода которых необходимо фиксировать по крайней мере одну входящую в их состав лексему и которые поэтому обрабатываются не общими, а трафаретными правилами TRADUCT. Сгруппируем правила перевода этих конструкций по типам лексем, фиксируемым в правиле.

1. Правила с лексически ограниченной позицией для глагола.

Начнем с перевода вводной конструкции, образованной глаголами типа HOPE, BELIEVE, UNDERSTAND, SUSPECT в пассиве: He would come in time, it was hoped => *Он, как можно было надеяться, придет вовремя.*

Необходимость в таком правиле вызвана тем, что русские соответствия указанных английских глаголов либо не имеют пассивной формы (например, *надеяться, полагать*), либо, будучи в пассиве, не употребляются во вводной конструкции (например, *понимать*).

Второе правило обработки пассивных конструкций предназначено для перевода некоторых конструкций типа *nominativus cum infinitivo*, не имеющих прямого соответствия в русском языке и формируемых глаголами SAY 'говорить', STATE 'заявлять', REPORT 'сообщать', BELIEVE 'полагать', THINK 'думать', EXPECT 'ожидать' и т.п. Так же как и предыдущее правило, оно превращает пассивную конструкцию в неопределенно-личную, понижая статус подлежащего. Отличие состоит в том, как это происходит. Если в предыдущем правиле подлежащее страдательного залога становилось его же вторым дополнением, то здесь оно остается подлежащим, но спускается в придаточное предложение: *The delegation is reported to have left Moscow => сообщают, что делегация уехала из Москвы.*

К конструкции *nominativus cum infinitivo* примыкает конструкция *accusativus cum infinitivo*, образуемая глаголами волеизъявления и восприятия (WANT 'хотеть', WISH 'желать', EXPECT 'ожидать', SEE 'видеть', NOTICE 'замечать' и т.п.): *They wanted plastic to replace natural rubber => они хотели, чтобы пластик заменил натуральный каучук.*

Правило перевода этой конструкции превращает инфинитивную группу в придаточное предложение, подлежащим которого становится первое дополнение исходной конструкции.

Следующее правило перевода касается глаголов типа APPEAR 'оказываться', SEEM 'казаться', HAPPEN 'случаться', которые допускают инфинитивное дополнение, невозможное (для соответствующих глаголов) в русском языке. Задача правила – превратить инфинитивную конструкцию в сложно-подчиненное предложение с союзом ЧТО1: *John seems to be ill => кажется, что Джон болен.*

2. Правила с лексически ограниченной позицией для предлога.

Правила перевода призваны компенсировать следующие свойства английских предлогов:

а) многозначность, разрешаемую с помощью синтаксического или лексического контекста. Например, предлог BY1 имеет стандартный перевод 'посредством', реализующийся в контекстах типа *engines driven by electricity* 'двигатели, приводимые в движение посредством электричества'. Однако в кратной конструкции он переводится иначе: *four by four* => *четыре на четыре*. Предлоги OVER1 и ABOVE1 нормально переводятся как 'над' (*over <above> the table* => *над столом*), однако если они управляют существительным с синтаксическим признаком MESUR ("единица измерения"), то их значение меняется: *over <above> three hundred dollars* => *свыше трехсот долларов*;

б) способность формировать вместе с герундием предложно-герундиальные обороты, переводимые причастными оборотами (вообще без помощи предлога). Первый тип таких оборотов формируется предлогами IN1 и BY1 и имеет значение одновременности. Переводным эквивалентом служит форма деепричастия несовершенного вида, ср.: *By having the correct interpole voltage it is possible to obtain sparkless commutation* => *имея правильное напряжение на вспомогательных полюсах, можно обеспечить безыскровую коммутацию*;

в) способность управлять наречием. Поскольку русские предлоги в такой конструкции невозможны, приходится преобразовывать наречие в именную группу: *since then* => *с того времени*; *before now* => *до настоящего времени*; *until recently* => *до недавнего времени*;

г) способность присоединять препозитивную именную группу со значением временного интервала. Соответствующие русские предлоги, имеющие семантическую валентность на такой интервал, реализуют ее с помощью предлога, ср.: *three years after the reconstruction* => *через три года после перестройки*.

3. Правила с лексически ограниченной позицией для наречия.

Правило такого же типа, как для предлогов AFTER1 'после' и BEFORE1 'до', существует для наречий типа AWAY 'прочь', OFF 'прочь', MORE 'больше' и LESS 'меньше'. Эти наречия также способны присоединять препозитивную беспредложную именную группу с количественным значением, заполняющую их семантическую валентность, а их русские эквиваленты также присоединяют соответствующий актант с помощью предлога: *two miles away from the shore* => *в двух милях от берега*; *two miles off* => *в двух милях отсюда*; *five tons more <less>* => *на пять тонн больше <меньше>*.

4. Правила с лексически ограниченной позицией для прилагательного.

Прежде всего следует отметить прилагательные с синтаксическим признаком OBJINF, присоединяющие инфинитив с объектным значением: *This problem is easy to solve* => *эту проблему легко решить*.

Существуют и такие прилагательные, которые формируют конструкции с субъектным инфинитивом: в них определяемое существительное выполняет роль субъекта при инфинитиве. Таковы, например, прилагательные FAST 'быстрый', QUICK 'быстрый', SLOW 'медленный', CERTAIN 'опреде-

ленный' и др.: Stored programs are certain to simplify the calculations => *встроенные программы, конечно, упрощают вычисления*. Правило, обрабатывающее конструкции с такими прилагательными, выносит в вершину предложения глагол, имевший форму инфинитива, а прилагательное превращает в наречие и подчиняет этому глаголу.

Следующее правило, ориентированное на перевод конструкций с прилагательными, обрабатывает параметрические прилагательные типа LONG 'длинный', DEEP 'глубокий', HIGH 'высокий' и др.: Two inches long => *длинной в два дюйма*.

Особое правило необходимо для перевода композитных конструкций, формируемых прилагательными с синтаксическим признаком A-ED – прилагательными, образованными от существительных с помощью окончания -ed и способными выступать в композитной конструкции типа two-barreled 'с двумя стволами', blue-eyed 'с голубыми глазами'.

Наконец, последнее правило с лексически ограниченной позицией для прилагательного касается всего двух лексем – MANY и MUCH. С его помощью переводится несколько идиоматичных конструкций, содержащих эти лексемы, ср.: Ten years were needed to develop new instrumental methods and nearly **as much** – to make them commercially available => *десять лет потребовалось для разработки новых инструментальных методов и почти столько же – для того, чтобы сделать их коммерчески доступными*.

5. Правила с лексически ограниченной позицией для существительного. Эти правила предназначены для обработки двух конструкций. Первая из них формируется существительным, обозначающим единицу измерения (синтаксический признак MESUR) и подчиняющимся слову, стоящему в сравнительной позиции, ср.: ten tons more => *на десять тонн больше*; three years older => *на пять лет старше*.

Вторая конструкция формируется существительным – названием месяца, подчиняющимся порядковому прилагательному с предлогом ON1; ср.: on the tenth of May => *десятого мая*.

6. Правило с лексически ограниченной позицией для союза. Особое правило требуется для перевода абсолютно-союзных оборотов, вводимых союзами WHEN 'когда', WHENEVER 'когда бы ни' и некоторыми другими, имеющими синтаксический признак ABSOL: While examining the goods, the buyers found then a considerable part of them was faulty => *осматривая товары, покупатели обнаружили, что значительная часть их была с дефектами*.

7. Правила с лексически ограниченной позицией для местоимения.

Например, неопределенные местоимения ANY 'какой-нибудь', ANYTHING 'что-нибудь', ANYONE 'кто-нибудь', ANYBODY 'кто-нибудь', ANYWHERE 'где-нибудь' переводятся соответствующими отрицательными местоимениями (НИКАКОЙ, НИЧЕГО, НИКТО, НИГДЕ) в одном и том же контексте: при наличии отрицания NOT или отрицательного местоимения при глаголе в том предложении или обороте, в который входит данное местоимение. Если отрицание стоит при глаголе, то обсуждаемое правило пере-

вода заменит неопределенное местоимение отрицательным, ср.: He didn't say anything => *он ничего не сказал*.

Если же отрицание выражено в местоимении (NOBODY 'никто', NOTHING 'ничего', NOWHERE 'нигде' и т.п.), то оно переводится тривиальным образом, а недостающее приглагольное отрицание вырабатывается в ходе этапа развертывания русской СинтС: He said nothing => *он ничего не сказал*.

Естественно, что в одном предложении могут понадобиться оба правила: He never says anything => *он никогда ничего не говорит*.

Следующее правило обрабатывает несколько конструкций, содержащих субстантивные анафорические местоимения ONE2 и THAT3. Соответствующие этим местоимениям русские слова ставятся всегда в мужском роде, что, естественно, может приводить к ошибке, ср.: the one on the left => *тот, что слева*; that having eighty pages => *тот, что имеет восемьдесят страниц*.

Нелегки для перевода конструкции с местоимениями типа WHICHEVER 'кто бы ни/ какой бы ни', HOWEVER 'как бы ни/ какой бы ни'. Представляет интерес их перевод в контексте предикативного, абсолютно-предикативного и определительного СинтО.

Тот оборот, внутри которого слово типа WHOEVER выступает в качестве подлежащего, в свою очередь, может иметь разные синтаксические функции. Например, оборот со словом WHOEVER может выполнять роль подлежащего: Whoever did it assumes full responsibility => *кто бы это ни сделал, он несет полную ответственность*; роль дополнения: We sell the machines to whoever needs them => *мы продаем машины любому, кто нуждается в них*; роль обстоятельства: Whoever signed the agreement, the problem remains => *кто бы и подписал соглашение, проблема остается*.

8.3.3. Перевод лексики

8.3.3.1. Общее строение массива правил перевода лексики

С точки зрения синтаксического класса все выражения, нуждающиеся в нетривиальном переводе, можно разбить на четыре группы в соответствии с тем, что стоит в их вершине – существительное, глагол, прилагательное или предлог. Сочетания, вершины которых относятся к другим частям речи, не образуют достаточно широких классов.

Начнем с двусловных сочетаний. Подчеркнем, что, говоря о количестве слов в словосочетании, мы имеем в виду лишь однозначные слова, поскольку семантически незначимые слова (например, пустые предлоги) были опущены на этапе нормализации и в структуру, подлежащую переводу, не входят.

Потенциально идиоматичные *субстантивные сочетания* строятся с помощью следующих СинтО: а) композитного (knife heater 'нагреватель ножевого типа'); б) определительного (canned fish <meat, fruit> 'рыбные <мясные, фруктовые> консервы'); в) атрибутивного (line (of) force 'силовая ли-

ния'); г) первого комплетивного (law (of) magnetism 'закон взаимодействия одноименных и разноименных магнитных полюсов'); д) квазиагентивного (moment (of) momentum 'кинетический момент'; sharpness (of) resonance 'острота резонансной кривой') и е) посессивного (warper's bobbin 'сновальная катушка'). Идиоматичные сочетания, в вершине которых стоит существительное, в научно-технических текстах особенно многочисленны и разнообразны.

Глагольные сочетания представлены меньшим числом типов, но тоже составляют внушительную группу. Они образованы а) предикативным СинТО (The war broke out 'разразилась война'); б) первым комплетивным (lift the ban 'снять запрет'; look for (something) 'искать (что-либо)'); в) вторым комплетивным (take into consideration 'принимать во внимание'); г) коллигативным (the solution was pointed at 'на решение было указано'); д) фразово-юнктивным (factor out 'выносить за скобку'; hold up with somebody 'одобрять кого-либо'); е) композитным (ore dressing 'обогащение руды') и ж) обстоятельственным (highly trained 'хорошо обученный').

Для предлогов имеется только одно СинТО, дающее идиоматичные сочетания – *предложное* (in importance 'по важности'; under construction 'строющийся'), а идиоматичные сочетания, в вершине которых стояли бы *прилагательные*, вообще нетипичны.

Слова, входящие в состав двусловных сочетаний всех перечисленных типов, в свою очередь, могут подвергаться распространению с помощью тех же механизмов. В результате получаются сочетания из трех, четырех и более слов, например: insertion phase shift 'вносимый фазовый сдвиг'; preserved milk products 'молочные консервы'; law (of) induced current 'закон Ленца'; difference limen for frequency 'дифференцированный порог частоты'; multiple-beam-forming matrix 'матрица образования многолучевой диаграммы направленности антенны'.

Таким образом, появляется список типовых английских структур, охватывающих основную часть идиоматичных выражений. Из этого списка были отобраны выражения, состоящие не более чем из трех слов, поскольку более длинные выражения (так же как и выражения, имеющие нетипичную или чересчур сложную структуру) целесообразно переводить индивидуально с помощью словарных правил.

Аналогичным образом формируется список типовых русских структур.

8.3.3.2. Классификация правил

Правила перевода лексики можно классифицировать по нескольким независимым основаниям.

Первое деление правил связано с тем, переводится ли данное слово или словосочетание в *произвольном* или *фиксированном контексте*. К первой группе относятся, например, правила, переводящие такие выражения, как heat resistance 'жаростойкость', fast hardening 'быстротвердеющий', dust ex-

hausting fan ‘пылесос’, recorder ‘записывающее устройство’, electrolyze ‘подвергать электролизу’, route locking release ‘размыкатель маршрута’ и др.

Ко второй группе относятся, например, правила, осуществляющие следующие преобразования (слово, служащее контекстом для данного правила, но само переводящееся стандартным образом, заключено в скобки): (common) factor => (общий) множитель (в свободных сочетаниях FACTOR переводится как ‘коэффициент’ или ‘фактор’); irrotational (field) => безвихревое (поле) (в свободных сочетаниях IRROTATIONAL переводится как ‘неротационный’); pay (attention) => обращать (внимание) (в свободных сочетаниях PAY переводится как ‘платить’); in (importance) => по (важности) (в свободных сочетаниях IN переводится как ‘в’); communication (band) => (полоса) частот связи (в свободных сочетаниях COMMUNICATION переводится как ‘связь’).

Второй признак, по которому можно классифицировать правила, – произвольность/фиксированность семантически нагруженных морфологических характеристик исходного или результирующего выражения.

Большинство правил не накладывает ограничений на форму этих выражений. Например, сочетание look for переводится как ‘искать’ независимо от того, в какой форме стоит глагол. В тех же правилах, где морфологические ограничения есть, можно выделить два случая.

1. Правило фиксирует форму *вершинного элемента* английского или русского выражения. Здесь речь идет о выборе той или иной формы глагола. Например, глагол HACK в форме пассивного причастия, выполняющего функцию определения, переводится не так, как в других формах: hacked (tree) => (дерево) со сделанной зарубкой, но: hack (a tree) => делать зарубку (на дереве). Несколько правил переводят английские прилагательные или предложные группы русскими причастиями. В этом случае форма причастия должна быть фиксирована: SUPERIOR => ПРЕВОСХОДИТЬ_{прич, несов, непрош} (т.е. *превосходящий*); CLAD => ПОКРЫВАТЬ_{прич, прош, сов, страд} (т.е. *покрытый*); VARIABLE => ИЗМЕНЯТЬ_{прич, несов, непрош, страд} (т.е. *изменяемый*); under construction => СТРОИТЬ_{прич, непрош, несов, страд} (т.е. *строящийся*).

2. Правило фиксирует форму *невершинного элемента*. В этом случае речь обычно идет о закреплении форм причастия или существительного, выступающих в функции определения. Например, преобразования в следующих парах различаются только тем, какие характеристики приписываются причастию и существительному, ср.: raise face => *восстающий* (несов, непрош) *забой* и branching fault => *разветвленный* (сов, прош, страд) *сброс*; route locking release => *размыкатель маршрута* (ед) и trouble record log => *журнал повреждений* (мн).

Третий признак, классифицирующий правила перевода лексики, отражает количество слов, участвующих в переводе с той и другой стороны. С этой точки зрения все правила можно рассматривать как переводящие одно (или два, или три) английских слова одним (или двумя, или тремя) русскими словами. Тем самым возникает девять групп правил: 1=>1, 2=>1, 3=>1, ...,

3=>3. Внутри каждой из этих групп правила различаются типом структуры, стоящей на входе и выходе правила.

Проиллюстрируем каждую из девяти групп несколькими разноструктурными примерами. В скобки будем заключать либо ключевые слова, в контексте которых производится данное преобразование, либо элементы служебного характера (something, что-либо и т.п.).

1=>1: anterior => *предшествующий*, lift (the ban) => *снимать (запрет)*, week's (delay) => *недельная (задержка)*, (point) at (something) => *(указывать) на (что-либо)*;

2=>1: heat resistance => *жаростойкость*, give (something) up => *отказываться (от чего-либо)*, look for (something) => *искать (что-либо)*, low duty (fan) => *низконапорный (вентилятор)*, force-distance (diagram) => *индикаторная (диаграмма)*;

3=>1: hold up with (something) => *одобрять (что-либо)*, bidirectional triode thyristor => *диак*, four-terminal network => *четырёхполюсник*, reactive near-field (region) => *ближняя (зона)*;

1=>2: electrolyze => *подвергать электролизу*, preheat => *предварительно нагревать*, infinitesimal => *бесконечно малый*, frictionless => *без трения*, dial => *круговая шкала*, communication (band) => *(полоса) частот связи*, photoelectric (fatigue) => *(усталость) фоточувствительной поверхности*, clock (diagram) => *круговая векторная (диаграмма)*, bearing (friction) => *(трение) в подшипнике*, phrase (programmable) => *(программируемый) по фразам*;

2=>2: hold potential => *иметь перспективу*, take for (something) => *принимать за (что-либо)*, drainage factor => *водопронусная способность*, heavy chemicals (plant) => *(завод) основной химии*, controller construction (diagram) => *конструктивная (диаграмма) контроллера*, (capacitance) to case => *(емкость) на землю*, take into (account) => *принимать в (расчет)*;

3=>2: hold back on (something) => *воздерживаться от (чего-либо)*, wet suction fan => *мокрый эксгаустер*, trouble record log => *журнал повреждений*, linear time-base (oscillator) => *(генератор) линейной развертки*, solid-state image (panel) => *твердотельная индикаторная (панель)*;

1=>3: fare => *плата за проезд*, feberite => *вольфрамвокислая соль железа*, ferry => *перевозить на пароме*, fireproof => *с огнестойкой изоляцией*, degeneration (factor) => *(фактор) отрицательной обратной связи*, primary (current) => *(ток) в первичной обмотке*;

2=>3: factor (something) out => *выносить (что-либо) за скобку*, face (of) drum => *лобовая поверхность барабана*, cutting face => *лезвие коронки бура*, earth fault => *замыкание на землю*, range delay (dial) => *(круговая шкала) задержки по дальности*, dielectric power (coefficient) => *(коэффициент) потерь в диэлектрике*, motor-drive (interrupter) => *(прерыватель) с моторным приводом*;

3=>3: compressed-air feed => *подача сжатого воздуха*, negative voltage feedback => *противосвязь по напряжению*, take (something) into consideration => *принимать (что-либо) во внимание*, exchange-coupled ion (pair) => *обмен-*

но-связанная ионная (пара), constant failure rate (period) => (период) постоянной интенсивности отказов.

Саратовский государственный университет имени Н. Г. Чернышевского

8.3.3.3. Сопутствующие синтаксические преобразования

Главная задача правил перевода лексики заключается, естественно, в замене английских слов и словосочетаний их русскими соответствиями. Однако во многих случаях этим дело не ограничивается. Лексическая замена часто требует тех или иных сопутствующих синтаксических преобразований, затрагивающих окружающий контекст. Основные типы таких преобразований следующие.

Первый тип связан с ситуацией, когда при переводе словосочетания какое-то из входящих в него слов, имеющее собственные зависимые, не получает прямого русского «наследника» со сходной синтаксической функцией, который бы автоматически перенял эти синтаксические связи. В этом случае следует специально позаботиться о том, чтобы передать эти связи другими словами. Например, сочетание *high quality* ‘высокое качество’, оказавшись в позиции композитного зависимого существительного, переводится одним словом – прилагательным **ВЫСОКОКАЧЕСТВЕННЫЙ**: *high quality translation* => *высококачественный перевод*. У прилагательного **HIGH** ‘высокий’, входящего в состав сочетания *high quality*, могут быть собственные атрибутивные зависимые, которые «по наследству» должны перейти к русскому слову, переводящему все сочетание: *very high quality translation* => *очень высококачественный перевод*. Существенно, что так можно поступить не со всеми зависимыми слова **HIGH**. Сочинительные зависимые, например, не допускают такого перевешивания, и поэтому наличие таких зависимых вообще блокирует перевод *high quality* => *высококачественный*: *high or low quality translation* => *перевод высокого или низкого качества* (а не: **высококачественный или низкий перевод*).

Другой пример: сочетание *get hold (of something)* переводится как *браться (за что-либо)*. В составе этого сочетания существительное **HOLD** ‘захват’ может определяться некоторыми прилагательными. Такие прилагательные должны, превратившись в наречие, переподчиниться глаголу: *get firm hold* => *крепко братья*.

Второй тип синтаксических преобразований вступает в действие тогда, когда при лексической замене появляется русское слово, не имеющее прямого английского прототипа. В этом случае может возникнуть необходимость передать ему некоторые из имеющихся синтаксических связей. Например, глагол **RAIN** ‘дождить’ при переводе расщепляется на полувспомогательный глагол **ИДТИ** и полнозначное существительное **ДОЖДЬ**. Параллельно этому расщеплению разные типы приглагольных обстоятельств ждет разная участь. Обстоятельства, характеризующие действие с внешней стороны (т.е. указывающие время, место, продолжительность, причину и т.п.), обычно остаются при глаголе, а обстоятельства, описывающие способ протекания действия, обычно переходят к полнозначному существительному: *Yesterday it rained very hard for three hours* => *Вчера шел очень сильный дождь в течение трех часов*.

Третий тип синтаксических преобразований, сопутствующих лексическим заменам, не связан с появлением или исчезновением синтаксически самостоятельных слов. Отметим два важнейших подкласса таких преобразований.

Первый подкласс образуют конверсные преобразования, меняющие местами некоторые актанты предикатной лексемы. Чаще всего возникает необходимость менять местами первый и второй актанты.

Грамматикализованный способ такой замены – это пассивизация. Поэтому важнейшей разновидностью конверсных правил являются такие, которые переводят пассивную форму глагола в активную. Актуальность этих правил обусловлена тем, что в английском языке страдательный залог распространён гораздо шире, чем в русском. Так, русские эквиваленты глаголов FOLLOW ‘следовать (за)’, AFFECT ‘влиять’, ACT (UPON) ‘воздействовать’, INSIST ‘настаивать’ и др. не имеют страдательного залога. Поэтому, если английский глагол стоит в пассиве, то при переходе к его русскому эквиваленту происходит замена пассива на актив с необходимой меной актантов: *Seven fat fleshed kine were followed by seven lean fleshed ones => За семью тучными коровами следовали семь тощих коров.*

Конверсное преобразование, меняющее первый и второй актанты, не ограничивается депассивизацией. Оно может сопровождать и обычную замену лексики в активном залоге.

Например, глагол HAVE ‘иметь’ переводится как БЫТЬ (У), если в позиции дополнения стоит название предмета: *We had spare parts for motors in large quantities => У нас были запасные части для моторов в большом количестве* (а не: *??Мы имели запасные части...*)

При переводе может возникнуть необходимость и в других типах конверсных преобразований. Ярким примером глагола, требующего мены второго и третьего актанта, служит глагол SUBSTITUTE ‘заменять’: *She often substitutes margarine for butter => Она часто заменяет сливочное масло маргарином.*

Второй важнейший подкласс синтаксических преобразований, которые, подобно конверсным преобразованиям, сопровождают лексические замены, состоит в изменении направления синтаксической связи. Такого преобразования, например, систематически требуют английские количественные слова MANY ‘много’, MUCH ‘много’, A FEW ‘мало’, SEVERAL ‘несколько’. Будучи прилагательными, они подчиняются существительному, которое они квантифицируют, по определительному СинтО, в то время как их русские соответствия являются наречиями и подчиняют существительное по (первому) комплетивному СинтО, навязывая ему родительный падеж: *many nuts => много гаек, several properties => несколько свойств.*

Вопросы и задания к Теме 8

1. Назовите задачи этапа перевода.
2. В чем состоят особенности перевода лексики, перевода морфологических характеристик и перевода синтаксических отношений?
3. Приведите примеры тривиального и нетривиального перевода лексики.
4. Какими параметрами определяется мера нетривиальности правила?
5. Приведите пример предельного случая тривиальности.
6. Каким образом исследуется контекст переводимой единицы?
7. Какие правила добавляются для перевода морфологических характеристик?
8. Приведите примеры перевода видо-временных характеристик у глаголов пассивного залога и перевода активных причастий.
9. Назовите правила перевода, не зависящие от лексического заполнения конструкции.
10. Приведите примеры абсолютно-предикативной конструкции с глаголом в вершине и композитной конструкции.
11. Перечислите правила перевода, зависящие от лексического заполнения конструкции.
12. Приведите примеры правил с лексически ограниченной позицией для глагола, правил с лексически ограниченной позицией для предлога, для наречия, для прилагательного, для существительного, для союза, для местоимения.
13. С помощью каких синтаксических отношений строятся потенциально идиоматичные субстантивные, глагольные и предложные сочетания?
14. Каким образом формируется список типовых английских и русских структур?
15. По каким принципам возможна классификация правил перевода лексики?
16. Приведите примеры перевода слов в произвольном и фиксированном контекстах.
17. Приведите примеры произвольности и фиксированности семантически нагруженных морфологических характеристик исходного и результирующего выражения.
18. Сформулируйте правило, фиксирующее форму вершинного элемента английского или русского выражения.
19. В чём состоит правило, фиксирующее форму невершинного элемента.
20. Приведите примеры сопутствующих синтаксических преобразований, затрагивающих окружающий контекст.
21. Перечислите конверсные преобразования, меняющие местами актанты предикатной системы.
22. Охарактеризуйте преобразования, изменяющие направления синтаксической связи.

СОДЕРЖАНИЕ

| | |
|---------------------------------------------------------------------|----|
| Тема 5. Компьютерное моделирование пословицы..... | 3 |
| Вопросы и задания к Теме 5..... | 8 |
| Тема 6. Общие сведения о системе ЭТАП-2..... | 9 |
| Вопросы и задания к Теме 6..... | 16 |
| Тема 7. Лингвистическая модель анализа текста..... | 17 |
| Вопросы и задания к Теме 7..... | 34 |
| Тема 8. Лингвистическое обеспечение собственно перевода..... | 36 |
| Вопросы и задания к Теме 8..... | 49 |