

ГРАММАТИЧЕСКИЕ ТРАНСФОРМАЦИИ ПРИ ЛАТЫШСКО-РУССКОМ МАШИННОМ ПЕРЕВОДЕ

*Работа представлена кафедрой прикладной лингвистики РГПУ им. А. И. Герцена.
Научный руководитель – доктор филологических наук, профессор Л. Н. Беляева*

В статье рассматриваются грамматические трансформации при латышско-русском машинном переводе на примере морфологических и синтаксических преобразований при переводе имени существительного.

***Ключевые слова:** машинный перевод, трансфер, грамматические трансформации.*

T. Gornostay

GRAMMATICAL TRANSFORMATIONS IN LATVIAN-RUSSIAN MACHINE TRANSLATION

The article studies grammatical transformations in Latvian-Russian machine translation by the example of morphological and syntactical transformations in noun translation.

***Key words:** machine translation, transfer, grammatical transformations.*

Перевод текста с одного естественного языка на другой – сложный процесс. Перевод с латышского языка на русский язык не явля-

ется исключением и предполагает необходимость постановки и решения разнообразных задач. Эти задачи можно разделить на экст-

ралингвистические, связанные с социальными, культурными, этническими, историческими, географическими аспектами языка, и собственно лингвистические – лексические и грамматические особенности сопоставляемых языков, которые приводят к необходимости преобразований в процессе перевода [1; 2; 3].

В статье рассматриваются грамматические преобразования (трансформации), необходимые при машинном переводе с латышского на русский язык. Работа конкретной системы машинного перевода и то, как в ней решаются конкретные проблемы, определяются принятой в системе стратегией перевода. В разработанной системе латышско-русского машинного перевода используется стратегия трансфера. В самом общем виде, алгоритм перевода в системе машинного перевода с трансфером представляет собой трехуровневый процесс: анализ входной структуры исходного языка, преобразование этой структуры в эквивалентную структуру переводного языка и синтез выходной структуры переводного языка.

При более детальном рассмотрении перевод предстает как многоуровневый процесс, при котором происходит последовательное преобразование текста. Так, в системах машинного перевода с латышским языком [5; 7] на уровне трансфера можно выделить три этапа:

1) грамматическое преобразование входной структуры латышского языка в выходную структуру русского языка (морфологические и синтаксические трансформации);

2) лексическое преобразование входной структуры (собственно перевод с извлечением переводных эквивалентов из двуязычного словаря);

3) морфологическое оформление выходной структуры русского языка (морфологические трансформации).

Грамматические (морфологические и синтаксические) трансформации проводятся с использованием правил, которые применяются на первом и третьем этапах трансфера.

Для разработки подобных правил необходим сопоставительный анализ грамматических особенностей латышского языка и их перевода на русский язык. Ввиду отсутствия представительного латышско-русского параллельного корпуса, сопоставительный анализ грамматических свойств латышского и русского языков проводился на материале практических грамматик обоих языков [4; 6]. В ходе такого анализа были составлены сопоставительные таблицы грамматических категорий для каждой части речи латышского и русского языков и основных структур, требующих грамматических трансформаций. На основе данных таблиц были разработаны правила для следующих грамматических трансформаций: замена, добавление или удаление морфологических характеристик; изменение порядка слов; скрывание вершины дерева; добавление новой вершины дерева; перенос грамматической информации с одной вершины дерева на другую. Рассмотрим примеры грамматических трансформаций при машинном переводе на материале сопоставительного анализа систем латышского и русского существительных.

В обоих языках совпадает набор грамматических категорий существительного, включающий категории рода, числа и падежа. Наибольшее количество различий в системах латышского и русского существительных наблюдается в наборе падежей каждого из языков, комплексе грамматических значений конкретного падежа и способе их выражения. Так, беспредложная форма локатива в пространственном значении в латышском языке не соответствует форме предложного падежа в русском языке, которая всегда употребляется с предлогом, например, *skolā – в школе*. В данном случае необходима синтаксическая трансформация добавления новой вершины дерева – предлога к русскому существительному в предложном падеже. На рис. 1 приведен пример такого правила, в котором функция *insert_leftmost* добавляет новую вершину *NewNode* (предлог *в*) и назначает ей характеристику части речи *PREP* (предлог).

```

Rule(N) // locative_2
{
    Target.Case == locative;
    Target.KindOf == "locative_2";
    insert_leftmost(NewNode, Target, pcomp);
    NewNode.TargetSpelling = "в";
    NewNode.POS = PREP;
}

```

Рис. 1. Пример правила синтаксической трансформации добавления предлога к русскому существительному в форме предложного падежа

Более того, в русском языке ограниченная группа слов мужского рода I склонения образует вторую форму предложного падежа в единственном числе в пространственном значении в сочетании с предлогами *в* и *на* с флексией *-|y|*, например, *в лесу*. Соответственно для данной группы слов необходимо синтезировать именно вторую форму предложного падежа русского переводного эквивалента, которая будет соответствовать локативу в латышском языке. Для этого необходимо передать системе машинного перевода знания о том, какие существительные в русском языке могут иметь вторую форму предложного падежа. Список таких существительных русского языка с дополнительной характеристикой *locative_2* хранится в системе латышско-русского машинного перевода в отдельном словарном модуле. При применении правила синтаксической трансформации добавления предлога к конкретному русскому существительному в предложном падеже, функция *KindOf* проверяет принадлежность данного существительного к этому списку. В случае совпадения будет синтезирована вторая форма предложного падежа для русского переводного эквивалента.

В латышском языке также есть отдельная падежная форма – вокатив (звательный падеж), которая соответствует обращению в русском языке, не имеющему морфологиче-

ского выражения, и переводится на русский язык формой именительного падежа, например, *Skolotāj!* – *Учитель!* Для корректного синтеза русского существительного необходима трансформация замены морфологической характеристики звательного падежа латышского существительного на значение именительного падежа русского переводного эквивалента.

При латышско-русском машинном переводе именных групп нередко приходится прибегать к изменению порядка слов в выходной структуре русской именной группы. Такая синтаксическая трансформация необходима, например, при переводе латышских именных генитивных конструкций (*adnomiņālais ģenitīvs*), которые представляют особую сложность для машинного перевода на русский язык. Дело в том, что в латышском языке в группе существительного все падежные формы следуют за существительным в генитиве, т. е. в латышской двухкомпонентной именной генитивной конструкции определяемое существительное в родительном падеже находится в препозиции к определяющему существительному в именительном падеже. Рассмотрим подобную трансформацию на примере латышской двухкомпонентной генитивной конструкции, состоящей из двух существительных, одно из которых употреблено в форме генитива (родительного падежа) *lasītāju skaitis* ($N^1_{gen} N^2$). Для успешного синтеза русской именной группы необходима синтаксическая трансформация изменения порядка слов в исходной структуре латышского языка $N^1_{gen} N^2 \rightarrow N^2 N^1_{gen}$. Для этого применяется правило, пример которого приведен на рис. 2. При этом с помощью функции *move_to_right* латышское существительное в родительном падеже *lasītāju* будет перемещено в положение постпозиции по отношению к ядру латышской именной генитивной конструкции *skaitis*, соответственно в момент формирования русского перевода порядок слов будет правильным.

```
TransferRule(N<-mod-N) //NgenN
{
    move_to_right(Child,Parent);
}
```

Рис. 2. Пример правила синтаксической трансформации изменения порядка слов в латышской именной генитивной конструкции

В результате сопоставительного анализа грамматик латышского и русского языков, а также на основе выделения структур, требующих преобразований, разработаны правила грамматических трансформаций. Эти правила, применяемые при латышско-русском

машинном переводе с трансфером, примеры которых рассмотрены в статье на материале существительного, позволяют внести изменения в грамматическую информацию отдельной лексической единицы или в структуру конкретной синтаксической конструкции для успешного их синтеза на последнем этапе машинного перевода. Таким образом, достигается основная цель работы системы машинного перевода и обеспечивается ее эффективность: система позволяет получить перевод текста, сохраняя в процессе преобразования его общее содержание и передавая его на язык перевода так, что его может воспринять пользователь, обладающий знаниями предметной области или темы сообщения.

СПИСОК ЛИТЕРАТУРЫ

1. *Беляева Л. Н.* Теория и практика перевода: учеб. пособие. СПб.: ООО «Книжный Дом», 2007. 212 с.
2. *Беляева Л. Н.* Лингвистические автоматы в современных гуманитарных технологиях: учеб. пособие. СПб.: ООО «Книжный Дом», 2007. 192 с.
3. *Пиотровский П. Г.* Новые горизонты машинного перевода // Научно-техническая информация. Сер. 2. Информационные процессы и системы. М., 2002. № 1. С. 17–29.
4. Русская грамматика: в 2 т. 1980. Т. 1: Фонетика. Фонология. Ударение. Интонация. Словообразование. Морфология. М.: Наука. 783 с.
5. *Deksne D., Skadiņa I., Skadiņš R., Vasiljevs A.* Foreign language reading tool – first step towards English-Latvian commercial machine translation system // Proceeding of the Second Baltic Conference on Human Language Technologies. Tallinn: Tallinna Raamatutrukikoda, 2005. P. 113–118.
6. *Paegle D.* Latviešu literārās valodas morfoloģija: 1. d. Rīga: Zinātne, 2003.
7. *Skadiņš R., Skadiņa I., Deksne D., Gornostay T.* English/Russian-Latvian Machine Translation System // Proceeding of the Third Baltic Conference on Human Language Technologies. Vilnius: Institute of the Lithuanian Language. 2003. P. 287–295.

REFERENCES

1. *Belyayeva L. N.* Teoriya i praktika perevoda: ucheb. posobiye. SPb.: ООО «Knizhny Dom», 2007. 212 s.
2. *Belyayeva L. N.* Lingvisticheskiye avtomaty v sovremennykh gumanitarnykh tekhnologiyakh: ucheb. posobiye. SPb.: ООО «Knizhny Dom», 2007. 192 s.
3. *Piotrovsky R. G.* Novye gorizonty mashinnogo perevoda // Nauchno-tekhnicheskaya informatiya. Ser. 2. Informatsionnye protsessy i sistemy. M., 2002. N 1. S. 17–29.
4. Russkaya grammatika: v 2 t. 1980. T. 1: Fonetika. Fonologiya. Udarenie. Intonatsiya. Slovoobrazovaniye. Morfoloģiya. M.: Nauka. 783 s.
5. *Deksne D., Skadiņa I., Skadiņš R., Vasiljevs A.* Foreign language reading tool – first step towards English-Latvian commercial machine translation system // Proceeding of the Second Baltic Conference on Human Language Technologies. Tallinn: Tallinna Raamatutrukikoda, 2005. P. 113–118.
6. *Paegle D.* Latviešu literārās valodas morfoloģija: 1. d. Rīga: Zinātne, 2003.

7. *Skadiņš R., Skadiņa I., Deksnis D., Gornostay T.* English/Russian-Latvian Machine Translation System // *Proceeding of the Third Baltic Conference on Human Language Technologies*. Vilnius: Institute of the Lithuanian Language. 2003. P. 287–295.