

ПОДХОДЫ К ФОРМАЛЬНОМУ ВЫДЕЛЕНИЮ ИДИОМАТИЧЕСКИХ ВЫРАЖЕНИЙ

Идиома – это словосочетание, значение которого не следует из значения слов, входящих в его состав, а также не имеет прямого значения их сочетания. Идиома представляет собой цельное словосочетание, смысл которого не очевиден для не знающего ее человека и характеризуется почти полной или полной неизменяемостью. Основные трудности при переводе идиом вызывает совокупность лингвистического и культурного аспектов.

При переводе с иностранного языка существует риск буквального перевода идиом, что может привести к искажению смысла или полному его отсутствию в переводе. Поэтому очень важно уметь распознать идиому в тексте оригинала. Одной из проблем перевода идиом является то, что многие идиомы плохо узнаваемы, в отличие от широко известных и распространенных идиом. Наличие в тексте непонятого по значению словосочетания, которое сложно вписать в смысловую картину всего текста, может быть признаком того, что это идиома. Так же слова и словосочетания, окружающие идиому, могут подсказать ее присутствие.

При машинном переводе идиом не всегда удастся дать точный перевод, руководствуясь обычными правилами. Проблема в том, что идиомы могут быть записаны в нескольких формах и эти формы могут не всегда образовываться по обычным грамматическим правилам. Идиомы должны выявляться в тексте на начальном этапе перевода, чтобы они не были потеряны, и обрабатываться системой как одно слово. Еще одной проблемой является возникновение трудности привязки идиомы к слову, по которому будет производиться ее поиск.

Разрешение идиоматичности является одной из формальных операций в системах машинного перевода. Такая операция производится либо с помощью стандартных грамматических и лексических программ анализа текста, работающих с автоматическим словарем, либо выполняется прямое соотнесение входного и выходного сегментов текста. В случае прямого соотнесения и входной, и выходной сегменты рассматриваются как неделимые обороты. Один или несколько выходных оборотов или словоформ, поставленных в соответствие каждому входному обороту, составляют автоматический словарь оборотов. В результате анализа идиоматическим выражениям приписывается определенный цифровой эквивалент, и они исключаются из дальнейшего грамматического анализа. Перед составлением алгоритма поиска и перевода оборотов в тексте выполняется лингвистическое исследование их дистрибуции. Необходимо выяснить посредством синтаксического анализа, является ли оборот цельным и включает ли он в себя изменяемые формы. Оборот считается цельным, неизменным, если между его элементами нельзя вставить другие единицы. Если оборот не является цельным, следует учи-

тывать это при составлении алгоритма. Для обработки оборотов, разорванных другими членами предложения, необходимы данные синтаксического анализа.

При машинном переводе текстов, содержащих идиоматические выражения, фразеологические единицы включаются в машинный словарь, как основные единицы языка и речи, а перевод отдельных слов используется только, когда не удастся осуществить перевод, опираясь на фразеологические единицы. Так же учитывается, что тексты, имеющие большой объем, обычно бывают политематическими и отличаются друг от друга распределениями вероятностей появления в них различных слов из общенационального словарного фонда. Поэтому машинный словарь, предназначенный для перевода текстов из одной тематической области, как и из различных предметных областей, является политематическим. В системах фразеологического машинного перевода наряду с основным политематическим словарем большого объема используются наборы небольших по объему дополнительных тематических словарей. Такие словари содержат информацию, которая отсутствует в основном словаре. Например, информацию о приоритетных переводных эквивалентах словосочетаний и слов для различных предметных областей, когда они не совпадают с приоритетными переводными эквивалентами основного словаря. В системах фразеологического машинного перевода используется и интерактивный режим работы. В этом режиме пользователь имеет возможность вмешиваться в процесс перевода и настраивать дополнительные машинные словари на тематику переводимых текстов. Для получения качественного машинного перевода текстов, особенно политематических, необходимы машинные словари большого объема, которые могут создаваться на основе автоматизированной обработки двуязычных параллельных текстов и в процессе функционирования систем перевода.

А. В. Зубов

ФОРМАЛИЗАЦИЯ СОДЕРЖАНИЯ РУССКОГО ЛИРИЧЕСКОГО СТИХОТВОРЕНИЯ

Под стихотворением понимается разбитые на строфы последовательности семантически связанных предложений. Поль Валери заметил, что «стихотворение создается не из идей и не из чувств, оно создается из слов». Естественно, что без идеи и чувств создать стихотворение невозможно, но когда речь идет об анализе уже готового стихотворения главным становится слово.

Известно, что каждый текст выражает индивидуальные лингвистические и экстралингвистические знания автора. Лингвистические знания связаны со спецификой употребляемых автором слов, словосочетаний, типов предложений и их последовательностей (абзацев, строф). Экстралингвистические