

и словосочетаний, отражающая наиболее существенные составляющие описываемой в тексте ситуации и наиболее важные отношения между этими составляющими. Это как раз то инвариантное, что содержится в тексте и одинаково выделяется и автором текста, и воспринимающими текст. При выделении этого понятийного ядра текста, то есть темы текста, осуществляется так называемое свертывание текста.

Сегодня компьютер в состоянии определить основное содержание любого текста, хранящегося в его памяти. Исходя из этого, задача компьютерного анализа и определения основного содержания текста на естественном языке остается одной из ведущих тем многочисленных теоретических и практических исследований в области компьютерной лингвистики. В нашей работе мы используем статистический метод для определения основного содержания текста.

Всего для анализа мы отобрали 80 англоязычных новостных публицистических текстов по теме «Культура», выделив более узкие тематики в данной области. Источниками послужили англоязычные интернет-ресурсы, посвященные публицистическим текстам, такие как [www.bbc.com](http://www.bbc.com) и [www.theguardian.com](http://www.theguardian.com). Все текстовые единицы по данной тематике были разделены на десять групп на основании тематической принадлежности публицистических текстов. Более узкие тематики были выделены на основании тематической классификации текстов на специализированном интернет-портале [www.theguardian.com](http://www.theguardian.com).

В отобранном нами материале тексты представлены довольно однородно. Наиболее многочисленная группа состоит из монотематических текстов – «Film» (15 %), «Music» (13 %), «Theater» (13 %).

Политематические тексты составляют вторую по численности группу (15 %): «TV & Dance», «TV & Music», «Photography & Books», «Photography & Film», «Film & Fashion».

**Д. В. Степанова**

#### ПРЕИМУЩЕСТВА И ВОЗМОЖНЫЕ ЗАТРУДНЕНИЯ ПРИ ИСПОЛЬЗОВАНИИ ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ В ПЕРЕВОДЧЕСКОЙ ДЕЯТЕЛЬНОСТИ

Появление новых отраслей науки и знания, стремительный рост информационного обмена на иностранных языках, повышение требований к качеству перевода при сокращении сроков на его выполнение ставит переводчика в достаточно жесткие условия. В целях повышения качества перевода и решения отдельных переводческих задач в последнее время в мировой практике все большую популярность приобретает корпусный подход, который основывается на использовании электронных корпусов параллельных текстов. Возможность применения такого подхода к проведению лингвистических исследований в общем и к переводу в частности обусловлена тем фактом, что благодаря постоянному совершенствованию компьютерных технологий и обеспечению широкого доступа к персо-

нальным компьютерам и Интернету, специалисты имеют возможность оперативно выполнять различные операции с объемными электронными текстами.

Отличительной чертой корпусного подхода является возможность обращения исследователей к реальному употреблению языковых единиц. В последнее десятилетие автоматический перевод научно-технических текстов получил новый серьезный импульс благодаря разработке и внедрению новой технологии машинного перевода Translation Memory, или технологии переводческой памяти, которая позволяет накапливать и использовать ранее переведенные текстовые фрагменты. Данная технология опирается на алгоритм поиска словоформ, словосочетаний или целых фрагментов текстов исходного языка в базе переведенных ранее текстов, или в корпусе параллельных текстов, с целью нахождения переводных эквивалентов. Специфика рассматриваемого метода связана с тем, что источником переводных эквивалентов является не некоторая база заранее переведенных словосочетаний, а конкретный перевод некоторого исходного предложения при условии параллельной сегментации исходного текста и его перевода по предложениям. При этом специалисты основываются на положении, что для достижения качественного перевода при наименьших затратах времени невозможно обойтись без знаний и навыков переводчика-человека. В связи с этим для нахождения верных эквивалентов в языке перевода лексическим единицам в исходном языке необходимо обратиться к уже готовым переводам текстов различной тематики. При работе над переводом текста человек не только использует словарь, но и учитывает экстралингвистическую ситуацию, микро- и макроконтексты, устанавливает, в каких значениях употреблены многозначные лексические единицы.

Использование корпусного подхода при переводе имеет ряд преимуществ:

1) корпус параллельных текстов может служить репрезентативным источником переводных эквивалентов для новых терминов, которые еще не зафиксированы в бумажных словарях, но уже активно используются специалистами определенной области знания;

2) электронный корпус параллельных текстов является удобным средством по оперативному, объективному и точному выбору вариантов перевода, поскольку такие тексты могут постоянно пополняться и отражать последние тенденции в языке;

3) в состав корпуса параллельных текстов входят оригинальные, естественные тексты, которые порождаются для координации деятельности людей в той или иной референтной ситуации, что определяет такое важное достоинство такого корпуса, как реалистичность;

4) благодаря реалистичности корпусов параллельных текстов, которые содержат актуальную информацию о некоторой области знания, переводчик при обращении к корпусу может не только узнать значение слова и его сочетаемость с другими словами, но и получить информацию об употреблении этой лексической единицы в определенном контексте;

5) корпуса параллельных текстов помогают разрешить проблему многозначности терминов при переводе, а также позволяют при наличии синонимичных вариантов переводов найти наиболее распространенный среди специалистов термин;

6) технология переводческой памяти избавляет переводчика от необходимости повторно переводить идентичные фрагменты текста, входящие в разные документы;

7) системы с переводческой памятью дают возможность установления единства терминологии при работе нескольких переводчиков над текстами определенной тематики.

Несмотря на все положительные характеристики рассматриваемого подхода к автоматическому переводу, существует ряд затруднений при разработке и применении систем с переводческой памятью.

Одной из основных проблем таких систем машинного перевода является параллельная сегментация, или установление соответствий между текстом оригинала и его переводом или переводами на уровне слов, предложений, абзацев или целых текстов. Стыковка текстов может осуществляться вручную или автоматически. Специалисты, занимающиеся проблемами разработки программного обеспечения для решения этой задачи, пока еще делают первые шаги в этой области и опираются, в основном, на структурное сходство оригинального текста и его перевода. Специалисты выделяют три подхода к автоматическому или полуавтоматическому выравниванию параллельных текстов: лингвистический, статистико-вероятностный и параллельная сегментация исходного текста и текста перевода.

Среди прочих проблем использования технологии переводческой памяти можно выделить необходимость постоянного пополнения баз данных достаточно большими массивами параллельных текстов разной тематики, а также автоматическое распространение неточностей, при их наличии, на весь проект.

Все вышперечисленное дает нам право утверждать, что использование технологии переводческой памяти при автоматическом переводе научно-технических текстов способствует повышению качества машинного перевода, основанного на правилах, за счет, в первую очередь, точности перевода терминологической лексики, при общем сокращении времени, необходимого на выполнение перевода, и уменьшении объема работы переводчика.

**М. В. Чернышевич**

## АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ НЕДОСТОВЕРНЫХ МНЕНИЙ В СООБЩЕНИЯХ ПОЛЬЗОВАТЕЛЕЙ НА АНГЛИЙСКОМ ЯЗЫКЕ

В связи со стремительным ростом популярности социальных платформ для общения пользователей коммерческие компании и публичные деятели стали уделять большое внимание присутствию собственных названий и имен в социальных сетях. По данным компании Altimeter в 2016 г. 38 % коммерческих компаний инвестировали в рекламу в социальных сетях более